



Modelação Ecológica

Componente Teórica - Prática

Ficha de trabalho

7b

1. Efectue uma análise de regressão (Função `lm`) sobre os dados *DataTP7densidade.csv*, relativos à densidade de uma espécie em função de variáveis ambientais, considerando a densidade como a variável dependente e a temperatura como independente. Faça uma representação gráfica adequada dos dados e do modelo. Explore os resultados e efectue uma avaliação do cumprimento dos pressupostos do modelo linear. Qual o valor da correlação e do coeficiente de determinação entre densidade e temperatura?
2. Com base nos dados da alínea anterior, efectue uma análise de regressão múltipla usando a função `lm`, sendo a variável dependente a densidade e as independentes todas as restantes. Comente os resultados (quais as variáveis importantes para explicar a densidade?) e verifique se os pressupostos são cumpridos. (Há funções muito práticas para extrair sub-componentes de modelos. Sobre o modelo que usar, experimente usar as funções `coef`, `fitted`, `predict.lm` e `AIC`). Com base no modelo encontrado, represente visualmente a relação entre a variável resposta e o(s) preditor(es) relevante(s). Represente também essa mesma relação caso estivesse interessado em prever a Densidade em função da Turbidez, mas sendo que a Salinidade era igual ao valor mínimo presente nos dados. Interprete os resultados obtidos.
3. Estimou a mortalidade de aves em alguns parques eólicos, em função de variáveis ambientais (*DataTP7parqueseolicos.csv*). Verifique se a variável "altitude" é um bom preditor da mortalidade, através de um modelo de regressão simples (Função `lm`). Caracterize a relação entre as variáveis. Elabore uma função que gere estimativas de mortalidade em função da altitude (Relembrar como criar funções, tutorial 1). Avalie o desempenho do modelo de regressão (Por exemplo, faça um `plot` do objecto de regressão criado e avalie os pressupostos do modelo de regressão).
4. Realize uma regressão múltipla aos dados *DataTP7parqueseolicos.csv* e interprete os resultados. Pretende obter um sub-modelo para utilizar noutras regiões com parques eólicos. Obtenha um sub-modelo para esse efeito. (Dica: começar por criar um modelo com todas as variáveis disponíveis e depois retirar as que não parecem importantes).
5. Aplique um GLM aos dados *DataTP7presenca*, os quais descrevem a abundância de uma espécie em função de algumas variáveis ambientais. Utilize a distribuição Gama como família de distribuição do erro (Gamma) e a função inversa como função de ligação (inverse, in a `glm` call, use argumente `family=Gamma(link="inverse")`). Compare os resultados com um GLM Poisson (in `glm` call, `family=poisson`). Compare as predições dos dois modelos num gráfico. Qual seria o valor previsto num caso em que estradas tomasse o valor 7. Qual dos dois modelos escolheria para modelar os dados (ver a função `AIC`)?
6. Aplique um GLM aos dados *DataTP7anchova.csv*, referentes à abundância de anchova (variável dependente, ANC) em função de várias variáveis ambientais (variáveis independentes). Explore os resultados.
7. Find a GLM that best fits the data "Owls.txt", where you are trying to explain the begging behavior of owls offspring when the parents are absent from the nest. The variable "SiblingNegotiation" represents the number of calls produced by the chicks in the nest during a 30 second period, while "BroodSize" represents the size of the brood. More details about this data can be found in Zuur et al. 2009.

- 8.** The data “DeerEcervi.txt” contains the incidence of *E. cervi* parasites in deer pellets, and we have also the corresponding sex, length and farm the deer were on. How many farms were available? Ignore them for now, and model the presence/absence of parasites in pellets as a function of deer characteristics. This is a dataset also used by Zuur et al. 2009.
- 9.** Explore os dados disponíveis em <https://data.mendeley.com/datasets/r3xpn3mccc/2>, usados num trabalho para prever o tamanho de grupos de baleias de bico em função da sua pegada acústica. Há uma boa descrição dos dados em “data4Mendeley.pdf”, mas por agora ignore as variáveis “glD”, “conf” e “cs0” como variáveis explicativas (na realidade, $cs0=cs-1$, o que pode ser interessante porque não há grupos de tamanho 0, mas uma Poisson assume que existem 0’s). Use o ficheiro “modeldata.txt” para tentar explicar o tamanho do grupo de animais (“cs”) em função das outras variáveis disponíveis.
- 10.** Usando os dados *data4GAM.txt*, ajuste um modelo linear, um GLM e um GAM e escolha o melhor modelo para representar a relação que explica “y” em função de “x”.