

# Trabalho Final

Nesta página estão as duas propostas de trabalho final.

## Proposta Principal

A contagem de estruturas em animais, principalmente quando estas são em grande número e/ou com pequenas dimensões, pode acarretar em incerteza e carregar consigo possíveis erros. Por essa razão, é convenção realizar mais de uma contagem, seja ela feita por um mesmo leitor (chamamos assim de contagens “cegas”, pois a pessoa não pode ter acesso ao número que contou anteriormente), ou por leitores diferentes. Um exemplo prático deste problema está em uma das abordagens do meu projeto de mestrado: contar anéis de crescimento em estatólitos de lulas. Os anéis apresentam deposição diária e portanto, a partir do número de anéis contados, é possível inferir a idade aproximada do indivíduo. O grande problema é que estes estatólitos possuem cerca de 1mm de diâmetro e as contagens devem ser feitas sob microscópio, em aumento de 1000x e com óleo de imersão. Uma forma de verificar a precisão de contagens ocorre por meio do cálculo de dois índices quantitativos: (1) o erro médio percentual (Bearnish & Fournier, 1981) e (2) o coeficiente de variabilidade (Chang, 1982), cujas equações são mostradas abaixo.

(1)



(2)



nas quais:

$N$  = número de indivíduos

$L$  = número de contagens

$S$  = desvio padrão

$X_{ij}$  =  $i$ ésima contagem do  $j$ ésimo indivíduo

$\bar{X}_j$  = leitura média do  $j$ ésimo indivíduo

Minha proposta é criar uma função que calcule os dois índices mostrados acima (o usuário poderá escolher se deseja calcular somente um dos índices ou os dois simultaneamente) a partir dos dados de contagem fornecidos pelo usuário. As contagens de cada indivíduo também serão comparadas duas a duas através de seus coeficientes de correlação. Neste caso, será testada a hipótese de que contagens semelhantes originariam coeficientes iguais a 1. Esta hipótese será testada por meio de um teste  $t$  ao nível de 5% de significância. No meu plano inicial, a função permitirá a entrada de três argumentos, sendo eles: o conjunto de dados (os valores de cada contagem), o número de indivíduos total e o número de contagens feita para cada indivíduo. O primeiro argumento poderá ser um data-frame ou uma matriz. Em ambos os casos, as linhas devem representar cada um dos indivíduos e as colunas devem representar os valores de cada contagem. O objeto de saída da função será uma lista contendo os valores calculados de cada índice. Os valores serão retornados como porcentagem. Por convenção, valores abaixo de 10% são considerados como aceitáveis, ou seja, conclui-se que as contagens não são significativamente diferentes entre si. Assim, ao final da função, propõe-se também a saída de uma mensagem na tela, juntamente ao resultado, alertando o usuário se os

valores obtidos estão dentro ou acima do limite aceitável.

Função relativamente simples, talvez até demais. A maior dificuldade acho que seria lidar com dados faltantes ou com entradas em formatos diferentes. Como vc lidaria com a entrada se cada linha tivesse um valor de contagens diferentes? Qual seria o melhor formato da entrada nesse caso?

—*Diogo Melo*

Diogo, não entendi sua pergunta. Cada linha corresponde a um indivíduo e portanto as linhas devem conter valores diferentes. Pode ser que para o indivíduo 1 sejam contados 100,110 e 90 aneis (exemplo do meu trabalho), e que o indivíduo 2 tenha 250,230,235 aneis. Eu vou aplicar a média para cada linha e comparar com os valores daquela linha específica. Acho que isso não ficou claro com a segunda fórmula, pois parece que é um valor único de desvio padrão, mas na verdade esse valor é para cada indivíduo. Nos dois coeficientes, os resultados são gerados para cada linha. Espero que tenha sido essa a sua dúvida em relação à função, e que eu tenha conseguido explicar melhor.

—*Lígia H.A.*

É, não fui muito claro mesmo... O problema seria ter número de contagens diferentes por linha, por exemplo, 4 contagens pro indivíduo 1 (250, 230, 235, 240) e 3 contagens pro indivíduo 2 (100,110 e 90). O melhor é deixar a entrada no formato que se adeque ao indivíduo com o maior numero de contagens e considerar os outros faltantes, ou usar um formato diferente? Por exemplo colocando uma contagem por linha e incluindo uma coluna de indivíduo? Talvez seja interessante a função aceitar os dois jeitos? Só estou pensando em como deixar a proposta mais interessante.

—*Diogo Melo*

## Proposta Alternativa (Plano B)

Nos processos de seleção sexual, a forte competição entre machos pode resultar em dimorfismo intrassexual, caracterizado pela descontinuidade de traços morfológicos, fisiológicos e de ciclo de vida entre indivíduos do mesmo sexo. Esse dimorfismo é encontrado em diversos grupos de animais – inclusive em lulas, meu objeto de estudo no mestrado – , resultando em uma diversidade fenotípica que inclui machos grandes e pequenos. Para uma investigação inicial sobre a possível existência de dimorfismo intrassexual entre machos em determinada população, proponho a criação de uma função, descrita abaixo. No plano inicial, a função permitirá a entrada de um argumento principal: um data-frame ou matriz que contenha duas colunas, nas quais estarão inseridos dados morfométricos (quantitativos) a respeito do grupo de interesse. As linhas devem representar cada indivíduo amostrado e as colunas devem seguir uma ordem específica: a primeira deve conter as medidas de tamanho corporal dos espécimes e a segunda deve conter medidas de interesse que serão associadas ao tamanho do corpo (peso gonadal, tamanho de espinhos ou armamentos, etc). A função irá gerar, como objetos de saída, um conjunto de gráficos, sendo eles: (1) um histograma com a distribuição dos valores da primeira coluna, para que o usuário consiga visualizar a presença (ou não) de duas classes de tamanho, (2) gráficos de dispersão entre as variáveis inseridas nas duas colunas (3) o resultado do modelo alométrico de dimorfismo, descrito por Eberhard & Gutiérrez (1991). Este modelo parte da premissa de que o dimorfismo intrassexual entre machos é caracterizado pela descontinuidade de traços morfológicos, como mencionado anteriormente.

### Modelo de dimorfismo intrassexual (Eberhard & Gutiérrez, 1991)

A proposta da função é fazer uma investigação inicial sobre a possível existência de dimorfismo intrassexual entre machos. Para isso, será aplicado um modelo linear para determinar se a relação entre tamanho do corpo e a característica escolhida é não-linear. A equação do modelo linear será:

$$\ln Y = \alpha_0 + \alpha_1 \ln X + \alpha_2 \ln X^2 + \epsilon$$

no qual: Y é a característica analisada, X é o tamanho corporal,  $\alpha$  é o coeficiente de regressão e  $\epsilon$  é o erro associado, assumindo uma distribuição normal com variância constante. Caso o coeficiente  $\alpha_2$  não seja significativamente diferente de zero, conclui-se que a relação entre as variáveis não apresenta desvios significantes da linearidade e que não há descontinuidade nos traços morfológicos. Neste caso, a função retornará ao usuário o sumário da equação e uma mensagem na tela resumindo a conclusão final do modelo. Caso o coeficiente  $\alpha_2$  seja significativamente diferente de zero, i.e., a relação seja não-linear, conclui-se que possivelmente existe dimorfismo e descontinuidade nos traços morfológicos. Assim como no caso anterior, a função retornará ao usuário o sumário da equação e uma mensagem na tela resumindo a conclusão final do modelo.

Observação: caso seja viável e possível dentro do prazo estipulado para enviar a função final (10/04/2015), uma nova análise será acrescentada à função descrita. Devido a sua complexidade, não me comprometo a inseri-la no projeto inicial. Para deixar a proposta mais concisa, os passos seguintes dessa nova análise foram colocados no arquivo abaixo.

Link para descrição do modelo: [modelos\\_eberhard\\_gutierrez.pdf](#)

Essa é mais interessante, mas não ficou claro pra mim como vc vai ajustar esses modelos. Só usando um modelo linear simples ou fazendo máxima verossimilhança? Só comparar

os modelos com e sem o segundo coeficiente me parece simples demais, a proposta estendida é melhor.

No primeiro modelo, oq é a variável X2 que vc associa ao coeficiente que indica dimorfismo intrasexual? Sem saber isso eu não consegui entender a ideia da comparação dos modelos.

Quanto aos gráficos diagnósticos, histogramas são horríveis pra avaliação visual de distribuições. Todas parecem iguais. É melhor usar um gráfico de densidade cumulativo (cdf) ou um qqplot.

Outra coisa é o intervalo que vc usa pra procurar os pontos de transição. Usar um intervalo fixo de busca pra todos os conjuntos de dados possíveis é uma ideia ruim, já que o intervalo de variação ou mesmo a escala podem ser muito diferentes. Tente fazer a função mais geral possível. Quanto à comparação, usar  $R^2$  não é a melhor coisa do mundo, e até difícil de definir em modelos mais complicados. Algum critério baseado em informação como AIC seria melhor do ponto de vista estatístico.

—[Diogo Melo](#)

Minha ideia é ajustá-los usando modelos lineares simples. Vou tentar fazer a análise completa, só achei que talvez fosse muito complexa com o que tivemos no curso. Mas, se acha que conseguirei fazê-la com o conhecimento que obtive no curso, tudo bem. Sobre seus comentários e sugestões: (1) na primeira equação, a variável é  $X^2$ . Foi um erro de digitação deixar como X2. Já consertei na proposta inicial. (2) vou alterar a saída gráfica para os tipos de gráfico que sugeriu, excluindo histogramas. (3) sobre os intervalos, também vou alterá-los. Talvez seja possível que o usuário entre o número de valores que deseja testar. Com isso, a função calcula a amplitude dos dados (máximo - mínimo) e divide essa amplitude pelo número dado pelo usuário, gerando os intervalos. (4) eu sugeri o uso do  $R^2$  porque o modelo é descrito dessa forma.

—[Lígia H.A.](#)

Estou achando a proposta B mais interessante, não tanto

pela implementação, que é relativamente simples, mas pq vc pode comparar os dois métodos de detectar dimorfismo (regressão quadrática ou presença de switch-point). É bem plausível de implementar os dois, e se vc tiver conjuntos de dados simulados (com e sem dimorfismo) vc pode comparar a eficiência de cada um em detectar o dimorfismo. Que tal? Eu posso ajudar com a implementação da simulação pelo forum.

—*Diogo Melo*

Tudo bem. Começarei a trabalhar na proposta B então. Assim que eu tiver o código mais ou menos pronto, crio um tópico no fórum para discutirmos a simulação de dados para testar a função. Obrigada pela ajuda e pelas sugestões.

—*Lígia H.A.*

### Referências:

Beamish, R.J. & Fournier, D.A. 1981. Method for comparing the precision of a set of age determinations. Canadian Journal of Fisheries and Aquatic Sciences, 38: 982-983.

Chang, W.Y.B. 1982. A statistical method for evaluating the reproducibility of age determination. Canadian Journal of Fisheries and Aquatic Sciences, 39:1208-1210.

Eberhard WG, Gutiérrez EE (1991) Male dimorphism in beetles and earwigs and the question of developmental constraints. Evolution 45:18-28

From:

<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:

[http://ecor.ib.usp.br/doku.php?id=05\\_curso\\_antigo:r2015:alunos:trabalho\\_final:ligia.apostolico:trabfinal](http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2015:alunos:trabalho_final:ligia.apostolico:trabfinal) 

Last update: **2020/08/12 06:04**