

Plano A

Função para análise de dados de prevalência obtidos por meio da técnica de respostas randomizadas

Minha função irá estimar a prevalência de comportamentos [sensitivos](#) a partir de dados obtidos por entrevistas. Por prevalência entende-se a proporção da população praticando o comportamento de interesse. O usuário poderá escolher se a estimativa será baseada em dados obtidos pela técnica de questionamento direto ou pela [técnica de respostas randomizadas](#). Ele poderá ainda realizar um teste estatístico comparando as estimativas de prevalência obtidas pelos dois métodos.

Dados de entrada: vetor numérico com valores de zero e um, indicando se os entrevistados responderam “não” ou “sim”, respectivamente, à questão sobre a prática do comportamento de interesse. Serão necessários argumentos indicando a origem dos dados (questionamento direto ou técnica de respostas randomizadas). Caso tenha sido utilizada a técnica de respostas randomizadas, será necessário um argumento indicando as probabilidades de o entrevistado responder a verdade ou responder “sim” independente da verdade. O usuário poderá também comparar estimativas obtidas simultaneamente pelos dois métodos, neste caso ele deverá inserir argumentos indicando dois vetores numéricos, um para cada método. Os vetores poderão conter NAs.

Dados de saída: uma lista com NAs omitidos, resultados da estimativa de prevalência para cada método e intervalo de confiança gerado por simulações de bootstrap, gráfico whisker-plot indicando a estimativa de prevalência e o intervalo de confiança, resultado de teste estatístico comparando as estimativas obtidas pelos dois métodos.

Plano B

Função para estimar o tamanho de populações com base em “contagens totais” incompletas

Estimativas de tamanho populacional são extremamente importantes em biologia da conservação. Existem vários métodos disponíveis para estimar o tamanho populacional, os quais possuem diferentes premissas e metodologias (captura-recaptura, contagens, etc.).

Minha função irá utilizar o método de contagem dupla (“double count” ou “two counts”) para estimar o tamanho de uma população a partir de “contagens totais” incompletas (maiores detalhes em [Sinclair et al. 2006:240](#)). Este método é apropriado para diversas situações onde não é conveniente realizar capturas e quando as entidades são estacionárias e suas posições podem ser mapeadas. Possíveis aplicações incluem estimativas do número de ninhos de quelônios, crocodilianos e aves em determinada área, ou estimativas populacionais de mamíferos a partir de fotografias tiradas ao longo de transectos aéreos, conduzidos em aviões convencionais ou em veículos aéreos não-tripulados (VANTS).

O método guarda semelhanças com a técnica de marcação e recaptura, mas não há necessidade de capturar ou marcar as entidades. As entidades são consideradas estacionárias e suas posições individuais podem ser mapeadas. Se determinada área é amostrada duas vezes, de forma independente, então as entidades presentes podem ser divididas em quatro categorias:

S1 = número de entidades visto na primeira amostragem, mas perdido na segunda

S2 = número de entidades visto na segunda amostragem, mas perdido na primeira

B = número de entidades visto nas duas amostragens

M = número de entidades perdido nas duas amostragens

Se P1 é a probabilidade de uma entidade ser vista na primeira amostragem e P2 a probabilidade de ser vista na segunda amostragem:

$$P1 = B/(B+S2)$$

$$P2 = (B/(B+S1))$$

$$M = S1S2/B$$

$$Y = (B+S1)(B+S2)/B$$

Onde Y é a estimativa de tamanho da população. Existem fórmulas para estimar a variância da estimativa ([Sinclair et al. 2006](#))), mas esta também pode ser estimada por simulações de bootstrap.

A função irá calcular P1, P2, M e Y e estimará o intervalo de confiança de Y a partir de simulações de bootstrap.

Dados de entrada: data-frame onde cada coluna é um vetor numérico correspondente aos dados de uma amostragem independente, e onde cada linha corresponde a uma entidade. O número total de linhas deve ser igual ao número de entidades encontradas nas duas amostragens (B), e as linhas deverão ser preenchidas com zeros e uns representando avistamento ou não da entidade na amostragem.

Dados de saída: Lista com estimativa de tamanho populacional e intervalo de confiança.

Gostei das duas propostas, ambas estão bem estruturadas!
O plano A me parece bem interessante e, supondo que você prefira investir nele, sugiro que vá em frente.

- Tenho uma sugestão: por que, em vez de colocar um vetor numérico de 0 e 1 como objeto de entrada, você não dá também a opção de inserir o arquivo externo (em algum formato que você costume usar, como .csv, por exemplo) com os dados mais “brutos” da pesquisa? Você poderia transformar esses dados em um vetor de 0 e 1 dentro da sua função.. facilitaria a vida do usuário e a função ficaria mais completa!
- Tenho uma dúvida: caso a escolha do usuário seja a técnica de respostas randomizadas, há mais de um valor de probabilidade de resposta SIM ou de resposta verdadeira? Não é sempre a mesma, a partir do cálculo que você mostrou [aqui](#)?

—[Luísa Novara](#)

Oi Luísa, obrigado pelos comentários. Estou trabalhando no plano A, mas já estou com algumas dificuldades. Parece que, sem perceber, propus duas funções diferentes: (i) uma função onde o

usuário entra apenas com um vetor de dados e escolhe o tipo de análise (questionamento direto ou técnica de respostas randomizadas); (ii) uma função onde o usuário entra com dois vetores de dados (um para cada método) e compara os resultados, com opção de realizar teste estatístico (pensei num teste de permutação). Observe que a lista de argumentos é diferente para as duas propostas, portanto são funções diferentes. Estou trabalhando na segunda alternativa, onde o usuário entra com dois vetores de dados. Na versão final, pretendo simular os dados para teste, mas por enquanto estou usando dados de uma situação real. Sobre sua dúvida, as probabilidades resposta verdadeira e SIM podem ser manipuladas pelo pesquisador, as probabilidades que usei no meu link explicativo e como default na função são apenas um exemplo.

Oi, Elildo! Achei que você tivesse percebido que a função seria executada de duas formas diferentes, uma quando se seleciona apenas uma técnica e outra quando se selecionam duas (a fim de compará-las). Não acho que seja uma boa você restringir sua função apenas à segunda situação, porque ela se torna limitada para a maior parte dos casos (em geral, os pesquisadores só coletam os dados uma vez, usando um só tipo de técnica). Mas você pode juntar as duas na mesma função sim! Você pode comparar os resultados advindos das duas técnicas só quando for colocado mais de um vetor como objeto de entrada, algo do gênero. E não entendi o que vc falou sobre o objeto de entrada.. você não acha uma boa possibilitar que ele seja em outro formato, já que é necessário um esforço grande pra passar as informações brutas da pesquisa para um vetor de 0 e 1?

—[Luísa Novara](#)

Oi Luisa, mais uma vez obrigado. Concordo que é melhor permitir que a função seja flexível e permita a entrada de diferentes tipos de dados, incluindo um único tipo. Estou trabalhando nisso. Sobre o objeto de entrada, acho que o ideal é manter a entrada como está, com vetores de zero e um. Os dados obtidos nessas pesquisas são sempre binários e em geral os pesquisadores já digitam os dados neste formato. Incluir na própria função a tarefa de carregar os dados externos seria possível, mas a tornaria deslegante e até mesmo confusa, pois teria argumentos adicionais para uma tarefa que o usuário pode executar facilmente fora da função, por meio de manipulação de dados (supondo que ele já não tenha inserido os dados neste formato).

Oi, Elildo! Apesar de não ter respondido antes, eu li seu comentário faz tempo (desculpe-me por isso)! E você me convenceu quanto ao objeto de entrada!

—[Luísa Novara](#)

From:
<http://labtrop.ib.usp.br/> - **Laboratório de Ecologia de Florestas Tropicais**

Permanent link:
http://labtrop.ib.usp.br/doku.php?id=cursos:ecor:05_curso_antigo:r2015:alunos:trabalho_final:elildojr:trabalho_final 

Last update: **2020/07/27 18:48**