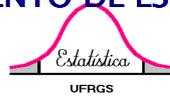




UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Avaliação da distribuição da estatística R e nível descritivo amostral na Análise de Similaridade – ANOSIM: um estudo de caso do Projeto MAPEM

Autora: Cristiane Bündchen

Orientador: Professor Dr. Fernando Hepp Pulgati

Porto Alegre, 08 de Dezembro de 2010.

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

Avaliação da distribuição da estatística R e
nível descritivo amostral na Análise de
Similaridade – ANOSIM: um estudo de caso
do Projeto MAPEM

Autora: Cristiane Bündchen

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:

Professor Dr. Fernando Hepp Pulgati (orientador)

Professora Dra. Jandyra Maria Guimarães Fachel

Porto Alegre, 08 de Dezembro de 2010.

“A mente que se abre a uma nova idéia jamais voltará ao seu tamanho original”

Albert Einstein

ÍNDICE GERAL

AGRADECIMENTOS	4
ÍNDICE DE FIGURAS	5
ÍNDICE DE TABELAS	6
RESUMO	8
ABSTRACT	9
<i>CAPÍTULO 1. INTRODUÇÃO</i>	10
<i>CAPÍTULO 2. OBJETIVO</i>	11
2.1.OBJETIVO PRINCIPAL	11
2.2. OBJETIVOS SECUNDÁRIOS	11
<i>CAPÍTULO 3. REVISÃO BIBLIOGRÁFICA</i>	12
<i>CAPÍTULO 4. MATERIAIS E MÉTODO</i>	14
4.1 COEFICIENTE DE SIMILARIDADE (S)	14
4.1.1 Matriz de similaridades	14
4.1.1.1 Transformação dos dados	15
4.1.2 Coeficiente de Bray-Curtis	15
4.1.3 Distância Euclidiana	16
4.2 ESTATÍSTICA R	17
4.3 ESTATÍSTICA R PERMUTADA	19
4.4 CÁLCULO DO P-VALOR	20
4.5 ANÁLISE DAS COMPARAÇÕES MÚLTIPLAS	21
4.6 PROJETO MAPEM	21
4.6 SOFTWARE	23
<i>CAPÍTULO 5. ESTUDO DE CASO – ANÁLISE DE DADOS E RESULTADOS</i>	27
5.1 DELINEAMENTO COM DOIS FATORES CRUZADOS	28
5.2 DELINEAMENTO UNIFATORIAL	32
<i>CAPÍTULO 6. CONCLUSÃO</i>	45
<i>CAPÍTULO 7. REFERÊNCIAS BIBLIOGRÁFICAS</i>	47

Agradecimentos

Agradeço a Claudino e Ivoni Bündchen, meus pais, que fizeram tudo o que estava ao alcance deles para me proporcionar a melhor formação possível.

A Nelton, Mara Rosanne e Gustavo Pinheiro Machado pelo acolhimento, apoio e carinho que recebi desde o primeiro dia que cheguei a Porto Alegre.

Aos meus amigos que estão longe e que me ensinaram o significado de uma amizade verdadeira e aos que tenho por perto e posso desfrutar de agradáveis momentos juntos.

Aos meus colegas pelo companheirismo.

E, ao meu orientador professor doutor Fernando Hepp Pulgati por ter aceitado me orientar neste trabalho.

ÍNDICE DE FIGURAS

Figura 1: Exemplo de distância Euclidiana entre as observações j e k _____	17
Figura 2: Diagramas ilustrando as similaridades entre e dentro dos grupos a e b _____	18
Figura 3: Exemplo de uma distribuição de R^* sob a hipótese nula _____	19
Figura 4: Exemplo de uma distribuição de R^* sob a hipótese nula _____	20
Figura 5: Delineamento amostral do projeto MAPEM _____	23
Figura 6: Banco de dados no software Primer _____	24
Figura 7: Escolha da matriz de similaridades _____	24
Figura 8: Escolha da técnica multivariada realizada com base na matriz de similaridades obtida _____	25
Figura 9: Definição do delineamento experimental, fatores e número de permutações _____	25
Figura 10: Parte da saída obtida para uma análise com dois fatores _____	26
Figura 11: Fluxograma do processo realizado no presente estudo _____	27
Figura 12: Histogramas do nível descritivo amostral referente ao efeito principal do fator Fluído _____	29
Figura 13: Histogramas do nível descritivo amostral referente ao efeito principal do fator Tempo _____	31
Figura 14: Histogramas do nível descritivo amostral do fator Fluído dentro do Tempo 3 _____	37
Figura 15: Histograma do nível descritivo amostral das comparações múltiplas entre os tempos 1 e 2 _____	44

ÍNDICE DE TABELAS

Tabela 1: Exemplo de uma matriz de similaridades _____	15
Tabela 2: Análise do nível descritivo amostral referente ao efeito principal do fator Fluído_	28
Tabela 3: Análise do nível descritivo amostral referente às comparações múltiplas dentro do fator Fluído _____	30
Tabela 4: Análise do nível descritivo amostral referente ao efeito principal do fator Tempo_	31
Tabela 5: Análise do nível descritivo amostral referente às comparações múltiplas dentro do fator Tempo _____	32
Tabela 6: Análise do nível descritivo amostral referente ao efeito do fator Fluído dentro do Tempo 1 _____	33
Tabela 7: Análise do nível descritivo amostral referente às comparações múltiplas do fator Fluído dentro do Tempo 1 _____	34
Tabela 8: Análise do nível descritivo amostral referente ao efeito do fator Fluído dentro do Tempo 2 _____	35
Tabela 9: Análise do nível descritivo amostral referente às comparações múltiplas do fator Fluído dentro do Tempo 2 _____	36
Tabela 10: Análise do nível descritivo amostral referente ao efeito do fator Fluído dentro do Tempo 3 _____	37
Tabela 11: Análise do nível descritivo amostral referente às comparações múltiplas do fator Fluído dentro do Tempo 3 _____	38
Tabela 12: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluído EB _____	39
Tabela 13: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluído EB _____	39
Tabela 14: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluído WBF_NAF _____	40

Tabela 15: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluído WBF_NAF _____	40
Tabela 16: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluído EB_WBF_NAF _____	41
Tabela 17: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluído EB_WBF_NAF _____	41
Tabela 18: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluído WBF _____	42
Tabela 19: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluído WBF _____	42
Tabela 20: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluído CONTROLE _____	43
Tabela 21: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluído CONTROLE _____	43

Resumo

Testes estatísticos com base na normalidade multivariada muitas vezes não são as melhores alternativas quando há grandes e esparsas matrizes de espécies decorrentes de estudos de campo e de comunidade experimental. Trabalhos recentes e amplamente utilizados descrevem como alternativa procedimentos de permutação ou aleatorização sem que seja necessário estabelecer as devidas suposições sobre a distribuição teórica. Tais procedimentos são baseados em uma definição motivada biologicamente na similaridade de pares de observações em termos da composição de espécies. Em 1993, Clarke denomina de teste ANOSIM (devido à sua dependência somente das similarites e a analogia com a ANOVA de um e dois fatores) um procedimento para estabelecer se as diferenças existentes entre grupos definidos *a priori* são significativas.

Por se tratar de um teste cujo nível descritivo amostral muitas vezes não é calculado com todas as permutações possíveis, devido à demanda computacional bem como das limitações do software comumente empregado, algumas inconsistências podem ser observadas na aplicação desta técnica. Utilizando um estudo de caso de uma variável aleatória biológica do banco de dados do Projeto MAPEM, algumas considerações serão feitas acerca destas inconsistências observadas.

Abstract

Multivariate normality-based statistical tests are not, most of the time, the best alternatives when there are large and sparse matrices originated in field and experimental communities studies. Recent and broadly-used articles describe as an alternative permutation or randomization procedures, without the necessity of establishing specific assumptions on the theoretical distribution. Such procedures are based on a biologically-defined motivation in the pairwise similarity, in terms of species composition. In 1993, Clarke denominates ANOSIM test (owing to it depending only on the similarities and the analogy to one and two-away ANOVA), a procedure to establish if the existing differences between *a priori* defined groups are significant.

Being a test with a descriptive sampling level mostly not calculated because of software limitations and high computational demands, a few inconsistencies may be observed in the application of this technique. Utilizing a case study of a biological random variable of the MAPEM Project database, considerations were made on these observed inconsistencies.

1. INTRODUÇÃO

As técnicas para a análise de dados multivariados, sobretudo na ecologia, são cada vez mais importantes. Ecologistas freqüentemente precisam testar hipóteses acerca dos efeitos de fatores experimentais sobre a estrutura de comunidades como um todo. Isto é importante em pesquisas ecológicas como em estudos de biodiversidade ou impactos ambientais em muitos habitats, incluindo nestes casos ambientes marinhos (Warwick et al. 1988; Gray et al. 1990; Chapman et al. 1995; Glasby 1997 *apud* Marti 2001), manguezais (Skilleter 1996; Kelaher et al. 1998 *apud* Marti 2001), sistemas de água doce (Faith et al. 1995; Quinn et al. 1996 *apud* Marti 2001) e/ou terrestres (Oliver & Beattie 1996 *apud* Marti 2001).

Métodos de análise multivariada de variância como a MANOVA acabam sendo inválidos quando os dados consistem tipicamente da contagem de muitas dezenas ou centenas de taxons registrados em uma matriz de espécies dominada por zeros, onde resultados diferentes de zero possuem grande variação e distribuições altamente assimétricas a direita, muitas vezes identificadas como uma distribuição de Poisson. A análise deste tipo de dados é realizada por meio de métodos não paramétricos e a representação geralmente inicia, explícita ou implicitamente, a partir de uma matriz triangular de similaridades definida, por exemplo, pelas distâncias de Bray-Curtis, Euclidiana entre outras. A escolha da medida utilizada varia conforme o tipo de dado estudado e o conhecimento do pesquisador.

Com base na análise da matriz de similaridades, Clarke & Green (1988) descrevem um procedimento simples de aleatorização para delineamentos unifatoriais, um tipo de teste de Mantel (1967) baseado em uma estatística de teste que reflete a diferença na classificação média das similaridades entre e dentro grupos definidos *a priori*. Clarke denominou tal teste de análise de similaridades (ANOSIM) em alusão à análise de variâncias (ANOVA).

Como muitos experimentos (a grande maioria) envolvem o estudo de dois ou mais fatores, Clarke (1993) descreveu uma extensão da ANOSIM para um delineamento com dois fatores, ambos com repetições no nível mais baixo. Um tipo de efeito de grande interesse quando dois fatores são analisados, surge nos casos em que um fator não produz o mesmo efeito sobre a resposta para todos os níveis do outro fator testado, ou seja, quando há um efeito de interação entre os fatores (Montgomery 2001). Entretanto, a ANOSIM testa apenas o efeito principal de cada fator, isto é, o efeito que o fator produz na variável resposta que é

devido exclusivamente a ele, pois tal efeito é constante em todos os níveis do outro fator testado.

Com a utilização de um banco de dados do Projeto MAPEM, algumas análises de similaridade serão realizadas e através da exposição dos resultados obtidos para a estatística de teste (R) e seu nível descritivo amostral associado (p-valor), considerando para isto, diferentes combinações nos número de permutações possíveis. O software que é amplamente utilizado na comunidade científica para realizar esta técnica é o PRIMER (*Plymouth Routines In Multivariate Ecological Research*) e a versão 5 será adotada neste trabalho.

2. OBJETIVO

2.1.OBJETIVO PRINCIPAL

Avaliar a consistência da análise de similaridades através da estatística de teste e os níveis descritivos amostrais (p-valores) resultantes de um esforço exploratório em uma variável aleatória biológica do banco de dados do Projeto MAPEM, considerando para isto diferentes números de permutações condicionadas às limitações impostas pelo software PRIMER.

2.2. OBJETIVOS SECUNDÁRIOS

- Avaliar a similaridade da estrutura de comunidade biológica observada no âmbito do projeto MAPEM quanto à distribuição entre os grupos formados pelos indicadores de fluídos de perfuração;
- Avaliar a similaridade da estrutura de comunidade biológica observada no âmbito do projeto MAPEM quanto à distribuição referente aos intervalos de tempo em que as coletas foram observadas.
- Constatar se é plausível estipular um número mínimo de permutações necessárias para se atingir a significância estatística em 100% das análises.

3. REVISÃO BIBLIOGRÁFICA

As técnicas multivariadas empregadas na análise de estrutura de comunidades tal como agrupamento ou ordenação, foram originalmente concebidas como ferramentas exploratórias de dados (Field et al. 1982 *apud* Clarke & Warwick 1993). Mais recentemente, elas vêm sendo aplicadas em delineamento de experimentos ou de forma análoga, em estudos estruturados de campo onde procedimentos de testes de hipóteses estatísticos são empregados para avaliar mudanças na estrutura de comunidade, através da avaliação dos efeitos de tratamentos repetidos ou não no tempo (Gee et al. 1985, Warwick et al. 1990c, Dawson-Shepherd et al. 1992 *apud* Clarke & Warwick 1993).

Métodos estatísticos multivariados como a tradicional Análise Multivariada de Variância (MANOVA) já existem há décadas, no entanto para os resultados serem válidos, a maioria destes testes se baseia em alguns pressupostos, tais como a linearidade e a normalidade multivariada dos dados. A estatística de teste da MANOVA, quando utilizada para analisar matrizes de estrutura de comunidades, não é robusta em relação à ausência da normalidade multivariada (Mardia 1971; Olson 1974; Johnson & Field 1993 *apud* Marti 2001). Tipicamente, se todas as espécies são consideradas na análise, as matrizes geralmente apresentam-se muito extensas, sendo predominantemente constituídas de valores iguais a zero. Uma alternativa para solucionar tal problema, descrita por Clarke & Green em 1988, é utilizar-se de uma transformação combinada com uma redução substancial das espécies consideradas. A redução na matriz original dos dados (redução de variáveis aleatórias) pode contribuir para que haja alguma chance de validade nas distribuições assintóticas das estatísticas de testes multivariados. Considera-se para isto, que o número total de observações (n) vezes o número de espécies (p) deve ser grande o suficiente quando comparado ao número de parâmetros a serem estimados na modelagem, em uma razão de $2p + [p(p-1)]/2$.

Em 1988, Clarke sugere um teste baseado nos princípios de permutação e aleatorização (Hope, 1968) que atua sobre uma matriz de similaridades, ou dissimilaridades, que mostra o quanto cada observação é similar (ou dissimilar) em termos da estrutura de comunidade através de uma medida escolhida. A matriz tem como colunas e linhas as observações, e os valores das células representam as suas respectivas distâncias. A diagonal é composta de valores iguais a 1 (ou 100%) (a similaridade entre uma observação e ela própria é 100%) e a matriz é, normalmente, simétrica (a similaridade entre as observações j e k é a

mesma que entre as observações k e j). Desse modo, podemos representá-la como uma matriz triangular inferior. Tais matrizes são início para muitas análises descritivas, tanto em técnicas de agrupamento quanto de ordenação.

A escolha da medida de similaridade mais apropriada varia de acordo com o tipo de dados que serão analisados (Legendre & Legendre 1988 *apud* Marti 2001), uma vez que há muitos tipos de medidas de similaridade cada uma delas com sua especificidade. Por exemplo, para expressar diferenças em estruturas de comunidades, a medida de simetria de Bray-Curtis (Bray & Curtis 1957 *apud* Marti 2001) ou a medida de Kulczynski (1928 *apud* Marti 2001) são geralmente preferidas dentre outras medidas simétricas como, por exemplo, a distância Euclidiana (Odum 1950; Hajdu 1981; Faith 1987; Clarke 1993 *apud* Marti 2001).

Suponha que n observações consistam de r repetições em cada um dos k grupos definidos *a priori*, sendo que o número de repetições, segundo Clarke & Green (1988), não é necessariamente igual em cada um dos grupos. Pode-se construir uma estatística de teste que provavelmente refletirá a similaridade das repetições dentro de um grupo em contraste com as similaridades entre os grupos. A estatística calculada é comparada com o seu valor sob um grande número de permutações aleatórias das possíveis classificações dos tratamentos atribuídos às observações, de acordo com uma hipótese nula de que "não há diferença entre tratamentos". Se a estatística calculada para os dados originais é mais extrema do que 95% das reclassificações aleatórias, então a hipótese nula é rejeitada por um teste aleatório com p -valor $< 0,05$. Alternativamente, se o número de grupos e repetições é pequeno, pode ser factível enumerar todas as possíveis reclassificações e assim construir um teste com todas as permutações possíveis, que é o desejável. Clarke & Green (1988) denominaram este procedimento de teste ANOSIM (análise de similaridades), em analogia à análise de variância (ANOVA).

A estatística de teste, denominada R , pode ser padronizada de modo que o valor zero reflita a hipótese nula de não diferença nos grupos, e quanto mais próximo de +1, mais próximo se estará de rejeitar a hipótese nula e admitir-se a hipótese alternativa, na qual todas as r repetições dentro de pelo menos um grupo são mais similares entre si do que quaisquer repetições entre grupos. Se as diferenças globais são encontradas, comparações de grupos podem ser feitas utilizando-se precisamente os mesmos princípios de aleatorização/permutação, desde que haja repetições suficientes em cada grupo para gerar um conjunto grande o suficiente de permutações possíveis (Clarke & Green 1988). Clarke (1993)

define R global como a estatística que testa a diferença entre e dentro todos os grupos e R como a estatística de teste para cada comparação múltipla.

4. MATERIAIS E MÉTODO

Segundo Clarke & Warwick (1994), para testar a veracidade da hipótese nula de que não há diferença entre os grupos, três etapas devem ser cumpridas:

- Calcular uma estatística de teste;
- Recalcular a estatística de teste mediante permutações;
- Calcular o nível descritivo amostral (p-valor).

Previamente ao cálculo da estatística de teste, é relevante mencionar alguns conceitos sobre a medida que a estatística utiliza como base.

4.1 COEFICIENTE DE SIMILARIDADE (S)

O conceito de similaridade (S) entre algum par de observações é o ponto de partida para muitas análises. Inevitavelmente, como a informação para cada observação é multivariada (muitas espécies), há muitas formas de definir similaridade, sendo que para cada forma há um critério de ponderação diferente, considerando diferentes aspectos de comunidade. Um coeficiente de similaridade é convencionalmente definido para valores no intervalo $[0,100\%]$, onde 100% representa que as duas observações são totalmente similares e 0% representa que as duas observações são totalmente dissimilares (Clarke & Warwick, 1994).

4.1.1 Matriz de similaridades

As similaridades são calculadas entre cada par de observações e, por convenção, apresentam-se os $n(n-1)/2$ valores em uma matriz triangular inferior. A matriz de similaridades é uma matriz quadrada cujas linhas e colunas são as observações (1 a n), entretanto não é necessário preencher tanto a diagonal (a similaridade da observação j com ela mesma é sempre 100%) quanto a parte triangular superior (a similaridade da observação j com a observação k é a mesma da observação k com a j) (Clarke & Warwick, 1994). A Tabela 1 apresenta um exemplo de uma matriz de similaridades para 10 observações.

Tabela 2: Exemplo de uma matriz de similaridades

	1	2	3	4	5	6	7	8	9	10
1										
2	43,9									
3	33,33	41,38								
4	76,19	46,15	50							
5	54,54	66,64	41,18	53						
6	81,81	58,53	64	54,14	26,62					
7	40	29,63	55,55	43,75	66,58	44,44				
8	53,65	68,42	27	77,78	26,64	52,62	26,08			
9	23	17	60	66,15	70,97	28,28	25	26,07		
10	47,62	82,05	22,22	26,08	42,86	54,45	57,14	75	30	

4.1.1.1 Transformação dos dados

Em alguns casos as similaridades podem ser indevidamente dominadas pela soma das espécies mais frequentes. A maior frequência nos dados da matriz original será muitas vezes extremamente variável em observações idênticas e é indesejável basear uma avaliação da similaridade das comunidades somente sobre a soma de espécies muito abundantes.

A alternativa para que isto não ocorra, é transformar os valores originais de y antes de calcular o coeficiente de similaridade. Transformações como $\log(1 + y)$ e $\sqrt[4]{y}$ têm um efeito bastante importante em sub-ponderar a importância das espécies muito abundantes de modo que espécies menos dominantes, e mesmo espécies raras, tenham algum papel na determinação da similaridade entre duas observações.

A escolha da transformação pode afetar as conclusões das análises, portanto ela acaba sendo mais uma questão biológica do que estatística. De fato, para dados muito variáveis, a escolha da transformação pode ser mais crítica do que a escolha do coeficiente de similaridade.

Entre os coeficientes de similaridade destacam-se o coeficiente de Bray-Curtis e a distância Euclidiana, cujas formas de cálculo serão detalhadas a seguir.

4.1.2 Coeficiente de Bray-Curtis

O coeficiente de Bray-Curtis, também é conhecido como coeficiente de Czekanowski e é um dos coeficientes de similaridade que tem sido amplamente utilizado em trabalhos

ecológicos. (Bray & Curtis, 1957 *apud* Clarke & Warwick 1994). A similaridade entre as observações j e k tem duas definições equivalentes:

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

Equação 1

$$= 100 \left(\frac{\sum_{i=1}^p 2 \min(y_{ij}, y_{ik})}{\sum_{i=1}^p (y_{ij} + y_{ik})} \right)$$

Aqui, as observações estão definidas nas colunas e as espécies nas linhas, então y_{ij} representa a entrada da i -ésima linha e da j -ésima coluna da matriz de dados. Do mesmo modo que y_{ik} é o valor para i -ésima espécie na k -ésima observação. $|...|$ representa o valor absoluto da diferença (o sinal é ignorado) e $\min(...)$ o mínimo dos dois valores; as somas do numerador e denominador são sobre todas as linhas (espécies) na matriz.

Uma propriedade que o coeficiente possui é que uma mudança na escala de medida não muda o seu valor; todos os valores de y são multiplicados pela mesma constante e então cancelados nos termos do numerador e denominador da Equação 1. Outra característica é que ausências conjuntas não têm efeito no coeficiente. Algumas vezes pode acontecer de ter espécies ausentes em todas as observações; a omissão dessas espécies não faz diferença nos dois somatórios na Equação 1. De acordo com Clarke & Warwick (1994) esta propriedade de a similaridade depender apenas de espécies presentes em uma ou outra (ou ambas) observação é normalmente desejável.

4.1.3 Distância Euclidiana

A distância natural entre quaisquer dois pontos no espaço é referida como distância Euclidiana (da geometria Euclidiana ou clássica). No contexto de matriz de espécies, a distância Euclidiana entre as observações j e k é definida algebricamente como:

$$d_{jk} = \sqrt{\sum_{i=1}^p (y_{ij} - y_{ik})^2}$$

Equação 2

Isto pode ser melhor compreendido, geometricamente, tomando o caso especial onde há somente duas espécies para que as repetições possam ser representadas por pontos no espaço bidimensional, ou seja, a posição delas sobre os dois eixos da contagem da Espécie 1 e da Espécie 2.

Isso é ilustrado na Figura 1 para uma matriz com duas repetições de duas espécies abundantes. As coordenadas (2,3) e (5,1) sobre os eixos (Espécie 1, Espécie 2) são as duas observações j e k . A distância direta d_{jk} entre elas corresponde à Equação 2.

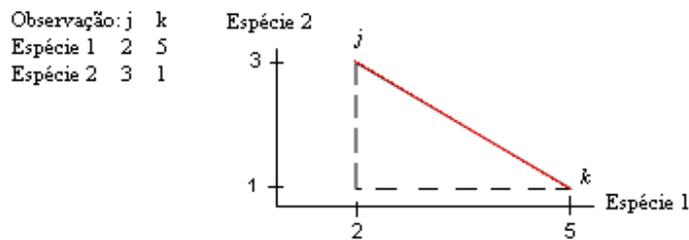


Figura 1: Exemplo de distância Euclidiana entre as observações j e k

Fonte: Clarke & Warwick (1994)

É fácil imaginar a extensão para uma matriz com três espécies; os dois pontos estão localizados em eixos tridimensionais e sua distância em linha reta é um conceito geométrico natural. Algebricamente, ela é a raiz da soma das distâncias ao quadrado ao longo dos três eixos (Equação 2). O conceito permanece inalterado em dimensões superiores a três: distâncias ao quadrado adicionais para cada nova espécie são adicionadas ao somatório sob a raiz quadrada (Equação 2) (Clarke & Warwick 1994)

4.2 ESTATÍSTICA R

Na ANOSIM, a matriz de similaridades é utilizada para discriminar grupos cuja similaridade entre as suas repetições são maiores que as similaridades entre as repetições de grupos diferentes. A estatística de teste proposta por Clarke (1988, 1993) é dada por:

$$R = \frac{\bar{r}_B - \bar{r}_W}{n(n-1)/4} \quad \text{Equação 3}$$

onde

\bar{r}_B, \bar{r}_W são as médias das similaridades entre grupos e dentro dos grupos, respectivamente e n o número total de observações (Clarke & Warwick 1994). A Figura 2 ilustra a medida de similaridade dentro de cada um de dois grupos genéricos com 4 repetições

cada – grupo preto (diagrama a) e grupo branco (diagrama b) – e a medida de similaridade entre os dois grupos (diagrama c).

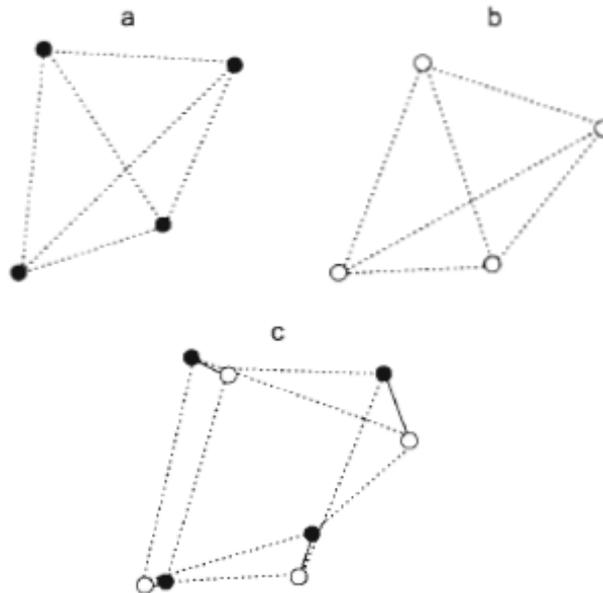


Figura 2: Diagramas ilustrando as similaridades entre e dentro dos grupos a e b
Fonte: Chapman & Underwood (1999)

A estatística R está definida para o intervalo $[-1$ a $1]$, onde valores próximos a zero indicam a aceitação da hipótese nula, ou seja, não há diferenças significativas nos grupos. Valores iguais a $+1$ são obtidos apenas quando todas as réplicas dentro dos grupos são mais similares entre si do que qualquer réplica de grupos diferentes (Masi & Zalmon 2008). A estatística normalmente se situa entre zero e um, indicando algum grau de discriminação entre os grupos.

Embora valores negativos ou muito próximos de -1 da estatística R sejam considerados pouco prováveis, uma vez que eles representariam uma variabilidade maior dentro dos grupos do que entre os grupos, em alguns estudos tais valores são, contudo, relativamente comuns (Somerfield & Clarke 1995, Olsgard et al. 1997, Chapman & Underwood unpubl. data *apud* Chapman & Underwood 1999) e podem inclusive ser substancialmente menores que zero. No entanto, eles têm sido geralmente ignorados na qualidade de valores irrelevantes, anômalos ou ainda, indicativos de uma “classificação incorreta dos grupos” (Clarke & Warwick 1994). Segundo Chapman & Underwood (1999), tais valores podem ser um indicativo sobre os processos ecológicos criadores de diferenças em potencial.

4.3 ESTATÍSTICA R PERMUTADA

Suponha que a hipótese nula é verdadeira, isto é, não há diferença entre os grupos definidos *a priori*. As classificações das observações que as identificam como pertencentes a um determinado grupo podem ser aleatoriamente misturadas (permutadas) e um novo valor da estatística R , denominado R^* , é obtido. Este cálculo aleatório de R^* é repetido para todas as possíveis reordenações das observações em relação às classificações gerando um valor de R^* para cada nova permutação.

Segundo Clarke (1993), para experimentos balanceados, com k grupos e r repetições por grupo, o número de diferentes resultados possíveis para a estatística R^* , em um teste com um fator, é dado por

$$\frac{(kr)!}{(r!)^k k!} \quad \text{Equação 4}$$

Em muitos casos, contudo, o cálculo de todas as permutações possíveis pode ser muito demorado ou até mesmo inviável. Neste caso, realiza-se uma amostragem aleatória com reposição para as permutações possíveis, que estima a distribuição da estatística R^* sob a hipótese nula estabelecida para a população.

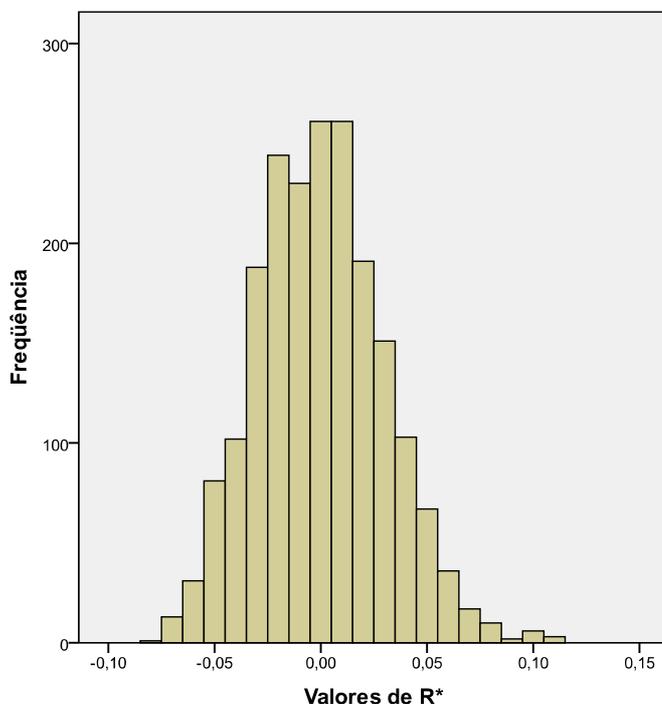


Figura 3: Exemplo de uma distribuição de R^* sob a hipótese nula.

4.4 CÁLCULO DO P-VALOR

Comparando o valor de R obtido considerando a alocação original das observações em cada um dos grupos definidos com a distribuição obtida pela permutação das similaridades esperadas para o caso em que a hipótese nula é verdadeira (R^*), um p-valor $(t + 1)/(T + 1)$ é calculado, onde T é o número de permutações calculadas (R^*) e t é a quantidade destas permutações que foram maiores que a estatística R observada.

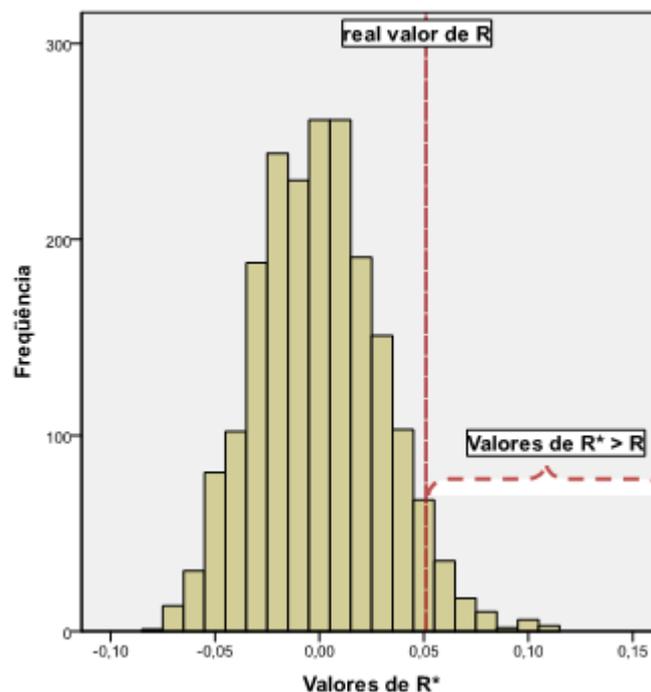


Figura 4: Exemplo de uma distribuição de R^* sob a hipótese nula

Note que o valor original observado de R é considerado parte da distribuição de R^* sob permutação logo, uma das possíveis ordenações das classificações nas linhas. Um nível de significância *a priori* geralmente dado por $\alpha = 0,05$ é utilizado para interpretar a significância do resultado, de forma análoga aos testes de hipóteses clássicos. Se o verdadeiro valor de R revelar-se improvável de ser originário da distribuição R^* , haverá evidências para rejeitar a hipótese nula (Clarke & Warwick 1994).

Como normalmente não é prático calcular todas as permutações possíveis em decorrência da demanda computacional envolvida, p pode ser calculado utilizando um grande subconjunto aleatório de todas as permutações possíveis. É preferível utilizar um teste conhecido de boa eficiência ao invés de um procedimento de teste de Monte Carlo assumindo que a hipótese estatística alternativa pode ser completamente especificada. No entanto, como exposto, nem sempre é possível utilizar o teste porque as condições necessárias para a sua aplicação não podem ser satisfeitas ou a distribuição subjacente pode ser desconhecida ou pode ser difícil decidir um critério de teste apropriado. Nestes casos, o uso de testes de Monte Carlo de significância torna possível obter algumas informações sobre uma situação que não seria possível ou facilmente atingível de outra forma. (Hope 1968).

A precisão do p-valor aumentará com o acréscimo do número de permutações. Geralmente, pelo menos 1000 permutações devem ser feitas para testes com significância de $\alpha = 0,05$ e pelo menos 5000 permutações devem ser feitas para testes com significância de $\alpha = 0,01$ (Manly 1997 *apud* Marti 2001).

4.5 ANÁLISE DAS COMPARAÇÕES MÚLTIPLAS

Uma vez que o teste rejeita a hipótese nula, é interessante descobrir entre quais grupos estão as diferenças significativas. O procedimento de comparações múltiplas, em analogia ao teste complementar da Análise de Variância tradicional, caracteriza-se pela comparação dois a dois de todos os níveis do fator. Para a comparação em questão, é calculada a estatística R (Equação 3) e o correspondente nível descritivo amostral. Para fim de diferenciação, a estatística R relacionada a cada comparação será apenas R enquanto que a relacionada a análise do conjunto inteiro de dados é denominada de R global.

4.6 PROJETO MAPEM

O MAPEM – Monitoramento Ambiental em Atividades de Perfuração Exploratória Marinha – é um projeto de monitoramento ambiental desenvolvido por iniciativa de Centros de Pesquisa das Universidades Federais do Rio Grande do Sul e de Santa Catarina, em parceria com a indústria brasileira de óleo e gás, representadas pelo IBP.

O desenvolvimento do projeto MAPEM teve por objetivo fornecer dados essenciais para validar os modelos de descargas, particularmente a distribuição dos cascalhos no fundo

do mar em águas profundas. e estabelecer os tipos de biota que vive nesses ambientes, o tipo e o grau de impacto sobre essa comunidade e a velocidade de recuperação da comunidade. Tais dados são muito úteis para as agências reguladoras à medida que forem desenvolvidos os regulamentos que governam o descarte de resíduos de perfuração.

A atividade de perfuração de poços de óleo e gás implica na utilização de fluídos genericamente denominados fluídos de perfuração e a indústria utiliza os fluídos base água (WBFs) e os não aquosos (NAFs).

Pressupõe-se uma região localizada próxima ao poço exploratório de amplitude máxima definida como o limite de alcance dos registros de atividade de perfuração, conhecida como região de impacto. A região referência está localizada a uma distância suficiente para preservar as condições ambientais naturais, independentes da atividade antrópica.

A região de impacto foi delimitada em 500 metros em torno do poço, observada através de 48 estações amostrais. A região utilizada para controle foi definida como região de referência sendo localizada a 2500 metros do centro de poço. Três estações foram situadas ao norte e três estações ao sul do poço. O esquema amostral seguiu um delineamento sistemático com pontos amostrais igualmente espaçados nos raios de 50, 100, 150, 300, 3500 e 2500 metros.

A distribuição espacial das observações foi repetida três vezes no tempo. A primeira observação da área de estudo ocorreu antes da atividade de perfuração. A segunda deu-se um mês o término da perfuração e a terceira ocorreu aproximadamente um ano após a perfuração. A sigla MD identifica o estudo em águas profundas e significa “MAPEM DEEP”. As observações realizadas antes da perfuração foram denominadas MD1, e um mês após a perfuração MD2. As observações realizadas um ano após a perfuração, foram denominadas de MD3 (Figura 5) (E.E.Toldo Jr et al. 2004)

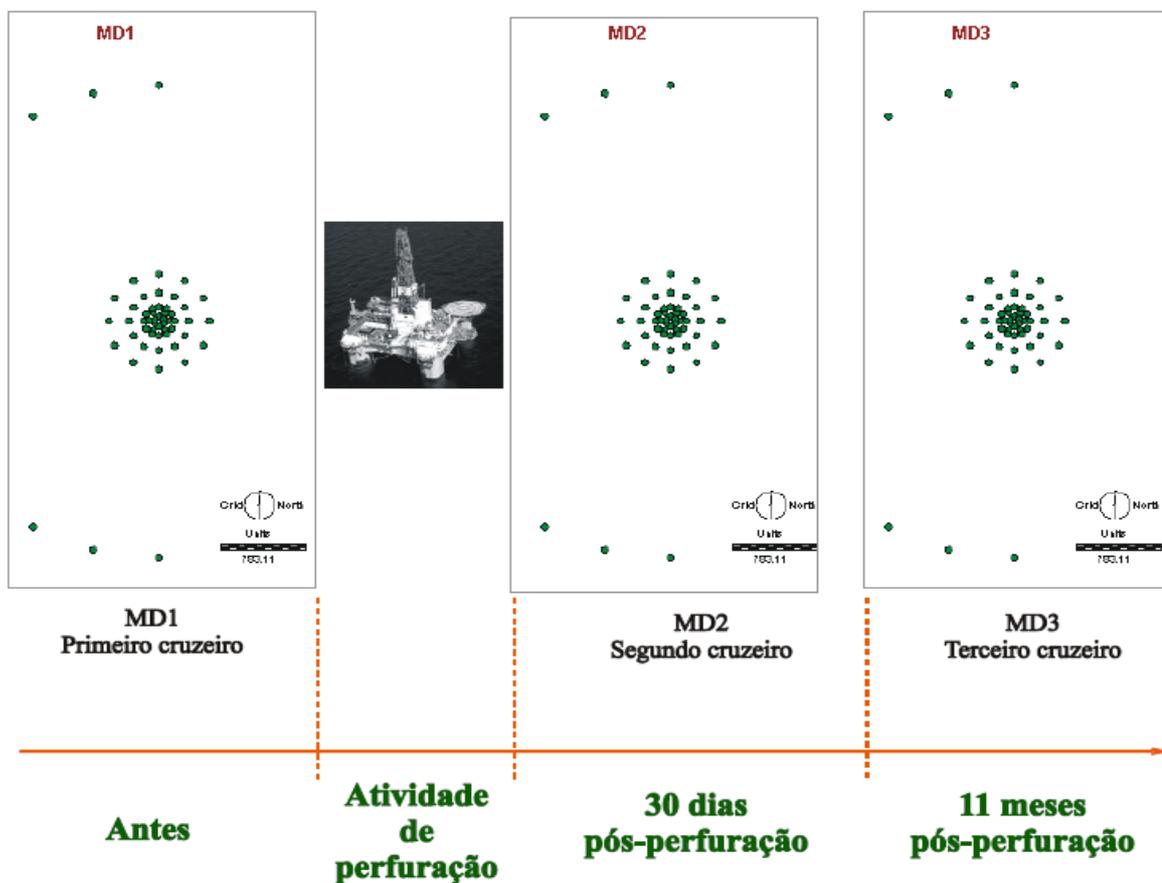


Figura 5: Delineamento amostral do projeto MAPEM
 Fonte: E.E.Toldo Jr et al. 2004

4.7 SOFTWARE

A ANOSIM está incluída no software PRIMER (*Plymouth Routines In Multivariate Ecological Research*), que compreende um conjunto de programas computacionais desenvolvidos no Laboratório Marinho de Plymouth. Para termos de curiosidade a seguir, mediante ilustrações, serão descritas as rotinas utilizadas, tais como a leitura do banco de dados pelo software (Figura 6), a escolha da medida de similaridade adotada, se a similaridade será entre as observações ou entre as espécies e, se será feita alguma transformação nos dados (Figura 7), a partir da matriz de similaridades gerada, a escolha da técnica multivariada que será adotada (Figura 8), a definição do delineamento experimental, número de fatores (no máximo dois) e o número de permutações (Figura 9) e por fim, a saída obtida que apresenta a estatística de teste global (R), o nível descritivo amostral encontrado com o número de permutações definido na etapa anterior e as comparações múltiplas (Figura 10).

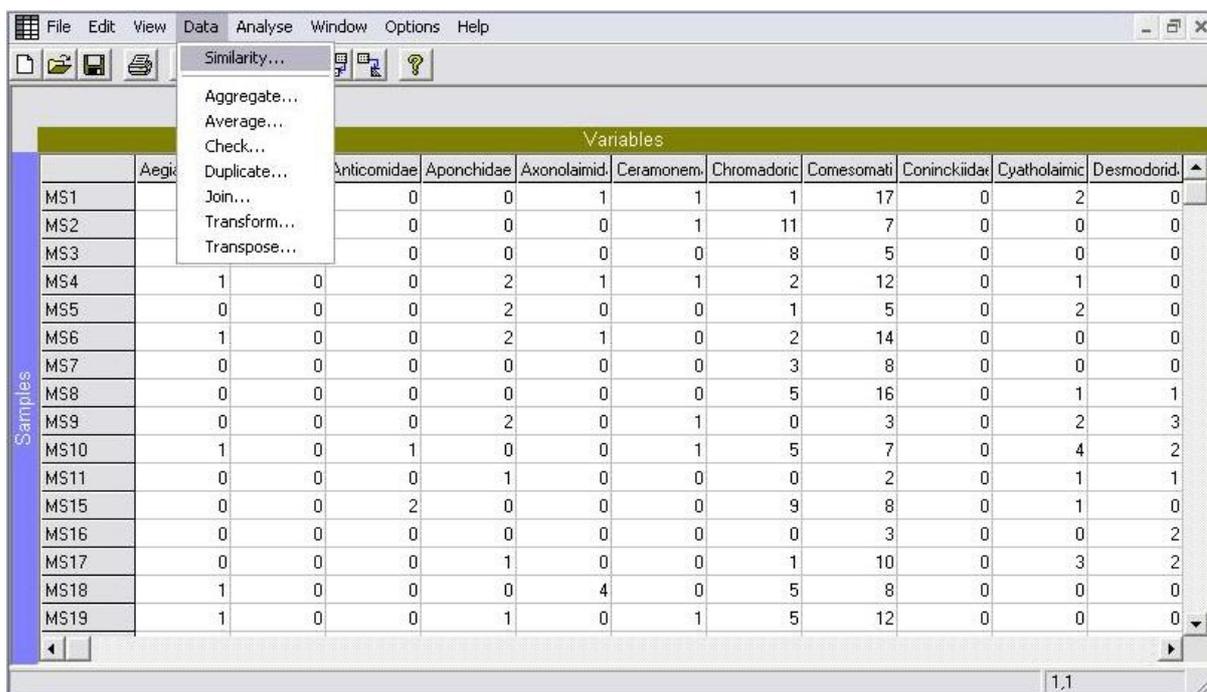


Figura 6: Banco de dados no software Primer

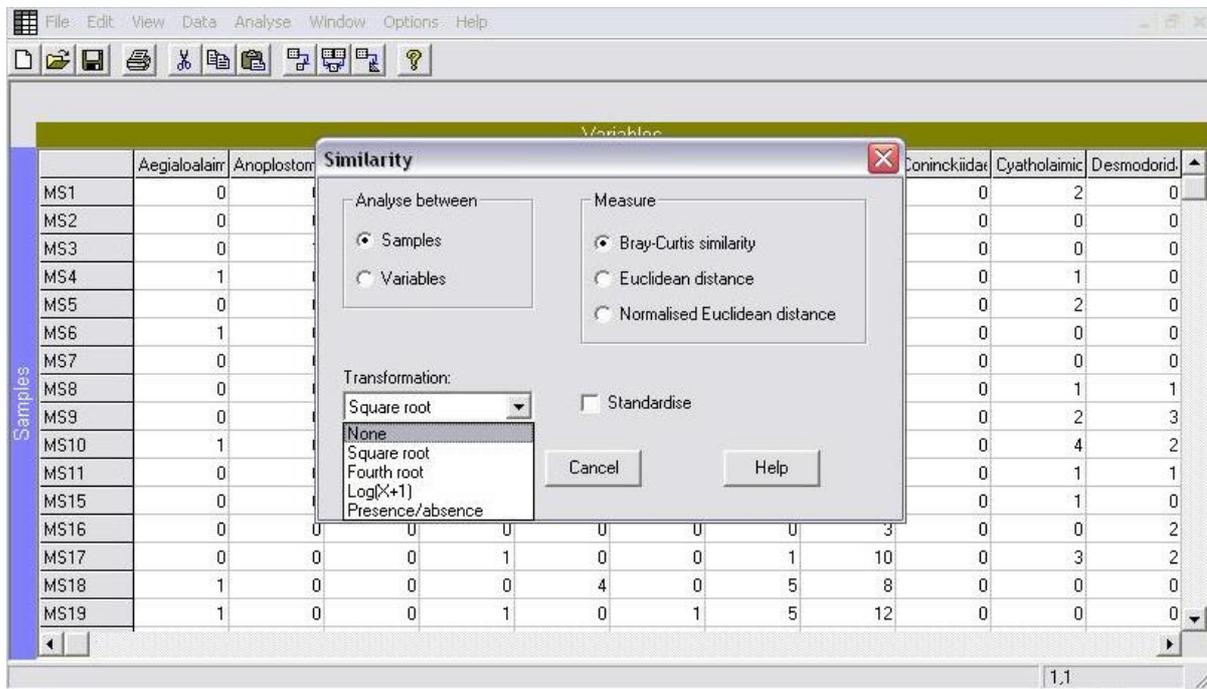


Figura 7: Escolha da matriz de similaridades

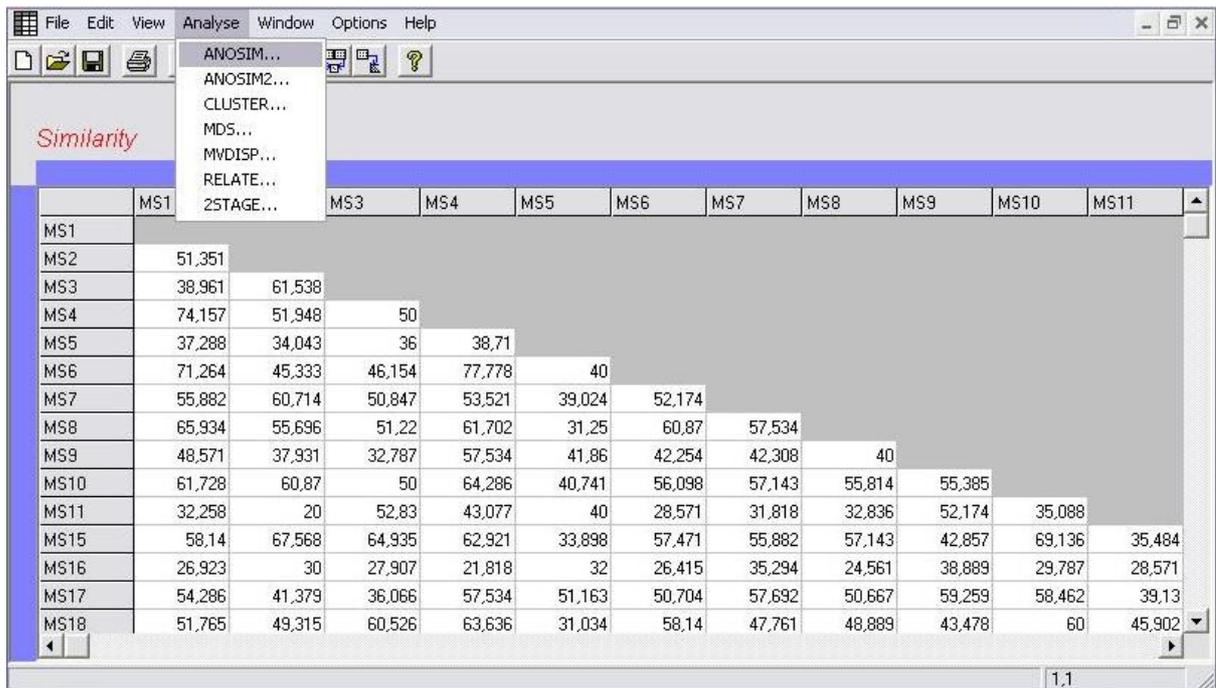


Figura 8: Escolha da técnica multivariada realizada com base na matriz de similaridades obtida

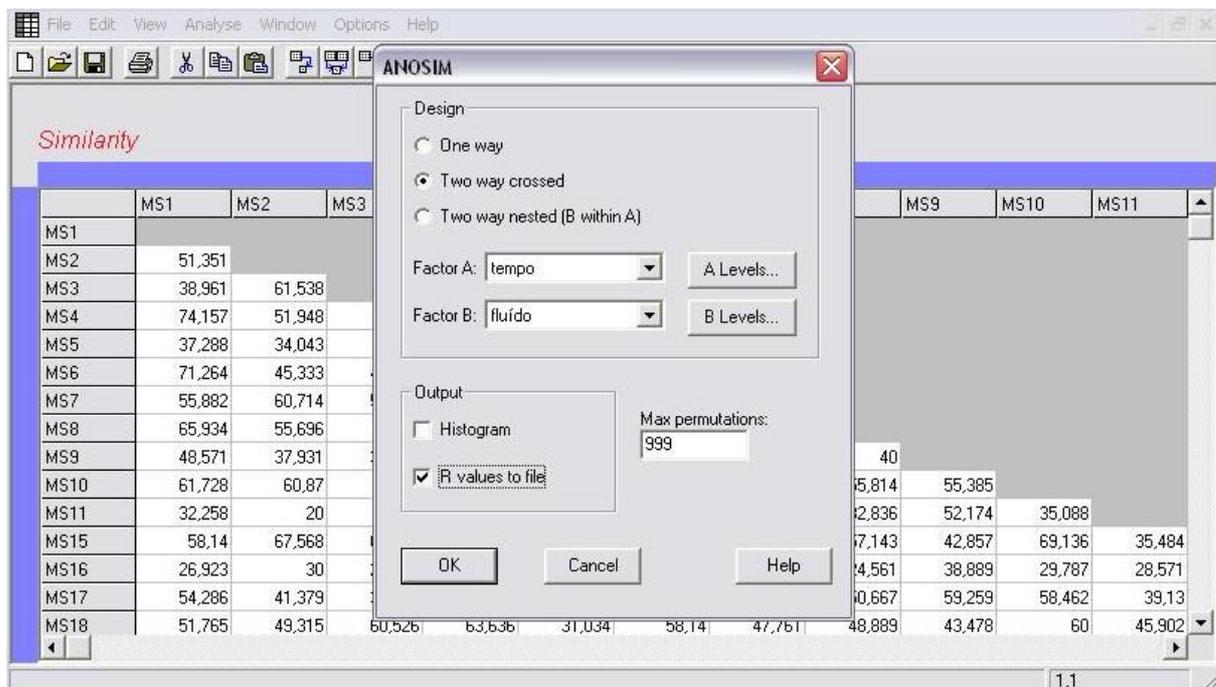


Figura 9: Definição do delineamento experimental, fatores e número de permutações.

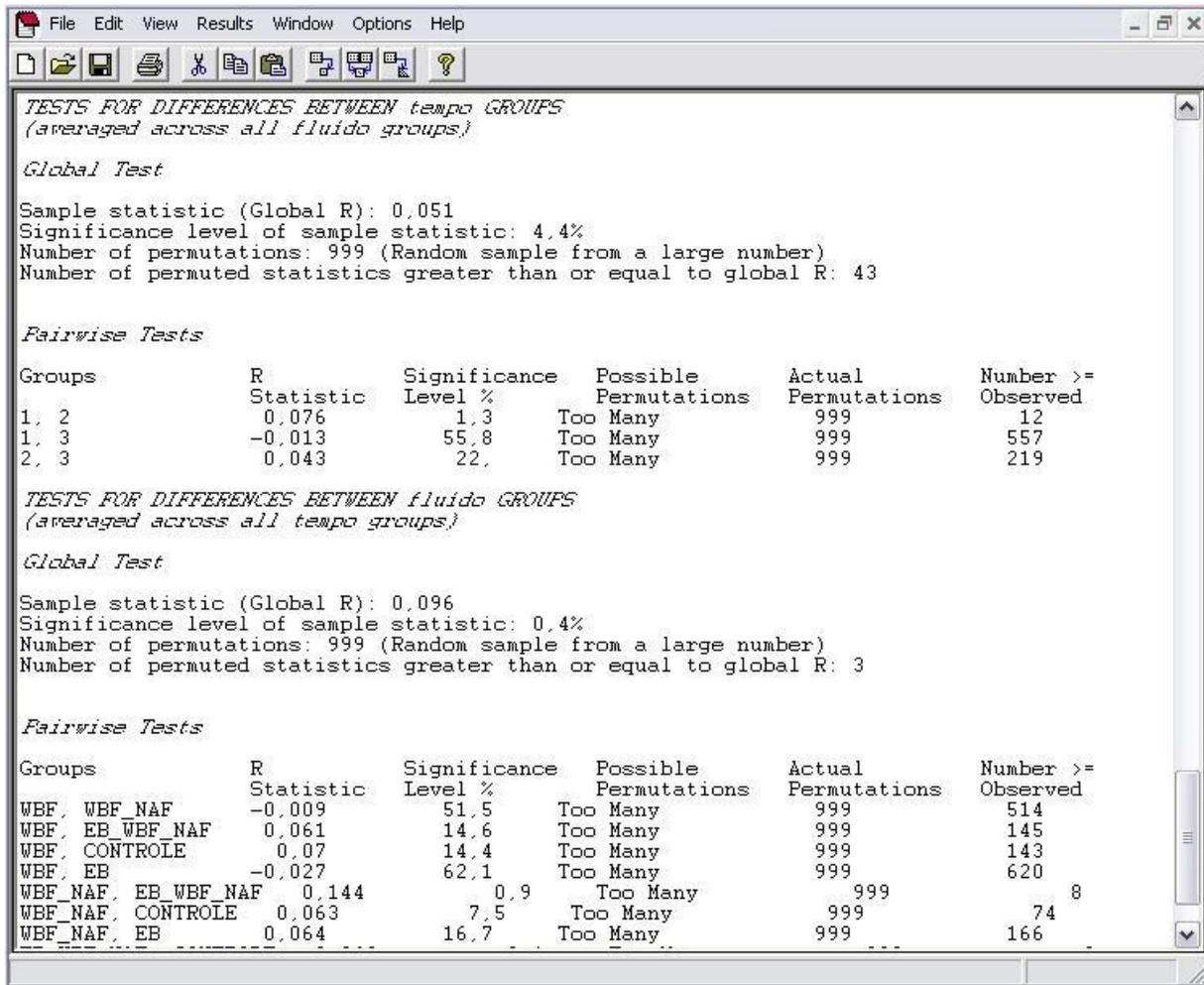


Figura 10: Parte da saída obtida para uma análise com dois fatores.

5. ESTUDO DE CASO – ANÁLISE DE DADOS E RESULTADOS

Visto que a ANOSIM é uma técnica baseada em permutações, para cada repetição da mesma análise é possível encontrar um resultado diferente uma vez que é feita uma amostragem do número total de permutações possíveis definido na Equação 2. Assim, o valor encontrado em outra permutação possível que não a primeira, pode levar a uma decisão estatística diferente da tomada inicialmente. A fim de avaliar este comportamento, estipularam-se para cada análise de interesse 999, 4999, 9999 e 49999 permutações mínimas e 50 repetições (amostras) de cada análise para cada um dos níveis definidos. A Figura 11 ilustra o processo realizado, desde a escolha da medida de similaridade até a 50ª repetição da análise em questão.

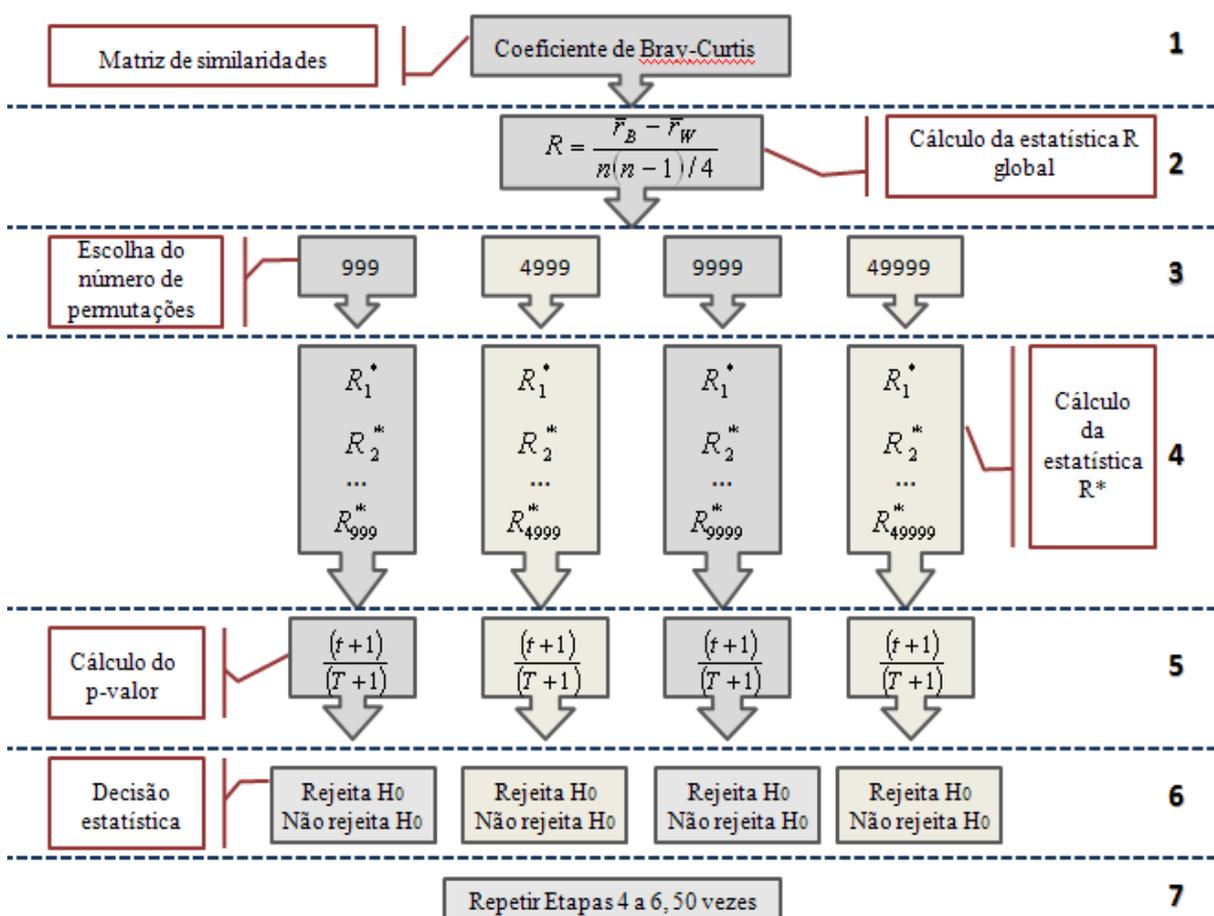


Figura 11: Fluxograma do processo realizado no presente estudo.

5.1 DELINEAMENTO COM DOIS FATORES CRUZADOS

Inicialmente, adotou-se o delineamento fatorial cruzado a fim de testar o efeito dos fatores Tempo e Fluido. Entretanto como a técnica não testa o efeito de interação, apenas o efeito principal de cada fator foi avaliado.

5.1.1 Efeito principal do fator Fluido

Na Tabela 2 é possível ver que na primeira análise utilizando 999 permutações possíveis observou-se diferença significativa em pelo menos um tipo de fluido ($R = 0,096$, $p\text{-valor} = 0,007$). A diferença também foi observada na primeira análise utilizando 4999 ($p\text{-valor} = 0,004$), 9999 ($p\text{-valor} = 0,003$) e 49999 permutações ($p\text{-valor} = 0,003$). Em 100% das análises realizadas, não há mudança na decisão estatística de rejeitar a hipótese nula de igualdade entre os fluidos. Através dos histogramas (Figura 12) é possível notar que com 999 permutações, o intervalo de variação dos níveis descritivos amostrais encontrados é mais amplo [0,001; 0,01]. Com o aumento da precisão do teste (aumento do número de permutações) o intervalo diminui.

Tabela 2: Análise do nível descritivo amostral referente ao efeito principal do fator Fluido

Fator	R	p-valor 1ª análise	Decisão estatística	% rejeição H_0	Nº permutações
Fluido	0,096	0,007	Rejeitar H_0	100	999
		0,004	Rejeitar H_0	100	4999
		0,003	Rejeitar H_0	100	9999
		0,003	Rejeitar H_0	100	49999

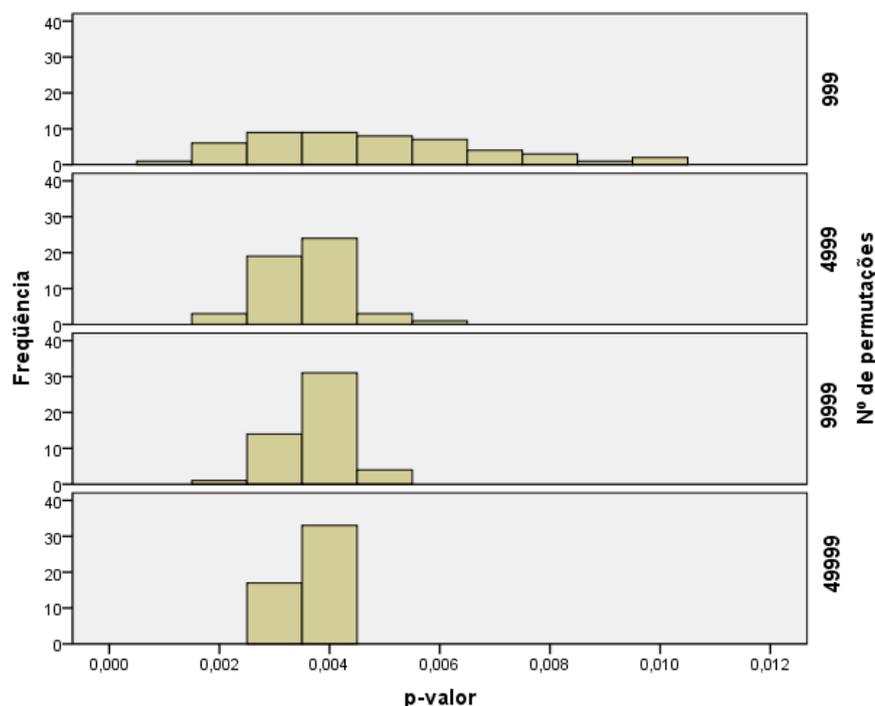


Figura 12: Histogramas do nível descritivo amostral referente ao efeito principal do fator Fluído

5.1.1.1 Comparações múltiplas

Os testes de comparações múltiplas permitem identificar qual (quais) nível(s) dos fatores diferem dos demais. A avaliação dessas comparações foi feita da mesma forma que do efeito principal dos fatores.

Na Tabela 3 observa-se que a primeira análise de comparações múltiplas do fator Fluído realizada com 999 permutações indicou que o fluído EB_WBF_NAF é estatisticamente diferente dos fluídos WBF_NAF ($R = 0,144$, $p\text{-valor} = 0,008$) e CONTROLE ($R = 0,263$, $p\text{-valor} = 0,001$). Esta mesma diferença é observada também com 4999 ($p\text{-valor} = 0,006$ e $0,000$), 9999 permutações ($p\text{-valor} = 0,007$ e $0,000$) e 49999 permutações ($p\text{-valor} = 0,006$ e $0,000$), os demais fluídos não diferem entre si. Das 50 vezes que as comparações foram repetidas, em 100% delas estas duas comparações apresentaram diferença significativa, tanto com 999, 4999, 9999 e 49999 permutações.

Tabela 3: Análise do nível descritivo amostral referente às comparações múltiplas dentro do fator Fluido

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº permutações
WBF x WBF_NAF	-0,009	0,515	Não rejeitar H ₀	0	999
		0,54	Não rejeitar H ₀	0	4999
		0,529	Não rejeitar H ₀	0	9999
		0,543	Não rejeitar H ₀	0	49999
WBF x EB_WBF_NAF	0,061	0,129	Não rejeitar H ₀	0	999
		0,132	Não rejeitar H ₀	0	4999
		0,137	Não rejeitar H ₀	0	9999
		0,133	Não rejeitar H ₀	0	49999
WBF x CONTROLE	0,07	0,17	Não rejeitar H ₀	0	999
		0,168	Não rejeitar H ₀	0	4999
		0,163	Não rejeitar H ₀	0	9999
		0,165	Não rejeitar H ₀	0	49999
WBF x EB	-0,027	0,639	Não rejeitar H ₀	0	999
		0,63	Não rejeitar H ₀	0	4999
		0,640	Não rejeitar H ₀	0	9999
		0,637	Não rejeitar H ₀	0	49999
WBF_NAF x EB_WBF_NAF	0,144	0,008	Rejeitar H₀	100	999
		0,006	Rejeitar H₀	100	4999
		0,007	Rejeitar H₀	100	9999
		0,006	Rejeitar H₀	100	49999
WBF_NAF x CONTROLE	0,063	0,093	Não rejeitar H ₀	0	999
		0,084	Não rejeitar H ₀	0	4999
		0,082	Não rejeitar H ₀	0	9999
		0,084	Não rejeitar H ₀	0	49999
WBF_NAF x EB	0,064	0,178	Não rejeitar H ₀	0	999
		0,17	Não rejeitar H ₀	0	4999
		0,167	Não rejeitar H ₀	0	9999
		0,167	Não rejeitar H ₀	0	49999
EB_WBF_NAF x CONTROLE	0,263	0,001	Rejeitar H₀	100	999
		0,000	Rejeitar H₀	100	4999
		0,000	Rejeitar H₀	100	9999
		0,000	Rejeitar H₀	100	49999
EB_WBF_NAF x EB	0,016	0,339	Não rejeitar H ₀	0	999
		0,365	Não rejeitar H ₀	0	4999
		0,378	Não rejeitar H ₀	0	9999
		0,362	Não rejeitar H ₀	0	49999
CONTROLE x EB	0,137	0,061	Não rejeitar H ₀	0	999
		0,075	Não rejeitar H ₀	0	4999
		0,067	Não rejeitar H ₀	0	9999
		0,069	Não rejeitar H ₀	0	49999

5.1.2 Efeito principal do fator Tempo

Na Tabela 4 observa-se que na primeira análise utilizando 999 permutações possíveis, há diferença significativa em pelo menos um tempo (R = 0,051, p-valor = 0,043). A repetição

desta análise apontou que em 20% das vezes rejeitou-se a hipótese nula e em 80% não. Aumentando o número de permutações para 4999, a porcentagem de rejeição da hipótese nula cai para 18%, sendo que a primeira análise não indicou tal rejeição (p-valor = 0,057), com 9999 permutações a porcentagem cai para 12% e com 49999 permutações chega a 0%. Ou seja, nas 50 análises realizadas com 49999 permutações possíveis, em nenhuma houve rejeição da hipótese de não diferença entre os tempos. Os histogramas da Figura 13 ajudam a ver como o p-valor tende a ser maior do que o nível de significância adotado ($\alpha = 0,05$).

Tabela 4: Análise do nível descritivo amostral referente ao efeito principal do fator Tempo

Fator	R	p-valor 1ª análise	Decisão estatística	% rejeição H_0	Nº Permutações
Tempo	0,051	0,043	Rejeitar H_0	20	999
		0,057	Não rejeitar H_0	18	4999
		0,055	Não rejeitar H_0	12	9999
		0,053	Não rejeitar H_0	0	49999

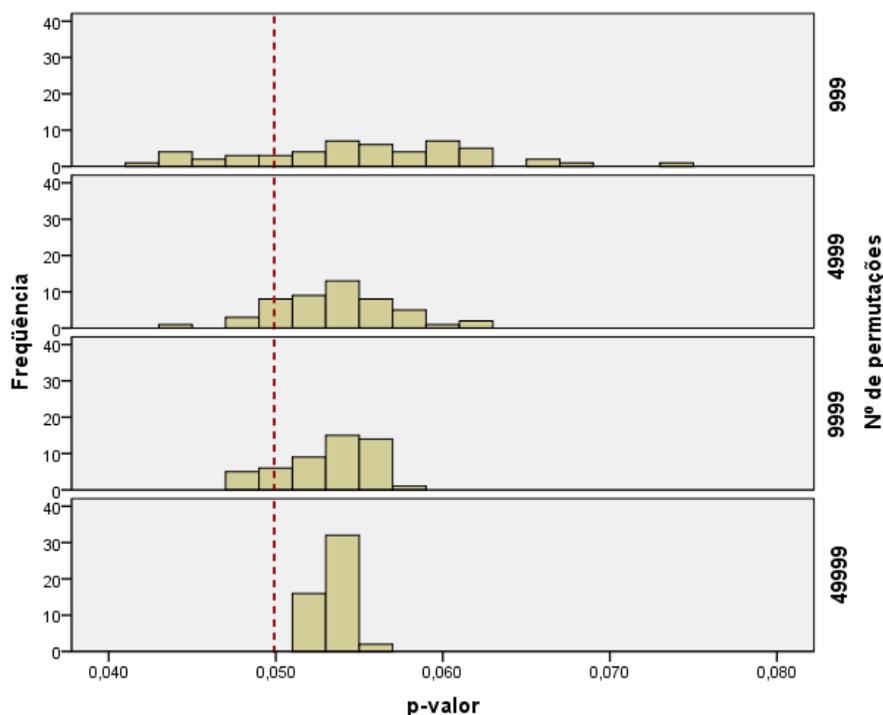


Figura 13: Histogramas do nível descritivo amostral referente ao efeito principal do fator Tempo

5.1.2.1 Comparações múltiplas

Com relação as comparações do fator Tempo, na Tabela 5 observa-se que na primeira análise realizada com 999 permutações há diferença significativa entre os Tempos 1 e 2 ($R =$

0,076; p-valor = 0,009), assim como quando realizada com 4999 (p-valor = 0,011), 9999 (p-valor = 0,009) e 49999 permutações (p-valor = 0,008). Os demais tempos não diferem entre si. Em 100% das análises, esta comparação apresentou diferença significativa, tanto com 999, 4999, 9999 e 49999 permutações.

Tabela 5: Análise do nível descritivo amostral referente às comparações múltiplas dentro do fator Tempo

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
Tempo 1 – 2	0,076	0,009	Rejeitar H₀	100	999
		0,011	Rejeitar H₀	100	4999
		0,009	Rejeitar H₀	100	9999
		0,008	Rejeitar H₀	100	49999
Tempo 1 – 3	-0,13	0,546	Não rejeitar H ₀	0	999
		0,561	Não rejeitar H ₀	0	4999
		0,554	Não rejeitar H ₀	0	9999
		0,558	Não rejeitar H ₀	0	49999
Tempo 2 – 3	0,043	0,227	Não rejeitar H ₀	0	999
		0,227	Não rejeitar H ₀	0	4999
		0,221	Não rejeitar H ₀	0	9999
		0,223	Não rejeitar H ₀	0	49999

5.2 DELINEAMENTO UNIFATORIAL

Como forma de identificar um possível efeito de interação, o banco de dados foi separado em três tempos, ou seja, foi testado o efeito do fator Fluido dentro dos tempos 1, 2 e 3, separadamente. Da mesma forma também foi dividido em cinco partes para analisar o efeito do fator Tempo dentro de cada um dos cinco tipos de fluídos.

5.2.1 Fator fluido dentro do Tempo 1

Na Tabela 6 observa-se que na primeira análise dentro do Tempo 1 não há diferença significativa com relação ao fator Fluido, tanto com 999 (R = -0,012, p-valor = 0,557), 4999 (p-valor = 0,554), 9999 (p-valor = 0,551) e 49999 (p-valor = 0,552). Em nenhuma das análises realizadas constatou-se diferença significativa entre os fluídos, tanto com 999, 4999, 9999 e 49999 permutações.

Tabela 6: Análise do nível descritivo amostral referente ao efeito do fator Fluído dentro do Tempo 1

Tempo	R	p-valor 1ª análise	Decisão estatística	% rejeição H_0	Nº Permutações
1	-0,012	0,557	Não rejeitar H_0	0	999
		0,554	Não rejeitar H_0	0	4999
		0,551	Não rejeitar H_0	0	9999
		0,552	Não rejeitar H_0	0	49999

5.2.1.1 Comparações múltiplas do fator fluído dentro do Tempo 1

A Tabela 7 apresenta as comparações múltiplas do fator Fluído dentro do Tempo 1, realizadas com 999, 4999, 9999 e 49999 permutações. Na primeira análise realizada com 999 permutações, não se observou comparações significativas. Com 4999, 9999 e 49999 permutações também não foram observadas comparações significativas. Em todas as análises realizadas, 100% das comparações foram não significativas, o que era esperado uma vez que os testes globais não indicaram diferença entre os Fluídos dentro deste tempo.

Tabela 7: Análise do nível descritivo amostral referente às comparações múltiplas do fator Fluído dentro do Tempo 1

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
WBF x WBF_NAF	0,078	0,204	Não rejeitar H ₀	0	999
		0,190	Não rejeitar H ₀	0	4999
		0,203	Não rejeitar H ₀	0	9999
		0,203	Não rejeitar H ₀	0	49999
WBF x EB_WBF_NAF	-0,068	0,724	Não rejeitar H ₀	0	999
		0,713	Não rejeitar H ₀	0	4999
		0,714	Não rejeitar H ₀	0	9999
		0,714	Não rejeitar H ₀	0	49999
WBF x CONTROLE	0,044	0,365	Não rejeitar H ₀	0	999
		0,342	Não rejeitar H ₀	0	4999
		0,337	Não rejeitar H ₀	0	9999
		0,342	Não rejeitar H ₀	0	49999
WBF x EB	-0,067	0,853	Não rejeitar H ₀	0	999
		0,853	Não rejeitar H ₀	0	4999
		0,853	Não rejeitar H ₀	0	9999
		0,853	Não rejeitar H ₀	0	49999
WBF_NAF x EB_WBF_NAF	-0,008	0,488	Não rejeitar H ₀	0	999
		0,493	Não rejeitar H ₀	0	4999
		0,490	Não rejeitar H ₀	0	9999
		0,487	Não rejeitar H ₀	0	49999
WBF_NAF x CONTROLE	0,017	0,398	Não rejeitar H ₀	0	999
		0,383	Não rejeitar H ₀	0	4999
		0,393	Não rejeitar H ₀	0	9999
		0,389	Não rejeitar H ₀	0	49999
EBF_NAF x EB	-0,026	0,561	Não rejeitar H ₀	0	999
		0,561	Não rejeitar H ₀	0	4999
		0,558	Não rejeitar H ₀	0	9999
		0,558	Não rejeitar H ₀	0	49999
EB_WBF_NAF x CONTROLE	0,006	0,398	Não rejeitar H ₀	0	999
		0,421	Não rejeitar H ₀	0	4999
		0,415	Não rejeitar H ₀	0	9999
		0,413	Não rejeitar H ₀	0	49999
EB_WBF_NAF x EB	-0,077	0,742	Não rejeitar H ₀	0	999
		0,746	Não rejeitar H ₀	0	4999
		0,730	Não rejeitar H ₀	0	9999
		0,730	Não rejeitar H ₀	0	49999
CONTROLE x EB	-0,075	0,672	Não rejeitar H ₀	0	999
		0,683	Não rejeitar H ₀	0	4999
		0,680	Não rejeitar H ₀	0	9999
		0,679	Não rejeitar H ₀	0	49999

5.2.2 Fator fluído dentro do Tempo 2

Com relação a primeira análise realizada dentro do Tempo 2, na Tabela 8 observa-se que o fator Fluído foi significativo tanto com 999 (R = 0,195, p-valor = 0,001), 4999 (p-valor

< 0,001), 999 (p-valor = 0,001) e 49999 permutações (p-valor < 0,001). Em 100% das análises realizadas, tanto com 999, 4999, 9999 e 49999 permutações, houve diferença significativa entre os fluídos.

Tabela 8: Análise do nível descritivo amostral referente ao efeito do fator Fluído dentro do Tempo 2

Tempo	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
2	0,195	0,001	Rejeitar H ₀	100	999
		0,000	Rejeitar H ₀	100	4999
		0,001	Rejeitar H ₀	100	9999
		0,000	Rejeitar H ₀	100	49999

5.2.2.1 Comparações múltiplas do fator fluído dentro do Tempo 2

Analisando as comparações múltiplas dentro do Tempo 2 (Tabela 9), a primeira análise realizada com 999 permutações indica que o fluído EB_WBF_NAF difere significativamente dos fluídos WBF (R = 0,255, p-valor=0,007), WBF_NAF (R = 0,289, p-valor = 0,007) e CONTROLE (R = 0,633, p-valor = 0,001) e o fluído CONTROLE difere do EB (R = 0,299, p-valor = 0,023). A primeira análise realizada com 4999, 9999 e 49999 permutações manteve significativas as mesmas comparações das apresentadas com 999 permutações. Em 100% das análises realizadas, as quatro comparações mantiveram-se significativas e os demais níveis não se diferiram.

Tabela 9: Análise do nível descritivo amostral referente às comparações múltiplas do fator Flúido dentro do Tempo 2

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H_0	Nº Permutações
WBF x WBF_NAF	-0,055	0,771	Não rejeitar H_0	0	999
		0,773	Não rejeitar H_0	0	4999
		0,763	Não rejeitar H_0	0	9999
		0,769	Não rejeitar H_0	0	49999
WBF x EB_WBF_NAF	0,255	0,007	Rejeitar H_0	100	999
		0,008	Rejeitar H_0	100	4999
		0,007	Rejeitar H_0	100	9999
		0,008	Rejeitar H_0	100	49999
WBF x CONTROLE	0,069	0,082	Não rejeitar H_0	0	999
		0,076	Não rejeitar H_0	0	4999
		0,075	Não rejeitar H_0	0	9999
		0,071	Não rejeitar H_0	0	49999
WBF x EB	0,155	0,081	Não rejeitar H_0	0	999
		0,084	Não rejeitar H_0	0	4999
		0,079	Não rejeitar H_0	0	9999
		0,071	Não rejeitar H_0	0	49999
WBF_NAF x EB_WBF_NAF	0,289	0,007	Rejeitar H_0	100	999
		0,005	Rejeitar H_0	100	4999
		0,004	Rejeitar H_0	100	9999
		0,004	Rejeitar H_0	100	49999
WBF_NAF x CONTROLE	0,082	0,145	Não rejeitar H_0	0	999
		0,157	Não rejeitar H_0	0	4999
		0,172	Não rejeitar H_0	0	9999
		0,165	Não rejeitar H_0	0	49999
EBF_NAF x EB	0,023	0,299	Não rejeitar H_0	0	999
		0,326	Não rejeitar H_0	0	4999
		0,311	Não rejeitar H_0	0	9999
		0,314	Não rejeitar H_0	0	49999
EB_WBF_NAF x CONTROLE	0,633	0,001	Rejeitar H_0	100	999
		0,000	Rejeitar H_0	100	4999
		0,000	Rejeitar H_0	100	9999
		0,000	Rejeitar H_0	100	49999
EB_WBF_NAF x EB	0,114	0,112	Não rejeitar H_0	0	999
		0,105	Não rejeitar H_0	0	4999
		0,112	Não rejeitar H_0	0	9999
		0,112	Não rejeitar H_0	0	49999
CONTROLE x EB	0,299	0,023	Rejeitar H_0	100	999
		0,023	Rejeitar H_0	100	4999
		0,019	Rejeitar H_0	100	9999
		0,022	Rejeitar H_0	100	49999

5.2.3 Fator flúido dentro do Tempo 3

Na Tabela 10 observa-se que dentro do Tempo 3 a primeira análise não apresentou diferença significativa entre os flúidos, tanto com 999 ($R = 0,089$, p -valor = 0,06), 4999 (p -

valor = 0,053), 9999 (p-valor = 0,054) e 49999 permutações (p-valor=0,055). Entretanto 22% das repetições da análise com 999 permutações apresentaram rejeição da hipótese nula de igualdade entre os fluídos e, com 4999, 9999 e 49999 permutações, nenhuma análise rejeitou a hipótese nula dentro do Tempo 3. A Figura 14 ajuda a ver como o p-valor tende a ser maior que o nível de significância adotado (5%) a partir de 4999 permutações.

Tabela 10: Análise do nível descritivo amostral referente ao efeito do fator Fluido dentro do Tempo 3

Tempo	R	p-valor 1ª análise	Decisão estatística	% rejeição H_0	Nº Permutações
3	0,089	0,06	Não rejeitar H_0	22	999
		0,053	Não rejeitar H_0	0	4999
		0,054	Não rejeitar H_0	0	9999
		0,055	Não rejeitar H_0	0	49999

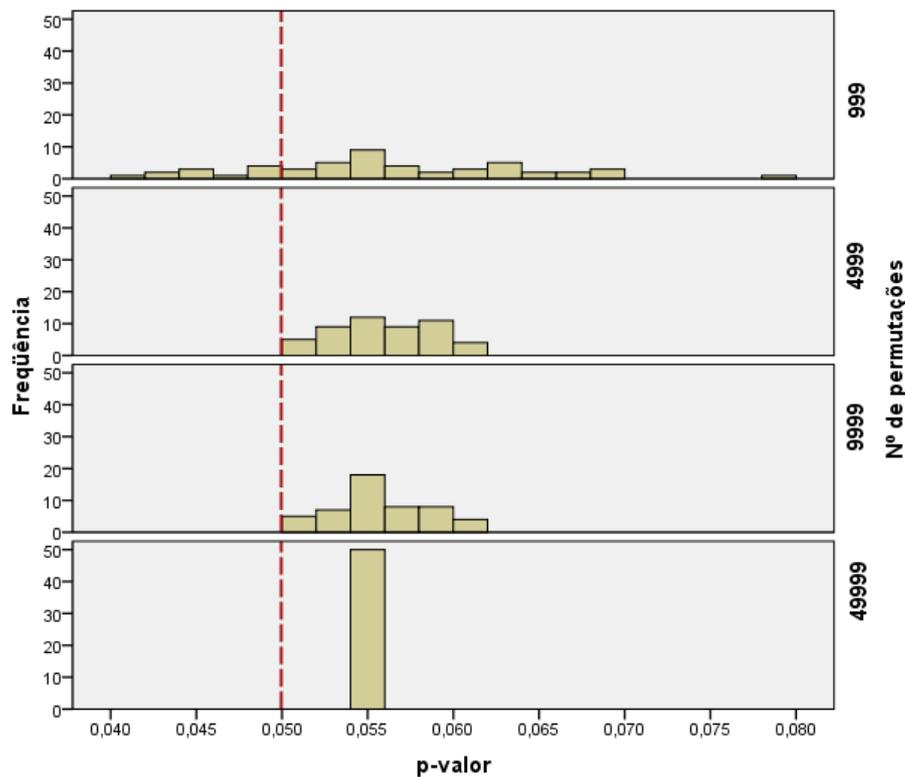


Figura 14: Histogramas do nível descritivo amostral do fator Fluido dentro do Tempo 3

5.2.3.1 Comparações múltiplas do fator fluido dentro do Tempo 3

Com relação as comparações dentro do Tempo 3 (Tabela 11), a primeira análise com 999 permutações indica que o fluido CONTROLE difere significativamente dos fluídos

WBF_NAF (R = 0,287, p-valor = 0,006) e EB (R = 0,578, p-valor=0,006). A primeira análise realizada com 4999, 9999 e 49999 permutações manteve significativas as mesmas comparações das apresentadas com 999 permutações. Em 100% das análises realizadas, as duas comparações mantiveram-se significativas e os demais níveis não se diferiram.

Tabela 11: Análise do nível descritivo amostral referente às comparações múltiplas do fator Flúido dentro do Tempo 3

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº permutações
WBF x WBF_NAF	0,016	0,379	Não rejeitar H ₀	0	999
		0,381	Não rejeitar H ₀	0	4999
		0,381	Não rejeitar H ₀	0	9999
		0,381	Não rejeitar H ₀	0	49999
WBF x EB_WBF_NAF	-0,058	0,757	Não rejeitar H ₀	0	999
		0,738	Não rejeitar H ₀	0	4999
		0,729	Não rejeitar H ₀	0	9999
		0,729	Não rejeitar H ₀	0	49999
WBF x CONTROLE	0,126	0,059	Não rejeitar H ₀	0	999
		0,077	Não rejeitar H ₀	0	4999
		0,071	Não rejeitar H ₀	0	9999
		0,071	Não rejeitar H ₀	0	49999
WBF x EB	-0,075	0,648	Não rejeitar H ₀	0	999
		0,648	Não rejeitar H ₀	0	4999
		0,648	Não rejeitar H ₀	0	9999
		0,648	Não rejeitar H ₀	0	49999
WBF_NAF x EB_WBF_NAF	0,139	0,082	Não rejeitar H ₀	0	999
		0,076	Não rejeitar H ₀	0	4999
		0,076	Não rejeitar H ₀	0	9999
		0,076	Não rejeitar H ₀	0	49999
WBF_NAF x CONTROLE	0,287	0,006	Rejeitar H₀	100	999
		0,005	Rejeitar H₀	100	4999
		0,005	Rejeitar H₀	100	9999
		0,005	Rejeitar H₀	100	49999
EBF_NAF x EB	-0,194	0,929	Não rejeitar H ₀	0	999
		0,929	Não rejeitar H ₀	0	4999
		0,929	Não rejeitar H ₀	0	9999
		0,929	Não rejeitar H ₀	0	49999
EB_WBF_NAF x CONTROLE	0,044	0,218	Não rejeitar H ₀	0	999
		0,240	Não rejeitar H ₀	0	4999
		0,237	Não rejeitar H ₀	0	9999
		0,237	Não rejeitar H ₀	0	49999
EB_WBF_NAF x EB	0,046	0,333	Não rejeitar H ₀	0	999
		0,333	Não rejeitar H ₀	0	4999
		0,333	Não rejeitar H ₀	0	9999
		0,333	Não rejeitar H ₀	0	49999
CONTROLE x EB	0,578	0,006	Rejeitar H₀	100	999
		0,006	Rejeitar H₀	100	4999
		0,006	Rejeitar H₀	100	9999
		0,006	Rejeitar H₀	100	49999

5.2.4 Fator tempo dentro do Fluido EB

Na Tabela 12 observa-se que na primeira amostra realizada dentro do Fluido EB não há diferença significativa entre os Tempos tanto com 999 ($R = 0,008$, $p\text{-valor} = 0,396$), 4999 ($p\text{-valor} = 0,402$), 9999 ($p\text{-valor} = 0,403$) e 49999 ($p\text{-valor} = 0,402$). Em todas as análises realizadas dentro deste fluido, nenhuma apresentou diferença significativa.

Tabela 12: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluido EB

Fluido	R	p-valor 1ª análise	Decisão estatística	% rejeição H_0	Nº Permutações
EB	0,008	0,396	Não rejeitar H_0	0	999
		0,402	Não rejeitar H_0	0	4999
		0,403	Não rejeitar H_0	0	9999
		0,402	Não rejeitar H_0	0	49999

5.2.4.1 Comparações múltiplas do fator tempo dentro do Fluido EB

A Tabela 13 apresenta as comparações múltiplas do fator Tempo dentro do Fluido EB realizadas com 999, 4999, 9999 e 49999 permutações. Em todas as análises realizadas, nenhuma comparação foi significativa para a igualdade de tempos, o que era esperado uma vez que os testes globais não indicaram diferença entre os Tempos dentro deste Fluido.

Tabela 13: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluido EB

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H_0	Nº Permutações
Tempos 1 – 2	0,061	0,176	Não rejeitar H_0	0	999
		0,203	Não rejeitar H_0	0	4999
		0,203	Não rejeitar H_0	0	9999
		0,203	Não rejeitar H_0	0	49999
Tempos 1 – 3	0,204	0,100	Não rejeitar H_0	0	999
		0,100	Não rejeitar H_0	0	4999
		0,100	Não rejeitar H_0	0	9999
		0,100	Não rejeitar H_0	0	49999
Tempos 2 - 3	-0,122	0,803	Não rejeitar H_0	0	999
		0,803	Não rejeitar H_0	0	4999
		0,803	Não rejeitar H_0	0	9999
		0,803	Não rejeitar H_0	0	49999

5.2.5 Fator tempo dentro do Fluido WBF_NAF

Na Tabela 14 observa-se que para a primeira análise realizada dentro do Fluido WBF_NAF não há diferença significativa, tanto com 999 ($R = 0,048$, $p\text{-valor} = 0,233$),

4999(p-valor = 0,239), 9999 (p-valor = 0,242) e 49999 permutações (p-valor = 0,246). Em todas as análises realizadas dentro deste fluido, nenhuma apresentou diferença significativa.

Tabela 14: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluido WBF_NAF

Fluido	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
WBF_NAF	0,048	0,233	Não rejeitar H ₀	0	999
		0,239	Não rejeitar H ₀	0	4999
		0,242	Não rejeitar H ₀	0	9999
		0,246	Não rejeitar H ₀	0	49999

5.2.5.1 Comparações múltiplas do fator Tempo dentro do Fluido WBF_NAF

Analisando o comportamento das comparações múltiplas dentro do fluido WBF_NAF (Tabela 15), a primeira análise realizada com 999 permutações indica que há diferença significativa entre os tempos 1 e 3 (R = 0,341, p-valor = 0,018), assim como com 4999 (p-valor = 0,017), 9999 (p-valor = 0,017) e 49999 permutações (p-valor = 0,017). 100% das análises realizadas com 999, 4999, 9999 e 49999 permutações apresentaram diferença significativa entre os Tempos 1 e 3. Os demais tempos não diferiram entre si.

Tabela 15: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluido WBF_NAF

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº permutações
Tempos 1 – 2	-0,1	0,965	Não rejeitar H ₀	0	999
		0,970	Não rejeitar H ₀	0	4999
		0,972	Não rejeitar H ₀	0	9999
		0,971	Não rejeitar H ₀	0	49999
Tempos 1 – 3	0,341	0,018	Rejeitar H₀	100	999
		0,017	Rejeitar H₀	100	4999
		0,017	Rejeitar H₀	100	9999
		0,017	Rejeitar H₀	100	49999
Tempos 2—3	0,156	0,167	Não rejeitar H ₀	0	999
		0,158	Não rejeitar H ₀	0	4999
		0,153	Não rejeitar H ₀	0	9999
		0,153	Não rejeitar H ₀	0	49999

5.2.6 Fator tempo dentro do Fluido EB_WBF_NAF

Na Tabela 16 observa-se que na primeira análise realizada dentro do fluido EB_WBF_NAF há diferença significativa entre os Tempos, tanto com 999(R = 0,317, p-valor = 0,001), 4999 (p-valor < 0,001), 9999 (p-valor < 0,001) e 49999 permutações (p-valor <

0,001). Em todas as análises realizadas, 100% delas indicaram diferença entre os Tempos dentro deste fluido.

Tabela 16: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluido EB_WBF_NAF

Fluido	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
EB_WBF_NAF	0,317	0,001	Rejeitar H ₀	100	999
		0,000	Rejeitar H ₀	100	4999
		0,000	Rejeitar H ₀	100	9999
		0,000	Rejeitar H ₀	100	49999

5.2.6.1 Comparações múltiplas do fator tempo dentro do Fluido EB_WBF_NAF

Analisando o comportamento das comparações múltiplas dentro do fluido EB_WBF_NAF (Tabela 17), a primeira análise realizada com 999 permutações indica que há diferença significativa do Tempo 2 com o Tempo 1 (R = 0,530, p-valor = 0,001) e com o Tempo 3 (R = 0,531, p-valor = 0,001), assim como com 4999 (p-valor < 0,001 e 0,001), 9999 (p-valor < 0,001 e 0,001) e 49999 permutações (p-valor < 0,001 e 0,001). 100% das análises realizadas com 999, 4999, 9999 e 49999 permutações apresentaram diferença significativa do Tempo 2 com o Tempo 1 e 3. Os demais tempos não diferiram entre si.

Tabela 17: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluido EB_WBF_NAF

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
Tempos 1 – 2	0,53	0,001	Rejeitar H ₀	100	999
		0,000	Rejeitar H ₀	100	4999
		0,000	Rejeitar H ₀	100	9999
		0,000	Rejeitar H ₀	100	49999
Tempos 1 – 3	-0,037	0,693	Não rejeitar H ₀	0	999
		0,676	Não rejeitar H ₀	0	4999
		0,674	Não rejeitar H ₀	0	9999
		0,68	Não rejeitar H ₀	0	49999
Tempos 2—3	0,531	0,001	Rejeitar H ₀	100	999
		0,000	Rejeitar H ₀	100	4999
		0,000	Rejeitar H ₀	100	9999
		0,000	Rejeitar H ₀	100	49999

5.2.7 Fator tempo dentro do Fluido WBF

A primeira amostra realizada dentro do Fluido WBF (Tabela 18) não apresentou diferença significativa entre os Tempos tanto com 999 (R = -0,067, p-valor = 0,875), 4999 (p-

valor = 0,862), 9999 (p-valor = 0,860) e 49999 permutações (p-valor = 0,858). Em todas as análises realizadas dentro deste fluido, nenhuma apresentou diferença significativa.

Tabela 18: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluido WBF

Fluido	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
WBF	-0,067	0,875	Não rejeitar H ₀	0	999
		0,862	Não rejeitar H ₀	0	4999
		0,860	Não rejeitar H ₀	0	9999
		0,858	Não rejeitar H ₀	0	49999

5.2.7.1 Comparações múltiplas do fator tempo dentro do Fluido WBF

Analisando o comportamento das comparações múltiplas dentro do fluido WBF (Tabela 19), a primeira análise realizada com 999 permutações indica que não há diferença significativa entre os tempos. Em todas as análises realizadas, nenhuma comparação foi significativa para a igualdade de tempos, o que era esperado uma vez que os testes globais não indicaram diferença entre os Tempos dentro deste Fluido.

Tabela 19: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluido WBF

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
Tempos 1 – 2	-0,127	0,902	Não rejeitar H ₀	0	999
		0,91	Não rejeitar H ₀	0	4999
		0,917	Não rejeitar H ₀	0	9999
		0,917	Não rejeitar H ₀	0	49999
Tempos 1 – 3	0,016	0,408	Não rejeitar H ₀	0	999
		0,395	Não rejeitar H ₀	0	4999
		0,395	Não rejeitar H ₀	0	9999
		0,395	Não rejeitar H ₀	0	49999
Tempos 2—3	-0,061	0,805	Não rejeitar H ₀	0	999
		0,796	Não rejeitar H ₀	0	4999
		0,795	Não rejeitar H ₀	0	9999
		0,797	Não rejeitar H ₀	0	49999

5.2.8 Fator tempo dentro do Fluido CONTROLE

Na Tabela 20 observa-se que na primeira análise realizada dentro do Fluido CONTROLE não há diferença significativa entre os Tempos, tanto com 999 (R = -0,03, p-valor=0,728), 4999(p-valor = 0,700), 9999 (p-valor = 0,703) e 49999 permutações (p-valor = 0,705). Em todas as análises realizadas dentro deste tratamento, nenhuma apresentou diferença significativa.

Tabela 20: Análise do nível descritivo amostral referente ao efeito do fator Tempo dentro do Fluido CONTROLE

Fluido	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
CONTROLE	-0,03	0,728	Não rejeitar H ₀	0	999
		0,700	Não rejeitar H ₀	0	4999
		0,703	Não rejeitar H ₀	0	9999
		0,705	Não rejeitar H ₀	0	49999

5.2.8.1 Comparações múltiplas do fator tempo dentro do Fluido CONTROLE

Analisando o comportamento das comparações múltiplas dentro do fluido CONTROLE (Tabela 21), a primeira análise realizada com 999 permutações indica que há diferença significativa do Tempo 1 com o Tempo 2 (R =0,062, p-valor = 0,046), assim como com 4999 (p-valor = 0,046) e 49999 permutações (p-valor = 0,049). Entretanto, a diferença observada na primeira análise realizada com 9999 permutações, não foi detectada diferença significativa (p-valor = 0,05). Com 999 permutações, das 50 análises realizadas, em 42% delas a hipótese de não diferença significativa entre os tempos 1 e 2 foi rejeitada; com 4999 permutações, 44% das análises, com 9999 permutações, 44% das análises e com 49999 permutações, em 60% a hipótese de não diferença significativa foi rejeitada. A Figura 15 ajuda a ver a dispersão dos p-valores encontrados.

Tabela 21: Análise do nível descritivo amostral referente às comparações múltiplas do fator Tempo dentro do fluido CONTROLE

Comparação	R	p-valor 1ª análise	Decisão estatística	% rejeição H ₀	Nº Permutações
Tempos 1 – 2	0,062	0,046	Rejeitar H₀	42	999
		0,046	Rejeitar H₀	44	4999
		0,05	Não rejeitar H ₀	44	9999
		0,049	Rejeitar H₀	60	49999
Tempos 1 – 3	-0,159	0,916	Não rejeitar H ₀	0	999
		0,920	Não rejeitar H ₀	0	4999
		0,920	Não rejeitar H ₀	0	9999
		0,920	Não rejeitar H ₀	0	49999
Tempos 2—3	-0,133	0,878	Não rejeitar H ₀	0	999
		0,875	Não rejeitar H ₀	0	4999
		0,879	Não rejeitar H ₀	0	9999
		0,878	Não rejeitar H ₀	0	49999

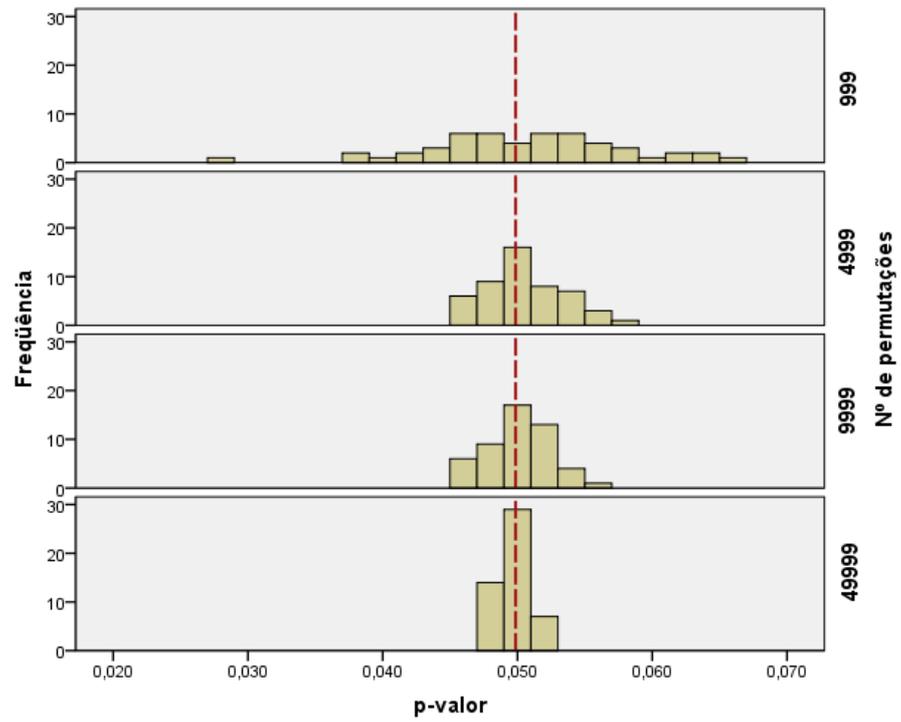


Figura 15: Histograma do nível descritivo amostral das comparações múltiplas entre os tempos 1 e 2

6. CONCLUSÃO

É esperado que uma técnica estatística consistente e robusta apresente resultados (estatística e nível descritivo amostral) que levem a mesma decisão estatística, o que não tem se evidenciado neste caso abordado.

No delineamento com dois fatores, se a análise é realizada apenas com 999 permutações, que é o que a literatura sugere como número mínimo de permutações para ter uma significância de 5%, constatou-se que das 50 vezes que a análise foi repetida, em 10 dessas repetições observou-se um p-valor significativo e em 40 não para o efeito principal do fator Tempo. Ou seja, a decisão estatística de confirmar ou não que há diferença entre os tempos de coleta está dependendo da escolha do número de permutações para a análise e posterior decisão, uma vez que essas 999 permutações nada mais são que o tamanho da amostra utilizada para gerar o nível descritivo amostral de cada análise. Como cada repetição dessa análise seleciona permutações diferentes, a consequência é gerar p-valores distintos. Entretanto, se o número de permutações realizadas é alterado para 49999, em nenhuma das repetições da análise foi encontrado um nível descritivo amostral significativo. O que evidencia que o tamanho mínimo de 999 permutações não era representativo da população de todas as permutações possíveis para determinada amostra.

É possível destacar ainda com relação ao fator Tempo, que em todas as repetições das análises de comparações múltiplas realizadas tanto com 999, 4999, 9999 e 49999 permutações, independente da análise global rejeitar ou não a hipótese de igualdade entre os tempos, os tempos 1 e 2 são significativamente distintos, com nível descritivo amostral abaixo de 1%, o que se caracteriza como uma incongruência entre o teste global e as comparações múltiplas.

Outro ponto a ser observado, ainda com relação ao delineamento fatorial, é o fato de não ser possível testar o efeito conjunto dos fatores (interação). Interações entre fatores são muito comuns e detectar este efeito é a grande vantagem deste tipo de delineamento, constatando-se desta forma, que a análise de similaridades ainda é carente neste tipo de delineamento.

Com relação ao delineamento unifatorial, a análise do fator fluído observado no terceiro período de coleta de dados (Tempo 3), das 50 análises realizadas com 999 permutações, 11 delas rejeitaram a hipótese nula de igualdade entre os tipos de fluídos.

Constata-se que com 49999 permutações, todas as permutações possíveis foram realizadas e, por consequência, o nível descritivo amostral verdadeiro para esta análise dentro do Tempo 3 baseada nesta amostra de dados coletados é 0,055, ou seja, não há diferença entre os fluídos de perfuração no 3º tempo de coleta. Contudo, a análise de comparações múltiplas aponta que o fluído CONTROLE difere significativamente dos fluídos EB e WBF_NAF. Mais uma incongruência entre o teste global e a análise de comparações múltiplas.

Por fim, a análise de diferença entre os tempos de coleta dentro do fluído WBF_NAF, 100% das repetições desta análise não indicaram diferença significativa, entretanto a análise de comparações múltiplas aponta que há diferença significativa entre os tempos 1 e 3.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- Clarke, K.R., Warwick, R.M., Similarity-based testing for community pattern: the two-way layout with no replication. Plymouth Marine Laboratory, 1993
- Clarke, K.R., Warwick, R.M., Change in marine communities: an approach to statistical analysis and interpretation. Plymouth Marine Laboratory, 1994
- Clarke, K.R., Gorley R.N. (2001 & 2006) PRIMER v5 (& v6): User manual/tutorial, PRIMER-E, Plymouth UK
- Clarke, K.R., Green, R.H. 1988. Statistical design and analysis for a “biological effects” study. *Mar Ecol Prog Ser* 46: 213-226
- Chapman M.G., Underwood A.J. 1999. Ecological patterns in multivariate assemblages: information and interpretation of negative values in ANOSIM tests. *Mar Ecol Prog Ser* 180: 257-265.
- Hope, A.C.A. 1968. A simplified Monte Carlo significance test procedure. *J. Roy. Stat. Soc. Ser. B* 30: 582-598
- Masi B.P. & Zalmon I.R. 2008. Zonação de comunidade bêntica do entremarés em molhes sobre diferente hidrodinamismo na costa norte do estado do Rio de Janeiro, Brasil. *Revista brasileira de zoologia* 25: 662-673.
- Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209-220.
- Marti J.A. 2001. A new method for non-parametric multivariate analysis of variance. *Aus Eco* 26: 32-49
- Montgomery D. C. 2001. Design and analysis of experiments.
- E.E. Toldo Jr.; R.N. Ayup Zouain; I.C.S. Corrêa; M.C.R. Perlaba; C.M.D.S. Freitas; J.M.G. Fachel & E.J.S. Sierra. 2004 Monitoramento Ambiental em Atividades de Perfuração Exploratória Marítima – MAPEM.