**Trabalho prático de Sistemas de Apoio à Decisão n.º 1**

**2019/2020**

Numa tentativa de se determinar a idade e o sexo das espécies pescadas em certas embarcações, a Marinha de um certo país expressou a necessidade de se começar por estabelecer quais os parâmetros que poderão identificar rapidamente essas duas características no abalone (Haliotis rubra).

A base de dados a que temos acesso são medidas retiradas de vários abalones estudados em laboratório. O ficheiro **abalone\_train\_[número atribuído ao grupo].xlsx** contém essa base de dados com o registo das seguintes características:

Comprimento – Em mm, representa a medida maior da concha

Diâmetro – Em mm, representa a medida perpendicular ao comprimento

Altura – Em mm, representa a altura da concha quando pousada

Peso Total – Em g.

Peso sem concha – Em g.

Peso das vísceras – Em g.

Peso da concha – Em g.

N.º de anéis – que pode ser usado para estimar a idade do abalone adicionando 1.5 ao número de anéis.

Sexo/Infantil – M (Macho), F (Feminino) e I (Infantil)

Para um técnico experiente o sexo e o número de anéis é ‘fácil de identificar’, no entanto, para um olhar destreinado isso já não é verdade.

O objetivo final do trabalho será o de definir que parâmetros poderão mais facilmente identificar os abalones infantis ou até mesmo as fêmeas, se possível, para controle da excessiva pesca de abalones, permitindo que sobrevivam infantis e fêmeas para manter a população.

Para tal será importante considerar as seguintes tarefas:

1. Criação de gráficos preliminares (globais ou por sexo) para poder descrever os dados e para tomar conhecimento da informação disponível. (histogramas, diagramas de dispersão, coordenadas paralelas, box-plots, etc) Não é pretendido que se mostre mil e um gráficos, mas sim que selecione os que visualizam melhor os dados.
2. Descrever quais as diferenças que observa para cada uma das classes do Sexo/Infantil, ou semelhanças.

**Questão: Pode-se dizer que alguma das variáveis são normais (para alguns testes de correlação linear existe a exigência de que a distribuição seja normal)?**

1. Existem valores estranhos em determinados dados? Quais e porquê? O que fazer quando um indivíduo tem valores estranhos?
2. Consegue-se observar ou aferir correlações lineares entre variáveis? Existem correlações que não as lineares?
3. Por vezes as variáveis não são normais mas poderão ser log-normais. Crie novas variáveis, que serão os logaritmos das variáveis originais, e estude a relação entre estas e também com as anteriores. Experimente criar outras variáveis!

**Questão: Pode-se criar uma variável para o volume? Se sim, faça-o.**

1. É possível efetuar projeções da base de dados para tentar dividir os dados. Como visualizar as novas variáveis e apurar se está bem dividido?
2. O ficheiro **abalone\_[número atribuído ao grupo].xlsx** contem alguns abalones diferentes onde há variáveis que supostamente um fiscal não saberia detetar nem ter tempo de avaliar (sexo, anéis, peso das vísceras, peso da concha, peso sem concha). Será possível identificar as variáveis objetivo (ou seja, sexo e idade) com apenas alguma das variáveis presente nestes ficheiros?
3. [Dificuldade acrescida]
   1. Utilize uma técnica de agrupamento (*clustering*) à sua escolha que permita dividir os dados por entre infantil e não infantil.

Ou

* 1. Crie um sistema de lógica difusa que poderia ser implementado para auxiliar um inspetor na catalogação da espécie aquando da inspeção.

Elabore um relatório do procedimento que adotou, das conclusões que retirou e defina medidas a tomar por um fiscal para detetar infantis e para detetar fêmeas. Será possível?

|  |  |  |  |
| --- | --- | --- | --- |
| Grupo | Cadetes | Ficheiro abalone\_train\_ | Ficheiro abalone |
| 1 | 401\_410\_428\_433 | 1 | 1 |
| 2 | 403\_416\_418\_426 | 2 | 2 |
| 3 | 404\_425\_429\_435\_440 | 3 | 3 |
| 4 | 405\_407\_417\_427 | 4 | 4 |
| 5 | 409\_421\_442\_445 | 5 | 5 |
| 6 | 411\_412\_424\_447 | 6 | 6 |
| 7 | 414\_438\_451 | 7 | 7 |
| 8 | 419\_443\_444\_454 | 8 | 8 |
| 9 | 430\_436\_480\_483 | 9 | 9 |
| 10 | 481\_482\_484\_486 | 9 | 1 |
| 11 | A preencher | 8 | 2 |
| 12 | A preencher | 7 | 3 |
| 13 | A preencher | 6 | 4 |
| 14 | A preencher | 5 | 5 |
| 15 | A preencher | 4 | 6 |
| 16 | A preencher | 3 | 7 |

PS: A atualizar com os grupos dos alunos n.ºs 402, 406, 408, 413, 415, 420, 422, 423, 431, 432, 434, 437, 439, 441, 446, 448, 449, 450, 452, 453, 485, se não houver equívoco.