

Bruno Henrique de Moraes e Souza

Mestrando em oceanografia biológica pelo Instituto Oceanográfico da Universidade de São Paulo.

e-mail: brunohms@usp.br

EXERCÍCIOS

Oi Bruno, acho que a propsta A, incorporando as sugestões do Ogro está de bom tamanho. — [Alexandre Adalardo de Oliveira](#) 2016/04/29 10:52

Trabalho Final

Proposta 1

Estatística Hipergeométrica

Introdução

Um experimento hipergeométrico possui duas características principais, a primeira é que cada indivíduo pode ser considerado como um sucesso ou falha na amostra. Outro ponto, é que a amostra é selecionada aleatoriamente, com um número qualquer de indivíduos de uma população conhecida e finita, onde não há reposição em caso de falha ou sucesso.

Descrição

A função tem o objetivo de mostrar a probabilidade de se ter sucesso em uma amostragem de uma população total e específica conhecidas, ou seja, quais são as chances de sucesso em um determinado tamanho amostral, para isso serão fornecidas todas as possibilidades amostrais, pré definidas de acordo com o intervalo e o número máximo amostral desejados, onde uma amostra com 0 indivíduos represente o mínimo (0%), e uma amostra do tamanho da população total representa o máximo de chances de sucesso (100%), para que com esses dados se encontre o melhor valor amostral para obter o número desejado de coletas positivas dentro de um número pré definido de possibilidades amostrais viáveis.

Argumentos

(x, y, z, w, a, b)

x = vetor do tamanho da população total (todas as espécies)

y = tamanho da população específica (espécie de interesse)

z = número de sucessos desejado (capturas da espécie de interesse)

w = vetor do intervalo de tamanhos amostrais viáveis (ex: de 100-150 indivíduos)

a = porcentagem mínima desejada de sucesso (ex: 40% - 0,4)

b = aumento mínimo percentual relevante para a captura (ex: 1% - 0,01; pois valores inferiores a esse incremento na probabilidade não representariam vantagens para a amostragem)

Input

Tamanho da população total, tamanho da população de interesse, intervalo de tamanho das amostras e número de sucessos desejados, porcentagem de sucesso desejada e incremento mínimo aceito que represente uma vantagem nas capturas.

Output

Um gráfico com as probabilidades de se ter sucesso em uma análise hipergeométrica de diferentes tamanhos amostrais, com o intuito de identificar intervalo amostral, levando em conta o percentual desejado de capturas e o valor mínimo de incremento de chances de sucesso.

Exemplo

Em uma comunidade hipotética com um número conhecido ou aproximado de indivíduos de diferentes espécies é necessário coletar representantes de uma única espécie de interesse, que também possui uma abundância conhecida ou aproximada, qual é o número necessário de capturas para se obter um número estipulado de sucessos (captura da espécie de interesse)? A função colocaria as probabilidades de captura de acordo com o número amostral de algum intervalo viável para se definir a amostra ideal.

Parece estar havendo uma certa confusão conceitual nessa proposta. Acho que um exemplo vai ajudar. Do jeito que está vc descreveu a função `phyper()` ou alguma variação muito simples dela. A ideia é reescrever a função do R? Com qual propósito?

Valeu!

—Ogro

Res: Oi Ogro tudo bem? Obrigado pelos comentários. Tenho preferência por essa função para o meu trabalho final. E meu objetivo com ela é na verdade seguir a ideia do exemplo, ou seja, ao utilizar uma análise hipergeométrica gostaria de saber qual o melhor tamanho amostral para obter sucesso na coleta, o gráfico daria as respostas das probabilidades de acertos para cada número amostral definido, é diferente da função `phyper()` que utiliza um único número amostral. Quanto maior o número de capturas, maior será a possibilidade de êxito, porém é possível que percentualmente esse aumento não seja significativo e não exista a necessidade de um número elevado de capturas. Redefini os argumentos da função de forma mais coesa.

Ficou um pouco mais claro. Então a ideia é que vc tem uma população com composição conhecida, e vc pretende fazer amostras aleatórias dessa população visando capturar um numero determinado de uma determinada espécie. Vc pretende escrever uma função que calcula a probabilidade de vc atingir o numero de capturas (sucessos) desejada com certo número de amostras, ou mesmo calcular a probabilidade para vários números de amostras.

Me parece ok, mas talvez fosse legal colocar uma saída gráfica, só pra deixar mais bonitinho. Tipo uma região onde a chance de chegar no numero de capturas seja maior que uma certa probabilidade determinada pela usuário. Que tal? ilidade valeu!

---Ogro

Res: Gostei da sugestão, mas sem dúvidas quanto maior o número amostral dessa população maior será a probabilidade de sucesso, mas é interessante sim para ressaltar uma parte do gráfico. Pensei em colocar uma outra saída nesse gráfico, uma área em que a variação da probabilidade fosse menor do que um valor determinada pelo usuário, assim seria possível determinar um intervalo em que há uma probabilidade de sucesso mínima deseja (previamente estabelecida como na sua sugestão) com um esforço amostral mínimo, em que o aumento de capturas não represente vantagem. Essa proposta está boa o suficiente em sua opinião Diogo?

Proposta 2

Análise Dependente

Introdução

Muitas análises realizadas são dependentes de um quadro negativo ou positivo de uma situação anterior, dessa forma é necessário realizar algumas dessas análises estatísticas levando em conta esse pré-requisito para que funcionem adequadamente.

Descrição

Uma função que analise a ocorrência de eventos de acordo com a dependência de outros fatores, em

que uma parte binária influenciará no resultado da função alterando os possíveis resultados.

Argumentos

$(x, y, z1, z2, z3, w1, w2, w3)$

x = conjunto de dados da variável preditora (ex: número de indivíduos de uma espécie ou população de interesse)

y = conjunto de dados da variável resposta quantitativa (ex: taxa de recrutamento por indivíduo)

$z1, z2, z3$ = variáveis binárias (true or false)

$w1, w2, w3$ = conjunto de dados associados à alguma alteração devido à uma variável binária.

Input

Dois vetores para os eixos x e y , além de uma ou mais variáveis binárias.

Output

Um gráfico com a relação dos vetores de acordo com variáveis binárias, que podem influenciar no resultado final.

Não entendi o propósito da função, acho que falta contexto. O output é um gráfico, mas a probabilidade é só um número, certo? Não seria só a fração de sucessos?

Ou vc está pensando mais num tipo de modelo mistura binomial/alguma coisa? Um coisa pode ou não ocorrer, dependendo de alguns preditores, e se ocorrer segue alguma outra distribuição. É essa a ideia?

Do jeito que está não faz sentido, tente pensar num exemplo ou numa descrição mais clara e reescreva.

qq coisa pode comentar no forum ou aqui mesmo.

Valeu!

—Ogro

Res: A ideia é o que você colocou mesmo, uma mistura binomial junto de uma tabela de x e y , em que algo pode ou não ocorrer devido à uma circunstância anterior, podendo seguir ou não uma outra distribuição, ou tornando-se inexistente (NA). Tinha colocado como probabilidade mas na verdade não

é. Quanto ao exemplo e uma descrição mais minuciosa farei o mais breve possível.

Ok, então vc gostaria de implementar algum tipo de modelo de mistura. Mas ainda não está claro com qual propósito, e como seriam os dados. Geralmente modelos de mistura desse tipo visam lidar com um excesso de zeros numa amostra. É esse tipo de situação que vc está pensando? De novo, um exemplo ajudaria.

Valeu!

—Ogro

Res: Pensei em alguma relação com taxas de recrutamento e disponibilidade de substratos adequados para algumas espécies, onde esse ambiente poderia fornecer um condições ideais de crescimento, ou condições que não fossem adequadas mas que não impedissem o desenvolvimento do animal e a possibilidade de o substrato impedir o crescimento. Faria isso com mais de uma variável binária (se sobrevive ou não, e depois se a taxa é ótima ou não). É possível pensar em espécies que se desenvolvem bem em alguns substratos, razoavelmente em outros e não se fixam em alguns. Para isso em seria necessário um conjunto com novos valores para as quantidades não ideias de desenvolvimento. Essa ideia não está totalmente clara para mim ainda, mas esse é um exemplo que consigo apontar e sei que são necessários outros argumentos ainda.

From:

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

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2016:alunos:trabalho_final:brunohms:start 

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