

**Universidade de São Paulo
Instituto de Matemática e Estatística**

Centro de Estatística Aplicada

Relatório de Análise Estatística

RAE-CEA–20P04

RELATÓRIO DE ANÁLISE ESTATÍSTICA SOBRE O PROJETO:

**Otimização da previsão de risco geológico utilizando a simulação de Monte Carlo
aplicado às estruturas geológicas**

Anatoly Yambarstev

Bruno de Assis Silva

Victor Junji Takara

São Paulo, julho de 2020

CENTRO DE ESTATÍSTICA APLICADA - CEA – USP

TÍTULO: Relatório de Análise Estatística sobre o Projeto: “Otimização da previsão de risco geológico utilizando a simulação de Monte Carlo aplicado às estruturas geológicas”.

PESQUISADORA: Bruna Catarino Xavier

ORIENTADOR: Prof. Dr. Marcos Egydio da Silva

INSTITUIÇÃO: Instituto de Geociências da USP

FINALIDADE DO PROJETO: Doutorado

RESPONSÁVEIS PELA ANÁLISE: Anatoly Yambarstev

Bruno de Assis Silva

Victor Junji Takara

REFERÊNCIA DESTE TRABALHO: SILVA, B.A.; TAKARA, V.J.; YAMBARSTEV, A.

Relatório de análise estatística sobre o projeto: “Otimização da previsão de risco geológico utilizando a simulação de Monte Carlo aplicado às estruturas geológicas”. São Paulo, IME-USP, 2020. (RAE–CEA-20P04)

FICHA TÉCNICA

REFERÊNCIAS BIBLIOGRÁFICAS:

ANDERSON, M.J. (2001) A new method for non-parametric multivariate analysis of variance, **Austral Ecology**, 26, 32–46.

ANDERSON, M.J.; WALSH, D.C.I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing?, **Ecological Monographs**, 83(4), pp. 557–574.

EDGINGTON, E. S. (1995). **Randomization Tests**, 3rd edn. Marcel Dekker, New York.

HÄRDLE, W.; SIMAR, L. (2012). **Applied Multivariate Analysis**, Springer, 3rd edition, p.193-200.

JOHNSON, R.A.; WISCHERN, D.W. (2007). **Applied Multivariate Statistical Analysis**, Pearson Prentice Hall, 6th edition, p. 284-312.

MANLY, B. F. J. (1997). **Randomization, Bootstrap and Monte Carlo Methods in Biology**, 2nd edn. Chapman & Hall, London.

KRUEGER, J.I.; PATRICK, R.H. (2019). Putting the p-value in its Place, **The American Statistician**, vol.73, NO. S1, 122-128: Statistical Inference in the 21st Century.

MARDIA, K.V.; KENT, J.T.; BIBBY, J.M. (1995). **Multivariate Analysis**, Academic Press, 2nd edition.

PEWSEY, A.; NEUHÄUSER, M.; RUXTON, G.D. (2013). **Circular Statistics in R**. 1.ed. Oxford, 198p.

VEYRET, Y. (2007). **Os Riscos – O homem como agressor e vítima do meio ambiente**. 1.ed., Contexto, 322p.

PROGRAMAS COMPUTACIONAIS UTILIZADOS:

Microsoft Word for Windows (versão 2016)

Microsoft Excel for Windows (versão 2016)

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Multidimensional (03:020)

Análise de Variância Não Paramétrica (08:050)

Análise de Variância Multivariada (08:110)

Outros(08:990)

ÁREA DE APLICAÇÃO

Ciências Físicas e Geoestatística (14:010)

Resumo

Variáveis geológicas são frequentemente coletadas para o estudo do risco de uma obra a ser realizada. A análise contida neste relatório estatístico teve como objetivo analisar se os centroides dos Cossenos diretores (transformações das variáveis Dip e Dipdirection) coletados na superfície e em outras profundidades de um maciço rochoso são similares para as estruturas de foliação e fratura. Uma vez que o ajuste pelo modelo MANOVA não teve a hipótese de normalidade satisfeita, foi realizada uma PERMANOVA e foram encontradas diferenças significativas entre os grupos. Com o intuito de se analisar a amplitude dessas diferenças, foram utilizados resultados assintóticos contidos em MARDIA et al (2013) e, assim, construídas regiões de confiança para os centroides por um método de simulação baseado em HARDLE e SILMAR (2012). A partir dessas regiões, verificou-se que as diferenças no geral não são grandes do ponto de vista prático e que a diferença na estrutura de fratura é menor do que nas estrutura de foliação.

Sumário

2. Objetivo	9
3. Descrição do estudo	9
4. Descrição das variáveis	10
5. Análise descritiva	10
6. Análise inferencial	12
7. Conclusões.....	13
APÊNDICE A	15
APÊNDICE B	17

1. Introdução

Para serem objetos de gestão aplicada, os riscos necessitam ser calculáveis, ou seja, é necessária a definição do par probabilidade/consequência ou frequência/gravidade para que sejam estabelecidos e geridos (VEYRET, 2007).

Quando uma obra de infraestrutura, sobretudo de grande porte, é planejada, há a necessidade de entendimento desses riscos, sejam em relação à vida, ao ambiente ou às implicações financeiras, a fim de que uma boa gestão desse risco possa ser realizada, podendo até mesmo inviabilizar obras. Visando o entendimento da estabilidade das obras, duas medidas das estruturas geológicas podem ser analisadas a partir de sua posição espacial, valores de direção (*dipdirection*) e mergulho (*dip*), as quais estão relacionadas às condições de risco do maciço rochoso.

Dessa forma, são feitas diversas escavações em diferentes regiões do maciço rochoso para se determinar os valores dessas variáveis em diversas profundidades. No entanto, isso pode ser bem custoso e, por vezes, até mesmo inviável, dadas as características do local.

2. Objetivo

O objetivo do estudo é averiguar se os comportamentos médios das variáveis associadas ao risco de deslizamento da obra (*Dip* e *Dipdirection* nas suas transformações em cossenos diretores X, Y, Z) são os mesmos em superfície ou em profundidades maiores.

3. Descrição do estudo

A coleta dos dados foi realizada num estudo de viabilidade para as obras das usinas hidrelétricas de Capim Branco I e de Capim Branco II, em Minas Gerais. Nesse estudo, as variáveis *Dip* e *Dipdirection* (ambas em graus) foram mensuradas com uma bússola geológica em diversos pontos e profundidades diferentes do maciço rochoso.

Por meio dessas duas variáveis combinadas, foram calculados cossenos diretores e convertidas em medidas adimensionais nas direções x,y,z, representadas na Figura B.2.

4. Descrição das variáveis

Variáveis que caracterizam a amostra:

- *Dip* (graus)
- *Dipdirection* (graus)
- Distância (a partir da superfície): 0m, de 5m a 45m, >45m
- Cota de profundidade (categórica, superfície se o a Distância é 0m, profundidade, caso contrário)
- Cossenos diretores x,y,z (adimensionais, variando entre [-1,1])
- Estrutura (fratura ou foliação)

5. Análise descritiva

Com base em PEWSEY et al.(2013), temos que nos atentar às variáveis *dip* e *dipdirection*, uma vez que são variáveis circulares, pois valores linearmente mais distantes podem estar mais próximos, como é o caso de ângulos próximos de 360° e de 0°.

Utilizando o gráfico *rose diagrams*, que tem interpretação análoga à do histograma para variáveis circulares, foram comparadas as variáveis circulares nos dados da superfície (Cota de profundidade de 0m) e nos dados em maior profundidade (Cota de profundidade superior a 0m) (Figuras B.3 e B.4).

Com base nesses gráficos, vemos uma similaridade entre a superfície e a profundidade tanto para a variável *dip* quanto para a variável *dipdirection*. Também é possível notar esta similaridade através das medidas descritivas, pois a média e desvio padrão do *dip* em superfície são semelhantes a essas medidas em profundidade, e o mesmo comportamento é observado para o *dipdirection*.(Tabelas A.1 e A.2). Observamos também que a quantidade de observações em profundidade é maior que em superfície.

Essas variáveis, contudo, devem ser analisadas de forma bivariada, uma vez que *dip* e *dipdirection* são duas componentes relacionadas que só fazem sentido do ponto de vista geológico quando analisadas conjuntamente. Essa representação para variáveis circulares, todavia, se faz complicada, pois seriam necessárias representações esféricas para a caracterização dos dados.

Uma vez que os cossenos diretores são transformações dos vetores compostos por *dip* e *dipdirection* que trazem também as informações desejadas para a análise do comportamento dos dados da superfície e da profundidade, optou-se por seguir com a análise através deles, havendo assim uma forma de representação multivariada descritiva.

Observando a tabela com as medidas descritivas dos cossenos diretores, notamos que essas medidas na superfície e em profundidade são semelhantes. (Tabelas A.3 e A4)

Inicialmente, verificamos pelos *box plots* que a componente x é a que apresenta maior variabilidade e z a menor. Há a presença de muitos *outliers* para a componente y, contudo, ressalta-se que as três componentes devem ser vistas de forma conjunta, por se tratarem de amostras de uma única variável aleatória multivariada (Figura B.5 e B.6). Esses *box plots* mostram também que há indícios de similaridade entre as distribuições oriundas da superfície e da profundidade.

Com base na análise do diagrama de dispersão dos dados tridimensionais para os cossenos diretores, verificamos também um padrão similar entre os dados em superfície e os dados em profundidade. (Figura B.7).

Foram também feitas análises por distância. Por meio das Figuras B.8 e B.9, vemos que os valores dos cossenos diretores nos três níveis de distância parecem ter o mesmo comportamento, tanto pelo diagrama de dispersão quanto pelos *box plots*. Esse resultado justifica o uso da variável profundidade ao invés da variável distância na análise dos dados.

Por meio das Figura B.10 a B.12, podemos visualizar que há um comportamento diferente dos dados nas estruturas de foliação e fratura. Vale ressaltar que o comportamento da superfície e da profundidade em cada estrutura parece o mesmo.

6. Análise inferencial

Devido às características dos dados, as interpretações devem ser feitas conjuntamente. Inicialmente aplicamos um modelo de MANOVA (JOHNSON e WICHERN, 2007) para os cossenos diretores, considerando estrutura e cota de profundidade como variáveis explicativas. No entanto, a suposição de normalidade não foi atendida, mesmo que a homocedasticidade tenha sido verificada.

Visto que a MANOVA não pode ser aplicada, utilizamos um modelo de PERMANOVA (do inglês *Permutational multivariate analysis of variance*), que é um modelo não-paramétrico para testar a hipótese de que os centroides dos grupos sejam iguais (ANDERSON, 2013).

Verificamos estatisticamente a existência de efeito de interação entre cota de profundidade e estrutura (valor- $p < 0,01$), ou seja, o efeito da Cota de profundidade não é o mesmo em cada estrutura. Também foram realizados testes do efeito de Cota de profundidade para cada estrutura, sendo, em ambos os casos, significativo (valor- $p < 0,01$).

Embora os resultados da PERMANOVA nos permitam dizer que existe diferença significativa entre os cossenos diretores de superfície e de profundidade do ponto de vista estatístico, ele não nos dá a magnitude dessa diferença do ponto de vista prático, o que foi entendido como um resultado incompleto para o que se desejava investigar. Visto que é sabido que o p-valor naturalmente decresce conforme o tamanho amostral aumenta (KRUEGER e HECK, 2019), optou-se por complementar o problema sob outra óptica.

Em MARDIA et al. (1995), é apresentada uma versão para o teorema central do limite multivariado, no qual o vetor de médias de uma sequência de vetores aleatórios

independentes tem distribuição assintótica normal multivariada. Uma vez que a amostra é grande, assumimos então que essa aproximação assintótica é razoável e, utilizando esse resultado, HARDLE e SILMAR (2012) nos apresenta um algoritmo para a construção de uma região de confiança fixado um nível de significância. Por meio de simulação, esses autores mostraram que a região de confiança com significância $\alpha\%$ significância é tal que:

$$A = \left\{ \mu \in \mathbb{R}^p \mid (\bar{X} - \mu)^T S_0^{-1} (\bar{X} - \mu) \leq \frac{p}{n-p} F_{1-\alpha; p; n-p} \right\}$$

Utilizando o resultado de HARDLE e SILMAR (2012) para os cossenos diretores, por cota de profundidade (Figura B.13), observa-se que não existem intersecções nas regiões de confiança, isto é, os resultados corroboram com a diferença significativa que concluímos no teste de PERMANOVA. Assim, podemos avaliar a magnitude dessa diferença. Comparando os elipsoides de superfície e profundidade, notamos que a maior diferença está no cosseno diretor x num intervalo de tamanho próximo de 0,1.

Separando os dados pelas estruturas foliação e fratura, temos o mesmo comportamento de que as diferenças são significativas conforme observamos anteriormente e que as regiões de confiança estão dentro de aproximadamente um cubo de tamanho 0,1. (Figuras B.14 e B.15)

7. Conclusões

Após as análises podemos ver que descritivamente não temos muitos indícios de que existem diferenças entre os cossenos diretores de profundidade e superfície. Contudo via o modelo de PERMANOVA obtemos que existe diferença significativa a 1% entre as duas cotas de profundidade. Mensurando esta diferença através da construção de regiões de confiança notamos que de forma geral as discrepâncias não são maiores que 0,1. Além disso, o tipo de estrutura parece ter efeito sobre como os cossenos diretores, pois nas fraturas a magnitude das diferenças são menores em comparação com as foliações.

APÊNDICE A

Tabelas

Tabela A.1 Medidas descritivas para *Dip* e *Dipdirection*, em Superfície

	<i>Dip</i>	<i>Dipdirection</i>
n	404	404
Média	59,03	208,71
Desvio padrão	20,33	95,86
Mínimo	4	8
Mediana	57	248
Máximo	89	355

Tabela A.2 Medidas descritivas para *Dip* e *Dipdirection*, em Profundidade

	<i>Dip</i>	<i>Dipdirection</i>
n	953	953
Média	62,04	212,66
Desvio padrão	18,76	94,12
Mínimo	15	2
Mediana	61	253
Máximo	92	365

Tabela A.3 Medidas descritivas para cossenos diretores, em Superfície

	<i>x</i>	<i>y</i>	<i>z</i>
n	404	404	404
Média	0,28	-0,02	0,81
Desvio padrão	0,42	0,24	0,19
Mínimo	-0,83	-0,99	0,07
Mediana	0,4	-0,03	0,84
Máximo	0,98	0,97	1

Tabela A.4 Medidas descritivas para cossenos diretores, em Profundidade

	<i>x</i>	<i>y</i>	<i>z</i>
n	953	953	953
Média	0,24	0	0,84
Desvio padrão	0,4	0,24	0,16
Mínimo	-0,93	-0,81	0,26
Mediana	0,18	-0,02	0,87
Máximo	0,9	0,9	1

APÊNDICE B

Figuras

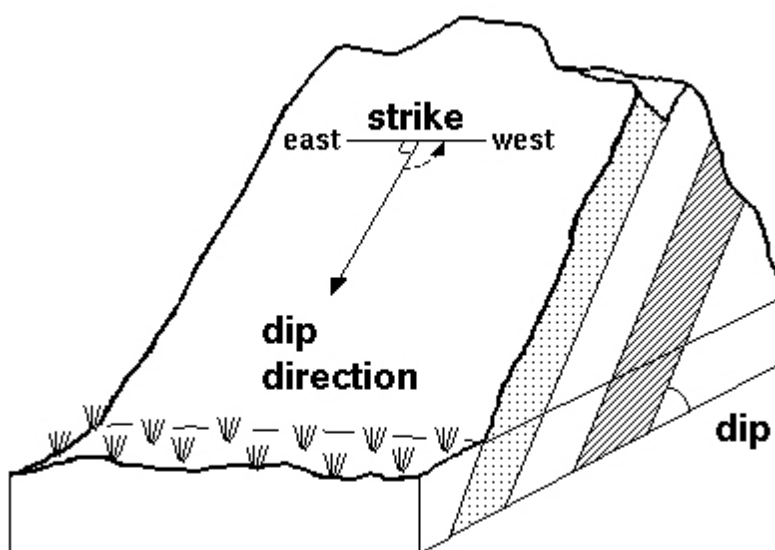


Figura B.1 Esquema explicativo para as variáveis *Dip* e *Dipdirection*

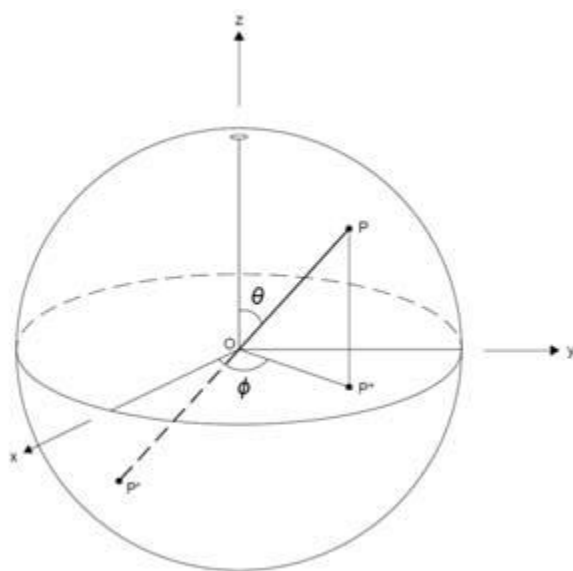


Figura B.2 Esquema explicativo para as variáveis transformadas pelos cossenos diretores (X, Y, Z)

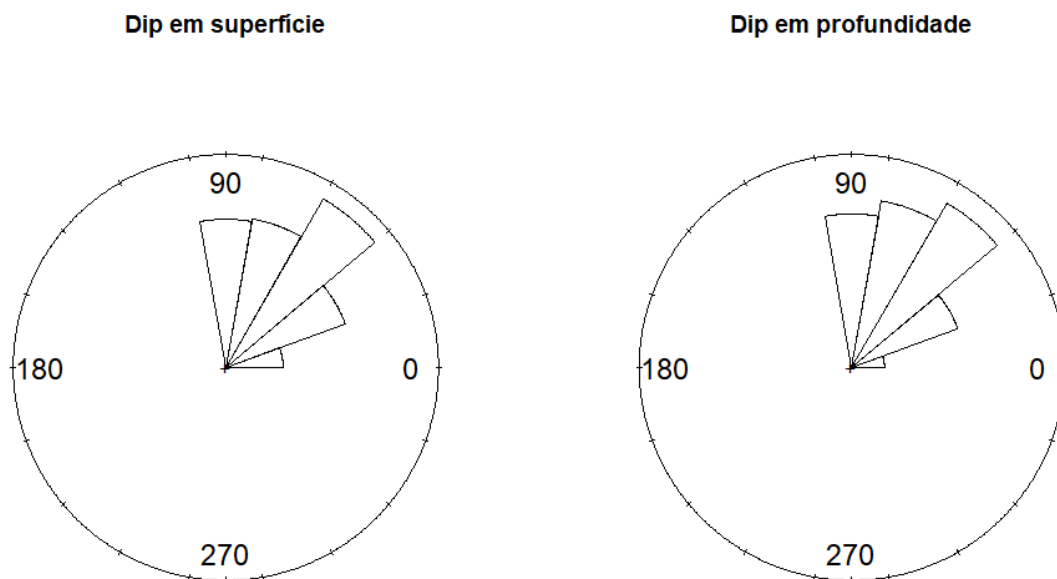


Figura B.3 Gráfico *rose diagram* para a variável Dip em Superfície e em Profundidade

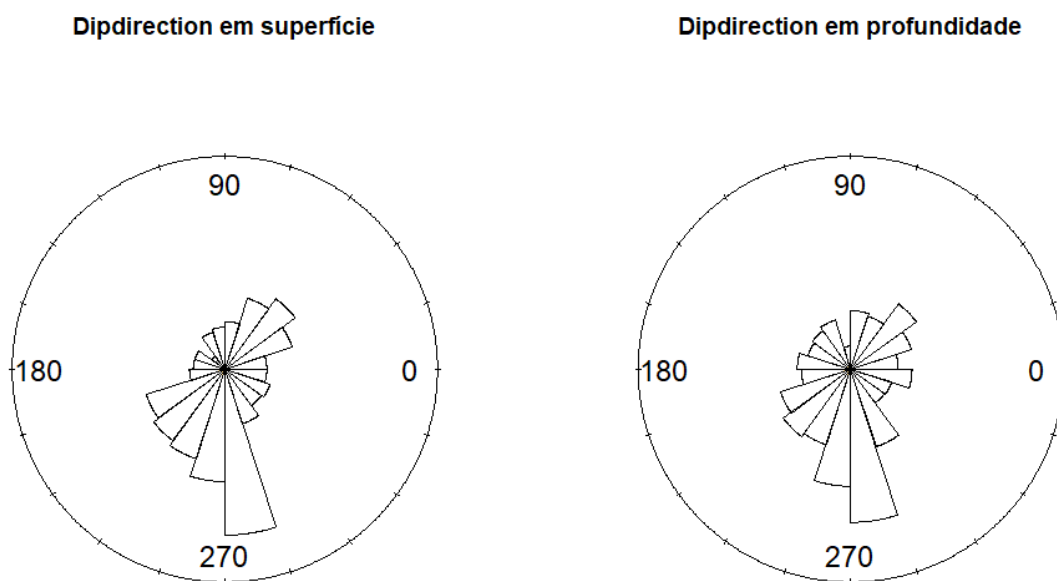


Figura B.4 Gráfico *rose diagram* para a variável Dipdirection em Superfície e em Profundidade

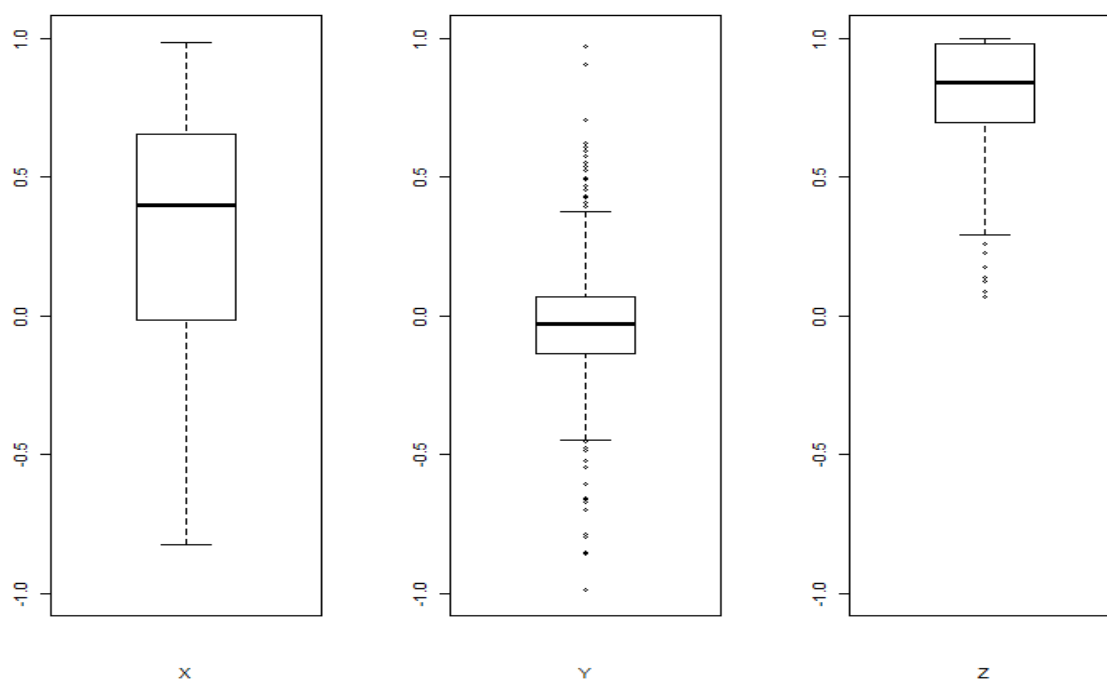


Figura B.5 *Box plots* dos cossenos diretores em superfície

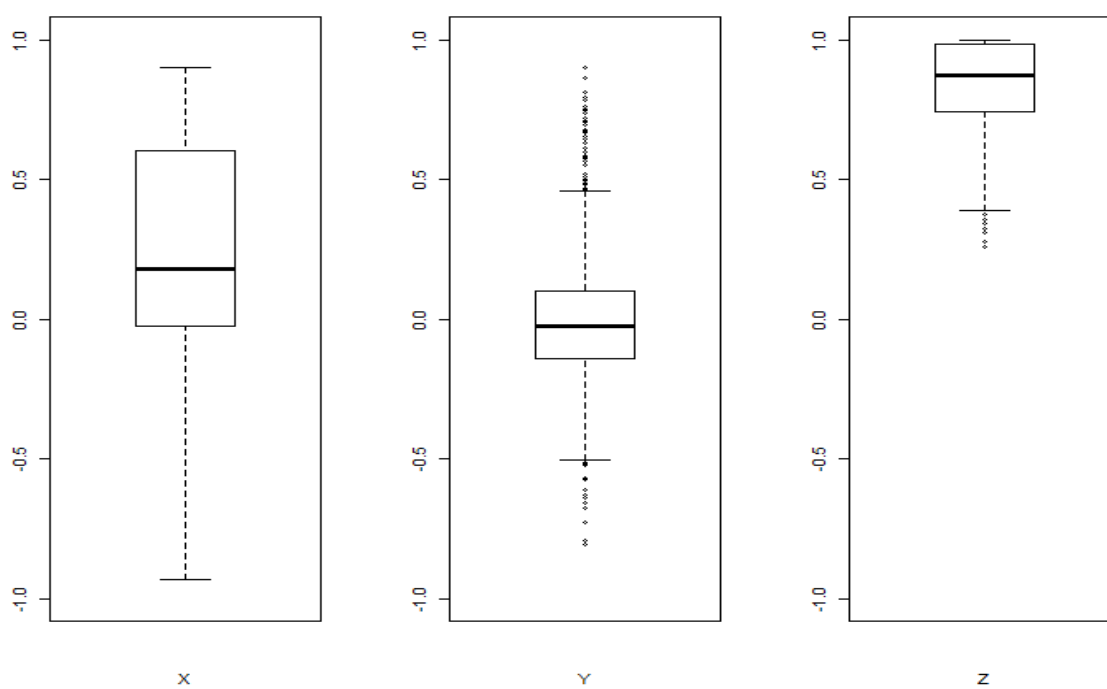


Figura B.6 *Box plots* dos Cossenos diretores em profundidade

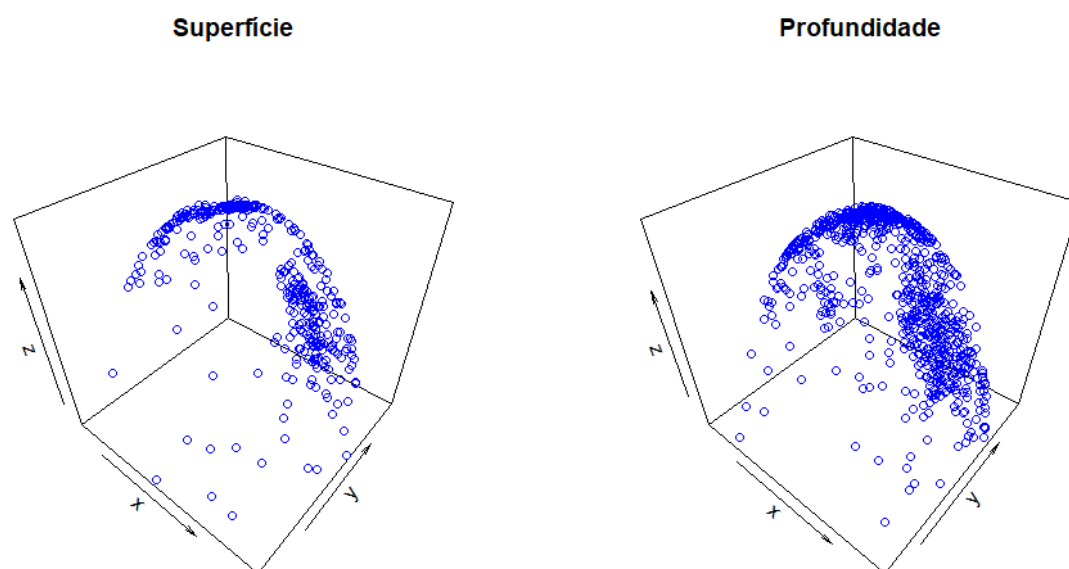


Figura B.7 Diagrama de dispersão dos Cossenos diretores por Cota de profundidade

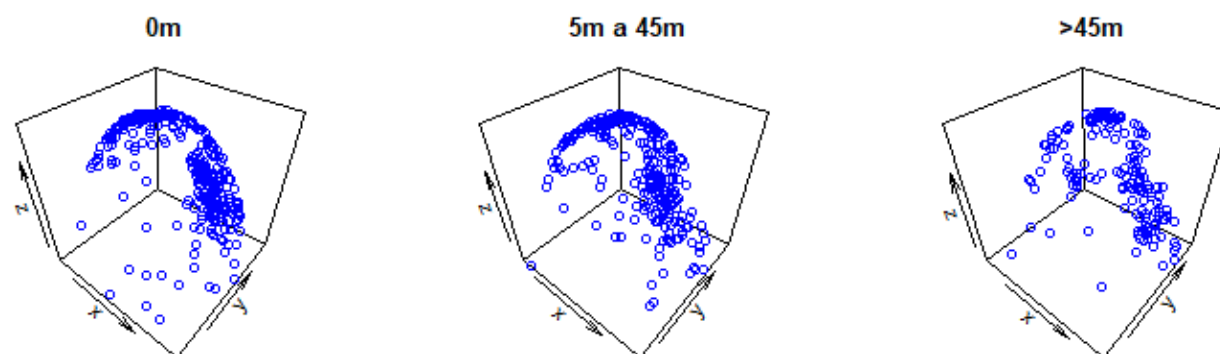


Figura B.8 Diagrama de dispersão dos Cossenos diretores por Distância

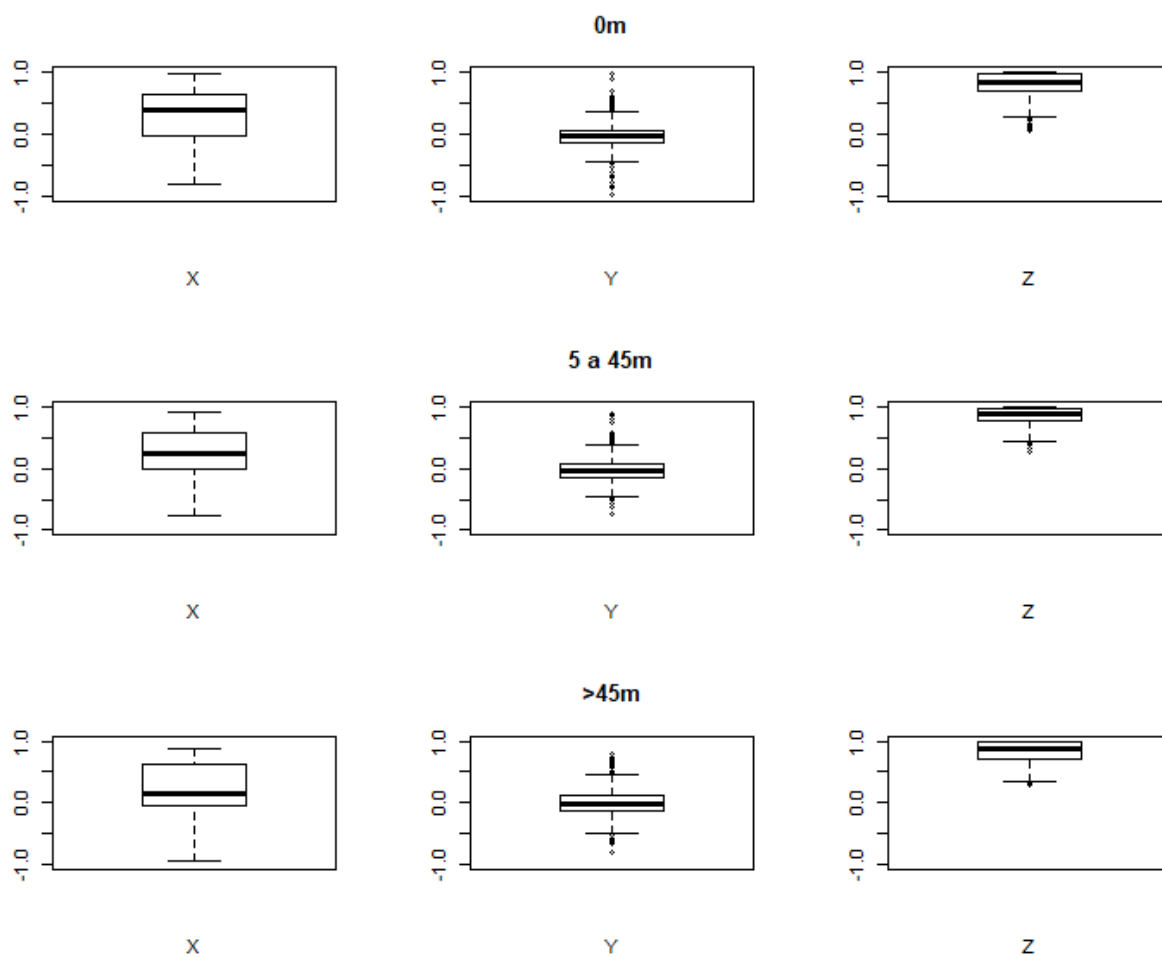


Figura B.9 *Box plots* dos dados dos Cossenos diretores por Distância

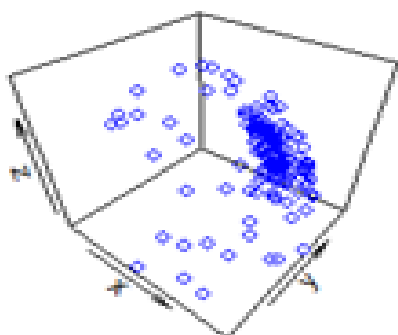
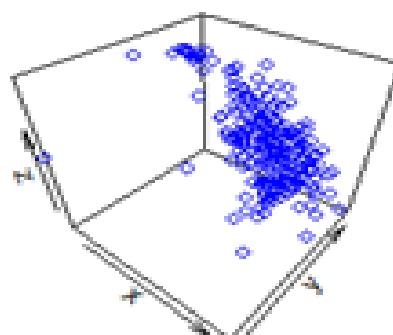
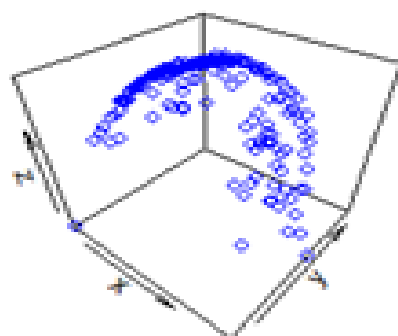
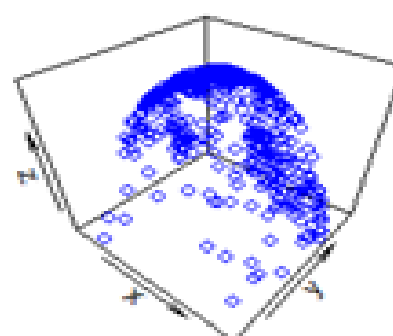
Superfície e foliação**Profundidade e foliação****Superfície e fratura****Profundidade e fratura**

Figura B.10 Diagrama de dispersão de dados dos Cossenos diretores por Estrutura e Cotas de profundidade

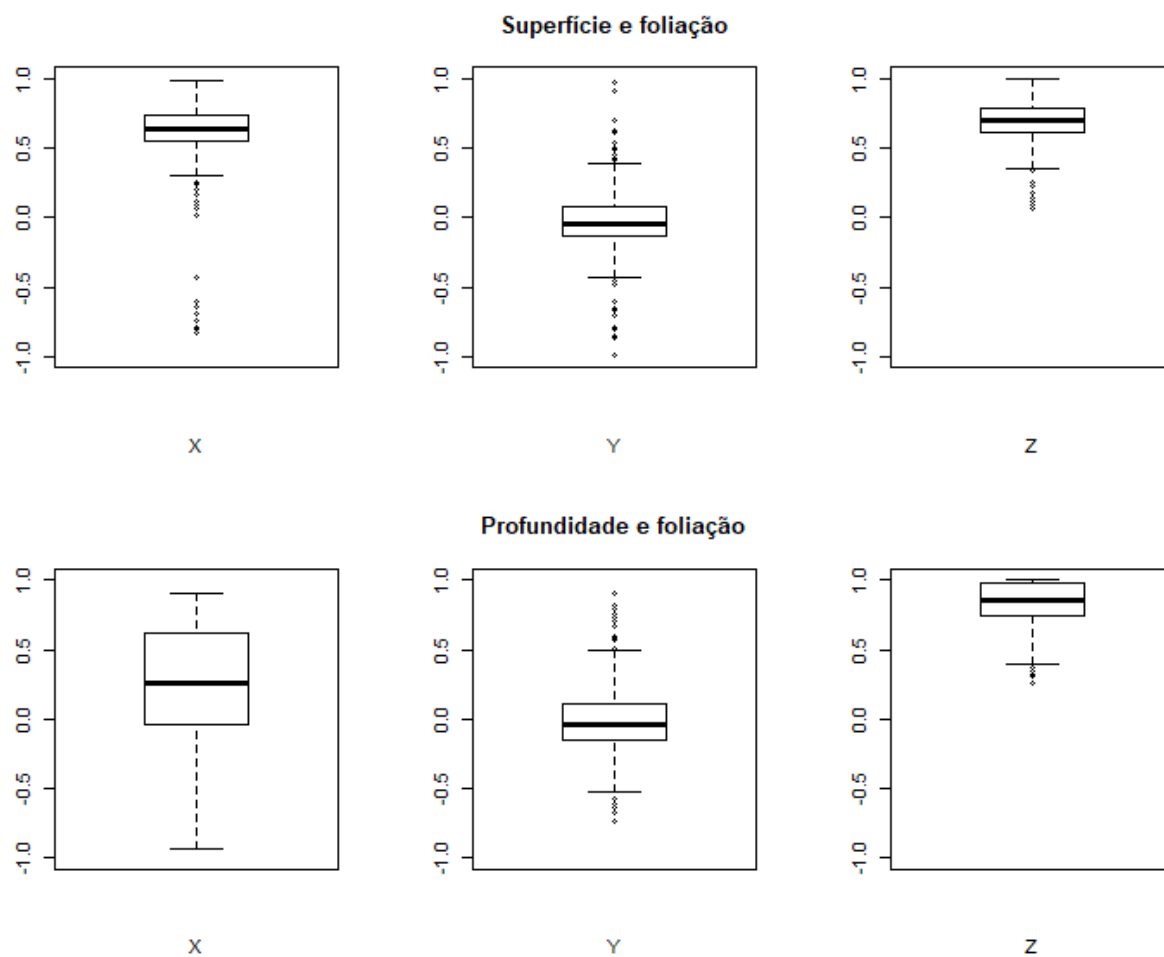


Figura B.11 *Box plots* dos Cossenos diretores por estrutura foliação

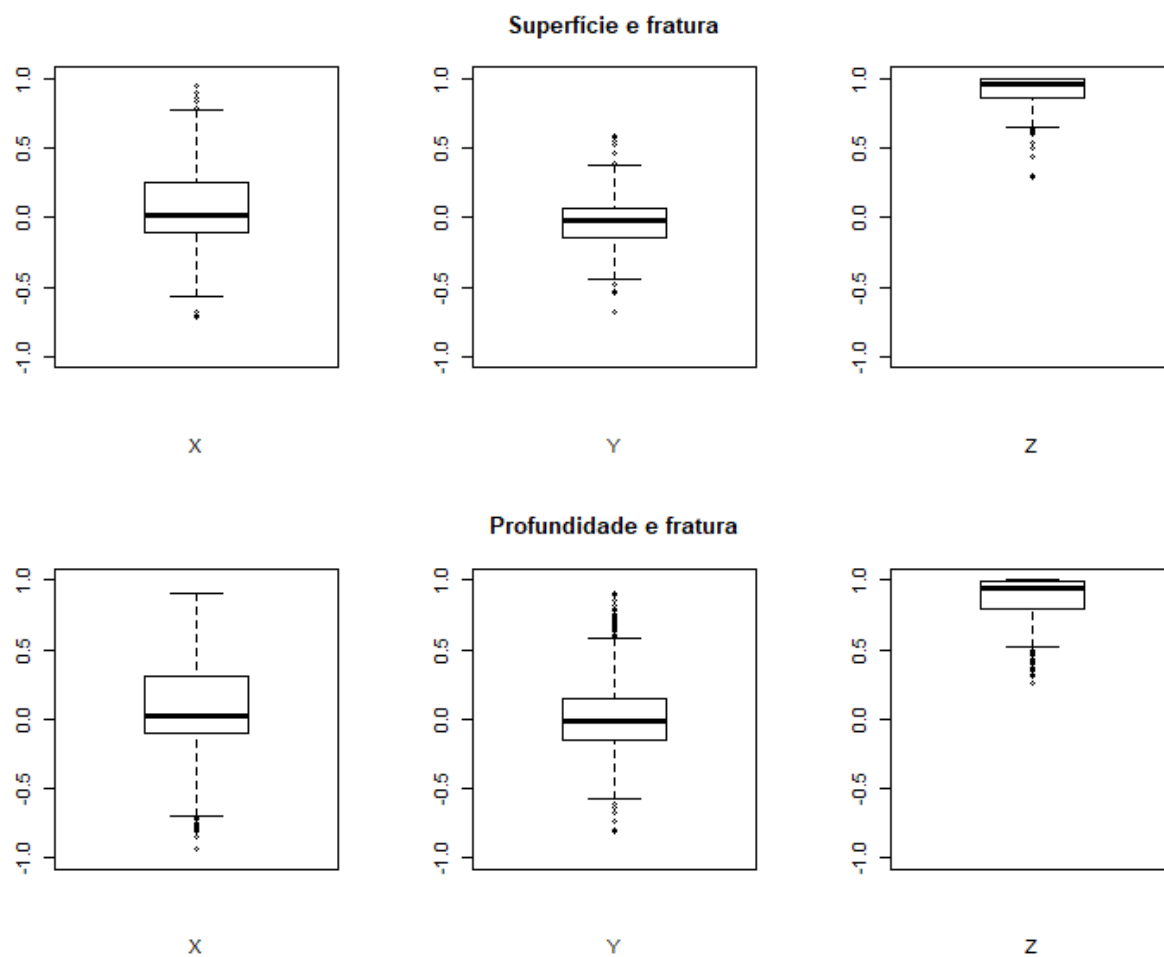


Figura B.12 *Box plots* dos Cossenos diretores por estrutura fratura

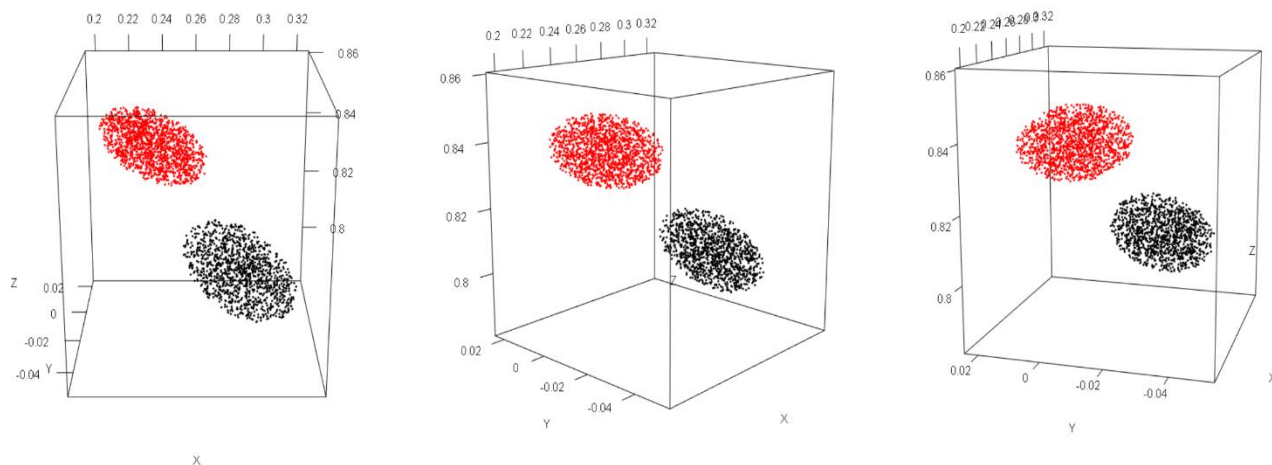


Figura B.13 Região de confiança para a média dos cossenos diretores, com $\alpha = 0,05$ (vermelho em profundidade e preto em superfície)

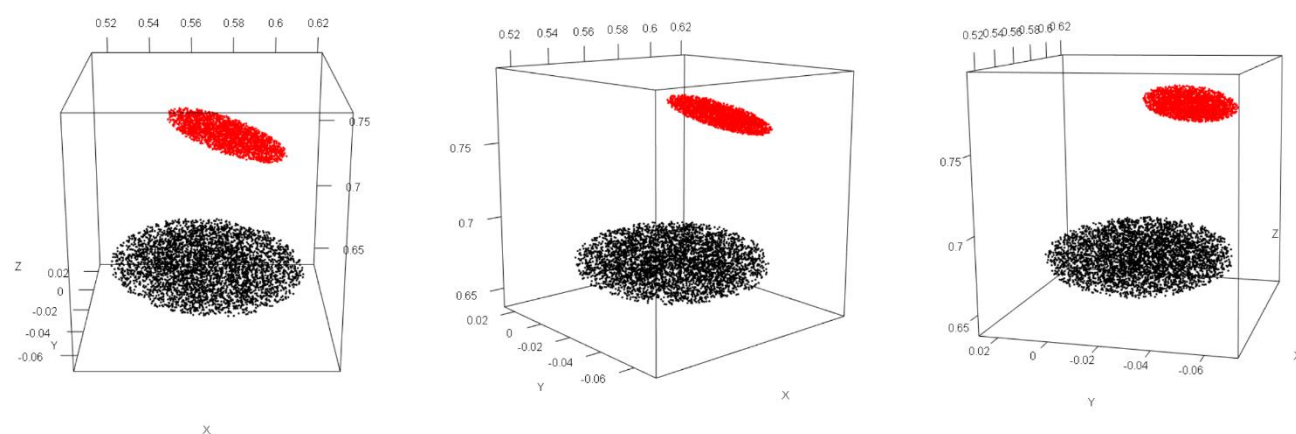


Figura B.14 Região de confiança para a média dos cossenos diretores de foliação, com $\alpha = 0,05$ (vermelho em profundidade e preto em superfície)

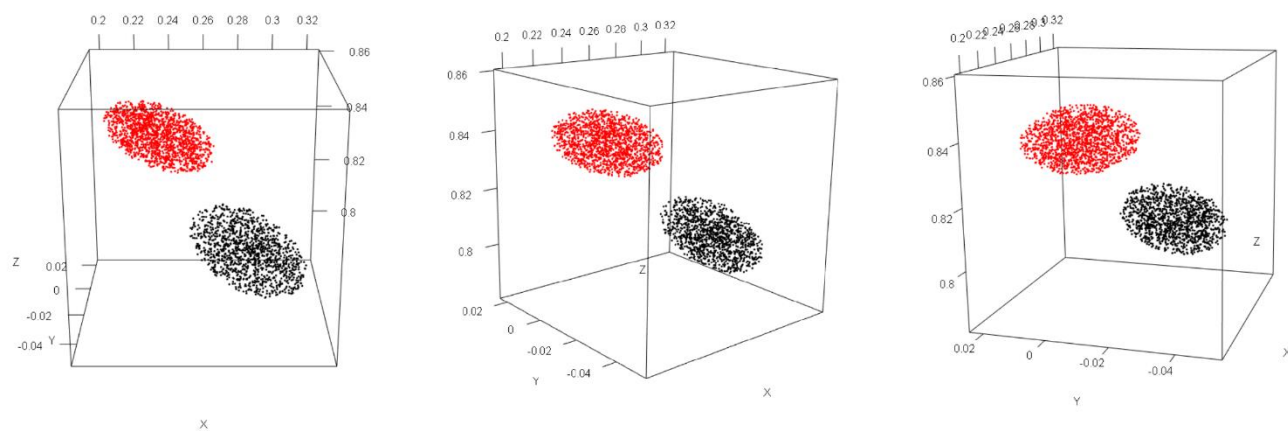


Figura B.15 Região de confiança para a média dos cossenos diretores de fratura, com $\alpha = 0,05$ (vermelho em profundidade e preto em superfície)