



# **ESTUDO COMPARATIVO ENTRE TÉCNICAS DE SAZONALIDADE PARA A PREVISÃO DE VAZÕES UTILIZANDO REDES NEURAIS.**

**Trabalho de Conclusão de Curso**  
**Engenharia da Computação**

**Bruno Monteiro Figueiredo**  
**Orientador: Prof. Dr. Mêuser Jorge da Silva Valença**





**Universidade de Pernambuco  
Escola Politécnica de Pernambuco  
Graduação em Engenharia de Computação**

**BRUNO MONTEIRO FIGUEIREDO**

**ESTUDO COMPARATIVO ENTRE TÉCNICAS DE  
SAZONALIDADE PARA A PREVISÃO DE VAZÕES  
UTILIZANDO REDES NEURAIS.**

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

**Recife, dezembro de 2014.**



**De acordo**

**Recife**

\_\_\_\_/\_\_\_\_/\_\_\_\_

---

**Orientador da Monografia**

Dedico este trabalho aos meus pais, Fred e Elizabete, aos meus avós, Mauri e Valda, aos meus irmãos, Tata e Báu, à minha linda namorada, Julianne Bezerra e a Deus.

# Agradecimentos

Agradeço em primeiro lugar, a Deus. Ele é o único responsável por eu chegar até aqui e sabe de tudo o que eu passei. Sem Sua vontade e Seu apoio eu não conseguiria finalizar esta etapa.

Agradeço à minha família. Meus pais, Fred e Elizabete, e meus avós, Mauri e Valda, por sempre me apoiarem e acreditarem em mim. Por darem todas as ferramentas que eu precisava para chegar aqui e por apostarem no meu sonho.

Agradeço à minha melhor amiga e namorada, Julianne Bezerra. Por estar sempre ao meu lado nos momentos bons e nos ruins também. Seu incentivo e suporte me permitiram chegar até aqui.

Agradeço à minha amiga Sabrina Andrade por toda sua ajuda e suporte durante a graduação. Além da amizade e companheirismo, foi uma guia quando me encontrava desorientado.

Agradeço ao meu orientador, Prof. Dr. Mêuser Jorge Valença, pela oportunidade de realização deste trabalho, por sua disponibilidade para orientação e por sempre mostrar o melhor caminho.

Agradeço às amigas Bruna Galle e Lara Dantas, por toda a ajuda para conclusão deste trabalho. Sem elas, que assumiram um papel de quase co-orientadores, nada disto estaria concretizado.

Por fim, agradeço a todos que contribuíram para minha formação acadêmica e aos bons professores que proporcionaram uma boa formação em assuntos os quais me identifico: Bruno Fernandes, Carmelo Bastos, Daniel Chaves, Gustavo Carvalho, Sérgio Campelo e Sérgio Murilo.

# Resumo

Um dos maiores desafios de hoje em dia é aproveitar, de maneira eficiente, a produção de energia elétrica. O demasiado crescimento do consumo requer cada vez mais uma produção eficiente. No Brasil, a maior parte da energia elétrica é produzida por usinas hidroelétricas, sendo complementadas por termoelétricas. A quantidade de energia produzida por uma usina hidroelétrica está diretamente ligada ao volume de água e às vazões das bacias hidrográficas onde atuam. Hoje em dia, existem diversos métodos de previsão de vazão das bacias. Este trabalho propõe realizar diversas previsões mensais de vazão de bacias hidrográficas utilizando Redes Neurais Artificiais. Várias configurações de redes serão treinadas com um mesmo conjunto de dados por um *Multi-Layer Perceptron*, tipo de rede comumente utilizada para este fim. Comparações serão realizadas entre todos os resultados e por fim, testes estatísticos comprovarão qual das configurações testadas apresentou melhor desempenho para as previsões.

**Palavras-chave:** Previsão de vazão, MLP, Redes Neurais Artificiais

# Abstract

One of the greatest challenges nowadays is to avail the electric power produce in the best suitable way. The huge growing of consumption requires increasingly an efficient planning for electric power production. In Brazil, the most of electric power is produced by hydroelectric power plants. Thus, the amount of electric power produced depends on the water flow and the volume of the watershed where it is located. Nowadays there are plenty of methods for water flow predictions. Therefore, this work aims to make distinct water flow predictions using artificial neural networks. All the network configurations is trained with the same database by the Multi-Layer Perceptron, a type of artificial neural network widely used for this purpose. At the end, comparisons will be made between the results and statistical tests will prove which configuration has a better performance.

**Keywords:** Water Flow Prediciton, MLP, Artificial Neural Networks.

# Sumário

<b>Capítulo 1 Introdução</b>	<b>1</b>
1.1 Motivação e Problema	1
1.2 Objetivos	3
1.2.1 Objetivos Gerais	3
1.2.2 Objetivos Específicos	3
1.3 Estrutura da Monografia	3
<b>Capítulo 2 Fundamentação Teórica</b>	<b>2</b>
2.1 Planejamento Energético do Brasil	2
2.1.1 Modelos de operação	2
2.1.2 Modelos de curto prazo	3
2.1.3 Modelos de médio prazo	3
2.2 Redes Neurais Artificiais	3
2.3 <i>Multi-Layer Perceptron</i>	7
2.3.1 <i>BackPropagation</i>	8
2.3.1.1 Fase <i>Forward</i>	8
2.3.1.2 Fase <i>Backward</i>	9
2.3.2 Critério de Parada	10
<b>Capítulo 3 Metodologia</b>	<b>11</b>
3.1 Base de Dados	11
3.2 Pré-processamento dos dados	12
3.3 Predizendo os valores de vazão mensal com a MLP	13
3.4 Configurações das redes MLPs	14
3.4.1 Divisão dos dados	14
	ix

3.4.2	Arquitetura da rede	15
3.5	Simulações	19
3.6	Testes Estatísticos	20
3.6.1	Teste de Shiapiro-Wilk	20
3.6.2	Teste F	21
3.6.3	Teste t-Student	21
3.6.4	Teste da Soma dos Postos de Wilcoxon	22
<b>Capítulo 4</b>		<b>24</b>
<b>Resultados</b>		<b>24</b>
4.1	Base de Três Marias	25
4.1.1	Cenário 1: Últimos 12 meses	25
4.1.2	Cenário 2: Últimos 2 meses	27
4.1.3	Cenário 3: Meses com alta correlação	30
4.1.4	Comparação entre os três cenários	32
4.2	Base Tucuruí	33
4.2.1	Cenário 1: Últimos Doze Meses	34
4.2.1	Cenário 2: Últimos 2 meses	36
4.2.3	Cenário 3: Coeficientes de alta correlação	38
4.2.4	Comparação entre os 3 cenários	41
4.3	Comparativo dos resultados entre as bases	42
<b>5</b>	<b>Considerações Finais</b>	<b>45</b>
5.1	Conclusões	45
5.2	Trabalhos Futuros	45
<b>Referências</b>		<b>46</b>
<b>Apêndice A</b>	<b>Scripts utilizados para a execução da Análise Estatística no Software R</b>	<b>48</b>

# Índice de Figuras

<b>Figura 1. Exemplo de usina hidroelétrica.....</b>	<b>1</b>
<b>Figura 2. Exemplo de usina termoelétrica.....</b>	<b>2</b>
<b>Figura 3. Neurônio Biológico .....</b>	<b>4</b>
<b>Figura 4. Ilustração do modelo proposto por McCulloch e Pitts para representar o neurônio artificial. ....</b>	<b>5</b>
<b>Figura 5. Exemplo da arquitetura de uma Rede Neural Artificial do tipo MLP ..</b>	<b>8</b>
<b>Figura 6. Usina Hidroelétrica de Três Marias .....</b>	<b>11</b>
<b>Figura 7. Usina Hidroelétrica de Tucuruí.....</b>	<b>12</b>
<b>Figura 8. Resultados de todas as configurações para cada cenário da base de Três Marias. ....</b>	<b>43</b>
<b>Figura 9. Resultado de todas as configurações para cada cenário da base de Tucuruí. ....</b>	<b>44</b>

# Índice de Tabelas

1	Correlação linear entre a vazão e a mesma variável em tempos anteriores da base de Três Marias	16
2	Correlação linear entre a vazão e a mesma variável em tempos anteriores da base de Tucuruí	16
3	Média do EPMA para cada configuração do cenário 1 da base Três Marias, após 30 execuções da MLP	25
4	Resultados dos testes estatísticos do cenário 1 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2)	26
5	Resultados dos testes estatísticos do cenário 1 entre amostras de 4 bits de sazonalidade (1) e com 12 bits de sazonalidade (2)	26
6	Resultados dos testes estatísticos do cenário 1 entre amostras de 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2)	27
7	Média do EPMA para cada configuração do cenário 2 da base Três Marias, após 30 execuções da MLP	27
8	Resultados dos testes estatísticos do cenário 2 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade	28
9	Resultados dos testes estatísticos do cenário 2 entre amostras de 4 bits de sazonalidade (1) e com 12 bits de sazonalidade (2)	28
10	Resultados dos testes estatísticos do cenário 2 entre amostras de 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2)	29
11	Média do EPMA para cada configuração do cenário 3 da base Três Marias, após 30 execuções da MLP	30
12	Resultados dos testes estatísticos do cenário 3 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2)	30
13	Resultados dos testes estatísticos do cenário 3 entre amostras sem bits de sazonalidade (1) e com 12 bits de sazonalidade (2)	31
14	Resultados dos testes estatísticos do cenário 3 entre amostras de 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2)	31
15	Resultados dos testes estatísticos entre o melhor resultado do cenário 1 e o melhor do cenário 2	32

16 Resultados dos testes estatísticos entre o melhor resultado do cenário 2 e o melhor do cenário 3	33
17 Média do EPMA para cada configuração do cenário 1 da base Tucuruí, após 30 execuções da MLP	34
18 Resultados dos testes estatísticos do cenário 1 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2)	35
19 Resultados dos testes estatísticos do cenário 1 entre amostras com 4 bits de sazonalidade (1) e com 12 bits de sazonalidade (2)	35
20 Resultados dos testes estatísticos do cenário 1 entre amostras com 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2)	36
21 Média do EPMA para cada configuração do cenário 2 da base Tucuruí, após 30 execuções da MLP	36
22 Resultados dos testes estatísticos do cenário 2 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2)	37
23 Resultados dos testes estatísticos do cenário 2 entre amostras com 04 bits de sazonalidade (1) e com 12 bits de sazonalidade (2)	37
24 Resultados dos testes estatísticos do cenário 2 entre amostras com 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2)	38
25 Média do EPMA para cada configuração do cenário 3 da base Tucuruí, após 30 execuções da MLP	38
26 Resultados dos testes estatísticos do cenário 3 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2)	39
27 Resultados dos testes estatísticos do cenário 3 entre amostras com 4 bits de sazonalidade (1) e com 12 bits de sazonalidade (2)	39
28 Resultados dos testes estatísticos do cenário 3 entre amostras de 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2)	40
29 Resultados dos testes estatísticos do melhor do cenário 1 e o melhor do cenário 2	41
30 Resultados dos testes estatísticos do melhor do cenário 2 e o melhor do cenário 3	42

## Lista de Abreviaturas e Siglas

CEPEL	Centro de Energia e Pesquisas Elétricas
CI	Computação Inteligente
EPMA	Erro Percentual Médio Absoluto
MLP	<i>Multi-Layer Perceptron</i>
ONS	Operador Nacional do Sistema Elétrico
RNA	Redes Neurais Artificiais

# Capítulo 1

## Introdução

Este capítulo encontra-se dividido em três seções. Inicia descrevendo o problema e a motivação do desenvolvimento deste trabalho. Em seguida, são elencados seus principais objetivos. Ao final, é mostrada a estrutura dos próximos capítulos da monografia.

### 1.1 Motivação e Problema

Por ser um país que possui diversos recursos hídricos, o Brasil faz uso de boa parte deles para compor sua matriz energética. Uma característica desses recursos é que eles apresentam acentuada sazonalidade (NETO, 2005) e certas irregularidades em suas vazões. De tal forma, a fim de garantir o fornecimento elétrico sem interrupção, a produção de energia é complementada com outras fontes, como termoelétricas (COSTA; PIÑERO; DAMÁZIO, 2007).

De acordo com a Agência Nacional de Energia Elétrica (ANEEL), em 2014, a matriz energética do país é composta, em sua maioria, de usinas hidroelétricas (SOUZA, 2010) e usinas termoelétricas. As hidroelétricas representam 66,99% e as termoelétricas 28,58%.

**Figura 1. Exemplo de usina hidroelétrica**



[fonte: <http://www.brasilecola.com/geografia/energia-hidreletrica.htm>]

**Figura 2. Exemplo de usina termoeétrica**



[fonte: <http://www.brasilecola.com/geografia/energia-termoeletrica.htm>]

Devido ao fato da energia proveniente de hidroelétricas ser mais abundante e mais economicamente viável algumas estratégias são aplicadas. Uma destas é o uso de reservatórios com grande capacidade para armazenar água em períodos de abundância e utilizá-los durante os períodos de escassez.

A previsão de vazão dos recursos hídricos é uma das estratégias para minimizar o impacto da irregularidade dos níveis armazenados nos reservatórios. Prever valores de vazão portanto, é descobrir antecipadamente a quantidade de água que irá afluir nestes determinados pontos de medição com base nas séries históricas já conhecidas. (NETO; COELHO; CHIGANER; MEZA; MONTEIRO, 2005)

Anteriormente as previsões diárias, semanais ou mensais, eram realizadas através de modelos estatísticos. Entretanto, vários pesquisadores realizaram estudos na tentativa de modelar a relação entre chuva e vazão utilizando Redes Neurais Artificiais (RNAs) (HAYKIN, 2007). Nestes estudos foram encontrados bons resultados para várias usinas hidroelétricas, melhores que os obtidos através dos modelos estatísticos utilizados até então (SOUZA, 2010).

A rede neural pode apresentar diferentes resultados para a previsão dependendo da configuração adotada (VALENÇA, 2005). Assim, um estudo comparativo para verificar qual a configuração a ser utilizada resultará em uma rede neural que mostre um resultado eficiente, mais próximo da realidade e com um erro

minimizado, proporcionando um melhor planejamento energético e consequentemente melhor aproveitamento dos recursos.

## 1.2 Objetivos

### 1.2.1 Objetivos Gerais

- Realizar um estudo comparativo entre as diferentes configurações para uma Rede Neural Artificial e determinar quais destas melhor se aproximam do resultado esperado.

### 1.2.2 Objetivos Específicos

- Implementar a RNA na linguagem de programação JAVA.
- Realizar previsões mensais de vazão em determinado tempo “t” utilizando como entrada valores passados ‘t-1’, ‘t-2’, ..., ‘t-11’, ‘t-12’.
- Escolher, também, as melhores variáveis de entrada, através do coeficiente de correlação, para realizar o treinamento da rede.
- Investigar se resultados obtidos são coerentes de acordo com a realidade de cada usina hidroelétrica observada.

## 1.3 Estrutura da Monografia

O capítulo 2 apresenta a fundamentação teórica e reúne todo o conteúdo necessário para a compreensão do trabalho proposto, a saber: O planejamento energético brasileiro e as Redes Neurais Artificiais, com enfoque na arquitetura proposta para esta monografia, a *Multi Layer Perceptron* (MLP). Logo em seguida, o capítulo 3 demonstra toda a metodologia utilizada para atingir o objetivo geral do projeto, tais como o pré-processamento da base de dados e escolha das variáveis de entrada de cada configuração observada. Em seguida, o capítulo 4 mostra os resultados obtidos a partir das aplicações e simulações das técnicas desenvolvidas. Por fim, o capítulo 5 apresenta um resumo do trabalho final, expõe as dificuldades encontradas, discussões geradas, conclusões obtidas assim como melhorias propostas para trabalhos futuros.

# Capítulo 2

## Fundamentação Teórica

Este capítulo apresenta todo o conceito teórico necessário para tentar resolver o problema descrito no capítulo 1. A seção 2.1 mostra como é realizado o planejamento energético brasileiro e seus modelos. Logo após, na seção 2.2, são abordadas as RNAs, técnica de Computação Inteligente (CI) que vem mostrando uma eficiente solução para problemas de previsão e classificação e por fim, a seção 2.3 apresenta uma arquitetura específica de RNA, a MLP com seus conceitos teóricos e aplicações.

### 2.1 Planejamento Energético do Brasil

No Brasil, mais de 30 hidroelétricas compõem a matriz de energia elétrica brasileira, que são apenas complementadas pelas usinas termoeletricas. Logo, com a maximização das usinas hidroelétricas instaladas no país, deve-se obter uma maior e mais eficiente produção de energia. Este é o objetivo do planejamento energético do país. Aumentar a produção de energia através de fontes limpas e renováveis, como as hidroelétricas e minimizando assim, o uso de energia proveniente de termoeletricas.

O Centro de Energia e Pesquisas Elétricas (CEPEL) desenvolveu três modelos do planejamento do sistema elétrico: modelos de médio prazo, modelos de curto prazo e modelos de operação.

#### 2.1.1 Modelos de operação

No modelo de Operação, as previsões são semanais, podendo chegar até a duas semanas. Atualmente, o modelo de previsões utilizadas é o Previsão de Vazões Horárias (PREVIVAZH). É um modelo considerado determinístico e utiliza tantos fatores internos quanto fatores externos para realizar as previsões. Exemplos de fatores internos são vazões de dias anteriores e exemplos de fatores externos são

precipitação de chuva ou umidade do ar, entre outros (COSTA; DAMÁZIO; CARVALHO, 2000) .

### **2.1.2 Modelos de curto prazo**

No modelo de previsão de curto prazo, a previsão é realizada para semanas a frente. Um dos modelos mensais utilizados é o modelo estocástico PREVIVAZM do Operador Nacional do Sistema Elétrico (ONS). Existe também o modelo Geração de Energia e Vazões Sintéticas, ou GEVAZP, que realiza previsões mensais para os dois primeiros meses mas pode chegar até 12 meses. Todos os modelos de curto prazo consideram as vazões como determinísticas (COSTA, 2003).

### **2.1.3 Modelos de médio prazo**

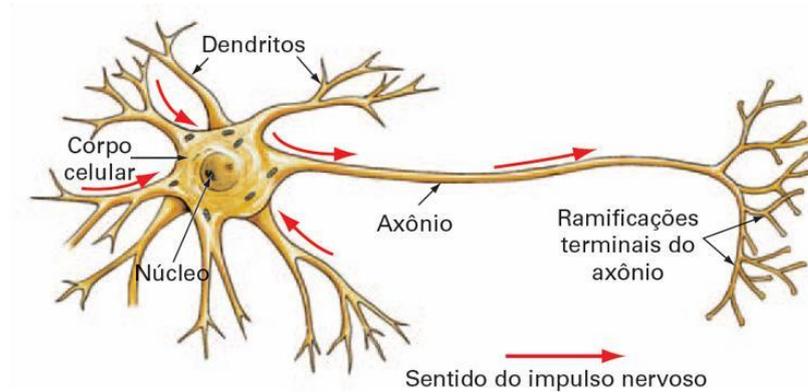
Em modelos de médio prazo, o horizonte de planejamento é maior. Normalmente se utiliza 12 meses à frente, porém este horizonte pode chegar até a 8 anos. A partir deste modelo é definida a quantidade de energia a ser utilizada pelo país. As vazões são consideradas estocásticas e o modelo mais comum é o GEVAZP (MACIEIRA; FALCÃO; PENNA, 2001).

## **2.2 Redes Neurais Artificiais**

O cérebro humano, composto aproximadamente por 10 bilhões de neurônios, possui uma capacidade de processamento maior que qualquer computador na atualidade. O neurônio é a célula do sistema nervoso. Cada neurônio possui um corpo celular e a partir disso se observam ramificações, os dendritos, e um prolongamento, o axônio. O axônio possui terminais chamados de telodendro.

A troca de informação entre dois neurônios ocorre na região de sinapse composta pelo axônio do neurônio que deseja passar a informação e o dendrito do que irá recebe-la. Diante de um mecanismo simples e de alta capacidade de processamento, foi proposto um modelo matemático do neurônio artificial que simula o funcionamento de um neurônio biológico e seu comportamento.

**Figura 3. Neurônio Biológico**



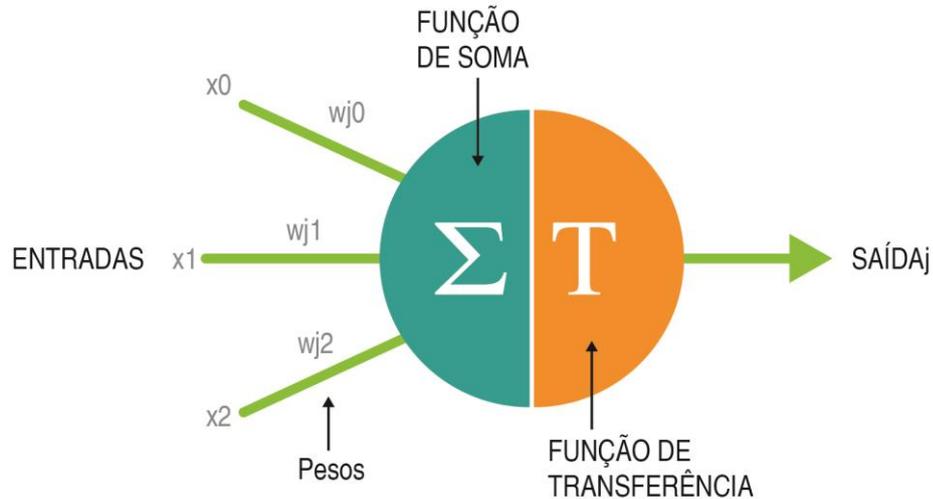
[fonte: <http://janelambulante.blogspot.com.br/2012/05/licao-do-neuronio.html>]

Os neurônios de uma Rede Neural Artificial (RNA) constituem elementos de processamento interconectados e se encontram organizados em camadas. (Valença, 2010).

O funcionamento de um neurônio é regido pela lei do Tudo ou Nada. Esta lei afirma que a informação será passada adiante se a intensidade do estímulo atingir um limiar excitatório mínimo. De tal modo, não importa o quão intenso é o estímulo, o potencial de ação do neurônio será sempre o mesmo. (VALENÇA, 2010).

O primeiro modelo matemático para representação de um neurônio foi proposto por McCulloch e Pitts (MCCULLOCH, PITTS. 1943) e encontra-se descrito na Figura 4.

**Figura 4. Ilustração do modelo proposto por McCulloch e Pitts para representar o neurônio artificial.**



fonte: elaboração própria]

O modelo mostrado é bastante simples. É formado por um conjunto de entradas, uma unidade de processamento e uma, ou mais, unidades de saída correspondendo, respectivamente, aos dendritos, corpo celular e telodendro de um neurônio biológico. A rede possui uma regra de propagação e uma função de ativação.

Cada entrada  $x_i$  é multiplicada por um peso  $w_j$  e a soma destes produtos é a entrada do neurônio. A entrada líquida de um neurônio  $i$  e com um limiar  $\theta$ , é calculada de acordo com a Equação 2.1.

$$net_i = \sum_{j=1}^n w_{ij}x_j - \theta \quad (2.1)$$

Entre os primeiros modelos propostos, surgiram o *Perceptron* e o *Adaline*. O primeiro foi proposto Frank Roseblatt (1957). Como o *Perceptron* lida apenas com saídas discretas, Bernard Widrow e Ted Hoff propuseram o *Adaline* que permite trabalhar com saídas em um universo contínuo. Basicamente a diferença entre os dois se encontra na função de ativação. A saída do *Perceptron* é definida pela função

degrau. Se o neurônio for ativado, emite o sinal 1, caso contrário emitirá o sinal 0. A Equação 2.2 demonstra a função de saída do *Perceptron*.

$$y_i = f(net) \begin{cases} 1, & \text{para } net_i \geq 0 \\ 0, & \text{para } net_i < 0 \end{cases} \quad 2.2$$

Para a saída do *Adaline*, pode ser utilizada uma função contínua. Segundo Valença existem três que são mais utilizadas: A linear, a sigmoideal logística e a tangente hiperbólica. As Equações 2.3, 2.4 e 2.5 representam, respectivamente, a função linear, a sigmoide logística e a tangente hiperbólica.

$$y = f(net_i) = net_i \quad (2.3)$$

$$y = f(net_i) = \frac{1}{1 + e^{-net_i}} \quad (2.4)$$

$$y = f(net_i) = \frac{e^{net_i} - e^{-net_i}}{e^{net_i} + e^{-net_i}} \quad (2.5)$$

A rede neural artificial surge quando neurônios artificiais são unidos, em duas ou mais camadas. Semelhantemente ao cérebro humano as RNAs possuem uma capacidade de aprendizado. Isto ocorre através de regras de aprendizagem que modificam os pesos sinápticos de acordo com as entradas apresentadas. Existem basicamente dois tipos de aprendizado: O Aprendizado supervisionado e o aprendizado não supervisionado. No primeiro exemplo, são fornecidos dados de entrada para a rede, e já se conhece o resultado desejado. Assim, a rede vai se ajustando para conseguir chegar o mais próximo possível do resultado esperado. No aprendizado não supervisionado, a rede aprende com suas próprias entradas. (VALENÇA, 2010). São apresentadas somente as entradas e não se faz necessário conhecer as saídas da rede referência.

O *Perceptron* e o *Adaline* possuem a limitação de só poderem resolver problemas linearmente separáveis. Para tentar buscar uma alternativa que resolva este problema, baseado no aprendizado supervisionado, surgiu o *Perceptron* de Múltiplas Camadas (do inglês, *Multi Layer Perceptron*, MLP).

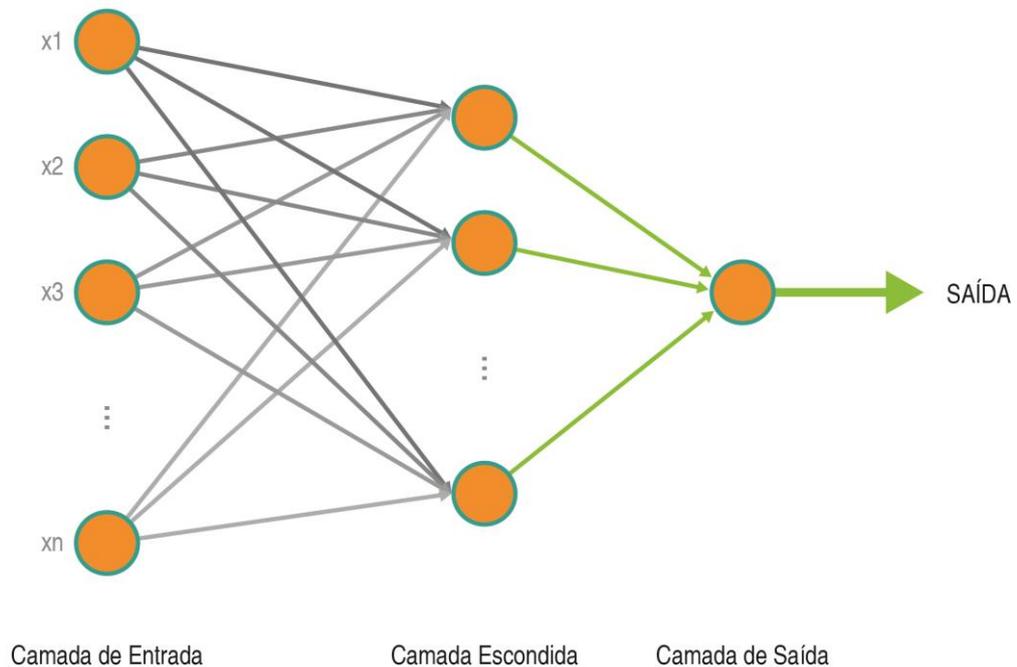
## 2.3 Multi-Layer Perceptron

Redes MLP podem ser consideradas como uma generalização do *Perceptron* e do *Adeline*. Possuem uma camada de entrada que contém as variáveis de entrada da rede e uma camada de saída representando a resposta da rede e o resultado esperado.

A novidade é que neste novo modelo é proposta a inserção de pelo menos uma camada a mais. Esta é denominada camada intermediária ou camada escondida. Como o próprio nome sugere, esta camada é posicionada entre as camadas de entrada e saída. Até então, os modelos apresentados não conseguiam resolver a maioria dos problemas do mundo real, que são não linearmente separáveis. Com mais de uma camada intermediária, é possível fazer uma aproximação de qualquer função matemática. (HAYKIN, 2007).

Na maioria das redes MLP, a função de ativação é do tipo sigmoidal logística ou tangente hiperbólica. A Figura 5 demonstra o modelo de uma rede MLP de três camadas e com apenas uma saída. A MLP possui um aprendizado supervisionado. Para tal a mesma precisa calcular o erro na saída, comparando o resultado obtido com o resultado esperado. A partir deste ponto, deve corrigir os pesos do neurônio com o intuito de minimizar o erro.

Figura 5. Exemplo da arquitetura de uma Rede Neural Artificial do tipo MLP



[fonte: elaboração própria]

Alguns algoritmos de aprendizado supervisionado já existem. Um dos mais antigos e utilizados em redes do tipo MLP é o algoritmo de Propagação recursiva (do inglês, *BackPropagation*).

### 2.3.1 *BackPropagation*

Foi proposto na década de 70 por Paul Werbos. (WERBOS J., 1974). O algoritmo consiste na retropropagação do erro e busca no seu processo de aprendizagem a minimização de uma função objetivo, que no geral é o erro médio quadrático (EMQ). Esse algoritmo é dividido em duas fases e ambas serão descritas nas seções 2.3.1.1 e 2.3.1.2.

#### 2.3.1.1 Fase *Forward*

Nesta fase ocorre a propagação progressiva do sinal. Ela ocorre da camada de entrada para a camada de saída. Ao final desta fase, os pesos permanecem inalterados e o erro é calculado observando o valor obtido e o valor esperado.

### 2.3.1.2 Fase *Backward*

Nesta fase ocorre a retropropagação do erro da camada de saída para a camada de entrada. Neste momento os pesos são alterados de forma recursiva. O cálculo do ajuste dos pesos obedece a equação a seguir:

$$W_{ij}^m(t+1) = W_{ij}^m(t) + \alpha \delta_i^m f^{m-1}(net_j^{m-1}) + \beta \Delta W_{ij}^m(t-1) \quad (2.6)$$

Onde,

- $\beta$  é a taxa de momento
- $\delta_i^m$  representa a sensibilidade do neurônio
- $\alpha$  é a taxa de aprendizado
- $W_{ij}^m(t)$  se refere ao i-ésimo peso do neurônio j.

O cálculo da sensibilidade pode variar de acordo com a camada que o neurônio se encontra. Na camada de saída, a sensibilidade é obtida por meio da Equação 2.7.

$$\delta_i^m = (d_i - y_i) f'(net_i) \quad (2.7)$$

Onde,

- $d_i$  é a saída esperada
- $y_i$  é a saída calculada após o treinamento
- $f'(net_i)$  é a derivada da função de ativação dos neurônios presentes na camada de saída.

Para as demais camadas, a sensibilidade é calculada de acordo com a Equação 2.8.

$$\delta_i^{m-1} = f'^{m-1}(net_j^{m-1}) \sum_{i=1}^N w_{ij}^m \delta_i^m \quad (2.8)$$

Onde,

- N é o número de neurônios na camada que recebe o sinal.
- $f'^{m-1}(net_j^{m-1})$  é a derivada da função de ativação do neurônio da camada m-1.
- $w_{ij}^m$  é o peso i referente ao neurônio j.

- $\delta_i^m$  é a sensibilidade transmitida pelo i-ésimo do termo imediatamente a frente.

### 2.3.2 Critério de Parada

Durante o treinamento da rede, vários padrões são mostrados à rede e então é possível definir a quantidade de simulações utilizadas. Determinar este número exige cuidado, pois ele pode apresentar dois tipos de problemas.

Quando a rede é treinada por mais tempo que o necessário, ela começa a decorar padrões e perde sua capacidade de generalização. Este acontecimento é denominado *overfitting*. O segundo problema é quando a rede treina por menos tempo que o necessário e então ela não consegue generalizar o suficiente. Para determinar o momento ideal de parar o treinamento, um dos critérios mais utilizados é o de validação cruzada.

Os dados são divididos em três partes: treinamento, validação e teste. Geralmente, 50% dos dados são separados para treinar a rede, 25% para a validação cruzadas e os 25% restantes para o processo de teste. Após cada ciclo, ou seja a apresentação de todos os exemplos de treinamento, os exemplos do conjunto de validação são apresentados à rede e o erro dessa etapa é calculado. Como estes são inéditos, nunca são usados para ajuste dos pesos. Desta maneira, enquanto o erro da validação estiver diminuindo, é sinal que a rede continua generalizando.

A partir do momento em que o erro de validação cruzada começar a aumentar e o erro do treinamento começar a diminuir, é sinal que a rede está começando a decorar padrões. Este é um bom momento para encerrar o treinamento enquanto a rede ainda consegue fazer generalizações (VALENÇA, 2010). Após este momento a rede é submetida ao conjunto de testes onde seu desempenho e precisão serão avaliados.

# Capítulo 3

## Metodologia

Este capítulo apresenta a metodologia utilizada para resolver o problema proposto neste trabalho.

### 3.1 Base de Dados

A base de dados utilizada neste projeto utiliza dados reais pertencentes a duas usinas hidroelétricas. Uma delas é a usina de Três Marias, localizada no rio São Francisco, em Minas Gerais. Esta foi inaugurada em 1962. Sua barragem possui uma área alagada de 1040 km<sup>2</sup>. A potência instalada na usina é de 396 MW. Os dados das vazões mensais da usina de Três Marias foram coletados em um período de 01 de janeiro de 1931 a 01 de dezembro de 2003.

**Figura 6. Usina Hidroelétrica de Três Marias**



[fonte: <http://blogsalaouazeiro.com.br/usina-de-tres-marias-no-rio-sao-francisco-deve-parar-de-gerar-energia/>]

A outra é a usina Tucuruí, localizada no rio Tocantins, no estado do Pará. Esta foi inaugurada em 1984. A barragem possui uma área alagada de 2850 km<sup>2</sup>. A capacidade desta usina é de 8.370 MW. Os dados das vazões mensais da usina de Tucuruí foram coletados em um período de 01 de janeiro de 1931 a 01 de dezembro de 2003.

**Figura 7. Usina Hidroelétrica de Tucuruí**



[fonte: <http://jornaldetucuruui.blogspot.com.br/2013/01/sonho-da-3-fase-da-usina-hidreletrica.html>]

## 3.2 Pré-processamento dos dados

Para dar início ao pré-processamento, deve-se normalizar os dados. O processo de normalização faz com que todas as variáveis recebam a mesma atenção durante o treinamento. Ao definir um intervalo numérico, variáveis com valores altos recebem a mesma importância de variáveis de valores mais baixos.

O valor de cada variável deve ser proporcional ao limite das funções de ativação utilizada na camada de saída. De acordo com Valença (2010), “caso utilizemos na saída a função logística como função de transferência, seus valores estão limitados entre [0 e 1], logo os dados são geralmente normalizados entre os intervalos [0,10 e 0,90] ou [0,15 e 0,85]” (VALENÇA, 2010).

Embora a RNA possua apenas uma variável, é importante realizar a normalização pois com o passar do tempo a mesma pode ser modificada e outras variáveis incluídas. Além disso, a normalização poderá evitar a que a rede sofra saturação. (VALENÇA, 2010).

O tipo de normalização escolhida foi a transformação linear, cuja fórmula está descrita na Equação 3.1.

$$y = \frac{(b-a)(x_i-x_{min})}{(x_{max}-x_{min})} + a \quad (3.1)$$

onde,

- $y$  é o valor normalizado.
- $x_i$  é o valor original.
- $x_{min}$  é o valor mínimo de  $x$ .
- $x_{max}$  é o valor máximo de  $x$ .
- $a$  e  $b$  são limites escolhidos.

Para este trabalho, foi definido  $a = 0,10$  e  $b = 0,90$ .

### 3.3 Predizendo os valores de vazão mensal com a MLP

Embora sejam frequentemente usadas em pesquisas, a rede MLP necessita que cada um de seus parâmetros sejam configurados. Dependendo de cada tipo de configuração, o resultado final irá variar. Abaixo, seguem os parâmetros para a rede MLP e o algoritmo *Backpropagation*:

- Número de neurônios da camada de entrada
- Número de neurônios da camada escondida
- Número de neurônios da camada de saída
- Função de ativação
- Critério de parada
- Taxa de aprendizado
- Momento

- Número máximo de ciclos

O número de entradas depende da configuração da rede escolhida. A saída contém um valor, referente à previsão de 1 mês à frente. O algoritmo para treinamento é o *Backpropagation*, descrito na subseção 2.3.1 do capítulo anterior e a função escolhida para ativação é a sigmoide logística, descrita na Equação 3.2.

$$y = \frac{1}{1 + e^{-net_i}} \quad (3.2)$$

onde,

- $y$  representa o valor da saída.
- $net_i$  representa a soma dos produtos de cada entrada por seus respectivos pesos no  $i$ -ésimo neurônio.

Os valores de  $y$  estão limitados dentro do intervalo  $[0; 1]$ .

Para a camada de saída, foi escolhida a função linear como função de ativação, descrita na Equação 3.3.

$$y = f(net_i) = net_i \quad (3.3)$$

onde  $net_i$  é a entrada líquida do neurônio da camada de saída.

## 3.4 Configurações das redes MLPs

Os experimentos com a MLP foram realizados utilizando diferentes valores para os parâmetros da rede neural. Todas as configurações adotadas estão descritas nas próximas seções.

### 3.4.1 Divisão dos dados

O conjunto de dados foi dividido em três grupos. O primeiro representado por 50% dos dados foi separado para o treinamento da rede. O segundo e terceiro grupo ficaram com 25% dos dados, cada. O segundo grupo foi destinado para realizar validação cruzada, isto é, informar quando a rede deve parar de treinar, a fim de não perder sua capacidade de generalização. O último grupo foi reservado para testes e avaliação do desempenho da rede.

### 3.4.2 Arquitetura da rede

Para a camada de entrada, o número de neurônios variou para cada configuração. Cada neurônio foi representado por um valor de vazão passado. Foram adotadas basicamente três possíveis cenários. O primeiro utiliza os doze últimos meses para realizar a previsão de um mês a frente. O segundo utiliza apenas os valores de vazão dos dois últimos meses para realizar a previsão. E o terceiro seleciona os meses que apresentaram uma forte dependência com a variável que se deseja prever.

Para verificar o grau de dependência das variáveis, foi utilizada uma métrica chamada correlação linear. A correlação entre duas variáveis indica a força e direção do relacionamento linear das mesmas. A correlação entre duas variáveis  $x$  e  $y$  podem ser medidas de acordo com a Equação 3.4.

$$r = \left( \frac{\sum \frac{(x - \bar{x})(y - \bar{y})}{S_x S_y}}{n - 1} \right) \quad (3.4)$$

onde,

- $r$  é a correlação linear.
- $x$  é o valor da variável  $x$ .
- $y$  é o valor da variável  $y$ .
- $\bar{x}$  é o valor médio de  $x$ .
- $\bar{y}$  é o valor médio de  $y$ .
- $S_x$  é a soma dos valores de  $x$ .
- $S_y$  é a soma dos valores de  $y$ .
- $n$  é a quantidade de valores.

A correlação linear sempre ocorrerá no intervalo de  $[-1;1]$ . Quanto mais próximo a 1, maior a relação entre as duas variáveis. A Tabela 1 mostra as correlações lineares da vazão em tempo  $t$  e a mesma vazão nos tempos 't-1', 't-2', 't-3', ..., 't-11', 't-12'. Os dados são da usina de Três Marias.

**Tabela 1: Correlação linear entre a vazão e a mesma variável em tempos anteriores da base de Três Marias.**

	t-12	t-11	t-10	t-9	t-8	t-7	t-6	t-5	t-4	t-3	t-2	t-1	t
t-12	1												
t-11	0,691103	1											
t-10	0,369806	0,692103	1										
t-9	0,042185	0,369744	0,690661	1									
t-8	-0,20068	0,041114	0,368315	0,690355	1								
t-7	-0,35282	-0,20147	0,040617	0,368796	0,690677	1							
t-6	-0,40777	-0,35343	-0,20168	0,041225	0,369464	0,690879	1						
t-5	-0,36565	-0,40833	-0,35362	-0,20122	0,042026	0,369856	0,691025	1					
t-4	-0,22718	-0,36636	-0,40842	-0,35323	-0,20031	0,042609	0,370163	0,691155	1				
t-3	0,003269	-0,2282	-0,36663	-0,40835	-0,3524	-0,19959	0,043125	0,370475	0,691319	1			
t-2	0,276153	0,001689	-0,22831	-0,36651	-0,40746	-0,3515	-0,19885	0,043683	0,370861	0,691534	1		
t-1	0,515095	0,274421	0,001821	-0,22822	-0,36589	-0,40673	-0,35082	-0,19826	0,044176	0,371231	0,691731	1	
t	0,585101	0,514572	0,275475	0,002241	-0,22816	-0,36586	-0,40676	-0,35091	-0,19841	0,044	0,370921	0,69143	1

[fonte: elaboração própria]

As variáveis que apresentaram forte correlação, acima de 0,5, foram ‘t-12’, ‘t-11’ e ‘t-1’.

**Tabela 2: Correlação linear entre a vazão e a mesma variável em tempos anteriores da base de Tucuruí.**

	t-12	t-11	t-10	t-9	t-8	t-7	t-6	t-5	t-4	t-3	t-2	t-1	t
t-12	1												
t-11	0,80979	1											
t-10	0,44142	0,8095	1										
t-9	0,02992	0,44064	0,80964	1									
t-8	-0,33151	0,02782	0,4396	0,80862	1								
t-7	-0,56947	-0,33201	0,02743	0,43989	0,80991	1							
t-6	-0,65176	-0,56929	-0,33219	0,02744	0,44078	0,8098	1						
t-5	-0,57523	-0,65127	-0,56932	-0,33261	0,02769	0,44071	0,80987	1					
t-4	-0,34777	-0,5746	-0,65117	-0,56993	-0,33284	0,02775	0,44095	0,80997	1				
t-3	-0,00813	-0,34705	-0,57439	-0,65178	-0,57053	-0,33271	0,02814	0,44121	0,81005	1			
t-2	0,36542	-0,00738	-0,34676	-0,57482	-0,65258	-0,57037	-0,3322	0,02858	0,44144	0,81012	1		
t-1	0,67121	0,36589	-0,00728	-0,34733	-0,57611	-0,65249	-0,56977	-0,33155	0,02905	0,44172	0,81025	1	
t	0,79448	0,67154	0,36598	-0,00738	-0,34841	-0,57612	-0,65197	-0,56911	-0,33099	0,02948	0,44201	0,81033	1

[fonte: elaboração própria]

Para cada entrada foram testados 3 tipos de codificação para representar a sazonalidade da vazão. Essas codificações dizem respeito à quantidade de neurônios que foi acrescentada na camada de entrada. Essas quantidades estão descritas a seguir.

- 4 neurônios a mais: Essa quantidade foi escolhida devido ao fato de se poder representar o número 12, quantidade de meses no ano, com apenas 4 bits. Neste caso, poderia adotar 0000 para representar o mês

de janeiro, 0001 para representar o mês de fevereiro e assim sucessivamente. Ressalva-se que cada mês deve ter uma codificação diferente, ou seja, uma mesma codificação não pode ser utilizada para dois meses distintos.

- 12 neurônios a mais: Optou-se por testar essa codificação para verificar se utilizar a representação com 12 bits indicaria alguma melhora no desempenho da rede neural. Neste caso, o mês de Janeiro poderia ser codificado com 100000000000, o mês de Fevereiro com 010000000000 e assim sucessivamente. Da mesma forma que o item anterior, uma mesma codificação não pode ser utilizada para meses distintos.
- 2 neurônios a mais: Esses dois neurônios são representados pelo seno e cosseno do mês que esteja sendo codificado. Isso está representado pelas Equações 3.5 e 3.6.

$$x_1 = \text{sen}\left(2\pi \frac{N}{180}\right) \quad (3.5)$$

$$x_2 = \text{cos}\left(2\pi \frac{N}{180}\right) \quad (3.6)$$

Onde,

N: representa o número de 1 a 12 do mês que esteja sendo codificado.

Para cada usina hidroelétrica foram considerados 3 cenários de previsão:

- Previsão de 1 mês a frente levando em consideração 12 valores passados - Cenário 1;
- Previsão de 1 mês a frente levando em consideração 2 valores passados - Cenário 2;
- Previsão de 1 mês a frente levando em consideração os 3 meses passados que tenham um coeficiente de correlação maior que 0,5 - Cenário 3.

Para o Cenário 1 foram utilizadas as seguintes configurações para a quantidade de neurônios na camada de entrada:

- Doze neurônios na entrada, representando os doze valores passados para previsão de um mês à frente.
- Dezesesseis neurônios na entrada, representados pelos meses passados e quatro bits para representar a sazonalidade.
- Vinte e quatro neurônios na entrada, representados pelos meses passados da vazão e doze bits para representar a sazonalidade.
- Quatorze neurônios na entrada, representados pelos doze meses passados e dois bits para representar a sazonalidade.

Para o Cenário 2 foram utilizadas as seguintes configurações para a quantidade de neurônios na camada de entrada:

- Dois neurônios na entrada, representando os dois meses passados para previsão de um à frente.
- Seis neurônios na entrada, representados pelos dois meses passados e quatro bits para representar a sazonalidade.
- Quatorze neurônios na entrada, representados pelos dois meses passados e doze bits para representar a sazonalidade.
- Quatro neurônios na entrada, representados pelos dois meses passados e dois bits para representar a sazonalidade.

Para o Cenário 3 foram utilizadas as seguintes configurações para a quantidade de neurônios na camada de entrada:

- Três neurônios na entrada, representando os três valores que apresentaram alta correlação, para previsão de um mês à frente.
- Sete neurônios na entrada, representados por três valores da vazão e quatro bits para representar a sazonalidade.
- Quinze neurônios na entrada, representados por três valores da vazão e doze bits para representar a sazonalidade.
- Cinco neurônios na entrada, representados por três valores de vazão e dois bits para representar a sazonalidade.

Esses cenários foram criados para ambas as usinas hidroelétricas.

### 3.5 Simulações

Para cada configuração da rede neural, foram realizadas 30 simulações. Esse número é considerado satisfatório para que os resultados se aproximem de uma distribuição normal. Posteriormente, os resultados obtidos foram comparados estatisticamente a fim de concluir qual configuração apresentava um melhor resultado.

A rede MLP foi implementada na linguagem de programação JAVA e no ambiente de desenvolvimento NETBEANS. Para todas as execuções da MLP, foram utilizados os parâmetros a seguir:

- Número de ciclos: 600
- Algoritmo de treinamento: *BackPropagation*
- Taxa de aprendizado: 0,8
- Taxa de Momento: 0,2
- Número de neurônios da camada escondida: 6
- Função de ativação da camada escondida: Sigmóide logística
- Função de ativação da camada de saída: função linear

Após cada execução, com o resultado da vazão previsto, foi calculado o Erro Percentual Médio Absoluto (EPMA), em relação à vazão desejada. A equação 3.19 mostra como é realizado o cálculo do EPMA:

$$EPMA = \frac{\sum_{i=1}^n \left| \frac{d_i - y_i}{d_i} \right| 100}{n} \quad (3.7)$$

onde,

- $d_i$  é a saída desejada para a i-ésima previsão
- $y_i$  é a saída prevista para a i-ésima previsão
- $n$  é a quantidade de previsões realizadas

Após o cálculo de todos os EPMAs, calculou-se a média entre eles e a partir desse indicativo é analisado o desempenho de cada configuração da rede.

## 3.6 Testes Estatísticos

Após as simulações de cada uma das configurações terem sido finalizadas, foram realizados testes estatísticos. Existem diversos tipos de testes estatísticos entre eles o t-Student e o Teste da Soma dos Postos de Wilcoxon são os mais comuns. Para utilizar o t-Student, deve-se utilizar antes dois testes, visto que trata-se de um teste paramétrico.

O primeiro deles é o teste de Shapiro-Wilk para verificar se as amostras estão normalmente distribuídas e em caso positivo, se realiza o teste F para verificar se as variâncias são estatisticamente iguais. Caso ambos os testes sejam satisfatórios, o teste t-Student pode ser aplicado. Caso contrário, aplica-se o teste de Wilcoxon.

### 3.6.1 Teste de Shapiro-Wilk

Para realizar o teste de Shapiro-Wilk, deve-se realizar os passos descritos abaixo:

1. Formulação da hipótese:

$H_0$ : a amostra é normalmente distribuída

$H_1$ : amostra não é normalmente distribuída

2. É estabelecido o  $\alpha$ , nível de significância do teste. Para este trabalho foi definido que  $\alpha = 0,05$
3. Calcula a estatística do teste:
  - a. Ordenar as  $n$  observações da amostra:  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ ,
  - b. Calcular o valor de  $b$ :

$$b = \begin{cases} \sum_{i=1}^{n/2} a_{n-i+1}(x_{n-i+1} - x_1), & \text{se } n \text{ é par} \\ \frac{(n+1)}{2} \sum_{i=1}^{n/2} a_{n-i+1}(x_{n-i+1} - x_i), & \text{se } n \text{ é ímpar} \end{cases} \quad (3.8)$$

- c. Calcular  $W$ :

$$W = \frac{b^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2} \quad (3.9)$$

Quanto menor o  $W$ , maior a certeza de que os dados são normalmente distribuídos.

4. Tomada de decisão: se o  $W_{calculado} < W_{\alpha}$ ,  $H_0$  é rejeitado. (ao nível de significância  $\alpha$ ).

### 3.6.2 Teste F

Tem como objetivo verificar se a variância de duas amostras são estatisticamente iguais. Para isso os seguintes passos devem ser executados:

1. Determinar a maior e a menor variância entre as amostras.
2. Calcula-se a razão entre elas. No numerador é colocada a maior ( $S_1^2$ ) entre as duas e no denominador a menor ( $S_2^2$ ). O valor calculado é atribuído a uma variável  $F$  segundo a Equação 3.22.

$$F = \frac{S_1^2}{S_2^2} \quad (3.10)$$

O valor de  $F_{calculado}$  é comparado a um valor de  $F_{critico}$  obtido através de uma tabela de distribuição de  $F$ , considerando o grau de liberdade tanto do numerador quanto do denominador e o nível de significância adotado.

Se  $F_{calculado} < F_{critico}$ , não se pode rejeitar a hipótese de que as variâncias são as mesmas. Neste caso a probabilidade de  $F$  é maior que o nível de significância adotado. Caso  $F_{calculado} \geq F_{critico}$ , é aceita a hipótese de que as variâncias são diferentes e não pertencem à mesma população.

### 3.6.3 Teste t-Student

O teste t-Student é utilizado para comparação de médias de amostras independentes com distribuição normal. Antes de aplicá-lo, deve-se anteriormente ter garantido de que as variâncias são estatisticamente iguais. Primeiramente, calcula-se o valor do erro padrão das duas amostras de acordo com a equação 3.23:

$$EP_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}} \quad (3.11)$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (3.12)$$

Após este passo, calcula-se o valor de t de acordo com a Equação 3.25:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{EP_{(x_1 - x_2)}} \quad (3.13)$$

O valor de t calculado deve ser comparado com o valor de  $t_{critico}$  obtido na tabela de distribuição de t-Student, considerando  $(n_1 + n_2 - 2)$  como graus de liberdade. Se o valor de t calculado for maior que o valor de  $t_{critico}$  encontrado na tabela, significa que a probabilidade de t é maior que o nível de significância adotado. Quando t é menor, esta hipótese é rejeitada.

#### 3.6.4 Teste da Soma dos Postos de Wilcoxon

Este teste utiliza as posições ocupadas por valores dos dados de duas populações independentes, quando colocados em ordem crescente. É verificado a hipótese de que duas amostras independentes provêm de populações com medianas iguais. A segunda hipótese é se tais amostras são de populações com medianas diferentes.

O teste é do tipo não paramétrico e nestes casos supõe-se uma não normalidade na distribuição dos dados experimentais, ou não existe informação necessária para afirmar a normalidade da distribuição. Os dados precisam ser ordenáveis.

As duas amostras independentes devem ser de dados selecionados aleatoriamente. As amostras não precisam ser normalmente distribuídas. Se ambos os tamanhos forem maiores que 10, a distribuição amostral (R) será, aproximadamente, normal com média  $\mu$  e desvio padrão  $\alpha$ .

A partir dos dados, aplicam-se seus valores nas Equações 3.26, 3.27 e 3.28.

$$Z = \frac{R - \mu_R}{\sigma_R} \quad (3.14)$$

$$\mu_R = \frac{n_1(n_1+n_2+1)}{2} \quad (3.15)$$

$$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}} \quad (3.16)$$

Onde  $n_1$  é o tamanho da amostra 1,  $n_2$  é o tamanho da amostra 2,  $R$  é a soma das posições da amostra de tamanho  $n_1$ ,  $\sigma_R$  é o desvio padrão dos valores amostrais de  $R$  e  $\mu_R$  é a média dos valores amostrais de  $R$ .

# Capítulo 4

## Resultados

Este capítulo tem como objetivo apresentar todos os resultados encontrados a partir da execução da metodologia apresentada no capítulo anterior. Previamente, foram feitas simulações para definir os valores de algumas variáveis necessárias para o treinamento da rede MLP. Estes valores foram definidos na seção 3.5.

Primeiramente, na seção 4.1, foram realizadas simulações com a base de Três Marias, aplicando todas as técnicas de sazonalidade e comprovando os resultados através dos testes estatísticos. Em seguida, na seção 4.2, foi feito o mesmo para a base de Tucuruí. A seção 4.3 comenta os dois resultados e faz um comparativo com o intuito de observar se existe alguma similaridade entre eles.

Após a realização de todos os treinamentos de cada base, foram feitos testes estatísticos. Para isso, foi utilizado o software R, com um nível de significância igual a 0,05. Para cada um dos quatro conjuntos de dados, foram realizados os testes de Shapiro-Wilk, com intuito de verificar a normalidade das amostras. Na hipótese nula, a amostra analisada é normalmente distribuída e na hipótese alternativa, não é normalmente distribuída. Após isto, caso as duas amostras sejam normalmente distribuídas, é realizado o Teste F para verificar a igualdade entre as variâncias das amostras. Caso o resultado do Teste F encontre-se de acordo com o nível de significância, é realizado, por fim, o teste t-Student. Ele serve para definir se as amostras normais apresentam desempenho estatisticamente iguais. Caso as amostras não sejam normalmente distribuídas ou o nível de significância não for atingido no teste F, é realizado o teste não paramétrico Soma dos Postos de Wilcoxon com intuito de verificar se as amostras apresentam desempenho estatisticamente iguais ou não.

---

## 4.1 Base de Três Marias

Como explicado anteriormente, todas as simulações de cada base foram divididas em três cenários. Cada cenário foram realizadas 30 execuções por configuração e em cada uma foi calculado o erro percentual médio absoluto. Após isto, obteve-se a média dos EPMA's encontrados. Cada cenário se encontra explicitado abaixo.

### 4.1.1 Cenário 1: Últimos 12 meses

No primeiro cenário, foram levados em conta os últimos doze meses para realizar a previsão de um mês à frente. Dentro deste cenário, foram consideradas quatro configurações, como descrito na seção 3.2.2. A Tabela 3 exibe a média do erro percentual médio absoluto de cada uma das quatro configurações.

**Tabela 3: Média do EPMA para cada configuração do cenário 1 da base Três Marias, após 30 execuções da MLP.**

Configuração	EPMA
Últimos 12 Meses	41,31 %
4 bits de Sazonalidade	39,99 %
12 bits de Sazonaliadade	38,01 %
2 bits de Sazonaldiade	40,43 %

[fonte: elaboração própria]

Em primeiro lugar, foram realizados testes com os dados da configuração sem bits de sazonalidade e com 4 bits de sazonalidade. O primeiro teste, a fim de verificar a normalidade das amostras, de Shapiro-Wilk, apresentou um *p-value* maior que 0,05 (nível de significância adotado pelo R) para as duas amostras. Com isso, as amostras puderam ser consideradas normalmente distribuídas. Em seguida, o teste de variância apresentou nível de significância maior que 0,05 e por fim, foi realizado o teste t-Student. Os resultados são mostrados na Tabela 4.:

---

**Tabela 4: Resultados dos testes estatísticos do cenário 1 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,52
Shapiro-Wilk 2	0,14
Teste F	0,69
Teste t-Student	0,11
Teste da Soma dos Postos de Wilcoxon	Não se aplica

[fonte: elaboração própria]

Como o teste t-Student apresentou *p-value* maior que 0,05, a hipótese nula foi aceita, ou seja, as amostras possuem desempenho estatisticamente iguais. Em seguida, seleciona-se a amostra com menor EPMA e ela com a amostra contendo 12 bits de sazonalidade. Foi realizado o mesmo procedimento para execução dos testes. Neste caso, como o *p-value* do teste de variância foi muito menor que o nível de significância, foi feito o teste da Soma dos Postos de Wilcoxon, ao invés do teste t-Student, citado no exemplo anterior. A Tabela 5 retrata os resultados deste caso:

**Tabela 5: Resultados dos testes estatísticos do cenário 1 entre amostras de 4 bits de sazonalidade (1) e com 12 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,14
Shapiro-Wilk 2	0,08
Teste F	0,0009
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,38

[fonte: elaboração própria]

Como pode ser observado, o *p-value* com o teste da Soma dos Postos de Wilcoxon deu maior que 0,05, ou seja, as amostras podem ser consideradas estatisticamente iguais. Com isto, seleciona-se a amostra de menor EPMA e realiza novos testes com a amostra de 2 bits de sazonalidade. Os resultados obtidos são

---

mostrados na Tabela 6 e são semelhantes aos resultados encontrados na Tabela 5, ou seja, comprovam que as duas configurações possuem desempenho estatisticamente iguais:

**Tabela 6: Resultados dos testes estatísticos do cenário 1 entre amostras de 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,08
Shapiro-Wilk 2	0,10
Teste F	0,01529
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,1304

[fonte: elaboração própria]

Após todas essas análises estatísticas, pode-se concluir que a melhor configuração, para o cenário 1, é a que possui 12 bits de sazonalidade, por possuir menor EPMA.

#### 4.1.2 Cenário 2: Últimos 2 meses

Semelhante ao cenário 1, a tabela 7 ilustra a média do EPMA de todas as configurações do cenário 2.

**Tabela 7: Média do EPMA para cada configuração do cenário 2 da base Três Marias, após 30 execuções da MLP.**

Configuração	EPMA
Últimos 12 Meses	40,50 %
4 bits de Sazonalidade	38,52 %
12 bits de Sazonalidade	37,47 %
2 bits de Sazonalidade	40,26 %

[fonte: elaboração própria]

Repetindo a sequência de testes do cenário 1, as amostras sem bits de sazonalidade são comparadas com as amostras com 4 bits de sazonalidade. Depois, seleciona-se a amostra de melhor desempenho e é feita a comparação com a amostra com 12 bits de sazonalidade. Por último, repete o procedimento com a amostra com 2 bits de sazonalidade. Os resultados são exibidos nas Tabelas 8, 9 e 10.

**Tabela 8: Resultados dos testes estatísticos do cenário 2 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,0178
Shapiro-Wilk 2	0,096
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,1872

[fonte: elaboração própria]

**Tabela 9: Resultados dos testes estatísticos do cenário 2 entre amostras de 4 bits de sazonalidade (1) e com 12 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,096
Shapiro-Wilk 2	0,5944
Teste F	0,1733
Teste t-Student	0,4356
Teste da Soma dos Postos de Wilcoxon	Não se aplica

[fonte: elaboração própria]

---

**Tabela 10: Resultados dos testes estatísticos do cenário 2 entre amostras de 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,5944
Shapiro-Wilk 2	0,0565
Teste F	0,9475
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,0181

[fonte: elaboração própria]

A Tabela 8 mostra que a amostra sem bits de sazonalidade não está normalmente distribuída, pois seu *p-value* foi menor que 0,05. Assim, o teste de variância não pode ser realizado, nem o teste t-Student. Dessa forma, faz-se o teste de Wilcoxon. Este último apresentou um valor maior que o nível de significância e com isso as amostras possuem desempenhos estatisticamente iguais.

A Tabela 9 mostra o resultado dos testes da amostra de 4 bits de sazonalidade com a amostra de 12 bits de sazonalidade. Todos os testes (de normalidade das amostras, de variância entre as amostras e o t-Student) apresentaram um *p-value* maior que o nível de significância. Diante disto, as amostras apresentem desempenho estatisticamente iguais e, neste caso, foi selecionada a amostra com 12 bits de sazonalidade, por apresentar um menor EPMA.

A Tabela 10 exhibe os resultados dos testes com as amostras de 12 bits de sazonalidade e 2 bits de sazonalidade. As amostras, após o teste de Shapiro-Wilk, foram consideradas normalmente distribuídas. O teste de variância também apresentou um *p-value* maior que o nível de significância, todavia o teste t-Student apresentou um *p-value* bem abaixo do nível de significância. Diante disto, é aceita a hipótese alternativa em que amostra com 12 bits de sazonalidade é considerada como a amostra com melhor desempenho. Ao analisar os resultados de todos os testes do cenário 2, conclui-se que a configuração de melhor desempenho é a configuração com 12 bits de sazonalidade.

---

#### 4.1.3 Cenário 3: Meses com alta correlação

A Tabela 11 demonstra a média dos EPMA para todas as configurações do cenário 3.

**Tabela 11: Média do EPMA para cada configuração do cenário 3 da base Três Marias, após 30 execuções da MLP.**

Configuração	EPMA
Últimos 12 Meses	37,48 %
4 bits de Sazonalidade	37,54 %
12 bits de Sazonalidade	36,03 %
2 bits de Sazonalidade	37,40 %

[fonte: elaboração própria]

A mesma sequência de testes foi realizada para o cenário 3. Primeiro ocorreram os testes entre as amostras sem bits de sazonalidade e com 4 bits de sazonalidade. Posteriormente, a amostra de melhor desempenho foi testada com a amostra de 12 bits de sazonalidade. E por fim, a amostra de melhor desempenho, deste último teste, foi testada com a amostra de 2 bits de sazonalidade. As Tabelas 12, 13 e 14 explicitam os resultados de cada um destes casos:

**Tabela 12: Resultados dos testes estatísticos do cenário 3 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,2826
Shapiro-Wilk 2	0,4999
Teste F	0,9589
Teste t-Student	0,9577
Teste da Soma dos Postos de Wilcoxon	Não se aplica

[fonte: elaboração própria]

---

**Tabela 13: Resultados dos testes estatísticos do cenário 3 entre amostras sem bits de sazonalidade (1) e com 12 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,2826
Shapiro-Wilk 2	0,2608
Teste F	0,8401
Teste t-Student	0,2022
Teste da Soma dos Postos de Wilcoxon	Não se aplica

[fonte: elaboração própria]

**Tabela 14: Resultados dos testes estatísticos do cenário 3 entre amostras de 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,2608
Shapiro-Wilk 2	0,0463
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos Wilcoxon	0,4147

[fonte: elaboração própria]

A Tabela 12 mostra os resultados dos testes da amostra sem bits de sazonalidade com a amostra com 4 bits de sazonalidade. Os testes de Shapiro-Wilk para as duas amostras, o teste F para a variância e o teste t-Student apresentaram todos um *p-value* maior que o nível de significância. Isto indica que as amostras possuem desempenho estaticamente iguas. Neste caso, é selecionada a amostra sem bits de sazonalidade, pois apresenta menor EPMA.

A Tabela 13 apresenta os resultados dos testes entre as amostras sem bits de sazonalidade e a amostra com 12 bits de sazonalidade. Todos os resultados são semelhantes aos do último teste. Assim, as amostras possuem mesmo desempenho estatístico e é escolhida àquela com menor EPMA.

---

A Tabela 14 apresenta os resultados dos testes entre as amostras com *12 bits* de sazonalidade e a amostra com 2 bits de sazonalidade. Neste caso, como a amostra de 2 bits de sazonalidade, após teste de Shapiro-Wilk, foi considerada como não normalmente distribuída, aplicou-se logo o teste de Wilcoxon. Este último apresentou resultados que mostram desempenhos estatísticos iguais entre as amostras. Assim, entre as duas amostras, é escolhida a de *12 bits* de sazonalidade, por apresentar um menor EPMA. Diante disto, no cenário 3, a configuração que apresentou melhor desempenho foi a configuração com *12 bits* de sazonalidade.

#### 4.1.4 Comparação entre os três cenários

Os mesmos testes estatísticos realizados em cada cenário, para obter a melhor configuração de cada um, foi feito entre os três cenários. Para isto, seleciona-se a amostra de melhor desempenho de cada cenário para representá-lo.

Primeiramente, foram realizados testes entre a melhor configuração do cenário 1 com a melhor configuração do cenário 2. Após isto, aquela que apresentou melhor desempenho, foi testada com a amostra de melhor desempenho do cenário 3. Ao finalizar estes testes, pode-se afirmar qual melhor configuração, entre todas as possíveis, para a base de Três Marias. Após a realização dos testes entre a amostra de melhor desempenho do cenário 1 e a amostra de melhor desempenho do cenário 2, são obtidos os resultados encontrados na Tabela 15:

**Tabela 15: Resultados dos testes estatísticos entre o melhor resultado do cenário 1 e o melhor do cenário 2:**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,0861
Shapiro-Wilk 2	0,5944
Teste F	0,1696
Teste t-Student	0,6892
Teste da Soma dos Postos Wilcoxon	Não se aplica

[fonte: elaboração própria]

---

A Tabela 5 já havia mostrado que a amostra de sazonalidade de *12 bits* do cenário 1 é normalmente distribuída. O mesmo ocorreu para o cenário 2 na Tabela 10. O teste de variância apresentou *p-value* maior que o nível de significância, tornando possível a execução do teste t-Student. Este último mostrou que as amostras possuem desempenho estatisticamente iguais. Assim, a amostra selecionada foi a do cenário 2, por apresentar menor EPMA.

A amostra de melhor desempenho do cenário três, a de *12 bits* de sazonalidade, já havia sido definida como normalmente distribuída, de acordo com a Tabela 13. Após realização dos testes entre as amostras de melhor desempenho do cenário 2 com o cenário 3, foram obtidos resultados semelhantes aos resultados do último teste, entre o cenário 1 e o cenário 2. Com isto, é selecionada a amostra com menor EPMA. :

**Tabela 16: Resultados dos testes estatísticos entre o melhor resultado do cenário 2 e o melhor do cenário 3:**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,5944
Shapiro-Wilk 2	0,2608
Teste F	0,7983
Teste t-Student	0,2089
Teste da Soma dos Postos de Wilcoxon	Não se aplica

[fonte: elaboração própria]

Diante disto, verifica-se que a melhor configuração para realizar a previsão para a base de Três Marias, é a configuração com *12 bits* de sazonalidade do terceiro cenário, ou seja, as entradas são selecionadas a partir do coeficiente de correlação entre elas e são utilizados *12 bits* para representar a sazonalidade da vazão.

## 4.2 Base Tucuruí

Semelhante à base de Três Marias, os dados foram divididos em três cenários. Cada cenário também possui as mesmas quatro configurações de representação da sazonalidade. Após as 30 execuções para cada amostra, foi calculado o erro

---

percentual médio absoluto de cada uma delas e depois calculado o EPMA médio de cada amostra. Abaixo, pode-se encontrar o resultado de cada cenário.

#### 4.2.1 Cenário 1: Últimos Doze Meses

Após as simulações, foi calculado o EPMA médio para cada uma das quatro configurações, como mostra a Tabela 17:

**Tabela 17: Média do EPMA para cada configuração do cenário 1 da base Tucuruí, após 30 execuções da MLP.**

Configuração	EPMA
Últimos 12 Meses	29,97 %
4 bits de Sazonalidade	25,15 %
12 bits de Sazonaliadade	24,46 %
2 bits de Sazonaldiade	28,42 %

[fonte: elaboração própria]

Para realização dos testes estatísticos, foram utilizadas, em primeiro lugar, as amostras sem bits de sazonalidade e com 4 bits de sazonalidade. Ao aplicar o teste de Shapiro-Wilk em cada uma das amostras, observou-se que a amostra com 4 bits de sazonalidade apresentou um *p-value* muito menor que o nível de significância. Assim, conclui-se que a amostra não é normalmente distribuída. Sabendo-se disto, foi feito o teste de Wilcoxon. Este último apresentou um *p-value* muito menor que o nível de significância, conforme a Tabela 18. Logo, a amostra com sazonalidade de 4 bits apresenta um desempenho melhor e será selecionada para dar continuidade aos testes.

---

**Tabela 18: Resultados dos testes estatísticos do cenário 1 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,8791
Shapiro-Wilk 2	0,0275
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,00035

[fonte: elaboração própria]

Em um segundo momento, realizou-se os testes das amostras de 4 bits de sazonalidade e a amostra com 12 bits de sazonalidade. A partir da Tabela 18, sabe-se que a amostra de 4 bits não é normalmente distribuída. Logo, parte-se para o teste de Wilcoxon. O resultado deste último apresenta um *p-value* maior que o nível de significância. Assim, as amostras possuem desempenho estatisticamente iguais e é escolhida a amostra de menor EPMA.

**Tabela 19: Resultados dos testes estatísticos do cenário 1 entre amostras com 4 bits de sazonalidade (1) e com 12 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,0275
Shapiro-Wilk 2	0,0126
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,2928

[fonte: elaboração própria]

Para finalizar o cenário 1, são realizados testes entre as amostras de 12 bits de sazonalidade e 2 bits de sazonalidade. Após aplicar o teste de Shapiro-Wilk na amostra de sazonalidade de 12 bits, sabe-se que a amostra não possui uma distribuição normal. Assim, faz-se necessário o teste da Soma dos Postos de Wilcoxon. A Tabela 20 mostra que este último apresenta um *p-value* bem menor que o nível de significância. Logo, a amostra de 12 bits de sazonalidade possui um

---

desempenho melhor que a amostra de 2 bits de sazonalidade. A amostra com 12 bits é, então, a melhor configuração possível para o cenário 1.

**Tabela 20: Resultados dos testes estatísticos do cenário 1 entre amostras com 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,01261
Shapiro-Wilk 2	0,3517
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,00336

[fonte: elaboração própria]

#### 4.2.1 Cenário 2: Últimos 2 meses

Semelhante ao cenário 1 e à base de Três Marias, a tabela 21 demonstra a média do EPMA de todas as configurações do cenário 2.

**Tabela 21: Média do EPMA para cada configuração do cenário 2 da base Tucuruí, após 30 execuções da MLP.**

Configuração	EPMA
Últimos 12 Meses	35,90 %
4 bits de Sazonalidade	24,03 %
12 bits de Sazonalidade	22,02 %
2 bits de Sazonalidade	37,16 %

[fonte: elaboração própria]

Após o treinamento e cálculo do erro para as quatro configurações, são realizados testes estatísticos. Para comparar as amostras sem bits de sazonalidade e com 4 bits de sazonalidade, foi realizado o teste da normalidade e ambas encontram-se normalmente distribuídas. Após o teste da variância, o *p-value* encontrado é maior que o nível de significância. A partir deste ponto, pode ser realizado o teste t-Student. Este último apresentou um *p-value* muito inferior ao nível de significância. Com isso,

---

sabe-se que a amostra com 4 bits possui um desempenho melhor, uma vez que apresenta um EPMA menor. A Tabela 22 mostra o resultado dos testes.

**Tabela 22: Resultados dos testes estatísticos do cenário 2 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,0529
Shapiro-Wilk 2	0,1712
Teste F	0,2439
Teste t-Student	$6,8 * 10^{-11}$
Teste da Soma dos Postos de Wilcoxon	Não se aplica

[fonte: elaboração própria]

Em seguida, os testes são realizados para as amostras com 4 bits de sazonalidade e com 12 bits de sazonalidade. Como a amostra de 12 bits não se encontra normalmente distribuída após o teste de Shapiro-Wilk, realiza-se o teste da Soma dos Postos de Wilcoxon. Este último apresenta um *p-value* maior que o nível de significância, logo as amostras possuem um desempenho estatisticamente iguais. Escolhe-se então a amostra com menor EPMA. A Tabela 23 retrata o resultado destes testes.

**Tabela 23: Resultados dos testes estatísticos do cenário 2 entre amostras com 04 bits de sazonalidade (1) e com 12 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,1712
Shapiro-Wilk 2	0,00705
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,0877

[fonte: elaboração própria]

Por fim, realizam-se os testes entre as amostras com *12 bits* de sazonalidade e com *2 bits* de sazonalidade. Como a amostra com *12 bits* de sazonalidade já havia sido classificada como não normal, parte-se para o teste da Soma dos Postos de Wilcoxon. Este último apresentou um resultado do *p-value* muito menor que o nível de significância, conforme a Tabela 24. Com isto, a amostra com *12 bits* de sazonalidade apresenta o melhor desempenho entre estas.

**Tabela 24: Resultados dos testes estatísticos do cenário 2 entre amostras com *12 bits* de sazonalidade (1) e com *2 bits* de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,0070
Shapiro-Wilk 2	0,00203
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	$1,98 * 10^{-13}$

[fonte: elaboração própria]

Considerando todos os testes do cenário 2, a configuração que apresentou melhor desempenho foi a com *12 bits* de sazonalidade.

#### 4.2.3 Cenário 3: Coeficientes de alta correlação

A tabela 25 demonstra a média dos EPMA para todas as configurações do cenário 3.

**Tabela 25: Média do EPMA para cada configuração do cenário 3 da base Tucuruí, após 30 execuções da MLP.**

Configuração	EPMA
Últimos 12 Meses	27,16 %
<i>4 bits</i> de Sazonalidade	26,38 %
<i>12 bits</i> de Sazonalidade	21,51 %
<i>2 bits</i> de Sazonalidade	28,33 %

[fonte: elaboração própria]

A mesma sequência de testes foi realizada para o cenário 3. Primeiro ocorreram os testes entre as amostras sem bits de sazonalidade e com 4 bits de sazonalidade. Posteriormente, a amostra de melhor desempenho foi colocada à teste com a amostra de sazonalidade com 12 bits. E por fim, o procedimento se repetiu para a sazonalidade com 2 bits. Os resultados dos testes encontram-se nas Tabelas 26, 27 e 28.

**Tabela 26: Resultados dos testes estatísticos do cenário 3 entre amostras sem bits de sazonalidade (1) e com 4 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,00023
Shapiro-Wilk 2	0,0030
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,2188

[fonte: elaboração própria]

**Tabela 27: Resultados dos testes estatísticos do cenário 3 entre amostras com 4 bits de sazonalidade (1) e com 12 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,0030
Shapiro-Wilk 2	0,000356
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,00353

[fonte: elaboração própria]

---

**Tabela 28: Resultados dos testes estatísticos do cenário 3 entre amostras de 12 bits de sazonalidade (1) e com 2 bits de sazonalidade (2):**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,0659
Shapiro-Wilk 2	0,0030
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	$3,43 \cdot 10^{-8}$

[fonte: elaboração própria]

Para verificar a normalidade das amostras, ao realizar o teste de Shapiro-Wilk, verificou-se que as amostras sem bits de sazonalidade, com 4 bits de sazonalidade e com 12 bits de sazonalidade, apresentaram um *p-value* muito menor que o nível de significância e com isso não foram consideradas normalmente distribuídas.

Para testar as amostras sem bits de sazonalidade e com 4 bits de sazonalidade (Tabela 26), realizou-se o teste da Soma dos Postos de Wilcoxon. Este apresentou um *p-value* maior que 0,05 e as amostras puderam ser consideradas com mesmo desempenho estatístico. Assim, foi selecionada a amostra com menor EPMA, a de 4 bits de sazonalidade.

Em seguida, a amostra de 4 bits de sazonalidade foi testada com a amostra de 12 bits de sazonalidade, através do teste de Wilcoxon. Neste caso, conforme mostra a Tabela 28, o resultado de *p-value* foi muito inferior a 0,05. Neste caso, a amostra de 12 bits apresenta melhor desempenho.

Repete-se o mesmo procedimento para as amostras com 12 bits de sazonalidade e com 2 bits de sazonalidade. Como a primeira não é normalmente distribuída, realiza-se o teste da Soma dos Postos Wilcoxon. Este, por sua vez, apresentou um *p-value* bem menor que o nível de significância. Assim, a configuração com 12 bits possui melhor desempenho que a configuração com 2 bits. Os resultados se encontram na Tabela 28.

Assim, dentro do cenário 3, a configuração que obteve melhor desempenho foi a configuração com 12 bits de sazonalidade.

---

#### 4.2.4 Comparação entre os 3 cenários

De maneira semelhante aos testes realizados na seção 4.1.3, compara-se o melhor resultado de cada cenário com a intenção de averiguar qual o melhor entre eles. Primeiro, a amostra com 12 bits de sazonalidade do cenário 1 e a amostra com 12 bits de sazonalidade do cenário 2. Nos dois casos, ao realizar o teste de Shapiro-Wilk, o valor do *p-value* calculado foi menor que 0,05. Logo, as amostras não são normalmente distribuídas e o teste a ser aplicado é o da Soma dos Postos de Wilcoxon. Ao aplicar este teste, verificou-se que as amostras possuem um desempenho estatisticamente igual, pois o *p-value* foi maior que 0,05. Os valores exatos se encontram na Tabela 29.

**Tabela 29: Resultados dos testes estatísticos do melhor do cenário 1 e o melhor do cenário 2:**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,0126
Shapiro-Wilk 2	0,00705
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,0699

[fonte: elaboração própria]

Com isto, a configuração do cenário 2 é selecionada, por apresentar um EPMA menor.

Finalmente, é realizado o teste com a amostra selecionada no teste anterior, de 12 bits de sazonalidade do cenário 2, com a amostra de 12 bits de sazonalidade do cenário 3. Os valores dos testes podem ser vistos na Tabela 30. Como as amostras não são normalmente distribuídas, foi aplicado o teste da Soma dos Postos de Wilcoxon. O valor do *p-value* obtido foi maior que o nível de significância, provando que as duas amostras possuem um desempenho estatisticamente igual. Assim, a amostra escolhida foi a que apresentou menor EPMA. Assim, na base de Tucuruí,

---

entre todas as configurações possíveis, a que apresentou melhor desempenho foi a de 12 bits de sazonalidade do cenário 3, ou seja, a configuração que seleciona as entradas a partir do coeficiente de correlação e utiliza 12 bits para representar a sazonalidade da vazão.

**Tabela 30: Resultados dos testes estatísticos do melhor do cenário 1 e o melhor do cenário 2:**

Teste	<i>p-value</i>
Shapiro-Wilk 1	0,0070
Shapiro-Wilk 2	0,00035
Teste F	Não se aplica
Teste t-Student	Não se aplica
Teste da Soma dos Postos de Wilcoxon	0,5819

[fonte: elaboração própria]

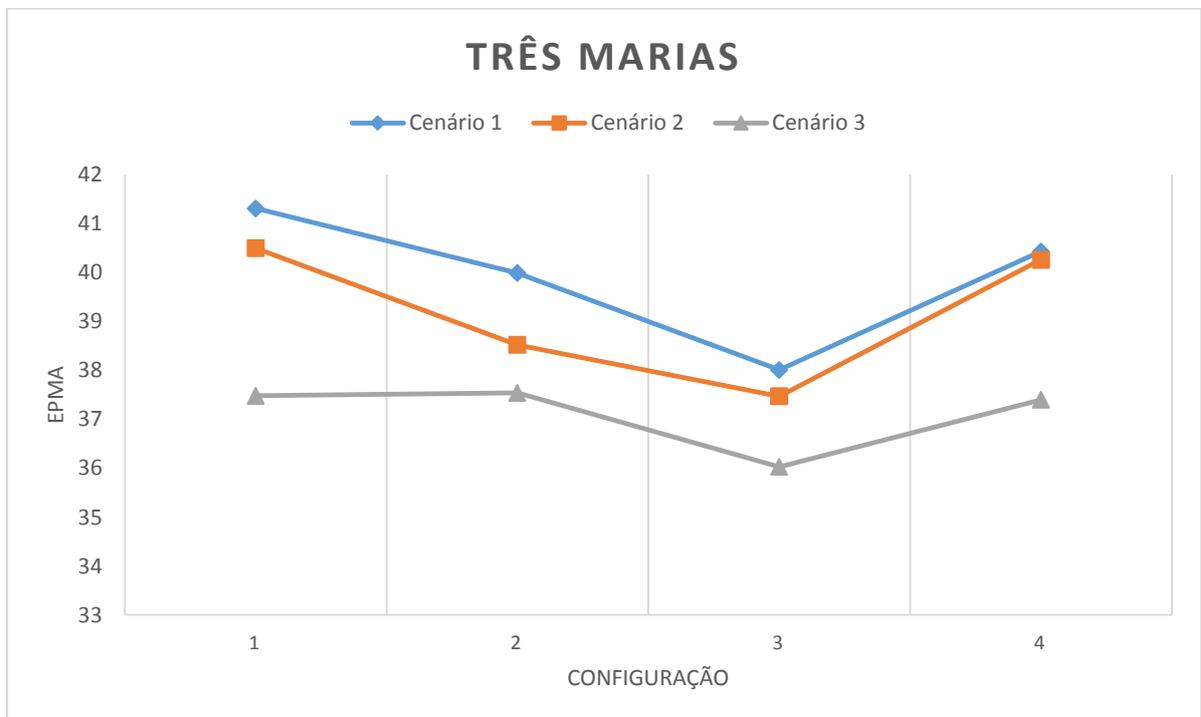
### 4.3 Comparativo dos resultados entre as bases

Após realizar todos os treinamentos das MLPs de todas as configurações possíveis para cada base, e em seguida aplicar todos os testes estatísticos, algumas semelhanças entre os resultados podem ser observadas. Dentro de cada cenário da base de Três Marias, ao finalizar os testes estatísticos, foi comprovado que a configuração que apresentava melhor desempenho era a que utilizava 12 bits para representar a sazonalidade. Independente se o cenário era utilizando os 12 últimos meses, os 2 últimos meses ou os meses de correlação mais alta. O mesmo resultado se aplica à base de Tucuruí.

Após este fato, ao se verificar dentro de cada base qual dos cenários apresentava o melhor desempenho, a resposta foi coerente e coincidiu de ser a mesma para as duas bases. A configuração de melhor desempenho, além de possuir 12 bits para representar a sazonalidade, foi a do cenário 3 para as duas bases. Entre os 12 últimos meses, seleciona-se àqueles que possuem um coeficiente de correlação superior a 0,5. Tais coeficientes foram os mesmos nas duas bases. São eles: 't-12'(valor da vazão de 12 meses atrás), 't-11'(valor da vazão de 11 meses atrás) e 't-1'(valor da vazão de 1 mês atrás). Utilizando estas variáveis como entrada e

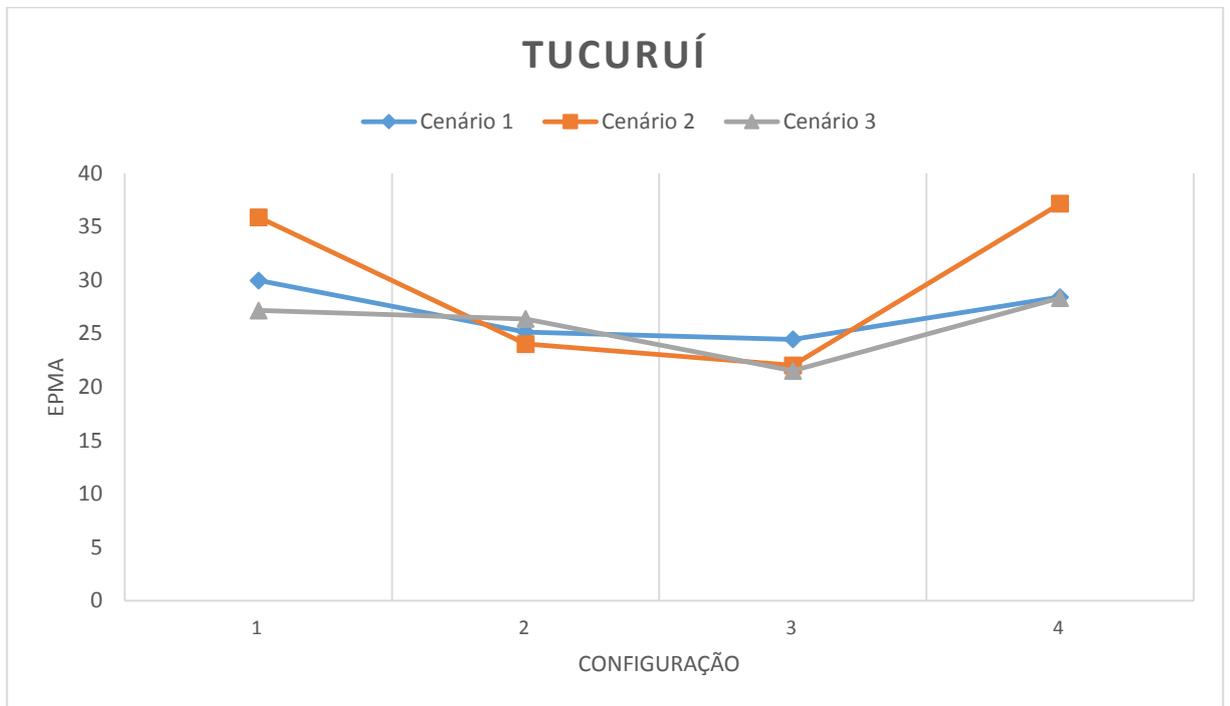
adicionando 12 bits para representar a sazonalidade, para estas duas usinas hidroelétricas, irá prover um melhor desempenho na previsão das vazões mensais. As figuras 8 e 9 abaixo retratam os resultados das base de Três Marias e Tuturucí, respectivamente. A configuração 1 represente a amostra sem bits de sazonalidade. A configuração 2 contém a amostra com 4 bits de sazonalidade. Na configuração 3, as amostras possuem 12 bits de sazonalidade e por fim a configuração 4 possui 2 bits de sazonalidade.

**Figura 8. Resultados de todas as configurações para cada cenário da base de Três Marias.**



[fonte: elaboração própria]

**Figura 9. Resultado de todas as configurações para cada cenário da base de Tucuruí.**



[fonte: elaboração própria]

# 5 Considerações Finais

## 5.1 Conclusões

Previsões mensais de bacias hidrográficas no país são realizadas frequentemente para o planejamento energético do Brasil, uma vez que este tipo de energia é responsável pela maioria da produção de energia utilizada no país. Diante disto, é desejado e necessário cada vez mais precisão nos resultados obtidos com as previsões.

O principal objetivo deste trabalho foi investigar diversas configurações de Redes Neurais Artificiais a fim de observar quais delas apresentavam melhor desempenho para realizar previsões mensais de bacias hidrográficas. Cada configuração averiguada foi simulada 30 vezes em uma RNA do tipo MLP e o algoritmo de treinamento utilizado foi o *BackPropagation*. Após todas as comparações, foram feitos testes estatísticos utilizando o software R, e ficou comprovado que a configuração de melhor desempenho da base de Três Marias foi a mesma na base de Tucuruí, isto é, a configuração com 12 bits de sazonalidade e utilizando variáveis com maior coeficiente de correlação.

## 5.2 Trabalhos Futuros

Para trabalhos futuros, pretende-se aprofundar os estudos deste trabalho utilizando outras técnicas de redes neurais artificiais para avaliar seus desempenhos na previsão mensal de vazões.

No lugar do *BackPropagation*, outros algoritmos para treinar a MLP podem ser testados, como o *Levenberg-Marquardt*. Os parâmetros da rede como o número de ciclos, taxa de aprendizado e *momentum* podem ser alterados para verificar se existe alguma melhora de desempenho.

Finalmente, obter dados de outras bases de previsões mensais de vazão e comparar os resultados alcançados neste trabalho a fim de verificar se a configuração com melhor desempenho foi a mesma.

---

# Referências

[1] BRAGA, A. D. P.; PONCE DE LEON, F. D. C. A.; BERNARDA LUDERMIR, T. **Redes Neurais Artificiais: Teorias e Aplicações**. 2. ed. [S.I.]: LTC, 2000.

[2] COSTA, F. D. S.; PIÑERO, M. E.; DAMÁZIO, J. M. Modelos de previsão hidrológica aplicados ao planejamento da operação do sistema elétrico brasileiro. **Revista Brasileira de Recursos Hídricos**, v. 12, p. 21-30, 2007.

[3] COSTA, F. S. Modelagem estocástica de previsão de vazões mensais: Previvazm. **XV Simpósio Brasileiro de Recursos Hídricos**, Curitiba, 2003.

[4] COSTA, F. S.; DAMÁZIO, J. M.; CARVALHO, A. L. Modelo etocástico de previsão de vazões diárias: Previvazh. **Seminário Internacional sobre Hidrologia Operativa**, Concordia, Argentina, 2000.

[5] HAYKIN, S. **Redes Neurais: Princípios e Práticas**. [S.I.]: Bookman, 2007.

[6] KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e Aplicações**. [S.I.]: Livraria da Física, 2006.

[7] MACIEIRA, M. E. P.; FALCÃO, D. M.; PENNA, D. D. J. Chain of optimization models for setting the energy dispatch and spot price in the brazilian system. **IEEE Porto Power Tech Proceedings**, Porto, v. 3, n. 6, 2001.

[8] MCCULLOCH, W.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, p. 115-133, 1943.

[9] ONS. Operador Nacional do Sistema Elétrico. Disponível em: <[www.ons.org.br/home](http://www.ons.org.br/home)>. Acesso em: 14 out. 2014.

---

[10] SOUZA, W. D. S.; SOUZA, F. D. A. S. Rede Neural Artificial aplicada à previsão de vazão da Bacia Hidrográfica do Rio Piancó. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, n. 2, p. 173-180, 2010.

[11] VALENÇA SILVA, J. M. **Fundamentos de Redes Neurais**. [S.l.]: Livro Rápido, 2010.

---

# Apêndice A

## Scripts utilizados para a execução da Análise Estatística no Software R

```
read.csv("dadosR.csv", header = TRUE, sep = ";", quote="\"", dec="," , fill=FALSE, comment.char="")
```

```
#LER DADOS TRES MARIAS (TM)
```

```
#Cenario 1 -> ultimos 12 meses
```

```
tm_cenarioUm_SemSazonalidade<-c(dados[,1])
```

```
tm_cenarioUm_4Bits<-c(dados[,2])
```

```
tm_cenarioUm_12Bits<-c(dados[,3])
```

```
tm_cenarioUm_SenCos<-c(dados[,4])
```

```
#Cenario 2 -> ultimos 2 meses
```

```
tm_cenarioDois_SemSazonalidade<-c(dados[,5])
```

```
tm_cenarioDois_4Bits<-c(dados[,6])
```

```
tm_cenarioDois_12Bits<-c(dados[,7])
```

```
tm_cenarioDois_SenCos<-c(dados[,8])
```

```
#Cenario 3 -> correlacao > 0,5
```

```
tm_cenarioTres_SemSazonalidade<-c(dados[,9])
```

```
tm_cenarioTres_4Bits<-c(dados[,10])
```

```
tm_cenarioTres_12Bits<-c(dados[,11])
```

```
tm_cenarioTres_SenCos<-c(dados[,12])
```

```
#LER DADOS TUCURUI (TU)
```

```
#Cenario 1 -> ultimos 12 meses
```

```
tu_cenarioUm_SemSazonalidade<-c(dados[,13])
```

```
tu_cenarioUm_4Bits<-c(dados[,14])
```

```
tu_cenarioUm_12Bits<-c(dados[,15])
```

```
tu_cenarioUm_SenCos<-c(dados[,16])
```

```
#Cenario 2 -> ultimos 2 meses
```

```
tu_cenarioDois_SemSazonalidade<-c(dados[,17])
```

```
tu_cenarioDois_4Bits<-c(dados[,18])
```

```
tu_cenarioDois_12Bits<-c(dados[,19])
```

```
tu_cenarioDois_SenCos<-c(dados[,20])
```

```
#Cenario 3 -> correlacao > 0,5
```

```
tu_cenarioTres_SemSazonalidade<-c(dados[,21])
```

```
tu_cenarioTres_4Bits<-c(dados[,22])
```

```
tu_cenarioTres_12Bits<-c(dados[,23])
```

```
tu_cenarioTres_SenCos<-c(dados[,24])
```

---

### #BASE TRES MARIAS

#### #CENARIO 1

```
shapiro.test(tm_cenarioUm_SemSazonalidade)
shapiro.test(tm_cenarioUm_4Bits)
shapiro.test(tm_cenarioUm_2Bits)
shapiro.test(tm_cenarioUm_SenCos)
```

```
var.test(tm_cenarioUm_SemSazonalidade, tm_cenarioUm_4Bits)
wilcox.test(tm_cenarioUm_SemSazonalidade, tm_cenarioUm_4Bits)
```

```
var.test(tm_cenarioUm_4Bits, tm_cenarioUm_12Bits)
wilcox.test(tm_cenarioUm_4Bits, tm_cenarioUm_12Bits)
```

```
var.test(tm_cenarioUm_12Bits, tm_cenarioUm_SenCos)
wilcox.test(tm_cenarioUm_12Bits, tm_cenarioUm_SenCos)
```

#### #CENARIO 2

```
shapiro.test(tm_cenarioDois_SemSazonalidade)
shapiro.test(tm_cenarioDois_4Bits)
shapiro.test(tm_cenarioDois_2Bits)
shapiro.test(tm_cenarioDois_SenCos)
```

```
var.test(tm_cenarioDois_SemSazonalidade, tm_cenarioDois_4Bits)
wilcox.test(tm_cenarioDois_SemSazonalidade, tm_cenarioDois_4Bits)
```

```
var.test(tm_cenarioDois_4Bits, tm_cenarioDois_12Bits)
wilcox.test(tm_cenarioDois_4Bits, tm_cenarioDois_12Bits)
```

```
var.test(tm_cenarioDois_12Bits, tm_cenarioDois_SenCos)
wilcox.test(tm_cenarioDois_12Bits, tm_cenarioDois_SenCos)
```

#### #CENARIO 3

```
shapiro.test(tm_cenarioTres_SemSazonalidade)
shapiro.test(tm_cenarioTres_4Bits)
shapiro.test(tm_cenarioTres_2Bits)
shapiro.test(tm_cenarioTres_SenCos)
```

```
var.test(tm_cenarioTres_SemSazonalidade, tm_cenarioTres_4Bits)
wilcox.test(tm_cenarioTres_SemSazonalidade, tm_cenarioTres_4Bits)
```

```
var.test(tm_cenarioTres_4Bits, tm_cenarioTres_12Bits)
wilcox.test(tm_cenarioTres_4Bits, tm_cenarioTres_12Bits)
```

```
var.test(tm_cenarioTres_12Bits, tm_cenarioTres_SenCos)
wilcox.test(tm_cenarioTres_12Bits, tm_cenarioTres_SenCos)
```

#### #COMPARACAO ENTRE OS TRES CENARIOS

```
var.test(tm_cenarioUm_12Bits, tm_cenarioDois_12Bits)
wilcox.test(tm_cenarioUm_12Bits, tm_cenarioDois_12Bits)
```

---

```
var.test(tm_cenarioDois_12Bits, tm_cenarioTres_12Bits)
wilcox.test(tm_cenarioDois_12Bits, tm_cenarioTres_12Bits)
```

```
#BASE TUCURUI
```

```
#CENARIO 1
```

```
shapiro.test(tu_cenarioUm_SemSazonalidade)
shapiro.test(tu_cenarioUm_4Bits)
shapiro.test(tu_cenarioUm_2Bits)
shapiro.test(tu_cenarioUm_SenCos)
```

```
var.test(tu_cenarioUm_SemSazonalidade, tu_cenarioUm_4Bits)
wilcox.test(tu_cenarioUm_SemSazonalidade, tu_cenarioUm_4Bits)
```

```
var.test(tu_cenarioUm_4Bits, tu_cenarioUm_12Bits)
wilcox.test(tu_cenarioUm_4Bits, tu_cenarioUm_12Bits)
```

```
var.test(tu_cenarioUm_12Bits, tu_cenarioUm_SenCos)
wilcox.test(tu_cenarioUm_12Bits, tu_cenarioUm_SenCos)
```

```
#CENARIO 2
```

```
shapiro.test(tu_cenarioDois_SemSazonalidade)
shapiro.test(tu_cenarioDois_4Bits)
shapiro.test(tu_cenarioDois_2Bits)
shapiro.test(tu_cenarioDois_SenCos)
```

```
var.test(tu_cenarioDois_SemSazonalidade, tu_cenarioDois_4Bits)
wilcox.test(tu_cenarioDois_SemSazonalidade, tu_cenarioDois_4Bits)
```

```
var.test(tu_cenarioDois_4Bits, tu_cenarioDois_12Bits)
wilcox.test(tu_cenarioDois_4Bits, tu_cenarioDois_12Bits)
```

```
var.test(tu_cenarioDois_12Bits, tu_cenarioDois_SenCos)
wilcox.test(tu_cenarioDois_12Bits, tu_cenarioDois_SenCos)
```

```
#CENARIO 3
```

```
shapiro.test(tu_cenarioTres_SemSazonalidade)
shapiro.test(tu_cenarioTres_4Bits)
shapiro.test(tu_cenarioTres_2Bits)
shapiro.test(tu_cenarioTres_SenCos)
```

```
var.test(tu_cenarioTres_SemSazonalidade, tu_cenarioTres_4Bits)
wilcox.test(tu_cenarioTres_SemSazonalidade, tu_cenarioTres_4Bits)
```

```
var.test(tu_cenarioTres_4Bits, tu_cenarioTres_12Bits)
wilcox.test(tu_cenarioTres_4Bits, tu_cenarioTres_12Bits)
```

```
var.test(tu_cenarioTres_12Bits, tu_cenarioTres_SenCos)
wilcox.test(tu_cenarioTres_12Bits, tu_cenarioTres_SenCos)
```

---

### #COMPARACAO ENTRE OS TRES CENARIOS

```
var.test(tu_cenarioUm_12Bits, tu_cenarioDois_12Bits)  
wilcox.test(tu_cenarioUm_12Bits, tu_cenarioDois_12Bits)
```

```
var.test(tu_cenarioDois_12Bits, tu_cenarioTres_12Bits)  
wilcox.test(tu_cenarioDois_12Bits, tu_cenarioTres_12Bits)
```