

**PROGRAMA DE PÓS-GRADUAÇÃO EM PRODUÇÃO VEGETAL**

Ana Luíza Medrado Monteiro

**PACOTE DO R *MULTIVARIATEANALYSIS* E SUAS APLICAÇÕES NA ANÁLISE  
ESTATÍSTICA DE DADOS**

Montes Claros  
2024

Ana Luíza Medrado Monteiro

**PACOTE DO R *MULTIVARIATE* ANALYSIS E SUAS APLICAÇÕES NA ANALÍSE  
ESTATÍSTICA DE DADOS**

Dissertação apresentada ao Programa de Pós-Graduação em Produção Vegetal da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Produção Vegetal.

Orientador: Alcinei Místico Azevedo.

Montes Claros  
2024

Monteiro, Ana Luiza Medrado.

M772p  
2024

Pacote do R Multivariate Analysis e suas aplicações na análise estatística de dados [manuscrito] / Ana Luiza Medrado Monteiro. Montes Claros, 2024.  
130 f. : il.

Dissertação (mestrado) - Área de concentração em Produção Vegetal. Universidade Federal de Minas Gerais / Instituto de Ciências Agrárias.

Orientador(a): Alcinei Místico Azevedo.

Banca examinadora: Alcinei Místico Azevedo, Anderson Rodrigo da Silva, Moysés Nascimento, Jailson Ramos Magalhães.

Inclui referências: f. 22-26; 110-111; 127-129.

1. Análise multivariada - Teses. 2. Agricultura - Experimentação - Teses 3. Agricultura - Estatística - Teses. I. Azevedo, Alcinei Místico. II. Universidade Federal de Minas Gerais. Instituto de Ciências Agrárias. III. Título.

CDU: 633

## ATA DE DEFESA DE DISSERTAÇÃO

Aos 05 dias do mês de julho do ano de dois mil e vinte e quatro, às 14:10 horas, sob a Presidência do Professor Alcinei Místico Azevedo, D. Sc. (Orientador – UFMG/ICA) e com a participação dos Professores Anderson Rodrigo da Silva, D. Sc. (Instituto Federal Goiano), Moysés Nascimento, D. Sc. (UFV) e Jailson Ramos Magalhães, D. Sc. (UFMG/ICA), reuniu-se, por videoconferência, a Banca de Defesa de Dissertação de **Ana Luiza Medrado Monteiro**, aluna do Curso de Mestrado em Produção Vegetal. Após avaliação da defesa de Dissertação da referida aluna, a Banca Examinadora procedeu à publicação do resultado da defesa de Dissertação intitulada: **“PACOTE DO R *Multivariate Analysis* E SUAS APLICAÇÕES NA ANÁLISE ESTATÍSTICA DE DADOS”** sendo a aluna considerada (aprovada/reprovada) **APROVADA**. E, para constar, eu, Professor Alcinei Místico Azevedo, Presidente da Banca, lavrei a presente ata que depois de lida e aprovada, será assinada por mim e pelos demais membros da Banca examinadora.

OBS.: A aluna somente receberá o título após cumprir as exigências do ARTIGO 65 do regulamento do Curso de Mestrado em Produção Vegetal, conforme apresentado a seguir:

**Art. 65 Para dar andamento ao processo de efetivação do grau obtido, o candidato deverá, após a aprovação de sua Dissertação ou Tese e da realização das modificações propostas pela banca examinadora, se houver, encaminhar à secretaria do Colegiado do Programa, com a anuência do orientador, 1 (um) exemplar impresso e 1 (um) exemplar eletrônico da dissertação ou, tese, no prazo de 60 (sessenta) dias.**

Montes Claros, 05 de julho de 2024.

Documento assinado digitalmente  
 **ALCINEI MISTICO AZEVEDO**  
Data: 10/07/2024 08:45:11-0300  
Verifique em <https://validar.iti.gov.br>

Alcinei Místico Azevedo  
Orientador

Documento assinado digitalmente  
 **JAILSON RAMOS MAGALHAES**  
Data: 10/07/2024 10:31:38-0300  
Verifique em <https://validar.iti.gov.br>

Jailson Ramos Magalhães  
Membro

Documento assinado digitalmente  
 **ANDERSON RODRIGO DA SILVA**  
Data: 10/07/2024 18:13:24-0300  
Verifique em <https://validar.iti.gov.br>

Anderson Rodrigo da Silva  
Membro

Documento assinado digitalmente  
 **MOYSES NASCIMENTO**  
Data: 15/07/2024 10:05:54-0300  
Verifique em <https://validar.iti.gov.br>

Moysés Nascimento  
Membro

## Dedico

A Deus, por estar do meu lado em cada passo; aos meus pais, Lourdes e Marcus, pelo apoio incondicional; a meu irmão, pelo suporte e a todos meus colegas.

## AGRADECIMENTOS

Primeiramente, a Deus, por guiar todos os meus passos e ser minha fortaleza durante essa jornada.

Aos meus pais, Lourdes e Marcus, pelo apoio incondicional durante essa caminhada.

Ao meu irmão, por todo o apoio.

Aos meus avós, Didi Medrado (*in memoriam*), por ser um exemplo de fé e determinação, João Lourenço, por toda sua sabedoria, e Dora, pelo exemplo de mulher.

Aos meus familiares, especialmente a tia Araceli e a Maricelma, pelo apoio desde o primeiro dia, me auxiliando nessa etapa, meu muito obrigada.

Aos meus colegas de Pós-Graduação em Produção Vegetal do ICA, Janete, Janaína, Natália e Murilo, por todo companheirismo, amizade e apoio durante a pesquisa. Com vocês, essa etapa se tornou mais fácil de ser cumprida. À Sandra, que nos momentos mais difíceis da pós-graduação foi como um anjo, minha eterna gratidão.

Ao Professor Alcinei, sempre solícito, pela orientação e pela disponibilidade. Agradeço também pela dedicação a este trabalho e pela confiança depositada em mim, por toda paciência e pelos ensinamentos compartilhados que me possibilitaram chegar ao fim desta etapa.

Aos colegas do grupo de estudos GEEA, em especial a Guilherme, pela grande amizade e pelo apoio.

Aos meus amigos e amigas, Mayara, Victória, Winy, Amanda, João Paulo, César, Karol, Jorge e Pedro Ivo, que me auxiliaram das mais diversas formas, sou muito grata pelo apoio e pela amizade.

À banca examinadora, pela disponibilidade e pelas contribuições.

À Universidade Federal de Minas Gerais, especialmente ao Programa de Pós-Graduação em Produção Vegetal do Instituto de Ciências Agrárias, pela oportunidade e da realização do meu mestrado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

**MEU MUITO OBRIGADA!**

## PACOTE DO R *MULTIVARIATEANALYSIS* E SUAS APLICAÇÕES NA ANÁLISE ESTATÍSTICA DE DADOS

### RESUMO

A análise multivariada consiste em um conjunto de métodos estatísticos utilizados em situações em que diversas variáveis são medidas simultaneamente em cada elemento amostral e pode ser utilizada em pesquisas de diferentes áreas do conhecimento. Nas Ciências Agrárias, é aplicada com frequência em pesquisas associadas a entomologia, fertilidade, nutrição de plantas, plantas medicinais, controle de plantas daninhas, melhoramento genético, olericultura, irrigação, fruticultura, entre outras áreas. Além de possibilitar o estudo dos efeitos dos “tratamentos” (variáveis independentes) considerando diversas características (variáveis dependentes) simultaneamente, a análise multivariada possibilita trabalhar de forma mais simples quando se tem um grande banco de dados, envolvendo muitas variáveis e dados amostrais. Entre os *softwares* utilizados atualmente, destaca-se o R, em sua interface RStudio, por ser gratuito e *open source*. Para a realização de análises multivariadas, como estimativa de medidas de dissimilaridade, análise de componentes principais, dendrogramas, variáveis canônicas, análise discriminante e análise de variância multivariada (MANOVA) há um pacote intitulado “*MultivariateAnalysis*”. Logo, objetivou-se divulgar o pacote R *MultivariateAnalysis* por meio de um artigo e *e-book*, demonstrando suas aplicabilidades em pesquisas associadas à produção vegetal. O pacote está disponível para *download* no link <<https://cran.r-project.org/package=MultivariateAnalysis>>. Esse pacote foi lançado no ano de 2021 pelo pesquisador Alcinei Místico Azevedo e, no ano de 2024, atingiu a marca de 17.325 downloads. Com o *MultivariateAnalysis* é possível a realização de técnicas como medidas de dissimilaridade, agrupamentos hierárquicos, análises de componentes principais, análises coordenadas principais, correlação mantel, dendrograma, análises de variância multivariada, variáveis canônicas, entre outras. O pacote conseguiu reunir várias funções de diversos programas em um único, facilitando a execução de análises multivariadas. O *e-book* constituiu-se de 11 capítulos e exemplifica as funções do pacote *MultivariateAnalysis*.

**Palavras-Chave:** análise multivariada; experimentação agrícola; ciências agrárias; pacote R.

## R MULTIVARIATE ANALYSIS PACKAGE AND ITS APPLICATIONS IN STATISTICAL DATA ANALYSIS

### ABSTRACT

Multivariate analysis comprises a set of statistical methods used when multiple variables are measured simultaneously in each sample element and can be applied across various fields of research. In Agricultural Sciences, it is frequently employed in studies related to entomology, soil fertility, plant nutrition, medicinal plants, weed control, genetic improvement, vegetable production, irrigation, fruit cultivation, among others. Besides enabling the study of the effects of "treatments" (independent variables) by considering multiple characteristics (dependent variables) simultaneously, multivariate analysis simplifies data handling, especially when dealing with large datasets involving many variables and sample data. Among the software currently used, R, with its RStudio interface, stands out for being free and open-source. For performing multivariate analyses, such as dissimilarity measurements, principal component analysis, dendrograms, canonical variables, discriminant analysis, and multivariate analysis of variance (MANOVA), there is a package called "MultivariateAnalysis". This paper aims to disseminate the R MultivariateAnalysis package through an article and an e-book, demonstrating its applications in plant production research. The package is available for download at <https://cran.r-project.org/package=MultivariateAnalysis>. Released in 2021 by researcher Alcinei Místico Azevedo, the package reached 17,325 downloads by 2024. With MultivariateAnalysis, it is possible to perform techniques such as dissimilarity measurements, hierarchical clustering, principal component analyses, principal coordinate analyses, Mantel correlation, dendrograms, multivariate variance analyses, canonical variables, and others. The package consolidates several functions from various programs into one, facilitating the execution of multivariate analyses. The e-book consists of 11 chapters, providing examples of the MultivariateAnalysis package functions.

**Keywords:** multivariate analysis; agricultural experimentation; agricultural sciences; R package.

## LISTA DE ILUSTRAÇÕES

### FIGURAS

<b>Figura 1.</b> Downloads do pacote <i>MultivariateAnalysis</i> .....	21
<b>Figura 2.</b> Interface do software GENES .....	31
<b>Figura 3:</b> Interface do software SAS .....	31
<b>Figura 4.</b> Interface do software SAEG.....	32
<b>Figura 5.</b> Interface do software R.....	33
<b>Figura 6.</b> Interface do software RStudio .....	33
<b>Figura 7.</b> Interface do site CRAN .....	36
<b>Figura 8.</b> Interface do site CRAN para download .....	36
<b>Figura 9.</b> Interface do RStudio .....	37
<b>Figura 10.</b> Início da instalação do <i>MultivariateAnalysis</i> .....	38
<b>Figura 11.</b> Função Install instalação de pacotes .....	38
<b>Figura 12.</b> Resumo das funções de medidas de dissimilaridade .....	110
<b>Figura 13.</b> Resumo das funções de análises de matriz de dissimilaridade .....	110
<b>Figura 14.</b> Resumo das funções de análise de variância multivariada.....	111
<b>Figura 15.</b> Resumo das funções de métodos de dispersão .....	111
<b>Figura 16.</b> Exemplos de resultados gráficos do pacote <i>MultivariateAnalysis</i> : A) Componentes Principais;.....	119
<b>Figura 17.</b> Dendrogramas obtidos no <i>MultivariateAnalysis</i> para dados de dissimilaridade genética de 60 genótipos de batata doce para A) dados moleculares; B) dados morfológicos e C) média ponderada entre as matrizes.....	120
<b>Figura 18.</b> Distâncias inter- e intracluster de dados de 60 genótipos de batata doce para A) dados moleculares; B) dados morfológicos e C) média ponderada entre as matrizes. ....	121
<b>Figura 19.</b> Modelos de medidas de dissimilaridade do pacote <i>MultivariateAnalysis</i> .....	123
<b>Figura 20.</b> Tipos de análises com matriz de dissimilaridade do pacote <i>MultivariateAnalysis</i> .....	124
<b>Figura 21.</b> Resultado de análises de Coordenadas Principais com os dados Iris no pacote <i>MultivariateAnalysis</i> .....	125
<b>Figura 22.</b> Métodos de dispersão realizados pelo <i>MultivariateAnalysis</i> .....	126
<b>Figura 23.</b> Análises de variância multivariada realizadas pelo <i>MultivariateAnalysis</i> .....	127

### TABELAS

<b>Tabela 1</b> Exemplo de comandos no pacote <i>MultivariateAnalysis</i> .....	20
<b>Tabela 2.</b> Teor de proteína e a altura do capim Tifton 85 em função de seis variedades.....	48
<b>Tabela 3.</b> Acessos de Jabuticabeira .....	66
<b>Tabela 4.</b> Agrupamento de 60 genótipos de batata doce ( <i>Ipomoea batatas</i> ) pelo método de Tocher. ....	121

## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO</b>	<b>14</b>
<b>2.</b>	<b>OBJETIVOS</b>	<b>16</b>
2.1.	Objetivo geral	16
2.2.	Objetivos específicos	16
<b>3.</b>	<b>REVISÃO DE LITERATURA</b>	<b>17</b>
3.1.	Software R	17
3.2.	Análise Multivariada	18
3.3.	Pacote MultivariateAnalysis	20
3.4.	Análise Multivariada aplicada a produção vegetal	21
<b>4.</b>	<b>REFERÊNCIAS</b>	<b>23</b>
<b>5.</b>	<b>INTRODUÇÃO</b>	<b>29</b>
5.1.	Análise multivariada e sua importância nas ciências agrárias	29
5.2.	Ferramentas computacionais para a análise multivariada	30
5.3.	Software R para análises multivariadas	33
5.4.	Package MultivariateAnalysis	34
<b>6.</b>	<b>SOFTWARE R E RSTUDIO</b>	<b>35</b>
6.1.	Instalação do R	35
6.2.	Instalação do RStudio	36
6.3.	Interface Rstudio	37
6.4.	Instalação do MultivariateAnalysis	37
<b>7.</b>	<b>MEDIDAS DE DISTÂNCIAS</b>	<b>39</b>
7.1.	Medidas para dados quantitativos	39
7.1.1	Distância Euclidiana	40
7.2.	Distância de Mahalanobis	41
7.3.	Distâncias de Cole Rodgers	42
7.4.	Distâncias para dados binários	43
7.5.	Dados Mistos	45
7.6.	Visualizações das matrizes de dissimilaridade	46
<b>8.</b>	<b>DENDROGRAMA</b>	<b>47</b>
8.1.	Algoritmo	47
8.2.	Análise em R	49
8.3.	Função HeatPlot	51
8.4.	Correlação cofenética	52
8.5.	Critério de Corte	53
8.5.1	Pontos de corte pelo método visual	53
8.5.2	Pontos de corte – Critério Mojena	54
<b>9.</b>	<b>MÉTODOS DE AGRUPAMENTO POR OTIMIZAÇÃO</b>	<b>55</b>

9.1.	Agrupamento Tocher.....	55
9.2.	Agrupamento Tocher Modificado.....	55
9.3.	Análise em R.....	55
9.3.1	Distância Intra e Intercluster no agrupamento Tocher.....	57
9.3.1.1	Distância Intercluster.....	57
9.3.1.2	Distância Intracluster.....	57
9.4.	Análise em R.....	57
9.5.	Método K-means.....	57
9.6.	Análises em R.....	57
10.	<b>COMPONENTES PRINCIPAIS.....</b>	<b>65</b>
10.1.	Algoritmo.....	66
11.	<b>COORDENADAS PRINCIPAIS.....</b>	<b>82</b>
11.1.	Análise em R.....	83
12.	<b>MANOVA – Análise de variância Multivariada.....</b>	<b>86</b>
12.1.	Delineamento inteiramente casualizado.....	86
12.2.	Delineamento em Blocos Casualizados.....	87
12.3.	Delineamento em quadrado latino.....	88
12.4.	Esquema Fatorial.....	89
13.	<b>VARIÁVEIS DISCRIMINANTES CANÔNICAS.....</b>	<b>94</b>
13.1.	Análise em R.....	95
13.1.1	Variáveis Discriminantes em DIC.....	95
13.1.2	Variáveis Discriminantes em DBC.....	96
13.1.3	Variáveis Discriminantes em DQL.....	97
13.1.4	Variáveis Discriminantes em fatorial duplo DIC.....	98
13.1.5	Variáveis Discriminantes em fatorial duplo DBC.....	101
13.1.6	Variáveis Discriminantes em fatorial duplo DIC.....	105
13.1.7	Variáveis Discriminantes em fatorial duplo DBC.....	107
14.	<b>RESUMO DAS PRINCIPAIS FUNÇÕES.....</b>	<b>110</b>
15.	<b>REFERÊNCIAS.....</b>	<b>112</b>
16.	<b>ARTIGO.....</b>	<b>114</b>
17.	<b>INTRODUÇÃO.....</b>	<b>117</b>
18.	<b>O PACOTE MULTIVARIATEANALYSIS.....</b>	<b>118</b>
19.	<b>VERIFICAÇÃO DOS DADOS.....</b>	<b>120</b>
20.	<b>MEDIDAS DE DISSIMILARIDADE.....</b>	<b>122</b>
21.	<b>ANÁLISE COM MATRIZ DE DISSIMILARIDADE.....</b>	<b>123</b>
22.	<b>MÉTODOS DE DISPERSÃO.....</b>	<b>125</b>
23.	<b>ANÁLISE DE VARIÂNCIA MULTIVARIADA.....</b>	<b>126</b>
24.	<b>CONSIDERAÇÕES FINAIS E MELHORIAS FUTURAS.....</b>	<b>127</b>

25.	REFERÊNCIAS .....	129
-----	-------------------	-----

## 1. INTRODUÇÃO

A análise multivariada consiste em um conjunto de métodos estatísticos utilizados em situações em que diversas variáveis são medidas simultaneamente em cada elemento amostral. Na maioria das vezes, as variáveis são correlacionadas entre si e quanto maior o número de variáveis, mais complexa torna-se a análise por métodos comuns. Em geral, os métodos de análise multivariada são utilizados com o propósito de simplificar ou facilitar a interpretação do fenômeno que está sendo estudado. Em essência, a análise multivariada possui dois grupos, um primeiro consistindo em técnicas exploratórias de simplificação da estrutura de variabilidade dos dados, e um segundo, consistindo em técnicas de inferência (HAIR *et al.*, 2004).

A aplicação da análise multivariada se estende a várias áreas do conhecimento e nas Ciências Agrárias é bastante utilizada em áreas como a entomologia (BARIONI JUNIOR *et al.*, 2015), a adubação (HACKENHAAR *et al.*, 2022), em plantas medicinais (GREGORIO; OLIVEIRA, 2021), aplicação em plantas daninhas (MIRANDA *et al.*, 2020), em genética de plantas (ALVES, 2021), em olericultura e irrigação (SANTOS *et al.*, 2020) e na fruticultura (SANTOS *et al.*, 2020).

Entre os *softwares* utilizados na atualidade, destaca-se o R, caracterizado por ser um *software* livre, assim como o RStudio, que tem por essência ser um projeto *open source* (HENNING *et al.*, 2013). O *software* R, desenvolvido por Ross Ihaka e Robert Gentleman em 1992, tem por objetivo fornecer um ambiente estatístico baseado na linguagem S, com código aberto e gratuito. Baseado em linguagem de programação, esse *software* fornece inúmeras vantagens em sua utilização, sendo uma delas o fato de a linguagem ser bastante intuitiva e ser possível a realização de atividade em lotes, o que facilita a análise dos dados.

O ambiente de desenvolvimento integrado RStudio é um *software* para a linguagem de programação destinado a combinar os componentes do R em um ambiente de trabalho produtivo e contínuo (ALLAIRE, 2012). Esses pacotes servem como uma espécie de biblioteca dentro do RStudio e estão disponíveis na plataforma do R, conhecida como CRAN (<https://cran.r-project.org/>). Esses pacotes R são reuniões de funções e dados que organizam em um local algumas funções extras, expandindo a linguagem do R, e são compostos obrigatoriamente por dois diretórios e dois arquivos de meta dados.

O repositório CRAN é a fonte principal dos pacotes do R, mas podem ser encontrados pacotes em outros repositórios, tais como BioConductor, R-Forge, rOpenSci e repositórios individuais que podem ser encontrados em plataformas, como GitHub.

Existe uma diversidade de pacotes disponíveis dentro do repositório CRAN, em muitas áreas do conhecimento. Quando instalado, o R acompanha um conjunto de pacotes básicos para a execução das atividades. Para as atividades específicas, é necessária a utilização de pacotes. Dessa forma, o R e o RStudio tornam-se ferramentas tecnológicas úteis e bastante aplicáveis para realização de análises estatísticas. Dentro do repositório CRAN, existe uma seção de visualizações de tarefas (*task views*) de pacotes disponíveis, separadas por tópicos, fornecendo uma visão global das funções. Existe uma diversidade de pacotes disponíveis para operações matemáticas, análise de variância, testes estatísticos, análises de sobrevivência e também análises multivariadas.

Para a realização de análise multivariada, como medida de dissimilaridade, análise de componentes principais, realização de dendograma e análise de variância multivariada (MANOVA), há um pacote intitulado *MultivariateAnalysis*, desenvolvido no ano de 2021, disponível no site da iniciativa do software R (<https://cran.r-project.org/>), na aba correspondente a pacotes.

O pacote *MultivariateAnalysis* é caracterizado por ser um pacote livre, para ser executado no RStudio. Dentro do pacote, é possível a realização de vários tipos de análises estatísticas, através da linguagem de programação R. Todo o conteúdo dessa dissertação terá como foco a apresentação de funcionalidades desse pacote, com aplicações associadas à produção vegetal.

## 2. OBJETIVOS

### **Objetivo geral**

Apresentar o pacote do R *MultivariateAnalysis*, por meio de um artigo e um *e-book*, demonstrando suas aplicabilidades em pesquisas associadas à produção vegetal.

### **Objetivos específicos**

Divulgar o pacote *MultivariateAnalysis*, por meio de um *e-book* e um artigo científico, demonstrar suas principais funcionalidades e apresentar suas aplicabilidades nas análises de pesquisas associadas à produção vegetal.

### 3. REVISÃO DE LITERATURA

#### **Software R**

Entre as consequências do processo de aceleração da humanidade como um todo, está a utilização de *softwares* que têm por base o desenvolvimento tecnológico e otimizam a produção. Nas últimas décadas, a utilização de *softwares* se tornou um facilitador na rotina das pessoas e na Estatística não seria diferente, com a utilização de programas computacionais e *softwares* estatísticos como ferramentas importantes para realização de análises estatísticas, promovendo avanços e melhor utilização do tempo. Operações que anteriormente eram impossíveis de serem realizadas manualmente são viabilizadas com a utilização de *softwares* e programas computacionais.

De acordo com Martins e Theóphilo (2009), nas últimas duas décadas, a área de Estatística sofreu grandes mudanças, graças ao desenvolvimento de *softwares* que possibilitam a realização de análises dessa natureza. A lista de programas computacionais estatísticos disponíveis aumenta constantemente e tende a continuar em expansão, pois a cada dia estão sendo desenvolvidos novos sistemas com maiores aplicações, por exemplo, com tratamentos de dados e aplicações gráficas.

Diversas são as opções de *softwares* estatísticos disponíveis para a realização de análises de experimentos na área das Ciências Agrárias. Por exemplo, Sisvar (FERREIRA, 2019), *Statistical Analysis System - SAS (Statistical Analysis System, 1995)*, SAEG (FUNARBE, 2007), GENES (CRUZ, 2006), R (R Core Team, 2016), cada um com características particulares, cabendo ao pesquisador determinar qual melhor se adapta às necessidades da pesquisa.

O R é um *software* livre para análise de dados. Foi desenvolvido em 1996 pelos professores de estatística Ross Ihaka e Robert Gentleman, da Universidade de Auckland, que criaram uma nova linguagem computacional, similar à linguagem S desenvolvida por John Chambers (SOUZA; PETERNELLI; MELLO, 2014).

Pode-se obter gratuitamente o *software* R, que se encontra disponível na plataforma *Comprehensive R Archive Network* (CRAN), sendo essa um núcleo da comunidade R e um repositório de materiais relacionados ao R (CALENGE, 2006), disponível no *site* <<http://cran.r-project.org>>. O *software* R é apresentado em versões compatíveis com os sistemas operacionais LINUX, Windows ou Macintosh. Além disso, encontra-se nesse *site* mais informações sobre a sua utilização e uma central de correspondências em que profissionais de várias áreas do conhecimento podem contribuir na implementação de novos recursos, assim como responder a dúvidas dos demais usuários.

O R é caracterizado por ser um *software* gratuito com código aberto (HENNING *et al.*, 2013), com uma linguagem acessível, de expansão exponencial entre pesquisadores, engenheiros e estatísticos, o qual se reinventa constantemente através de novas aplicações. Em meados no mês de junho de 2024, o *site* oficial do R (<<https://cran.r-project.org/>>) apresentava 20.935 pacotes disponíveis para *download*. Por apresentar uma linguagem amigável, ele é perfeitamente adequado para o uso entre os acadêmicos (SILVA, 2018).

Segundo Nedel (2019), pode-se definir pacotes no R como união de funções com a finalidade de executar comandos. De acordo com Souza *et al.* (2012), a utilização de pacotes estatísticos livres para a realização de análises de dados é importante no que se refere à interpretação desses dados. Alguns pacotes são para uso geral, como é o caso do pacote *multcomp* (HOTHORN; BRETZ; WESTFALL, 2008), e outros para uso específico em várias áreas do conhecimento.

O uso de pacotes estatísticos para o tratamento de dados é de grande importância para a análise e a interpretação de resultados. Contudo, observa-se que sua aquisição apresenta um custo relativamente elevado, o que dificulta sua utilização pelo pesquisador. Dentre os *softwares* de domínio público, livres, que podem ser utilizados para análise de dados em geral, encontra-se o Ambiente R, ou simplesmente R, conforme usualmente chamado pelos seus usuários, que apresenta código fonte aberto, podendo ser modificado ou implementado com novos procedimentos desenvolvidos por qualquer usuário, a qualquer momento. Pode-se destacar o *software* R como um instrumento de grande importância tecnológica para análise de dados com ampla aplicação na estatística (ARNHOLD, 2014).

O R torna-se, portanto, uma importante ferramenta na análise e na manipulação de dados, com testes paramétricos e não paramétricos, modelagem linear e não linear, análise de séries temporais, análise de sobrevivência, simulação e estatística espacial, entre outros recursos, além de apresentar facilidade na elaboração de diversos tipos de gráficos, sobre os quais o usuário tem pleno controle.

### **Análise Multivariada**

A análise multivariada consiste em um conjunto de métodos estatísticos utilizados em situações em que várias variáveis são medidas simultaneamente em cada elemento amostral. Na maioria das vezes, as variáveis são correlacionadas entre si e quanto maior o número de variáveis, mais complexa torna-se a análise por métodos comuns.

Para Hocheim (2001), as análises multivariadas de dados possibilitam estudar as evidências, as semelhanças, as diferenças e as ligações entre as variáveis envolvidas no processo. Já para Steiner (1995), a análise multivariada vem da necessidade de entender as

relações entre as diversas variáveis aleatórias, e dessa maneira essa metodologia se apresenta com um grande potencial de uso.

Dentro da fundamentação teórica dos métodos de análise multivariada, existem cinco categorias baseadas nos objetivos. Johnson e Wichern (1998) propõem uma classificação composta por: investigação das dependências entre as variáveis; predição; construção de hipóteses e testes; redução dos dados ou simplificação estrutural; e agrupamento de objetos ou variáveis. Algumas técnicas podem ser classificadas em mais de uma classe.

Para a escolha da técnica que melhor atenda ao objetivo da pesquisa, deve-se considerar, além do objetivo, o tipo de relação examinada, o número de variáveis dependentes e o tipo de escala utilizada (GOUVEA *et al.*, 2011). Como todas as técnicas estatísticas, a análise multivariada apresenta suposições intrínsecas, estatísticas e matemáticas, que, quando não respeitadas, podem causar efeitos negativos aos resultados da análise. Além desse efeito, essas suposições fazem parte da construção teórica da técnica, ou seja, quando não respeitadas, não é possível assegurar que o algoritmo realmente represente o resultado almejado (PREARO *et al.*, 2011).

Existem sete principais premissas que devem ser respeitadas quando se objetiva trabalhar com técnicas de análise multivariada, sendo elas: sensibilidade ao tamanho da amostra; influência de *outliers*; linearidade das relações; normalidade multivariada; multicolinearidade; homoscedasticidade; e autocorrelação dos resíduos (PREARO *et al.*, 2011). Cada uma dessas, com suas características e alguns testes que comprovem sua significância.

Em geral, os métodos de análise multivariada são utilizados com o propósito de simplificar ou facilitar a interpretação do fenômeno que está sendo estudado (ANDRADE *et al.*, 2007). Basicamente, a análise multivariada se divide em dois grupos: um primeiro consistindo em técnicas exploratórias de simplificação da estrutura de variabilidade dos dados, em uma tentativa de sintetizar as variáveis, e um segundo, consistindo em técnicas de inferência (HAIR *et al.*, 2004).

As técnicas de análise multivariada sofrem das mesmas limitações de todas as estatísticas (MOITA NETO, 2004) e podem ser consideradas mais complexas que as análises estatísticas univariadas (GOUVEA *et al.*, 2011). Apesar dessas considerações, essas técnicas podem ser realizadas com o apoio de *softwares* e sem o auxílio de um especialista em estatística. O uso de *softwares* estatísticos para análises multivariadas transforma os trabalhos de realização de cálculos, geralmente cansativos e trabalhosos, em uma atividade mais simples, devido à sua interface amigável e intuitiva (MOITA NETO, 2004).

## Pacote *MultivariateAnalysis*

Para a realização de análises multivariadas como estimativa de medidas de dissimilaridade, análise de componentes principais, dendrogramas, variáveis canônicas, análise discriminante e análise de variância multivariada (MANOVA), há um pacote intitulado *MultivariateAnalysis*, desenvolvido em 2021 pelo pesquisador Alcinei Místico Azevedo, docente da Universidade Federal de Minas Gerais. O pacote *MultivariateAnalysis* foi concebido para facilitar a realização de análise multivariada dentro do *software* R, pode ser executado dentro do RStudio e está disponível na plataforma CRAN (<<https://cran.r-project.org/>>).

Esse pacote reúne diversas funções que estão disponíveis de forma separada em um mesmo pacote. Por exemplo, a função estimativas de medidas de dissimilaridades, pacote alternativo TSclust (MONTERO *et al.*, 2014); realização de agrupamentos, como os pacotes smacof (DE LEEUW; MAIR, 2013) ape (PARADIS, 2004) e ade4 (DRAY; DUFOUR, 2007); execução de correlações canônicas, como o pacote metan (OLIVOTO, 2020) e o pacote car (FOX; WEISBERG, 2011); dendrograma, como o pacote factoextra (KASSAMBARA; MUNDT, 2017); análises de agrupamentos, como o pacote biotools (DA SILVA, 2017); componentes principais, como os pacotes FactoMineR (HUSSON *et al.*, 2016), ade4 (DRAY; DUFOUR, 2014) e amap (LUCAS, 2014); MANOVA, como o pacote car (FOX; WEISBERG, 2011); e correlação canônicas, como os pacotes CCA (GONZÁLEZ; DÉJAN, 2013) e yacca (BUTTS, 2012). Esse pacote realiza a análise de experimentos por abordagem multivariada e a integração dessas diversas funções facilita o trabalho do pesquisador.

O '*MultivariateAnalysis*' é um pacote livre para a realização de vários tipos de análises estatísticas multivariadas para experimentos com e sem delineamento estatístico. Ele se destaca pela sua facilidade de utilização, dependendo de poucos comandos, e, conseqüentemente, sendo útil para usuários iniciantes. Em 2024, o pacote se encontra na sua versão 5.0, constando com 41 funções presentes. Assim como todos os pacotes do R, o *MultivariateAnalysis* é orientado por linhas de comando, dentro das quais são inseridos argumentos para a execução das funções.

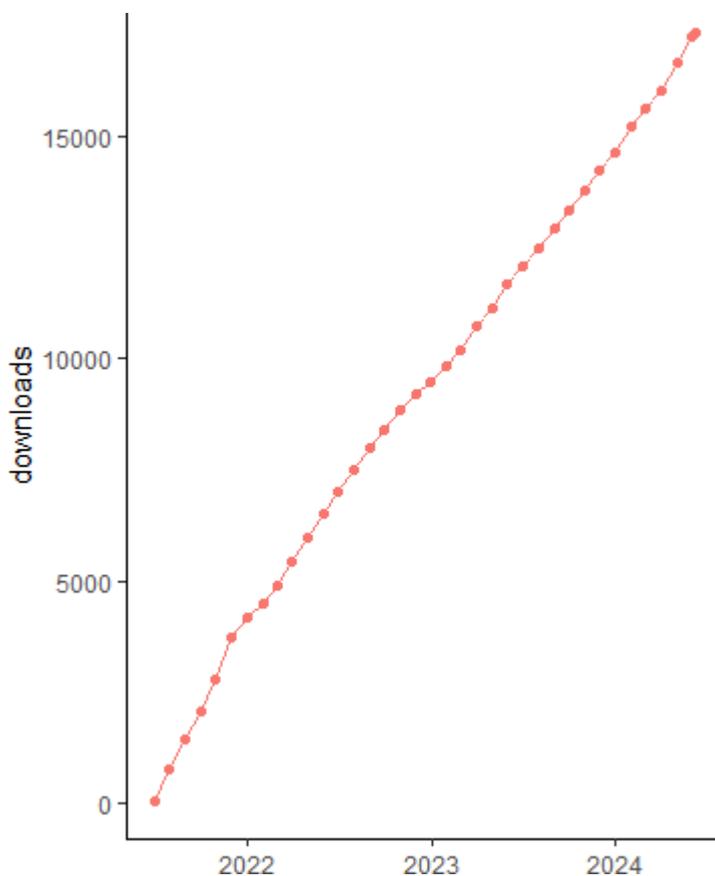
**Tabela 1** Exemplo de comandos no pacote *MultivariateAnalysis*

<b>Funções</b>	<b>Comandos</b>
Agrupamento Kmeans	Kmeans () ou Kmeans_NumeroOtimo ()
Agrupamento Tocher	Tocher()
Componentes Principais	ComponentesPrincipais()
Contribuição Relativa	ContribuicaoRelativa ()
Coordenadas Principais	CoordenadasPrincipais()
Correlação Mantel	CorrelacaoMantel()

Dendrograma	Dendrograma()
Manova	MANOVA()
Medidas de distância	Distancia()
Variáveis Canônicas	VariaveisCanonicas()

Um dado importante sobre o *MultivariateAnalysis* está no crescente número de *downloads* realizados entre o ano do lançamento – 2021 – e 2024, totalizando em junho de 2024 17.325 *downloads* realizados, como mostra a Figura 1.

**Figura 1.** *Downloads* do pacote *MultivariateAnalysis*



### **Análise Multivariada aplicada a produção vegetal**

Assim como em muitas áreas do conhecimento a análise multivariada é aplicada, na produção vegetal não seria diferente. Essas técnicas de análise podem estar presentes nas mais diversas áreas das Ciências Agrárias. De acordo com Chaves *et al.* (2020), a utilização de métodos estatísticos com destaque para componentes principais e análise hierárquica de agrupamentos tem sido bastante recorrente nos últimos anos.

Considerando sua ampla aplicabilidade, as técnicas de análises multivariadas conseguem abranger pesquisas relacionadas às grandes culturas, como é o caso da soja (LOURES *et al.*, 2022), da seleção de genótipos de feijão (CARBONARI *et al.*, 2023) e da avaliação de nutrientes na cultura do milho (ANDRADE *et al.*, 2023). Também encontramos sua aplicação em estudos de consórcio milho-girassol (SOUZA *et al.*, 2020), e na cultura do algodão, estudando mecanismos de resistência de genótipos (PEREIRA, 2022).

Ainda sobre seleção de cultivares, tem-se relato de técnicas de análise multivariada utilizadas na seleção de novas cultivares de pêssego com adaptações a diferentes regiões (SOBIERAJSKI *et al.*, 2024). Já sobre o tema de fruticultura, Firetti *et al.* (2024), em estudo com análise multivariada do crescimento vegetativo e produtividade em bananeiras, aplicaram algumas técnicas que se mostraram uma boa opção para esse tipo de pesquisa. Também Góes *et al.* (2020) relatam que a utilização de métodos estatísticos multivariados para avaliar a disponibilidade de nutrientes no solo tem se mostrado uma ferramenta eficaz para a tomada de decisão pelo produtor. Da mesma maneira, a utilização combinada de análise multivariada e agrupamento hierárquico possibilitou o estudo quantitativo de progênies, tornando-se assim uma ferramenta precisa e eficaz (YOKOMIZO *et al.*, 2022).

#### 4. REFERÊNCIAS

ALLAIRE, J. **RStudio**: integrated development environment for R. Boston, MA, v. 770, n. 394, p. 165-171, 2012.

ALVES, F. A. L. Utilização de técnicas de análise multivariadas no estudo da diversidade genética em quixabeira (*Sideroxylon obtusifolium*). **Pesquisa Agropecuária Pernambucana**, v. 26, n. 1, 2021.

ANDRADE, E. M. D. *et al.* Seleção dos indicadores da qualidade das águas superficiais pelo emprego da análise multivariada. *Engenharia agrícola*, 27, 683-690. 2007.

ANDRADE, M. F. *et al.* Análise multivariada do efeito de diferentes doses de zinco no tratamento de sementes de milho, cv. Chico Modesto. **CONTRIBUCIONES A LAS CIENCIAS SOCIALES**, [S. l.], v. 16, n. 11, p. 27836–27850, 2023. DOI: 10.55905/revconv.16n.11-184. Disponível em: <https://ojs.revistacontribuciones.com/ojs/index.php/clcs/article/view/3296>. Acesso em: 4 jun. 2024.

ARNHOLD, E. Pacote em ambiente R para automatizar estatísticas descritivas. **Sigmae**, v. 3, n. 1, p. 36-42, 2014.

BARIONI JUNIOR, W. G. *et al.* **Aplicação de técnicas multivariadas na seleção de genótipos de plantas para gramados resistentes a insetos**. Brasília-DF: EMBRAPA, 2015.

BUTTS, Carter T. Yacca: Yet another canonical correlation analysis package. **R package version**, v. 1, 2012.

CALENGE, Clément. The package “adehabitat” for the R software: a tool for the analysis of space and habitat use by animals. **Ecological modelling**, v. 197, n. 3-4, p. 516-519, 2006.

CARBONARI, Luan Tiago *et al.* Análise multivariada aplicada na discriminação de genótipos em caracteres do tempo de cozimento em feijão (*Phaseolus vulgaris* L.). **Revista de Ciências Agroveterinárias**, v. 22, n. 3, p. 358-366, 2023.

CHAVES, Henrique Sousa *et al.* Aplicação estatística multivariada para a avaliação físico-química na qualidade da água subterrânea na cidade de Parauapebas (Sudeste do Estado do Pará). **Revista Ibero-Americana de Ciências Ambientais**, v. 11, n. 5, p. 261-272, 2020.

CRUZ, C. D. **Programa Genes - Biometria**. 1st ed. Viçosa: UFV, 2006

DRAY, S. DUFOUR, A. B. The ade4 package: implementing the duality diagram for ecologists. **Journal of statistical software**, v. 22, p. 1-20, 2007.

DRAY, Stéphane; DUFOUR, Anne B.; THIOLOUSE, J. ade4. **CRAN| L”(K)**, 2014.

FERREIRA, D. F. SISVAR: A COMPUTER ANALYSIS SYSTEM TO FIXED EFFECTS SPLIT PLOT TYPE DESIGNS. REVISTA BRASILEIRA DE BIOMETRIA, [S.l.], v. 37, n. 4, p. 529-535, dec. 2019. ISSN 1983-0823. Available at: <<http://www.biometria.ufla.br/index.php/BBJ/article/view/450>>. Data de acesso: 14 jul. 2022.

FIRETTI, R. *et al.* Multivariate analysis of vegetative growth and productivity in cultivars of the Cavendish subgroup of the banana. **Revista Ciência Agrônômica**, v. 55, p. e20228571, 2023.

FOX, J. *et al.* Package ‘car’. **Vienna: R Foundation for Statistical Computing**, v. 16, n. 332, p. 333, 2012.

FUNARBE. **SAEG – Sistema para análises estatísticas – versão 5.0**. Viçosa: Fundação Arthur Bernardes. 2007. 80p

GÓES, B. C. *et al.* Multivariate analysis of soy cultivated under no-tilling systems in plant covers of millet and sorghum. **Revista Brasileira de Engenharia de Biosistemas**, Tup, v. 14, n. 3, p. 283–298, 2020.

GONZÁLEZ, I.; DEJÉAN, S. (2012) CCA package version 1.2. Available online at [http:// www.lsp.ups-tlse.fr/CCA](http://www.lsp.ups-tlse.fr/CCA).

GOUVEA, M. A. *et al.* **Avaliação do emprego da técnica de análise multivariada de variância em teses e dissertações dos programas de pós-graduação em administração da Universidade de São Paulo e da Universidade Federal do Grande ABC**. 2011.

GREGORIO, W.; OLIVEIRA, V. B. Análise quimiométrica de infusões medicinais utilizadas popularmente. **Scientific Electronic Archives**, v. 14, n. 6, 2021.

HACKENHAAR, C. *et al.* Análise multivariada em diferentes épocas de semeadura e doses de potássio em soja. **Revista em Agronegócio e Meio Ambiente**, v. 15, n. 2, p. 1-12, 2022.

HAIR, J. F. *et al.* **Análise multivariada de dados**. São Paulo: Bookman editora. 2009.

HENNING, E. *et al.* RStudio como suporte no ensino de planejamento de experimentos. In **Congresso Brasileiro de Educação em Engenharia-COBENGE**, Gramado, 2013.

HOSKULDSSON, A.; **J. Chemom**, v. 2, 211, 1988.

HOTHORN, T.. *et al.* Simultaneous Inference in General Parametric Models. **Biometrical Journal**, v.50, p.346-363, 2008.

HUSSON, Francois *et al.* Package ‘factominer’. **An R package**, v. 96, n. 96, p. 698, 2016.

JOHNSON, R. A. e WICHERN, D. W. **Applied multivariate statistical analysis**. Fourth Edition. New Jersey. Printice Hall – 1998.

KASSAMBARA, A. MUNDT, F. Package ‘factoextra’. **Extract and visualize the results of multivariate data analyses**, v. 76, n. 2, 2017.

LEAO, P. C. S.; CARVALHO, J. N.. **Diversidade genética entre progênies de uvas sem sementes no Vale do São Francisco**. 2023.

LEEuw, J. *et al.* SMACOF for Multidimensional Scaling. **Maintainer Jan. Package ‘smacof’**. 2013.

LOURES, E. M. S. *et al.* **Aplicação da análise multivariada para a produção de soja brasileira**. Brasília-DF: EMBRAPA. 2022.

LUCAS, A.. **Amap**: Another multidimensional analysis package. [http://CRAN.R-project.org/package= amap](http://CRAN.R-project.org/package=amap), 2014.

MARTINS, Gilberto de Andrade; THEÓPHILO, Carlos Renato. **Metodo da investigação científica para ciências sociais aplicadas**. 2. ed. São Paulo: Atlas, 2009.

MIRANDA, D. A. *et al.* Estudo de seleção da comunidade infestante por herbicidas utilizando técnicas de análise multivariada. **Revista Brasileira de Herbicidas**, v. 19, n. 2, p. 688-1, 2020.

MOITA NETO, J. M. **Análise multivariada: uma visão didática metodológica**. **Crítica**, p. 1-13, 2004.

MOITINHO, A. C. R. **Estimativa de parâmetros genéticos e análise multivariada na seleção de genótipos de soja oriundos do cruzamento de genitores tipo grão x tipo alimento**. Dissertação (Mestrado em Agronomia) - Unesp, São Paulo, 2021.

MONTERO, P.; VILAR, J. A. TSclust: An R Package for Time Series Clustering. **Journal of Statistical Software**, [S. l.], v. 62, n. 1, p. 1–43, 2014.

**Multivariados com Resposta Dicotômica**. 1995. Tese (Doutorado em Engenharia),

NEDEL, F. B. Pacote csapAIH: a lista brasileira de internações por condições sensíveis à atenção primária no programa. **Epidemiologia e Serviços de Saúde**, v. 28, p. 2019.

NÓBREGA, V. G. **Perfil gerencial de propriedades da agricultura familiar da região metropolitana de Ribeirão Preto/SP: análise multivariada e qualitativa sobre a eficiência de gestão**. 2017. Tese (Doutorado) - Universidade de São Paulo, São Paulo, 2017.

OLIVOTO, T.; LÚCIO, A. D. metan: An R package for multi-environment trial analysis. **Methods in Ecology and Evolution**, v. 11, n. 6, p. 783-789, 2020.

PARADIS, E. *et al.* APE: analyses of phylogenetics and evolution in R language. **Bioinformatics**, v. 20, n. 2, p. 289-290, 2004.

PEREIRA, E. T. V. **Estudos dos mecanismos de resistência de genótipos de algodão submetidos à diferentes condições bióticas e abióticas.** Tese (Doutorado) – UFSCAR, São Carlos – SP, 2022.

PREARO, L. C. *et al.* Avaliação do emprego da técnica de análise fatorial em teses e dissertações de algumas instituições de ensino superior. **REGE-Revista de Gestão**, v. 18, n. 4, p. 621-637, 2011. R Core Team. 2016.

ROSA, Thalles R. *et al.* Redgim como aplicativo de smartphone para aplicações quimiométricas por meio de análise de imagens: um uso em pls. **Química Nova**, v. 45, n. 05, p. 550-559, 2022.

SANTOS G. B. *et al.* Tipificação de produtores por nível de inovação: análise na fruticultura do vale do São Francisco. **Revista de Economia e Agronegócio**, v. 18, n. 3, p. 1-21, 2020.

SANTOS, M. A. L. *et al.* Desempenho agrônomo e análise multivariada na produção da cebolinha verde em resposta a lâminas de irrigação e níveis de adubação sintética. **Research, Society and Development**, v. 9, n. 12, 2020.

SAS Institute. **INC.SAS/STAT<sup>TM</sup> SAS user's guide for windows environment.** 6.11 ed. Cary : SAS Institute, 1995.

Science. **Package 'biotools'**, v. 12. Available online at: <https://CRAN.R-project.org/package=~biotools>, 2017.

SILVA M. T. *et al.* Geotecnologias e análise multivariada para investigação da tipologia do uso agropecuário do estado de Mato Grosso. **Geo UERJ**, v. 33, 2018.

SILVA, A. R.; SILVA, M. A. R. Tools for Biometry and Applied Statistics in Agricultural

SILVA, H. A. *et al.* **A mudança na atitude frente à estatística em graduandos de Medicina em um curso introdutório de estatística utilizando o RSTUDIO.** Dissertação (Mestrado) – Unfenas, Belo Horizonte, 2018.

SILVA, J. A. M. **Análise multivariada aplicada a projeto de desenvolvimento da produção agroecológica no DF.** 2023 36 f., il. Trabalho de Conclusão de Curso (Bacharelado em Estatística) — Universidade de Brasília, Brasília, 2023.

SOBIERAJSKI, G. A. *et al.* Cultivares e novas seleções IAC de pêssigo para regiões de inverno ameno. **Revista Brasileira de Fruticultura**, v. 46, p. e-025, 2024.

SOUZA, D. C. *et al.* Agronomic variability between commercial and experimental carrot genotypes with emphasis on multivariate analysis. **Research, Society and Development**, [S. l.], v. 10, n. 13, 2021.

SOUZA, E. F. M. *et al.* **Software Livre R**: aplicação estatística. 2014.

SOUZA, J. W. S. *et al.* Análise multivariada dos atributos químicos de um argissolo sob sistema rotacional (girassol/milho) submetidos à diferentes níveis de salinidade e nitrogênio. **Agronomia: jornadas científicas**, v. 1, n. 1, p. 118-129, 2020.

STEINER, M. T. A. **Uma Metodologia para o Reconhecimento de Padrões**  
UFSC, Florianópolis. Disponível em  
<<https://repositorio.ufsc.br/xmlui/handle/123456789/157952>>. Acesso em 02/06/2024.

YOKOMIZO, G. K. I. *et al.* Eficiência do uso simultâneo entre dendrograma e AMMI na seleção de progênies de açaizeiro, Yokomizo et al Científic@ Multidisciplinary Journal, v. 9, n. 1, 2022.

***Ebook: Aprendendo análise multivariada no R com `MultivariateAnalysis`***

## 5. INTRODUÇÃO

Neste capítulo será destacado o que é a análise multivariada, sua importância nas Ciências Agrárias e os principais *softwares* utilizados para esse fim. Além disso, será feita uma descrição básica do pacote *MultivariateAnalysis* – pacote do R destinado à realização dos principais métodos de análise multivariada. Vale ressaltar que, para a execução desse pacote, é necessária a instalação prévia do *software* R, bem como ter conhecimento de seus principais comandos.

Denomina-se análise multivariada um conjunto de métodos estatísticos que analisam diversas variáveis de forma simultânea. No cotidiano, é bastante comum que, para uma tomada de decisão, haja diversas variáveis respostas que devem ser analisadas de forma conjunta. Dentre as vantagens de se utilizar a abordagem multivariada em comparação com a univariada, destacam-se a obtenção de tomadas de decisão mais robustas, a compreensão da associação existente entre as variáveis dependentes e a possibilidade de sumarização das informações oriundas de um grande conjunto de dados.

Várias ferramentas computacionais podem ser utilizadas para a análise multivariada de dados experimentais. Dentre elas, o pacote *MultivariateAnalysis* se destaca por ser de uso livre, possuir várias funções, considerar a presença de delineamento estatístico e ser de fácil utilização.

Este *e-book* é composto por 10 capítulos. No primeiro apresenta-se a introdução sobre o tema central do material, a importância da análise multivariada nas Ciências Agrárias e ferramentas computacionais para análise multivariada. Já no segundo, descreve-se o processo de instalação do *software* R e Rstudio e a ativação de pacotes. Na sequência, o capítulo três trata das medidas de dissimilaridade tanto para dados quantitativos quanto para dados qualitativos, assim como para dados mistos. No capítulo quatro, descreve-se dendrogramas e suas metodologias para obtenção. O quinto capítulo apresenta a correlação cofenética. O sexto capítulo é relacionado a métodos de otimização, como, por exemplo, agrupamento Tocher. No sétimo, apresentam-se componentes principais. No oitavo, coordenadas principais. No nono, análise de variância multivariada (MANOVA). E no décimo capítulo, apresentam-se variáveis canônicas.

### **Análise multivariada e sua importância nas ciências agrárias**

A aplicação da análise multivariada é verificada em vários trabalhos, em diferentes áreas do conhecimento. Nas Ciências Agrárias, é aplicada com frequência em pesquisas

associadas à entomologia, fertilidade, nutrição de plantas, plantas medicinais, controle de plantas daninhas, no melhoramento genético, olericultura, irrigação, fruticultura etc.

Além de possibilitar o estudo dos efeitos dos “tratamentos” (variáveis independentes), considerando várias características (variáveis dependentes) simultaneamente, a análise multivariada possibilita trabalhar de forma mais simples quando se tem um grande banco de dados a ser utilizado, envolvendo muitas variáveis e dados amostrais, como, por exemplo, em pesquisas envolvendo a área de geotecnologia.

Muitas informações importantes podem ser obtidas a partir da análise multivariada, sendo possível identificar grupos de tratamentos similares, conhecer o que caracteriza cada grupo, identificar a associação entre grupos de características, compreender como as variáveis se relacionam na explicação de fenômenos biológicos, selecionar genitores em programas de melhoramento etc. Dessa forma, mesmo sendo a análise univariada de grande importância, a análise multivariada tem seu uso crescente nas pesquisas, possibilitando a obtenção de conclusões mais robustas e completas.

### **Ferramentas computacionais para a análise multivariada**

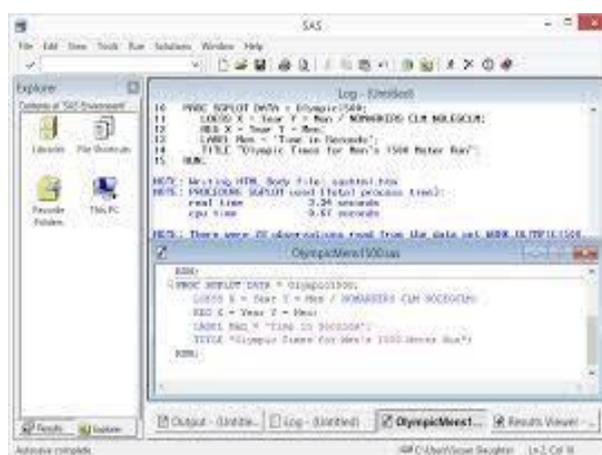
Para a realização de análise multivariada, vários *softwares* podem ser utilizados. Cada um tem suas especificidades, reunindo vantagens e desvantagens comparativamente. Nesse sentido, o *software* utilizado depende fortemente das experiências do pesquisador, das metodologias que deseja utilizar, de recursos financeiros para dispor de opções pagas, do tamanho do conjunto de dados etc.

Um dos *softwares* por meio do qual é possível a realização de análise multivariada é o GENES (CRUZ, 2013). Esse programa estatístico foi desenvolvido com o objetivo de atender às necessidades dos pesquisadores da área de Genética e Estatística Experimental. Na parte superior de sua interface podem-se encontrar várias ferramentas de análise multivariada, nas abas “A. Multivariada” e “Diversidade Genética”. O *software* tem enfoque em aplicações para o melhoramento genético, porém pode ser utilizado em várias áreas do conhecimento. Esse *software* possui interface gráfica, dispensando a necessidade de conhecimento de linguagens de programação, sendo por isso preferido por muitos usuários.



**Figura 2.** Interface do software GENES

Outro *software* possível para a execução de análise multivariada é o *Statistical Analysis System –SAS* (RODRIGUES, 2011). Esse programa computacional está disponível no mercado desde 1976, sendo um dos mais utilizados em todo o mundo, pois permite a realização de processamento e análise estatística de dados. Uma característica desse programa estatístico está no amplo espectro de aplicações. A realização de análise multivariada no *software* SAS depende do uso de linguagem de programação, o que dificulta sua utilização. O SAS é um programa pago, o que é uma grande desvantagem, pois não é acessível a todos.



**Figura 3:** Interface do *software* SAS

As opções para realização de análise multivariada são bastante vastas, indo de programas computacionais mais antigos até programas mais atuais. Outro exemplo de *software* computacional para execução de dados estatísticos é o SAEG (UFV, 1997). Com ele, é possível a realização de mais de 70 procedimentos estatísticos, incluindo as análises multivariadas. A interface do *software* SAEG é bastante amigável e simples, assim como mostra a Figura 4. As limitações de uso desse programa estatístico estão principalmente no

fato de ele ser executado somente em versões antigas do Windows, como a versão XP, e ser pago. Na versão de demonstração (gratuita), há a possibilidade de utilizar até 40 linhas de dados, o que é insuficiente para a análise da maioria dos experimentos.



**Figura 4.** Interface do *software* SAEG

O *software* STATISTICA (HILBE, 2007) é um programa que visualiza e executa análise de dados. Suas características estão na combinação de análises que cobrem milhares de algoritmos, testes, entre muitas outras funções, desde as mais simples até as mais complexas, sendo possível também a realização de análise multivariada. Esse programa pode ser obtido no *site* do repositório IBM SPSS ([Site IBM](#)), em que se encontra também o tutorial de instalação, assim como a versão mais atualizada do programa.

A lista de *softwares* que permitem a realização de análise multivariada é bastante longa. Além dos já citados anteriormente, podemos citar, ainda, Matlab, python, JMP, Minitab, SPSS, R. Muitos desses são usados em todo o mundo e utilizam linguagem de programação. Podemos destacar entre eles os programas computacionais que têm distribuição gratuita, como por exemplo o R e o python.

O *software* R, o qual será abordado no decorrer do *e-book*, destaca-se dos demais por vários motivos: é gratuito, depende de conhecimentos superficiais de programação, possui pacotes com interface gráfica e é aberto para programação. Por ser um *software* aberto, qualquer pessoa consegue criar pacotes para a solução dos mais diversos problemas. Posteriormente, esse pacote pode ser disponibilizado para instalação em todo o mundo, seja pela plataforma do R-CRAN, github ou várias outras opções. Devido a isso, o *software* R tem sido cada vez mais utilizado e reúne diversas funcionalidades para as mais variadas áreas do conhecimento. No momento, o *software* possui 18.648 pacotes, sendo vários desses úteis para estudos envolvendo a análise multivariada.

## Software R para análises multivariadas

Como dito anteriormente, dentre os *softwares* livre mais utilizados na atualidade, destaca-se o R, o qual pode ser utilizado com um visual mais agradável, por meio do RStudio. Desta forma, no RStudio é possível a utilização de linguagens de programação e pacotes do *software* R. Esses pacotes são bibliotecas, as quais reúnem várias funções e estão disponíveis na plataforma do R, conhecida como CRAN ([SITE CRAN](#)), podendo ser utilizados de forma gratuita.

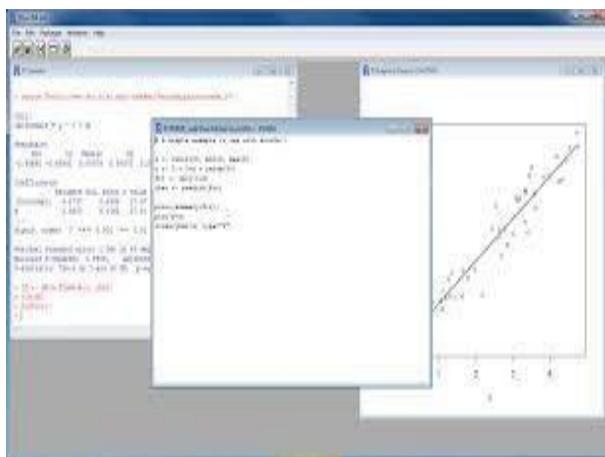


Figura 5. Interface do software R

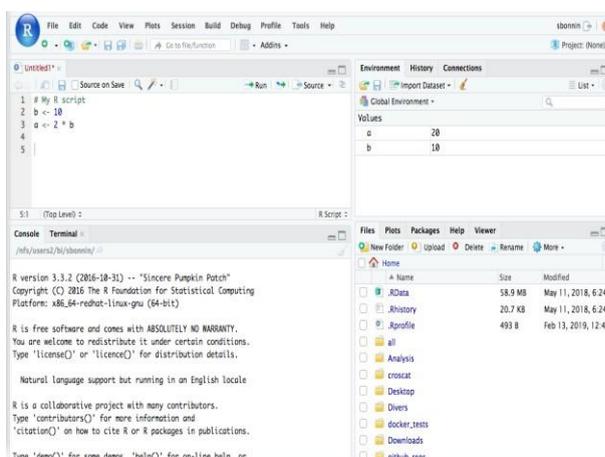


Figura 6. Interface do *software* RStudio

Na plataforma CRAN, encontram-se vários pacotes que realizam análises multivariadas. Um exemplo é o pacote *candisc* (FRIENDLY; FOX, 2017), criado no ano de 2008 por Michael Friendly e John Fox, tendo como principal função a execução de análises de variáveis canônicas. Outro pacote disponível para execução de análise multivariada, com enfoque principal na realização de análises filogenéticas e evolução, chama-se *ape*

(PARADIS *et al.*, 2019). Esse pacote foi publicado no ano de 2002, pelo pesquisador Emmanuel Paradis, e atualmente está na sua quinta atualização, lançada no ano de 2023.

O *biotools* (DA SILVA, 2017) é um importante pacote para realização de análises de biometria e estatística aplicada às Ciências Agrárias. Publicado por Anderson Rodrigo da Silva em 2014, esse pacote foi projetado para realizar e avaliar análises de cluster e análises discriminantes, possui ferramentas para realização de cálculos de tamanho de amostra e tamanho ótimo de parcela e já se encontra disponível o teste para a heterogeneidade da amostra de sementes.

Um pacote chamado *MVar* (OSSANI; CIRILLO, 2022) está disponível na plataforma CRAN para *download*. Nesse pacote é possível a realização de análise de correspondência simples e múltipla, análise de componentes principais, análise de correlação canônica, análise fatorial, escalonamento multidimensional, análise discriminante linear e quadrática, análise de cluster hierárquico e não hierárquico, regressão linear simples e múltipla, análise de múltiplos fatores para dados quantitativos, qualitativos, de frequência e dados mistos, *biplot*, *scatter plot*, *projection pursuit*, *grant tour* e outras funções úteis para a análise multivariada. A última atualização desse pacote foi lançada em 2022 pelo autor, Paulo Cesar Ossani Marcelo Angelo Cirillo, contendo versão em português e em inglês.

Para a realização de análises de componentes principais, rotação ortogonal e análise de múltiplos fatores existe o pacote *PCAmixdata* (CHAVENT *et al.*, 2014). Sua publicação ocorreu em 2014 por seus autores Marie Chavent (autora e criadora), Vanessa Kuentz, Amaury Labenne, Benoit Liquet e Jerome Saracco. Em 2023, esse pacote se encontrava na versão 3.1.

Para a realização de análise multivariada, há um pacote intitulado *MultivariateAnalysis*, desenvolvido em 2021, disponível no *site* da iniciativa do *software* R ([SITE CRAN](#)). Ele foi desenvolvido pelo pesquisador Alcinei Místico Azevedo, docente da Universidade Federal de Minas Gerais. Neste *e-book*, serão demonstradas várias funcionalidades desse pacote, com enfoque em pesquisas da área de produção vegetal.

### **Package MultivariateAnalysis**

O *MultivariateAnalysis* foi criado para ser um pacote que abranja metodologias de análise multivariada para avaliação de experimentos. Ele foi criado em 2021 pelo pesquisador Alcinei Místico Azevedo, professor da Universidade Federal de Minas Gerais (UFMG), cujas disciplinas ministradas são Estatística Experimental, Tópicos Avançados em Estatística Experimental, Inteligência Computacional e Visão Computacional, para os cursos de Agronomia, Zootecnia, Engenharia de Alimentos e para o Programa de Pós-Graduação em Produção Vegetal.

Em maio de 2024, o pacote se encontra na versão 0.5, sendo essa sua versão mais atual. O *MultivariateAnalysis* se encontra no idioma português, assim como seu manual, facilitando sua utilização por pesquisadores brasileiros.

Dentro do *MultivariateAnalysis*, é possível a realização de estimativa de medidas de dissimilaridade, dendrogramas, MANOVA, estimativa de análise de componentes principais, variáveis canônicas e agrupamento Tocher. Além desses, é possível a plotagem de gráficos relacionados a análises multivariadas, entre outras possibilidades. A descrição das análises e a orientação de como devem ser realizadas podem ser encontradas no manual do pacote, disponível na plataforma CRAN ([SITE CRAN](#)).

## 6. SOFTWARE R E RSTUDIO

Nos últimos anos, tem aumentado o número de usuários do *software* R, tanto no ambiente acadêmico quanto no ambiente corporativo, devido essencialmente ao fato de ser um *software* livre, de fácil aprendizado, gratuito e de fácil implementação. Uma característica do R é a possibilidade de realizar cálculos simples, mas também executar tarefas complexas. Com a finalidade de execução de tarefas mais elaboradas em determinadas áreas, é possível a utilização de pacotes disponíveis na plataforma CRAN. Esses pacotes têm por objetivo agrupar e preservar uma série de funções, já criadas, que são conhecidas como rotinas. Para obter maiores informações sobre o *software* R, visite a página do *The R Project for Statistical Computing R* ([SITE CRAN](#)).

### Instalação do R

O R está disponível de forma gratuita no *site* oficial ([SITE CRAN](#)) para as plataformas LINUX, Windows e MacOS. Para iniciar o *download* do R, deve-se acessar o *site* do CRAN. Ao acessar, a interface do *site* dá destaque para o *download* do programa. Na lateral esquerda, encontram-se informações estatísticas do *software*, novidades, pesquisas e equipe CRAN, no primeiro item chamado CRAN. O segundo item da lateral esquerda contém informações sobre o R e é intitulado *About R*. No terceiro item, apresentam-se informações do *software* que contemplam: Fontes R, Binários R, Pacotes, Visualizações de tarefas e Outros.

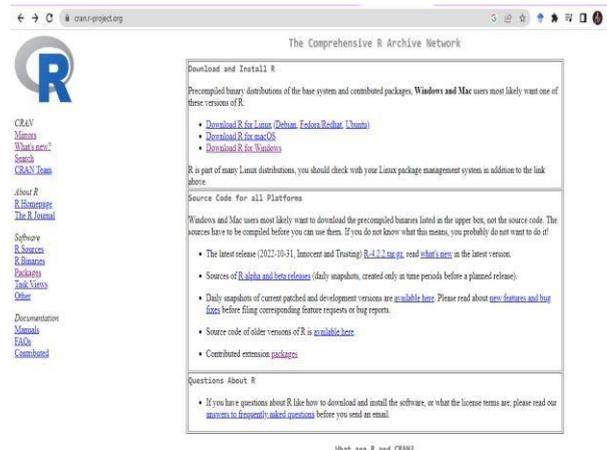


Figura 7. Interface do *site* CRAN

Após selecionar a opção desejada e clicar sobre ela, o *site* encaminha para outra página, para seleção de subdiretórios. Nessa parte, deve-se selecionar a opção: BASE. Após selecionar o item BASE, deve-se escolher a versão mais recente do R para *download*, clicando na opção de destaque para baixar o programa.

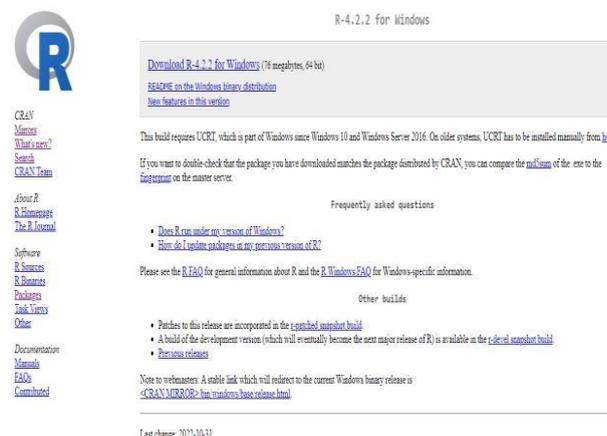


Figura 8. Interface do *site* CRAN para download

Em seguida, é necessário aguardar o *download*. Após concluído, deve-se executar e salvar o programa. Para a execução do programa, deve-se clicar sobre ele após *download*, selecionar o idioma e clicar em “avançar” várias vezes, para selecionar onde será salvo e realizar algumas escolhas pessoais que o *software* solicita. Em caso de dúvidas, existe um documento com perguntas frequentes no *site* do CRAN ([SITE CRAN](#)) e as dificuldades podem ser sanadas de forma *on-line*.

## Instalação do RStudio

O RStudio é um ambiente de desenvolvimento integrado (IDE) para o R. Assim como o R, o RStudio tem sua distribuição e utilização gratuitas, é uma ferramenta

facilitadora para desenvolvimento no R, devido à sua aparência mais amigável e extremamente visual, apresentando uma barra de comandos chamada console, que é muito simples e de fácil utilização. Nesta seção será demonstrado o processo de instalação do *software* RStudio.

Para iniciar o *download* do Rstudio, o primeiro passo consiste em acessar ao *site* do CRAN ([SITE CRAN](#)) e localizar na parte superior o *link* para *download*. Em seguida, deve-se clicar em *Products->RStudio*, ir em *Download RStudio Desktop* e selecionar seu sistema operacional. Para que não haja problemas futuros, é recomendável realizar a instalação do RStudio no mesmo diretório em que foi instalado o R e, em seguida, configurar o RStudio para *Executar como Administrador*.

### Interface Rstudio

O RStudio apresenta uma interface mais amigável e mais intuitiva que o R. A interface do RStudio é composta por console que apresenta os resultados dos comandos e possíveis erros, área de comandos (*script*) e, no canto superior direito, encontram-se *workspace* e *history*, que mostram os objetos criados, suas configurações e relatório dos últimos comandos executados. No canto inferior direito, são apresentadas as saídas gráficas do R. Nessa parte, é possível realizar a exportação, utilizando o botão *Export*, visualizar os pacotes instalados e obter informações através da função *Help*.

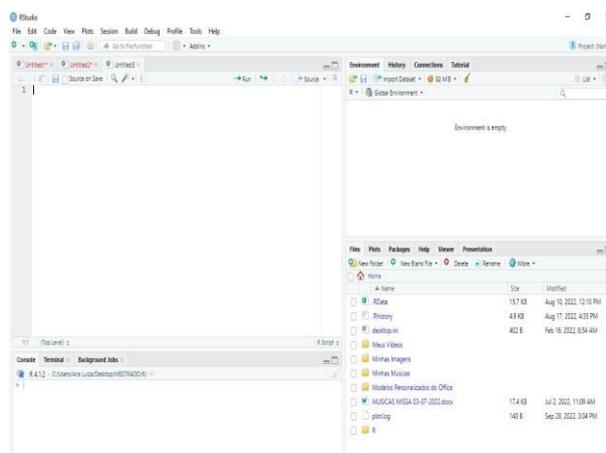


Figura 9. Interface do RStudio

### Instalação do MultivariateAnalysis

Para a realização da instalação do *MultivariateAnalysis*, deve-se seguir o

procedimento padrão para instalação de pacotes no RStudio. Para a realização desse procedimento, deve-se iniciar na parte inferior direita do RStudio, em que consta a função intitulada *packages*.

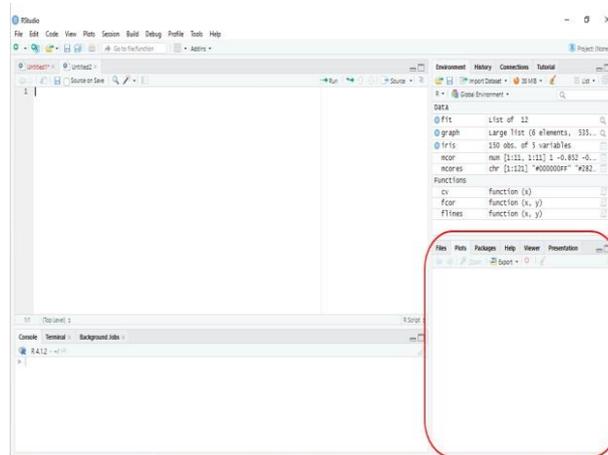


Figura 10. Início da instalação do *MultivariateAnalysis*

Nessa aba, deve-se clicar em *Install* e escrever o nome do pacote na caixa de diálogo. Essa opção é a mais usual para a realização de instalação do pacote, pois facilita encontrar o pacote.

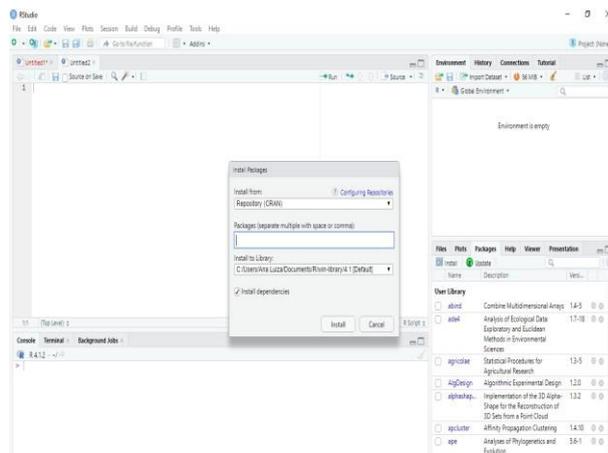


Figura 11. Função *Install* instalação de pacotes

Após concluído o processo de instalação do pacote, ele deve ser ativado para execução. Para isso, deve-se utilizar a função *library*, que em português significa biblioteca.

`library("MultivariateAnalysis")`

Mesmo após a instalação, sempre é necessário ativar os pacotes. Isto é uma estratégia utilizada pelo *software* para o funcionamento com um menor gasto de recurso computacional, pois, caso contrário, quando se têm muitos pacotes instalados, o *software* se torna muito lento.

Além da forma apresentada para a instalação do pacote, outra possibilidade é a utilização do seguinte comando: `install.packages("MultivariateAnalysis")`.

Após a instalação dos pacotes, é necessária a execução deles. Para a realização desse procedimento, pode-se utilizar a função `Run`, presente na parte superior da janela do *script* do programa R. Outra opção é a utilização das teclas `ctrl + enter`, logo após a execução da linha de comando.

Uma forma simples de importar dados no R é a partir de arquivos com a extensão `.txt`. Para a obtenção desses arquivos, pode-se tabular os dados no Excel e depois copiá-los e colá-los no bloco de notas. A forma como os dados devem ser tabulados depende da análise que se pretende fazer, e maiores informações podem ser obtidas nos arquivos de ajuda do manual do pacote *MultivariateAnalysis*. Inicialmente, deve-se indicar ao R a pasta em que o arquivo `.txt` está salvo. Para isso, pode-se utilizar a função `setwd()`. Dentro dos parênteses, entre aspas, deve ser colado o endereço da pasta.

Para ler o arquivo de trabalho, utiliza-se o comando `read.table()`. E entre os parênteses desse comando, deve-se conter o nome do arquivo entres aspas duplas. Uma maneira prática para localizar o arquivo de trabalho é digitar as aspas duplas e, entre elas, clicar na tecla TAB do teclado do computador. Dessa forma, serão mostrados todos os arquivos dentro da pasta indicada na função `setwd()`.

Quando realizada de forma correta a leitura dos dados, eles aparecerão no console do RStudio. Caso haja alguma incoerência, seja na linha de comandos ou na leitura dos dados, no console aparecerá uma mensagem vermelha indicando que houve algum erro e qual tipo de erro, em inglês. Traduzindo-o para o português, será possível identificar qual o problema.

Quando os dados apresentam cabeçalho (identificação dos tratamentos, repetições e variáveis dependentes analisadas), é necessária a identificação e a leitura do cabeçalho. Para realizá-las, deve adicionar ao comando `read.table()` o argumento chamado `HEAD`, que em português significa cabeçalho. Nesse argumento, pode-se atribuir o valor verdadeiro ou falso (em relação à presença ou ausência do cabeçalho no arquivo de dados). Quando verdadeiro, deve-se utilizar `head=TRUE`, e, em caso de falso, deve-se utilizar `head=FALSE`.

## 7. MEDIDAS DE DISTÂNCIAS

### Medidas para dados quantitativos

Uma das principais funções do *MultivariateAnalysis* é a realização de medidas de distâncias. Essas se referem à distância entre diferentes indivíduos, considerando-se múltiplas características simultaneamente. Esse primeiro momento se refere a dados quantitativos, ou seja, variáveis expressas por números, como diâmetro de caule, altura de planta, peso, entre outras. Quanto aos dados qualitativos, serão explicados posteriormente.

### 7.1.1 Distância Euclidiana

Pode-se definir distância euclidiana como o intervalo entre dois pontos, generalizando o Teorema de Pitágoras para diversas variáveis simultaneamente. Matematicamente, a distância euclidiana pode ser expressa pela fórmula:

$$D_{ij} = \sqrt{\sum_{k=1}^n (x_{ki} - x_{kj})^2}$$

Em que:

$D_{ij}$  é a distância entre os indivíduos  $i$  e  $j$ ;

$X_{ki}$  é o valor observado para a variável  $k$  no indivíduo  $i$ ;

$X_{kj}$  é o valor observado para a variável  $k$  no indivíduo  $j$ ; e

$n$  é o número de variáveis.

Para a execução de medidas euclidianas utilizando o *MultivariateAnalysis*, deve-se ativar o pacote com a função *library (MultivariateAnalysis)* e importar os dados. Como exemplo, pode-se utilizar os próprios dados presentes no pacote. Para conhecê-los, pode-se acessar o manual, recorrendo ao comando *?MultivariateAnalysis*. No tópico presente, serão utilizados os dados apresentados quando se executa *?Distancia*.

O primeiro arquivo a ser utilizado é denominado *dados.MED*, que pode ser importado executando *data(dados.MED)*. Para visualizar os dados, uma opção é executar o comando *View(dados.Med)*. Uma observação que se deve levar em conta ao trabalhar com dados quantitativos é que os valores presentes nas colunas devem ser preenchidos apenas com números, para que não haja nenhum erro na execução. Além disso, os nomes de indivíduos e variáveis não podem estar contidos entre dados, e sim no cabeçalho ou no nome de linhas. Outra informação importante é que no R deve-se considerar ponto como separador de casas decimais, não vírgula.

```
library(MultivariateAnalysis)
```

```
## Warning: package 'MultivariateAnalysis' was built under R version 4.2.3
```

```
## #####
```

```
## Obrigado por utilizar o MultivariateAnalysis
```

```
## Author: Alcinei Místico Azevedo (ICA-UFG)
```

```
## Veja tutoriais sobre este e outros pacotes no youtube:
```

```
## https://www.youtube.com/channel/UCDGyvLCJnv9RtTY1YMBMVNQ
```

```
## Se inscreva e compartilhe para ajudar o canal a crescer.
```

```
## #####
```

```
D=DadosQuanti=matrix(rnorm(100,100,5),ncol=4)
D
##           [,1]      [,2]      [,3]      [,4]
## [1,]  91.78409 102.70402  95.96916  99.73029
## [2,]  96.15618 114.76415 101.44407  98.23381
## [3,]  90.99494 100.19211 103.07133 103.89199
## [4,] 112.70158 108.27965  96.91413  90.58947
## [5,]  98.23677  99.33533  95.58782  97.11083
## [6,]  96.01129 103.68571  94.19513 100.67658
## [7,]  90.27355 105.07046  99.65735 109.03108
## [8,] 101.04291  95.68009  97.58652  99.93627
## [9,]  96.52213 103.35770 108.40286  98.60540
## [10,] 95.51543  97.24190  96.29098  96.32692
## [11,] 103.27352 109.58493  95.24489 108.96634
## [12,]  96.48374  98.48797 104.14740  89.51865
## [13,]  95.20413 109.39584 103.12847  98.48159
## [14,]  98.25038 101.52879 107.55859  94.67650
## [15,] 104.44259  95.14784 101.79726  91.74657
## [16,] 104.14288  96.10089  92.86230 103.85943
## [17,]  97.01049  92.90945  94.79612 102.78717
## [18,]  98.30997  99.83741 108.21474  99.76357
## [19,]  96.05814  97.96481 112.00179 107.71967
## [20,] 102.26230 103.34532  90.31231  92.20346
## [21,]  97.85938 103.76877  90.26248  90.10778
## [22,]  97.48277 107.81956 103.41881  98.04549
## [23,]  92.82818  97.81890 103.41117  98.06662
## [24,] 100.65615  89.58268 107.07108  91.08845
## [25,] 101.37531 102.77939 100.43039 100.40611
```

Dentro do pacote, existem várias formas de execução de distâncias para dados quantitativos, em que se considera a distância euclidiana, sendo elas:

```
+Distância euclidiana média
+Distância euclidiana padronizada
+Distância euclidiana padronizada média
```

```
library(MultivariateAnalysis)
?MultivariateAnalysis

## starting httpd help server ... done

data(Dados.MED)
#####> Dados quantitativos
#1 = Distancia euclidiana.
Distancia(Dados.MED,1)

## Medida de dissimilaridade: 1 = Distancia euclidiana.
##
## Menor Distancia: 0.4474513
## Maior Distancia: 10.54687
## Media das Distancias: 5.182184
## Amplitude das Distancias: 10.09941
## Desvio Padrao das Distancias: 2.622392
## Coeficiente de variacao das Distancias: 50.60399
## Indivíduos mais proximos: 2 4
## Indivíduos mais distantes: 6 10
```

## Distância de Mahalanobis

A distância de Mahalanobis pode ser utilizada quando há delineamento estatístico,

empregando-se a matriz de variância e covariância. Para obter essa matriz, pode-se utilizar a função MANOVA (análise de variância para análise multivariada). Essa função retorna uma matriz com os valores de média e outra com a covariância residual (CRE). A distância de Mahalanobis pode ser obtida considerando-se o método 7 na função `Distancia`. Para realizar o cálculo da distância euclidiana e distância de Mahalanobis, são utilizados estimadores, como mostra a equação a seguir:

$$D_{ii'}^2 = V'R^{-1}V$$

Em que:

$D_{ii'}^2$  = representa a distância de Mahalanobis, que não é representada pela raiz quadrada como a distância euclidiana e sim pelo valor ao quadrado;

V = vetor resultado da diferença entre as variáveis respostas;

V' = transposta do vetor resultado da diferença entre as variáveis respostas;

$R^{-1}$  = matriz de variância e covariância residual.

```
library(MultivariateAnalysis)
?MultivariateAnalysis
data(Dados.DBC)
m=MANOVA(Dados.DBC,2)
Med=apply(Dados.DBC[, -c(1:2)],2,function(x) tapply(x,as.factor(Dados.DBC[,1]),mean))
CRE=m$CovarianciaResidual
Distancia(Med,7,CRE)

## Medida de dissimilaridade: 7 = Distancia de Mahalanobis.
##
## Menor Distancia: 3.657931
## Maior Distancia: 47.26671
## Media das Distancias: 23.07077
## Amplitude das Distancias: 43.60877
## Desvio Padrao das Distancias: 11.57168
## Coeficiente de variacao das Distancias: 50.15731
## Individuos mais proximos: Variedade7 Variedade9
## Individuos mais distantes: Variedade6 Variedade7
```

### Distâncias de Cole Rodgers

A distância de Cole Rodgers é um método estatístico também utilizado para dados quantitativos, podendo ser calculada a partir da expressão abaixo, de acordo com Cruz (2006):

$$D_{ii'} = \sum_{j=1}^n (r_{ij} - r_{i'j})^2$$

Em que:

$$r_{ij} = \frac{Y_{ij} - Y_{jmin}}{Y_{jmax} - Y_{jmin}}$$

Sendo:

$Y_{ij}$ : corresponde ao valor do  $i$ -ésimo tratamento em relação à  $j$ -ésima variável;

$Y_{j\min}$ : valor mínimo encontrado para  $j$ -ésima característica;

$Y_{j\max}$ : valor máximo encontrado para  $j$ -ésima característica.

Para Cole-Rodgers *et al.* (2004), para descritores qualitativos, o modelo da equação é a seguinte:

$$d_{ij} = \sum_{j=1}^n \frac{D_j}{C_j D_j}$$

Em que:

$D_j$ : é o número de descritores entre os grupos para o  $j$ -ésimo descritor;

$C_j$ : é o número de concordância entre os grupos para o  $j$ -ésimo descritor.

### **Distâncias para dados binários**

Os dados binários são caracterizados por terem apenas duas classificações, por exemplo: sim ou não, morto ou vivo, doente ou sadio, aprovado ou reprovado. Uma maneira de representar esses dados binários é utilizando os números 0 e 1, em que 0 pode representar ausência e 1 presença, por exemplo. Para a estimativa da similaridade nesse caso, pode-se considerar a porcentagem de concordância nas classificações dos indivíduos como medida de similaridade.

$$Co incidência = \frac{a + d}{a + b + c + d}$$

Em que:

$a$ = número de coincidência entre os indivíduos do tipo 1-1

$b$ = número de discordância entre os indivíduos do tipo 1-0

$c$ = número de discordância entre os indivíduos do tipo 0-1

$d$ = número da coincidência entre os indivíduos do tipo 0-0

Quanto mais próximo for a medida de 1, mais similar são os indivíduos. Já para obter a dissimilaridade, ou seja, o quanto os indivíduos analisados são diferentes, é necessário fazer uma correção da fórmula, que é o complemento da coincidência.

$$Discordância = 1 - \frac{a + d}{a + b + c + d}$$

Os valores de discordância geram uma matriz de dissimilaridade. Essa, por sua vez, é uma matriz quadrada, apresentando na diagonal principal o valor de 0, que representa a distância do indivíduo a ele mesmo, e fora da diagonal principal constam os valores das distâncias. A partir dessa matriz, é possível realizar dendrogramas, agrupamentos Tocher etc.

O mesmo princípio utilizado para a coincidência deve ser aplicado para a coeficiente de Jacard. Esse é um método que desconsidera como similares os indivíduos que apresentam valores de 0 e 0 para uma variável. Isso é muito importante para pesquisas com marcadores moleculares, por exemplo, quando há ausência de uma banda em dois indivíduos, o que não quer dizer que necessariamente há diferença entre os indivíduos.

$$\text{ComplementoJacard} = \frac{a + a}{a + b + c}$$

No *MultivariateAnalysis*, para a análise de dados binários, é preciso uma matriz em que constem, nas colunas, as variáveis respostas, e nas linhas, os indivíduos, sendo que, dentro da linha, devem constar apenas os valores numéricos (0 ou 1). Deve-se inicialmente indicar quais dados serão trabalhados e, utilizando o método 9 na função *Distancia*, obter-se a frequência de coincidência. Para medidas de discordância, deve-se utilizar o método 10. Para índice inverso de coincidência, basta utilizar o método 11.

```
library(MultivariateAnalysis)
?MultivariateAnalysis
#>Dados qualitativos: binarios ou multicategoricos
#9 = Frequencia de coincidencia.
data(Dados.CAT)
Distancia(Dados.CAT,9)

## Medida de dissimilaridade: 9 = Frequencia de coincidencia.
##
## Menor Distancia: 0
## Maior Distancia: 0.8461538
## Media das Distancias: 0.3653846
## Amplitude das Distancias: 0.8461538
## Desvio Padrao das Distancias: 0.194058
## Coeficiente de variacao das Distancias: 53.1106
## Indivíduos mais proximos: 4 5
## Indivíduos mais distantes: 1 9

#10 = Frequencia de discordancia.
Distancia(Dados.CAT,10)

## Medida de dissimilaridade: 10 = Frequencia de discordancia.
##
## Menor Distancia: 0.1538462
## Maior Distancia: 1
## Media das Distancias: 0.6346154
## Amplitude das Distancias: 0.8461538
## Desvio Padrao das Distancias: 0.194058
## Coeficiente de variacao das Distancias: 30.57883
## Indivíduos mais proximos: 1 9
## Indivíduos mais distantes: 4 5

data(Dados.BIN)
Distancia(Dados.BIN,10)
```

```

## Medida de dissimilaridade: 10 = Frequencia de discordancia.
##
## Menor Distancia: 0.2222222
## Maior Distancia: 0.8888889
## Media das Distancias: 0.5777778
## Amplitude das Distancias: 0.6666667
## Desvio Padrao das Distancias: 0.2146871
## Coeficiente de variacao das Distancias: 37.15738
## Indivíduos mais proximos: 1 4
## Indivíduos mais distantes: 4 5

#11 = indice Inverso de 1+coincidencia > 1/(1+c)
Distancia(Dados.CAT,11)

## Medida de dissimilaridade: 11 = indice Inverso de 1+coincidencia = 1/(1+c)
##
## Menor Distancia: 0.5416667
## Maior Distancia: 1
## Media das Distancias: 0.7466749
## Amplitude das Distancias: 0.4583333
## Desvio Padrao das Distancias: 0.1048636
## Coeficiente de variacao das Distancias: 14.04408
## Indivíduos mais proximos: 1 9
## Indivíduos mais distantes: 4 5

```

## Dados Mistos

Em pesquisas, pode haver tanto dados qualitativos como quantitativos. Nesse caso, denominamos os dados como mistos. Existem diferentes formas para se trabalhar com dados mistos. Uma delas é utilizar a metodologia apropriada para as diferentes naturezas dos dados. Por exemplo, realizar a distância de Mahalanobis para os dados quantitativos e o índice de discordância para os dados qualitativos e, posteriormente, fazer a média ponderada dessas matrizes, considerando-se o número de variáveis contempladas em cada um. Nesse caso, um cuidado importante antes de efetuar a média ponderada das matrizes é a normalização delas, fazendo com que se tornem adimensionais (variando entre zero e um).

Outra possibilidade para se trabalhar com dados mistos é utilizando o Índice Gower, o qual, semelhante aos casos anteriores, é um estimador de similaridade, sendo necessária a correção para a obtenção da dissimilaridade.

$$S_{ij} = \frac{\sum_{k=1}^p W_{ijk} S_{ijk}}{\sum_{k=1}^p W_{ijk}}$$

Em que:

i e j são dois indivíduos quaisquer;

k é a característica; p é o número total de características avaliadas;

$W_{ijk}$  é um peso dado à comparação ijk, atribuindo valor 1 para comparações válidas e valor 0 para comparações inválidas;

$S_{ijk}$  é a contribuição da variável k na similaridade entre os indivíduos i e j possuindo valores entre 0 e 1;

Para uma variável qualitativa, se o valor da variável  $k$  é o mesmo para ambos os indivíduos  $i$  e  $j$ , então  $S_{ijk} = 1$ ; caso contrário é igual a zero; para uma variável contínua  $S_{ijk} = 1 - |x_{ik} - x_{jk}| / R_k$  em que  $x_{ik}$  e  $x_{jk}$  são os valores da variável  $k$  para os indivíduos  $i$  e  $j$ , respectivamente.

Para a execução de dados mistos no *MultivariateAnalysis*, deve ser utilizado o modelo de número 21 na função *Distancia*, como mostra o exemplo abaixo. Para a organização dos dados, deve-se seguir a ordem: nas linhas, os indivíduos; nas colunas, as variáveis respostas, não importando a ordem das variáveis respostas. No mesmo arquivo, podem constar tanto variáveis qualitativas, como quantitativas e binárias.

```
library(MultivariateAnalysis)
?MultivariateAnalysis
#' #####>Dados mistos (quantitativos, binarios e multicategoricos)
data(Dados.Misto)
Distancia(Dados.Misto,21)

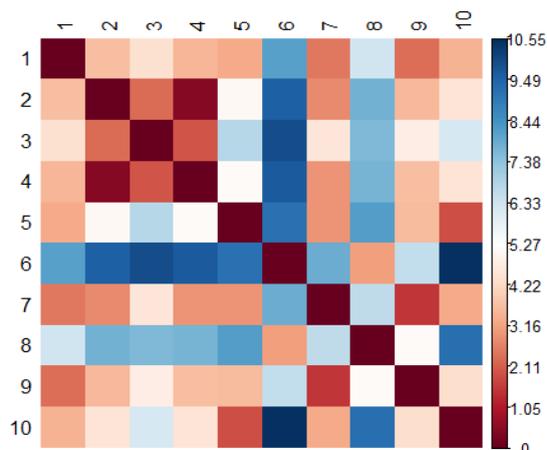
## Medida de dissimilaridade: 21 =Dissimilaridade de Gower
##
## Menor Distancia: 0.3249251
## Maior Distancia: 0.7883117
## Media das Distancias: 0.5741259
## Amplitude das Distancias: 0.4633866
## Desvio Padrao das Distancias: 0.1364548
## Coeficiente de variacao das Distancias: 23.7674
## Indivíduos mais proximos: 1 2
## Indivíduos mais distantes: 4 5
```

### Visualizações das matrizes de dissimilaridade

Existe uma função dentro do *MultivariateAnalysis* chamada de *SummaryDistancia*, que possibilita a obtenção de informações relacionadas à matriz de dissimilaridade. Os argumentos utilizados para realização dessa função se iniciam com a matriz de dissimilaridade, seguida por valor numérico das casas decimais, *layout* do gráfico a ser plotado, seguido de tipo, título e características da fonte. Por fim, o argumento se encerra com a paleta de cores que se tem interesse.

Além disso, a função possibilita a dispersão da matriz de dissimilaridade em uma figura, facilitando a visualização dos indivíduos mais dissimilares.

```
library(MultivariateAnalysis)
data(Dados.MED)
Dist=Distancia(Dados.MED,1)
SummaryDistancia(Dist)
```



```
##
##      Medio  Minimo  Maximo   sd  MaisProximo  MaisDistante
## 1    4.18    2.37    8.13  1.87           9           6
## 2    4.41    0.45    9.57  2.79           4           6
## 3    5.40    2.00    9.91  2.53           4           6
## 4    4.40    0.45    9.66  2.84           2           6
## 5    5.12    1.86    9.21  2.50          10           6
## 6    8.28    3.08   10.55  2.30           8          10
## 7    3.89    1.50    7.90  2.09           9           6
## 8    6.86    3.08    9.25  1.84           6          10
## 9    3.95    1.50    6.54  1.50           7           6
## 10   5.33    1.86   10.55  2.86           5           6
##
```

```
## Menor Distancia: 0.4474513
## Maior Distancia: 10.54687
## Media das Distancias: 5.182184
## Amplitude das Distancias: 10.09941
## Desvio Padrao das Distancias: 2.622392
## Coeficiente de variacao das Distancias: 50.60399
## Indivíduos mais proximos: 2 4
## Indivíduos mais distantes: 6 10
##
```

## 8. DENDROGRAMA

O dendrograma é uma representação gráfica das distâncias estimadas pelos métodos hierárquicos de agrupamento, facilitando a visualização e a interpretação. Existem muitas metodologias para se obterem dendrogramas. Além disso, sua apresentação pode ser circular ou em árvores.

### Algoritmo

Um dos métodos para a criação do dendrograma é o método do vizinho mais próximo ou método da ligação simples. Para se obter o dendrograma, deve-se observar na matriz de dissimilaridade quais são os indivíduos mais parecidos (menor distância), estabelecendo-se o ponto de fusão. O passo seguinte é a união entre os indivíduos para se obter uma nova matriz de dissimilaridade, ou seja, a atualização da matriz anterior. Para isso, observam-se os valores dos indivíduos mais similares e escolhe-se o menor valor entre eles – daí o nome

vizinho mais próximo –, o qual será substituído na nova matriz. Da mesma forma, devem-se observar as novas distâncias, ver qual delas tem maior proximidade e depois escolher o menor valor. Os passos devem se repetir até a finalização do dendrograma. Para melhor entendimento, pode-se observar o exemplo a seguir.

**Tabela 2.** Teor de proteína e a altura do capim Tifton 85 em função de seis variedades

Variedade	Teor de proteína (%)	Altura (cm)
A (1)	9.6	28
B (2)	8.4	31
C (3)	2.4	42
D (4)	18.2	38
E (5)	3.9	25
F (6)	6.4	41

A partir desses cálculos, obteve-se a seguinte matriz de distância euclidiana:

	A	B	C	D	E	F
A	0.00	–	–	–	–	–
B	3,23	0.00	–	–	–	–
C	15,74	12,53	0.00	–	–	–
D	13,19	12,04	16,29	0.00	–	–
E	6,44	7,50	17,06	19,33	0.00	–
F	13,39	10,19	4,12	12,18	16,19	0.0

**Passo 1:** Identificar os tratamentos mais próximos na matriz anterior, para formar o primeiro grupo e atualizá-lo para uma nova matriz, considerando a distância mínima. Nesse caso, os tratamentos mais próximos foram A e B ( $D^2=3,23$ ). Portanto, a matriz atualizada será:

	A B	C	D	E	F
A B	0.00	–	–	–	–
C	12,53	0.00	–	–	–
D	12,04	16,29	0.00	–	–
E	6,44	17,06	19,33	0.00	–
F	10,19	4,12	12,18	16,19	0.00

**Passo 2:** Identificar, na matriz anterior, os tratamentos mais próximos, para formar o primeiro grupo e atualizá-lo para uma nova matriz, considerando a distância mínima. Nesse caso, os tratamentos mais próximos foram C e F ( $D^2=4,12$ ). Portanto, a matriz atualizada será:

	A B	C F	D	E
A B	0,00	-	-	-
C F	10,19	0,00	-	-
D	12,04	12,18	0,00	-
E	6,44	16,19	19,33	0,00

**Passo 3:** Identificar, na matriz anterior, os tratamentos mais próximos, para formar o primeiro grupo e atualizá-lo para uma nova matriz, considerando a distância mínima. Nesse caso, os tratamentos mais próximos foram “A,B” e E ( $D^2=6,44$ ). Portanto, a matriz atualizada será:

	A,B,E	C,F	D
A,B,E	0,00	-	-
C,F	10,19	0,00	-
D	12,04	12,18	0

**Passo 4:** Identificar, na matriz anterior, os tratamentos mais próximos, para formar o primeiro grupo e atualizá-lo para uma nova matriz, considerando a distância mínima. Nesse caso, os tratamentos mais próximos foram “A,B,E” e “C,F” ( $D^2=10,19$ ). Portanto, a matriz atualizada será:

	A,B,C,F	F
A,B,E,C,F	0	-
D	12,04	0

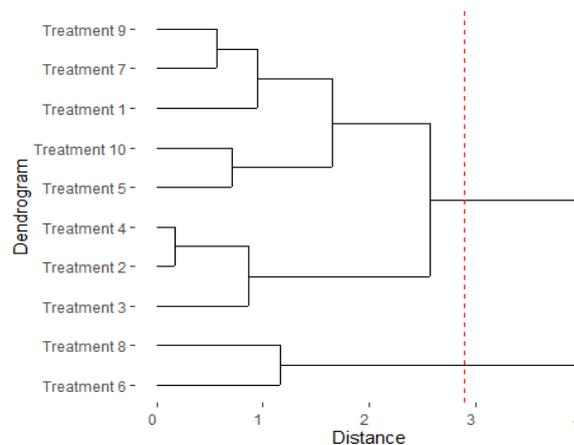
Assemelhando-se ao método do vizinho mais próximo, o vizinho mais distante tem uma pequena diferença, pois, no passo 2, ao atualizar a matriz, seria selecionada a maior distância. Já no método UPGMA, é considerado o valor da média entre as distâncias nessa etapa. Além dos métodos destacados acima, há vários outros que se baseiam em outros princípios, como os métodos de Ward, do Centróide, Mediana, Mcquitty etc.

### Análise em R

No *MultivariateAnalysis*, é possível a realização de todos os métodos de construção de dendrogramas acima citados. Para a realização dos dendrogramas, necessita-se de dois objetos, sendo um obtendo uma matriz de dissimilaridade e o número que representa a metodologia escolhida. Na parte inferior do manual, na opção de dendrograma, consta um exemplo de como se deve realizar a função para o processamento dos dados para obtenção do dendrograma.

Seguindo o exemplo dentro pacote, inicia-se limpando a memória do R, utilizando o comando `remove(list=ls())`. Após este, deve-se ativar o pacote previamente instalado, utilizando-se o comando `library("MultivariateAnalysis")`. Para a finalidade de consulta do manual do pacote, utiliza-se a interrogação (`?`), seguida do nome do pacote – `?MultivariateAnalysis` – e, no índice, identifica-se a função dendrograma, ou se executa o comando `?Dendrograma`. Como todos os outros comandos, deve-se começar indicando quais são os dados e sua localização. Em seguida, obtêm-se as distâncias, utilizando-se a função `Distancia(Dados.MED,2)`. Obtidas as distâncias, pode-se seguir com a obtenção do dendrograma, utilizando-se a função `Dendrograma`, em que o primeiro objeto contempla o objeto de distâncias obtidas anteriormente.

```
library(MultivariateAnalysis)
data(Dados.MED)
rownames(Dados.MED)=paste0("Treatment ",1:10)
Dist=Distancia(Dados.MED,2)
Dendrograma(Dist,Metodo = 2)
```



```
## -----
## Estimativa de correlacao cofenetica:
## [1] 0.8914551
## Significancia da correlacao cofenetica pelo teste Mantel
## pvalor: 0.001
## Hipotese alternativa: A correlacao e maior que 0
##
## Criterio de Corte
## k=1.25
## 2.891794
##
## Agrupamentos
## Cluster
## Treatment 1 1
## Treatment 2 1
## Treatment 3 1
## Treatment 4 1
## Treatment 5 1
## Treatment 6 2
## Treatment 7 1
## Treatment 8 2
## Treatment 9 1
## Treatment 10 1
## -----
```

Além do exemplo acima retratado, dentro do *MultivariateAnalysis* é possível a realização de oito métodos para obtenção de dendrogramas, sendo eles:

- Método 1 = Ligação simples (Vizinho mais próximo);
- Método 2 = Ligação completa (Vizinho distante);
- Método 3 = Ligação média entre grupos (UPGMA);
- Método 4 = Método de Ward;
- Método 5 = Método de Ward (d2);
- Método 6 = Método da Mediana (WPGMC);
- Método 7 = Método do Centróide (UPGMC);
- Método 8 = Método Mcquitty (UPGMC);

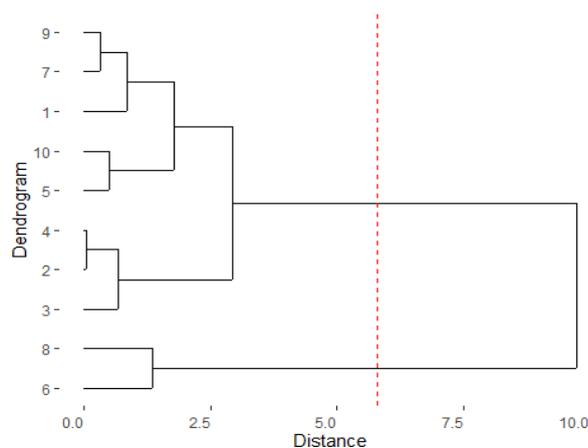
### Função HeatPlot

A função *HeatPlot*, disponível no *MultivariateAnalysis*, tem por finalidade a apresentação de um mapa de calor junto com o dendrograma, de forma simples e objetiva, utilizando poucos argumentos. Para sua realização, utiliza-se o comando *HeatPlot*, com os argumentos *dendo* – que está relacionado ao objeto ‘dendrograma’ utilizado anteriormente – e o argumento *col* – relacionado à paleta de cores que será utilizada.

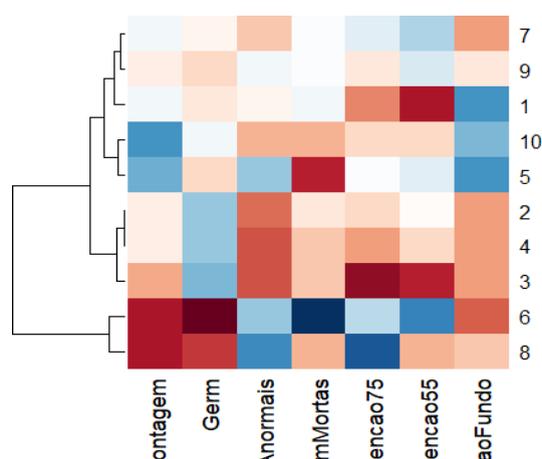
```
#PlotHeat
#Distancia euclidiana
data("Dados.MED")
dist=Distancia(Dados.MED,Metodo = 3)
dist

## Medida de dissimilaridade: 3 = Quadrado da distancia euclidiana media.
##
## Menor Distancia: 0.02860181
## Maior Distancia: 15.89091
## Media das Distancias: 4.797021
## Amplitude das Distancias: 15.86231
## Desvio Padrao das Distancias: 4.383965
## Coeficiente de variacao das Distancias: 91.38932
## Individuos mais proximos: 2 4
## Individuos mais distantes: 6 10

Dendo=Dendrograma(dist)
```



HeatPlot (Dendo)



O exemplo de um gráfico de calor gerado pelo *MultivariateAnalysis*. O gráfico de calor adiciona ao dendrograma informações sobre o que caracteriza cada tratamento. Desta forma, é possível identificar as características que fazem os genótipos se diferenciarem entre si.

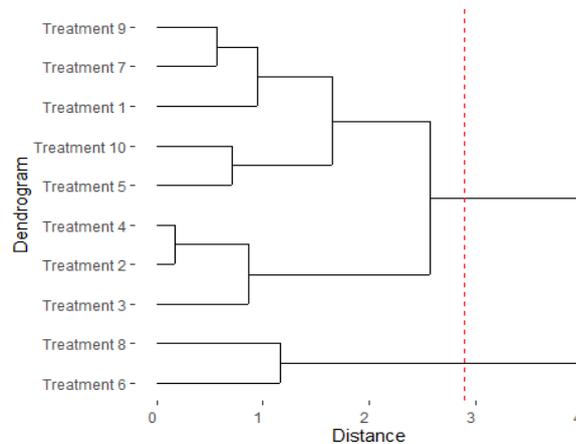
### Correlação cofenética

A matriz de dissimilaridade pode ser resumida em um dendrograma, conforme apresentado nos tópicos anteriores. E a partir dos pontos de fusão presentes no dendrograma é possível obter uma matriz de distâncias chamada matriz fenética. Assim, uma forma de verificar se o dendrograma representa bem a matriz de dissimilaridade é estimar a correlação entre a matriz de dissimilaridade e a matriz fenética, obtendo-se o que chamamos de correlação cofenética.

Pode-se definir a correlação cofenética como coeficiente que tem por objetivo quantificar a similaridade entre a matriz de dissimilaridade e o dendrograma (CRISPIM *et al.*, 2019, CARVALHO *et al.*, 2009; GOBO, GALVANI, WOLLMANN, 2018). Ou seja, a correlação cofenética indica se o dendrograma está sendo eficiente ou não em refletir, linearmente, a matriz de dissimilaridade. Pode-se utilizar como referência o valor de 0,7. Valores menores

que essa referência não explicam de forma satisfatória a matriz de dissimilaridade; valores maiores que 0,7 explicam de forma mais satisfatória. No *MultivariateAnalysis*, a correlação cofenética é obtida junto com o dendrograma.

```
library(MultivariateAnalysis)
?MultivariateAnalysis
data(Dados.MED)
rownames(Dados.MED)=paste0("Treatment ",1:10)
Dist=Distancia(Dados.MED,2)
Dendrograma(Dist,Metodo = 2)
```



```
## -----
## Estimativa de correlacao cofenetica:
## [1] 0.8914551
## Significancia da correlacao cofenetica pelo teste Mantel
## pvalor: 0.001
## Hipotese alternativa: A correlacao e maior que 0
##
## Criterio de Corte
## k=1.25
## 2.891794
##
## Agrupamentos
##      Cluster
## Treatment 1      1
## Treatment 2      1
## Treatment 3      1
## Treatment 4      1
## Treatment 5      1
## Treatment 6      2
## Treatment 7      1
## Treatment 8      2
## Treatment 9      1
## Treatment 10     1
## -----
```

## Critério de Corte

### 8.1.1 Pontos de corte pelo método visual

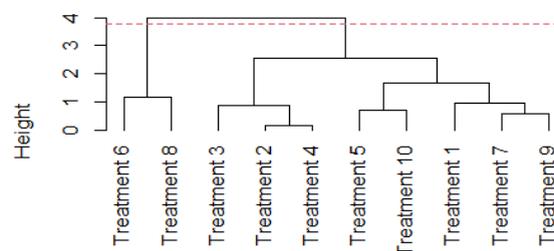
Após realizado o dendrograma, é preciso determinar quantos grupos devem ser formados, a partir de uma linha conhecida como ponto de corte (determinante da quantidade

de grupos). O método subjetivo pode ser feito, considerando a formação de grupos de forma abrupta dentro do dendrograma ou a experiência do pesquisador. Por ser subjetivo, pode ser realizado de diferentes formas por cada pessoa. Para a utilização desse método, basta colocar no argumento corte, na função Dendrograma, o valor desejado. Esse critério visual torna-se mais difícil ainda na medida que a quantidade de indivíduos aumenta. Uma solução para esse problema está na utilização de metodologias objetivas de determinação do ponto de corte, por exemplo, o método Mojena.

### 8.1.2 Pontos de corte – Critério Mojena

Para a estimativa do ponto ótimo de corte, por meio teste de Mojena (1977), basta fazer a média dos valores de ponto de fusão e somar com o desvio-padrão de todos os valores multiplicado por uma constante “k”, como mostra a equação abaixo. Para os valores da constante “k” em trabalhos de simulação, indicaram-se dois valores, sendo igual a 1,25 ou 2,00. Desses valores, o mais considerado nas pesquisas é o valor de 1,25.

```
library(MultivariateAnalysis)
?MultivariateAnalysis
data(Dados.MED)
rownames(Dados.MED)=paste0("Treatment ",1:10)
Dist=Distancia(Dados.MED,2)
Dendrograma(Dist,2,layout =1,corte = "Mojena2")
```



```
## -----
## Estimativa de correlacao cofenetica:
## [1] 0.8914551
## Significancia da correlacao cofenetica pelo teste Mantel
## pvalor: 0.001
## Hipotese alternativa: A correlacao e maior que 0
##
## Critério de Corte
##      k=2
## 3.785391
##
## Agrupamentos
##      Cluster
## Treatment 1      1
## Treatment 2      1
```

```
## Treatment 3      1
## Treatment 4      1
## Treatment 5      1
## Treatment 6      2
## Treatment 7      1
## Treatment 8      2
## Treatment 9      1
## Treatment 10     1
## _____
```

Além do critério Mojena, o *MultivariateAnalysis* realiza outros tipos de corte, como Cindex, Frey, Mcclain e Dunn. Para a utilização desses cortes, basta adicionar o argumento corte seguido do nome do método de corte que se deseja, e este entre aspas duplas.

## 9. MÉTODOS DE AGRUPAMENTO POR OTIMIZAÇÃO

### Agrupamento Tocher

O método de agrupamento Tocher (CRUZ & CARNEIRO, 2006) se baseia em formação de grupos em que as distâncias dentro deles sejam menores que a distâncias entre eles (FARIA *et al.*, 2012). Ou seja, esse método é um procedimento utilizado para agrupar em grupos homogêneos, baseando-se na distância entre eles.

Com esse método, observa-se homogeneidade dentro do grupo e apresenta-se heterogeneidade fora dos grupos. Inicialmente, os tratamentos mais similares formam o primeiro grupo e, para a inclusão de tratamentos dentro desse grupo, é verificado um critério de agrupamento.

### Agrupamento Tocher Modificado

O método de agrupamento chamado de Tocher Modificado ou Tocher Invertido foi proposto por Vasconcelos *et al.* (2007) e tem diferença em relação ao original. A diferença está em adotar um critério diferente para incluir um tratamento em cada grupo em formação durante o processo de agrupamento, de forma sequencial. Para incluir um tratamento a um grupo, compara-se o acréscimo ocorrido no valor da média das distâncias do grupo a um valor máximo da medida de dissimilaridade encontrado nos valores das distâncias de cada tratamento.

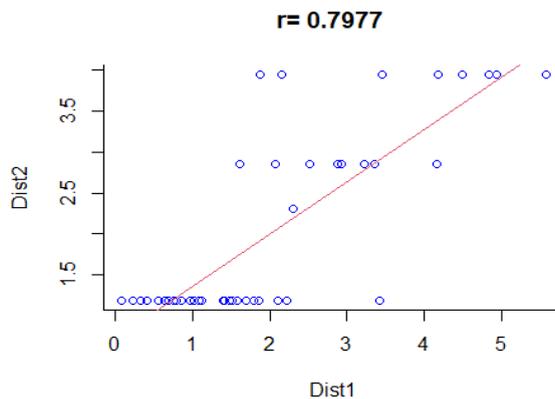
### Análise em R

Para a obtenção desse agrupamento, inicia-se de maneira igual ao método original. Primeiro, com a determinação do critério de agrupamento, seguido pela inclusão do tratamento ao primeiro grupo, para o terceiro passo; diferindo do método original, é escolhido um novo critério de agrupamento.

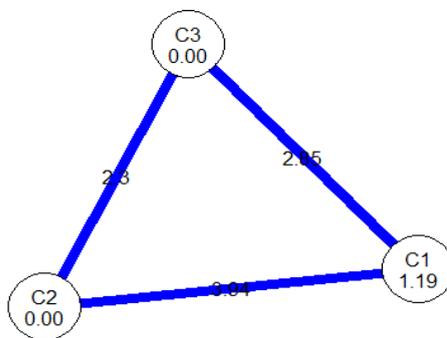
```

library(MultivariateAnalysis)
?MultivariateAnalysis
data("Dados.MED")
Dist=Distancia(Dados.MED,Metodo=6)
Tocher(Dist)

```



**Distancia intra e intercluster**



```

##
## Agrupamento Tocher
## Cluster1:
## 2 4 7 9 10 1 3 5
##
## Cluster2:
## 6
##
## Cluster3:
## 8
##
## Distancia intra e intercluster:
##      Cluster1 Cluster2 Cluster3
## Cluster1 1.193409 3.939134 2.845845
## Cluster2 3.939134 0.000000 2.304709
## Cluster3 2.845845 2.304709 0.000000
##
##
## Correlacao Cofenetica: 0.7977306
## pvalor: 0.01 baseado no teste Mantel
## Hipotese alternativa: A correlacao e maior que 0
##

```

### 9.1.1 Distância Intra e Intercluster no agrupamento Tocher

#### 9.1.1.1 Distância Intercluster

Essa representa a distância entre dois tratamentos pertencentes a dois *clusters* diferentes.

#### 9.1.1.2 Distância Intracluster

Representa a distância entre dois tratamentos pertencentes ao mesmo grupo, ou seja, ao mesmo cluster.

### Análise em R

Os comandos para a realização de distância inter e intracluster dentro do *MultivariateAnalysis* está dentro da rotina para a obtenção de dendrogramas.

### Método K-means

O método K-Means ou K-Means Clustering ou C-Means (DU, 2010) é um algoritmo bastante difundido na academia, com aplicação na indústria, com o qual é possível o agrupamento de objetos semelhantes, caracterizado por ser um algoritmo de agrupamento não supervisionado, que agrupa de acordo com a similaridade.

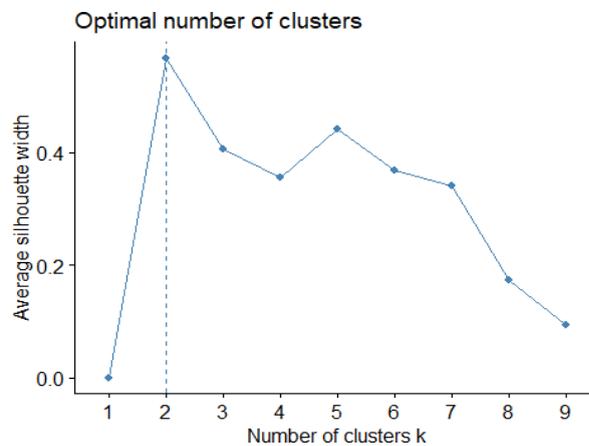
Para a realização do método, previamente define-se a constante “k”, o número de *clusters* e, a partir destes, são feitos centroides iniciais. A partir da posição inicial dos centroides, são formados grupos associados a todos os pontos dos centroides mais próximos e a partir desse é calculado o centroide de cada grupo obtido, sendo esse processo repetido até que os centroides mudem.

### Análises em R

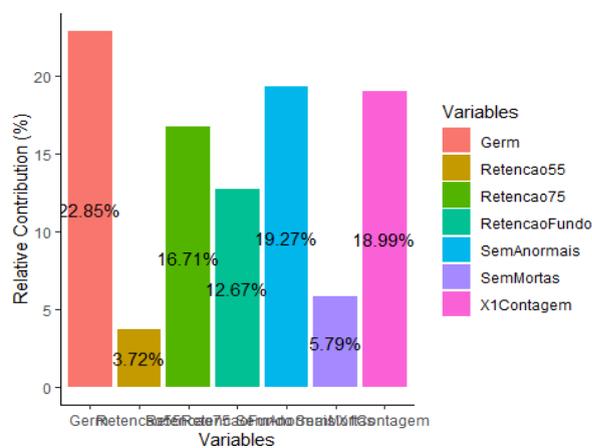
A realização do método K-Means no *MultivariateAnalysis* se inicia com a importação dos dados, seguida da função *Kmeans\_NumeroOtimo* com os argumentos de identificação dos dados, o design correspondente ao delineamento, o número de *cluster* com o argumento *nclusters*, o número de iterações permitidas com *iter.max* e o método de agrupamento desejado com o argumento *algorithm*. Um exemplo de execução do método K-Means em DIC dentro do *MultivariateAnalysis*:

```
#####
library(MultivariateAnalysis)
?MultivariateAnalysis
#Dados sem repeticao
data("Dados.MED")
```

```
Dados=Dados.MED
rownames(Dados)=paste("Genotipo",1:10,sep="_")
Kmeans_NumeroOtimo(Dados,design=1,nboot=100,method="silhouette",NumMax=NULL)
```



```
Kmeans(Dados,design=1,nclusters=3,iter.max = 10,nstart = 1,
algorithm = "Hartigan-Wong")
```



```
## $model
## K-means clustering with 3 clusters of sizes 3, 5, 2
##
## Cluster means:
##   X1Contagem   Germ SemAnormais SemMortas Retencao75 Retencao55 RetencaoFundo
## 1      85.5 94.66667    2.236801   1.408027   16.17678    5.954656    1.103553
## 2      87.9 91.60000    2.723609   1.423190   16.38422    6.468867    1.500462
## 3      82.0 87.00000    3.206335   1.820386   16.78241    6.975948    1.103553
##
## Clustering vector:
##   Genotipo_1 Genotipo_2 Genotipo_3 Genotipo_4 Genotipo_5 Genotipo_6
##           2           1           1           1           2           3
##   Genotipo_7 Genotipo_8 Genotipo_9 Genotipo_10
##           2           3           2           2
##
## Within cluster sum of squares by cluster:
## [1] 3.148080 18.359945 4.735592
## (between_SS / total_SS = 82.6 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"       "withinss"    "tot.withinss"
## [6] "betweenss"    "size"        "iter"        "ifault"
##
## $predict
```

```

##          Trat Cluster
## 1  Genotipo_1      2
## 2  Genotipo_2      1
## 3  Genotipo_3      1
## 4  Genotipo_4      1
## 5  Genotipo_5      2
## 6  Genotipo_6      3
## 7  Genotipo_7      2
## 8  Genotipo_8      3
## 9  Genotipo_9      2
## 10 Genotipo_10     2
##
## $Classe
##  Genotipo_1  Genotipo_2  Genotipo_3  Genotipo_4  Genotipo_5  Genotipo_6
##          2          1          1          1          2          3
##  Genotipo_7  Genotipo_8  Genotipo_9  Genotipo_10
##          2          3          2          2
##
## $Anova
## $Anova[[1]]
## Analysis of Variance Table
##
## Response: Dados[, i]
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Between_Clusters  2   50.8  25.4000  12.978 0.004416 **
## Residuals         7   13.7   1.9571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## $Anova[[2]]
## Analysis of Variance Table
##
## Response: Dados[, i]
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Between_Clusters  2 70.533  35.267  63.845 3.2e-05 ***
## Residuals         7   3.867   0.552
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## $Anova[[3]]
## Analysis of Variance Table
##
## Response: Dados[, i]
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Between_Clusters  2  1.15249  0.57625  13.943 0.003619 **
## Residuals         7  0.28931  0.04133
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## $Anova[[4]]
## Analysis of Variance Table
##
## Response: Dados[, i]
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Between_Clusters  2  0.26013  0.13007   1.1072 0.3821
## Residuals         7  0.82233  0.11748
##
## $Anova[[5]]
## Analysis of Variance Table
##
## Response: Dados[, i]
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Between_Clusters  2  0.44317  0.22159   7.8997 0.01604 *
## Residuals         7  0.19635  0.02805
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

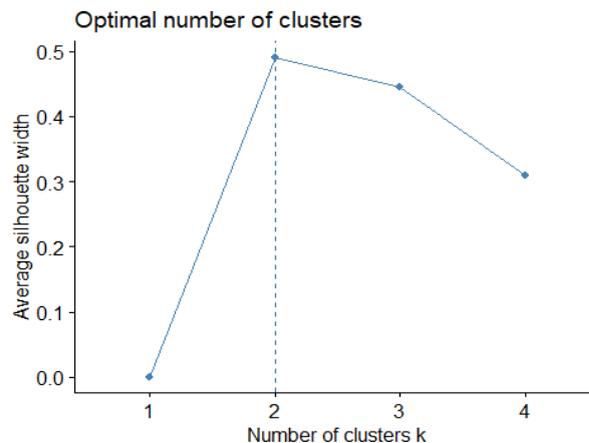
```

```

##
## $Anova[[6]]
## Analysis of Variance Table
##
## Response: Dados[, i]
##           Df Sum Sq Mean Sq F value Pr(>F)
## Between_Clusters  2  1.2796  0.63979  0.6385 0.5563
## Residuals        7  7.0136  1.00195
##
## $Anova[[7]]
## Analysis of Variance Table
##
## Response: Dados[, i]
##           Df Sum Sq Mean Sq F value Pr(>F)
## Between_Clusters  2  0.39384  0.196920  3.8792 0.07349 .
## Residuals        7  0.35535  0.050764
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## $Explication
## [1] "82.6%"
##
## $RelativeContribution
##      X1Contagem      Germ      SemAnormais      SemMortas      Retencao75
##      18.986274      22.853737      19.269447      5.793191      16.705260
##      Retencao55 RetencaoFundo
##      3.719455      12.672637
##
## attr("class")
## [1] "Kmeans"

#Dados de experimento em dic
data("Dados.DIC")
Kmeans_NumeroOtimo(Dados=Dados.DIC,design=2,nboot=100,method="silhouette",NumMax=NULL)

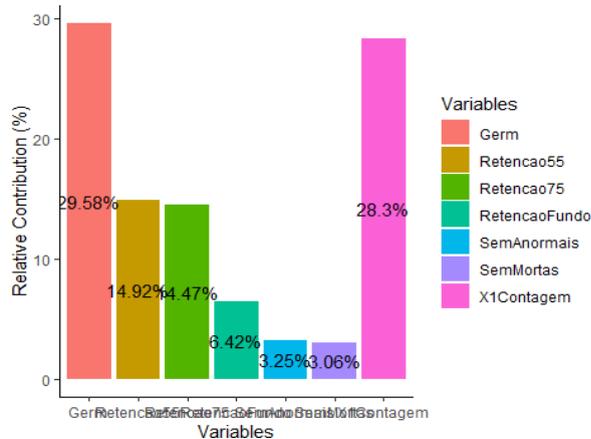
```



```

Kmeans(Dados=Dados.DIC,design=2,nclusters=2,iter.max = 20,nstart = 1,
algorithm = "Hartigan-Wong")

```



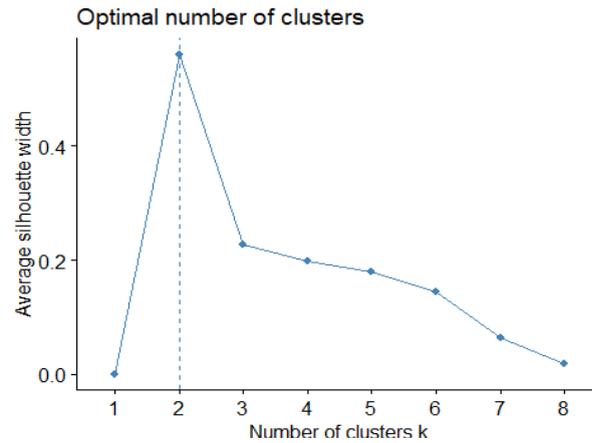
```
## $model
## K-means clustering with 2 clusters of sizes 2, 3
##
## Cluster means:
##   X1Contagem   Germ SemAnormais SemMortas Retencao75 Retencao55 RetencaoFundo
## 1  82.50000 89.50000   2.913917  1.546924  16.44147   5.738562   1.156526
## 2  87.66667 94.11111   2.683229  1.860425  16.28646   7.021071   1.391431
##
## Clustering vector:
## Trat1 Trat2 Trat3 Trat4 Trat5
##   1    2    1    2    2
##
## Within cluster sum of squares by cluster:
## [1] 7.503887 16.129102
## (between_SS / total_SS = 71.7 %)
##
## Available components:
##
## [1] "cluster"      "centers"        "totss"          "withinss"      "tot.withinss"
## [6] "betweenss"     "size"           "iter"           "ifault"
##
## $predict
##   Trat Cluster
## 1 Trat1      1
## 2 Trat2      2
## 3 Trat3      1
## 4 Trat4      2
## 5 Trat5      2
##
## $Classe
## Trat1 Trat2 Trat3 Trat4 Trat5
##   1    2    1    2    2
##
## $Anova
## $Anova[[1]]
##           GL          SQ          QM      Fc pValue
## Treatments    4 257.866667  64.466667  0.8319 0.5176
## Within_Cluster 1  1   0.333333  0.333333  0.0043 0.9482
## Within_Cluster 2  2  65.333333  32.666667  0.4215 0.6606
## Between_Clusters 1    192.2    192.2  2.4802 0.1279
## Residuals    25 1937.333333  77.493333
##
## $Anova[[2]]
##           GL          SQ          QM      Fc pValue
## Treatments    4 196.533333  49.133333  1.1113 0.3733
## Within_Cluster 1  1  40.333333  40.333333  0.9122 0.3487
## Within_Cluster 2  2   3.111111  1.555556  0.0352 0.9655
## Between_Clusters 1 153.088889 153.088889  3.4625 0.0746
## Residuals    25 1105.333333  44.213333
```

```

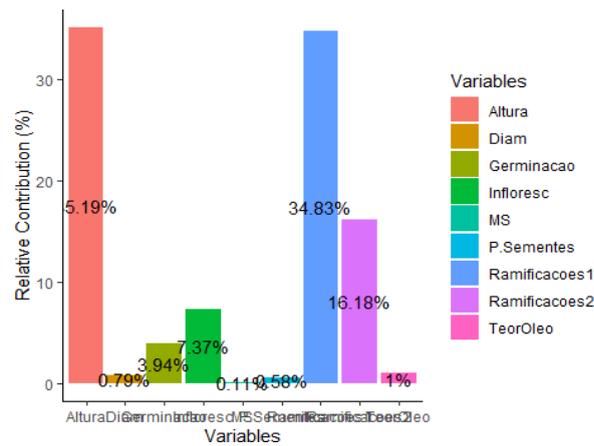
##
## $Anova[[3]]
##          GL          SQ          QM          Fc pValue
## Treatments      4      4.47591 1.118978 1.1165 0.371
## Within_Cluster 1 1 0.001831 0.001831 0.0018 0.9665
## Within_Cluster 2 2 4.090915 2.045458 2.041 0.151
## Between_Clusters 1 0.383164 0.383164 0.3823 0.542
## Residuals      25 25.054704 1.002188
##
## $Anova[[4]]
##          GL          SQ          QM          Fc pValue
## Treatments      4      8.76916 2.19229 0.5655 0.6899
## Within_Cluster 1 1 2.166637 2.166637 0.5589 0.4617
## Within_Cluster 2 2 5.894885 2.947442 0.7603 0.478
## Between_Clusters 1 0.707639 0.707639 0.1825 0.6729
## Residuals      25 96.921519 3.876861
##
## $Anova[[5]]
##          GL          SQ          QM          Fc pValue
## Treatments      4 0.454139 0.113535 0.5276 0.7165
## Within_Cluster 1 1 0.038947 0.038947 0.181 0.6742
## Within_Cluster 2 2 0.242183 0.121091 0.5627 0.5767
## Between_Clusters 1 0.173009 0.173009 0.804 0.3785
## Residuals      25 5.379782 0.215191
##
## $Anova[[6]]
##          GL          SQ          QM          Fc pValue
## Treatments      4 30.140936 7.535234 1.4698 0.2412
## Within_Cluster 1 1 2.057384 2.057384 0.4013 0.5322
## Within_Cluster 2 2 16.240787 8.120394 1.5839 0.2251
## Between_Clusters 1 11.842765 11.842765 2.31 0.1411
## Residuals      25 128.170197 5.126808
##
## $Anova[[7]]
##          GL          SQ          QM          Fc pValue
## Treatments      4 2.350548 0.587637 3.6957 0.017
## Within_Cluster 1 1 0.091855 0.091855 0.5777 0.4543
## Within_Cluster 2 2 1.861395 0.930697 5.8533 0.0082
## Between_Clusters 1 0.397298 0.397298 2.4987 0.1265
## Residuals      25 3.975093 0.159004
##
##
## $Explication
## [1] "71.7%"
##
## $RelativeContribution
##      X1Contagem      Germ      SemAnormais      SemMortas      Retencao75
##      28.302592      29.578455      3.250659      3.064233      14.465982
## Retencao55 RetencaoFundo ##
14.919848      6.418230 ##
## attr(,"class")
## [1] "Kmeans"

#Dados de experimento em dbc
data("Dados.DBC")
Kmeans_NumeroOtimo(Dados=Dados.DBC,design=3,nboot=100,method="silhouette",NumMax=NULL)

```



```
Kmeans(Dados=Dados.DBC,design=3,nclusters=2,iter.max = 20,nstart = 1,
algorithm = "Hartigan-Wong")
```



```
## $model
## K-means clustering with 2 clusters of sizes 8, 1
##
## Cluster means:
##      Altura      MS      Diam Ramificacoes1 Ramificacoes2 Infloresc P.Sementes
## 1  77.94975 16.3585 8.18175      26.996      5.95425  10.62925    6.952
## 2 103.86600 16.5940 7.65000      34.622      3.33200   9.06600    7.580
##      Germinacao TeorOleo
## 1      6.35 26.14061
## 2      3.80 26.91610
##
## Clustering vector:
## Variedade1 Variedade2 Variedade3 Variedade4 Variedade5 Variedade6 Variedade7
##          1          1          1          1          1          1          1
## Variedade8 Variedade9
##          2          1
##
## Within cluster sum of squares by cluster:
## [1] 377.7279  0.0000
## (between_SS / total_SS = 63.7 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"      "withinss"    "tot.withinss"
## [6] "betweenss"    "size"        "iter"      "ifault"
##
## $predict
##      Trat Cluster
## 1 Variedade1    1
```

```

## 2 Variedade2      1
## 3 Variedade3      1
## 4 Variedade4      1
## 5 Variedade5      1
## 6 Variedade6      1
## 7 Variedade7      1
## 8 Variedade8      2
## 9 Variedade9      1
##
## $Classe
## Variedade1 Variedade2 Variedade3 Variedade4 Variedade5 Variedade6 Variedade7
##          1          1          1          1          1          1          1
## Variedade8 Variedade9
##          2          1
##
## $Anova
## $Anova[[1]]
##          GL          SQ          QM          Fc pValue
## Treatments      8 3930.50132 491.312665 6.2492 1e-04
## Within_Cluster 1 7 945.381258 135.054465 1.7178 0.14
## Between_Clusters 1 2985.120062 2985.120062 37.9691 0
## Block           4 683.768324 170.942081 2.1743 0.0943
## Residuals      32 2515.828836 78.619651
##
## $Anova[[2]]
##          GL          SQ          QM          Fc pValue
## Treatments      8 101.29776 12.66222 1.5873 0.1677
## Within_Cluster 1 7 101.05127 14.435896 1.8096 0.1196
## Between_Clusters 1 0.24649 0.24649 0.0309 0.8616
## Block           4 24.537498 6.134374 0.769 0.5534
## Residuals      32 255.276062 7.977377
##
## $Anova[[3]]
##          GL          SQ          QM          Fc pValue
## Treatments      8 73.32528 9.16566 12.8257 0
## Within_Cluster 1 7 72.068578 10.295511 14.4068 0
## Between_Clusters 1 1.256702 1.256702 1.7585 0.1942
## Block           4 2.878236 0.719559 1.0069 0.4184
## Residuals      32 22.868164 0.71463
##
## $Anova[[4]]
##          GL          SQ          QM          Fc pValue
## Treatments      8 343.80368 42.97546 1.638 0.153
## Within_Cluster 1 7 85.33312 12.190446 0.4646 0.8527
## Between_Clusters 1 258.47056 258.47056 9.8516 0.0036
## Block           4 133.794089 33.448522 1.2749 0.3003
## Residuals      32 839.567631 26.236488
##
## $Anova[[5]]
##          GL          SQ          QM          Fc pValue
## Treatments      8 87.533444 10.941681 2.8859 0.0154
## Within_Cluster 1 7 56.972577 8.13894 2.1467 0.0667
## Between_Clusters 1 30.560867 30.560867 8.0606 0.0078
## Block           4 22.600902 5.650226 1.4903 0.2283
## Residuals      32 121.324178 3.791381
##
## $Anova[[6]]
##          GL          SQ          QM          Fc pValue
## Treatments      8 68.253591 8.531699 1.4833 0.2022
## Within_Cluster 1 7 57.392478 8.198925 1.4255 0.2296
## Between_Clusters 1 10.861114 10.861114 1.8883 0.1789
## Block           4 34.488667 8.622167 1.4991 0.2258
## Residuals      32 184.053653 5.751677
##
## $Anova[[7]]
##          GL          SQ          QM          Fc pValue

```

```

## Treatments      8 140.140498 17.517562 3.9148 0.0026
## Within_Cluster 1 7 138.38768 19.769669 4.4181 0.0016
## Between_Clusters 1 1.752818 1.752818 0.3917 0.5358
## Block          4 4.11028 1.02757 0.2296 0.9197
## Residuals     32 143.18908 4.474659
##
## $Anova[[8]]
##           GL      SQ      QM      Fc pValue
## Treatments 8      339.6    42.45 5.3164 3e-04
## Within_Cluster 1 7      310.7 44.385714 5.5588 3e-04
## Between_Clusters 1 1      28.9    28.9 3.6194 0.0661
## Block      4 43.688889 10.922222 1.3679 0.267
## Residuals 32 255.511111 7.984722
##
## $Anova[[9]]
##           GL      SQ      QM      Fc pValue
## Treatments 8 124.025501 15.503188 5.3301 3e-04
## Within_Cluster 1 7 121.352691 17.336099 5.9603 2e-04
## Between_Clusters 1 1 2.672809 2.672809 0.9189 0.345
## Block      4 9.653775 2.413444 0.8298 0.5162
## Residuals 32 93.075755 2.908617
##
##
## $Explication
## [1] "63.7%"
##
## $RelativeContribution
##           Altura      MS      Diam Ramificacoes1 Ramificacoes2
## 35.1891707 0.1127443 0.7940980 34.8333897 16.1765845
## Infloresc P.Sementes Germinacao TeorOleo
## 7.3729967 0.5795199 3.9429869 0.9985093
##
## attr(,"class")
## [1] "Kmeans"

```

## 10. COMPONENTES PRINCIPAIS

Componentes principais são uma técnica para a obtenção da combinação linear de um grupo de  $p$  variáveis dependentes (respostas), a fim de obter  $p$  variáveis latentes (componentes) não correlacionadas, sendo os primeiros componentes principais os que concentram maior variação explicada. A análise de componentes principais consiste em transformar um conjunto original de variáveis em outro de dimensão equivalente, mas com propriedades importantes, que são de grande interesse.

Quando as variáveis respostas possuem diferentes escalas de medida, recomenda-se fazer a padronização dos dados. Quando se trabalha com massa, ou diâmetro, quantidade ou porcentagem, cada um tem a sua unidade de medida e isso pode fazer com que, dependendo da magnitude da variabilidade dos conjuntos de dados de uma determinada variável resposta, esta poderá ter maior peso do que as outras, apenas pelo fato de apresentar uma escala de medida diferente. Para evitar esse problema, deve-se realizar a padronização dos dados.

Para haver relevância no uso dos componentes principais, deve existir dependência entre as variáveis respostas, ou seja, as correlações devem ser diferentes de zero, com magnitudes consideráveis. A técnica de componentes principais não requer o uso de

repetições. Além disso, só faz sentido usar essa técnica se houver mais de duas variáveis respostas.

Em geral, os primeiros componentes principais em estudo de dissimilaridade têm sido utilizados quando retêm pelo menos 80% da variação total. Se os dois primeiros componentes somados à variação de cada um forem maiores que 80%, então pode-se estudar apenas esses dois componentes principais e desconsiderar os demais. Outras situações são quando os dois primeiros componentes não forem 80% da variância total, em que se pode incluir mais um componente principal, e caso a variação dos três primeiros seja superior ou igual a 80%.

A importância de cada variável resposta pode ser calculada considerando-se a correlação entre seus valores e os dos primeiros componentes principais. Por exemplo: no caso da utilização de dois componentes principais com dez variáveis respostas, pode ser feita a correlação de cada uma das variáveis respostas medidas originalmente com o primeiro e segundo componente principal. Quanto maior for essa correlação, mais importante é a variável resposta.

### Algoritmo

Como exemplo, considerem-se sete variáveis analisadas em 25 acessos de jabuticabeira, como mostra a Tabela 3:

**Tabela 3.** Acessos de Jabuticabeira

Acessos	SSTPolpa	Acidezpolpa	Ratiopolpa	MassaFruto	DiamLat	DiamTran
A1	16,10	0,69	23,22	3,86	2,16	18,27
A2	15,10	0,55	28,14	3,84	5,11	18,26
A3	13,90	0,67	20,72	3,65	5,29	18,29
A4	15,36	0,42	36,00	3,93	3,47	17,91
A5	13,60	0,67	20,17	3,67	2,59	16,62
A6	14,03	0,77	18,08	3,73	2,32	15,93
A7	15,06	0,58	25,90	3,53	2,21	15,38
A8	14,13	0,70	20,07	3,72	2,24	15,40
A9	14,53	0,90	16,07	3,47	5,28	20,69
A10	13,43	1,07	12,59	3,44	5,97	21,57
A11	13,16	1,52	8,730	3,33	6,51	22,56
A12	13,00	0,93	13,95	3,34	4,17	18,82
A13	12,60	1,24	10,16	3,11	3,56	17,97
A14	14,90	0,82	18,20	3,39	2,30	15,58
A15	14,90	1,13	13,12	3,23	4,77	18,96
A16	14,86	0,85	17,53	3,60	2,99	15,23
A17	14,76	1,33	11,02	3,41	4,07	20,21

Acessos	SSTPolpa	Acidezpolpa	Ratiopolpa	MassaFruto	DiamLat	DiamTran
A18	15,23	0,69	22,00	3,49	4,86	18,73
A19	16,16	1,07	15,10	3,53	4,82	17,81
A20	15,00	0,81	18,39	3,61	4,77	19,62
A21	15,00	0,97	15,38	3,60	3,80	19,48
A22	15,66	0,73	21,40	3,53	4,60	19,44
A23	15,30	0,82	19,20	3,73	2,98	17,89
A24	13,13	0,56	23,79	3,39	2,11	19,02
A25	13,56	0,66	20,44	3,33	3,70	16,62

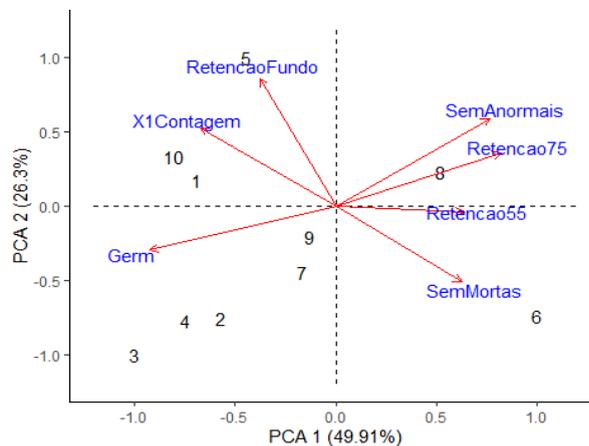
Como se pode observar, os dados apresentam escalas diferentes. Caso não seja feita a padronização dos dados, algumas variáveis podem ter maior relevância que as outras, apenas devido à escala de medida. Já quando se realiza a padronização, retira-se a escala das variáveis. Após o processo de padronização, obtém-se uma tabela, em que cada variável apresenta média com valor de zero e desvio padrão igual a um.

Acessos	SSTPolpa	Acidezpolpa	Ratiopolpa	MassaFruto	DiamLat	DiamTran
A1	1,58	-0,59	0,74	1,59	-1,31	0,01
A2	1,08	-1,13	1,55	1,49	0,95	0,00
A3	-0,62	-0,67	0,32	0,55	1,09	0,02
A4	0,84	-1,62	2,86	1,94	-0,30	-0,18
A5	-0,91	-0,67	0,23	0,65	-0,98	-0,84
A6	-0,49	-0,29	-0,12	0,95	-1,18	-1,20
A7	0,54	-1,01	1,18	-0,04	-1,27	-1,48
A8	-0,39	-0,56	0,21	0,90	-1,25	-1,47
A9	0,01	0,21	-0,45	-0,34	1,08	1,26
A10	-1,08	0,85	-1,03	-0,49	1,61	1,72
A11	-1,35	2,56	-1,67	-1,03	2,03	2,23
A12	-1,51	0,32	-0,80	-0,98	0,23	0,29
A13	-1,91	1,50	-1,43	-2,12	-0,23	-0,15
A14	0,38	-0,10	-0,10	-0,73	-1,20	-1,38
A15	0,38	1,08	-0,94	-1,53	0,69	0,37
A16	0,34	0,02	-0,21	0,31	-0,67	-1,56
A17	0,24	1,84	-1,29	-0,64	0,16	1,01
A18	0,71	-0,59	0,54	-0,24	0,76	0,25
A19	1,63	0,85	-0,61	-0,04	0,73	-0,23
A20	0,48	-0,14	-0,06	0,35	0,69	0,71
A21	0,48	0,47	-0,56	0,31	-0,05	0,64

Acessos	SSTPolpa	Acidezpolpa	Ratiopolpa	MassaFruto	DiamLat	DiamTran
A22	1,14	-0,44	0,44	-0,04	0,56	0,62
A23	0,78	-0,10	0,07	0,95	-0,68	-0,19
A24	-1,38	-1,09	0,83	-0,73	-1,35	0,40
A25	-0,95	-0,71	0,28	-1,03	-0,13	-0,84
Média	0,00	0,00	0,00	0,00	0,00	0,00
SD	1	1	1	1	1	1

Para trabalhar com componentes principais, necessita-se de um conjunto de dados em que nas linhas tem-se os indivíduos, ou seja, os tratamentos, e nas colunas, as variáveis respostas. Dentro do pacote ativado com o comando `?ComponentesPrincipais`, pode ser consultado o manual do pacote. Dentro dos parênteses, devem ser utilizados os argumentos necessários.

```
library(MultivariateAnalysis)
?MultivariateAnalysis
data(Dados.MED)
ComponentesPrincipais(Dados.MED)
```



```
## $`Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## $`Autovetor da matriz de covariancia`
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,] -0.3587 0.3871 -0.5429 0.0533 -0.2356 0.2281 0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838 0.1067
## [3,] 0.4091 0.4345 0.2058 0.0466 0.4720 -0.3574 0.4988
## [4,] 0.3344 -0.3781 -0.2023 0.6802 -0.3444 -0.2953 0.1892
## [5,] 0.4385 0.2653 0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,] 0.3411 -0.0281 -0.7842 -0.2093 0.3362 -0.1151 -0.3127
## [7,] -0.2012 0.6319 -0.0085 0.4338 -0.0868 -0.2750 -0.5374
##
## $`Escores dos componentes principais`
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,] -1.1330 0.8300 0.9098 1.6047 -0.2636 -0.1154 0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615 1.1314 0.1047 0.4035 0.1094 0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326 0.0327 0.0559 -0.0734
## [5,] -0.4205 2.5876 -0.5824 -0.3718 0.5429 -0.0251 -0.0141
```

```

## [6,] 3.8556 -1.1019 -0.5778 0.9658 0.1362 0.0621 -0.0567
## [7,] 0.3980 -0.4825 -1.0258 -0.5582 -0.3213 0.0263 0.1676
## [8,] 2.4308 0.9543 1.8224 -1.0289 -0.2246 0.0160 -0.0071
## [9,] 0.5147 0.0240 -0.4226 0.1525 0.3173 -0.0825 0.0558
## [10,] -1.4816 1.1705 -0.7204 0.2445 -0.4270 0.1826 -0.0654
##
## `$Correlacao entre as variaveis e os comp. principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## X1Contagem -0.6705 0.5252 -0.5137 0.0426 -0.0794 0.0270 0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929 0.0083
## SemAnormais 0.7646 0.5894 0.1947 0.0372 0.1590 -0.0424 0.0386
## SemMortas  0.6250 -0.5129 -0.1914 0.5428 -0.1160 -0.0350 0.0146
## Retencao75 0.8197 0.3599 0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55 0.6375 -0.0381 -0.7421 -0.1670 0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761 0.8573 -0.0080 0.3461 -0.0292 -0.0326 -0.0416
##
## `$Explicacao dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## Autovalor   3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
## % Explicacao 49.9080 26.2955 12.7910 9.0972 1.6220 0.2009 0.0854
## % Explicacao Acumulada 49.9080 76.2035 88.9946 98.0917 99.7137 99.9146 100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"

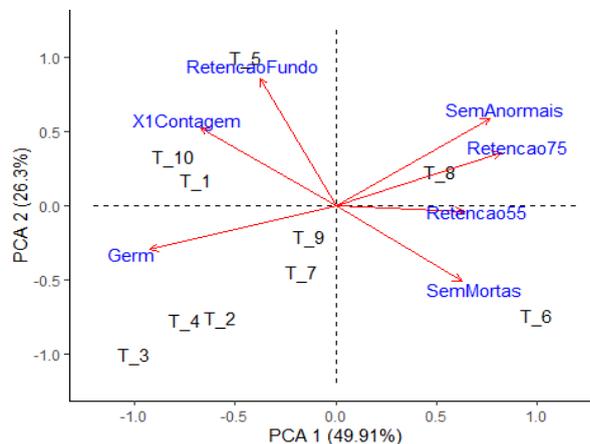
```

*#Atribuindo nome aos tratamentos*

```

Trat=paste("T_",1:nrow(Dados.MED),sep="")
ComponentesPrincipais(Dados.MED,NomeTrat = Trat)

```



```

## `$Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## `$Autovetor da matriz de covariancia`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587 0.3871 -0.5429 0.0533 -0.2356 0.2281 0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838 0.1067
## [3,] 0.4091 0.4345 0.2058 0.0466 0.4720 -0.3574 0.4988
## [4,] 0.3344 -0.3781 -0.2023 0.6802 -0.3444 -0.2953 0.1892
## [5,] 0.4385 0.2653 0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,] 0.3411 -0.0281 -0.7842 -0.2093 0.3362 -0.1151 -0.3127
## [7,] -0.2012 0.6319 -0.0085 0.4338 -0.0868 -0.2750 -0.5374
##
## `$Escores dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330 0.8300 0.9098 1.6047 -0.2636 -0.1154 0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615 1.1314 0.1047 0.4035 0.1094 0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326 0.0327 0.0559 -0.0734
## [5,] -0.4205 2.5876 -0.5824 -0.3718 0.5429 -0.0251 -0.0141

```

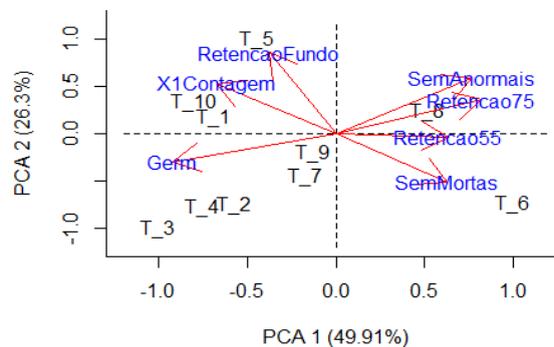
```

## [6,] 3.8556 -1.1019 -0.5778 0.9658 0.1362 0.0621 -0.0567
## [7,] 0.3980 -0.4825 -1.0258 -0.5582 -0.3213 0.0263 0.1676
## [8,] 2.4308 0.9543 1.8224 -1.0289 -0.2246 0.0160 -0.0071
## [9,] 0.5147 0.0240 -0.4226 0.1525 0.3173 -0.0825 0.0558
## [10,] -1.4816 1.1705 -0.7204 0.2445 -0.4270 0.1826 -0.0654
##
## `$Correlacao entre as variaveis e os comp. principais`
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## X1Contagem -0.6705 0.5252 -0.5137 0.0426 -0.0794 0.0270 0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929 0.0083
## SemAnormais 0.7646 0.5894 0.1947 0.0372 0.1590 -0.0424 0.0386
## SemMortas  0.6250 -0.5129 -0.1914 0.5428 -0.1160 -0.0350 0.0146
## Retencao75 0.8197 0.3599 0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55 0.6375 -0.0381 -0.7421 -0.1670 0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761 0.8573 -0.0080 0.3461 -0.0292 -0.0326 -0.0416
##
## `$Explicacao dos componentes principais`
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## Autovalor  3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
## % Explicacao 49.9080 26.2955 12.7910 9.0972 1.6220 0.2009 0.0854
## % Explicacao Acumulada 49.9080 76.2035 88.9946 98.0917 99.7137 99.9146 100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"

```

Dentro do pacote *MultivariateAnalysis*, é possível realizar algumas alterações com a finalidade de personalizar a análise. Pode-se adicionar o nome dos tratamentos:

```
ComponentesPrincipais(Dados.MED, NomeTrat = Trat, layout=1)
```



```

## `$Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## `$Autovetor da matriz de covariancia`
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,] -0.3587 0.3871 -0.5429 0.0533 -0.2356 0.2281 0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838 0.1067
## [3,] 0.4091 0.4345 0.2058 0.0466 0.4720 -0.3574 0.4988
## [4,] 0.3344 -0.3781 -0.2023 0.6802 -0.3444 -0.2953 0.1892
## [5,] 0.4385 0.2653 0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,] 0.3411 -0.0281 -0.7842 -0.2093 0.3362 -0.1151 -0.3127
## [7,] -0.2012 0.6319 -0.0085 0.4338 -0.0868 -0.2750 -0.5374
##
## `$Escores dos componentes principais`
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,] -1.1330 0.8300 0.9098 1.6047 -0.2636 -0.1154 0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815

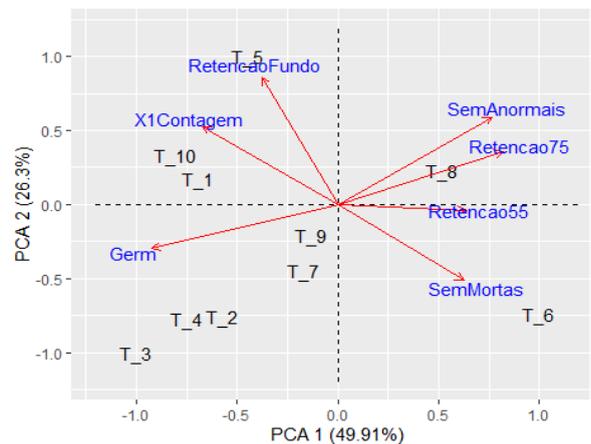
```

```

## [3,] -2.0573 -1.6615 1.1314 0.1047 0.4035 0.1094 0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326 0.0327 0.0559 -0.0734
## [5,] -0.4205 2.5876 -0.5824 -0.3718 0.5429 -0.0251 -0.0141
## [6,] 3.8556 -1.1019 -0.5778 0.9658 0.1362 0.0621 -0.0567
## [7,] 0.3980 -0.4825 -1.0258 -0.5582 -0.3213 0.0263 0.1676
## [8,] 2.4308 0.9543 1.8224 -1.0289 -0.2246 0.0160 -0.0071
## [9,] 0.5147 0.0240 -0.4226 0.1525 0.3173 -0.0825 0.0558
## [10,] -1.4816 1.1705 -0.7204 0.2445 -0.4270 0.1826 -0.0654
##
## `$Correlacao entre as variaveis e os comp. principais`
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## X1Contagem -0.6705 0.5252 -0.5137 0.0426 -0.0794 0.0270 0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929 0.0083
## SemAnormais 0.7646 0.5894 0.1947 0.0372 0.1590 -0.0424 0.0386
## SemMortas   0.6250 -0.5129 -0.1914 0.5428 -0.1160 -0.0350 0.0146
## Retencao75  0.8197 0.3599 0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55  0.6375 -0.0381 -0.7421 -0.1670 0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761 0.8573 -0.0080 0.3461 -0.0292 -0.0326 -0.0416
##
## `$Explicacao dos componentes principais`
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## Autovalor      3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
## % Explicacao   49.9080 26.2955 12.7910 9.0972 1.6220 0.2009 0.0854
## % Explicacao Acumulada 49.9080 76.2035 88.9946 98.0917 99.7137 99.9146 100.0000
##
## attr("class")
## [1] "ComponentesPrincipais"

```

**ComponentesPrincipais(Dados.MED, NomeTrat = Trat, layout=2)**



```

## `$Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## `$Autovetor da matriz de covariancia`
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,] -0.3587 0.3871 -0.5429 0.0533 -0.2356 0.2281 0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838 0.1067
## [3,] 0.4091 0.4345 0.2058 0.0466 0.4720 -0.3574 0.4988
## [4,] 0.3344 -0.3781 -0.2023 0.6802 -0.3444 -0.2953 0.1892
## [5,] 0.4385 0.2653 0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,] 0.3411 -0.0281 -0.7842 -0.2093 0.3362 -0.1151 -0.3127
## [7,] -0.2012 0.6319 -0.0085 0.4338 -0.0868 -0.2750 -0.5374
##
## `$Escores dos componentes principais`
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## [1,] -1.1330 0.8300 0.9098 1.6047 -0.2636 -0.1154 0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615 1.1314 0.1047 0.4035 0.1094 0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326 0.0327 0.0559 -0.0734

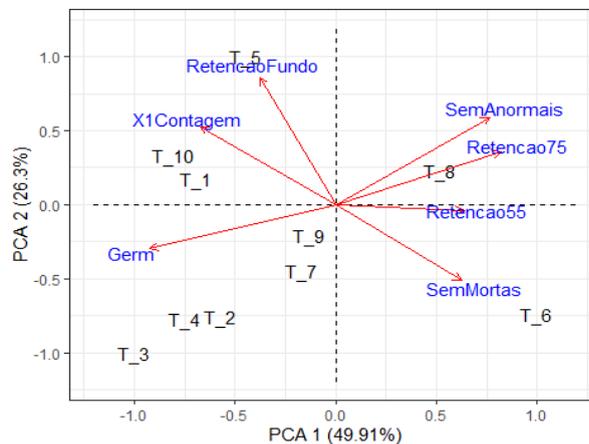
```

```

## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071
## [9,]  0.5147  0.0240 -0.4226  0.1525  0.3173 -0.0825  0.0558
## [10,] -1.4816  1.1705 -0.7204  0.2445 -0.4270  0.1826 -0.0654
##
## `$Correlacao entre as variaveis e os comp. principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais  0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas   0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75  0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55  0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## `$Explicacao dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## Autovalor   3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao 49.9080 26.2955 12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080 76.2035 88.9946 98.0917 99.7137 99.9146 100.0000
##
## attr("class")
## [1] "ComponentesPrincipais"

```

**ComponentesPrincipais(Dados.MED, NomeTrat = Trat, layout=3)**



```

## `$Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## `$Autovetor da matriz de covariancia`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## `$Escores dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567

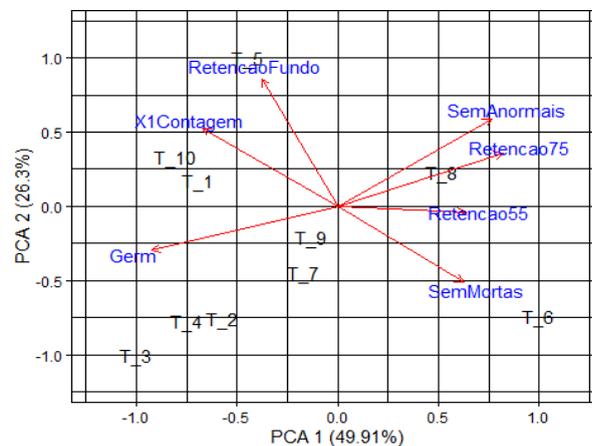
```

```

## [7,] 0.3980 -0.4825 -1.0258 -0.5582 -0.3213 0.0263 0.1676
## [8,] 2.4308 0.9543 1.8224 -1.0289 -0.2246 0.0160 -0.0071
## [9,] 0.5147 0.0240 -0.4226 0.1525 0.3173 -0.0825 0.0558
## [10,] -1.4816 1.1705 -0.7204 0.2445 -0.4270 0.1826 -0.0654
##
## `$Correlacao entre as variaveis e os comp. principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais 0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas   0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75  0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55  0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## `$Explicacao dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao    49.9080  26.2955  12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080  76.2035  88.9946  98.0917  99.7137  99.9146  100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"

```

```
ComponentesPrincipais(Dados.MED, NomeTrat = Trat, layout=4)
```



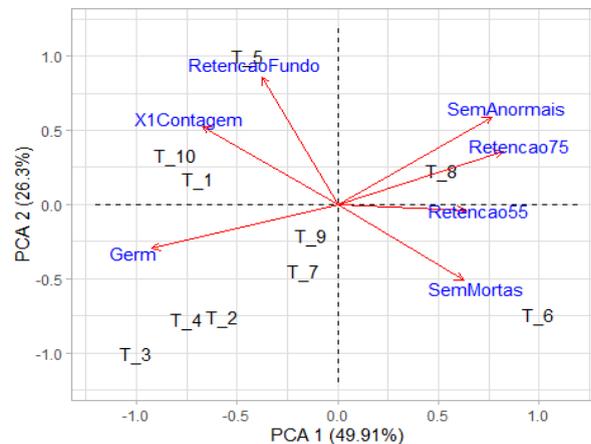
```

## `$Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## `$Autovetor da matriz de covariancia`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## `$Escores dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071

```

```
## [9,] 0.5147 0.0240 -0.4226 0.1525 0.3173 -0.0825 0.0558
## [10,] -1.4816 1.1705 -0.7204 0.2445 -0.4270 0.1826 -0.0654
##
## `$Correlacao entre as variaveis e os comp. principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## X1Contagem -0.6705 0.5252 -0.5137 0.0426 -0.0794 0.0270 0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929 0.0083
## SemAnormais 0.7646 0.5894 0.1947 0.0372 0.1590 -0.0424 0.0386
## SemMortas   0.6250 -0.5129 -0.1914 0.5428 -0.1160 -0.0350 0.0146
## Retencao75  0.8197 0.3599 0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55  0.6375 -0.0381 -0.7421 -0.1670 0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761 0.8573 -0.0080 0.3461 -0.0292 -0.0326 -0.0416
##
## `$Explicacao dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## Autovalor      3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
## % Explicacao    49.9080 26.2955 12.7910 9.0972 1.6220 0.2009 0.0854
## % Explicacao Acumulada 49.9080 76.2035 88.9946 98.0917 99.7137 99.9146 100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"
```

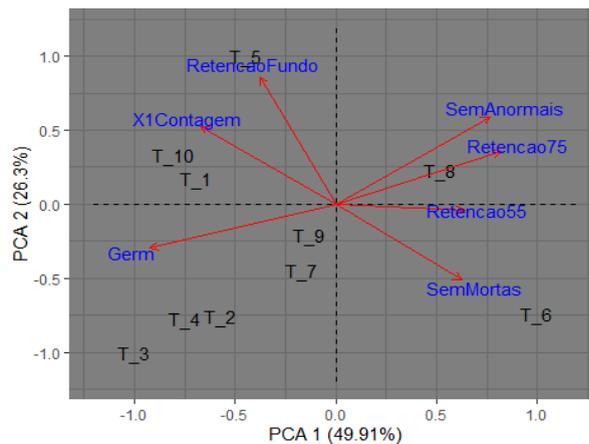
**ComponentesPrincipais(Dados.MED, NomeTrat = Trat, layout=5)**



```
## `$Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## `$Autovetor da matriz de covariancia`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587 0.3871 -0.5429 0.0533 -0.2356 0.2281 0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838 0.1067
## [3,] 0.4091 0.4345 0.2058 0.0466 0.4720 -0.3574 0.4988
## [4,] 0.3344 -0.3781 -0.2023 0.6802 -0.3444 -0.2953 0.1892
## [5,] 0.4385 0.2653 0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,] 0.3411 -0.0281 -0.7842 -0.2093 0.3362 -0.1151 -0.3127
## [7,] -0.2012 0.6319 -0.0085 0.4338 -0.0868 -0.2750 -0.5374
##
## `$Escores dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330 0.8300 0.9098 1.6047 -0.2636 -0.1154 0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615 1.1314 0.1047 0.4035 0.1094 0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326 0.0327 0.0559 -0.0734
## [5,] -0.4205 2.5876 -0.5824 -0.3718 0.5429 -0.0251 -0.0141
## [6,] 3.8556 -1.1019 -0.5778 0.9658 0.1362 0.0621 -0.0567
## [7,] 0.3980 -0.4825 -1.0258 -0.5582 -0.3213 0.0263 0.1676
## [8,] 2.4308 0.9543 1.8224 -1.0289 -0.2246 0.0160 -0.0071
## [9,] 0.5147 0.0240 -0.4226 0.1525 0.3173 -0.0825 0.0558
## [10,] -1.4816 1.1705 -0.7204 0.2445 -0.4270 0.1826 -0.0654
```

```
##
## $`Correlacao entre as variaveis e os comp. principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais  0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas   0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75  0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55  0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## $`Explicacao dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao   49.9080 26.2955 12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080 76.2035 88.9946 98.0917 99.7137 99.9146 100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"
```

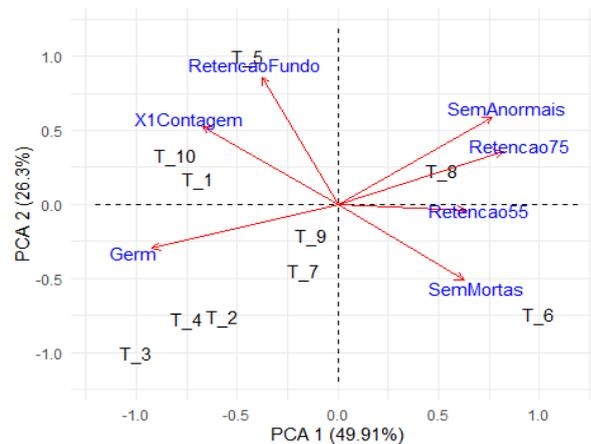
```
ComponentesPrincipais(Dados.MED, NomeTrat = Trat, layout=6)
```



```
## $`Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## $`Autovetor da matriz de covariancia`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## $`Escores dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071
## [9,]  0.5147  0.0240 -0.4226  0.1525  0.3173 -0.0825  0.0558
## [10,] -1.4816  1.1705 -0.7204  0.2445 -0.4270  0.1826 -0.0654
##
## $`Correlacao entre as variaveis e os comp. principais`
```

```
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais 0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas  0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75 0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55 0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## $`Explicacao dos componentes principais`
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao    49.9080  26.2955  12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080  76.2035  88.9946  98.0917  99.7137  99.9146  100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"
```

```
ComponentesPrincipais(Dados.MED, NomeTrat = Trat, layout=7)
```



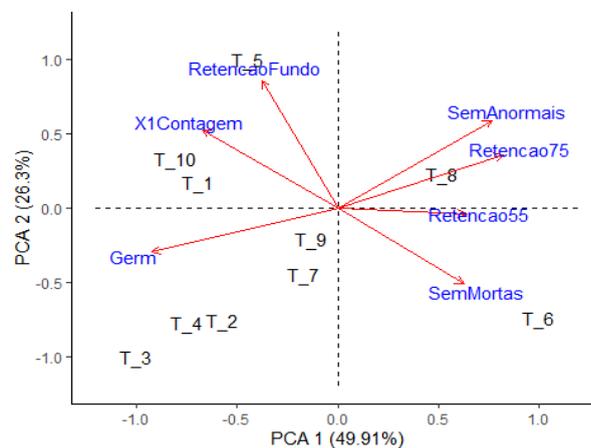
```
## $`Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## $`Autovetor da matriz de covariancia`
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## $`Escores dos componentes principais`
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071
## [9,]  0.5147  0.0240 -0.4226  0.1525  0.3173 -0.0825  0.0558
## [10,] -1.4816  1.1705 -0.7204  0.2445 -0.4270  0.1826 -0.0654
##
## $`Correlacao entre as variaveis e os comp. principais`
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
```

```

## Germ          -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais   0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas     0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75    0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55    0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## $`Explicacao dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao   49.9080 26.2955 12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080 76.2035 88.9946 98.0917 99.7137 99.9146 100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"

```

**ComponentesPrincipais(Dados.MED, NomeTrat = Trat, layout=8)**



```

## $`Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## $`Autovetor da matriz de covariancia`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## $`Escores dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071
## [9,]  0.5147  0.0240 -0.4226  0.1525  0.3173 -0.0825  0.0558
## [10,] -1.4816  1.1705 -0.7204  0.2445 -0.4270  0.1826 -0.0654
##
## $`Correlacao entre as variaveis e os comp. principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais  0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386

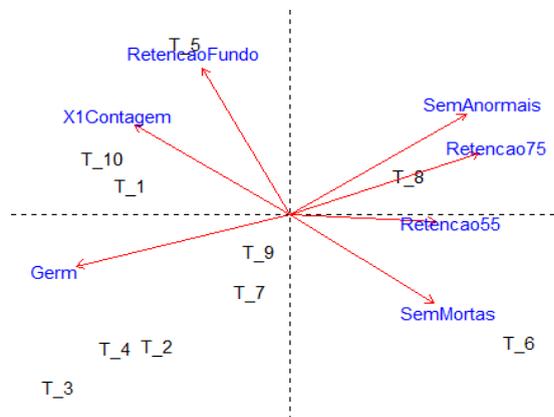
```

```

## SemMortas      0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75     0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55     0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## $`Explicacao dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao    49.9080  26.2955  12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080  76.2035  88.9946  98.0917  99.7137  99.9146  100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"

```

**ComponentesPrincipais(Dados.MED, NomeTrat = Trat, layout=9)**



```

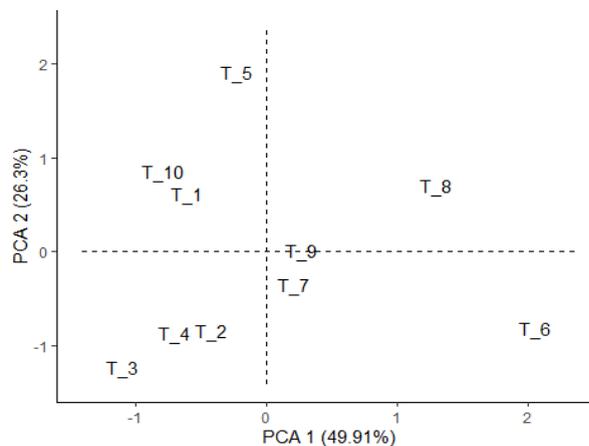
## $`Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## $`Autovetor da matriz de covariancia`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## $`Escores dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071
## [9,]  0.5147  0.0240 -0.4226  0.1525  0.3173 -0.0825  0.0558
## [10,] -1.4816  1.1705 -0.7204  0.2445 -0.4270  0.1826 -0.0654
##
## $`Correlacao entre as variaveis e os comp. principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais  0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas   0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75  0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025

```

```
## Retencao55      0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## $`Explicacao dos componentes principais`
##           [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao    49.9080  26.2955  12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080  76.2035  88.9946  98.0917  99.7137  99.9146  100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"
```

Outra personalização possível de ser realizada é a retirada das setas que indicam as correlações das variáveis com os componentes principais. Para isso, basta colocar a função `CorPlot = FALSE` e aparecerá apenas a dispersão, como mostra a rotina abaixo. Podem ser alteradas as bordas do gráfico com o argumento `bty`, podendo ser consultadas no manual do pacote.

```
ComponentesPrincipais(Dados.MED, NomeTrat = Trat, CorPlot = FALSE)
```



```
## $`Autovalor da matriz de covariância`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## $`Autovetor da matriz de covariância`
##           [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## $`Escores dos componentes principais`
##           [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071
## [9,]  0.5147  0.0240 -0.4226  0.1525  0.3173 -0.0825  0.0558
```

```

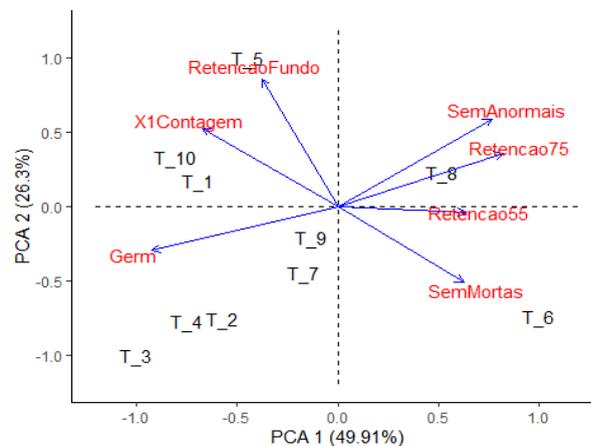
## [10,] -1.4816  1.1705 -0.7204  0.2445 -0.4270  0.1826 -0.0654
##
## $`Correlacao entre as variaveis e os comp. principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## X1Contagem  -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ        -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais  0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas    0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75   0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55   0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## $`Explicacao dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao   49.9080 26.2955 12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080 76.2035 88.9946 98.0917 99.7137 99.9146 100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"

```

```

ComponentesPrincipais(Dados.MED, NomeTrat = Trat, CorPlot = TRUE,
  CorCol = "blue", VarCol="red" )

```



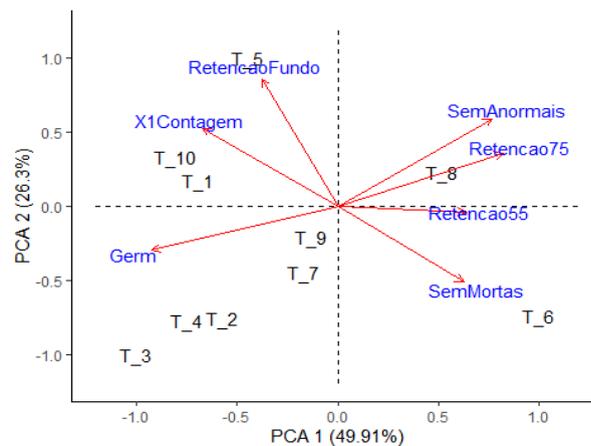
```

## $`Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## $`Autovetor da matriz de covariancia`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## $`Escores dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071
## [9,]  0.5147  0.0240 -0.4226  0.1525  0.3173 -0.0825  0.0558
## [10,] -1.4816  1.1705 -0.7204  0.2445 -0.4270  0.1826 -0.0654

```

```
##
## $`Correlacao entre as variaveis e os comp. principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais  0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas   0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75  0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55  0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## $`Explicacao dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao   49.9080 26.2955 12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080 76.2035 88.9946 98.0917 99.7137 99.9146 100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"
```

```
ComponentesPrincipais(Dados.MED, NomeTrat = Trat, CorPlot = TRUE, bty = "n")
```



```
## $`Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## $`Autovetor da matriz de covariancia`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## $`Escores dos componentes principais`
##           [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071
## [9,]  0.5147  0.0240 -0.4226  0.1525  0.3173 -0.0825  0.0558
## [10,] -1.4816  1.1705 -0.7204  0.2445 -0.4270  0.1826 -0.0654
##
## $`Correlacao entre as variaveis e os comp. principais`
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais 0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas  0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75 0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55 0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## $`Explicacao dos componentes principais`
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao    49.9080  26.2955  12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080  76.2035  88.9946  98.0917  99.7137  99.9146  100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"
```

O estudo das correlações pode ser apresentado por meio de uma tabela que também é gerada nessa mesma rotina de componentes principais, dentro do *MultivariateAnalysis*. Como dito anteriormente, é importante a padronização dos dados, mas caso não seja necessário, pode-se adicionar um argumento retirando a padronização, com o argumento: *padronizar=FALSE*.

Para mudar o nome dos eixos, usa-se a função *xlab=""* para x e *ylab=""* para Y. Para a contribuição relativa, utiliza-se o argumento *CR=*, em que aparecerá para *TRUE* e para *FALSE* não aparecerá.

## 11. COORDENADAS PRINCIPAIS

As análises de coordenadas principais (PCoA), ou escalonamento multidimensional métrico ou clássico (MD), assemelham-se às análises de Componentes Principais (PCA), mas com algumas diferenças. Dentro da PCoA, é possível usar qualquer coeficiente de similaridade. Para a PCA, é possível utilizar apenas distância euclidiana e dados quantitativos. Outra diferença está nos objetivos: para a PCoA: o objetivo principal é a exploração de dados para ilustrar os padrões e gerar hipóteses.

Por apresentar uma flexibilidade, a PCoA é indicada para uma quantidade de aplicação maior de ordenação, com o objetivo de preservar as distâncias multivariadas originais entre as observações no espaço da ordenação.

Dentre as vantagens da PCoA, podemos citar quatro principais. A primeira é que essa pode ser realizada quando se tem apenas a matriz de dissimilaridade. A segunda vantagem corresponde à adequação quando o número de variáveis é maior que o número de amostras. A terceira vantagem está relacionada à robustez para os valores ausentes. E, por último, a PCoA é útil para analisar variações sazonais e gradientes de diversidade ou quando existem poucas unidades amostrais.

Existem algumas desvantagens na utilização da análise PCoA. Dentre elas, citam-se o fato de não informar quais variáveis influenciam a distribuição dos objetos e de não fornecer a relação entre as variáveis e os eixos principais, apresentando apenas para as unidades amostrais. Outra desvantagem é a impossibilidade de interpretar os eixos com base na projeção dos escritores ou em subconjuntos.

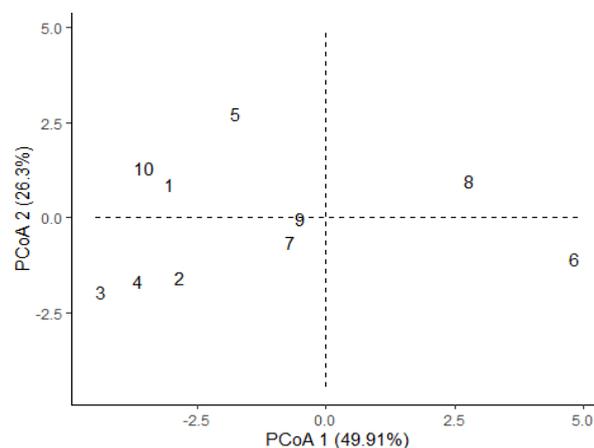
Para verificar se a PCoA é a análise mais adequada, observa-se se foi produzido autovalores negativos e altos. Quando isso acontece, é possível que a representação cartesiana esteja distorcida, indicando que a matriz de distância utilizada pode não ser apropriada.

### Análise em R

Para a realização de PCoA dentro do *MultivariateAnalysis*, deve-se iniciar com sua ativação com a função *library*. Os dados necessários para execução devem ser uma matriz contendo informações para calcular a distância ou o coeficiente que se desejar. A função que calcula as distâncias dentro do *MultivariateAnalysis* é conhecida como *Distancia* e dentro dessa deve-se ser indicar qual método de distância se utilizará. Na execução de PCoA após obtenção da matriz de similaridade utiliza-se a função *CoordenadasPrincipais* composta pelo argumento de matriz de similaridade.

```
#Exemplo com dados Quantitativos
data("Dados.MED")
Dist=Distancia(Dados.MED,Metodo=5)
CoordenadasPrincipais(Dist)
```

```
## [1] 1
```



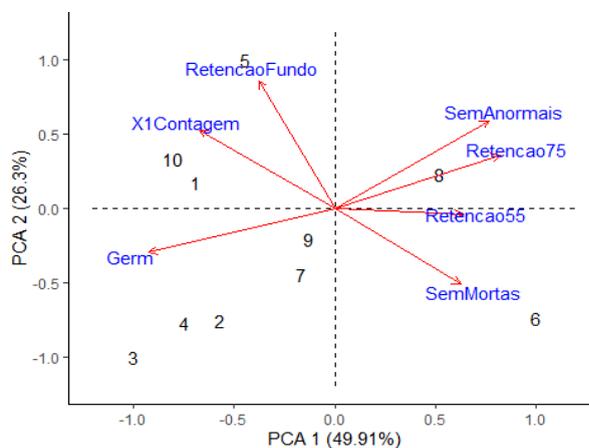
```
## $values
## [1] 4.491718e+00 2.366598e+00 1.151194e+00 8.187457e-01 1.459765e-01
## [6] 1.807946e-02 7.688523e-03 8.901369e-17 0.000000e+00 0.000000e+00
##
## $vectors
##      X1      X2      X3      X4      X5      X6
## 1  0.4282280 -0.313698426 0.34387831 0.60651615 0.09961776 0.043611731
## 2  0.2992957 0.432394896 -0.16829809 -0.21951305 0.07415144 0.086655414
```

```

## 3  0.7775888  0.627978568  0.42763702  0.03956147 -0.15250258 -0.041340385
## 4  0.4969843  0.444673656 -0.03374901 -0.20128506 -0.01235869 -0.021123670
## 5  0.1589410 -0.978031491 -0.22014316 -0.14050902 -0.20519743  0.009485604
## 6 -1.4572936  0.416471042 -0.21839937  0.36505527 -0.05149491 -0.023475614
## 7 -0.1504424  0.182350974 -0.38773011 -0.21099132  0.12143018 -0.009930074
## 8 -0.9187556 -0.360689094  0.68880938 -0.38888233  0.08489360 -0.006029379
## 9 -0.1945219 -0.009058457 -0.15971681  0.05763115 -0.11993801  0.031176053
## 10 0.5599754 -0.442391668 -0.27228816  0.09241674  0.16139864 -0.069029679
##
##          X7          X8 X9 X10
## 1  0.012349853  0.000000e+00  0  0
## 2 -0.030811467 -3.033992e-17  0  0
## 3  0.015928941 -3.043002e-17  0  0
## 4 -0.027760268  1.862443e-17  0  0
## 5 -0.005315811 -4.375586e-19  0  0
## 6 -0.021420458 -1.099523e-17  0  0
## 7  0.063343703 -2.596874e-18  0  0
## 8 -0.002664980 -2.317652e-17  0  0
## 9  0.021077355 -5.550049e-17  0  0
## 10 -0.024726868 -4.455093e-17  0  0
##
## attr(,"class")
## [1] "pcoa"

```

*#Compare os resultados com os componentes principais*  
**ComponentesPrincipais(Dados.MED, padronizar = TRUE)**



```

## $`Autovalor da matriz de covariancia`
## [1] 3.4936 1.8407 0.8954 0.6368 0.1135 0.0141 0.0060
##
## $`Autovetor da matriz de covariancia`
##      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -0.3587  0.3871 -0.5429  0.0533 -0.2356  0.2281  0.5624
## [2,] -0.4930 -0.2190 -0.0680 -0.2797 -0.0202 -0.7838  0.1067
## [3,]  0.4091  0.4345  0.2058  0.0466  0.4720 -0.3574  0.4988
## [4,]  0.3344 -0.3781 -0.2023  0.6802 -0.3444 -0.2953  0.1892
## [5,]  0.4385  0.2653  0.0478 -0.4712 -0.6944 -0.1725 -0.0330
## [6,]  0.3411 -0.0281 -0.7842 -0.2093  0.3362 -0.1151 -0.3127
## [7,] -0.2012  0.6319 -0.0085  0.4338 -0.0868 -0.2750 -0.5374
##
## $`Escores dos componentes principais`
##      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]
## [1,] -1.1330  0.8300  0.9098  1.6047 -0.2636 -0.1154  0.0327
## [2,] -0.7919 -1.1440 -0.4453 -0.5808 -0.1962 -0.2293 -0.0815
## [3,] -2.0573 -1.6615  1.1314  0.1047  0.4035  0.1094  0.0421
## [4,] -1.3149 -1.1765 -0.0893 -0.5326  0.0327  0.0559 -0.0734
## [5,] -0.4205  2.5876 -0.5824 -0.3718  0.5429 -0.0251 -0.0141
## [6,]  3.8556 -1.1019 -0.5778  0.9658  0.1362  0.0621 -0.0567
## [7,]  0.3980 -0.4825 -1.0258 -0.5582 -0.3213  0.0263  0.1676
## [8,]  2.4308  0.9543  1.8224 -1.0289 -0.2246  0.0160 -0.0071

```

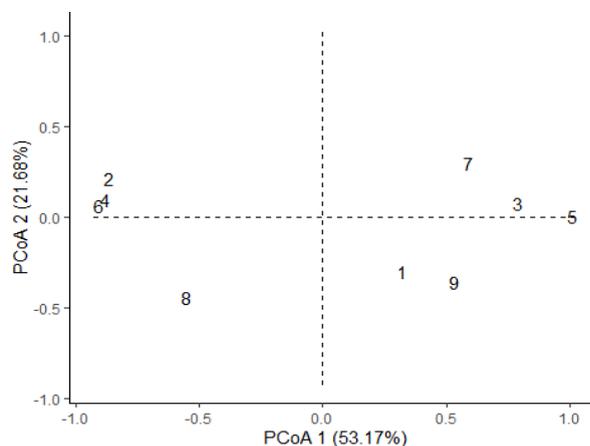
```
## [9,] 0.5147 0.0240 -0.4226 0.1525 0.3173 -0.0825 0.0558
## [10,] -1.4816 1.1705 -0.7204 0.2445 -0.4270 0.1826 -0.0654
##
## `$Correlacao entre as variaveis e os comp. principais`
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## X1Contagem -0.6705  0.5252 -0.5137  0.0426 -0.0794  0.0270  0.0435
## Germ       -0.9214 -0.2972 -0.0643 -0.2232 -0.0068 -0.0929  0.0083
## SemAnormais 0.7646  0.5894  0.1947  0.0372  0.1590 -0.0424  0.0386
## SemMortas   0.6250 -0.5129 -0.1914  0.5428 -0.1160 -0.0350  0.0146
## Retencao75  0.8197  0.3599  0.0452 -0.3760 -0.2340 -0.0205 -0.0025
## Retencao55  0.6375 -0.0381 -0.7421 -0.1670  0.1133 -0.0137 -0.0242
## RetencaoFundo -0.3761  0.8573 -0.0080  0.3461 -0.0292 -0.0326 -0.0416
##
## `$Explicacao dos componentes principais`
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## Autovalor      3.4936  1.8407  0.8954  0.6368  0.1135  0.0141  0.0060
## % Explicacao    49.9080  26.2955  12.7910  9.0972  1.6220  0.2009  0.0854
## % Explicacao Acumulada 49.9080  76.2035  88.9946  98.0917  99.7137  99.9146  100.0000
##
## attr(,"class")
## [1] "ComponentesPrincipais"
```

Além de ser possível a realização de PCoA com dados quantitativos dentro do *MultivariateAnalysis*, pode-se realizá-la com dados de natureza qualitativos. A diferença na execução para os dados qualitativos está na escolha no argumento de método de similaridade; as demais funções seguem como o modelo de dados quantitativos.

#### #Exemplo com dados Qualitativos

```
data=data("Dados.CAT")
Dist=Distancia(Dados.CAT,Metodo=10)
CoordenadasPrincipais(Dist)

## [1] 1
```



```
## $values
## [1] 9.962303e-01 4.061528e-01 2.498189e-01 1.764132e-01 3.306009e-02
## [6] 1.212331e-02 1.110223e-16 0.000000e+00 0.000000e+00
##
## $vectors
##           X1           X2           X3           X4           X5           X6
## 1 -0.1182268  0.18058233  0.19870590  0.036276834 -0.076259210  0.052287726
## 2  0.3659021 -0.20430214 -0.05968898 -0.099539812 -0.097277781 -0.054587540
## 3 -0.3605134 -0.10009657 -0.22573090  0.280049458 -0.028957970 -0.000213903
## 4  0.4096876 -0.10690338  0.17368106  0.201707361  0.059779916 -0.010718830
```

```

## 5 -0.4641318 -0.00906798 -0.15413366 -0.144476741 0.042281906 -0.008586722
## 6 0.3892343 -0.06530542 -0.17322014 -0.111552237 -0.000828563 0.069440949
## 7 -0.2574707 -0.31974735 0.19478200 -0.097739688 0.065580580 0.004498400
## 8 0.2593281 0.37783258 -0.10547513 -0.008801214 0.075475742 -0.022873745
## 9 -0.2238094 0.24700794 0.15107984 -0.055923961 -0.039794621 -0.029246335
##
##          X7 X8 X9
## 1 3.700743e-17 0 0
## 2 3.700743e-17 0 0
## 3 3.700743e-17 0 0
## 4 3.700743e-17 0 0
## 5 3.700743e-17 0 0
## 6 3.700743e-17 0 0
## 7 3.700743e-17 0 0
## 8 3.700743e-17 0 0
## 9 3.700743e-17 0 0
##
## attr(,"class")
## [1] "pcoa"

```

## 12. MANOVA – Análise de variância Multivariada

Quando se trabalha com análise multivariada e se tem delineamento estatístico, pode-se fazer a Análise Multivariada de Variância, ou MANOVA. A MANOVA pode ser compreendida como uma extensão da análise de variância, quando se apresenta mais de uma variável dependente. Nesse caso, o delineamento estatístico pode ser considerado na análise multivariada, também com a vantagem de permitir melhores condições para teste de hipóteses. Então utiliza-se a análise de variância multivariada, também conhecida como MANOVA, que possibilita também a obtenção de covariâncias/correlações entre as variáveis respostas de diferentes naturezas, sendo essas fenotípicas, genéticas ou residual, por exemplo.

Todos os conceitos e princípios básicos envolvidos para a determinação e uso de delineamentos na estatística experimental sob aspecto univariado também são válidos para a estatística multivariada. Logo, pode-se dizer que a MANOVA é uma generalização da ANOVA, para casos em que diversas variáveis respostas são consideradas simultaneamente. As mesmas fontes de variação que são consideradas na ANOVA (em caso univariado) também são presentes na MANOVA.

Considerando o teste de hipótese para MANOVA, a hipótese nula indica que os vetores de médias de tratamentos são iguais, ou não há diferença nos tratamentos do ponto de vista multivariado. Já para hipótese alternativa, pelo menos um dos vetores de médias de um dos tratamentos se difere dos demais, do ponto de vista multivariado.

### Delineamento interamente casualizado

A execução de análise de variância multivariada (MANOVA) no *MultivariateAnalysis* se inicia semelhantemente às demais funções, com importação dos dados e ativação do pacote.

Após a importação dos dados, deve-se ativar a função `MANOVA`. Nessa função, devem ser colocados os argumentos, como a identificação dos dados e o modelo de delineamento.

```
#Delineamento inteiramente casualizado (DIC)
data("Dados.DIC")
MANOVA(Dados.DIC,1)

##
## _____
## MANOVA com o teste Pillai
##          Df          Pillai          approx F num Df den Df
## Trat      4 0.923557138089452 0.943494899326211      28      88
## Residuals 25
##          Pr(>F)
## Trat      0.553532575460337
## Residuals
##
## MANOVA com o teste Wilks
##          Df          Wilks          approx F num Df          den Df
## Trat      4 0.320682382551752 0.926168578503273      28 69.9276793356718
## Residuals 25
##          Pr(>F)
## Trat      0.577118761177976
## Residuals
##
## MANOVA com o teste Hotelling
##          Df Hotelling-Lawley          approx F num Df den Df          Pr(>F)
## Trat      4 1.44731832078461 0.90457395049038      28      70 0.60555702095796
## Residuals 25
##
## MANOVA com o teste Roy
##          Df          Roy          approx F num Df den Df          Pr(>F)
## Trat      4 0.89501778044243 2.81291302424764      7      22 0.0298030163187348
## Residuals 25
##
## As medias dos tratamentos podem ser acessados com o $Med
## Os Graus de liberdade do residuo podem ser acessados com o $GLres
## A matriz de (co)variancias residuais pode ser acessada com o $CovarianciaResidual
## _____
```

## Delineamento em Blocos Casualizados

Para a realização de MANOVA no *MultivariateAnalysis* inicia-se com a importação dos dados, que devem ser tabulados em colunas, com a primeira constando os tratamentos, a segunda com o bloco e, a partir da terceira, as variáveis analisadas. Após a importação dos dados, deve-se realizar a ativação do pacote. A função de `MANOVA` para DBC assemelha-se à DIC; a diferença está no tipo do modelo trabalhado. A função `MANOVA` se inicia com a identificação dos dados seguida do modelo; para o DBC, deve-se utilizar o modelo 2.

```
#Delineamento em blocos casualizados (DBC)
data(Dados.DBC)
MANOVA(Dados.DBC,2)

##
## _____
## MANOVA com o teste Pillai
##          Df          Pillai          approx F num Df den Df
## Trat      8 3.41112687003393 2.56041879215572      72      248
## Bloco     4 1.01952256443478 1.02620058679435      36      108
```

```

## Residuals 32
##                               Pr(>F)
## Trat      3.88033554581756e-08
## Bloco     0.444201237736759
## Residuals
##
## MANOVA com o teste Wilks
##      Df      Wilks      approx F num Df      den Df
## Trat   8 0.00224676346620771 3.67964225576572    72 153.565638439245
## Bloco  4 0.283651652720517 1.01776238907217    36 91.6764243863006
## Residuals 32
##                               Pr(>F)
## Trat   6.79854759058188e-12
## Bloco  0.458726231922828
## Residuals
##
## MANOVA com o teste Hotelling
##      Df Hotelling-Lawley      approx F num Df den Df
## Trat   8 14.4192295284348 4.45594245843992    72 178
## Bloco  4 1.59818345907489 0.998864661921807    36 90
## Residuals 32
##                               Pr(>F)
## Trat   3.24533729991487e-16
## Bloco  0.485684435194678
## Residuals
##
## MANOVA com o teste Roy
##      Df      Roy      approx F num Df den Df
## Trat   8 7.04048563648178 24.2505616367706    9 31
## Bloco  4 0.79426617101982 2.38279851305946    9 27
## Residuals 32
##                               Pr(>F)
## Trat   1.21249826306718e-11
## Bloco  0.0392110037699888
## Residuals
##
## As medias dos tratamentos podem ser acessados com o $Med
## Os Graus de liberdade do residuo podem ser acessados com o $GLres
## A matriz de (co)variancias residuais pode ser acessada com o $CovarianciaResidual
##

```

### Delineamento em quadrado latino

Para a realização de MANOVA em delineamento em quadrado latino (DQL) dentro do *MultivariateAnalysis*, deve-se realizar os passos iniciais de importação, leitura de dados e ativação do pacote. Dentro da função *MANOVA*, são necessários argumentos que indiquem quais são os dados e qual o modelo escolhido. No caso DQL, o modelo deve ser o de número 3.

```

#Delineamento em quadrado Latino (DQL)
data(Dados.DQL)
MANOVA(Dados.DQL,3)

##
## MANOVA com o teste Pillai
##      Df      Pillai      approx F num Df den Df
## Trat   3 2.483024676264 6.00373113149789    12 15
## Linha  3 1.53516410078449 1.31001372031386    12 15
## Coluna 3 0.937035892208126 0.567772779388717    12 15
## Residuals 6

```

```

##                               Pr(>F)
## Trat      0.000837983402915417
## Linha     0.306388308606897
## Coluna    0.835782340165989
## Residuals
##
## MANOVA com o teste Wilks
##           Df           Wilks           approx F num Df           den Df
## Trat      3 0.000119219702606937  20.1665727415217    12 8.22875655532295
## Linha     3 0.00966848009208786   3.27354823931747    12 8.22875655532295
## Coluna    3 0.278751292746002  0.425595772150771    12 8.22875655532295
## Residuals 6
##                               Pr(>F)
## Trat      9.49213421685706e-05
## Linha     0.048192625305916
## Coluna    0.912639046667421
## Residuals
##
## MANOVA com o teste Hotelling
##           Df Hotelling-Lawley           approx F num Df den Df
## Trat      3 303.338251133409  42.1303126574179    12 5
## Linha     3 47.0030715347101  6.52820437982085    12 5
## Coluna    3 1.84092794239893  0.255684436444296    12 5
## Residuals 6
##                               Pr(>F)
## Trat      0.000322487509359036
## Linha     0.024970012391082
## Coluna    0.975414051712585
## Residuals
##
## MANOVA com o teste Roy
##           Df           Roy           approx F num Df den Df
## Trat      3 290.564302502825  363.205378128531     4 5
## Linha     3 45.8139135901858  57.2673919877323     4 5
## Coluna    3 1.3071259486546  1.63390743581825     4 5
## Residuals 6
##                               Pr(>F)
## Trat      2.40528704098824e-06
## Linha     0.00022985456951168
## Coluna    0.298875763296089
## Residuals
##
## As medias dos tratamentos podem ser acessados com o $Med
## Os Graus de liberdade do residuo podem ser acessados com o $GLres
## A matriz de (co)variancias residuais pode ser acessada com o $CovarianciaResidual
##

```

## Esquema Fatorial

No *MultivariateAnalysis*, é possível a realização de análise de variância multivariada em esquema fatorial, de forma simples e prática. Inicia-se com a introdução dos dados através da função `data`, seguida do número do esquema que se deseja utilizar. Os esquemas disponíveis são:

- Fatorial em duplo DIC;
- Fatorial em duplo DBC;
- Fatorial em triplo DIC;
- Fatorial em triplo DBC.

*#Esquema fatorial duplo em DIC*`data(Dados.Fat2.DIC)``MANOVA(Dados.Fat2.DIC,4)`

```
##
## -----
## MANOVA com o teste Pillai
##              Df              Pillai              approx F num Df den Df
## FatorA        2 0.245732967832992 0.875483048358393          8    50
## FatorB        2 0.357341009855392  1.35961348338013          8    50
## FatorA:FatorB  4 0.733589330598066  1.51595389640446         16   108
## Residuals     27
##              Pr(>F)
## FatorA        0.543379313238653
## FatorB        0.237222140971121
## FatorA:FatorB 0.107038862842308
## Residuals
##
## MANOVA com o teste Wilks
##              Df              Wilks              approx F num Df          den Df
## FatorA        2 0.757146042754169 0.895431078741142          8          48
## FatorB        2 0.671814949475775  1.32025851656713          8          48
## FatorA:FatorB  4 0.415621988192898  1.53898538514806         16 73.9588372775532
## Residuals     27
##              Pr(>F)
## FatorA        0.527767312975845
## FatorB        0.256461061241195
## FatorA:FatorB 0.108981870704289
## Residuals
##
## MANOVA com o teste Hotelling
##              Df Hotelling-Lawley              approx F num Df den Df
## FatorA        2 0.31694670923161 0.911221789040878          8    46
## FatorB        2 0.445106336836346  1.27968071840449          8    46
## FatorA:FatorB  4 1.07391891914613  1.51019848004924         16    90
## Residuals     27
##              Pr(>F)
## FatorA        0.515733765057234
## FatorB        0.277467619369256
## FatorA:FatorB 0.11347535133299
## Residuals
##
## MANOVA com o teste Roy
##              Df              Roy              approx F num Df den Df
## FatorA        2 0.304457440421171  1.90285900263232          4    25
## FatorB        2 0.300854622314435  1.88034138946522          4    25
## FatorA:FatorB  4 0.675997659238373  4.56298419985901          4    27
## Residuals     27
##              Pr(>F)
## FatorA        0.141296683533622
## FatorB        0.145276085786805
## FatorA:FatorB 0.00604697810454517
## Residuals
##
## As medias dos tratamentos podem ser acessados com o $Med
## Os Graus de liberdade do residuo podem ser acessados com o $GLres
## A matriz de (co)variancias residuais pode ser acessada com o $CovarianciaResidual
## -----
```

*#Esquema fatorial duplo em DBC*`data(Dados.Fat2.DBC)``MANOVA(Dados.Fat2.DBC,5)`

```
##
## -----
## MANOVA com o teste Pillai
##              Df              Pillai              approx F num Df den Df
## FatorA        5 1.38567191384126  4.34500723670916         30   340
```

```

## FatorB      3  1.61506974111905 12.8279146465213      18      198
## Bloco       3  0.283747716435623  1.14909240930156      18      198
## FatorA:FatorB 15  2.60997171266997  3.54152498465953      90      414
## Residuals   69
##
##                               Pr(>F)
## FatorA      7.69578454816936e-12
## FatorB      1.4408706225226e-24
## Bloco       0.307795399091048
## FatorA:FatorB 2.12066700979853e-18
## Residuals
##
## MANOVA com o teste Wilks
##              Df      Wilks      approx F num Df      den Df
## FatorA      5  0.0660773486619989  8.36232183583444      30      258
## FatorB      3  0.0355105470437555  22.73661157558      18 181.504617357995
## Bloco       3  0.729510797377149  1.18944295355249      18 181.504617357995
## FatorA:FatorB 15  0.0112233707025844  4.97629080560899      90 366.523598590873
## Residuals   69
##
##                               Pr(>F)
## FatorA      4.42318578502632e-24
## FatorB      2.83964120237213e-37
## Bloco       0.273680895339913
## FatorA:FatorB 2.83226079131227e-28
## Residuals
##
## MANOVA com o teste Hotelling
##              Df  Hotelling-Lawley      approx F num Df den Df
## FatorA      5  7.70792434537239  16.0324826383746      30      312
## FatorB      3  11.8198789958883  41.1506898375371      18      188
## Bloco       3  0.352778433490539  1.22819158326336      18      188
## FatorA:FatorB 15  10.8012525728159  7.48086752265397      90      374
## Residuals   69
##
##                               Pr(>F)
## FatorA      6.86282254581754e-47
## FatorB      2.26220194695508e-55
## Bloco       0.24213213785258
## FatorA:FatorB 4.63145836964843e-45
## Residuals
##
## MANOVA com o teste Roy
##              Df      Roy      approx F num Df den Df
## FatorA      5  6.83546069933376  77.4685545924493      6      68
## FatorB      3  10.6896240434465  117.585864477912      6      66
## Bloco       3  0.294133778212201  3.23547156033421      6      66
## FatorA:FatorB 15  7.35788643192169  33.8462775868398      15      69
## Residuals   69
##
##                               Pr(>F)
## FatorA      1.93278235492605e-28
## FatorB      2.89609270159161e-33
## Bloco       0.00755940161343434
## FatorA:FatorB 7.08733364292308e-26
## Residuals
##
## As medias dos tratamentos podem ser acessados com o $Med
## Os Graus de liberdade do residuo podem ser acessados com o $GLres
## A matriz de (co)variancias residuais pode ser acessada com o $CovarianciaResidual
##
## -----
## #Esquema fatorial triplo em DIC
## data(Dados.Fat3.DIC)
## MANOVA(Dados.Fat3.DIC,6)
##
## -----
## MANOVA com o teste Pillai
##              Df      Pillai      approx F num Df den Df
## FatorA      1  0.960206220108884  178.187486708254      13      96

```

```

## FatorB          2  1.60972929527311  30.7762199516127    26    194
## FatorC          5  1.84594999700205  4.50202607696689    65    500
## FatorA:FatorB  2  0.201977934428724    0.83817999510656    26    194
## FatorA:FatorC  5  0.426097535716481    0.716603245753442    65    500
## FatorB:FatorC 10  2.91022316363641    3.31542856311219    130   1050
## FatorA:FatorB:FatorC 10 0.626822779764576  0.540136952077645    130   1050
## Residuals     108
##
##                                     Pr(>F)
## FatorA          4.35657092173829e-61
## FatorB          3.47023168264642e-55
## FatorC          1.47724767277165e-22
## FatorA:FatorB    0.693886522276918
## FatorA:FatorC    0.95174996345531
## FatorB:FatorC    1.16339432723015e-26
## FatorA:FatorB:FatorC 0.999991052208372
## Residuals
##
## MANOVA com o teste Wilks
##                                     Df          Wilks          approx F num Df
## FatorA          1  0.0397937798911089   178.187486708288    13
## FatorB          2  0.0354862952519344   31.8164210907679    26
## FatorC          5  0.0654464357407146    5.49558666210577    65
## FatorA:FatorB  2  0.806801574282513    0.836760641463571    26
## FatorA:FatorC  5  0.628053990604712    0.728178428681124    65
## FatorB:FatorC 10  0.013201040895852    4.32772642679382   130
## FatorA:FatorB:FatorC 10 0.51411261379977  0.522989178066418   130
## Residuals     108
##                                     den Df          Pr(>F)
## FatorA          96 4.35657092170058e-61
## FatorB         192 5.93151456489571e-56
## FatorC        457.621917317145 1.1401597689199e-28
## FatorA:FatorB  192    0.695742412308276
## FatorA:FatorC  457.621917317145 0.942582581211975
## FatorB:FatorC  784 6.56067307034753e-38
## FatorA:FatorB:FatorC 784 0.999995549711553
## Residuals
##
## MANOVA com o teste Hotelling
##                                     Df Hotelling-Lawley          approx F num Df den Df
## FatorA          1 24.1295554917475   178.187486708289    13    96
## FatorB          2  8.99778666542026   32.876528200574    26   190
## FatorC          5  4.71658610457623    6.84993428110763    65   472
## FatorA:FatorB  2  0.228580264199725    0.835197119191302    26   190
## FatorA:FatorC  5  0.510279749978923    0.741083206123236    65   472
## FatorB:FatorC 10  7.45593603436092    5.40268595720614   130   942
## FatorA:FatorB:FatorC 10 0.707606462400988  0.512742528909024   130   942
## Residuals     108
##                                     Pr(>F)
## FatorA          4.35657092169898e-61
## FatorB          1.08206879703831e-56
## FatorC          1.1606855395863e-37
## FatorA:FatorB    0.697786421220441
## FatorA:FatorC    0.932207241766169
## FatorB:FatorC    1.96414739469998e-54
## FatorA:FatorB:FatorC 0.999997937195785
## Residuals
##
## MANOVA com o teste Roy
##                                     Df          Roy          approx F num Df den Df
## FatorA          1 24.1295554917475   178.187486708289    13    96
## FatorB          2  5.93344377578955   44.2726189424297    13    97
## FatorC          5  3.09334846579847   23.7949881984498    13   100
## FatorA:FatorB  2  0.160984518307523    1.20119217506383    13    97
## FatorA:FatorC  5  0.287515986468934    2.21166143437641    13   100
## FatorB:FatorC 10  3.54312885154435   28.6175791855505    13   105
## FatorA:FatorB:FatorC 10 0.196805250390139  1.58958086853574    13   105

```

```

## Residuals          108
##
##                               Pr(>F)
## FatorA              4.35657092169898e-61
## FatorB              6.5511876421458e-35
## FatorC              5.9813831498478e-25
## FatorA:FatorB      0.290213931372678
## FatorA:FatorC      0.0140376504327233
## FatorB:FatorC      1.1270345437437e-28 ##
FatorA:FatorB:FatorC 0.0996350730151482 ##
Residuals
##
## As medias dos tratamentos podem ser acessados com o $Med
## Os Graus de liberdade do residuo podem ser acessados com o $GLres
## A matriz de (co)variancias residuais pode ser acessada com o $CovarianciaResidual
##

```

### *#Esquema fatorial triplo em DBC*

**data**(Dados.Fat3.DBC)

**MANOVA**(Dados.Fat3.DBC,7)

```

##
## _____
## MANOVA com o teste Pillai
##                               Df          Pillai          approx F num Df den Df
## FatorA              1 0.192841222924299    1.95909165939844         5    41
## FatorB              3 0.307437078921249    0.981948781224196        15   129
## FatorC              1 0.0995884414832218    0.906946620618267         5    41
## Bloco              3 0.469035295609686    1.59374152284553        15   129
## FatorA:FatorB      3 0.323197389350911    1.03836477794821        15   129
## FatorA:FatorC      1 0.0106177801892459 0.0880001639492468         5    41
## FatorB:FatorC      3 0.597221231415094    2.13756782660214        15   129
## FatorA:FatorB:FatorC 3 0.788692666991692    3.06730631010984        15   129
## Residuals          45
##
##                               Pr(>F)
## FatorA              0.105340654014232
## FatorB              0.477910034239824
## FatorC              0.485946184285941
## Bloco              0.0839409058012104
## FatorA:FatorB      0.420856890370886
## FatorA:FatorC      0.993729287574112
## FatorB:FatorC      0.0117874014195729
## FatorA:FatorB:FatorC 0.000291839559606956
## Residuals
##
## MANOVA com o teste Wilks
##                               Df          Wilks          approx F num Df
## FatorA              1 0.807158777075817    1.95909165939699         5
## FatorB              3 0.712289335335364    0.990233175464823        15
## FatorC              1 0.900411558516778    0.906946620618267         5
## Bloco              3 0.586536485909877    1.61444713973575        15
## FatorA:FatorB      3 0.694611372113746    1.06854090110892        15
## FatorA:FatorC      1 0.989382219810754    0.0880001639492471         5
## FatorB:FatorC      3 0.483162331922382    2.28285598929587        15
## FatorA:FatorB:FatorC 3 0.337511608366129    3.65054218034042        15
## Residuals          45
##
##                               den Df          Pr(>F)
## FatorA              41    0.105340654014465
## FatorB             113.584345357753    0.470252833786583
## FatorC              41    0.485946184285941
## Bloco             113.584345357753    0.0805335137820936
## FatorA:FatorB      113.584345357753    0.393307294284436
## FatorA:FatorC      41    0.993729287574112
## FatorB:FatorC      113.584345357753    0.00733098925179534
## FatorA:FatorB:FatorC 113.584345357753    3.50452821426466e-05
## Residuals
##
## MANOVA com o teste Hotelling

```

```

##                               Df   Hotelling-Lawley          approx F num Df den Df
## FatorA                        1  0.23891361699949    1.95909165939582      5    41
## FatorB                        3  0.376650238213382    0.996030629942056     15   119
## FatorC                        1  0.110603246416862    0.906946620618267      5    41
## Bloco                         3  0.610838648283745    1.6153288699059      15   119
## FatorA:FatorB                 3  0.414128090548233    1.09513872833866     15   119
## FatorA:FatorC                 1  0.0107317273108838    0.0880001639492468      5    41
## FatorB:FatorC                 3  0.910010256976406    2.40647156844872     15   119
## FatorA:FatorB:FatorC         3  1.6033487731649     4.23996675570274     15   119
## Residuals                    45
##                               Pr(>F)
## FatorA                        0.105340654014654
## FatorB                        0.463959295079621
## FatorC                        0.485946184285941
## Bloco                         0.0794492388978126
## FatorA:FatorB                 0.368244581999515
## FatorA:FatorC                 0.993729287574112
## FatorB:FatorC                 0.00441647552283236
## FatorA:FatorB:FatorC         3.05761932413084e-06
## Residuals
##
## MANOVA com o teste Roy
##                               Df   Roy          approx F num Df den Df
## FatorA                        1  0.238913616998895    1.95909165939094      5    41
## FatorB                        3  0.288679385904918    2.4826427187823      5    43
## FatorC                        1  0.110603246416862    0.906946620618265     5    41
## Bloco                         3  0.347347645133278    2.98718974814619     5    43
## FatorA:FatorB                 3  0.340454489257216    2.92790860761206     5    43
## FatorA:FatorC                 1  0.0107317273108837    0.0880001639492467     5    41
## FatorB:FatorC                 3  0.706950425354478    6.07977365804851     5    43
## FatorA:FatorB:FatorC         3  1.35683726395917    11.6688004700488     5    43
## Residuals                    45
##                               Pr(>F)
## FatorA                        0.105340654015442
## FatorB                        0.0462283287242857
## FatorC                        0.485946184285942
## Bloco                         0.0211437785791049
## FatorA:FatorB                 0.0231696839439272
## FatorA:FatorC                 0.993729287574112
## FatorB:FatorC                 0.000242663727172713
## FatorA:FatorB:FatorC         3.69935841295858e-07
## Residuals
##
## As medias dos tratamentos podem ser acessados com o $Med
## Os Graus de liberdade do residuo podem ser acessados com o $GLres
## A matriz de (co)variancias residuais pode ser acessada com o $CovarianciaResidual
##

```

### 13. VARIÁVEIS DISCRIMINANTES CANÔNICAS

A técnica de análise de variáveis canônicas dentro da estatística multivariada é a que possibilita a redução da dimensão dos dados, bem como os componentes principais e correlações canônicas. Dessa forma, simplifica em poucas informações os dados de um determinado grupo de indivíduos e essas informações têm a capacidade de reter o máximo da variação originalmente disponível e de serem independentes entre si (CRUZ, 2006).

Além dessas simplificações, as variáveis canônicas podem representar uma diversidade de populações em um subespaço de menor dimensão, pois essa técnica busca, em uma base, um grande número de características originais correlacionadas, e obtém

combinações lineares dessas características chamadas variáveis canônicas, de tal maneira que a correlação entre essas variáveis seja nula (KHATTREE & NAIK, 2000).

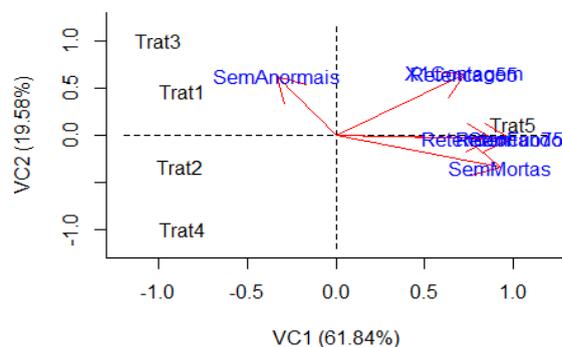
Essa técnica possibilita capturar o efeito simultâneo das variáveis originais, permitindo identificar as variações que passariam despercebidas, ao serem utilizadas características de forma isolada. Para o processamento da técnica de variáveis canônicas em delineamento inteiramente casualizado, são necessários dados que contenham, em sua primeira coluna, os tratamentos; na segunda, as repetições; e a partir da terceira, as variáveis analisadas.

## Análise em R

### 13.1.1 Variáveis Discriminantes em DIC

Dentro do *MultivariateAnalysis*, para realização de análise de variáveis canônicas, após a ativação do pacote e importação dos dados, basta utilizar a função `VariaveisCanonicas` e, nessa função, contemplar os argumentos de dados e modelo de delineamento. Para DIC, deve ser escolhido o modelo 1

```
#Delineamento inteiramente casualizado (DIC)
data("Dados.DIC")
VariaveisCanonicas(Dados.DIC,1)
```



```
## [1] "a"

## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##   CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.4723   0.8950   61.8397   61.8397
## VC2 0.2208   0.2833   19.5759   81.4156
## VC3 0.1693   0.2038   14.0820   95.4976
## VC4 0.0612   0.0652    4.5024  100.0000
##
## Escores das variaveis canonicas
##   Can1   Can2   Can3   Can4
## Trat1 -0.3856  0.2967 -0.7598  0.0394
## Trat2 -0.4078 -0.2691  0.1257 -0.4283
## Trat3 -0.5403  0.6761  0.4942  0.1129
```

```

## Trat4 -0.3897 -0.7490 0.1022 0.2716
## Trat5 1.7235 0.0453 0.0378 0.0043
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##
##          Can1      Can2 X1Contagem      Germ
## SemAnormais  0.7280  0.6537   0.2164 -0.1511
## SemMortas    0.8859 -0.0369   0.6394  0.1197
## Retencao75   -0.3330  0.6241  -0.6370 -0.6574
## Retencao55   0.9277 -0.3389   0.8989  0.5117
## RetencaoFundo 0.9882 -0.0337   0.7945  0.3764
## _____

```

### 13.1.2 Variáveis Discriminantes em DBC

A realização de variáveis canônicas para delineamento em blocos causalizados no *MultivariateAnalysis* requer, inicialmente, um conjunto de dados que contemplem colunas tratamentos e blocos nas duas primeiras colunas e nas demais as variáveis dependentes.

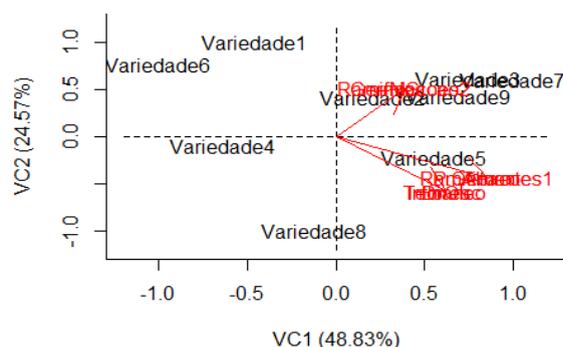
Para realizar a execução de variáveis canônicas em DBC no *MultivariateAnalysis*, após a realização sua ativação, realiza-se a importação dos dados, como mostra o item 2.5. A função execução se chama *VariaveisCanonicas*, composta pelos argumentos de dados e pelo modelo. No caso de DBC, deve ser escolhido o número 2.

Para o gráfico de variáveis canônicas, existem argumentos que podem ser colocados para edição do gráfico, por exemplo, o *layout*, constando no pacote 9 tipos, variando do número 1 ao 9. *xlab* e *ylab* correspondem aos nomes dos eixos x e y das variáveis canônicas e *cols* representa o número de variáveis canônicas que se pretende apresentar no gráfico. *CR* está relacionado ao valor lógico e, quando indicado no gráfico, aparecerá a contribuição relativa de cada eixo. *CorPlot* representa valor lógico para as correlações. Já *CorCol* é o argumento que mostra a cor das setas da dispersão gráfica da correlação, bem como *VarCol* está relacionado ao nome das variáveis. *bty* é a função que indica o tipo da borda desejada no gráfico e *length* refere-se ao tamanho da seta.

```

#Delineamento em quadrado Latino (DQL)
data(Dados.DBC)
VARIAVEISCANONICAS(Dados.DBC,2,CorCol = "red",VarCol = "red")

```



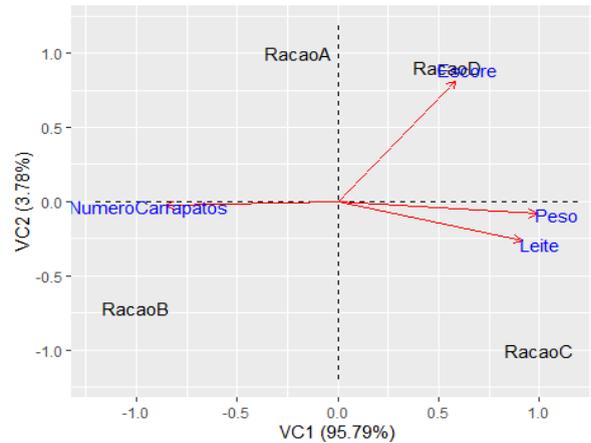
```
## [1] "a"

##
## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##   CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.8756   7.0405   48.8271   48.8271
## VC2 0.7799   3.5429   24.5708   73.3979
## VC3 0.6219   1.6445   11.4048   84.8027
## VC4 0.5861   1.4160   9.8203   94.6229
## VC5 0.3631   0.5701   3.9537   98.5766
## VC6 0.1265   0.1448   1.0041   99.5807
## VC7 0.0458   0.0480   0.3332   99.9138
## VC8 0.0123   0.0124   0.0862   100.0000
##
## Escores das variaveis canonicas
##           Can1   Can2   Can3   Can4   Can5   Can6   Can7   Can8
## Variedade1 -1.9220  1.9665  0.7113  0.0915  0.2706  0.3454 -0.0806  0.1822
## Variedade2  0.3394  0.3677 -0.7531 -0.1462  1.5286  0.0764  0.1822 -0.0743
## Variedade3  2.1634  0.9320 -2.4979  0.2026 -0.4863  0.0084 -0.1527  0.0262
## Variedade4 -2.5448 -1.0317 -0.5459 -1.2742 -0.4535 -0.3795  0.2576  0.0708
## Variedade5  1.5205 -1.3708  0.7359 -1.9637 -0.1280  0.3503 -0.1783 -0.0546
## Variedade6 -3.7660  1.3411  0.2066  0.4908 -0.4275 -0.0175 -0.1064 -0.1731
## Variedade7  3.0204  0.8747  0.9996  0.7279 -0.6974  0.1760  0.3149 -0.0281
## Variedade8 -0.7784 -3.4833 -0.0100  1.6393  0.0669  0.1367 -0.0561  0.0311
## Variedade9  1.9673  0.4037  1.1536  0.2320  0.3266 -0.6962 -0.1807  0.0198
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##           Can1   Can2  Altura   MS   Diam Ramificacoes1 Ramificacoes2
## P.Sementes  0.8577 -0.4350 -0.3177 -0.8085 -0.9331   -0.1062   0.2735
## Germinacao  0.3916  0.5167 -0.2737 -0.3785 -0.2578   -0.2084  -0.0922
## TeorOleo    0.6211 -0.5821 -0.2198 -0.7197 -0.6809   0.0471   0.3620
##
##           Infloresc
## P.Sementes  0.5100
## Germinacao  0.0005
## TeorOleo    0.5888
##
```

### 13.1.3 Variáveis Discriminantes em DQL

Bem como o exemplo em DBC, é possível a realização de variáveis canônicas em delineamento em quadrado latino (DQL). O processamento assemelha-se ao procedimento em DBC, diferindo apenas no modelo de delineamento. Para o DQL, deve ser escolhido o modelo 3.

```
#Delineamento em quadrado Latino (DQL)
data(Dados.DQL)
VariaveisCanonicas(Dados.DQL,3,layout=2)
```



```
## [1] "a"

## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##   CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.9966 290.5643 95.7889 95.7889
## VC2 0.9198 11.4662 3.7800 99.5689
## VC3 0.5667 1.3077 0.4311 100.0000
##
## Escores das variaveis canonicas
##   Can1 Can2 Can3
## RacaoA -3.9000 2.1640 0.9320
## RacaoB -14.9457 -1.7411 -0.3466
## RacaoC 12.5598 -2.3789 0.3403
## RacaoD 6.2860 1.9560 -0.9257
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##   Can1 Can2 NumeroCarrapatos
## Can1 1.0000 0.0000 -0.8519
## Can2 0.0000 1.0000 -0.0283
## NumeroCarrapatos -0.8519 -0.0283 1.0000
## Peso 0.9891 -0.0817 -0.7764
## Escore 0.5842 0.8113 -0.5104
## Leite 0.9140 -0.2614 -0.6090
## -----
```

#### 13.1.4 Variáveis Discriminantes em fatorial duplo DIC

Quando se deseja realizar a análise de variáveis canônicas em esquema fatorial, para qualquer delineamento deve-se iniciar com o formato dos dados em que, nas três primeiras colunas, constam os fatores e as repetições. Para execução de variáveis canônicas em fatorial duplo em DIC inicia-se com a função `VariaveisCanonicas` composta pelos argumentos de dados e indicando o modelo. Nesse caso, utiliza-se o número 4.

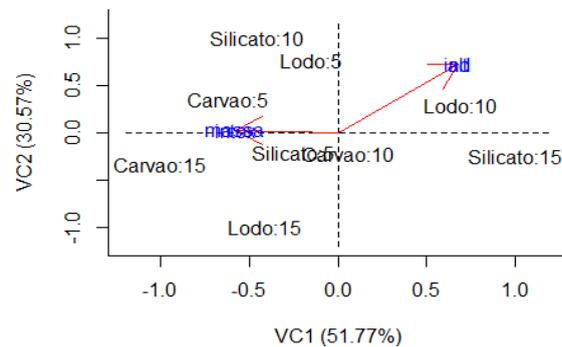
Nesse caso, deve ser adicionado um argumento indicando o que se deseja apresentar no gráfico. Por exemplo, se dentro do argumento Fator for utilizado A:B, serão apresentadas todas as combinações entre os fatores, ou seja, todos os tratamentos. Quando utilizado o argumento Fator="A", serão apresentados apenas os níveis do fator A, bem como se utilizado

Fator="B", serão demonstrados apenas os níveis do fator B. Existem vários modelos *layouts* para gráficos dentro do pacote, variando do número 1 ao 9, adequados a diversas situações.

```
#Esquema fatorial duplo em DIC
```

```
data(Dados.Fat2.DIC)
```

```
VariaveisCanonicas(Dados.Fat2.DIC,4,Fator="A:B")
```



```
## [1] "a"
```

```
##
```

```
## Estudo das variáveis canônicas
```

```
##
```

```
## Explicação das variáveis canônicas
```

##	CanRsqr	Autovalor	Porcentagem	PorcentagemAcumulada
## VC1	0.4873	0.9506	51.7737	51.7737
## VC2	0.3595	0.5612	30.5681	82.3417
## VC3	0.1943	0.2412	13.1352	95.4769
## VC4	0.0767	0.0830	4.5231	100.0000

```
##
```

```
## Escores das variáveis canônicas
```

##		Can1	Can2	Can3	Can4
## Carvão:10		0.2604	-0.3029	-0.3524	0.0606
## Carvão:15		-1.2379	-0.4202	-0.4033	0.4406
## Carvão:5		-0.7022	0.3440	-0.2345	-0.5006
## Lodo:10		1.1623	0.2705	-0.7799	-0.0479
## Lodo:15		-0.4063	-1.1664	0.3839	-0.1193
## Lodo:5		-0.0273	0.8092	0.4960	0.0956
## Silicato:10		-0.4703	1.0644	0.2438	0.1476
## Silicato:15		1.5981	-0.3218	0.4128	0.1374
## Silicato:5		-0.1767	-0.2770	0.2336	-0.2140

```
##
```

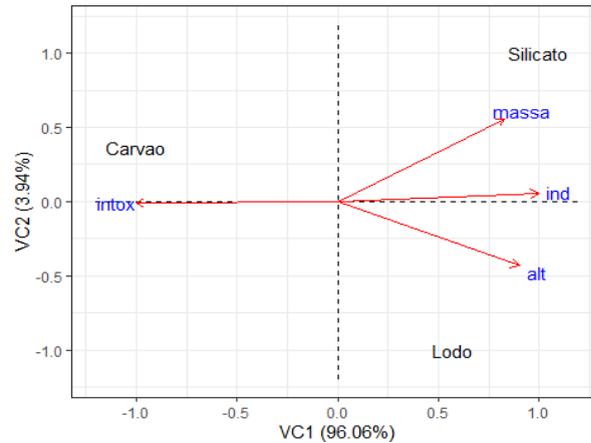
```
## Importância
```

```
## Correlação das características com os escores das variáveis canônicas
```

##		Can1	Can2	intox	alt
## massa		-0.5785	0.0259	0.1810	-0.1942
## ind		0.6752	0.7298	-0.3293	0.1359

```
##
```

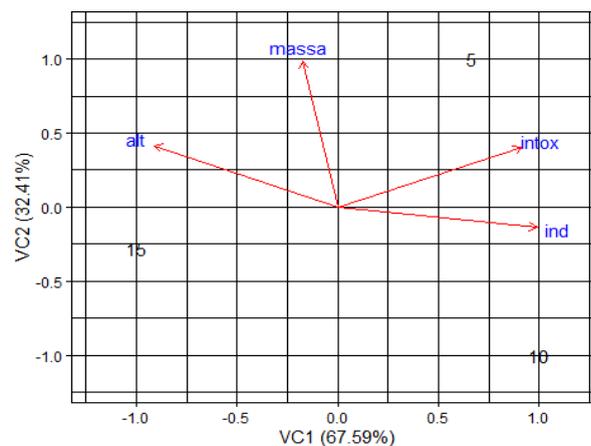
```
VariaveisCanonicas(Dados.Fat2.DIC,4,Fator="A",layout=3)
```



```
## [1] "a"
```

```
## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##   CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.2334   0.3045   96.0595       96.0595
## VC2 0.0123   0.0125    3.9405      100.0000
##
## Escores das variaveis canonicas
##           Can1   Can2
## Carvao  -0.6615  0.0280
## Lodo     0.2112 -0.1300
## Silicato 0.4503  0.1021
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##           Can1   Can2
## Can1     1.0000  0.0000
## Can2     0.0000  1.0000
## intox   -1.0000 -0.0075
## alt      0.9012 -0.4334
## massa    0.8305  0.5570
## ind      0.9983  0.0574
## -----
```

**VariaveisCanonicas**(Dados.Fat2.DIC,4,Fator="B",layout=4)



```
## [1] "a"
```

```

##
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##      CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.2313    0.3009    67.5916          67.5916
## VC2 0.1261    0.1443    32.4084          100.0000
##
## Escores das variaveis canonicas
##      Can1    Can2
## 5    0.2400  0.4345
## 10   0.4234 -0.3611
## 15  -0.6634 -0.0733
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##      Can1    Can2
## Can1  1.0000  0.0000
## Can2  0.0000  1.0000
## intox 0.9151  0.4032
## alt  -0.9092  0.4163
## massa -0.1762  0.9844
## ind   0.9904 -0.1379
##

```

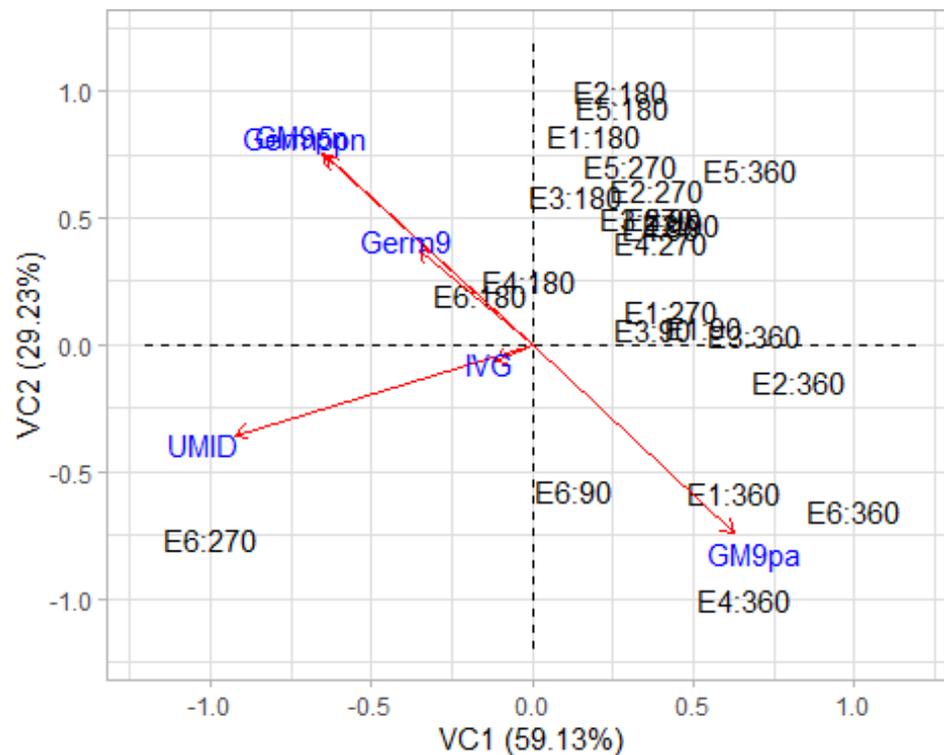
### 13.1.5 Variáveis Discriminantes em fatorial duplo DBC

A ideia de execução de dados em esquema fatorial duplo em DBC segue o mesmo processamento do exemplo em DIC, como mostra o exemplo anterior desse material. A pequena diferença está na escolha do modelo de delineamento. Para dados em DBC, deve ser escolhido o modelo de número 5.

```

#Esquema fatorial duplo em DBC
data(Dados.Fat2.DBC)
VariaveisCanonicas(Dados.Fat2.DBC,5,Fator="A:B",layout=5)

```



```
## [1] "a"
```

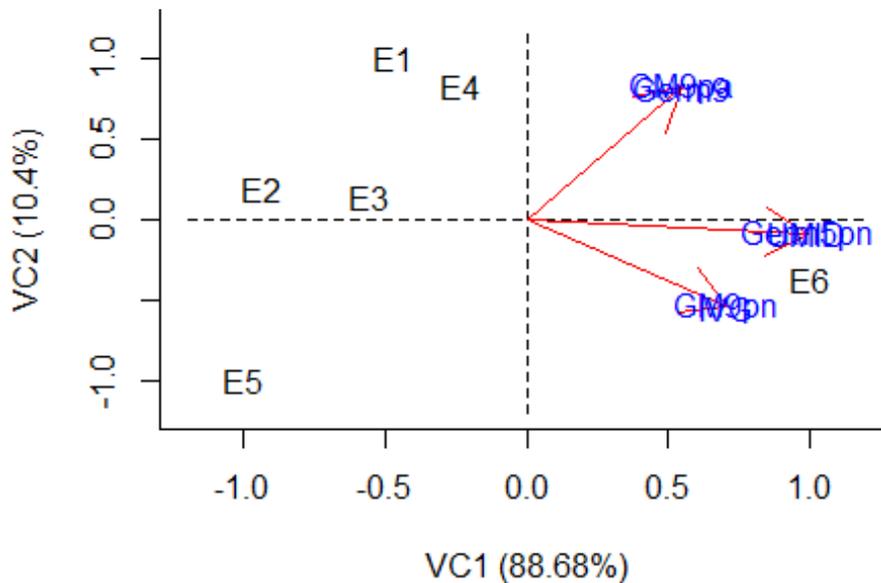
```
##
## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##   CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.9472  17.9326    59.1268      59.1268
## VC2 0.8986   8.8665    29.2342     88.3610
## VC3 0.6685   2.0166     6.6491     95.0101
## VC4 0.4153   0.7104     2.3424     97.3524
## VC5 0.3716   0.5913     1.9496     99.3021
## VC6 0.1747   0.2117     0.6979    100.0000
##
## Escores das variaveis canonicas
##           Can1   Can2   Can3   Can4   Can5   Can6
## E1:180 -1.4654  2.8846 -1.4562  0.8070 -0.0224  0.3080
## E1:270  0.7661 -0.2808  0.2903 -1.5566  0.0968 -0.6158
## E1:360  2.6463 -3.5730  0.9152  0.2667  0.4389 -0.5898
## E1:90   1.7377 -0.5884  0.8477 -0.4182  0.6153  0.8250
## E2:180 -0.7372  3.6788 -0.6842  0.0058  0.1615  0.0815
## E2:270  0.3761  1.9116 -0.7116  0.9181 -0.5818 -0.3564
## E2:360  4.5298 -1.5635  0.4726  0.6651  1.1841 -0.5001
## E2:90   1.0842  1.2902  1.4345 -0.5865  0.2592  0.1993
## E3:180 -1.9981  1.7863 -0.9216  0.3700  0.0405  0.0103
## E3:270  0.0796  1.3954  0.6244  0.3316 -0.6977 -0.4085
## E3:360  3.2466 -0.7166  0.0354  0.8473  0.2042  0.0714
## E3:90   0.2289 -0.6098  1.3213 -0.6329  0.2762  0.6271
## E4:180 -3.4192  0.2450 -0.6247 -0.8917  0.8761  0.0925
## E4:270  0.4811  0.9443  0.3613  0.2377  0.5239 -0.2461
## E4:360  2.9004 -5.5002  0.0556  1.2259 -0.5320  0.7620
## E4:90   0.4749  1.2149  1.4092 -0.7102  0.4896  0.2416
## E5:180 -0.6602  3.3933 -0.2000 -0.1789 -0.3809  0.2076
## E5:270 -0.3904  2.3259 -0.4290  0.3441  0.4539 -0.2688
## E5:360  3.0904  2.2444 -0.6518  0.6960 -0.1952  0.1695
## E5:90   0.5295  1.3961  0.0536 -0.3591 -0.6451 -0.3243
## E6:180 -4.8202  0.0126  1.0246 -0.4872 -1.7018  0.1876
## E6:270 -12.7699 -4.4290 -1.4508  0.3031  0.7485 -0.0665
```

```

## E6:360  6.1655 -3.9114 -3.7731 -1.3958 -0.6433 -0.0036
## E6:90  -2.0763 -3.5507  2.0570  0.1988 -0.9685 -0.4038
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##          Can1    Can2    Germ9  Germ5pn  GM9pn  GM9pa
## Can1      1.0000  0.0000 -0.3512 -0.6407 -0.6518  0.6260
## Can2      0.0000  1.0000  0.3759  0.7433  0.7540 -0.7422
## Germ9     -0.3512  0.3759  1.0000  0.6216  0.5410 -0.4956
## Germ5pn   -0.6407  0.7433  0.6216  1.0000  0.9916 -0.9777
## GM9pn     -0.6518  0.7540  0.5410  0.9916  1.0000 -0.9847
## GM9pa      0.6260 -0.7422 -0.4956 -0.9777 -0.9847  1.0000
## UMID      -0.9261 -0.3548  0.2223  0.3223  0.3313 -0.3013
## IVG       -0.1186 -0.0623  0.4305  0.0323  0.0093  0.0677
##

```

VariaveisCanonicas(Dados.Fat2.DBC,5,Fator="A")



```

## [1] "a"
##
## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##      CanRsqr Autovalor Porcentagem PorcentagemAcumulada
## VC1  0.8724    6.8355    88.6810    88.6810
## VC2  0.4449    0.8016    10.3995    99.0804
## VC3  0.0464    0.0486    0.6308    99.7112
## VC4  0.0138    0.0140    0.1816    99.8928
## VC5  0.0082    0.0083    0.1072    100.0000
##
## Escores das variaveis canonicas
##      Can1    Can2    Can3    Can4    Can5
## E1 -0.3538  0.9669 -0.2971  0.0076  0.0698
## E2 -1.9025  0.0641 -0.0056  0.1375 -0.1188
## E3 -0.6313  0.0048  0.2836  0.0743  0.1108
## E4  0.4387  0.7834  0.1685 -0.1563 -0.0623
## E5 -2.1029 -1.2628 -0.1006 -0.1002  0.0160

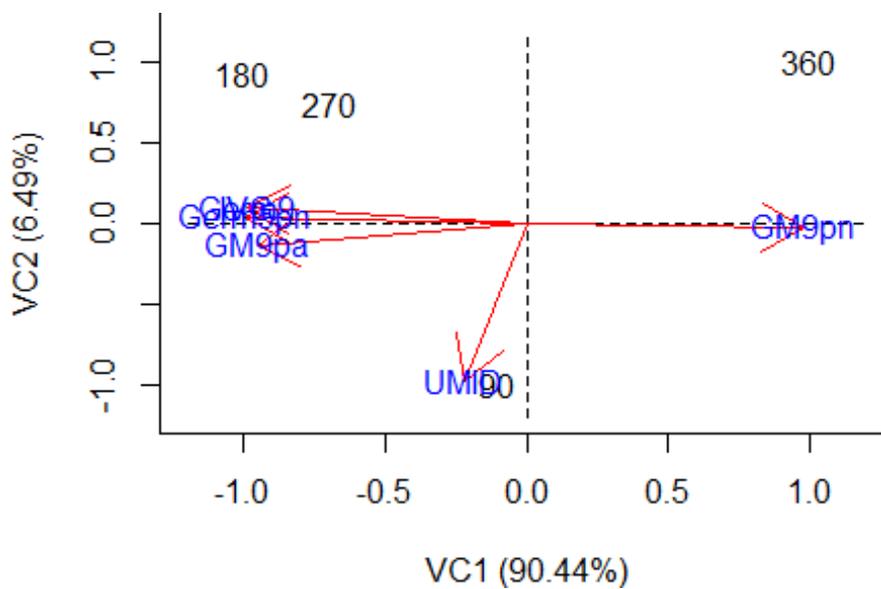
```

```

## E6 4.5519 -0.5563 -0.0488 0.0371 -0.0155
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##      Can1      Can2      Germ9 Germ5pn  GM9pn
## GM9pa 0.5457 0.8236 -0.7720 -0.9802 -0.9919
## UMID  0.9955 -0.0838 0.1067 -0.3624 -0.5330
## IVG   0.7069 -0.5395 0.6397 0.1800 0.0049
##

```

VariaveisCanonicas(Dados.Fat2.DBC,5,Fator="B")



```

## [1] "a"
##
## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##      CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.9145 10.6896 90.4377 90.4377

```

```

## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##      Can1      Can2      Germ9
## Germ5pn -0.9886 0.1011 0.8630
## GM9pn   -0.9930 0.0362 0.8285
## GM9pa   0.9813 -0.0275 -0.8240
## UMID    -0.9554 -0.1414 0.6460

```

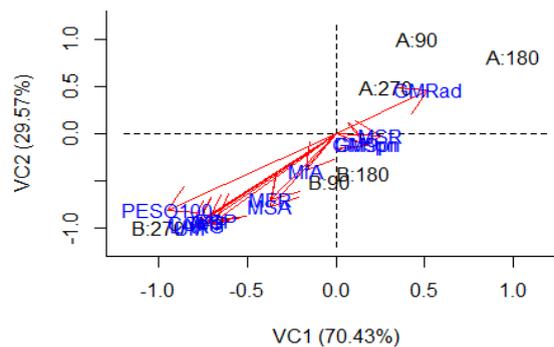
```
## IVG      -0.2219 -0.9717 -0.3840
##
```

### 13.1.6 Variáveis Discriminantes em fatorial duplo DIC

Executar variáveis canônicas em esquema fatorial triplo requer um conjunto de dados diferente dos citados anteriormente. Para fatorial triplo, em qualquer dos delineamentos, necessita-se que as três primeiras colunas contêm os fatores, a quarta coluna deve conter a repetição e as demais contêm as variáveis dependentes.

A execução de dados em esquema de fatorial triplo em DBC, dentro do *MultivariateAnalysis*, segue a mesma ideia dos demais fatores, iniciando-se com ativação do pacote e importação dos dados e finalizando com os argumentos de *layout* do gráfico e interações entre os fatores. O único detalhe a que se deve atentar é a escolha do modelo adequado. Para exemplo desse tópico, deve ser escolhido o modelo 6.

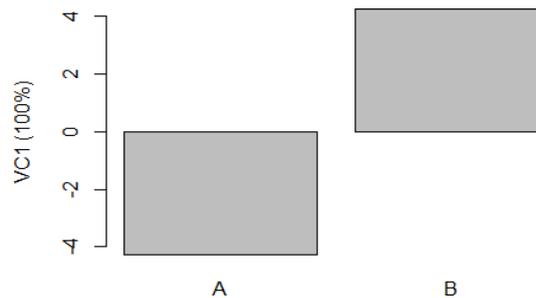
```
#Esquema fatorial triplo em DIC
data(Dados.Fat3.DIC)
VariaveisCanonicas(Dados.Fat3.DIC,6,Fator="A:B")
```



```
## [1] "a"
##
## -----
## Estudo das variáveis canônicas
##
## Explicação das variáveis canônicas
##      CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.1387    0.1610    70.428      70.428
## VC2 0.0633    0.0676    29.572     100.000
##
## Escores das variáveis canônicas
##      Can1    Can2
## A:90  0.6954  3.4394
## A:180 1.8574  2.7632
## A:270 0.2997  1.5699
## B:90  -0.3914 -2.0983
## B:180 0.0242 -1.7831
## B:270 -2.4852 -3.8910
##
## Importância
## Correlação das características com os escores das variáveis canônicas
```

```
##          Can1    Can2
## GMRad    0.5204  0.4661
## GM5pn    0.1690 -0.1224
## GM9pn    0.1795 -0.1279
## UM       -0.8224 -0.9904
## IVG      -0.7206 -0.9303
## MfA      -0.1728 -0.3897
## MFR      -0.3744 -0.7001
## MSA      -0.3729 -0.7754
## MSR       0.2529 -0.0230
## PES0100 -0.9540 -0.8134
## COMP     -0.7823 -0.9502
## LARG     -0.7805 -0.9652
## ESP      -0.6752 -0.9185
##
```

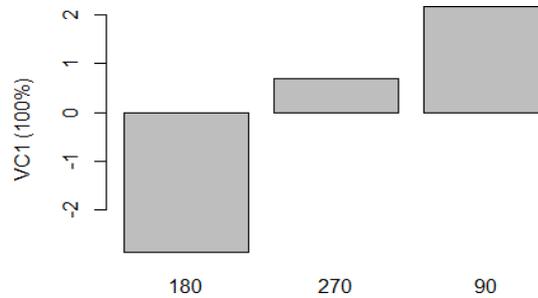
VariaveisCanonicas(Dados.Fat3.DIC,6,Fator="A")



```
## [1] "a"
```

```
##
## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##      CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.9602    24.1296          100          100
##
## Escores das variaveis canonicas
##      A      B
## -4.2541  4.2541
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##      GMRad  GM5pn  GM9pn    UM    IVG    MfA    MFR    MSA    MSR PES0100
##      -1      1      1      1      1      1      1      1      NA      1
##      COMP   LARG   ESP
##      1      1      1
##
```

VariaveisCanonicas(Dados.Fat3.DIC,6,Fator="B")



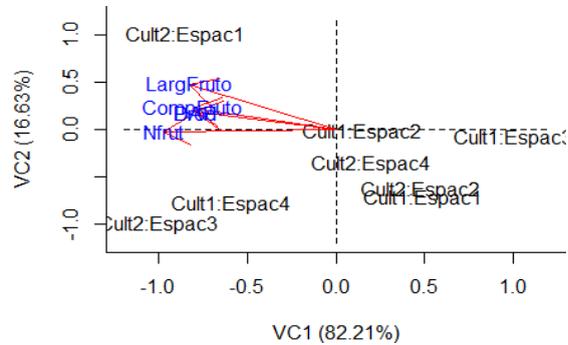
```
## [1] "a"

##
## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##   CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.8558    5.9334    65.9434            65.9434
## VC2 0.7540    3.0643    34.0566            100.0000
##
## Escores das variaveis canonicas
##   90    180    270
## 2.1636 -2.8606 0.6970
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##   GMRad  GM5pn  GM9pn    UM    IVG    MfA    MFR    MSA    MSR PES0100
## -0.9149 -0.9695 -0.9797 -0.2070 0.6057 -0.8447 -0.9234 -0.9838 -0.9753 0.5928
##   COMP    LARG    ESP
## 0.7401 0.7650 0.5923
##
## -----
```

### 13.1.7 Variáveis Discriminantes em fatorial duplo DBC

A ideia de execução de dados em esquema fatorial triplo em DBC segue o mesmo processamento do exemplo em DIC, como mostra o exemplo anterior desse material. A pequena diferença está na escolha do modelo de delineamento. Para dados em DBC, deve ser escolhido o modelo de número 7.

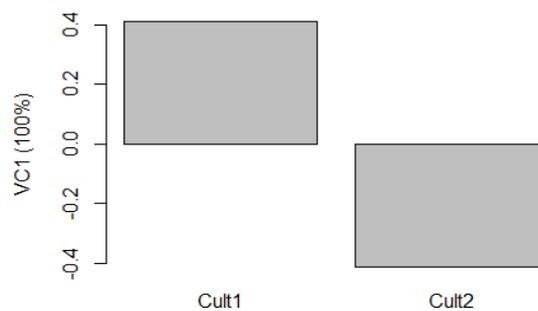
```
#Esquema fatorial triplo em DBC
data(Dados.Fat3.DBC)
VariaveisCanonicas(Dados.Fat3.DBC,7,Fator="A:B")
```



```
## [1] "a"
```

```
##
## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##   CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.2540    0.3405    82.2099          82.2099
## VC2 0.0644    0.0689    16.6304          98.8404
## VC3 0.0048    0.0048     1.1596          100.0000
##
## Escores das variaveis canonicas
##           Can1    Can2    Can3
## Cult1:Espac1  0.4306 -0.1705  0.5526
## Cult1:Espac2  0.1383  0.1269 -0.5428
## Cult1:Espac3  0.8650  0.1043  0.4723
## Cult1:Espac4 -0.4913 -0.1962  0.0116
## Cult2:Espac1 -0.7097  0.5655  0.3612
## Cult2:Espac2  0.4201 -0.1313 -0.6312
## Cult2:Espac3 -0.8341 -0.2850  0.1569
## Cult2:Espac4  0.1810 -0.0136 -0.3806
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##           Can1    Can2    Prod
## Nfrut      -0.7976  0.1815  1.0000
## LargFruto  -0.9764 -0.0217  0.8783
## CompFruto  -0.8303  0.4798  0.9085
## DAP        -0.8074  0.2232  0.9983
##
## -----
```

```
VariaveisCanonicas(Dados.Fat3.DBC,7,Fator="A")
```

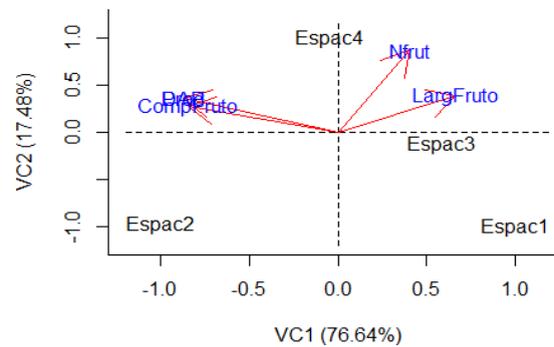


```
## [1] "a"

## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##   CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.1928   0.2389       100           100
##
## Escores das variaveis canonicas
## Cult1 Cult2 ##
0.4099 -0.4099 ##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas

##   Prod   Nfrut LargFruto CompFruto   DAP
##   -1     -1     -1         -1         -1
## -----
```

**VariaveisCanonicas**(Dados.Fat3.DBC,7,Fator="B")



```
## [1] "a"

## -----
## Estudo das variaveis canonicas
##
## Explicacao das variaveis canonicas
##   CanRsq Autovalor Porcentagem PorcentagemAcumulada
## VC1 0.2240   0.2887   76.6439           76.6439
## VC2 0.0618   0.0658   17.4779           94.1218
## VC3 0.0217   0.0221    5.8782           100.0000
##
## Escores das variaveis canonicas
##   Can1   Can2   Can3
## Espac1 0.5180 -0.1919 -0.1172
## Espac2 -0.6783 -0.1841  0.0043
## Espac3 0.2682  0.0402  0.2016
## Espac4 -0.1079  0.3358 -0.0887
##
## Importancia
## Correlacao das caracteristicas com os escores das variaveis canonicas
##   Can1   Can2   Prod
## Nfrut  -0.8711  0.3627  0.9999
## LargFruto 0.4007  0.8719  0.0530
## CompFruto 0.6616  0.3824 -0.2243
## DAP      -0.8464  0.2763  0.9899
## -----
```

## 14. RESUMO DAS PRINCIPAIS FUNÇÕES

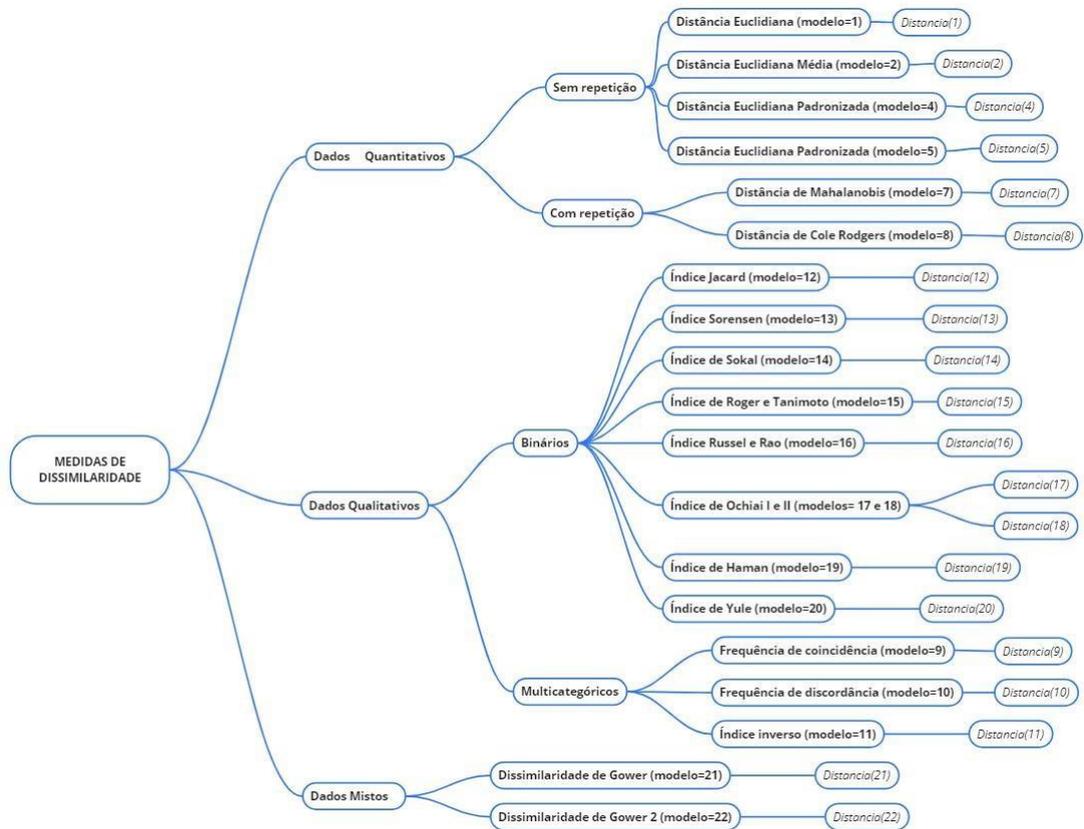


Figura 12. Resumo das funções de medidas de dissimilaridade

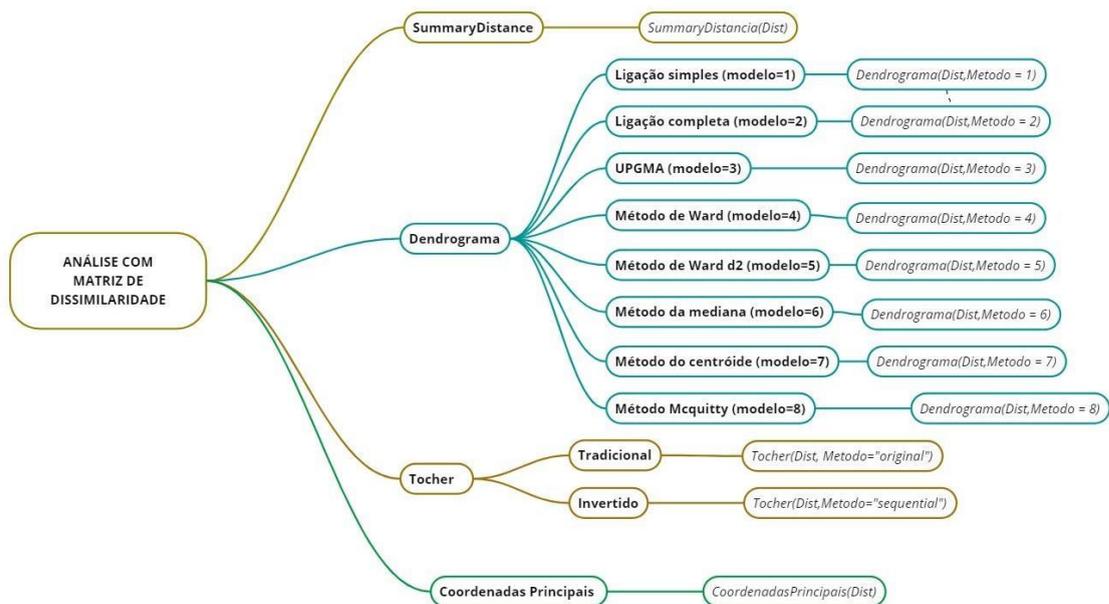


Figura 13. Resumo das funções de análises de matriz de dissimilaridade

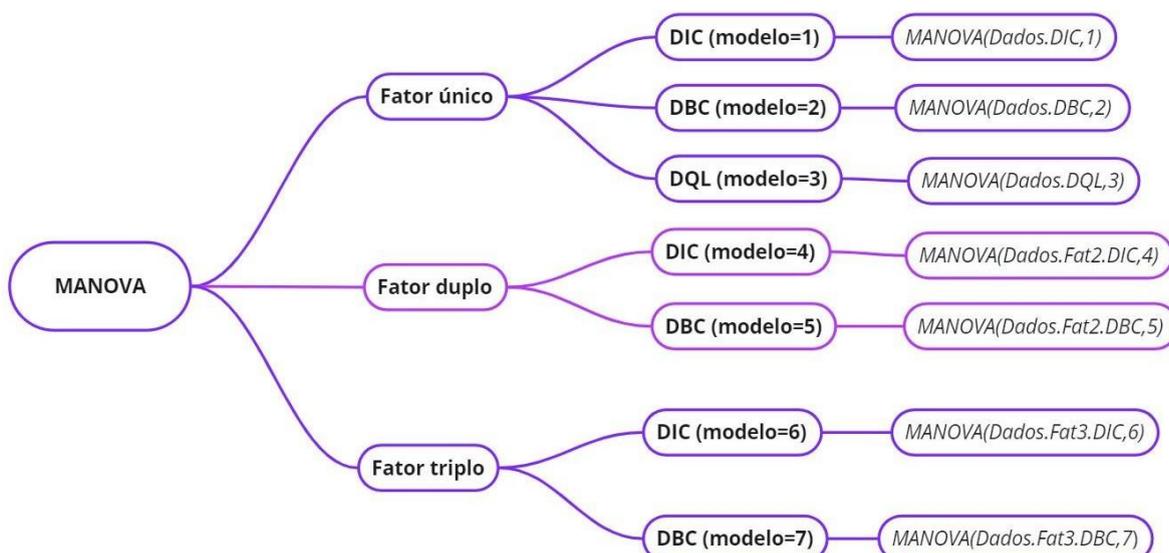


Figura 14. Resumo das funções de análise de variância multivariada

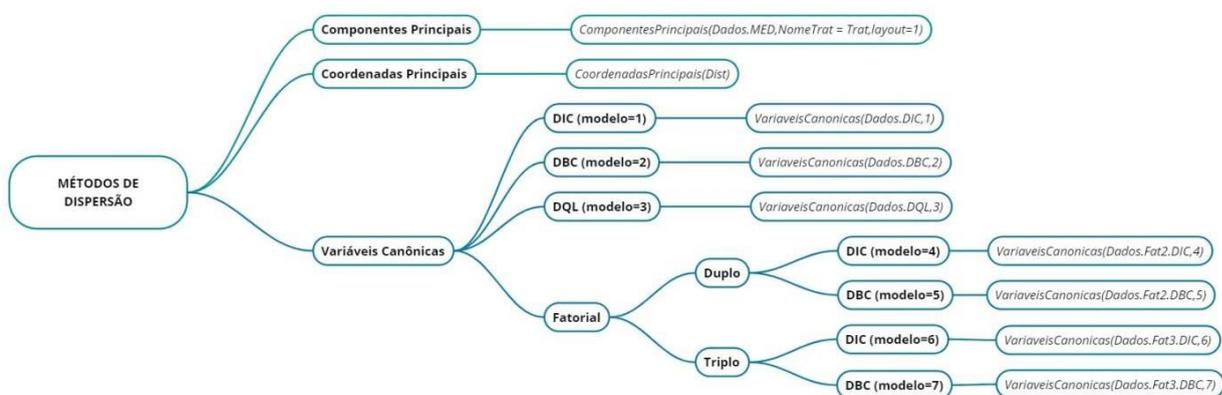


Figura 15. Resumo das funções de métodos de dispersão

## 15. REFERÊNCIAS

- ALBUQUERQUE, Mácio Augusto de *et al.* Estabilidade em análise de agrupamento (cluster analysis). 2005.
- BORDIGNON, Fernando Luis. Técnicas Inteligentes para Análise de Agrupamento de Dados. 2010.
- CARVALHO, M. F. *et al.* Aplicação da análise multivariada em avaliações de divergência genética através de marcadores moleculares dominantes em plantas medicinais. *Revista Brasileira de Plantas Medicinais*, v. 11, n. 3, p. 339-346, 2009.
- CHAVENT, Marie *et al.* Package 'PCAmixdata'. **The R Project for Statistical Computing**, 2017.
- COLE-RODGERS, P.; SMITH, D. W.; BOSLAND, P. W. A novel statistical approach to analyze genetic resource evaluations using Capsicum as an example. **Crop Science**, v. 37, n. 3, p. 1000-1002, 1997.
- COLE-RODGERS, P. *et al.* A novel statistical approach to analyze genetic resource evaluations using Capsicum as an example. *Crop Science*, v.37, p.1000-1002, 1997. Disponível em: <<http://dx.doi.org/10.2135/cropsci1997.0011183X003700030050x>>. Acesso em 08/06/2024
- CRISPIM, Diêgo Lima; FERNANDES, Lindemberg Lima; DE OLIVEIRA ALBUQUERQUE, Roberta Luiza. Aplicação de técnica estatística multivariada em indicadores de sustentabilidade nos municípios do Marajó-PA. **Revista Principia-Divulgação Científica e Tecnológica do IFPB**, n. 46, p. 145-154, 2019.
- Cruz CD (2006) Programa Genes: análise multivariada e simulação. Viçosa, Editora UFV. 175p
- CRUZ, C. D.; CARNEIRO, P. C. S. (2006). Modelos biométricos aplicado ao melhoramento genético. Viçosa: editora UFV. 2.ed. v.2, 585 p.
- CRUZ, Cosme Damião. GENES: software para análise de dados em estatística experimental e em genética quantitativa. **Acta Scientiarum. Agronomy**, v. 35, p. 271-276, 2013.
- DA SILVA, Anderson Rodrigo; DA SILVA, Maintainer Anderson Rodrigo. Package 'biotools'. **Available online at: <https://CRAN.R-project.org/package=~biotools> (accessed December 27, 2020)**, 2017.
- DU, K.-L. Clustering: A neural network approach. *Neural Networks*, v. 23, n. 1, p. 89–107, jan. 2010. ISSN 08936080. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S089360800900207X?via%3Dihub>
- EMMANUEL PARADIS, KLAUS SCHLIEP, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R, *Bioinformatics*, Volume 35, Issue 3, 01 February 2019, Pages 526–528, <https://doi.org/10.1093/bioinformatics/bty633>

FARIA, Priscila N. *et al.* Métodos de agrupamento em estudo de divergência genética de pimentas. **Horticultura Brasileira**, v. 30, p. 428-432, 2012.

FRIENDLY, Michael; FOX, John; FRIENDLY, Maintainer Michael. Package 'candisc'. **R package**, 2017.

GOBO, J. P. A.; GALVANI, E.; WOLLMANN, C. A. Influência do clima regional sobre o clima local a partir do diagnóstico de abrangência espacial e extrapolação escalar. *Revista Brasileira de Climatologia*, v. 22, p. 210-228, 2018.

[https://www.puc-rio.br/ensinopesq/ccpg/pibic/relatorio\\_resumo2016/relatorios\\_pdf/ctc/IND/IND-Andr%C3%A9%20Provenzano%20Naveiro.pdf](https://www.puc-rio.br/ensinopesq/ccpg/pibic/relatorio_resumo2016/relatorios_pdf/ctc/IND/IND-Andr%C3%A9%20Provenzano%20Naveiro.pdf)

HILBE, Joseph M. STATISTICA 7: an overview. **The American Statistician**, v. 61, n. 1, p. 91-94, 2007.

JOHNSON, R. A.; WICHERN, D. W. Applied multivariate statistical analysis. 4th ed. Upper Saddle River, New Jersey: Prentice-Hall, 1999, 815 p.

KHATTREE, R. & NAIK, D.N. Multivariate data reduction and discrimination with SAS software. Cary, NC, USA: SAS Institute Inc., 2000. 558 p.

OSSANI, Paulo Cesar; CIRILLO, Marcelo Angelo; OSSANI, Maintainer Paulo Cesar. Package 'MVar. pt'. 2022.

PARADIS E. *et al.* (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20, 289–290.

PRADO, Bruna Queiroz de Melo. Análise de agrupamentos das taxas de incidência de dengue nos estados brasileiros. 2015.

RODRIGUEZ, Robert N. Sas. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 3, n. 1, p. 1-11, 2011.

UNIVERSIDADE FEDERAL DE VIÇOSA – UFV. **Sistema para análises estatísticas e genéticas –SAEG** Versão 7.1. Viçosa, MG, 1997. 150p

VARELLA, Carlos Alberto Alves. Análise multivariada aplicada as ciencias agrárias. **Seropédica: Universidade Federal Rural do Rio de Janeiro**, 2008.

VASCONCELOS, Edmar Soares de *et al.* Método alternativo para análise de agrupamento. **Pesquisa agropecuária brasileira**, v. 42, p. 1421-1428, 2007.

## 16. ARTIGO

***MultivariateAnalysis*: um pacote R para análise multivariada.**

Este artigo foi elaborado conforme as normas da revista *Scientia Agricola*

## ***MultivariateAnalysis*: um pacote R para análise multivariada.**

### **RESUMO**

As análises estatísticas são parte importante na pesquisa. Com a modernização do processo produtivo, tem aumentado o volume de dados a serem processados. Assim, o uso de técnicas como análise multivariada tem proporcionado a execução de dados de forma simultânea. As execuções desses tipos de análises geralmente são complexas, necessitando do auxílio de *softwares* estatísticos. O pacote *MultivariateAnalysis* é um pacote R desenvolvido com o objetivo de realizar análises multivariadas e se encontra disponível gratuitamente na plataforma CRAN. Esse pacote foi lançado em 2021 pelo pesquisador Alcinei Místico Azevedo e contém técnicas como análises de componentes principais, análises de coordenadas principais, agrupamentos hierárquicos, correlação mantel, dendrogramas, variáveis canônicas, medidas de dissimilaridade, análise de variância multivariada, entre outras. O objetivo deste trabalho foi descrever o pacote *MultivariateAnalysis*, com base na sua execução. Neste artigo, apresentamos um resumo das funções apresentadas pelo pacote, guiando o usuário de forma prática, desde as primeiras funções até a obtenção dos resultados, a fim de propiciar o conhecimento sobre a realização de atividades neste pacote e as potencialidades de sua aplicação, por apresentar um código aberto que permite adicionar funções. Em 2024, o *MultivariateAnalysis* se encontra em sua versão 5.0, que conta com melhoria nas funções gráficas, oferecendo um ambiente de trabalho simples, bastante flexível, com funções bastante intuitivas e com ferramentas que facilitarão a implementação deste pacote pelo usuário, em qualquer área do conhecimento.

**Palavras-chave:** Análise de dados; pacotes R; análise multivariada.

## **MultivariateAnalysis: an R package for multivariate analysis**

### **Abstract**

Statistical analyses are a crucial part of research. With the modernization of the production process, the volume of data to be processed has increased. Thus, the use of techniques such as multivariate analysis has enabled simultaneous data processing. The execution of these analyses is generally complex, requiring the aid of statistical software. The MultivariateAnalysis package is an R package developed to perform multivariate analyses and is freely available on the CRAN platform. Released in 2021 by researcher Alcinei Místico Azevedo, this package includes techniques such as principal component analyses, principal coordinate analyses, hierarchical clustering, Mantel correlation, dendrograms, canonical variables, dissimilarity measurements, multivariate variance analysis, and more. The objective of this paper is to describe the MultivariateAnalysis package based on its execution. In this article, we provide a summary of the package's functions, guiding the user in a practical way from the initial steps to the final results, thus promoting knowledge on how to use the package and its potential applications, as it presents an open-source code that allows users to add functions. In 2024, MultivariateAnalysis is in its version 5.0, which includes improvements in graphic functions, offering a simple and flexible work environment with highly intuitive functions and tools that facilitate the implementation of the package in any fields of knowledge.

**Keywords:** data analysis, R packages, multivariate analysis.

## 17. INTRODUÇÃO

Os avanços tecnológicos têm modificado muitos processos nas últimas décadas e, entre esses avanços, a agricultura vem se destacando. Diversas atividades que anteriormente eram feitas manualmente, hoje podem ser automatizadas, por exemplo, com a utilização de imagens de satélites para identificação de manchas de solo, ou o uso de inteligência artificial na mecanização agrícola, auxiliando desde o plantio até a colheita. Essa evolução foi impulsionada pela agricultura 4.0, que revolucionou a agricultura brasileira. As operações tecnológicas na agricultura vão além do campo; operações e decisões do processo produtivo, como relativas ao clima, podem ser extraídas em tempo real (SILVA & CAVICHIOLI, 2020).

O processo evolutivo da agricultura também gerou a informatização dos dados e o aumento do volume de dados para serem interpretados. Esse processo gera consequências nas tomadas de decisões, devido à abundância de dados gerados. Além da agricultura, os dados de produção científica também cresceram com a informatização da agricultura. A utilização de técnicas que possibilitem considerar diversas variáveis de forma simultânea torna-se uma ferramenta importante na tomada de decisão. As análises multivariadas são técnicas estatísticas que, de forma simultânea, analisam várias medidas sobre indivíduos ou objetos em análise (HAIR *et al.*, 2009).

As principais técnicas de análises multivariadas são análises de componentes principais, análises discriminantes, agrupamentos hierárquicos, análises de variância multivariada e correlações canônicas. As técnicas de análise multivariada são divididas em cinco categorias, sendo elas: investigação de dependências entre as variáveis; predição; construção de hipóteses e testes; redução dos dados ou simplificação estrutural; e agrupamentos de objetos ou variáveis (GOUVEA *et al.*, 2011). Cada uma dessas categorias é composta por técnicas específicas e existem técnicas que podem estar em mais de uma categoria, podendo ser realizadas de forma manual ou com o uso de *softwares*. Dentre os *softwares* livres pode ser citado o R (R CORE TEAM, 2019) e sua interface RStudio (ALLAIRE, 2012).

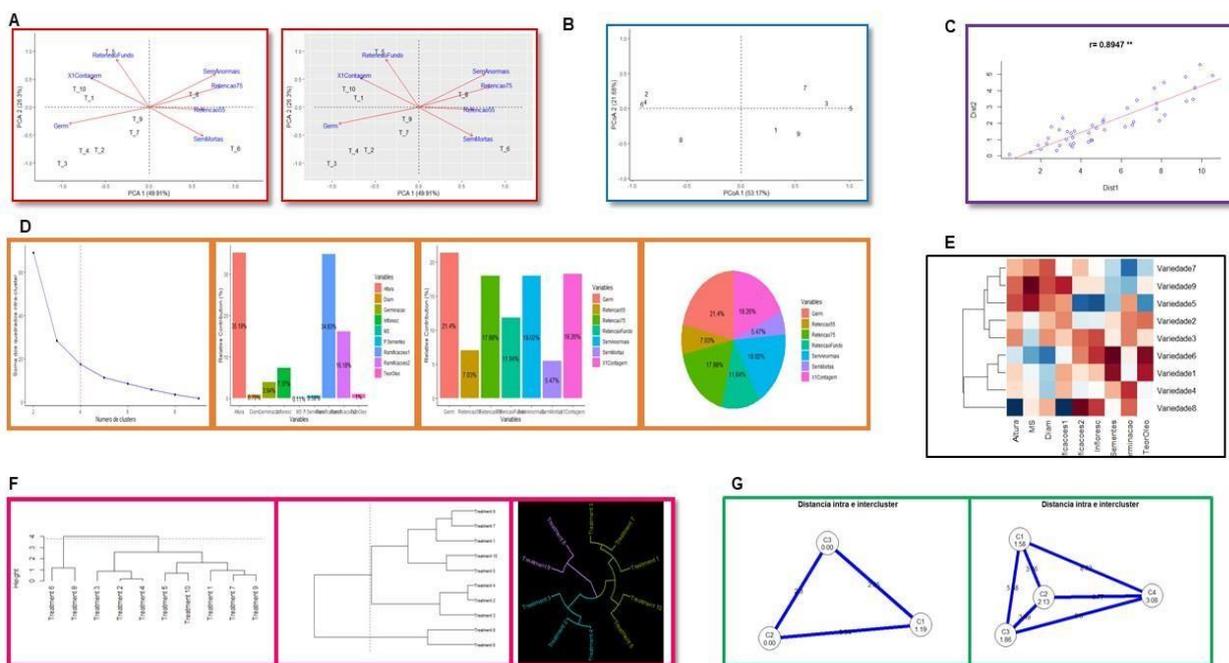
Uma das vantagens da utilização do R e RStudio é a utilização de pacotes que facilitam o trabalho de muitos pesquisadores, devido à diversidade de pacotes disponíveis com muitas funções e ao fato de ser gratuito. Esses pacotes estão disponíveis no repositório *Comprehensive R Archive Network* (CRAN). Para a realização de análises multivariadas, existem alguns pacotes disponíveis, como o pacote Vegan (OKSANEN *et al.*, 2018), que realiza análises multivariadas aplicadas a ecologia, o pacote biotools (DA SILVA, 2016), com as funções de análise de agrupamentos, o pacote Candisc (FRIENDLY & FOX, 2013), com as análises de variância multivariada, análise discriminante e correlação canônica, e o pacote metan (OLIVOTO & LÚCIO, 2020), para a realização de análises multiambientais.

Esses pacotes são bastante úteis para realização de análises multivariadas, porém não reúnem todas as técnicas das análises ou apresentam rotinas bastante complexas e frequentemente requerem a utilização de vários pacotes para execução de uma atividade. Além dos pacotes acima citados, existe o denominado *MultivariateAnalysis*. Desenvolvido em 2021, esse pacote R caracteriza-se por possibilitar a realização de análise multivariada de formas simples e eficiente.

No *MultivariateAnalysis*, é possível a realização de técnicas como agrupamentos hierárquicos, análises de componentes principais, análises de coordenadas principais, correlação mantel, dendrograma, análise de variância multivariada, medidas de distâncias, variáveis canônicas, entre outras. Caracterizado por ser simples e intuitivo, esse pacote pode ser aplicado em várias áreas do conhecimento. Neste trabalho, descreve-se o pacote do R *MultivariateAnalysis*, destinado à realização de análises multivariadas com uma linguagem de programação simples e eficaz. O objetivo deste artigo é apresentar o *MultivariateAnalysis* e demonstrar as principais funções e sua aplicabilidade nas Ciências Agrárias.

## **18. O PACOTE MULTIVARIATEANALYSIS**

A essência do *MultivariateAnalysis* está na realização de técnicas de análises multivariadas, como: medidas de dissimilaridade; dendrograma; análise de variância multivariada; componentes principais; variáveis canônicas; coordenadas principais; agrupamento Tocher; métodos K-means etc. Alguns exemplos de resultados gráficos obtidos pelo pacote *MultivariateAnalysis* são mostrados na figura abaixo (Figura 16).



**Figura 16.** Exemplos de resultados gráficos do pacote *MultivariateAnalysis*: A) Componentes Principais; B) Coordenadas Principais; C) Correlação Cofenética do Agrupamento Tocher; D) Agrupamentos kmeans; E) HeatPlot; F) Dendrograma; G) Distância intra e inter cluster

Em 2024, o pacote *MultivariateAnalysis* se encontra na sua versão 5.0, disponível para *download* no repositório CRAN (<https://cran.r-project.org/web/packages/MultivariateAnalysis>), podendo ser instalado diretamente no RStudio, utilizando o comando dentro do R chamado *install.packages("MultivariateAnalysis")*. Para melhor compreensão das funções do pacote, está disponível no YouTube uma playlist com videoaulas no canal Alcinei Azevedo – Dicas e Aulas (<https://www.youtube.com/@alcineiazevedo-dicaseaulas9398>) e no site [www.expstat.com](http://www.expstat.com).

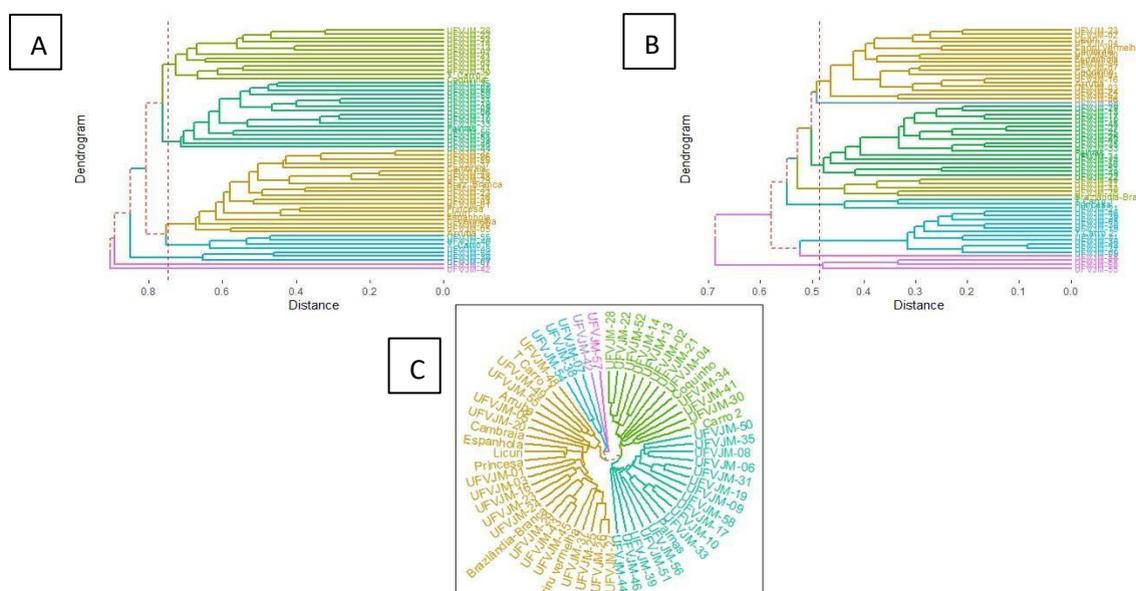
Com a finalidade de ilustrar e exemplificar as funções dentro do pacote, foi utilizado o conjunto de dados Iris, sendo esse bastante utilizado em exemplo de aplicação de *softwares* e algoritmos. Para início da realização das operações utilizando o *MultivariateAnalysis*, deve-se ativar o pacote com a função *library("MultivariateAnalysis")* e, depois desse processo, deve-se importar os dados que se pretende analisar com a função *data()*, com o argumento informando quais são os dados. Dependendo da função que se deseja executar, deve-se atentar ao formato de tabulação dos dados. Em caso de dúvidas de execução ou de tabulação dos dados, podem ser consultados os exemplos no manual do *MultivariateAnalysis*.

## 19. VERIFICAÇÃO DOS DADOS

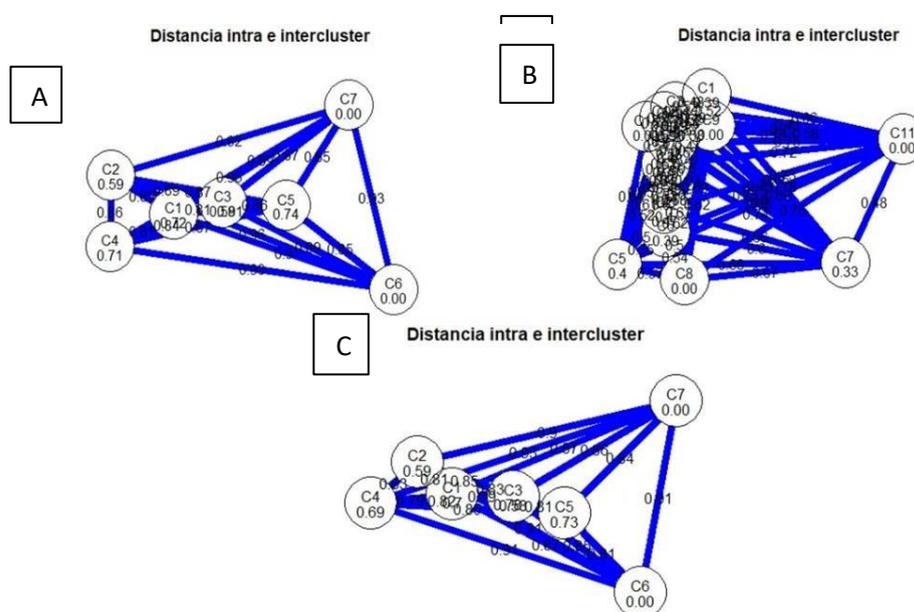
Algo importante na demonstração das funções de um pacote está na utilização de dados já publicados. A demonstração das funcionalidades do *MultivariateAnalysis* se dará utilizando os dados de dissimilaridade genética entre genótipos de batata-doce (ANDRADE *et al.*, 2017). Os dados foram divididos em dados moleculares, dados morfológicos e dados de médias ponderadas das matrizes.

Para medidas de dissimilaridade dos dados realizados no *MultivariateAnalysis*, pode-se observar para os dados moleculares que, para o método de Jacard, os indivíduos mais próximos são UFVJM-26 e UFVJM-27 e, em contrapartida, os mais distantes são Espanhola e UFVJM-42, como mostram os dados da Tabela 4. Os mesmos dados podem ser representados de forma gráfica, como mostra a Figura 18. Para os dados de morfologia, os indivíduos mais próximos, de acordo com os resultados do *MultivariateAnalysis*, foram UFVJM-37 e UFVJM-42 e os mais distantes foram T-carro 1 e UFVJM-55, como mostram os dados da Tabela 4 e Figura 18.

Os Dendrogramas obtidos a partir das matrizes de dissimilaridade são apresentados na Figura 7 e podem ser observados para os dados moleculares (Figura 17A), dados morfológicos (Figura 17B) e dados de média ponderada entre as matrizes (Figura 17C). Assim como os dendrogramas, as distâncias inter e intra cluster formaram sete grupos, como mostra a Figura 8. No *MultivariateAnalysis*, foram formados sete clusters, como os dados encontrados por Andrade *et al.*, (2017), demonstrando a acurácia do *MultivariateAnalysis* em realizar dendrogramas e medir distâncias inter e intracluster.



**Figura 17.** Dendrogramas obtidos no *MultivariateAnalysis* para dados de dissimilaridade genética de 60 genótipos de batata doce para A) dados moleculares; B) dados morfológicos e C) média ponderada entre as matrizes.



**Figura 18.** Distâncias inter- e intracluster de dados de 60 genótipos de batata doce para A) dados moleculares; B) dados morfológicos e C) média ponderada entre as matrizes.

O agrupamento Tocher dos dados de dissimilaridade genética entre genótipos de batata-doce realizado no *MultivariateAnalysis* resultou em sete clusters, sendo o grupo um composto pela maior quantidade de genótipos em um grupo, como mostram os dados da Tabela 4. Essa quantidade de cluster também foi observada no artigo de referência Andrade *et al.*, (2017). Essa semelhança entre os resultados encontrados neste trabalho e os autores anteriormente citados ressalta a precisão e a acurácia dos resultados no *MultivariateAnalysis*.

**Tabela 4.** Agrupamento de 60 genótipos de batata doce (*Ipomoea batatas*) pelo método de Tocher.

CLUSTER	GENÓTIPOS
1	UFVJM-26, UfvJM-27, UfvJM-25, UfvJM-29, UfvJM-45, UfvJM-43, UfvJM-37 Braz. Branca, UfvJM-24, UfvJM-23, Cambraia, UfvJM-16, UfvJM-03, Princesa, Licuri, Espanhola, UfvJM-01, C.vermelha, UfvJM-05, UfvJM-20, Arruba, UfvJM-48, UfvJM-55 UfvJM-49, UfvJM-08, T.Carro1, UfvJM-13, UfvJM-35, UfvJM-58, UfvJM-10, UfvJM-33, UfvJM-17, UfvJM-19, UfvJM-09, UfvJM-06, UfvJM-50, UfvJM-56, Palmas, UfvJM-31, UfvJM-52, UfvJM-22, UfvJM-02, UfvJM-14, UfvJM-28, UfvJM-34, Coquinha, UfvJM-51, UfvJM-39,
2	UfvJM-38, UfvJM-54, UfvJM-07
3	UfvJM-30, UfvJM-41, T. Carro 2
4	UfvJM-04, UfvJM-21
5	UfvJM-44, UfvJM-46
6	UfvJM-42
7	UfvJM-57

## 20. MEDIDAS DE DISSIMILARIDADE

O *MultivariateAnalysis* executa medidas de dissimilaridade com dados qualitativos, quantitativos e mistos, assim como o *software* GENES (CRUZ, 2016) e algumas funções dentro dos pacotes *DisimForMixed* (PATHBERIYA, 2016), *distancia* (BENITO & BIRKS, 2020) e *TSclust* (MONTERO & VILAR, 2015). A função principal para obtenção de distâncias tem por nome *Distancia* (). Os dados são o primeiro argumento, e o segundo argumento é a escolha do modelo escolhido, como mostra a Figura 19.

Alguns resultados são obtidos a partir dessa função, como a matriz de dissimilaridade e informações como a menor e maior distância, média e amplitude das distâncias, desvio padrão, coeficiente de variação das distâncias, além de apresentar os indivíduos mais próximos e mais distantes. Em caso de dúvidas sobre a execução, podem ser consultados os exemplos com o comando *?Distancia*. Para dados qualitativos binários ou multicategóricos, os dados podem ser compostos por números ou textos, os quais compõem as informações para verificar a similaridade ou dissimilaridade dos indivíduos. Os dados binários são compostos apenas pelos números 0 ou 1. Já os dados mistos são compostos por textos e números. Os índices de dissimilaridades presentes no pacote *MultivariateAnalysis* estão apresentados na Figura 9.

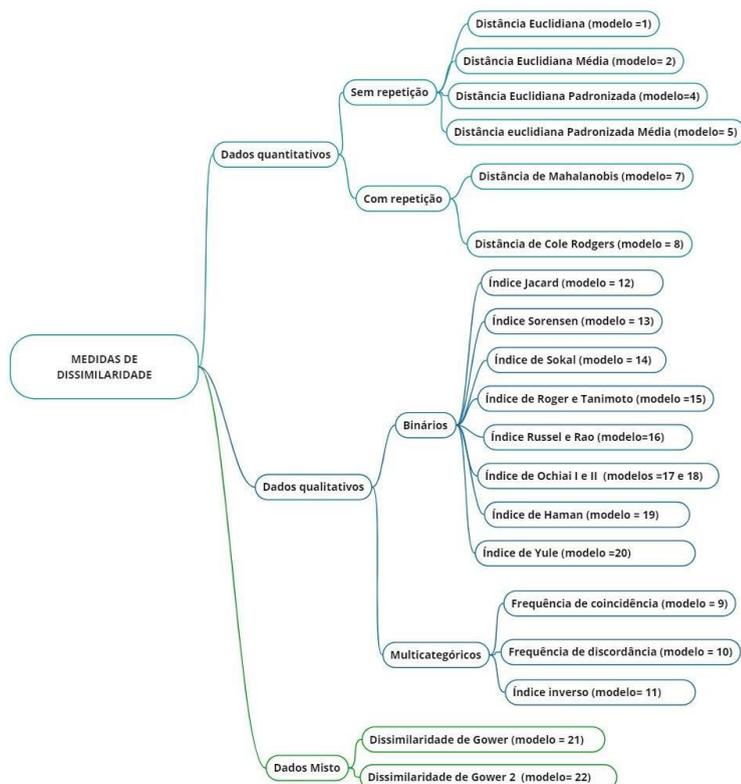
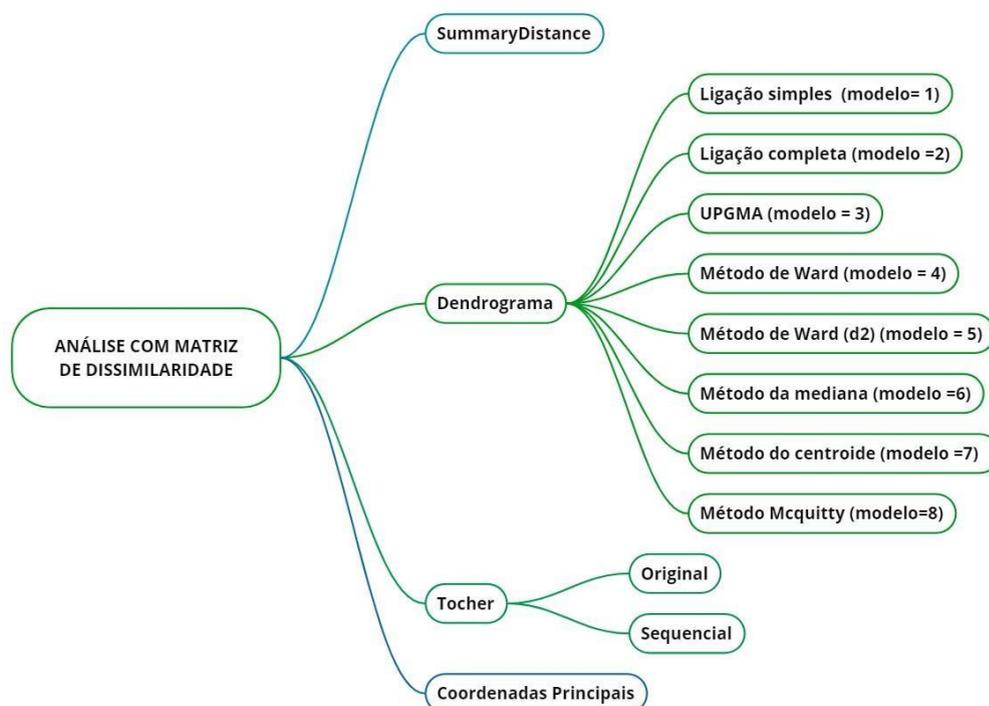


Figura 19. Modelos de medidas de dissimilaridade do pacote *MultivariateAnalysis*

## 21. ANÁLISE COM MATRIZ DE DISSIMILARIDADE

As funções *SummaryDistance ()*, *Dendrograma ()*, *Tocher ()* e *CoordenadasPrincipais ()* são os métodos que utilizam as informações das matrizes de dissimilaridade no *MultivariateAnalysis*. A execução de dendrograma pode ser realizada também nos pacotes *factoextra* (KASSAMBARA & MUNDT, 2017), *dendextend* (GALILI, 2015), *DendSer* (HURLEY & EARLE, 2022) e *ggdendro* (DE VRIES *et al.*, 2024). *Dendrograma ()* é a função para essa execução no *MultivariateAnalysis*, com quatro argumentos: o primeiro, os dados; o segundo, o método; o terceiro, o *layout*; e o quarto, o tipo de corte. Os métodos disponíveis estão exemplificados na Figura 20, bem como os tipos de análises com matrizes de dissimilaridade.



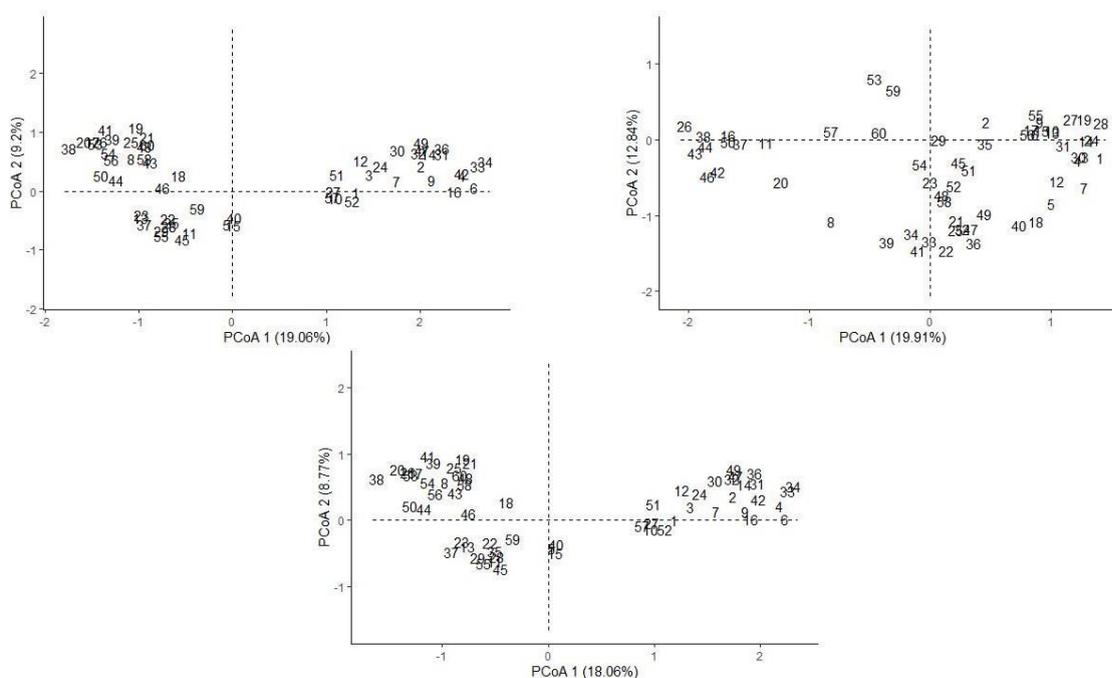
**Figura 20.** Tipos de análises com matriz de dissimilaridade do pacote *MultivariateAnalysis*

Os resultados dos dendrogramas no *MultivariateAnalysis* são compostos por representação gráfica do dendrograma, estimativa de correlação cofenética e significância da correlação cofenética pelo teste de Mantel, critério de corte e cluster formados. A partir desse pacote é possível estabelecer o ponto de corte no dendrograma a partir do método Mojena, Cindex, Frey, Mcclain e Dunn.

Para execução do método Tocher no *MultivariateAnalysis*, existem dois métodos disponíveis: o original e o sequencial, como mostra a Figura 21. A função *Tocher()* executa esse método, composta basicamente por dois argumentos: o primeiro, a matriz de dissimilaridade, e o segundo, o método, *sequencial* ou *original*. *Xlab*, *ylab* e *bty* são argumentos adicionais para modificações das expressões gráficas do eixo x, y e as bordas respectivamente no gráfico de dispersão da correlação cofenética.

As análises de coordenadas principais (PCoA) podem ser consideradas uma generalização dos métodos de componentes principais com os autovalores e autovetores, obtidos através de uma matriz de dissimilaridades (JONGMAN *et al.*, 1995). Esse tipo de análise pode ser realizado no pacote *vegan* (OKSANEN *et al.*, 2018), além de pelo *MultivariateAnalysis*. Os argumentos necessários nessa função são a matriz de dissimilaridade, o estilo do *layout* e o argumento para plotagem (*plot=T*). Na Figura 21 é

demonstrado o resultado da análise de coordenadas principais para os dados de divergência genética de 60 genótipos de batata doce.



**Figura 21.** Resultado de análises de Coordenadas Principais com os dados Iris no pacote *MultivariateAnalysis*

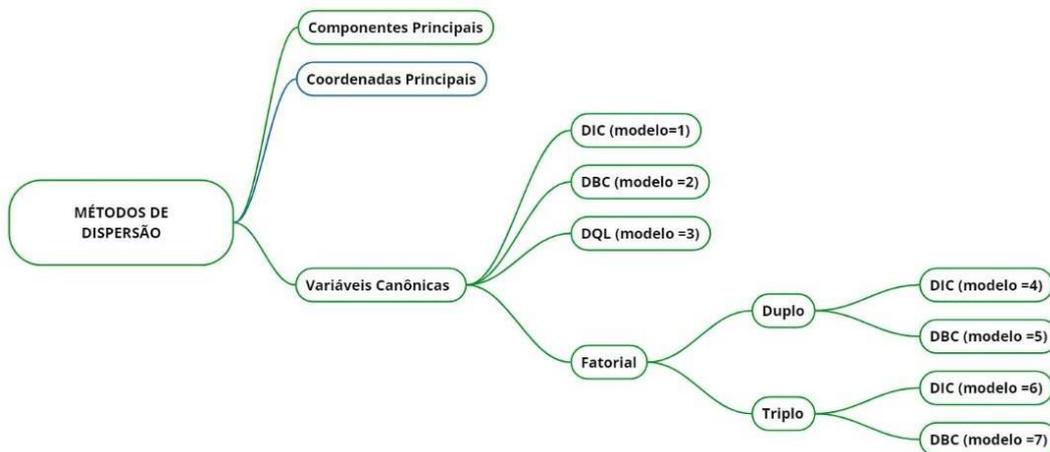
## 22. MÉTODOS DE DISPERSÃO

As medidas de dispersão que podem ser realizadas no *MultivariateAnalysis* são: Componentes Principais (PCA), Coordenadas Principais (PCoA) e Variáveis Canônicas. As análises de componentes principais também podem ser realizadas nos pacotes FactoMineR (HUSSON *et al.*, 2014), ade4 (DRAY & DUFOUR, 2007) e amap (LUCAS, 2014). A execução dos componentes principais é composta por uma função chamada *ComponentesPrincipais* (). Além das várias funções presentes no *MultivariateAnalysis*, esse pacote se caracteriza pela simplicidade de execução das funções, comandos com nomes bastantes intuitivos e acurácia dos dados.

No *MultivariateAnalysis*, é possível a realização de Variáveis Canônicas com os diversos delineamentos experimentais, como mostra a Figura 22. Pacotes como CCA (GONZÁLEZ & DÉJAN, 2013) e yacca (BUTTS, 2012) realizam análises de variáveis canônicas e cada um com suas características.

A função de variáveis canônicas tem a execução bastante simples. Nomeada *VariáveisCanônicas* (), é composta pelo argumento de dados e o modelo referente ao delineamento utilizado. Além de alguns argumentos acessórios para melhorar a qualidade

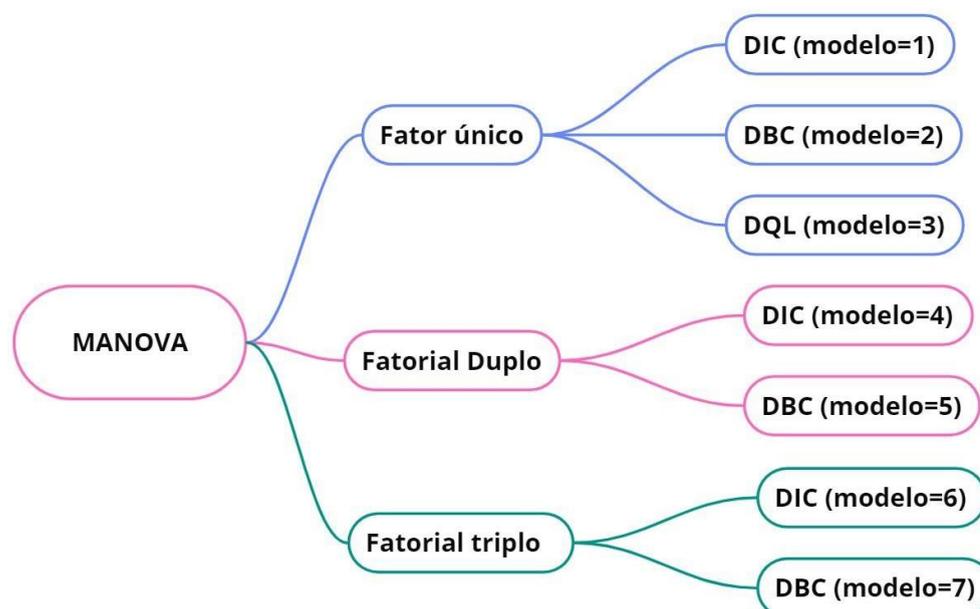
gráfica dos resultados, como as funções *CorCol*, *CorPlot*, *VarCol* e *bty* – que representam as cores, presença de seta, cores do texto e bordas, respectivamente –, além do gráfico de componentes principais, os resultados gerados no *MultivariateAnalysis* também apresentam o estudo delas, com a explicação, os escores e a importância deles.



**Figura 22.** Métodos de dispersão realizados pelo *MultivariateAnalysis*

### 23. ANÁLISE DE VARIÂNCIA MULTIVARIADA

Dentre as análises de dados multivariados, existe a análise de variância multivariada (MANOVA- *multivariate analysis of variance*), e essa, de acordo Fávero *et al.* (2009), pode ser definida como uma extensão da análise de variância univariada, com a introdução de mais variáveis dependentes. De acordo com Hair *et al.* (2009), essa técnica tem aplicação na realização de comparações entre grupos em múltiplas variáveis dependentes. No *MultivariateAnalysis*, é possível a realização de MANOVA fator único, duplo ou triplo, em quaisquer delineamentos (Figura 23).



**Figura 23.** Análises de variância multivariada realizadas pelo *MultivariateAnalysis*

Além do *MultivariateAnalysis*, existem pacotes nos quais podem ser realizadas MANOVA; entre eles, o *car* (FOX & WEISBERG, 2012) e o *Candisc* (FRIENDLY & FOX, 2013). Uma das principais características do *MultivariateAnalysis* é a simplicidade de realização das análises. Esse fato também se aplica a MANOVA, pois os comandos para execução dessa função são dois: a importação dos dados e a função *MANOVA()*, com os argumentos de identificação dos dados e o modelo de fatorial e delineamento escolhido. Abaixo, apresenta-se exemplo, para fator único, em delineamento inteiramente casualizado (DIC).

```

data("DADOS")
MANOVA(DADOS, 1)

```

## 24. CONSIDERAÇÕES FINAIS E MELHORIAS FUTURAS

O pacote *MultivariateAnalysis* foi desenvolvido com o intuito de facilitar a realização de análises multivariadas, de uma maneira fácil, eficaz e bastante intuitiva, agilizando as atividades do pesquisador em relação à realização de análises estatísticas. No pacote, os usuários poderão encontrar as mais diversas análises para o nicho multivariado. O *MultivariateAnalysis* conseguiu reunir em um único pacote diversas funções que estavam dispersas em vários outros.

Em 2024, o *MultivariateAnalysis* se encontra em sua versão 5.0 e em processo de atualização para o inglês, possibilitando o acesso por pesquisadores de todas as partes do mundo. Esse pacote está disponível de forma online e gratuita, dentro do repositório CRAN,

com registros transparentes e reproduzíveis, o que possibilita que novas funções sejam implementadas em breve.

## 25. REFERÊNCIAS

- ALLAIRE, J. RStudio: integrated development environment for R. **Boston, MA**, v. 770, n. 394, p. 165-171, 2012.
- ANDERSON, E. Edgar Anderson's Iris Data- R. 1935. Disponível em: <<https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/iris.html>>.
- ANDRADE, Elisângela Knoblauch Viega de *et al.* Genetic dissimilarity among sweet potato genotypes using morphological and molecular descriptors. **Acta Scientiarum. Agronomy**, v. 39, n. 4, p. 447-455, 2017.
- BENITO, Blas M.; BIRKS, H. John B. distantia: an open-source toolset to quantify dissimilarity between multivariate ecological time-series. **Ecography**, v. 43, n. 5, p. 660-667, 2020.
- BUTTS, Carter T.; BUTTS, Maintainer Carter T. Package 'yacca'. 2022.
- COMPARAÇÃO ENTRE AS REDES NEURAIS ADALINE E PERCEPTRON UTILIZANDO O CONJUNTO DE DADOS IRIS. **Colloquium Exactarum**. ISSN: 2178-8332, [S. l.], v. 13, n. 1, p. 1-8, 2021. Disponível em: <https://journal.unoeste.br/index.php/ce/article/view/3769>. Acesso em: 9 jun. 2024.
- CRUZ, C. D. GENES. Programa Genes - Ampliado e integrado aos aplicativos R, Matlab e Selegen. **Acta Scientiarum Agronomy**, Maringá, v.38, n.4, p.547-552, 2016.
- DA SILVA, Anderson Rodrigo; DA SILVA, Maintainer Anderson Rodrigo. Package 'biotools'. **Available online at: <https://CRAN.R-project.org/package=~biotools>**, 2017.
- DE VRIES, Andrie; RIPLEY, Brian D.; DE VRIES, Maintainer Andrie. Package 'ggdendro'. 2024.
- DRAY, Stéphane; DUFOUR, Anne-Béatrice. The ade4 package: implementing the duality diagram for ecologists. **Journal of statistical software**, v. 22, p. 1-20, 2007.
- FÁVERO, L. P. L.; BELFIORE P.P.; SILVA, F.L.; CHAN, B.L. Análise de dados: modelagem multivariada para tomada de decisões. 2009.
- FERREIRA, Márcia Miguel Castro. Quimiometria III-Revisitando a análise exploratória dos dados multivariados. **Química Nova**, v. 45, n. 10, p. 1251-1264, 2022.
- FOX, John *et al.* Package 'car'. **Vienna: R Foundation for Statistical Computing**, v. 16, n. 332, p. 333, 2012.
- FRIENDLY, M.; FOX, J. (2013) candisc: Visualizing generalized canonical discriminant and canonical correlation analysis. R package version 0.6-5. Disponível em: <http://CRAN.R-project.org/package=candisc>
- GALILI, Tal. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. **Bioinformatics**, v. 31, n. 22, p. 3718-3720, 2015.

- GONZÁLEZ, I. .; DÉJEAN, S.; MARTIN, P. . G. P.; BACCINI, A. CCA: An R Package to Extend Canonical Correlation Analysis. **Journal of Statistical Software**, [S. l.], v. 23, n. 12, p. 1–14, 2008. DOI: 10.18637/jss.v023.i12. Disponível em: <https://www.jstatsoft.org/index.php/jss/article/view/v023i12>. Acesso em: 10 jun. 2024.
- GOUVEA, M. A.; PREARO, L. C.; ROMEIRO, M. do C.. Avaliação do emprego da técnica de análise multivariada de variância em teses e dissertações dos programas de pós-graduação em administração da Universidade de São Paulo e da Universidade Federal do Grande ABC. 2011.
- HAIR, J. F.; BLACK, W.C.; BABIN, B.J.; ANDERSON R.E.; TATHAM, R.L. Análise multivariada de dados. **Bookman** editora, 2009.
- HURLEY, Catherine B.; EARLE, Denise; HURLEY, Maintainer Catherine. Package ‘DendSer’. 2022.
- HUSSON, F.; JOSSE, F.; LÊ, S. Package ‘factominer’. **An R package**, v. 96, n. 96, p. 698, 2016.
- JONGMAN, R.H.G.; TER BRAAK, C.J.F. & VAN TONGEREN, O.F.R. Data analysis in community and landscape ecology. 2nd ed., Cambridge, **Cambridge Univ. Press**, 1995.
- KASSAMBARA, Alboukadel; MUNDT, Fabian. Package ‘factoextra’. **Extract and visualize the results of multivariate data analyses**, v. 76, n. 2, 2017.
- LUCAS, Antoine. amap: Another multidimensional analysis package. [http://CRAN.R-project.org/package= amap](http://CRAN.R-project.org/package=amap), 2014.
- MONTERO, Pablo; VILAR, José A. TSclust: An R package for time series clustering. **Journal of Statistical Software**, v. 62, p. 1-43, 2015.
- OKSANEN, J., BLANCHET, F. G., FRIENDLY, M., KINDT, R., LEGENDRE, P., MCGLINN, D., MINCHIN, P. R., O'HARA, R. B., SIMPSON, G. L., SOLYMOS, P., STEVENS, M. H. H., SZOECES, E., & WAGNER, H. (2018). Vegan: **Community Ecology Package**. R package version 2.5-3.
- OLIVOTO, T.; LÚCIO, A. D. metan: An R package for multi-environment trial analysis. **Methods in Ecology and Evolution**, v. 11, n. 6, p. 783-789, 2020.
- PATHBERIYA, H. A. Disimformixed: Calculate Dissimilarity Matrix For Dataset With Mixed Attributes [Computer Software Manual]: R Foundation For Statistical Computing. 2016.
- R CORE TEAM. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>
- RODRIGUES, Francisco A.; DE ARRUDA, Guilherme Ferraz; COSTA, Luciano da Fontoura. A complex networks approach for data clustering. **arXiv preprint arXiv:1101.5141**, 2011.
- SILVA, J. M. P. .; CAVICHIOLI, F. A. . O USO DA AGRICULTURA 4.0 COMO PERSPECTIVA DO AUMENTO DA PRODUTIVIDADE NO CAMPO. **Revista Interface Tecnológica**, [S. l.], v.

17, n. 2, p. 616–629, 2020. DOI: 10.31510/infa.v17i2.1068. Disponível em: <https://revista.fatectq.edu.br/interfacetecnologica/article/view/1068>. Acesso em: 6 jun. 2024.

