IBI5086 – Métodos Estatísticos para a Bioinformática 2º Sem/2020

Alguns Exercícios

1. Considere os dados a seguir de um Estudo de Bioequivalência de um Medicamento Novo relativamente a um Medicamento Padrão no tratamento de uma doença. Respostas de pacientes a um Placebo também são apresentadas.



Neste tipo de estudo, órgãos reguladores adotam o seguinte critério para comprovação da bioequivalência do medicamento novo:



Construa um intervalo bootstrap de 95% de confiança para o parâmetro θ. Que conclusão é obtida? Na solução, considere:

a) Delineamento completamente aleatorizado (DCA) em que os três tratamentos foram atribuídos aleatoriamente a 8 pacientes de uma amostra de 32 voluntários, que aceitaram participar do estudo e cumpriram a todos os critérios de inclusão.

b) Delineamento Aleatorizado em Blocos Completos em que os três tratamentos foram aplicados a 8 pacientes voluntários, que aceitaram participar do estudo e cumpriram a todos os critérios de inclusão. Para cada paciente a ordem do tratamento foi aleatorizada, sendo que entre os tratamentos foi respeitado um período de “limpeza” (washout) do organismo.

2. Considere os dados a seguir apresentados em Manly (2007), que correspondem à distribuição (em porcentagem) de indivíduos de 17 colônias de caracol de acordo com o tipo de concha. As colônias estão classificadas de acordo com o tipo de habitat dos caracóis.

Como trata-se de dados composicionais (cada linha soma 100%), na realização das análises a seguir exclua a última coluna de dados.

a) Em Análise de Componentes Principais (CP), por exemplo, da matriz Y17x9 dos dados dos caracóis, há interesse em avaliar quantos componentes devem ser retidos na redução de dimensionalidade dos dados. Um dos critérios é adotar os primeiros CP extraídos dos dados originais com variância (autovalor) maior ou igual à média dos autovalores. No caso de dados padronizados esse critério corresponde a considerar os CP com autovalores maiores ou iguais a 1. Para os dados dos caracóis obtenha os CP e os correspondentes intervalos bootstrap a 95% de confiança para os autovalores.

b) A Função Discriminante de Fisher é um classificador ótimo em muitas situações práticas. É obtida a partir da decomposição espectral de W-1B, com W e B matrizes de covariância entre e dentro de grupo, respectivamente. Os autovalores desta decomposição espectral podem ser usados no seguinte teste de significância do poder discriminante das funções de Fisher:



Em que n é o tamanho amostral (n=17, no caso dos dados dos caracóis), p é o número de variáveis (p=9), G é o número de grupos (G=6), λj é o j-ésimo autovalor (j=,1,2,...,min(n,p,G-1)) e χ2 é a distribuição Qui-Quadrado com o correspondente número de graus de liberdade. A significância desta estatística para um conjunto de dados amostrais é uma indicação de que a discriminação não é aleatória (esporádica). Para os dados dos caracóis obtenha as duas primeiras funções discriminantes e represente as 17 colônias. Os 6 habitats podem ser bem preditos? Com base na estatística Qui-Quadrado o que é possível concluir sobre o poder discriminante das funções de Fisher neste caso? Construa também os correspondentes Testes de Aleatorização.



3. Agora, considere os dados dos caracóis com o objetivo de calcular a correlação canônica entre os dois conjuntos de variáveis definidas para conchas yellow e pink (com p=q=4). Calcule os coeficientes de correlação canônicos para estes dados. Quais variáveis mais contribuíram para a correlação? Obtenha também, para cada coeficiente, intervalos bootstrap de 95% de confiança. Esses intervalos incluem o zero?

4. Com base em sua área de trabalho e pesquisa, contextualize um conjunto de dados com estrutura similar aos dados dos caracóis, isto é, resposta composicional, muitas variáveis sendo avaliadas e podendo ser particionadas em dois conjuntos e, finalmente, com os indivíduos sendo classificados em diferentes grupos.

Boa Sorte ☺