Rodolfo Liporoni Dias

Mestrando em Ecologia, Instituto de Biociências, USP.

Laboratório de Abelhas

Meu Currículo Lattes

Contato: rld@usp.br

Trabalho com a comunidade de abelhas noturnas da Mata Atlântica, buscando entender aspectos básicos de sua biologia (morfologia, comportamento, nidificação), abundância, riqueza, composição e sazonalidade.

Especificamente, pretendo elucidar inicialmente como a lua influencia os padrões de atividade noturna das abelhas. Será que elas forrageiam de acordo com o ciclo da lua?

Meus exercícios

exec

Proposta de Trabalho Final

Plano A: Biologia reprodutiva vegetal

Tarefa a ser executada

- Analisar dados de diferentes tratamentos de biologia reprodutiva vegetal.

- Apresentar um sumário com os principais índices usados pela área.

- Construir um gráfico que mostre a relação entre dois índices, seja para nível populacional ou de comunidades.

Utilidade da função

- Automatiza as análises exploratórias e os testes estatísticos para estudos de biologia reprodutiva vegetal básicos e rotineiros. Especialmente importante visto o recente esforço global de caracterizar a polinização e os sistemas reprodutivos das espécies vegetais para entender quais espécies são

mais vulneráveis e evitar o colapso das populações dependentes de polinizadores.

Entrada da função

- Primeiro argumento: um objeto da classe data.frame que contenha a proporção de flores convertidas em frutos (número de frutos por número de flores) em vários tratamentos usuais de biologia reprodutiva, no seguinte formato:

1. observações são as proporções de flores convertidas em frutos, e as colunas são diferentes tratamentos de polinização;
2. primeira coluna: espécies (se for uma só, todas as linhas devem conter o nome da espécie);
3. segunda coluna: número de identificação de cada indivíduo;
4. terceira: tratamento controle (polinização natural);
5. quarta: tratamento de autopolinização manual;
6. quinta: tratamento de polinização cruzada manual.

- Segundo argumento: qual o nível ecológico do gráfico de dispersão? População ou comunidade (para data.frames com mais de uma espécie). Se comunidade, gráfico rotula cada ponto (que representa a média dos índices) com os nomes das espécies.

- Terceiro argumento: qual teste estatístico usar para diferenciar os tratamentos? Opções mais usadas pelos pesquisadores na área são o teste t e o qui-quadrado.

Saída da função

- Sumário com os testes estatísticos para os tratamentos, par a par.

- Sumário com os 3 índices calculados por espécie: índice de autoincompatibilidade (ISI), índice de limitação polínica (IPL) e índice de eficiência reprodutiva (REI). Fórmulas dos índices mais usados

- Gráfico de dispersão que relaciona ISI e IPL (cada ponto ou é um indivíduo ou é a média dos índices entre os indivíduos de cada espécie, de acordo com a escolha feita no segundo argumento).

Dúvidas e desafios

- Encontrar uma maneira genérica de filtrar cada espécie para fazer as análises e construir os gráficos, visto que são objeto da classe caráter. Função subset seria útil para isso?

- Como construir o terceiro argumento, já que ele está relacionado a todos os outros. Seria usando if dentro de if?

> Oi, Rodolfo!
>
> Adoro sua organização! <3 Está tudo muito claro, parabéns e obrigada, hehe. Seu plano A está bom (mas veja meu

comentário sobre o plano B! *.*). Sobre as duas dificuldades:

- Não entendi porque os nomes das espécies serem caracteres impossibilita o uso de funções que vcs já aprenderam (aggregate e família apply), talvez eu tenha entendido alguma coisa errada sobre o que vc quer fazer então. O que vc quer é calcular uma média por espécie (e depois fazer os testes estatísticos) de cada índice para cada tratamento, certo? Imagino que de algo assim (aqui vou calcular a média das proporções de fruto, mas poderia ser o cálculo dos índices por espécie) :

```
  especie individuo controle auto cruzada
1     sp1         1      0.5  0.2
0.8
2     sp1         2      0.6  0.1
0.9
3     sp2         3      0.4  0.1
0.7
4     sp2         4      0.5  0.0
0.8
5     sp3         5      0.4  0.0
0.7
```

  Você queira chegar em algo assim:

```
  Group.1 controle auto cruzada
1     sp1     0.55 0.15    0.85
2     sp2     0.45 0.05    0.75
3     sp3     0.40 0.00    0.70
```

  Não é isso? Você pode usar a função subset para gerar um data frame para cada espécie e depois calcular o que vc quiser tb, mas não me parece necessário. Me avise caso eu esteja deixando passar alguma coisa!

- Pelo que entendi, vc que deve decidir quando fazer um teste qui-quadrado ou um teste t, certo? Sendo assim, vc não precisa deixar isso como um argumento da função, pode colocar esse if dentro dela mesmo, e avisar qual teste foi feito (e esclarecer no help quando cada tipo de teste será realizado). É isso?

—*Luísa Novara*

Oi Lu!

Obrigado pelos comentários e elogios! E pela devolutiva ligeira também, assim já posso ir adiantando a função hehe

Então, é exatamente isso que quero fazer. A dica do aggregate() ajudou muito, já descobri como usar. Valeu!

Já os testes estatísticos, eu pensei em deixar o usuário escolher mesmo. Nunca consegui detectar um padrão nos estudos de biologia reprodutiva vegetal que li, alguns autores optam por usar o qui-quadrado, outros testes t, sem justificar muito. Então, na dúvida, pensei em deixar o usuário escolher mesmo, até porque deve ser possível optar por ambos mesmo, a depender da pergunta do trabalho né? A função precisa fazer uma comparação entre as médias de cada tratamento, par a par (natural vs cruzada, natural vs auto e auto vs cruzada). Pensei em criar a função para o teste t e depois repetir todo o código, só que para o teste qui-quadrado. Isso é aceitável/desejável?

Pensei agora que poderia gerar um gráfico boxplot também para cada tratamento por espécie indicando com letras se há diferença significativa entre os tratamentos. Só que fiquei pensando como operacionalizar isso para um data frame que tenha muitas espécies (além da melhor forma de mostrar inúmeros gráficos, sabe?)

Abraços,

R.

Oi, Rodolfo!

Sobre os testes estatísticos: ok, tinha entendido errado. Neste caso, é melhor deixar a opção pro usuário mesmo. Vc antes tinha sugerido a opção de usar o if(), não é? Neste caso, não precisaria repetir o código todo não, porque os dois “caminhos possíveis” só precisariam existir no momento de realização do teste. Ainda não entendi o porquê precisaria ser um if() dentro do if() como vc tinha falando antes (na verdade, não entendi como o argumento do teste estatístico está relacionado aos outros argumentos, isso que me gerou a confusão..).

Sobre os gráficos, acho que vc já conhece as funções que podem te ajudar a cria-los a partir de um data frame que tenha várias espécies. A lógica de aplicar funções a linhas específicas de um data frame é a mesma. Em relação ao

problema de talvez ter que gerar vários gráficos (a Lu está com a mesma questão, né?), ainda não sei o que fica mais bonito/viável... Talvez apresentar várias janelas com alguns gráficos em cada, não tenho opinião formada ainda, hehe. Mas acho bem bacana a ideia de fazer esses gráficos sim!

Beijo,

—*Luísa Novara*

Plano B: Como montar uma restauração?

Tarefa a ser executada

- Sugerir um mapa de plantio de mudas para restaurações.

Utilidade da função

- Uma das partes mais importantes de qualquer projeto de restauração é escolher o protocolo de plantio e então, de forma árdua, pensar na melhor maneira de dispor as mudas das espécies disponíveis na área de plantio. A função busca automatizar esse processo seguindo um protocolo bastante utilizado que determina que o plantio ocorra em linhas igualmente espaçadas e que linhas de plantio com espécies pioneiras devem ser intercaladas com linhas com espécies não pioneiras (Figura ilustrativa), e indivíduos da mesma espécie devem preferencialmente não ser plantados lado a lado.

Entrada da função

- Primeiro argumento: objeto da classe data.frame que contenha as espécies que se deseja usar na restauração, o número de mudas disponíveis por espécie e o grupo sucessional a que pertencem, nessa ordem (3 colunas, portanto).

- Segundo argumento: quais as dimensões (comprimento e largura) da área de plantio? A função trabalhará com áreas que sejam retângulos aproximados. Variações da área real para outros polígonos são complicações que dificultariam muito a preparação desta função. Contudo, como quadriláteros são um formato comum de propriedades rurais, a aplicabilidade da função está garantida.

Saída da função

- Uma matriz que mostra a disposição das mudas na área de plantio que seja a mais adequada para o protocolo especificado acima.

* Alternativa seria um gráfico de dispersão que fizesse a mesma coisa, onde cada ponto representasse uma muda no terreno, rotulada pela espécie e pelo grupo sucessional. Seria uma alternativa mais fácil?

Dúvidas e desafios

1. Encontrar uma maneira de organizar um data frame intercalando os dados de espécies pioneiras com não pioneiras sem perder a informação à qual espécie pertence para jogar na matriz.
2. Encontrar uma forma mais prática de inserir cada muda numa posição da matriz respeitando as regras do protocolo (uma complicação extra é fazer a função entender que cada espécie tem número de mudas diferentes e precisa aprender a repetir espécies mais numerosas mais vezes).
3. Pensei em transformar a coluna do grupo sucessional num vetor lógico para operacionalizar a intercalação do mesmo. Será que isso facilita ou dificulta minha tarefa?

> Rodolfo! Sua proposta A é boa, mas sei que está na sua zona de conforto. A sua proposta B é muito empolgante! Acho que vc iria aprender bastante com seus desafios (vc teria que quebrar um pouco cabeça com indexação, e pra mim aprender R é aprender indexação, hehe), e vi que vc conseguiria resolver com as funções que vcs aprenderam no curso, o que é ótimo! Sobre seus desafios:
>
> 1. Dica: dá uma olhada no capítulo 3 da apostila
> 2. Depois que vc resolver o desafio 1, a parte de preencher a matriz com os dados fica fácil. Sobre a complicação extra: em vez de exigir que o usuário indique a quantidade de mudas de cada espécie, fica mais fácil se vc também apresentar como output o número de mudas de cada espécie que ele precisa plantar (já que o usuário vai indicar a área do plantio). Assim vc não precisaria se preocupar com a diferença no número entre pioneiras e não-pioneiras ou entre espécies de um mesmo grupo sucessional. De qualquer forma, se vc optar por deixar como está, é possível que este problema extra se resolva como efeito colateral de resolver o desafio 1, hehe. Porque se vc tem dois vetores de tamanhos diferentes (por exemplo, um de indivíduos do grupo pioneiro P e outro de indivíduos do grupo não-pioneiro NP) e vc fizer com que os elementos desses vetores se intercalem (desafio 1), o que o R vai fazer é deixar que no final se acumulem vários elementos do vetor maior (P NP P NP P NP P NP P P P P P P). E isso não é um problema, imagino..
> 3. (relacionado ao desafio 1) Acho que me soa um

> caminho bom. A forma como pensei em resolver o desafio 1 não envolve transformar a coluna do grupo sucessional em um vetor lógico, mas acho pode ser um bom ponto de partida pra pensar as coisas..

Sobre a forma de saída da função, acho que a matriz está bom, mas seria mais legal ainda se vc fizesse um gráfico a partir da matriz e apresentasse os dois. Porque com a matriz a informação sobre a distância entre as mudas se perde, e vc teria que acrescentar uma saída em forma de texto também indicando a distância entre as mudas pro plantio.

Se vc estiver com dúvidas/inquietações, podemos conversar bastante ainda! Me escreva por aqui, mas me mande e-mail me avisando que vc atualizou o wiki pra agilizarmos a comunicação, ok? (: Beijo

—*Luísa Novara*

Oi Lu, concordo com você que estou um pouco na zona de conforto com a proposta A. Gostei das duas, mas tentarei me empenhar mais na proposta B então hehe. Indexação é massa mesmo!

Obrigado pela dica, verei o capítulo 3 com calma. Eu consegui intercalar P e NP por linha do data frame criando uma coluna extra com um índice artificial de 1 a 10 (para P) e de 0.5 a 10.5 (para NP), e depois pedindo para organizar o data frame por esse índice (com o order). Mas foi forçar a barra né? haha. Depois pensei melhor e pesquisei uns protocolos e decidi que fazia mais sentido intercalar linhas de plantio de P e NP (como no link da figura ilustrativa que criei ai em cima), então meu problema mudou um pouco. Você notou essa mudança?

Sobre indicar o número necessário de P e NP para preencher a área desejada, acho uma boa! Pensei nisso também para eliminar essa complicação extra (além de ter uma justificativa biológica forte né para tanto). Acho que vou seguir a sua sugestão, e depois, se sobrar tempo, tentar acrescentar essa complicação extra (um opcional), pode ser? Quanto a acumular mudas do maior vetor no final, imagino que seja indesejável para o usuário, porque faria as espécies ficarem concentradas numa faixa do plantio, que é justamente o que a função tentaria evitar. Mas, isso já são extrapolações da proposta original, imagino, porque se estiver sobrando mudas de uma espécie ou grupo sucessional, a ordem de plantio para usar todas as mudas deveria ser revista para intercalá-las melhor. Vou adotar o

que você sugeriu então, ok?

Vetor lógico: mantido por hora então hehe. Se sentir que é desnecessário, corto depois.

Tá, acho que um gráfico traz mais informações mesmo. Pensei inicialmente em manter as distâncias entre mudas e entre linhas de plantio conforme manda o protocolo que estou adotando (2m x 3m), o que viria explicitado na página de ajuda da função, então essa informação complementaria a matriz gerada. Se eu conseguir fazer até aqui já está ótimo, e tendo ideias para expandir a função depois

Obrigado pelos valiosos comentários, Lu!

Abraços,

R.

Oie!

Quanto ao comentário “Depois pensei melhor e pesquisei uns protocolos e decidi que fazia mais sentido intercalar linhas de plantio de P e NP (como no link da figura ilustrativa que criei ai em cima), então meu problema mudou um pouco. Você notou essa mudança?”: sim, notei a mudança. Entendo que isso ajuda a resolver a questão de intercalar P e NP na montagem da matriz, mas vc continua com o problema de intercalar espécies de um mesmo grupo em uma mesma linha para deixá-las o mais distante possível, não? Até onde entendi, o “desafio 1” (só que pra intercalação de espécies) continua aí. Neste sentido, acho sim que o order() deve te ajudar bastante, sugiro vc explorar mais essa função (talvez vc encontre uma forma menos “forçação de barra”, mas vc está indo por um caminho legal). Neste caso, vc não precisaria nem transformar o grupo sucessional em vetor lógico, não é?

De fato, acho que se houvesse um número muito maior de P ou de NP, o método de intercalação que vc mostrou na figura não deve ser o mais recomendado, então foge do escopo da sua função mesmo. Concordamos então que a solução mais prática é não colocar o número de mudas de cada espécie como input do modelo! Mas acho legal o que vc falou de, se sobrar tempo, vc colocar a opção do usuário informar o número de mudas que ele tem disponível!

> Sobre colocar o gráfico ou não, vc pode decidir o que acha melhor e mais informativo!
>
> Beijo,
>
> —*Luísa Novara*

> gentem, adorei esta discussão toda aqui. Gostei da adesão à proposta B. Só pra dar um palpite, acho que o gráfico é sempre bom deixar à escolha do usuário e não do Rodolfo 😉
> — *Sara Mortara*

> Oi Sara, obrigado pelo comentário! Dei uma adaptada na sua sugestão e criei um argumento para o usuário escolher se deseja ver a matriz ou não, e o gráfico ficará como saída constante da função por ser mais informativo. 😉
>
> Abraços, R.

Minha função

Função do Rodolfo

From:
http://ecor.ib.usp.br/ - **ecoR**

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

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