

AVALIAÇÃO DE EVENTOS EXTREMOS DE PRECIPITAÇÃO POR MEIO DE UMA ANÁLISE DE VARIÂNCIA NA FRONTEIRA OESTE DO RIO GRANDE DO SUL

Thiago Ferreira Gomes¹, Rafael Maroneze².

*¹Discente do Programa Pós-Graduação em Engenharia - UNIPAMPA, Alegrete-RS, Brasil
(thiagofgomes17@gmail.com)*

²Docente do magistério superior - UNIPAMPA, Alegrete - RS, Brasil

Resumo: O objetivo do trabalho é realizar uma análise de variância sobre a fronteira oeste do Rio Grande do Sul por meio de seis municípios localizados na região, canalizando um intervalo de 10 anos em eventualidade de precipitações, entre os anos de 2011 e 2021. As camadas de dados utilizados são disponibilizadas de forma virtual pelo Instituto Nacional de Meteorologia (INMET), que analisa a ocorrência de algum evento para a captação desses valores obtidos. Foi aplicado o teste de Shapiro-Wilk para determinar a margem aceitável da hipótese para as distribuições de gama e weibull, implementado para essa análise o software RStudio para o cálculo estatístico por meio de pacotes e bibliotecas disponibilizadas, comparando o seu erro quadrático médio desses conjuntos. Todos os grupos destacam uma homogeneidade em seus dados, confirmando que podem ser empregadas essas distribuições por uma boa ascendência acerca desses fenômenos naturais. Assim, o teste de Shapiro-Wilk contém valores relativamente bons e pode ser empregado para outras áreas e regiões, que possuam a mesma disponibilidade de dados.

Palavras-chave: ANOVA; Distribuição Gama; Distribuição Weibull; Dados pluviométricos.

INTRODUÇÃO

Em testes de verificação, deve-se ser verificado os altos índices de precipitação, bem como a organização territorial em zonas urbanas, sendo empregadas práticas para uma ordenação referente à construção civil em regiões ribeirinhas (Braga, 2016). Na análise das cheias, muitos fatores são associados, e um dos principais é a questão de grandes quantidades de precipitação por fenômenos meteorológicos, sendo comum essas ocorrências em todos os continentes. Para tal, empregadas formas de rastreio que podem servir de alerta para esses eventos, por intermédio de dados de satélites que fornecem todos os aspectos das diversas localidades e suas características, bem como uma grande conectividade de informações que são disponibilizadas por órgãos públicos e privados (Maggioni e Massari, 2018).

Essa variabilidade de precipitação provém de análises em um intervalo de tempo, podendo ser de dias ou anos, sendo analisada a quantidade de chuvas ocasionada em áreas demarcadas para o estudo, por dados de satélite ou pluviômetros situados em determinados pontos. Consequentemente, para a pesquisa hidráulica, deve-se olhar para o percurso que a bacia faz em toda a região, pois, antes de

chegar ao local, haverá precipitações cumulativas - ou não - que ajudam a totalizar a quantidade de vazão d'água (Lima et al., 2022).

Um dos principais fatores é o escoamento superficial, o qual altera a vazão necessária resultante sobre o curso do rio, considerando que o termo para a vazão é dado pelo volume de água em canais ao longo do rio, tomando o tempo como um fator. Podendo a modelagem terrestre atuar sobre a formação principal do percurso que um rio percorre e, por efeitos atmosféricos, é alterado o relevo, podendo ser considerado um dos principais causadores de grandes cheias, pois denota-se que este acumula locais com grandes concentrações de água, tendo a ajuda do tipo de solo e eventos antrópicos relacionados à região (Liang e Xie, 2001).

Com a aplicação de uma análise de variância, devem ser elaboradas sequências para se gerar um bom produto, utilizando-se de métodos estatísticos e matemáticos. Relacionando-se o teste do ajustamento de Shapiro Wilk, um dos meios para se calcular uma distribuição sobre os dados propostos, por conta de sua média, e verificando sua igualdade, preferencialmente em pesquisas elaboradas acerca de fenômenos naturais (Hanusz, Tarasinska e Zielinski, 2016).

Torna-se o objetivo principal deste trabalho a realização de uma análise de variância das bacias encontradas na fronteira oeste do estado do Rio Grande do Sul, nos municípios de Alegrete, Quaraí, Sant'Ana do Livramento, São Borja, São Gabriel e Uruguaiana, mediante dados de precipitações diárias de estações localizadas sobre as regiões, utilizando métodos de distribuição para esses eventos e relacionando com o teste de Shapiro Wilk entre os dias de 01 de janeiro de 2011 até 01 de janeiro de 2021.

MATERIAL E MÉTODOS

Foram utilizadas camadas que possuem as precipitações diárias nas regiões estudadas, por análises observadas pelo Instituto Nacional de Meteorologia (INMET), expressados na tabela 1 seus códigos. Após a captação dos dados, foram checadas as datas de colocação das estações para que se pudesse escolher um intervalo de tempo, sendo este do dia 01 de janeiro de 2011 até 01 de janeiro de 2021, gerando, assim, um intervalo de dez anos de precipitações diárias, temáticos que se tem por uma influência estatística.

Tabela 1. Dados das estações do INMET.

Município	Ano de início	Código
Alegrete	27/09/2006	A826
Quaraí	16/10/2007	A831
Sant'Ana do Livramento	21/11/2001	A804
São Borja	20/07/2007	A826
São Gabriel	10/09/2007	A832
Uruguaiana	27/09/2006	A826

Uma ANOVA (ANalysis Of VAriance) é o tratamento estatístico de fatores sistemáticos que se tem por uma influência estatística no conjunto de dados, ou aleatórios, que não possuem agrupamento, efetuados sobre a captação de fatores envolvendo diversas aplicações em variabilidade, podendo ser inserida em algumas distribuições (Sawyer, 2009).

Nas diversas formas de análise pela distribuição gama, pode-se prever um tempo de espera para análises futuras em suas parametrizações, sendo frequentemente usada para um tempo de espera, até que ele aconteça, por ser aplicada em variáveis contínuas, e expressadas por Tripathi, Gupta e Pair (1993) como:

$$f(y) = \frac{(y/\alpha)^{1-n} e^{-(y/\alpha)}}{\alpha \Gamma(\beta)} \quad (1)$$

sendo α o parâmetro de forma, β o parâmetro para a escala, γ o parâmetro de localização e Γ a função de gama (definida pela integral), resultando nos valores para a distribuição gama. Aplicada e seguindo a média dos parâmetros de escala e forma apresentados no processamento de dados é expresso pela equação:

$$E[x] = \frac{\alpha}{\beta} \quad (2)$$

A apresentação do método de Weibull (1951), como a distribuição gama e a lognormal, não possui um começo definido e como deve-se encontrar a melhor distribuição a aplicar aos seus dados experimentais, elaborar mais de duas análises é necessário para essa compreensão. Tendo em vista esses aspectos é importante analisar estatisticamente suas diversas aplicações em áreas de estudos. Sua formulação é expressada por Rocha et al. (2015) como:

$$f(y) = \frac{\beta \cdot y^{\beta-1}}{\alpha^{\beta}} \cdot \exp\left[-\left(\frac{y}{\alpha}\right)^{\beta}\right] \quad (3)$$

o primeiro parâmetro α , conhecido como parâmetro de escala, tem o domínio definido e possui a mesma unidade que t . Já o segundo parâmetro β , conhecido como parâmetro de forma, possui o domínio e não possui dimensão, nos quais os parâmetros apresentam variáveis infinitesimais (Rocha et al., 2015). Em que se aplica a média seguindo os parâmetros de escala e forma apresentados no processamento de dados e expressado pela equação:

$$E[y] = \beta \Gamma\left(1 + \frac{1}{\alpha}\right) \quad (4)$$

O método elaborado por Lucambio (2008 apud Shapiro e Wilk, 1965), utiliza a formulação apresentada como:

$$W = \frac{(\sum_{i=1}^n a_i y_{(i)})^2}{(\sum_{i=1}^n y_i - \bar{y})} \quad (5)$$

onde são calculadas as variáveis de a_i , que são dadas pela equação:

$$(a_1, a_2, \dots, a_1) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}} \quad (6)$$

sendo $m = (m_1, m_2, \dots, m_n)^T$

o vetor dos valores esperados das aplicações em estatística na ordem da amostra e V é dada como uma matriz de covariância dessas aplicações. O p-valor é apresentado como o teste que é calculado exatamente para $n = 3$, em outras situações utilizam-se aproximações diferentes para $4 \leq n \leq 11$ e para $n \leq 12$ (Lucambio, 2008 apud Shapiro e Wilk, 195).

Tem-se que o Erro Quadrático Médio, também conhecido como Mean Squared Error (MSE) é responsável por verificar a acurácia de dados e aplica um valor maior para aqueles pontos que possuem maiores erros, tendo em vista que cada um dos elementos é elevado ao quadrado individualmente, assim, depois que é feito esse processo, gera-se a média destes erros. Sua formulação representa a variância dos dados, e é dada através:

$$MSE = \frac{(\sum_{i=1}^n y_i - \bar{y})^2}{n - 1} \quad (7)$$

tendo em vista que o somatório é feito em relação a y_i representando cada variável encontrada no conjunto de dados, $\bar{y}$ a média de todo o conjunto de variáveis e n é dado como o número de variáveis totais (Fernandez, 2017).

Para um teste em relação às distribuições, é feito uma comparação na variável observada (pelo método de Shapiro-Wilk) e na teórica (por meio das distribuições) para que sejam feitas perspectivas em relação à aplicação dos métodos por meio de estatística em favor dos seus dados a fim de descobrir seu nível de significância de 5% (ou 0,05).

RESULTADOS E DISCUSSÃO

Com testes realizados para descobrir os resultados da tabela 2, foram feitas comparações em relação a todo esse intervalo de tempo de dez anos, contando também os dias que não possuíam precipitações.

Contudo, em uma comparação com os resultados obtidos, mostrou-se que deveriam ser retirados esses dias que não obtiveram nenhuma precipitação,

deixando, assim, não nulos os resultados da mediana da tabela 2.

Seguindo a análise visual sobre a tabela 2, foram feitos novos processos sobre os dias que apenas sofreram com precipitações, somando todos os que foram pegos nesses 10 anos de contabilidade sobre as estações de cada região. Na tabela 3, são apresentados os dias que apenas apresentaram alguma quantidade de precipitação, aplicando os mesmos métodos que já foram utilizados para obtenção dos dados da tabela 2, sendo que foram contabilizados, em dias, 1192 (Alegrete), 1236 (Quaraí), 841 (Sant'Ana do Livramento), 1047 (São Borja), 1296 (São Gabriel) e 1113 (Uruguiana), que obtiveram alguma porcentagem de precipitação.

Analisando a tabela 3, nota-se que é alterada bruscamente a variância e a curtose em relação à tabela 2, efetuando essas diferenças e, pelo fato de haverem apenas dados com os dias que foram captadas chuvas, foram alterados todos esses dados, tornando mais precisa essa média em relação às demais pesquisas.

Na tabela 4, foram feitas as comparações relativas à média calculada em sua forma padrão (representada como distribuição calculada) e as distribuições elaboradas de Weibull e Gama.

Tabela 2. Resultados da primeira análise no software RStudio.

Município	Média (mm/dia)	Mediana (mm/dia)	Variância (mm/dia)	Desvio padrão	Curtose	Assimetria
Alegrete	4,42	0	163,30	12,89	47,28	5,22
Quaraí	4,25	0	164,80	12,84	46,24	5,41
Sant'Ana do Livramento	4,24	0	162,30	12,74	41,50	5,23
São Borja	4,04	0	141,97	11,92	35,33	4,81
São Gabriel	4,29	0	143,60	11,98	34,22	4,74
Uruguiana	3,92	0	147,90	12,16	49,90	5,66

Tabela 3. Resultados da segunda análise no software RStudio.

Município	Média (mm/dia)	Mediana (mm/dia)	Variância (mm/dia)	Desvio padrão	Curtose	Assimetria
Alegrete	12,43	4,40	359,90	18,97	20,95	3,17
Quaraí	11,14	2,80	355,10	18,84	20,39	3,35
Sant'Ana do Livramento	12,02	3,80	366,90	19,15	16,99	3,11
São Borja	11,18	3,20	312,80	17,69	14,59	2,81
São Gabriel	11,45	4	301,80	17,37	14,92	2,87
Uruguaiana	11,99	4	355,50	18,85	19,64	3,30

Tabela 4. Resultado da média calculada e das distribuições em função dos termos apresentados pela variância.

Município	Calculada	Weibull	Gama
Alegrete	12,43	12,59	12,44
Quaraí	11,14	11,07	11,14
Sant'Ana do Livramento	12,02	12,06	12,02
São Borja	11,18	11,02	11,18
São Gabriel	11,45	11,55	11,46
Uruguaiana	11,99	11,99	11,99

Assim, é expressado que o cálculo da média pelo teste da distribuição calculada e a distribuição gama se assemelham muito por possuir alguns dados exatamente iguais, e havendo discrepância em relação aos dados pela distribuição de Weibull. Também é notável que o município de Uruguaiana apresenta os mesmos valores para as 3 distribuições analisadas em relação aos dias que apenas obtiveram algum valor de precipitação.

Na tabela 5, são apresentadas as aplicações referentes aos teste de variância efetuados sobre as distribuições apresentadas, dadas como Observada (Shapiro-Wilk), Teórica I (Gama) e Teórica II (Weibull).

Após a análise, nota-se que as distribuições seguem o nível de significância e podem ser aplicadas como boas para realizar os testes para caracterização das hipóteses de rejeição e de aceitação dos dados propostos, nos quais foram aceitas dentro da margem. Para isto, diz-se que as aplicações teóricas devem ser apresentadas em menor valor ao compará-las com a observada, o que ocorreu muito bem para as duas distribuições propostas pelos autores.

Tabela 5. Resultado das distribuições aplicadas para a hipótese.

Município	Observada	Teórica I	Teórica II
Alegrete	0,67	0,46	0,59
Quaraí	0,62	0,42	0,55
Sant'Ana do Livramento	0,65	0,45	0,58
São Borja	0,66	0,47	0,60
São Gabriel	0,67	0,47	0,59
Uruguaiana	0,65	0,47	0,60

A análise efetuada em relação à quantidade de precipitação segue a figura 1, que apresenta o histograma da variação das chuvas dos municípios, levando em fator a frequência da ocorrência de chuvas encontradas em cada região, para os dados que contabilizaram apenas dias em que ocorreram contagem de precipitações, não correndo anormalidade em suas precipitações no conjunto de dados contínuos estudados.

Os dados apresentados nas figuras 2 e 3 demonstram a normalidade nas ocorrências encontradas, sendo o Q-Q plot o responsável por relacionar a homogeneidade conhecida para os aspectos referentes às diferentes distinções entre os dados.

Assim, adverte-se que as aplicações relacionadas a esses testes apresentam uma boa distribuição ao checar as imagens demonstradas pelas funções. Pode-se analisar que o grupo de Alegrete (representação) denota diferença apenas em seu Q-Q plot, e constata que se assemelha em suas outras 2 representações.

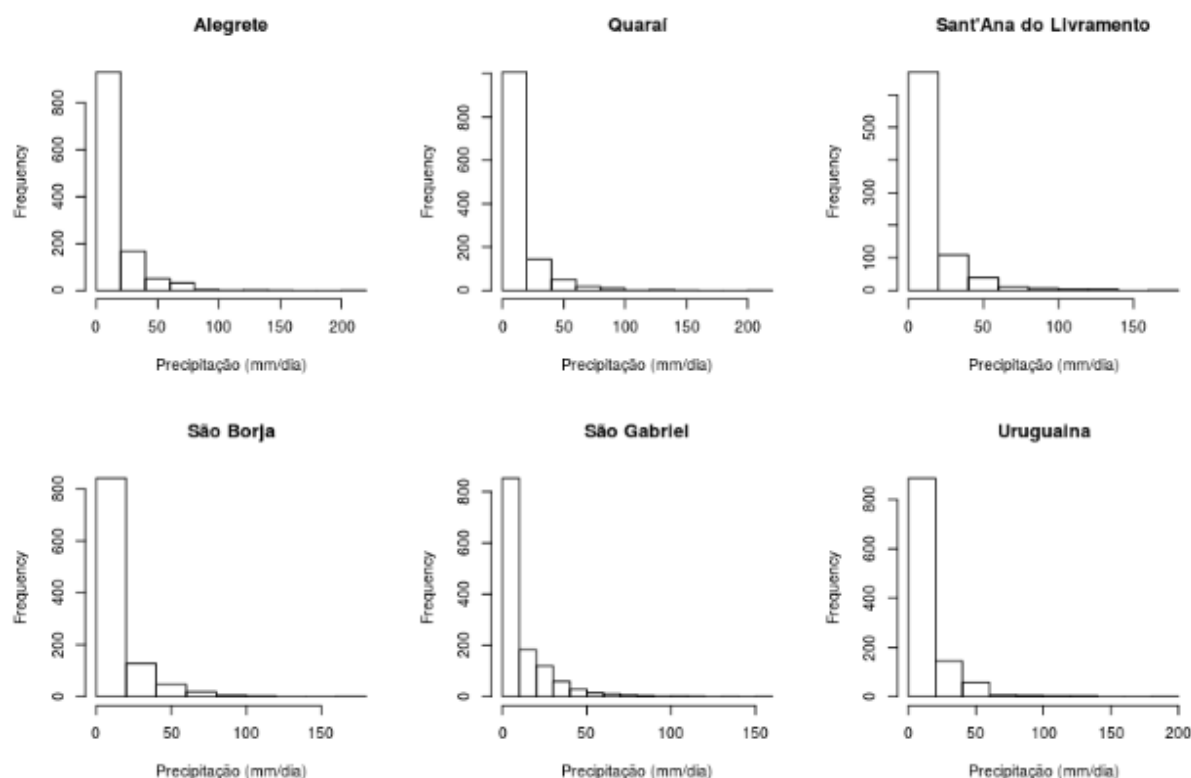


Figura 1. Histograma dos dados de precipitação (mm/dia) do INMET.

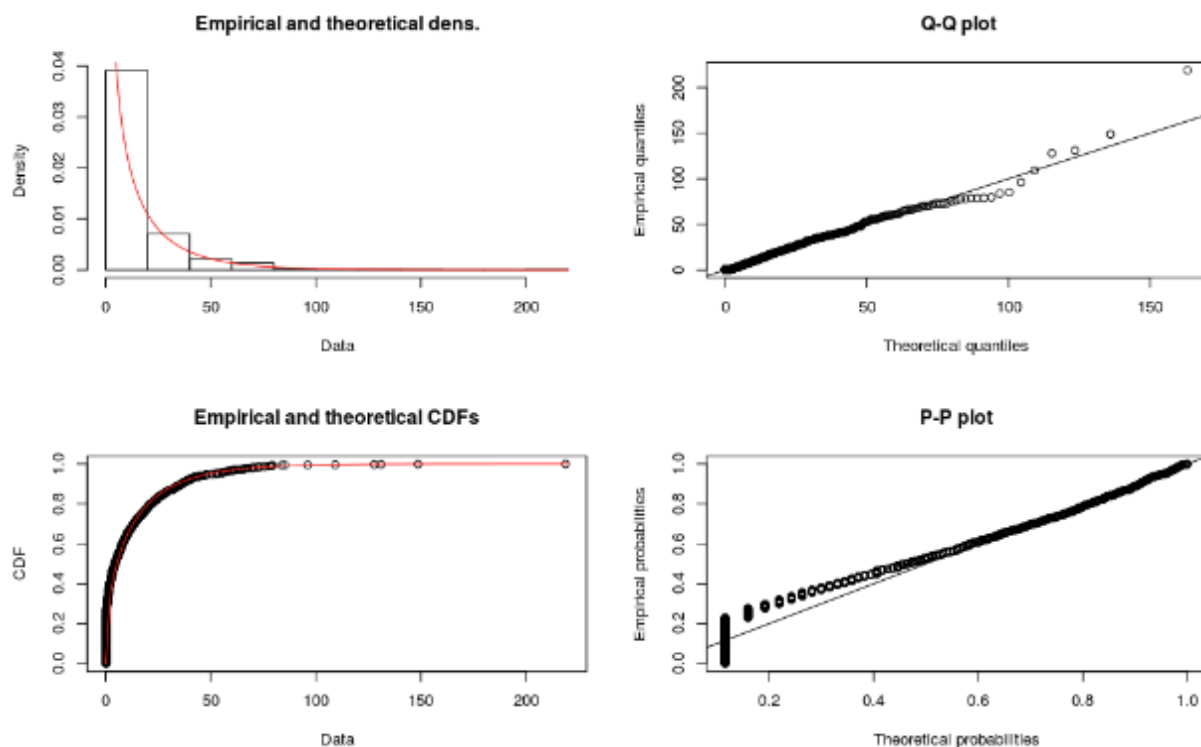


Figura 2. Representação da distribuição de gama para o município de Alegrete.

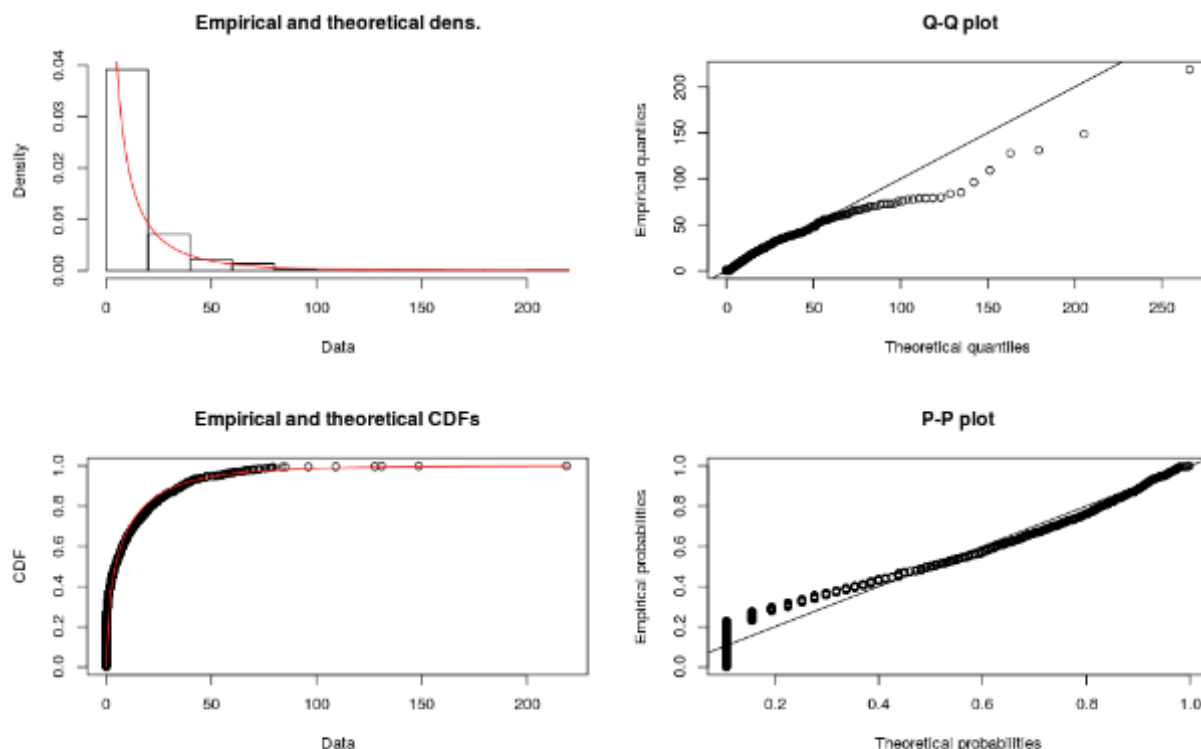


Figura 3. Representação da distribuição weibull para o município de Alegrete.

CONCLUSÃO

É possível afirmar que, ao impor o teste de Shapiro-Wilk para essas distribuições teóricas, são observadas similaridades aceitáveis para aplicação desse teste de ajustamento a dados de observação sobre fenômenos naturais e, assim, são representadas significativamente as suas variâncias (MSE).

Constata-se que os dados apresentados assemelham-se entre as suas médias, pelas checagens pluviométricas realizadas em cada região compreendida no processamento dos 10 anos norteados. E com a aplicação destes testes, em que foi estudada a variância das precipitações sobre a região da fronteira oeste do Rio Grande do Sul, as análises retratam uma boa homogeneidade que se refere aos dados empregados. Logo, mostra-se necessário analisar, frequentemente, essas distribuições, em proporcionalidade de futuras pesquisas acerca dos temas propostos neste artigo.

AGRADECIMENTOS

Agradeço ao apoio financeiro da CAPES, ao Laboratório de Fluidodinâmica Computacional e Turbulência Atmosférica (LFCTA) e à UNIPAMPA.

REFERÊNCIAS

- BRAGA, J. O. **ALAGAMENTOS E INUNDAÇÕES EM ÁREAS URBANAS: ESTUDO DE CASO NA CIDADE DE SANTA MARIA** - DF. Orientadora: Ruth Elias de Paula Laranja. 2016. 42 f. TCC (graduação) - Curso de Geografia, Universidade Federal de Brasília, Brasília, 2016.
- FERNANDEZ, L. H. O. **Método Genérico para Estimção e Modelagem do Erro RMS em Dados de Profundidade de Sensores para Visão 3D**. Orientador: Luiz Marcos Garcia Gonçalves. 2017. 74 f. Dissertação (mestrado) - Curso de Engenharia Mecatrônica, Universidade Federal do Rio Grande do Norte, Natal, 2017.

- HANUSZ, Z.; TARASINSKA, J.; ZIELINSKI, W. Shapiro-wilk test with known mean. **REVSTAT-StatisticRal Journal**, v. 14, n. 1, p. 89-100, 2016.
- LIANG, X.; XIE, Z. A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models. **Advances in Water Resources**, Elsevier, v. 24, n. 9-10, p. 1173-1193, 2001.
- LIMA, C. A. S. et al. Multicriteria analysis for identification of flood control mechanisms: Application to extreme events in cities of different brazilian regions. **International Journal of Disaster Risk Reduction**, Elsevier, v. 71, p. 102769, 2022.
- LUCAMBIO, F. Diferentes testes para verificar a normalidade de uma amostra aleatória. **Statistic Research of Paraná**, v.1, p. 1-12, 2008.
- MAGIONI, V.; MASSARI, C. On the performance of satellite precipitation products in riverine flood modeling: A review. **Journal of Hydrology**, Elsevier, v. 558, p. 214-224, 2018.
- ROCHA, R. J. Estudo sobre métodos de estimação de $F(t)$ para o cálculo dos parâmetros de distribuições de Weibull com a aplicação em uma análise de falhas de garrafas PET. Orientador: [S.I.]. 2015. TCC (graduação) - Curso de Engenharia Mecânica, Universidade Federal de Santa Catarina, Santa Catarina, 2015.
- SAWYER, S. F. Analysis of variance: the fundamental concepts. **Journal of Manual & Manipulative Therapy**, Taylor & Francis, v. 17, n. 2, p. 27E-38E, 2009.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, JSTOR, v. 52, n. 3/4, p. 591-611, 1965.
- TRIPATHI, R. C.; GUPTA, R. C.; PAIR, R. K. Statistical tests involving several independent gamma distributions. **Annals of the institute of Statistical Mathematics**, Springer, v. 45, n. 4, p. 773-786, 1993.