

Exame de Bioinformática Prática

Prova “I”

26-01-2024

As respostas ao exame devem ser entregues em formato de texto simples (pode ser no próprio *script* de R utilizado para obter os resultados), “PDF” ou “Rmarkdown” (neste caso por favor apresente no final o ficheiro “.Rmd” e o respetivo “PDF”). Estes ficheiros podem ser produzidos em qualquer software que tenha disponível e considere adequado (mas recomenda-se o uso do *Rstudio*).

Apresente todo o código que escreveu num ficheiro denominado *ExameX.R* (ou *ExameX.Rmd* para o caso de “Rmarkdown”), onde **X** é a letra da prova, por exemplo, *ExameB.R*.

Inicie o código relativo a cada alínea do exercício com uma linha contendo “### Y”, onde **Y** é a alínea do exercício a resolver, por exemplo: “### b)” antes do código usado para resolver a alínea “b)”.

Note que código automatizado e/ou funcional (onde aplicável) será mais valorizado que uma solução repetitiva e/ou imperativa.

As figuras devem ser guardadas no formato *png* com as dimensões que achar adequadas.

Todos os ficheiros produzidos deverão ser colocados numa pasta denominada “**PrimeiroNome_UltimoNome_NumeroDeAluno**”. Ex.: “Francisco_PinaMartins_29609”. Esta pasta e o respetivo conteúdo será então entregue ao docente para avaliação. Certifique-se que esta contém todos os ficheiros que considere relevantes.

É altamente recomendado ler a prova toda antes de começar a responder.

Diversidade e morfologia de *Lacerta schreiberi*

Nas pitorescas e diversificadas encostas de montanhas e zonas húmidas da Península Iberica, uma equipa de biólogos iniciou um estudo sobre *Lacerta schreiberi*. Este réptil, endémico desta região, é aqui objeto de um estudo que incide sobre as respetivas características morfológicas. Os dados medidos estão presentes **neste ficheiro**.

Cada entrada neste conjunto de dados representa um indivíduo identificado por um número de ID exclusivo. As medições morfológicas estão detalhadas na Tabela 1.

Tabela 1: Significado dos nomes das colunas do ficheiro CSV.

Coluna CSV	Descrição
ID	Identificador do indivíduo
Population_number	Identificação numérica da região amostrada
Population_name	Identificação da região amostrada
Capture	Técnica de captura utilizada (1=Armadilha, 2=Laço)
Snout-Vent_length	Comprimento focinho-ventre (mm)
Head_width	Largura da cabeça (mm)
Head_length	Comprimento da cabeça (mm)
Melanization_index	Índice de melanização (UA)
Right flank	Número de "eyespots" no flanco direito
Left flank	Número de "eyespots" no flanco esquerdo
Eyespot_avg_length	Diâmetro médio dos "eyespots" (mm)
Dorsal_Brightness	"Brightness" do dorso (UA)
Dorsal_chroma	"Chroma" do dorso (UA)
Dorsal_Hue	"Hue" do dorso (UA)
Throat_Brightness	"Brightness" da garganta (UA)
Throat_chroma	"Chroma" da garganta (UA)
Throat_hue	"Hue" da garganta (UA)
Head_scales	Número de escamas na cabeças

Espera-se que o conjunto de dados possa fornecer *insights* valiosos sobre as populações de lagartos amostradas. Neste exame irá assumir o papel de consultor(a) de bioinformática e ajudar a equipa de investigação a resolver alguns dos conflitos através de técnicas de análise de dados. Deve justificar todas as repostas com base nos dados disponíveis. Sempre que recorra a testes de hipótese deve indicar todas as H_0 consideradas e deixar explícito se estas são ou não rejeitadas.

- a) Comece a analisar os dados usando uma abordagem exploratória. Com base em todas as características medidas, encontra indícios de que os indivíduos se agrupam por região?
- b) Com base em todas as características, é possível sugerir que as técnicas de “armadilha” e “laço” capturam indivíduos com características diferentes?
- c) Quais as duas características dos indivíduos que podem ser consideradas mais importantes na eventual diferenciação de populações? E qual a menos importante para esta distinção?
- d) Encontra diferenças significativas entre as populações das regiões “Sul” e “Serra da Estrela” para as três características identificadas em “c)? Caso não tenha chegado a nenhuma conclusão em “c)” considere as variáveis “Throat_hue”, “Eyespot_avg_length” e “Dorsal_Hue” para testar as diferenças entre populações.
- e) Obtenha um gráfico que lhe permita visualizar as diferenças expressas em “d)”.
- f) O número de “eyespots” do flanco esquerdo dos indivíduos é significativamente diferente do número de “eyespots” do flanco direito?
- g) Alguns membros da equipa sugerem que indivíduos oriundos de regiões diferentes são capturados de forma diferente pelas diferentes técnicas usadas. Encontra evidências nesse sentido?
- h) Encontra evidências de que as variáveis “Head_width” e “Head_length” estão relacionadas? Se sim, qual a força dessa relação? Grafique essa relação da forma mais completa que conseguir.
- i) E as variáveis “Dorsal_Brightness” e “Throat_Brightness”? Grafique também esta relação.