

Ecologia Numérica

Exame Teórico – 2018/2019 – 1ª Época - 22-01-2018

Proposta de Resolução

(Todas as respostas abaixo foram retiradas de exames reais entregues a 22-01-2018 – estão naturalmente anonimizadas por causa do RGPD; não seria necessariamente a minha escolha de palavras sempre, mas seriam respostas que recebiam a cotação total)

1.1

A amostra recolhida teria uma dimensão muito grande pois permite detetar diferenças mínimas em relação ao número de crias, sendo estas diferenças na ordem de 0.002. Ora tendo em conta que uma fêmea não pode ter 0.35 ou 0.36 filhos médios por ano, mas sim apenas valores inteiros desde 0, 1, etc, então a dimensão da amostra seria muito grande.

1.2

O P-value representa uma probabilidade, sob a veracidade de H_0 , de obter um valor da estatística de teste igual ou mais extremo do que o observado nos cálculos com base numa determinada amostra.

1.3

Existem vários valores usuais de significância - 0.1, 0.05, 0.01, e independentemente do nível que escolhêssemos rejeitaríamos a hipótese nula de que as duas espécies de cetáceos têm o mesmo número de crias. Ou seja, estes resultados tem uma grande significância estatística. Em relação à relevância biológica, tendo em conta que o valor médio entre as 2 populações difere em 0.002 e que é de 0.3/0.4 e que um indivíduo independentemente da espécie nunca poderá ter 0.3 crias por ano, mas sim ou zero ou 1 cria sendo sempre um número inteiro, a significância biológica pode não ser muito grande pois para voltar a diferenciar estas duas espécies quanto ao número médio de crias por ano era necessário ter dados relativamente a um número enorme de fêmeas em cada uma das espécies.

Significância Estatística – Imensa! Especialmente em casos de biologia, é improvável encontrar p-values assim tão baixos. Significância Biológica – pode não ser muita, depende de quanto realmente a diferença de 0.002 altera a biologia. Uma diferença que se tem a certeza que existe não equivale a uma diferença importante. No limite, com “n” altíssimo fica difícil encontrar diferenças não relevantes matematicamente, poderão é ter amplitudes práticas negligenciáveis.

2.1

Variáveis nominais são de valor discreto e medidas sem qualquer ordem aparente (por exemplo género, cor dos olhos). Variáveis ordinais são também discretas mas com ordem intrínseca (por exemplo, lugares no pódio).

2.3

A classe dos objectos que podemos usar para representar este tipo de variáveis é um factor

3.1

É uma variável que varia consoante eventos estocásticos e que apesar de sabermos que valores pode tomar apenas sabemos os valores realizados depois da observação

3.2

$f(x)$ é sempre maior ou igual a 0 e o integral da função densidade de probabilidade é 1.

3.3

Nº de ovos por ninho de uma espécie de aves,

Poisson

Comprimento de indivíduos adultos de uma espécie de golfinhos

Gaussiana

4.1

O primeiro pressuposto é o de que os valores são independentes. Outro pressuposto é o de que as amostras provem de populações normais. E por ultimo a homogeneidade das variâncias ou seja que as populações são homocedásticas.

4.2

Quando os pressupostos não são verificados devo analisar os dados de modo a verificar se uma transformação dos dados pode levar a que estes cumpram os pressupostos, se não resultar então devo usar o teste equivalente não-paramétrico.

4.3

Caso rejeitemos H_0 , deveríamos proceder de seguida a um teste à posteriori (ex: Tukey para ANOVA, Dunn para KW). Se não rejeitamos H_0 , concluímos que não há razões para acreditar que as médias sejam significativamente diferentes (sendo que nunca podemos "aceitar" H_0).

5.1

Escolheria um nível de significância de 0.05 uma vez que me permitiria ter um intervalo de confiança aceitável de 95% e uma vez que estava a tentar verificar a existência de uma relação e assim poderia posteriormente aprofundar o estudo. As que aparentam uma relação significativa é a profundidade, a largura e a temperatura.

5.2

(18 em 20)

$2.01+0.19*6-0.066*8+0.09*15.3-0.003998*75.3$

(20 em 20, ninguém conseguiu)

a função de ligação era o logaritmo, logo era preciso inverter com a exponencial

$\exp(2.01+0.19*6-0.066*8+0.09*15.3-0.003998*75.3)$

5.3

Como a largura apresenta um declive negativo, logo uma relação negativa, seria de esperar que com o aumento da largura a concentração da hormona diminua. Ou seja, é esperado que num local muito largo a concentração seja menor logo maior concentração da hormona num local pouco largo.

6

Poderia realizar um escalamento multidimensional para avaliar se os dados se agrupam de acordo com um dos gradientes em estudo. Para tal, após uma análise do gráfico iria ver se os agrupamentos formados estavam mais de acordo com a latitude ou com a distância das poças entre si. Se poças a distâncias semelhantes estivessem mais próximas no agrupamento, o gradiente seria o segundo, caso os agrupamentos fossem por locais, seria um gradiente latitudinal.

7.1

Os locais dos grupos 3 e 4 apresentam um gradiente de alterações de características, tais como os locais dos grupos 1 e 2. Curiosamente, os locais dos grupos 2 e 4 apresentam características próximas podendo dar a impressão de que estando a diferentes distâncias da foz, possuem características ambientais semelhantes.

7.2

Parece-me que tenha sido realizado sobre matriz de correlação, uma vez que quando se quer analisar muitas variáveis, na maioria das vezes estas apresentam diferentes escalas, logo é necessário utilizar uma matriz de correlações para que todas as variáveis apresentem o mesmo peso quando a análise é realizada.

7.3

Tendo em conta que tínhamos 15 variáveis ambientais e que os dois primeiros eixos explicam 92% da percentagem de variação isto quer dizer que a eficiência da análise é bastante elevada, pois o nosso objectivo é que o máximo possível de variância seja explicada pelas 1as componentes principais pois à medida que são acrescentados eixos há uma % da variância que é perdida e não é explicada, ora 92% é um valor bastante grande que permite ter confiança na nossa análise sendo que adicionar mais eixos não vai levar a uma melhor explicação da variância. Para além disso na verdade apenas temos capacidade para analisar a 2 dimensões, logo acrescentar componentes principais impede a correcta análise da PCA.

Commented [TM1]: incorrecto

Commented [TM2]: Dificulta uma eficiente interpretação da pca

8

Os dados tanto dos animais da primeira como da segunda espécie são emparelhados (nota TAM: são não independentes), pois as medições são efetuadas em 10 animais (um número não muito elevado) ao longo do tempo, são dependentes. Ele compara os dados de ambas as espécies, mas esqueceu-se deste pormenor.