

Exame de Bioinformática Prática

Prova “H”

17-11-2023

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.

Biodiversidade e Dinâmica de *Habitat* em Regiões Ecológicas Diversas

Para compreender a intrincada interação entre fatores ambientais, biodiversidade e dinâmica de *habitats*, uma equipa de investigadores interdisciplinares embarcou num estudo que abrangeu diversas regiões ecológicas. O principal objetivo era desvendar as relações complexas entre vários parâmetros ambientais e a flora e fauna em *habitats* distintos. Os dados recolhidos estão disponíveis [neste ficheiro](#) ([link alternativa](#)). O significado de cada coluna do ficheiro pode ser consultado na Tabela 1.

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

Coluna CSV	Descrição
Region	Identificação da Localização geográfica
Land use type	Categorização com base no tipo de utilização humana
Habitat type	Categorização com base no <i>habitat</i> predominante
Solar radiation intensity	Intensidade da radiação solar
Plant species	Contagem e variedade de espécies
Animal species	Contagem e variedade de espécies
Temperature	Valores de temperatura média diária (°C)
Rainfall	Quantidade de precipitação (mm)
Soil pH	Valor médio de pH do solo
Elevation	Elevação média (m)
Tree coverage	Extensão da presença de árvores (Km²)
Predator species	Quantidade de espécies predadoras no local
Pollinator species	Quantidade de espécies polinizadores no local
Built road	Extensão de estrada construída (m)
Topographic diversity	Reflete a diversidade de elevação e topográfica
Agricultural intensity	Medida da influência da agricultura na paisagem

Espera-se que o conjunto de dados possa fornecer *insights* valiosos sobre a resiliência de ecossistemas diversos, o impacto de atividades humanas na biodiversidade e o papel de fatores ambientais na configuração de *habitats*. Infelizmente, o tamanho e *background* da equipa levou à discórdia de opiniões entre os membros da equipa. 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) Um dos membros da equipa afirma que da sua experiência o tipo de habitat depende da intensidade da radiação solar. Os dados suportam esta afirmação?
- b) Crie um gráfico que permita à equipa visualizar a "Solar radiation intensity" por cada "Habitat type" presente no *dataset*.
- c) Construa dois gráficos que permitam à equipa de investigação comparar os valores de, "Elevation" e "Agricultural intensity" entre tipos de diferentes de "Land use type".
- d) Toda a equipa considera que é importante saber se as regiões do tipo "Agricultural" e "Wild" se podem distinguir quanto a "Temperature", "Elevation", "Rainfall", "Agricultural intensity". Utilize testes de hipótese para inferir para quais destas variáveis há diferenças entre estes tipos de uso de terra.
- e) Alguns membros da equipa sugerem que a variável "Animal species" está relacionada com a variável "Built road". Mostre se essa relação existe e indique a respetiva força.
- f) Mostre graficamente a relação testada em e). Inclua uma *trendline*.
- g) Alguns dos investigadores afirmam que com base nas características quantitativas disponíveis, o "Land use type" será o melhor discriminante para agrupar as regiões. Um outro grupo sugere que o "Habitat type" será melhor para o efeito. Efetue uma abordagem que lhe permita explorar estas perspetivas através de gráficos.
- h) Qual é a percentagem de variação dos dados explicada por cada eixo?
- i) Indique aos investigadores (através de um gráfico) quais as duas variáveis mais importantes para distinguir o tipo de habitat "Mountainous" dos demais.