UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE CIÊNCIAS

DEPARTAMENTO DE ESTATISTICA E MATEMÁTICA APLICADA

CURSO DE GRADUAÇÃO EM ESTATÍSTICA

VITÓRIA DE ANDRADE ALVES

TESTES DE PERMUTAÇÃO OU ALEATORIZAÇÃO NO R- PERMANOVA

FORTALEZA

2022

VITÓRIA DE ANDRADE ALVES

TESTES DE PERMUTAÇÃO OU ALEATORIZAÇÃO NO R- PERMANOVA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Estatística do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Estatística.

Orientadora: Prof. Dra. Sílvia Maria de Freitas

FORTALEZA

2022

VITÓRIA DE ANDRADE ALVES

TESTES DE PERMUTAÇÃO OU ALEATORIZAÇÃO NO R- PERMANOVA

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Estatística do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Estatística.

Aprovada em:

BANCA EXAMINADORA

Prof. Dra. Sílvia Maria de Freitas (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Gualberto Segundo Agamez Montalvo
Universidade Federal do Ceará (UFC)

Prof. Dr. Luis Gustavo Bastos Pinho
Universidade Federal do Ceará (UFC)

À Deus.

À minha família, em especial minhas irmãs Camila e Vanessa, pois sem a ajuda delas eu não teria chegado até aqui.

AGRADECIMENTOS

Agradeço primeiramente à Deus por ter me dado forças para chegar até aqui. À minha família, em especial as minhas irmãs por terem me dado suporte quando precisei.

Aos meus colegas que fizeram parte do curso, Maria Bianca e Guilherme por terem estado sempre comigo nos momentos mais complicados durante o começo da faculdade. Também aos amigos que fiz e me deram apoio na reta final do curso.

Ao professor Gualberto por ter confiado em mim e me ajudado nessa reta final. E, a minha orientadora professora Silvia, por ter aceitado me orientar neste trabalho e por toda compreensão até aqui.

"O coração do homem traça o seu caminho, mas o Senhor lhe dirige os passos."

(Provérbios-16:9)

RESUMO

Testes de permutação e aleatorização são muito utilizados em estudos de ecologia, onde as variáveis geralmente consistem de contagens de abundâncias (ou porcentagem de cobertura, frequências ou biomassa) para um grande número de espécies, porém também pode ser utilizados em outras àreas. Métodos computacionalmente intensivos para inferência estatística são usados quando as abordagens tradicionais não são adequadas. Sendo estes, os testes de permutação ou aleatorização, onde a ideia básica implícita nestes testes é a de calcular uma estatística de teste para cada permutação dos dados, de modo a construir o conjunto de referência a partir do qual se determina a probabilidade associada à estatística de teste observada. Para tais procedimentos de permutação e aleatorização não se faz necessário estabelecer suposições sobre a distribuição teórica, sendo considerados testes livres de distribuições. Porém possuem limitações, só podem ser usados para hipóteses que envolvam comparações (trocar observações entre grupos) ou desalinhar registros (como em correlação, por exemplo) ou hipóteses que dizem que as observações para um grupo estão em uma ordem aleatória (onde a aleatorização envolve a geração de ordens aleatórias alternativas). Portanto, não podem ser usados para testar hipóteses sobre parâmetros individuais. Estes testes são baseados na Análise de Variância (ANOVA) para análise multivariada com base na permutação. No software R temos os testes "Análise de Variância Permutacional Multivariada (PERMANOVA)"e a "Análise de Similaridade (ANOSIM)"para teste de permutação de uma matriz de distância envolvendo uma variável categórica. Além do "teste de Mantel", que consiste em calcular a correlação entre duas matrizes, com base na permutação. Estes testes podem ser encontrados no pacote “vegan: Community Ecology Package”. Para modelos lineares, os testes de permutação são úteis quando a suposição de normalidade é violada ou quando o tamanho da amostra é muito pequeno para aplicar a teoria assintótica. Estes são tipicos testes de significância onde a distribuição estatistica testada é obtida calculando-se todos os possíveis valores desta estatística rearranjando-se os valores da amostra considerando uma hipótese nula verdadeira, sendo uma maneira de determinar se o valor da hipótese nula é razoável. Assim, o objetivo principal deste trabalho é apresentar alguns testes baseados em permutação ou aleatorização, aplicados em dois casos. Em situação para um grupo-matriz e a segunda para dois grupos, com duas matrizes.

Palavras-chave: Teste de Permutação, Permanova, Anosim, Teste de Mantel.

ABSTRACT

Permutation and randomization tests are widely used in ecology studies, where the variables usually consist of abundance counts (or percentage coverage, frequencies or biomass) for a large number of species, but can also be used in other areas. Computationally intensive methods for statistical inference are used when traditional approaches are not adequate. These being the permutation or randomization tests, where the basic idea implicit in these tests is to calculate a test statistic for each permutation of the data, in order to build the reference set from which the probability associated with the statistic is determined. observed test. For such permutation and randomization procedures, it is not necessary to establish assumptions about the theoretical distribution, being considered distribution-free tests. However, they have limitations, they can only be used for hypotheses that involve comparisons (swapping observations between groups) or misaligned records (as in correlation, for example) or hypotheses that say that the observations for a group are in a random order (where randomization involves the generation of alternative random orders). Therefore, they cannot be used to test hypotheses about individual parameters. These tests are based on ANOVA for multivariate analysis based on permutation. In the R software we have the tests "PERMANOVA"and "ANOSIM"for testing the permutation of a distance matrix involving a categorical variable. In addition to the "Mantel test", which consists of calculating the correlation between two matrices, based on the permutation. These tests can be found in the “vegan: Community Ecology Package”. For linear models, permutation tests are useful when the assumption of normality is violated or when the sample size is too small to apply asymptotic theory. These are typical significance tests where the tested statistical distribution is obtained by calculating all possible values of this statistic rearranging the sample values considering a true null hypothesis, being a way to determine if the value of the null hypothesis is reasonable. Thus, the main objective of this work is to present some tests based on permutation or randomization, applied in two cases. In situation for a matrix-group and the second for two groups, with two matrices.

Keywords: Permutation Test, Permanova, Anosim, Mantel Test.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

ANOSIM	Análise de Similaridade
ANOVA	Análise de Variância
MANOVA	Análise Multivariada da variância
NMDS	Escalonamento multidimensional não-métrico
PCO	Análise de coordenadas principais
PERMANOVA	Análise de Variância Permutacional Multivariada
PERMDISP	Teste de homogeneidade de dispersões multivariadas
RDA	Análise de redundância clássica

SUMÁRIO

1 INTRODUÇÃO

Existem vários métodos de análise multivariada, com finalidades bem diversas entre si e o pesquisador deve avaliar de forma cautelosa ao trabalhar com as técnicas de análise multivariada, para que a opção escolhida possa detectar os padrões esperados nos dados em relação ao método a ser escolhido de acordo com o tipo de pesquisa. Pesquisas metodológicas recentes produziram métodos de permutação para testar parâmetros em presença de variáveis incômodas em modelos lineares ou ANOVA de medidas repetidas. Existem muitas variações da ANOVA devido aos diferentes tipos de delineamentos que podem ser realizados.

A ANOVA Baseia-se na decomposição da variação total da variável resposta em partes que podem ser atribuídas aos tratamentos(variância entre) e ao erro experimental (variância dentro). Em uma ANOVA, examinamos as diferenças estatísticas de uma variável dependente contínua por uma variável de agrupamento independente. E também pode determinar se as médias de n grupos são diferentes. É utilizado o teste F para verificar a igualdade entre as médias. Existe alumas pressuposições básicas para se utilizar a Anova: As amostras são aleatórias e independentes; As populações tem distribuição normal. As variâncias populacionais são iguais. Para saber se as médias dos grupos são iguais. Ela pode ser utilizada para dois ou mais grupos. A partir da ANOVA surgiram testes de análise multivariada baseados na permutação, chamados de testes de aleatorização que possuem abordagem baseada em permutação das observações, em reamostragem e/ou simulação.

Um teste de aleatorização é um teste estatístico cuja validade tem por base a distribuição aleatória das unidades experimentais pelos tratamentos. Assim, o teste de aleatorização tem por base o modelo de distribuição aleatória, enquanto os testes estatísticos clássicos, como o teste t de Student ou o teste F da Análise de Variância, têm por base o modelo de amostragem aleatória. Também é neste mesmo modelo da amostragem aleatória, que se baseiam os testes de permutação (BRANCO, 2010).

O teste de aleatorização é uma maneira de determinar se o valor da hipótese nula é razoável neste tipo de situação. Uma estatística *S* é escolhida para medir até que ponto os dados mostram o padrão em questão. O valor *s* de *S* para os dados observados é então comparado com a distribuição de *S* que é obtida pela reordenação aleatória dos dados. O argumento feito é que se a hipótese nula for verdadeira, então todas as ordens possíveis para os dados eram igualmente prováveis de terem ocorrido. Os dados observados ordem é então apenas uma das ordens igualmente prováveis, e s deve aparecer como um valor típico da distribuição de aleatorização

de *S*. Se isso não não parece ser o caso (de modo que s é significativo), então a hipótese nula é desacreditada até certo ponto e, por implicação, a alternativa hipótese é considerada mais razoável (ROBINSON *et al.*, 2007).

Como citado anteriormente, temos alguns testes multivariados baseados na permutação disponível no software R. No pacote “vegan”, encontramos a “permanova” e ”anosim”, essas com base na anova. E o teste de mantel, que também é um teste de permutação. A PERMANOVA = ANOVA por Permutação, pode ser adequado quando a intenção é de permutar a matriz de distância, gerando os valores das classes ao acaso. Portanto se o resultado da matriz original for muito improvável de ser encontrado ao acaso, rejeitamos a hipótese nula, e aceitamos que de fato existe uma associação entre as classes ou grupos com a matriz de distâncias. Como o Escalonamento multidimensional não-métrico (NMDS) é a representação da matriz de distâncias, pode-se utilizar o NMDS para representar os resultados da PERMANOVA, o mesmo vale para um Cluster. "PERMANOVA"= ANOVA adaptada para uma Matriz de Distâncias. Pode-se utilizar para Matrizes de Distâncias criadas pelos Métodos: Euclidiano, Jaccard e Bray-Curtis (REFFATTI, 2019).

Para testes envolvendo permutação também temos o ANOSIM - teste de permutação de uma matriz de distância envolvendo uma variável categórica, com rankeamento dos dados. Objetivo: Aplicar o teste estatístico ANOSIM para dar rigor estatístico aos os agrupamentos formados por meio de uma variável categórica e visualizados através de um NMDS/PCA/Cluster de uma matriz de distância. A lógica da ANOSIM é a de permutar a matriz de distância, com os dados RANKEADOS, gerando os valores das variáveis categóricas ao acaso. Portanto se o resultado da matriz original for muito improvável de ser encontrado ao acaso, rejeitamos a hipótese nula, e aceitamos que de fato existe uma associação entre as categorias (fatores/classes) com a matriz de distâncias visualizados através do NMDS/Cluster/PCA (REFFATTI, 2019).

Já teste de Mantel teste de permutação, compara duas matrizes de distância. Sendo uma a matriz de distância da composição de espécies pelo método Jaccard (Presença/Ausência) e a outra a matriz de distâncias das variáveis ambientais padronizadas pelo método Euclidiano. Objetivo: Realizar o Teste de Mantel para comparar duas matrizes de distância, sendo uma a matriz de distâncias das variáveis espécies através do método Jaccard e a outra da matriz de distâncias das variáveis ambientais através do método euclidiano. Este teste calcula a correlação entre as duas matrizes (correlação de Pearson). O detalhe é que este teste faz inúmeras permutações nos valores das duas matrizes e calcula o valor de R(correlação) em cada uma

permutação de matrizes. Após isso, busca o valor de R(correlação) real dos dados e faz um teste estatístico buscando comparar o R real dos dados com todos os R originados das inúmeras permutações. Isto busca saber se o valor de R(correlação) realmente existe, ou pode ser resultado simplesmente do acaso (REFFATTI, 2019).

A proposta do trabalho é fazer uma aplicação de teste de aleatorização ou permutação no software R para dois casos envolvendo matrizes, onde no primeiro caso seria aplicado um grupo um teste com uma matriz e outro para dois grupos, com duas matrizes. Essa monografia esta subdividida de acordo com as ordens de relevância do assunto para o melhor entendimento ao final, sendo da seguinte forma no Capítulo 2 são apresentados os testes de aleatorização e permutação, para nos capítulos seguintes apresentar os testes de permutação e aleatorização disponíveis no $software R$ dos capítulos 3 ao 5, posteriormente as aplicações e as considerações finais.

2 TESTE DE PERMUTAÇÃO OU ALEATORIZAÇÃO

Muitas vezes o pesquisador está interessado em comparar médias ou a forma da distribuição de dois grupos. Uma maneira para compará-los seria aplicando testes paramétricos, tais como o Teste T ou Teste Z (no caso de duas amostras independentes) ou o Teste T pareado. Porém, tais testes apresentam certas exigências que frequentemente podem não ser atendidas. Neste caso, é indicada a utilização de testes não paramétricos ou o teste de aleatorização. Este teste é baseado na suposição de que, se a hipótese nula é verdadeira, todas as possíveis ordens dos dados são igualmente prováveis. O teste de aleatorização é um procedimento em que se comparam valores de uma estatística observada para os dados no arranjo original com os valores desta estatística após a aleatorização das observações. A regra de decisão é baseada no p-valor - proporção de vezes em que a estatística de teste com os aleatorizados é maior ou igual a estatística de teste com os dados do arranjo original (FILHO *et al.*, 2010).

Um teste de aleatorização é válido para qualquer tipo de amostra, independentemente de como o amostra é selecionada. Esta é uma propriedade extremamente importante porque o uso de amostras não aleatórias é comum na experimentação, e tabelas estatísticas paramétricas (por exemplo, tabelas t e F) não são válidas para tais amostras. As tabelas estatísticas paramétricas são aplicáveis apenas a amostras aleatórias, e sua invalidade de aplicação a amostras não aleatórias é amplamente reconhecida. Os testes de aleatorização são extremamente versáteis devido ao seu potencial para garantir a validade dos testes estatísticos existentes e para desenvolver novos testes especiais. Os testes de aleatorização oferecem a oportunidade de desenvolver novas estatísticas de teste personalizadas e analisar dados de experimentos envolvendo procedimentos de atribuição aleatória não convencionais. A versatilidade dos testes de aleatorização está relacionada ao aumento do poder ou sensibilidade dos testes estatísticos de várias maneiras (EDGINGTON; ONGHENA, 2007). Na literatura os termos aleatorização e permutação, são associados à um só. Muitos autores utilizam a expresão "Testes de aleatorização também conhecido como testes de permutação"ou vice-versa.

2.1 Vantagens e desvantagens dos teste de aleatorização ou permutação:

Em comparação com métodos estatísticos mais padronizados, os testes de aleatorização têm duas vantagens principais. Primeiro, eles são válidos mesmo sem amostras aleatórias. Em segundo lugar, muitas vezes é relativamente fácil levar em conta as peculiaridades da situação de interesse e usar teste fora do padrão Estatisticas (ROBINSON *et al.*, 2007).

Há uma desvantagem com testes de aleatorização que podem aparecer à primeira vista grave: não é necessariamente possível generalizar as conclusões de um teste de aleatorização para uma população de interesse. Há uma limitação óbvia com testes de aleatorização nos mais estritos sentido em que eles só podem ser usados para testar hipóteses envolvendo comparações entre dois ou mais grupos (onde a aleatorização o envolve trocando observações entre grupos), ou hipóteses que dizem que as observações para um grupo estão em uma ordem aleatória (onde a aleatorização envolve a geração de ordens aleatórias alternativas). Por sua própria natureza, não são aplicáveis para testar hipóteses sobre valores absolutos de parâmetros de populações (ROBINSON *et al.*, 2007).

2.2 Calculo do P-valor

O teste de aleatorização é uma maneira de determinar se o valor da hipótese nula é razoável neste tipo de situação. Uma estatística *S* é escolhida para medir até que ponto os dados mostram o padrão em questão. O valor *s* de *S* para os dados observados é então comparado com a distribuição de *S* que é obtida pela reordenação aleatória dos dados. O nível de significância de *s* é a proporção ou porcentagem de valores que são tão extremos ou mais extremos que esse valor na distribuição de aleatorização. Isso pode ser interpretado da mesma forma que para testes convencionais de significância. Se s for menor que 5%, isso fornece alguma evidência de que a hipótese nula não é verdadeira; se *s* for menor que 1%, então fornece forte evidência de que a hipótese nula não é verdadeira; e se *s* for menor que 0,1%, então fornece uma evidência muito forte que a hipótese nula não é verdadeira. Para evitar a caracterização pertencer a “aquele grupo de pessoas cujo objetivo na vida é estar errado 5% da época” (Kempthorne e Doerfler, 1969), é melhor considerar o nível de significância como uma medida da força da evidência contra (ROBINSON *et al.*, 2007).

2.3 Matrizes de distância

Uma matriz de distância é uma matriz não negativa, quadrada e simétrica com elementos correspondentes a estimativas de alguma distância pareada entre as sequências em um conjunto. (WEYENBERG; YOSHIDA, 2015). Quanto menor for essas distâncias entre os pontos, maior será a similaridade entre elas. Já um índice de distância corresponde a uma dissimilaridade. Alguns índices de dissimilaridade citados na função *vergdist* do pacote *vegan* são, "manhattan", "euclidean", "canberra", "clark", "bray", "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford", "raup", "binomial", "chao", "cao", "mahalanobis", "chisq", "chord", "aitchison", ou "robust.aitchison". Onde será abordado a seguir os mais utilizados, como "jaccard", "euclidean"e "bray"(OKSANEN *et al.*, 2022).

2.3.1 Coeficiente de Jaccard

O índice de Jaccard (ou coeficiente de Jaccard) indica a proporção de espécies compartilhadas entre duas amostras em relação ao total de espécies (ECOVIRTUAL,). Uma forma de calculá-lo é:

$$J = \frac{Scom}{s1 + s2Scom}$$

O que é o mesmo que:

$$J = \frac{Scom}{S}$$

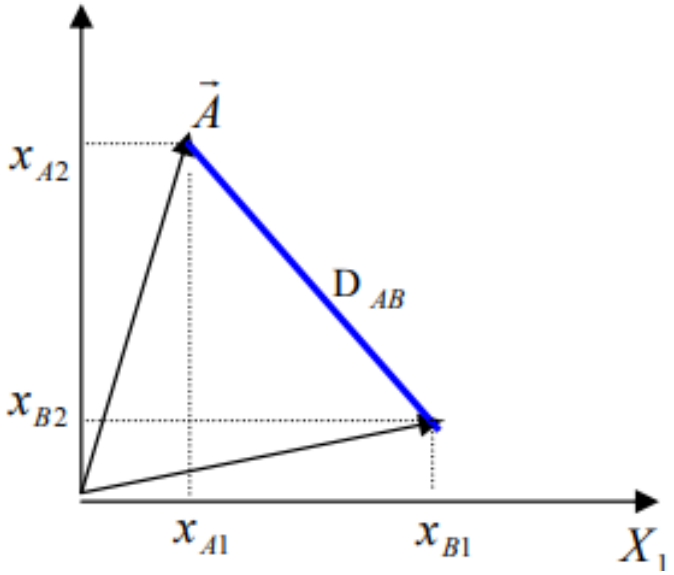
onde:

- S_{com} é o número de espécies em comum nas duas amostras.
- s_1 e s_2 é o número total de espécies em cada uma das amostras.
- S é o total de espécies no conjunto de amostras.

2.3.2 Distância Euclidiana

Considerando o caso mais simples, no qual existem n indivíduos, onde cada um dos quais possuem valores para p variáveis, a distância euclidiana entre eles é obtida mediante o teorema de Pitágoras para um espaço multidimensional. Esta distância é uma medida de semelhança e pode ser expressa pela distância D entre as extremidades de dois vetores (VICINI, 2005). Como mostra a Figura 1.

Figura 1 – Distância euclidiana D_{AB} entre dois vetores A e B.



Fonte: Valentin (2000).

A distância euclidiana é calculada com base no teorema de Pitágoras:

$$D_{A,B} = \sqrt{(x_{A1} - x_{B1})^2 + (x_{A2} - x_{B2})^2}$$

ou ainda, generalizando para duas amostras contendo m espécies, a distância euclidiana é dada por:

$$D_{A,B} = \sqrt{\sum_{j=1}^{m} (x_{A,J} - x_{B,J})}$$

2.3.3 Coeficiente de Bray-Curtis

A Distância de Bray & Curtis(1957) é de uso frequente, por ser disponível na maioria dos pacotes estatísticos. Ela varia entre 0 (similaridade) e 1 (dissimilaridade). Este índice não considera as duplas-ausências e é fortemente influenciado pelas espécies dominantes. As espécies raras acrescentam muito pouco ao seu valor. Seu cálculo é baseado nas diferenças absolutas e nas somas das abundâncias de cada espécie (i) nas duas amostras:

$$D_{A-B} = \frac{\sum |x_{Ai} - x_{Bi}|}{\sum (x_{Ai} + x_{Bi})}$$

Vários autores preferem definir esta medida como "Similaridade", fazendo $(1 - D)$. Neste caso, o índice de Bray & Curtis equivale ao coeficiente de similaridade de Czekanowski.

Outras medidas de distância podem ser encontradas na literatura, com formulação parecida à de Bray & Curtis, tais como as distâncias de Manhattan e de Camberra (VALENTIN, 1995).

3 PERMANOVA

A análise de variância permutacional multivariada (PERMANOVA) é um particionamento de variação em uma nuvem de dados multivariada, definida explicitamente no espaço de uma medida de dissimilaridade escolhida, em resposta a um ou mais fatores em uma análise de projeto de variância. O método é semiparamétrico, motivado pelo desejo de realizar um particionamento clássico, como na ANOVA (portanto, permitindo testes e estimativas de tamanhos de efeitos principais, termos de interação, estruturas hierárquicas, componentes aleatórios em modelos mistos, etc.), mantendo simultaneamente importantes estatísticas robustas propriedades de métodos multivariados não paramétricos baseados em classificação, como a análise de semelhanças (ANOSIM), ou seja, a flexibilidade para basear a análise em uma medida de dissimilaridade de escolha (tal como Bray–Curtis, Jaccard, etc) (ANDERSON, 2014). Uma extensão da ANOVA é a Análise Multivariada da variância (MANOVA), porém a MANOVA usual (paramétrica) tem suposições relativamente restritivas e para que os resultados sejam válidos a maiorias destes testes se baseiam em alguns pressupostos, tais como a linearidade e a normalidade multivariada dos dados, não sendo robusta quanto a falta de normalidade dos dados. Uma alternativa para tal problema é a PERMANOVA ou NPMANOVA que é um método livre de distribuição.

3.1 Abordagem do teste no Software R

Podemos encontrar no pacote *"vegan"* a função *"adonis2"* que corresponde a PERMANOVA. Onde usa matrizes de distância para particionar matrizes de distância entre fontes de variação e ajuste de modelos lineares, para matrizes de distância usa um teste de permutação com razões pseudo-F (ANDERSON; WALSH, 2013).

adonis2 é uma função para a análise e particionamento de somas de quadrados usando dissimilaridades. A função é baseada nos princípios de McArdle Anderson (2001) e pode realizar funções sequenciais, testes marginais e globais. A função também permite usar constantes aditivas ou raiz quadrada de dissimilaridades para evitar autovalores negativos, mas também pode manipular índices semimétricos (como Bray-Curtis) que produzem autovalores negativos. Os testes do *Adonis2* são idênticos ao *anova.cca* do *dbrda*. Com distâncias euclidianas, os testes também são idênticos ao *anova.cca* de *rda* (OKSANEN *et al.*, 2013).

A função particiona somas de quadrados de um conjunto de dados multivariado, e elas

são diretamente análogas a MANOVA. McArdle e Anderson (2001) e Anderson (2001) referir-se ao método como "MANOVA permutacional"(anteriormente "MANOVA não paramétrica"). Além disso, como as entradas são preditores lineares e uma matriz de resposta de um número arbitrário de colunas, eles são uma alternativa robusta tanto para MANOVA paramétrica quanto para métodos de ordenação para descrever como a variação é atribuída a diferentes tratamentos experimentais ou covariáveis não controladas (OKSANEN *et al.*, 2013).

3.2 Pseudo Estatística F

A PERMANOVA usa estatística pseudo F para testar a hipótese nula de não haver diferenças nas posições dos centróides do grupo no espaço da medida de dissimilaridade escolhida é dada por

$$F = (SS_A/SS_R) * [(N-g)/(g-1)]$$

Quando as distâncias euclidianas são usadas, as somas dos quadrados PERMANOVA são cada uma igual à soma das somas univariadas clássicas dos quadrados entre as variáveis originais; isto é, $SS_A = \sum_{k=1}^{p} SS_A^{[k]}$, onde $SS_A^{[k]}$ é a soma de quadrados univariada entre grupos para a variável $k = 1, \ldots, p$ (e similarmente para cada um de SS_R e SS_T). Portanto, pseudo F em distâncias euclidianas é o mesmo que a estatística F usada em Análise de redundância clássica (RDA). Além disso, PERMANOVA em uma variável de resposta usando distância euclidiana produz o clássico estatística F univariada. Então, PERMANOVA também pode ser usado para fazer ANOVA univariada, mas onde os valores de p são obtidos por permutação, evitando assim a suposição de normalidade (ANDERSON, 2014).

3.3 Inferência

Para casos onde há muito poucas permutações possíveis para alcançar um valor-p preciso para inferências em um nível de significância adequadamente pequeno, valores de p aproximados podem ser obtidos usando o método aleatório de Monte Carlo que extrai da distribuição de permutação assintótica. A PERMANOVA não faz suposições explícitas sobre as distribuições de variáveis originais em Y ou as distribuições de dissimilaridades em D (Matriz). Para um determinado teste, a PERMANOVA assume apenas permutabilidade de unidades permutáveis sob uma hipótese nula verdadeira. PERMANOVA é apenas “não paramétrico” para

o caso unidirecional. As inferências permanecem livres de distribuição, mas a PERMANOVA se aplica um modelo linear para o espaço de dissimilaridade; as interações são definidas por referência aos efeitos principais aditivos. De fato, uma motivação chave para o desenvolvimento do PERMANOVA foi realizar testes de interação (ANDERSON, 2014).

3.4 Homogeneidade de Dispersões Multivariadas

Tanto o ANOSIM quanto o teste de Mantel são extremamente sensíveis às diferenças nas dispersões entre os grupos; no entanto, PERMANOVA (como ANOVA) é muito robusto à heterogeneidade para designs balanceados, mas não designs desbalanceados. Além disso, a PERMANOVA, ao contrário do ANOSIM ou da MANOVA tradicional, não é sensível às diferenças na estrutura de correlação (forma) entre os grupos.

Um Teste de homogeneidade de dispersões multivariadas (PERMDISP) no espaço da dissimilaridade escolhida medida pode ser feita, quer para acompanhar a PERMANOVA, quer por si só. Este teste compara a dispersão dentro do grupo entre grupos usando o valor médio das distâncias de observações individuais para seu próprio centroide de grupo. As direções específicas das distâncias não são levadas em consideração, então o PERMDISP, assim como o PERMANOVA, não identifica diferenças nas formas das nuvens de dados entre os grupos, apenas suas propagação relativa (ANDERSON, 2014).

3.5 Distâncias entre Centróides

Chama-se centroide ou baricentro o ponto de interseção das medianas que fazem parte de um triângulo. Pode-se dizer que o centroide é o ponto no qual, se uma linha o cruza, fica dividido em dois segmentos da mesma proporção em relação à reta em questão.

Seja D uma matriz que contém distâncias euclidianas, então as distâncias entre os centróides são equivalentes às distâncias euclidianas entre as médias aritméticas calculadas separadamente para cada variável. Essa equivalência não valem, no entanto, para diferenças não euclidianas. Distâncias entre centroides com base em alguma outra medida de dissimilaridade escolhida são calculadas da seguinte forma:

- (i) obter o conjunto completo de eixos Análise de coordenadas principais (PCO) da matriz G, com cada eixo tendo sido padronizado pelo valor absoluto de seu respectivo autovalor;
- (ii) calcular aritmética médias para cada grupo separadamente ao longo de cada eixo PCO;

- (iii) para cada par de centróides$(\ell,\ell),(\ell=1,\ldots,g)$ e $(\ell'=1,\ldots,g)$, calcule as distâncias euclidianas separadamente em cada um dos dois conjuntos: um baseado nos eixos PCO correspondente a autovalores não negativos $(d_\ell^+\ell)$ e um baseado naqueles correspondentes a valores próprios negativos $(d^{\ell\ell)}$, se houver; e
- (iv) a matriz (g × g) de distâncias entre centróides no espaço da dissimilaridade medida é então $D^{[C]}=d_\ell^{[C]}\ell'$, onde $d_\ell^{[C]}\ell' = \sqrt{(d_\ell^+\ell)^2-(d_\ell^+\ell)^2}$.

As distâncias entre centroides também podem ser calculadas diretamente da matriz G (ANDERSON, 2014).

4 APLICAÇÕES

Os dados utilizados para o estudo é o *iris dataset* de (ANDERSON, 1935), disponível no *software R*. Este conjunto de dados fornece as medidas em centímetros das variáveis comprimento e largura das sépalas e comprimento e largura das pétalas, respectivamente, para 50 flores de 3 espécies (setosa, versicolor e virginica). O banco de dados contém 150 casos (linhas) e 5 variáveis (colunas) denominadas Sepal.Length, Sepal.Width, Petal.Length, Petal.Width e Species.

A princípio iremos usar a MANOVA para a análise dos dados originais. A MANOVA no R usa o teste *Pillai's Trace* para os cálculos, que é então convertido em uma estatística F quando queremos verificar a significância das diferenças médias dos grupos (RADEčIć, 2022). Como MANOVA usa mais de uma variável dependente, as hipóteses nula e alternativa são definidas por:

H_0: Os vetores médios são os mesmos para todos os grupos de espécie ou não diferem significativamente.
H_1: Pelo menos um dos vetores médios é diferente dos demais.

O teste estatístico MANOVA tem algumas suposições como: Normalidade multivariada, Linearidade, Ausência de multicolinearidade e Homogeneidade.

4.0.0.0.1

Primeiramente será apresentado a exploratória dos dados, matrizes de var-cov entre os grupos, para ter uma noção das características da base de dados estudada, e para tirar conclusões definitivas sobre esta, faremos teste de normalidade e de homogeneidade para avaliar se as matrizes de variância-covariância entre os grupos(espécies) são iguais.

4.1 Análise Exploratória dos Dados:

Na Tabela 1 está apresentado um resumo geral do comportamento descritivo das variáveis. Aqui observa-se a estatística descritiva de cada variável independentemente de cada espécie.

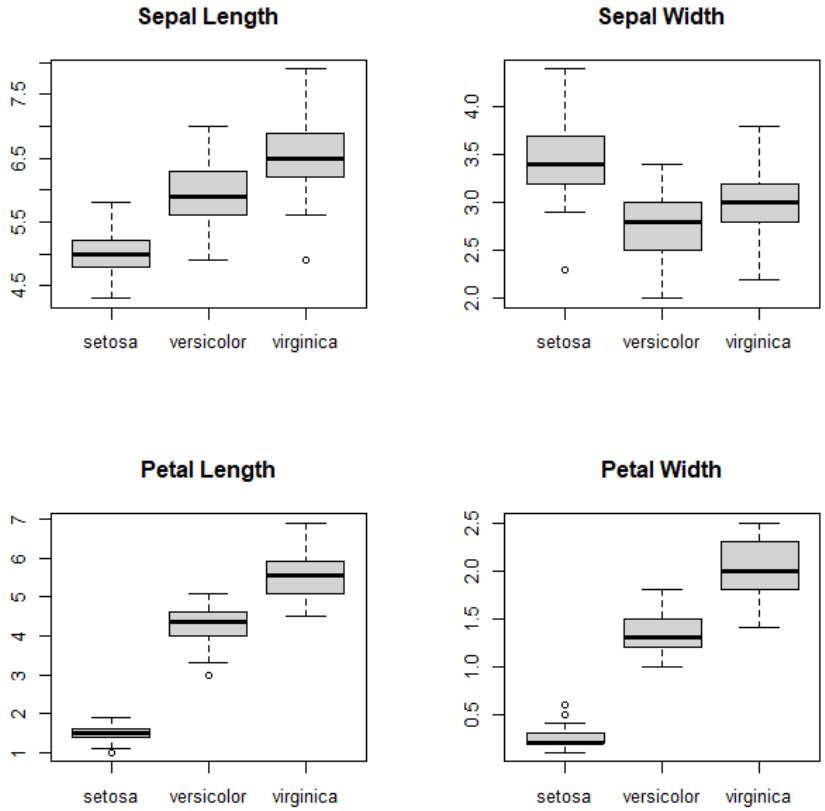
Tabela 1 – Medidas descritivas do comprimento e largura da sépala e comprimento e largura da pétala

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Mínimo	4,300	2,000	1,000	0,100
1ª Quartil	5,100	2,800	1,600	0,300
Mediana	5,800	3,000	4,350	1,300
Média	5,843	3,057	3,758	1,199
3ª Quartil	6,400	3,300	5,100	1,800
Máximo	7,900	4,400	6,900	2,500
Desvio-padrão	0,828	0,436	1,765	0,762
Coeficiente de Variação (CV %)	14,171	14,256	46,974	63,555

Fonte: elaborado pela autora (2022).

Na Figura 2, o gráfico boxplot das variáveis por espécie, mostra que existem alguns pontos de outliers. Uma das suposições para a MANOVA é de que nos dados não devem haver *outliers* nas variáveis, o que é possível observar abaixo, para quase todas as espécies.

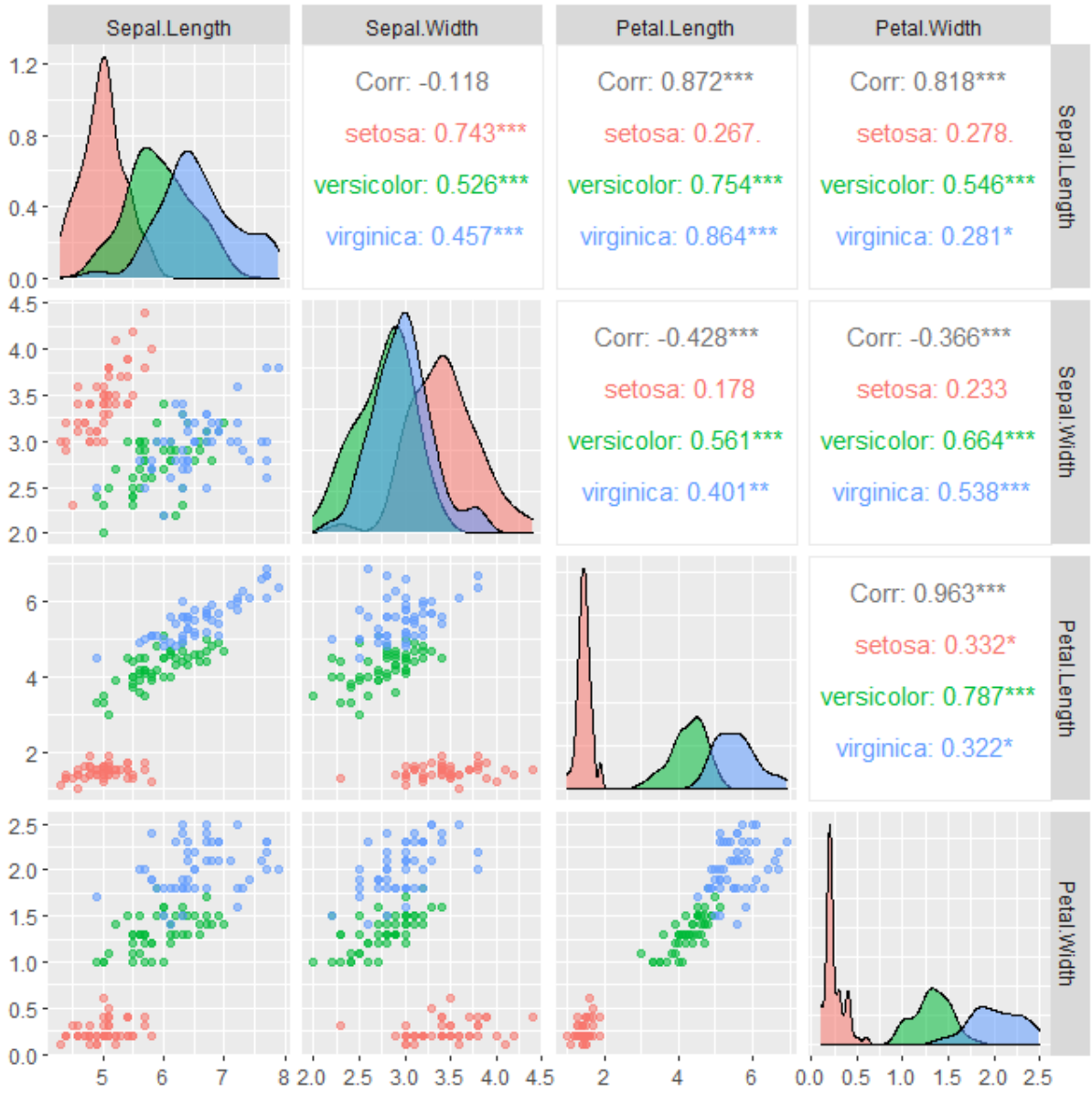
Figura 2 – Gráfico Boxplot para as variáveis, por espécie.



Fonte: Elaborado pela autora (2022).

Na Figura 3 estão representados os gráficos de dispersão, densidades e correlação entre as variáveis, por espécie.

Figura 3 – Gráficos de dispersão e densidade por espécies e correlação entre as variáveis, por espécie.



Fonte: Elaborado pela autora (2022).

4.1.0.0.1

Através da Figura 3 observa-se uma forte correlação positiva entre Petal.Length e Sepal.Length, no global (r= 0,872) e nas espécies Versicolor e Virginica, também para Petal.Width e Sepal.Length no global (r= 0,818) e para Petal.Width e Petal.Length no global (r= 0,963) e na espécie Versicolor.

Agora com a correlação das espécies em relação ao comprimento e largura de sépalas e pétalas. Pode-se notar que existe uma correlação maior nas Setosas entre a largura e comprimento, na espécie Versicolor também teve maior correlação entre o comprimento e largura das pétalas.

Observando os gráficos das densidades da Figura 3 por espécies em relação as variáveis tamanho e largura de sépalas e pétalas. Nas medidas das pétalas, a espécie setosa se destaca bem das demais com as menores medidas, tanto de comprimento quanto de largura, enquanto que as maiores medidas ficam com o grupo de virginicas. Nas medidas das Sépalas, principalmente as larguras, as setosas apresentam terem Sépalas mais largas e menos compridas que as demais, enquanto que as virginicas aparentam ter Sépalas mais compridas.

4.2 Teste para normal multivariada:

Outra suposição para a MANOVA é a normalidade multivariada, onde o conjunto de variáveis dependente deve ter uma distribuição Normal multivariada. Utilizando a função "*mvn*"no *R*, temos alguns testes para avaliar a normalidade multivariada como; "mardia"para o teste de Mardia, "hz"para o teste de Henze-Zirkler, "royston"para o teste de Royston, "dh"para o teste de Doornik-Hansen e *energy* para a estatística E. O padrão é o teste "hz"de Henze-Zirkler (KORKMAZ *et al.*, 2014).

Ao realizar o teste de *Royston* com as variáveis dependentes (Figura 4), em que a hipótese nula seria os dados virem de uma Normal de dimensão 4, o valor-p foi inferior a 0,05 dessa forma rejeitamos H_0 ou seja, os dados não possuem distribuição Normal.

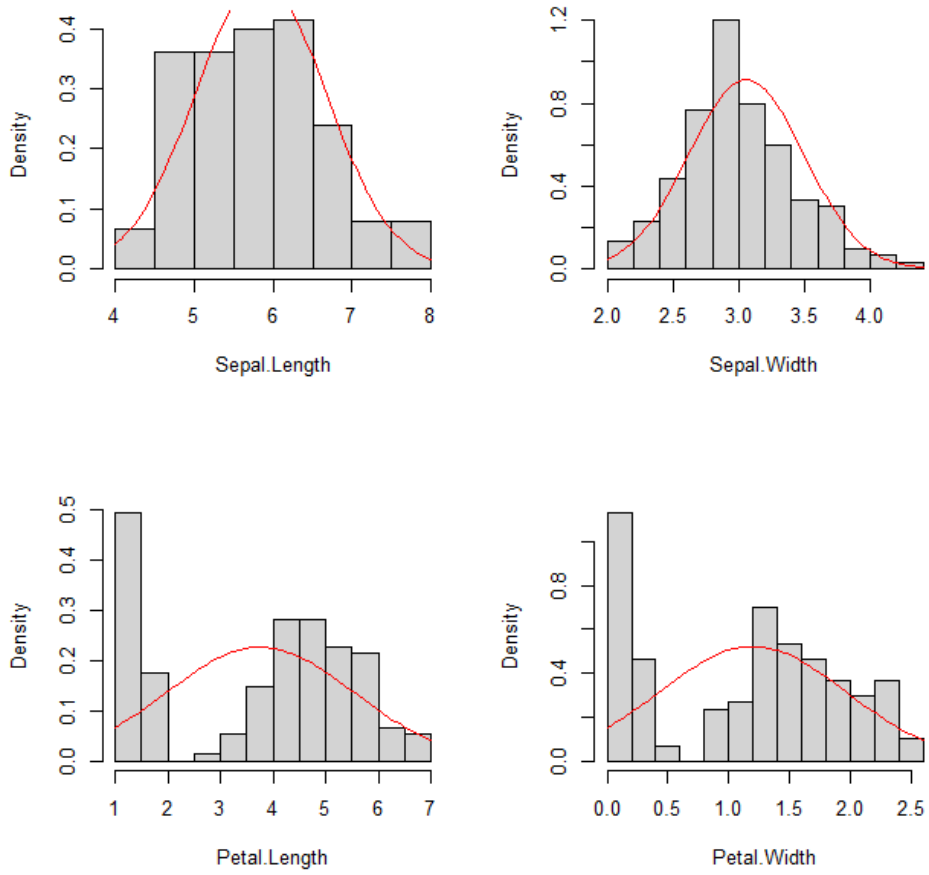
Figura 4 – Teste de Normalidade Multivariada gerado pela função "mvn"do *R*

```
> mvn(data = iris[,-5], mvnTest = "royston")
$multivariateNormality
     Test         H       p value MVN
1 Royston 50.39667 3.098229e-11  NO

$univariateNormality
               Test      Variable Statistic   p value
                  Normality
1 Anderson-Darling Sepal.Length     0.8892  0.0225      NO
2 Anderson-Darling Sepal.Width      0.9080  0.0202      NO
3 Anderson-Darling Petal.Length     7.6785  <0.001      NO
4 Anderson-Darling Petal.Width      5.1057  <0.001      NO
```

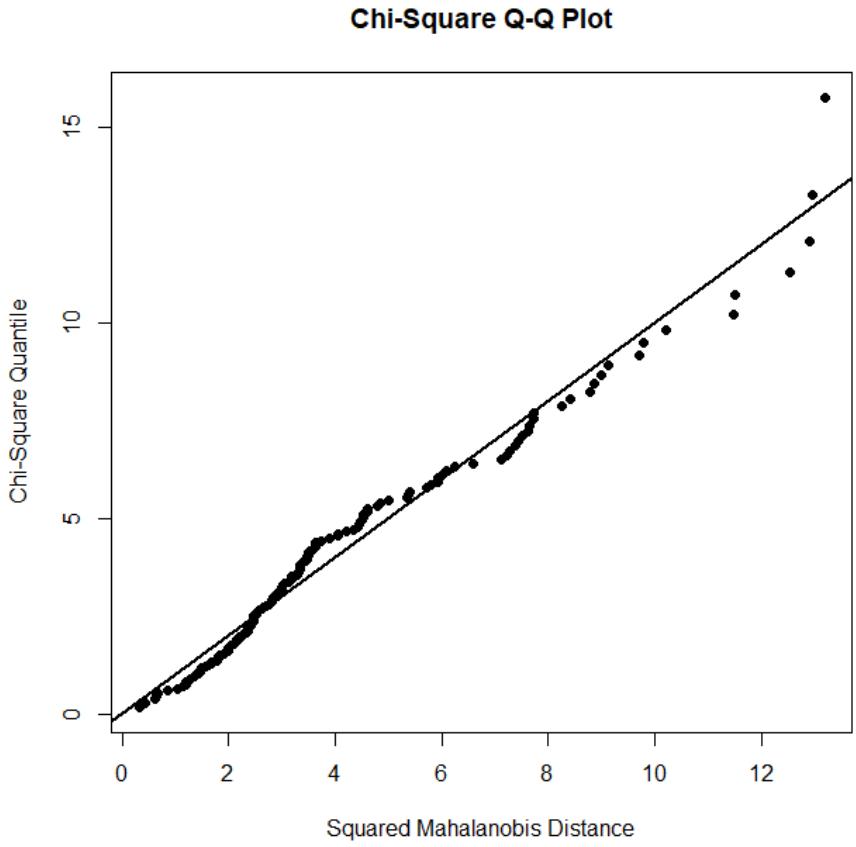
Os testes mostram a normalidade multivariada e a normalidade univariada dos dados *Iris*. O teste de *Royston* baseia-se na estatística de *Shapiro-Wilk* para testar a normalidade multivariada (LEITE, 2018). Para testar a normalidade univariada foi utilizado o teste de *Anderson-Darling*, podemos ver o comportamento graficamente dessas variáveis na Figura 5, através dos respectivos histogramas.

Figura 5 – Histogramas e curvas de densidade da largura e comprimento das sépalas e pétalas.



Como visto nos testes da Figura 4 os dados não provém de uma Normal (multivariada ou univariada, individualmente). Na Figura 5 estão os histogramas para cada variável, gerados pela função *"mvn"* do *R*, na opção "univariatePlot = histogram". Na Figura 6 podemos ver o gráfico para a normalidade multivariada dos dados, que indica graficamente a falta de normalidade multivariada.

Figura 6 – qqplot para a normalidade multivariada.



Observando a Figura 6 acima, com o Q-Q plot para *outliers* multivariados, gerado pela função *"mvn"* na opção *"multivariatePlot = qq"* (que tem por padrão o método de quantil "quan"baseado na distância de Mahalanobis e método de quantil ajustado "adj"baseado na distância de Mahalanobis) pode-se supor então a existência de dados extremos que estejam levando à distribuição multivariada não se conformar com a normalidade.

As distâncias de Mahalanobis são distâncias euclidianas de uma matriz onde as colunas são centradas, têm variância unitária e não são correlacionadas. O cálculo é baseado na transformação da matriz de dados e, em seguida, usando distâncias euclidianas, para maiores detalhes em "vegdist{vegan}"(OKSANEN *et al.*, 2022).

4.3 Análise via PERMANOVA:

Como visto anteriormente, uma alternativa para quando os dados não provém de uma distribuição normal e as outras suposições não são atendidas para se fazer uma análise por MANOVA, é a PERMANOVA ou NPMANOVA que são livres de distribuição, e faz em apenas uma suposição explícita a saber: a permutabilidade de unidades sob a hipótese nula (ou seja, os indivíduos devem ser independentes, sem medidas repetidas ou autocorrelação de qualquer tipo).

No Capitulo 3 foi abordado como a PERMANOVA é calculada em sua teória e prática no *Software R*. A estatística de teste é uma "pseudo (F) razão", que é essencialmente uma razão de dispersão entre grupos sobre dispersão dentro do grupo. Essencialmente uma razão de dispersão entre grupos sobre a dispersão dentro do grupo. Utilizando o pacote *vegan* com a função *adonis2()* precisamos também escolher o método a ser aplicado, ou seja, uma dissimilaridade que será usada para calcular distâncias pareadas entre indivíduos e permuta-las até que um valor p seja significativo. As hipóteses para este teste, são:

H_0: Os centróides ou média multivariada de todos os grupos são iguais ou não diferem significativamente.
H_1: Há pelo menos um par de grupos com centróides significativamente diferentes.

Para o caso específico agora da PERMANOVA ou NPMANOVA, como apresentada em (SANTOS', 2020), se esta hipótese nula acima fosse verdadeira, as diferenças observadas entre os três centróides deveriam ser relativamente "pequenas". Pequeno significa que eles devem ter a mesma magnitude que o que seria obtido com a realocação aleatória de indivíduos para um dos três grupos. É exatamente assim que o teste procede: é chamado de MANOVA permutacional porque simplesmente verifica se as diferenças realmente observadas são compatíveis com as diferenças que obtemos permutando os rótulos dos grupos de indivíduos.

Nas Figuras 7 e 8 tem-se o teste da permanova com análise entre os grupos, os três tipos de espécies para 999 e 9999 números de permutações.

Figura 7 – Teste PERMANOVA entre os grupos de espécie com 999 permutações.

```
adonis2(formula = dependent_vars ~ iris$Species, data =
   iris, permutations = 999, method = "euclidian")
             Df SumOfSqs      R2      F Pr(>F)
iris$Species   2   592.07 0.86894 487.33  0.001 ***
Residual     147    89.30 0.13106
Total        149   681.37 1.00000
---
Signif. codes:  0    ***    0.001    **    0.01    *
   0.05    .    0.1         1
```

Figura 8 – Teste PERMANOVA entre os grupos de espécie com 9999 permutações.

```
adonis2(formula = dependent_vars ~ iris$Species, data =
   iris, permutations = 9999, method = "euclidian")
             Df SumOfSqs      R2      F Pr(>F)
iris$Species   2   592.07 0.86894 487.33  1e-04 ***
Residual     147    89.30 0.13106
Total        149   681.37 1.00000
---
Signif. codes:  0    ***    0.001    **    0.01    *
   0.05    .    0.1         1
>
```

Pelos resultados acima pode-ses concluir que será rejeitado a hipótese nula H_0 de que as médias não são iguais entre os grupos. A partir da estatistica F e do valor-p baixo podemos tirar essa conclusão. Uma importante observação é que o valor-p vai ficando menor a partir do maior número de permutações, porém os outros valores permanecem iguais. Temos o valor R2, coeficiente de determinação, que fornece uma medida de quão bem os resultados observados são replicados pelo modelo, sendo este um bom resultado para o teste realizado. Para o caso em que estudo R2 = 0,86894, indica que 86,89% dos dados são replicados pelo modelo.

Para a Tabela 2 temos a comparação entre os grupos para ter uma ideia de como as médias se diferem quando comparadas grupo a grupo de espécies.

Tabela 2 – Comparações dos vetores de médias por grupo de espécie.

Espécie	R2	estatística F	valor-p	Decisão estatística	Nº de Permutações
Setosa e Versicolor	0,849	551,00	0,001 ***	Rejeita H_0	999
	0,849	551,00	1e-04 ***	Rejeita H_0	9999
Setosa e Virginica	0,905	943,8	0,001 ***	Rejeita H_0	999
	0,905	943,8	1e-04 ***	Rejeita H_0	9999
Versicolor e Virginica	0,469	86,77	0,001 ***	Rejeita H_0	999
	0,469	86,77	1e-04 ***	Rejeita H_0	9999

Fonte: elaborado pela autora (2022).

Pelos resultados apresentados na Tabela 2, temos a mesma conclusão já dita anteriormente, em todas as comparações as médias dos grupos diferem (valor-p $\leq$ 0,001). Pelo resultado do valor-p podemos notar o mesmo comportamento em todas as permutações realizadas, de 999 e 9999, que a medida que cresce o número de permutações o valor-p fica menor.

Na Tabela 3 temos as médias e desvio-padrão por espécies. Aqui podemos ver como de fato as médias se diferem para cada grupo de espécies em cada variável, bem como a média geral. Reforçando o que já foi apresentado e assim podemos concluir que para a PERMANOVA as médias multivariadas dos dados *Iris* para os três grupos, rejeitaremos H_0, a hipótese nula apresentada em todas as situações análisadas aqui, isto é, o vetor de médias (de dimensão quatro) difere para todos os quatro grupos de espécie.

Tabela 3 – Médias das variáveis (desvio-padrão).

Variáveis	Setosa	Versicolor	Virginica	Média Geral
Comprimento da Sépala	5,01 (0,35)	5,94 (0,51)	6,59 (0,63)	5,84 (0,82)
Largura da Sépala	3,43 (0,37)	2,77 (0,31)	2,97 (0,32)	3,06 (0,43)
Comprimento da Pétala	1,46 (0,17)	4,26 (0,46)	5,55 (0,55)	3,76 (1,76)
Largura da Pétala	0,25 (0,10)	1,33 (0,19)	2,03 (0,27)	1,20 (0,76)

Fonte: elaborado pela autora (2022).

5 CONSIDERAÇÕES FINAIS

REFERÊNCIAS

ANDERSON, E. **Edgar Anderson's Iris Data - R**. 1935. Disponível em: <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/iris.html>.

ANDERSON, M. J. Permutational multivariate analysis of variance (permanova). **Wiley statsref: statistics reference online**, Wiley Online Library, p. 1–15, 2014.

ANDERSON, M. J.; WALSH, D. C. Permanova, anosim, and the mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? **Ecological monographs**, Wiley Online Library, v. 83, n. 4, p. 557–574, 2013.

BRANCO, F. J. dos S. **Investigação Experimental: Potência estatística dos testes de aleatorização na comparação de dois grupos independentes**. Tese (Doutorado) — Universidade Aberta (Portugal), 2010.

ECOVIRTUAL. **Análise de Classificação - Roteiro em R**. EcoVirtual - USP. Disponível em: <http://ecovirtual.ib.usp.br/doku.php?id=ecovirt:roteiro:comuni:comuni_classr>. Acesso em: 20 out. 2022.

EDGINGTON, E.; ONGHENA, P. **Randomization tests**. [S.l.]: Chapman and Hall/CRC, 2007.

FILHO, J. P.; VIOLA, D. N.; FERNANDES, G. Uso de teste de aleatorização para comparar dois grupos considerando teste não paramétrico. **Simpósio Nacional de Probabilidade e Estatística**, v. 19, 2010.

KORKMAZ, S.; GOKSULUK, D.; ZARARSIZ, G. Mvn: An r package for assessing multivariate normality. **The R Journal**, v. 6, n. 2, p. 151–162, 2014. Disponível em: <https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf>.

LEITE, M. de S. **Testando normalidade multivariada**. 2018. RPubs by RStudio. Disponível em: <https://rpubs.com/melinatarituba/356739>. Acesso em: 02 nov. 2022.

OKSANEN, J.; BLANCHET, F. G.; KINDT, R.; LEGENDRE, P.; MINCHIN, P. R.; O'HARA, R.; SIMPSON, G. L.; SOLYMOS, P.; STEVENS, M. H. H.; WAGNER, H. *et al.* Package 'vegan'. **Community ecology package, version**, v. 2, n. 9, p. 1–295, 2013.

OKSANEN, J.; SIMPSON, G. L.; BLANCHET, F. G.; KINDT, R.; LEGENDRE, P.; MINCHIN, P. R.; O'HARA, R.; SOLYMOS, P.; STEVENS, M. H. H.; SZOECS, E.; WAGNER, H.; BARBOUR, M.; BEDWARD, M.; BOLKER, B.; BORCARD, D.; CARVALHO, G.; CHIRICO, M.; De Caceres, M.; DURAND, S.; EVANGELISTA, H. B. A.; FITZJOHN, R.; FRIENDLY, M.; FURNEAUX, B.; HANNIGAN, G.; HILL, M. O.; LAHTI, L.; MCGLINN, D.; OUELLETTE, M.-H.; Ribeiro Cunha, E.; SMITH, T.; STIER, A.; Ter Braak, C. J.; WEEDON, J. **vegan: Community Ecology Package**. [S.l.], 2022. R package version 2.6-2. Disponível em: <https://CRAN.R-project.org/package=vegan>.

RADEčIć, D. **MANOVA in R – How To Implement and Interpret One-Way MANOVA**. 2022. MANOVA in R Explained. Disponível em: <https://appsilon-com.translate.goog/manova-in-r/?_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt-BR&_x_tr_pto=sc>. Acesso em: 26 out. 2022.

REFFATTI, L. **MODULO 5 - Estatística Multivariada**. 2019. RPubs by RStudio. Disponível em: <http://rpubs.com/leonardoreffatti>. Acesso em: 20 out. 2022.

ROBINSON, A. *et al.* Randomization, bootstrap and monte carlo methods in biology. **Journal of the Royal Statistical Society-Series A**, London: Royal Statistical Society, 1988-, v. 170, n. 3, p. 856, 2007.

SANTOS', F. **How to perform a NPMANOVA with R?** 2020. Caderno de Frédéric Santos Truques Emacs e R para antropólogos e arqueólogos. Disponível em: <https://f-santos.gitlab.io/2020-05-07-npmanova.html>. Acesso em: 24 out. 2022.

VALENTIN, J. L. Agrupamento e ordenação. **Oecologia brasiliensis**, Universidade Federal do Estado do Rio de Janeiro (UNIRIO), v. 2, n. 1, p. 2, 1995.

VICINI, L. Análise multivariada: da teoria à prática. Universidade Federal de Santa Maria, 2005.

WEYENBERG, G.; YOSHIDA, R. Chapter 12 - reconstructing the phylogeny: Computational methods. In: ROBEVA, R. S. (Ed.). **Algebraic and Discrete Mathematical Methods for Modern Biology**. Boston: Academic Press, 2015. p. 293–319. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780128012130000125>.

APÊNDICE A – CÓDIGO UTILIZADO NAS APLICAÇÕES

```
#_____________ PACOTE IRIS_____________##


rm(list=ls(all=TRUE))
install.packages("vegan")
require(vegan)
library(MASS)
library(MVN)
library (lattice)
install.packages("ggplot2")
require(ggplot2)
library(ggplot2)
is.data.frame(iris)
View(iris)
head(iris)
attach(iris)
require(xtable)

# Grupos

setosa_g1<-iris[iris$Species=='setosa',-5];setosa_g1
versicolor_g2<-iris[iris$Species=='versicolor',-5];
   versicolor_g2
virginica_g3<-iris[iris$Species=='virginica',-5];
   virginica_g3

#descritiva da distribui  o dos dados

xtable(summary(iris[,-5]))
variancia<-c(var(iris$Sepal.Length),
```

```
29              var(iris$Sepal.Width),
30              var(iris$Petal.Length),
31              var(iris$Petal.Width));variancia
32
33
34 #esp cies:
35 windows()
36 par(mfrow=c(2,2))
37 boxplot(Petal.Length ~ Species,  main=" Comprimento da
      P tala",xlab = "",ylab = "")
38 boxplot(Petal.Width ~ Species, main=" Largura da P tala",
      xlab = "",ylab = "")
39 boxplot(Sepal.Length ~ Species,main=" Comprimento da
      S pala",xlab = "",ylab = "")
40 boxplot(Sepal.Width ~ Species, main=" Largura da S pala",
      xlab = "",ylab = "")
41
42
43 #Histograma
44 windows()
45 qplot(x=Sepal.Length,data = iris,geom = 'histogram',fill=
      Species,color=Species,
46       xlab='Sepal Length (cm)') +scale_color_manual(values
           = c("orange", "black", "red"))
47 qplot(x=Petal.Length,data = iris,geom = 'histogram',fill=
      Species,color=Species,
48       xlab='Petal Length (cm)') +scale_color_manual(values
           = c("orange", "black", "red"))
49 qplot(x=Sepal.Width,data = iris,geom = 'histogram',fill=
      Species,color=Species,
50       xlab='Sepal Width (cm)') +scale_color_manual(values =
           c("orange", "black", "red"))
```

```
51 qplot(x=Petal.Width,data = iris,geom = 'histogram',fill=
      Species,color=Species,
52       xlab='Petal Width (cm)') +scale_color_manual(values =
           c("orange", "black", "red"))
53
54 install.packages("GGally")
55 library(GGally)
56
57 windows()
58 ggpairs(iris,
59         columns = 1:4,
60         aes(color = Species,
61             alpha = 0.5))
62 #-------------------------------------------
63 dependent_vars <- cbind(iris$Sepal.Length, iris$Sepal.Width
      , iris$Petal.Length, iris$Petal.Width)
64 independent_var <- iris$Species
65
66
67
68
69 #teste de Normal Multivariada:
70 mvn(data = iris[,-5], mvnTest = "royston")
71 windows()
72 mvn(data = iris[,-5], mvnTest = "royston", univariatePlot =
73       "histogram")
74 windows()
75 mvn(data = iris[,-5], mvnTest = "royston",
76     multivariatePlot = "qq")
77
78 par(mfrow=c(1,3)) ### Normais bidimensionais X1 versus X2,
      X3,X4, depois fazer para as demais
```

```
79 mvn(data = iris[1:150,c(1,2)], mvnTest = "royston",
80     multivariatePlot = "persp")
81 mvn(data = iris[1:150,c(1,3)], mvnTest = "royston",
82     multivariatePlot = "persp")
83 mvn(data = iris[1:150,c(1,4)], mvnTest = "royston",
84     multivariatePlot = "persp")
85
86
87 #-----------------------------------------
88 #manova
89 modelo = manova(dependent_vars ~ iris$Species);modelo
90 names(modelo)
91
92 summary.aov(modelo)
93 summary(modelo)
94 summary(modelo, test = "Wilks")
95 summary(modelo, test = "Pillai")#teste de pillai
96 summary(modelo, test = "Hotelling-Lawley")
97 summary(modelo, test = "Roy")
98
99
100 #correla  o das variaveis dependentes
101 #sepalas
102 cor.test(setosa_g1$Sepal.Length,setosa_g1$Sepal.Width)#
       setosa
103 cor.test(versicolor_g2$Sepal.Length,versicolor_g2$Sepal.
       Width)#versicolor
104 cor.test(virginica_g3$Sepal.Length,virginica_g3$Sepal.Width
       )#virginica
105 cor.test(iris$Sepal.Width,iris$Sepal.Length)
106 cor.test(iris$Petal.Length,iris$Sepal.Width)
107 cor.test(iris$Petal.Width,iris$Petal.Length)
```

```
108
109 #petalas
110 cor.test(setosa_g1$Petal.Length,setosa_g1$Petal.Width)#
       setosa
111 cor.test(versicolor_g2$Petal.Length,versicolor_g2$Petal.
       Width)#versicolor
112 cor.test(virginica_g3$Petal.Length,virginica_g3$Petal.Width
       ) #virginica
113
114 windows()
115 pairs(iris[,1:4],col=iris[,5],oma=c(4,4,6,12))
116 par(xpd=TRUE)
117 legend(0.85,0.6, as.vector(unique(iris$Species)),fill=c
       (1,2,3))
118 cor=summarise()
119
120
121
122 #-------------------------------------------
123 # permanova
124
125
126 Medias_Geral=mvn(data = iris[1:150,1:4], mvnTest = "royston
       ")$Descriptives[,2];Medias_Geral
127
128 Medias_Especie1=mvn(data = iris[1:50,1:4], mvnTest = "
       royston")$Descriptives[,2];Medias_Especie1
129
130 Medias_Especie2=mvn(data = iris[51:100,1:4], mvnTest = "
       royston")$Descriptives[,2];Medias_Especie2
131
```

```
Medias_Especie3=mvn(data = iris[101:150,1:4], mvnTest = "
   royston")$Descriptives[,2];Medias_Especie3

# Geral
?adonis2

adonis2(dependent_vars ~ iris$Species, iris, permutations =
    999, method = "euclidian")
adonis2(dependent_vars ~ iris$Species, iris, permutations =
    999, method = "jaccard")
adonis2(dependent_vars ~ iris$Species, iris, permutations =
    999, method = "bray")

adonis2(dependent_vars ~ iris$Species, iris, permutations =
    9999, method = "euclidian")




#Comparando entre as especies 1 e 2 # setosa e versicolor

adonis2(dependent_vars[1:100,] ~ iris$Species[1:100], iris,
    permutations = 999, method = "euclidian",
        strata = NULL, contr.unordered = "contr.sum",
        contr.ordered = "contr.poly", parallel = getOption
           ("mc.cores"))
adonis2(dependent_vars[1:100,] ~ iris$Species[1:100], iris,
    permutations = 9999, method = "euclidian",
        strata = NULL, contr.unordered = "contr.sum",
        contr.ordered = "contr.poly", parallel = getOption
           ("mc.cores"))
Medias_Especie1
```

```
155 Medias_Especie2
156
157 #Comparando entre as especies 1 e 3 # setosa e virginica
158
159 adonis2(dependent_vars[c(1:50,101:150),] ~ iris$Species[c
    (1:50,101:150)], iris, permutations = 999, method = "
    euclidian",
160         strata = NULL, contr.unordered = "contr.sum",
161         contr.ordered = "contr.poly", parallel = getOption
            ("mc.cores"))
162
163 adonis2(dependent_vars[c(1:50,101:150),] ~ iris$Species[c
    (1:50,101:150)], iris, permutations = 9999, method = "
    euclidian",
164         strata = NULL, contr.unordered = "contr.sum",
165         contr.ordered = "contr.poly", parallel = getOption
            ("mc.cores"))
166 Medias_Especie1
167 Medias_Especie3
168
169 #Comparando entre as especies 2 e 3 # versicolor e
    virginica
170
171 adonis2(dependent_vars[c(51:150),] ~ iris$Species[c(51:150)
    ], iris, permutations = 999, method = "euclidian",
172         strata = NULL, contr.unordered = "contr.sum",
173         contr.ordered = "contr.poly", parallel = getOption
            ("mc.cores"))
174 adonis2(dependent_vars[c(51:150),] ~ iris$Species[c(51:150)
    ], iris, permutations = 9999, method = "euclidian",
175         strata = NULL, contr.unordered = "contr.sum",
```

```
176         contr.ordered = "contr.poly", parallel = getOption
               ("mc.cores"))
177 Medias_Especie2
178 Medias_Especie3
179
180 medias<-cbind(Medias_Especie1,Medias_Especie2,
       Medias_Especie3);medias
181 xtable(medias)
```