Ciências ULisboa

ECOLOGIA NUMÉRICA

Exame teórico-prático – 1ª época 2018/2019
22/01/2018 – 3º ano Biologia

Este é um exame realizado exclusivamente em R. Todo o código utilizado deverá estar visível no HTML resultante da compilação do .Rmd que é entregue. Todo o código utilizado deverá estar visível no HTML resultante da compilação do .RMD que é entregue.

1. (cotação 0.25) Crie uma pasta onde deverá colocar todos os ficheiros e objetos criados durante o exame, incluindo os ficheiros de dados que sejam utilizados no decorrer do mesmo. Crie um relatório dinâmico em RMarkdown, com um título apropriado, a identificação dos autores (nome e respetivo número), onde deverá apresentar todo o código necessário e output para a realização do exame.
O nome do ficheiro deverá ser EPENE1A1819A*****.Rmd. No caso de ser um par deverá ser EPENE1A1819A*****A*****.Rmd, onde ***** representa(m) o(s) número(s) de aluno correspondente(s). Este documento terá de ser compilável e será o documento que entregam via e-mail (tamarques@fc.ul.pt) e que será avaliado. Cada exercício/alínea deverá estar claramente identificada com
Exercício ?
Exercício ?.!

(onde ? vai de 1 a 8 e ! vai de 1 a 5, potencialmente!)

2. (cotação 0.25) Crie um conjunto de dados simulados de uma variável aleatória Y, os ys, Y representa um índice de complexidade de habitat que pode tomar qualquer valor real, e que é uma função potencial de 6 variáveis independentes (x1 a x6). Os dados foram recolhidos em dois tipos de habitats ("hab") e em 4 estações do ano ("est") utilizando o código abaixo (para evitar enganos, aconselho vivamente o copy-paste).

```
set.seed(****)
# **** dias do mês em que o aluno faz anos,
#e.g. Carlos e Maria com anos a 1 e 13 de Maio, 0113
b0=rnorm(1,0,0.5)
b1=rnorm(1,0,1)
b2=rnorm(1,0,0.5)
b3=rnorm(1,0,1)
b4=rnorm(1,0,0.5)
b5=rnorm(1,0,1)
b6=rnorm(1,0,2)
n1=rpois(1,30)
n2=2*n1
n=4*n1
hab= sample(x=c("H1", "H2"),size=n,prob=c(0.5,0.5),replace=T)
est= sample(x=c("P", "V", "O", "I"),size=n,prob=rep(0.25,4),replace=T)
x1=c(runif(n2,0,10), runif(n2,10,20))
x2=c(runif(n2,0,10), runif(n2,5,15))
x3= c(rnorm(n1,0,1), rnorm(n1,1,1),rnorm(n1,2,1),rnorm(n1,3,1))
x4= c(rnorm(n1,0,1), rnorm(n1,0,1),rnorm(n1,0,1),rnorm(n1,1,1))
x5=rnorm(n,15,2)
x6=rnorm(n,0,5)
torf=sample(x=0:1,size=6,prob=c(0.2,0.8),replace=T)
ys=b0+b1*x1*torf[1]+b2*x2*torf[2]+b3*x3*torf[3]+b4*x4*torf[4]+b5*x5*torf[5]+b6*x6
*torf[6]+rnorm(n,0,5)
```

2.1 (cotação 0.25) Quantas observações de ys gerou?

2.2 (cotação 0.5) O tamanho da amostra vai ser igual para os seus colegas, ou não? Justifique

3. Com base nos dados gerados no exercício anterior**,** há interesse em saber se o habitat tem uma influência nos valores do índice de complexidade

3.1 (cotação 0.25) Quantas observações foram recolhidas em cada tipo de habitat "H1" e "H2", tal como definidos na variável hab

3.2 (cotação 1) Compare com um gráfico adequado para o efeito os valores do índice em cada habitat e interprete os resultados.

3.3 (cotação 2) Teste formalmente se há diferenças entre os valores dos índices nos 2 habitats.

4. Com base nos dados gerados no exercício **2,** há interesse em saber se a estação do ano tem uma influência nos valores do índice de complexidade

4.1 (cotação 1) Compare com um gráfico adequado para o efeito os valores do índice em cada estação e interprete os resultados.

4.2 (cotação 2) Teste formalmente se há diferenças entre os valores dos índices nas 4 estações, usando uma metodologia paramétrica.

4.3 (cotação 1) Se encontrou diferenças significativas realize o teste à posteriori correspondente e interprete os resultados. Se não encontrou diferenças significativas, refira os testes que poderia realizar para comparações à posteriori num contexto não paramétrico.

5. (cotação 0.25) Com base nos dados gerados no exercício **2,** crie uma `data.frame` adequada, a que chama `datays`, para realizar uma análise de regressão múltipla em que explica a variável dependente `ys` em função das variáveis independentes (apenas `x1` a `x6`, ignore a estação e o habitat).

5.1 (cotação 0.75) Implemente um modelo linear múltiplo para explicar a variável dependente em função das independentes.

5.2 (cotação 1) Com base nos valores simulados, quais das variáveis lhe parecem importantes para explicar os `ys`?

5.3 (cotação 1) Qual o valor esperado do índice para um local cujos valores observados de `x1` a `x6` sejam os valores médios observados na sua amostra?

5.4 (cotação 1) Qual o valor do R-quadrado e o que lhe permite concluir sobre a capacidade de prever o índice em função das variáveis disponíveis?

5.5 (cotação 0.75) No código acima, o que representava o objecto `torf`? Com base neste, quais as variáveis que de facto influenciavam o valor de `ys`? Cometeu algum erro de tipo I ou de tipo II com a sua análise? Justifique

6. (cotação 0.25) Execute o código

```
set.seed(****)
# **** dias do mês em que o aluno faz anos,
#e.g. Carlos e Maria com anos a 1 e 13 de Maio, 0113
file=ceiling(runif(1,0,100))
```

6.1 (cotação 0.25) Quais os valores que o objeto `file` pode tomar?

6.2 (cotação 0.25) Qual a família de distribuições da variável aleatória gerada?

6.3 (cotação 0.25) Qual a probabilidade de observar um valor menor ou igual a 50?

6.4 (cotação 0.25) Que número que se encontra dentro do objecto file?

7. (cotação 0.25) Leia o ficheiro "data4EPENg*.txt", onde substitui o * pelo número que tem no objecto `file` criado no Exercício 6. Neste conjunto de dados temos as abundâncias de 9 espécies de 3 géneros de aves detectadas em pontos de escuta de 10 minutos. A primeira coluna contém o tipo de habitat do local correspondente.

7.1 (cotação 0. 5) Quantos locais tem em cada habitat?

7.2 (cotação 0.5) Calcule a matriz de distâncias euclidianas entre os locais. Quais os dois locais com maior semelhança entre si e qual o valor dessa semelhança?

7.3 (cotação 1) Realize uma análise de agrupamento considerando os método de agrupamento "single" e "complete".

7.4 (cotação 1) Qual das análises lhe dá uma melhor separação dos locais por habitat?

8. (cotação 0.25) Leia o ficheiro "data4EPENdt*.txt" , onde substitui o * pelo número que tem no objeto `file` criado no Exercício 6. Neste conjunto de dados temos os valores de variáveis ambientais em pontos ao longo dum rio, em que os locais foram ordenados pela distância à foz ou à nascente. As variáveis recolhidas em cada local foram profundidade (prof), altitude (alt), oxigénio (O2), pH (pH), salinidade (sal), partículas em suspensão (sus), concentração de Mercúrio (Mg) e Chumbo (Pb).

8.1 (cotação 0.5) Realize uma análise em componentes principais adequada para descrever os locais em função das suas características.

8.2 (cotação 0.5) Qual a proporção de variação explicada pelos 2 primeiros eixos?

8.3 (cotação 0. 5) Quantos eixos recomendaria reter para interpretação?

8.4 (cotação 1) Interprete o biplot da PCA. Consegue identificar se os lugares com números baixos correspondem à foz ou à nascente do rio?

8.5 (cotação 0.5) Se tivesse de enviar a GNR para investigar a existência de uma fábrica poluidora com libertação de Mercúrio para o ambiente, em que locais começaria as suas investigações? Justifique

A minha nota vai ser:

Cotação total
0.25+
0.25+0.25+0.5+
0.25+1+2+
1+2+1+
0.5+0.5+1+1+1+0.75+
0.25+0.25+0.25+0.25+0.25+
0.25+0.5+0.5+1+1+
0.25+0.5+0.5+0.5+1+0.5
=21