



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Faculdade de Engenharia Elétrica
Graduação em Engenharia Biomédica

IZABELLA NONATO OLIVEIRA LIMA

**DESENVOLVIMENTO DE UMA PLATAFORMA PARA O
PROCESSAMENTO ESTATÍSTICO DE SINAL
ELETROENCEFALOGRÁFICO (EEG)**

Uberlândia
Julho, 2018.

IZABELLA NONATO OLIVEIRA LIMA

**DESENVOLVIMENTO DE UMA PLATAFORMA PARA O
PROCESSAMENTO ESTATÍSTICO DE SINAL
ELETROENCEFALOGRÁFICO (EEG)**

Trabalho apresentado como requisito parcial de avaliação
na disciplina Trabalho de Conclusão de Curso de
Engenharia Biomédica da Universidade Federal de
Uberlândia.

Orientador: João Batista Destro Filho

Assinatura do Orientador

Uberlândia
Julho, 2018.

Dedico este trabalho aos meus pais, pelo estímulo,
carinho e compreensão.

AGRADECIMENTOS

Agradeço primeiramente à Deus por ter me dado forças para trilhar este caminho que muitas das vezes se mostra árduo e sinuoso, à minha família pelo apoio, paciência, compreensão, motivação e principalmente pelo incentivo.

À minha coorientadora Camila Davi Ramos pela ajuda, paciência, suporte e companheirismo, ao meu orientador João Batista Destro Filho incentivo, motivação e orientação deste trabalho e por último, mas não menos importante, agradeço aos meus amigos por fazer essa etapa da minha vida ser mais leve e inesquecível.

RESUMO

Este trabalho tem como finalidade o desenvolvimento de um software de análise estatística de sinais de eletroencefalografia (EEG) tendo seu enfoque na Análise Estatística Descritiva e também em Testes Estatísticos, como Testes de Hipóteses e Testes de Normalidade, sendo assim, todos os registros de EEG (contendo 20 canais e N épocas) utilizados no software foram submetidos à essas análises estatísticas que à priori já continham informações sobre três quantificadores distintos, os quais são: Porcentagem de Contribuição de Potência (PCP) (informa a quantidade de energia contida em uma faixa de frequência); Frequência Mediana (FM) (valor da frequência média na faixa de frequência) e por último Coerência (grau de similaridade entre o hemisfério direito e esquerdo). Como decorrência do estudo, foi obtida uma plataforma com sete interfaces no total, distribuídas em: Interface Principal, Interface de Análise Estatística Descritiva, Interface de Visualização e Interfaces de Testes (Hipóteses e Normalidade). O software aqui apresentado foi testado apenas com registros reais de EEG de pacientes comatosos e de voluntários normais, e todos os dados adquiridos na realização da análise estatística foram comparados com softwares estatísticos renomados existentes no mercado. A validação do software ocorreu confrontando os dados obtidos neste trabalho com os resultados obtidos em outros dois softwares, foram confrontados valores oriundos do processamento estatístico descritivo, dos testes de normalidade e de hipóteses, sendo que todas as comparações realizadas se mostraram satisfatórias e iguais aos outros softwares comparados. Portanto, o software deste trabalho se mostra de grande valia, dado que este é voltado exclusivamente para análise estatística de registros de EEG, obteve um tempo de processamento melhor quando comparado aos outros softwares utilizados, se mostra de fácil utilização e de grande praticidade, além de, evitar a utilização de diversas plataformas para realizar a análise estatística, visto que, neste software é possível realizar uma análise estatística satisfatória e completa.

Palavras-chaves: Bioestatística, eletroencefalografia, software.

ABSTRACT

The aim of this work is the development of a software for the statistical analysis of electroencephalography (EEG) signals, focusing on Descriptive Statistical Analysis and also on Statistical Tests, such as Hypothesis Testing and Normality Tests, thus, all EEG registers (containing 20 channels and N periods) used in the software were submitted to these statistical analyzes that previously had information on three different quantifiers, which are: Power Contribution Percentage (PCP) (informs the amount of energy contained in a band of frequency); Median Frequency (FM) (value of the average frequency in the frequency range) and finally Coherence (degree of similarity between the right and left hemispheres). As a result of the study, a platform was obtained with seven interfaces in total, distributed in: Main Interface, Descriptive Statistical Analysis Interface, Visualization Interface and Test Interfaces (Hypotheses and Normality). The software presented here was tested only with actual EEG records of comatose patients and normal volunteers, and all data acquired in the statistical analysis were compared with renowned statistical software on the market. The validation of the software was performed by comparing the data obtained in this work with the results obtained in two other software, comparing values from descriptive statistical processing, normality tests and hypotheses, and all comparisons were satisfactory and the same as the others comparative software. Therefore, the software of this work proves to be of great value, given that it is exclusively focused on statistical analysis of EEG records, obtained a better processing time when compared to other software used, is easy to use and very practical, besides to avoid the use of several platforms to perform the statistical analysis, since in this software it is possible to perform a satisfactory and complete statistical analysis.

Keywords: Biostatistics, electroencephalography, software.

LISTA DE ILUSTRAÇÕES

Figura 1: Interface inicial do software de estatística.	29
Figura 2: Interface Estatística Descritiva.....	30
Figura 3: Exemplo do nome do arquivo gerado ao final do cálculo da estatística descritiva. .	31
Figura 4: Interface Visualizar Tabela.	32
Figura 5: Interface Visualização em Gráficos.	32
Figura 6: Interface de Visualização de Topografias.....	33
Figura 7: Interface do Teste de Normalidade.	34
Figura 8: Exemplo do arquivo gerado após o término do processamento do teste de normalidade.	34
Figura 9: Interface de Teste de Comparação.	35
Figura 10: Exemplo de nome do arquivo que armazena os resultados advindos do processamento dos testes de comparação.....	36

LISTA DE TABELAS

Tabela 1: Softwares estatísticos existentes e seus respectivos valores de aquisição.....	19
Tabela 2: Resultados obtidos da comparação do software estatístico, software A e B para o cálculo da estatística descritiva, para a situação de um indivíduo com grau de consciência normal e quantificador PCP.....	37
Tabela 3: Comparação dos resultados obtidos dos softwares estatísticos, A e B para os dados do cálculo estatístico descrito de um grupo de dois indivíduos normais para o quantificador FM.	39
Tabela 4: Resultados obtidos da comparação do software estatístico, software A e B para o cálculo da estatística descritiva, para a situação de um grupo de três indivíduos com grau de consciência coma, par de eletrodos Fp1-Fp2 e quantificador coerência.	41
Tabela 5: Resultados obtidos da comparação do software estatístico e o software C para o teste de Kolmogorov-Smirnov, para a situação de um indivíduo com grau de consciência coma e quantificador FM.	43
Tabela 6: Resultados obtidos da comparação do software estatístico e o software A para o teste de Mann-Whitney, para a situação de comparação entre um grupo de registros de indivíduos (normal x coma) e quantificador coerência.	43

LISTA DE ABREVIATURAS E SIGLAS

AOS - Apneia Obstrutiva do Sono

CPAP - Pressão Positiva Contínua nas Vias Aéreas

DAC - Doença Arterial Coronariana

EEG - Eletroencefalografia

MEGS - Encefalografia Eletromagnética

EMG – Eletromiografia

fMRI - Ressonância Magnética

ICA - Componentes Independentes

OA - Osteoartrite do joelho

TS -Torácica Superior

TI - Torácica Inferior

UFU – Universidade Federal de Uberlândia

USG - Ultrassonografia

SUMÁRIO

1 INTRODUÇÃO	11
2 DESENVOLVIMENTO.....	17
2.1 REVISÃO BIBLIOGRÁFICA.....	18
2.2 METODOLOGIA	20
2.2.1 Arquivos de Dados Avaliados	20
2.2.2 Funcionalidades do Software.....	21
2.2.2.1 Estatística Descritiva	21
2.2.2.3 Teste de Normalidade.....	25
2.2.2.4 Teste de Comparação	27
2.3 RESULTADOS.....	29
2.3.1 Interfaces Estatística Descritiva	30
2.3.2 Interfaces Testes de Normalidade	33
2.3.3 Interfaces Testes de Hipóteses	34
2.4 DISCUSSÃO DOS RESULTADOS	36
2.4.1 Situação 1	36
2.4.2 Situação 2	38
2.4.3 Situação 3	40
2.4.4 Situação 4	42
2.4.5 Situação 5	43
3 CONCLUSÕES	44
4 REFERÊNCIAS	45

1 INTRODUÇÃO

Relatos de levantamentos estatísticos são encontrados desde 2000 anos antes de Cristo no continente asiático, precisamente na China, onde nestes levantamentos se encontravam informações sobre a população, riquezas, tributos e outros. Já os registros egípcios mostravam informações sistemáticas de natureza estatística realizadas por faraós. Outras civilizações utilizavam também de dados estatísticos para levantamentos sobre suas populações, riquezas, militares, etc. Na Era Renascentista as informações estatísticas eram prioritariamente para aplicações de administração pública. Deste modo, nota-se que a Estatística propriamente dita teve seu desenvolvimento advindo de aplicações, ou seja, é uma ciência que tem como caráter o significado e os dados (MEMÓRIA, 2004). Mas o que é especificamente Estatística?

A estatística tem como objetivo, observar, coletar, classificar, resumir, organizar, analisar e interpretar um conjunto de dados. A área estatística é de suma importância porque através desta é possível delimitar as incertezas advindas dos resultados obtidos do processamento dos dados, com isso, é concebível replicar uma determinada pesquisa em mesmas condições, obtendo os mesmos resultados em todas as repetições executáveis (FONTELLES, 2012).

Para a realização de uma análise estatística correta, é necessário realizar etapas que auxiliam na organização e futura interpretação (correta) dos dados em observação. O primeiro passo é definir o problema que se deseja resolver, ou seja, é necessário encontrar um problema para qual se pretende encontrar futuras respostas. O segundo passo é levantar dados necessários o suficiente em tempo hábil e completo para que estes contribuam para a resolução do problema em análise. O terceiro passo faz menção à organização dos dados coletados no segundo passo ao realizar essa etapa a análise da base de dados ficará mais fácil do que se os dados estivessem bagunçados. Os dois últimos passos fazem referência a análise dos dados e interpretação e por fim, a escolha da decisão correta em base dos resultados da análise estatística (REIS et al., 1999).

Realizando todos os passos descritos, a estatística descritiva pode e é aplicada em diversas áreas, seja na agrícola, na econômica, na de análise da população de um país ou até mesmo a população mundial, outro âmbito que a estatística vem ganhando espaço é área da saúde, o crescimento é tamanho que quando a estatística se refere à saúde, é denominada de bioestatística (FONTELLES, 2012).

A bioestatística normalmente é utilizada para informar a eficiência de um determinado medicamento, incidência de mortalidade de uma certa doença, prevalências de uma doença, análise de um conjunto de dados biológicos, levantamento de uma população com uma doença, incidência de uma doença epidemiológica, regressão logística, das tabelas de sobrevivência, das técnicas de ensaios biológicos, entre outros. Na área de saúde a bioestatística não se restringe somente à um conjunto de cálculos e resultados

obtidos a partir da base de dados, na medicina, a bioestatística auxilia na escolha das situações experimentais e na determinação da quantidade de indivíduos a serem examinados (ZAROS; MEDEIROS, 2011).

Na bioestatística, existem testes que auxiliam na interpretação dos dados coletados. Um desses testes é denominado teste de hipóteses, que são questionamentos erguidos que possuem relação ao problema em análise e que ao serem respondidas, auxiliam na solução do mesmo. Nos testes de hipóteses são geradas duas hipóteses chamadas hipótese nula (H_0) e hipótese alternativa (H_1). Sendo que a hipótese nula é referente àquela que é colocada à prova, enquanto que hipótese alternativa é aquela que será aceitável, caso a hipótese nula seja rejeitada (PAES, 1998).

Normalmente na área de saúde quando se tratam de problemas mais simples, a hipótese nula está relacionada à uma igualdade entre médias ou proporções que indicam a independência entre os motivos em questão. Um exemplo simples pode ser encontrado quando se estuda fatores de risco para doenças cardiovasculares, a hipótese nula seria “a proporção de doentes cardiovasculares entre hipertensos é igual a não hipertensos” e a hipótese alternativa diria o contrário, ou seja, “a proporção de doentes cardiovasculares entre hipertensos e não hipertensos são diferentes” (PAES, 1998).

Outro exemplo é dado quando se tenta avaliar se os coeficientes de mortalidade estão relacionados com os níveis socioeconômicos da população de São José do Rio Preto, essa hipótese seria a nula, onde esse questionamento foi colocado à prova na realização do estudo (GODOY et al., 2007). Já na área de odontologia a hipótese colocada em questão foi se o medicamento diazepam tem relação com o comportamento (choro, movimento do corpo e/ou cabeça, fuga, esquivas) de crianças com a história de não colaboração no tratamento odontológico (POSSOBON et al., 2003).

Ao analisar uma eletromiografia do músculo isquiotibiais, o estudo (BRANCO et al., 2006) levantou como hipótese nula que a tensão que é aplicada a esses músculos e o desconforto existente após o alongamento dos mesmos estão diretamente relacionados. Já o estudo (DESLANDES et al., 2010) levantou durante um ano a hipótese que existe efeito do treinamento aeróbico na assimetria do ritmo alfa no sinal de eletroencefalografia (EEG) e sintomas depressivos em idosos. Observando esses estudos aqui listados, nota-se o quão importante e necessário é a estatística aplicada na área de saúde, já que esta auxilia em descobertas que passam desde um pequeno problema até um problema que atinge toda a população (ZANOBETTI, 2000).

Como citado anteriormente, a bioestatística pode ser empregada na análise quantitativa de registros de sinal de eletroencefalografia, que nada mais é que o registro da atividade cerebral. Os primeiros registros de EEG da história ocorreram por volta de 1875, quando o cientista inglês Richard Caton (1842-1926) usando um galvanômetro, colocou sobre o couro cabeludo de um indivíduo dois eletrodos, registrando assim a primeira atividade cerebral na forma de sinais elétricos. Depois disso, com o passar dos anos vários cientistas de diversas nacionalidades começaram ou desenvolveram novas técnicas para o aperfeiçoamento

desse registro, até que Hans Berger (1873-1941) em 1920, começou a estudar sinais de EEG em humanos, sendo o responsável por realizar o primeiro registro de uma gravação de desse sinal que durou cerca de três minutos e em seu primeiro relatório (1929), Hans Berger incluiu o ritmo alfabético como componente principal, sendo assim Berger deu o nome de alfa para o primeiro ritmo que registrou ao realizar o seu primeiro registro de EEG (SANEI; CHAMBERS, 2007).

Um eletroencefalográfico é uma medida das correntes que fluem durante as excitações sinápticas dos dendritos de muitos neurônios piramidais no córtex cerebral. O córtex é a camada mais externa do cérebro e tem uma espessura de 2-3 mm. A superfície cortical é altamente convoluta por sulcos e vales de tamanhos variados sendo constituído por dois hemisférios simétricos - esquerdo e direito - que são separados pela profunda fissura sagital (o sulco central). Cada hemisfério é dividido em quatro lobos diferentes: os lobos frontais, temporais, parietais e occipitais. A informação sensorial é processada em várias partes dos lobos: o córtex auditivo está localizado na parte superior dos lobos occipitais, e o córtex sensorial somático está localizado logo depois do sulco central do lobo parietal (SÖRNMO; LAGUNA, 2005). Quando as células cerebrais (neurônios) são ativadas, as correntes sinápticas são produzidas dentro dos dendritos. Essa corrente gera um campo magnético mensurável por eletromiografia (EMG) e um campo elétrico secundário sobre o couro cabeludo mensurável pelos sistemas EEG. A cabeça humana consiste em diferentes camadas, incluindo o couro cabeludo, crânio, cérebro e muitas outras camadas finas no meio, o crânio atenua os sinais aproximadamente cem vezes mais do que o tecido mole. Portanto, grandes populações de neurônios ativos podem gerar potencial suficiente para serem registradas usando os eletrodos do couro cabeludo. Estes sinais são posteriormente amplificados para fins de grande interesse (SANEI; CHAMBERS, 2007).

A diversidade de ritmos do EEG é enorme e depende, entre muitas outras coisas, do estado mental do sujeito, como o grau de atenção, despertar e sono. Os ritmos são convencionalmente caracterizados por sua faixa de frequência e amplitude relativa. A amplitude do sinal EEG está relacionada ao grau de sincronia com o qual os neurônios corticais interagem. A frequência, ou a taxa oscilatória, de um ritmo de EEG é parcialmente sustentada pela atividade inicial do tálamo e este ritmo é mais uma expressão de um mecanismo de feedback que pode ocorrer em um circuito neuronal. Os ritmos de alta frequência / baixa amplitude refletem um cérebro ativo associado ao estado de alerta ou sono onírico, enquanto os ritmos de baixa frequência / grande amplitude estão associados à sonolência e a estados de sono não soníferos (SÖRNMO; LAGUNA, 2005).

Ritmos eletroencefalográficos, também conhecidos como ritmos de fundo, são convencionalmente classificados em cinco bandas de frequência (variando entre 0,5 a 30-40 Hz) diferentes. As interpretações dessas bandas em termos de "normal" ou "anormal" são relativas e dependem da idade e do

estado mental do sujeito. As cinco bandas de frequências mais conhecidas são: Delta, Teta, Alfa, Beta e Gama (SÖRNMO; LAGUNA, 2005).

- Ritmo Delta (<4Hz), sendo um ritmo tipicamente encontrado durante o sono profundo com uma grande amplitude e normalmente não é observado em adultos (normais) acordados, seu aparecimento para este caso é indicativo de, por exemplo, lesão cerebral ou doença cerebral (encefalopatia) (SÖRNMO; LAGUNA, 2005).
- Ritmo Teta (4-7Hz), este ritmo ocorre durante a sonolência e em certos estágios do sono (SÖRNMO; LAGUNA, 2005).
- Ritmo Alfa (8-13 Hz), este ritmo é mais proeminente em indivíduos normais que estão relaxados e acordados com os olhos fechados, a atividade é suprimida quando os olhos estão abertos (SÖRNMO; LAGUNA, 2005).
- Ritmo Beta (14-30Hz), é um ritmo rápido com baixa amplitude, associado a um córtex ativado e que pode ser observado, por exemplo, durante certos estágios do sono. O ritmo beta é observado principalmente nas regiões frontal e central do couro cabeludo (SÖRNMO; LAGUNA, 2005).
- Ritmo Gama (> 30Hz), este ritmo é relacionado a um estado de processamento ativo de informações do córtex, o ritmo gama pode ser observado durante os movimentos dos dedos (SÖRNMO; LAGUNA, 2005).

O EEG clínico é comumente registrado usando o sistema internacional 10/20, que é um sistema padronizado para colocação de eletrodos. Este sistema de gravação particular (montagem de eletrodos) emprega 21 eletrodos ligados à superfície do couro cabeludo em locais definidos por certos pontos de referência anatômicos, o espaçamento dos eletrodos é de aproximadamente 4,5 cm em uma cabeça adulta típica. Para se evitar efeitos de *aliasing* no domínio da frequência, é normalmente recomendado que a taxa de amostragem para a aquisição do sinal EEG é geralmente selecionada para ser pelo menos 200 Hz (SÖRNMO; LAGUNA, 2005).

O projeto de uma interface cérebro-computador é outra aplicação do EEG que é considerada, até agora, estudada principalmente a partir de uma perspectiva orientada para a pesquisa, por isso que o processamento de sinais em várias aplicações de EEG se mostram de suma importância. É graças a esses processamentos é possível diagnosticar diversas doenças e/ou anomalias advindas de registros de eletroencefalografia, como por exemplo, epilepsia, insônia, distúrbios do ritmo circadiano, coma, entre outros (SÖRNMO; LAGUNA, 2005).

- Epilepsia: uma pessoa com epilepsia sofre convulsões durante as quais explosões súbitas de atividade elétrica descontrolada ocorrem em um grupo de neurônios do córtex cerebral. Crises epilêpticas se manifestam de muitas maneiras diferentes, dependendo de onde a origem (foco) da atividade elétrica

está localizada e de como diferentes áreas do cérebro se tornam sucessivamente recrutadas durante uma convulsão. Acredita-se que um desequilíbrio entre as duas atividades seja uma importante causa de epilepsia (SÖRNMO; LAGUNA, 2005).

- Insônia: distúrbios em iniciar ou manter o sono. A maioria das pessoas, em algum momento de suas vidas, sofreu de insônia devido a um evento agonizante ou a um exame próximo, esta condição é normalmente transitória e não é tratada (SÖRNMO; LAGUNA, 2005).
- Distúrbios do ritmo circadiano: distúrbios no horário de sono-vigília. O exemplo mais conhecido de tais desordens resulta de voar através de vários fusos horários (SÖRNMO; LAGUNA, 2005).
- Coma: a palavra coma significa um estado patológico de consciência suspensa e falta de resposta a estímulos externos ou internos. O estado pode durar de horas a semanas e mudar para a recuperação da consciência com ou sem sequelas, para a morte ou para um estado permanente em que o paciente possa estar “acordado”, mas sem (ou com uma quantidade mínima) de consciência, conhecido como estado vegetativo crônico. Existem diferentes tipos de coma, dentre eles estão: coma isquêmico, coma por morte encefálica, coma em estado vegetativo crônico, dentre outros (BUZSÁKI, 2009).
 - Coma isquêmico: neste tipo de coma uma ampla gama de alterações pode ser observada no EEG. Medicamentos diferentes podem afetar a suscetibilidade do cérebro à isquemia, a temperatura corporal pode ser um fator importante, como em acidentes por afogamento ou outras formas de hipotermia, algumas áreas do cérebro também podem ser mais suscetíveis devido à arteriosclerose com perfusão reduzida. Em um EEG, isso pode resultar em diferentes graus de comprometimento funcional cortical difuso sem elementos excitatórios. Supõe-se que essas variações expressem o fato de que as camadas corticais são afetadas de maneiras diferentes (BUZSÁKI, 2009).
 - Coma por morte encefálica: a destruição cortical total é manifestada pela ausência de atividade em um EEG do couro cabeludo. Quando um EEG é administrado, hipotermia abaixo de 32 ° C, o choque circulatório e a sedação devem ser excluídos. A causa deve ser conhecida e sinais clínicos de morte encefálica devem ter existido por 12 horas, mas isso pode ser reduzido para seis horas se for estabelecido que a causa é irreversível (BUZSÁKI, 2009).
 - Coma em estado vegetativo crônico: pacientes em estado vegetativo crônico, isto é, sem consciência, mas com estruturas preservadas do tronco cerebral que mantêm funções autonômicas, às vezes podem não ter atividade cortical visível em um EEG do couro cabeludo. No entanto, a maioria mostra atividade de EEG com baixa voltagem, lenta, irregular e não reativa a estímulos sensoriais. Não há variação no EEG em um período de 24

horas, mesmo que o paciente exibe clinicamente ciclos diários de sono e vigília (BUZSÁKI, 2009).

Portanto, o objetivo deste trabalho é desenvolver um software capaz de processar, utilizando ferramentas estatísticas, registros de EEG.

2 DESENVOLVIMENTO

O desenvolvimento do trabalho foi subdividido em quatro tópicos, sendo eles: Revisão Bibliográfica, Metodologia, Resultados e Discussão dos Resultados.

No tópico “Revisão Bibliográfica”, são discutidos o que há de novo em quesito de softwares estatísticos para a área de saúde, abordando assim, artigos que trazem progressos sobre o desenvolvimento de softwares que contribuem para a análise estatísticas de sinais biomédicos e o motivo pelo qual neste trabalho foi desenvolvido um software diferente dos existentes.

Em “Metodologia”, existem 2 subtópicos apresentando os métodos utilizados para o desenvolvimento do software estatístico exibido neste trabalho. O primeiro subtópico “Arquivos de Dados Avaliados” diz respeito aos dados utilizados para a realização da parte estatística, ou seja, dos dados que são avaliados e quais quantificadores são utilizados no software aqui apresentado. O último subtópico “Funcionalidades do Software” tem em seu interior mais três tópicos que fazem menção à Estatística Descritiva, Teste de Normalidade e Teste de Comparação, nesses itens são apresentados todas as variáveis estatísticas e testes utilizados no software de estatística em questão, sendo assim, no item “Estatística Descritiva” são apresentadas as 14 variáveis utilizadas no processamento estatístico descritivo dos registros de EEG e nas temáticas “Teste de Normalidade” e “Teste de Comparação” são apresentados os testes de normalidade e de hipótese, respectivamente, que são calculados no software desenvolvido.

Nos Resultados, mencionado às interfaces que o software em questão neste trabalho possui, sendo que esta temática está dividida em mais três subtópicos referentes a cada interface. O primeiro subtópico, Interface Estatística Descritiva, cita quatro interfaces, sendo a primeira é referente à interface onde se realiza o cálculo da estatística descrita e a três últimas fazem menção à visualização dos resultados obtidos através do processamento estatístico descritivo, sendo possível visualizar em formato de tabela, gráficos e topografias. Os dois últimos subtópicos apresentam as interfaces referentes aos testes de normalidade e teste de comparação, onde neste, é possível escolher qual teste de normalidade ou teste de comparação, respectivamente, deseja-se realizar e consequentemente gerar um arquivo resultante desde processamento.

O último tópico desenvolvido, é referente à Discussão dos Resultados, onde nesta temática são apresentadas as comparações dos resultados obtidos no software deste trabalho com softwares estatísticos já existentes. Dentro ainda deste tópico, é possível notar que existem mais 5 subtópicos que fazem menção as situações escolhidas, entre elas, para um grupo normal ou coma, para um indivíduo normal ou coma e para algum dos testes de normalidade e testes de hipóteses existentes no software desenvolvido.

2.1 Revisão Bibliográfica

A análise estatística ajuda na produção de uma visão geral dos dados, permitindo colocar limites nas incertezas dos resultados obtidos no processamento, além de envolver comparações entre tratamentos ou procedimentos, entre os dados em análise. Existem também, por exemplo, comparações de uma característica de um grupo com um valor numérico teórico. A análise estatística pode ser feita analisando gráficos, tabelas, calculando valores estatístico, resultados de testes, entre outros, sempre buscando realizar uma análise geral dos dados estudados (FONTELLES, 2012).

Em (CACIOPPO et al., 2014), a análise estatística foi utilizada na comparação de duas abordagens de ferramentas estatísticas e apresentar nova ferramenta que inclui quatro novos métodos analíticos diferentes para o mesmo fim, com o intuito de investigar processos periódicos de registros de EEG. A análise estatística foi de suma importância, em (SUGRUE et al., 2017) onde foi realizado a análise arquitetural da onda T, para a identificação do risco de arritmia cardíaca e para uma futura terapia para a síndrome de QT (taquiarritmia ventricular congênita) Tipo 1 e 2.

Já em (TISON et al., 2018) a estatística foi utilizada para mostrar que é possível detectar fibrilação atrial passiva utilizando um *smartwatches* (relógio que capta sinais biológicos, principalmente sinal cardíaco) ao comparar os valores que o relógio apresenta com os valores de referência para essa mesma patologia. Em (LI; COOMBS, 2018) a estatística foi utilizada para comprovar de maneira cruzada que existe um impacto das comorbidades psicológicas em pacientes com queimaduras graves. (DONG; ZHANG; QIN, 2013) comprovou que pacientes com doença arterial coronariana (DAC) e apneia obstrutiva do sono (AOS) e que fazem tratamento com pressão positiva contínua nas vias aéreas (CPAP) podem prevenir eventos cardiovasculares subsequentes.

Por meio da estatística comprovou-se que formulações que contém extrato das plantas curcumina e boswellias podem ser valiosas no tratamento de osteoartrite do joelho (OA) reduzindo os sintomas da doença e se tornando uma alternativa mais segura de medicamento, como é mostrado em (BANNURU et al., 2018). O estudo (GUTTIKONDA; VADAPALLI, 2018) conclui por meio da análise estatística que a integração da ultrassonografia (USG) de múltiplos órgãos focada por meio da veia cava inferior (VCI) cardíaca e ultrassonografia renal na avaliação clínica de rotina de pacientes com dispneia tem maior acurácia para diferenciar causas de dispneia no departamento de emergência.

Em (WALDMAN et al., 2018) foi desenvolvido um software estatístico que utiliza análise topográfica de parcelas de tempo-frequência que possui a função de distinguir as oscilações que são frequentemente associados com ondas de EEG aguçadas. O software de encefalografia eletromagnética (EMEGS) foi desenvolvida em (PEYK; DE CESAREI; JUNGHÖFER, 2011) para fornecer usuários novatos, especialistas no campo de neurociência para a realização de dados de EEG e EMG, as funcionalidades que o

software oferece são um emaranhado de interfaces gráficas dedicadas ao pré-processamento, análise e visualização de dados eletromagnéticos. Já em (TADEL et al., 2011) foi utilizado o aplicativo *Brainstorm* que tem como objetivo a visualização e processamento de dados também de (EMG), tendo ênfase nas técnicas de estimativa de fonte cortical e sua integração com dados de ressonância magnética (fMRI) anatômica.

Uma revisão bibliográfica realizada em (KANG et al., 2018) foi realizado uma meta-análise a partir da análise estatística realizada no software estatístico R, sobre os dados que avaliavam os resultados clínicos e radiográficos da parte torácica superior (TS) versus a parte torácica inferior (TI) para adultos com escoliose, afirmando que não há diferença significativa entre os resultados. Em seu estudo (DELORME; SEJNOWSKI; MAKEIG, 2007) demonstrou resultados da decomposição de análise de componentes independentes (ICA) em simulações realizadas em sinais de EE. Já (BRUNET; MURRAY; MICHEL, 2011) apresentou o software *CARTOOL*, centralizando suas análises nas avaliações quantitativas de mudanças topográficas de campo ao longo do tempo, e nas propriedades espaciais de campos elétricos do cérebro. O software *CARTOOL* fornece diversas opções para a exibição de EEG, mapas com escalas dinâmicas em 2D e 3D, incluindo também a ferramenta denominada *TANOVA* (ANOVA topográfica).

O aplicativo *Brainstorm* foi utilizado em (TADEL et al., 2011) com o objetivo de visualizar e processar dados de EMG, com ênfase nas técnicas de estimativa de fonte cortical com a integração de dados de ressonância magnética (fMRI). Com o intuito de distinguir oscilações frequentemente associadas com ondas de EEG, (WALDMAN et al., 2018) desenvolveu um software estatístico que utiliza a análise topográfica de parcelas tempo-frequência. O *EEG-LAB* desenvolvido por (DELORME et al., 2011) permite que o usuário combine e compare todos os tipos de rejeições estatísticas, além de outras ferramentas complementares de coleta e processamento de dados de EEG, permitindo flexibilidade e facilidade na realização da análise estatísticas dos dados.

São inúmeros softwares existentes no mercado capazes de analisar estatisticamente sinais biomédicos, como mostrado nos estudos citados acima. Dentre esses foram destacados 5 softwares, dispostos na Tabela 1, que já foram utilizados anteriormente nas análises estatísticas dos sinais biomédicos trabalhados nesse projeto. Foram informados os nomes e os valores de aquisição de cada software.

Tabela 1: Softwares estatísticos existentes e seus respectivos valores de aquisição.

Softwares	Valores
BioEstat versão 5.3	Gratuito
R	Gratuito
Pacote Office 365 Personal - Excel	R\$239,00
WolframAlpha versão para estudantes	U\$4,75/mês
Sisvar	Gratuito
MATLAB licença individual	U\$940/ano

Fonte: Própria, 2018.

O software BioEstat (URL1) é um software gratuito voltado, principalmente, para a área de ciências biológicas e médicas, sendo destinado essencialmente a estudantes de graduação, pós-graduação, professores e pesquisados dessas áreas citadas acima.

O software R (URL2) é destinado à computação estatística e gráfica, sendo uma plataforma gratuita e que está disponível para uma vasta variedade de sistemas operacionais.

O Excel, pertencente ao pacote Office (URL3) sendo este um editor de planilhas onde é possível realizar desde operações básicas até operações aritméticas de maior complexidade, nesta plataforma é exequível realizar a construção de gráficos e utilizar a versão online do software. Este software é pago (URL4), sendo possível adquirir o pacote que inclui outros softwares ou realizar a aquisição apenas do Excel.

WolframAlpha é uma plataforma online que realiza cálculos desde as operações básicas até as mais complexas, passando por derivadas à integrais de múltiplas variáveis, sendo que também é possível construir gráficos em até três dimensões. O site apresenta um limite livre de realizações de cálculos gratuitos, e ao atingir este limiar é necessário que o usuário adquira uma versão paga, devendo assim escolher dentre as opções oferecidas, qual é aquela que melhor se adequa ao seu objetivo (URL5).

O software estatístico Sisvar é gratuito e tem como o objetivo realizar análises estatísticas (FERREIRA, 2015) passando pela análise descritiva, testes de normalidade e comparação, cálculos de probabilidade, estimadores de densidades entre outros.

MATLAB é uma plataforma que utiliza matrizes para que os cálculos desejados sejam realizados, essas operações passam desde uma simples operação aritmética a programações com comunicações com outros softwares de programação existentes, ou seja, o MATLAB é utilizado para processamentos de sinais até mesmo para utilização na área de robótica. É uma plataforma que disponibiliza 30 dias gratuitamente para um teste e que posteriormente é necessário que o usuário faça a aquisição de algum pacote oferecido pela empresa desenvolvedora (URL6).

Dentre os softwares listados na Tabela 1, o MATLAB foi utilizado para desenvolver o software de estatística proposto neste trabalho e foram utilizados BioEstat, R e Excel para validação dos resultados obtidos por meio da plataforma desenvolvida.

2.2 Metodologia

2.2.1 Arquivos de Dados Avaliados

A plataforma estatística desenvolvida apresentou como objetivo central o processamento de quantificadores realizados em (RAMOS, 2017) específicos para sinais eletroencefalográficos (EEG) comatosos e controles (normais). Todos esses quantificadores foram desenvolvidos por meio do software MATLAB.

Os arquivos de EEG usados no software do processamento estatístico estão em formato “.mat”, contendo informações de três quantificadores distintos, os quais são: Porcentagem de Contribuição de Potência (PCP), Frequência Mediana (FM) e Coerência. Portanto, cada exame avaliado apresenta pelo menos três informações distintas que devem ser analisadas sob a ótica estatística.

Porcentagem de Contribuição de Potência (PCP) é um quantificador que informa a quantidade de energia compreendida em uma determinada faixa de frequência. Frequência Mediana (FM) é o valor de frequência em que 50% da potência calculada estão nas frequências mais baixas e 50% estão em frequências mais altas. Coerência é referente ao grau de similaridade de fase entre dois sinais, em particular para um sinal EEG, é baseada nos hemisférios direito e esquerdo.

Em relação aos registros analisados, cada registro de EEG apresenta 20 eletrodos que devem ser submetidos aos cálculos de PCP, FM e Coerência. Além disso, são avaliados diferentes trechos do mesmo indivíduo, com o objetivo de ampliar a análise do sinal e ao mesmo tempo garantir que os critérios de estacionariedade sejam mantidos ao serem submetidos aos cálculos dos quantificadores. Considera-se, portanto, que cada registro de EEG é composto por 20 canais e N trechos (ou épocas). Devido à essas características se faz imprescindível a análise estatística.

2.2.2 Funcionalidades do Software

O software apresentado neste trabalho foi desenvolvido para realizar análise estatística descritiva, testes de normalidade e testes de hipótese.

2.2.2.1 Estatística Descritiva

O intuito da análise descritiva é gerar conclusões a serem adotadas sobre uma determinada população, podendo também testar se hipóteses levantadas sobre esta população são aceitáveis ou não (SCHMILDT, 2007). Sendo assim para análise de dados quantitativos foram escolhidas as principais ferramentas estatísticas que podem caracterizar de forma eficaz uma determinada população, totalizando 14 informações listadas de a) à n).

a) *Tamanho da Amostra*: É a quantidade de épocas que aquele arquivo em análise possui. A função utilizada na programação para essa variável foi *length()*.

b) *Mínimo*: Refere-se ao menor elemento da amostra. Foi utilizada a função *min()*, para realizar o cálculo de mínimo da amostra em análise.

c) *Máximo*: A medida de máximo mostra o maior elemento da amostra. Para a realização do cálculo de máximo da amostra, foi utilizado a função *max()*.

d) *Amplitude Total*: É a medida definida como sendo a diferença entre o maior e o menor valor do conjunto de dados.

$$AmpTotal = X_{Máximo} - X_{Mínimo} \quad (4)$$

Onde:

- AmpTotal: Amplitude Total do conjunto de dados
- $X_{Máximo}$: Valor máximo da população
- $X_{Mínimo}$: Valor mínimo da população

e) *Mediana*: valor que separa a metade maior e a metade menor de uma amostra, uma população ou uma distribuição de probabilidade. Por isso que a mediana é considerada o segundo quartil.

$$P_{Md} = 0,5 (n + 1) \quad (5)$$

$$Md = X_{P_{Md}} \quad (6)$$

Onde:

- PMd: Posição da mediana
- n: Número de elementos da população
- Md: Mediana
- XPMd: Elemento da posição mediana

Para que o cálculo de mediana da amostra fosse realizado, foi utilizada a função *median()*.

f) *Primeiro Quartil*: é a medida que deixa 25% das observações abaixo e 75% acima.

$$Q_j = X_k + \left(\frac{n+1}{4} - k \right) (X_{k+1} - X_k) \quad (7)$$

Onde:

- Q_j : Quartil analisado, neste caso, o primeiro quartil
- X_k : Elemento do conjunto de dados na posição determinada pela variável k.
- n: Número de elementos da população
- k: Variável maior inteiro ou igual a $\frac{j(n+1)}{4}$
- X_{k+1} : Elemento do conjunto de dados na posição determinada pela variável k+1.

A realização do cálculo do primeiro quartil da amostra, foi utilizando a função *quantile()*, para uma probabilidade $p = 0,25$ ou 25%.

g) *Terceiro Quartil*: é a medida que deixa 75% das observações abaixo e 25% acima.

$$Q_j = X_k + \left(\frac{n+1}{4} - k \right) (X_{k+1} - X_k) \quad (8)$$

Onde:

- Q_j : Quartil analisado, neste caso, o terceiro quartil
- X_k : Elemento do conjunto de dados na posição determinada pela variável k.
- n: Número de elementos da população

- k: Variável maior inteiro ou igual a $\frac{j(n+1)}{4}$

- X_{k+1} : Elemento do conjunto de dados na posição determinada pela variável k+1

A realização do cálculo do terceiro quartil da amostra, foi utilizando a função *quantile()*, para uma probabilidade $p = 0,75$ ou 75%. A maneira como a plataforma de programação calcula a probabilidade dos quartis é determinada para equação (9).

$$quantiles = \left(\frac{0.5}{n}\right), \left(\frac{1.5}{n}\right), \dots, \left(\frac{[n-0.5]}{n}\right) \quad (9)$$

Onde:

- *quantiles*: Quartil em análise

- n: número de elementos da amostra analisada

h) *Média*: medida que mostra para onde se concentram os dados de uma distribuição como o ponto de equilíbrio, podendo ser interpretada como um valor significativo de uma lista de números. Para o cálculo da média da amostra em análise, foi utilizado a função *mean()*.

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (10)$$

Onde:

- $\bar{X}$: Média Aritmética

- n: Número de elementos do conjunto de dados

- x_i : Elementos da amostra

i) *Variância*: medida de dispersão estatística que indica o “quão longe” em geral os valores encontrados estão do valor esperado. A função *var()*, foi utilizada para a realização da variância da amostra que estava em análise.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1} \quad (11)$$

Onde:

- S^2 : Variância

- n: Número de elementos da população em análise

- x_i : Elementos da população em análise

- $\bar{X}$: Média Aritmética

j) *Desvio Padrão*: medida de dispersão em torno da média populacional ou da média amostral de uma variável aleatória. Para se obter o valor de desvio padrão da amostra, foi utilizada a função *std()*, da variância, essa função tem o intuito de realizar a raiz quadrada da variável que está dentro dos parênteses.

$$S = \sqrt{S^2} \quad (12)$$

Onde:

- S: Desvio Padrão

- S²: Variância

k) *Desvio Padrão da Mediana*: medida de dispersão em torno da mediana populacional ou da mediana amostral de uma variável aleatória.

$$DesPMd = \sqrt{\frac{\sum_{i=1}^n (x_i - Md)^2}{n-1}} \quad (13)$$

Onde:

- DesPMd : Desvio Padrão da Mediana
- n: Número de elementos da população em análise
- x_i : Elementos da população em análise
- Md: Mediana

l) *Coefficiente de Variação*: medida padronizada de distribuição de probabilidade ou de uma distribuição de frequências, frequentemente expresso em porcentagem (%).

$$CV = \frac{S}{\bar{X}} \text{ ou } CV = \frac{S}{\bar{X}} \times 100\% \quad (14)$$

Onde:

- CV: Coeficiente de Variação
- S: Desvio Padrão
- $\bar{X}$: Média Aritmética

m) *Assimetria*: medida que caracteriza como e quanto à distribuição de dados ou frequências se afasta da condição de assimetria. Assimetrias positivas são alongadas a direita e assimetrias negativas são alongadas a esquerda. A função utilizada para a realização do cálculo da assimetria da amostra analisada é *skewness()*. A faixa padrão do *default* da plataforma de programação utilizada é de 0 à 1, sendo que o 1 é o padrão utilizado quando o usuário não define o viés sistemático.

$$As = \frac{\bar{X} - Mo}{S} \quad (15)$$

Onde:

- As: Assimetria
- S: Desvio Padrão
- $\bar{X}$: Média Aritmética
- Mo: Moda

n) *Curtose*: medidas que caracterizam a forma da distribuição quanto ao seu achatamento, tendo achatamentos como platicúrticas (que possui o centro mais planar), mesocúrticas (que possui uma distribuição considerada normal) e leptocúrticas (que possui o centro mais pontudo). A função *kurtosis()*, é utilizada para

que o cálculo da variável curtose seja realizada, a faixa de normalização padrão adotada pela plataforma de programação para este cálculo, varia numa faixa de 0 à 3.

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} \quad (16)$$

Onde:

- K : Curtose
- Q_3 : Terceiro quartil
- Q_1 : Primeiro quartil
- P_{90} : 90º percentil
- P_{10} : 10º percentil

2.2.2.3 Teste de Normalidade

Os testes de normalidade têm como objetivo verificar se a distribuição de probabilidade em associação a um determinado conjunto de dados se aproxima ou não de uma distribuição normal (MONTGOMERY, 2003). A distribuição normal mostra-se vantajosa, porque seu modelo populacional é conexo à diversos fenômenos aleatórios, tornando mais simples a obtenção de estimadores por intervalo e testes exatos (CIRILLO; FERREIRA, 2003). Os principais testes usados em análises biomédicas são Shapiro-Wilk, Anderson-Darling e Kolmogorov-Smirnov (LEOTTI; BIRCK; RIBOLDI, 2005), portanto esses foram os escolhidos para constar na plataforma estatística aqui desenvolvida.

O Teste de Anderson-Darling mede se os dados seguem uma distribuição específica. Em geral, quanto melhor a distribuição se ajusta aos dados, menor a estatística AD. Os p -valores calculados a partir da estatística ajudam a determinar qual modelo de distribuição deverá ser usado para uma análise de capacidade ou uma análise de confiabilidade. A estatística AD também é utilizada para testar se uma amostra de dados é proveniente de uma população com uma distribuição especificada. Por exemplo, para testar se os dados analisados atendem à suposição de normalidade para um teste t. A função utilizada para a realização do teste de Anderson-Darling utilizada foi *adtest()*, onde a hipótese nula (resultado do teste igual a $h = 0$) é que a população em análise possui uma distribuição normal e a hipótese contrária defende que a população possui uma distribuição não normal, a hipótese nula pode ser aceita utilizando um nível de significância de 5%.

$$AD^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]}{F(x)(1 - F(x))} dF(x) \quad (17)$$

Onde:

- AD^2 : Anderson-Darling
- n: Número de elementos
- $F_n(x)$: Função de distribuição acumulada empírica definida como

$$F_n(x) = \begin{cases} 0, & \text{se } x < x_{(1)} \\ \frac{k}{n}, & \text{se } x_{(k)} \leq x < x_{(k+1)} \\ 1, & \text{se } x \geq x_{(n)} \end{cases}$$

- Sendo $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$, são as estatísticas de ordem da amostra.

O teste de Jarque-Bera é um teste de bondade de ajuste se os dados da amostra têm a afinidade e a segmentação correspondentes a uma distribuição normal. Se os dados provêm de uma distribuição normal, a estatística JB tem, assintoticamente, uma distribuição de qui-quadrado com dois graus de liberdade, de modo que a estatística pode ser usada para testar a hipótese de que os dados são de uma distribuição normal. A função utilizada para a realização do cálculo do teste de Jarque-Bera foi *jbtest()*, possuindo com hipótese nula (resultado do teste igual a $h = 0$) uma população com distribuição normal com uma variância e média desconhecida e a hipótese alternativa é contrária a hipótese nula, ao rejeitar ou aceitar a hipótese nula é utilizado um nível de significância de 5%.

$$JB = n \left[\frac{As}{6} + \frac{(K-3)^2}{24} \right] \quad (18)$$

Onde:

- JB : Jarque-Bera
- n : Número de elementos
- As: Assimetria
- K : Curtose

O teste Kolmogorov–Smirnov (também conhecido como teste KS ou teste K–S) é um teste não paramétrico sobre a igualdade de distribuições de probabilidade contínuas e unidimensionais que pode ser usado para comparar uma amostra com uma distribuição de probabilidade de referência (teste K–S uniamostrai) ou duas amostras uma com a outra (teste K–S biamostrai). A estatística de Kolmogorov–Smirnov quantifica a distância entre a função distribuição empírica da amostra e a função distribuição acumulada da distribuição de referência ou entre as funções distribuição empírica de duas amostras. O teste de Kolmogorov–Smirnov pode ser modificado para servir como um teste da qualidade do ajuste. Isto equivale a tornar a média e a variância da distribuição de referência iguais aos estimados das amostras, sabendo que usar isto para definir a distribuição de referência específica muda a distribuição nula da estatística. A função *kstest()* foi utilizada para a realização do teste em questão, onde este possui como hipótese nula a distribuição padrão normal da população em análise e a hipótese contrária afirma que a distribuição da população possui uma

distribuição padrão não normal, a aceitação ou rejeição da hipótese nula é realizada em um nível de significância de 5%.

$$F_n(n) = \frac{1}{n} \sum_{i=1}^n I_{[-\infty, x]}(X_i) \quad (19)$$

$$K - S = \sup_x |F_x(x) - F(x)| \quad (20)$$

Onde:

- K-S: Kolmogorov-Smirnov
- $F_n(x)$: Função Distribuição Empírica
- n: Número de elementos
- X_i : Elementos independentes e identicamente distribuídos
- $I_{[-\infty, x]}(X_i)$: Função indicadora, igual a 1 se $X_i \leq x$ e igual a 0
- $\sup_x$: Supremo do conjunto de distâncias
- $F(x)$: Função acumulada

2.2.2.4 Teste de Comparação

Os testes de hipóteses são utilizados para a tomada de decisão de aceitar ou rejeitar uma hipótese estatística, levando em consideração a base de dados amostrais, em estatística a hipótese é uma suposição quanto ao valor de um parâmetro da população, a qual será verificada por meio da utilização de um teste estatístico, esses testes são denominados testes de hipótese ou testes de comparação (SCHMILDT, 2007). Foram definidos três testes de hipóteses: Teste de Friedman, Teste de Mann-Whitney e Teste T. O Teste de Friedman é um teste estatístico não paramétrico utilizado para comparar dados amostrais vinculados (avaliar mais de uma vez o mesmo indivíduo), ou seja, no teste de Friedman, o estudo dos dados é feito analisando as linhas (postos/blocos) em relação às colunas que são ordenadas de acordo com o que se deseja investigar, por exemplo, a classificação de qualidade de x tochas de soldagem em relação aos y soldadores que as utilizam, ou seja, um soldador (linha) utiliza 3 tochas de soldagem (colunas) diferentes e classifica dentre elas qual a melhor, outro soldador utiliza as mesmas tochas e classifica a melhor e assim por diante. Este teste só poderá ser utilizado pelo usuário quando os dados a serem comparados forem homogêneos, ou seja, a mesma quantidade de linhas e colunas da matriz dos dados em comparação deve ser igual. Utilizando a função *friedman()*, foi possível realizar o teste de Friedman.

$$Fried = \frac{12b}{k(k+1)} \sum_{j=1}^k \left(\frac{R_j}{b} - \frac{k+1}{2} \right)^2 = \left[\frac{12}{bk(k+1)} \sum_{j=1}^k R_j^2 \right] - 3b(k+1) \quad (21)$$

Onde:

- Fried: Teste de Friedman
- b : Blocos separados

- k : Número de elementos ordenados do menor para o maior
- R_j : Rank dos elementos
- j: Números de coluna (cada tratamento)

O Teste de Mann-Whitney é um teste não paramétrico que normalmente é aplicado para duas amostras independentes, é utilizado para testar a heterogeneidade de duas amostras ordinais, ou seja, é utilizado para testar se uma população tende a ter valores maiores do que a outra, ou se elas têm a mesma mediana. A função *ranksum()* foi a utilizada para a realização do teste de Mann-Whitney. O teste de hipótese afirma que a amostra analisada possui uma distribuição contínua com mediana igual (hipótese nula, $h = 0$) e a hipótese alternativa é contrária a hipótese nula. O resultado encontrado do teste de hipótese possui um nível de significância de 5%.

$$U_m = S_m - \frac{1}{2}m(m + 1) \quad (22)$$

$$U_n = S_n - \frac{1}{2}n(n + 1) \quad (23)$$

Onde:

- U_m : Teste de Mann-Whitney para a primeira amostra ou amostra X.
- U_n Teste de Mann-Whitney para a segunda amostra ou amostra Y.
- S_m Soma das fileiras das observações da primeira amostra ou amostra X.
- S_n : Soma das fileiras das observações da segunda amostra ou amostra Y.
- m: Tamanho da amostra da população da primeira amostra ou amostra X.
- n: Tamanho da amostra da população da segunda amostra ou amostra Y.

Por fim o Teste T ou Teste T de Student, refere-se a um teste de hipótese que usa conceitos estatísticos para rejeitar ou não uma hipótese nula quando a estatística de teste (t) segue uma distribuição t de Student, normalmente é utilizado quando a estatística do teste segue uma distribuição normal, mas a variância da população é desconhecida, usando neste caso, a variância amostral. O Teste t também pode ser utilizado para testar a significância de coeficientes de regressões. Em geral esse teste é usado para confirmar se a variável que está sendo usada na regressão está realmente contribuindo para a estimativa. A função utilizada para a realização do teste T foi *ttest()*, tendo como retorno de hipótese nula (resultado $h = 0$) dados da população em análise proveniente de uma distribuição normal possuindo média nula e variância desconhecida e tendo com hipótese contrária, uma população com uma média diferente de zero. O resultado do teste é dado em um nível de significância de 5%.

$$t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \quad (24)$$

Onde:

- t : Teste T

- $\bar{X}$: Média Aritmética
- μ_0 : Valor fixo usado para comparação com a média da amostra
- S: Desvio padrão
- n: Número de elementos da amostra

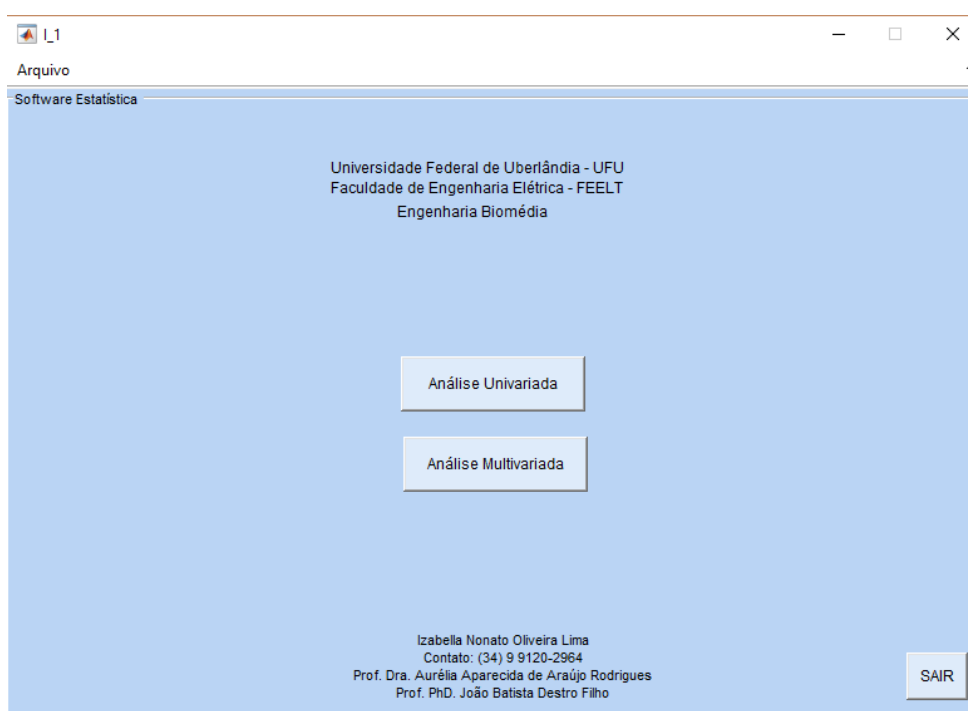
É importante destacar que para o Teste de Friedman e o Teste T, os vetores a serem comparadas precisam obrigatoriamente apresentar o mesmo tamanho, ou seja, devem possuir o mesmo número de linhas ou colunas.

2.3 Resultados

O software de estatística demonstrado neste trabalho foi baseado em cinco interfaces sendo essas separadas em interface inicial, interface de estatística descritiva, interfaces de visualização e interface de testes. A interface inicial (Figura 1) é aquela que será mostrada ao usuário assim que este começar a utilizar o software, nela existem dois botões, sendo o primeiro destinado à Análise Univariada (seta vermelha) que foi programada no desenvolvimento desse trabalho e o outro Análise Multivariada (seta verde), que será programado futuramente.

Ainda na interface inicial, no canto superior esquerdo, é possível através da opção “Arquivo”, “Tutorial”, ler o tutorial onde se explica todo o funcionamento do software, passando por todas as interfaces, antes de começar o processamento em si.

Figura 1: Interface inicial do software de estatística.



Fonte: Própria, 2018.

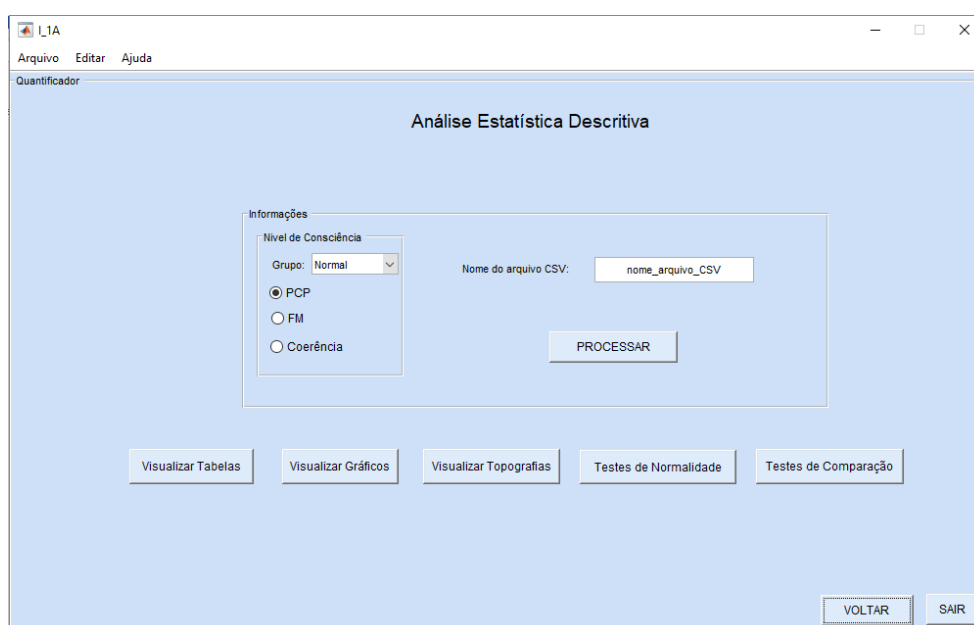
2.3.1 Interfaces Estatística Descritiva

Ao clicar no botão “Análise Univariada”, a interface Estatística Descritiva será apresentada, como é demonstrado pela Figura 2. Nessa existe um botão principal (denominado “Processar”) e cinco botões secundários (visualizar tabelas, visualizar gráficos, visualizar topografias, teste de normalidade e teste de comparação). Os botões “Voltar” e “Sair” se referem a voltar à interface inicial (Figura 2) ou fechar a interface atual, respectivamente.

O botão “Processar” permite que todo o cálculo descritivo seja realizado, portanto são geradas pelo menos 14 informações distintas (Tamanho da amostra, Mínimo, Máximo, Amplitude Total, Mediana, Primeiro Quartil, Terceiro Quartil, Média, Variância, Desvio Padrão, Desvio Padrão da Mediana, Coeficiente de Variação, Assimetria e Curtose). Antes do usuário clicar nesse botão deve-se informar a qual grupo o registro (ou os registros) a ser analisado estatisticamente se refere: Grupo coma ou Grupo normal. Destaca-se que a análise individual se refere ao resumo descritivo daquele registro que apresentou N informações do quantificador. Já a análise em grupo leva em consideração não apenas as N informações de cada indivíduo, bem como a quantidade de registros analisadas. Além de definir o grupo, o usuário precisa também informar qual quantificador ele deseja avaliar (PCP, FM ou Coerência).

Posteriormente é necessário que no campo “Nome do arquivo CSV”, seja digitado o nome da planilha que contém a(s) referência(s) do(s) arquivo(s) ‘.mat’ que se deseja realizar a avaliação descritiva. Por fim, para que o cálculo da estatística descritiva seja realizado, é necessário clicar no botão “Processar”. Assim que esse for pressionado o arquivo digitado no campo destinado ao nome do arquivo, como mostra a Figura 2, será aberto e o usuário poderá selecionar os registros que deseja processar.

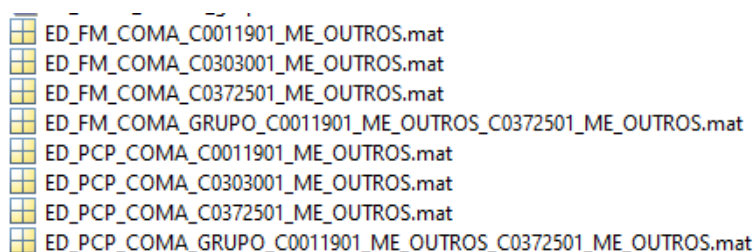
Figura 2: Interface Estatística Descritiva.



Fonte: Própria, 2018.

Ao final do processamento uma mensagem de alerta de término será mostrada e então o usuário poderá localizar os arquivos que foram gerados. Caso o processamento tenha sido feito apenas para um registro será gerado um arquivo “.mat” com todas as informações descritivas calculadas, cujo nome é padronizado no formato “ED_Quantificador_Grupo_NomeRegistro. Por outro lado, se o processamento reunir um grupo de Y registros, serão gerados Y+1 arquivos (um arquivo referente ao estudo do grupo e portanto o nome será acrescido da palavra “GRUPO” e Y arquivos referentes ao processamento individual de cada registro que faz parte do grupo, seguindo o padrão de nomenclatura. Um exemplo de como os arquivos resultantes do processamento da estatística descritiva ficará está demonstrado na Figura 3.

Figura 3: Exemplo do nome do arquivo gerado ao final do cálculo da estatística descritiva.



ED_FM_COMA_C0011901_ME_OUTROS.mat
 ED_FM_COMA_C0303001_ME_OUTROS.mat
 ED_FM_COMA_C0372501_ME_OUTROS.mat
 ED_FM_COMA_GRUPO_C0011901_ME_OUTROS_C0372501_ME_OUTROS.mat
 ED_PCP_COMA_C0011901_ME_OUTROS.mat
 ED_PCP_COMA_C0303001_ME_OUTROS.mat
 ED_PCP_COMA_C0372501_ME_OUTROS.mat
 ED_PCP_COMA_GRUPO_C0011901_ME_OUTROS_C0372501_ME_OUTROS.mat

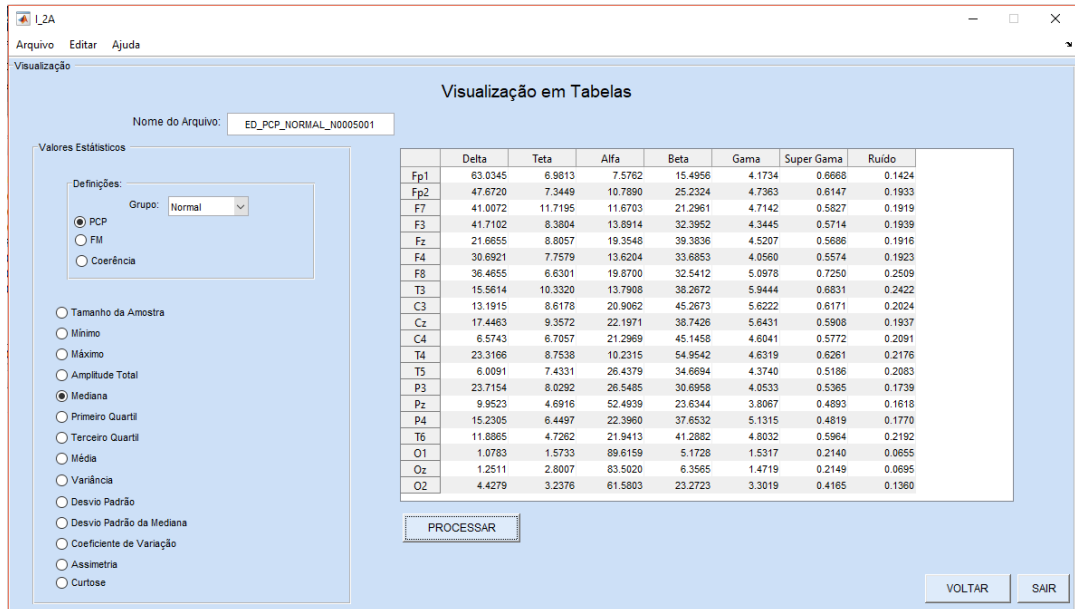
Fonte: Própria, 2018.

O botão “Visualizar Tabelas” permite que o usuário visualize todos resultados das variáveis obtidas através do processamento estatístico, em formato de tabela. Para isso é necessário que no campo destinado “Nome do Arquivo” seja digitado o nome do arquivo advindo do cálculo da estatística descritiva, em seguida, deve-se escolher o grupo (normal/coma), e depois o quantificador de análise. O processo passo é a escolha da variável que se deseja visualizar em formato de tabela, sendo assim, no canto esquerdo é possível escolher dentre as 14 opções àquela que seja de interesse do usuário, por fim deve-se clicar no botão “Processar” que se encontra na interface. Para visualizar outras variáveis em formato de tabela, basta escolher a variável de desejo e clicar no botão “PROCESSAR”. A Figura 4, mostra um exemplo de como é a visualização dos resultados no formato de tabela. Caso o usuário queira voltar para a interface apresentada na Figura 2 ou sair da interface atual, basta clicar em “voltar” ou em sair, respectivamente.

O botão “Visualizar Gráficos” permite que seja feita a visualização dos dados das variáveis advindas do processamento estatístico, em formato de gráficos. Portanto, primeiramente deve-se digitar o nome do arquivo advindo do processamento estatístico descritivo. Logo após, deve-se escolher o grupo (normal/coma) e em seguida escolher o quantificador em questão o usuário deseja analisar. A próxima etapa é a escolha de um dos três gráficos (boxplot, histograma e dispersão), depois da escolha do tipo de gráfico a ser analisado, deve-se escolher qual variável estatística será observado em forma de gráfico. Por fim, para que os resultados sejam demonstrados em gráficos, deve-se clicar no botão “PROCESSAR”, um exemplo de como

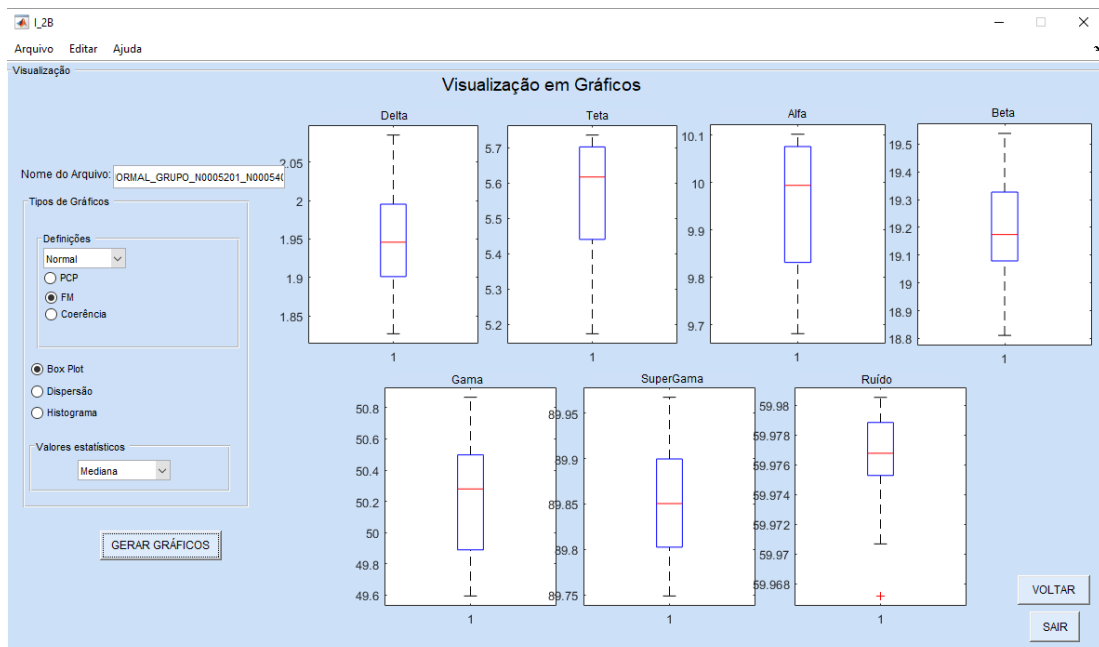
é a visualização é mostrado na Figura 5. Caso o usuário queira voltar para a interface apresentada na Figura 2 ou sair da interface atual, basta clicar em “voltar” ou em sair, respectivamente.

Figura 4: Interface Visualizar Tabela.



Fonte: Própria, 2018.

Figura 5: Interface Visualização em Gráficos.

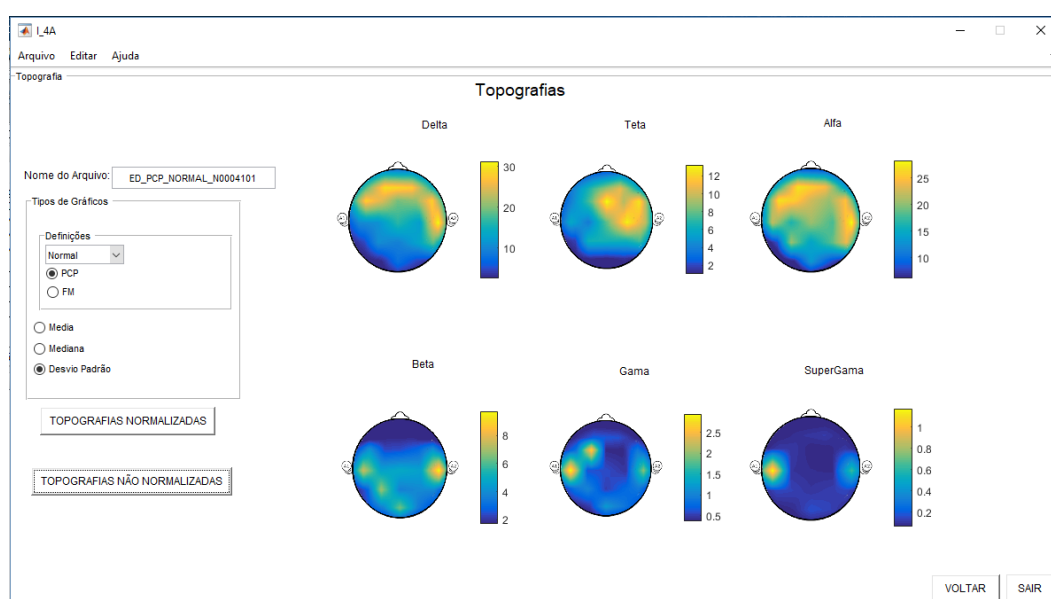


Fonte: Própria, 2018.

O botão “Visualizar Topografias” permite que a interface de visualização de topografias seja aberta e consequentemente permite que o usuário visualize os dados das variáveis estatísticas advindas do processamento estatístico em formato topográfico, porém é necessário primeiramente, como nas duas outras

interfaces de visualização, digitar o nome do arquivo gerado no processamento estatístico. Depois, deve-se escolher o grupo (normal/coma) e em seguida, escolher o quantificador (PCP, FM, Coerência). Em seguida, deve-se escolher dentre as opções qual variável estatística deseja-se visualizar em forma topográfica, e por fim, escolher se deseja uma visualização de topografia normalizada ou topografia não normalizada. Na Figura 6 é demonstrado um exemplo do resultado para a visualização de topografias. Caso o usuário queira voltar para a interface apresentada na Figura 2 ou sair da interface atual, basta clicar em “voltar” ou em sair, respectivamente.

Figura 6: Interface de Visualização de Topografias.



Fonte: Própria, 2018.

2.3.2 Interfaces Testes de Normalidade

Ao clicar no botão de Teste de Normalidade que está na interface mostrada pela Figura 2, a interface de teste de normalidade aparecerá na tela do computador, e nesta interface pode-se realizar três testes de normalidade diferentes. Para realização do cálculo do teste de normalidade é necessário que primeiramente se insira o nome da planilha em formato “.CSV” onde contém o nome do(s) arquivo(s) que será(serão) processados, na Figura 33 é possível ver uma seta indicando o local onde o nome desta planilha deve ser digitado. O segundo passo é escolher o grupo que está sendo analisado (normal/coma), e o quantificador em análise (PCP, FM e Coerência). O próximo passo é a escolha do teste de normalidade a ser feito, na interface existem três tipos diferentes de testes de normalidades (Teste Anderson Darling, Teste Jarque-Bera e Teste Kolmogorov) o último passo é a escolha de qual botão apertar, a interface oferece dois botões para o processamento, um destinado ao cálculo do teste de normalidade para um indivíduo ou para um grupo de indivíduo, segundo botão. Na Figura 7 é possível observar a interface de Teste de Normalidade.

Figura 7: Interface do Teste de Normalidade.

Fonte: Própria, 2018.

Ao final do processamento uma mensagem de alerta de término será mostrada e então o usuário poderá localizar os arquivos que foram gerados. Caso o processamento tenha sido feito apenas para um registro será gerado um arquivo “.mat” com todas as informações descritivas calculadas, cujo nome é padronizado no formato “TesteRealizado_Quantificador_Grupo_NomeRegistro”, onde este “grupo” faz referência à normal/coma. Por outro lado, se o processamento reunir um grupo de Y registros, serão gerados Y+1 arquivos (um arquivo referente ao estudo do grupo) e, portanto, o nome será acrescido da palavra “GRUPO” e Y arquivos referentes ao processamento individual de cada registro que faz parte do grupo, seguindo o padrão de nomenclatura.

A Figura 8 mostra um exemplo de como será salvo o arquivo advindo do cálculo do teste de normalidade, é possível observar um exemplo do nome do arquivo para o processamento individual e em grupo.

Figura 8: Exemplo do arquivo gerado após o término do processamento do teste de normalidade.

Anderson-Darling_PCP_NORMAL_GRUPO_N0004701_N0005201.mat
Anderson-Darling_PCP_NORMAL_N0005501.mat

Fonte: Própria, 2018.

2.3.3 Interfaces Testes de Hipóteses

O último botão da interface mostrada na Figura 2 é referente ao cálculo dos testes de hipótese/comparação, ao clicar neste botão a interface de Teste de Comparação será mostrada no visor do

computador. Sendo assim, o primeiro passo a ser tomado antes do processamento do teste de comparação ser realizado é escolher o nome que o arquivo com os resultados advindos da realização do teste de comparação terá. Em seguida deve-se escolher qual(is) planilha(s) em formato “.CSV” que deseja-se analisar, sendo necessário digitar na região “Arquivo_CSV_1”, o nome da primeira planilha em formato “.CSV” que será analisada e em seguida na área “Arquivo_CSV_2”, transcrever o nome da segunda planilha que deseja-se analisar. Lembrando que é possível digitar nomes iguais nas duas áreas, ou seja, se o objetivo for analisar arquivos que estejam na mesma planilha é possível digitar o nome da mesma planilha nos dois campos, sendo assim, tem-se que digitar o nome do mesmo arquivo na região “Arquivo_CSV_1” e em seguida no local “Arquivo_CSV_2”.

Logo após o nome das planilhas serem digitados no local específico, deve-se escolher o grupo de análise (normal/coma), escolher o quantificador a ser analisado e escolher o teste de comparação que se deseja realizar existem três tipos de testes de comparação diferentes, sendo eles, Teste de Friedman, Mann-Whitney e o Teste T. Por fim, para que o cálculo do teste de comparação seja efetuado é necessário que o botão “PROCESSAR” seja clicado.

Observação 1: Se o intuito for comparar exames do tipo “normal” com o tipo “coma”, necessariamente deve-se escolher a opção “coma”.

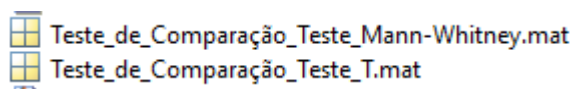
Observação 2: Para que o teste de Friedman e o teste T é obrigatoriamente necessário que os registros sejam homogêneos, ou seja, as matrizes precisam possuir o mesmo número de linhas e de colunas.

A Figura 10 mostra um exemplo de como será salvo o arquivo advindo do cálculo do teste de normalidade, é possível observar um exemplo do nome do arquivo para o processamento individual e em grupo.

Figura 9: Interface de Teste de Comparação.

Fonte: Própria, 2018.

Figura 10: Exemplo de nome do arquivo que armazena os resultados advindos do processamento dos testes de comparação.



Fonte: Própria, 2018.

2.4 Discussão dos Resultados

A validação dos dados advindos do software estatístico descrito neste trabalho foi realizada confrontando os resultados do processamento encontrados na plataforma desenvolvida com softwares estatísticos existentes. Para evitar qualquer tipo de favorecimento à algum software estatístico existente, os softwares aqui utilizados para a validação receberão o nome de Software A e software B.

Para âmbito de comparação todos as ferramentas programadas foram validadas comparando com softwares já renomados, contudo foram definidas apenas cinco situações de processamento para constar nesse documento, sendo elas:

- Situação 1: Estatística descritiva, Grupo normal, registro individual, quantificador PCP.
- Situação 2: Estatística descritiva, Grupo normal, dois registros agrupados, quantificador FM.
- Situação 3: Estatística descritiva, Grupo coma, com três indivíduos, quantificador Coerência.
- Situação 4: Teste de normalidade, Kolmogorov-Smirnov, Grupo coma, para três eletrodos, para um indivíduo e quantificador FM.
- Situação 5: Comparação entre o grupo Coma (cinco indivíduos) e o grupo Controle (três indivíduos) por meio do Teste de Mann-Whitney, para o quantificador coerência e para dois pares de eletrodos.

2.4.1 Situação 1

Nesta validação foram confrontados valores oriundos do processamento da estatística descritiva, comparando assim todos os resultados das variáveis estatísticas disponíveis. Na Tabela 2 estão dispostos os resultados advindos do processamento do software estatístico proposto nesse trabalho e dos softwares A e B, para uma situação de grau de consciência normal, quantificador PCP, para o eletrodo F7 (escolhido de forma aleatória) e para um indivíduo. Da comparação entre os três softwares é possível notar que todos os resultados deram iguais e/ou próximos com diferença na quarta casa decimal após a vírgula, essa diferença se dá pelo método de arredondamento que cada software utiliza. As únicas variáveis estatísticas que deram valores distintos foram, “Primeiro Quartil”, “Terceiro Quartil”, “Assimetria” e “Curtose”, isso ocorreu porque cada software estatístico tem um *default* (limiar) diferente para essas variáveis, tendo em mente o limiar que o MATLAB utiliza, sabe-se que para os quartis é utilizado uma equação para realização do cálculo para quando

a matriz de dados é par ou ímpar, como pode-se observar na equação (9). Para Assimetria o limiar que o MATLAB utiliza varia de 0 a 1 e para Curtose esse limiar pode variar de 0 a 3.

Tabela 2: Resultados obtidos da comparação do software estatístico, software A e B para o cálculo da estatística descritiva, para a situação de um indivíduo com grau de consciência normal e quantificador PCP.

Variáveis Estatísticas	Softwares	Ritmos Cerebrais						
		Delta	Teta	Alfa	Beta	Gama	Supergama	Ruído
Tamanho da Amostra	Estatístico	10	10	10	10	10	10	10
	A	10	10	10	10	10	10	10
	B	10	10	10	10	10	10	10
Mínimo	Estatístico	1.6798	1.3961	9.5333	1.8491	0.7199	0.1030	0.0371
	A	1.6798	1.3960	9.5333	1.8491	0.7199	0.1030	0.0371
	B	1.6797	1.3960	9.5333	1.8490	0.7198	0.1029	0.0371
Máximo	Estatístico	84.9327	15.5950	90.7425	15.4732	3.5761	0.4889	0.1860
	A	84.9327	15.5950	90.7425	15.4732	3.5761	0.4889	0.1860
	B	84.9327	15.5950	90.7424	15.4732	3.5760	0.4888	0.1860
Amplitude Total	Estatístico	83.2529	14.1990	81.2091	13.6241	2.8562	0.3859	0.1489
	A	83.2529	14.1990	81.2092	13.6241	2.8562	0.3859	0.1489
	B	83.2529	14.1990	81.2014	13.6241	2.8561	0.3859	0.1488
Mediana	Estatístico	23.3360	5.4940	50.3060	4.3318	1.4786	0.1836	0.0566
	A	23.3361	5.4940	50.3060	4.3318	1.4787	0.1836	0.0566
	B	23.3360	5.4940	50.3059	4.3318	1.4786	0.1836	0.0565
Primeiro Quartil	Estatístico	8.9736	3.581	29.3774	2.4554	1.1493	0.1162	0.0438
	A	9.6437	3.6468	33.3055	2.5826	1.1629	0.1224	0.0459
	B	8.3907	3.3890	29.1842	2.4437	1.0832	0.1151	0.0426
Terceiro Quartil	Estatístico	59.2016	12.0928	81.5918	6.1400	2.0114	0.2641	0.0907
	A	53.9059	11.7843	76.5956	5.7519	1.9169	0.2544	0.0872
	B	59.3684	12.3539	83.0705	6.9427	2.1365	0.2105	0.0740
Média	Estatístico	31.7644	7.4304	53.4666	5.3893	1.6648	0.2105	0.0741
	A	31.7644	7.4304	53.4666	5.3893	1.6648	0.2105	0.0741
	B	31.7643	7.4304	53.4665	5.3892	1.6648	0.2105	0.0740
Variância	Estatístico	793.0477	25.3601	736.0737	17.4159	0.7276	0.0137	0.0020
	A	793.0470	25.3601	736.0743	17.4158	0.7276	0.0137	0.0020
	B	793.0477	25.3601	736.0737	17.4159	0.7276	0.01366	0.0020

Variáveis Estatísticas	Softwares	Ritmos Cerebrais						
		Delta	Teta	Alfa	Beta	Gama	Supergama	Ruído
Desvio Padrão	Estatístico	28.1611	5.0359	27.1307	4.1732	0.8530	0.1169	0.0450
	A	28.1611	5.0359	27.1307	4.1732	0.8530	0.1169	0.0450
	B	28.1611	5.0358	27.1306	4.1732	0.8530	0.1168	0.0449
Desvio Padrão da Mediana	Estatístico	29.5293	5.4338	27.3345	4.3195	0.8753	0.1203	0.0486
	A	-	-	-	-	-	-	-
	B	29.5292	5.4338	27.3344	4.3195	0.8752	0.1202	0.0485
Coeficiente de Variação	Estatístico	88.65%	67.77%	50.74%	77.43%	51.23%	55.53%	60.72%
	A	88.66%	67.77%	50.74%	77.44%	51.24%	55.53%	60.73%
	B	88.65%	67.77%	50.74%	77.43%	51.23%	55.53%	60.72%
Assimetria	Estatístico	0.6605	0.3859	-0.045	1.5707	1.1437	1.3761	1.6568
	A	0.7832	0.4576	-0.0054	1.8626	1.3562	1.6320	1.9647
	B	0.7832	0.4575	-0.0053	1.8626	1.3562	1.6319	1.9647
Curtose	Estatístico	2.1409	1.6069	1.9033	4.4506	3.5350	4.2177	4.9004
	A	-0.5545	-1.4985	-0.9745	3.5287	1.9102	3.1173	4.3235
	B	-0.5545	-1.4985	-0.9745	3.5287	1.9100	3.1169	4.3239

Fonte: Própria, 2018.

Estatístico: Software desenvolvido no trabalho.
A e B: Softwares estatísticos já existentes.

2.4.2 Situação 2

A segunda situação da validação dos dados é referente ao quantificador frequência mediana (FM), onde a *Tabela 3* faz menção aos resultados da comparação entre os três softwares para uma situação em que o conjunto de dados comparado é advindo do cálculo estatístico descritivo de um grupo de dois indivíduos normais. Observando os resultados obtidos da comparação entre os três softwares é possível notar que todos os resultados deram iguais e/ou próximos com diferença na quarta casa decimal após a vírgula, essa diferença se dá pelo método de arredondamento que cada software utiliza. As únicas variáveis estatísticas que deram valores distintos foram, “Primeiro Quartil”, “Terceiro Quartil”, “Assimetria” e “Curtose”, isso ocorreu porque cada software estatístico tem um *default* (limiar) diferente para essas variáveis, tendo em mente o limiar que o MATLAB utiliza, sabe-se que para os quartis é utilizado uma equação para realização do cálculo para quando a matriz de dados é par ou ímpar, como pode-se observar na equação (9). Para Assimetria o limiar que o MATLAB utiliza varia de 0 à 1 e para Curtose esse limiar pode variar de 0 à 3.

Tabela 3: Comparação dos resultados obtidos dos softwares estatísticos, A e B para os dados do cálculo estatístico descrito de um grupo de dois indivíduos normais para o quantificador FM.

Variáveis Estatísticas	Softwares	Ritmos Cerebrais						
		Delta	Teta	Alfa	Beta	Gama	Supergama	Ruído
Tamanho da Amostra	Estatístico	20	20	20	20	20	20	20
	A	20	20	20	20	20	20	20
	B	20	20	20	20	20	20	20
Mínimo	Estatístico	1.5500	4.5195	9.6144	18.8051	43.6383	89.4461	59.9048
	A	1.5500	4.5194	9.6144	18.8051	43.6383	89.4461	59.9048
	B	1.5500	4.5194	9.6143	18.8050	43.6383	89.4461	59.9048
Máximo	Estatístico	2.4100	5.9593	10.2536	20.2891	52.3045	91.6039	60.0138
	A	2.4079	5.9593	10.2536	20.2891	52.3045	91.6039	60.0138
	B	2.4100	5.9592	10.2536	20.2891	52.3046	91.6038	60.0139
Amplitude Total	Estatístico	0.8600	1.4398	0.6392	1.4840	8.6662	2.1578	0.1090
	A	0.8600	1.4398	0.6392	1.4840	8.6662	2.1578	0.1090
	B	0.8599	1.4398	0.6392	1.4840	8.6662	2.1577	0.1090
Mediana	Estatístico	1.9635	5.0364	9.9103	19.4115	49.9594	89.8699	59.9757
	A	1.9635	5.0364	9.9103	19.4115	49.9594	89.8699	59.9757
	B	1.9634	5.0364	9.9103	19.4115	49.9593	89.8698	59.9757
Primeiro Quartil	Estatístico	1.6869	4.9051	9.7174	19.1908	47.4248	89.7507	59.9409
	A	1.7284	4.9003	9.7152	19.1876	47.4210	89.7490	59.9399
	B	1.6699	4.8997	9.6952	19.1496	47.3210	89.7470	59.9389
Terceiro Quartil	Estatístico	2.0996	5.2362	10.0924	19.6257	50.4923	90.3221	59.9807
	A	2.0933	5.2402	10.0942	19.6257	50.4925	90.2591	59.9808
	B	2.1032	5.2417	10.0946	19.6261	50.4928	90.1071	59.9808
Média	Estatístico	1.9296	5.0758	9.9105	19.4448	49.0071	90.0542	59.9636
	A	1.9279	5.0758	9.9105	19.4448	49.0071	90.0542	59.9636
	B	1.9295	5.0758	9.9105	19.4447	49.0070	90.0542	59.9635
Variância	Estatístico	0.0639	0.0936	0.0444	0.1483	5.1482	0.2423	0.0008
	A	0.0638	0.0936	0.0444	0.1483	5.1482	0.2423	0.0008
	B	0.0627	0.0936	0.0443	0.1483	5.1482	0.2423	0.0008
Desvio Padrão	Estatístico	0.2529	0.3060	0.2107	0.3851	2.2690	0.4923	0.0294
	A	0.2526	0.3060	0.2107	0.3851	2.2690	0.4923	0.0294
	B	0.2528	0.3059	0.2106	0.3851	2.2689	0.4922	0.0293

Variáveis Estatísticas	Softwares	Ritmos Cerebrais						
		Delta	Teta	Alfa	Beta	Gama	Supergama	Ruído
Desvio Padrão da Mediana	Estatístico	0.2552	0.3086	0.2107	0.3866	2.4704	0.5274	0.0319
	A	-	-	-	-	-	-	-
	B	0.2552	0.3086	0.2106	0.3866	2.4703	0.5273	0.0319
Coeficiente de Variação	Estatístico	13.10%	6.02%	2.12%	1.98%	4.62%	0.54%	0.04%
	A	13.10%	6.02%	2.12%	1.98%	4.62%	0.54%	0.04%
	B	13.10%	6.02%	2.12%	1.98%	4.62%	0.54%	0.04%
Assimetria	Estatístico	0.0563	0.8249	0.1837	0.4441	-0.7061	1.5553	-0.3683
	A	0.0433	0.8659	0.1855	0.4575	-0.7584	1.6812	-0.3562
	B	0.0609	0.8932	0.1989	0.4809	-0.7646	1.6822	-0.3988
Curtose	Estatístico	2.1264	4.8706	1.8482	2.7558	2.7345	5.8861	2.2441
	A	-0.6656	3.5545	1.4985	2.8745	2.5287	5.9102	2.1173
	B	-0.7665	2.8116	-1.1293	0.0540	0.0263	4.1357	-0.6131

Fonte: Própria, 2018.

Estatístico: Software desenvolvido no trabalho.

A e B: Softwares estatísticos já existentes.

2.4.3 Situação 3

Os resultados obtidos dos três softwares quando se confrontam os valores oriundos do processamento estatístico descritivo para o quantificador coerência de um grupo de três indivíduos do grupo coma e para o par de eletrodos Fp1-Fp2 estão listados na Tabela 4 fazendo referência a terceira situação da validação dos dados.

Observando os resultados obtidos da comparação entre os três softwares é possível notar que todos os resultados deram iguais e/ou próximos com diferença na quarta casa decimal após a vírgula, essa diferença se dá pelo método de arredondamento que cada software utiliza. As únicas variáveis estatísticas que deram valores distintos foram, “Primeiro Quartil”, “Terceiro Quartil”, “Assimetria” e “Curtose”, isso ocorreu porque cada software estatístico tem um *default* (limiar) diferente para essas variáveis, tendo em mente o limiar que o MATLAB utiliza, sabe-se que para os quartis é utilizado uma equação para realização do cálculo para quando a matriz de dados é par ou ímpar, como pode-se observar na equação (9). Para Assimetria o limiar que o MATLAB utiliza varia de 0 à 1 e para Curtose esse limiar pode variar de 0 à 3.

Tabela 4: Resultados obtidos da comparação do software estatístico, software A e B para o cálculo da estatística descritiva, para a situação de um grupo de três indivíduos com grau de consciência coma, par de eletrodos Fp1-Fp2 e quantificador coerência.

Variáveis Estatísticas	Softwares	Ritmos Cerebrais			
		Delta	Teta	Alfa	Beta
Tamanho da Amostra	Estatístico	90	150	180	690
	A	90	150	180	690
	B	90	150	180	690
Mínimo	Estatístico	0.0022	0.0011	0.0021	0.0001
	A	0.0022	0.0011	0.0021	0.0001
	B	0.0021	0.0011	0.0020	0.0001
Máximo	Estatístico	0.9282	0.8781	0.7697	0.8752
	A	0.9282	0.8781	0.7697	0.8752
	B	0.9282	0.8781	0.7697	0.8751
Amplitude Total	Estatístico	0.9261	0.8770	0.7677	0.8750
	A	0.9261	0.8770	0.7677	0.8750
	B	0.9260	0.8770	0.7676	0.8749
Mediana	Estatístico	0.2406	0.1905	0.2876	0.2583
	A	0.2406	0.1905	0.2876	0.2583
	B	0.2405	0.1905	0.2876	0.2583
Primeiro Quartil	Estatístico	0.1476	0.0933	0.1239	0.1062
	A	0.1507	0.0937	0.1257	0.1063
	B	0.1502	0.0941	0.1262	0.1065
Terceiro Quartil	Estatístico	0.3998	0.3992	0.5005	0.4457
	A	0.3991	0.3984	0.5000	0.4455
	B	0.3981	0.3975	0.4994	0.4450
Média	Estatístico	0.2832	0.2543	0.3160	0.2911
	A	0.2832	0.2543	0.3160	0.2911
	B	0.2832	0.2542	0.3160	0.2910
Variância	Estatístico	0.0421	0.0423	0.0458	0.0448
	A	0.0421	0.0423	0.0458	0.0448
	B	0.0421	0.0422	0.0457	0.0447
Desvio Padrão	Estatístico	0.2053	0.2056	0.2140	0.2116
	A	0.2053	0.2056	0.2140	0.2116
	B	0.2052	0.2055	0.2193	0.2116

Variáveis Estatísticas	Softwares	Ritmos Cerebrais			
		Delta	Teta	Alfa	Beta
Desvio Padrão da Mediana	Estatístico	0.2097	0.2153	0.2159	0.2141
	A	-	-	-	-
	B	0.2097	0.2153	0.2159	0.2141
Coeficiente de Variação	Estatístico	72.48%	80.84%	67.70%	72.69%
	A	72.48%	80.84%	67.70%	72.69%
	B	72.48%	80.84%	67.70%	72.69%
Assimetria	Estatístico	1.0771	0.7653	0.3871	0.5366
	A	1.0845	0.7750	0.3471	0.5667
	B	1.0954	0.7730	0.3904	0.5377
Curtose	Estatístico	4.1874	2.7058	1.9224	2.8222
	A	3.548	1.045	1.4581	2.4751
	B	1.3258	-0.2631	-1.0741	-0.7142

Fonte: Própria, 2018.

Estatístico: Software desenvolvido no trabalho.

A e B: Softwares estatísticos já existentes.

2.4.4 Situação 4

A quarta situação traz o resultado da comparação entre os softwares estatístico e C, para o quantificador FM, confrontando o resultado para o de teste de normalidade de Kolmogorov-Smirnov, a comparação dos dados ocorreu escolhendo para se analisar três eletrodos aleatórios e um ritmo cerebral para ser observado, sendo este também escolhido de forma aleatória, portanto para o quantificador FM foram escolhidos os eletrodos Fp1, C4 e O2. Sendo assim, a *Tabela 5* apresenta os resultados obtidos da comparação entre os softwares para uma situação individual com o grau de consciência coma. Confrontando os valores de *p-valor* obtidos tanto no software estatístico quando no A, é possível notar que não houve diferença significativa, para um intervalo de 95%, entre os valores dos dois softwares, sendo que NS significa “Não significativo” o que se pode concluir é que o valor não é significativo ou que ele é tão pequeno que se aproxima de 0 (zero), o que ocorre tanto no software estatístico quando no A. Portanto para essa situação, o software estatístico está apto a realizar testes de normalidade quando comparado à outro software.

Tabela 5: Resultados obtidos da comparação do software estatístico e o software A para o teste de Kolmogorov-Smirnov, para a situação de um indivíduo com grau de consciência coma e quantificador FM.

Eletrodos	Softwares	Delta	Teta	Alfa	Beta
Fp1	Estatístico	0	0	0	0
	A:	NS	NS	<0.01	NS
C4	Estatístico:	0	0	0	0
	A:	NS	NS	NS	NS
O2	Estatístico:	0	0	0	0
	A:	NS	NS	NS	NS

Fonte: Própria, 2018.

Estatístico: Software desenvolvido no trabalho.

A: Software estatístico já existente.

NS: Não Significativo.

2.4.5 Situação 5

Na situação 5, escolheu-se dois pares de eletrodos dentre os oito existentes, de forma aleatória, selecionando os pares 1 e 5, que fazem correspondência aos pares Fp1-Fp2 e C3-C4. Com isso, a análise comparativa entre os softwares estatístico e o A, em relação ao teste de Mann-Whitney está disposta na Tabela, para uma situação de comparação entre um grupo de registros de indivíduos em estado de coma. Confrontando os valores de *p-valor* obtidos tanto no software estatístico quando no A, é possível notar que não houve diferença significativa, para um intervalo de 95%, entre os valores dos dois softwares, sendo que NS significa “Não significativo” o que se pode concluir é que o valor não é significativo ou que ele é tão pequeno que se aproxima de 0 (zero), o que ocorre tanto no software estatístico quando no A. Portanto para essa situação o software estatístico está apto a realizar testes de comparação quando comparado à outro software.

Tabela 6: Resultados obtidos da comparação do software estatístico e o software A para o teste de Mann-Whitney, para a situação de comparação entre um grupo de registros de indivíduos (normal x coma) e quantificador coerência.

Eletrodos	Softwares	Delta	Teta	Alfa	Beta
Fp1-Fp2	Estatístico	0	0	0	0
	A:	0	0	0	0
C3-C4	Estatístico:	0	0	0	0.0185
	A:	0	0	0	<0.01

Fonte: Própria, 2018.

Estatístico: Software desenvolvido no trabalho.

A: Software estatístico já existente.

3 CONCLUSÕES

O desenvolvimento deste software se mostrou importante porque é voltado exclusivamente para processamento estatístico de registros de EEG. Com isso, quantificadores, estatística descritiva, testes de normalidade, testes de hipóteses, visualizações de resultados, foram escolhidos exclusivamente pensando neste tipo de análise e para este tipo de exame, sendo assim este software em questão se destaca dentro os outros softwares estatísticos por essa razão, já que tudo foi pensado, montado e desenvolvido para essa área específica da saúde, desde modo, se for do interesse do usuário realizar o processamento estatístico descritivo apenas de registros de EEG, este software aqui apresentado se mostra de grande valia e desempenho.

Posterior a realização de vários testes de processamentos de diversas formas, conjuntos e análises, o processamento dos dados se mostrou de maneira eficiente. Na parte de validação dos resultados, os resultados obtidos no software em questão se revelaram semelhantes aos softwares estatísticos utilizados para a comparação dos dados, com isso é possível concluir que o software desenvolvido aqui apresentado se mostra equivalente ou até mesmo superior a alguns softwares estatísticos devido a exclusividade que este software apresenta de ser específico para análise estatística de registros de EEG. Com isso o usuário não precisa necessariamente utilizar exclusivamente softwares estatísticos para realizar o processamento desses registros, podendo assim utilizar este software aqui demonstrado, tendo a garantia de resultados semelhantes à outros softwares, porém em interfaces mais específicas e fáceis de se mexer. Outro ponto importante, foi o tempo de processamento dos dados, este se mostrou curto e rápido quando comparado aos softwares existentes utilizados na validação dos dados e o software aqui apresentado utiliza pouca memória do computador em que este está instalado, mostrando assim outra vantagem em sua utilização quando comparado com as outras plataformas existentes.

Para trabalhos futuros, pretende-se desenvolver a parte de análise de multivariada, visando deixar mais completo a parte do software já desenvolvida até o momento e agregando mais uma parte da estatística em que visa investigar simultaneamente um emaranhado de múltiplas medidas em cada objeto ou indivíduo sob estudo (HAIR, 2009). Sendo assim, seria possível analisar de forma mais profunda um grupo de registros e poder levantar questões, sugestões e até possíveis conclusões através do(s) resultado(s) obtido(s) através da análise de multivariada.

Outra ideia para trabalhos futuros, seria deixar o software aqui mostrado, mais completo, onde neste seria feito toda a parte de tratamento do sinal, desde a escolha das épocas, separação das mesmas, processamento do sinal, análise estatística descritiva, análise de multivariada, até a visualização dos resultados obtidos. Com isso, o software seria totalmente independente de outros softwares, evitando assim, a utilização de diversos programas que em alguns pontos podem gerar confusão posteriormente.

4 REFERÊNCIAS

1. BANNURU, R. R. et al. Efficacy of curcumin and Boswellia for knee osteoarthritis: Systematic review and meta-analysis. **Seminars in Arthritis and Rheumatism**, 2018.
2. BRANCO, V. et al. Relação entre a tensão aplicada e a sensação de desconforto nos músculos isquiotibiais durante o alongamento. **Revista Brasileira de Fisioterapia**, 2006.
3. BRUNET, D.; MURRAY, M. M.; MICHEL, C. M. Spatiotemporal analysis of multichannel EEG: CARTOOL. **Computational Intelligence and Neuroscience**, v. 2011, 2011.
4. BUZSÁKI, G. **Rhythms of the Brain**. [s.l: s.n.].
5. CACIOPPO, S. et al. Dynamic spatiotemporal brain analyses using high performance electrical neuroimaging: Theoretical framework and validation. **Journal of Neuroscience Methods**, v. 238, p. 11–34, 2014.
6. CIRILLO, M. A.; FERREIRA, D. F. Extensão do Teste para Normalidade Univariado Baseado no Coeficiente de Correlação Quantil-Quantil para o Caso Multivariado. **Revista de Matemática e Estatística**, v. 21, n. 3, p. 67–84, 2003.
7. DELORME, A. et al. EEGLAB, SIFT, NFT, BCILAB, and ERICA: New tools for advanced EEG processing. **Computational Intelligence and Neuroscience**, v. 2011, 2011.
8. DELORME, A.; SEJNOWSKI, T.; MAKEIG, S. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. **NeuroImage**, v. 34, n. 4, p. 1443–1449, 2007.
9. DESLANDES, A. C. et al. Effect of aerobic training on EEG alpha asymmetry and depressive symptoms in the elderly: A 1-year follow-up study. **Brazilian Journal of Medical and Biological Research**, 2010.
10. DONG, J.-Y.; ZHANG, Y.-H.; QIN, L.-Q. Obstructive sleep apnea and cardiovascular risk: Meta-analysis of prospective cohort studies. **Atherosclerosis**, 2013.
11. FERREIRA, D. F. Sisvar. In: **Daniel's Website**. Lavras: DES – UFLA, 2015. Disponível em: <<http://www.dex.ufla.br/~danielff/programas/sisvar.html>>. Acessado em: 18 mar. 2018.
12. FONTELLES, M. J. **BIOESTATÍSTICA Aplicada à Pesquisa Experimental: volume 1**. [s.l: s.n.].
13. GODOY, M. F. DE et al. Mortalidade por doenças cardiovasculares e níveis

- socioeconômicos na população de São José do Rio Preto, estado de São Paulo, Brasil. **Arquivos Brasileiros de Cardiologia**, 2007.
14. GUTTIKONDA, S. N. R.; VADAPALLI, K. Approach to undifferentiated dyspnea in emergency department: aids in rapid clinical decision-making. **International Journal of Emergency Medicine**, 2018.
 15. KANG, X. et al. Clinical and radiographic outcomes of upper thoracic versus lower thoracic upper instrumented vertebrae for adult scoliosis: A meta-analysis. **Brazilian Journal of Medical and Biological Research**, 2018.
 16. LEOTTI, V. B.; BIRCK, A. R.; RIBOLDI, J. Comparação dos testes de aderência à normalidade Kolmogorov-Smirnov, Anderson-Darling, Cramer-Von Mises e Shapiro-Wilk por simulação. **11.º Simpósio de Estatística Aplicada à Experimentação Agrônômica e a 50ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria (RBRAS)**, 2005.
 17. LI, F.; COOMBS, D. Mental health history—a contributing factor for poorer outcomes in burn survivors. **Burns & Trauma**, 2018.
 18. MEMÓRIA, J. M. P. Breve história da Estatística. **Embrapa Informação Tecnológica**, p. 111, 2004.
 19. MONTGOMERY, D. C. **Applied Statistics and Probability for Engineers Third Edition**. [s.l: s.n.].
 20. PAES, Â. T. Itens essenciais em bioestatística. **Arquivos Brasileiros de Cardiologia**, 1998.
 21. PEYK, P.; DE CESAREI, A.; JUNGHÖFER, M. Electromagnetic encephalography software: Overview and integration with other EEG/MEG toolboxes. **Computational Intelligence and Neuroscience**, v. 2011, 2011.
 22. POSSOBON, R. D. F. et al. O comportamento de crianças durante atendimento odontológico. **Psicologia: Teoria e Pesquisa**, 2003.
 23. RAMOS, C. D. Caracterização do eletroencefalograma normal em situação de vigília: elaboração da base de dados e análise quantitativa. p. 156, 2017.
 24. REIS, E. et al. **Estatística Aplicada 1**. [s.l: s.n.].
 25. SANEI, S.; CHAMBERS, J. A. **EEG Signal Processing**. [s.l: s.n.].
 26. SCHMILDT, E. R. Bioestatística: estatística descritiva. **Vide Science Technique Et Applications**, 2007.
 27. SÖRNMO, L.; LAGUNA, P. **Bioelectrical Signal Processing in Cardiac and**

- Neurological Applications.** [s.l: s.n.].
28. SUGRUE, A. et al. Architectural T-Wave Analysis and Identification of On-Therapy Breakthrough Arrhythmic Risk in Type 1 and Type 2 Long-QT Syndrome. **Circulation. Arrhythmia and electrophysiology**, 2017.
 29. TADEL, F. et al. Brainstorm: A user-friendly application for MEG/EEG analysis. **Computational Intelligence and Neuroscience**, v. 2011, 2011.
 30. TISON, G. H. et al. Passive Detection of Atrial Fibrillation Using a Commercially Available Smartwatch. **JAMA Cardiology**, 2018.
 31. URL 1: Bioestat - versão 5.3 “Mamiraua”. Disponível em: <<https://www.mamiraua.org.br/pt-br/downloads/programas/bioestat-versao-53/>>. Acessado em: 18 mar. 2018.
 32. URL 2: The R Project for Statistical Computing. “The R-Project”. Disponível em: <<https://www.r-project.org/>>. Acessado em: 18 mar. 2018.
 33. URL 3: Products Office Excel. “Microsoft”. Disponível em <<https://products.office.com/pt-br/excel>>. Acessado em: 18 mar. 2018.
 34. URL 4: Support Office. “Microsoft”. Disponível em <<https://support.office.com/pt-br/article/operadores-de-c%C3%A1lculo-e-preced%C3%Aancia-no-excel-48be406d-4975-4d31-b2b8-7af9e0e2878a>>. Acessado em: 18 mar. 2018.
 35. URL 5: What Is Wolfram|Alpha?. “Wolfram Alpha”. Disponível em: <<https://www.wolframalpha.com/tour/>>. Acessado em: 18 mar. 2018.
 36. URL 6: MathWorks Products. “MathWorks”. Disponível em: <<https://www.mathworks.com/products/matlab.html>>. Acessado em: 18 março 2018.
 37. WALDMAN, Z. J. et al. A method for the topographical identification and quantification of high frequency oscillations in intracranial electroencephalography recordings. **Clinical Neurophysiology**, v. 129, n. 1, p. 308–318, 2018.
 38. ZANOBETTI, A. Generalized additive distributed lag models: quantifying mortality displacement. **Biostatistics**, 2000.
 39. ZAROS, L. G.; MEDEIROS, H. R. DE. Bioestatística. In: [s.l: s.n.]. p. 214.