

**JOSÉ AMÉRICO TRIVELLATO MESSIAS**

**REDES NEURAIS ARTIFICIAIS: NOVA ABORDAGEM  
PARA IDENTIFICAÇÃO DA QUALIDADE DA BEBIDA DO  
CAFÉ ARÁBICA UTILIZANDO-SE RESULTADOS DE  
ANÁLISES QUÍMICAS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

**VIÇOSA  
MINAS GERAIS – BRASIL  
2009**

**Ficha catalográfica preparada pela Seção de Catalogação e  
Classificação da Biblioteca Central da UFV**

T

M585r  
2009

Messias, José Américo Trivellato, 1960-

Redes neurais artificiais: nova abordagem para identificação da qualidade da bebida do café arábica utilizando-se resultados de análises químicas / José Américo Trivellato Messias.

– Viçosa, MG, 2009.

vii, 76f. : il. ; 29cm.

Inclui apêndices.

Orientador: Evandro de Castro Melo.

Tese (doutorado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 52-58.

1. *Coffea arabica*. 2. Café - Qualidade. 3. Café - Classificação.  
4. Café - Análise. 5. Redes neurais (Computação).

I. Universidade Federal de Viçosa. II. Título.


CDD 22.ed. 663.93


**JOSÉ AMÉRICO TRIVELLATO MESSIAS**


**REDES NEURAIS ARTIFICIAIS: NOVA ABORDAGEM  
PARA IDENTIFICAÇÃO DA QUALIDADE DA BEBIDA DO  
CAFÉ ARÁBICA UTILIZANDO-SE RESULTADOS DE  
ANÁLISES QUÍMICAS**

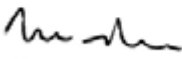
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Engenharia Agrícola, para obtenção do título de *Doctor Scientiae*.

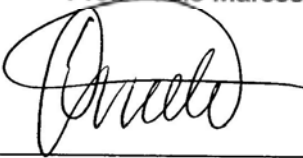
APROVADA: 7 de dezembro de 2009.

  
Prof. José Helvecio Martins  
(Coorientador)

  
Prof. Flávio Meira Borém

  
Prof. Jacir Nogueira da Silva

  
Prof. Paulo Marcos de Barros Monteiro

  
Prof. Evandro de Castro Melo  
(Orientador)

Aos meus pais, Marcus (*in memoriam*) e Maria Ignês, à minha filha, Aline, esposa, Maria Cristina, e aos irmãos, Marcus, Ana e Helena, que sempre me acolheram com carinho e amor, dedico.

## AGRADECIMENTOS

Ao término deste trabalho, gostaria de registrar meu agradecimento a todos que auxiliaram, direta ou indiretamente, durante sua realização.

Ao Departamento de Computação (DECOM) do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto, por ter me liberado para fazer este curso.

À Universidade Federal de Viçosa, por intermédio do Departamento de Engenharia Agrícola, pela acolhida.

A CAPES, pelo apoio financeiro.

Aos amigos e colegas da pós-graduação Paulo Pinto, Daniela, José Eduardo, André, Mauro, Rony, Cristiano, Juliana, Rafael, Paulo Coradi, Silvia, Fernando, Ana Paula, Ernandes e José Vitor, pelo companheirismo e auxílio. Aos demais colegas, pelo convívio e amizade ao longo desses anos de estudos.

Aos meus conselheiros, professores Adílio Flauzino de Lacerda Filho, José Helvécio Martins, José Luis Braga e Paulo Roberto Cecon, pelo estímulo, acompanhamento e sugestões.

Ao professor Paulo Marcos de Barros Monteiro, pelo incentivo, apoio e amizade, sem os quais este trabalho não poderia ser realizado.

Ao professor Flávio Meira Borém, pelo apoio e compreensão.

Ao professor Evandro de Castro Melo, pela orientação, com paciência, empenho e amizade.

A Deus, sempre presente.

## BIOGRAFIA

José Américo Trivellato Messias, filho de Marcos Lanna Messias e Maria Ignês Trivellato Messias, nasceu em Ponte Nova, MG, em 28 de abril de 1960. Graduiu-se em Engenharia Civil na Universidade Federal de Viçosa. Mestre em Ciências da Computação pela Universidade Federal de Minas Gerais em 1994, concluiu, em dezembro de 2009, o doutorado em Engenharia Agrícola na Universidade Federal de Viçosa.

Profissionalmente, é professor do Departamento de Computação do Instituto de Ciências Exatas e Biológicas da Universidade Federal de Ouro Preto, desde 1989.

# ÍNDICE

RESUMO .....	vi
ABSTRACT.....	vii
1. INTRODUÇÃO.....	1
2. REVISÃO BIBLIOGRÁFICA .....	3
2.1. Rede Neural Artificial.....	3
2.2. Qualidade do Café.....	10
2.2.1. Classificação .....	10
2.2.2. Composição Química e Qualidade.....	13
3. MATERIAL E MÉTODOS .....	20
3.1. Dados Experimentais.....	20
3.2. Análise Estatística .....	21
3.3. Rede Neural Artificial.....	21
3.4. Análise dos Dados.....	25
4. RESULTADOS E DISCUSSÕES.....	26
4.1. Regressão linear múltipla <i>versus</i> Rede Neural Artificial .....	26
4.1.1. Equação polinomial .....	31
4.1.2. Equação polinomial com interferência de $\epsilon$ .....	31
4.1.3. Equação polinomial com interferência de $2\epsilon$ .....	32
4.1.4. Equação Mista.....	32
4.1.5. Equação Mista com interferência de $\epsilon$ .....	33
4.1.6. Equação Mista com interferência de $2\epsilon$ .....	34
4.2. Regressão Linear Simples.....	35
4.2.1. Açúcar Redutor.....	36
4.2.2. Açúcar não Redutor.....	37
4.2.3. Acidez Titulável Total .....	39
4.2.4. Condutividade Elétrica.....	40
4.2.5. Lixiviação de Potássio .....	42
4.2.6. Polifenóis.....	44
4.3. Análise Sensorial em Função dos resultados de Acidez Titulável Total, Condutividade Elétrica e Lixiviação de Potássio .....	46
4.4. Análise Sensorial em Função de Todas as Análises Químicas. 47	
4.5. Teste da Rede Neural Utilizando Dados de Campo .....	48
4.6. Discussão dos resultados.....	51
5. CONCLUSÃO .....	53
6. TRABALHOS FUTUROS.....	53
7. REFERÊNCIAS BIBLIOGRÁFICAS .....	54
APÊNDICE.....	61

## RESUMO

MESSIAS, José Américo Trivellato, D. Sc., Universidade Federal de Viçosa, dezembro de 2009. **Redes neurais artificiais: Nova abordagem para identificação da qualidade da bebida do Café Arábica utilizando-se resultados de análises químicas.** Orientador: Evandro de Castro Melo. Coorientadores: Adílio Flauzino de Lacerda Filho, José Helvecio Martins, José Luis Braga e Paulo Roberto Cecon.

Este trabalho foi realizado com o objetivo de avaliar a utilização de rede neural artificial para correlacionar os valores resultantes da análise sensorial com os das análises químicas de amostras de café, com vistas à classificação da bebida de café. As amostras de café utilizadas são referentes ao café *Coffea arabica* L., cultivares Acaia do Cerrado, Topázio, Catuaí vermelho 99, Acaia 474-19 e Bourbon, coletadas na região Sul do Estado de Minas Gerais, com processamento pós-colheita controlado. As análises químicas utilizadas foram de açúcar não redutor, açúcar redutor, acidez titulável total, condutividade elétrica, lixiviação de potássio e polifenóis. A qualidade da bebida do café foi avaliada pela análise sensorial. A regressão linear múltipla dos valores de análise sensorial em função dos valores das análises químicas teve coeficiente de determinação igual a 0,6681, enquanto o método de rede neural artificial empregado obteve um nível de acerto na classificação dos valores da análise sensorial em torno de 85%. A rede neural artificial obtida foi testada com amostras de café coletadas em propriedades da mesma região, com processamento pós-colheita sem controle, obtendo acerto de 75% na classificação da bebida. O resultado obtido é indicador de que o método de rede neural artificial é uma ferramenta adequada para correlacionar os valores resultantes da análise sensorial com os das análises químicas.



## ABSTRACT

MESSIAS, José Américo Trivellato, D. Sc., Universidade Federal de Viçosa, December of 2009. **Artificial neural network: new approach for identification of the quality of drinking Coffee Arabic using chemical analysis results.** Adviser: Evandro de Castro Melo. Co-advisers: Adílio Flauzino de Lacerda Filho, José Helvecio Martins, José Luis Braga and Paulo Roberto Cecon.

The present study was performed with the objective of evaluating the utilization of an artificial neural network to correlate results from sensorial analysis with the chemical analysis of coffee samples. The coffee samples utilized belonged to the coffee plant *Coffea arabica* L., cultivars Acaiá do Cerrado, Topázio, Catuaí vermelho 99, Acaiá 474-19 and Bourbon, collected in the southern region of Minas Gerais State, Brazil, with controlled post-harvest processing. Performed chemical analyses were those of non-reducing sugars, reducing sugars, total titratable acidity, electrical conductivity, leaching of potassium and polyphenols. Coffee quality was evaluated by the sensorial analysis. The multiple linear regression fits of the experimental data from sensorial analysis as a function of the data from the chemical analysis presented a determination coefficient of 0.6681, while the employed artificial neural network reached a classification success rate of, approximately, 85% for the values of sensorial analysis. The artificial neural network obtained was tested with coffee samples collected from properties in the same region, with uncontrolled postharvest processing, reaching a 75% success rate for drink classification. The obtained result is an indicator that the artificial neural network is a suitable tool to correlate results from sensorial analysis with the chemical analysis.

## 1. INTRODUÇÃO

A qualidade do café está relacionada a diversos constituintes físicos e químicos que determinam o aspecto dos grãos, o aroma e sabor da bebida. Segundo Goulart et al. (2003), a constituição química do grão do café interfere diretamente na qualidade da bebida, destacando-se os seguintes constituintes: aminoácidos, proteínas, cafeína, triglicerídios, açúcares, dextrina, pentosanas, galactomananas, celulose, ácido cafeico, ácido clorogênico e minerais. A dificuldade na definição de um protocolo de determinação laboratorial para a qualidade do café decorre da complexidade da composição química dos grãos. A qualidade da bebida é de suma importância, pois dela depende o valor de comercialização do café, podendo, inclusive, ser fator determinante de exclusão do mercado por não apresentar padrão mínimo de comercialização.

O estudo de processos que visam à determinação dos fatores que interferem na qualidade da bebida do café é foco de inúmeros trabalhos acadêmicos, que estabelecem várias características que interferem direta e indiretamente na determinação dessa qualidade. A análise dessas variáveis é feita, tradicionalmente, com a utilização de ferramentas estatísticas, que têm o foco na obtenção de um modelo que explique a variação das características. Porém, essas ferramentas normalmente não fornecem um modelo que explique a correlação entre as características e classificação da bebida de forma eficiente, tornando-se necessário o estudo de modelos mais eficazes.

A rede neural artificial é ferramenta computacional que provê um modelo que, a partir dos valores de entrada, permite obter valores de saída. O uso da rede neural artificial neste trabalho se deve às suas características de modelagem, sendo possível a criação e reconhecimento de classes de objetos em função de suas características. Os ambientes de programação de rede neural artificial permitem sua adequação ao caso em estudo e a modificação de sua estrutura, tendo em vista o aumento de sua eficiência. O objeto de rede, equivalente a

uma função matemática, é obtido com o treinamento da rede com a utilização de um conjunto de dados de entrada, em que se conhecem os valores de saída. Posteriormente, esse objeto de rede gerado pode ser utilizado para obter o valor de saída, em que se conhecem os valores de entrada e não o valor de saída. A utilização da rede neural artificial contempla a solução do problema em estudo, qual seja estabelecer um método que determina a correspondência entre as variações dos valores da análise sensorial e as variações dos valores das análises químicas.

O objetivo deste trabalho foi verificar o desempenho obtido do método de rede neural artificial no estabelecimento da correlação entre os valores da análise sensorial e os das análises químicas, em comparação com o resultado obtido por meio da análise de regressão linear. Além disso, foi realizado o estudo do comportamento do método de rede neural artificial com a utilização de amostras de cafés diferentes das amostras que deram origem aos dados utilizados no treinamento da rede.

## 2. REVISÃO BIBLIOGRÁFICA

### 2.1. Rede Neural Artificial

Uma rede neural artificial é um processador fortemente paralelo e distribuído, constituído de unidades de processamento simples que têm a propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso (HAYKIN, 2000). Ela se assemelha ao cérebro humano em dois aspectos:

- O conhecimento é adquirido pela rede a partir de seu ambiente, por meio de um processo de aprendizagem.
- As forças de conexão entre neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido.

O funcionamento de uma rede neural artificial tenta simular o funcionamento de uma rede neural humana. Na Figura 2.1, ilustra-se o neurônio, célula do sistema nervoso humano responsável pela condução do impulso nervoso.

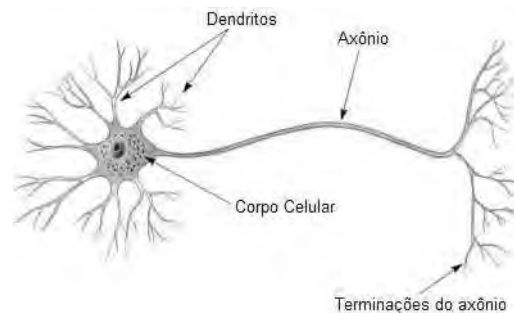


Figura 2.1 - Neurônio humano.

O neurônio é construído de soma (corpo), dendrito e axônio. O dendrito é a conexão entre dois neurônios vizinhos, segundo as formações que fazem o contato entre essas células, para que se propague o impulso. Os sinais são transportados de um neurônio para outro, por meio das sinapses, por uma variedade de substâncias químicas, chamadas de neurotransmissores.

A rede neural biológica é constituída por cerca de  $10^{11}$  neurônios com  $10^4$  conexões para cada neurônio. As pessoas nascem com determinadas estruturas neurais, enquanto outras estruturas são estabelecidas por experiência. As funções biológicas dos neurônios são armazenadas neles e nas conexões. Aprender é um processo de estabelecimento de novas conexões ou modificações das existentes.

A Figura 2.2 ilustra o esquema de um neurônio de uma rede neural artificial.

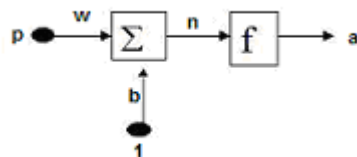


Figura 2.2 - Representação do neurônio artificial.

O valor de saída de cada neurônio pode ser estimado conforme a Figura 2.2, utilizando-se a Equação 2.1 (HAGAN, 2002).

$$a = f(w \times p + b) \quad (2.1)$$

em que:

$a$  = saída;

$p$  = peso sináptico, que é um valor de entrada do neurônio;

$w$  = peso associado ao valor de entrada;

$b$  = tendência<sup>1</sup>; e

$f$  = função de transferência.

O peso sináptico ( $w$ ) é multiplicado pelo valor de entrada ( $p$ ), podendo assumir valores negativos ou positivos. Um combinador linear ( $\Sigma$ ) é utilizado para somar os valores de entrada e tendência ( $b$ ) do

<sup>1</sup> Tendência, polarização ou inclinação, o que leva a preferir ou privilegiar um elemento em detrimento de outro (bias).

neurônio. A função de ativação ( $f$ ), ou de transferência, é utilizada normalmente para limitar a amplitude da saída.

A Figura 2.3 representa a arquitetura de uma rede neural artificial com quatro camadas, em que ocorrem  $R$  valores de entrada,  $S_i$  neurônios na camada intermediária  $i$  e um neurônio na camada de saída.

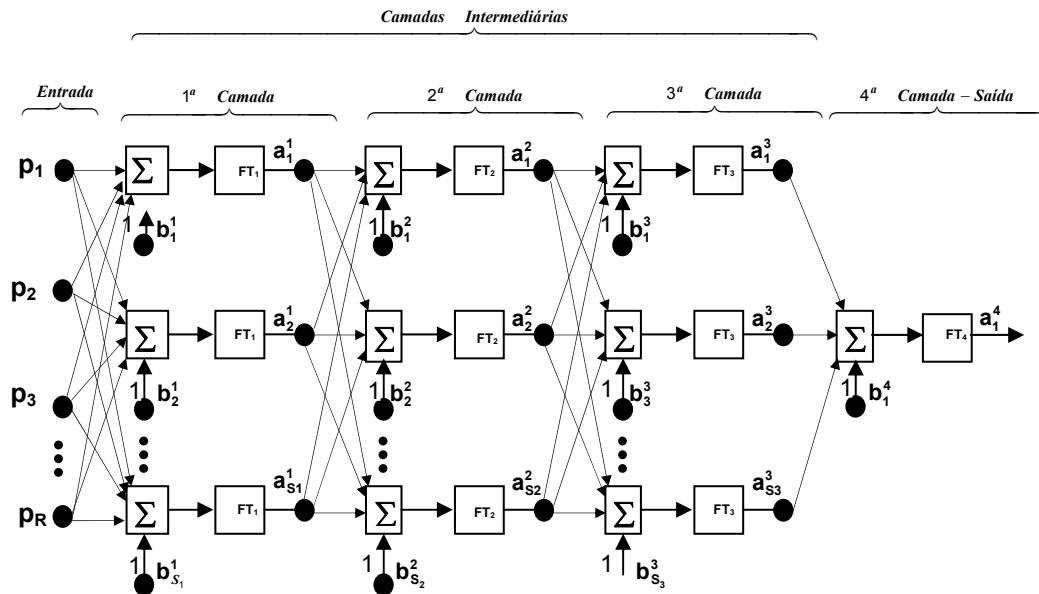


Figura 2.3 - Arquitetura de uma rede neural artificial, em que:

$FT_i$  = função de transferência da cada camada  $i$ ;

$a_j^i$  = valor de saída em que  $i$  é o número da camada e  $j$ , o número do neurônio;

$p_i$  = valor de entrada sendo  $i$  o número da entrada;

$b_j^i$  = valor da tendência em que  $i$  indica o número da camada e  $j$ , o neurônio de destino.

O valor de saída da rede ( $a = a_1^4$ ) pode ser obtido por meio da Equação 2.2.

$$a = FT_4(W_4.FT_3(W_3.FT_2(W_2.FT_1(W_1.p + b^1) + b^2) + b^3) + b^4) \quad (2.2)$$

em que:

$FT_i$  = função de transferência ou de ativação de cada camada  $i$ ;

$p$  = vetor dos valores de entrada da primeira camada;

$W_i$  = matriz de pesos da camada  $i$ ; e

$b^i$  = vetor dos valores de tendência da camada  $i$ .

Os parâmetros  $p$ ,  $w$  e  $b$  da Equação 2.2 são representados matricialmente, conforme as Equações 2.3, 2.4 e 2.5, respectivamente.

$$p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_R \end{bmatrix} \quad (2.3)$$

$$w_1 = \begin{bmatrix} w_{11}, w_{12}, w_{13}, \dots, w_{1R} \\ w_{21}, w_{22}, w_{23}, \dots, w_{2R} \\ \vdots \\ w_{S_1,1}, w_{S_1,2}, w_{S_1,3}, \dots, w_{S_1,R} \end{bmatrix}, \quad w_2 = \begin{bmatrix} w_{11}, w_{12}, w_{13}, \dots, w_{1S_1} \\ w_{21}, w_{22}, w_{23}, \dots, w_{2S_1} \\ \vdots \\ w_{S_2,1}, w_{S_2,2}, w_{S_2,3}, \dots, w_{S_2,S_1} \end{bmatrix}, \quad (2.4)$$

$$w_3 = \begin{bmatrix} w_{11}, w_{12}, w_{13}, \dots, w_{1S_2} \\ w_{21}, w_{22}, w_{23}, \dots, w_{2S_2} \\ \vdots \\ w_{S_3,1}, w_{S_3,2}, w_{S_3,3}, \dots, w_{S_3,S_2} \end{bmatrix}, \quad w_4 = [w_{11}, w_{12}, w_{13}, \dots, w_{1S_3}]$$

$$b_1 = \begin{bmatrix} b_1^1 \\ b_2^1 \\ b_3^1 \\ \vdots \\ b_{S_1}^1 \end{bmatrix}, \quad b_2 = \begin{bmatrix} b_1^2 \\ b_2^2 \\ b_3^2 \\ \vdots \\ b_{S_2}^2 \end{bmatrix}, \quad b_3 = \begin{bmatrix} b_1^3 \\ b_2^3 \\ b_3^3 \\ \vdots \\ b_{S_3}^3 \end{bmatrix}, \quad b_4 = [b_1^4] \quad (2.5)$$

em que:

$p_i$  = valores de entrada em que  $i$  representa o número da entrada;

$w_{ji}$  = peso em que  $j$  indica o neurônio de destino e  $i$  indica a origem do sinal; e

$b_j^i$  = tendência em que  $i$  indica número da camada e  $j$  indica o neurônio de destino.

O processo de criação da rede é constituído por três etapas: implementação, treinamento e teste. Na implementação, a rede é descrita de acordo com sua constituição: número de camadas, número de neurônios em cada camada e funções de ativação de cada camada, ou função de transferência (FT). As funções de ativação são do tipo  $a = f(n)$ .

Por exemplo:

$$\text{hardlimit}(n) = \begin{cases} 0 & \text{se } n < 0 \\ 1 & \text{se } n \geq 0 \end{cases};$$

$$\text{purelim}(n) = \{n, \quad \forall n;$$

$$\text{logsig}(n) = \left\{ \frac{1}{1 + e^{-n}}, \quad \forall n; e \right.$$

$$\text{tansig}(n) = \left\{ \frac{e^n - e^{-n}}{e^n + e^{-n}}, \quad \forall n. \right.$$

Na etapa de treinamento, um primeiro conjunto de dados, de valores de entrada e saída conhecidos, é passado para a rede, e ela é executada, utilizando um algoritmo de aprendizagem até que consiga se ajustar a esses dados. O algoritmo de aprendizagem, ou regras de aprendizagem, é constituído de um conjunto preestabelecido de regras bem definidas para solução de um problema. Existem classes de algoritmos que melhor se enquadram em cada tipo de aplicação.

Na etapa de teste, um segundo conjunto de dados é utilizado pela rede para testar se os valores gerados por ela estão compatíveis com os resultados experimentais. São passados os valores de entrada para a rede gerada no treinamento, que calcula os valores de saída. Então, esses valores são comparados com os valores conhecidos.

Uma variante do processo de desenvolvimento da rede é dividir o conjunto de dados de treinamento em dois subconjuntos: o primeiro é utilizado para o treinamento propriamente dito, enquanto o segundo, para fazer validação durante o processo de treinamento, corrigindo, com isso, distorções.

A propriedade de importância primordial para a rede neural artificial (BRAGA et al., 2000) é sua capacidade de *aprender* (processo de



treinamento da rede) a partir de seu ambiente e de melhorar o seu desempenho por meio da aprendizagem. Uma rede neural artificial aprende acerca do seu ambiente por meio de um processo iterativo de ajustes, aplicados a seus pesos sinápticos e à tendência (*bias*).

Uma rede neural progressiva multicamadas (*feedforward*) consiste em uma rede com várias camadas de neurônios, em que só há ligação no sentido entrada-saída, ligação em sentido único: o grafo formado não contém ciclo. O aprendizado supervisionado é aquele obtido por meio de tutor. O tutor detém o conhecimento do ambiente que a rede neural desconhece. Um conjunto entrada/saída pode representar esse conhecimento.

Um algoritmo de aprendizado com retropropagação (*backpropagation*) do erro utiliza um aprendizado supervisionado. É passado para a rede neural um conjunto de valores de entrada, em que se verifica a saída da rede, o qual é comparado com o valor esperado, determinando-se o erro. De acordo com o valor do gradiente de erro da última camada, os valores dos pesos sinápticos são ajustados, podendo-se, então, calcular o erro da penúltima camada, e assim por diante, até atingir a primeira camada. Esse processo é, então, repetido até que se tenha o valor do erro menor que a tolerância desejada.

Reconhecimento de padrões é o processo pelo qual um padrão é associado a alguma classe dentro de um número preestabelecido de classes. Primeiramente, a rede neural é submetida ao treinamento com um conjunto de dados com a caracterização das classes. Posteriormente, é apresentado um novo conjunto de dados à rede neural e, então, ela se associa às classes estabelecidas no treinamento.

A rede neural artificial é utilizada, atualmente, nas mais variadas áreas de conhecimento. Esse espectro amplo vem incrementando a sua pesquisa e o seu desenvolvimento. A título de exemplo, podem ser citadas as áreas de:

- Saúde, com o desenvolvimento de sistema de detecção e prevenção de doenças, diagnóstico de câncer de mama

(MAZUROWSKI et al., 2008), pesquisa de câncer (ODAJIMA et al., 2008), diagnóstico médico (PENG e KING, 2008) e doenças cardiovasculares (GUELI et al., 2005).

- Modelagem ambiental, consumo de água (FIRAT et al., 2009), concentração de poluentes ambientais (CAWLEY et al., 2007; GRIVAS e CHALOULAKOU, 2006), previsão de chuva (RAMÍREZ et al., 2004) e determinação das propriedades psicrométricas do ar de secagem (MITTAL e ZHANG, 2003).
- Estatística, função-alvo em problemas de regressão e classificação (CHERKASSKY e MA, 2009; KHOSHGOFTAAR e GAO, 2007), modelo de regressão (HERVÁS-MARTÍNEZ et al., 2008) e comparação de desempenho entre regressão linear múltipla e rede neural artificial (CIURANA et al., 2008; SOUSA et al., 2007; VERLINDEN et al., 2008).
- Agricultura, fertilidade do solo (DU e ZHOU, 2009), produção de enzima (TOKATLI et al., 2009), pesquisa de teor de nitrogênio em plantas (CHEN et al., 2008), detecção de teor de água no solo (LIU et al., 2008), cor do vinho armazenado em diferentes tonéis de madeira (RECAMALES et al., 2007), qualidade de chá (DUTTA et al., 2003), identificação de produtos químicos voláteis com o emprego do “nariz eletrônico” (PARDO et al., 2006; KERMANI et al., 2005; PARDO e SBERVEGLIERI, 2004; CHARUMPORN et al., 2003; TOMINAGA et al., 2001), maturação do café (PINTO et al., 2003) e secagem de plantas medicinais (ERENTUR et al., 2004).
- Controle e automação, controle proporcional-integral-derivativo PID (LU et al., 2009) e controle em tempo real (KORJANI et al., 2009).

A utilização de redes neurais artificiais para a classificação de produtos, utilizando sua composição química, é, atualmente, uma grande área de pesquisa. A “língua eletrônica” e o “nariz eletrônico” são utilizadas por meio de sondas imersas em amostras do produto, a fim de

detectar seus constituintes químicos, que são, então, relacionados à qualidade do produto. Outros trabalhos empregam análises químicas e relacionam seus resultados à qualidade do produto.

Suchacz e Wesolowski (2006) analisaram as concentrações dos elementos zinco, cobre, chumbo e cádmio presentes em medicamentos. Essas concentrações são utilizadas como variáveis de entrada para os modelos de redes neurais e empregadas para reconhecer a taxonomia das plantas que foram utilizadas em sua composição.

Pesquisa desenvolvida por Anderson e Smith (2002) demonstrou a viabilidade de um método para diferenciar as regiões geográficas produtoras de café. Análises elementares de K, Mg, Ca, Na, Al, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Mo, S, Cd, Pb e P das amostras de grãos de café foram utilizadas como entrada de um sistema que utiliza métodos de reconhecimento de padrão estatístico, incluindo a análise de componentes principais, a função discriminante e a análise e modelagem de redes neurais, para a determinação de origem geográfica dos alimentos.

## **2.2. Qualidade do Café**

A qualidade da bebida do café está relacionada a diversos constituintes físicos e químicos, que determinam aspectos dos grãos, o sabor e o aroma da bebida. Segundo Goulart et al. (2003), destacam-se os seguintes constituintes: aminoácidos, proteínas, cafeína, triglicerídios, açúcares, dextrina, pentosanas, galactomananas, celulose, ácido cafeico, ácido clorogênico e minerais. A dificuldade na definição de um protocolo de determinação laboratorial da qualidade do café é proveniente da complexidade da composição química do seu fruto.

### **2.2.1. Classificação**

A classificação da bebida é feita por meio de análise sensorial ou prova de xícara, e conforme a Instrução Normativa N° 8 (MAPA, 2003), tem a caracterização apresentada segundo a espécie, formato dos grãos,

granulometria, aroma, sabor e cor. O café beneficiado, de acordo com a bebida e com o grupo a que pertence, é classificado em sete subgrupos, como mostrado no Quadro 2.1.

Quadro 2.1 - Café Arábica – café beneficiado grão cru.

<b>Classificação</b>	<b>Características</b>
Estritamente Mole	Café que apresenta, em conjunto, todos os requisitos de aroma e sabor “mole”, porém mais acentuado.
Mole	Café que apresenta aroma e sabor agradável, brando e adocicado.
Apenas Mole	Café que apresenta sabor levemente doce e suave, mas sem adstringência ou aspereza de paladar.
Duro	Café que apresenta sabor acre, adstringente e áspero, porém não apresenta paladares estranhos.
Riado	Café que apresenta leve sabor típico de iodofórmio.
Rio	Café que apresenta sabor típico e acentuado de iodofórmio.
Rio Zona	Café que apresenta aroma e sabor muito acentuado, assemelhado ao iodofórmio ou ao ácido fênico, sendo repugnante ao paladar.

Fonte: MAPA, 2003.

A classificação da bebida por meio da prova de xícara é subjetiva, pois depende da aptidão do provador e está diretamente relacionada ao seu treinamento, às variedades de cafés com que já se trabalhou e à região de cultivo.

A classificação da bebida para cafés especiais definida pelo método da Associação Americana de Cafés Especiais (SCAA – *Specialty Coffee Association of America*) é baseada em dez atributos (LINGLE, 2001): fragrância e aroma, uniformidade, defeitos, doçura, sabor, acidez, corpo, finalização, equilíbrio e avaliação final. As notas são estabelecidas a partir

de uma escala que representa níveis de qualidade, como descrito no Quadro 2.2.

Quadro 2.2 - Escala de qualidade da metodologia SCAA

<b>Escala de Qualidade</b>	
<b>Nota</b>	<b>Descrição</b>
6 a 6,75	Bom
7 a 7,75	Muito Bom
8 a 8,75	Excelente
9 a 9,75	Excepcional

Fonte: CACCER, 2005.

A faixa de notas entre 2 e 6 é aplicada aos cafés comerciais. O resultado final da classificação, Quadro 2.3, é obtido pela soma dos resultados individuais, subtraindo-se os defeitos.

Quadro 2.3 - Escala de qualidade final da metodologia da SCAA

<b>Pontuação</b>	<b>Descrição</b>	<b>Classificação</b>
<b>Total</b>	<b>Especial</b>	
95 – 100	Exemplar	Super Prêmio de Especialidade
90 – 94	Excepcional	Prêmio de Especialidade
85 – 89	Excelente	Especialidade
80 – 84	Muito Bom	Especial
75 – 79	Bom	Qualidade Boa Normal
70 – 74	Fraco	Qualidade Média
60 – 70		Nota Exchange
50 – 60		Comercial
40 – 50		Nota Baixa
< 40		Sem Nota

Fonte: CACCER, 2005.

O método da Associação Brasileira de Cafés Especiais (BSCA – Brazil Specialty Coffee Association) baseia-se em uma classificação em que a nota atribuída varia de 0 a 8 (ISQUIERDO, 2008). Os atributos avaliados são: corpo, aroma, acidez, doçura, balanço, bebida limpa e sabor característico. Como são oito atributos, e a nota máxima é oito para cada um, a nota máxima total seria 64. Porém, é somado o valor 36 ao total da nota. Dessa forma, a nota total poderá atingir o valor total 100.

### **2.2.2. Composição Química e Qualidade**

O desenvolvimento de pesquisa envolvendo a qualidade da bebida do café é uma prática antiga. No trabalho de Garruti et al. (1968), foi feito um estudo para verificação da quantidade de mistura de café com qualidade dos tipos rio e mole para determinação de bebida com qualidade riado, sendo a qualidade da bebida determinada por avaliação sensorial. A análise dos resultados foi realizada por meio de regressão linear, apresentando como conclusão que a mistura de 10% de café tipo rio e 90% de café tipo mole foi a que teve resultado mais próximo da classificação riado.

A preocupação com a influência dos componentes químicos do café também foi objeto de pesquisa de Sabbagh et al. (1977), que analisaram os teores e tipos de açúcares do café torrado. Nesse estudo, observaram que a sacarose foi o açúcar predominante no café arábica, enquanto os açúcares redutores predominaram no café robusta verde.

Ao analisar o procedimento de secagem, Borém et al. (2008) observaram que, quando a temperatura utilizada for inferior a 40 °C, a membrana celular do grão de café é preservada; quando for utilizada a temperatura de 50 °C, há uma desorganização da membrana com perda da qualidade da bebida do café e, quando for utilizada a temperatura de 60 °C, há a ruptura das vesículas da membrana celular, causando oclusão dos espaços intercelulares. Nesse caso, a perda de qualidade é maior. Esse resultado pode ser observado, de forma indireta, em experimentos de secagem, em que há aumento da condutividade elétrica e lixiviação de

potássio, devido à ruptura da membrana e extravasamento do conteúdo celular.

Na avaliação físico-química e de qualidade do café (*Coffea arabica* L.) submetido a diferentes tempos de espera para secagem, Pimenta et al. (2008) utilizaram as análises, entre outras, de acidez titulável total, lixiviação de potássio e prova de xícara. Nos cafés de pior qualidade, observaram que a acidez titulável total e a lixiviação de potássio tinham maiores valores.

Para estudar o efeito das condições de secagem e armazenamento sobre a qualidade do café natural e despulpado, Coradi et al. (2007) trabalharam com as análises de condutividade elétrica, lixiviação de potássio, acidez titulável total, açúcares totais e redutores e análise sensorial da bebida pelo método da BSCA.

Para avaliar o potencial da região Sul de Minas Gerais para a produção de cafés especiais, Chagas et al. (2005) utilizaram as análises de condutividade elétrica, lixiviação de potássio e de polifenoloxidase. A qualidade da bebida foi estabelecida, de forma indireta, pelos valores resultantes dessas análises químicas. A menor atividade da polifenoloxidase indicou menor índice de danos aos grãos, conseqüentemente podendo-se obter melhor bebida. O menor valor de condutividade elétrica e de lixiviação de potássio indicou que as membranas celulares sofreram menos danos, o que pode possibilitar obter melhor qualidade da bebida.

Na verificação da influência da secagem, em diferentes tipos de terreiro, sobre a qualidade do café durante o armazenamento, Reinato et al. (2007) utilizaram as análises de condutividade elétrica, lixiviação de íons de potássio, açúcares totais e prova de xícara. Os maiores valores de condutividade elétrica e de lixiviação de potássio do café secado em camada espessa ou quando secado em camada fina, mas em terreiro de terra, foram associados à redução na qualidade da bebida do café. Foi observado, também, diminuição dos valores de açúcar total nos cafés secados em camada espessa, o que proporcionou melhores condições

para o desenvolvimento de microrganismos que utilizam os açúcares como um dos principais substratos para o seu desenvolvimento.

Na determinação da qualidade do café submetido a diferentes temperaturas, fluxo de ar e períodos de pré-secagem, Borém et al. (2006) utilizaram a análise sensorial e as análises químicas de condutividade elétrica, lixiviação de potássio, açúcares totais e redutores. Os menores valores de condutividade elétrica e de lixiviação de potássio e os maiores valores de açúcares totais e redutores foram indicadores de melhor qualidade da bebida. Os testes de condutividade elétrica e lixiviação de potássio foram indicadores da integridade das membranas celulares e os açúcares, os precursores do sabor e aroma do café torrado.

Com o objetivo de avaliarem cultivares com resistência à ferrugem, Mendonça et al. (2007) utilizaram análises de composição química e físico-química dos grãos crus de café. Foram empregadas as análises de polifenóis, açúcares totais, redutores e não-redutores. Observaram-se diferenças na composição química dos grãos crus e, dessa forma, concluíram ser possível inferir que existe influência do genótipo sobre essas características.

Malta et al. (2005), analisando os fatores que afetam a determinação da condutividade elétrica e lixiviação de potássio na determinação da qualidade da bebida do café, concluíram que a presença de grãos defeituosos influencia os valores de condutividade elétrica e lixiviação de potássio, podendo fornecer interpretações inadequadas quanto à qualidade do café analisado.

Com o objetivo de avaliarem a influência das diferentes operações de pré-processamento, da colheita até o armazenamento, sobre a qualidade do café, Afonso Júnior et al. (2004) utilizaram a análise de condutividade elétrica como indicador de qualidade. Foi observado elevação dos valores de condutividade elétrica com o prolongamento do tempo de armazenamento, indicando perda de qualidade do café e tendência para a perda da integridade das membranas celulares durante o armazenamento. O processo de retirada da casca do café contribuiu



para redução dos valores médios de condutividade elétrica dos grãos e, conseqüentemente, para a obtenção de cafés com qualidade superior. Houve aumento dos valores médios de condutividade elétrica no processo de secagem, sendo mais acentuado para aqueles em que a secagem do café foi realizada mecanicamente. Observou-se, também, que não houve influência significativa do processo de retirada mecânica da mucilagem na alteração dos valores médios de condutividade elétrica dos grãos de café.

O modo de analisar estatisticamente os resultados das análises químicas foi alterado com o passar dos anos. Inicialmente, as análises univariadas eram mais utilizadas. Posteriormente, as análises multivariadas passaram predominar nos trabalhos, principalmente quando o número de variáveis analisadas se torna grande, caso da análise de componentes voláteis.

Os trabalhos discutidos a seguir deram origem aos dados que foram utilizados no método de aplicação da rede neural artificial para correlacionar os resultados de análises químicas e físico-químicas com a qualidade da bebida neste trabalho.

Isquierdo (2008) apresentou uma avaliação dos efeitos de diferentes períodos de repouso durante a secagem e de diferentes teores de água em que a secagem do café cereja desmucilado foi interrompida, tendo influenciado na composição química e qualidade da bebida. Para caracterização da qualidade da bebida, foram realizadas as análises de condutividade elétrica, lixiviação de potássio, açúcares totais, redutores e não-redutores, acidez titulável total, polifenóis, sólidos solúveis totais e análise sensorial. Os dados de secagem e das análises químicas do café foram avaliados utilizando-se o programa *Sisvar 4.0*, sendo as médias comparadas pelo teste de *Tukey*. Os dados da análise sensorial foram analisados utilizando-se técnicas multivariadas. Entre os resultados apresentados, destacam-se a análise sensorial, que não foi afetada pelo período de repouso nem pelos níveis do teor de água quando a secagem foi interrompida; os cafés, que tiveram a secagem interrompida com teor de água de 24% (bu) para posterior secagem até 11% (bu), sofreram

menor lixiviação de potássio e condutividade elétrica, independentemente do período de repouso; e nas demais análises químicas não se observou influência do período de repouso e do teor de água com a interrupção da secagem.

Nobre (2008) propôs um método para o processamento de frutos imaturos de café. Para tanto, foram realizadas as análises de açúcares totais, açúcares redutores e não redutores, acidez titulável total, sólidos solúveis, lixiviação de potássio, condutividade elétrica, ácidos clorogênicos, classificação física, sensorial (método da BSCA) e prova de xícara. A análise de variância foi utilizada para analisar as variáveis químicas e físico-químicas e a classificação por defeitos do café. As médias foram comparadas entre si, por meio dos testes de *Student* e *Tukey*. Os dados da análise sensorial foram analisados pela análise de agrupamento, utilizando-se o programa *Minitab 15*. O tipo e a bebida foram avaliados em forma de gráficos de frequência. Como resultado destaca-se que o descascamento do café verde ocasionou melhor qualidade final do café, com melhora de suas características químicas e físico-químicas, além de diminuir o número de defeitos e melhorar a qualidade da bebida na prova de xícara.

Taveira (2008) propôs uma avaliação da qualidade sensorial e fisiológica dos grãos de café processados e secos utilizando diferentes métodos, bem como a identificação de marcadores bioquímicos capazes de diferenciar esses procedimentos pós-colheita. Foram utilizados café cereja maduro e café natural maduro. A análise sensorial da bebida do café foi feita pelo sistema proposto pela SCAA. Foram efetuadas as análises químicas do teor de açúcares, acidez titulável total e compostos fenólicos. Foram feitas, ainda, as análises fisiológicas de lixiviação de potássio, condutividade elétrica, germinação, teste de velocidade de emergência e análise eletroforética de proteínas. Os dados obtidos das análises químicas, sensorial e fisiológicas do café foram submetidos à análise de variância e as médias, comparadas pelo teste de *Scott-Knott*. Com base nos resultados observados, verificou-se que a secagem em

terreiro proporcionou melhor qualidade fisiológica e de bebida dos grãos de café. Observou-se melhor qualidade fisiológica no café despulpado que no natural, independentemente do método de secagem; e que a temperatura de 60 °C não é indicada para a secagem, sendo as temperaturas de 40 °C e 60 °C adequadas para o café despulpado, porém não são indicadas para o café natural.

Fernandes (2008) propôs um estudo sobre a qualidade sanitária, físico-química e sensorial de café. Foram utilizados cafés cereja descascado, cereja e verde, roça e bóia. Foi aplicado, nas folhas e frutos, cloreto de benzalcônico (amônia quaternária), que é um produto sanificante no controle de fungos e combate à ferrugem do cafeeiro. Foram realizadas análises microbiológicas, físico-químicas e sensoriais. As análises físico-químicas realizadas foram de teor de água, condutividade elétrica, lixiviação de íons potássio, sólidos solúveis, acidez total titulável, polifenóis, açúcares redutores, não redutores e totais. A análise estatística foi realizada segundo diferentes métodos, envolvendo análises estatísticas univariadas e multivariadas. A análise dos componentes principais foi utilizada para avaliar os dados referentes à qualidade do café, quanto à ausência e presença de defeitos nos grãos. A análise de correspondência foi utilizada para verificar as amostras mais representativas para cada tipo de café. A análise de variância e o teste de *Scott-Knott* foram utilizados para interpretação dos cruzamentos dos tipos de colheita e processamento para esclarecimento de ocorrência de fungos. Os dados da análise sensorial foram analisados, utilizando-se técnicas multivariadas. Com os resultados observados, verificou-se que a análise química de cafés com defeitos não discriminou diferenças entre os tipos de colheita e processamento; apenas o uso de sanificante na pré-colheita ou relacionado a diferentes tipos de colheitas não contribuiu para a melhoria da qualidade dos cafés; o método de processamento cereja descascado permitiu manter a melhor qualidade do café; e cafés com maior quantidade de grãos ardidos tiveram pior qualidade na análise sensorial.

Dois trabalhos estão em andamento: o de Figueiredo (2009) e o de Ribeiro (2009). Figueiredo (2009) utilizando café cereja descascado, desenvolveu um trabalho para descrever o perfil sensorial, químico e proteômico de cultivares de cafeeiro Bourbon (*Coffea arabica* L.), no Sul do Estado de Minas Gerais, visando à identificação de plantas promissoras para a produção de cafés de qualidade superior. As análises químicas utilizadas foram de condutividade elétrica, lixiviação de potássio, sólidos solúveis, acidez titulável total, açúcares redutores, não redutores e totais, polifenóis e análise sensorial, pelo método da SCAA. Já Ribeiro (2009), utilizando café cereja descascado, desenvolveu um trabalho para fazer análise técnica da utilização de atmosfera artificial no armazenamento de cafés especiais. As análises químicas empregadas foram de acidez titulável total, condutividade elétrica, lixiviação de potássio, açúcares totais, redutores e não redutores, compostos fenólicos totais e análise sensorial, pelo método SCAA.

### **3. MATERIAL E MÉTODOS**

Este trabalho foi desenvolvido na Área de Energia do Departamento de Engenharia Agrícola do Centro de Ciências Agrárias da Universidade Federal de Viçosa, em Viçosa, MG.

#### **3.1. Dados Experimentais**

Os dados utilizados foram disponibilizados pela equipe do professor Flávio Meira Borém, do Departamento de Engenharia da Universidade Federal de Lavras, em Lavras, MG.

Os dados estão caracterizados nas referências: ISQUIERDO (2008), 30 conjuntos; RIBEIRO (2009), 84 conjuntos; NOBRE (2009), 70 conjuntos; TAVEIRA (2009), 18 conjuntos; FIGUEIREDO (2009), 115 conjuntos; e FERNANDES (2008), 96 conjuntos. No total, são 413 conjuntos de dados a serem utilizados para implementação da Rede Neural Artificial. Esses dados se referem à espécie *Coffea arabica* L., cultivares Acaiá do Cerrado, cultivar Topázio, cultivar Catuaí Vermelho 99, cultivar Acaiá 474-19 e do grupo Bourbon. Os cafés foram coletados predominantemente na região Sul de Minas Gerais, nos municípios de Lavras, Três Pontas, Santo Antônio do Amparo, Guaxupé e Machado. Os procedimentos de colheita, secagem e armazenamento foram específicos em cada experimento. Cada experimento teve um período de colheita, porém as colheitas foram efetuadas nos últimos cinco anos. As análises foram apropriadas para o objetivo de estudo de cada experimento, porém as análises químicas de açúcar não redutor e açúcar redutor, acidez titulável total, condutividade elétrica, lixiviação de potássio e polifenóis foram comuns a todos. A análise sensorial, método da BSCA ou método da SCAA, também foi comum a todos os experimentos. Essas análises têm seus métodos de determinação descritos nos trabalhos referenciados anteriormente e foram efetuadas em diferentes laboratórios da região.

### 3.2. Análise Estatística

As análises de regressão linear simples e múltipla foram feitas utilizando o Programa SAEG, versão 9.1 (SAEG, 2007). Os resultados da execução dessas análises estão no Apêndice 1. Primeiro, foi realizada a regressão linear simples dos resultados da análise sensorial, em função dos resultados de cada uma das análises químicas. Em seguida, foi realizada a regressão linear múltipla dos resultados da análise sensorial, em função dos resultados das análises de Acidez Titulável Total, Condutividade Elétrica e Lixiviação de Potássio. Por fim, foi realizada a regressão linear múltipla dos resultados da análise sensorial, em função dos resultados de todas as análises químicas.

### 3.3. Rede Neural Artificial

A rede neural artificial desenvolvida foi implementada, treinada e testada, utilizando-se o programa MATLAB<sup>®</sup>, versão 6, *MathWorks Inc.*, com o pacote de ferramentas de redes neurais. Para criar um objeto rede, no MATLAB<sup>®</sup>, foi utilizado o comando *newff*, que possui o seguinte formato:

*Rede* = *newff* (*MM*, *TC*, {*FA*}, *LR*)

Os termos utilizados nesse comando são:

*Rede* – é uma variável que armazena o objeto de rede neural artificial a ser criada;

*MM* – matriz  $R \times 2$ , em que  $R$  é o número de entradas (constituintes físico-químicos do café) com maiores e menores valores dos parâmetros;

*TC* – matriz com o tamanho (número de neurônios) de cada camada;

*FA* – funções de ativação de cada camada, função responsável pela manipulação dos valores de entrada para obter os valores de saída.

O valor de saída **a** é obtido aplicando-se a função de transferência à soma da tendência **b**, *bias*, ao produto da matriz de peso **w** pelo vetor dos valores de entrada **p**, ou seja,  $a = f(w.p+b)$ ; e

*LR* – regra (algoritmo) de aprendizagem utilizada no treinamento da rede.

O treinamento da rede foi feito com a utilização do comando *train*, que possui o formato:

*train (Rede, p, t)*

Os parâmetros utilizados nesse comando são:

*p* – a matriz com a amostra dos vetores de entrada para treinamento, os quais têm que estar em coluna. A matriz possui dimensões  $R \times N$ .  $R$  é a dimensão do vetor de entrada (número de análises químicas utilizadas) e  $N$  o número de vetores da amostra (número de conjuntos de resultados das análises químicas); e

*T* – matriz com os valores corretos da classificação dos  $N$  vetores da amostra. A matriz possui dimensão  $S \times N$ , sendo  $S$  é o número de neurônios na saída da rede neural artificial. Foi utilizado um neurônio na camada de saída ( $S = 1$ ), sendo esse valor de saída associado à classe a que pertence cada conjunto de dados, de acordo com o valor da análise sensorial pertencente.

O teste da rede foi feito com a utilização do comando *sim*, que possui o formato:

*saida = sim (Rede, p')*

Os termos utilizados nesse comando são:

*p'* – matriz transposta de *p* com a amostra dos vetores de teste, valores de entrada, constituintes físico-químicos do café; e

*saida* – vetor contendo os valores associados à classe da qualidade da bebida classificada pela rede.

Os nomes *Rede*, *MM*, *TC*, *p*, *t* e *saida*, utilizados aqui, são variáveis, podendo ser substituídos por qualquer outro nome de variável. *FA* é o nome de uma função de ativação válida para o MATLAB®. Da mesma forma, *LR* é o nome do algoritmo de aprendizagem válido para o MATLAB®, utilizado no treinamento da rede. No desenvolvimento do trabalho, foram testados diferentes algoritmos de aprendizagem e funções de transferência para determinação da rede com melhor desempenho.

A rede neural artificial utilizada foi uma rede progressiva (*feedforward*) multicamadas, com algoritmo de aprendizado com retropropagação (*backpropagation*) do erro.

O valor da análise sensorial está compreendido, teoricamente, entre 0 e 100. Para implementação do método, os valores da análise sensorial foram agrupados da seguinte forma: classe 1 para valor maior que 79; classe 2 para valor inferior ou igual a 79 e maior que 74; classe 3 para valor inferior ou igual a 74 e maior que 69; e classe 4 para valor inferior ou igual a 69.

Os 413 conjuntos de dados, composição química e análise sensorial, utilizados para implementar a rede neural artificial, foram divididos, de forma aleatória, em três grupos: um primeiro, utilizado para treinamento da rede (em torno de 80% dos dados); um segundo, para validação durante o treinamento da rede (em torno de 10% dos dados); e um terceiro, para teste da rede gerada (em torno de 10% dos dados). A verificação de desempenho da rede foi feita comparando o resultado esperado, relativo ao terceiro grupo de dados, com o resultado estabelecido pela rede neural artificial. Cada conjunto possui os dados que foram utilizados como entrada da rede neural artificial, resultados da composição química e valores utilizados para comparar com os valores de saída da rede neural artificial, resultados da análise sensorial. A utilização dos resultados da análise sensorial se dá na fase de treinamento para determinação do erro que será utilizado na aferição do treinamento e na fase de teste para determinação do desempenho da rede neural artificial.

Para definição da arquitetura de rede neural artificial a ser utilizada, foi testado o desempenho de várias configurações de rede com uma camada de entrada, uma camada de saída e duas ou três camadas intermediárias. Foi definido que a camada de saída seria constituída de um neurônio e seu valor associado à classe da bebida do café, determinada pelo valor da análise sensorial. As redes com pequeno número de neurônios, de um a quatro neurônios, nas camadas de entrada e intermediárias não apresentaram bom desempenho. Da mesma forma,



as redes com mais de quinze neurônios por camada também não apresentaram desempenho que justificasse sua utilização. As redes com melhores desempenhos, escolhidas para utilização, foram aquelas com número de neurônios nas camadas de entrada e intermediárias com valores entre cinco e treze neurônios por camada.

Foram testadas as funções de ativação *logsig*, *tansig*, *hardlimit* e *purelim* para utilização nas camadas. O melhor desempenho foi obtido com a utilização da função de ativação *purelim* na camada de saída e *logsig* nas demais.

Foram testados os algoritmos de aprendizado: *trainlm* (*Levenberg-Marquardt*), *trainbfg* (*BFGS Quasi-Newton*), *trainrp* (*Resilient Backpropagation*), *trainscg* (*Scaled Conjugate Gradient*), *traincgb* (*Conjugate Gradient with Powell/Beale Restarts*), *traincgf* (*Fletcher-Powell Conjugate Gradient*), *traincgp* (*Polak-Ribière Conjugate Gradient*), *trainoss* (*One-Step Secant*) e *traingdx* (*Variable Learning Rate Backpropagation*). O algoritmo *trainlm* foi escolhido por prover melhor controle de utilização de memória e tempo de convergência. A utilização dos demais algoritmos não resultou em melhoria de desempenho que justificasse sua utilização.

O maior número de ciclos (*epochs*) utilizado auxilia na obtenção de um resultado em que a rede neural artificial possa ter menor erro. Inicialmente, o valor do erro reduz mais rapidamente e, com o passar dos ciclos, esse erro tende a estabilizar-se. O aumento do valor do número de ciclos faz que o treinamento da rede gaste mais tempo para terminar. Os valores testados para número de ciclos foram: 100, 200, 300, 500, 1.000, 5.000 e 10.000. Verificou-se que o valor 300 foi mais apropriado para utilização, apresentando erro final adequado e com tempo de processamento reduzido para sua obtenção.

A precisão do erro quadrático médio para critério de parada foi definido para que o treinamento fosse interrompido quando se atingisse o valor do erro dentro da faixa adequada ao método, sem que o número de ciclos fosse atingido. O valor 0,0001 foi o que permitiu definir um erro final que não interferisse na determinação das classes.

Depois que todos os testes iniciais foram efetuados, foi definido um conjunto de parâmetros para utilização na implementação da rede neural artificial para os testes de seu desempenho na definição da qualidade do café, a saber:

- Seis valores de entrada, um de cada variável associada aos valores das análises químicas.
- Duas camadas intermediárias, a primeira com 5, 7 ou 9 neurônios, e a segunda com 5, 7, 9 ou 13 neurônios.
- Camada de saída com um neurônio associado ao valor da análise sensorial
- *Função logsig*, utilizada como função de ativação das camadas intermediárias e *purelin* para última camada
- Algoritmo *trainlm*, utilizado como algoritmo de aprendizado;
- o número de ciclos (*epochs*) utilizados foi 300 para cada execução.
- A precisão do erro quadrático médio para critério de parada foi de 0,0001.
- Foram utilizadas 20 execuções para cada configuração de rede, selecionando a configuração que apresentou melhor desempenho.

### **3.4. Análise dos Dados**

A análise dos dados foi feita com a finalidade de verificar a interferência da variação dos dados relativos aos valores das análises químicas e aos valores da análise sensorial do café. Essas análises foram realizadas por meio da utilização de rede neural artificial e confrontada com a análise de regressão linear.

## 4. RESULTADOS E DISCUSSÕES

### 4.1. Regressão linear múltipla *versus* Rede Neural Artificial

Para entendimento dos parâmetros necessários à comparação dos métodos de regressão linear múltipla e rede neural artificial, foi feita, inicialmente, uma simulação teórica com utilização de duas equações, envolvendo seis variáveis independentes e uma dependente. Foram utilizadas seis variáveis independentes porque as análises químicas foram em número de seis e a pretensão era obter um resultado que pudesse ser comparado com o resultado obtido com os dados experimentais.

A primeira equação a ser utilizada, Equação 4.1, é polinomial de grau 1, em que se tem variação simples dos coeficientes. A segunda equação, Equação 4.2, envolve logaritmo na base 10 e polinômio de expoente positivo (+1, +2 e +3), negativo (-1) e fracionário (0,5). O objetivo da utilização dessa segunda equação foi estabelecer um modelo um pouco mais complexo para verificação de desempenho dos dois métodos.

$$Y = 65 + \frac{X_1}{100} - \frac{X_2}{10} + \frac{2 + X_3}{2} - X_4 + \frac{5 X_5}{2} - 7 X_6 \quad (4.1)$$

Os valores das variáveis independentes foram ajustados para que os valores da variável dependente possam ter comportamento próximo ao da análise sensorial. Os valores de Y variam entre 31 e 92, enquanto os valores da análise sensorial, entre 44 e 82. As variáveis independentes utilizadas nessas equações são hipotéticas e não têm correspondência com as variáveis experimentais.

Valores das variáveis para Equação 4.1:

$X_1 = (100; 303; 506; 709; 912);$

$X_2 = (21; 48; 75; 102; 129);$

$$X_3 = (5,0; 7,3; 9,6; 11,9; 14,2);$$

$$X_4 = (1,9; 3,6; 5,3; 7,0; 8,7);$$

$$X_5 = (0,7; 1,6; 2,5; 3,4; 4,3); \text{ e}$$

$$X_6 = (0,07; 0,20; 0,33; 0,46; 0,59).$$

Os valores de Y, Equação 4.1, resultaram na variação ilustrada na Figura 4.1. O menor valor foi 45, o maior 88, a média dos valores foi 66,5 e o desvio padrão 6,56.

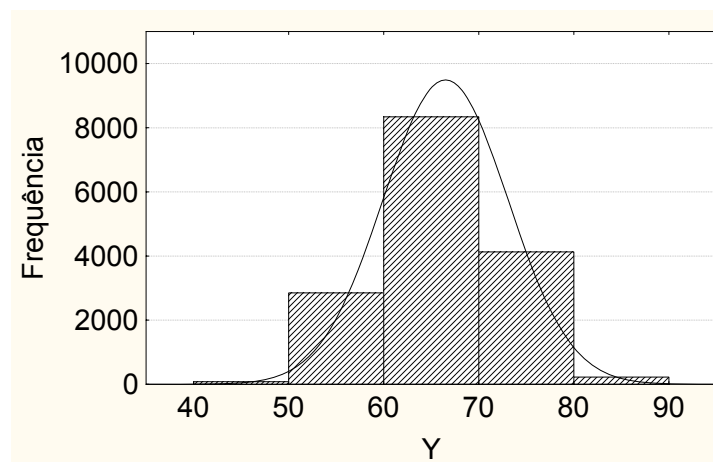


Figura 4.1 - Histograma dos valores de Y da Equação 4.1.

$$Y = 60 + \log X_1 + \frac{1}{X_2} - X_3^2 + X_4 + X_5^{0,5} - X_6^3 \quad (4.2)$$

Valores das variáveis para Equação 4.2:

$$X_1 = (200; 20000; 20000; 200000; 2000000);$$

$$X_2 = (2; 10,5; 0,25; 0,125);$$

$$X_3 = (0,1; 1; 2; 2,5; 3,5);$$

$$X_4 = (1,9; 3,6; 5,3; 7,0; 8,7);$$

$$X_5 = (2; 4; 15; 25; 39); \text{ e}$$

$$X_6 = (0,5; 1,2; 1,6; 2; 2,5).$$

Nos valores de Y, Equação 4.2, as variações contidas na Figura 4.2 foram observadas. O menor valor foi 38, o maior 89, a média dos valores foi 65,49 e o desvio padrão, 8,2578.

As duas equações foram utilizadas para gerar os conjuntos de dados para serem analisados. Para cada variável, assumiram-se cinco valores distintos. Combinando cada valor, foram gerados 15.625 conjuntos de dados, contendo os valores das variáveis independentes e o resultado da aplicação da equação (variável dependente), para cada equação. Para aplicação da rede neural artificial, os dados foram agrupados em classes, da mesma forma que se agruparam os valores da análise sensorial (<69, ≥69 e <74, ≥74 e <79 e ≥79).

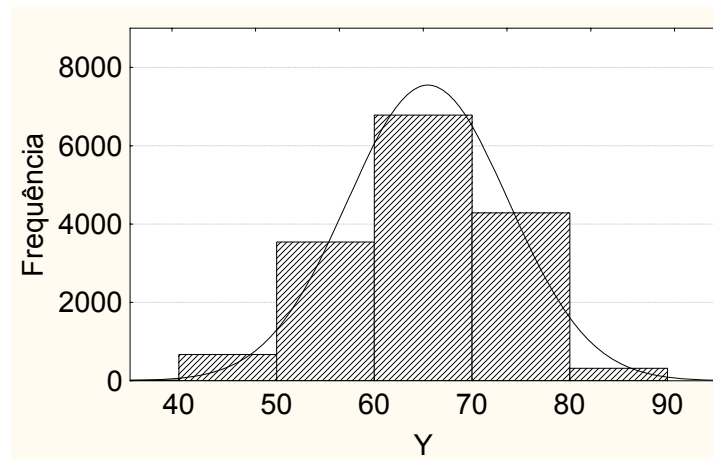


Figura 4.2 - Histograma dos valores de Y da Equação 4.2.

Em uma segunda situação, foi somado ao valor calculado por cada equação um valor aleatório ( $\epsilon$ ), que é uma variável aleatória com média zero e desvio padrão *um* variando no intervalo [-4,4], para simbolizar interferências na execução do experimento (ruído). Dessa forma, têm-se mais dois conjuntos com 15.625 valores.

Na Equação 4.1, sendo adicionado a ela o valor aleatório  $\epsilon$ , observou-se dispersão dos valores conforme a Figura 4.3. O menor valor foi 44, o maior 89, a média dos valores foi 66,49 e o desvio padrão 6,6618.

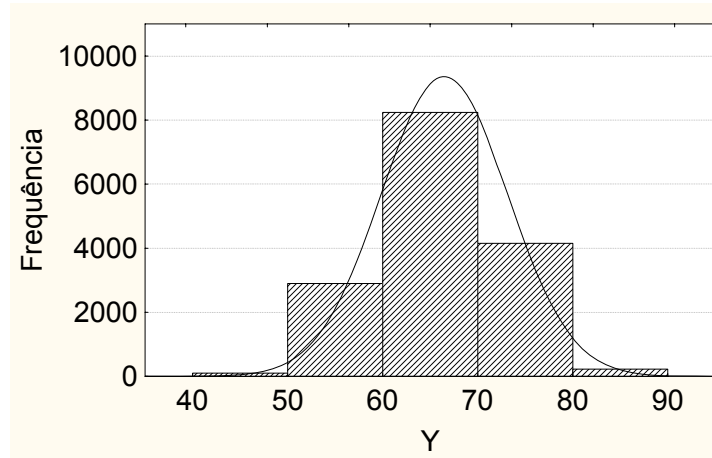


Figura 4.3. Histograma dos valores de Y da Equação 4.1 com adição da variável  $\epsilon$ .

A Equação 4.2, sendo adicionado a ela o valor aleatório ( $\epsilon$ ), apresentou dispersão dos valores como apresentado na Figura 4.4. O menor valor foi 37, o maior 89, a média dos valores foi 65,49 e o desvio padrão 8,3098.

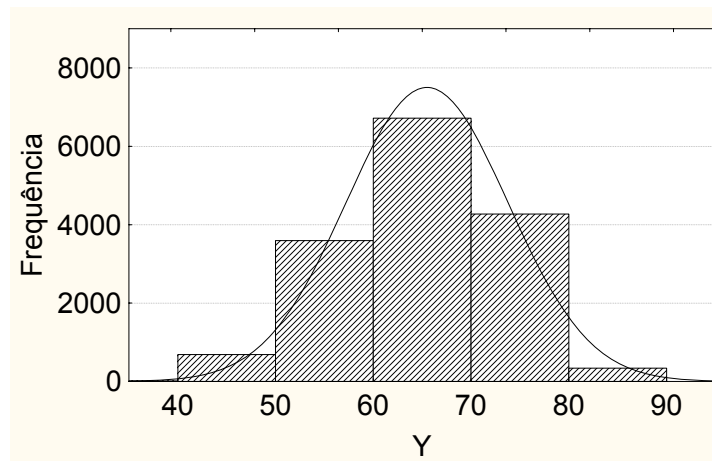


Figura 4.4 - Histograma dos valores de Y da Equação 4.2 com adição da variável  $\epsilon$ .

Na terceira situação, repetiu-se o procedimento anterior, mas somando o valor aleatório multiplicado por dois ( $2\epsilon$ ), representando uma interferência mais severa no experimento, sendo gerados, portanto, mais dois conjuntos com 15.625 valores.

A Equação 4.1, sendo adicionado a ela o valor aleatório multiplicado por dois, resultou na dispersão dos valores como os da Figura 4.5. O menor valor foi 44, o maior 92, a média dos valores foi 66,48 e o desvio padrão 6,8702.

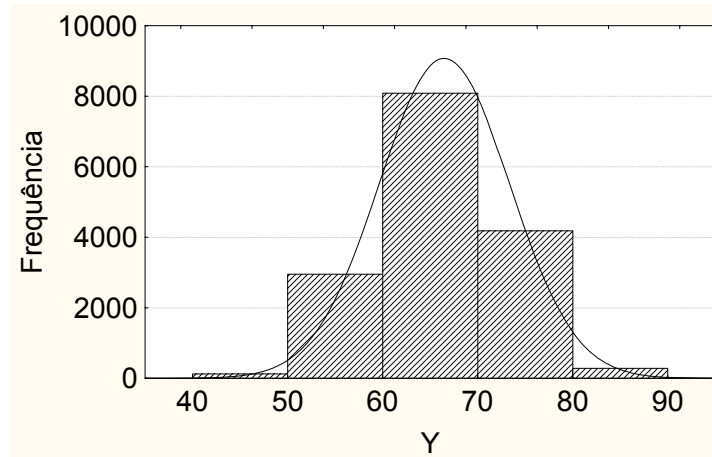


Figura 4.5 - Histograma dos valores de Y da Equação 4.1 adicionada ao valor aleatório multiplicado por dois ( $2 \epsilon$ ).

A Equação 4.2, sendo adicionado a ela o valor aleatório multiplicado por dois ( $2 \epsilon$ ), resultou na dispersão dos valores como os apresentados na Figura 4.6. O menor valor foi 31, o maior 86, a média dos valores foi 61,53 e o desvio padrão 8,4841.

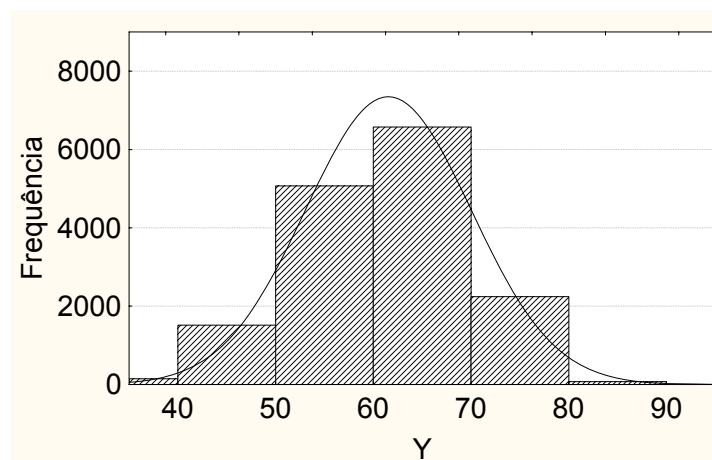


Figura 4.6 - Histograma dos valores de Y da Equação 4.2 adicionada ao valor aleatório multiplicado por dois ( $2 \epsilon$ ).

Aos seis conjuntos de dados foram aplicados os dois métodos de análises, regressão linear múltipla e rede neural artificial.

#### 4.1.1. Equação polinomial

Na execução para os dados gerados pela aplicação da Equação 4.1, obteve-se, com a utilização do método de rede neural artificial, acerto de 100% na classificação e valores calculados pela rede igual aos valores esperados. O número de neurônios nas camadas intermediárias não alterou o resultado, pois em todas as configurações houve acerto de 100% na classificação.

Efetuada a regressão linear múltipla do mesmo conjunto de dados, obteve-se a Equação 4.3 como função de regressão. O valor do coeficiente de determinação,  $R^2$ , foi 0,9999.

$$\hat{Y} = 66 + 0,01 \times X_1 - 0,1 \times X_2 + 0,5 \times X_3 - X_4 + 2,5 \times X_5 - 7 \times X_6 \quad (4.3)$$

O alto índice de acerto dos dois métodos era esperado, pois, além da simplicidade da Equação 4.1, houve coincidência entre os valores e o modelo representado por essa equação.

#### 4.1.2. Equação polinomial com interferência de $\epsilon$

A execução para os dados gerados pela Equação 4.1, adicionada do valor de  $\epsilon$ , resultou, com a utilização do método de rede neural artificial, acerto de 94,43% na classificação, para a configuração de rede (7, 5, 1) de neurônios por camada. Com a variação do número de neurônios nas camadas intermediárias, o menor valor de acerto na classificação foi de 94,11%, para a configuração de rede (5, 9, 1) de neurônios por camada.

A regressão linear múltipla teve como função de regressão a Equação 4.4, e o valor do coeficiente de determinação,  $R^2$ , foi 0,9773.

$$\hat{Y} = 65,9798 + 0,0100175 \times X_1 - 0,100113 \times X_2 + 0,500816 \times X_3 - 1,00308 \times X_4 + 2,51286 \times X_5 - 7,0288 \times X_6 \quad (4.4)$$



Ocorreu redução mais acentuada no resultado apresentado pela rede neural artificial que na regressão linear múltipla, indicando que, nesse caso, o método de rede neural artificial é mais sensível às interferências nos valores experimentais, porém os dois métodos têm desempenho satisfatório.

#### **4.1.3. Equação polinomial com interferência de $2 \varepsilon$**

A execução para os dados gerados pela Equação 4.1 adicionada do valor de  $2 \varepsilon$  teve, com a utilização do método de rede neural artificial, acerto de 89,00% na classificação, para a configuração de rede (7, 13, 1) de neurônios por camada. Com a variação do número de neurônios nas camadas intermediárias, o menor valor de acerto na classificação foi de 88,68%, para as configurações de rede (5, 7, 1), (5, 13, 1) e (7, 9, 1) de neurônios por camada.

A regressão linear múltipla teve como função de regressão a Equação 4.5. O valor do coeficiente de determinação,  $R^2$ , foi 0,9143.

$$\hat{Y} = 66,0342 + 0,0100147 \times X_1 - 0,10048 \times X_2 + 0,497407 \times X_3 - 0,99575 \times X_4 + 2,49374 \times X_5 - 7,01633 \times X_6 \quad (4.5)$$

A interferência de  $2 \varepsilon$  no experimento implicou redução de 11% (de 100% para 89%) de acerto no método de rede neural artificial. Na regressão linear múltipla, a redução do  $R^2$  foi de 8,56% (de 0,9999 para 0,9143). Para simulação dos valores da Equação 4.1, os dois métodos apresentaram desempenhos equiparados para utilização.

#### **4.1.4. Equação mista**

A execução para os dados gerados pela Equação 4.2 teve, com a utilização do método de rede neural artificial, acerto de 98,08% na classificação, para as configurações de rede (7, 13, 1) e (9, 9, 1) de neurônios por camada. Com a variação do número de neurônios nas

camadas intermediárias, o menor valor de acerto na classificação foi de 96,23%, para a configuração de rede (5, 9, 1) de neurônios por camada.

A regressão linear múltipla teve, como função de regressão, a Equação 4.6. O valor do coeficiente de determinação,  $R^2$ , foi 0,8584.

$$\hat{Y} = 78,5234 + 0,0000012352 \times X_1 - 3,01612 \times X_2 - 3,55682 \times X_3 + 1,00196 \times X_4 + 0,129419 \times X_5 - 7,55434 \times X_6 \quad (4.6)$$

O desempenho do método de rede neural artificial foi melhor quando aplicado aos dados decorrentes da aplicação da Equação 4.2, apresentando índice de acerto de 98,08%. O desempenho da regressão linear múltipla, aplicada ao mesmo conjunto de dados, foi inferior, com coeficiente de determinação,  $R^2$ , de 0,8584, indicando que 85,84% dos resultados podem ser explicados pela regressão linear múltipla.

#### 4.1.5. Equação mista com interferência de $\epsilon$

A execução para os dados gerados pela Equação 4.2 adicionada do valor  $\epsilon$  apresentou, com a utilização do método de rede neural artificial, acerto de 94,82% na classificação, para a configuração de rede (9, 13, 1) de neurônios por camada. Com a variação do número de neurônios nas camadas intermediárias, o menor valor de acerto na classificação foi de 93,35%, para a configuração de rede (5, 13, 1) de neurônios por camada.

A aplicação da regressão linear múltipla ao mesmo conjunto de dados teve como função de regressão a Equação 4.7. O valor do coeficiente de determinação,  $R^2$ , foi 0,8458.

$$\hat{Y} = 78,4818 + 0,00000125006 \times X_1 - 3,01346 \times X_2 - 3,54965 \times X_3 + 1,00533 \times X_4 + 0,129618 \times X_5 - 7,55224 \times X_6 \quad (4.7)$$

A interferência de  $\epsilon$  causou pequena redução no desempenho dos dois métodos, indicando que, nos dois casos, o modelo assimila bem as interferências nos experimentos.

#### 4.1.6. Equação mista com interferência de $2 \epsilon$

A execução para os dados gerados pela Equação 4.2 adicionada do valor de  $2 \epsilon$ , teve, com a utilização do método de rede neural artificial, acerto de 96,67% na classificação, para as configurações de rede (5, 5, 1) e (7, 9, 1) de neurônios por camada. Com a variação do número de neurônios nas camadas intermediárias, o menor valor de acerto na classificação foi de 96,29%, para a configuração de rede (9, 9, 1) de neurônios por camada.

A aplicação da regressão linear múltipla ao mesmo conjunto de dados teve como função de regressão a Equação 4.8. O valor do coeficiente de determinação,  $R^2$ , foi 0,8093.

$$\hat{Y} = 74,5849 + 0,00000122589 \times X_1 - 3,00292 \times X_2 - 3,54384 \times X_3 + 0,997552 \times X_4 + 0,129325 \times X_5 - 7,56911 \times X_6 \quad (4.8)$$

A interferência de  $2 \epsilon$  influenciou mais o método de regressão linear múltipla, com redução de 5,72% (de 0,8584 para 0,8093) no valor de  $R^2$ . O método da rede neural artificial apresentou redução de 1,82% (de 98,08 para 96,29).

Analisando os resultados, pôde-se verificar que, no caso de um experimento em que um modelo de equação como a Equação 4.1 para representação dos valores, os métodos de regressão linear múltipla e de rede neural artificial podem ser igualmente utilizados.

Em um experimento cujo modelo pode ser representado por uma equação equiparada à Equação 4.2, envolvendo logaritmo e polinômio com potências diversas, o modelo de regressão linear tem desempenho pior que o método de rede neural artificial. O uso do método de rede neural artificial permitiu assimilar um modelo mais complexo e filtrar a interferência nos dados desse modelo.

No caso de modelagem de eventos biológicos, a complexidade do modelo é maior e envolve, geralmente, modelos exponenciais. Nesse

caso, é de se esperar que os modelos de rede neurais artificiais tenham desempenho melhor que o desempenho da regressão linear múltipla.

Esses resultados de desempenho do método de rede neural artificial, superiores aos obtidos por meio de regressão linear múltipla, são compatíveis com os apresentados por Ciurana et al. (2008), Sousa et al. (2007) e Verlinden, et al. (2008), que demonstraram, em outras aplicações, melhor utilização do método de rede neural artificial.

## 4.2. Regressão Linear Simples

O desempenho do método de rede neural artificial foi, inicialmente, comparado à regressão linear simples. Foi realizada análise individual de interferência da alteração nos valores da análise sensorial com os valores das análises químicas.

A análise sensorial dos 413 conjuntos de dados utilizados resultou na dispersão de valores, como os apresentados no gráfico de frequência da Figura 4.7.

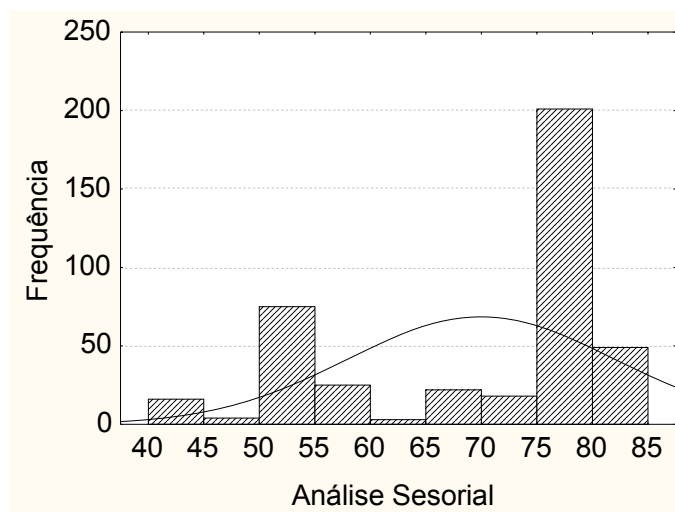


Figura 4.7 - Histograma dos resultados de análise sensorial.

Os resultados da análise sensorial resultaram em menor valor igual a 44, maior valor igual a 82, média dos valores igual a 69,98 e desvio padrão igual a 12,01.

#### 4.2.1. Açúcar Redutor

A Figura 4.8 representa a frequência dos resultados da análise de açúcar redutor e a linha de distribuição normal.

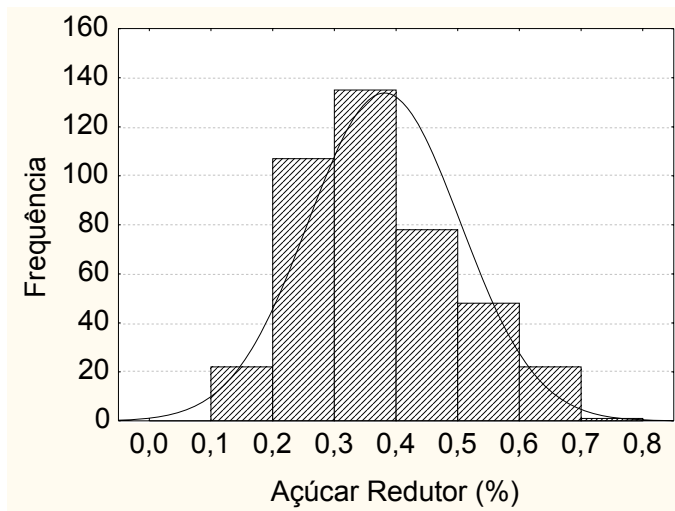


Figura 4.8 - Histograma dos resultados da análise de açúcar redutor.

Foram observados os seguintes valores dos resultados da análise de açúcar redutor: menor valor igual a 0,15%, maior valor igual a 0,71%, média dos valores igual a 0,3809% e desvio padrão igual a 0,1232%.

A regressão linear dos resultados de análise sensorial ( $A_{Se}$ ) em função de açúcar redutor ( $A_{Rd}$ ), nos 413 conjuntos de dados experimentais, resultou na Equação 4.9, cujo coeficiente de determinação,  $r^2$ , foi de 0,1396. Na Figura 4.9, apresentam-se os resultados da regressão linear.

$$\hat{A}_{Se} = 83,8569 - 36,4338 \times A_{Rd} \quad (4.9)$$

A utilização do método de rede neural artificial aplicado aos resultados da análise de açúcar redutor e análise sensorial resultou em melhor acerto de classificação o valor de 43,90% para a configuração de (9, 9, 1) neurônios por camada.

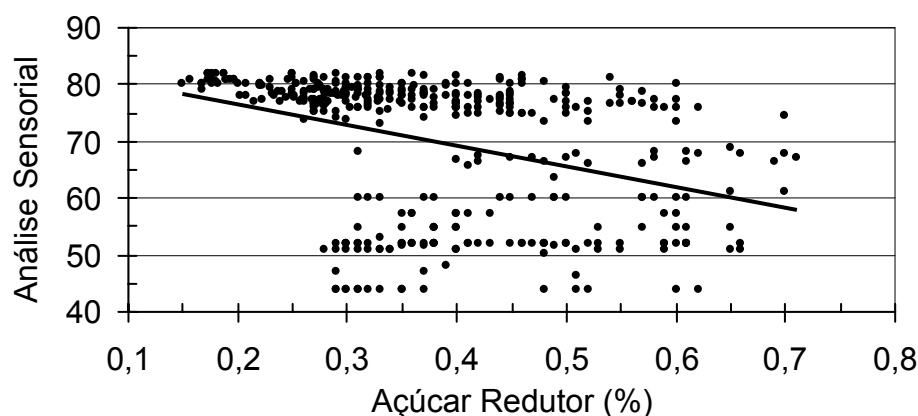


Figura 4.9 - Regressão linear da análise sensorial em função dos valores da análise de açúcar redutor.

A aplicação do método de rede neural artificial teve melhor desempenho que a regressão linear, uma vez que o coeficiente de determinação indica que somente 13,96% dos resultados de análise sensorial são explicados pela regressão. O desempenho do método, apesar de indicar um índice de acerto pequeno, abaixo da metade dos casos, mesmo assim foi melhor que a regressão linear.

#### 4.2.2. Açúcar não redutor

Na Figura 4.10, observam-se a frequência dos resultados da análise do açúcar não redutor e a linha de distribuição normal.

Foram observados os seguintes valores dos resultados da análise de açúcar não redutor: menor valor igual a 1,95%, maior valor igual a 10,38%, média dos valores igual a 6,5125% e desvio padrão igual a 1,5987%.

A regressão linear dos resultados de análise sensorial ( $A_{Se}$ ) em função do açúcar não redutor ( $A_{nR}$ ), para os 413 conjuntos de dados experimentais resultou na Equação 4.10, cujo coeficiente de determinação,  $r^2$ , foi de 0,0865. Na Figura 4.11, observam-se os resultados da regressão linear.

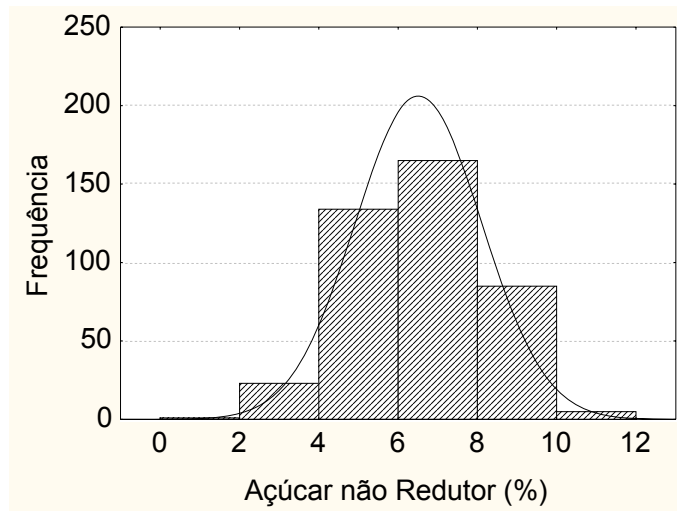


Figura 4.10 - Histograma dos resultados da análise de açúcar não redutor.

$$\hat{A}_{Se} = 55,5830 + 2,21072 \times A_{nR} \quad (4.10)$$

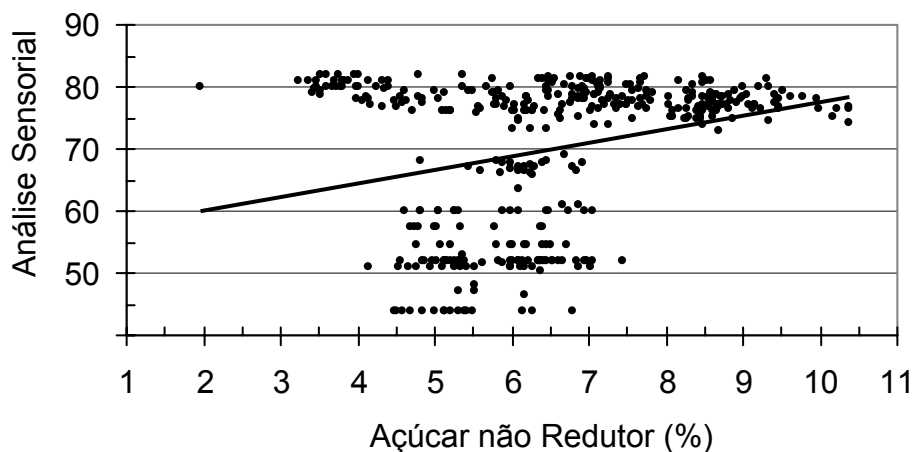


Figura 4.11 - Regressão linear da análise sensorial em função dos valores da análise de açúcar não redutor.

A utilização do método de rede neural artificial aplicado aos resultados da análise de açúcar não redutor e da análise sensorial resultou em melhor acerto de classificação o valor de 41,46% nas configurações de (5, 7, 1), (5, 13, 1) e (9, 7, 1) neurônios por camada.

Nesse caso, também a aplicação do método de rede neural artificial teve melhor desempenho que a regressão linear, uma vez que o

coeficiente de determinação indica que somente 8,65% dos resultados de análise sensorial são explicados pela regressão.

### 4.2.3. Acidez Titulável Total

Na Figura 4.12, observam-se a frequência dos resultados da análise de acidez titulável total e a linha de distribuição normal.

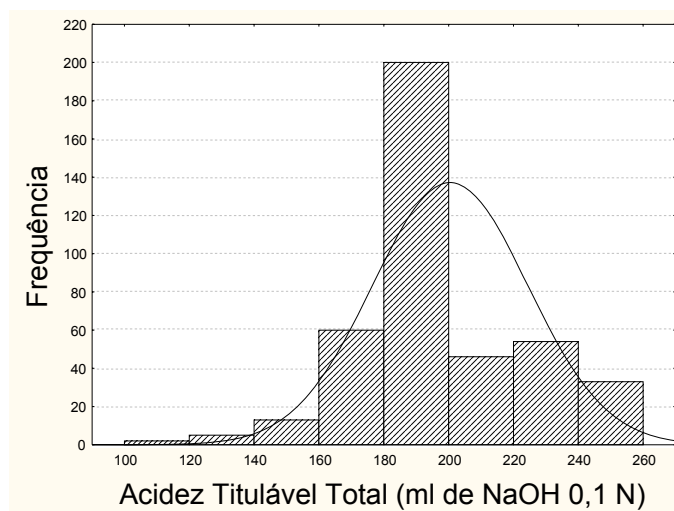


Figura 4.12 - Histograma dos resultados da análise de acidez titulável total.

Foram observados os seguintes valores dos resultados da análise de acidez titulável total: menor valor igual a 115,83 ml de NaOH 0,1 N, maior valor igual a 255 ml de NaOH 0,1 N, média dos valores igual a 200,44 ml de NaOH 0,1 N e desvio padrão igual a 24,02 ml de NaOH 0,1 N.

A regressão linear dos resultados de análise sensorial ( $A_{Se}$ ) em função da acidez titulável total ( $A_{TT}$ ), para os 413 conjuntos de dados experimentais, resultou na Equação 4.11, cujo coeficiente de determinação,  $r^2$ , foi 0,2677. Na Figura 4.13, observa-se o resultado da regressão linear.

$$\hat{A}_{Se} = 121,840 - 0,258718 \times A_{TT} \quad (4.11)$$



A utilização do método de rede neural artificial aplicado aos resultados da análise de acidez titulável total e da análise sensorial teve melhor acerto de classificação o valor de 53,66% nas configurações de (5, 9, 1), (5, 13, 1), (7, 7, 1), (7, 9, 1), (7, 13, 1), (9, 7, 1) e (9, 9, 1) neurônios por camada.

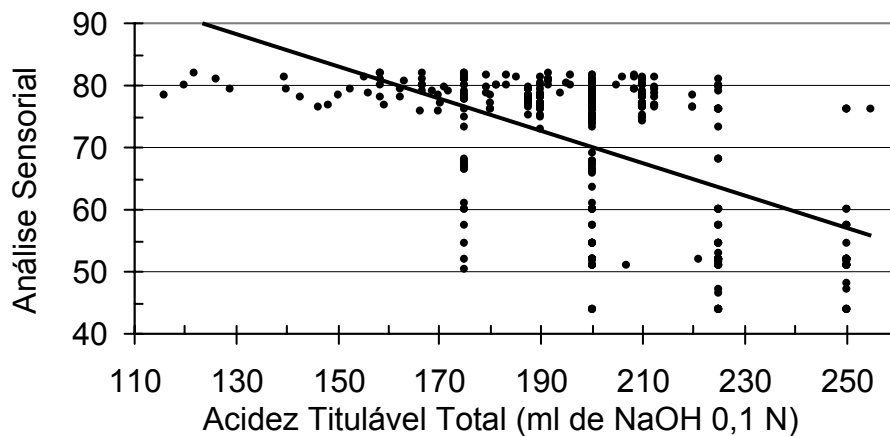


Figura 4.13 - Regressão linear da análise sensorial em função dos valores da análise da acidez titulável total.

A aplicação do método de rede neural artificial teve melhor desempenho, com acerto de 53,66%, uma vez que o coeficiente de determinação da regressão linear indicou que 26,77% dos resultados de análise sensorial foram explicados pela regressão.

#### 4.2.4. Condutividade Elétrica

Na Figura 4.14, observam-se a frequência dos resultados da análise da condutividade elétrica e a linha de distribuição normal.

Foram verificados os seguintes valores dos resultados da análise de condutividade elétrica: menor valor igual a  $36,76 \mu\text{S cm}^{-1} \text{g}^{-1}$ , maior valor igual a  $297 \mu\text{S cm}^{-1} \text{g}^{-1}$ , média dos valores igual a  $155,9102 \mu\text{S cm}^{-1} \text{g}^{-1}$  e desvio padrão igual a  $49,8011 \mu\text{S cm}^{-1} \text{g}^{-1}$ .

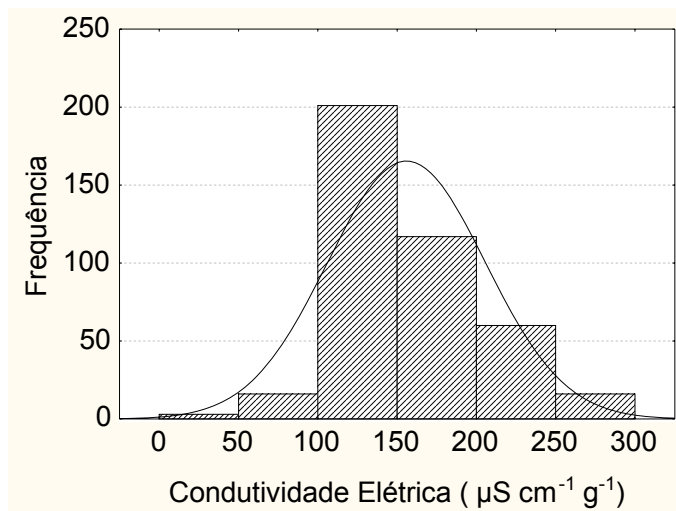


Figura 4.14 - Histograma dos resultados da análise de condutividade elétrica.

A regressão linear dos resultados de análise sensorial ( $A_{Se}$ ) em função da condutividade elétrica ( $C_{EI}$ ), para os 413 conjuntos de dados experimentais, resultou na Equação 4.12, cujo coeficiente de determinação,  $r^2$ , foi 0,5412. A Figura 4.15 contém os resultados da regressão linear.

$$\hat{A}_{Se} = 97,6475 - 0,177456 \times C_{EI} \quad (4.12)$$

A utilização do método de rede neural artificial aplicado aos resultados da análise de condutividade elétrica e da análise sensorial teve como resultado de melhor acerto de classificação o valor de 68,29% para a configuração de (5, 7, 1) neurônios por camada.

O acerto pela aplicação do método de rede neural artificial, de 68,29%, foi melhor que pela regressão linear, uma vez que o coeficiente de determinação indicou que 54,12% dos resultados da análise sensorial foram explicados pela regressão. Esses dois índices indicam que os valores da análise sensorial se correlacionam melhor com os valores da análise de condutividade elétrica.

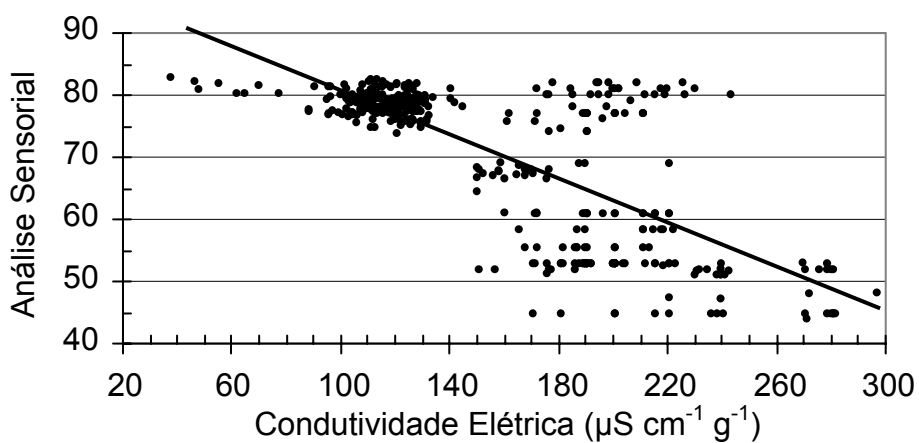


Figura 4.15 - Regressão linear da análise sensorial em função dos valores da análise de condutividade elétrica.

#### 4.2.5. Lixiviação de Potássio

Na Figura 4.16, representam-se a frequência dos resultados da análise de lixiviação de potássio e a linha de distribuição normal.

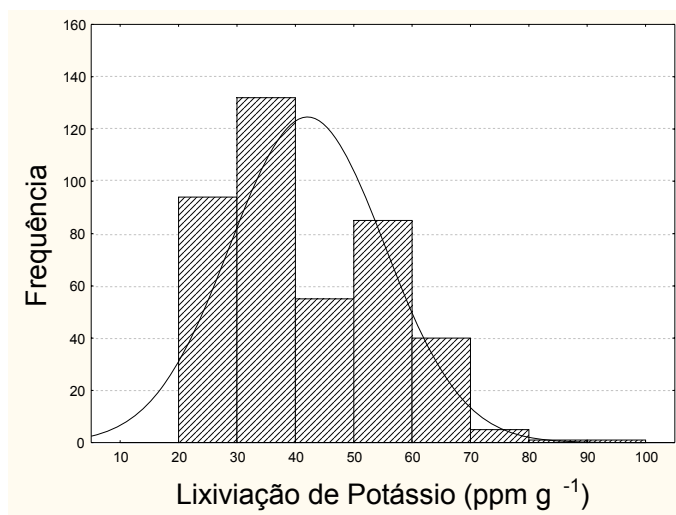


Figura 4.16 - Histograma dos resultados da análise de lixiviação de potássio.

Foram observados os seguintes valores dos resultados da análise de lixiviação de potássio: menor valor igual a 24,18 ppm g<sup>-1</sup>, maior valor

igual a 91,71 ppm g<sup>-1</sup>, média dos valores igual a 42,0599 ppm g<sup>-1</sup> e desvio padrão igual a 13,2269 ppm g<sup>-1</sup>.

A regressão linear dos resultados de análise sensorial ( $A_{Se}$ ) em função da lixiviação de potássio ( $L_{Po}$ ), para os 413 conjuntos de dados experimentais, resultou na Equação 4.13, cujo coeficiente de determinação,  $r^2$ , foi de 0,2906. Na Figura 4.17, observa-se o resultado dessa regressão linear.

$$\hat{A}_{Se} = 90,5751 - 0,489655 \times L_{Po} \quad (4.13)$$

A utilização do método de rede neural artificial aplicado aos resultados da análise de lixiviação de potássio e da análise sensorial teve como resultado de melhor acerto de classificação o valor de 51,22% para a configuração de (5, 5, 1) neurônios por camada.

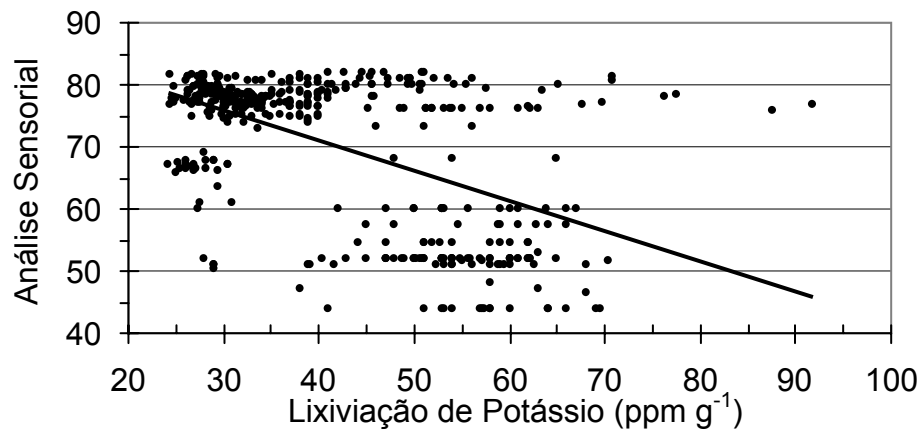


Figura 4.17 - Regressão linear da análise sensorial em função dos valores da análise de lixiviação de potássio.

O acerto de 51,22%, obtido pela utilização do método de rede neural artificial, foi melhor que o resultado com a regressão linear, uma vez que o coeficiente de determinação indicou que somente 29,06% dos resultados da análise sensorial foram explicados pela regressão.

#### 4.2.6. Polifenóis

Na Figura 4.18, observam-se a frequência dos resultados da análise de polifenóis e a linha de distribuição normal.

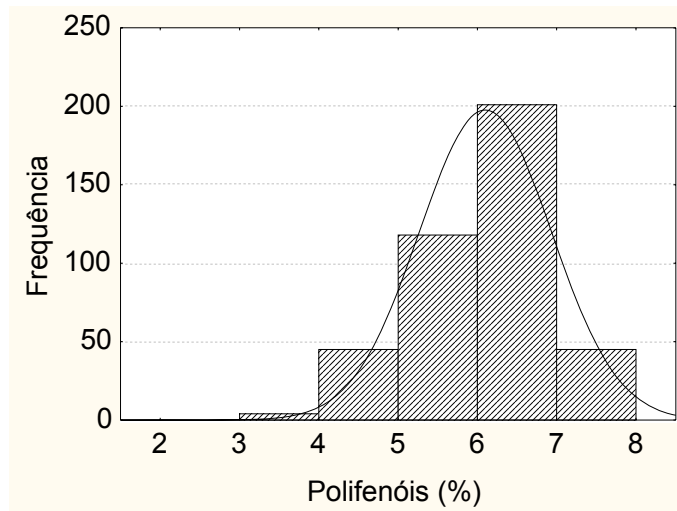


Figura 4.18 - Histograma dos resultados da análise de polifenóis.

Foram observados os seguintes valores dos resultados da análise de polifenóis: tiveram o menor valor igual a 3,42%, o maior valor igual a 7,9%, a média dos valores igual a 6,10% e o desvio padrão igual a 0,83%.

A regressão linear dos resultados da análise sensorial ( $A_{Se}$ ) em função de polifenóis ( $P_{ol}$ ), para os 413 conjuntos de dados experimentais, resultou na Equação 4.14, cujo coeficiente de determinação,  $r^2$ , foi de 0,1373. A Figura 4.19 contém os resultados dessa regressão linear.

$$\hat{A}_{Se} = 102,556 - 5,33970 \times P_{ol} \quad (4.14)$$

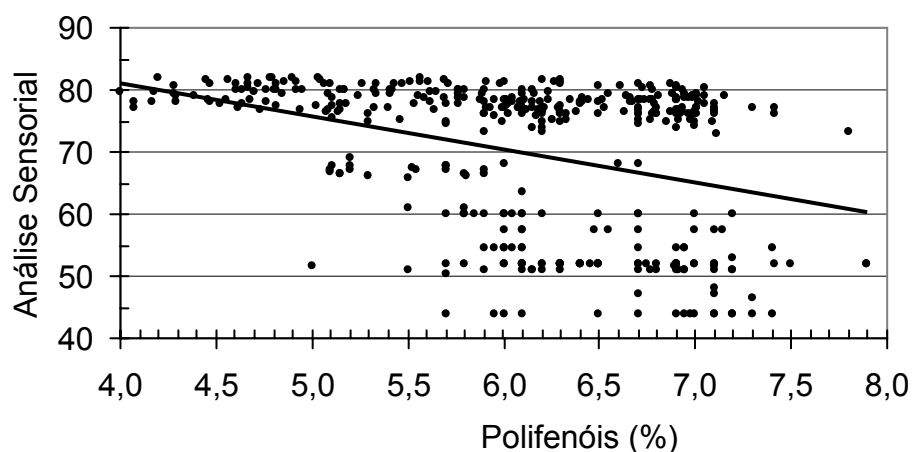


Figura 4.19 - Regressão linear da análise sensorial em função dos valores da análise de polifenóis.

A utilização do método de rede neural artificial aplicado aos resultados da análise de polifenóis e da análise sensorial teve como resultado de melhor acerto de classificação o valor de 39,02% para a configuração de (5, 5, 1) neurônios por camada.

O acerto de 39,02%, obtido pela aplicação do método de rede neural artificial, é melhor que o resultado da regressão linear, uma vez que o coeficiente de determinação indica que somente 13,73% dos resultados de análise sensorial são explicados pela regressão.

O resultado do resíduo, constante no Apêndice 1, e do coeficiente de determinação da análise de regressão linear simples dos valores da análise sensorial em função dos valores de cada análise química utilizada de forma individual reforça a busca por modelos que possam apresentar melhor desempenho na explicação das variações nos dados experimentais.

### 4.3. Análise Sensorial em Função dos resultados de Acidez Titulável Total, Condutividade Elétrica e Lixiviação de Potássio

A tentativa de correlacionar, de forma isolada, os valores da análise sensorial aos das análises químicas não teve resultados adequados. Portanto, verificou-se a necessidade de trabalhar com os valores das análises de forma conjunta.

A utilização do método de rede neural artificial aplicado aos resultados das análises de acidez titulável total, da condutividade elétrica e da lixiviação de potássio, de forma conjunta, proporcionou acerto de classificação, para análises sensoriais, com valores acima de 50%. Então, fez-se a aplicação do método de rede neural artificial envolvendo os valores resultantes dessas três análises, assim como uma análise de regressão linear múltipla para verificação de desempenho dos dois métodos nesse tipo de associação de dados.

A regressão linear múltipla desses dados resultou na Equação 4.15, cujo coeficiente de determinação,  $R^2$ , foi 0,5736.

$$\hat{A}_{Se} = 114,806 - 0,0953072 \times A_{TT} - 0,131843 \times C_{EI} - 0,122815 \times L_{Po} \quad (4.15)$$

Esse desempenho da regressão linear múltipla não indicou melhoria significativa na composição das três análises, uma vez que o valor do coeficiente de determinação,  $R^2$ , para a análise de regressão dos valores da análise sensorial em função da análise de condutividade elétrica foi 0,5412.

A utilização do método de rede neural artificial aplicado aos resultados das análises de acidez titulável total, da condutividade elétrica, da lixiviação de potássio e da análise sensorial proporcionou acerto de classificação de 75,61% para as configurações de (5, 7, 1), (5, 13, 1), (7, 5, 1), (7, 13, 1) e (9, 9, 1) neurônios por camada. Esse resultado, de forma contrária ao da regressão linear múltipla, indica que houve melhoria na

composição das três análises. Isso porque, de forma isolada, a aplicação do método de rede neural artificial, de melhor desempenho, foi a rede obtida com os valores da análise de condutividade elétrica, com acerto de 68,29%.

#### **4.4. Análise Sensorial em Função de Todas as Análises Químicas**

A correspondência entre os resultados da análise sensorial e os resultados das análises químicas foi verificada com uma configuração, envolvendo todas as análises químicas e a análise sensorial.

A regressão linear múltipla dos resultados da análise sensorial ( $A_{Se}$ ) em função das análises de açúcares redutor ( $A_{Rd}$ ) e não redutor ( $A_{nR}$ ), de acidez titulável total ( $A_{TT}$ ), da condutividade elétrica ( $C_{el}$ ), da lixiviação de potássio ( $L_{Po}$ ) e de polifenóis ( $P_{oi}$ ), para os 413 conjuntos de dados experimentais, resultou na Equação 4.16, cujo coeficiente de determinação,  $R^2$ , foi de 0,6681.

$$\begin{aligned} \hat{A}_{Se} = & 121,553 - 28,0864 \times A_{Rd} + 0,707124 \times A_{nR} - 1,71888 \times P_{oi} \\ & - 0,0631717 \times A_{TT} - 0,127175 \times C_{El} - 0,0595256 \times L_{Po} \end{aligned} \quad (4.16)$$

A utilização do método de rede neural artificial aplicado aos resultados de todas as análises químicas e análise sensorial resultou em um acerto de classificação de 85,37%, para a configuração de (5, 9, 1) neurônios por camada. Nesse caso foram utilizados os 413 conjuntos de dados, sendo 331 conjuntos utilizados no treinamento, 41 na validação durante o treinamento e 41 nos testes após o treinamento. O resultado de acerto de 85,37% nos testes indica que, dos 41 conjuntos de dados de teste, houve acerto da classe, definida pela análise sensorial, em 35 e erro em 6 conjuntos de dados.

Nessa configuração, comparada com a anterior (item 4.3), que teve melhor desempenho naquele teste, a melhora na regressão linear múltipla



foi expressiva, pois o coeficiente de determinação,  $R^2$ , passou de 0,5736 para 0,6681.

O resultado de classificação (85,37%) pelo uso do método de rede neural artificial foi superior ao valor de 75,61% da execução anterior. Esse resultado indica que houve correspondência entre a variação dos valores da análise sensorial com a variação dos valores das análises químicas. Segundo o modelo de rede neural artificial estabelecido, no caso em estudo a classificação do tipo de bebida do café é explicada em mais de 85% pela variação dos valores das análises químicas quando analisados em conjunto.

#### **4.5. Teste da Rede Neural Utilizando Dados de Campo**

Na aplicação do método de rede neural artificial, utilizaram-se dados de treinamento e de testes provenientes dos procedimentos da colheita, secagem e armazenamento controlados, possibilitando a não inserção de variáveis adicionais. Um teste a ser efetuado é verificar o desempenho do método de rede neural artificial utilizando amostra de dados provenientes dos procedimentos de colheita, secagem e armazenamento não controlados.

Foram obtidas amostras de cafés provenientes de várias propriedades da região Sul de Minas Gerais, no Município de Machado. Essas amostras foram submetidas às mesmas análises, tendo sido obtidos 86 conjuntos de dados. A Tabela 4.1 contém as informações estatísticas desses dados e, a Figura 4.20, a distribuição deles.

Com a utilização da rede neural artificial, obtida com o treinamento com os 413 conjuntos de dados, que resultou em acerto 85,37%, foi feita a simulação com os 86 conjuntos de dados, obtendo-se classificação correta de 75,58%.

Foram gerados os valores da análise sensorial, utilizando-se a equação de regressão linear múltipla, Equação 4.16, e, em seguida, esses valores foram transformados em classes, utilizadas pela rede

neural artificial. Esses valores coincidiram em 18,60% com os valores da análise sensorial da amostra de teste.

Tabela 4.1 - Estatística dos dados da amostra de teste

	$A_{Rd}^*$	$A_{nR}^*$	$P_{ol}^*$	$A_{TT}^*$	$C_{EI}^*$	$L_{Po}^*$	$A_{Se}^*$
Menor	0,305	4,935	4,950	170,000	72,584	22,421	43,250
Maior	0,735	6,983	6,750	250,000	156,257	56,839	83,167
Média	0,519	5,970	5,756	203,000	106,239	36,861	76,076
Desvio padrão	0,101	0,527	0,387	16,367	16,452	7,348	9,289

\*  $A_{Rd}$  – Açúcar redutor (%)

$A_{nR}$  – Açúcar não redutor (%)

$P_{ol}$  – Polifenóis (%)

$A_{TT}$  – Acidez titulável total (ml de NaOH 0,1 N)

$C_{EI}$  – Condutividade elétrica ( $\mu S\ cm^{-1}\ g^{-1}$ )

$L_{Po}$  – Lixiviação de potássio (ppm  $g^{-1}$ )

$A_{Se}$  – Análise sensorial

Com a utilização do método de rede neural artificial, obteve-se melhor resultado de classificação do tipo da bebida do café do que o apresentado pela regressão linear múltipla. Todavia, esse resultado foi inferior ao observado quando se utilizou o método de rede neural artificial aplicada aos dados experimentais, em condições controladas. Isso indica que o método é sensível aos dados para os quais a rede neural artificial foi treinada, o que é uma característica positiva, pois demonstra que há diferença nos valores das análises químicas e da análise sensorial utilizadas para gerar a rede neural artificial. Para se ter o desempenho desejado do método de rede neural artificial, a rede deverá ser treinada com dados apropriados para sua utilização posterior, ou seja, deve ser treinada com o mesmo tipo de dados em que ela for utilizada.

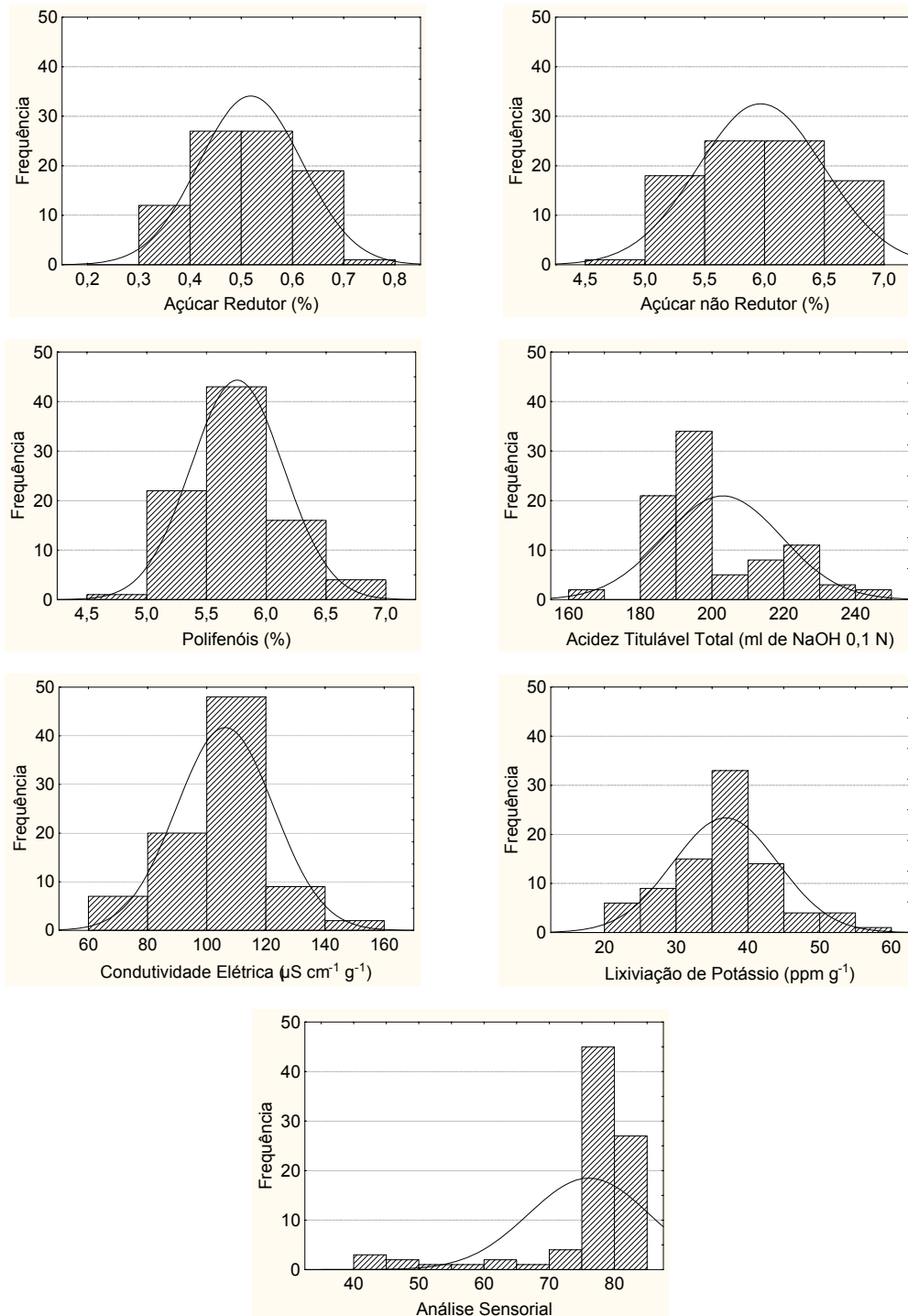


Figura 4.20 - Histograma dos resultados das análises químicas e análise sensorial da amostra de campo.

#### 4.6. Discussão dos resultados

Os resultados de desempenho, de acerto na classificação do tipo da bebida do café, da aplicação do método de rede neural artificial (RNA) para relacionar os valores da análise sensorial e os valores das análises químicas de amostras de café em que os procedimentos pós-colheita foram controlados em comparação aos não controlados, 85,37% e 75,58%, respectivamente, nos permite concluir que esse método é sensível ao tipo de amostra utilizado. Pode-se então obter diferentes redes neurais artificiais treinadas com resultados de análises sensorial e químicas, de diferentes tipos de amostras de café. Quando o teste de classificação de uma determinada amostra for efetuado, o índice de acerto na utilização da RNA será melhor para a rede que foi treinada para aquele tipo de amostra. Conclui-se com isso que a utilização do método de RNA é sensível à amostra do material utilizada para seu treinamento.

A tabela 4.2 contém os resultados de desempenho da aplicação do método de RNA e o método de regressão linear (RL). Há relação direta entre os resultados dos desempenhos da aplicação desses métodos no estabelecimento da correspondência entre a variação dos valores da análise sensorial e os valores das análises químicas. Observou-se que os melhores desempenhos da aplicação do método de RNA ocorreram onde também houve os melhores desempenhos da aplicação do método de RL.

A análise dos dados, quando se tem um número variado de fatores condicionantes, utilizando regressão linear simples (RLS), não se mostrou eficiente. A análise, utilizando o método de RNA, apresentou melhor desempenho que a RLS. A análise conjunta, utilizando os resultados de todas as análises químicas das amostras de café, se mostrou mais eficiente. Com a utilização do método de regressão linear múltipla (RLM), conseguiu-se elevar o coeficiente de determinação para o valor de 0,6681 e, no método de RNA, conseguiu-se um índice de acerto de 85,37% na correlação entre os valores da análise sensorial e os valores resultantes das análises químicas.

Tabela 4.2 - Comparativo de desempenho da RL e RNA.

Análises químicas utilizadas	Coefficiente de Determinação (RL)	Acerto na RNA
Polifenóis	0,1373	39,02%
Açúcar não Redutor	0,0865	41,46%
Açúcar Redutor	0,1396	43,90%
Lixiviação de Potássio	0,2906	51,22%
Acidez Titulável Total	0,2677	53,66%
Condutividade Elétrica	0,5412	68,29%
Acidez Titulável Total, Condutividade Elétrica e Lixiviação de Potássio	0,5736	75,61%
Todas	0,6681	85,37%

O desempenho de 85,37% de acerto, na classificação do tipo da bebida do café, obtido pelo método de RNA, é um forte argumento para afirmar que existe uma relação entre a variação dos resultados de análise sensorial e os resultados das análises químicas de amostras de café. Como o valor do coeficiente de determinação, resultante da aplicação do método de RLM, no mesmo experimento, foi de 0,6681, isso nos permite concluir que o método de RNA apresenta desempenho superior ao método de RNA.

## **5. CONCLUSÃO**

Esse trabalho demonstra o potencial de utilização do método de rede neural artificial na identificação da qualidade da bebida do café arábica, utilizando resultados de análises químicas.

O método de rede neural artificial é capaz de detectar variações em um mesmo tipo de bebida para diferentes amostras de café.

Há relação direta entre os resultados das simulações obtidos pelo método de rede neural artificial e aqueles obtidos por regressão linear.

A aplicação do método de rede neural artificial aos dados experimentais demonstra haver correspondência entre a variação dos resultados da análise sensorial e o resultado das análises químicas efetuadas.

O método de rede neural artificial é mais eficiente que o método de regressão linear múltipla para classificar o tipo de bebida do café em função dos resultados das análises químicas das amostras.

## **6. TRABALHOS FUTUROS**

Desenvolver um sistema de denominação de origem de café, utilizando o método de redes neurais artificiais treinadas com resultados de análises químicas de amostras de café de diferentes regiões produtoras.

Comparar o desempenho do método de redes neurais artificiais, na classificação do tipo de bebida do café, os métodos que utilizem outras técnicas computacionais: rede neural probabilística e rede neuro-fuzzy, entre outras.

## 7. REFERÊNCIAS

AFONSO JÚNIOR, P. C.; CORRÊA, P. C.; GONELI, A. L. D.; SILVA, F. S. Contribuição das etapas do pré-processamento para qualidade do café. **Revista Brasileira de Armazenamento**, Viçosa, Especial Café, n. 8, p. 46-53, 2004.

ANDERSON, K. A.; SMITH, B. W. Chemical Profiling To Differentiate Geographic Growing Origins of Coffee. **Journal of Agricultural and Food Chemistry**, v. 50, p. 2068-2075, 2002. DOI: 10.1021/jf011056v.

BORÉM, F. M.; MARQUES, E. R.; ALVES, E. Ultrastructural analysis of drying damage in parchment Arabica coffee endosperm cells. **Biosystems Engineering**. v. 99, n. 1, p. 62-66, jan. 2008. DOI: 10.1016/j.biosystemseng.2007.09.027.

BORÉM, F. M.; RIBEIRO, D. M.; PEREIRA, R. G. F. A.; ROSA, S. D. V. F.; MORAIS, A. R. Qualidade do café submetido a diferentes temperaturas, fluxos de ar e períodos de pré-secagem. **Coffee Science**, v. 1, n. 1, p. 55-63, Lavras, abr./jun. 2006.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais teoria e aplicações**. Rio de Janeiro, RJ: LTC, 2000. 262 p.

CACCER. **Protocolos para análise sensorial de café, metodologia SCAA**. Fev. 2005. Disponível em: [www.cafedocerrado.org](http://www.cafedocerrado.org). Acesso em 04/set./2009.

CAWLEY, G. C.; JANACEK, G. J.; HAYLOCK, M. R.; DORLING, S. R. Predictive uncertainty in environmental modeling. **Neural Networks**, v. 20, p. 537-549, 2007. DOI:10.1016/j.neunet.2007.04.024.

CHAGAS, S. J. R.; MALTA, M. R.; PEREIRA, R. G. F. A. Potencial da região Sul de Minas Gerais para a produção de cafés especiais. **Ciência e tecnologia**, Lavras, v. 29, n. 3, p. 590-597, MG: maio/jun. 2005.

CHARUMPORN, B.; YOSHIOKA, M.; FUJINAKA, T.; OMATU, S. Identify Household Burning Smell Using an Electronic Nose System with Artificial Neural Networks. In: **Computational Intelligence in Robotics and Automation**. Japan: Kobe, 2003. p. 1070-1074.

CHEN, C.; CHEN, S.; WANG, C.; YANG, I.; HSIAO S.; TSAI, C. Evaluation of nitrogen content in cabbage seedlings using hyper-spectral

images. Sensing and Instrumentation for **Food Quality and Safety**, v. 2, n. 2, p. 97-102, jun. 2008. DOI: 10.1007/s11694-008-9041-5.

CHERKASSKY, V.; MA, Y. Another look at statistical learning theory and regularization. **Neural Networks**, v. 22, p. 958-969, 2009. DOI:10.1016/j.neunet.2009.04.005.

CIURANA, J.; QUINTANA, G.; GARCIA-ROMEU, M. L. Estimating the cost of vertical high-speed machining centers, a comparison between multiple regression analysis and the neural networks approach. **International Journal of Production Economics**, v. 115, 171-178, 2008. DOI: 10.1016/j.ijpe.2008.05.009.

CORADI, P. C.; BORÉM, F. M.; SAATH, R.; MARQUES, E. R. Effect drying and storage conditions on the quality of natural and washed coffee. **Coffee Science**, Lavras, v. 2, n.1, p. 38-47, jan.-jun. 2007.

DU, C.; ZHOU, J. Evaluation of soil fertility using infrared spectroscopy: a review. **Environmental Chemistry Letters**, v. 7, n. 2, p. 97-113, jun. 2009. DOI: 10.1007/s10311-008-0166-x.

DUTTA, R.; KASHWAN, K. R.; BHUYAN, M.; HINES, E. L.; GARDNER J. W. Electronic nose based tea quality standardization. **Neural Networks**, v. 16, p. 847-853, 2003. DOI:10.1016/S0893-6080(03)00092-3.

ERENTUR, K.; ERENTUR, S.; TABIL, L. G. A comparative study for the estimation of dynamical drying behavior of *Echinacea angustifolia*: regression analysis and neural network. **Computers and electronics in agriculture**, v. 45, p. 71-90, 2004.

FERNANDES, M. **Uso de cloreto de benzalcônio na pré-colheita e suas interações com a colheita e pós-colheita na qualidade do café**. 2008. 91 f. Tese (Doutorado em Ciência dos Alimentos) – Universidade Federal de Lavras, Lavras, MG, 2008.

FIGUEIREDO, L. P. **Perfil sensorial, químico e proteômico relacionados ao processamento pós-colheita de cultivares de cafeeiro**. Lavras, MG: UFLA, 2009. (Projeto de Pesquisa).

FIRAT, M.; YURDUSEV, M. A.; TURAN, M. E. Evaluation of Artificial Neural Network Techniques for Municipal Water Consumption Modeling. **Water Resources Management**, v. 23, n. 4, 617-632, mar. 2009. DOI 10.1007/s11269-008-9291-3.



GARRUTI, R. S.; PUPO, L. M.; TEXEIRA, A. A.; PEREIRA, L. S. P. Determinação da bebida de café “riada”. **Coletânea do Instituto Tecnológico de Alimentos**, v. 2, p. 243-249, 1968.

GOULART, P. F. P.; ALVES, J. D.; MALTA, M. R.; MAGALHÃES, M. M.; PEREIRA, R. G. F. A.; MEYER, L. E. Análise comparativa entre lixiviação de potássio, condutividade elétrica, ácido clorogênico e métodos de quantificação da atividade da polifenol oxidase em extrato semipurificado de amostras de café de diferentes padrões de qualidade. 3, 2003, Porto Seguro. In: **Simpósio de pesquisa dos cafés do Brasil e workshop internacional de café & saúde**. Brasília: Embrapa Café, 2003. 447 p. p. 249-250.

GRIVAS, G.; CHALOULAKOU, A. Artificial neural network models for prediction of PM10 hourly concentrations, in the Greater Area of Athens, Greece. **Atmospheric Environment**, v. 40, p. 1216-1229, 2006. DOI: 10.1016/j.atmosenv.2005.10.036.

GUELI, N.; PICCIRILLO G.; TROISIA, G.; CICONETTI, P.; MELONI, F.; ETTORRE, E.; VERICO, P.; D'ARCANGELO, E.; CACCIAFFESTA, M. The influence of lifestyle on cardiovascular risk factors Analysis using a neural network. **Archives of Gerontology and Geriatrics**, v. 40, p. 157-172, 2005. DOI:10.1016/j.archger.2004.08.002.

HAGAN, M. T. **Neural network design**. 2. ed. Boston- MA, EUA: PWS Publishing Company, 2002.

HAYKIN, S. **Redes neurais**. 2. ed. Rio de Janeiro: Bookman Companhia, 2000. 900 p.

HERVÁS-MARTÍNEZ, C.; MARTÍNEZ-ESTUDILLO, F. J.; CARBONERO-RUZ, M. Multilogistic regression by means of evolutionary product-unit neural networks. **Neural Networks**, v. 21, p. 951-961, 2008. DOI: 10.1016/j.neunet.2007.12.052.

ISQUIERDO, E. P. **Secagem e qualidade do café cereja desmucilado submetido a diferentes períodos de repouso**. 2008. 98 f. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Federal de Lavras, Lavras, MG, 2008.

KERMANI, B. G.; SCHIFFMAN, S. S.; NAGLE, H. T. Performance of the Levenberg–Marquardt neural network training method in electronic nose applications. **Sensors and Actuators B Chemical**, v. 110, p. 13-22, 2005. doi:10.1016/j.snb.2005.01.008.

KHOSHGOFTAAR, T. M.; GAO, K. Count Models for Software Quality Estimation. **IEEE Transactions on Reliability**, v. 56, n. 2, jun. 2007. Digital Object Identifier 10.1109/TR.2007.896757.

KORJANI, M. M.; BAZZAZ, O.; MENHAJ, M. B. Real time identification and control of dynamic systems using recurrent neural networks. **Artificial Intelligence Review**, 2009. DOI 10.1007/s10462-009-9111-z.

LINGLE, T. R. **The coffee cupper's handbook: systematic guide to the sensory evaluation of coffee's flavor**. 3. ed. Long Beach – California, EUA: Specialty Coffee Association of America, 2001. 71 p.

LIU, H.; XIE, D.; WU, W. Soil water content forecasting by ANN and SVM hybrid architecture. **Environmental Monitoring and Assessment**, v. 143, n. 1-3, p.187-193, aug. 2008. DOI 10.1007/s10661-007-9967-9.

LU, J.; HUANG, G.; LI, S. A Study of Maneuvering Control for an Air Cushion Vehicle Based on Back Propagation Neural Network. **Journal of Shanghai Jiaotong University (Science)**, v. 14, n. 4, p. 482-485, aug. 2009. DOI: 10.1007/ s12204-009-0482-8.

MALTA, M. R.; PEREIRA, R. G. F. A.; CHAGAS, S. J. R. Condutividade elétrica e lixiviação de potássio do escudado de grãos de café: alguns fatores que podem influenciar essas avaliações. **Ciência e Agrotecnologia**, Lavras, v. 29, n. 5, p. 1015-1020, set./out. 2005.

MAPA - Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa No 8 de 11 de junho de 2003. Aprova o Regulamento Técnico de Identidade e de Qualidade para a Classificação do Café Beneficiado Grão Cru. Ministro Roberto Rodrigues. **Diário Oficial** (da República federativa do Brasil). Brasília, p. 4, 13 de Junho de 2003, Seção 1.

MAZUROWSKI, M. A., HABASA P. A., ZURADAA J. M., LOB J. Y., BAKERB J. A., TOURASSIB G. D. Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. **Neural Networks**, v. 21, p. 427-436, 2008. DOI: 10.1016/j.neunet.2007.12.031.

MENDONÇA, L. M. V. L.; PEREIRA, R. G. F. A.; MENDES, A. N. G.; BORÉM, F. M.; MARQUES, E. R. Composição química de grãos crus de cultivares de *Coffea arabica* L. suscetíveis e resistentes à *Hemileia vastatrix* Berg et Br. **Ciência e Agrotecnologia**, Lavras, v. 31, n. 2, p. 413-419, mar.-abr. 2007.

MITTAL, G. S.; ZHANG, J. Artificial Neural Network-based Psychometric Predictor. **Biosystems Engineering. Information Technology and the Human Interface**, v. 85, n. 3, p. 283-289, 2003. DOI: 10.1016/S1537-5110(03)00071-0.

NOBRE, G. W. **Processamento e qualidade de frutos verdes de café arábica**. 2009. 85 f. Tese (Doutorado em Agronomia) – Universidade Federal de Lavras, Lavras, MG, 2009.

ODAJIMA, K.; HAYASHI, Y.; TIANXIA, G.; SETIONO, R. Greedy rule generation from discrete data and its use in neural network rule extraction. **Neural Networks**, v. 21, p. 1020-1028, 2008. DOI: 10.1016/j.neunet.2008.01.003.

PARDO, M.; SISK, B. C.; SBERVEGLIERI, G.; LEWIS, N. S. Comparison of Fisher's linear discriminant to multilayer perceptron networks in the classification of vapors using sensor array data. **Sensors and Actuators B Chemical**, v. 115, p. 647-655, 2006. DOI: 10.1016/j.snb.2005.10.033.

PARDO, M.; SBERVEGLIERI, G. Remarks on the use of multilayer perceptrons for the analysis of chemical sensor array data. **IEEE Sensors Journal**, v. 4, n. 3, p. 355-363, 2004. DOI:10.1109/JSEN.2004.827207.

PENG, X.; KING, I. Robust BMPM training based on second-order cone programming and its application in medical diagnosis. **Neural Networks**, v. 21, p. 450-457, 2008. DOI: 10.1016/j.neunet.2007.12.051.

PIMENTA, C. J.; CHALFOUN, S. M.; PEREIRA, M. C.; ANGÉLICO, C. L.; TAVARES, L. S.; MARTINS, R. T. Avaliação físico-química e de qualidade do café (*Coffea arabica* L.) submetido a diferentes tempos de espera para secagem. **Revista Brasileira de Armazenamento**, Viçosa, n.10, p.36-41, 2008. Edição Especial.

PINTO, F. A. C.; CORRÊA, P. C.; QUEIROZ, D. M.; OLIVEIRA, A. S. C. Utilização de um sistema de visão artificial para estimativa do teor de umidade de frutos de café. **Revista Brasileira de Armazenamento**, Viçosa, n. 6, p. 30-36, 2003. Especial Café.

RAMÍREZ, M. C. V.; VELHO, H. F. C.; FERREIRA, N. J. Artificial neural network technique for rainfall forecasting applied to the São Paulo region. **Journal of Hydrology**, v. 301, p. 146-162, 2005. DOI: 10.1016/j.jhydrol.2004.06.028.

RECAMALES, A. F.; HERNANZ, D.; ÁLVAREZ, C.; GONZÁLEZ-MIRET, M. L.; HEREDIA, FRANCISCO J. Colour of Amontillado wines aged in two

oak barrel types. **European Food Research and Technology**, v. 224, n. 3, p. 321-327, jan. 2007. DOI: 10.1007/s00217-006-0399-5.

REINATO, C. H. R.; BORÉM, F. M.; SILVA, PABLO J.; OLIVEIRA, E. C. Influência da secagem, em diferentes tipos de terreiro, sobre a qualidade do café ao longo do armazenamento. **Coffee Science**, Lavras, v. 2, n. 1, p. 48-60, jan.-jun. 2007.

RIBEIRO, F. C. **Análise técnica da utilização do silo bolsa no armazenamento de cafés especiais**. Lavras, MG: UFLA, 2009. (Projeto de Pesquisa).

SABBAGH, N. K.; YOKOMIZO, Y.; FARIA, J. B. Influência da torração nos conteúdos de monossacárides de cafés arábica, robusta e do híbrido Icatu. **Coletânea do ITAL**, Campinas, 8, t.1, p.111-130, 1977.

SAEG - **Sistema para Análises Estatísticas**, Versão 9.1. Viçosa, MG: UFV/Fundação Arthur Bernardes, 2007.

SOUSA, S. I. V.; MARTINS, F. G.; ALVIM-FERRAZ, M. C. M.; PEREIRA, M. C. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. **Environmental Modeling & Software**, v. 22, p. 97-103, 2007. DOI: 10.1016/j.envsoft.2005.12.002.

SUCHACZ, B.; WESOLOWSKI, M. The recognition of similarities in trace elements content in medicinal plants using MLP and RBF neural networks. **Talanta**, v. 69, p. 37-42, 2006. DOI: 10.1016/j.talanta. 2005.08.026.

TAVEIRA, J. H. S. **Aspectos fisiológicos e bioquímicos associados à qualidade da bebida de café submetido a diferentes métodos de processamento e secagem**. 2009. 67 f. Dissertação (Mestrado em dos Alimentos) – Universidade Federal de Lavras, Lavras, MG, 2009.

TOKATLI, F.; TARI, C.; UNLUTURK, S. M.; BAYSAL, N. G. Modeling of polygalacturonase enzyme activity and biomass production by *Aspergillus sojae* ATCC 20235. **Journal of Industrial Microbiology and Biotechnology**, v. 36, n. 9, p. 1139-1148, sep. 2009. DOI 10.1007/s10295-009-0595-y.

TOMINAGA, O.; ITO, F.; HANAI, T.; HONDA, H.; KOBAYASHI, T. Sensory Modeling of Coffee with a Fuzzy Neural Network. **Journal of Food Science**, v. 67, n. 1, p. 363-368, 2001.

VERLINDEN, B.; DUFLOU, J. R.; COLLIN, P.; CATTRYSSE, D. Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study. **International Journal of Production Economics**, v. 111, p. 484-492, 2008. DOI: 10.1016/j.ijpe. 2007.02.004.

## **APÊNDICE**

## Apêndice 1. Análise Estatística

### Regressão Linear Múltipla dos dados Equação 4.1

```

-----
Data                DEA                Hora
16/06/2009         Evandro Castro          09:34:18
                    Sistema para Análises Estatísticas
-----
Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = R
Independentes = X1      C2      C3      C4      C5      C6
E s t a t í s t i c a s   S i m p l e s
Nome          Média      Desvio padrão
R              66.5030      6.5675
X1             506.0000     287.0945
C2             75.0000      38.1850
C3             9.6000      3.2528
C4             5.7970      2.4215
C5             2.5000      1.2728
C6             0.3300      0.1839
M a t r i z   d e   C o r r e l a ç õ e s
R              X1      C2      C3      C4      C5      C6
R              1.00000  0.43735 -0.58148  0.24729 -0.36928  0.48480 -0.19614
X1              0.43735  1.00000  0.00000  0.00000 -0.00057  0.00000  0.00000
C2              -0.58148  0.00000  1.00000  0.00000  0.00015  0.00000  0.00000
C3              0.24729  0.00000  0.00000  1.00000  0.00094  0.00000  0.00000
C4              -0.36928 -0.00057  0.00015  0.00094  1.00000 -0.00078  0.00048
C5              0.48480  0.00000  0.00000  0.00000 -0.00078  1.00000  0.00000
C6              -0.19614  0.00000  0.00000  0.00000  0.00048  0.00000  1.00000
Variável Dependente = R

P a r â m e t r o s   d a   R e g r e s s ã o
Nome                Coeficiente      Desvio padrão      Valor de T      Coef. Beta      Probab.
X1                  0.100000E-01     0.311836E-10      *****          0.437143         0.0001
C2                  -0.100000E+00     0.234454E-09      *****          -0.581422        0.0001
C3                  0.500000E+00     0.275229E-08      *****          0.247643         0.0001
C4                  -0.100000E+01     0.369714E-08      *****          -0.368709        0.0001
C5                  0.250000E+01     0.703363E-08      *****          0.484519         0.0001
C6                  -0.700000E+01     0.486944E-07      *****          -0.195961        0.0001
CONSTANTE          0.660000E+02

R2                  = 1.000000
R2 ajustado         = 1.000000
A n á l i s e   d e   V a r i â n c i a
Fontes de Variação  GL      Soma de Quadrados      Quadrado Médio      F      Signif.
Devido a Regressão  6       673898.1                112316.4            *****          0.0001
Independente        ****    -0.195577E-07          0.1000000E-05

T E S T E S
DURBIN-WATSON      = 0.7110959E-04
THEIL-NAGAR        = 0.9999648
Coeficiente RHO T  = -0.7114933E-01
Coeficiente RHO T-1 = 0.9998351

```

### Regressão Linear Múltipla dos dados Equação 4.1 + ε

```

-----
Data                DEA                Hora
16/06/2009         Evandro Castro          09:36:18
                    Sistema para Análises Estatísticas
-----
Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = R
Independentes = X1      X2      X3      X4      X5      X6

```

Estatísticas Simples

Nome	Média	Desvio padrão
R	66.4902	6.6619
X1	506.0000	287.0945
X2	75.0000	38.1850
X3	9.6000	3.2528
X4	5.8026	2.4216
X5	2.5000	1.2728
X6	0.3300	0.1839

Matriz de Correlações

	R	X1	X2	X3	X4	X5	X6
R	1.00000	0.43193	-0.57412	0.24448	-0.36573	0.48044	-0.19398
X1	0.43193	1.00000	0.00000	0.00000	-0.00063	0.00000	0.00000
X2	-0.57412	0.00000	1.00000	0.00000	0.00080	0.00000	0.00000
X3	0.24448	0.00000	0.00000	1.00000	0.00015	0.00000	0.00000
X4	-0.36573	-0.00063	0.00080	0.00015	1.00000	-0.00089	-0.00001
X5	0.48044	0.00000	0.00000	0.00000	-0.00089	1.00000	0.00000
X6	-0.19398	0.00000	0.00000	0.00000	-0.00001	0.00000	1.00000

Variável Dependente = R

Parâmetros da Regressão

Nome	Coefficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
X1	0.100175E-01	0.279462E-04	358.455383	0.431705	0.0001
X2	-0.100113E+00	0.210114E-03	-476.469402	-0.573834	0.0001
X3	0.500816E+00	0.246655E-02	203.042916	0.244534	0.0001
X4	-0.100308E+01	0.331324E-02	-302.748535	-0.364614	0.0001
X5	0.251286E+01	0.630342E-02	398.650691	0.480114	0.0001
X6	-0.702880E+01	0.436390E-01	-161.066852	-0.193980	0.0001
CONSTANTE	0.659798E+02				

R2 = 0.977347

R2 ajustado = 0.977338

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	6	677693.1	112948.8	*****	0.0000
Independente	****	15707.70	1.005743		

## Regressão Linear Múltipla dos dados Equação 4.1 + 2 ε

DEA							
Data	Evandro Castro				Hora		
16/06/2009	Sistema para Análises Estatísticas				09:37:06		
-----							
Procedimento	= Regressão Linear						
Objetivo	= Regressão linear simples e múltipla						
Dependentes	= R						
Independentes	= X1	X2	X3	X4	X5	X6	
Estatísticas Simples							
Nome	Média	Desvio padrão					
R	66.4895	6.8702					
X1	506.0000	287.0945					
X2	75.0000	38.1850					
X3	9.6000	3.2528					
X4	5.7949	2.4214					
X5	2.5000	1.2728					
X6	0.3300	0.1839					
Matriz de Correlações							
	R	X1	X2	X3	X4	X5	X6
R	1.00000	0.41824	-0.55884	0.23548	-0.35096	0.46175	-0.18794
X1	0.41824	1.00000	0.00000	0.00000	0.00075	0.00000	0.00000
X2	-0.55884	0.00000	1.00000	0.00000	0.00106	0.00000	0.00000
X3	0.23548	0.00000	0.00000	1.00000	0.00008	0.00000	0.00000
X4	-0.35096	0.00075	0.00106	0.00008	1.00000	0.00076	0.00051
X5	0.46175	0.00000	0.00000	0.00000	0.00076	1.00000	0.00000
X6	-0.18794	0.00000	0.00000	0.00000	0.00051	0.00000	1.00000
Variável Dependente = R							
-----							



Parâmetros da Regressão

Nome	Coefficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
X1	0.100147E-01	0.560304E-04	178.737808	0.418500	0.0001
X2	-0.100480E+00	0.421266E-03	-238.518433	-0.558471	0.0001
X3	0.497407E+00	0.494529E-02	100.581884	0.235504	0.0001
X4	-0.995750E+00	0.664314E-02	-149.891499	-0.350959	0.0001
X5	0.249374E+01	0.126380E-01	197.321516	0.462012	0.0001
X6	-0.701633E+01	0.874936E-01	-80.192528	-0.187764	0.0001
CONSTANTE	0.660342E+02				

R2 = 0.914379  
R2 ajustado = 0.914346

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	6	674307.1	112384.5	*****	0.0000
Independente	****	63141.44	4.042864		

## Regressão Linear Múltipla dos dados Equação 4.2

```

-----
Data                      DEA                      Hora
16/06/2009                Evandro Castro                09:38:22
                          Sistema para Análises Estatísticas
-----
Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = R
Independentes = X1          C2          C3          C4          C5          C6
Estatísticas Simples
Nome         Média         Desvio padrão
R            65.4912        8.2578
X1           448040.0000     779396.5895
C2            0.7750         0.6819
C3            1.8200         1.1789
C4            4.8005         2.4213
C5            17.0000        13.7554
C6            1.5600         0.6830
Matriz de Correlações
R            X1          C2          C3          C4          C5          C6
R            1.00000  0.11659  -0.24925  -0.50822  0.29690  0.21637  -0.62552
X1           0.11659  1.00000  0.00000  0.00000  0.00000  0.00000  0.00000
C2           -0.24925  0.00000  1.00000  0.00000  -0.00059  0.00000  0.00000
C3           -0.50822  0.00000  0.00000  1.00000  -0.00159  0.00000  0.00000
C4            0.29690  0.00000  -0.00059  -0.00159  1.00000  0.00271  -0.00252
C5            0.21637  0.00000  0.00000  0.00000  0.00271  1.00000  0.00000
C6           -0.62552  0.00000  0.00000  0.00000  -0.00252  0.00000  1.00000
Variável Dependente = R
-----
Parâmetros da Regressão
Nome         Coeficiente  Desvio padrão  Valor de T  Coef. Beta  Probab.
X1           0.123524E-05  0.319023E-07   38.719575   0.116586    0.0001
C2           -0.301612E+01  0.364620E-01  -82.719711  -0.249072    0.0001
C3           -0.355682E+01  0.210922E-01  -168.631616 -0.507756    0.0001
C4           0.100196E+01  0.102694E-01   97.568002   0.293783    0.0001
C5           0.129419E+00  0.180762E-02   71.595978   0.215579    0.0001
C6           -0.755434E+01  0.364073E-01  -207.495013 -0.624776    0.0001
CONSTANTE   0.785234E+02
-----
R2            = 0.858402
R2 ajustado  = 0.858348
Análise de Variância
Fontes de Variação  GL  Soma de Quadrados  Quadrado Médio  F  Signif.
Devido a Regressão  6   914563.4            152427.2        *****  0.0000
Independente        **** 150861.7            9.659473

```

## Regressão Linear Múltipla dos dados Equação 4.2 + $\epsilon$

```

-----
                DEA
Data            Evandro Castro                Hora
16/06/2009     Sistema para Análises Estatísticas 09:40:38
-----

Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = R
Independentes = X1      C2      C3      C4      C5      C6

E s t a t í s t i c a s   S i m p l e s
Nome                    Média      Desvio padrão
R                        65.4913    8.3099
X1                       448040.0000 779396.5895
C2                        0.7750    0.6819
C3                        1.8200    1.1789
C4                        4.7977    2.4234
C5                        17.0000   13.7554
C6                        1.5600    0.6830
M a t r i z   d e   C o r r e l a ç õ e s
      R      X1      C2      C3      C4      C5      C6
R      1.00000  0.11712 -0.24720 -0.50358  0.29347  0.21482 -0.62078
X1     0.11712  1.00000  0.00000  0.00000 -0.00042  0.00000  0.00000
C2    -0.24720  0.00000  1.00000  0.00000  0.00031  0.00000  0.00000
C3    -0.50358  0.00000  0.00000  1.00000 -0.00008  0.00000  0.00000
C4     0.29347 -0.00042  0.00031 -0.00008  1.00000  0.00088 -0.00030
C5     0.21482  0.00000  0.00000  0.00000  0.00088  1.00000  0.00000
C6    -0.62078  0.00000  0.00000  0.00000 -0.00030  0.00000  1.00000
Variável Dependente = R

P a r â m e t r o s   d a   R e g r e s s ã o
Nome                Coeficiente  Desvio padrão  Valor de T      Coef. Beta  Probab.
X1                   0.125006E-05  0.334921E-07   37.323898       0.117245    0.0001
C2                   -0.301346E+01  0.382790E-01  -78.723597     -0.247293    0.0001
C3                   -0.354965E+01  0.221433E-01  -160.303369    -0.503558    0.0001
C4                    0.100533E+01  0.107717E-01   93.330458      0.293178    0.0001
C5                    0.129618E+00  0.189770E-02   68.302889      0.214559    0.0001
C6                   -0.755224E+01  0.382215E-01  -197.591294    -0.620690    0.0001
CONSTANTE           0.784818E+02

R2                   = 0.845887
R2 ajustado          = 0.845828
A n á l i s e   d e   V a r i a n c i a
Fontes de Variação  GL      Soma de Quadrados  Quadrado Médio  F      Signif.
Devido a Regressão  6       912624.8           152104.1        ***** 0.0000
Independente        ****    166272.3           10.64620


```

## Regressão Linear Múltipla dos dados Equação 4.1 + 2 $\epsilon$

```

-----
                DEA
Data            Evandro Castro                Hora
16/06/2009     Sistema para Análises Estatísticas 09:41:18
-----

Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = R
Independentes = X1      C2      C3      C4      C5      C6

E s t a t í s t i c a s   S i m p l e s
Nome                    Média      Desvio padrão
R                        61.5364    8.4842
X1                       448040.0000 779396.5895
C2                        0.7750    0.6819
C3                        1.8200    1.1789
C4                        4.8003    2.4175
C5                        17.0000   13.7554
C6                        1.5600    0.6830

```

Matriz de Correlações

	R	X1	C2	C3	C4	C5	C6
R	1.00000	0.11262	-0.24123	-0.49236	0.28416	0.21004	-0.60922
X1	0.11262	1.00000	0.00000	0.00000	0.00000	0.00000	0.00000
C2	-0.24123	0.00000	1.00000	0.00000	0.00050	0.00000	0.00000
C3	-0.49236	0.00000	0.00000	1.00000	0.00015	0.00000	0.00000
C4	0.28416	0.00000	0.00050	0.00015	1.00000	0.00129	0.00026
C5	0.21004	0.00000	0.00000	0.00000	0.00129	1.00000	0.00000
C6	-0.60922	0.00000	0.00000	0.00000	0.00026	0.00000	1.00000

Variável Dependente = R

Parâmetros da Regressão

Nome	Coefficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
X1	0.122589E-05	0.380320E-07	32.233214	0.112617	0.0001
C2	-0.300292E+01	0.434678E-01	-69.083859	-0.241366	0.0001
C3	-0.354384E+01	0.251449E-01	-140.936793	-0.492406	0.0001
C4	0.997552E+00	0.122617E-01	81.355350	0.284240	0.0001
C5	0.129325E+00	0.215493E-02	60.013383	0.209675	0.0001
C6	-0.756911E+01	0.434025E-01	-174.393228	-0.609297	0.0001
CONSTANTE	0.745849E+02				

R2 = 0.809356

R2 ajustado = 0.809283

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	6	910227.1	151704.5	*****	0.0000
Independente	****	214404.5	13.72804		

## Leitura dos resultados Análises Química e Sensorial

----- Data 22/09/2009	DEA Evandro Castro Regressao	----- Hora 08:20:46
-----------------------------	------------------------------------	---------------------------

### D E S C R I Ç Ã O   D O   A R Q U I V O

Tipo de Leitura      -      Microsoft Excel

Variáveis	Mínimos	Máximos	Perdidos	Válidos
ARD	0.150000	0.710000	0	413
ANR	1.950000	10.38000	0	413
POL	3.420000	7.900000	0	413
ATT	115.8300	255.0000	0	413
CEL	36.76000	297.0000	0	413
LPO	24.18000	91.71000	0	413
ASE	44.00000	82.00000	0	413
Observações Gravadas...		413		
Variáveis Totais.....		7		
Valores Perdidos.....		0		

## Regressão Linear Simples ASE = f(ARd)

---

```

Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = ASE
Independentes = ARD
Estatísticas Simples
Nome         Média      Desvio padrão
ASE         69.9803    12.0123
ARD         0.3809     0.1232
Matriz de Correlações
           ASE      ARD
ASE       1.00000  -0.37371
ARD      -0.37371  1.00000
Variável Dependente = ASE

```

---

```

Parâmetros da Regressão
Nome Coeficiente Desvio padrão Valor de T Coef. Beta Probab.
ARD  -0.364338E+02 0.446054E+01  -8.168027 -0.373708 0.0001
CONSTANTE 0.838569E+02

```

```

R2 = 0.139657
R2 ajustado = 0.137564

```

```

Análise de Variância
Fontes de Variação GL Soma de Quadrados Quadrado Médio F Signif.
Devido a Regressão 1 8302.655 8302.655 66.72 0.0000
Independente 411 51147.51 124.4465

```

## Regressão Linear Simples ASE = f(AnR)

---

```

Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = ASE
Independentes = ANR
Estatísticas Simples
Nome         Média      Desvio padrão
ASE         69.9803    12.0123
ANR         6.5125     1.5987
Matriz de Correlações
           ASE      ANR
ASE       1.00000  0.29422
ANR       0.29422  1.00000
Variável Dependente = ASE

```

---

```

Parâmetros da Regressão
Nome Coeficiente Desvio padrão Valor de T Coef. Beta Probab.
ANR  0.221072E+01 0.354228E+00  6.240937 0.294217 0.0001
CONSTANTE 0.555830E+02

```

```

R2 = 0.086564
R2 ajustado = 0.084341

```

```

Análise de Variância
Fontes de Variação GL Soma de Quadrados Quadrado Médio F Signif.
Devido a Regressão 1 5146.228 5146.228 38.95 0.0000
Independente 411 54303.93 132.1264

```

## Regressão Linear Simples ASE = f(Pol)

---

```

Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = ASE
Independentes = POL
Estatísticas Simples
Nome         Média      Desvio padrão
ASE         69.9803    12.0123
POL         6.1007     0.8336
Matriz de Correlações
              ASE      POL
ASE          1.00000  -0.37054
POL          -0.37054  1.00000
Variável Dependente = ASE
    
```

---

```

Parâmetros da Regressão
Nome Coeficiente Desvio padrão Valor de T Coef. Beta Probab.
POL  -0.533970E+01 0.660223E+00 -8.087717 -0.370540 0.0001
CONSTANTE 0.102556E+03
    
```

R2 = 0.137300

R2 ajustado = 0.135201

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	1	8162.496	8162.496	65.41	0.0000
Independente	411	51287.67	124.7875		

## Regressão Linear Simples ASE = f(ATT)

---

```

Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = ASE
Independentes = ATT
Estatísticas Simples
Nome         Média      Desvio padrão
ASE         69.9803    12.0123
ATT         200.4497   24.0243
Matriz de Correlações
              ASE      ATT
ASE          1.00000  -0.51743
ATT          -0.51743  1.00000
Variável Dependente = ASE
    
```

---

```

Parâmetros da Regressão
Nome Coeficiente Desvio padrão Valor de T Coef. Beta Probab.
ATT  -0.258718E+00 0.211053E-01 -12.258436 -0.517427 0.0001
CONSTANTE 0.121840E+03
    
```

R2 = 0.267731

R2 ajustado = 0.265949

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	1	15916.66	15916.66	150.27	0.0000
Independente	411	43533.50	105.9209		

## Regressão Linear Simples ASE = f(CEI)

---

```

Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = ASE
Independentes = CEL
Estatísticas Simples
Nome         Média      Desvio padrão
ASE         69.9803    12.0123
CEL        155.9102    49.8011
Matriz de Correlações
           ASE      CEL
ASE      1.00000  -0.73570
CEL     -0.73570  1.00000
Variável Dependente = ASE
    
```

---

```

Parâmetros da Regressão
Nome      Coeficiente Desvio padrão Valor de T   Coef. Beta  Probab.
CEL     -0.177456E+00 0.805846E-02  -22.021085  -0.735702  0.0001
CONSTANTE 0.976475E+02
    
```

```

R2          = 0.541258
R2 ajustado = 0.540142
Análise de Variância
Fontes de Variação GL Soma de Quadrados Quadrado Médio   F   Signif.
Devido a Regressão  1    32177.87    32177.87    484.93 0.0000
Independente       411   27272.29     66.35595
    
```

## Regressão Linear Simples ASE = f(LPo)

---

```

Procedimento = Regressão Linear
Objetivo     = Regressão linear simples e múltipla
Dependentes  = ASE
Independentes = LPO
Estatísticas Simples
Nome         Média      Desvio padrão
ASE         69.9803    12.0123
LPO         42.0599    13.2269
Matriz de Correlações
           ASE      LPO
ASE      1.00000  -0.53916
LPO     -0.53916  1.00000
Variável Dependente = ASE
    
```

---

```

Parâmetros da Regressão
Nome      Coeficiente Desvio padrão Valor de T   Coef. Beta  Probab.
LPO     -0.489655E+00 0.377279E-01  -12.978578  -0.539165  0.0001
CONSTANTE 0.905751E+02
    
```

```

R2          = 0.290699
R2 ajustado = 0.288973
Análise de Variância
Fontes de Variação GL Soma de Quadrados Quadrado Médio   F   Signif.
Devido a Regressão  1    17282.09    17282.09    168.44 0.0000
Independente       411   42168.08     102.5987
    
```

## Regressão Linear Múltipla ASE = f(ATT, CEI, LPO)

---

Procedimento = Regressão Linear  
 Objetivo = Regressão linear simples e múltipla  
 Dependentes = ASE  
 Independentes = ATT            CEL            LPO

**E s t a t í s t i c a s       S i m p l e s**  
 Nome            Média            Desvio padrão  
 ASE            69.9803            12.0123  
 ATT            200.4497            24.0243  
 CEL            155.9102            49.8011  
 LPO            42.0599            13.2269

**M a t r i z   d e   C o r r e l a ç õ e s**  
                  ASE            ATT            CEL            LPO  
 ASE            1.00000    -0.51743    -0.73570    -0.53916  
 ATT            -0.51743    1.00000    0.53285    0.26295  
 CEL            -0.73570    0.53285    1.00000    0.64729  
 LPO            -0.53916    0.26295    0.64729    1.00000

Variável Dependente = ASE

---

**P a r â m e t r o s   d a   R e g r e s s ã o**  
 Nome    Coeficiente    Desvio padrão    Valor de T    Coef. Beta    Probab.  
 ATT    -0.953072E-01    0.192329E-01    -4.955430    -0.190611    0.0001  
 CEL    -0.131843E+00    0.117437E-01    -11.226750    -0.546599    0.0001  
 LPO    -0.122815E+00    0.387811E-01    -3.166883    -0.135233    0.0008  
 CONSTANTE    0.114806E+03

R2            = 0.573675  
 R2 ajustado = 0.570548

**A n á l i s e   d e   V a r i â n c i a**  
 Fontes de Variação    GL    Soma de Quadrados    Quadrado Médio    F    Signif.  
 Devido a Regressão    3       34105.05       11368.35    183.45    0.0000  
 Independente            409       25345.11       61.96849

**T E S T E S**  
 DURBIN-WATSON            =       1.046431  
 THEIL-NAGAR              =       0.4769184  
 Coeficiente RHO T        =       0.8475147  
 Coeficiente RHO T-1      =       0.4483559

Variável Dependente = ASE  
 Após a Retirada de    = LPO

---

**P a r â m e t r o s   d a   R e g r e s s ã o**  
 Nome    Coeficiente    Desvio padrão    Valor de T    Coef. Beta    Probab.  
 CEL    -0.154947E+00    0.930359E-02    -16.654513    -0.642383    0.0001  
 ATT    -0.875678E-01    0.192859E-01    -4.540511    -0.175132    0.0001  
 CONSTANTE    0.111691E+03

R2            = 0.563221  
 R2 ajustado = 0.561090



Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	2	33483.56	16741.78	264.34	0.0000
Independente	410	25966.60	63.33318		

T E S T E S

DURBIN-WATSON	=	1.063564
THEIL-NAGAR	=	0.4682911
Coeficiente RHO T	=	-0.5617717E-01
Coeficiente RHO T-1	=	0.4582044

Análise de Variância para a Eliminação

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Efeito da retirada	409	621.4911	61.96849	3.34	0.1740
Modelo Anterior	3	25966.60	63.33318		

Variável Dependente = ASE

Após a Retirada de = ATT

Parâmetros da Regressão

Nome	Coeficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
CEL	-0.177456E+00	0.805846E-02	-22.021085	-0.735702	0.0001
CONSTANTE	0.976475E+02				

R2 = 0.541258

R2 ajustado = 0.540142

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	1	32177.87	32177.87	484.93	0.0000
Independente	411	27272.29	66.35595		

T E S T E S

DURBIN-WATSON	=	0.2857451
THEIL-NAGAR	=	0.8571676
Coeficiente RHO T	=	2.593358
Coeficiente RHO T-1	=	0.7480323

Análise de Variância para a Eliminação

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Efeito da retirada	3	1305.692	63.33318	10.31	0.0897
Modelo Anterior	2	27272.29	66.35595		

## Regressão Linear Múltipla ASE = f(ARd, AnR, Pol, ATT, CEI, LPO)

Procedimento = Regressão Linear  
 Objetivo = Regressão linear simples e múltipla  
 Dependentes = ASE  
 Independentes = ARD ANR POL ATT CEL LPO

### Estadísticas Simples

Nome	Média	Desvio padrão
ASE	69.9803	12.0123
ARD	0.3809	0.1232
ANR	6.5125	1.5987
POL	6.1007	0.8336
ATT	200.4497	24.0243
CEL	155.9102	49.8011
LPO	42.0599	13.2269

### Matriz de Correlações

	ASE	ARD	ANR	POL	ATT	CEL	LPO
ASE	1.00000	-0.37371	0.29422	-0.37054	-0.51743	-0.73570	-0.53916
ARD	-0.37371	1.00000	0.22865	0.32837	0.16341	0.08336	0.05153
ANR	0.29422	0.22865	1.00000	0.31093	0.02131	-0.52580	-0.43531
POL	-0.37054	0.32837	0.31093	1.00000	0.40106	0.21139	0.36306
ATT	-0.51743	0.16341	0.02131	0.40106	1.00000	0.53285	0.26295
CEL	-0.73570	0.08336	-0.52580	0.21139	0.53285	1.00000	0.64729
LPO	-0.53916	0.05153	-0.43531	0.36306	0.26295	0.64729	1.00000

Variável Dependente = ASE

### Parâmetros da Regressão

Nome	Coefficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
ARD	-0.280864E+02	0.301377E+01	-9.319330	-0.28808	0.0001
ANR	0.707124E+00	0.328283E+00	2.154009	0.094109	0.0156
POL	-0.171888E+01	0.571789E+00	-3.006141	-0.119279	0.0013
ATT	-0.631717E-01	0.191830E-01	-3.293109	-0.126341	0.0005
CEL	-0.127175E+00	0.121616E-01	-10.457090	-0.527244	0.0001
LPO	-0.595256E-01	0.390355E-01	-1.524909	-0.065544	0.0636
CONSTANTE	0.121553E+03				

R2 = 0.668152

R2 ajustado = 0.663248

### Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	6	39721.74	6620.290	136.24	0.0000
Independente	406	19728.42	48.59217		

### TESTES

DURBIN-WATSON	=	0.2015498E-01
THEIL-NAGAR	=	0.9904914
Coefficiente RHO T	=	557.0811
Coefficiente RHO T-1	=	0.8800392

Variável Dependente = ASE

Após a Retirada de = LPO

---

Parâmetros da Regressão

Nome	Coeficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
CEL	-0.134780E+00	0.111100E-01	-12.131416	-0.558773	0.0001
ARD	-0.278643E+02	0.301515E+01	-9.241439	-0.285809	0.0001
ATT	-0.581500E-01	0.189290E-01	-3.072011	-0.116298	0.0011
POL	-0.212401E+01	0.507143E+00	-4.188189	-0.147392	0.0001
ANR	0.857104E+00	0.313713E+00	2.732130	0.114069	0.0031
CONSTANTE	0.120639E+03				

R2 = 0.666251  
R2 ajustado = 0.662151

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	5	39608.75	7921.750	162.50	0.0000
Independente	407	19841.41	48.75040		

T E S T E S

DURBIN-WATSON	=	0.1105304E-01
THEIL-NAGAR	=	0.9948916
Coeficiente RHO T	=	128.0242
Coeficiente RHO T-1	=	0.6803505

Análise de Variância para a Eliminação

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Efeito da retirada	406	112.9937	48.59217	0.39	*****
Modelo Anterior	6	19841.41	48.75040		

Variável Dependente = ASE  
Após a Retirada de = ANR

---

Parâmetros da Regressão

Nome	Coeficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
CEL	-0.155371E+00	0.822702E-02	-18.885464	-0.644142	0.0001
ARD	-0.262640E+02	0.298105E+01	-8.810328	-0.269394	0.0001
POL	-0.160915E+01	0.474547E+00	-3.390925	-0.111665	0.0003
ATT	-0.426953E-01	0.182065E-01	-2.345051	-0.085389	0.0095
CONSTANTE	0.122583E+03				

R2 = 0.660130  
R2 ajustado = 0.656798

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	4	39244.85	9811.212	198.11	0.0000
Independente	408	20205.31	49.52283		

T E S T E S

DURBIN-WATSON	=	0.4186535
THEIL-NAGAR	=	0.7909322
Coeficiente RHO T	=	3.372066
Coeficiente RHO T-1	=	0.6291964

Análise de Variância para a Eliminação

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Efeito da retirada	6	363.8991	48.75040	1.49	0.3384
Modelo Anterior	5	20205.31	49.52283		

Variável Dependente = ASE  
Após a Retirada de = ATT

---

Parâmetros da Regressão

Nome	Coefficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
CEL	-0.165023E+00	0.716255E-02	-23.039703	-0.684157	0.0001
ARD	-0.264990E+02	0.299571E+01	-8.845647	-0.271804	0.0001
POL	-0.196936E+01	0.451463E+00	-4.362181	-0.136661	0.0001
CONSTANTE	0.117816E+03				

R2 = 0.655549  
R2 ajustado = 0.653023

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	3	38972.51	12990.84	259.47	0.0000
Independente	409	20477.65	50.06761		

T E S T E S

DURBIN-WATSON = 0.3887892  
THEIL-NAGAR = 0.8057713  
Coeficiente RHO T = 3.394383  
Coeficiente RHO T-1 = 0.6485069

Análise de Variância para a Eliminação

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Efeito da retirada	5	272.3391	49.52283	1.37	0.3901
Modelo Anterior	4	20477.65	50.06761		

Variável Dependente = ASE  
Após a Retirada de = POL

---

Parâmetros da Regressão

Nome	Coefficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
CEL	-0.171131E+00	0.717710E-02	-23.844081	-0.709481	0.0001
ARD	-0.306682E+02	0.290090E+01	-10.571949	-0.314569	0.0001
CONSTANTE	0.108342E+03				

R2 = 0.639524  
R2 ajustado = 0.637765

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	2	38019.79	19009.90	363.69	0.0000
Independente	410	21430.37	52.26920		

T E S T E S

DURBIN-WATSON = 0.3892647  
THEIL-NAGAR = 0.8054594  
Coeficiente RHO T = 3.543902  
Coeficiente RHO T-1 = 0.6461844

Análise de Variância para a Eliminação

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Efeito da retirada	4	952.7175	50.06761	6.34	0.0805
Modelo Anterior	3	21430.37	52.26920		

Variável Dependente = ASE  
Após a Retirada de = ARD

---

Parâmetros da Regressão

Nome	Coefficiente	Desvio padrão	Valor de T	Coef. Beta	Probab.
CEL	-0.177456E+00	0.805846E-02	-22.021085	-0.735702	0.0001
CONSTANTE	0.976475E+02				

R2 = 0.541258

R2 ajustado = 0.540142

Análise de Variância

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Devido a Regressão	1	32177.87	32177.87	484.93	0.0000
Independente	411	27272.29	66.35595		

TESTES

DURBIN-WATSON = 0.2857451

THEIL-NAGAR = 0.8571676

Coefficiente RHO T = 2.593358

Coefficiente RHO T-1 = 0.7480323

Análise de Variância para a Eliminação

Fontes de Variação	GL	Soma de Quadrados	Quadrado Médio	F	Signif.
Efeito da retirada	3	5841.924	52.26920	55.88	0.0176
Modelo Anterior	2	27272.29	66.35595		

\*\*\* Atenção \*\*\* Variável com efeito Significativo