Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Wanessa Luana de Brito Costa

Aplicação dos testes não paramétricos em dados do ENEM 2011

Campina Grande
Dezembro de 2012

Wanessa Luana de Brito Costa

Aplicação dos testes não paramétricos em dados do ENEM 2011

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientadora:
Divanilda Maia Esteves

Co-orientadora:
Ana Patricia Bastos Peixoto

Campina Grande
Dezembro de 2012

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

C837a Costa, Wanessa Luana de Brito.

Aplicação dos testes não paramétricos em dados do Enem 2011 [manuscrito] / Wanessa Luana de Brito Costa. – 2012.

47f..: il. color.

Trabalho de Conclusão de Curso (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2012.

"Orientação: Prof. Dr. Divanilda Maia Esteves, Departamento de Estatística".

1. Estatística - Testes Não-Paramétricos. 2. Enem - Provas. 3. Comparação de Médias. I. Título.

21. ed. CDD 519.5

Wanessa Luana de Brito Costa

Aplicação dos testes não paramétricos em dados do ENEM 2011

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 14 / 12 / 2012

Banca Examinadora:

Ana Patrícia Bastos Peixoto

Prof. Msc. Ana Patrícia Bastos Peixoto
Universidade Estadual da Paraíba UEPB -
Departamento de Estatística/CCT
Co-Orientadora

Tiago Almeida de Oliveira

Prof. Dr. Tiago Almeida de Oliveira
Universidade Estadual da Paraíba UEPB -
Departamento de Estatística/CCT
Examinador

Prof. Dr. Gustavo H. Esteves
Universidade Estadual da Paraíba UEPB -
Departamento de Estatística/CCT
Examinador

Dedicatória

A Veridiana de Brito Costa e João Bosco Alves Costa, meus pais, a razão da minha existência.

Agradecimentos

Primeiramente a Deus, que me proporcionou a realização deste sonho, capacitou-me para que eu chegasse até aqui, pois sem ele eu não conseguiria, toda honra e toda glória sejam dadas a ti Senhor.

A minha mãe Veridiana de Brito que acreditou no meu potencial, me ensinou a lutar pelos meus objetivos com garra, humildade e determinação, sempre esteve ao meu lado torcendo pela minha realização, incentivando-me nos momentos mais difíceis para que eu chegasse até aqui. A meu pai João Bosco, aos meus irmãos Bruna Vaneska e Jordy Luan, aos meus avós Manoel e Maria das Dores, Maria e Francisco, as minhas tias Rosilene, Núbia, Margarida e Irenir e a todos os meus familiares, por todo apoio concedido.

A minha orientadora Diana Maia e, principalmente, a minha co-orientadora Ana Patrícia Bastos, que esteve sempre disposta a me ajudar, sempre depositando confiança e incentivo.

A UEPB, por me conceder oportunidade de formação. A todos os meus professores, especialmente aos professores, Ana Patricia, Tiago de Almeida, Ricardo Alves, Ana Cristina, Juarez, Diana Maia e Edwirde Luís, pela dedicação, compreensão, confiança, ética e profissionalismo. E por todo conhecimento adquirido para a minnha formação.

As minhas amigas e companheiras de estudo, Adriana Souza, Samara Rilda, Arielly Arethuza, Valneli Silva, Nathielly Lima e Analu Cabral, por tudo que passamos. Juntas não dividimos apenas conhecimentos, mas também alegrias, companheirismo, dificuldades e forças, esses momentos serão sempre lembrados.

A Nyedja Fialho, Salete Morais e Fátima Correia, por toda confiança depositada em mim, pelo incentivo e compreensão.

A Superintendência de Trânsito e Transportes Públicos (STTP), pela oportunidade de estágio.

Enfim, a todos que me ajudaram a tornar este sonho realidade, o meu muito obrigada!

Resumo

Este trabalho teve por objetivo verificar se existem diferenças significativas entre médias das provas do ENEM 2011, entre alunos da rede pública e privada do estado da Paraíba. Além disso, comparou-se a existência de diferença nas médias de alunos do estado da Paraíba com os demais estados do Nordeste. Para o mesmo utilizou-se dados do ENEM 2011 cuja taxa de participação das escolas foi igual a 50%, obtidos através do site do INEP. As variáveis analisadas foras as médias das provas de Ciências Humanas (CH), Ciências da Natureza (CN), Matemática (M), Linguagens de Códigos (LC), Objetivas (O) e Redação (R). Para analisar estas diferenças utilizou-se testes estatísticos não-paramétricos, testes cujo modelo não especificam condições sobre os parâmetros da população da qual extraiu-se a amostra. A amostra em estudo apresentou diferenças significativas entre as médias das provas do ENEM 2011 no ensino das redes pública e privada em todo Nordeste. Observou-se ainda que o estado da Paraíba está no mesmo caminho que os demais estados da região, em todos os estados o ensino da rede pública foi desfavorecido.

Palavras-chave: Testes Não-Paramétricos, ENEM, Comparação de Médias

Abstract

This study aimed to verify whether there are significant differences between means of evidence ENEM 2011, between students in public and private state of Paraíba. Furthermore, we compared the existence of difference in average student in the state of Paraíba with other Northeastern states. For the same we used data from 2011 ENEM whose participation rate of schools was equal to 50 %, obtained through the site INEP. The analyzed variables forces the means of proof of Humanities (CH), Natural Sciences (CN), Mathematics (M), Languages Codes (LC), targeted (O) and Writing (R). To analyze these differences, we used nonparametric statistical tests, tests the model does not specify conditions on the parameters of the population from which the sample was extracted. The study sample showed significant differences between the means of evidence ENEM 2011 in education from public and private throughout the Northeast. It was also observed that the state of Paraíba is the same path as the other states of the region, in every state of public education was disadvantaged.

Key-words: Non-parametric Tests, ENEM, Mean-Comparison

Sumário

Lista de Figuras

Lista de Tabelas

1 Introdução

O Exame Nacional do Ensino Médio (ENEN) foi criado pelo Ministério da Educação para avaliar o desempenho dos alunos que estão em fase de conclusão ou concluíram o ensino médio, sendo uma alternativa para o ingresso nas instituições de ensino superior. O ingresso nas instituições públicas usando, total ou parcialmente a nota do ENEM, pode facilitar a vinda de estudantes de outros estados para a Paraíba, já que o Estado possui três grandes universidades públicas que passaram a adotar a nota do ENEM em seus processos seletivos.

No desenvolvimento dos métodos da estatística moderna, as primeiras técnicas da inferência que apareceram foram aquelas que faziam diversas hipóteses sobre a natureza da população da qual os dados foram extraídos. Como os valores relacionados com a população são parâmetros, tais técnicas estatísticas são chamadas paramétricas. Os testes de hipóteses paramétricos incidem explicitamente sobre um parâmetro de uma ou mais populações, pressupõe-se que as amostras foram retiradas de uma população que se distribui segundo uma função de probabilidade normal, além de pressupor também que as variâncias sejam homogenêas.

Assim, se algum desses pressupostos são violados, então os tradicionais testes paramétricos não tem rigor estatístico e deverão ser evitados e substituídos por testes que não exigem o cumprimento de tais pressupostos. Estes são ditos testes não paramétricos ou testes de distribuição livre. Essas técnicas tem como resultado conclusões que exigem menos qualificações, a lógica desses testes é simples e mais fácil do que seus equivalentes paramétricos, porém estes são menos poderosos, ou seja, têm menor probabilidade, do que os paramétricos, de rejeitar a hipótese de nulidade quando ela é falsa.

Este trabalho tem por objetivo verificar se existe diferenças entre médias das provas do ENEM 2011, entre alunos da rede pública e privada do estado da Paraíba, com taxa de participação das escolas igual a 50%, como também comparar se existe diferença nas médias de alunos do estado da Paraíba com a região Nordeste no geral.

2 Fundamentação Teórica

O conteúdo desta seção relata os principais aspectos da utilização dos testes não paramétricos, por meio de artigos práticos e teóricos relacionados ao objetivo da pesquisa.

2.1 Marco Histórico

Segundo Oliveira et al. (2009) a estatística é um conjunto de métodos de obtenção e utilização de informações, para auxiliar a tomada de decisões em uma situação prática envolvendo incerteza. Os Estados, desde tempos remotos, precisaram conhecer determinadas características da população, efetuar a sua contagem e saber a sua composição ou os seus rendimentos. Realizaram-se as primeiras estatísticas, nomeadamente para determinarem leis sobre impostos e números de homens disponíveis para combater.

Até ao início do século XVII, a Estatística limitou-se ao estudo dos assuntos de Estado. Usada pelas autoridades políticas na inventariação ou arrolamento dos recursos disponíveis, limitava-se a uma simples técnica de contagem, traduzindo numericamente fatos ou fenômenos observados (Estatística Descritiva). No século XVII, com os aritméticos políticos, nomeadamente John Graunt (1620-1674) e Sir William Petty (1623-1687), inicia-se na Inglaterra uma nova fase de desenvolvimento da Estatística, virada para a análise dos fenômenos observados fase da Estatística Analítica.

O desenvolvimento do cálculo das probabilidades surge também no século XVII. A ligação das probabilidades com os conhecimentos estatísticos veio dar uma nova dimensão à Estatística, que progressivamente foi se tornando um instrumento científico poderoso e indispensável. Considera-se assim uma nova fase, a terceira, em que se começa a fazer inferência estatística: quando a partir de observações se procurou deduzir relações causais, entre variáveis, realizando-se previsões a partir daquelas relações.

A designada "estatística não-paramétrica" baseia-se num conjunto de processos de inferência, que são válidos para um grupo mais vasto e diversificado de distribuições, que não a Normal. O termo inferência não-paramétrica, deriva do fato de não ser necessário desenvolver um modelo populacional em termos de uma função densidade de probabilidade, dependente dos parâmetros, como é o caso da distribuição normal. O primeiro livro-texto voltado aos métodos não-paramétricos foi escrito por Siegel (1975) e é usado até hoje pela facilidade de leitura.

2.2 Exame Nacional do Ensino Médio

O exame nacional do ensino médio (ENEM) foi criado em 1998, durante a gestão do ministro da educação Paulo Renato Souza, no governo Fernando Henrique Cardoso, com o objetivo de avaliar o desempenho dos estudantes do ensino médio em todo país. Na sua primeira edição em 1998, o ENEM contou com um número de 157,2 mil inscritos e de 115,6 mil participantes. Na quarta edição, em 2001, já alcançava a marca expressiva de 1,6 milhão de inscritos e de 1,2 milhão de participantes. Inicialmente, a prova era composta por 63 questões havendo apenas um dia de prova, em 2009, a prova passou a ter 180 questões, uma redação e agora sendo dois dias de prova. Porém, a mudança não foi apenas nas provas. O ENEM passou a ter outras funções além de observar o grau de conhecimento dos alunos. Servia também como certificação de conclusão do ensino médio em cursos de educação de jovens e adultos (EJA), antigo supletivo subistituindo o exame nacional para certificação de competência de jovens e adultos (Encceja).

Outra mudança que ocorreu, foi que a prova passou a ser uma via de ingresso no ensino superior. Com a criação do SISU (Sistema de Seleção Unificada), algumas instituições públicas de ensino superior passaram a disponibilizar vagas a serem preenchidas diretamente através da nota do ENEM, sem ser necessário se submeter ao vestibular. Isso propicia aos estudantes a oportunidade de buscarem o ingresso em cursos de interesse em várias universidades, inclusive aquelas mais distantes de sua instituição de origem, sem que seja necessário ir até lá para prestar vestibular. Além disso a nota do ENEM tem sido utilizada na primeira fase do vestibular como parte da nota final e para preencher vagas remanescentes nas instituições. A partir de 2004, nas instituições privadas, a nota do ENEM pode ser usada como parâmetro para o aluno que deseja obter bolsa de estudo parcial ou integral, através do programa Universidade para Todos (ProUni) do Ministério da Educação. Essas mudanças além da isenção do pagamento da taxa de inscrição para os alunos da escola pública foram fatores que certamente contribuíram para a popularização do exame.

O que se vê, é um número crescente de instituições cadastradas no INEP para utilizar os resultados do ENEM em seus processos seletivos. Muitas instituições de ensino superior já usam a nota do ENEM como fator influente nas notas do vestibular e outras tantas já aboliram o vestibular, sendo a seleção feita unicamente pela nota do exame.

A teoria de resposta ao item (TRI) tem sido bastante divulgada no meio acadêmico. Embora este tema não seja o abordado neste trabalho, vale ressaltar que a pequena sigla TRI tornou-se popular nas falas, artigos e reportagens sobre o ENEM. A TRI é a metodologia que dá suporte a elaboração e correção do novo modelo de provas adotado pelo ENEM, envolve aspectos como psicologia, estatística e informática. Há três parâmetros importantes considerados em cada item: o grau de dificuldade, a discriminação do item e o acerto casual, afirma Dalton Francisco de Andrade professor titular do departamento de informática e estatística da UFSC (BORGES, 2010).

O modelo logístico da TRI calcula a probabilidade de um candidato acertar aquele item a partir do conhecimento que possui (dificuldade) e o conhecimento mínimo necessário para responder a questão (discriminação), e avalia o padrão de respostas do aluno na prova, para ter certeza de que ele não está acertando ao acaso. O INEP já estuda como adaptar o ENEM para testes computacionais. Assim a avaliação se adaptaria ao nível de conhecimento dos estudantes. A cada acerto o estudante receberia o próximo quesito com um nível de dificuldade um pouco maior, em caso de erro um mais fácil. Para isso o sistema precisa de um banco de dados variado e numeroso. Tal estudo pode ser levado em consideração futuramente, já que a teoria vem sendo repercutida em todo país (BORGES, 2010).

2.3 Teste de hipóteses

A inferência estatística aborda dois tipos de problemas fundamentais: a estimação de parâmetros de uma população e os testes de hipóteses. Os testes de hipóteses são utilizados em estatística quando um pesquisador deseja verificar se algum parâmetro da distribuição de uma variável de interesse é condizente com o valor por ele estipulado. Para isso, algum estimador obtido a partir dos dados amostrais é usado para avaliar essa suposição. Uma hipótese é uma pressuposição a respeito de um determinado problema. Uma vez estabelecida uma hipótese, ela estará sujeita a uma comprovação. Para tanto é necessário a utilização de testes estatísticos que são regras de decisões, vinculadas a um fenômeno da população que nos possibilitam avaliar, por meio de uma amostra, se determinadas hipóteses estabelecidas por um pesquisador possam ser rejeitadas ou não.

No campo da inferência estatística a busca por respostas acerca de certas características de uma população estudada é de fundamental importância, apenas com base nessas características é que se devem estabelecer regras e tomar decisões sobre qualquer hipótese formulada no que se refere a população. Dessa forma, escolhida uma variável X e colhida uma amostra aleatória da população, pode-se estar interessado em inferir a respeito de alguns de seus parâmetros, como por exemplo média, variância e proporção, e também sobre o comportamento da variável (a sua distribuição de probabilidade). A realização de testes de hipóteses nos fornecem meios para que se possa, com determinado grau de certeza, concluir se os valores dos parâmetros ou mesmo as distribuições associadas à população considerada, podem representá-la de forma satisfatória.

De acordo com Siegel (1975), para se chegar a uma decisão objetiva sobre uma determinada hipótese, devemos dispor de um processo objetivo que nos permita rejeitar ou aceitar determinada hipótese. Esse procedimento objetivo deve basear-se na informação que obtemos em nossa pesquisa e no risco que queiramos correr de que nossa decisão em relação à hipótese não seja correta. O procedimento usualmente seguido possui os seguintes estágios:

i) Definir a hipótese de nulidade (H_0).

ii) Escolher uma função da amostra (estatística) para testar (H_0). Tal estatística, chamada **Estatística de teste**, deve representar bem a quantidade de interesse e ter um modelo probabilístico completamente especificado.

iii) Especificar um nível de significância α e um tamanho de amostra (n).

iv) Determinar (ou supor determinada) a distribuição amostral da prova estatística sob a hipótese de nulidade (H_0).

v) Decisão, onde rejeitamos a hipótese H_0 se o valor-p é menor que o nível de significância α caso contrário não rejeitamos H_0.

O primeiro passo a ser realizado na tomada de decisão é definir a hipótese de nulidade (H_0). A hipótese de nulidade é uma afirmativa de que o valor de um determinado parâmetro populacional é igual a algum valor especificado.Se rejeitada, pode-se aceitar a hipótese alternativa (H_1). A hipótese alternativa é a afirmativa de que um parâmetro tem um valor que, de alguma maneira difere da hipótese nula, é a definição operacional da hipótese de pesquisa do pesquisador, isto é, é a predição deduzida da teoria que esta sendo comprovada.

De acordo com Bolfarine e Sandoval (2001), quando deseja-se tomar uma decisão sobre diferenças, testa-se H_0 contra H_1. Ao formularmos uma decisão sobre H_0, podem ocorrer dois erros distintos, o primeiro designado por "Erro do tipo I" e o segundo é o "Erro do tipo II". Na Tabela 1, pode-se observar uma representação das decisões possíveis e de quais situações representam cada tipo de erro.

Tabela 1: Tipos de erros que podem ser cometidos em um teste de hipóteses

Decisão	H_0 verdadeira	H_0 falso
Aceitar H_0	Decisão correta	Erro tipo II
Rejeitar H_0	Erro tipo I	Decisão correta

Formuladas as hipóteses H_0 e H_1, escolhida uma estatística de teste adequada, o próximo passo consiste em especificar um nível de significância, representado por α, que é a probabilidade de a estatística de teste cair na região crítica quando a hipótese H_0 realmente for verdadeira (erro tipo I). A estes erros são associados valores de probabilidade:

$$P(\text{rejeitar } H_0|H_0 \text{ verdadeira}) = \alpha,$$

e

$$P(\text{aceitar } H_0|H_0 \text{ falsa}) = \beta.$$

Qualquer que seja a escolha da estatística de teste a ser aplicada a seus dados, o pesquisador deve, logo a seguir, determinar qual a distribuição amostral de tal estatística. De modo geral o conhecimento da distribuição amostral de determinada estatística permite fazer afirmações de caráter probabilístico sobre a ocorrência de certos valores numéricos da mesma (SIEGEL, 1975). Em outras palavras, para que seja possível calcular as probabilidades de erros, é fundamental que a estatística de teste seja completamente especificada, ou seja, que se saiba a qual família de distribuições a mesma está associada e que não dependa de parâmetros desconhecidos. Disto depende a aplicabilidade do teste.

2.3.1 Região de rejeição ou região crítica

A região de rejeição é uma região da distribuição amostral. A distribuição amostral inclui todos os valores possíveis que um teste estatístico pode assumir sob H_0, a região de rejeição consiste de um subconjunto desses valores possíveis, e é definida de modo que seja igual a probabilidade α, sob H_0, da ocorrência de um valor da estatística de teste naquele subconjunto. A localização da região de rejeição é afetada pela natureza de H_1. Se H_1 indica o sentido da diferença, usa-se um teste unilateral. Se H_1 não indica o sentido da diferença, usa-se um teste bilateral. Em outras palavras, em um teste unilateral, a região de rejeição está inteiramente localizada em uma das caudas da distribuição amostral. Em um teste bilateral, a região de rejeição está localizada em ambas as extremidades (ou caudas) da distribuição. A Figura 1 mostra um exemplo de representação da região crítica de um teste junto da curva da distribuição da estatística de teste.

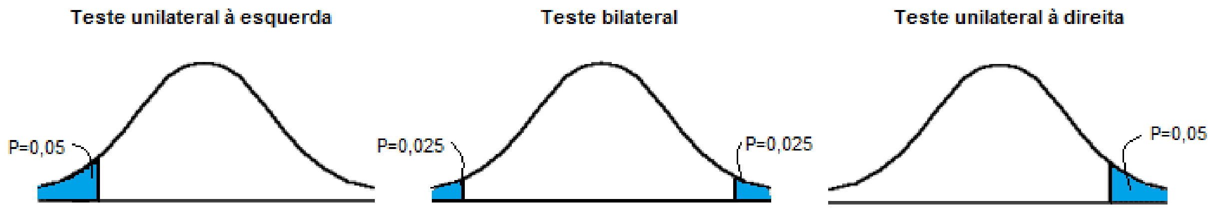

Figura 1: Exemplo de representação da região crítica de testes uni e bilaterais, conjuntamente com a curva da distribuição da estatística de teste.

Se o valor da estatística de teste calculado na amostra observada possui um valor pertencente à região de rejeição, rejeitamos H_0. Além desse criterio, pode-se utilizar o valor-p. Quando a probabilidade associada a um valor observado da estatística é igual ou menor de que o valor previamente determinado de α, concluímos que H_0 é falsa. Tal valor observado é chamado "significativo". A hipótese de nulidade a ser comprovada é rejeitada sempre que ocorre um valor "significativo". Valor significativo é um valor cuja probabilidade de ocorrência, sob H_0, é igual a ou menor do que α (SIEGEL, 1975).

Os testes de hipóteses se dividem em paramétricos e não- paramétricos. Os paramétricos são aqueles que utilizam os parâmetros da distribuição, ou uma estimativa destes, para o cálculo da sua estatística. Normalmente estes testes são mais rigorosos e possuem mais pressuposições para sua validação. Já os não-paramétricos utilizam, para o cálculo de sua estatística, postos (nem todos) atribuídos aos dados ordenados e são livres da distribuição de probabilidades dos dados estudados.

2.3.2 Testes paramétricos

Em termos gerais, uma hipótese é uma conjectura sobre algum fenômeno ou conjunto de fatos. Em estatística inferencial o termo hipótese tem um significado bastante específico. É uma conjectura sobre um ou mais parâmetros populacionais. O teste de hipóteses paramétrico envolve fazer inferências sobre a natureza da população com base nas observações de uma amostra extraída desta população. O processo segue um esquema como o representado na Figura 2.

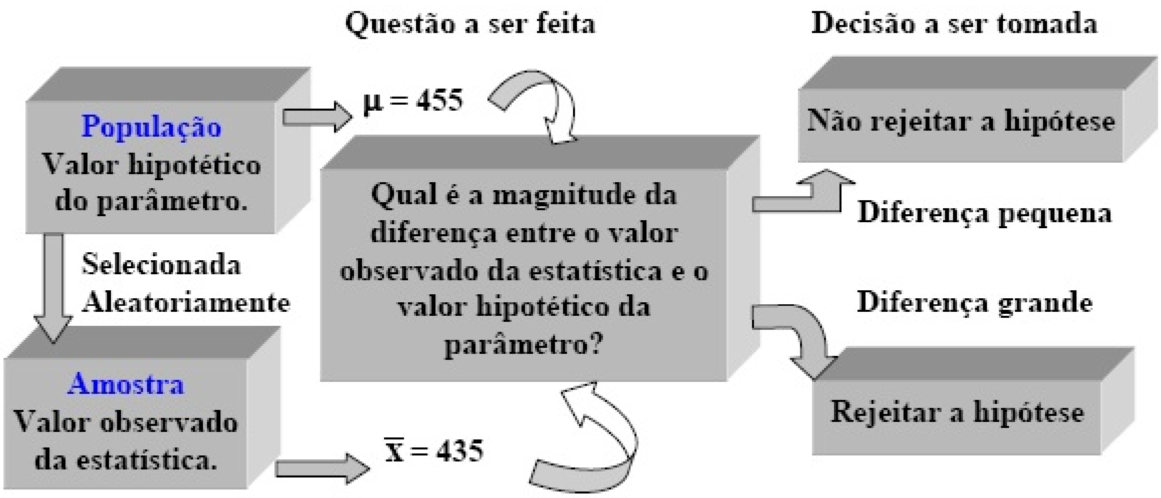

Figura 2: Esquema representativo da lógica dos testes de hipóteses

Um teste estatístico paramétrico é aquele cujo modelo especifica certas condições sobre os parâmetros da população da qual se extraiu a amostra em estudo. Abaixo estão listadas algumas das condições mais frequentemente exigidas:

i) As amostras consideradas devem ser independentes, ou seja, a escolha de determinado elemento para inclusão na amostra não deve influenciar na escolha de outros elementos.

ii) As observações devem ser extraídas de populações com distribuição normal.

iii) As variâncias são homogêneas.

iv) As variáveis em análise devem ser medidas pelo menos em escala intervalar.

v) No caso de Análise de Variância (teste F), exige-se ainda que as médias dessas populações normais e homocedásticas devem ser combinações lineares de efeitos devidos a colunas e/ou linhas. Ou seja os efeitos devem ser aditivos (SIEGEL, 1975).

Quando tem-se a certeza de que tais pressupostos são satisfeitos pelos dados em análise, optar por um teste paramétrico para analisar a amostra em estudo é uma boa escolha. Isto porque o teste paramétrico é o de maior poder para a rejeição da hipótese H_0 quando de fato H_0 deve ser rejeitada. Se algumas dessas condições não forem atendidas, então os testes paramétricos tem sua confiabilidade comprometida. Portanto, se existirem dúvidas a respeito da validade das suposições consideradas para o teste, os testes paramétricos devem ser evitados e substituídos por testes que exijam hipóteses menos rigorosas sobre os parâmetros. Esses são os testes não paramétricos, os quais também são chamados de "distribuição livre".

2.4 Métodos estatísticos não paramétricos

Segundo Siegel (1975), um teste não paramétrico é aquele cujo modelo não especifica condições sobre os parâmetros da população da qual extraiu-se a amostra. Há certas suposições básicas associadas à maioria dos testes não-paramétricos. Em geral, supõe-se que as observações sejam independentes, alguns exigem simetria e que a variável em estudo tenha continuidade básica, suposições essas em menor número e mais fracas do que aquelas associadas aos testes paramétricos.

Os testes indicados quando a variável qualitativa é medida em escala ordinal são não paramétricos. Também os dados quantitativos que não apresentam distribuição normal devem ser analisados por meio dos mesmos testes utilizados para as variáveis qualitativas ordinais. Quando os dados são qualitativos ou quantitativos sem distribuição normal, os testes não paramétricos são mais indicados.

Para dados qualitativos ordinais ou quantitativos sem distribuição normal, a estatística não paramétrica é baseada na conversão dos dados originais da amostra em postos *(ranks)* ou sinais positivos ou negativos. Um posto é considerado como sendo a posição de um determinado dado quando todos os valores que compõem a amostra estão ordenados de forma crescente. Nos testes de hipóteses as "estatísticas" dos testes não paramétricos não dependem diretamente dos valores das observações mas sim de outras características, como por exemplo a relação de ordem e a graduação das observações.

O princípio da comparação entre grupos utilizando postos se baseia na ideia de que se as amostras a serem comparadas são iguais, a ordenação de todos os dados tende a misturá-los de maneira uniforme. Porém, ao contrário, se na medida em que os valores ordenados das amostras separam os dados de cada grupo, a probabilidade de igualdade entre eles se torna mais remota. Construir um teste usando postos é uma forma de medir a posição relativa da observação sem usar o valor observado diretamente. Os postos relacionados às observações de uma variável X_1, X_2,...,X_n são obtidos através:

i) As observações devem ser ordenadas em ordem crescente, $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$.

ii) Associam-se valores, correspondendo às suas posições relativas na amostra. A primeira observação recebe o valor 1, a segunda o valor 2 e assim por diante até a n-ésima observação.

iii) No caso de valores repetidos, será considerado empate, deverá ser feito uma média dos postos que seria ocupado por essas observações.

Os testes não paramétricos não são tão eficientes quanto os paramétricos para encontrar diferenças na população, pois eles excluem algumas das informações disponíveis ao converter números em postos ou diferenças entre números em sinais positivos e negativos simplesmente. Assim sendo, os testes não paramétricos apresentam algumas vantagens e desvantagens.

i) Vantagens dos testes não paramétricos

1. São menos exigentes do que os testes paramétricos. Dispensam, por exemplo a normalidade entre os dados.
2. Em geral, as probabilidades das afirmativas obtidas da maioria dos testes não-paramétricos, são probabilidades exatas, salvo quando se usam aproximações para grandes amostras.
3. Independem da forma da população da qual a amostra foi obtida.
4. São, em geral de mais fácil aplicação e exigem, quase sempre, menor volume de cálculos.
5. Existem testes não paramétricos que nos permitem trabalhar com dados de diferentes populações, o que não é possível com os testes paramétricos.
6. São úteis nos casos em que é difícil estabelecer uma escala de valores quantitativos para os dados. O pesquisador pode apenas dizer que um dado tem mais ou menos da característica que está sendo analisada, sem poder precisar ou quantificar as diferenças. Os dados se encontram numa certa ordem de classificação: mais ou menos; melhor ou pior; maior ou menor, enfim.
7. São mais eficientes do que os paramétricos, quando os dados da população não têm distribuição normal. E quando a população é normalmente distribuída,

sua eficiência, em alguns casos, é levemente inferior à dos seus competidores (CAMPOS, 1983).

ii) Desvantagens dos testes não paramétricos

1. Em geral não levam em consideração a magnitude dos dados. É muito comum transformar os dados, de valores para simples ordens ou sinais. Em muitos casos isso se traduz num desperdício de informações.
2. Quando todas as exigências do modelo estatístico são satisfeitas, o teste paramétrico é mais poderoso. Para se obter a mesma eficiência, se faz necessário um maior tamanho de amostra para o não-paramétrico.
3. Em geral, não nos permite testar interações, salvo sob condições especiais sobre aditividade. Isso restringe o seu uso em modelos mais complicados.
4. A obtenção, utilização e interpretação das tabelas são, em geral, mais complexas (CAMPOS, 1983).

2.4.1 Testes não paramétricos

É importante a definição de critérios que nos ajudem a decidir qual o teste ideal para determinado problema. Um desses critérios, sem dúvida é o Poder do Teste. Diz-se que um teste é o mais poderoso de tamanho α, quando, entre todos os testes que tem nível de significância menor ou igual a α, ele é o que apresenta uma maior probabilidade de rejeitar H_0 quando H_0 é falsa. Em outras palavras, limita-se a probabilidade de erro tipo I e procura-se por aquele teste que tenha a menor probabilidade de erro tipo II. Na Inferência paramétrica, há uma série de resultados que nos indicam como encontrar o teste mais poderoso, como por exemplo, o Lema de Neyman-Pearson. No entanto, tais resultados são válidos sob certas condições que nem sempre são observadas. Pra começar, é necessário que tenhamos uma amostra obtida a partir de uma distribuição de probabilidade conhecida.

2.4.1.1 Testando a normalidade

Os testes de normalidade são utilizados ao iniciar-se uma análise de dados, para verificar se os dados seguem ou não uma distribuição normal. Eles têm por objetivo direcionar o pesquisador a saber qual o tipo de teste será utilizado, se um teste paramétrico ou não paramétrico. Uma questão que pode ser levantada primeiramente é se a maioria das variáveis é normalmente distribuída e, portanto podem ser empregados testes paramétricos sem preocupação quantos às suas restrições. Testes estatísticos com grandes amostras mostram que nem sempre as suposições de normalidades se confirmam. Por outro lado, como nem sempre se dispõe de um número elevado de casos para estudo, às vezes nem é possível decidir se determinada variável possui ou não distribuição normal. Um dos principais testes para normalidade e também um dos mais utilizados e citados na bibliografia estatística é o teste de normalidade de *Shapiro-Wilk* (OLIVEIRA et al., 2009).

O teste de *Shapiro-Wilk* é baseado em estatísticas de ordem da distribuição normal e de seus respectivos valores esperados ou médios. Para esclarecer os conceitos dessas estatísticas, suponha que de uma população normal sejam repetidamente retiradas amostras aleatórias de tamanho n, ou seja, X_1, X_2,..., X_n. Os n valores são ordenados de forma crescente em cada amostra retirada e cada número individualmente é denominado de estatística de ordem. Denota-se a i-ésima estatística de ordem por $X_{(i)}$. O valor esperado da i-ésima estatística de ordem é a média do valor $X_{(i)}$ em repetidas amostras aleatórias de tamanho n retiradas da população. Seja m_i a média dessa estatística de ordem para uma população normal padrão, ou seja, de uma normal com média $\mu = 0$ e variância $\sigma^2 = 1$ e seja $Z_{(1)}$, $Z_{(2)}$,...,$Z_{(n)}$ uma amostra aleatória da distribuição normal (FERREIRA, 2009) então,

$$E(Z_{(i)}) = m_i,$$

é denominado de valor esperado da i-*ésima* estatística de ordem da distribuição normal padrão.

A estatística W do teste de Shapiro-Wilk é dependente do coeficiente a_i, que é denominado por melhor estimador linear não-viesado normalizado do valor esperado das estatísticas de ordem da distribuição normal padrão. A determinação exata desses coeficientes é dependente da determinação exata dos valores esperados e da matriz de variâncias e covariâncias das estatísticas de ordem da distribuição normal padrão. Uma aproximação será considerada posteriormente para possibilitar a aplicação também do teste de normalidade para amostras de tamanhos superiores a 20, sem necessidade de calcular os valores das variâncias e covariâncias das estatísticas de ordem (FERREIRA, 2009). A estatística

W é dada por

$$W = \frac{\left|\sum_{i=1}^{n} a_i x_{(i)}\right|^2}{(n-1)S^2} = \frac{\left|\sum_{i=1}^{n} a_i x_{(i)}\right|^2}{\sum_{i=1}^{n} X_i^2 - \frac{\left|\sum_{i=1}^{n} X_i^2\right|^2}{n}}. \tag{2.1}$$

Se o valor calculado de W é estatisticamente significativo (para $\alpha = 0,05$) rejeita-se a hipótese que a distribuição estudada é normal, ou seja, para a distribuição ser considerada o valor de p deve ser maior que $0,05$.

2.4.2 Testes para o caso de uma amostra

Segundo Siegel (1975), os testes para o caso de uma amostra podem ser aplicados para comprovar uma hipótese que exige apenas a extração de uma amostra. Tais testes nos mostram se determinada amostra pode derivar de uma população especificada, e contrastam com o teste de duas amostras, que comparam e comprovam a probabilidade de as duas serem provenientes da mesma população. Esse tipo de teste de uma amostra é usualmente chamado teste de aderência, tais testes provam ainda se existem diferenças significativas na locação (tendência central) entre amostra e população, se existem diferenças significativas entre as frequências observadas e as frequências que poderíamos esperar com base em um determinado príncipio, se há diferenças significativas entre proporções observadas e proporções esperadas, tal como, se é razoável admitir que a amostra seja uma amostra aleatória de alguma população conhecida.

2.4.2.1 Teste binomial

O teste binomial é aplicado a amostras provenientes de populações consistindo de somente duas classes (dados dicotômicos). Esses resultados podem ser, por exemplo, sim ou não; macho ou fêmea; positivo ou negativo. No entanto, frequentemente, identifica-se tais resultados como **sucesso** e **fracasso**. Aqui, considere que sucesso equivale ao resultado que é objeto de interesse principal do estudo. Assim, pode-se considerar que os dados constituem o resultado de n repetições independentes de um ensaio de Bernoulli com probabilidade de sucesso p. Para facilitar os cálculos, considera-se a notação $q = 1-p$ (CAMPOS, 1983).

A construção do teste se baseia em uma ideia simples. Num experimento como foi descrito acima, a variável aleatória associada ao número de sucessos observados tem, claramente, uma distribuição binomial com parâmetros n e p. Portanto, a probabilidade da amostra conter x sucessos e $n - x$ fracassos é dada por:

$$p(x) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, ..., n \tag{2.2}$$

É claro que, a probabilidade de sucesso é teórica aqui. Para o teste, usa-se um estimador de tal probabilidade, que é a frequência relativa de sucessos. O teste será construído usando a função de distribuição acumulada da binomial, para sabermos qual a probabilidade de se obter valores mais extremos do que os observados. Relembre que a função de distribuição acumulada da distribuição binomial é:

$$F(x) = \sum_{i=0}^{x} \binom{n}{i} p^i q^{n-i}. \tag{2.3}$$

O Método

O procedimento do teste pode ser resumido nos passos a seguir.

1. Como em qualquer procedimento de teste de hipóteses, o primeiro passo é definir as hipóteses que se quer testar. Neste caso, para algum $p_0 \in (0, 1)$, a hipótese nula será

   $$H_0 : p = p_0$$

   e a alternativa

   - $H_1 : p \neq p_0$, para um teste bilateral;
   - $H_1 : p < p_0$ ou $H_1 : p > p_0$, para os testes unilaterais.

2. Determinar o número de casos observados n;

3. Determinar as frequências das ocorrências em cada umas das categorias, $\hat{p}$;

4. Sob H_0, o número de sucessos na amostra tem distribuição $Bin(n, p_0)$. O interesse está em calcular a probabilidade de, sob H_0, se observar um número de sucessos mais extremo do que aquele observado na amostra. Esse "extremo" será determinado pela hipótese alternativa. Por exemplo, se $H_1 : p < p_0$, então será calculada a probabilidade de se obter valores menores do que o observado. O tamanho da amostra dirá que tipo de abordagem será usada.

   (a) Se $n \leq 25$, deve-se determinar a probabilidade sob H_0, de ocorrência do valor observado de x, ou de um valor ainda mais extremo, usando a fórmula da função de distribuição acumulada

$$F(x) = \sum_{i=0}^{x} \binom{n}{i} p_0^i q_0^{n-i}. \tag{2.4}$$

(b) Se ($n > 25$) pode se demonstrar que quando n cresce, a distribuição binomial tende para a distribuição normal. Tal convergência será mais acelerada se p estiver próximo de 1/2. Os parâmetros a serem utilizados serão a média $\mu_x = np_0$ e o desvio padrão $\sigma_x = \sqrt{np_0q_0}$. Assim, se Z tem distribuição aproximadamente normal com média 0 e variância 1, e H_0 pode ser comprovada mediante:

$$z = \frac{x - \mu_x}{\sigma_x} = \frac{x - np_0}{\sqrt{np_0q_0}} \tag{2.5}$$

A aproximação torna-se excelente quando se incorpora uma correção de continuidade. A correção é necessária porque a distribuição normal se refere a uma variável contínua, enquanto que a distribuição binomial envolve uma variável discreta. A correção de continuidade consiste em diminuir de 0,5 a diferença entre o valor observado de x e o valor esperado $\mu_x = np_0$. Portanto, quando $x < \mu_x$ somamos 0,5 a x, e quando $x > \mu_x$, subtraímos 0,5 de x. Ou seja, a diferença observada fica reduzida de 0,5. Z é dado por:

$$Z = \frac{x \pm 0,5 - np_0}{\sqrt{np_0q_0}} \tag{2.6}$$

Os valores da estatística de teste devem ser comparados com os quantis de uma distribuição normal padrão, escolhidos de acordo com o tipo de teste que está sendo feito (uni ou bilateral).

Se o valor de p associado ao valor observado de x, ou ainda a um valor mais extremo, não supera α, rejeitamos H_0 (SIEGEL, 1975).

2.4.2.2 Teste de qui-quadrado (χ^2)

Este é um teste amplamente utilizado quando se tem dados divididos em $k \geq 2$ categorias e sua origem é devida a Karl Pearson. Ele compara as frequências de elementos de cada categoria na amostra observada com valores teóricos, e então, testa-se se há diferença significativa entre os valores. Siegel (1975) descreve que esta técnica é do tipo teste de aderência, no sentido que pode ser empregada para comprovar se existe diferença significativa entre o número observado de indivíduos, ou respostas, em determinada categoria, e o respectivo número esperado, baseado na hipótese de nulidade. Ou seja, a técnica do

χ^2 testa se as frequências observadas estão suficientemente próximas das esperadas para justificar sua ocorrência sob a hipótese nula.

É válido observar que usa-se uma estatística de teste com uma distribuição aproximada, então é importante que não se tenha uma amostra muito pequena. É comum na literatura usar-se a seguinte regra:

- Se o número de classes é igual a dois, isto é, $k = 2$, a frequência esperada mínima deve ser superior ou igual a 5;
- Se $K > 2$, o teste (χ^2) não deve ser usado se mais de 20% das frequências esperadas forem menores do que 5 ou se qualquer uma delas for inferior a 1.

O Método

O método envolve os seguintes passos:

i) Enquadrar as frequências observadas nas k categorias. A soma das frequências deve ser n, número de observações independentes.

ii) Por meio de H_0, determinar as frequências esperadas ($E_i's$) para cada uma das k células.

iii) Calcular o valor de (χ^2) por meio da seguinte fórmula:

$$\chi^2 = \frac{\sum_{i=1}^{k} (O_i - E_i)^2}{E_i} \tag{2.7}$$

em que, O_i é o número de casos observados na i-ésima categoria; E_i é o número de casos esperados na i-ésima categoria sobre H_0; K é o número de categoria na classificação.

iv) Determinar o número de graus de liberdade, $gl = k - 1$;

v) Com base na tabela dos valores críticos do χ^2, determinar a probabilidade associada à ocorrência sob a hipótese H_0 de um valor tão grande quanto o valor observado de χ^2 para o valor do grau de liberdade considerado. Se o valor de p assim obtido for igual a, ou menor do que α rejeitar H_0 (SIEGEL, 1975).

2.4.2.3 Teste de Kolmogorov-Smirnov

Este teste foi introduzido para adaptação de uma específica e bem conhecida distribuição $S_N(x) = F(x)$ a dados provenientes de uma distribuição desconhecida $F_0(x)$. A hipótese de nulidade especifica alguma distribuição $F(x)$. Uma amostra X_1, X_2,...,X_n, é retirada de alguma população cuja distribuição $F_0(x)$ é desconhecida, estabelecendo-se o confronto com $F(x)$.

Sua vantagem sobre o teste do χ^2 é que ele pode ser aplicado, sem restrição para pequenas amostras, além disso, ele trata dados individualmente, não perdendo informações devido a agrupamentos, como ocorre no teste do χ^2 (CAMPOS, 1983).

O Método

Seja $F_0(x)$ uma distribuição de frequência acumulada completamente especificada, ou seja, a distribuição acumulada, teórica, sob H_0. Isto é para qualquer valor de x, o valor de $F_0(x)$ é a proporção de casos esperados com *scores* iguais a, ou menores de que, x.

Seja $S_N(x) = F(x)$ a distribuição de frequência acumulada de uma amostra aleatória de N observações. Quando x é qualquer *score* possível, $S_N(x) = k/n$ onde k é o número de observações inferiores a x.

Pela hipótese de nulidade, de que a amostra tenha sido extraída da distribuição teórica especificada, espera-se que, para cada valor de x, $S_N(x)$ esteja suficientemente próximo de $F_0(x)$. Ou seja, espera-se sob H_0, que as diferenças entre S_N(x) e $F_0(x)$ sejam pequenas e estejam dentro dos limites dos erros aleatórios. O teste foca a maior entre as diferenças. O maior valor de $F_0(x) - S_N(x)$ é chamado desvio máximo (D) (SIEGEL, 1975), dado por

$$D = \max|F_0(x) - S_N(x)|$$

Passos para aplicação do teste:

1. Especificar a distribuição acumulada esperada sob H_0;

2. Dispor os *scores* observados em uma distribuiçãao cumulativa, fazendo corresponder cada intervalo de S_N(x) com o intervalo comparável de $F_0(x)$;

3. Para cada posto da distribuição cumulativa, subtrair $S_N(x)$ de $F_0(x)$.

4. Determinar o valor de D:

$$D = \max |F_0(x) - S_N(x)|$$

5. Mediante referência a tabela de valores críticos de Kolmogorov-Smirnov, determinar a probabilidade bilateral associada à ocorrência, sob H_0, de valores tão grandes quanto o valor observado de D. Se o valor de p correspondente é igual, ou inferior, a α, rejeitar H_0.

2.4.3 Testes para o caso de duas amostras relacionadas

Os testes estatísticos não paramétricos de duas amostras são empregados quando se tem interesse em decidir se dois tratamentos diferem entre si, ou ainda se determinado tratamento é mais eficaz que outro. Inúmeras vezes nos deparamos com situações, em que estamos interessados em estabelecer comparações entre dois tratamentos. Nesse caso, duas situações devem ser consideradas (CAMPOS, 1983):

i) Os indivíduos são tomados ao acaso, em duas amostras, sendo uma delas um grupo controle.

ii) Os indivíduos são subdivididos ao acaso, em duas amostras, cada uma recebendo um determinado tratamento.

Em ambas as alternativas, considerando as amostras independentes, uma grande dispersão de dados poderá comprometer seriamente a validade acerca das conclusões obtidas. A fim de contornar o problema, visando a redução da heterogeneidade, recomenda-se estruturar os dados em pares, onde cada elemento receberá um dos tratamentos.

2.4.3.1 Teste de Wilcoxon

O teste dos postos assinalados de Wilcoxon é utilizado quando se quer comparar médianas entre dois grupos, no caso em que se tem dados pareados. Ele é uma alternativa (não paramétrica) para o teste t pareado. Assim como acontece com os demais procedimentos não paramétricos, ele deve ser usado se as suposições do teste t estiverem seriamente comprometidas.

Segundo Siegel (1975), este teste é extremamente útil para as ciências do comportamento, com dados sobre o comportamento, não são raros os casos em que o pesquisador pode dizer qual membro de um par é *maior do que* o outro e dispor as diferenças por ordem de seu valor absoluto. Ou seja, o pesquisador pode fazer o julgamento do tipo *maior do que* entre os resultados d (diferenças) e qualquer par, bem como julgar em relação às diferenças relativas a dois pares quaisquer.

Para proceder com o teste, considera-se primeiramente d_i como sendo a diferença entre os elementos de determinado par de observações$(X_i,\, Y_i)$, isto é, a diferença entre os dois *scores*, sob dois tratamentos. Cada par tem um d_i. Para empregar o teste de Wilcoxon, atribui-se postos a todos os $d_i's$ independentemente de sinal, ao menor d_i, o posto 1, posto 2 ao mais próximo e assim por diante. Quando se atribuem postos independentemente de sinal, um valor de d_i= -1 tem posto inferior tanto a um d_i = -2 como igual a +2, por exemplo. Em seguida, a cada posto atribui-se o sinal da diferença. Ou seja, indicar quais postos decorrem de $d_i's$ negativos e quais de $d_i's$ positivos (SIEGEL, 1975).

Ocasionalmente, podem ocorrer empates, ou seja, não se observou diferença entre os resultados entre os tratamentos, de forma que $d = 0$. Neste caso, tais pares são excluídos da análise, e considera-se n como o número de pares com $d \neq 0$. Outra situação que deve ser citada é aquela em que dois ou mais pares apresentam a mesma diferença , isto é, $d_i's$ que apresentam o mesmo valor. O que se faz nesse caso, é atribuir a cada um desses pares um posto igual que é a média dos postos que teriam sido atríbuidos se os $d_i's$ fossem diferentes.

O procedimento geral do teste pode ser resumido nos passos a seguir.

1. Fixe o nível de significância α.
2. A hipótese nula é que a diferenças entre as medidas é igual ($H_0 : \mu_1 = \mu_2$)
3. Calcule a diferença entre a primeira e a segunda medida, ou seja, a "amostra das diferenças".
4. Exclua os valores que deram iguais a 0, se existirem.
5. Ordene a "amostra das diferenças" sem levar em consideração o sinal e a cada valor atribua seu respectivo posto.
6. Coloque sinais nos postos, de acordo com o sinal das diferenças. Os p assinalados serão indicados por R.
7. Sob H_0, as medidas no mesmo par deveriam ser iguais. Então, sob H_0, $\sum R$ deveria ser igual a 0. Assim, se esta soma estiver "muito longe" de 0, podemos concluir que

as medidas são diferentes.

8. Calcule a estatística do teste:

$$Z = \frac{\sum R}{\sqrt{\sum R^2}} \sim N(0,1).$$

9. Desta forma, o teste consiste em comparar o valor calculado de Z com o valor z_c.

Quando as amostras são pequenas, a distribuição de Z não se aproxima satisfatoriamente da normal. Nesses casos, não é adequado usar a fórmula vista. O que é feito, então, é somar os postos com sinal positivo e os postos com sinais negativos. O valor absoluto da menor soma de postos será chamado de T e deverá ser comparado com valores críticos encontrados em tabelas específicas. Tais tabelas estão disponíveis em diversos livros, como por exemplo Kanji(1999) e Zar(1999).

2.4.4 Teste para o caso de duas amostras independentes

Assim como os testes de duas amostras relacionadas, os testes de duas amostras independentes determinam se as diferenças nas amostras constituem evidências convincentes de uma diferença nos processos, ou tratamentos, aplicados a elas. A principal diferença entre elas é de que as amostras são independentes. Em nenhum dos casos se exige que as amostras tenham mesmo tamanho.

2.4.4.1 Teste U de Mann-Whitney

Este teste foi introduzido em 1947 e é uma alternativa não paramétrica para o teste t para duas amostras. Ele é usado quando se quer comparar médias de duas populações, e temos duas amostras independentes das mesmas.

Considere duas amostras independentes de duas populações A e B. Considere ainda que há interesse em se testar a hipótese nula

$$H_0 : \mu_A = \mu_B$$

contra hipótese alternativa que pode ser de uma das formas

- $H_1 : \mu_A \neq \mu_B$, para testes bilaterais,
- $H_1 : \mu_A > \mu_B$ ou $H_1 : \mu_A < \mu_B$, para testes unilaterais.

O teste começa juntando as duas amostras como se fosse uma grande amostra única e assinalando o posto de cada observação. Depois, somam-se os postos do grupo 1 ($\sum R_1$) e do grupo 2 ($\sum R_2$). Aqui, vale ressaltar que o grupo 1 é a amostra menor. Se as duas amostras tiverem o mesmo tamanho, o grupo 1 será aquele cuja soma de postos for menor.

O restante do procedimento do teste será discutido a seguir e leva em consideração o tamanho das amostras que se tem. Quando o tamanho da menor amostra não ultrapassa 20 e o tamanho da maior amostra não ultrapassa 40, Zar(1999) traz em um de seus anexos uma tabela com valores críticos para a tomada de decisão.

Para valores que não são encontrados na tabela, usa-se o fato de que à medida que n_1, n_2 aumentam, a distribuição amostral de U tende rapidamente para a distribuição normal, com média (2.8)

$$\mu_U = \frac{n_1 n_2}{2},$$

e desvio Padrão:

$$\sigma_U = \sqrt{\frac{(n_1)(n_2)(n_1 + n_2 + 1)}{12}}$$

Assim, para n_1 e n_2 "grandes", pode-se determinar a significância de um valor observado de U mediante a relação:

$$z = \frac{U - \mu_U}{\sigma_U} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{(n_1)(n_2)(n_1+n_2+1)}{12}}} \sim N(0, 1) \tag{2.8}$$

sob H_0. Ou seja a probabilidade associada a ocorrência, sob H_0 de valores tão extremos quanto um z observado pode ser determinada com o auxílio de uma tabela da distribuição normal padrão.

Quando ocorrem *scores* empatados, atribuímos a cada uma das observações empatadas a média dos postos que lhes seriam atribuídos se não houvesse empate. Se os empates ocorrem entre duas ou mais observações do mesmo grupo, o valor de U não é afetado. Mas se os empates ocorrem entre duas ou mais observações envolvendo ambos os grupos, o valor de U é afetado. Conquanto os efeitos práticos de tais empates sejam desprezíveis, existe uma correção para empates para uso com a aproximação normal empregada para grandes amostras. O efeito de postos empatados consiste em modificar a variabilidade do conjunto de postos, assim, a correção para empates deve ser aplicada ao desvio-padrão da distribuição amostral de U. Com essa correção para empates, o desvio-padrão se apresenta como:

$$\sigma_U = \sqrt{\left(\frac{n_1 n_2}{N(N-1)}\right)\left(\frac{N^3-N}{12} - \sum_{i=1}^{N} T\right)}$$

em que, N = $n_1 + n_2$ e

$$T = \frac{t^3 - t}{12}$$

$\sum_{i=1}^{N} T$ obtém-se somando os $T's$ sobre todos os grupos de observações empatadas.

Com a correção para empates, obtemos o valor de z mediante:

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\left(\frac{n_1 n_2}{N(N-1)}\right)\left(\frac{N^3-N}{12} - \sum_{i=1}^{N} T\right)}} \sim N(0,1) \tag{2.9}$$

e também aqui a decisão será feita com base em valores críticos de uma distribuição normal padrão.

2.5 Testes para o caso de *k* amostras relacionadas

A finalidade desse teste é apresentar técnicas para comprovar a significância de diferenças entre três ou mais grupos relacionados, isto é, para comprovar a hipótese de que três ou mais amostras, tenham sido extraídas da mesma população ou de populações idênticas.

2.5.1 Teste Q de Cochran

De acordo com Siegel (1975), o teste Q de Cochram para k amostras relacionadas, proporciona um método para comprovar se três ou mais conjuntos correspondentes de frequências diferem significativamente entre sí. O teste Q adapta-se especialmente ao caso em que os dados se apresentam em escala nominal ou sob forma de informação ordinal dicotomizada.

O método

Se os dados de uma pesquisa se dispõem em uma tabela de dupla entrada com n linhas e k colunas é possível testar a hipótese de nulidade, de que a proporção ou frequência de respostas de determinado tipo seja a mesma em cada coluna, executadas as diferenças devidas ao acaso. Cochram mostrou (1950) que a hipótese de nulidade é verdadeira, isto é, se não há diferença na probabilidade de sucesso sob cada condição (o que equivale a dizer que os sucessos ou fracasso se distribuem aleatoriamente entre as linhas e colunas da tabela de dupla entrada), então se o número de linhas não é muito pequeno, tem distribuição aproximadamente qui-quadrado com $gl = k - 1$,

$$Q = \frac{(k-1)\left[k\sum_{j=1}^{k} G_j^2 - \left(\sum_{j=1}^{k} G_j\right)^2\right]}{k\sum_{i=1}^{N} L_i - \sum_{i=1}^{N} L_i^2} \tag{2.10}$$

sendo, G_j o número total de sucessos na coluna j; G a média dos G_j, L_i o número total de sucessos na linha i.

Como a distribuição amostral de Q é aproximadamente qui-quadrado com $gl = k - 1$, a probabilidade associada à ocorrência, sob H_0, de valores tão grandes quanto um Q observado pode ser determinada através de uma tabela da distribuição qui-quadrado, a qual pode ser encontrada em diversos livros de Estatística, por exemplo, Siegel(1975). Se o valor observado de Q calculado por (2.10), não é inferior ao valor crítico para um dado nível de significância α e detrminado valor de $gl = k - 1$, implica que a proporção ou frequência de sucessos difere significamente entre as várias amostras. Isto é, H_0 pode ser rejeitada ao nível α (SIEGEL, 1975).

O Método se resume nos seguintes passos

1. Para dados dicotomizados, atribuir o *score* 1 a cada sucesso, e o *score* 0 a cada fracasso.
2. Dispor os dados em uma tabela N x K, com k colunas e N linhas. N é o número de casos em cada um dos k grupos.
3. Determinar o valor de Q através da fórmula (2.10)
4. Decisão, se a probabilidade associada à ocorrência, sob H_0, de um valor tão grande quanto um valor observado de Q não supera α, rejeita-se a hipótese H_0.

2.5.2 Teste de Friedman

Segundo Campos (1983) o teste de Friedman pode ser considerado uma extensão do teste bilateral do sinal para amostras pareadas. O teste é útil para comprovar a hipótese de nulidade, de que as k amostras tenham sido extraídas da mesma população, com relação aos postos médios. Pode ser utilizado quando a mensuração da variável se apresenta no minímo em escala ordinal.

O método

Para o teste de Friedman, os dados se dispõem em uma tabela de dupla entrada com N linhas e K colunas. As linhas representam o conjunto correspondente de indivíduos, e as colunas as diversas condições. Se estão sendo estudados os *scores* de indivíduos observados sob todas as condições, então cada linha fornece o *score* de um indivíduo sob as K condições. Os dados do teste são postos. Aos *scores* de cada linha atribue-se postos separadamente. Ou seja, com K condições em estudo, os postos em qualquer linha vão de 1 a K, o teste determina se é provável que as diferentes colunas de postos (amostras) provenham da mesma população (SIEGEL, 1975).

Para aplicar esse teste, calculamos uma estatística que Friedman denotou por χ_r^2. Quando o número de linhas e/ou colunas não é muito pequeno, pode-se mostrar que χ_r^2 tem distribuição aproximadamente qui-quadrado, com $gl = k - 1$, sendo:

$$\chi_r^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} (R_j)^2 - 3N(k+1) \tag{2.11}$$

em que, N é número de linhas, K é número de colunas, R_j é soma de postos da coluna j e $\sum_{j=1}^{k} (R_j)^2$ é o somátorio dos quadrados das somas de postos sobre todas as k condições.

O Método se resume nos seguintes passos

1. Dispor os *scores* em uma tabela de dupla entrada NxK com N linhas (indivíduos do grupo) e K colunas (condições).
2. Atribuir postos de 1 a k aos *scores* em cada linha.
3. Determinar a soma dos postos em cada coluna R_j.
4. Calcular o valor de χ_r^2, pela fórmula (2.11)
5. Comparar o valor observado da estatística com valores teóricos da distribuição χ^2.

6. Se a probabilidade obtida pelo método adequado indicado no item 5 não supera o nível de significância α, rejeita-se H_0.

2.6 Teste para o caso de k amostras independentes

Os testes para o caso de k amostras independentes, apresentam técnicas para comprovar a significância de diferenças entre três ou mais grupos independentes de amostras, ou seja, para comprovar a hipótese de nulidade, de que k amostras independendetes tenham a mesma média. A técnica paramétrica usual para tal comprovação é a análise de variância ou teste F. As suposições associadas ao modelo estatístico básico do teste F são que as observações tenham sido extraídas independentemente de populações normalmente distribuídas, todas com a mesma variância. Exige-se ainda que a mensuração da variável em estudo tenha sido feita no mínimo em escala intervalar.

Se o pesquisador acha que tais suposições são irreais para seus dados, ou ainda se ele deseja evitar as suposições restritivas inerentes ao teste F, tornando, assim, suas conclusões mais gerais, poderá então empregar um teste não paramétrico para k amostras independentes. Tais testes tem ainda a vantagem de permitir estudar quanto a significância, dados que são inerentemente apenas classificativos em escala nominal ou que se apresentem em postos, escala ordinal (SIEGEL, 1975).

2.6.1 Teste de Kruskal-Wallis

O teste de Kruskal-Wallis foi introduzido por estes autores em 1952, como um competidor ou um substituto do teste F do campo paramétrico. Sua finalidade é estabelecer confronto entre k amostras independentes (CAMPOS, 1983). O teste é extremamente útil para decidir se k amostras independentes provêm de populações com médias diferentes. Os valores amostrais quase que invariavelmente diferem entre si, e o problema é decidir se essas diferenças entre as amostras significam diferenças efetivas entre as populações, ou se representam apenas variações casuais, que podem ser esperadas entre amostras aleatórias de uma mesma população (SIEGEL, 1975).

Pressuposições

Segundo Campos (1983) as pressuposições gerais a serem consideradas para a realização desses testes são:
i) As observações são todas independentes;
ii) Dentro de uma dada amostra, todas as observações são provenientes de uma mesma população;
iii) as k populações são aproximadamente da mesma forma e contínuas.

Hipóteses

De acordo com a estrutura do teste pode-se considerar:

$$\begin{cases} H_0 : \mu_1 = \mu_2 = ... = \mu_k \\ vs \\ H_1 : \mu_i \neq \mu_j \quad \text{para pelo menos um par de índices, com } i, j = 1, 2, \ldots, k, i \neq j \end{cases}$$

em que pelo menos duas médias diferem entre si.

O Método

No cálculo do teste de Kruskal-Wallis cada uma das n observações é substituída por um posto. Ou seja, todos os *scores* de todas as k amostras combinadas são dispostos em uma única série de postos. Ao menor *score* atribui-se o posto 1, ao seguinte o posto 2, e assim por diante, ao maior o posto n. n é o número total de observações independentes nas k amostras. Feito isto determina-se a soma dos postos em cada amostra (coluna). O teste determina se essas somas são tão díspares que não seja provável que a amostra seja extraída da mesma população (SIEGEL, 1975).

Pode-se mostrar que as k amostras provêm efetivamente de populações com a mesma mediana, isto é se H_0 é verdadeira, então H a estatística do teste tem distribuição qui-quadrado com $gl = k - 1$, desde que os tamanhos das k amostras não sejam muitos pequenos, H é calculada pela seguinte fórmula:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1) \tag{2.12}$$

em que, k é o número de amostras, n_j é o número de casos na amostra j, R_j é a soma de postos na amostra (coluna) j, $n = \sum n_j$ é número de casos em todas as amostras combinadas, $\sum\limits_{j=1}^{k}$ indica somatório sobre todas as k amostras (colunas), tem distribuição aproximadamente qui-quadrado com $gl = k - 1$, para tamanhos de amostras $(n_j's)$ suficientemente grandes (SIEGEL, 1975).

Tamanho de amostras

Quando há mais de 5 casos nos diversos grupos, isto é n_j >5, a probabilidade associada a ocorrência sob H_0, de valores tão grandes quanto um H observado pode ser determinada com o auxílio da tabela de valores críticos do qui-quadrado. Se o valor observado de H é igual ou superior ao valor de qui-quadrado constante para o nível de significância α fixado para $gl = k - 1$, então H_0 pode ser rejeitada ao nível α.

Quando $k = 3$ e o número de casos em cada uma das 3 amostras é 5 ou menos, a aproximação qui-quadrado para a distribuição amostral de H não é suficiente. Para tais casos tabelaram-se probabilidades exatas apartir da fórmula (2.12); essas probabilidades constam na tabela de Kruskal-Wallis. A primeira coluna da tabela dá o número de casos nas 3 amostras, isto é, os diversos valores possíveis de n_1, n_2 e n_3. A segunda coluna dá diversos valores de H, calculados pela fórmula (2.12). A terceira dá a probabilidade associada a ocorrência , sob H_0 de valores tão grandes quanto um H observado (SIEGEL, 1975).

Caso de empates

Quando ocorrem empates entre dois ou mais *scores*, atribui-se a cada um deles a média dos postos respectivos. Como o valor de H é de certo modo influenciado pelos empates, pode-se querer introduzir uma correção para empates no cálculo de H. Para este fim o valor de H é calculado por:

$$\mathrm{H} = \frac{\frac{12}{\mathrm{n(n+1)}} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)}{1 - \frac{\sum \mathrm{T}}{\mathrm{n}^3 - n}} \tag{2.13}$$

em que, $T = t^3 - t$ (t sendo o número de observações empatadas em um grupo de *scores* empatados), n é o número de observações em todas a k amostras conjuntamente, isto é $n = \sum n_j$, $\sum T$ indica somatório sobre todos os grupos de empates.

De acordo com (SIEGEL, 1975), o efeito da introdução da correção para empates é de aumentar o valor de H, tornando o valor de H mais significativo. Portanto o pesquisador pode rejeitar a hipótese de nulidade H_0 com uma razão mais forte.

O método se resume nos seguintes passos

1. Dispor, em postos as observações de todos os k grupos em uma única série, atribuindo-lhes postos de 1 a N.
2. Determinar o valor de R (soma dos postos) para cada um dos k grupos de postos.
3. Se houver grande proporção de empates, calcular o valor de H pela fórmula (2.13). Em caso contrário utilizar a fórmula (2.13).
4. O método para determinar a significância do valor observado de H depende do tamanho d k e do tamanho dos grupos

i) Se $k = 3$ e n_1, n_2, n_3 $\leq$5, pode-se utilizar a tabela de kruskal-Wallis para determinar a probabilidade associada, sob H_0 de um H tão grande quanto o observado.

ii) Em outros casos, a significância de um valor tão grande quanto o valor observado de H pode ser determinada mediante referência a tabela de valores críticos do qui-quadrado, com $gl = k - 1$.

5. Se a probabilidade associada ao valor observado de H não supera o nível de significância α fixado, rejeitar a hipótese H_0 em favor de H_1 (SIEGEL, 1975).

3 Aplicação

Diante da crescente demanda por métodos estatísticos cada vez mais eficientes que possam elucidar respostas mais confiáveis, neste trabalho aplicou-se métodos não- paramétricos para os dados do ENEM ano de 2011 na região nordeste, uma vez que a ausência de normalidade entre os dados foi detectada.

Encontram-se nesta seção as principais metodologias que serviram de base para este trabalho, explorando o uso dos testes não paramétricos mais apropriados aos dados.

3.1 Análise da região Nordeste

Os dados utilizados neste trabalho foram obtidos no site http://www.dados.gov.com.br/. O interesse foi o de avaliar as notas do ENEM das escolas do ensino médio da rede pública e privada da região Nordeste, cuja participação foi maior ou igual a 50%, e as variáveis analisadas foram as médias das provas de Ciências da Natureza (CN), Ciências Humanas (CH), Matemática (M), Linguagens e Códigos (LC), Objetivas (O) e Redação (R). A caracterização das variáveis pode ser representada por meio da Figura 3

Vale salientar que a amostra utilizada para avaliar cada estado foi a seguinte: para o estado de Alagoas $n_1 = 44$, Bahia $n_2 = 167$, Ceará $n_3 = 219$, Maranhão $n_4 = 104$ Paraíba $n_5 = 88$, Pernambuco $n_6 = 91$, Piauí $n_7 = 77$, Rio Grande do Norte $n_8 = 92$ e Sergipe $n_9 = 54$, totalizando no geral uma amostra de n = 934 escolas. Todas as amostras são consideradas independentes.

Com o intuito de conhecer o perfil dos dados, aplicou-se o teste de Shapiro-Wilk, para verificar se os dados em análise apresentam normalidade. As hipósteses testadas foram H_0: A amostra provém de uma população Normal *versus* H_1: A amostra não provém de uma população o Normal. O teste de Shapiro-Wilk foi utilizado para cada variável considerada no banco de dados, com isso verificou-se que todas as disciplinas avaliadas nos 9 estados da região nordeste não poderiam ser considerados provenientes de uma

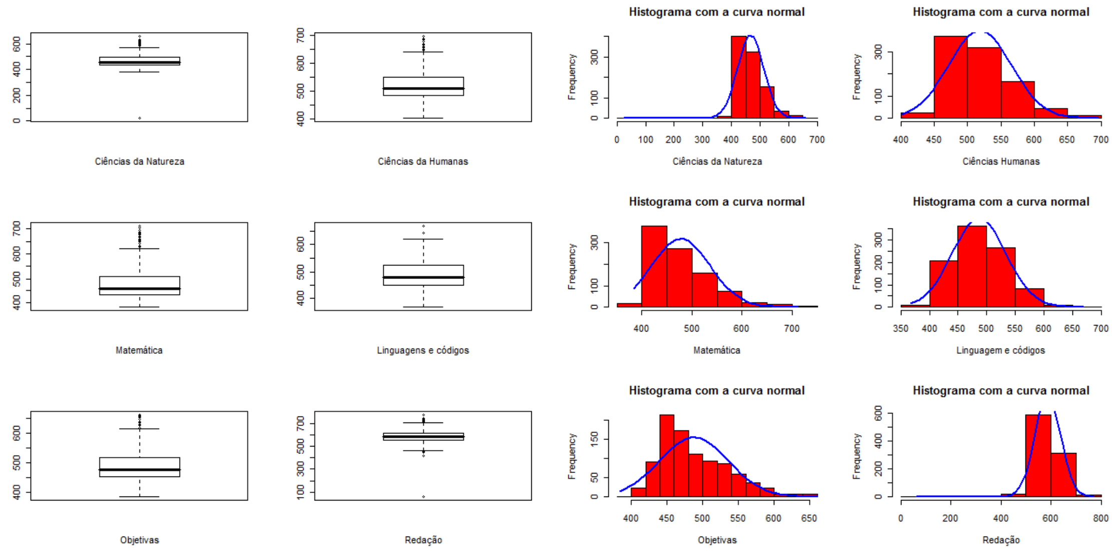

Figura 3: Box-Plot e Histograma com sobreposição da densidade normal para as variáveis CN, CH, M, LC, O, R para as provas do ENEM de 2011 na Região Nordeste.

distribuição normal, independente da rede de ensino. O valor da estatística W e o Valor P correspondente encontra-se na Tabela 2.

Tabela 2: Teste de Shapiro-Wilk para as variáveis Ciências da Natureza (CN), Ciências Humanas (CH), Matemática (M), Linguagens e Códigos (LC), Objetivas (O) e Redação (R) da Região Nordeste referentes ao ENEM de 2011

Variáveis	Estatística W	Valor P
CN	0,8930	$< 0,001$
CH	0,9535	$< 0,001$
M	0,9034	$< 0,001$
LC	0,9807	$< 0,001$
O	0,9420	$< 0,001$
R	0,9348	$< 0,001$

Após o teste constatamos que os dados do Nordeste em geral não apresentaram normalidade, assim a aplicação de testes não paramétricos pode ser realizada, uma vez que uma das principais suposições dos testes paramétricos foi violada.

3.1.1 Análise não paramétrica

Considerando-se o fato do teste de Kruskal-Wallis ser uma alternativa não paramétrica à análise que se faz por recorrência à estatística F e por poder comparar amostras independentes optou-se por aplicar o mesmo com o intuito de verificar se as amostras analisadas provêm de populações diferentes. A hipótese nula neste caso indica que todas as amostras vem de uma população com médias iguais, porém no caso em questão verifica-se que existem evidências estatísticas altamente significantes (valor p $< 0,001$) de uma diferença entre as variáveis analisadas em função das redes de ensino (Tabela 3).

Tabela 3: Teste de Kruskal-Wallis para as variáveis Ciências da Natureza (CN), Ciências Humanas (CH), Matemática (M), Linguagens e Códigos (LC), Objetivas (O) e Redação (R) da Região Nordeste referentes ao ENEM de 2011

Variáveis	Estatística χ^2	gl	Valor P
CN	479,5684	1	$< 0,001$
CH	474,9506	1	$< 0,001$
M	476,7995	1	$< 0,001$
LC	517,5340	1	$< 0,001$
O	507,3667	1	$< 0,001$
R	387,7919	1	$< 0,001$

O teste U de Mann-Whitney é apresentado na Tabela 4. Neste caso estamos interessados em contestar as seguintes hipóteses:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ vs \\ H_{1:} : \mu_1 \neq \mu_2 \end{cases}$$

em que, a hipótese de nulidade H_0, é a hipótese de não existir diferenças entre as provas da rede pública e privada e a hipótese alternativa H_1 alega que existe diferenças entre as redes. Sendo μ_1 a média das provas das escolas públicas e μ_2 a média das provas das escolas privadas. Logo ao nível de significância de 0,05, pode-se concluir que dentre as variáveis analisadas os dados provém de populações não idênticas. O que implica que as médias das duas redes são diferentes, sendo a média da rede pública menor que da privada.

Tabela 4: Teste U de Mann-Whitney para as variáveis Ciências da Natureza (CN), Ciências Humanas (CH), Matemática (M), Linguagens e Códigos (LC), Objetivas (O) e Redação (R) da Região Nordeste referentes ao ENEM de 2011

Variáveis	Estatística U	Valor P
CN	185961	$< 0,001$
CH	185544,5	$< 0,001$
M	185711,5	$< 0,001$
LC	189312	$< 0,001$
O	188427,	$< 0,001$
R	177265,	$< 0,001$

3.2 Análise para o estado da Paraíba

De maneira análoga ao que foi feito para a região nordeste, todos os testes foram aplicados ao estado da Paraíba com o objetivo de comparar o estado em relação a região. O objetivo é verificar e comparar o que as notas das disciplinas Ciências da Natureza (CN), Ciências Humanas (CH), Matemática (M), Linguagens e Códigos (LC), Objetivas (O) e Redação (R) do estado da Paraíba referentes ao ENEM de 2011 em relação a região Nordeste. O primeiro aspecto averiguado diz respeito a amostra possuirem distribuição normal ou não aplicando-se teste de Shapiro-Wilk (Tabela 5)

Tabela 5: Teste de Shapiro-Wilk para as variáveis Ciências da Natureza (CN), Ciências Humanas (CH), Matemática (M), Linguagens e Códigos (LC), Objetivas (O) e Redação (R) do estado da Paraíba referentes ao ENEM de 2011

Variáveis	Estatística W	Valor P
CN	0,6464	$< 0,001$
CH	0,969	$< 0,05$
M	0,9034	$< 0,001$
LC	0,9807	$0,2581$
O	0,9420	$< 0,001$
R	0,9348	$0,28$

Pelo teste apresentado na Tabela 5, dentre todas as variáveis avaliadas, percebe-se que as variáveis Linguagen e Códigos (LC) e Redação (R), apresentaram um valor de W não significativo, logo não rejeita-se a hipótese H_0 de que os dados provêm de uma população normal. Esse comportamento pode ser observado por meio da Figura 4.

O teste U de Mann-Whitney é apresentado na Tabela 6 observa-se que a hipótese de nulidade H_0 foi rejeitada pois existem evidências estatísticas altamente significantes (valor p $<$ 0,001). Logo concluímos ao nível de significância α = 0,05 que existem diferenças entre as médias das redes públicas e privadas, ou seja, as variáveis analisadas são de populações não idênticas.

Com a finalidade de verificar se as amostras analisadas do estado da Paraíba vem de populações diferentes aplicou-se o teste de Kruskal-Wallis (Tabela 7). A hipótese nula neste caso indica que todas as amostras vem de uma mesma população, ou seja, não existe diferenças entre as médias das redes de ensino. Logo concluímos que existem evidências estatísticas altamente significantes (valor p $< 0,001$) de uma diferença entre as variáveis analisadas em função das redes de ensino.

Por meio dos testes realizados, o que se pode observar é que tanto na região Nordeste

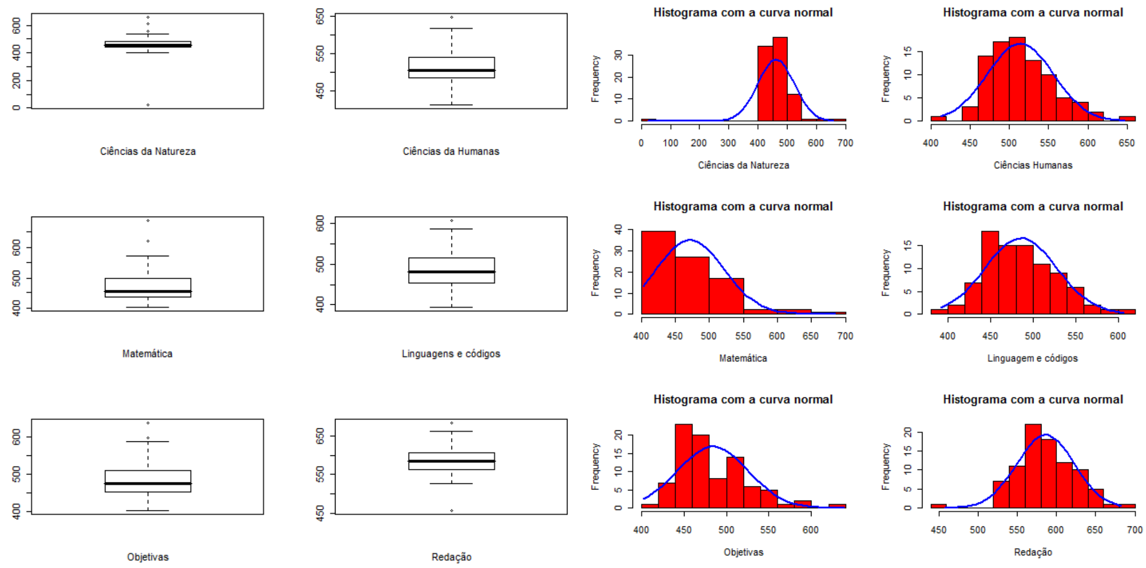

Figura 4: Box-Plot e Histograma com sobreposição da densidade normal para as variáveis CN, CH, M, LC, O, R para as provas do ENEM de 2011 no estado da Paraíba

Tabela 6: Teste U de Mann-Whitney para as variáveis Ciências da Natureza (CN), Ciências Humanas (CH), Matemática (M), Linguagens e Códigos (LC), Objetivas (O) e Redação (R) do Estado da Paraíba referentes ao ENEM de 2011

Variáveis	Estatística U	Valor P
CN	1578	$< 0,001$
CH	1543	$< 0,001$
M	1556	$< 0,001$
LC	1598	$< 0,001$
O	1587	$< 0,001$
R	1521	$< 0,001$

Tabela 7: Teste de Kruskal-Wallis para as variáveis Ciências da Natureza (CN), Ciências Humanas (CH), Matemática (M), Linguagens e Códigos (LC), Objetivas (O) e Redação (R) do estado da Paraíba referentes ao ENEM de 2011

Variáveis	Estatística χ^2	gl	Valor P
CN	46,6037	1	$< 0,001$
CH	42,3798	1	$< 0,001$
M	43,9251	1	$< 0,001$
LC	49,1067	1	$< 0,001$
O	47,7217	1	$< 0,001$
R	39,8288	1	$< 0,001$

em geral, quanto para o estado da Paraíba, as hipóteses de nulidade de que não existiam diferenças entre as médias das provas de alunos de escolas de rede públicas e privadas foram rejeitadas ao nível de significância $\alpha = 0{,}05$. Encontrando assim um resultado que já era esperado, poís o ensino de escolas da rede pública infelizmente ainda é muito falho, o que pode desfavorecer alunos que dependem deste ensino em tais exames.

Por meio destes resultados, percebemos que a Paraíba não possui inferioridade nas redes de ensino, com relação aos demais estados da região Nordeste, ou seja o nível de conhecimento dos alunos de redes públicas e privadas, são basicamente aproximados para toda região, uma vez que foram encontradas diferenças significativas entre as redes em todo Nordeste inclusive na Paraíba, fator que nos diz que o ensino público é falho em todo Nordeste.

4 Conclusão

Os testes de hipóteses são uma importante ferramenta em análises estatísticas. Eles nos fornecem um critério de tomada de decisão, mas, em geral, para garantirmos que os mesmos são confiáveis, precisamos observar uma série de condições. Uma das principais suposições consideradas nos testes paramétricos mais populares é a de que os dados provém de uma distribuição normal. Tal condição nem sempre é plausível, o que torna necessário o uso de outro tipo de abordagem. Os testes não-paramétricos são testes construídos sem uma exigência sobre o modelo probabilístico da população da qual a amostra foi extraída. Em geral eles são baseados em estatísticas de ordem. A vantagem destes é que podem ser usados quando não se estiver seguro sobre a validade dos pressupostos necessários à realização dos testes paramétricos. A desvantagem é que eles são menos poderosos.

Neste trabalho, foi feita uma comparação entre notas do ENEM 2011 do estado da Paraíba com outros estados do Nordeste, além da comparação entre escolas públicas e privadas. A ausência de normalidade entre os dados foi detectada, e, portanto, optou-se pela utilização de uma abordagem não-paramétrica.

Com base nos valores analisados através dos testes não-paramétricos utilizados, observou-se que existe uma diferença significativa entre as médias do ENEM 2011 entre as escolas da rede pública e privada de ensino em todo Nordeste. Observou-se ainda que o estado da Paraíba esta no mesmo caminho que os demais estados da região, em todos os estados o ensino da rede pública foi desfavorecido.

Referências

BOLFARINE, H.; SANDOVAL, M. C. **Introdução à Inferência Estatística**. Rio de Janeiro: Sociedade Brasileira de Matemática, 2001.

BORGES, P. **Tri a teoria por trás do novo ENEM**. Disponível em: http://ultimosegundo.ig.com.br/enem/tri+a+teoria+por+tras+do+novo+enem/n-1237789481686.html. Acesso em: 22 novembro. 2012.

CAMPOS, H. **Estatística Experimental Não-Paramétrica**. 4ª ed. Piracicaba, 1983.

FERREIRA, D.F. **Estatística Básica**. 2ª ed. Lavras: UFLA, 2009.

KANJI, G.K. **100 Statistical Tests**. SAGE Publications, 1999.

OLIVEIRA, M.S. et. al. **Introdução à Estatística** . Lavras: UFLA, 2009.

SIEGEL, S. **Estatística Não-Paramétrica**. Brasil: Mc Graw-Hill, 1975.

ZAR, J.H. **Biostatistical Analysis**. 4ª Ed. Prentice Hall, 1999.