UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE MATEMÁTICA

DEPARTAMENTO DE ESTATÍSTICA

# Estabilidade da estatística R sob condições de desbalanceamento na ANOSIM

Autora: Giselle Spindler

Orientador: Professor Dr. Fernando Hepp Pulgati

Porto Alegre, 16 de Julho de 2013.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

# Estabilidade da estatística R sob condições de desbalanceamento na ANOSIM

Autora: Giselle Spindler

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professor Dr. Fernando Hepp Pulgati (orientador)
Professor Dr. Álvaro Vigo (convidado)

Porto Alegre, 16 de Julho de 2013.

*Dedico este trabalho a Maria Isabel Müller, minha mãe.*

*“Pensar sem aprender torna-nos caprichosos, e aprender sem pensar é um desastre.” (Confúcio)*

# Agradecimentos

A minha mãe, que sempre esteve na torcida e sempre acreditou na minha volta aos estudos.

À toda a minha família, pelas repetidas ausências em ocasiões especiais decorrentes de muito trabalho acumulado.

Aos meus amigos, por estar distante em muitas festas e muitas comemorações, mas sempre contando com a torcida e a compreensão de todos.

Aos meus chefes e meus colegas de trabalho, pela compreensão nas trocas de horários para que pudesse chegar até aqui.

A todos os meus professores, como sempre excelentes, que me auxiliaram e sempre me indicaram o caminho para melhorar como profissional e como pessoa.

À ComGrad do curso de Estatística que possibilitou trocas de horários das disciplinas para que eu pudesse cursá-las.

Ao Dr. João Cabrera, meu querido médico, por me manter saudável durante este período.

Ao meu orientador Fernando Hepp Pulgati, pela paciência confiança e dedicação em aperfeiçoar este trabalho.

Muito Obrigada!

# Resumo

Este trabalho apresenta um estudo sobre a estatística de teste R utilizada no teste ANOSIM. O procedimento de teste é baseado na matriz de similaridade entre observações geralmente estabelecida a partir da abundância de indivíduos de diferentes espécies que comumente resultam em matrizes esparsas. O objetivo é verificar a estabilidade de tal estatística sob condições de desbalanceamento considerando a quantidade de permutações recomendada na literatura. Para atingir o objetivo foram derivadas expressões importantes para compreensão dos cálculos que envolvem as estimativas. Concomitantemente, foram realizadas simulações forçando alguns desbalanceamentos de interesse. Os exercícios envolvem dois grupos que foram construídos com valores distintos para a média da distribuição, $\lambda$, considerando proporções de 50%, 60%, 66,67% 40% e 25% entre grupos. Ainda, para efeito de comparação, foram gerados grupos desbalanceados com mesmo valor de $\lambda$. Os resultados permitiram observar tanto para os grupos com valores distintos para $\lambda$, como para valores iguais, um comportamento distinto do esperado. Além disto, e de acordo com o esperado, verificou-se que quanto maior a quantidade de permutações mais estável se torna o p-valor da estatística de teste R.

**Palavras-Chave:** ANOSIM, permutações, desbalanceamentos.

# Abstract

This work presents a study about the statistics of test R used in ANOSIM test. The test procedure is based on the similarity matrices between data generally established from the abundance of individuals from different species that costumary results in sparse matrices. The objective is to verify the stability of such statistics under unbalanced conditions being considered the amount of permutations recommended in literature. To reach the objective important expressions for understanding had been derived from the calculations that involve the estimates. Concurrently some unbalanced of interest had been carried through simulation forcing. The exercises involve two groups that had been constructed with distinct values for the average of the distribution, λ, considering proportions of 50%, 60%, 66.67% 40% and 25% between groups. Still, for comparison effect, groups unbalanced with the same value for λ had been generated. The results had allowed observing as much for the groups with distinct values for λ, as for equal values, a distinct behavior from the waited one. Moreover, and in accordance with the waited one, was verified that how much bigger the amount of permutations is more stable becomes the p-value of the statistics of test R.

**Key Words:** ANOSIM, permutations, unbalanced.

# Sumário

# 1. INTRODUÇÃO

Ecologistas ao redor do mundo necessitam realizar análises acerca de comunidades ecológicas, bem como estudos de impactos ambientais. Em algumas situações, a estatística paramétrica não é a mais apropriada por necessitar de uma série de suposições *a priori* que na maioria dos casos não é satisfeita pelas distribuições destas comunidades. Portanto é necessário fazer uso de técnicas não paramétricas.

Muitas teorias neste sentido têm sido desenvolvidas por pesquisadores com o intuito de permitir tais estudos. Clarke e Green (1988) desenvolveram um procedimento denominado ANOSIM com o objetivo de realizar análises flexibilizando as suposições acerca das distribuições das variáveis aleatórias. Clarke e Warwick (2001) desenvolveram o software *Plymouth Routines in Multivariate Ecological Research* (PRIMER), com o intuito de disseminar esta nova técnica aos pesquisadores na área da ecologia.

A estatística de teste denominada R tem dimensão multivariada e é baseada nas médias das similaridades dentro e entre os grupos.

Sob a hipótese nula, de que não existem diferenças entre os grupos, ao gerar todas as permutações possíveis se obtém a distribuição da estatística R. Porém muitas vezes, gerar todas as permutações possíveis nem sempre é um procedimento viável no aspecto tempo/processamento. Desta forma, em muitos casos (estudos) se faz uso de uma quantidade bem menor, mas suficientemente grande, de permutações para gerar a distribuição amostral de R. Cada valor obtido para R gerado através destas permutações é denominado neste trabalho por R*. O p-valor para a estatística R é calculado com base na localização de R na distribuição amostral de R, comparando-se quantos valores de R* são maiores ou iguais a R na referida distribuição.

O problema é que ao se trabalhar com grupos desbalanceados, fato comum em estudos ambientais, a estatística R pode conduzir a resultados duvidosos. Surge então a questão: ***O estimador para R sob condições de desbalanceamento e/ou com um número inapropriado de permutações é uma estatística estável?***

O presente tema foi escolhido, pois é um assunto relevante com ampla aplicação em ciências biológicas. Desta forma o tema mostra-se oportuno, pois gera conhecimento sobre técnicas estatísticas não paramétricas para análise de dados multivariados, amplamente utilizadas por ecologistas (ANOSIM); importante, pois avaliará a estabilidade do estimador sob condições de desbalanceamento através de uma série de exercícios com desdobramentos algébricos; atual, uma vez que assunto impacto ambiental faz parte das

discussões governamentais da atualidade; e viável, pois a pesquisadora possui acesso ao software PRIMER (e ao R) e computadores, e este assunto já foi trabalhado, através de simulação, no estágio.

Este trabalho tem como objetivo avaliar a estabilidade da estatística R considerando o número de permutações e as condições de desbalanceamento das réplicas através da técnica não paramétrica ANOSIM.

David L. Jones, em sua *homepage,* comenta que a ANOSIM embora muito utilizada entre os pesquisadores na área biológica, nada mais é que uma "simples versão do Teste de Mantel[1] baseado na padronização dos *rankings* das correlações entre duas matrizes de distâncias". O autor também faz menção ao fato de que o método ANOSIM não deveria realizar comparações diretamente sobre a diversidade Beta (variação na abundância de espécies e composição entre unidades amostrais) em termos de algum fator de agrupamento ou nível de tratamento experimental, mas sim analisar a variação da diversidade Beta. Desta forma pode-se dizer que o método ANOSIM vem sendo questionado, quanto a seus resultados, por alguns pesquisadores.

Cabe informar ao leitor que os trabalhos que envolvem ausência/abundância de espécies admitem expressões próprias. Desta forma neste trabalho serão considerados grupos os locais, ou tratamentos, distintos onde se obteve as amostras para as contagens das diferentes espécies. Considera-se uma réplica cada uma das amostras coletadas em cada local. Pode-se obter em um local 2 réplicas e em outro local 10 réplicas. A quantidade de réplicas coletadas dependerá do grau de dificuldade na obtenção destas.

1 Nathan Mantel (1919–2002) estatístico americano. Ficou conhecido por, juntamente com William Haenszel (1910-1998), ter desenvolvido o conhecido Teste de Mantel Haenszel.

# 2. OBJETIVOS

Aqui estão detalhados os objetivos, geral e específicos, deste trabalho.

## Objetivo Geral

Este trabalho teve o objetivo de avaliar a estabilidade da estatística de teste R considerando o número de permutações e condições de desbalanceamento dos grupos na ANOSIM.

## Objetivos Específicos

O objetivo geral foi alcançado através da avaliação da:

- Estabilidade da estatística como função no número de permutação;
- Estabilidade da estatística em experimentos desbalanceados.

# 3. REVISÃO BIBLIOGRÁFICA

## A Análise Multivariada

Em muitos momentos, ao examinar dados amostrais, pode ser de interesse do pesquisador analisar as correlações existentes entre as variáveis observadas. Em várias áreas, pesquisadores se vêem diante de muitas variáveis ao mesmo tempo e desejam encontrar estruturas mais simples, ou menores (em dimensões menores) sem perder as informações contidas nas variáveis originais.

A área da Estatística que aborda o estudo e análise de dados em várias dimensões chama-se Análise Multivariada. Segundo Johnson e Wichern (2007), os objetivos das investigações científicas para os quais os métodos multivariados se aplicam incluem:

> Redução de dados ou simplificação na estrutura: O fenômeno a ser estudado é representado da forma mais simples possível sem sacrificar substancialmente as informações contidas nas variáveis. Espera-se que este procedimento torne a interpretação mais fácil.
> Classificação e agrupamento: Grupos de variáveis ou objetos similares são criados baseados em medidas características. Alternativamente, regras de classificação para os objetos em grupos bem definidos podem ser necessárias.
> Investigação de dependência entre variáveis: A natureza das relações entre as variáveis é de interesse. OU seja, deseja-se verificar se todas as variáveis são mutuamente independentes ou existem uma ou mais variáveis que dependem de outras. Ainda, para o caso de existir tal dependência, como ela se dá.
> Predição: Relações entre variáveis podem ser determinadas com o propósito de predizer os valores de uma ou mais variáveis baseados nas observações de outras variáveis.
> Testando hipóteses: São testadas hipóteses formuladas em termos de parâmetros de populações multivariadas. Isso pode ser feito para validar as suposições ou reforçar convicções prévias.

Cada um dos objetivos citados anteriormente é avaliado por técnicas e modelos específicos, embora todos apresentem um mesmo ponto de partida: a matriz de dados. A partir desta matriz, são produzidas análises exploratórias importantes para que se possa estabelecer um caminho a ser traçado utilizando-se as técnicas de Análise Multivariada. Neste contexto surgem aspectos relacionados a possíveis *outliers*[2] e valores extremos (*extreme values*[3]) que poderão afetar fortemente os resultados.

[2] *Outliers*: são valores, em geral, localizados entre 1,5 e 3,0 distâncias interquartílicas (Barnett e Lewis, 1994). Para o caso de dados multivariados pode-se usar a Distância de Mahalanobis (3.6). Desta forma, quando se está diante de uma distribuição Normal Multivariada, o quadrado da Distância de Mahalanobis se aproxima de uma distribuição $\chi^2$ com p

No presente trabalho não se abordará todas as técnicas disponíveis, mas sim aquelas que se referem ao assunto em estudo. Por este motivo, optou-se por descrever de forma sucinta o conjunto de técnicas amplamente utilizadas em trabalhos que envolvam dados provenientes de ausência/abundância de espécies em sítios ecológicos, ou que tenham alguma relação com estes.

**ANOVA (*Analysis of Variance*)**

É uma técnica da estatística paramétrica utilizada para comparação de médias entre *g* grupos. Esta técnica é utilizada quando existe apenas uma variável dependente. O teste, cuja hipótese nula afirma não haver diferença entre as médias dos grupos, realizado para esta comparação faz uso da estatística,

$$F = \frac{SQ_{trat}/(g-1)}{SQ_{erro}/(\sum_{l=1}^{g} n_l - g)} \tag{3.1}$$

que tem distribuição $F_{g-1,\sum n_l - g}$. Ainda observa-se que,

$$SQ_{trat} = \sum_{l=1}^{g} n_l (\bar{x}_l - \bar{x})^2 \tag{3.2}$$

e

$$SQ_{erro} = \sum_{l=1}^{g} \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2 \tag{3.3}$$

Para que se possa realizar este teste é necessário a verificação de algumas suposições sobre os erros:

i) Os erros das variáveis aleatórias devem ser independentes;

ii) As variâncias populacionais devem ser iguais (constante) nos *g* grupos;

iii) A distribuição dos erros em cada grupo deve ser normal ou aproximadamente normal. Simbolicamente representa-se: $\varepsilon_j \sim N(0, \sigma^2)$.

graus de liberdade. Assim define-se *outlier* como medidas que se localizam em determinado intervalo quartílico da distribuição qui-quadrado (Filzmoser et al, 2005, *apud* Valadares e outros, 2012).

[3] *Extreme Values:* são valores, em geral, localizados a mais de 3,0 distâncias interquartílicas (Barnett e Lewis, 1994).

Pode-se observar que esta técnica não pode ser utilizada para analisar as diferentes comunidades ecológicas, pois nem sempre as variâncias populacionais serão iguais (na maioria das vezes serão muito diferentes) e ainda a distribuição das observações em cada grupo não segue uma normal, mas sim uma distribuição Poisson. Embora o teste F seja robusto com respeito à violação da normalidade, ou seja, ainda é válido com pequenas violações desta suposição, ainda esta técnica não é suficiente, pois as comunidades avaliadas em estudos ambientais se apresentam com respostas multivariadas na estrutura de comunidades.

**MANOVA *(Multivariate Analysis of Variance)***

É uma extensão da técnica ANOVA, é utilizada quando há mais de uma variável dependente em estudo, ou seja, quando as respostas podem ser avaliadas na forma multivariada. Através desta técnica podem-se avaliar diferenças entre *g* grupos com relação a duas ou mais variáveis dependentes combinadas entre si.

O teste para a hipótese nula de que não existe efeito significativo pode ser produzido utilizando-se a estatística lambda de Wilks[4], obtida por:

$$\Lambda^* = \frac{|W|}{|B+W|} = \frac{\left|\sum_{l=1}^{g}\sum_{j=1}^{n_l}(x_{lj}-\bar{x}_l)(x_{lj}-\bar{x}_l)'\right|}{\left|\sum_{l=1}^{g}\sum_{j=1}^{n_l}(x_{lj}-\bar{x})(x_{lj}-\bar{x})'\right|} \tag{3.4}$$

Temos que, segundo Bartlett[5] (apud Jonhson e Wichern, 2007, p.304), quando $H_0$ é verdadeira e $\sum n_l$ é grande então, $-\left(n-1-\frac{(p+g)}{2}\right)\ln\Lambda^*$ tem distribuição aproximada por uma $\chi^2_{p(g-1)}$, onde *p* é o número de variáveis e *g* é o número de grupos.

Outras estatísticas de teste são também utilizadas, como o traço de Pillai[6], traço de Lawley-Hotelling[7], e maior raiz de Roy[8], embora para grandes amostras todos sejam equivalentes.

4 Samuel Stanley Wilks (1906-1964) matemático americano criador do teste Lambda de Wilks que é uma estatística cuja função é denotar a significância estatística do poder discriminatório da função discriminante em questão.
5 Maurice Stevenson Bartlett (1910-2002) estatístico inglês desenvolveu o teste para homogeneidade de variâncias, sendo este mais eficiente que o teste de Levene quando não se rejeita a hipótese de normalidade dos dados.
6 K. C. Sreedharam Pillai (1920-1985) estatístico indiano desenvolveu um teste multivariado para verificar igualdade de médias entre grupos a partir do traço de uma matriz de covariâncias. Utilizando um procedimento não paramétrico onde os dados são transformados em postos, de forma independente para cada variável, onde a estatística de teste é comparada

Para que se possa utilizar a MANOVA com o propósito de analisar os grupos e suas diferenças de médias devem-se considerar algumas suposições antes de realizar os testes:

i) Os dados observados devem ser provenientes de uma população com distribuição normal p-variada;

ii) As amostras deverão pertencer a populações cujos grupos possuem a mesma variância;

iii) As observações devem ser independentes;

iv) A linearidade e multicolinearidade entre as variáveis dependentes devem ser avaliadas cuidadosamente. No caso da segunda, não deve ser muito elevada.

Utilizar esta técnica para analisar dados provenientes de observações em comunidades ecológicas também pode não ser a mais apropriada, pois as populações não costumam apresentar distribuições normais. As matrizes de observações são geralmente muito grandes e compostas por muitas entradas nulas, e principalmente porque as interações entre as variáveis de interesse em cada caso não é considerada nesta estrutura de modelo, diferentemente de uma técnica que considere como informação relevante, a similaridade p-variada entre casos.

**ANOSIM (*Analysis of Similarities*)**

De forma análoga a ANOVA, Clarke e Green em 1988 criaram um método não-paramétrico com o objetivo de realizar análises multivariadas quando as hipóteses das técnicas da análise clássica são violadas ou não respondem as hipóteses de pesquisa estabelecidas. A necessidade de desenvolver tal procedimento originou-se da insuficiência de métodos para analisar dados provenientes de comunidades ecológicas onde a suposição de normalidade multivariada (Clarke e Warwick, 1993) é violada, mesmo com transformações nos dados. Ou seja, existem muitas dezenas (ou centenas) de dados constituídos de entradas nulas ou, quando não-nulas, impondo constantemente variabilidade muito grande.

---

com o valor de uma distribuição Qui-quadrado com $p(c-1)$ graus de liberdade, onde $p$ = número de variáveis e $c$= número de grupos ou tratamento.

7 Harold Hotelling (1895-1973) matemático americano desenvolveu uma estatística usada para testar a igualdade entre as médias de $k$ vetores $p$ –variados com distribuições normais e matriz de covariância comum, mas desconhecida. A hipótese nula tem distribuição assintótica qui-quadrado com $k.p$ graus de liberdade

8 Samarendra Nath Roy (1906-1964) matemático indiano desenvolveu uma estatística que mede as diferenças entre as médias de $k$ vetores $p$-variados através do maior autovalor da matriz de covariâncias.

Esta técnica é baseada em matrizes de similaridades (ou distâncias), a partir do qual se realiza o ordenamento (*ranking*) dos valores obtidos com o objetivo de gerar uma estatística de teste denominada R. Existem muitas maneiras de calcular tais distâncias (ou coeficientes de similaridades) entre pares de observações p-dimensionais $(\vec{x}, \vec{y})$. A seguir estão listadas as mais comuns:

**Distância Euclideana**

Esta é a medida de distância mais conhecida entre quaisquer dois pontos no espaço, pois é aquela estudada na escola através da Geometria Clássica (no plano Euclidiana). Adaptada a esta teoria se define algebricamente distância Euclidiana entre duas observações p-dimensionais $\vec{x}$ e $\vec{y}$ como:

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})'(\vec{x} - \vec{y})} \tag{3.5}$$

**Distância de Mahalanobis**

A distância de Mahalanobis[9] entre dois vetores de observações $\vec{x} = (x_1, x_2, ..., x_p)^T$ e $\vec{y} = (y_1, y_2, ..., y_p)^T$ com matriz de covariâncias $\sum$ , é definida como:

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})' \sum^{-1} (\vec{x} - \vec{y})} \tag{3.6}$$

Observa-se que quando a matriz de covariâncias é a matriz identidade então a distância de Mahalanobis coincide com a Distância Euclidiana.

**Distância de Minkowski**

Outra distância conhecida é a distância (métrica) de Minkowski[10], dada pela seguinte expressão:

[9] Prasanta Chandra Mahalanobis (1893 – 1972). Matemático e estatístico indiano. Fundador do Instituto Indiano de Estatística. Desenvolveu, a partir de estudos antropométricos, uma medida de distância multivariada connhecida como Distância de Mahalanobis.
[10] Hermann Minkowski (1864 – 1909). Matemático alemão criou e desenvolveu a geometria dos números.

$$d(\vec{x},\vec{y}) = \left[\sum_{i=1}^{p} |x_i - y_i|^m\right]^{1/m},\ m > 0. \tag{3.7}$$

Cabe salientar que quando $m = 2$ esta distância coincide com a Distância Euclidiana. Em geral, ao modificar os valores de m altera-se o peso atribuído a diferenças maiores ou menores.

**Distância de Bray-Curtis**[11]

Esta distância, entr dois vetores de observações $\vec{x} = (x_1, x_2, ..., x_p)^T$ e $\vec{y} = (y_1, y_2, ..., y_p)^T$, é muito utilizada por ecologistas em seus trabalhos e segundo Johnson e Wichern (2007) pode ser representada pela expressão:

$$d(\vec{x},\vec{y}) = 1 - \frac{2\sum_{i=1}^{p} \min(x_i, y_i)}{\sum_{i=1}^{p} (x_i + y_i)} \tag{3.8}$$

De acordo com Clarke e Warwick (2001) pode ser interpretada como a similaridade entre as observações *j* e *k* e escrita como segue.

$$S_{jk} = 100\left(1 - \frac{\sum_{i=1}^{p} |y_{ij} - y_{ik}|}{\sum_{i=1}^{p} (y_{ij} + y_{ik})}\right) = 100\left(\frac{\sum_{i=1}^{p} 2\min(y_{ij}, y_{ik})}{\sum_{i=1}^{p} (y_{ij} + y_{ik})}\right) \tag{3.9}$$

onde as observações estão definidas nas linhas da matriz e as espécies estão dispostas nas colunas. Desta forma $y_{ij}$ representa a entrada da i-ésima coluna e da j-ésima linha da matriz de dados, ou seja, a j-ésima observação da i-ésima espécie. Para realizar o cálculo considera-se |.| como o valor absoluto da diferença entre duas observações diferentes de mesma espécie e, min(.,.) é o mínimo entre estes dois valores. No denominador apresenta-se a soma com relação a estes valores para cada uma das espécies (colunas da matriz).

Das distâncias apresentadas anteriormente aquela que efetivamente é mais utilizada nos trabalhos sobre comunidades ecológicas é a Distância de Bray-Curtis, pois as matrizes

[11] John Thomas Curtis (1913 – 1961). Botânico e ecologista americano e John Roger Bray ( - ) conhecidos por suas contribuições para o desenvolvimento de métodos numéricos em ecologia.

de dados são geralmente constituídas por muitos zeros, que indicam, por exemplo, ausência da espécie i no k-ésimo caso.

No Quadro 1, apresenta-se um exemplo de uma matriz de dados envolvendo seis espécies diferentes em quatro locais distintos. Utilizando a distância de Bray-Curtis e a Distância Euclideana.

*Quadro 1: Matriz de observações de 6 espécies em 4 locais distintos*

| | Espécie 1 | Espécie 2 | Espécie 3 | Espécie 4 | Espécie 5 | Espécie 6 |
|---|---|---|---|---|---|---|
| Local 1 | 9 | 19 | 9 | 0 | 0 | 0 |
| Local 2 | 0 | 0 | **37** | 12 | 128 | 0 |
| Local 3 | 0 | 0 | 0 | 144 | 344 | 0 |
| Local 4 | 0 | 3 | 10 | 9 | 2 | 0 |

*FONTE: Adaptado de Clarke e Warwick, 2001, p2-2*

Observe que a matriz de dados no Quadro 1 apresenta as espécies nas colunas e os locais nas linhas, desta forma reproduz os dados como aparecem nos trabalhos científicos da área.

Por exemplo, na segunda linha da matriz temos o vetor $\vec{x}_2 = (0,0,37,12,128,0)^T$ indicando a abundância/ausência de 6 espécies no local 2 (ou grupo 2, ou ainda tratamento 2). Ainda com relação ao Quadro 1, pode-se observar que o valor $x_{23} = 37$ indica que foram encontradas 37 unidades da espécie 3 no local 2.

No Quadro 2 é apresentada a Matriz de Similaridades usando a Distância de Bray-Curtis adaptada por Clarke e Warwick (2001) através da expressão constante da Fórmula (3.9).

*Quadro 2: Matriz de similaridades utilizando o Coeficiente de Bray-Curtis*

| | Local 1 | Local 2 | Local 3 | Local 4 |
|---|---|---|---|---|
| Local 1 | 100 | 8 | 0 | 39 |
| Local 2 | 8 | 100 | **42** | 21 |
| Local 3 | 0 | **42** | 100 | 4 |
| Local 4 | 39 | 21 | 4 | 100 |

*FONTE: Adaptado de Clarke e Warwick, 2001, p2-2*

Observe como foi calculada a entrada da linha 2 com a coluna 3 da matriz de similaridades, que indica a similaridade entre os locais 2 e 3 com relação a todas as espécies observadas (nota-se que o valor é o mesmo obtido através do cálculo da linha 3 com a coluna 2, pois a matriz de similaridades é simétrica).

$$S_{23} = 100\left(1 - \frac{\sum_{i=1}^{6}|y_{i2} - y_{i3}|}{\sum_{i=1}^{6}(y_{12} + y_{i3})}\right) = 100\left(1 - \frac{0+0+37+132+216+0}{0+0+37+156+472+0}\right) = 100\left(1 - \frac{385}{665}\right) =$$

$$= 42{,}11 \approx 42$$

Pode-se observar também que a matriz é simétrica e suas entradas na diagonal principal são de valor 100, indicando a similaridade máxima entre as observações das espécies no mesmo local.

No Quadro 3 apresenta-se a matriz de similaridades usando a Distância Euclideana. Pode-se observar que da mesma forma que no Quadro 2 tem-se a diagonal principal representando as similaridades entre espécies no mesmo local, porém constituído de entradas com valor zero indicando a máxima similaridade (menor distância).

*Quadro 3: Matriz de similaridades utilizando a Distância Euclidiana*

| | L1 | L2 | L3 | L4 |
|---|---|---|---|---|
| L1 | 0 | 133 | 374 | 21 |
| L2 | 133 | 0 | **256** | 129 |
| L3 | 374 | **256** | 0 | 368 |
| L4 | 21 | 129 | 368 | 0 |

*FONTE: Adaptado de Clarke e Warwick, 2001, p2-2*

Assim como anteriormente descreve-se a forma de cálculo para a entrada da linha 2 com a coluna 3 da matriz de similaridades, que indica a similaridade entre os locais 2 e 3 com relação a todas as espécies observadas utilizando a distância Euclidiana.

$$d(\vec{x}, \vec{y}) = \sqrt{(\vec{x} - \vec{y})'(\vec{x} - \vec{y})} = \sqrt{(0-0)^2(0-0)^2(37-0)^2(12-144)^2(128-344)^2(0-0)^2}$$

$$= \sqrt{65449} = 255{,}83 \approx 256$$

A seguir apresenta-se as etapas para obtenção da estatística de teste R, bem como para obtenção da distribuição onde esta estatística será alocada com o objetivo de obter um p-valor.

**Teste de Hipótese**

A hipótese nula afirma não haver diferenças significativas entre os grupos definidos *a priori*, isto é, o efeito de grupo não é significativo.

**A Estatística de teste R**

Neste procedimento, sobre os valores obtidos na matriz de similaridades, define-se $\bar{r}_w$ como a média de todos os *ranks* de similaridades entre todas as réplicas dentro dos grupos e, $\bar{r}_B$ é a média de todos os *ranks* de similaridades entre todas as réplicas entre os grupos, então a estatística de teste (Clarke e Warwick, 1988 e 1993) é dada por:

$$R = \frac{\left(\bar{r}_B - \bar{r}_w\right)}{\frac{1}{2}M} \qquad (3.10)$$

onde M = n(n-1)/2 e n é o número total de observações.

Observa-se que a estatística R pertence ao intervalo [-1,1], onde valores próximos de zero indicam que a hipótese nula é verdadeira, isto é as similaridades entre e dentro dos grupos possuem a mesma média. E, ainda, R=1 somente se todas as réplicas dentro dos grupos são mais similares que quaisquer outras provenientes de grupos diferentes.

Normalmente R se situa entre 0 e 1, indicando algum grau de discriminação entre os grupos. Valores de R menores que zero são pouco prováveis, pois indicam que as similaridades entre os diferentes grupos são maiores que as similaridades dentro dos grupos. Casos extremos como este são citados em Chapman e Underwood (1999 *apud* Clarke e Warwick, 2001).

A estatística R é ela própria uma medida comparativa do grau de separação dos grupos e, seu valor costuma ser menos importante do que sua significância estatística. Como em testes univariados padrão, é perfeitamente possível para R ser significativamente diferente de zero ainda que seja muito pequeno o seu valor, se houver muitas repetições em cada local (Clarke e Warwick, 2001, p. 6-2).

**Calculando e Estatística R sobre permutações**

Sob a hipótese nula de que não existem diferenças significativas entre os grupos definidos *a priori*, é provável que exista um pequeno efeito sobre o valor de R se as observações forem rearranjadas (permutadas) aleatoriamente entre os grupos. Desta forma a estatística R é recalculada após estas permutações, dando origem ao valor R*. Assim para cada nova permutação obtém-se um valor para R* originando a distribuição amostral (pois na maioria dos casos não se consegue gerar todas as permutações possíveis) para os valores de R*, onde o valor de R será alocado.

Tem-se T maneiras diferentes de permutar as n réplicas de cada um dos k grupos, sempre que os grupos forem balanceados, isto é, sempre que o número n de réplicas em cada grupo for o mesmo. O valor T é representado pela expressão a seguir:

$$T = \frac{(kn)!}{[(n!)^k k!]} \tag{3.11}$$

Por exemplo, para um estudo onde existem 3 grupos balanceados com 4 réplicas cada um, obtém-se um total de 5775 permutações possíveis. Para este exemplo é computacionalmente possível calcular todos os valores para R* e examinar esta distribuição. Porém nem sempre este procedimento é possível na prática, pois pequenos aumentos no número de réplicas e de grupos podem gerar uma quantidade muito grande de permutações possíveis demandando muito tempo para o seu cálculo. Desta forma o processo de permutações é realizado para um número T suficientemente grande de vezes.

**Nível de significância do teste**

Compara-se o valor obtido para a estatística de teste R com a distribuição dos valores R* obtidos através de um número T de permutações.

Se $H_0$ for verdadeira, a provável distribuição dos valores de R* é dada pelas permutações aleatórias, da mesma forma se o verdadeiro valor da R parecer improvável que venha desta distribuição, há evidências para rejeitar a hipótese nula. Formalmente, se apenas uma quantidade t dos T valores de R* são superiores ao valor calculado para R então $H_0$ pode ser rejeitada a um nível de significância de (t+1)/(T+1), ou em termos de porcentagem de 100(t+1)/(T+1) %.

Quanto mais permutações forem realizadas maior será a precisão do p-valor (nível descritivo amostral), De acordo com Manly (1997, p.82)

> [...]É fácil ver que 1000 aleatorização nos apresentam quase o mesmo resultado do que com a distribuição completa, exceto em casos-limite com *p* muito próximo de 0,05. [...] mas com nível de significância de 1% são apresentados [...] Aqui com 5000 aleatorizações é praticamente certo que apresentam o mesmo nível de significância do que com a distribuição completa, exceto em casos-limite.

Este método é utilizado em vários estudos, isto é, os pesquisadores se utilizam destes resultados para, dependendo do nível de significância $\alpha$ desejado, realizar 1000 ou 5000 permutações. Porém se os dados não se apresentam de forma balanceada talvez estas

quantidades de permutações não sejam suficientes, como já foi mostrado previamente por Bündchen (2012).

**Testes pareados**

O teste descrito anteriormente é conhecido como teste global, isto é, ele apenas indica que pode haver (ou não) diferença significativa em algum grupo, mas não indica, quando a hipótese nula é rejeitada, em qual grupo isto ocorre. Desta forma para que se conheça qual grupo difere significativamente dos demais deve-se realizar o teste comparando os grupos dois a dois. Assim a estatística R é calculada para cada par de grupos bem como a sua distribuição a partir das permutações geradas pelos valores de R*.

Para que se possam distinguir os valores em questão considera-se R global o valor da estatística de teste calculada para todas as observações em todos os grupos e, simplesmente de R a estatística de teste calculada para cada par de grupos em questão.

# 4. METODOLOGIA

Neste capítulo apresenta-se a metodologia utilizada para avaliar a estabilidade da estatística R da análise de similaridades (ANOSIM) como função no número de permutações e do desbalanceamento através de um conjunto de exercícios com desdobramento algébrico do resultado. Nos casos para os quais não se encontrou na literatura as expressões que representassem as quantidades de interesse, as expressões foram deduzidas através de comparações com resultados obtidos no softwares. O trabalho seguiu as etapas a seguir descritas.

**Etapa 1:** Avaliar a estabilidade da estatística como função no número de permutações e desbalanceamentos através de um conjunto de exercícios com desdobramento algébrico do resultado.

**Etapa 2:** Realizar simulações com o propósito de confirmar os resultados dos exercícios desenvolvidos na Etapa 1. Para cada série de exercícios foram analisados dois locais com três espécies cada um. As simulações foram construídas de acordo com a seguinte composição dos grupos:

**a)** Série onde o primeiro grupo contém 2 réplicas e o outro contém 4 réplicas. As réplicas são geradas aleatoriamente e os valores de $\lambda$ são alocados da seguinte forma: primeiro 2 réplicas com $\lambda=5$ e 4 réplicas com $\lambda=8$ e também o contrário. Desta forma se construiu uma desbalanceamento onde um grupo apresenta 50% das réplicas do outro;

**b)** Série onde o primeiro grupo contém 3 réplicas e o outro contém 5 réplicas. As réplicas são geradas aleatoriamente e os valores de $\lambda$ são alocados da seguinte forma: primeiro 3 réplicas com $\lambda=5$ e 5 réplicas com $\lambda=8$ e também o contrário. Aqui se construiu um desbalanceamento onde um grupo apresenta 60% das réplicas do outro;

**c)** Série onde o primeiro grupo contém 2 réplicas e o outro contém 3 réplicas. As réplicas são geradas aleatoriamente e os valores de $\lambda$ são alocados da seguinte forma: primeiro 2 réplicas com $\lambda=5$ e 3 réplicas com $\lambda=8$ e também o contrário. Aqui se construiu um desbalanceamento onde um grupo apresenta 66,67% das réplicas do outro;

**d)** Série onde o primeiro grupo contém 2 réplicas e o outro contém 5 réplicas. As réplicas são geradas aleatoriamente e os valores de λ são alocados da seguinte forma: primeiro 2 réplicas com λ=5 e 5 réplicas com λ=8 e também o contrário. Aqui se construiu um desbalanceamento onde um grupo apresenta 40% das réplicas do outro;

**e)** Série onde o primeiro grupo contém 2 réplicas e o outro contém 8 réplicas. As réplicas são geradas aleatoriamente e os valores de λ são alocados da seguinte forma: primeiro 2 réplicas com λ=5 e 8 réplicas com λ=8 e também o contrário. Aqui se construiu um desbalanceamento onde um grupo apresenta 25% das réplicas do outro.

Para estas primeiras séries de exercícios cada uma das espécies foi construída separadamente através da geração de números aleatórios segundo uma distribuição Poisson(λ), através do software R, versão 2.13.2, com o comando *rpois*.

A partir de cada série inicial, duplicou-se a quantidade de réplicas dentro de cada série até um total máximo de 40 réplicas, sendo que as únicas séries a atingir este valor foram as iniciadas pelos grupos de 2 e 3 réplicas e pelos grupos de 2 e 8 réplicas. Desta forma foram gerados 6 ou 8 exercícios em cada série (pois em cada caso se considera o grupo menor com λ=5 e com λ=8). Foram geradas, quando possível, 999 permutações e 4999 permutações para cada série de exercícios com o intuito de comparar resultados.

Mantendo a estrutura de desbalanceamento anterior construiu-se grupos com o valor λ=8 para ambos os grupos com o objetivo de verificar se o teste é capaz de detectar, com mais facilidade, se os grupos são iguais ou não quando estes realmente o são.

Em outro momento, com o objetivo de se aproximar mais da realidade, construiu-se seis séries com 100 valores cada, onde as três primeiras apresentam uma distribuição Poisson(5) e as outras três, uma Poisson(8). As séries são realizadas em grupos de três, pois representam as três espécies. Desta forma, para obter as distribuições em cada caso, realizou-se um sorteio de uma amostra aleatória simples sem reposição para cada grupo. Assim cada valor sorteado estará representando uma réplica que contém as três espécies que estão sendo observadas e que provém de distribuições Poisson(λ).

Numa nova etapa, como já foi feito para os grupos gerados aleatoriamente, realizou-se o mesmo tipo de exercício, com todas as séries, porém com λ=8 fixo para cada grupo de réplicas, sorteados deste grupo de 100. Os sorteios, embora sejam de uma mesma população, são realizados separadamente para cada grupo.

Tanto para as simulações dos grupos gerados separadamente como para os grupos obtidos através de sorteios realizou-se o teste de Kolmogorov-Smirnov[12] com o objetivo de verificar se as distribuições das espécies seguiam uma distribuição Poisson(λ) para λ=5 e λ=8.

# 5. ANÁLISES E RESULTADOS

Antes de realizar as investigações realizou-se um pequeno exemplo, apresentado a seguir, mostrando a forma de cálculo e estabelecendo algumas questões que não foram encontradas nos artigos consultados para este trabalho.

Construiu-se através de simulação (utilizando o software R, através do comando *rpois*) dois grupos desbalanceados, cada um com três espécies. O primeiro grupo (ou local, ou tratamento) é composto por duas réplicas (ou amostras) com λ = 15 e o segundo grupo com quatro réplicas com λ = 5. A seguir estão os valores obtidos e a sua representação, tal como os ecologistas utilizam.

*Quadro 4: Simulação das três espécies com dois grupos contendo duas e quatro réplicas respectivamente.*

| RÉPLICAS | | | ESPÉCIES | | |
|---|---|---|---|---|---|
| | | | ESPÉCIE 1 | ESPÉCIE 2 | ESPÉCIE 3 |
| | Grupo Um (Fator Um) | **S11** | 11 | 11 | 20 |
| | | **S12** | 13 | 14 | 21 |
| | Grupo Dois (Fator Dois) | **S21** | 6 | 4 | 8 |
| | | **S22** | 2 | 7 | 6 |
| | | **S23** | 5 | 7 | 3 |
| | | **S24** | 7 | 4 | 2 |

*FONTE: Elaborado pela autora*

A partir dos dados simulados é obtida então a matriz de similaridades, através da distância de Bray-Curtis, bem como o cálculo do R Global que leva em consideração o *ranking* das similaridades.

A primeira dúvida surge com relação à quantidade de permutações possíveis quando se está diante de um quadro de desbalanceamento. Para esta dúvida não foi encontrada resposta na bibliografia consultada, por este motivo deduziu-se a expressão 5.1 apresentada a seguir:

[12] Andrey Kolmogorov (1903-1987), matemático soviético e Vladimir Ivanivich Smirnov (1887-1974), matemático russo, desenvolveram o teste usado para determinar se duas distribuições de probabilidade subjacentes diferem uma da outra ou se uma das distribuições de probabilidade subjacentes difere da distribuição em hipótese, em qualquer dos casos com base em amostras finitas.

$$T = \frac{\left(\sum_{i=1}^{k} n_i\right)!}{\prod_{i=1}^{k} n_i!} \qquad (5.1)$$

onde: k = número de grupos; $\left(\sum_{i=1}^{k} n_i\right)!$ = fatorial do total de elementos, isto é ($n_1 + n_2 + ... + n_K$)! e $\prod_{i=1}^{k} n_i!$ é o produto dos fatoriais do número de elementos em cada grupo, isto é, $n_1!$ $n_2!$ ... $n_K!$ Este número é bem menor que o total de permutações para grupos balanceados, mas ainda pode ser um valor muitíssimo alto dependendo da quantidade de réplicas existentes em cada grupo bem como da quantidade de grupos.

Voltando ao exemplo, observa-se que para estes dados originais estão salientados três grupos distintos a partir dos quais são calculados os valores para a distribuição da estatística R obtida através das permutações. O número total de permutações possíveis para este exemplo, onde $n_1 = 2$ e $n_2 = 4$, fazendo uso da expressão 5.1 é:

$$T = \frac{\left(\sum_{i=2}^{2} n_i\right)!}{\prod_{i=1}^{2} n_i!} = \frac{(2+4)!}{2!.4!} = \frac{6!}{2!.4!} = \frac{6.5.4!}{2.4!} = 15 \text{ permutações}$$

Cabe lembrar que os três grupos distintos em questão são: a) as similaridades dentro do grupo 1; b) as similaridades dentro do grupo 2; e c) as similaridades entre os dois grupos.

O Quadro 5 apresenta a matriz de similaridades utilizando a distância de Bray-Curtis (Fórmula 3.9). Através desta matriz pode-se perceber que o maior valor é 93,33, que pela legenda é justamente a similaridade entre as réplicas 1 e 2 dentro do grupo 1. Este fato fica mais evidente quando se observa o Quadro 6 composto pelos *rankings* destas similaridades.

*Quadro5: Similaridades baseadas na Distância de Bray-Curtis sem transformações*

| | S11 | S12 | S21 | S22 | S23 | S24 |
|---|---|---|---|---|---|---|
| **S11** | | | | | | |
| **S12** | 93,33333 | | | | | |
| **S21** | 60 | 54,54545 | | | | |
| **S22** | 52,63158 | 47,61905 | 72,72727 | | | |
| **S23** | 52,63158 | 47,61905 | 72,72727 | 80 | | |
| **S24** | 47,27273 | 42,62295 | 77,41935 | 57,14286 | 78,57143 | |

Legenda:
- Similaridades dentro do Grupo 1
- Similaridades dentro do Grupo 2
- Similaridades entre os Grupos 1 e 2

*FONTE: Elaborado pela autora*

Apresentam-se então, no Quadro 6, os *rankings*, em ordem decrescente, das similaridades e então calcula-se as médias dos *rankings* das similaridades dentro ($\overline{R}_W$) e entre ($\overline{R}_B$) os grupos, por este motivo os valores estão apresentados com cores distintas para que se possa diferenciar os valores dentro de cada um dos grupos e entre estes. Cabe salientar que quanto menor o *ranking* maior é a similaridade entre as réplicas envolvidas.

*Quadro 6: Similaridades em ordem decrescente e cálculos das médias das similaridades dentro e entre os grupos*

| Similaridades | Ordem |
|---|---|
| 93,33333333 | 1 |
| 80 | 2 |
| 78,57142857 | 3 |
| 77,41935484 | 4 |
| 72,72727273 | 5,5 |
| 72,72727273 | 5,5 |
| 60 | 7 |
| 57,14285714 | 8 |
| 54,54545455 | 9 |
| 52,63157895 | 10,5 |
| 52,63157895 | 10,5 |
| 47,61904762 | 12,5 |
| 47,61904762 | 12,5 |
| 47,27272727 | 14 |
| 42,62295082 | 15 |

Cálculo das médias dos *rankings* das similaridades entre os Grupos:

$$\overline{r}_B = \frac{7+9+10{,}5+10{,}5+12{,}5+12{,}5+14+14}{8} = 11{,}375$$

Cálculo das médias dos *rankings* das similaridades dentro dos Grupos:

$$\overline{r}_W = \frac{1+2+3+4+5{,}5+5{,}5+8}{7} = 4{,}142857$$

*FONTE: Elaborado pela autora*

No decorrer da realização do exemplo, surgiu o primeiro questionamento relacionado aos empates resultantes de *rankings* iguais, muito comuns em testes não-paramétricos que fazem uso de postos em suas estatísticas. Para resolver esta dúvida, foi utilizado o software PRIMER, onde através de simulações, comparando-se os resultados obtidos, deduziu-se que é realizada a média aritmética simples dos *rankings* correspondentes. E, por este motivo, na presença de empates, utilizou-se a média aritmética simples dos *rankings* correspondentes (Quadro 6).

Considerando um grupo com 2 réplicas e um grupo com 4 réplicas, tem-se um total de n = 6 réplicas. Desta forma é possível calcular o valor total de *rankings* (ou de similaridades), dado por M = 6.(6-1)/2 = 15 necessário para o cálculo do valor de R Global obtido através da expressão 3.10:

$$R = \frac{(\bar{r}_B - \bar{r}_w)}{\frac{1}{2}M} = \frac{11{,}375 - 4{,}142857}{\frac{1}{2}15} = 0{,}964286 \approx 0{,}96$$

Este é o valor da estatística de teste R, mas para rejeitar (ou não rejeitar) a hipótese nula, de que os grupos são similares, é necessário que se tenha uma distribuição que é obtida através das permutações, isto é, as réplicas do grupo 1 são permutadas com as réplicas do grupo 2 e o cálculo de R é realizado para cada caso. Neste exercício foi possível realizar todas as 15 permutações e gerar a distribuição da estatística R completa apresentada na Figura 1 a seguir, onde já está destacado o valor da estatística R obtida.

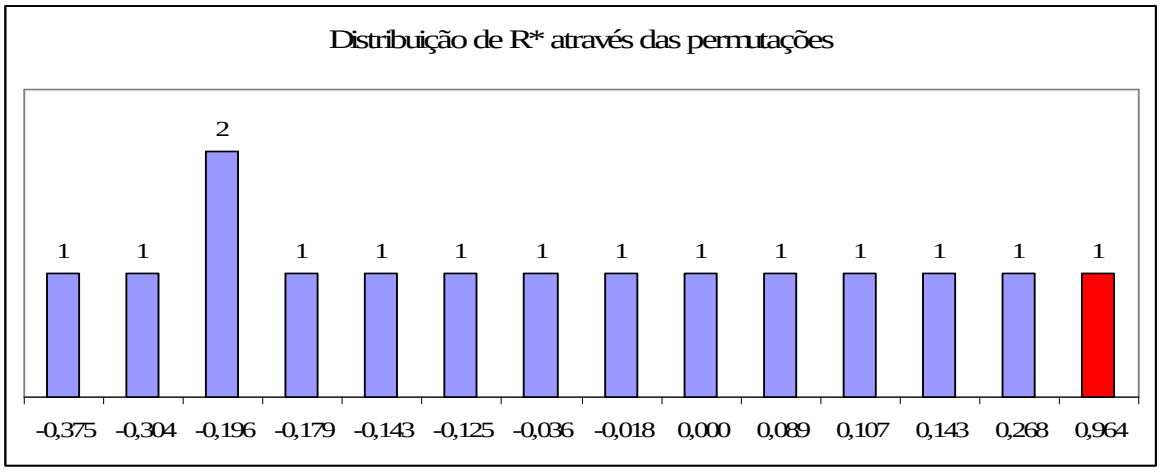


*Figura 1: Gráfico de Barras com a Distribuição de R**
*Fonte: Dados obtidos através das 15 permutações das três espécies em grupos de 2 e 4 réplicas*

A partir do gráfico apresentado na Figura 1, pode-se então alocar a estatística R obtida e calcular seu p-valor através da expressão t/T (se utiliza esta expressão pois todas as permutações possíveis foram consideradas), onde T é o número total de permutações e t é o número de permutações maiores ou iguais à estatística R. Para este caso tem-se que t = 1 e T = 15. Desta forma o p-valor para o R Global é 1/15 = 0,0667 (ou em percentual, como é utilizado pelos criadores da estatística R, p-valor = 6,7%). Este é um exemplo apenas para ilustrar a técnica e seus cálculos. Neste caso, então, embora não se devesse rejeitar a Hipótese Nula, deve-se ter claro que não existe a possibilidade de um p-valor menor que este. Desta forma se utilizarmos um nível de significância de 0,05, quando se está diante de poucas réplicas não se poderá rejeitar a Hipótese Nula de igualdade entre os grupos mesmo que tal diferença seja significativa.

Surge uma nova questão no decorrer dos estudos, a saber, quando se está diante de tais permutações e se deve recalcular as similaridades na realidade se está trocando de lugar os valores dos *rankings*. É importante saber se a quantidade que a média entre grupos "perde" é a mesma quantidade que a média dentro dos grupos "ganha". Também para esta dúvida não se encontrou resposta na literatura, então deduziu-se a expressão 5.2. A cada permutação a média dos *rankings* dentro (e respectivamente entre) dos grupos aumenta (ou diminui) da seguinte quantidade:

$$\Delta = 2 * \left( \frac{\text{diferença dos ranks perdidos entre}}{\text{Número de similaridades entre}} - \frac{\text{diferença dos ranks adquiridos dentro}}{\text{Número de similaridades dentro}} \right) \tag{5.2}$$

Este valor pode ser positivo, ou negativo, dependendo dos *rankings* calculados dentro e entre os grupos. Assim, a partir do valor R original, a cada permutação tem-se para o valor de $R^* = R + \Delta$. Cabe também observar que o denominador não tem um valor fixo, mesmo para grupos balanceados, pois existem valores distintos para a quantidade de *rankings* dentro e entre os grupos. Por exemplo, para uma caso onde se tenha dois grupos cada um com 5 réplicas tem-se M = 45 similaridades onde 25 são entre os grupos e apenas 20 são similaridades dentro dos grupos.

A seguir são apresentadas as expressões 5.3 e 5.4 que indicam o número de *rankings* dentro e entre grupos, quando se está diante de situações tanto de balanceamento como de desbalanceamento. Sejam $n_1$, $n_2$, ... $n_k$ a quantidade de réplicas em cada um dos k grupos. O número de *rankings* dentro dos grupos é dado por:

$$\sum_{i=1}^{k} \frac{n_k.(n_k - 1)}{2} \tag{5.3}$$

E o número de *rankings* entre grupos é dado por:

$$\sum_{i=1}^{k} \left( n_i \sum_{j=i+1}^{k} n_j \right) \tag{5.4}$$

Utilizando as expressões 5.3 e 5.4, para este exemplo, obtém-se que o número de *rankings* dentro os grupos é 7:

$$\sum_{i=1}^{k} \frac{n_k.(n_k - 1)}{2} = \sum_{i=1}^{2} \frac{n_k.(n_k - 1)}{2} = \frac{n_1(n_1 - 1)}{2} + \frac{n_2(n_2 - 1)}{2} = \frac{2(2-1)}{2} + \frac{4(4-1)}{2} = 1 + 6 = 7$$

Da mesma forma, o número de *rankings* entre os grupos é 8, obtido da seguinte forma:

$$\sum_{i=1}^{k}\left(n_i\sum_{j=i+1}^{k}n_j\right)=\sum_{i=1}^{2}\left(n_i\sum_{j=i+1}^{2}n_j\right)=n_1(n_2)=2.4=8$$

A partir daqui são apresentados os resultados das simulações realizadas onde se trabalha com apenas dois grupos, foco principal deste estudo. Nestes grupos os dados se encontram desbalanceados da várias formas distintas.

Como as espécies se distribuem de acordo com uma distribuição de Poisson, foram escolhidos dois valores distintos para λ, cinco e oito. A partir desta escolha construiu-se cinco séries de exercícios com desbalanceamentos que são apresentados a seguir:

## 5.1 EXERCÍCIOS COM GRUPOS DESBALANCEADOS E VALORES PARA Λ DISTINTOS GERADOS SEPARADAMENTE

Para esta série de exercícios geraram-se as três espécies nos grupos separadamente através do comando *rpois* no software R.

### 5.1.1 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 4$

O Quadro 7 apresenta o resumo dos exercícios realizados para esta série com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor (o número de permutações com valor para R* igual ou superior ao R Global) e a decisão estatística a  um nível de 5% de significância. Cabe relembrar que o p-valor é obtido através da expressão t/T quando se realizou todas as permutações possíveis, ou (t+1)/(T+1) quando isso não foi possível. Desta forma o p-valor 0,667 = 10/15 onde 10 é o número de permutações com valor para R* igual ou superior ao R original e 15 é o número de permutações possíveis, sendo todas elas realizadas.

*Quadro 7: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 4$*

| Tamanho dos Grupos | Valores de λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 4$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 15 | 15 | - 0,143 | 0,667 (10) | Os grupos são iguais ** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 15 | 15 | 0, 357 | 0,267 (4) | Os grupos são iguais ** |
| $n_1 = 4$ e $n_2 = 8$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 495 | 495 | 0,302 | 0,057 (21) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 495 | 495 | 0,267 | 0,071 (35) | **Os grupos não são iguais** |
| $n_1 = 8$ e $n_2 = 16$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 735.471 | 999 | 0,335 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,335 | 0,0008 (3) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 735.471 | 999 | 0,203 | 0,035 (34) | **Os grupos não são iguais** |
| | | | 4999 | 0,203 | 0,035 (153) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Observa-se que embora se tenha construído os grupos com valores para λ distintos e que 495 permutações seja um número razoável de permutações (e sendo o número total de permutações possíveis), mesmo assim não se conseguiu concluir que os grupos eram realmente distintos. Ao se averiguar os valores da estatística K-S obteve-se, para o grupo onde $n_1$ apresenta 4 réplicas com λ =5, os seguintes valores para cada uma das espécies: E1 apresentou p-valor = 0,4195 para a estatística de teste K-S, a espécie E2 p-valor = 0,4195 e a espécie E3 p-valor = 0,6576. Enquanto que as espécies do grupo onde $n_2$ apresentava 8 réplicas com λ =8 obteve-se os seguintes p-valores para cada uma das três espécies respectivamente: 0,6371, 0,9326 e 0,3558. Para auxiliar na compreensão e visualização destes resultados apresenta-se a seguir a tabela com a distribuição das espécies geradas bem como a matriz de similaridades entre as espécies.

*Tabela 1: Distribuição das espécies para o caso em que $n_1 = 4$ réplicas com λ =5*

| | | | ESPÉCIES | | |
|---|---|---|---|---|---|
| | | | ESPÉCIE 1 | ESPÉCIE 2 | ESPÉCIE 3 |
| RÉPLICAS | Grupo 1 λ =5 | S11 | 5 | 7 | 5 |
| | | S12 | 8 | 5 | 1 |
| | | S13 | 4 | 5 | 8 |
| | | S14 | 8 | 4 | 5 |
| | Grupo 2 λ =8 | S21 | 9 | 4 | 7 |
| | | S22 | 6 | 11 | 12 |
| | | S23 | 6 | 7 | 8 |
| | | S24 | 7 | 6 | 7 |
| | | S25 | 14 | 10 | 15 |
| | | S26 | 11 | 6 | 11 |
| | | S27 | 5 | 10 | 6 |
| | | S28 | 13 | 6 | 8 |

*FONTE: Dados obtidos através do comando rpois(n, λ) no software R.*

| | S11 | S12 | S13 | S14 | S21 | S22 | S23 | S24 | S25 | S26 | S27 | S28 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **S11** | | | | | | | | | | | | |
| **S12** | 70,97 | | | | | | | | | | | |
| **S13** | 82,35 | 64,52 | | | | | | | | | | |
| **S14** | 82,35 | 83,87 | 76,47 | | | | | | | | | |
| **S21** | 75,68 | 76,47 | 81,08 | 91,89 | | | | | | | | |
| **S22** | 73,91 | 55,81 | 73,91 | 65,22 | 69,39 | | | | | | | |
| **S23** | 89,47 | 68,57 | 89,47 | 78,95 | 82,93 | 84,00 | | | | | | |
| **S24** | 86,49 | 76,47 | 86,49 | 86,49 | 90,00 | 77,55 | 92,68 | | | | | |
| **S25** | 60,71 | 52,83 | 60,71 | 60,71 | 67,80 | 82,35 | 70,00 | 67,80 | | | | |
| **S26** | 71,11 | 66,67 | 75,56 | 75,56 | 83,33 | 80,70 | 81,63 | 83,33 | 83,58 | | | |
| **S27** | 89,47 | 62,86 | 78,95 | 73,68 | 73,17 | 84,00 | 85,71 | 82,93 | 70,00 | 69,39 | | |
| **S28** | 72,73 | 68,29 | 77,27 | 77,27 | 85,11 | 71,43 | 83,33 | 85,11 | 81,82 | 90,91 | 70,83 | |

*Figura 2: Matriz de similaridade, com a distância de Bray-Curtis, entre as espécies para o caso em que $n_1$ = 4 réplicas com $\lambda$ =5*
*FONTE: Dados obtidos através do software PRIMER.*

Pode-se observar através destes resultados que apenas quando se tem grupos com $n_1$ = 8 e $n_2$ = 16 réplicas se consegue efetivamente rejeitar a hipótese nula que afirma que os grupos são iguais. Desta forma decidiu-se realizar 50 repetições com 999 e 4999 permutações para verificar se o cálculo do p-valor se mantém. Nas figuras 3 e 4 se encontram as distribuições dos p-valores encontrados para cada uma destas 50 repetições. Em cada uma das figuras está destacado o valor obtido na primeira repetição.

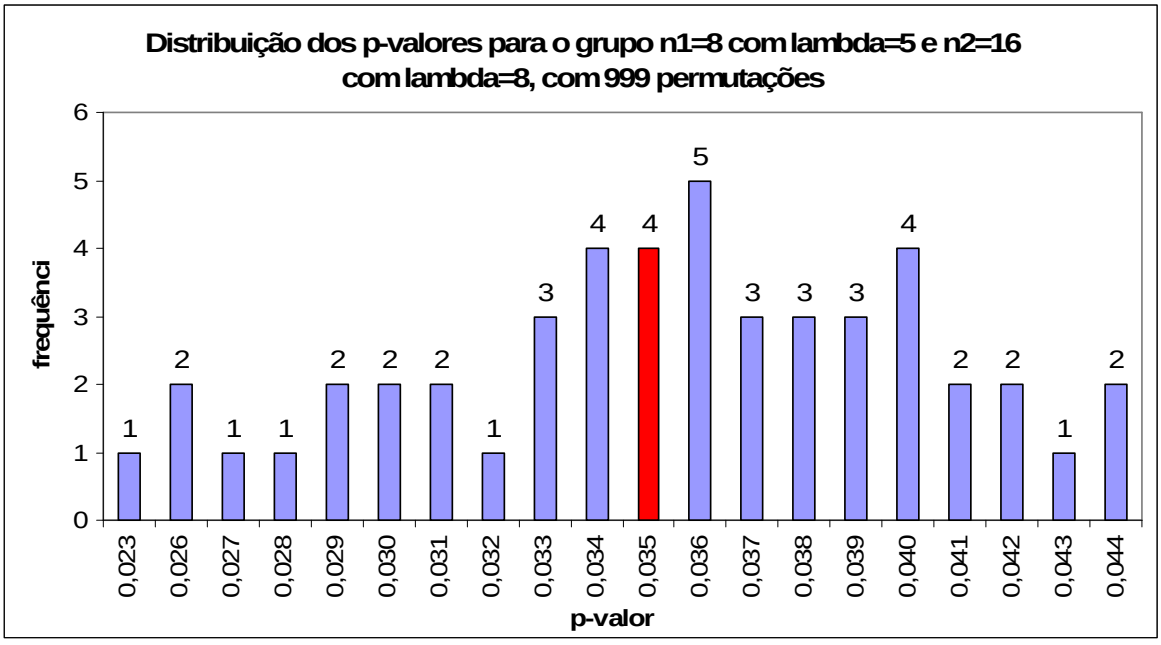


*Figura 3: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8($\lambda$=5) x 16($\lambda$=8) réplicas com 999 permutações, com 20 repetições..*

Observa-se, através das figuras 2 e 3 que ao aumentar o número de permutações o intervalo dentro do qual variam os valores encontrados para o p-valor diminui, porém

mantém como valor central 0,035. Jackson e Somers (1988) comprovaram que na medida em que o número de matrizes de permutações cresce a instabilidade no teste de Mantel (que também se baseia em matrizes de similaridades) decresce, diminuído o intervalo de variação do p-valor. No mesmo trabalho uma das recomendações é que quando o p-valor está muito próximo do limite de 0,05 se realize de 50.000 até 100.000 permutações.

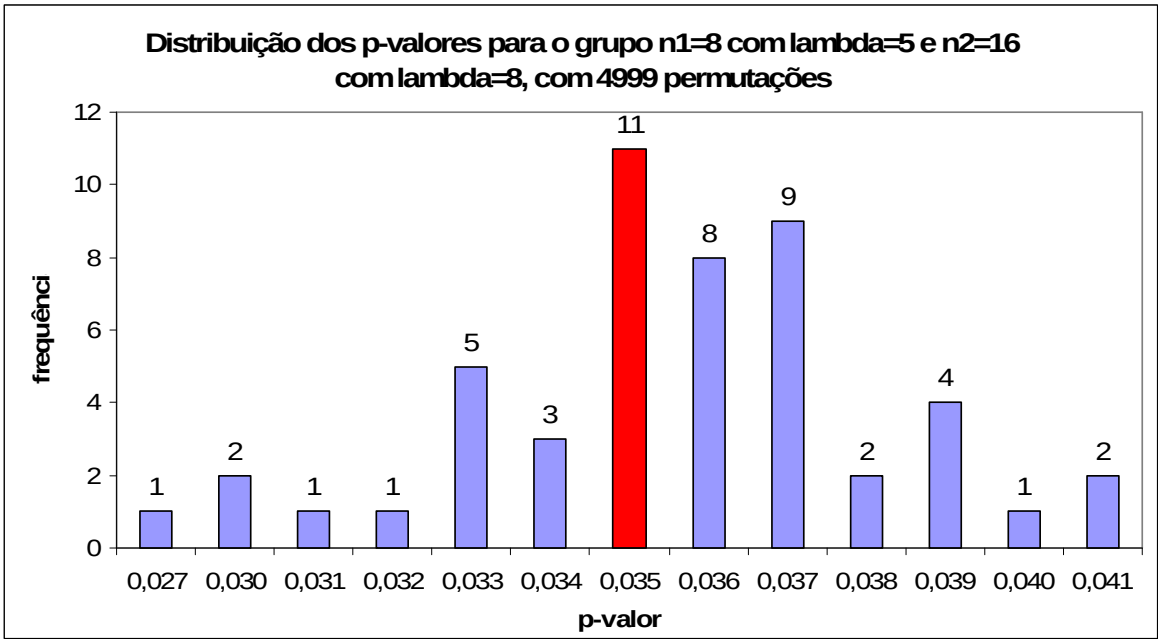


*Figura 4: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8(λ=5) x 16(λ=8) réplicas com 4999 permutações, com 20 repetições.*

### 5.1.2 Série de Exercícios com grupos de $n_1 = 3$ e $n_2 = 5$

A seguir, no quadro 8, apresenta-se os exercícios realizados para esta série com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a um nível de 5% de significância.

*Quadro 8: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 4$*

| Tamanho dos Grupos | Valores de λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 3$ e $n_2 = 5$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 56 | 56 | 0,374 | 0,071 (4) | Os grupos são iguais ** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 56 | 56 | 0,938 | 0,018 (1) | **Os grupos não são iguais** |
| $n_1 = 6$ e $n_2 = 10$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 8.008 | 999 | 0,436 | 0,003 (2) | **Os grupos não são iguais** |
| | | | 4999 | 0,436 | 0,002 (10) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 8.008 | 999 | 0,421 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,421 | 0,001 (2) | **Os grupos não são iguais** |
| $n_1 = 12$ e $n_2 = 20$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 225.792.840 | 999 | 0,306 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,306 | <0,0001 (0) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 225.792.840 | 999 | 0,409 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,409 | 0,0004 (1) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Para o único caso em que o teste ANOSIM não rejeitou a hipótese nula quando deveria tê-la rejeitado foi justamente o caso onde se tem a menor amostra. Para este caso não se deve levar em consideração os resultados contrários ao que seria esperado, mesmo porque o teste é realizado levando em consideração todas as permutações possíveis e as amostras são muito pequenas gerando assim um número relativamente pequeno de permutações.

### 5.1.3 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 3$

O Quadro 9 apresenta a comparação dos exercícios realizados para esta série com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância..

*Quadro 9: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 3$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 3$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 10 | 10 | 0,167 | 0,40 (4) | Os grupos são iguais ** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 10 | 10 | 0,25 | 0,30 (3) | Os grupos são iguais ** |
| $n_1 = 4$ e $n_2 = 6$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 210 | 210 | 0,706 | 0,005 (1) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 210 | 210 | 0,456 | 0,019 (4) | **Os grupos não são iguais** |
| $n_1 = 8$ e $n_2 = 12$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 125.970 | 999 | 0,322 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,322 | 0,002 (7) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 125.970 | 999 | 0,325 | 0,002 (1) | **Os grupos não são iguais** |
| | | | 4999 | 0,325 | 0,002 (9) | **Os grupos não são iguais** |
| $n_1 = 16$ e $n_2 = 24$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 62.852.101.650 | 999 | 0,258 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,258 | 0,001 (2) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 62.852.101.650 | 999 | 0,282 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,282 | 0,0002 (0) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Observa-se neste caso que também apenas para o caso das pequenas amostras é que houve comportamento diferente daquele esperado.

### 5.1.4 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 5$

A seguir o Quadro 10 apresenta o comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância..

*Quadro 10: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 5$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 5$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 21 | 21 | 0,309 | 0,143 (3) | Os grupos são iguais ** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 21 | 21 | 0,1 | 0,381 (8) | Os grupos são iguais ** |
| $n_1 = 4$ e $n_2 = 10$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 1.001 | 999 | 0,634 | 0,001 (0) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 1.001 | 999 | 0,834 | 0,001 (0) | **Os grupos não são iguais** |
| $n_1 = 8$ e $n_2 = 20$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 3.108.105 | 999 | 0,055 | 0,213 (212) | **Os grupos são iguais **** |
| | | | 4999 | 0,055 | 0,253 (1265) | **Os grupos são iguais **** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 3.108.105 | 999 | 0,334 | 0,005 (4) | **Os grupos não são iguais** |
| | | | 4999 | 0,334 | 0,005 (24) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Novamente, como no exercício anterior, para o caso das pequenas amostras houve comportamento contrário daquele esperado na interpretação dos dados. Quando as amostras começam a aumentar que é o caso em que $n_1=4$ e $n_2=10$, observa-se que a estatística de teste decide de forma correta (a 0,05 de significância) que os grupos são distintos. Porém para o caso em que $n_1=8$ ($\lambda=8$) e $n_2=20$ ($\lambda=5$) o teste viesa novamente.

A seguir, na Tabela 2, as espécies obtidas através da simulação para realizar este experimento. A matriz de similaridades para este exercício encontra-se no Anexo 1 ao final deste trabalho.

*Tabela 2:Distribuição das espécies para o caso em que $n_1$ = 8 réplicas com $\lambda$ =8*

| | | | ESPÉCIES | | |
|---|---|---|---|---|---|
| | | | ESPÉCIE 1 | ESPÉCIE 2 | ESPÉCIE 3 |
| RÉPLICAS | Grupo 1 | S11 | 6 | 9 | 5 |
| | | S12 | 6 | 5 | 8 |
| | | S13 | 7 | 9 | 8 |
| | | S14 | 7 | 13 | 6 |
| | | S15 | 11 | 4 | 10 |
| | | S16 | 8 | 7 | 9 |
| | | S17 | 9 | 6 | 8 |
| | | S18 | 6 | 6 | 10 |
| | Grupo 2 | S21 | 5 | 0 | 6 |
| | | S22 | 4 | 6 | 4 |
| | | S23 | 6 | 2 | 2 |
| | | S24 | 9 | 5 | 7 |
| | | S25 | 8 | 5 | 5 |
| | | S26 | 9 | 8 | 6 |
| | | S27 | 7 | 10 | 8 |
| | | S28 | 3 | 5 | 9 |
| | | S29 | 4 | 4 | 3 |
| | | S210 | 9 | 4 | 3 |
| | | S211 | 5 | 4 | 4 |
| | | S212 | 3 | 2 | 2 |
| | | S213 | 6 | 2 | 7 |
| | | S214 | 4 | 7 | 7 |
| | | S215 | 1 | 2 | 2 |
| | | S216 | 9 | 7 | 3 |
| | | S217 | 4 | 6 | 2 |
| | | S218 | 3 | 5 | 3 |
| | | S219 | 6 | 0 | 4 |
| | | S220 | 5 | 5 | 4 |

*FONTE: Dados obtidos através do comando rpois(n, λ) no software R.*

A Tabela 3 apresenta os valores da estatística K-S para cada espécie em cada um dos dois grupos, tanto para o caso em que se tem $n_1$=8 com $\lambda$ =8 como quando $n_2$=8 com $\lambda$ =5.

*Tabela 3: Valores da estatística K-S para cada espécie nos grupos $n_1$= 8 e $n_2$= 20.*

| | K-S grupo $n_1$=8 ($\lambda$ =8) | K-S grupo $n_2$=20 ($\lambda$=5) | K-S grupo $n_1$=20($\lambda$ =8) | K-S grupo $n_2$=8 ($\lambda$=5) |
|---|---|---|---|---|
| E1 | 0,4118 | 0,1976 | 0,3822 | 0,2324 |
| E2 | 0,9433 | 0,5701 | 0,6353 | 0,9482 |
| E3 | 0,3049 | 0,6889 | 0,489 | 0,7417 |

*FONTE: Valores obtidos através do software SPSS 20.0*

Como se trata de um caso em que o teste deveria ter rejeitado a Hipóteese nula de igualdade entre os grupos decidiu-se realizar 20 repetições com 999 e 4999 permutações para verificar se a estatística de teste se mantém estável. Cabe observar que ao seguir a recomendação de Jackson e Somers (1988) se realizou 49.999 permutações e o p-valor encontrado foi de 0,253. Valor este que foi obtido já com 4.999 permutações.

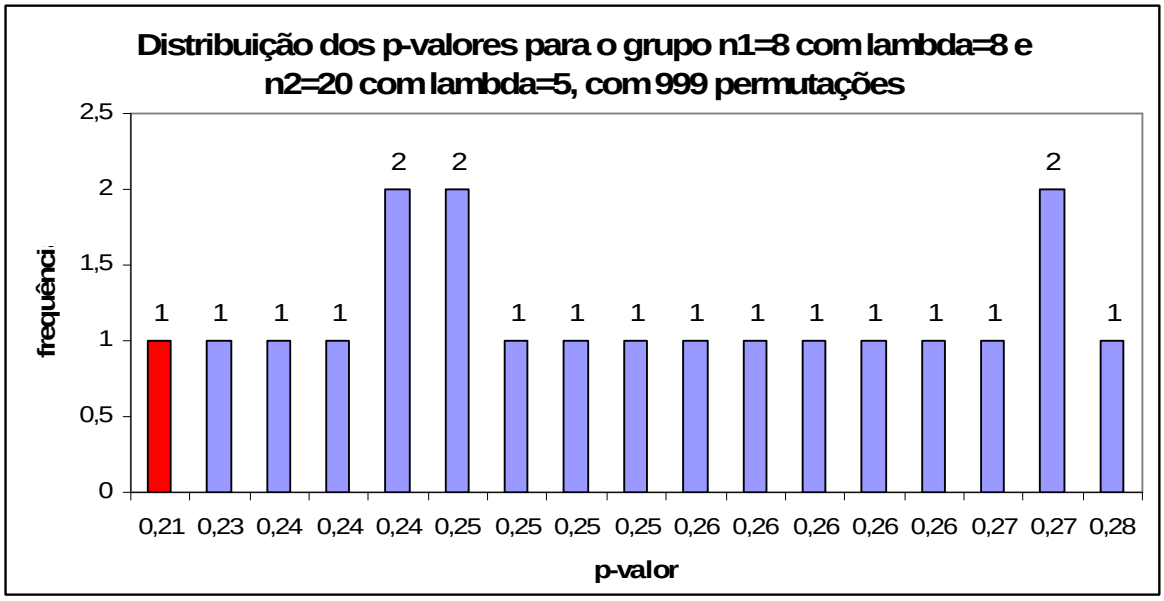


*Figura 5: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8(λ=8) x 20(λ=5) réplicas com 999 permutações, com 20 repetições..*

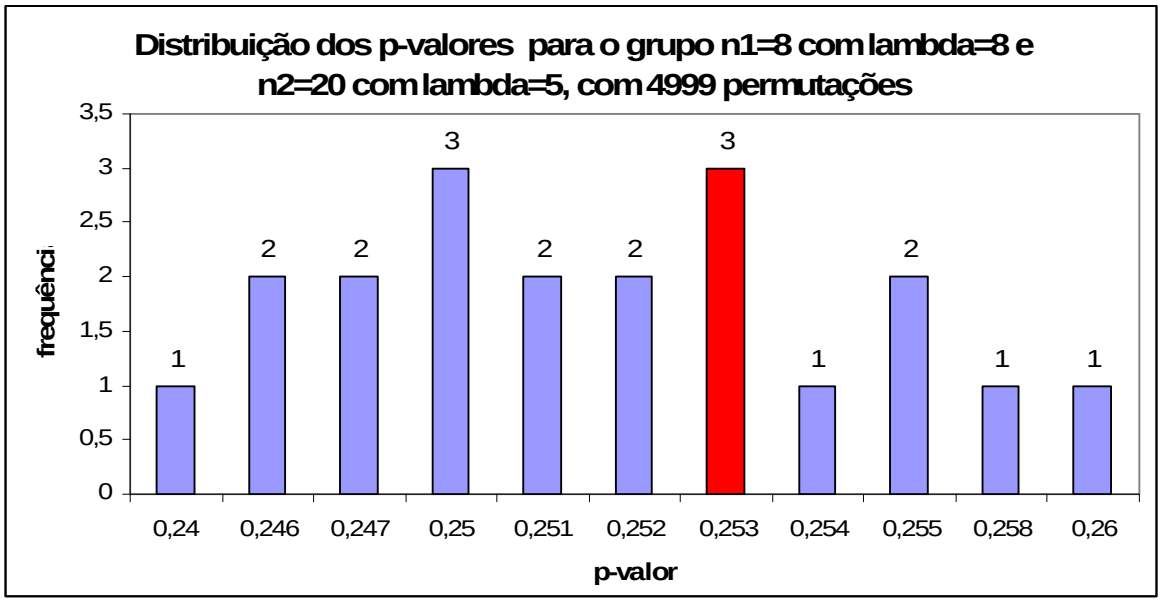


*Figura 6: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8(λ=8) x 20(λ=5) réplicas com 4999 permutações, com 20 repetições..*

### 5.1.5 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 8$

Quadro comparativo dos exercícios realizados para esta série com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância..

*Quadro 11: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 8$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 8$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 45 | 45 | 0,392 | 0,111 (5) | Os grupos são iguais ** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 45 | 45 | 0,013 | 0,489 (22) | Os grupos são iguais ** |
| $n_1 = 4$ e $n_2 = 16$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 4.845 | 999 | 0,371 | 0,028 (27) | **Os grupos não são iguais** |
| | | | 4845 | 0,371 | 0,025 (122) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 4.845 | 999 | 0,621 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4845 | 0,621 | 0,001 (3) | **Os grupos não são iguais** |
| $n_1 = 8$ e $n_2 = 32$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 76.904.685 | 999 | 0,256 | 0,021 (20) | **Os grupos não são iguais** |
| | | | 4999 | 0,256 | 0,014 (67) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 76.904.685 | 999 | 0,458 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,458 | 0,001 (3) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Neste bloco de simulações não houve resultados distintos do esperado na interpretação dos valores p, com exceção dos casos onde os grupos são muito pequenos e, consequentemente, não se pode realizar muitas permutações.

## 5.2 EXERCÍCIOS COM GRUPOS DESBALANCEADOS E VALORES PARA Λ = 8 EM AMBOS OS GRUPOS GERADOS SEPARADAMENTE

A seguir serão apresentados exercícios com grupos desbalanceados, porém com o mesmo lambda, apenas para verificar que o fato de haver uma quantidade menor de permutações possíveis quando os grupos são desbalanceados não impede que o teste detecte que os grupos são iguais quando estes realmente o são.

### 5.2.1 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 4$

Quadro comparativo dos exercícios realizados para esta série com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância..

*Quadro 12: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 4$*

| Tamanho dos Grupos | Valores de λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 4$ | $\lambda_1 = \lambda_2 = 8$ | 15 | 15 | 0,607 | 0,067 (1) | **Os grupos são iguais** |
| $n_1 = 4$ e $n_2 = 8$ | $\lambda_1 = \lambda_2 = 8$ | 495 | 495 | 0,136 | 0,194 (96) | **Os grupos são iguais** |
| $n_1 = 8$ e $n_2 = 16$ | $\lambda_1 = \lambda_2 = 8$ | 735.471 | 999 | - 0,071 | 0,788 (776) | **Os grupos são iguais** |
| | | | 4999 | - 0,071 | 0,787 (3935) | **Os grupos são iguais** |

*FONTE: Elaborado pela autora*

Cabe observar aqui que, mesmo quando os grupos são pequenos, nenhum caso apresentou comportamento distinto do esperado na decisão estatística.

### 5.2.2 Série de Exercícios com grupos de $n_1 = 3$ e $n_2 = 5$

Quadro comparativo dos exercícios realizados para esta série com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a um nível de 5% de significância.

*Quadro 13: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 3$ e $n_2 = 5$*

| Tamanho dos Grupos | Valores de λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 3$ e $n_2 = 5$ | $\lambda_1 = \lambda_2 = 8$ | 56 | 56 | 0,251 | 0,143 (8) | **Os grupos são iguais** |
| $n_1 = 6$ e $n_2 = 10$ | $\lambda_1 = \lambda_2 = 8$ | 8.008 | 999 | 0,029 | 0,322 (321) | **Os grupos são iguais** |
| | | | 4999 | 0,029 | 0,323 (1614) | **Os grupos são iguais** |
| $n_1 = 12$ e $n_2 = 20$ | $\lambda_1 = \lambda_2 = 8$ | 225.792.840 | 999 | -0,032 | 0,677 (676) | **Os grupos são iguais** |
| | | | 4999 | -0,032 | 0,668 (3341) | **Os grupos são iguais** |

*FONTE: Elaborado pela autora.*

Também neste caso não houve problemas em aceitar a hipótese nula quando ela deveria ser mesmo aceita.

### 5.2.3 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 3$

Quadro comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a um nível de 5% de significância..

*Quadro 14: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 3$*

| Tamanho dos Grupos | Valores de λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 3$ | $\lambda_1 = \lambda_2 = 8$ | 10 | 10 | - 0,417 | 0,90 (9) | **Os grupos são iguais** |
| $n_1 = 4$ e $n_2 = 6$ | $\lambda_1 = \lambda_2 = 8$ | 210 | 210 | 0,008 | 0,424 (89) | **Os grupos são iguais** |
| $n_1 = 8$ e $n_2 = 12$ | $\lambda_1 = \lambda_2 = 8$ | 125.970 | 999 | -0,094 | 0,896 (895) | **Os grupos são iguais** |
| | | | 4999 | -0,094 | 0,899 (4493) | **Os grupos são iguais** |
| $n_1 = 16$ e $n_2 = 24$ | $\lambda_1 = \lambda_2 = 8$ | 62.852.101.650 | 999 | -0,013 | 0,577 (576) | **Os grupos são iguais** |
| | | | 4999 | -0,013 | 0,558 (2788) | **Os grupos são iguais** |

*FONTE: Elaborado pela autora*

Para este exercício a hipótese nula não pode ser rejeitada em nenhum caso, indicando assim que os grupos apresentam a mesma distribuição com média sem diferença significativa a 0,05, isto é, os grupos são considerados iguais.

### 5.2.4 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 5$

O Quadro 15 apresenta o comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a um nível de 5% de significância.

*Quadro 15: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 5$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 5$ | $\lambda_1 = \lambda_2 = 8$ | 21 | 21 | -0,164 | 0,714 (15) | **Os grupos são iguais** |
| $n_1 = 4$ e $n_2 = 10$ | $\lambda_1 = \lambda_2 = 8$ | 999 | 999 | -0,125 | 0,798 (797) | **Os grupos são iguais** |
| $n_1 = 8$ e $n_2 = 20$ | $\lambda_1 = \lambda_2 = 8$ | 3.108.105 | 999 | 0,000 | 0,462 (461) | **Os grupos são iguais** |
| | | | 4999 | 0,000 | 0,472 (2360) | **Os grupos são iguais** |

*FONTE: Elaborado pela autora*

Para esta série de exercícios também não houve rejeição da hipótese nula de que os grupos são significativamente iguais.

### 5.2.5 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 8$

Quadro comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância.

*Quadro 16: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 8$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 8$ | $\lambda_1 = \lambda_2 = 8$ | 45 | 45 | -0,091 | 0,667 (30) | **Os grupos são iguais** |
| $n_1 = 4$ e $n_2 = 16$ | $\lambda_1 = \lambda_2 = 8$ | 4845 | 999 | -0,144 | 0,787 (786) | **Os grupos são iguais** |
| | | | 4845 | -0,114 | 0,774 (3750) | **Os grupos são iguais** |
| $n_1 = 8$ e $n_2 = 32$ | $\lambda_1 = \lambda_2 = 8$ | 76.904.685 | 999 | 0,266 | 0,022 (21) | **Os grupos não são iguais **** |
| | | | 4999 | 0,266 | 0,016 (87) | **Os grupos não são iguais **** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Observa-se que para o caso em que $n_1$=8 e $n_2$=32 o teste não conseguiu indicar que os grupos eram iguais. Confrontando com os valores da estatística K-S obtida para estas espécies obteve-se os valores constantes da tabela a seguir:

*Tabela 4: Valores da estatística K-S para cada espécie nos grupos $n_1$= 8 e $n_2$= 32.*

| | K-S grupo $n_1$=8 (λ =8) | K-S grupo $n_2$=32 (λ=8) |
|---|---|---|
| E1 | 1,0 | 0,990 |
| E2 | 0,881 | 0,55 |
| E3 | 0,881 | 0,217 |

*FONTE: Valores obtidos através do software SPSS 20.0*

Pode-se notar que as espécies apresentam distribuição Poisson (λ =8), mas mesmo assim o teste não foi capaz de detectar que as espécies teriam a mesma distribuição, indicando que o teste pode gerar decisões contarias ao esperado também para detectar que as espécies apresentam a mesma distribuição com médias sem diferenças significativas. Nas figuras 8 e 9 a seguir apresenta-se os gráficos com os p-valores obtidos através das 20 repetições das 999 e 4999 permutações. Nos gráficos das figuras 8 e 9, estão salientados os p-valores encontrados para a primeira das 20 repetições em cada caso. Ao realizar 49.999 permutações o p-valor encontrado para a estatística de teste R foi de 0,013.

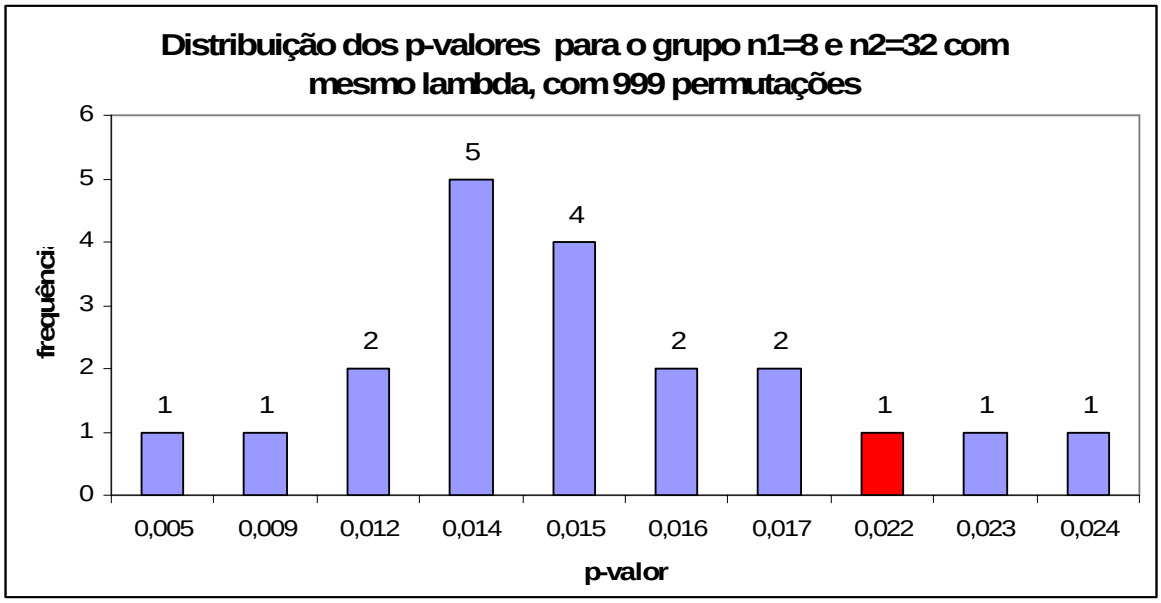


*Figura 8: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8x32 réplicas, ambas com λ=8, com 999 permutações, com 20 repetições.*

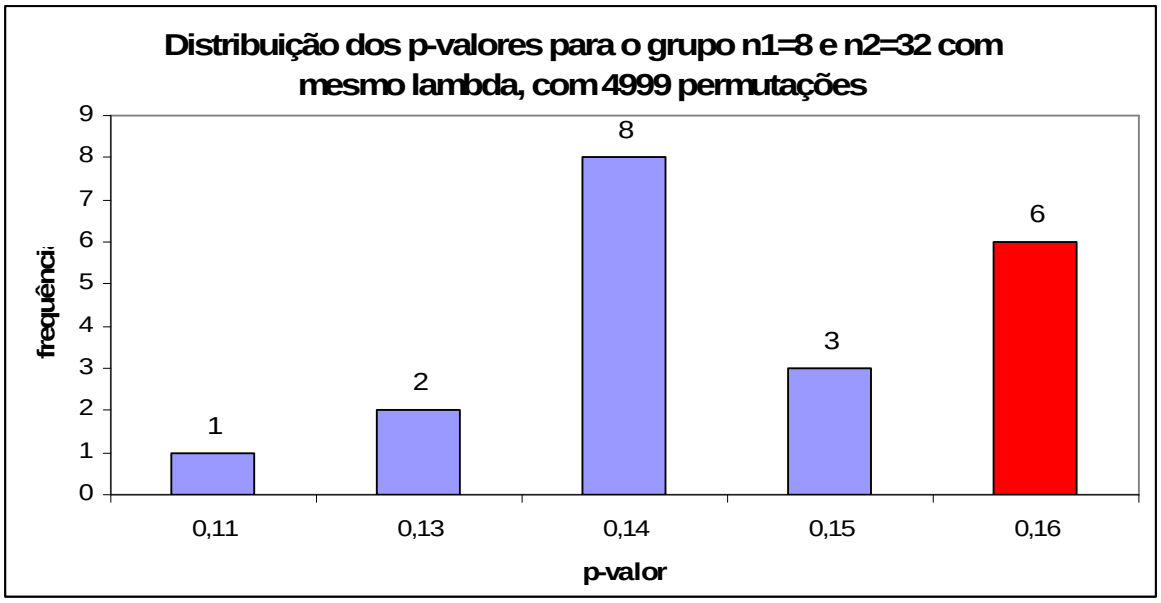


*Figura 9: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8x32 réplicas, ambas com λ=8, com 4999 permutações, com 20 repetições..*

## 5.3 EXERCÍCIOS COM GRUPOS DESBALANCEADOS E VALORES PARA Λ DISTINTOS GERADOS A PARTIR DE SORTEIOS

Para que se pudesse realmente simular a realidade da coleta de amostras em estudos ecológicos decidiu-se então realizar a simulação de uma população com três espécies com 100 observações segundo uma distribuição Poisson(λ) e então realizar o sorteio das amostras dentro destas populações. As 100 observações foram geradas separadamente e

alocadas em três colunas distintas para que cada valor sorteado (entre 1 e 100) representasse uma coleta (ou réplica). A seguir são apresentados os resultados.

### 5.3.1 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 4$

O Quadro 17 é o comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a um nível de 5% de significância.

*Quadro 17: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 4$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 4$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 15 | 15 | 0,5 | 0,133 (2) | Os grupos são iguais ** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 15 | 15 | 0,875 | 0,067 (1) | Os grupos são iguais ** |
| $n_1 = 4$ e $n_2 = 8$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 495 | 495 | 0,116 | 0,202 (100) | **Os grupos são iguais **** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 495 | 495 | 0,075 | 0,299 (148) | **Os grupos são iguais **** |
| $n_1 = 8$ e $n_2 = 16$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 735.471 | 999 | 0,304 | 0,006 (5) | **Os grupos não são iguais** |
| | | | 4999 | 0,304 | 0,004 (17) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 735.471 | 999 | 0,463 | 0,002 (1) | **Os grupos não são iguais** |
| | | | 4999 | 0,463 | 0,001 (2) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Observa-se que para estes dados o teste só conseguiu detectar que os grupos eram significativamente distintos a partir de amostras bem maiores. Na tabela 5 são apresentados os valores da estatística K-S para cada espécie em cada grupo, observa-se que as amostras que foram sorteadas de distribuições Poisson(λ) continuam apresentando distribuição Poisson(λ).

*Tabela 5: Valores da estatística K-S para cada espécie nos grupos $n_1 = 4$ e $n_2 = 8$.*

| | K-S grupo $n_1$=4 (λ =8) | K-S grupo $n_2$=8 (λ=5) | K-S grupo $n_1$=4 (λ =5) | K-S grupo $n_2$=8 (λ=8) |
|---|---|---|---|---|
| E1 | 0,2528 | 0,1816 | 0,09611 | 0,6371 |
| E2 | 0,476 | 0,2343 | 0,3178 | 0,4189 |
| E3 | 0,4307 | 0,1816 | 0,3036 | 0,4189 |

*FONTE: Valores obtidos através do software SPSS 20.0*

### 5.3.2 Série de Exercícios com grupos de $n_1 = 3$ e $n_2 = 5$

Quadro comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a um nível de 5% de significância.

*Quadro 18: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 3$ e $n_2 = 5$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 3$ e $n_2 = 5$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 56 | 56 | 0,041 | 0,429 (24) | Os grupos são iguais ** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 56 | 56 | 0,272 | 0,089 (5) | Os grupos são iguais ** |
| $n_1 = 6$ e $n_2 = 10$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 8.008 | 999 | 0,297 | 0,019 (18) | **Os grupos não são iguais** |
| | | | 4999 | 0,297 | 0,015 (75) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 8.008 | 999 | 0,684 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,684 | 0,0004 (1) | **Os grupos não são iguais** |
| $n_1 = 12$ e $n_2 = 20$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 225.792.840 | 999 | 0,233 | 0,007 (6) | **Os grupos não são iguais** |
| | | | 4999 | 0,233 | 0,002 (11) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 225.792.840 | 999 | 0,391 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,391 | 0,0002 (0) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Neste caso o teste só se apresentou distinto do esperado para o caso em que foram sorteadas poucas réplicas em cada grupo. Embora o teste K-S tenha indicado que as distribuições realmente apresentavam distribuições Poisson com λ distintos mesmo para as amostras pequenas.

### 5.3.3 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 3$

Quadro comparativo dos exercícios realizados para esta série com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a um nível de 5% de significância.

*Quadro 19: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 3$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 3$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 10 | 10 | 0,833 | 0,10 (1) | Os grupos são iguais ** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 10 | 10 | 0,083 | 0,50 (5) | Os grupos são iguais ** |
| $n_1 = 4$ e $n_2 = 6$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 210 | 210 | 0,165 | 0,162 (34) | **Os grupos são iguais**** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 210 | 210 | 0,536 | 0,005 (1) | **Os grupos não são iguais** |
| $n_1 = 8$ e $n_2 = 12$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 125.970 | 999 | 0,119 | 0,087 (86) | **Os grupos são iguais**** |
| | | | 4999 | 0,119 | 0,078 (392) | **Os grupos são iguais**** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 125.970 | 999 | 0,623 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,623 | 0,0002 (0) | **Os grupos não são iguais** |
| $n_1 = 16$ e $n_2 = 24$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 62.852.101.650 | 999 | 0,258 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,258 | 0,001 (2) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 62.852.101.650 | 999 | 0,561 | 0,001 (0) | **Os grupos não são iguais** |
| | | | 4999 | 0,561 | 0,0002 (0) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Neste exercício pode-se observar que em várias situações a decisão estatística a um nível de 5% de significância, está incorreta. Salienta-se o caso onde o grupo com $n_1 = 4$ ($\lambda_1 = 8$) e $n_2 = 6$ ($\lambda_2 = 5$). Num primeiro momento pode-se dizer que as amostras são pequenas então talvez existam respostas contrárias a esperada, porém foram geradas todas as permutações possíveis, portanto o teste deveria ter revelado que os grupos eram significativamente distintos, ou pelo menos ter apresentado um p-valor inferior ao encontrado.

Um outro caso que merece atenção é para o qual se tem $n_1 = 8$ ($\lambda_1 = 8$) e $n_2 = 12$ ($\lambda_2 = 5$), resolveu-se investigar mais detalhadamente calculando os valores da estatística K-S (Tabela6) e realizar as 20 repetições para 999 e 4.999 permutações (figuras 10 e 11, respectivamente). Também neste caso calculou-se o p-valor para 49.999 permutações se obtendo 0,085.

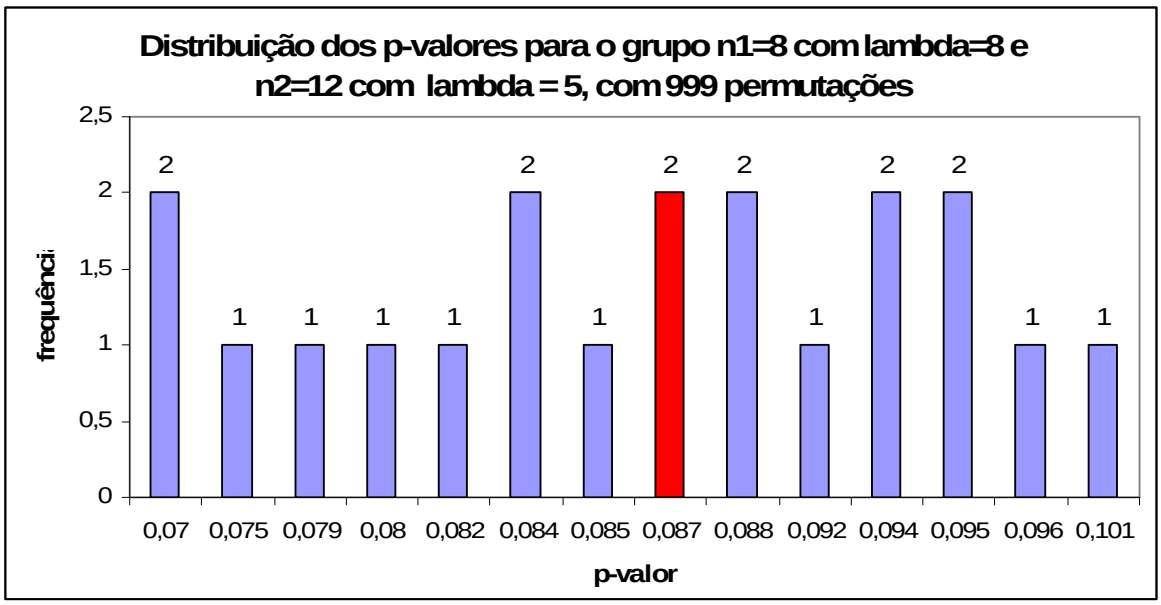


*Figura 10: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8(λ=8) x12(λ=5) réplicas com 999 permutações, com 20 repetições..*

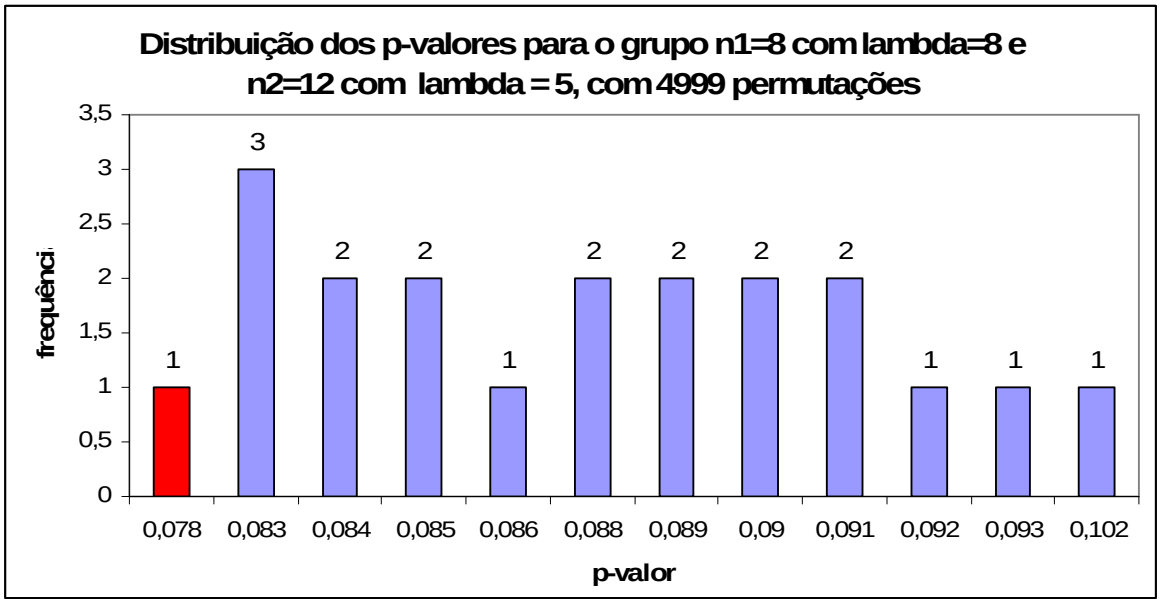


*Figura 11: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8(λ=8) x12(λ=5) réplicas com 4999 permutações, com 20 repetições..*

O teste K-S foi realizado para que não houvesse dúvida alguma de que esta amostra que foi retirada de uma distribuição Poisson(λ) continuava sendo uma distribuição Poisson com o mesmo lambda da população da qual foi retirada. Os valores para a estatística K-S se apresentam na Tabela 6 a seguir:

*Tabela 6: Valores da estatística K-S para cada espécie nos grupos $n_1$= 4 e $n_2$= 6 e também $n_1$= 8 e $n_2$= 12.*

| | K-S grupo $n_1$=4 (λ =8) | K-S grupo $n_2$=6 (λ=5) | K-S grupo $n_1$=8 (λ =8) | K-S grupo $n_2$=12 (λ=5) |
|---|---|---|---|---|
| E1 | 0,1205 | 0,1947 | 0,08919 | 0,3293 |
| E2 | 0,3283 | 0,7592 | 0,3451 | 0,7234 |
| E3 | 0,827 | 0,2198 | 0,3558 | 0,7234 |

*FONTE: Valores obtidos através do software SPSS 20.0*

É bom salientar que nesta série de exercício está-se muito próximo à realidade, pois realizou-se o experimento através de sorteio de populações com 3 espécies onde cada elemento sorteado representa uma réplica que contém 3 espécies.

### 5.3.4 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 5$

Quadro comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância.

*Quadro 20: Resultados obtidos para a série de exercícios dos grupos onde $n_1$ = 2 e $n_2$ = 5*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 5$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 21 | 21 | 0,064 | 0,381 (8) | Os grupos são iguais ** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 21 | 21 | 0,173 | 0,19 (4) | Os grupos são iguais** |
| $n_1 = 4$ e $n_2 = 10$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 1.001 | 999 | 0,204 | 0,085 (84) | **Os grupos são iguais**** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 1.001 | 999 | 0,224 | 0,121 (120) | **Os grupos são iguais**** |
| $n_1 = 8$ e $n_2 = 20$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 3.108.105 | 750.000 | 0,375 | 0,001 (0) | **Os grupos não são iguais** |
| | | | | 0,375 | 0,001 (2) | **Os grupos não são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 3.108.105 | 750.000 | 0,404 | 0,002 (1) | **Os grupos não são iguais** |
| | | | | 0,404 | 0,002 (10) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Para esta série de exercícios houve problemas com relação à decisão estatística, pois nos primeiros casos não se conseguia rejeitar a hipótese nula de igualdade dos grupos quando na realidade eles eram distintos, chegando a obter um p-valor de 0,121. Cabe salientar que foram realizadas praticamente todas as permutações possíveis. Para retirar qualquer dúvida com relação aos parâmetros e ao tipo de distribuição realizou-se o teste K-S e se obteve os resultados na Tabela 7 a seguir.

*Tabela 7: Valores da estatística K-S para cada espécie nos grupos $n_1$= 4 e $n_2$= 10.*

| | K-S grupo $n_1$=4 (λ =5) | K-S grupo $n_2$=10 (λ=8) | K-S grupo $n_1$=4 (λ =8) | K-S grupo $n_2$=10 (λ=5) |
|---|---|---|---|---|
| E1 | 0,5184 | 0,3981 | 0,2854 | 0,2709 |
| E2 | 0,8773 | 0,575 | 0,7799 | 0,9182 |
| E3 | 0,2448 | 0,9821 | 0,827 | 0,1966 |

*FONTE: Valores obtidos através do software SPSS 20.0*

Cabe salientar novamente que nesta série de exercícios obteve-se as espécies através de sorteios, desta forma esta simulação é bem semelhante à realidade e, mesmo assim, o teste não rejeitou a hipótese nula de igualdade dos grupos quando eles eram comprovadamente distintos.

### 5.3.5 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 8$

O Quadro 21 a seguir apresenta a comparação dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a um nível de 5% de significância.

*Quadro 21: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 8$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 8$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 45 | 45 | 0,08 | 0,267 (12) | Os grupos são iguais** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 45 | 45 | 0,797 | 0,002 (1) | **Os grupos não são iguais** |
| $n_1 = 4$ e $n_2 = 16$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 4.845 | 999 | 0,084 | 0,234 (233) | **Os grupos são iguais**** |
| | | | 4.845 | 0,084 | 0,246 (1194) | **Os grupos são iguais**** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 4.845 | 999 | 0,166 | 0,18 (179) | **Os grupos são iguais**** |
| | | | 4.845 | 0,166 | 0,179 (869) | **Os grupos são iguais**** |
| $n_1 = 8$ e $n_2 = 32$ | $\lambda_1 = 8$ e $\lambda_2 = 5$ | 76.904.685 | 999 | 0,024 | 0,353 (352) | **Os grupos são iguais**** |
| | | | 4999 | 0,024 | 0,379 (1894) | **Os grupos são iguais**** |
| | $\lambda_1 = 5$ e $\lambda_2 = 8$ | 76.904.685 | 999 | 0,354 | 0,004 (3) | **Os grupos não são iguais** |
| | | | 4999 | 0,354 | 0,003 (15) | **Os grupos não são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Neste caso, o comportamento distinto do esperado encontrado foi mais explícito, pois embora a amostra seja maior, pois se tem um total de 40 réplicas onde um grupo apresenta 25% das réplicas do outro não se conseguiu rejeitar a hipótese nula de que os grupos eram

iguais. Para o caso das 49.999 permutações no grupo com $n_1 = 8$ ($\lambda_1 = 8$) e $n_2 = 32$ ($\lambda_2 = 5$) obteve-se um p-valor de 0,372, que também não foi capaz de rejeitar a Hipótese Nula.

A seguir nas tabelas 8 e 9 apresenta-se os valores da estatística K-S para as distribuições das três espécies em cada grupo. Cabe relembrar que neste exercício realizou-se o sorteio das réplicas, indicando que a simulação é mais próxima da realidade.

*Tabela 8: Valores da estatística K-S para cada espécie nos grupos $n_1 = 4$ e $n_2 = 16$.*

| | K-S grupo $n_1$=4 ($\lambda$ =5) | K-S grupo $n_2$=16 ($\lambda$=8) | K-S grupo $n_1$=4 ($\lambda$ =8) | K-S grupo $n_2$=16 ($\lambda$=5) |
|---|---|---|---|---|
| E1 | 0,05491 | 0,06401 | 0,5834 | 0,6071 |
| E2 | 0,09548 | 0,3423 | 0,4307 | 0,3108 |
| E3 | 0,8382 | 0,9536 | 0,7799 | 0,2214 |

*FONTE: Valores obtidos através do software SPSS 20.0*

*Tabela 9: Valores da estatística K-S para cada espécie nos grupos $n_1 = 8$ e $n_2 = 32$.*

| | K-S grupo $n_1$=8 ($\lambda$ =5) | K-S grupo $n_2$=32 ($\lambda$=8) | K-S grupo $n_1$=8 ($\lambda$ =8) | K-S grupo $n_2$=32 ($\lambda$=5) |
|---|---|---|---|---|
| E1 | 0,9988 | 0,0875 | 0,8967 | 0,1959 |
| E2 | 0,0975 | 0,2046 | 0,4118 | 0,09051 |
| E3 | 0,09226 | 0,9303 | 0,4118 | 0,1005 |

*FONTE: Valores obtidos através do software SPSS 20.0*

As figuras 12 e 13 apresentam as 20 repetições com 999 e 4.999 permutações, respectivamente com o objetivo de comparar os p-valores obtidos com o p-valor obtido na primeira permutação.

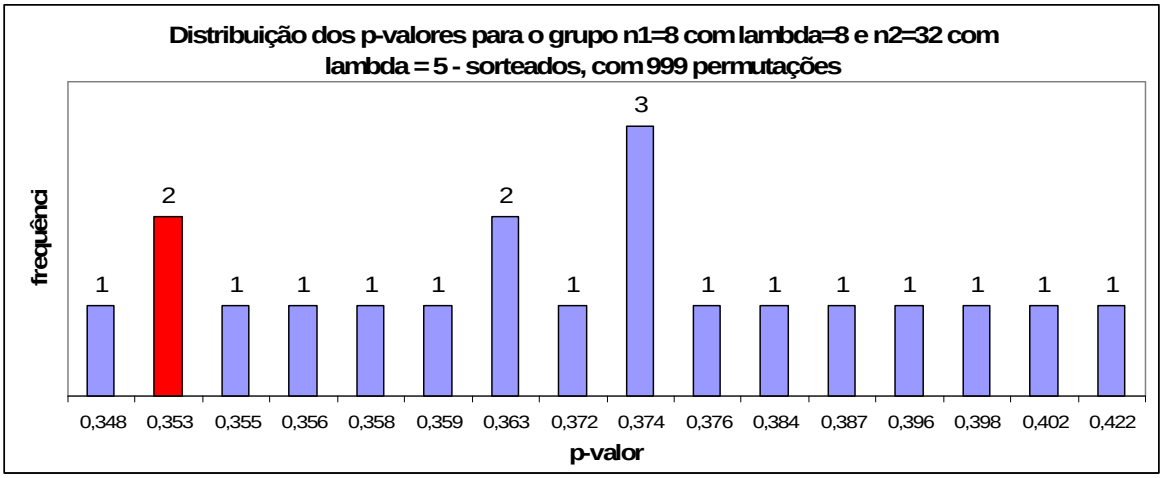


*Figura 12: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8($\lambda$=8) x32($\lambda$=5) réplicas com 999 permutações, com 20 repetições..*

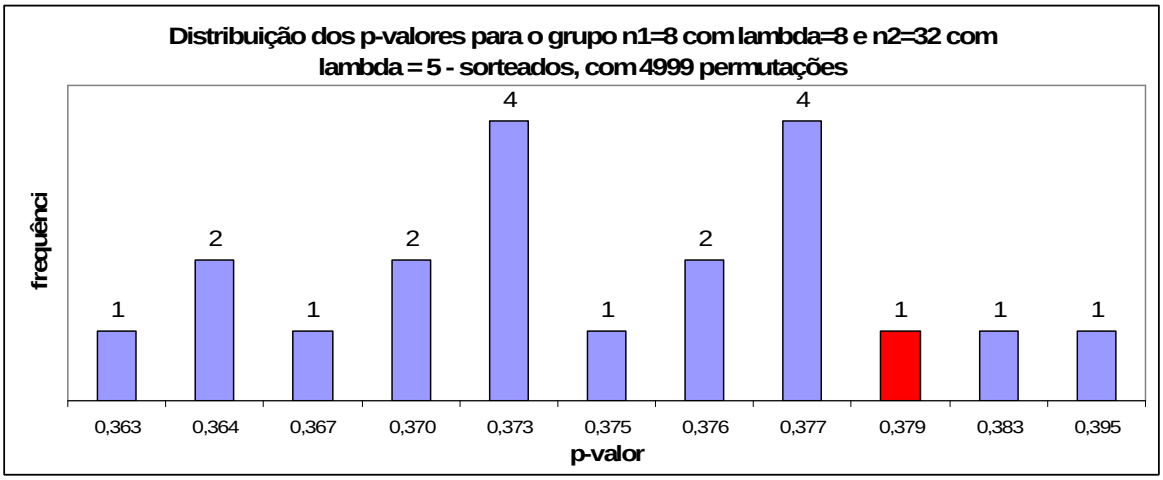


*Figura 13: Gráfico de barras com a distribuição do p-valor*
*Fonte: Dados obtidos através das simulações das três espécies em grupos de 8(λ=8) x32(λ=5) réplicas com 4999 permutações, com 20 repetições..*

## 5.4 EXERCÍCIOS COM GRUPOS DESBALANCEADOS E VALORES PARA Λ = 8 EM AMBOS OS GRUPOS OBTIDOS A PARTIR DE SORTEIOS

Novamente se realizou os exercícios para os grupos desbalanceados porém com o mesmo lambda sendo que estas réplicas foram desta vez retiradas, mediante sorteio, de populações geradas com 100 réplicas de distribuições Poisson(λ=8). O objetivo aqui é verificar se realmente o teste indica desde o início que os grupos são iguais.

### 5.4.1 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 4$

Quadro 22 mostra o comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a um nível de 5% de significância.

*Quadro 22: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 4$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 4$ | $\lambda_1 = \lambda_2 = 8$ | 15 | 15 | 0,00 | 0,467 (7) | **Os grupos são iguais** |
| $n_1 = 4$ e $n_2 = 8$ | $\lambda_1 = \lambda_2 = 8$ | 495 | 495 | -0,143 | 0,808 (400) | **Os grupos são iguais** |
| $n_1 = 8$ e $n_2 = 16$ | $\lambda_1 = \lambda_2 = 8$ | 735.471 | 999 | 0,096 | 0,166 (165) | **Os grupos são iguais** |
| | | | 4999 | 0,096 | 0,14 (700) | **Os grupos são iguais** |

*FONTE: Elaborado pela autora*

Para este exercício não houve problemas em indicar que os grupos eram iguais quando realmente deveriam ser. Nem mesmo para o caso em que a amostra é pequena e onde normalmente existem os comportamentos distintos do esperado.

### 5.4.2 Série de Exercícios com grupos de $n_1 = 3$ e $n_2 = 5$

Quadro comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância.

*Quadro 23: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 3$ e $n_2 = 5$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 3$ e $n_2 = 5$ | $\lambda_1 = \lambda_2 = 8$ | 56 | 56 | -0,123 | 0,714 (40) | **Os grupos são iguais** |
| $n_1 = 6$ e $n_2 = 10$ | $\lambda_1 = \lambda_2 = 8$ | 8.008 | 999 | -0,11 | 0,842 (841) | **Os grupos são iguais** |
| | | | 4999 | -0,11 | 0,842 (4207) | **Os grupos são iguais** |
| $n_1 = 12$ e $n_2 = 20$ | $\lambda_1 = \lambda_2 = 8$ | 225.792.840 | 999 | -0,023 | 0,581 (580) | **Os grupos são iguais** |
| | | | 4999 | -0,023 | 0,581 (2906) | **Os grupos são iguais** |

*FONTE: Elaborado pela autora*

Neste caso também não foram encontrados problemas mesmo para as pequenas amostras. O teste indicou que os grupos eram iguais quando de fato eram.

### 5.4.3 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 3$

No Quadro 24 a seguir mostra-se o comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância.

*Quadro 24: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 3$*

| Tamanho dos Grupos | Valores de λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 3$ | $\lambda_1 = \lambda_2 = 8$ | 10 | 10 | -0,292 | 0,90 (9) | **Os grupos são iguais** |
| $n_1 = 4$ e $n_2 = 6$ | $\lambda_1 = \lambda_2 = 8$ | 210 | 210 | -0,067 | 0,619 (130) | **Os grupos são iguais** |
| $n_1 = 8$ e $n_2 = 12$ | $\lambda_1 = \lambda_2 = 8$ | 125.970 | 999 | 0,071 | 0,175 (174) | **Os grupos são iguais** |
| | | | 4999 | 0,071 | 0,173 (862) | **Os grupos são iguais** |
| $n_1 = 16$ e $n_2 = 24$ | $\lambda_1 = \lambda_2 = 8$ | 62.852.101.650 | 999 | -0,043 | 0,819 (817) | **Os grupos são iguais** |
| | | | 4999 | -0,043 | 0,822 (4107) | **Os grupos são iguais** |

*FONTE: Elaborado pela autora*

Neste exercício onde um grupo apresenta 66,67% das réplicas do outro, também não foram detectados problemas com relação a decisão estatística de não rejeitar a hipótese de igualdade entre os grupos, pois todos foram sorteados de populações com a mesma distribuição Poisson(λ).

### 5.4.4 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 5$

Quadro comparativo dos exercícios realizados para esta série com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância.

*Quadro 25: Resultados obtidos para a série de exercícios dos grupos onde $n_1 = 2$ e $n_2 = 5$*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 5$ | $\lambda_1 = \lambda_2 = 8$ | 21 | 21 | 0,782 | 0,048 (1) | Os grupos não são iguais** |
| $n_1 = 4$ e $n_2 = 10$ | $\lambda_1 = \lambda_2 = 8$ | 1.001 | 999 | 0,354 | 0,03 (29) | **Os grupos não são iguais**** |
| $n_1 = 8$ e $n_2 = 20$ | $\lambda_1 = \lambda_2 = 8$ | 3.108.105 | 999 | 0,011 | 0,413 (412) | **Os grupos são iguais** |
| | | | 4999 | 0,011 | 0,429 (2142) | **Os grupos são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Neste exercício foram observados problemas na decisão, pois embora as distribuições sejam Poisson(λ=8), conforme informado  na Tabela 10, observa-se que mesmo para uma amostra um pouco maior o teste leva a decidir de forma distinta do esperado. Tem-se aqui um caso em que quando um grupo apresenta 40% das réplicas do

outro pode-se ter problemas inclusive para detectar que os grupos são iguais quando eles realmente o são.

*Tabela 10: Valores da estatística K-S para cada espécie nos grupos $n_1$= 4 e $n_2$= 10.*

| | K-S grupo $n_1$=4 (λ =8) | K-S grupo $n_2$=10 (λ=8) |
|---|---|---|
| E1 | 0,8403 | 0,3981 |
| E2 | 0,3485 | 0,2798 |
| E3 | 0,9729 | 0,384 |

*FONTE: Valores obtidos através do software SPSS 20.0*

### 5.4.5 Série de Exercícios com grupos de $n_1 = 2$ e $n_2 = 8$

Quadro comparativo dos exercícios realizados para esta série, com o número de permutações possíveis e realizadas, os valores da estatística R Global, seu respectivo p-valor e a decisão estatística a  um nível de 5% de significância..

*Quadro 26: Resultados obtidos para a série de exercícios dos grupos onde $n_1$ = 2 e $n_2$ = 8*

| Tamanho dos Grupos | Valores para λ | Nº permutações possíveis | Nº permutações realizadas | R Global | p-valor | Decisão |
|---|---|---|---|---|---|---|
| $n_1 = 2$ e $n_2 = 8$ | $\lambda_1 = \lambda_2 = 8$ | 45 | 45 | 0,703 | 0,022 (1) | **Os grupos não são iguais**** |
| $n_1 = 4$ e $n_2 = 16$ | $\lambda_1 = \lambda_2 = 8$ | 4.845 | 999 | -0,098 | 0,724 (723) | **Os grupos são iguais** |
| | | | 4.845 | -0,098 | 0,727 (3.520) | **Os grupos são iguais** |
| $n_1 = 8$ e $n_2 = 32$ | $\lambda_1 = \lambda_2 = 8$ | 76.904.685 | 999 | 0,131 | 0,08 (79) | **Os grupos são iguais** |
| | | | 4999 | 0,131 | 0,093 (466) | **Os grupos são iguais** |

*FONTE: Elaborado pela autora*
*** Indicando que a decisão estatística a 5% está incorreta.*

Para esta série de exercícios as decisões foram viesadas apenas para grupos com poucas réplicas. Observa-se, através da Tabela 11, os valores da estatística K-S para as espécies em cada grupo e pode-se inferir que um dos motivos deste comportamento é, além das amostras pequenas, que os p-valores encontrados foram muito pequenos, quase na área de rejeição da hipótese nula de que as espécies apresentavam distribuição Poisson(λ =8).

*Tabela 11: Valores da estatística K-S para cada espécie nos grupos $n_1$= 2 e $n_2$= 8.*

| | K-S grupo $n_1$=2 (λ =8) | K-S grupo $n_2$=8 (λ=8) |
|---|---|---|
| E1 | 0,8017 | 0,07504 |
| E2 | 0,8017 | 0,06138 |
| E3 | 0,05015 | 0,8434 |

*FONTE: Valores obtidos através do software SPSS 20.0*

**Limitações do Estudo**

Neste trabalho foram construídos dois grupos com três espécies cada um, embora os trabalhos na área ecológica envolvam matrizes com um número muito maior de espécies sendo estas pertencentes a um número maior de grupos. Esta abordagem foi escolhida para facilitar a visualização dos valores obtidos nas matrizes de similaridades e para desenvolvimento algébrico em alguns casos de interesse.

Também cabe salientar que os valores $\lambda = 5$ e $\lambda = 8$ foram escolhidos de modo que a variabilidade dos dados estivesse dentro de valores razoáveis, visto que se trata de distribuições Poisson. Sendo que estes valores se encontram fixos por grupo e não por espécie, isto é, por exemplo, o grupo 1 é formado por três espécies, sendo todas com distribuição Poisson($\lambda = 5$). Em estudos de campo, as espécies apresentem médias distintas dentro de cada grupo. Desta forma se entendeu que o controle seria maior sobre a enorme variabilidade da geração dos dados e valores para médias.

# 6. CONSIDERAÇÕES FINAIS

Assim como no trabalho desenvolvido por Bündchen (2010) se constatou que a estatística de teste R nem sempre é capaz de detectar as diferenças ou semelhanças entre os grupos. Desta forma nem sempre se mostra robusta por estar em condições de desbalanceamento.

A partir deste estudo também se pode derivar algumas expressões que não foram encontradas na literatura e que são importantes para o entendimento da estatística de teste R utilizada na técnica ANOSIM. Estas quantidades são as expressões que indicam as quantidades de permutações possíveis sob condições de desbalanceamento; a quantidade ganha/perdida a cada permutação para as médias entre e dentro dos grupos; e o número de *rankings*, entre e dentro dos grupos (de acordo com 5.1, 5.2, 5.3 e 5.4). Estas quantidades podem indicar que os *p*-valores da estatística de teste dependem da proporção do desbalanceamento.

Observou-se que ao realizar uma quantidade maior de permutações o p-valor tende a estabilizar com a diminuição da variabilidade. Segundo alguns autores (como Jackson e Somers, 1989), estatísticas não paramétricas que utilizam matrizes de similaridades e permutações, que produzam valores para o nível descritivo amostral *p* próximos de 0,05 ou 0,01, sugerem o aumento no número total de permutações, contrário às recomendações de muitos pesquisadores da área da ecológica que sugerem apenas 999 (para $\alpha$=0,05) ou 4.999 (para $\alpha$=0,01). Em alguns trabalhos é indicado até mesmo realizar 49.999 permutações ou mais.

Através das séries de exercícios realizados pode-se ter a dimensão dos desbalanceamentos que podem causar desvios na decisão estatística. Os casos observados apresentavam dois grupos onde o tamanho de um determinado grupo representava 50%, 60%, 66,67%, 40% e 25% do tamanho do outro grupo. Observou-se que o único caso para o qual não se encontrou problemas na interpretação dos resultados foi para o de 60% de proporção entre os tamanhos dos grupos.

Além disso, foi possível observar, através das simulações, resultados distintos do esperado, isto é, o teste não detectou a diferença quando ela de fato existia. Os desvios não foram exclusivamente nesta direção, mas de igual forma quando se tem efetivamente grupos iguais na população tendo o teste apontado que há diferença. Desta forma o desbalanceamento pode levar a decisões incorretas em qualquer sentido.

# 7. REFERÊNCIAS BIBLIOGRÁFICAS


Barnett,V.,Lewi, T. (1994) Outliers in Statistical Data, 3º Ed.. Willey. Pondicherry.

Bündchen, C. Avaliação da Distribuição da Estatística R e Nível Descritivo Amostral na Análise de Similaridade – ANOSIM: Um estudo de Caso do Projeto MAPEM. Monografia, 2010.

Clarke, K.R., Green, R.H. Statistical design and analysis for a "biological effects" study. Mar Ecol Prog Ser 46: 213-226, 1988.

_, K.R., Warwick, R.M. Similarity-based testing for community pattern: the two-way layout with no replication. Plymouth Marine Laboratory, 1993.

_, K.R., _, R.M. Change in marine communities: an approach to statistical analysis and interpretation, 2ºed. PRIMER-E: Plymouth, 2001.

Hair, J.F., Anderson, R.E., Tatham, R.L., Black, W.C. Multivariate Data Analysis, 5ª Edição, Prentice Hall, Upper Saddle River, 1998.

Jackson, D.A., Somers K.M. Are probability estimates from the permutation model of Mantel's test stable? Canadian Journal of Zoology, 67: 766-769,1989.

Johnson, R.A., Wichern, D.W. Applied Multivariate Statistical Analysis, 6ª ed. Pearson, London, 2007.

Jones, David. L., *homepage* disponível em <http://www.marine.usf.edu/user/djones/anosim/anosim.html>. Acessado em 21/07/2013.

Larson, H. J.,Introduction to Probability Theory and Statistical Inference, 2ª ed. John Wiley e sons, California, 1974.

Manly,B. F. J. Randomization, Bootstrap and Monte Carlo Methods in Biology, 2º edition, Chapman e Hall, London, 1997.

Marti, J.A. A new method for non-parametric multivariate analysis of variance. Aus Eco 26: 32-49, 2001.

Package 'Biodiversity R', 2012. Disponível em : <http://cran.r-project.org/web/packages/BiodiversityR/BiodiversityR.pdf>. Acessado em 16/07/2012.

Pagina da Universidade de St Andrews, Escola de Matemática e Estatística, página de Biografias: <http://www-history.mcs.st-and.ac.uk/Biographies/Bartlett.html>. Acessado em 16/04/2013.

_, <http://www-history.mcs.st-and.ac.uk/Biographies/Wilks.html>. Acessado em 16/04/2013.

_, <http://www-history.mcs.st-and.ac.uk/Biographies/Pillai.html>. Acessado em 16/04/2013.

_, <http://www-history.mcs.st-and.ac.uk/Biographies/Hotelling.html>. Acessado em 16/04/2013.

_, <http://www-history.mcs.st-and.ac.uk/Mathematicians/Roy.html>. Acessado em 16/04/2013.

_, <http://www-history.mcs.st-and.ac.uk/Biographies/Mahalanobis.html>. Acessado em 16/04/2013.

_, <http://www-history.mcs.st-and.ac.uk/Biographies/Minkowski.html>. Acessado em 16/04/2013.

Toldo Jr. E.E.; R.N. Zouain R.N.A. Estrutura, Organização e Premissas do Projeto MAPEM (Monitoramento Ambiental em Atividades de Perfuração Exploratória Marítima). Disponível em: <http://www.ufrgs.br/ceco/mapem/pdf/capitulo%201.pdf> Acessado em 15/08/2011.

Valadares, F.G., Aquino, A.L.L. e Pereira Jr., A.R. Detecção de outliers multivariados em redes de sensores. CLAIO. SBPO: 1350-1351, Setembro 2012. RJ. <http://www.din.uem.br/sbpo/sbpo2012/pdf/arq0272.pdf> Acessado em 17/07/2013.

## SOFTWARES UTILIZADOS

R versão 2.13.2,

PRIMER versão 5.1.2,

Planilha Excel do Microsoft Office,

SPSS versão 18.0 e versão 20.0.

# Anexos

# Anexo 1

## Matriz de similaridades do caso n1=8 com lambda =8 e n2 = 20 com lambda = 5

| | S11 | S12 | S13 | S14 | S15 | S16 | S17 | S18 | S21 | S22 | S23 | S24 | S25 | S26 | S27 | S28 | S29 | S210 | S211 | S212 | S213 | S214 | S215 | S216 | S217 | S218 | S219 | S220 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S11 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S12 | 82,05 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S13 | 90,91 | 88,37 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| S14 | 86,96 | 75,56 | 88,00 | | | | | | | | | | | | | | | | | | | | | | | | | |
| S15 | 66,67 | 81,82 | 77,55 | 66,67 | | | | | | | | | | | | | | | | | | | | | | | | |
| S16 | 81,82 | 88,37 | 91,67 | 80,00 | 85,71 | | | | | | | | | | | | | | | | | | | | | | | |
| S17 | 79,07 | 90,48 | 89,36 | 77,55 | 87,50 | 93,62 | | | | | | | | | | | | | | | | | | | | | | |
| S18 | 80,95 | 92,68 | 86,96 | 75,00 | 85,11 | 91,30 | 88,89 | | | | | | | | | | | | | | | | | | | | | |
| S21 | 64,52 | 73,33 | 62,86 | 59,46 | 61,11 | 62,86 | 64,71 | 66,67 | | | | | | | | | | | | | | | | | | | | |
| S22 | 82,35 | 78,79 | 73,68 | 70,00 | 61,54 | 73,68 | 75,68 | 77,78 | 64,00 | | | | | | | | | | | | | | | | | | | |
| S23 | 66,67 | 68,97 | 58,82 | 55,56 | 57,14 | 58,82 | 60,61 | 62,50 | 66,67 | 66,67 | | | | | | | | | | | | | | | | | | |
| S24 | 78,05 | 90,00 | 84,44 | 76,60 | 86,96 | 88,89 | 95,45 | 83,72 | 68,75 | 74,29 | 64,52 | | | | | | | | | | | | | | | | | |
| S25 | 84,21 | 86,49 | 80,95 | 77,27 | 79,07 | 85,71 | 87,80 | 80,00 | 68,97 | 81,25 | 71,43 | 92,31 | | | | | | | | | | | | | | | | |
| S26 | 88,37 | 80,95 | 89,36 | 85,71 | 79,17 | 89,36 | 91,30 | 80,00 | 64,71 | 75,68 | 60,61 | 90,91 | 87,80 | | | | | | | | | | | | | | | |
| S27 | 88,89 | 86,36 | 97,96 | 90,20 | 76,00 | 89,80 | 87,50 | 85,11 | 61,11 | 71,79 | 57,14 | 82,61 | 79,07 | 87,50 | | | | | | | | | | | | | | |
| S28 | 70,27 | 88,89 | 78,05 | 65,12 | 76,19 | 82,93 | 80,00 | 87,18 | 64,29 | 77,42 | 51,85 | 78,95 | 74,29 | 70,00 | 76,19 | | | | | | | | | | | | | |
| S29 | 70,97 | 73,33 | 62,86 | 59,46 | 61,11 | 62,86 | 64,71 | 66,67 | 63,64 | 88,00 | 76,19 | 68,75 | 75,86 | 64,71 | 61,11 | 71,43 | | | | | | | | | | | | |
| S210 | 72,22 | 74,29 | 70,00 | 66,67 | 78,05 | 75,00 | 82,05 | 68,42 | 59,26 | 73,33 | 76,92 | 86,49 | 88,24 | 82,05 | 68,29 | 60,61 | 81,48 | | | | | | | | | | | |
| S211 | 78,79 | 81,25 | 70,27 | 66,67 | 68,42 | 70,27 | 72,22 | 74,29 | 75,00 | 88,89 | 78,26 | 76,47 | 83,87 | 72,22 | 68,42 | 73,33 | 91,67 | 82,76 | | | | | | | | | | |
| S212 | 51,85 | 53,85 | 45,16 | 42,42 | 43,75 | 45,16 | 46,67 | 48,28 | 55,56 | 66,67 | 82,35 | 50,00 | 56,00 | 46,67 | 43,75 | 58,33 | 77,78 | 60,87 | 70,00 | | | | | | | | | |
| S213 | 74,29 | 88,24 | 76,92 | 68,29 | 75,00 | 76,92 | 78,95 | 81,08 | 84,62 | 68,97 | 80,00 | 83,33 | 78,79 | 73,68 | 75,00 | 75,00 | 69,23 | 70,97 | 78,57 | 63,64 | | | | | | | | |
| S214 | 84,21 | 86,49 | 85,71 | 77,27 | 69,77 | 85,71 | 82,93 | 85,00 | 68,97 | 87,50 | 57,14 | 82,05 | 77,78 | 82,93 | 83,72 | 85,71 | 75,86 | 64,71 | 77,42 | 56,00 | 78,79 | | | | | | | |
| S215 | 40,00 | 41,67 | 34,48 | 32,26 | 33,33 | 34,48 | 35,71 | 37,04 | 37,50 | 52,63 | 66,67 | 38,46 | 43,48 | 35,71 | 33,33 | 45,45 | 62,50 | 47,62 | 55,56 | 83,33 | 50,00 | 43,48 | | | | | | |
| S216 | 82,05 | 73,68 | 79,07 | 75,56 | 72,73 | 83,72 | 85,71 | 73,17 | 53,33 | 78,79 | 68,97 | 85,00 | 86,49 | 90,48 | 77,27 | 61,11 | 73,33 | 91,43 | 75,00 | 53,85 | 64,71 | 75,68 | 41,67 | | | | | |
| S217 | 75,00 | 70,97 | 66,67 | 63,16 | 54,05 | 66,67 | 68,57 | 70,59 | 52,17 | 92,31 | 72,73 | 66,67 | 73,33 | 68,57 | 64,86 | 68,97 | 86,96 | 71,43 | 80,00 | 73,68 | 59,26 | 80,00 | 58,82 | 77,42 | | | | |
| S218 | 70,97 | 73,33 | 62,86 | 59,46 | 55,56 | 62,86 | 64,71 | 66,67 | 54,55 | 88,00 | 66,67 | 68,75 | 75,86 | 64,71 | 61,11 | 78,57 | 90,91 | 74,07 | 83,33 | 77,78 | 61,54 | 75,86 | 62,50 | 73,33 | 86,96 | | | |
| S219 | 66,67 | 68,97 | 58,82 | 55,56 | 57,14 | 58,82 | 60,61 | 62,50 | 85,71 | 66,67 | 80,00 | 64,52 | 71,43 | 60,61 | 57,14 | 51,85 | 66,67 | 69,23 | 78,26 | 58,82 | 80,00 | 57,14 | 40,00 | 62,07 | 54,55 | 57,14 | | |
| S220 | 82,35 | 84,85 | 73,68 | 70,00 | 66,67 | 73,68 | 75,68 | 77,78 | 72,00 | 92,86 | 75,00 | 80,00 | 87,50 | 75,68 | 71,79 | 77,42 | 88,00 | 80,00 | 96,30 | 66,67 | 75,86 | 81,25 | 52,63 | 78,79 | 84,62 | 88,00 | 75,00 | |