

Table of Contents

Preparando dados para análises	1
Guia para análises de definição de grupos	2
Interpretação e validação dos grupos	3
Similaridades entre classificações	3
Como os grupos diferem?	3
Qual o nível de simpatria entre grupos?	3
Os grupos definidos podem conter mais de 1 espécie?	3
Gerando descrições taxonômicas dinamicamente	3

Guia para taxonomia numérica

Este protocolo visa orientar um processo analítico no R de delimitação de espécies através da exploração de dados morfológicos e/ou dados espectrais (NIR) coletados de amostras de herbário.

Preparando dados para análises

Cada conjunto de dados deve ter um IDENTIFICADOR comum entre todas as tabelas. Esse identificador não pode repetir entre linhas nas suas tabelas. Se os dados veem da(s) base(s) do laboratório esse identificador pode ser:

- o WikiEspecimenID (+ NOME DA BASE se os dados foram extraídos de mais de uma base de dados)
- paste(WikiEspecimenID,WikiPlantaID) se os dados forem uma mistura de dados atribuídos a plantas e a espécimes
- uma referência única para cada amostra (coletor+numero+data, geralmente é único).

Esse identificador deve ser o nome das linhas dos seguintes objetos:

1. **dados.morfo.bruto** = DADOS MORFOLÓGICOS COMPLETOS = planilha de dados com 1 linha por amostra e todas as variáveis medidas [nome das linhas é o identificador. Atribuir com rownames(); não deve ter outras colunas nessa tabela], ie. a) variáveis quantitativas; b) variáveis ordenais (ou semiquantitativas) convertidas em números inteiros; c) variáveis de presença e ausência ou verdadeiro e falso (binárias) codificadas com 0 e 1; d) variáveis categóricas não ordenais - estados de variação (categorias) como colunas codificadas como presente (1) ou ausente (0). Células terão valores numéricos, ou números separados por ';':

SCRIPT 00 - Preparando dados para análises - antigo

SCRIPT 001 - Preparando dados para análises - Explicitando similaridades e checando correlações 🚨 - usar este

1. O script acima permite gerar:
 1. **dados.media** (valores médios de dados.morfo.bruto)
 2. **dadosMORFOnmds** - dados em eixos ordenados de NMDS a partir de uma matriz de distância gower gerada com todos as suas amostras para as quais há alguma informação em *dados.media*.
2. **dados.amostras** = 1 linha por amostra e as colunas: taxonomia (genero, especie infraespecie, atual das amostras), grupos definidos *a priori* (sua classificação se houver), latitude e longitude, etc. Identificador como nome de linhas. Uma linha por amostra em dados.completos.
3. **dados.nir** = dados com todas as colunas NIR (X..) e o identificador como nome de linha.
4. **dados.shape** = dados com as PCAs informativas (pela interpretação das imagens produzidas pelo shape) e identificador como nome de linhas
5. **dados.grupos** = para definições *a priori*, se houver, da literatura e/ou se suas observações: 1 linha por grupo e uma coluna para cada variável diagnóstica (amplitude de valores no caso de quantitativas)

Guia para análises de definição de grupos


1. Testando grupos definidos **a priori** (da literatura ou seus)
 1. Tenho hipóteses de grupos (espécies) e de seus limites morfológicos definidos (i.e. tenho dados.grupos).
 1. Testar quantas amostras satisfazem as definições morfológicas dos seus grupos: para cada grupo calcular o número de amostras que satisfazem (a variação se enquadra) em 100% das variáveis, 90% das variáveis, (variação de % dependendo dos resultados. Produzir uma tabela GRUPO vs. COLUNAS DE % DE VARIÁVEIS, com número de amostras em cada célula correspondente. Quantas amostras satisfazem 100% das variáveis de cada grupo?

SCRIPT 01 - Testando Conceitos
 2. Testar via LDA ou SVM se meus dados suportam a existência de meus grupos para as variáveis que definem seus limites. Para isso precisa ter dados.completos e também um coluna com o agrupamento de todas as amostras a serem testadas. Pode usar o teste acima para classificar amostras segundo os conceitos antes de testar a classificação (ver última parte do [script 01](#) para isso). Produzir uma matriz de confusão GRUPOS vs GRUPOS com o número de amostras de cada grupo (linhas) e suas predições (colunas).

SCRIPT 02 - Testando Grupos**SCRIPT 03 - Classificando amostras**
 2. Eu não sei exatamente como meus grupos se separam morfológicamente mas classifiquei minhas amostras em grupos *a priori* de forma intuitiva, então tenho uma coluna no objeto dados.amostras que define essa categoria para cada amostra (ou pelo menos a maioria delas) mas não tenho dados.grupos.
 1. Todas amostras foram pre-classificadas (visualmente)
 1. Testar via LDA ou SVM se os dados (morfológicos e/ou NIR) suportam a existência de meus grupos. Precisa de uma coluna com o agrupamento de todas as amostras (de dados.amostras). Produzir uma matriz de confusão GRUPOS vs GRUPOS com o número de amostras de cada grupo (linhas) e suas predições (colunas).

SCRIPT 02 - Testando Grupos**SCRIPT 03 - Classificando amostras**
 2. Eu não classifiquei todas as minhas amostras nos grupos mas tenho amostras classificadas
 1. Mesma LDA ou SVM que quando tem amostras classificadas, mas apenas com as amostras para as quais tem classe definida *a priori*. Além disso, depois de gerar GRUPOS CORE fazer um modelo com as amostras em GRUPOS CORE (excluindo as incertezas) e classificar as amostras para as quais não há pre-classificação: aquelas preditas com 95% ou mais de probabilidade são incluídas no GRUPO CORE, as demais classificadas como incertezas.

SCRIPT 03 - Classificando amostras
2. Definir grupos **a posteriori**
 1. Qual o número de grupos K que meus dados sugerem? Que amostras definem GRUPOS CORE (?) e quais são incertezas?

SCRIPT 04 - Grupos Morfológicos à posteriori  novo

 1. Pode usar:
 1. apenas dados morfológicos
 1. e depois testar o agrupamento gerado usando NIR de forma independente com [SCRIPT 02 - Testando Grupos](#)

2. ou dados morfológicos + NIR - grupos morfológicos com > predição NIR

Interpretação e validação dos grupos

Similaridades entre classificações

Compare as classificações *a posteriori* (não supervisionada) com as classificações *a priori* (etiquetas ou seus grupos), usando matrizes de confusão.

Compara grupos & Refina classificação final 🚨 novo

Como os grupos diferem?

Como é a distribuição da variação morfológica original entre os grupos (qualquer classificação)?

Grupos vs. Variáveis 🚨 novo

Qual o nível de simpatria entre grupos?

Mapas para grupos com morfologia

Os grupos definidos podem conter mais de 1 espécie?

Estruturação da variação morfológica no espaço

Gerando descrições taxonômicas dinamicamente

Descrições taxonômicas usando Markdown

From:

<https://botanicaamazonica.wiki.br/labotam/> - **Ecologia e Evolução de Plantas Amazônicas**

Permanent link:

<https://botanicaamazonica.wiki.br/labotam/doku.php?id=analises:tax:inicio>

Last update: **19/06/2017 13:06**