



## ANÁLISE DE DADOS AMOSTRAIS COMPLEXOS UTILIZANDO REDES NEURAIS

Savano Sousa Pereira

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientadores: Luiz Pereira Calôba  
Pedro Luis do Nascimento Silva

Rio de Janeiro  
Setembro de 2017

ANÁLISE DE DADOS AMOSTRAIS COMPLEXOS UTILIZANDO REDES  
NEURAIS

Savano Sousa Pereira

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ  
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)  
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR  
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

---

Prof. Luiz Pereira Calôba, D.Sc.

---

Prof. Pedro Luis do Nascimento Silva, Ph.D.

---

Prof. Jorge Lopes de Sousa Leão, D.Sc.

---

Prof. Fernando Antônio da Silva Moura, Ph.D.

---

Prof. José André de Moura Brito, D.Sc.

---

Prof. Marcelo Albano Moret Simões Gonçalves, D.Sc.

RIO DE JANEIRO, RJ – BRASIL  
SETEMBRO DE 2017

Pereira, Savano Sousa

Análise de Dados Amostrais Complexos utilizando Redes Neurais/Savano Sousa Pereira. – Rio de Janeiro: UFRJ/COPPE, 2017.

XIV, 187 p.: il.; 29, 7cm.

Orientadores: Luiz Pereira Calôba

Pedro Luis do Nascimento Silva

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2017.

Referências Bibliográficas: p. 69 – 71.

1. Redes Neurais Artificiais. 2. Amostragem. 3. Abordagem de superpopulação. I. Calôba, Luiz Pereira *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

*Vocês são minha família e isso é  
suficiente para me sentir capaz  
de vencer qualquer desafio na  
vida.*

Anônimo.

# Agradecimentos

Há cinco anos a decisão de iniciar este trabalho foi escolhida em meio a muitas incertezas. Diversos fatores externos contribuíram de maneiras positivas e outras vezes de formas negativas. O principal desafio foi equacionar o nascimento de minha filha e a iniciativa de empreender com um libanês (que me foi apresentado por um judeu) com o início do curso de doutorado. Período de muito trabalho, ausências familiares e alguns desentendimentos. Porém, no final acho que tudo ocorreu de uma maneira aceitável.

Os agradecimentos se fazem necessários, porém, o simples registro nesta página do trabalho não expressa realmente o quanto tenho a agradecer as pessoas que fizeram parte destes 5 anos de aprendizado. Todavia, se faz necessário este registro formal para posteridade.

O primeiro grupo de pessoas que agradeço é a minha família sobretudo a minha esposa que sempre me apoiou nos momentos difíceis, a minha filha que ajudou a controlar minha ansiedade e proporcionou uma motivação a mais para continuar e concluir este trabalho e minha sogra que também sempre me apoiou. Aos meus pais e irmão pelo apoio e pelo incentivo aos estudos. Não poderia deixar de agradecer ao amigo Umberto pelas conversas sobre o meio acadêmico, política, entre outras.

O segundo bloco de agradecimentos se direciona para os sócios da *Mobi2buy* pelo apoio, incentivo e torcida pelo sucesso desta caminhada. Um agradecimento especial para o libanês pela amizade e pela sensibilidade dos seus conselhos.

Por último, porém, não menos importantes um ‘grande obrigado’ aos meus orientadores Prof. Calôba e Prof. Pedro que me guiaram nesta jornada.

Finalizo agradecendo o apoio financeiro do CNPq, o qual foi importante para sustentação deste projeto. E se por acaso deixei de agradecer alguém peço desculpas.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

## ANÁLISE DE DADOS AMOSTRAIS COMPLEXOS UTILIZANDO REDES NEURAIIS

Savano Sousa Pereira

Setembro/2017

Orientadores: Luiz Pereira Calôba  
Pedro Luis do Nascimento Silva

Programa: Engenharia Elétrica

O ajuste de um modelo de Redes Neurais Artificiais (RNA) considera que os dados são provenientes de uma amostra aleatória simples com reposição. Entretanto, na prática em pesquisas amostrais a seleção de uma amostra aleatória simples é raramente utilizada e esquemas amostrais mais complexos são utilizados.

A amostragem complexa reflete estruturas mais complexas da população. Estas estruturas provenientes do plano amostral necessitam ser incorporadas quando ajustamos uma RNA. Na literatura estatística, existem diferentes abordagens para modelagem de dados amostrais complexos. Entretanto, a literatura relacionada a RNA não existe menção de como proceder quando os dados são provenientes de um esquema amostral complexo.

Neste trabalho propomos uma abordagem de superpopulação para modelar os dados provenientes de um esquema amostral complexo fazendo o uso de RNA. A avaliação da metodologia proposta é realizada de forma empírica por meio de simulações e, posteriormente, são calculadas medidas de avaliação dos estimadores. A aplicação prática é feita utilizando os dados da Pesquisa Nacional por Amostra de Domicílio do ano de 2014. O objetivo é melhorar a estimativa da renda domiciliar *per capita* utilizada para a construção do Mapa de Pobreza em IBGE (2008).

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

## ARTIFICIAL NEURAL NETWORKS WITH COMPLEX SURVEY DATA

Savano Sousa Pereira

September/2017

Advisors: Luiz Pereira Calôba

Pedro Luis do Nascimento Silva

Department: Electrical Engineering

The fitting of an Artificial Neural Network (ANN) considers the data coming from an simple random sample with replacement. However, in practice the selection of simple random samples for surveys is rarely used and more complex sampling schemes are used.

The complex sampling schemes reflect complex structures from population. These structures of sampling scheme need to be incorporated when we fitting an ANN. In statistical literature, there are different approaches for modelling data from complex surveys. However in the literature related to ANN there is no mention of how proceed when the data come from complex sample surveys.

This work porpose an superpopulation approach for modelling data from complex survey using ANN. An evaluation of the proposed methodology is carried out empirically through simulations and, after, estimation measures are calculated. The practical application is done using data from the National Household Sample Survey for the year 2014. The objective is to improve the estimate of the per capita household income used to construct the Poverty Map in IBGE (2008).

# Sumário

<b>Lista de Figuras</b>	<b>xi</b>
<b>Lista de Tabelas</b>	<b>xiii</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Referencial e Conceitos Básicos</b>	<b>3</b>
2.1 Amostragem . . . . .	3
2.2 Redes Neurais Artificiais . . . . .	5
<b>3 Abordagens Clássica e Probabilística para Modelagem de Dados Amostrais</b>	<b>7</b>
3.1 Abordagens Clássica e Probabilística . . . . .	7
3.2 O estimador de Horwitz-Thompson . . . . .	10
<b>4 Abordagem de Superpopulação</b>	<b>11</b>
4.1 Modelagem de superpopulação . . . . .	11
4.2 Modelo de Regressão Linear . . . . .	14
4.2.1 Especificação do Modelo . . . . .	14
4.2.2 Pseudo-parâmetros do Modelo . . . . .	15
4.2.3 Estimadores de Máxima Pseudo-Verossimilhança dos Parâmetros do Modelo . . . . .	16
4.2.4 Estimadores da Variância de Estimadores de Máxima Pseudo-Verossimilhança . . . . .	17
<b>5 Modelo Clássico de Rede Neural</b>	<b>19</b>
5.1 Abordagem clássica de RNA . . . . .	19
5.2 Identificação em modelos de Redes Neurais . . . . .	25
5.2.1 Determinação do número de neurônios da camada intermediária	26
5.2.2 Teste-LM considerando amostragem aleatória simples . . . . .	27
<b>6 Redes Neurais para Dados Amostrais Complexos</b>	<b>31</b>
6.1 RNA sob o modelo de superpopulação . . . . .	31



6.2	Estimação da RNA sob o modelo de superpopulação . . . . .	33
6.2.1	Estimação dos parâmetros . . . . .	33
6.2.2	Estimação das variâncias . . . . .	36
<b>7</b>	<b>Avaliação da abordagem de superpopulação em RNA</b>	<b>41</b>
7.1	Estratégia geral . . . . .	41
7.2	População Hipotética . . . . .	42
7.3	Procedimento de seleção das amostras . . . . .	44
7.4	Ajuste dos modelos de RNA . . . . .	45
7.5	Medidas de avaliação dos estimadores . . . . .	46
7.5.1	Estimação por simulação das medidas de qualidade . . . . .	47
7.6	Análise dos coeficientes estimados pela RNA . . . . .	48
<b>8</b>	<b>Aplicação aos dados da PNAD da abordagem de superpopulação em RNA</b>	<b>60</b>
8.1	Construção da base de dados . . . . .	61
8.2	Aplicação e Resultados . . . . .	63
<b>9</b>	<b>Considerações Finais e Trabalhos Futuros</b>	<b>67</b>
9.1	Considerações Finais . . . . .	67
9.2	Trabalhos Futuros . . . . .	68
	<b>Referências Bibliográficas</b>	<b>69</b>
<b>A</b>	<b>Dicionário de variáveis</b>	<b>72</b>
A.1	Pesquisa Nacional por Amostra de Domicílios - PNAD - ano 2014 . . . . .	72
A.2	Variáveis Chave: Arquivos de Domicílio e de Pessoas . . . . .	72
A.3	Arquivo original: Domicílio . . . . .	73
A.4	Variáveis Dicotômicas . . . . .	79
A.5	Arquivo original: Pessoas . . . . .	81
<b>B</b>	<b>Resultados dos Modelos</b>	<b>91</b>
B.1	Modelo linear clássico - variável dependente renda . . . . .	91
B.2	Modelo linear clássico com logaritmo natural da renda como variável dependente . . . . .	92
B.3	Modelo linear de superpopulação com renda como variável dependente	94
B.4	Modelo linear de superpopulação com logaritmo natural da renda como variável dependente . . . . .	96
B.5	Coefficientes da RNA-Clássica com variável dependente renda . . . . .	98
B.6	Coefficientes da RNA-Clássica com variável dependente o logaritmo natural da renda . . . . .	99

B.7	Coeficientes da RNA-Superpopulação com variável dependente a renda	100
B.8	Coeficientes da RNA-Superpopulação da variável dependente o logaritmo natural da renda . . . . .	101
B.9	Gráficos dos resíduos dos modelos ajustados . . . . .	102
B.10	Gráficos da evolução do SQE dos modelos de RNA estimados . . . . .	106
<b>C</b>	<b>Scripts R</b>	<b>108</b>
C.1	RNA sob abordagem clássica identificável . . . . .	108
C.2	RNA sob abordagem de superpopulação identificável . . . . .	113
C.3	Construção da base da PNAD-2014 . . . . .	118
C.4	Modelo Não Linear Fraco - Abordagem Clássica . . . . .	128
C.5	Modelo Não Linear Fraco - Abordagem de Superpopulação . . . . .	155

# Lista de Figuras

3.1	Representação gráfica da abordagem clássica . . . . .	8
3.2	Representação gráfica da abordagem por amostragem probabilística . . . . .	9
4.1	Representação gráfica da abordagem de superpopulação . . . . .	12
5.1	Representação gráfica de uma RNA para um elemento. . . . .	20
7.1	População hipotética . . . . .	43
7.2	Log da população hipotética . . . . .	43
7.3	Histograma da variável $z$ . . . . .	43
7.4	Histograma da variável $y$ . . . . .	43
7.5	Histograma do $\ln(z)$ . . . . .	43
7.6	Histograma do $\ln(y)$ . . . . .	43
7.7	Representação gráfica da RNA a ser ajustada aos dados . . . . .	45
7.8	$\beta_{11} \times \beta_{21}$ - Estimativas da RNA clássica . . . . .	49
7.9	$B_1 \times B_2$ - Estimativas da RNA clássica . . . . .	50
7.10	$\beta_{11} \times \beta_{21}$ - Estimativas da RNA sob superpopulação . . . . .	50
7.11	$B_1 \times B_2$ - Estimativas da RNA sob superpopulação . . . . .	51
7.12	Variância estimada ( $\times 1000$ ) dos coeficientes da RNA por tipo e tamanho de amostra . . . . .	53
7.13	Viés estimado ( $\times 1000$ ) dos coeficientes estimados pela RNA por tipo e tamanho de amostra . . . . .	56
7.14	ERM estimados ( $\times 1000$ ) dos coeficientes estimados pela RNA por tipo e tamanho de amostra . . . . .	58
8.1	Estrutura da RNA para modelagem dos dados da PNAD 2014 . . . . .	65
B.1	Resíduos do modelo linear clássico da renda . . . . .	102
B.2	Resíduos do modelo linear clássico do log. natural da renda . . . . .	103
B.3	Resíduos do modelo linear de superpopulação da renda . . . . .	103
B.4	Resíduos do modelo linear de superpopulação do log. natural da renda	104
B.5	Resíduos da RNA clássica da renda . . . . .	104

B.6	Resíduos da RNA clássica do log. natural da renda . . . . .	105
B.7	Resíduos da RNA de superpopulação da renda . . . . .	105
B.8	Resíduos da RNA de superpopulação do log. natural da renda . . . . .	106
B.9	Evolução do SQE da RNA clássica da renda . . . . .	107
B.10	Evolução do SQE da RNA clássica da log. natural da renda . . . . .	107
B.11	Evolução do SQE da RNA de superpopulação da renda . . . . .	107
B.12	Evolução do SQE da RNA de superpopulação do log. natural da renda . . . . .	107

# Lista de Tabelas

7.1	Exemplo de amostragem estratificada simples com alocação igual e $n_h = 2$ . . . . .	45
7.2	Rótulos das amostras cujas estimativas apresentaram valores discrepantes . . . . .	52
7.3	Estatísticas resumo dos coeficientes estimados com as populações geradas . . . . .	52
7.4	Variância estimada ( $\times 1000$ ) dos coeficientes da RNA sob o modelo clássico (RNAC) e sob a abordagem de superpopulação (RNASP) por tamanho de amostra . . . . .	54
7.5	Viés estimado ( $\times 1000$ ) dos coeficientes da RNA sob o modelo clássico (RNAC) e sob a abordagem de superpopulação (RNASP) por tamanho de amostra . . . . .	57
7.6	ERM estimado ( $\times 1000$ ) dos coeficientes da RNA sob o modelo clássico (RNAC) e sob a abordagem de superpopulação (RNASP) por tamanho de amostra . . . . .	59
8.1	Variáveis independentes para estimação dos modelos . . . . .	61
8.2	Resultados do teste LM para o número de neurônios na camada intermediária . . . . .	64
8.3	Coefficiente de correlação múltipla, $R^2$ , em porcentagem, dos modelos estimados . . . . .	66
A.1	Brasil, Unidade da Federação e Região Metropolitana de Minas Gerais, fração de amostragem, composição da amostra por municípios, setores, unidades domiciliares e pessoas. PNAD - 2014 . . . . .	73
A.2	Tipo de domicílio . . . . .	74
A.3	Total de cômodos no domicílio . . . . .	74
A.4	Total de cômodos no domicílio servindo de dormitório . . . . .	75
A.5	Condição de ocupação do domicílio . . . . .	76
A.6	Número de banheiros ou sanitários no domicílio . . . . .	77
A.7	Esgotamento sanitário . . . . .	78

A.8	Existência de banheiro ou sanitário no domicílio . . . . .	79
A.9	Existência de rádio no domicílio . . . . .	80
A.10	Existência de televisão no domicílio . . . . .	80
A.11	Existência de máquina de lavar no domicílio . . . . .	80
A.12	Existência de geladeira ou freezer no domicílio . . . . .	81
A.13	Existência de computador no domicílio . . . . .	81
A.14	Existência de automóvel no domicílio . . . . .	81
A.15	Existência de cônjuge no domicílio . . . . .	82
A.16	Sexo do Responsável pelo domicílio . . . . .	83
A.17	Idade do responsável pelo domicílio . . . . .	85
A.18	Número de crianças de até sete anos presentes no domicílio . . . . .	86
A.19	Número de crianças de 7 a 12 anos presentes no domicílio . . . . .	86
A.20	Número de idosos acima de 60 anos no domicílio . . . . .	86
A.21	Cor ou raça do responsável pelo domicílio . . . . .	87
A.22	Número total de estudantes presentes no domicílio . . . . .	88
A.23	Posição na ocupação do responsável pelo domicílio . . . . .	90
B.1	Coeficientes da RNA - Clássica da com variável dependente renda . . .	98
B.2	Coeficientes da RNA - Clássica da com variável dependente o loga- ritmo natural da renda . . . . .	99
B.3	Coeficientes da RNA-Superpopulação com variável dependente a renda	100
B.4	Coeficientes da RNA-Superpopulação da variável dependente o loga- ritmo natural da renda . . . . .	101

# Capítulo 1

## Introdução

Diversos métodos estatísticos são utilizados para analisar dados provenientes de pesquisas amostrais. Tais métodos frequentemente pressupõem que os dados são provenientes de uma amostra aleatória simples, vide CASELLA e BERGER (2001). Porém, na prática, raramente a seleção de amostras para pesquisas é realizada por amostragem aleatória simples. Com frequência são utilizados esquemas mais complexos de amostragem, tais como amostragem estratificada e amostragem por conglomerados, por exemplo.

Os esquemas mais complexos de amostragem, frequentemente, refletem as estruturas complexas que compõe a população tais como hierarquia geográfica e/ou uma estratificação por sexo, por exemplo. Estas características da pesquisa, provenientes do esquema amostral utilizado, necessitam ser tratadas de forma adequada para a aplicação dos métodos estatísticos.

Os métodos estatísticos usuais ignoram o esquema de amostragem resultando na não adequação das suposições estocásticas. Por exemplo, assumimos que as observações para diferentes elementos são independentes, porém, na prática seria necessário considerar a correlação dos elementos obtidos por meio de amostragens por conglomerados. Ao pressupormos a independência dos elementos amostrais quando esta não ocorre somos levados a erros de interpretação em relação à análise de variabilidade das estimativas dos parâmetros estimados.

Nesses casos, inferências que se estendem à população devem ser baseadas no uso de informações sobre o plano amostral e dos pesos amostrais provenientes dos diferentes estágios de seleção, dependentes das probabilidades de inclusão dos elementos na amostra. Assim, pesquisas cujos dados são provenientes de um esquema amostral complexo ou refletem alguma estrutura complexa associada à população pesquisada são chamadas de *pesquisas amostrais complexas*.

Na literatura de estatística existem diferentes abordagens para a modelagem de dados provenientes de pesquisas amostrais complexas, como podem ser vistas em HANSEN *et al.* (1983) e PESSOA e DO NASCIMENTO SILVA (1998). Entretanto,

na literatura pertinente a redes neurais artificiais não existe menção sobre como proceder quando os dados são provenientes de pesquisas amostrais complexas como pode ser visto em SARLE (1994), HAYKIN (2001) e SILVA *et al.* (2010). Porém, o artigo de AMER (2009) propõe dois métodos para utilização de dados amostrais complexos na imputação de dados faltantes utilizando modelos de redes neurais artificiais. O primeiro é incluir os pesos amostrais na rede de forma similar ao que é feito quando utilizamos o método de estimação por mínimos quadrados ponderados, e o segundo é considerar a estrutura do esquema amostral complexo na construção da rede correspondente. Tais propostas não podem ser generalizadas para qualquer plano amostral complexo ou mesmo para redes neurais que não utilizam o método de estimação dos parâmetros por mínimos quadrados ponderados.

O objetivo deste trabalho é propor uma abordagem para a utilização de dados provenientes de pesquisas amostrais complexas em modelos de redes neurais artificiais cujo resultado poderá ser aplicado para diferentes esquemas amostrais complexos, não se limitando ao método de estimação dos parâmetros ou à incorporação do esquema amostral complexo a estrutura da rede como proposto em AMER (2009).

Assim, propomos uma abordagem de superpopulação para redes neurais, a qual permite a modelagem de dados provenientes de pesquisas amostrais complexas, como mostrado em PESSOA e DO NASCIMENTO SILVA (1998). No decorrer do trabalho obteremos a matriz de variâncias utilizando o método proposto em BINDER (1983) e a seguir apresentamos o cálculo da matriz de variâncias para amostras complexas. A simplificação do modelo será realizado utilizando o teste de hipótese descrito em WHITE (1989a) modificado de forma que possa ser aplicado ao problema de modelar dados amostrais complexos via redes neurais.

No Capítulo 2 fazemos uma introdução sobre conceitos iniciais de amostragem e de redes neurais artificiais, a fim de familiarizar o leitor com os temas tratados no trabalho. No Capítulo 3 apresentamos as duas abordagens usadas para modelagem estatística de dados amostrais: a abordagem clássica, a qual não considera a complexidade do plano amostral nas estimativas dos parâmetros de interesse do modelo; e a abordagem probabilística, que considera que os dados são obtidos de amostras complexas de elementos de uma população finita selecionados segundo um plano amostral probabilístico. O Capítulo 4 introduz a abordagem de superpopulação e exemplifica com um modelo de regressão linear. A rede neural clássica é apresentada no Capítulo 5. A abordagem de superpopulação para redes neurais é proposta no Capítulo 6. Uma avaliação empírica da abordagem de superpopulação para redes neurais por meio de simulação é realizada no Capítulo 7. Uma aplicação prática da abordagem proposta é realizada no Capítulo 8. As considerações finais são apresentadas no Capítulo 9.1.



# Capítulo 2

## Referencial e Conceitos Básicos

### 2.1 Amostragem

Amostragem é o processo de seleção de um subconjunto (denominado amostra) de elementos de uma *população finita*<sup>1</sup>, o qual é observado por meio de uma pesquisa e a seguir utilizado para estimar parâmetros de certas *características de interesse*<sup>2</sup> da população com determinada precisão. Tais parâmetros desconhecidos da população (média populacional, variância populacional, etc.), são os alvos de inferência usuais em pesquisas por amostragem.

O conjunto de procedimentos para inclusão de elementos da população na amostra é chamado plano amostral. As principais fases do plano amostral são: definição do objetivo do estudo, definição da população, escolha dos dados a pesquisar, escolha do método de amostragem e definição do tamanho da amostra.

Existem diversas técnicas de amostragem tais como:

- Amostragem probabilística: todos os elementos da população têm probabilidade conhecida e superior a zero de serem selecionados para fazer parte da amostra;
- Amostragem não probabilística:
  - Amostragem intencional: amostragem não probabilística subordinada a objetivos específicos do pesquisador;

---

<sup>1</sup>População finita: é a coleção de todos os elementos com uma dada característica comum referente ao público-alvo pretendido para a extração de informações usando uma amostra. Representa-se a população pelo conjunto  $U = \{1, \dots, N\}$  onde cada rótulo identifica um elemento da população e  $N$  é o tamanho da população, para mais detalhes vide PESSOA e DO NASCIMENTO SILVA (1998).

<sup>2</sup>Característica de Interesse é a nomenclatura utilizada para denotar a variável ou vetor de variáveis com informações observáveis associados a cada elemento da população.

- Amostragem não intencional: amostragem não probabilística regida por critérios de conveniência ou disponibilidade dos elementos a serem pesquisados.

No presente trabalho utilizaremos a técnica de amostragem probabilística, pois ela assegura que todos os subgrupos relevantes que constituem a população podem estar representados na amostra, e que os resultados obtidos com o estudo das características de interesse dos elementos da amostra podem ser generalizados para a população com determinada margem de erro.

Os tipos mais conhecidos de amostragem probabilística são listados a seguir.

**Amostragem Aleatória Simples:** É a forma mais simples de amostragem probabilística. A amostragem aleatória simples (sem reposição) de tamanho  $n$ , doravante referida como AAS, é realizada quando todos os possíveis subconjuntos de  $n$  unidades distintas na população possuem a mesma chance de pertencer à amostra. Por exemplo, se possuímos uma população formada por 6 elementos [A,B,C,D,E,F] e pretendemos selecionar uma AAS de 2 elementos, todos os subconjuntos de dois elementos possíveis [A,B], [A,C], [A,D], [A,E], [A,F], [B,C], [B,D], [B,E], [B,F], [C,D], [C,E], [C,F], [D,E], [D,F] e [E,F] terão a mesma probabilidade de seleção (neste caso, essa probabilidade é  $1/15$ ). Em consequência, cada um dos 6 elementos da população terá a mesma probabilidade de ser selecionado (neste caso, essa probabilidade é  $1/6$ ). Na AAS os valores amostrados não são independentes, ou seja, o sorteio de um elemento possui influência no sorteio do próximo elemento. Por outro lado, na amostragem aleatória simples com reposição, doravante referida como AASC, os valores são independentes, ou seja, o sorteio de um elemento não possui influência no sorteio do próximo elemento.

**Amostragem Estratificada Simples:** Nesse método, a população é dividida em subgrupos denominados estratos. Então uma AAS é selecionada independentemente de cada estrato. Os estratos são, frequentemente, subgrupos de interesse do pesquisador. Elementos pertencentes ao mesmo estrato muitas vezes possuem mais similaridades que elementos selecionados aleatoriamente de uma população. Nestes casos a estratificação aumenta a precisão. Na prática, os estratos podem ser definidos por qualquer variável, como por exemplo as características sociodemográficas dos indivíduos: sexo, idade, escolaridade, classe social, etc.

**Amostragem por Conglomerados:** Ocorre quando as unidades de interesse observadas na população são agregadas, segundo alguma característica conhecida na população sob estudo, em unidades maiores chamadas conglomerados. Em

seguida, sorteamos um determinado número de conglomerados e selecionamos alguns ou todos os elementos de cada conglomerado para pesquisas. Podemos citar como exemplo uma amostra de eleitores obtida pelo sorteio de um número de domicílios.

Qualquer amostra obtida por um método diferente da amostragem aleatória simples é dita uma amostra complexa, como mencionado em LEE (2006).

As vantagens da utilização de amostragem são: redução de custo, rapidez na obtenção dos resultados, e diminuição dos erros não amostrais na obtenção da informação (na coleta e tratamento da informação há sempre erros).

Na prática, a obtenção de estimativas relativas aos parâmetros de uma população por meio de amostragem é realizada utilizando *inferência estatística*, pois o parâmetro populacional é uma constante desconhecida. Assim, para estimar (obter um valor aproximado) um parâmetro populacional devemos considerar uma função que dependa somente dos valores da amostra. Tal função é denominada *estimador* do parâmetro de interesse. O valor do estimador calculado para uma determinada amostra é chamado *estimativa*.

## 2.2 Redes Neurais Artificiais

Redes neurais artificiais, doravante RNAs, são modelos estatísticos adaptativos baseados em uma estrutura análoga ao cérebro. Tais modelos são ditos adaptativos no sentido de que podem aprender a estimar os parâmetros de alguma população utilizando um pequeno número de elementos. As RNAs, essencialmente, não diferem de um modelo estatístico usual, pois podemos encontrar arquiteturas de redes neurais semelhantes à análise discriminante, análise de componentes principais, regressão logística e regressão linear, entre outras técnicas estatísticas. Exemplos destas arquiteturas podem ser encontrados em SARLE (1994). De fato, o mesmo arcabouço matemático pode ser usado para avaliação de ambos os modelos. As RNAs são usadas em aplicações nas áreas de psicologia, estatística, engenharia, econometria, entre outras.

Basicamente, as RNAs são formadas por unidades chamadas de neurônios por analogia ao cérebro. Essas unidades são interligadas por um conjunto de pesos chamados de pesos sinápticos. A aprendizagem é, geralmente, realizada pela modificação (atualização) dos pesos sinápticos. Cada neurônio corresponde a uma característica de um determinado padrão que desejamos analisar ou predizer. Os neurônios são organizados em camadas. O algoritmo de aprendizagem comumente usado é o algoritmo ‘*backpropagation*’, vide DU e SWAMY (2006) e HAYKIN (2001), o qual será descrito detalhadamente nos capítulos posteriores.

As RNAs possuem, frequentemente, múltiplas camadas. A primeira é sempre a camada de entrada e a última é a camada de saída. As camadas internas (entre a camada de saída e a camada de entrada) são chamadas de camadas intermediárias. A informação a ser analisada é apresentada aos neurônios da primeira camada e, a seguir, é propagada aos neurônios da camada intermediária para processamento e, a seguir, é novamente propagada até a camada de saída, resultando na informação processada. A informação processada é comparada ao valor real, resultando no cálculo do erro. A seguir este erro é retropropagado através da RNA para correção dos pesos sinápticos. Este processo é realizado iterativamente até que o erro possua um valor mínimo estipulado pelo modelista. Este processo é a base do algoritmo ‘*backpropagation*’.

O objetivo das RNA é o aprendizado, ou a descoberta, de alguma relação entre as variáveis de entrada e as variáveis de saída. O aprendizado é resultado da modificação (atualização) dos pesos sinápticos entre os neurônios. Em termos estatísticos isto é equivalente à estimação dos parâmetros do modelo.

As RNA são frequentemente utilizadas para modelar fenômenos de natureza não linear ou fenômenos onde os modelos estatísticos usuais não são adequados.

No Capítulo 3 introduziremos os conceitos de abordagem clássica e abordagem probabilística presentes no modelo estatístico e que, posteriormente, estenderemos para o caso dos modelos de redes neurais.

# Capítulo 3

## Abordagens Clássica e Probabilística para Modelagem de Dados Amostrais

### 3.1 Abordagens Clássica e Probabilística

Analistas frequentemente realizam estudos com dados obtidos por pesquisas amostrais considerando que as observações foram selecionadas de forma independente e equiprovável, ou seja, são provenientes de uma AASC. Porém, na prática, a seleção de uma amostra envolve mais complexidade que uma AASC. Algumas observações da amostra podem ser selecionadas com maior probabilidade que outras e algumas podem ser incluídas na amostra com probabilidade igual a 1, como mencionado em LOHR (1999).

A *abordagem clássica*<sup>1</sup>, onde se presume que os dados são provenientes de uma AASC, não leva em consideração a complexidade do plano amostral, e portanto, as estimativas dos parâmetros de interesse gerados pelo modelo poderão ser *enviesadas*<sup>2</sup> sob o plano amostral, quando forem ignorados os pesos distintos das observações. Além disso, as estimativas das variâncias são influenciadas pela conglomeração, estratificação e pesos. A pesquisa cuja amostragem é diferente da AASC é complexa na sua concepção e necessita de considerações analíticas especiais, como mencionado em LEE (2006).

Do ponto de vista da Inferência Estatística o modelo usual (na abordagem clássica), fig.(3.1), considera as observações  $y_1, \dots, y_n$  de uma variável  $Y$  na amostra

---

<sup>1</sup>A abordagem clássica é útil em estudos analíticos como, por exemplo, na investigação da natureza das associações entre variáveis.

<sup>2</sup>Segundo a definição presente em LARSON (1969), um estimador  $\hat{\theta}$  de um parâmetro desconhecido  $\theta$  é não enviesado se  $E(\hat{\theta}) = \theta$  para todos os valores de  $\theta \in \Theta$ , onde  $E(\bullet)$  é o operador média. A diferença  $B(\hat{\theta}) = E(\hat{\theta}) - \theta$  é chamada de *viés* de  $\hat{\theta}$ : se  $B(\hat{\theta}) \neq 0$  então  $\hat{\theta}$  é um estimador enviesado para  $\theta$ .

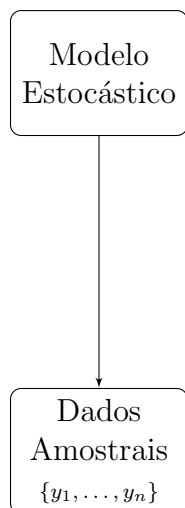


Figura 3.1: Representação gráfica da abordagem clássica

de tamanho  $n$ , como realizações de variáveis aleatórias  $Y_1, \dots, Y_n$  *independentes*<sup>3</sup> e *identicamente distribuídas*<sup>4</sup> com a mesma distribuição da variável de interesse  $Y$ . Considera também  $f(y; \theta)$ , a função de densidade de  $Y$ , onde  $\theta \in \Theta$  é o parâmetro indexador da distribuição  $f$  e  $\Theta$  é o espaço paramétrico. A partir das observações  $y_1, \dots, y_n$  são feitas inferências a respeito do parâmetro  $\theta$ . Na abordagem clássica o modelo estocástico é formulado para descrever diretamente os dados amostrais. Além disso, o parâmetro de interesse  $\theta$  é um atributo do modelo proposto, e não de uma população específica.

A *abordagem por amostragem probabilística* considera que os dados são obtidos de amostras de elementos de uma população finita selecionados segundo um plano amostral, possivelmente, com probabilidades de seleção distintas, vide BOLFARINE e BUSSAB (2005). A partir dos valores observados na amostra são realizadas inferências descritivas a respeito de funções particulares de valores populacionais tais como totais, médias, proporções, etc.

<sup>3</sup>As variáveis aleatórias  $Z_1$  e  $Z_2$  são ditas independentes se  $f(z_1, z_2) = f(z_1)f(z_2) \forall z$ .

<sup>4</sup>As variáveis aleatórias  $Z_1$  e  $Z_2$  são identicamente distribuídas se possuem a mesma distribuição de probabilidade.

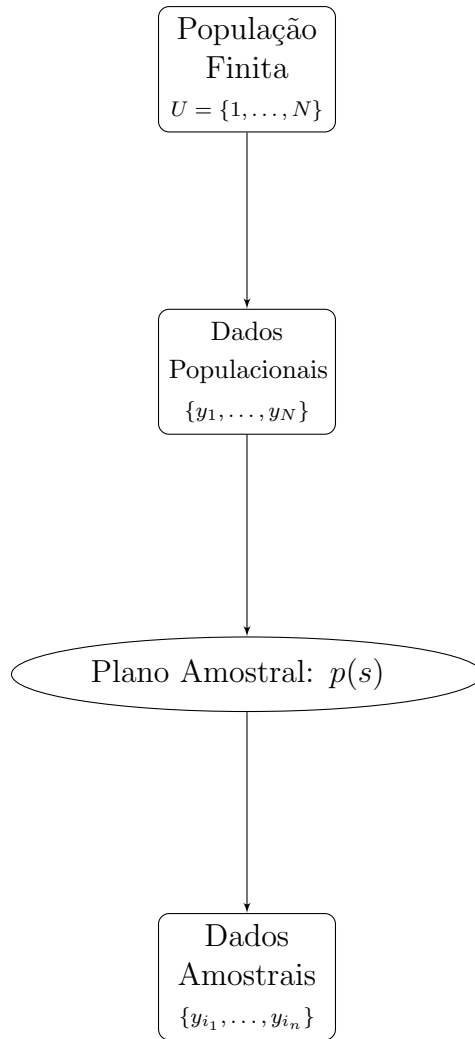


Figura 3.2: Representação gráfica da abordagem por amostragem probabilística

A fig.(3.2) contém a representação gráfica da abordagem por amostragem probabilística. Tal abordagem, do ponto de vista da Inferência Estatística, considera que a partir de valores da variável de interesse  $Y$  observados na amostra, denotados por  $y_{i_1}, \dots, y_{i_n}$ , podem ser feitas inferências a respeito de funções de valores populacionais, tal como  $g(y_1, \dots, y_N)$ . No geral, o objetivo é realizar estudos descritivos utilizando funções de  $g$  particulares, como por exemplo totais  $g(y_1, \dots, y_N) = \sum_{i=1}^N y_i$ , médias  $g(y_1, \dots, y_N) = \frac{1}{N} \sum_{i=1}^N y_i$ , entre outras.

Tal abordagem foi concebida para fins descritivos, porém é cada vez mais comum a utilização de dados obtidos através de pesquisas amostrais complexas para objetivos analíticos, com a aplicação de métodos de análise da modelagem clássica.

## 3.2 O estimador de Horwitz-Thompson

O estimador de Horwitz-Thompson, HORVITZ e THOMPSON (1952), é um estimador não tendencioso do total populacional, o qual é utilizado para tratar amostras retiradas de uma população finita com probabilidades desiguais de seleção. Ao atribuímos um vetor de probabilidades escolhido adequadamente, é possível reduzir as variâncias das estimativas não enviesadas se comparadas as estimativas obtidas com probabilidades iguais de inclusão na amostra.

O estimador de Horwitz-Thompson, doravante referido como HT, do total populacional é dado por:

$$\hat{\tau}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i} \quad (3.1)$$

onde  $\pi_i$  é a probabilidade de inclusão do  $i$ -ésimo elemento na amostra.

As propriedades do estimador, enunciadas e demonstradas em BOLFARINE e BUSSAB (2005), são:

**Teorema 3.2.1.** *Se as unidades amostrais são selecionados sem reposição, com probabilidades de inclusão  $\pi_i$  e  $\pi_{ij}$ , tem-se:*

$$E[\hat{\tau}_{HT}] = \sum_{i \in U} y_i \quad (3.2)$$

e

$$V[\hat{\tau}_{HT}] = \sum_{i \in U} \frac{1 - \pi_i}{\pi_i} y_i^2 + 2 \sum_{i \in U} \sum_{i < j \in U} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} y_i y_j \quad (3.3)$$

**Teorema 3.2.2.** *Um estimador não viciado de  $V[\hat{\tau}_{HT}]$  é:*

$$\hat{V} = \sum_{i \in s} \frac{1 - \pi_i}{\pi_i^2} y_i^2 + 2 \sum_{i \in s} \sum_{i < j \in s} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j \pi_{ij}} y_i y_j \quad (3.4)$$

O próximo capítulo apresenta uma abordagem estatística que propõe um modelo parametrizado que incorpora a estrutura do plano amostral. Tal abordagem é denominada de modelagem de superpopulação.



# Capítulo 4

## Abordagem de Superpopulação

### 4.1 Modelagem de superpopulação

A *modelagem de superpopulação* propõe um modelo parametrizado, como na abordagem clássica, incorporando na análise e ajuste do modelo a estrutura do plano amostral, como na abordagem da amostragem probabilística, vide PESSOA e DO NASCIMENTO SILVA (1998). Nessa abordagem, os valores  $y_1, \dots, y_N$  da variável de interesse  $Y$  na população finita são considerados observações ou realizações das variáveis aleatórias  $Y_1, \dots, Y_N$ , supostas *independentes e identicamente distribuídas*, doravante referido como IID<sup>1</sup>, com função de densidade  $f(y; \theta)$ , onde  $\theta \in \Theta$  é o parâmetro indexador da distribuição  $f$ , e  $\Theta$  é o espaço paramétrico. Este modelo é denominado *modelo de superpopulação*. Utilizando o plano amostral definido por  $p(s)$ <sup>2</sup>, obtemos os valores na amostra  $y_{i_1}, \dots, y_{i_n}$ . A partir de  $y_{i_1}, \dots, y_{i_n}$  (observações de variáveis aleatórias não consideradas IID, em geral) desejamos inferir sobre o parâmetro  $\theta$ . A fig.(4.1) contém a representação gráfica da abordagem de superpopulação.

---

<sup>1</sup>Para mais detalhes sobre variáveis aleatórias IID vide CASELLA e BERGER (2001).

<sup>2</sup>Considere a população finita  $U = \{1, \dots, N\}$ , da qual é selecionada uma amostra  $s = \{i_1, \dots, i_n\}$ , segundo um plano amostral caracterizado por  $p(s)$ , probabilidade de ser selecionada a amostra  $s$ , suposta calculável para todas as possíveis amostras.

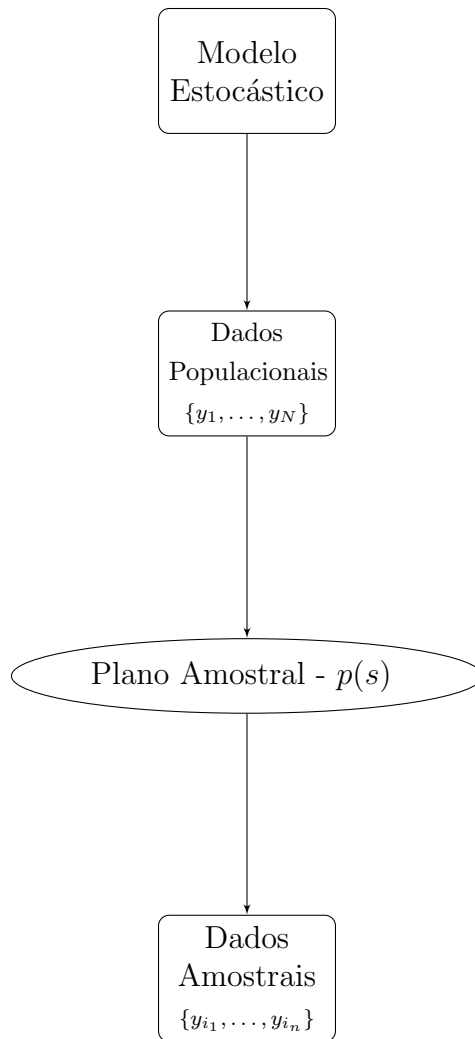


Figura 4.1: Representação gráfica da abordagem de superpopulação

O método de estimação que consideraremos para estimar os parâmetros do modelo de superpopulação é o *método de máxima pseudo-verossimilhança*. CHAMBERS e SKINNER (2003) apresenta uma boa discussão sobre esse método. Assim, suponha que os vetores observados  $\mathbf{y}_i$  das variáveis de pesquisa do elemento  $i$  são gerados por vetores aleatórios  $\mathbf{Y}_i$ , para  $i \in U$ . Além disso, suponha que  $\mathbf{Y}_1, \dots, \mathbf{Y}_N$  são IID com função de densidade  $f(\mathbf{y}, \boldsymbol{\theta})$ . Se todos os elementos da população finita  $U$  fossem conhecidos, as funções de verossimilhança e de log-verossimilhança *populacionais* seriam dadas respectivamente por:

$$l_U(\boldsymbol{\theta}) = \prod_{i \in U} f(\mathbf{y}_i; \boldsymbol{\theta}) \quad (4.1)$$

$$L_U(\boldsymbol{\theta}) = \sum_{i \in U} \log[f(\mathbf{y}_i; \boldsymbol{\theta})] \quad (4.2)$$

As equações de verossimilhança populacionais são dadas por:

$$\sum_{i \in U} \mathbf{u}_i(\boldsymbol{\theta}) = 0 \quad (4.3)$$

onde

$$\mathbf{u}_i(\boldsymbol{\theta}) = \frac{\partial \log[f(\mathbf{y}_i; \boldsymbol{\theta})]}{\partial \boldsymbol{\theta}} \quad (4.4)$$

é o vetor  $K \times 1$  dos *escores* do elemento  $i$ ,  $i \in U$ . Sob condições de regularidade, vide BINDER (1983), a solução  $\boldsymbol{\theta}_U$  do sistema da eq.(4.3) é o *Estimador de Máxima Verossimilhança* de  $\boldsymbol{\theta}$  no caso de um censo. Podemos considerar  $\boldsymbol{\theta}_U$  como uma *Quantidade Descritiva Populacional Correspondente a  $\boldsymbol{\theta}$* , doravante QDPC, sobre a qual se deseja fazer inferências com base em informações da amostra.

A QDPC  $\boldsymbol{\theta}_U$  definida com base na eq.(4.3) não é calculável a menos que um censo seja realizado. Entretanto, desempenha papel fundamental nessa abordagem inferencial, por constituir-se num pseudo-parâmetro, eleito como alvo da inferência num esquema que incorpora o planejamento amostral. Isto se justifica porque, sob certas condições de regularidade,  $\boldsymbol{\theta}_U - \boldsymbol{\theta} = o_p(1)$ . Como em pesquisas por amostragem o tamanho da população é geralmente grande, um estimador adequado para  $\boldsymbol{\theta}_U$  será geralmente adequado para  $\boldsymbol{\theta}$ .

Seja  $\mathbf{T}(\boldsymbol{\theta}) = \sum_{i \in U} \mathbf{u}_i(\boldsymbol{\theta})$  a soma dos vetores escores na população, o qual é um vetor de totais populacionais. Para estimar este vetor de totais, podemos usar um estimador linear ponderado da forma  $\widehat{\mathbf{T}}(\boldsymbol{\theta}) = \sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\theta})$  onde  $w_i$  são pesos apropriadamente definidos. Com essa notação, podemos agora obter um estimador para  $\boldsymbol{\theta}_U$  resolvendo o sistema de equações obtido igualando o estimador  $\widehat{\mathbf{T}}(\boldsymbol{\theta})$  do total  $\mathbf{T}(\boldsymbol{\theta})$  a zero.

**Definição 4.1.1.** O estimador de Máxima Pseudo-Verossimilhança  $\hat{\boldsymbol{\theta}}_{MPV}$  de  $\boldsymbol{\theta}_U$  será a solução das equações de Pseudo-Verossimilhança dadas por:

$$\hat{\mathbf{T}}(\boldsymbol{\theta}) = \sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\theta}) = \mathbf{0} \quad (4.5)$$

Através da linearização de Taylor, vide PESSOA e DO NASCIMENTO SILVA (1998), podemos obter a variância assintótica do estimador  $\hat{\boldsymbol{\theta}}_{MPV}$  e seu estimador correspondente:

$$V_p(\hat{\boldsymbol{\theta}}_{MPV}) \cong [J(\boldsymbol{\theta}_U)]^{-1} V_p \left[ \sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\theta}_U) \right] [J(\boldsymbol{\theta}_U)]^{-1} \quad (4.6)$$

e

$$\hat{V}_p(\hat{\boldsymbol{\theta}}_{MPV}) = [\hat{J}(\hat{\boldsymbol{\theta}}_{MPV})]^{-1} \hat{V}_p \left[ \sum_{i \in s} w_i \mathbf{u}_i(\hat{\boldsymbol{\theta}}_{MPV}) \right] [\hat{J}(\hat{\boldsymbol{\theta}}_{MPV})]^{-1} \quad (4.7)$$

onde

$$J(\boldsymbol{\theta}_U) = \frac{\partial \mathbf{T}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_U} = \sum_{i \in U} \frac{\partial \mathbf{u}_i(\boldsymbol{\theta})}{\partial(\boldsymbol{\theta})} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_U} \quad (4.8)$$

$$\hat{J}(\hat{\boldsymbol{\theta}}_{MPV}) = \frac{\partial \hat{\mathbf{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{MPV}} = \sum_{i \in s} w_i \frac{\partial \mathbf{u}_i(\boldsymbol{\theta})}{\partial(\boldsymbol{\theta})} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{MPV}} \quad (4.9)$$

$V_p[\sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\theta}_U)]$  é a matriz de variância do estimador do total populacional dos escores e  $\hat{V}_p[\sum_{i \in s} w_i \mathbf{u}_i(\hat{\boldsymbol{\theta}}_{MPV})]$  é um estimador consistente para esta variância. A distribuição assintótica do estimador  $\hat{\boldsymbol{\theta}}_{MPV}$  é normal multivariada, o que fornece base para a inferência, vide BINDER (1983).

## 4.2 Modelo de Regressão Linear

A fim de exemplificar a modelagem de superpopulação mostraremos como estimar os parâmetros de um modelo de regressão linear normal para um subconjunto das variáveis da pesquisa, utilizando o procedimento de *máxima pseudo-verossimilhança*. As seções seguintes, deste capítulo, têm o conteúdo baseado em PESSOA e DO NASCIMENTO SILVA (1998), capítulo 6.

### 4.2.1 Especificação do Modelo

Considere um vetor  $\mathbf{z}_i = (z_{i1}, \dots, z_{iP})'$  de dimensão  $P \times 1$  com os valores de variáveis explicativas  $\mathbf{z}$  de um modelo de regressão utilizado para explicar os valores de

uma variável  $Y$ , considerada como variável resposta, vide PESSOA e DO NASCIMENTO SILVA (1998). Denotamos por  $Y_i$  e  $\mathbf{Z}_i$  a variável e o vetor aleatórios que geram  $y_i$  e  $\mathbf{z}_i$  para  $i \in U$ , respectivamente.

Suponha que  $(\mathbf{Y}_i, \mathbf{Z}_i)'$ ,  $i \in U$ , são vetores aleatórios independentes tais que:

$$f(y_i|\mathbf{z}_i; \boldsymbol{\beta}, \sigma_e^2) = (2\pi\sigma_e^2)^{-1/2} \exp \left[ - (y_i - \mathbf{z}_i' \boldsymbol{\beta})^2 / 2\sigma_e^2 \right] \quad (4.10)$$

onde  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$  e  $\sigma_e^2 > 0$  são parâmetros desconhecidos do modelo.

As funções escore para  $\boldsymbol{\beta}$  e  $\sigma_e^2$  correspondentes ao modelo da eq.(4.10) são dadas por (vide PESSOA e DO NASCIMENTO SILVA (1998)):

$$\partial \log[f(y_i|\mathbf{z}_i; \boldsymbol{\beta}, \sigma_e^2)] / \partial \boldsymbol{\beta} = \mathbf{z}_i (y_i - \mathbf{z}_i' \boldsymbol{\beta}) / \sigma_e^2 \propto \mathbf{z}_i (y_i - \mathbf{z}_i' \boldsymbol{\beta}) = \mathbf{u}_i(\boldsymbol{\beta}) \quad (4.11)$$

e

$$\partial \log[f(y_i|\mathbf{z}_i; \boldsymbol{\beta}, \sigma_e^2)] / \partial \sigma_e^2 = [(y_i - \mathbf{z}_i' \boldsymbol{\beta})^2 - \sigma_e^2] / 2\sigma_e^4 \propto (y_i - \mathbf{z}_i' \boldsymbol{\beta})^2 - \sigma_e^2 = u_i(\sigma_e^2) \quad (4.12)$$

## 4.2.2 Pseudo-parâmetros do Modelo

Supondo que todos os elementos da população tivessem sido pesquisados (ou seja, que tivéssemos realizado um censo), os Estimadores de Máxima Verossimilhança, doravante EMV, de  $\boldsymbol{\beta}$  e  $\sigma_e^2$ , denotados por  $\mathbf{B}$  e  $S_e^2$  respectivamente, seriam obtidos resolvendo as equações de verossimilhança do censo:

$$\sum_{i \in U} u_i(\mathbf{B}) = \sum_{i \in U} \mathbf{z}_i (y_i - \mathbf{z}_i' \mathbf{B}) = \mathbf{z}'_U \mathbf{y}_U - (\mathbf{z}'_U \mathbf{z}_U) \mathbf{B} = \mathbf{0} \quad (4.13)$$

e

$$\sum_{i \in U} u_i(S_e^2) = \sum_{i \in U} [(y_i - \mathbf{z}_i' \mathbf{B})^2 - S_e^2] = (\mathbf{y}_U - \mathbf{z}'_U \mathbf{B})' (\mathbf{y}_U - \mathbf{z}'_U \mathbf{B}) - N S_e^2 = 0 \quad (4.14)$$

onde  $\mathbf{z}_U = (\mathbf{z}_1, \dots, \mathbf{z}_N)'$  e  $\mathbf{y}_U = (y_1, \dots, y_N)'$ .

Se  $\mathbf{z}'_U \mathbf{z}_U$  for não singular, as soluções para estas equações são:

$$\mathbf{B} = (\mathbf{z}'_U \mathbf{z}_U)^{-1} \mathbf{z}'_U \mathbf{y}_U \quad (4.15)$$

e

$$S_e^2 = N^{-1} \sum_{i \in U} (y_i - \mathbf{z}_i' \mathbf{B})^2 = N^{-1} (\mathbf{y}_U - \mathbf{z}'_U \mathbf{B})' (\mathbf{y}_U - \mathbf{z}'_U \mathbf{B}) \quad (4.16)$$

Com uma parametrização que isole o termo correspondente ao intercepto (primeira coluna do vetor  $\mathbf{z}_i$ ) do modelo de regressão eq.(4.10), como mostrado em WOOLDRIDGE (2008) os EMVs de  $\boldsymbol{\beta}_2$  (igual a  $\boldsymbol{\beta}$  excluído o primeiro componente),  $\beta_1$  e  $\sigma_e^2$  são dados por:

$$\mathbf{B}_2 = \mathbf{S}_z^{-1} \mathbf{S}_{zy}, \quad (4.17)$$

$$B_1 = \bar{Y} - \bar{\mathbf{Z}}' \mathbf{B}_2, \quad (4.18)$$

e

$$S_e^2 = N^{-1} \sum_{i \in U} (y_i - B_1 - \mathbf{z}_i' \mathbf{B}_2)^2 = N^{-1} \sum_{i \in U} e_i^2, \quad (4.19)$$

onde  $\bar{Y} = N^{-1} \sum_{i \in U} y_i$ ,  $\bar{\mathbf{Z}} = N^{-1} \sum_{i \in U} \mathbf{z}_i$ ,  $\mathbf{S}_z = N^{-1} \sum_{i \in U} (\mathbf{z}_i - \bar{\mathbf{Z}})(\mathbf{z}_i - \bar{\mathbf{Z}})'$ ,  $\mathbf{S}_{zy} = N^{-1} \sum_{i \in U} (\mathbf{z}_i - \bar{\mathbf{Z}})(y_i - \bar{Y})$  e  $e_i = y_i - B_1 - \mathbf{z}_i' \mathbf{B}_2 = (y_i - \bar{Y}) - (\mathbf{z}_i - \bar{\mathbf{Z}})' \mathbf{B}_2$ , sendo neste trecho os vetores de variáveis preditoras tomados sem o termo constante referente ao intercepto.

Os EMVs do censo, dados em eq.(4.17) a eq.(4.19), coincidem com os estimadores de mínimos quadrados ordinários, sob hipóteses mais fracas do modelo dadas por NATHAN e HOLT (1980) (apud PESSOA e DO NASCIMENTO SILVA (1998)), onde se dispensou a hipótese de normalidade dos erros, ou seja:

$$\begin{aligned} E_M(Y_i | \mathbf{Z}_i = \mathbf{z}_i) &= \beta_1 + \mathbf{z}_i' \boldsymbol{\beta}_2 \\ V_M(Y_i | \mathbf{Z}_i = \mathbf{z}_i) &= \sigma_e^2 \\ COV_M(Y_i, Y_j | \mathbf{Z}_i = \mathbf{z}_i, \mathbf{Z}_j = \mathbf{z}_j) &= 0, \quad \forall i \neq j \in U. \end{aligned} \quad (4.20)$$

### 4.2.3 Estimadores de Máxima Pseudo-Verossimilhança dos Parâmetros do Modelo

Quando apenas uma amostra de unidades da população é observada, são utilizados os pesos amostrais  $w_i$  para obtenção dos estimadores de máxima pseudo-verossimilhança de  $\boldsymbol{\beta}$  e  $\sigma_e^2$ , ou alternativamente de  $\mathbf{B}$  e  $S_e^2$ , se as quantidades descritivas populacionais correspondentes forem escolhidas como o alvo da inferência. Se os pesos  $w_i$  satisfizerem às condições de regularidade, vide PESSOA e DO NASCIMENTO SILVA (1998), será imediato obter as equações de pseudo-verossimilhança correspondentes ao modelo eq.(4.10) como:

$$\sum_{i \in s} w_i \mathbf{u}_i(\widehat{\mathbf{B}}_w) = \sum_{i \in s} w_i z_i (y_i - z_i' \widehat{\mathbf{B}}_w) = \mathbf{z}'_s \mathbf{W}_s \mathbf{y}_s - (\mathbf{z}'_s \mathbf{W}_s \mathbf{z}_s) \widehat{\mathbf{B}}_w = \mathbf{0} \quad (4.21)$$

e

$$\begin{aligned} \sum_{i \in s} w_i u_i(s_{e,w}^2) &= \sum_{i \in s} w_i \left[ (y_i - z_i' \widehat{\mathbf{B}}_w)^2 - s_{e,w}^2 \right] = \\ &= (\mathbf{y}_s - \mathbf{z}_s \widehat{\mathbf{B}}_w)' \mathbf{W}_s (\mathbf{y}_s - \mathbf{z}_s \widehat{\mathbf{B}}_w) - (\mathbf{1}'_s \mathbf{W}_s \mathbf{1}_s) s_{e,w}^2 = 0 \end{aligned} \quad (4.22)$$

onde  $\mathbf{z}_s$  e  $\mathbf{y}_s$  são os análogos amostrais de  $\mathbf{z}_U$  e  $\mathbf{y}_U$ , respectivamente,  $\mathbf{W}_s = \text{diag}[(w_{i_1}, \dots, w_{i_n})]$  é uma matriz diagonal  $n \times n$  com os pesos dos elementos da amostra na diagonal principal, e  $\widehat{\mathbf{B}}_w$  e  $s_{e,w}^2$  são estimadores de MPV de  $\boldsymbol{\beta}$  e  $\sigma_e^2$ , respectivamente.

Supondo que  $\mathbf{z}'_s \mathbf{W}_s \mathbf{z}_s$  é não singular e resolvendo as equações eq.(4.21) e eq.(4.22) em  $\widehat{\mathbf{B}}_w$  e  $s_{e,w}^2$ , obtemos as seguintes expressões para os estimadores MPV dos parâmetros do modelo:

$$\widehat{\mathbf{B}}_w = (\mathbf{z}'_s \mathbf{W}_s \mathbf{z}_s)^{-1} \mathbf{z}'_s \mathbf{W}_s \mathbf{y}_s \quad (4.23)$$

e

$$\begin{aligned} s_e^w &= (\mathbf{1}'_s \mathbf{W}_s \mathbf{1}_s)^{-1} (\mathbf{y}_s - \mathbf{z}'_s \widehat{\mathbf{B}}_w)' \mathbf{W}_s (\mathbf{y}_s - \mathbf{z}'_s \widehat{\mathbf{B}}_w) \\ &= (\mathbf{1}'_s \mathbf{W}_s \mathbf{1}_s)^{-1} \mathbf{y}'_s \left[ \mathbf{W}_s - \mathbf{W}_s \mathbf{z}_s (\mathbf{z}'_s \mathbf{W}_s \mathbf{z}_s)^{-1} \mathbf{z}'_s \mathbf{W}_s \right] \mathbf{y}_s \end{aligned} \quad (4.24)$$

sendo a segunda expressão para  $s_{e,w}^2$  obtida mediante substituição do valor de  $\widehat{\mathbf{B}}_w$  em eq.(4.23) na primeira eq.(4.24).

#### 4.2.4 Estimadores da Variância de Estimadores de Máxima Pseudo-Verossimilhança

A avaliação da precisão das estimativas dos parâmetros é realizada utilizando-se as estimativas das variâncias correspondentes. Concentraremos atenção na estimação das variâncias dos estimadores de MPV dos coeficientes de regressão  $\boldsymbol{\beta}$ . As expressões a seguir são obtidas por aplicação direta dos resultados gerais fornecidos na seção 4.1, observando-se que os escores correspondentes a  $\boldsymbol{\beta}$  no ajuste do censo do modelo eq.(4.10) são dados por  $\mathbf{u}_i(\mathbf{B}) = z_i(y_i - z_i' \mathbf{B}) = z_i e_i$ , onde  $e_i = (y_i - \bar{Y}) - (z_i - \bar{Z})' \mathbf{B}$  para  $i \in U$ , com o Jacobiano correspondente dado

por:

$$J(\mathbf{B}) = \sum_{i \in U} \frac{\partial \mathbf{u}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i \in U} \frac{\partial \mathbf{z}_i(y_i - \mathbf{z}'_i \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial (\mathbf{z}'_U \mathbf{y}_U - \mathbf{z}'_U \mathbf{z}_U \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\mathbf{z}'_U \mathbf{z}_U, \quad (4.25)$$

onde  $\boldsymbol{\beta} = \mathbf{B}$

Substituindo em eq.(4.6) e eq.(4.7) os valores dos escores, do jacobiano e dos estimadores  $\pi$  – ponderados correspondentes (também chamados de estimadores de Horwitz-Thompson) , obtemos as seguintes expressões para a variância assintótica de aleatorização do estimador de MPV padrão  $\widehat{\mathbf{B}}_\pi$  e seu estimador consistente, dadas por:

$$V_p(\widehat{\mathbf{B}}_\pi) = (-\mathbf{z}'_U \mathbf{z}_U)^{-1} V_p \left( \sum_{i \in s} \pi_i^{-1} \mathbf{z}_i e_i \right) (-\mathbf{z}'_U \mathbf{z}_U)^{-1} \quad (4.26)$$

e

$$\widehat{V}_p(\widehat{\mathbf{B}}_\pi) = (-\mathbf{z}'_s \Pi_s^{-1} \mathbf{z}_s)^{-1} \widehat{V}_p \left( \sum_{i \in s} \pi_i^{-1} \mathbf{z}_i e_i \right) (-\mathbf{z}'_s \Pi_s^{-1} \mathbf{z}_s)^{-1} \quad (4.27)$$

onde

$$V_p \left( \sum_{i \in s} \pi_i^{-1} \mathbf{z}_i e_i \right) = \sum_{i \in U} \sum_{j \in U} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} e_i \mathbf{z}_i \mathbf{z}'_j e_j, \quad (4.28)$$

$$\widehat{V}_p \left( \sum_{i \in s} \pi_i^{-1} \mathbf{z}_i \widehat{e}_i \right) = \sum_{i \in s} \sum_{j \in s} (\pi_i^{-1} \pi_j^{-1} - \pi_{ij}^{-1}) \widehat{e}_i \mathbf{z}_i \mathbf{z}'_j \widehat{e}_j, \quad (4.29)$$

$\widehat{e}_i = y_i - \mathbf{z}'_i \widehat{\mathbf{B}}_\pi$  para  $i \in s$  e  $\Pi_s = \text{diag}[(\pi_{i_1}, \pi_{i_2}, \dots, \pi_{i_n})]$  é uma matriz diagonal  $n \times n$  com as probabilidades de inclusão de primeira ordem na diagonal principal.



# Capítulo 5

## Modelo Clássico de Rede Neural

A seguir introduziremos o modelo clássico de Rede Neural Artificial, doravante RNA. O termo modelo clássico de rede neural será utilizado para nomear a abordagem que considera que os dados são provenientes de uma AASC, vide HAYKIN (2001). Apresentaremos a *nomenclatura*<sup>1</sup> e detalharemos o *algoritmo backpropagation*, vide SILVA *et al.* (2010), doravante denominado BP, utilizado para estimação dos *parâmetros ou coeficientes da rede*<sup>2</sup>.

### 5.1 Abordagem clássica de RNA

A RNA é um conjunto de unidades computacionais simples altamente interconectadas. As unidades são chamadas de nós e representam os neurônios biológicos. A rede abordada no presente trabalho para um elemento da amostra observada está representada na fig.(5.1). Os neurônios são representados graficamente pelos círculos. As conexões entre os neurônios são unidirecionais e são representadas pelas setas pontilhadas e representam as conexões sinápticas no cérebro. Cada conexão tem um peso associado chamado de peso sináptico que pode ser interpretado como a intensidade do sinal.

---

<sup>1</sup>A nomenclatura usada na literatura de RNA é diferente da usada na literatura estatística. Sendo assim, procuramos compatibilizá-las para que leitores de ambas as áreas possam ler o material sem dificuldades.

<sup>2</sup>Na nomenclatura de RNA os coeficientes da rede a serem estimados são conhecidos como pesos sinápticos.

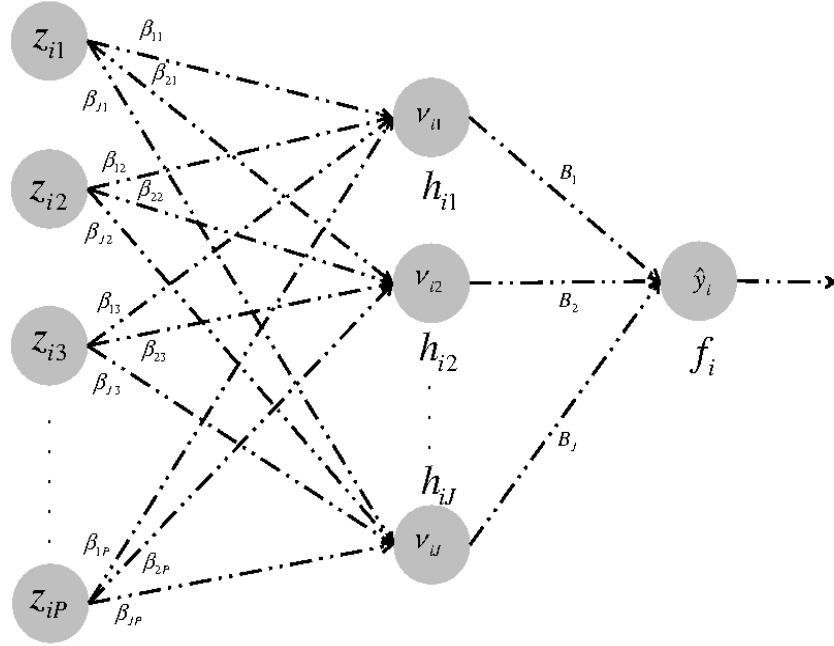


Figura 5.1: Representação gráfica de uma RNA para um elemento.

A estimação do conjunto de pesos sinápticos (ou coeficientes da RNA),  $\{B_1, \dots, B_J; \beta_{11}, \dots, \beta_{JP}\}$ , é realizada utilizando os dados disponíveis  $\{y_{i_1}, \dots, y_{i_n}\}$  de valores da resposta observados na amostra, e os valores  $\{z_{i_1}, \dots, z_{i_n}\}$  das variáveis preditoras da amostra observada.

O modelo para obtenção do valor predito pela RNA para ao elemento  $i$  se inicia com os valores das variáveis explicativas,  $\{z_{i1}, \dots, z_{iP}\}$ , sendo apresentados para a camada de entrada, a qual não realiza operações com estas informações, simplesmente, as passa para a camada intermediária. O valor  $v_{ij}$  do  $j$ -ésimo neurônio da camada intermediária para o elemento  $i$  é obtido em duas etapas. Primeiramente, calcula-se o valor da transformação linear:

$$h_{ij} = \sum_{p=1}^P \beta_{jp} z_{ip} \quad (5.1)$$

onde  $P$  é o número total de neurônios da camada de entrada,  $\beta_{jp}$  é o peso sináptico do neurônio  $p$  da camada de entrada para o neurônio  $j$  da camada intermediária, e  $z_{ip}$  é o valor da  $p$ -ésima variável explicativa para o elemento  $i$ . Em

seguida, aplica-se a esse valor intermediário uma *função de ativação*<sup>3</sup>, a qual assumimos ser tangente hiperbólica:

$$v_{ij} = g_1(h_{ij}) = \tanh(h_{ij}) \quad (5.2)$$

obtendo assim o valor do neurônio  $j$  da camada intermediária para o elemento  $i$ .

Similarmente, o valor do neurônio da camada de saída para o elemento  $i$  é também calculado em um processo de duas etapas. Primeiramente, calcula-se a transformação linear:

$$f_i = \sum_{j=1}^J B_j v_{ij} \quad (5.3)$$

onde  $J$  é o número de neurônios da camada intermediária,  $B_j$  denota o peso sináptico do  $j$ -ésimo neurônio da camada intermediária para o neurônio da camada de saída. Em seguida, o valor predito do neurônio da camada de saída para o elemento  $i$  é obtido com a aplicação da função de ativação  $g_2$ , que doravante assumimos ser a função identidade, a saber:

$$\hat{y}_i = g_2(f_i) = f_i \quad (5.4)$$

As funções de ativação utilizadas em RNA's são geralmente funções sigmóides, onde as mais usadas são as funções tangente hiperbólica e a logística. Na camada de saída podemos optar por uma função de ativação linear, sendo assim, a rede proposta é dita associadora. Por outro lado, quando a função de ativação do neurônio de saída é uma sigmoide, a rede é dita classificadora. Na camada intermediária podemos optar por uma das sigmóides citadas, dependendo do problema a ser solucionado ser uma classificação ou associação.

O ajuste do modelo de redes neurais corresponde à estimação dos pesos sinápticos  $\{\beta_{11}, \dots, \beta_{JP}\}$  e  $\{B_1, \dots, B_J\}$ , fixado o número  $J$  de neurônios da camada intermediária. O processo de estimação dos pesos sinápticos é chamado de *treinamento* ou *aprendizagem* da RNA. Existem dois tipos de treinamento: *supervisionado* e *não-supervisionado*. O treinamento supervisionado ocorre quando existe um valor predito,  $\hat{y}_i$ , associado com cada valor da camada de entrada,  $\{z_{i1}, \dots, z_{iP}\}$  e o valor observado  $y_i$ , no *conjunto de treinamento*.<sup>4</sup> Este valor predito será comparado ao valor observado,  $y_i$ , e as diferenças entre os valores preditos e observados serão usadas para treinamento da RNA. Existem diferentes algoritmos para o treinamento supervisionado de uma RNA, sendo o algoritmo BP o mais comumente usado.

<sup>3</sup>O objetivo da função de ativação é limitar a saída do neurônio dentro de um intervalo de valores razoáveis a serem assumidos pela sua própria imagem funcional.

<sup>4</sup>Na abordagem clássica de RNA, o conjunto de treinamento é formado pelos dados amostrais que são usados na modelagem.

No processo de treinamento não supervisionado não existem os valores preditos  $\hat{y}_i$  e observados  $y_i$  como no treinamento supervisionado. Conseqüentemente, a própria RNA deve se auto-organizar em relação às particularidades entre os elementos da amostra, identificando subconjuntos (*clusters*) de unidades que possuam similaridades. Os pesos sinápticos dos neurônios da rede são ajustados pelo algoritmo de aprendizado de forma a refletir esta representação internamente. O número a priori de subconjuntos pode ser especificado pelo analista levando em consideração seu conhecimento a respeito do problema. Na seqüência desta tese será considerado o caso do treinamento supervisionado.

O treinamento supervisionado é realizado considerando alguma métrica de desempenho. A mais utilizada é a média dos quadrados dos erros definida como:

$$D_s = \frac{1}{n} \sum_{i \in s} \sum_{k=1}^K (y_{ik} - \hat{y}_{ik})^2 \quad (5.5)$$

onde o índice  $i$  se refere aos elementos da amostra com total de  $n$ , o índice  $k$  aos neurônios da camada de saída com total de  $K$ ,  $y$  é a resposta observada e  $\hat{y}$  é a resposta predita. O caso mais simples, representado na fig.(5.1), é quando temos um único neurônio na camada de saída, ou seja,  $K = 1$ . Neste caso, simplificamos a notação mediante eliminação do índice  $k$ , e a eq.(5.5) se reduz a:

$$D_s = \frac{1}{n} \sum_{i \in s} (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_{i \in s} D_i^2 \quad (5.6)$$

onde  $D_i^2 = (y_i - \hat{y}_i)^2$ .

A fim de simplificar as expressões para o cálculo da matriz de covariâncias, multiplicamos a eq.(5.6) pelo fator de escala  $\frac{1}{2}$  obtendo a nova métrica de desempenho:

$$D_s = \frac{1}{2n} \sum_{i \in s} (y_i - \hat{y}_i)^2 = \frac{1}{2n} \sum_{i \in s} D_i^2 \quad (5.7)$$

Dado  $J$ , o objetivo é achar o conjunto de pesos sinápticos  $\{\beta_{11}, \dots, \beta_{JP}\}$  e  $\{B_1, \dots, B_J\}$  que minimizam a função objetivo, eq.(5.7). A minimização da função objetivo é realizada igualando-se as derivadas de  $D$  em relação a cada um dos parâmetros a serem estimados a zero e resolvendo o sistema de equações resultante. A derivada parcial da função objetivo em relação a um peso sináptico especificado representa a taxa de variação da função em relação a esse peso sináptico. Por exemplo:

$$\Delta_s B_j = \frac{\partial D_s}{\partial B_j} \quad (5.8)$$

Primeiramente, derivamos uma expressão para calcular o ajuste dos pesos

sinápticos das conexões da camada intermediária à camada de saída. Substituindo as eq.(5.1) a eq.(5.4) na eq.(5.7) temos:

$$D_s = \frac{1}{2n} \sum_{i \in s} \left( y_i - \sum_{j=1}^J B_j g_1 \left( \sum_{p=1}^P \beta_{jp} z_{ip} \right) \right)^2 \quad (5.9)$$

Expandindo a eq.(5.9) e utilizando a regra da cadeia:

$$\frac{\partial D_s}{\partial B_j} = \frac{\partial \left( \frac{1}{2n} \sum_{i \in s} D_i^2 \right)}{\partial B_j} = \frac{1}{2n} \sum_{i \in s} \frac{\partial D_i^2}{\partial B_j} = \frac{1}{2n} \sum_{i \in s} \frac{\partial D_i^2}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial f_i} \frac{\partial f_i}{\partial B_j} \quad (5.10)$$

mas

$$\frac{\partial D_i^2}{\partial \hat{y}_i} = -2(y_i - \hat{y}_i) \quad (5.11)$$

e

$$\frac{\partial \hat{y}_i}{\partial f_i} = g_2'(f_i) = 1 \quad (5.12)$$

no caso da função de ativação da eq.(5.4) e:

$$\frac{\partial f_i}{\partial B_j} = v_{ij} \quad (5.13)$$

Assim temos:

$$\frac{\partial D_i^2}{\partial B_j} = -2(y_i - \hat{y}_i)v_{ij} \quad (5.14)$$

Substituindo estes resultados na eq.(5.8) a variação do peso sináptico da camada intermediária para a camada de saída,  $\Delta_s B_j$ , é dada por:

$$\Delta_s B_j = -\frac{1}{n} \sum_{i \in s} (y_i - \hat{y}_i)v_{ij} \quad (5.15)$$

O ponto de mínimo da função de objetivo, eq.(5.7), é obtido resolvendo a equação  $\Delta_s B_j = -\frac{1}{n} \sum_{i \in s} (y_i - \hat{y}_i)v_{ij} = 0$ . Esta equação não tem solução em forma fechada, porém, a solução pode ser obtida por um procedimento iterativo de atualização dos pesos sinápticos da camada de saída:

$$B_j^{(t+1)} = B_j^{(t)} + \Delta_s B_j^{(t)} \quad (5.16)$$

A estimação dos pesos sinápticos,  $\beta_{jp}$ , da camada de entrada à camada intermediária segue a mesma lógica da eq.(5.16). Assim:

$$\Delta_s \beta_{jp} = \frac{\partial D_s}{\partial \beta_{jp}} = \frac{\partial \left( \frac{1}{2n} \sum_{i \in s} D_i^2 \right)}{\partial \beta_{jp}} = \frac{1}{2n} \sum_{i \in s} \frac{\partial D_i^2}{\partial \beta_{jp}} \quad (5.17)$$

Usando a regra da cadeia:

$$\frac{\partial D_i^2}{\partial \beta_{jp}} = \frac{\partial D_i^2}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial f_i} \frac{\partial f_i}{\partial v_{ij}} \frac{\partial v_{ij}}{\partial h_{ij}} \frac{\partial h_{ij}}{\partial \beta_{jp}} \quad (5.18)$$

onde  $\frac{\partial D_i^2}{\partial \hat{y}_i}$  e  $\frac{\partial \hat{y}_i}{\partial f_i}$  são dados pelas eq.(5.11) e eq.(5.12). Temos também que:

$$\frac{\partial f_i}{\partial v_{ij}} = B_j, \quad (5.19)$$

$$\frac{\partial v_{ij}}{\partial h_{ij}} = g'_1(h_{ij}) = 1 - v_{ij}^2, \quad (5.20)$$

e

$$\frac{\partial h_{ij}}{\partial \beta_{jp}} = z_{ip} \quad (5.21)$$

Assim temos:

$$\frac{\partial D_i^2}{\partial \beta_{jp}} = -2(y_i - \hat{y}_i)B_j(1 - v_{ij}^2)z_{ip} \quad (5.22)$$

Substituindo na eq.(5.17) temos a variação dos pesos sinápticos da camada de entrada à camada intermediária:

$$\Delta_s \beta_{jp} = -\frac{1}{n} \sum_{i \in s} (y_i - \hat{y}_i) B_j (1 - v_{ij}^2) z_{ip} \quad (5.23)$$

O ponto de mínimo da função objetivo, eq.(5.7), é obtido resolvendo a equação  $\Delta_s \beta_{jp} = -\frac{1}{n} \sum_{i \in s} (y_i - \hat{y}_i) B_j (1 - v_{ij}^2) z_{ip} = 0$ . Novamente, esta equação não tem solução em forma fechada, porém, a solução pode ser obtida por um procedimento iterativo de atualização dos pesos sinápticos da camada de saída. Sendo assim, a atualização dos pesos sinápticos  $\beta_{jp}$  é dada por:

$$\beta_{jp}^{(t+1)} = \beta_{jp}^{(t)} + \Delta_s \beta_{jp}^{(t)} \quad (5.24)$$

Os valores iniciais dos pesos sinápticos  $\{\beta_{11}, \dots, \beta_{JP}\}$  e  $\{B_1, \dots, B_J\}$  desempenham um papel significativo na convergência do treinamento, acelerando, desacelerando ou parando o processo de aprendizagem da rede. Os valores devem ser escolhidos de modo que a função sigmoide seja ativada na sua região linear. No caso da função *tanh* escolhemos valores na vizinhança de zero. Se os valores iniciais forem grandes a sigmoide irá saturar resultando em gradientes pequenos que fazem o treinamento ficar lento ou mesmo paralisando o treinamento da rede. Se os valores iniciais forem pequenos então os gradientes serão muito pequenos afetando o processo de treinamento. A escolha de valores intermediários que variam ao longo da região linear da sigmoide permitem que os gradientes sejam grandes o suficiente

para que o processo de treinamento possa avançar, sendo que a rede irá aprender a parte linear do mapeamento antes da parte não linear da função *tanh*, a qual é na vizinhança de  $\pm 1$ . Assim, de acordo com DU e SWAMY (2006), os valores iniciais dos coeficientes da rede devem ser gerados por uma distribuição uniforme no intervalo  $[-0,77;0,77]$ , pois empiricamente isto melhora o desempenho do algoritmo de aprendizagem.

O algoritmo de aprendizagem utilizado é o BP e o método de aprendizagem usado é o ‘em lotes’, isto é, todos dados são apresentados para RNA e os pesos são todos atualizados no final de cada *época*<sup>5</sup>. Este método é referido na literatura de RNA como método de otimização determinística. DU e SWAMY (2006) apresenta uma discussão sobre o algoritmo BP e o método de aprendizagem ‘em lotes’. A convergência do algoritmo é geralmente rápida, se a média de cada variável explicativa sobre o conjunto de treinamento é próxima de zero. Sendo assim, as variáveis são normalizadas antes de iniciarmos o treinamento da RNA. LECUN *et al.* (1998) apresenta uma discussão sobre a normalização das variáveis explicativas.

A taxa de aprendizagem, parâmetro que controla o quanto variam os coeficientes e os limiares no processo de aprendizagem do algoritmo, tem grande influência no processo de treinamento da rede neural. Sendo a taxa de aprendizagem baixa o processo de aprendizagem da rede neural se torna lento, ao passo que uma taxa de aprendizagem alta provoca oscilações no treinamento e impede a convergência do processo de aprendizado como mencionado em DU e SWAMY (2006). A taxa de aprendizagem usada nesta tese é constante de valor igual a 0,01.

A avaliação da RNA mais apropriada para o mapeamento de um problema é comumente feita de forma empírica. O dimensionamento da RNA depende do algoritmo de aprendizagem, da inicialização dos pesos e da qualidade do conjunto de treinamento disponível. A técnica mais utilizada para avaliação é a validação cruzada, vide SILVA *et al.* (2010). O propósito é avaliar a qualidade do ajuste (medida pelo menor erro de predição) de cada especificação quando aplicada a um conjunto de dados diferente do que foi usado no ajuste dos parâmetros. A validação cruzada não depende de qualquer hipótese probabilística e não é afetada por problemas de identificação do modelo de RNA.

## 5.2 Identificação em modelos de Redes Neurais

Os modelos clássicos de RNA são não-identificáveis globalmente. A consequência disto é que existem dois ou mais conjuntos de valores de parâmetros que minimizam, igualmente, a métrica de desempenho. Isso quer dizer que existem modelos que não podem ser discriminados um do outro com base nos dados. Em de-

---

<sup>5</sup>Consideramos uma época quando todo o conjunto de dados é apresentado para RNA uma vez.

corrência deste fato, não podemos estimar os parâmetros (pesos sinápticos) de um modelo não-identificável consistentemente, ou seja, as estimativas não convergirão para seus valores verdadeiros em probabilidade a medida que o número de observações tende para o infinito. Assim, não possuímos resultados assintóticos das distribuições de probabilidade dos parâmetros da RNA, como pode ser visto em MEDEIROS e TERASVIRTA (2001). Consequentemente, não podemos realizar inferências nos parâmetros estimados pela RNA. HWANG e DING (1997) e PAULINO e BRAGANÇA PEREIRA apresentam uma discussão detalhada sobre o problema de identificabilidade.

As características que implicam na não-identificabilidade do modelo de RNA são:

- (i) A permutação dos neurônios da camada intermediária não varia o valor da métrica de desempenho, ou seja, a permutação resulta em  $J!$  modelos distintos que não podem ser diferenciados entre si e a função de desempenho possui  $J!$  máximos locais;
- (ii) A relação  $\tanh(-x) = -\tanh(x)$  resulta em parametrizações equivalentes para cada neurônio da camada intermediária;
- (iii) A presença de neurônios irrelevantes na camada intermediária.

O problema (i) é corrigido fazendo-se  $B_1 \geq \dots \geq B_J$  onde  $j = 1, \dots, J$ . O item (ii) pode ser contornado, por exemplo, pela imposição da restrição  $\beta_{jp} > 0$  e  $B_j > 0$ ,  $j = 1, \dots, J$  e  $p = 1, \dots, P$ . No caso da rede com limiares sugerimos  $\beta_{j1} > 0$ , para os limiares da camada de entrada, e  $B_1 \neq 0$ , para o limiar da camada de saída. Para remediar o problema (iii) é necessário garantir que o modelo não possua neurônios irrelevantes na camada intermediária. Esta é a principal dificuldade para aplicação de inferência estatística segundo MEDEIROS e TERASVIRTA (2001). A próxima seção aborda o problema do item (iii).

### 5.2.1 Determinação do número de neurônios da camada intermediária

Sob o ponto de vista da estatística, podemos considerar que uma RNA é um modelo de regressão não linear e os procedimentos para teste de significância dos parâmetros, em princípio, podem ser realizados. Entretanto, para aplicação dos testes necessitamos da distribuição assintótica dos parâmetros, a qual foi derivada por WHITE (1989b). Dada a existência de um único conjunto de parâmetros  $\theta^*$ , o qual é a melhor aproximação da verdadeira função para uma RNA, WHITE (1989b) mostra que  $\theta^*$  pode ser consistentemente estimado através de um conjunto de estimadores  $\hat{\theta}$  obtido por máxima-quasi verossimilhança sob condições gerais. Além disso,



$\hat{\boldsymbol{\theta}}$  é assintoticamente normal com média  $\boldsymbol{\theta}^*$  e matriz de covariância  $\frac{1}{n}\Sigma$ , ou seja,  $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \rightarrow N(\mathbf{0}, \Sigma)$ , onde  $n$  é o número de observações na amostra.

Porém, continuamos com o problema da não-identificabilidade global do modelo. Sendo assim, o  $\boldsymbol{\theta}^*$  ótimo não é único prejudicando a aplicação de testes de hipóteses. A fim de realizarmos inferências nos parâmetros da RNA, tendo como base uma distribuição assintoticamente normal, devemos garantir que os parâmetros são localmente identificáveis não possuindo neurônios irrelevantes na camada intermediária. A investigação da existência ou não de neurônios irrelevantes na camada intermediária pode ser realizada, em princípio, por testes de significância dos parâmetros. Entretanto, os pesos  $\boldsymbol{\beta}$  e  $\mathbf{B}$  são mutuamente dependentes e existem dois modos de formular a hipótese nula. O primeiro é testar se  $\mathbf{B}$  é diferente de zero, ou então testar se  $\boldsymbol{\beta}$  é diferente de zero. No primeiro caso, a hipótese nula é  $H_0 : \mathbf{R}\mathbf{B} = 0$  com a alternativa  $H_A : \mathbf{R}\mathbf{B} \neq 0$ . No segundo caso, a hipótese nula é  $H_0 : \mathbf{R}\boldsymbol{\beta} = 0$  com alternativa  $H_A : \mathbf{R}\boldsymbol{\beta} \neq 0$ , onde  $\mathbf{R}$  é a matriz de restrição. Em ambos os casos, se a hipótese nula é verdadeira os parâmetros não são localmente identificáveis e o estimador não segue uma distribuição assintoticamente normal.

O artigo de WHITE (1989a) propõe um teste estatístico que se baseia em métodos que determinam se existe ou não melhora no modelo com a adição de um neurônio à camada intermediária. A descrição do teste se encontra na próxima seção.

## 5.2.2 Teste-LM considerando amostragem aleatória simples

Suponha que possuímos um conjunto de treinamento obtido de uma sequência aleatória  $X^n = \{X_i, i = 1, \dots, n\}$ ,  $X_i = (Y_i, \mathbf{Z}_i)$ , onde  $Y_i$  é um escalar e  $\mathbf{Z}_i$  é um vetor coluna de dimensão finita. Assumimos que os  $X_i$  são identicamente distribuídos mas não necessariamente independentes. Nosso interesse é estudar a relação entre  $Y_i$  e  $\mathbf{Z}_i$ , através da esperança condicional de  $Y_i$  dado  $\mathbf{Z}_i$ ,  $E(Y_i|\mathbf{Z}_i)$ . Tal relação pode ser escrita como uma função de regressão  $g(\mathbf{Z}_i) = E(Y_i|\mathbf{Z}_i)$ . O objetivo é investigar a adequação de uma RNA como uma representação de  $g$ .

Considere a RNA dada por:

$$f(z, \boldsymbol{\theta}) = \beta_{j1}z + \sum_{j=1}^J B_j \psi(z\boldsymbol{\beta}_j) \quad (5.25)$$

onde  $\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jP})$  e  $\boldsymbol{\theta} = (B_0, \dots, B_J, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_J)$  são os parâmetros desconhecidos.

A RNA sendo capaz de uma representação exata da função desconhecida  $g$ , então, existe um vetor de parâmetros  $\boldsymbol{\theta}^*$  tal que  $g(\mathbf{Z}_i) = f(\mathbf{Z}_i, \boldsymbol{\theta}^*)$  com probabilidade 1. Assim, a hipótese nula de interesse é:

$$H_0 : P[g(\mathbf{Z}_i) = f(\mathbf{Z}_i, \boldsymbol{\theta}^*)] = 1, \text{ para algum } \boldsymbol{\theta}^* \quad (5.26)$$

A hipótese alternativa é que a RNA não é capaz de uma representação exata de  $g$ , formalmente:

$$H_A : P[g(\mathbf{Z}_i) = f(\mathbf{Z}_i, \boldsymbol{\theta})] < 1, \text{ para todo } \boldsymbol{\theta} \quad (5.27)$$

O teste estatístico para testarmos a hipótese nula contra a alternativa está baseado na consequência de  $H_0$ . Quando a hipótese  $H_0$  é verdadeira temos que  $E(e_i^* | \mathbf{Z}_i) = 0$ , onde  $e_i^* = g(\mathbf{Z}_i) - f(\mathbf{Z}_i, \boldsymbol{\theta}^*)$  é o erro. Conseqüentemente,  $e_i^*$  é não correlacionado com qualquer função de  $\mathbf{Z}_i$ , como por exemplo a função  $\psi(\mathbf{Z}_i \boldsymbol{\Gamma}_j)$  para  $j = J+1, J+2, \dots, J+r$ ,  $r \in \mathbb{N}$  e  $\boldsymbol{\Gamma}_j$  é um vetor aleatório independente de  $X^n$ . Isto é  $E(\psi(\mathbf{Z}_i \boldsymbol{\Gamma}_j) e_i^* | \boldsymbol{\Gamma}_j) = 0$  sob  $H_0$ . Por outro lado, quando temos  $E(\psi(\mathbf{Z}_i \boldsymbol{\Gamma}_j) e_i^* | \boldsymbol{\Gamma}_j) \neq 0$  para algum  $j$ , então  $H_0$  é falsa.

Desta forma, podemos entender que a inclusão de um neurônio adicional na camada intermediária com função de ativação  $\psi(\mathbf{Z}_i \boldsymbol{\Gamma}_j)$  poderia aumentar o poder preditivo da RNA. Os pesos  $\boldsymbol{\Gamma}_j$  são escolhidos aleatoriamente dentre os pesos localizados entre a camada de entrada e a camada intermediária. Tal escolha é computacionalmente conveniente, pois não envolve um novo treinamento da rede o que poderia destruir as propriedades estatísticas do teste, segundo WHITE (1989a).

Fixamos  $r \in \mathbb{N} = \{1, 2, \dots\}$  e construímos o vetor  $\Psi_i = (\psi(\mathbf{Z}_i \boldsymbol{\Gamma}_{q+1}), \dots, \psi(\mathbf{Z}_i \boldsymbol{\Gamma}_{q+r}))$  onde  $\boldsymbol{\Gamma}_j$ ,  $j = J+1, \dots, J+r$ , são vetores aleatórios independente de  $X^n$ . Seja  $\boldsymbol{\theta}^*$  a solução para minimização do erro médio quadrático:

$$\min(E[(Y_i - f(\mathbf{Z}_i, \boldsymbol{\theta}))^2] = E[(Y_i - g(\mathbf{Z}_i))^2] + E[(g(\mathbf{Z}_i) - f(\mathbf{Z}_i, \boldsymbol{\theta}))^2]) \quad (5.28)$$

dado  $g(\mathbf{Z}_i) = E(Y_i | \mathbf{Z}_i)$ . Definimos  $e_i = Y_i - f(\mathbf{Z}_i, \boldsymbol{\theta}^*)$  e  $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_{J+1}, \dots, \boldsymbol{\Gamma}_{J+r})$  e, formalmente, testamos  $H_0^* : E(\Psi_i' e_i | \boldsymbol{\Gamma}) = 0$  contra a hipótese alternativa  $H_A^* : E(\Psi_i' e_i | \boldsymbol{\Gamma}) \neq 0$ . Dado que a hipótese  $H_0^*$  é verdadeira não existe aumento no desempenho da RNA com o neurônio adicional na camada intermediária. Quando a hipótese  $H_A^*$  é verdadeira o neurônio adicional na camada intermediária poderia ser usado para aumentar o desempenho da RNA, pois existe alguma não linearidade ainda a ser capturada pelo modelo. Como  $H_0^*$  é uma consequência de  $H_0$  a rejeição de  $H_0^*$  implica na rejeição de  $H_0$ .

Segundo WHITE (1989a) o teste estatístico da hipótese  $H_0^*$  contra  $H_A^*$  pode ser realizado como um teste multiplicador de Lagrange.

Suponha o vetor  $\boldsymbol{\theta}^*$  conhecido. Então,  $E(\Psi_i' e_i | \boldsymbol{\Gamma})$  pode ser estimado pela média  $\Psi_i' e_i$  do conjunto de treinamento,  $\frac{1}{n} \sum_{i=1}^n \Psi_i' e_i$ . Sendo a hipótese  $H_0^*$  verdadeira a

média estará próxima de zero, caso contrário estará longe de zero. A questão de quanto afastado de zero estando em conformidade com a variação aleatória sob a hipótese  $H_0^*$  pode ser resolvida notando que distribuição amostral de  $\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi'_i e_i$  sob  $H_0^*$ , é aproximadamente normal multivariada com média zero e matriz de covariância  $V_{r \times r}$ , sob condições gerais dadas pelo teorema central do limite, quando  $n$  tende a infinito, isto é:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi'_i e_i \xrightarrow{d} N(0, V) \quad (5.29)$$

onde  $\xrightarrow{d}$  representa a convergência em distribuição quando  $n \rightarrow \infty$ . Assim, se  $V$  é uma matriz conhecida com inversa  $V^{-1}$ , segundo WHITE (1989b), temos:

$$(n^{-1/2} \sum_{i=1}^n \Psi'_i e_i)' V^{-1} (n^{-1/2} \sum_{i=1}^n \Psi'_i e_i) \xrightarrow{d} \chi_r^2 \quad (5.30)$$

onde  $\chi_r^2$  é uma distribuição qui-quadrado com  $r$  graus de liberdade. O teste estatístico de  $H_0^*$  contra  $H_A^*$  com nível de significância  $\alpha$  é realizado comparando-se o valor da estatística de teste com o valor tabelado  $\chi_{r,1-\alpha}^2$ . Se a estatística de teste exceder o valor tabelado rejeitamos a hipótese  $H_0^*$ .

Na prática  $\theta^*$  e  $V$  são desconhecidos. Porém,  $\theta^*$  pode ser substituído pelo vetor  $\hat{\theta}$  obtido do aprendizado sobre a amostra de treinamento. Definimos  $u_i = Y_i - f(\mathbf{Z}_i, \hat{\theta})$ ,  $i = 1, \dots, n$ . Assim temos:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi'_i u_i \xrightarrow{d} N(0, W) \quad (5.31)$$

com  $V \neq W$  sendo que tais quantidades são também desconhecidas. Entretanto, é possível estimar  $W$  através da amostra de treinamento por  $\hat{W}$  tal que  $\hat{W} \xrightarrow{p} W$ , onde  $\xrightarrow{p}$  indica convergência em probabilidade. A estimativa de  $W$  poder ser realizada pela obtenção da matriz Hessiana, vide ROJAS (1993), seguida de sua inversa, ou seja,  $W^{-1} = H^{-1}$ , onde  $H$  denota a matriz Hessiana.

Assim, temos, segundo WHITE (1989b):

$$(n^{-1/2} \sum_{i=1}^n \Psi'_i u_i)' \hat{W}^{-1} (n^{-1/2} \sum_{i=1}^n \Psi'_i u_i) \xrightarrow{d} \chi_r^2 \quad (5.32)$$

O cálculo da estatística de White, eq.(5.32), é dificultado pelo cálculo da estimativa de  $\hat{W}^{-1}$ . Porém, existe um procedimento computacional que torna o cálculo do teste de White mais simples, resultando uma estatística assintoticamente equivalente a dada na eq.(5.32) como mostrado em WHITE (1989b). Assume-se que o vetor de parâmetros estimados  $\hat{\theta}$  é obtido via treinamento da RNA. Fixamos  $r$  e calculamos o vetor  $\Psi_i = (\psi(\mathbf{Z}_i \mathbf{\Gamma}_{q+1}), \dots, \psi(\mathbf{Z}_i \mathbf{\Gamma}_{q+r}))$  onde  $\mathbf{\Gamma}_j$ ,  $j = J + 1, \dots, J + r$ .

A seguir fazemos uma regressão linear cuja variável dependente é  $u_i$  e as variáveis exploratórias são  $\nabla f(\mathbf{Z}_i, \hat{\boldsymbol{\theta}})$  e  $\Psi_i$ , onde  $\nabla$  é o operador gradiente em relação a  $\boldsymbol{\theta}$ . Sendo  $R^2$  o coeficiente de correlação múltipla da regressão temos, segundo WHITE (1989a), que  $nR^2 \rightarrow \chi_r^2$ , onde  $n$  é o tamanho da amostra.

Assim, finalizamos o teste para identificar a existência de neurônios irrelevantes na camada intermediária.

O capítulo seguinte apresenta a abordagem de superpopulação para o ajuste RNA.

# Capítulo 6

## Redes Neurais para Dados Amostrais Complexos

No presente trabalho propomos usar a abordagem de modelagem de superpopulação, vide HAYKIN (2001), para o ajuste de Redes Neurais Artificiais quando os dados disponíveis para o treinamento da RNA são obtidos por amostragem complexa. As motivações no uso desta abordagem são:

1. Um modelo de RNA não assume forma funcional linear, como no modelo de regressão linear, vide eq.(4.10);
2. O emprego dos modelos de superpopulação permite lidar corretamente com os planos amostrais complexos empregados em grande parte das pesquisas amostrais usadas na prática.

A seção seguinte introduz a abordagem do modelo de superpopulação em RNA, a qual consideramos identificada, pois somente desta forma podemos calcular a variância assintótica do estimador corretamente.

### 6.1 RNA sob o modelo de superpopulação

A proposta da modelagem de superpopulação é utilizar a RNA descrita na fig.(5.1) para modelar os dados populacionais e não mais os dados amostrais observados como descrito em PESSOA e DO NASCIMENTO SILVA (1998) e ISAKI e FULLER. Assim, faremos uso da mesma notação utilizada em modelos estatísticos sob abordagem de superpopulação em RNA.

Definimos o vetor  $\boldsymbol{\theta} = [B_1, \dots, B_J, \beta_{11}, \dots, \beta_{JP}]$ , o qual é o vetor de todos os pesos sinápticos da RNA cuja dimensão é  $(J + JP) \times 1$ ,  $\mathbf{y}_U = [y_1, \dots, y_N]$  como o vetor de valores das respostas observadas na população e  $\mathbf{z}_U = [\mathbf{z}_1, \dots, \mathbf{z}_N]'$  como a matriz de valores das variáveis preditoras na população. Supondo que todos os

elementos da população finita  $U$  fossem conhecidos poderíamos definir a métrica de desempenho  $D$  em relação ao vetor de parâmetros  $\boldsymbol{\theta}$  da RNA como:

$$D_U(\boldsymbol{\theta}) = \frac{1}{2N} \sum_{i \in U} D(y_i, \boldsymbol{\theta}) \quad (6.1)$$

onde  $D(y_i, \boldsymbol{\theta}) = (y_i - \hat{y}_i)^2$  e  $\hat{y}_i = \sum_{j=1}^J B_j g_1(\sum_{p=1}^P \beta_{jp} z_{ip}) = \sum_{j=1}^J B_j v_{ij}$

Calculando as derivadas parciais de  $D_U(\boldsymbol{\theta})$  com relação a cada componente de  $\boldsymbol{\theta}$  e igualando a zero, obtemos as equações de ajuste dos pesos sinápticos para a RNA populacional dadas por:

$$\frac{1}{2N} \sum_{i \in U} \mathbf{u}_i(\boldsymbol{\theta}) = \frac{1}{2N} \mathbf{T} = \mathbf{0} \quad (6.2)$$

onde

$$\mathbf{u}_i(\boldsymbol{\theta}) = \frac{\partial D(y_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial D(y_i; \boldsymbol{\theta})}{\partial B_1} \\ \frac{\partial D(y_i; \boldsymbol{\theta})}{\partial B_2} \\ \vdots \\ \frac{\partial D(y_i; \boldsymbol{\theta})}{\partial B_J} \\ \frac{\partial D(y_i; \boldsymbol{\theta})}{\partial \beta_{11}} \\ \frac{\partial D(y_i; \boldsymbol{\theta})}{\partial \beta_{12}} \\ \vdots \\ \frac{\partial D(y_i; \boldsymbol{\theta})}{\partial \beta_{JP}} \end{bmatrix} \quad (6.3)$$

é o vetor  $(J + JP) \times 1$  dos escores do elemento  $i, i \in U$ .

Assim como descrito na seção 4.1 consideramos o vetor de totais populacionais  $\mathbf{T} = \sum_{i \in U} \mathbf{u}_i(\boldsymbol{\theta})$  cuja estimação é realizada utilizando um estimador linear ponderado da forma  $\hat{\mathbf{T}} = \sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\theta})$  onde  $w_i$  são pesos amostrais apropriadamente definidos. Com esta notação podemos obter um estimador para  $\boldsymbol{\theta}$  resolvendo o sistema de equações obtido igualando a função do estimador  $\hat{\mathbf{T}}$  do total  $\mathbf{T}$  a zero e definindo que o estimador de  $\boldsymbol{\theta}_U$  (ou de  $\boldsymbol{\theta}$ ) será a solução de  $\frac{1}{2N} \hat{\mathbf{T}} = \frac{1}{2N} \sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\theta}) = \mathbf{0}$ , analogamente, como feito na definição 4.1.1 na seção 4.1.

A obtenção da *variância assintótica do estimador*<sup>1</sup>  $\hat{\boldsymbol{\theta}}$  é feita através da linearização de Taylor como apresentado em BINDER (1983):

$$V_p(\hat{\boldsymbol{\theta}}) \cong [J(\boldsymbol{\theta}_U)]^{-1} V_p \left[ \sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\theta}_U) \right] [J(\boldsymbol{\theta}_U)]^{-1} \quad (6.4)$$

<sup>1</sup>A estimação das variâncias dos pesos sinápticos da RNA permite realização de inferências sobre os parâmetros estimados e, conseqüentemente, realização da poda da rede. A poda é um método de minimização do tamanho da RNA mantendo-se bom o nível de predição. É mais provável que uma RNA com tamanho mínimo não aprenda o ruído dos dados de treinamento, e possa assim generalizar melhor sobre novos dados.

e seu estimador é dado por:

$$\widehat{V}_p(\widehat{\boldsymbol{\theta}}) = [\widehat{J}(\widehat{\boldsymbol{\theta}})]^{-1} \widehat{V}_p \left[ \sum_{i \in S} w_i \mathbf{u}_i(\widehat{\boldsymbol{\theta}}) \right] [\widehat{J}(\widehat{\boldsymbol{\theta}})]^{-1} \quad (6.5)$$

onde

$$J(\boldsymbol{\theta}_U) = \frac{\partial \mathbf{T}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_U} = \sum_{i \in U} \frac{\partial \mathbf{u}_i(\boldsymbol{\theta})}{\partial (\boldsymbol{\theta})} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_U} \quad (6.6)$$

$$\widehat{J}(\widehat{\boldsymbol{\theta}}) = \frac{\partial \widehat{\mathbf{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \sum_{i \in S} w_i \frac{\partial \mathbf{u}_i(\boldsymbol{\theta})}{\partial (\boldsymbol{\theta})} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \quad (6.7)$$

A distribuição assintótica de  $\widehat{\boldsymbol{\theta}}$ , dadas as suposições feitas em BINDER (1983), é normal multivariada permitindo-se a inferência. O estimador  $\widehat{J}(\widehat{\boldsymbol{\theta}}) \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ , dado pela eq.(6.7), é um estimador consistente de eq.(6.6).

O cálculo da variância do total  $V_p[\sum_{i \in S} w_i \mathbf{u}_i(\boldsymbol{\theta}_U)]$ , eq.(6.4), é realizado utilizando-se as observações  $\{e_i, \mathbf{z}_i\}$ , para  $i \in U$ , onde  $e_i = y_i - f(\mathbf{z}_i, \boldsymbol{\theta})$ , dado  $\hat{y}_i = f(\mathbf{z}_i, \boldsymbol{\theta})$ . Supomos que  $\widehat{V}_p[\sum_{i \in S} w_i \mathbf{u}_i(\widehat{\boldsymbol{\theta}})]$  é um estimador consistente da variância do estimador do total considerando as observações  $\{\hat{e}_i, \mathbf{z}_i\}$ , para  $i \in S$ , onde  $\hat{e}_i = y_i - f(\mathbf{z}_i, \widehat{\boldsymbol{\theta}})$ .

Assim completamos a especificação do procedimento de estimação para ajustar modelos de RNA sob o modelo de superpopulação. Este procedimento é bastante flexível e aplicável a dados obtidos sob uma ampla gama de planos amostrais complexos e considerando variadas arquiteturas de RNA.

## 6.2 Estimação da RNA sob o modelo de superpopulação

### 6.2.1 Estimação dos parâmetros

Os pesos sinápticos serão estimados utilizando-se o treinamento *supervisionado* com um neurônio na camada de saída utilizando uma função de ativação linear. A métrica usada será o erro quadrático médio como definida na eq.(5.7), cuja minimização coincide com a estimação de máxima verossimilhança para uma variável aleatória com função de distribuição Gaussiana, vide DUNNE (2007).

Supondo que todos os elementos da população finita  $U$  fossem conhecidos temos o vetor de totais populacionais igual a:

$$\mathbf{T} = \sum_{i \in U} \mathbf{u}_i(\boldsymbol{\theta}) \quad (6.8)$$

A obtenção das equações de atualização populacionais é realizada igualando-se  $\frac{1}{2N} \mathbf{T}$  a zero. Assim utilizando a eq.(6.3) e o resultado eq.(5.14) temos:

$$\mathbf{u}_i(\boldsymbol{\theta}) = \frac{\partial D(y_i; \boldsymbol{\theta})}{\partial B_j} = -2(y_i - \hat{y}_i)v_{ij} \quad (6.9)$$

Somando, dividindo por  $2N$  e igualando a zero:

$$\frac{1}{2N} \sum_{i \in U} \mathbf{u}_i(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i \in U} (y_i - \hat{y}_i)v_{ij} = 0 \quad (6.10)$$

obtemos as equações de atualização dos pesos sinápticos para os dados da população finita da camada intermediária para a camada de saída dada por:

$$B_j^{(t+1)} = B_j^{(t)} + \Delta_U B_j^{(t)} \quad (6.11)$$

onde

$$\Delta_U B_j = -\frac{1}{N} \sum_{i \in U} (y_i - \hat{y}_i)v_{ij} \quad (6.12)$$

Para a obtenção dos pesos sinápticos da camada de entrada para camada intermediária temos:

$$\mathbf{u}_i(\boldsymbol{\theta}) = \frac{\partial D(\mathbf{y}_i; \boldsymbol{\theta})}{\partial \beta_{jp}} = -2(y_i - \hat{y}_i)B_j(1 - v_{ij}^2)z_{ip} \quad (6.13)$$

Somando, dividindo por  $2N$  e igualando a zero:

$$\frac{1}{2N} \sum_{i \in U} \mathbf{u}_i(\boldsymbol{\theta}) = -\frac{1}{N} \sum_{i \in U} (y_i - \hat{y}_i)B_j(1 - v_{ij}^2)z_{ip} = 0 \quad (6.14)$$

Assim temos:

$$\beta_{jp}^{(t+1)} = \beta_{jp}^{(t)} + \Delta_U \beta_{jp}^{(t)} \quad (6.15)$$

onde

$$\Delta_U \beta_{jp} = -\frac{1}{N} \sum_{i \in U} (y_i - \hat{y}_i)B_j(1 - v_{ij}^2)z_{ip} \quad (6.16)$$

Na prática não possuímos os dados populacionais somente os dados amostrais. Sendo assim, estimamos o vetor de totais, eq.(6.8), utilizando um estimador linear ponderado da forma:

$$\hat{\mathbf{T}} = \sum_{i \in s} w_i \mathbf{u}_i(\boldsymbol{\theta}) \quad (6.17)$$

em que  $w_i$  são pesos amostrais apropriadamente definidos. A obtenção do esti-



mador para  $\theta$  é realizada igualando-se a função do estimador  $\hat{\mathbf{T}}$  do total  $\mathbf{T}$  a zero e dividindo pelo estimador de N,  $\hat{N} = \sum_{i \in s} w_i$ .

Seguindo o procedimento de estimação dos coeficientes descrito na seção 5.1, as equações de atualização dos pesos sinápticos para os dados amostrais, sob um plano amostral complexo, são:

- Da camada intermediária para a camada de saída

$$B_j^{(t+1)} = B_j^{(t)} + \Delta_{sw} B_j^{(t)} \quad (6.18)$$

onde

$$\Delta_{sw} B_j = -\frac{1}{\hat{N}} \sum_{i \in s} w_i (y_i - \hat{y}_i) v_{ij} \quad (6.19)$$

- Da camada de entrada para a camada intermediária

$$\beta_{jp}^{(t+1)} = \beta_{jp}^{(t)} + \Delta_{sw} \beta_{jp}^{(t)} \quad (6.20)$$

onde

$$\Delta_{sw} \beta_{jp} = -\frac{1}{\hat{N}} \sum_{i \in s} w_i (y_i - \hat{y}_i) B_j (1 - v_{ij}^2) z_{ip} \quad (6.21)$$

A abordagem de superpopulação necessita do cálculo dos pesos amostrais para ser empregada. O estimador de Horvitz-Thompson para os totais, como mencionado na seção 3.2, considera os pesos amostrais como o inverso da probabilidade de inclusão dos indivíduos, ou seja,  $w_i = \pi_i^{-1}$ . Os pesos amostrais definidos desta forma causam a saturação dos neurônios e, conseqüentemente, a paralisação do processo de aprendizagem. Pois, se possuímos a probabilidade de inclusão sendo  $\pi = 0,001$  o peso amostral será  $w = 1000$  que multiplicado nas equações de atualização eq(6.18) e eq(6.20) faz com que o treinamento da RNA seja realizado somente nos extremos da função de ativação. A fim de contornar este problema padronizamos os pesos amostrais da seguinte forma,  $w_i^* = \frac{w_i}{\hat{N}} \times n$ , onde  $\sum_{i \in s} w_i = \hat{N}$  e  $n$  é o tamanho da amostra. Esta modificação implica que  $\sum_{i \in s} w_i^* = n$ .

Desta forma as novas equações de correção para a atualização dos pesos são:

$$\Delta_{sw} B_j = -\frac{1}{n} \sum_{i \in s} w_i^* (y_i - \hat{y}_i) v_{ij} \quad (6.22)$$

$$\Delta_{sw} \beta_{jp} = -\frac{1}{n} \sum_{i \in s} w_i^* (y_i - \hat{y}_i) B_j (1 - v_{ij}^2) z_{ip} \quad (6.23)$$

## 6.2.2 Estimação das variâncias

O cálculo das estimativas de variâncias dos estimadores dos parâmetros da RNA proposta na fig.(5.1) é feito derivando-se as eq.(6.12) e eq.(6.16) em relação ao vetor de parâmetros  $\boldsymbol{\theta}$  para obtermos a matriz Jacobiana  $J(\boldsymbol{\theta}_U) |_{\boldsymbol{\theta}=\boldsymbol{\theta}_U}$  quadrada com dimensão  $(J + JP) \times (J + JP)$  cuja representação por blocos é igual a:

$$J(\boldsymbol{\theta}_U) = \left[ \begin{array}{c|c} 1 & 2 \\ \hline 3 & 4 \end{array} \right]$$

em que:

Bloco 1:

$$\left[ \begin{array}{cccc} \frac{\partial \Delta_U B_1}{\partial B_1} & \frac{\partial \Delta_U B_1}{\partial B_2} & \cdots & \frac{\partial \Delta_U B_1}{\partial B_J} \\ \frac{\partial \Delta_U B_2}{\partial B_1} & \frac{\partial \Delta_U B_2}{\partial B_2} & \cdots & \frac{\partial \Delta_U B_2}{\partial B_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Delta_U B_J}{\partial B_1} & \frac{\partial \Delta_U B_J}{\partial B_2} & \cdots & \frac{\partial \Delta_U B_J}{\partial B_J} \end{array} \right]$$

As expressões dos termos do bloco 1 da matriz são:

$$\frac{\partial \Delta_U B_\ell}{\partial B_\ell} = \frac{1}{N} \sum_{i \in U} v_{i\ell}^2 \quad (6.24)$$

e

$$\frac{\partial \Delta_U B_\ell}{\partial B_j} = \frac{1}{N} \sum_{i \in U} v_{i\ell} v_{ij} \quad , \text{ para } \ell \neq j. \quad (6.25)$$

em que a expressão de  $v_{ij}$  é dada pela eq.(5.2).

Bloco 2:

$$\left[ \begin{array}{cccc} \frac{\partial \Delta_U B_1}{\partial \beta_{11}} & \frac{\partial \Delta_U B_1}{\partial \beta_{12}} & \cdots & \frac{\partial \Delta_U B_1}{\partial \beta_{JP}} \\ \frac{\partial \Delta_U B_2}{\partial \beta_{11}} & \frac{\partial \Delta_U B_2}{\partial \beta_{12}} & \cdots & \frac{\partial \Delta_U B_2}{\partial \beta_{JP}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Delta_U B_J}{\partial \beta_{11}} & \frac{\partial \Delta_U B_J}{\partial \beta_{12}} & \cdots & \frac{\partial \Delta_U B_J}{\partial \beta_{JP}} \end{array} \right]$$

As expressões dos termos do bloco 2 da matriz são:

$$\frac{\partial \Delta_U B_\ell}{\partial \beta_{jp}} = -\frac{1}{N} \sum_{i \in U} [z_{ip}(v_{ij}^2 - 1)(B_1 v_{i1} + \dots + B_J v_{iJ} - y_i) - B_j z_{ip} v_{ij}(v_{ij}^2 - 1)] \quad , \text{ para } \ell = j. \quad (6.26)$$

$$\frac{\partial \Delta_U B_\ell}{\partial \beta_{jp}} = -\frac{1}{N} \sum_{i \in U} B_j z_{ip} v_{i\ell}(v_{ij}^2 - 1) \quad , \text{ para } \ell \neq j. \quad (6.27)$$

Bloco 3:

$$\begin{bmatrix} \frac{\partial \Delta_U \beta_{11}}{\partial B_1} & \frac{\partial \Delta_U \beta_{11}}{\partial B_2} & \dots & \frac{\partial \Delta_U \beta_{11}}{\partial B_J} \\ \frac{\partial \Delta_U \beta_{12}}{\partial B_1} & \frac{\partial \Delta_U \beta_{12}}{\partial B_2} & \dots & \frac{\partial \Delta_U \beta_{12}}{\partial B_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Delta_U \beta_{JP}}{\partial B_1} & \frac{\partial \Delta_U \beta_{JP}}{\partial B_2} & \dots & \frac{\partial \Delta_U \beta_{JP}}{\partial B_J} \end{bmatrix}$$

As expressões dos termos do bloco 3 da matriz são:

$$\frac{\partial \Delta_U \beta_{jp}}{\partial B_\ell} = -\frac{1}{N} \sum_{i \in U} [z_{ip}(v_{ij}^2 - 1)(B_1 v_{ij} + \dots + B_J v_{iJ} - y_i) - B_j z_{ip} v_{ij}(v_{ij}^2 - 1)] \quad , \text{para } \ell = j. \quad (6.28)$$

$$\frac{\partial \Delta_U \beta_{jp}}{\partial B_\ell} = -\frac{1}{N} \sum_{i \in U} B_j z_{ip} v_{i\ell}(v_{ij}^2 - 1) \quad , \text{para } \ell \neq j. \quad (6.29)$$

Bloco 4:

$$\begin{bmatrix} \frac{\partial \Delta_U \beta_{11}}{\partial \beta_{11}} & \frac{\partial \Delta_U \beta_{11}}{\partial \beta_{12}} & \dots & \frac{\partial \Delta_U \beta_{11}}{\partial \beta_{JP}} \\ \frac{\partial \Delta_U \beta_{12}}{\partial \beta_{11}} & \frac{\partial \Delta_U \beta_{12}}{\partial \beta_{12}} & \dots & \frac{\partial \Delta_U \beta_{12}}{\partial \beta_{JP}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Delta_U \beta_{JP}}{\partial \beta_{11}} & \frac{\partial \Delta_U \beta_{JP}}{\partial \beta_{12}} & \dots & \frac{\partial \Delta_U \beta_{JP}}{\partial \beta_{JP}} \end{bmatrix}$$

onde a expressão do elemento da diagonal principal do bloco 4 da matriz é

$$\frac{\partial \Delta_U \beta_{jp}}{\partial \beta_{jp}} = \frac{1}{N} \sum_{i \in U} [B_j^2 z_{ip}^2 (v_{ij}^2 - 1)^2 + 2B_j z_{ip}^2 v_{ij}(v_{ij}^2 - 1)(B_1 v_{i1} + \dots + B_J v_{iJ} - y_i)] \quad (6.30)$$

e os termos fora da diagonal principal da matriz são:

$$\frac{\partial \Delta_U \beta_{jp}}{\partial \beta_{\ell q}} = \frac{1}{N} \sum_{i \in U} [B_j^2 z_{ip} z_{iq} (v_{ij}^2 - 1)^2 + 2B_j z_{ip} z_{iq} v_{ij}(v_{ij}^2 - 1)(B_1 v_{i1} + \dots + B_J v_{iJ} - y_i)] , \quad \text{para } \ell = j. \quad (6.31)$$

e

$$\frac{\partial \Delta_U \beta_{jp}}{\partial \beta_{\ell q}} = \frac{1}{N} \sum_{i \in U} B_j B_\ell z_{ip} z_{iq} (v_{ij}^2 - 1)(v_{i\ell}^2 - 1) \quad , \text{para } \ell \neq j. \quad (6.32)$$

em que  $\ell = 1, \dots, J$  e  $q = 1, \dots, P$ .

O estimador  $\hat{J}(\hat{\theta})$  é obtido derivando-se as equações eq.(6.19) e eq.(6.21) em relação ao vetor de parâmetros estimados  $\hat{\theta}$ :

$$\hat{J}(\hat{\theta}) = \left[ \begin{array}{c|c} 1 & 2 \\ \hline 3 & 4 \end{array} \right]$$

em que:

Bloco 1:

$$\left[ \begin{array}{cccc} \frac{\partial \Delta_{sw} B_1}{\partial B_1} & \frac{\partial \Delta_{sw} B_1}{\partial B_2} & \dots & \frac{\partial \Delta_{sw} B_1}{\partial B_J} \\ \frac{\partial \Delta_{sw} B_2}{\partial B_1} & \frac{\partial \Delta_{sw} B_2}{\partial B_2} & \dots & \frac{\partial \Delta_{sw} B_2}{\partial B_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Delta_{sw} B_J}{\partial B_1} & \frac{\partial \Delta_{sw} B_J}{\partial B_2} & \dots & \frac{\partial \Delta_{sw} B_J}{\partial B_J} \end{array} \right]$$

As expressões dos termos do bloco 1 da matriz são:

$$\frac{\partial \Delta_{sw} B_\ell}{\partial B_\ell} = \frac{1}{\hat{N}} \sum_{i \in s} w_i v_{i\ell}^2 \quad (6.33)$$

e

$$\frac{\partial \Delta_{sw} B_\ell}{\partial B_j} = \frac{1}{\hat{N}} \sum_{i \in s} w_i v_{i\ell} v_{ij} \quad , \text{ para } \ell \neq j. \quad (6.34)$$

em que a expressão de  $v_{ij}$  é dada pela eq.(5.2).

Bloco 2:

$$\left[ \begin{array}{cccc} \frac{\partial \Delta_{sw} B_1}{\partial \beta_{11}} & \frac{\partial \Delta_{sw} B_1}{\partial \beta_{12}} & \dots & \frac{\partial \Delta_{sw} B_1}{\partial \beta_{JP}} \\ \frac{\partial \Delta_{sw} B_2}{\partial \beta_{11}} & \frac{\partial \Delta_{sw} B_2}{\partial \beta_{12}} & \dots & \frac{\partial \Delta_{sw} B_2}{\partial \beta_{JP}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Delta_{sw} B_J}{\partial \beta_{11}} & \frac{\partial \Delta_{sw} B_J}{\partial \beta_{12}} & \dots & \frac{\partial \Delta_{sw} B_J}{\partial \beta_{JP}} \end{array} \right]$$

As expressões dos termos do bloco 2 da matriz são:

$$\frac{\partial \Delta_{sw} B_\ell}{\partial \beta_{jp}} = \frac{1}{\hat{N}} \sum_{i \in s} w_i [-z_{ip}(v_{ij}^2 - 1)(B_1 v_{i1} + \dots + B_J v_{iJ} - y_i) - B_j z_{ip} v_{ij}(v_{ij}^2 - 1)] \quad ,$$

*para*  $\ell = j.$

(6.35)

$$\frac{\partial \Delta_{sw} B_\ell}{\partial \beta_{jp}} = \frac{1}{\hat{N}} \sum_{i \in s} w_i [-B_j z_{ip} v_{i\ell}(v_{ij}^2 - 1)] \quad , \text{ para } \ell \neq j. \quad (6.36)$$

Bloco 3:

$$\left[ \begin{array}{cccc} \frac{\partial \Delta_{sw} \beta_{11}}{\partial B_1} & \frac{\partial \Delta_{sw} \beta_{11}}{\partial B_2} & \dots & \frac{\partial \Delta_{sw} \beta_{11}}{\partial B_J} \\ \frac{\partial \Delta_{sw} \beta_{12}}{\partial B_1} & \frac{\partial \Delta_{sw} \beta_{12}}{\partial B_2} & \dots & \frac{\partial \Delta_{sw} \beta_{12}}{\partial B_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Delta_{sw} \beta_{JP}}{\partial B_1} & \frac{\partial \Delta_{sw} \beta_{JP}}{\partial B_2} & \dots & \frac{\partial \Delta_{sw} \beta_{JP}}{\partial B_J} \end{array} \right]$$

As expressões dos termos do bloco 3 da matriz são:

$$\frac{\partial \Delta_{sw} \beta_{jp}}{\partial B_\ell} = \frac{1}{\hat{N}} \sum_{i \in s} w_i [-z_{ip}(v_{ij}^2 - 1)(B_1 v_{ij} + \dots + B_J v_{iJ} - y_i) - B_j z_{ip} v_{ij}(v_{ij}^2 - 1)],$$

*para*  $\ell = j$ .

(6.37)

$$\frac{\partial \Delta_{sw} \beta_{jp}}{\partial B_\ell} = \frac{1}{\hat{N}} \sum_{i \in s} w_i [-B_j z_{ip} v_{i\ell}(v_{ij}^2 - 1)] \quad , \textit{para } \ell \neq j. \quad (6.38)$$

Bloco 4:

$$\begin{bmatrix} \frac{\partial \Delta_{sw} \beta_{11}}{\partial \beta_{11}} & \frac{\partial \Delta_{sw} \beta_{11}}{\partial \beta_{12}} & \dots & \frac{\partial \Delta_{sw} \beta_{11}}{\partial \beta_{JP}} \\ \frac{\partial \Delta_{sw} \beta_{12}}{\partial \beta_{11}} & \frac{\partial \Delta_{sw} \beta_{12}}{\partial \beta_{12}} & \dots & \frac{\partial \Delta_{sw} \beta_{12}}{\partial \beta_{JP}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \Delta_{sw} \beta_{JP}}{\partial \beta_{11}} & \frac{\partial \Delta_{sw} \beta_{JP}}{\partial \beta_{12}} & \dots & \frac{\partial \Delta_{sw} \beta_{JP}}{\partial \beta_{JP}} \end{bmatrix}$$

onde a expressão do elemento da diagonal principal do bloco 4 da matriz é

$$\frac{\partial \Delta_{sw} \beta_{jp}}{\partial \beta_{jp}} = \frac{1}{\hat{N}} \sum_{i \in s} w_i [B_j^2 z_{ip}^2 (v_{ij}^2 - 1)^2 + 2B_j z_{ip}^2 v_{ij}(v_{ij}^2 - 1)(B_1 v_{i1} + \dots + B_J v_{iJ} - y_i)]$$

(6.39)

e os termos fora da diagonal principal da matriz são:

$$\frac{\partial \Delta_{sw} \beta_{jp}}{\partial \beta_{\ell q}} = \frac{1}{\hat{N}} \sum_{i \in s} w_i [B_j^2 z_{ip} z_{iq} (v_{ij}^2 - 1)^2 + 2B_j z_{ip} z_{iq} v_{ij}(v_{ij}^2 - 1)(B_1 v_{i1} + \dots + B_J v_{iJ} - y_i)],$$

*para*  $\ell = j$ . (6.40)

e

$$\frac{\partial \Delta_{sw} \beta_{jp}}{\partial \beta_{\ell q}} = \frac{1}{\hat{N}} \sum_{i \in s} w_i [B_j B_\ell z_{ip} z_{iq} (v_{ij}^2 - 1)(v_{i\ell}^2 - 1)] \quad , \textit{para } \ell \neq j. \quad (6.41)$$

em que  $\ell = 1, \dots, J$  e  $q = 1, \dots, P$ .

O cálculo da parcela  $\hat{V}_p \left[ \sum_{i \in s} w_i \mathbf{u}_i(\hat{\boldsymbol{\theta}}) \right]$  presente na eq.(6.5) é realizado utilizando-se técnicas usuais de estimação de variância. As mais utilizadas são o método do conglomerado primário e a reamostragem, cujas descrições detalhadas podem ser encontradas em WOLTER (2006). Os procedimentos apresentados nesta seção foram programados utilizando o software estatístico R, para o caso tratado no

capítulo 7.

# Capítulo 7

## Avaliação da abordagem de superpopulação em RNA

### 7.1 Estratégia geral

A abordagem para a avaliação da modelagem de superpopulação é fundamentada no trabalho de HANSEN *et al.* (1983). O procedimento se baseia na construção de 1000 populações hipotéticas, de tamanho  $N=14.000$  elementos, e na seleção de uma amostra de cada população. As amostras são selecionadas seguindo um esquema amostral complexo (utilizaremos um esquema de amostragem estratificada simples com alocação igual).

Após a construção das populações e seleção das amostras, ajustamos uma RNA para cada população e a cada uma das amostras utilizando a abordagem clássica, como descrita no capítulo 5. Tais ajustes fornecem estimativas dos parâmetros da RNA considerando os dados populacionais e estimativas dos parâmetros considerando os dados amostrais, supondo um esquema de amostragem simples com reposição, o que corresponde a ignorar o plano amostral que foi efetivamente usado para obter as amostras. Em seguida, para o mesmo conjunto de amostras ajustamos RNAs utilizando a abordagem de superpopulação descrita no capítulo 6, que leva em conta o plano amostral complexo de fato usado para seleção das amostras.

A comparação da qualidade do ajuste dos modelos de RNA estimados utilizando a abordagem clássica e a abordagem de superpopulação é realizada empregando-se medidas de avaliação dos estimadores. Tais medidas se baseiam em comparar os resultados amostrais obtidos com os resultados populacionais.

## 7.2 População Hipotética

Definimos uma população gerada como uma amostra aleatória de 14.000 elementos de uma superpopulação bivariada onde a variável  $z$  possui distribuição gama com função de densidade:

$$f(z) = 0,02 z \exp\left(\frac{-z}{5}\right)$$

e a variável  $y$  condicional em  $z$  tem distribuição gama com função de densidade:

$$g(y; z) = (1/b^c \Gamma(c)) y^{c-1} \exp(-y/b)$$

onde,

$$b = 1,25 z^{3/2} (8 + 5z^2)^{-1},$$
$$c = 0,04 z^{-3/2} (8 + 5z^2)^2 \tanh\left(\frac{z}{10} + 10\right)$$

cuja média e variância condicionais são respectivamente:

$$E(y|z) = 0,05(0,8 + 5z^2) \tanh\left(\frac{z}{10} + 10\right)$$
$$Var(y|z) = 0,0625 z^{3/2} \tanh\left(\frac{z}{10} + 10\right)$$

A fig.(7.1) apresenta o gráfico de um exemplo da população hipotética. As fig.(7.3) e fig(7.4) mostram os histogramas das variáveis  $y$  e  $z$ , respectivamente. Os histogramas mostram uma grande assimetria positiva em ambas as variáveis. Sendo assim, para amenizar a assimetria aplicamos o logaritmo neperiano nas variáveis  $y$  e  $z$  antes de ajustar a rede neural. A fig.(7.2) apresenta o gráfico do logaritmo das variáveis  $y$  e  $z$  da população. As fig.(7.5) e fig.(7.6) apresentam os histogramas das variáveis  $\ln(z)$  e  $\ln(y)$ , os quais demonstram que a forte assimetria à direita das variáveis foi amenizada e até revertida em ligeira assimetria à esquerda..



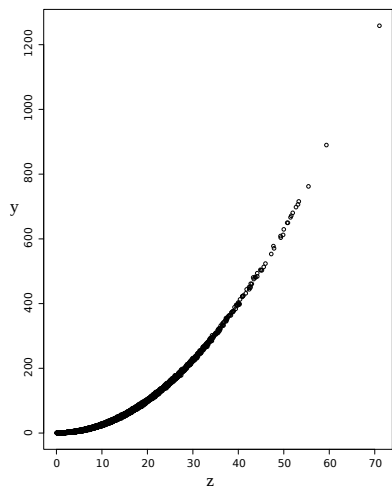


Figura 7.1: População hipotética

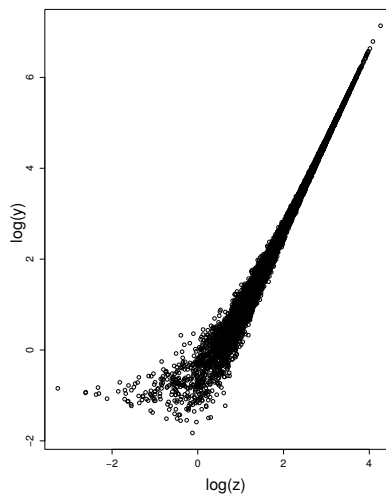


Figura 7.2: Log da população hipotética

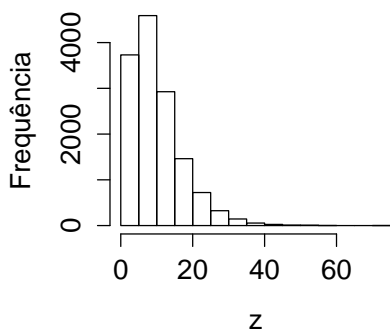


Figura 7.3: Histograma da variável  $z$

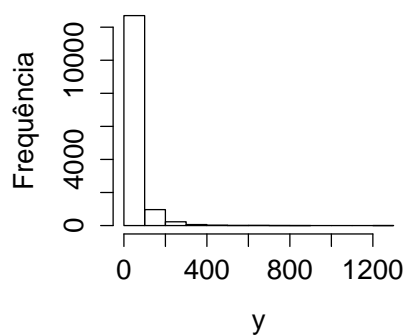


Figura 7.4: Histograma da variável  $y$

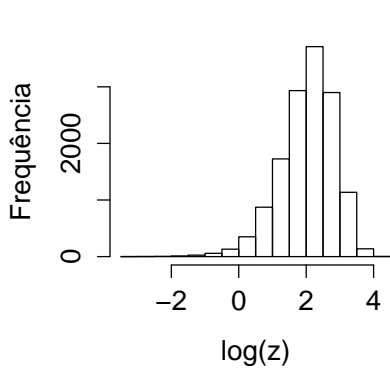


Figura 7.5: Histograma do  $\ln(z)$

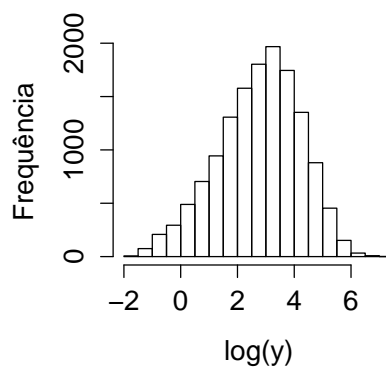


Figura 7.6: Histograma do  $\ln(y)$

### 7.3 Procedimento de seleção das amostras

A amostragem estratificada simples com alocação igual é utilizada para a melhoria de precisão das estimativas e por ser um esquema de amostragem complexa com um procedimento amostral relativamente simples. A amostragem estratificada foi implementada dividindo-se cada população gerada em 10 estratos definidos por intervalos da *variável*  $z$ <sup>1</sup>, tais que os totais da variável  $z$  são aproximadamente os mesmos para cada estrato. Em seguida, amostras aleatórias simples sem reposição de tamanhos iguais (2, 4, 10, 20 e 200 unidades selecionadas por estrato) são selecionadas de cada um dos estratos. Os cinco tamanhos totais de amostra são 20, 40, 100, 200 e 2000, respectivamente. O procedimento de seleção foi repetido, independentemente, para os cinco tamanhos de amostra. A tab.(7.1) mostra um exemplo de amostra estratificada simples com alocação igual de tamanho 20 com os respectivos pesos, onde  $N_h$  é o tamanho do estrato  $h$  e  $n_h$  é o tamanho da amostra selecionada do estrato  $h$ .

O procedimento de seleção da amostra implica em pesos desiguais para elementos em estratos diferentes. A alocação da amostra favorece a observação mais frequente de elementos com maiores valores de  $z$ , e portanto, também de  $y$ . Este tipo de esquema é comumente usado na amostragem de empresas ou estabelecimentos.

---

<sup>1</sup>A variável  $z$  é suposta conhecida para cada elemento da população finita e a variável  $y$  é suposta conhecida somente para aqueles elementos da população que são incluídos na amostra observada.

Estrato	y	z	$N_h$	$n_h$	$w_i$	$w_i^*$
1	0,0000676145	2,4241966151	4208	2	2104	3,0
1	0,0001364158	3,8424689340	4208	2	2104	3,0
2	0,0002023182	5,5817079833	2124	2	1062	1,5
2	0,0002767325	6,6734641442	2124	2	1062	1,5
3	0,0003512387	8,0060423360	1624	2	812	1,2
3	0,0002819080	8,0314199552	1624	2	812	1,2
4	0,0004039954	10,3505697992	1335	2	668	1,0
4	0,0004421166	10,9192679174	1335	2	668	1,0
5	0,0004822328	11,4047128963	1131	2	566	0,8
5	0,0004980336	11,6305508674	1131	2	566	0,8
6	0,0005084017	13,4141529557	976	2	488	0,7
6	0,0005457234	14,5039367530	976	2	488	0,7
7	0,0005713514	16,3375963475	846	2	423	0,6
7	0,0005767591	17,0155286031	846	2	423	0,6
8	0,0006045892	19,7130197084	719	2	360	0,5
8	0,0006063066	19,7857411939	719	2	360	0,5
9	0,0006067015	23,8255815619	597	2	299	0,4
9	0,0005867113	24,9213838214	597	2	299	0,4
10	0,0004900529	30,7408714181	441	2	221	0,3
10	0,0004150519	35,2425270043	441	2	221	0,3

Tabela 7.1: Exemplo de amostragem estratificada simples com alocação igual e  $n_h = 2$

## 7.4 Ajuste dos modelos de RNA

A arquitetura da RNA escolhida para o ajuste dos dados simulados (populacionais e amostrais) está representada na fig.(7.7). A preferência por esta arquitetura simples é por consequência da identificação da RNA, especificamente pelo fato de não possuir neurônios irrelevantes na camada intermediária. Assim, a identificabilidade da RNA é assegurada pelas restrições (i) e (ii) apresentadas na seção 5.2.

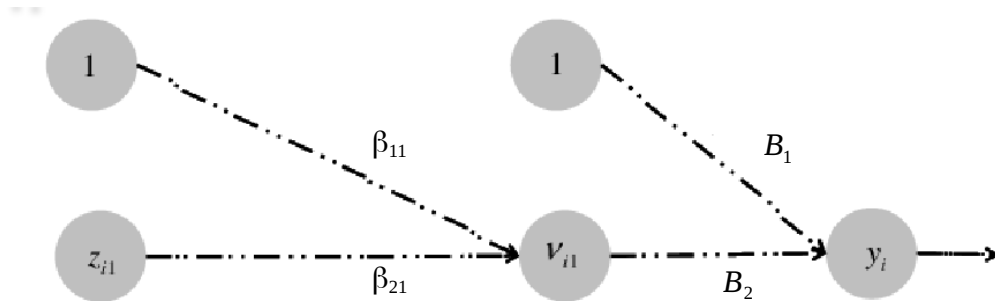


Figura 7.7: Representação gráfica da RNA a ser ajustada aos dados

A rede da fig.(7.7) possui um neurônio na camada de entrada,  $z_{i1}$ , um neurônio tangente hiperbólica na camada intermediária com um limiar,  $v_{i1} = \tanh(\beta_{01} + \beta_{11}z_{i1})$ , e um neurônio linear na camada de saída com um limiar,  $y_i = B_0 + B_1v_{i1}$ .

O limiar é representado por uma sinapse de peso fixo conectado a uma entrada fixa de valor unitário e pode ser visto como um intercepto do modelo (por analogia a um modelo de regressão linear). Assim, o limiar  $\beta_{01}$  é o valor onde a tanh intercepta o eixo das abscissas e o limiar  $B_0$  é o valor onde a tanh intercepta o eixo das ordenadas. Os coeficientes  $\beta_{11}$  e  $B_1$  fornecem a inclinação e a amplitude da tanh, respectivamente.

## 7.5 Medidas de avaliação dos estimadores

Considere um parâmetro  $\theta$  de um modelo de interesse proposto para descrever aspectos da população finita  $U$ . Seja  $A$  um plano amostral usado para obtenção de uma amostra da população finita  $U$  para a estimação de  $\theta$ , e seja  $S$  o conjunto formado por todas as amostras possíveis  $s = \{i_1, \dots, i_n\}$  que poderiam ser selecionadas da população  $U$  seguindo o plano amostral  $A$ .

A avaliação do estimador  $\hat{\theta}$  para  $\theta$ , sob o plano amostral  $A$ , é realizada utilizando o viés,  $B(\hat{\theta})$ , a variância  $V(\hat{\theta})$  e o erro relativo médio,  $ERM(\hat{\theta})$ , vide BOLFARINE e BUSSAB (2005).

**Definição 7.5.1.** O valor esperado de  $\hat{\theta}$ , sob um plano amostral  $A$ , denotado por  $E_A(\hat{\theta})$  é definido por:

$$E_A(\hat{\theta}) = \sum_{s \in S} Pr(s) \hat{\theta}(s)$$

onde  $Pr(s)$  é a probabilidade de seleção da amostra  $s$  e  $\hat{\theta}(s)$  é o valor de  $\hat{\theta}$  para amostra  $s$ .

**Definição 7.5.2.** A variância de  $\hat{\theta}$ , sob um plano amostral  $A$ , denotada por  $Var_A(\hat{\theta})$  é definida por:

$$Var_A(\hat{\theta}) = \sum_{s \in S} Pr(s) [\hat{\theta}(s) - E_A(\hat{\theta})]^2$$

**Definição 7.5.3.** Um estimador  $\hat{\theta}$  é não viciado sob um plano amostral  $A$ , se

$$E_A[\hat{\theta}] = \theta$$

**Definição 7.5.4.** O viés do estimador  $\hat{\theta}$ , sob o plano amostral  $A$ , é dado por

$$B_A[\hat{\theta}] = E_A[\hat{\theta} - \theta] = E_A[\hat{\theta}] - \theta$$

e o erro quadrático médio desse mesmo estimador é dado por

$$EQM_A[\hat{\theta}] = E_A[\hat{\theta} - \theta]^2$$

Assim temos

$$EQM_A[\hat{\theta}] = Var_A[\hat{\theta}] + B_A^2[\hat{\theta}]$$

A variância e o EQM são medidas de difícil interpretação, pois suas unidades de medida são expressas ao quadrado. Sendo assim, introduzimos duas medidas adimensionais frequentemente usadas que são o *coeficiente de variação* e *erro relativo médio*.

**Definição 7.5.5.** *O coeficiente de variação do estimador  $\hat{\theta}$ , sob um plano amostral  $A$ , é dado pela expressão*

$$CV_A[\hat{\theta}] = \frac{\sqrt{Var_A[\hat{\theta}]}}{|E_A[\hat{\theta}]|}$$

O coeficiente de variação mede a dispersão das estimativas de  $\theta$  em relação ao valor esperado de  $\hat{\theta}$ . É útil, principalmente, para estimadores não viciados com média não nula.

**Definição 7.5.6.** *O erro relativo médio do estimador  $\hat{\theta}$ , sob um plano amostral  $A$ , é dado pela expressão*

$$ERM_A[\hat{\theta}] = \frac{\sqrt{EQM_A[\hat{\theta}]}}{|\theta|}$$

O erro relativo médio pode ser interpretado como um coeficiente de variação considerando a componente de vício.

O erro amostral,  $\hat{\theta} - \theta$ , somente pode ser obtido se o parâmetro populacional  $\theta$  for conhecido<sup>2</sup>. Assim, a avaliação é realizada verificando-se as propriedades do estimador, segundo as medidas anteriores, sob um plano amostral  $A$ .

### 7.5.1 Estimação por simulação das medidas de qualidade

A avaliação dos estimadores, sob o plano amostral  $A$ , é realizada por simulação. O processo é realizado ajustando a RNA, como descrito na seção 7.4. Sejam  $s_1, s_2, \dots, s_R$  as réplicas independentes de amostras extraídas de  $U$  sob o plano amostral  $A$ . Assim,  $s_1, s_2, \dots, s_R$  formam uma amostra aleatória com reposição dos elementos do conjunto  $S$  formado por todas as amostras possíveis, sob o plano amostral  $A$ , aplicado à população finita  $U$ . Portanto, os valores  $\hat{\theta}(s_1), \hat{\theta}(s_2), \dots, \hat{\theta}(s_R)$

---

<sup>2</sup>O parâmetro populacional na prática não é conhecido.

formam uma amostra aleatória simples com reposição dos valores possíveis para o estimador  $\hat{\theta}$ , sob o plano amostral  $A$ , aplicado à população  $U$ .

A esperança do estimador  $\hat{\theta}$  é estimada pela média aritmética das  $R$  estimativas  $\hat{\theta}(s_1), \hat{\theta}(s_2), \dots, \hat{\theta}(s_R)$  correspondente às réplicas  $s_1, s_2, \dots, s_R$ :

$$\widehat{E}_A(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R \hat{\theta}(s_r) \quad (7.1)$$

A esperança do estimador  $\theta_{U_r}$  é estimada pela média aritmética das  $R$  estimativas correspondente às réplicas da população:

$$\widehat{E}(\theta_U) = \frac{1}{R} \sum_{r=1}^R \theta_{(U_r)} \quad (7.2)$$

A estimativa do vício do estimador  $\hat{\theta}$  é dada por:

$$\widehat{B}_A(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R [\hat{\theta}(s_r) - \theta_{(U_r)}] \quad (7.3)$$

O erro quadrático médio é dado por:

$$\widehat{EQM}_A(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R [\hat{\theta}(s_r) - \theta_{(U_r)}]^2 \quad (7.4)$$

Utilizando a definição 7.5.4 o cálculo da variância é dado por:

$$\widehat{Var}_A(\hat{\theta}) = \widehat{EQM}_A(\hat{\theta}) - \widehat{B}_A^2(\hat{\theta}) \quad (7.5)$$

O coeficiente de variação de  $\hat{\theta}$  é obtido pela razão entre o desvio padrão de  $\hat{\theta}$  e a estimativa da esperança de  $\hat{\theta}$ :

$$\widehat{CV}_A(\hat{\theta}) = \frac{\sqrt{\widehat{Var}_A(\hat{\theta})}}{|\widehat{E}_A(\hat{\theta})|} \quad (7.6)$$

A estimativa do erro relativo médio  $\hat{\theta}$  é dada por:

$$\widehat{ERM}_A(\hat{\theta}) = \frac{\sqrt{\widehat{EQM}_A(\hat{\theta})}}{|\widehat{E}(\theta_U)|} \quad (7.7)$$

## 7.6 Análise dos coeficientes estimados pela RNA

A presente seção tem por objetivo apresentar a avaliação empírica da abordagem de superpopulação para modelos de RNA. A avaliação é feita considerando a especificação mais simples da RNA identificável, como definida na fig.(7.7).

O algoritmo de treinamento BP foi programado utilizando-se o software de estatística  $R^3$ , o hardware utilizado foi Intel Core I5 com 16GB de memória RAM. Os resultados do algoritmo programado, sem considerar as restrições dos parâmetros para identificação, como mostra na seção 5.2, foram comparados aos obtidos utilizando o pacote ‘neuralnet’, do próprio software, com os mesmos parâmetros tais como taxa de aprendizagem, valores iniciais e método de treinamento. Os resultados ficaram muito próximos, ficando desta forma validado o algoritmo programado.

Após a validação do algoritmo inserimos as restrições, seção 5.2, para tornar a RNA identificável. As populações e amostras geradas foram numeradas de 1 a 1000 para identificação e foi ajustada uma RNA, como definida na fig.(7.7), para cada uma das populações e amostras. Em alguns casos não houve convergência em determinadas amostras quando foi usado o algoritmo clássico para ajuste da RNA, como ocorreu com as amostras de tamanho 100 e 200 cujos números de referência são 416/790 e 497, respectivamente. No caso do ajuste da RNA considerando a modificação para lidar com as amostras complexas houve convergência em todos os casos. As amostras para as quais não houve convergência foram retiradas da análise subsequente.

A identificabilidade das RNAs estimadas foi verificada pelos gráficos das estimativas dos parâmetros e observando se todos os pontos estavam presentes em uma mesma região do gráfico. As fig.(7.8) e fig.(7.9) apresentam os gráficos das estimativas  $\beta_{11} \times \beta_{21}$  e  $B_0 \times B_1$  obtidas pela RNA clássica para amostra de tamanho 200, por exemplo.

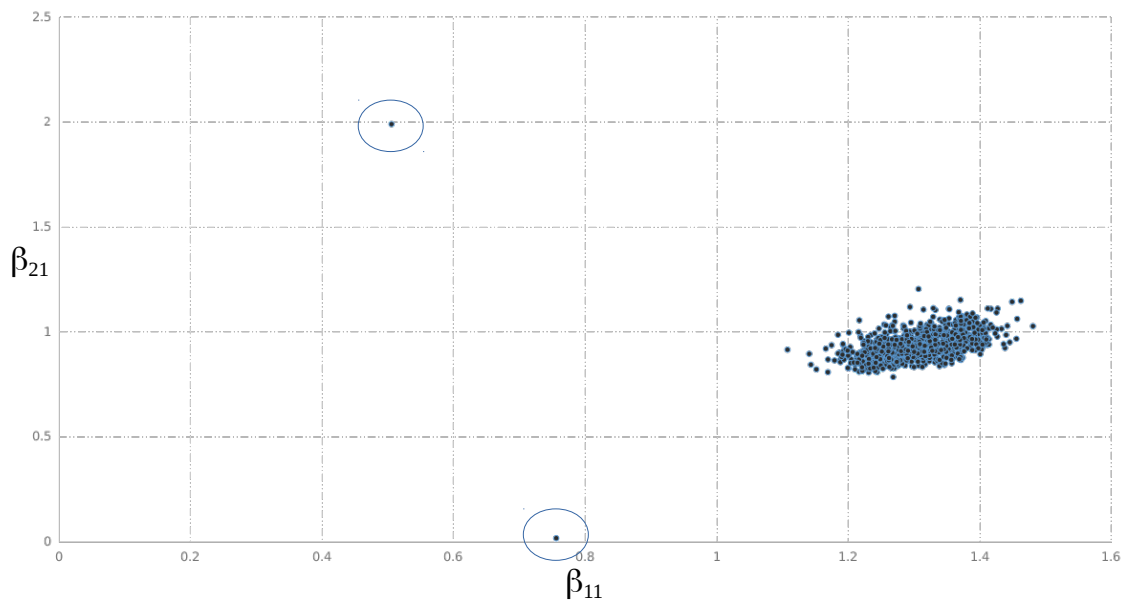


Figura 7.8:  $\beta_{11} \times \beta_{21}$  - Estimativas da RNA clássica

<sup>3</sup>sítio: [www.r-project.org](http://www.r-project.org)

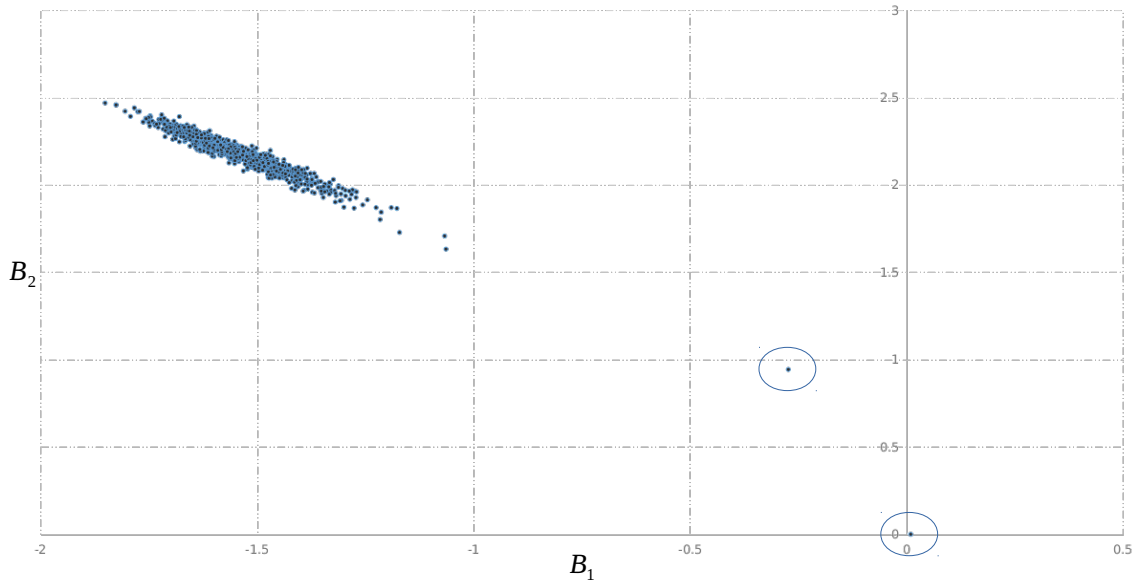


Figura 7.9:  $B_1 \times B_2$  - Estimativas da RNA clássica

As fig.(7.10) e fig.(7.11) apresentam os gráficos das estimativas sob o modelo de superpopulação para amostra de tamanho 200, por exemplo.

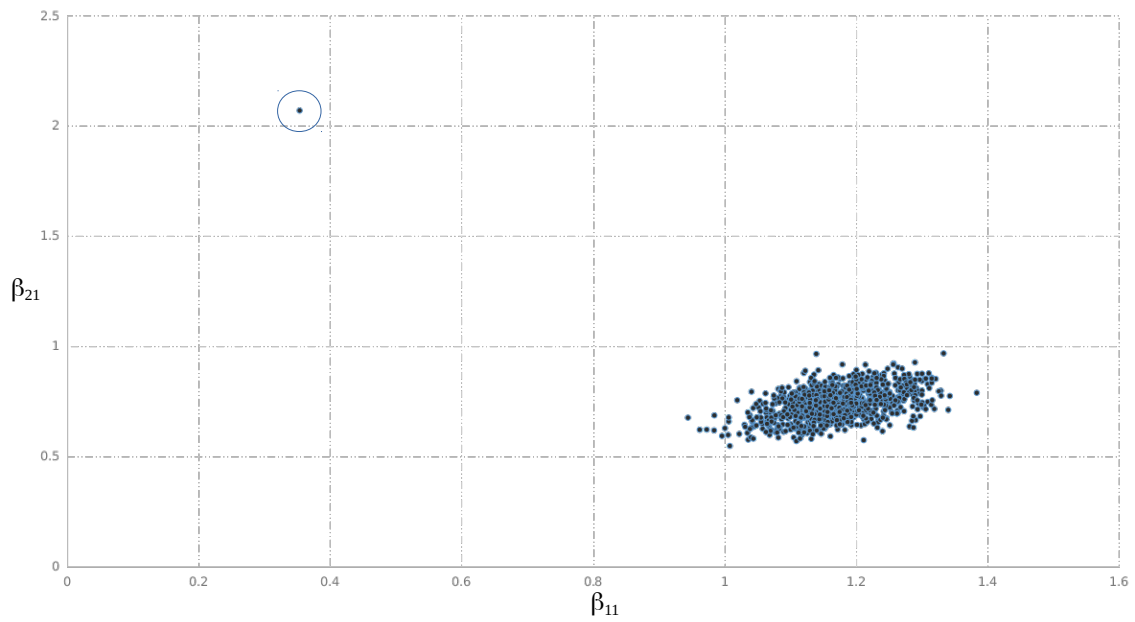


Figura 7.10:  $\beta_{11} \times \beta_{21}$  - Estimativas da RNA sob superpopulação



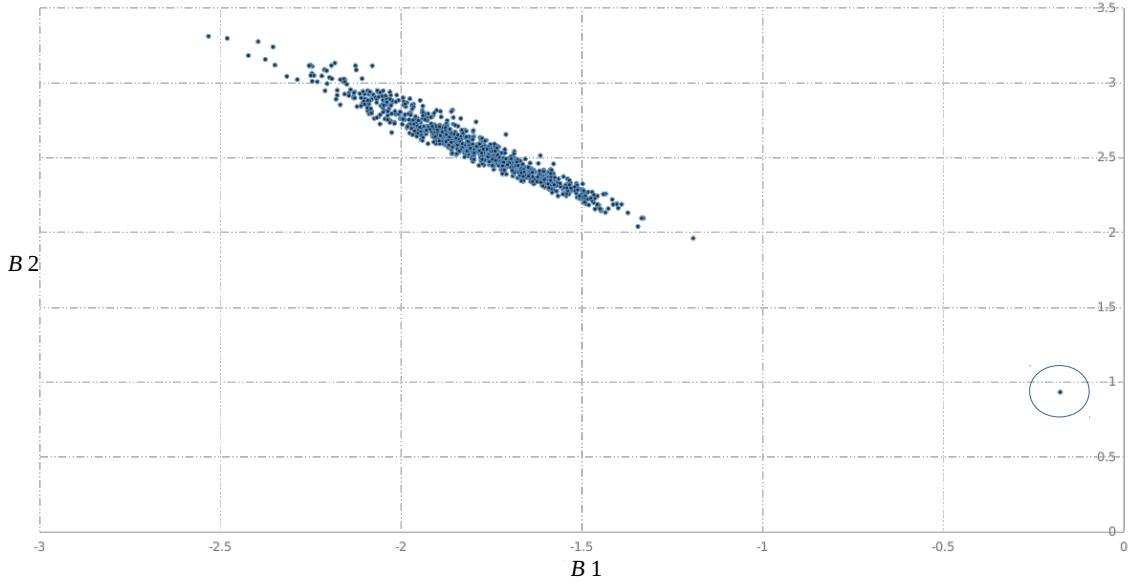


Figura 7.11:  $B_1 \times B_2$  - Estimativas da RNA sob superpopulação

Os gráficos apresentam alguns poucos pontos que estão afastados da nuvem de pontos. Isto indica que as restrições propostas pela literatura, como mencionadas na seção 5.2, não foram suficientes para garantir identificabilidade dos modelos de RNA em determinadas amostras.

A fim de identificar as amostras cujas estimativas ficaram afastadas da região de identificabilidade utilizamos a distância de Mahalanobis, a qual é definida como:

$$d_{mah}^2 = (\boldsymbol{\beta} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\beta} - \boldsymbol{\mu}) \quad (7.8)$$

onde  $\boldsymbol{\beta}$  é o vetor de estimativas com  $p$  colunas,  $\boldsymbol{\mu}$  é o vetor de médias das estimativas e  $\boldsymbol{\Sigma}$  é a matriz de covariâncias de dimensão  $p \times p$  dos valores estimados.

A distância de Mahalanobis foi calculada para as estimativas obtidas pela RNA clássica e sob o modelo de superpopulação. A medida de distância foi comparada a  $\chi_4^2(0,999)$ . Se a  $d_{mah}^2 > \chi_4^2(0,999)$  consideramos as estimativas geradas por determinada amostra fora da região de identificabilidade. A tab.(7.2) apresenta as amostras que geraram estimativas discrepantes. Este critério também foi aplicado aos ajustes considerando toda população, e constatou-se a inexistência de valores discrepantes.

Amostra n = 20		Amostra n = 40		Amostra n = 100		Amostra n = 200		Amostra n = 2000	
Clássica	Superpop.	Clássica	Superpop.	Clássica	Superpop.	Clássica	Superpop.	Clássica	Superpop.
32	32	36	114	93	36	50	69	312	34
140	140	52	164	114	93	69	360	573	
154	154	164	175	253	253	420	535		
184	184	192	192	416	380	497	570		
255	282	217	217	428	428		874		
281	298	394	250	681	450		991		
282	340	413	279	689	469				
381	381	422	291	727	573				
436	432	647	394	728	618				
460	460	872	410	790	681				
683	683	897	517		689				
811	782	910	647		727				
934	785	932	660		728				
947	844		754		777				
	934		799		792				
			897		925				
			910		955				
			914						
			932						
			957						
			966						

Tabela 7.2: Rótulos das amostras cujas estimativas apresentaram valores discrepantes

Após a identificação das amostras cujas estimativas apresentaram estimativas discrepantes, ajustamos 100 vezes a RNA para cada amostra e calculamos novamente a distância de Mahalanobis. Seleccionamos aleatoriamente um conjunto de estimativas cuja  $d_{mah}^2 \leq \chi_4^2(0,999)$  e substituímos no conjunto de estimativas utilizadas para as análises subsequentes. Tal procedimento foi adotado visando eliminar o efeito das estimativas discrepantes que ficaram fora da região de identificabilidade.

A tab.(7.3) apresenta as estatísticas dos valores dos coeficientes estimados pela RNA clássica a partir dos dados populacionais com 1.000 réplicas.

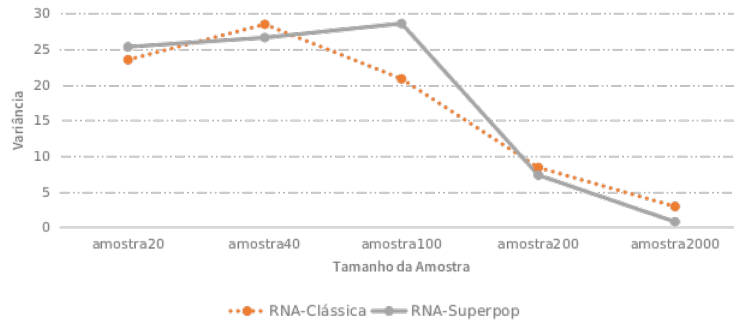
Estatística	$\beta_{11}$	$\beta_{21}$	$B_0$	$B_1$
Mínimo	0,6504	0,7890	-0,9229	1,8792
Média	0,6774	0,8141	-0,8788	1,9232
Máximo	0,7134	0,8448	-0,8359	1,9713
Desvio Padrão	0,0089	0,0087	0,0128	0,0136

Tabela 7.3: Estatísticas resumo dos coeficientes estimados com as populações geradas

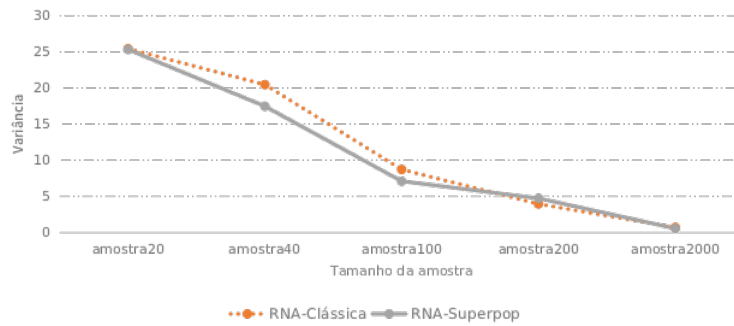
A avaliação do experimento se inicia com a análise da variância das estimativas dos parâmetros. Em seguida, analisamos o viés e finalizamos a análise do experimento com o ERM dos parâmetros estimados.

A fig.(7.12) apresenta a variância estimada, de acordo com eq.(7.5), dos coeficientes estimados pela RNA clássica (linha pontilhada) e sob o modelo de superpopulação (linha sólida) para cada tamanho de amostra. Podemos observar, em ambos os casos, que à medida que o tamanho da amostra cresce a variância decresce. Porém, as estimativas das variâncias da abordagem de superpopulação para todos os tamanhos de amostra são superiores às estimativas da RNA clássica. Tais fatos são compatíveis com a literatura pertinente como visto em CASELLA e BER-

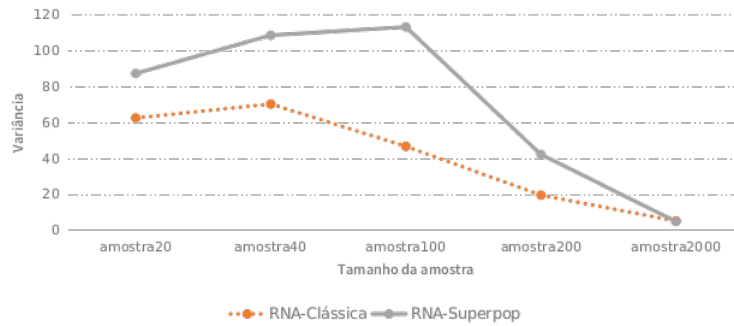
GER (2001) e LARSON (1969). A tab.(7.4) apresenta os valores das estimativas das variâncias dos coeficientes estimados da RNA clássica e sob o modelo de superpopulação, respectivamente.



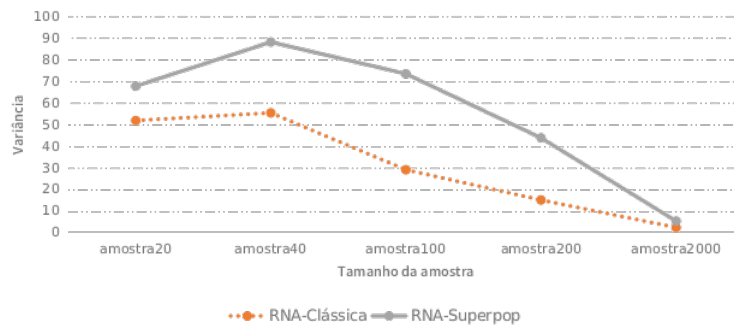
(a) Coeficiente  $\beta_{11}$



(b) Coeficiente  $\beta_{21}$



(c) Coeficiente  $B_0$



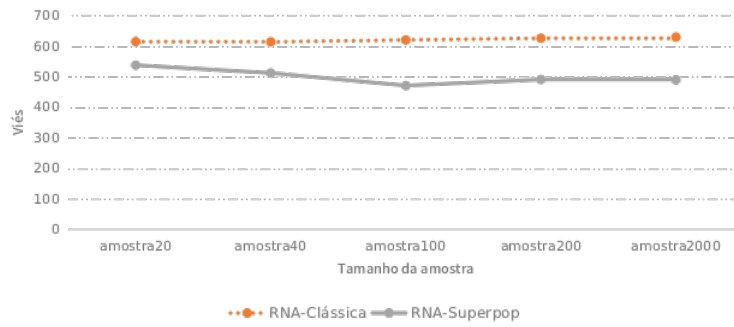
(d) Coeficiente  $B_1$

Figura 7.12: Variância estimada ( $\times 1000$ ) dos coeficientes da RNA por tipo e tamanho de amostra

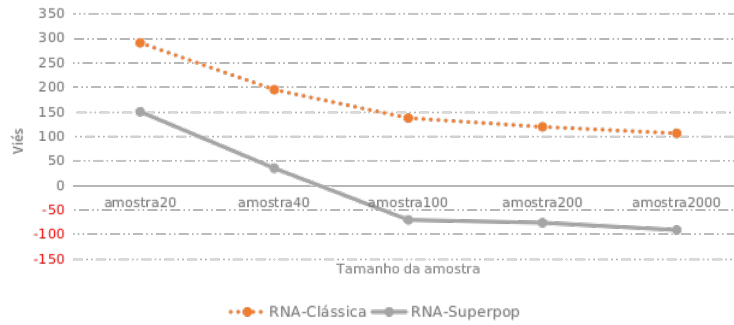
Sample size	$\beta_{11}$		$\beta_{21}$		$B_1$		$B_2$	
	RNAC	RNASP	RNAC	RNASP	RNAC	RNASP	RNAC	RNASP
<b>20</b>	23.50	25.31	25.38	25.29	62.38	87.14	51.74	67.67
<b>40</b>	28.46	26.60	20.43	17.44	70.13	108.41	55.27	88.12
<b>100</b>	20.82	28.58	8.68	7.07	46.63	112.98	28.92	73.38
<b>200</b>	8.40	7.33	3.92	4.70	19.49	42.08	14.91	43.70
<b>2000</b>	2.94	0.78	0.68	0.53	5.10	4.80	2.15	5.14

Tabela 7.4: Variância estimada ( $\times 1000$ ) dos coeficientes da RNA sob o modelo clássico (RNAC) e sob a abordagem de superpopulação (RNASP) por tamanho de amostra

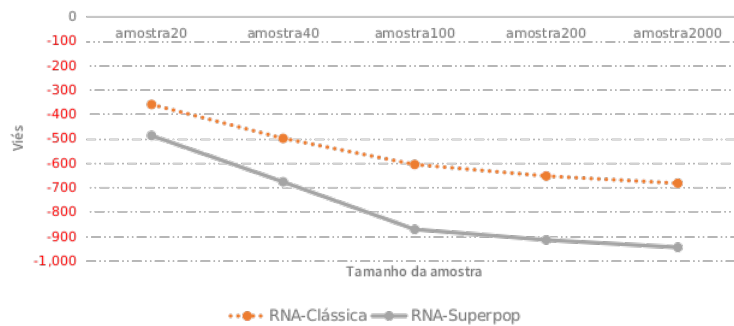
A fig.(7.13) apresenta o viés dos coeficientes estimados pela RNA clássica (linha pontilhada) e sob o modelo de superpopulação (linha sólida) para cada tamanho de amostra. Na análise do viés estamos interessados que ele se aproxime de zero. As estimativas dos coeficientes da primeira camada da RNA  $\beta_{11}$  e  $\beta_{21}$ , fig.(7.13(a)) e fig.(7.13(b)), respectivamente, apresentam menor viés quando as estimativas são obtidas pela abordagem de superpopulação. Tal fato é compatível com a literatura como visto em ISAKI e FULLER, CHAMBERS e SKINNER (2003) e NATHAN e HOLT (1980). Por outro lado, as estimativas dos coeficientes da segunda camada da RNA,  $B_0$  e  $B_1$ , fig.(7.13(c)) e fig.(7.13(d)), respectivamente, sob o modelo de superpopulação, apresentam viés maior que o obtido pela modelagem clássica. Esse achado contraria o esperado conforme a literatura pertinente. A tab.(7.5) apresenta os valores dos coeficientes do viés estimado dos coeficientes da RNA clássica e sob a abordagem de superpopulação, respectivamente.



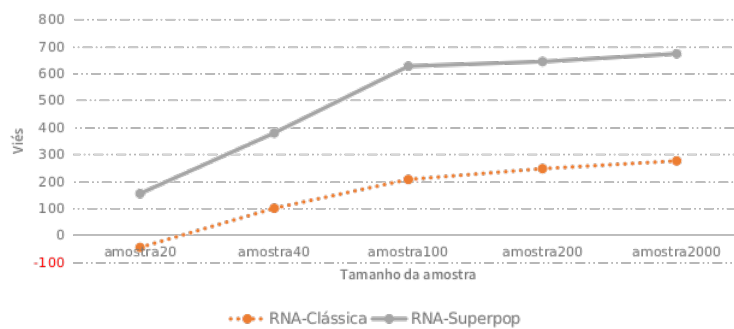
(a) Coeficiente  $\beta_{11}$



(b) Coeficiente  $\beta_{21}$



(c) Coeficiente  $B_0$



(d) Coeficiente  $B_1$

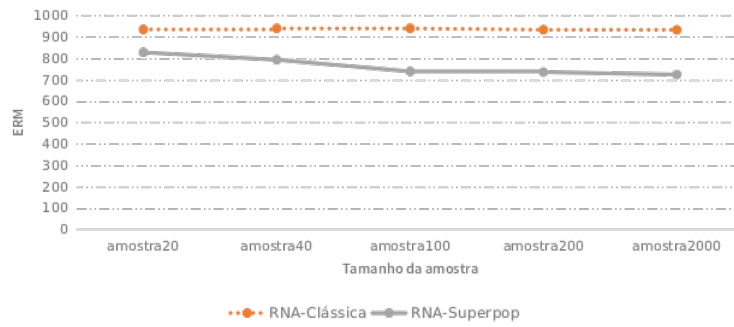
Figura 7.13: Viés estimado ( $\times 1000$ ) dos coeficientes estimados pela RNA por tipo e tamanho de amostra

Sample size	$\beta_{11}$		$\beta_{21}$		$B_1$		$B_2$	
	RNAC	RNASP	RNAC	RNASP	RNAC	RNASP	RNAC	RNASP
<b>20</b>	613.95	537.10	290.03	149.72	-360.81	-488.30	-45.29	154.37
<b>40</b>	613.20	511.38	194.50	34.49	-499.63	-677.69	100.02	378.92
<b>100</b>	619.56	470.55	136.72	-70.93	-606.72	-872.51	206.61	626.58
<b>200</b>	625.19	490.38	118.68	-76.75	-653.96	-915.85	246.68	643.46
<b>2000</b>	628.99	489.21	105.65	-91.48	-683.45	-946.09	275.42	672.66

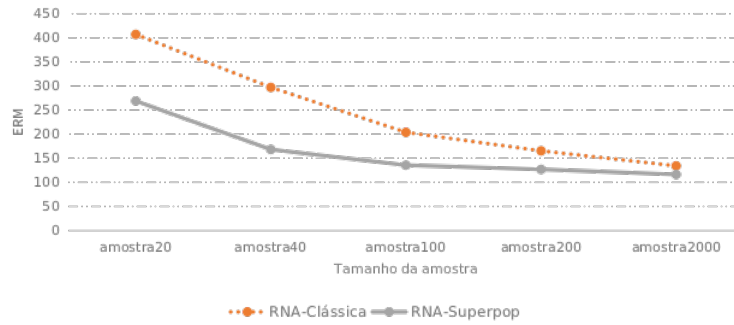
Tabela 7.5: Viés estimado ( $\times 1000$ ) dos coeficientes da RNA sob o modelo clássico (RNAC) e sob a abordagem de superpopulação (RNASP) por tamanho de amostra

A fig.(7.14) apresenta o ERM dos coeficientes estimados pela RNA clássica (linha pontilhada) e sob o modelo de superpopulação (linha sólida) para cada tamanho de amostra.

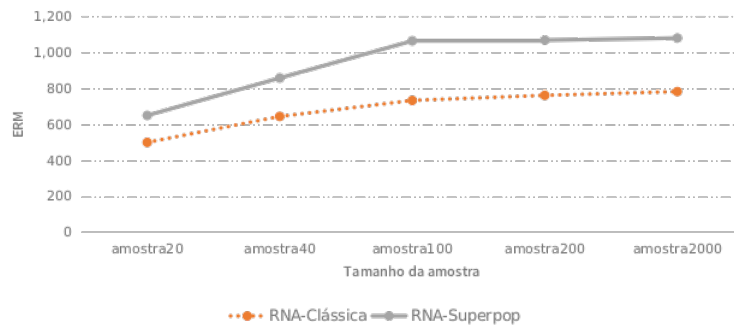
As fig.(7.14(a)) e fig.(7.14(b)) apresentam o ERM dos coeficientes  $\beta_{11}$  e  $\beta_{21}$  das abordagens clássica e de superpopulação, respectivamente. Podemos observar que sob a abordagem de superpopulação a variabilidade relativa a média é menor que da abordagem clássica e decresce à medida que o tamanho da amostra aumenta. Por outro lado, as fig.(7.14(c)) e fig.(7.14(d)) apresentam o ERM dos coeficientes,  $B_1$  e  $B_2$ , cuja variabilidade relativa sob o modelo de superpopulação é maior que da abordagem clássica. O ERM do coeficiente  $B_1$  diminui à medida que o tamanho da amostra aumenta tanto para a abordagem clássica como a de superpopulação, contrastando com o coeficiente  $B_2$  cujo ERM aumenta quando o tamanho da amostra cresce em ambas as abordagens.



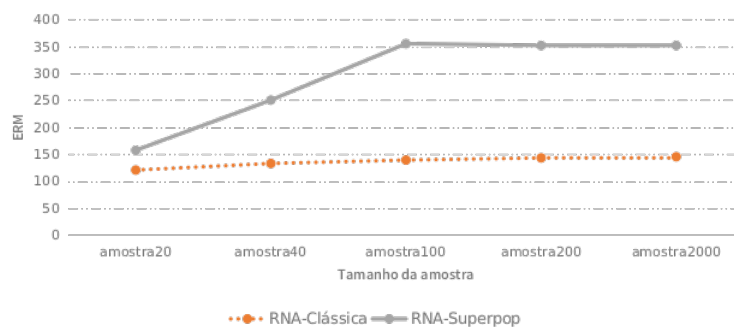
(a) Coeficiente  $\beta_0$



(b) Coeficiente  $\beta_1$



(c) Coeficiente  $B_0$



(d) Coeficiente  $B_1$

Figura 7.14: ERM estimados ( $\times 1000$ ) dos coeficientes estimados pela RNA por tipo e tamanho de amostra

Considerando a simplicidade e as limitações do experimento realizado podemos sugerir o uso da abordagem sob o modelo de superpopulação em RNA objetivando ter predições menos enviesadas. Porém, temos que atentar para alguns fatos impor-



Tamanho da amostra	$\beta_{11}$		$\beta_{21}$		$B_1$		$B_2$	
	RNAC	RNASP	RNAC	RNASP	RNAC	RNASP	RNAC	RNASP
<b>20</b>	934.16	826.93	406.50	268.29	499.33	649.27	120.60	157.28
<b>40</b>	938.85	792.38	296.50	167.65	643.45	857.33	132.85	250.29
<b>100</b>	939.09	738.11	203.26	135.11	732.80	1063.95	139.14	354.94
<b>200</b>	932.78	734.86	164.83	126.42	760.90	1067.96	143.12	351.79
<b>2000</b>	931.97	723.37	133.66	115.88	781.93	1079.43	145.23	351.74

Tabela 7.6: ERM estimado ( $\times 1000$ ) dos coeficientes da RNA sob o modelo clássico (RNAC) e sob a abordagem de superpopulação (RNASP) por tamanho de amostra

tantes. Primeiro, com relação à variância dos coeficientes estimados, o experimento mostrou que o ajuste do modelo de RNA sob a abordagem de superpopulação levou as variâncias maiores que sob a abordagem clássica, e que as variâncias diminuem com o aumento do tamanho da amostra. Tal fato é compatível com a literatura em estatística pertinente, como mostrado anteriormente. Segundo, o viés dos coeficientes estimados,  $B_1$  e  $B_2$ , da camada de saída é maior sob a abordagem de superpopulação que sob a abordagem clássica contrariando o esperado segundo a literatura estatística. Porém, os coeficientes  $\beta_{11}$  e  $\beta_{21}$  da camada de entrada apresentam viés menor sob o modelo de superpopulação. Finalizando, temos que o ERM dos coeficientes,  $\beta_{11}$  e  $\beta_{21}$ , da camada de entrada apresentam menor variabilidade relativa sob o modelo de superpopulação, diferentemente dos coeficientes,  $B_1$  e  $B_2$ , da camada de saída, os quais apresentam maior variabilidade relativa em relação a média sob a abordagem de superpopulação.

O próximo capítulo apresenta uma aplicação prática da abordagem de superpopulação utilizando dados de uma pesquisa que utiliza um esquema amostral complexo, provenientes do IBGE (Instituto Brasileiro de Geografia e Estatística).

## Capítulo 8

# Aplicação aos dados da PNAD da abordagem de superpopulação em RNA

A ilustração do uso da RNA sob a abordagem de superpopulação realizamos uma aplicação a um exemplo real. No presente trabalho utilizamos a *Pesquisa Nacional por Amostra de Domicílios - 2014*<sup>1</sup>, sigla PNAD 2014, de abrangência nacional, para modelagem da renda *per capita* domiciliar.

A RNA clássica e sob a abordagem de superpopulação foram ajustadas aos dados da PNAD 2014. Ajustamos, também, um modelo de regressão linear clássico e sob a abordagem de superpopulação os resultados dos ajustes dos vários modelos foram comparados.

A seção 8.1 descreve a construção da base dados seguindo o que foi proposto em IBGE (2008) para estimação da renda domiciliar *per capita*. Consideramos apenas o estado de Minas Gerais, o qual possui vasta extensão territorial e grande variação socioeconômica. Os municípios do sul do estado, os quais fazem fronteira com os estados do Rio de Janeiro e São Paulo, são influenciados pela economia ativa destes estados e, conseqüentemente, são mais ricos. Os municípios do norte e nordeste de Minas Gerais fazem fronteira com o interior do estado da Bahia e são reconhecidamente mais pobres. Por conta desta distinção o estado de Minas Gerais já foi intitulado “Brasil dentro do Brasil”, vide IBGE (2008).

---

<sup>1</sup>A Pesquisa Nacional por Amostra de Domicílios - PNAD investiga anualmente, de forma permanente, características gerais da população, de educação, trabalho, rendimento e habitação e outras, com periodicidade variável, de acordo com as necessidades de informação para o País, como as características sobre migração, fecundidade, nupcialidade, saúde, segurança alimentar, entre outros temas. O levantamento dessas estatísticas constitui, ao longo dos 48 anos de realização da pesquisa, um importante instrumento para formulação, validação e avaliação de políticas orientadas para o desenvolvimento socioeconômico e a melhoria das condições de vida no Brasil.

## 8.1 Construção da base de dados

As variáveis explicativas utilizadas para a modelagem da renda domiciliar *per capita* do estado de Minas Gerais provenientes da PNAD 2014 estão na tab.(8.1). O dicionário de variáveis se encontra no apêndice A.

Variável	Descrição
V0205	Número total de cômodos
V0206	Número de cômodos servindo como dormitório
V2016	Número de banheiros
V8005	Idade em anos
cri_ate7	Crianças até 7 anos
cri_7a14	Crianças com idade entre 7 e 14 anos
idoso	Número de idosos (pessoas com 60 anos ou mais)
tipofam	Tipo de família
V0215	Existência Banheiro
V0302	Sexo do responsável pelo domicílio
V0404	Cor/raça do responsável pelo domicílio
V0225	Existência de rádio no domicílio
TV	Existência de televisão no domicílio
V0230	Existência de máquina de lavar roupa no domicílio
V0228	Existência de geladeira no domicílio
V0231	Existência de microcomputador no domicílio
V2032	Existência de automóvel para uso particular
V4111	Vive em companhia de cônjuge ou companheiro(a)
V0207	Condição de ocupação do domicílio
V0217	Tipo de esgotamento sanitário
V0602	Frequência a escola
V4706	Posição na ocupação

Tabela 8.1: Variáveis independentes para estimação dos modelos

Excetuando a variável V8005, o restante das variáveis, tab.(8.1), foram dicotomizadas da seguinte forma; 1 para a característica presente e -1 para característica ausente. Algumas variáveis tiveram as categorias agrupadas:

V0205: o número total de cômodos do domicílio foi agrupado em três classes: domicílios que possuem 1 a 4 cômodos, 5 a 8 cômodos e 9 cômodos ou mais;

V0206: o número de cômodos servindo como dormitório foi agrupado em quatro classes: domicílios que possuem 1 dormitório, 2 dormitórios, 3 dormitórios e 4 ou mais dormitórios;

V2016: o número de banheiros foi agrupado em três classes: domicílios sem banheiro, com 1 banheiro e com 2 banheiros ou mais;

tipofam: o tipo de família foi agrupado em três classes: casal sem filhos, casal com filhos e outros;

V0404: na variável cor/raça consideramos somente 4 grupos: branca, preta, amarela e parda;

V0207: a variável condição de ocupação do domicílio foi agrupada em três classes: próprio ou de algum morador (já pago) , alugado e outros;

V0217: a variável de esgotamento sanitário foi agrupada em duas classes: possui esgotamento sanitário e não possui esgotamento sanitário.

## 8.2 Aplicação e Resultados

A aplicação do modelo de RNA requer especificação da estrutura da rede, das variáveis explicativas que serão usadas para a predição do fenômeno e o número total de neurônios na camada intermediária. Não utilizamos as restrições para identificação, como descrito na seção 5.2, pois a RNA estimada não satisfaz a restrição de monotonicidade crescente dos pesos,  $B_i$ , da camada de saída. Assim sendo, o objetivo desta aplicação é melhorar a predição da renda domiciliar *per capita* utilizando a abordagem de superpopulação em RNA e comparando com os modelos comumente utilizados (regressão linear e RNA clássica).

As variáveis utilizadas para a predição da renda domiciliar *per capita* foram apresentadas na tab.(8.1). A estratégia de construção da RNA se iniciou pela determinação do número de neurônios da camada intermediária utilizando o procedimento computacional do teste LM, seção 5.2.2. Assim, consideramos a RNA representada na eq.(5.25). Podemos observar que existe uma parte linear, o primeiro termo da equação, e uma parte não linear (neurônios adicionais) somada ao termo linear. O teste tem início com a estimação da parte linear da eq.(5.25) por mínimos quadrados ordinários utilizando as variáveis da tab.(8.1). O modelo foi estimado com as variáveis normalizadas e tendo como variável dependente o logaritmo natural da renda. Os resultados dessa estimação se encontram no apêndice B.4 . A seguir, testamos a hipótese nula de interesse, eq.(5.26), ou seja, se a RNA é capaz de uma representação exata da função desconhecida  $g$ .

A estatística de teste é obtida fazendo-se a regressão com a variável dependente  $u_i$  e as variáveis preditoras  $\nabla f(\mathbf{Z}_i, \hat{\boldsymbol{\theta}})$  ( $= \mathbf{z}$ , neste caso) e  $\boldsymbol{\Psi}_i$  o vetor de neurônios adicionais da camada intermediária. Sendo  $R^2$  o coeficiente de correlação múltipla da regressão temos, segundo WHITE (1989a), que  $nR^2 \rightarrow \chi_r^2$ , onde  $n$  é o tamanho da amostra.

O teste é realizado sequencialmente, isto é, acrescentamos um neurônio adicional por vez e realizamos o cálculo da estatística de teste  $\chi_{calc}^2$ , comparando com o valor tabelado da  $\chi_r^2$  com  $r$  graus de liberdade (o número de graus de liberdade é definido pelo número de neurônios adicionais da camada intermediária). Se  $\chi_{calc}^2 > \chi_r^2$  rejeitamos a hipótese nula. O teste é feito sequencialmente até que a hipótese  $H_0$  não seja rejeitada.

A tab.(8.2) apresenta o resultado do teste onde avaliamos uma rede com até quatro neurônios na camada intermediária. A estatística de teste calculada para a RNA com quatro neurônios na camada intermediária é menor do que a tabelada,  $\chi_{calc}^2 < \chi_4^2$ . Assim sendo, não possuímos evidências para rejeitar  $H_0$ , com um nível de significância de 95%, e concluímos que a camada intermediária da RNA possui quatro neurônios.

Neurônios adicionais	$R^2$	Tamanho da amostra	$\chi^2 - Calculado$	$\chi^2 - Tabela$
1	0,001230	6.228	7,60	3,84
2	0,001292	6.228	8,05	5,99
3	0,001295	6.228	8,07	7,82
4	0.001475	6.228	9.19	9,48

Tabela 8.2: Resultados do teste LM para o número de neurônios na camada intermediária

O teste LM para determinação do número de neurônios na camada intermediária da RNA não leva em consideração a complexidade do plano amostral. Porém, estimaremos a RNA sob o modelo de superpopulação com quatro neurônios na camada intermediária.

Para fins de comparação estimaremos os seguintes modelos:

Abordagem para ajuste	Modelo	Variável Dependente	
		Renda	ln(Renda)
Clássica	Linear	$M_1$	$M_2$
	RNA	$M_3$	$M_4$
Superpopulação	Linear	$M_5$	$M_6$
	RNA	$M_7$	$M_8$

Na literatura pertinente, o logaritmo natural da renda domiciliar *per capita* (doravante renda) é frequentemente utilizado para modelagem, como visto em WO-OLDRIDGE (2008). Porém, ajustamos modelos com a variável dependente renda e logaritmo natural da renda, pois sendo a RNA um modelo não linear, achamos necessário testar e confrontar os resultados com o modelo linear em ambos os casos.

A amostra de PNAD 2014 foi separada em treino (80%) e teste (20%) mantendo-se a estratificação e a conglomeração por unidade primária de amostragem (UPA), e selecionando aleatoriamente dentro de cada estrato a porcentagem de treino e teste. Desta forma, mantivemos a estrutura do plano amostral complexo para as amostras de treinamento e teste.

Os ajustes dos modelos foram realizados tendo como variáveis dependentes a renda e o logaritmo natural da renda, sob a abordagem clássica e de superpopulação, respectivamente. As variáveis contínuas (renda e idade em anos) são normalizadas para fins de comparação, pois os dados para modelagem da RNA, como descrito anteriormente, são normalizados. Os modelos de RNA ajustados seguem a estrutura descrita na eq.(5.25) com quatro neurônios na camada intermediária, fig.(8.2).

A RNA, fig.(8.2), possui dois limiares um na camada de entrada e outro na camada intermediária representados pelos neurônios unitários. O neurônio da camada intermediária,  $v_{i1}$ , é linear e os neurônios restantes da camada intermediária são tangente hiperbólicos. O treinamento foi realizado, primeiramente, na parte linear da RNA. Fazendo-se  $\beta_{11} = 0$ ,  $B_2 = 1$ ,  $B_1 \neq 0$  e os coeficientes restantes iguais a zero. Após a finalização do treinamento da parte linear da RNA os coeficientes estimados são fixados e, a seguir, iniciou-se o treinamento da parte não linear. O algoritmo utilizado para o treinamento da RNA é o BP. Mais detalhes sobre o procedimento de treinamento vide CALÔBA.

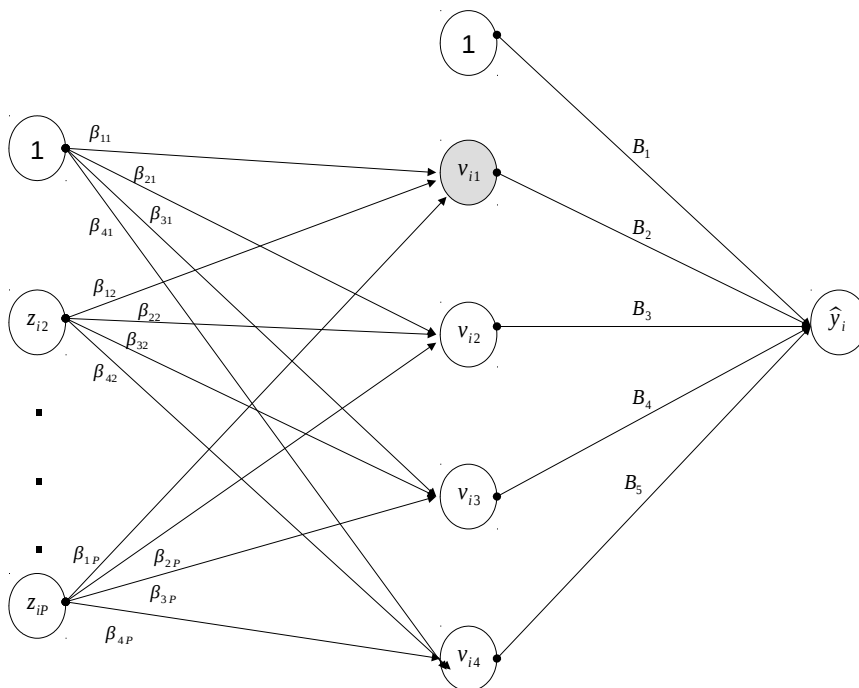


Figura 8.1: Estrutura da RNA para modelagem dos dados da PNAD 2014

No apêndice B temos os resultados do ajuste dos modelos de RNA. A tab.(B.1) apresenta os coeficientes estimados para RNA clássica cuja variável dependente é renda e a tab.(B.2) as estimativas dos coeficientes para variável dependente do logaritmo natural da renda.

A métrica de comparação dos modelos será o coeficiente de correlação múltipla:

$$R^2 = 1 - \frac{\sum_{i \in S} w_i (y_i - \hat{y}_i)^2}{\sum_{i \in S} w_i (y_i - \bar{y}_i)^2}$$

A tabela apresenta o resultado do  $R^2$  (%) para os modelos estimados:

Abordagem para ajuste	Modelo	R-quadrado	
		Renda	ln(Renda)
Clássica	Linear	16,1	16,2
	RNA	15,5	16,0
Superpopulação	Linear	16,8	19,5
	RNA	16,5	19,7

Tabela 8.3: Coeficiente de correlação múltipla,  $R^2$ , em porcentagem, dos modelos estimados

Observando a tab.(8.3), o  $R^2$  do modelo linear clássico para variável dependente renda ficou 3,87% maior do que o modelo de RNA clássico. Utilizando o logaritmo natural da renda o  $R^2$  da RNA clássica aumentou em 3,22% ficando semelhante ao da regressão linear clássica.

A abordagem de superpopulação do modelo linear clássico e da RNA clássica obtiveram resultados do  $R^2$  semelhantes para a variável dependente renda. A modelagem do logaritmo natural da renda ocasionou melhora em ambos os modelos sendo 16,07% para o modelo linear e 19,39% para a RNA. A RNA sob a abordagem de superpopulação apresentou o melhor resultado entre todos os modelos ajustados.



# Capítulo 9

## Considerações Finais e Trabalhos Futuros

### 9.1 Considerações Finais

Os métodos usuais de modelagem em RNA pressupõem que os dados são provenientes de uma amostra aleatória simples. Porém, na prática, em muitas pesquisas são utilizados esquemas amostrais mais complexos. O presente trabalho propôs uma abordagem de superpopulação em RNA com o objetivo de incorporar um esquema amostral complexo na modelagem de redes neurais.

O esquema amostral complexo foi incorporado para ajuste da RNA via equações de atualização, eq.(6.22) e eq.(6.23), como descrito na seção 6.2. A seguir, calculamos as estimativas das variâncias para os estimadores dos parâmetros, as quais são válidas para redes neurais identificáveis. A identificabilidade em redes neurais foi discutida na seção 5.2.

Uma avaliação empírica da abordagem de superpopulação em RNA foi feita utilizando um experimento muito simples. Construímos 1000 populações hipotéticas e selecionamos um conjunto de amostras de cada população, seguindo um esquema amostral complexo. Após a construção da população e seleção das amostras, ajustamos uma RNA identificável com um neurônio tangente hiperbólico na camada intermediária e um neurônio linear na camada de saída sob as abordagens clássica e de superpopulação, respectivamente. As medidas de avaliação de desempenho dos estimadores foram apresentadas na seção 7.5.

Dadas as limitações e simplicidade do experimento observamos, que as estimativas dos coeficientes  $\beta_{11}$  e  $\beta_{12}$ , fig.(7.13(a)) e fig.(7.13(b)) da camada de entrada apresentam menor viés sob a abordagem de superpopulação. Entretanto, as estimativas dos coeficientes  $B_1$  e  $B_2$ , fig.(7.13(c)) e fig.(7.13(d)), apresentam maior viés que o ajuste feito pela RNA clássica. Por outro lado, temos que o ERM dos coeficientes

$\beta_{11}$  e  $\beta_{12}$  da camada de entrada, apresentam uma menor variabilidade relativa sob a abordagem de superpopulação, fig.(7.14(a)) e fig.(7.14(b)), ao contrário os coeficientes,  $B_1$  e  $B_2$ , fig.(7.14(c)) e fig.(7.14(d)), da camada de saída apresentam maior variabilidade relativa sob a abordagem de superpopulação. As variâncias de todos os estimadores decrescem à medida que o tamanho da amostra aumenta, fig.(7.12).

A ilustração foi realizada com os dados da PNAD 2014, onde modelamos o logaritmo natural da renda utilizando o modelo linear e a RNA, ambos, sob a abordagem de superpopulação. Observamos que a RNA sob a abordagem de superpopulação obteve o maior valor do coeficiente de correlação múltipla,  $R^2 = 19,7$ , em relação aos outros modelos ao modelar o logaritmo natural da renda, vide tab.(8.3).

Tais resultados, do experimento e da aplicação prática, indicam que é necessário um trabalho adicional para colher mais evidências em relação ao desempenho da abordagem de superpopulação comparada à clássica.

## 9.2 Trabalhos Futuros

A continuação deste trabalho visa colher mais evidências em relação a modelagem de superpopulação em RNA.

A primeira etapa será a construção de populações com diferentes processos geradores e seleção de conjuntos de amostras com diferentes esquemas amostrais complexos. A fim de ajustarmos uma RNA identificável, fig.(7.7), e avaliar os estimadores utilizando as medidas apresentadas na seção 7.5. A seguir, realizar o mesmo procedimento, porém, utilizando diferentes estruturas de RNA.

A segunda etapa será derivar o teste de White, seção 5.2.2, para amostras complexas e realização de simulações visando a avaliação do teste.

Após a segunda etapa, nos concentraremos na estimação de variâncias dos estimadores da RNA identificável sob a abordagem de superpopulação para diferentes esquemas amostrais complexos.

Finalizando, ao final das etapas anteriores o objetivo é criar um pacote no software estatístico R para ajuste de RNA sob a abordagem de superpopulação e disponibilizá-lo.

# Referências Bibliográficas

- CASELLA, G., BERGER, R. *Statistical Inference*. Duxbury Resource Center, June 2001. ISBN: 0534243126.
- HANSEN, M. H., MADOW, W. G., TEPPING, B. J. “An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys”, *Journal of the American Statistical Association*, v. 78, n. 384, pp. 776:793, 1983.
- PESSOA, D. G. C., DO NASCIMENTO SILVA, P. L. *Análise de Dados Amostrais Complexos*. 1 ed. São Paulo, ABE, 1998.
- SARLE, W. S. “Neural Networks and Statistical Models”. In: *Proceedings of the Nineteenth Annual SAS Users Group International Conference, April, 1994*, p. 1538:1550, Cary, NC, 1994. SAS Institute.
- HAYKIN, S. *Redes Neurais: Princípios e prática*. 2 ed. Porto Alegre, Bookman, 2001.
- SILVA, I. N., SPATTI, D. H., FLAUZINO, R. A. *Redes Neurais Artificiais para engenharia e ciências aplicadas*. 1 ed. São Paulo, Artliber, 2010.
- AMER, S. R. “Neural Network in Complex Survey Design”, *International Journal of Computer Systems Science and Engineering*, v. 3, n. 1, pp. 12:17, fev. 2009.
- BINDER, D. “On the Variances of Asymptotically Normal Estimators from Complex Surveys”, *International Statistical Review*, v. 51, n. 3, pp. 279:292, 1983.
- WHITE, H. “An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks”. In: *Neural Networks, 1989. IJCNN., International Joint Conference on*, p. 451:455 vol.2, 1989a. doi: 10.1109/IJCNN.1989.118281.
- LEE, E. S. *Analyzing complex survey data*. Sage Publications, Inc., 2006.

- DU, K.-L., SWAMY, M. N. S. *Neural Networks in a Softcomputing Framework*. Springer, 2006.
- LOHR, S. L. *Sampling: Design and Analysis*. Duxbury Press, dez. 1999. ISBN: 0534353614.
- LARSON, H. *Introduction to probability theory and statistical inference*. Series in probability and mathematical statistics. New York, John Wiley, 1969. ISBN: 0-471-51780-1.
- BOLFARINE, H., BUSSAB, W. O. *Elementos de Amostragem*. 1 ed. São Paulo, Edgard Blücher, 2005.
- HORVITZ, D. G., THOMPSON, D. J. “A Generalization of Sampling Without Replacement From a Finite Universe”, *Journal of the American Statistical Association*, v. 47, n. 260, pp. 663:685, 1952. ISSN: 01621459.
- CHAMBERS, R., SKINNER, C. *Analysis of survey data*. John Wiley and Sons, 2003.
- WOOLDRIDGE, J. *Introductory Econometrics: A Modern Approach (with Economic Applications, Data Sets, Student Solutions Manual Printed Access Card)*. South-Western College Pub, 2008. ISBN: 0324581629.
- NATHAN, G., HOLT, D. “The Effect of Survey Design on Regression Analysis”, *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 42, n. 3, pp. 377:386, 1980. ISSN: 00359246.
- LECUN, Y., BOTTOU, L., ORR, G., et al. “Efficient BackProp”. In: Orr, G., K., M. (Eds.), *Neural Networks: Tricks of the trade*. Springer, 1998.
- MEDEIROS, M. C., TERASVIRTA, T. “Statistical Methods for Modelling Neural Networks”, *Engineering Intelligent Systems*, v. 9, pp. 227:235, 2001.
- HWANG, J. T. G., DING, A. A. “Prediction Intervals for Artificial Neural Networks”, *Journal of the American Statistical Association*, v. 92, n. 438, pp. 748:757, 1997. ISSN: 01621459.
- PAULINO, C. D. M., BRAGANÇA PEREIRA, C. A. “On identifiability of parametric statistical models”, *Journal of the Italian Statistical Society*, v. 3, n. 1, pp. 125:151. ISSN: 1613-981X.
- WHITE, H. “Learning in Artificial Neural Networks: A Statistical Perspective”, *Neural Computation*, v. 1, pp. 425:464, 1989b.

- ROJAS, R. *Second Order Backpropagation: Efficient Computation of the Hessian Matrix for Neural Networks*. International Computer Science Institute, 1993.
- ISAKI, C. T., FULLER, W. A. “Survey Design under the Regression Superpopulation Model”, *Journal of the American Statistical Association*, , n. 377, pp. 89:96. doi: 10.1080/01621459.1982.10477770.
- DUNNE, R. A. *A Statistical Approach to Neural Networks for Pattern Recognition (Wiley Series in Computational Statistics)*. Wiley-Interscience, 2007. ISBN: 0471741086.
- WOLTER, K. M. *Introduction to Variance Estimation*. Springer, 2006. ISBN: 978-0-387-26127-0.
- IBGE. *Mapa de pobreza e desigualdade: municípios brasileiros*. Relatório técnico, Instituto Brasileiro de Geografia e Estatística, 2008.
- CALÔBA, L. P. “Redes Neurais em Modelagem de Sistemas”. In: *Enciclopédia de Automática*, p. 325:344.

# Apêndice A

## Dicionário de variáveis

### A.1 Pesquisa Nacional por Amostra de Domicílios - PNAD - ano 2014

Conforme IBGE (2008), foram adotados os seguintes conceitos:

- Domicílio: Conceituou-se como local de moradia estruturalmente separado e independente, constituído por um ou mais cômodos. A separação fica caracterizada quando o local de moradia é limitado por paredes, muros, cercas, coberto por um teto, e permite que seus moradores se isolem, arcando com parte ou todas as suas despesas de alimentação ou moradia. A independência fica caracterizada quando o local de moradia tem acesso direto, permitindo que seus moradores possam entrar e sair sem passar por local de moradia de outras pessoas.
- População Residente: foi composta pelos moradores presentes e ausentes, ou seja, pelas pessoas que tinham a unidade domiciliar (domicílio particular ou unidade de habitação em domicílio coletivo) como local de residência habitual e, na data da entrevista, estavam presentes ou ausentes, temporariamente, por período não superior a 12 meses em relação àquela data.
- Data de referência: dia 27 de setembro de 2014.
- Semana de referência: 21 a 27 de setembro de 2014.

### A.2 Variáveis Chave: Arquivos de Domicílio e de Pessoas

- UF - Unidade da Federação (2 dígitos): Para esse banco de dados foi selecionado o código 31, que se refere a Minas Gerais. Variável lida como numérica.

- V0102 e V0103 : são códigos numéricos para organização do banco de dados e identificam o setor censitário e o domicílio, quando combinados.

## A.3 Arquivo original: Domicílio

Foram selecionados os registros de domicílios particulares permanentes.

PD\_PESO - Peso Domiciliar (5 dígitos): Variável de expansão da amostra.

Unidades da Federação e Regiões Metropolitanas	Fração de Amostragem	Composição da amostra			
		Municípios	Setores	Unidades domiciliares	Pessoas
<b>Brasil</b>		1 100	9 166	151 291	362 627
<b>Minas Gerais</b>	1/650	129	813	13 940	33 384
<b>Região Metropolitana de Belo Horizonte</b>	1/400	35	297	5 242	12 805

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa Nacional por Amostra de Domicílios 2014. Nota: A composição da amostra da Unidade da Federação inclui a Região Metropolitana e as áreas urbanas e rurais.

Tabela A.1: Brasil, Unidade da Federação e Região Metropolitana de Minas Gerais, fração de amostragem, composição da amostra por municípios, setores, unidades domiciliares e pessoas. PNAD - 2014

PD\_TDOM - Tipo de Domicílio (1 dígito):

Classificação da informação:

2 - Casa - Para o domicílio que: ocupasse totalmente um prédio, de um ou mais pavimentos, ou dois ou mais prédios, de um ou mais pavimentos, localizados no mesmo terreno; ou ocupasse parte de um prédio, de um pavimento, que não tivesse espaços comuns (tais como: vestíbulo, escada, corredor, portaria e outras dependências) para servir aos domicílios particulares permanentes ali existentes. Assim, também, foi considerado o domicílio situado em prédio de, no máximo, três pavimentos em que as demais unidades existentes não fossem domicílios particulares permanentes;

4 - Apartamento - Para o domicílio situado em prédio de: um ou mais pavimentos, com mais de um domicílio particular permanente, servidos por espaços comuns (vestíbulo, escada, corredor, portaria e outras dependências); dois ou mais pavimentos, com mais de um domicílio particular permanente, e com entradas independentes para os andares; ou mais de três pavimentos, em que as demais unidades fossem não residenciais; ou

6 - Cômodo - Para o domicílio que ocupasse um ou mais cômodos de uma casa de cômodos, cortiço, cabeça de porco etc.

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
Válido	2	9 640	86,0	86,0	86,0
	4	1 547	13,8	13,8	99,9
	6	16	0,1	1,0	100
	Total	11 203	100	100	

Tabela A.2: Tipo de domicílio

V0205 - Total de cômodos no domicílio (2 dígitos): Considerou-se como cômodo todo compartimento, coberto por um teto e limitado por paredes, que fosse parte integrante do domicílio particular permanente, com exceção de corredor, alpendre, varanda aberta, garagem, depósito e outros compartimentos utilizados para fins não residenciais.

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa	
Válido	1	11	,1	,1	,1	
	2	85	,8	,8	,9	
	3	357	3,2	3,2	4,0	
	4	1 006	9,0	9,0	13,0	
	5	3 055	27,3	27,3	40,3	
	6	2 411	21,5	21,5	61,8	
	7	1 840	16,4	16,4	78,2	
	8	1 075	9,6	9,6	87,8	
	9	577	5,2	5,2	93,0	
	10	334	3,0	3,0	96,0	
	11	168	1,5	1,5	97,5	
	12	106	,9	,9	98,4	
	13	62	,6	,6	99,0	
	14	46	,4	,4	99,4	
	15	23	,2	,2	99,6	
	16	14	,1	,1	99,7	
	17	7	,1	,1	99,8	
	18	6	,1	,1	99,8	
	19	4	,0	,0	99,9	
	20	6	,1	,1	99,9	
	21	1	,0	,0	99,9	
	22	4	,0	,0	100,0	
	23	1	,0	,0	100,0	
	24	1	,0	,0	100,0	
	25	1	,0	,0	100,0	
	30	2	,0	,0	100,0	
		<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.3: Total de cômodos no domicílio



V0206 - Total de cômodos no domicílio servindo de dormitório (2 dígitos): Considerou-se como dormitório o cômodo que estivesse, em caráter permanente, sendo utilizado para esta finalidade por morador do domicílio particular permanente.

		<b>Frequência</b>	<b>Porcentagem</b>	<b>Porcentagem Válida</b>	<b>Porcentagem Cumulativa</b>
<b>Válido</b>	1	4 060	36,2	36,2	36,2
	2	4 553	40,6	40,6	76,9
	3	2 273	20,3	20,3	97,2
	4	276	2,5	2,5	99,6
	5	35	,3	,3	99,9
	6	4	,0	,0	100,0
	7	2	,0	,0	100,0
	<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.4: Total de cômodos no domicílio servindo de dormitório

V0207 - Condição de ocupação do domicílio (1 dígito): Classificação da informação:

1 ou 2 - Próprio - Para o domicílio de propriedade, total ou parcial, de morador, estando integralmente quitado (1) ou não (2), e independentemente da condição de ocupação do terreno;

3 - Alugado - Para o domicílio cujo aluguel fosse, totalmente ou parcialmente, pago por morador;

4 ou 5 - Cedido - Para o domicílio cedido gratuitamente por empregador (4) de morador, instituição ou pessoa não moradora (parente ou não) (5), ainda que mediante uma taxa de ocupação ou conservação. Nesta condição, incluiu-se o domicílio cujo aluguel fosse integralmente pago, diretamente ou indiretamente, por empregador de morador, instituição ou pessoa não moradora; ou

6 - Outra - Para o domicílio ocupado em condição diferente das anteriormente arroladas, como, por exemplo, no caso de invasão.

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
<b>Válido</b>	1	7 591	67,8	67,8	67,8
	2	617	5,5	5,5	73,3
	3	2 020	18,0	18,0	91,3
	4	268	2,4	2,4	93,7
	5	645	5,8	5,8	99,4
	6	62	0,6	0,6	100
	<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.5: Condição de ocupação do domicílio

Banheiro ou sanitário Investigou-se a existência de banheiro ou sanitário, para uso dos moradores, no domicílio particular permanente ou no terreno, ou na propriedade em que estava situado. Pesquisou-se, também, se o banheiro ou sanitário era de uso exclusivo ou comum dos moradores de mais de um domicílio particular permanente. Considerou-se como banheiro o cômodo destinado a banho e que também dispusesse de vaso sanitário ou buraco para dejeções. Considerou-se como sanitário o cômodo ou o local limitado por paredes de qualquer material, coberto, ou não, por um teto e que dispusesse de vaso sanitário ou buraco para dejeções.

V2016 - Número de banheiros ou sanitários (2 dígitos): informado um número entre 1 a 15. Os casos omissos ocorrem por ter respondido à pergunta se “tinha ou não banheiro ou sanitário no domicílio ou propriedade” como não.

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
<b>Válido</b>	1	7 897	70,5	70,5	71,2
	2	2 356	21,0	21,3	92,5
	3	609	5,4	5,5	98,8
	4	155	1,4	1,4	99,4
	5	45	0,4	0,4	99,8
	6	16	0,1	0,1	99,9
	7	5	0,0	0,0	100
	8	3	0,0	0,0	100
	Total	11 086	99,0	<b>100</b>	
<b>Omisso</b>	Sistema	117	1,0		
<b>Total</b>		<b>11 203</b>	<b>100</b>		

Tabela A.6: Número de banheiros ou sanitários no domicílio

V0217 - Esgotamento sanitário (1 dígito): O escoadouro do banheiro ou sanitário de uso dos moradores dos domicílios particulares permanentes foi classificado, quanto ao tipo, em:

Classificação da informação:

1 - Rede coletora (de esgoto ou pluvial) - Quando a canalização das águas servidas e dos dejetos estivesse ligada a um sistema de coleta que os conduzisse para um desaguadouro geral da área, região ou município, mesmo que o sistema não dispusesse de estação de tratamento da matéria esgotada;

2 - Fossa séptica ligada à rede coletora de esgoto ou pluvial - Quando as águas servidas e os dejetos fossem esgotados para uma fossa, onde passavam por um processo de tratamento ou decantação, sendo a parte líquida canalizada para um desaguadouro geral da área, região ou município;

3 - Fossa séptica não ligada à rede coletora de esgoto ou pluvial - Quando as águas servidas e os dejetos fossem esgotados para uma fossa, onde passavam por um processo de tratamento ou decantação, sendo a parte líquida absorvida no próprio terreno;

4 - Fossa rudimentar - Quando os dejetos fossem esgotados para uma fossa rudimentar (fossa negra, poço, buraco etc.);

5 - Vala - Quando os dejetos fossem esgotados diretamente para uma vala;

6 - Direto para o rio, lago ou mar - Quando os dejetos fossem esgotados diretamente para uma vala, rio, lago ou mar; ou

7 - Outro - Quando os dejetos fossem esgotados para um escoadouro que não se enquadrasse em quaisquer dos tipos descritos anteriormente.

		<b>Frequência</b>	<b>Porcentagem</b>	<b>Porcentagem Válida</b>	<b>Porcentagem Cumulativa</b>
<b>Válido</b>	1	8 967	80,0	80,6	80,6
	2	77	0,7	0,7	81,3
	3	295	2,6	2,7	83,9
	4	1 314	11,7	11,8	95,7
	5	69	0,6	0,6	96,4
	6	397	3,5	3,6	99,9
	7	7	0,1	0,1	100
	Total	11 126	99,3	<b>100</b>	
<b>Omisso</b>	Sistema	77	0,7		
<b>Total</b>		<b>11 203</b>	<b>100</b>		

Tabela A.7: Esgotamento sanitário

## A.4 Variáveis Dicotômicas

Categorias: 1 = SIM; 0 = NÃO; 9=OMISSOS.

V0215 - Existência de banheiro ou sanitário (1 dígito): verifica a existência de banheiro ou sanitário, para uso dos moradores, no domicílio particular permanente ou no terreno, ou na propriedade em que estava situado.

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
<b>Válido</b>	0	77	0,7	0,7	0,7
	1	11 126	99,3	99,3	<b>100</b>
	<b>Total</b>	<b>11 203</b>	<b>100</b>		

Tabela A.8: Existência de banheiro ou sanitário no domicílio

V0225 - Existência de rádio (1 dígito): Nos domicílios particulares permanentes, pesquisou-se a existência de rádio, mesmo que fizesse parte de conjunto que aco-  
plasse outros aparelhos, tais como: radiogravador, rádio toca-fitas etc. Incluíram-se,  
também, os aparelhos de mp3 e mp4 com rádio.

		<b>Frequência</b>	<b>Porcentagem</b>	<b>Porcentagem Válida</b>	<b>Porcentagem Cumulativa</b>
<b>Válido</b>	0	2 399	21,4	21,4	21,4
	1	8 804	78,6	78,6	<b>100</b>
	<b>Total</b>	<b>11 203</b>	<b>100</b>		

Tabela A.9: Existência de rádio no domicílio

TV - Existência de televisão (1 dígito): Nos domicílios particulares permanentes,  
pesquisou-se a existência de televisão em cores e, para os que não tinham este tipo  
de aparelho, a existência de televisão em preto e branco.

		<b>Frequência</b>	<b>Porcentagem</b>	<b>Porcentagem Válida</b>	<b>Porcentagem Cumulativa</b>
<b>Válido</b>	0	247	2,2	2,2	2,2
	1	10 956	97,8	97,8	<b>100</b>
	<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.10: Existência de televisão no domicílio

V0230 - Existência de máquina de lavar (1 dígito): Nos domicílios particulares  
permanentes, pesquisou-se a existência de máquina de lavar roupa (aparelho que  
desenvolve, de forma automática, todas as etapas da lavagem de roupa, desde a  
entrada de água na máquina, passando pelos processos de agitação e enxágue, até  
o de centrifugação).

		<b>Frequência</b>	<b>Porcentagem</b>	<b>Porcentagem Válida</b>	<b>Porcentagem Cumulativa</b>
<b>Válido</b>	0	4 860	43,4	43,4	43,4
	1	6 343	56,6	56,6	<b>100</b>
	<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.11: Existência de máquina de lavar no domicílio

V0228 - Existência de geladeira ou freezer (1 dígito): Nos domicílios particulares  
permanentes, pesquisou-se a existência de geladeira de duas portas (ou seja, o apa-  
relho que acopla dois compartimentos independentes, sendo um de refrigeração e o  
outro de congelamento de alimentos) e, para os que não tivessem este tipo de apa-  
relho, a existência de geladeira de uma porta. Também, pesquisou-se a existência  
de freezer.

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
Válido	0	155	1,4	1,4	1,4
	1	11 048	98,6	98,6	100
	<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.12: Existência de geladeira ou freezer no domicílio

V0231 - Existência de computador (1 dígito): Nos domicílios particulares permanentes, pesquisou-se a existência de microcomputador, inclusive portátil.

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
Válido	0	5 412	48,3	48,3	48,3
	1	5 791	51,7	51,7	100
	<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.13: Existência de computador no domicílio

V2032 - Existência de automóvel (1 dígito): Nos domicílios particulares permanentes, pesquisou-se a existência de carro ou motocicleta de uso pessoal.

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
Válido	0	4 490	40,1	40,1	40,1
	1	6 713	59,9	59,9	100
	<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.14: Existência de automóvel no domicílio

## A.5 Arquivo original: Pessoas

A população residente foi composta pelos moradores presentes e ausentes, ou seja, pelas pessoas que tinham a unidade domiciliar (domicílio particular ou unidade de habitação em domicílio coletivo) como local de residência habitual e, na data da entrevista, estavam presentes ou ausentes, temporariamente, por período não superior a 12 meses em relação àquela data. Excluíram-se da pesquisa as pessoas residentes em embaixadas, consulados e legações e, também, as pessoas institucionalizadas residentes em domicílios coletivos de estabelecimentos institucionais, tais como: os militares em caserna ou dependências de instalações militares; os presos em penitenciárias; os internos em escolas, orfanatos, asilos, hospitais, etc.; e os religiosos em conventos, mosteiros etc. No entanto, para esse banco de dados foram considerados apenas os registros dos residentes em domicílios particulares permanentes. O

banco de dados foi montado em relação aos dados do responsável pelo domicílio. Os registros das demais pessoas moradoras do mesmo domicílio foram utilizadas para identificar o cônjuge (CONJU) e o número de moradores (PP\_NMOR). A variável utilizada para essa identificação foi a Condição na família (V0401). Essa variável indica que dentro de cada família, as pessoas foram classificadas em função da relação com a pessoa de referência ou com o seu cônjuge, de acordo com as seguintes definições:

1 - Pessoa de referência - Pessoa responsável pela família ou que assim fosse considerada pelos demais membros da família;

2 - Cônjuge - Pessoa que vivia conjugalmente com a pessoa de referência da família, existindo ou não o vínculo matrimonial;

3 - Filho - Pessoa que era filho, enteado, filho adotivo ou de criação da pessoa de referência da família ou do seu cônjuge;

4 - Outro parente - Pessoa que tinha qualquer outro grau de parentesco com a pessoa de referência da família ou com o seu cônjuge;

5 - Agregado - Pessoa que não era parente da pessoa de referência da família nem do seu cônjuge e não pagava hospedagem nem alimentação a membro da família;

6 - Pensionista - Pessoa que não era parente da pessoa de referência da família nem do seu cônjuge e pagava hospedagem ou alimentação a membro da família;

7 - Empregado doméstico - Pessoa que prestava serviço doméstico remunerado em dinheiro ou somente em benefícios a membro (s) da família; ou

8 - Parente do empregado doméstico - Pessoa que era parente do empregado doméstico e não prestava serviço doméstico remunerado a membro (s) da família. Para efeito de divulgação, os agregados, os pensionistas, os empregados domésticos e os parentes dos empregados domésticos constituíram o grupo denominado “sem parentesco”.

V4111 - Indica se há a existência de cônjuge no domicílio (1 dígito): Melhor explicado no texto referente a variável da PNAD V0401. Classificação da informação:

1 - Sim

. - Não

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
<b>Válido</b>	1	7 001	62,5	<b>100,0</b>	<b>100,0</b>
<b>Omisso</b>	Sistema	4 202	37,5		
<b>Total</b>		<b>11 203</b>	<b>100</b>		

Tabela A.15: Existência de cônjuge no domicílio

V0302 - Sexo do Responsável pelo domicílio (1 dígito): Classificação da informação:



1 - Masculino

2 - Feminino

		<b>Frequência</b>	<b>Porcentagem</b>	<b>Porcentagem Válida</b>	<b>Porcentagem Cumulativa</b>
<b>Válido</b>	1	7 113	63,5	63,5	63,5
	2	4 090	36,5	36,5	100,0
	<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.16: Sexo do Responsável pelo domicílio

V8005 - Idade do responsável pelo domicílio (3 dígitos): A investigação da idade foi feita por meio da pesquisa do dia, mês e ano de nascimento da pessoa ou da idade presumida da pessoa que não soubesse a data de nascimento. A idade foi calculada em relação à data de referência da pesquisa. Os registros das demais pessoas moradoras do mesmo domicílio foram utilizadas para identificar o número de moradores menores de 7 anos (cri\_ate7), de 7 a 14 anos (cri\_7a14) e acima de 60 anos (idoso).

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
Válido	11,00	1	,0	,0	,0
	16,00	1	,0	,0	,0
	17,00	10	,1	,1	,1
	18,00	22	,2	,2	,3
	19,00	22	,2	,2	,5
	20,00	52	,5	,5	1,0
	21,00	46	,4	,4	1,4
	22,00	70	,6	,6	2,0
	23,00	84	,7	,7	2,7
	24,00	86	,8	,8	3,5
	25,00	127	1,1	1,1	4,7
	26,00	137	1,2	1,2	5,9
	27,00	154	1,4	1,4	7,2
	28,00	164	1,5	1,5	8,7
	29,00	176	1,6	1,6	10,3
	30,00	186	1,7	1,7	11,9
	31,00	186	1,7	1,7	13,6
	32,00	206	1,8	1,8	15,4
	33,00	222	2,0	2,0	17,4
	34,00	236	2,1	2,1	19,5
	35,00	244	2,2	2,2	21,7
	36,00	240	2,1	2,1	23,9
	37,00	252	2,2	2,2	26,1
	38,00	247	2,2	2,2	28,3
	39,00	218	1,9	1,9	30,3
	40,00	252	2,2	2,2	32,5
	41,00	221	2,0	2,0	34,5
	42,00	298	2,7	2,7	37,1
	43,00	223	2,0	2,0	39,1
	44,00	251	2,2	2,2	41,4
	45,00	235	2,1	2,1	43,5
	46,00	243	2,2	2,2	45,6
47,00	218	1,9	1,9	47,6	
48,00	245	2,2	2,2	49,8	
49,00	247	2,2	2,2	52,0	
50,00	249	2,2	2,2	54,2	
51,00	251	2,2	2,2	56,4	
52,00	245	2,2	2,2	58,6	
53,00	231	2,1	2,1	60,7	
54,00	209	1,9	1,9	62,5	
55,00	195	1,7	1,7	64,3	

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
	56,00	270	2,4	2,4	66,7
	57,00	221	2,0	2,0	68,7
	58,00	215	1,9	1,9	70,6
	59,00	205	1,8	1,8	72,4
	60,00	194	1,7	1,7	74,1
	61,00	178	1,6	1,6	75,7
	62,00	181	1,6	1,6	77,4
	63,00	196	1,7	1,7	79,1
	64,00	190	1,7	1,7	80,8
	65,00	165	1,5	1,5	82,3
	66,00	179	1,6	1,6	83,9
	67,00	133	1,2	1,2	85,1
	68,00	134	1,2	1,2	86,3
	69,00	123	1,1	1,1	87,4
	70,00	138	1,2	1,2	88,6
	71,00	105	,9	,9	89,5
	72,00	112	1,0	1,0	90,5
	73,00	93	,8	,8	91,4
	74,00	119	1,1	1,1	92,4
	75,00	102	,9	,9	93,3
Válido	76,00	93	,8	,8	94,2
	77,00	67	,6	,6	94,8
	78,00	84	,7	,7	95,5
	79,00	76	,7	,7	96,2
	80,00	61	,5	,5	96,7
	81,00	54	,5	,5	97,2
	82,00	51	,5	,5	97,7
	83,00	45	,4	,4	98,1
	84,00	43	,4	,4	98,4
	85,00	33	,3	,3	98,7
	86,00	18	,2	,2	98,9
	87,00	33	,3	,3	99,2
	88,00	13	,1	,1	99,3
	89,00	18	,2	,2	99,5
	90,00	13	,1	,1	99,6
	91,00	14	,1	,1	99,7
	92,00	6	,1	,1	99,8
	93,00	6	,1	,1	99,8
	94,00	5	,0	,0	99,9
	95,00	6	,1	,1	99,9
	97,00	1	,0	,0	99,9
	98,00	3	,0	,0	100,0
	100	1	,0	,0	100,0
	101,00	3	,0	,0	100,0
	107,00	1	,0	,0	100,0
	<b>Total</b>	<b>11 203</b>	<b>100,0</b>	<b>100,0</b>	

Tabela A.17: Idade do responsável pelo domicílio

Os registros de pessoas moradoras do mesmo domicílio foram utilizadas para identificar o número de moradores, utilizando a variável de idade da pessoa do registro (V8005). As variáveis derivadas criadas foram; menores de 7 anos (cri\_ate7), de 7 a 14 anos (cri\_7a14) e acima de 60 anos (idoso).

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
<b>Válido</b>	1,00	1 907	17,0	80,2	80,2
	2,00	410	3,7	17,2	97,4
	3,00	54	0,5	2,3	99,7
	4,00	4	0,0	0,2	99,9
	5,00	3	0,0	0,1	100
	<b>Total</b>		2 378	21,2	100
<b>Omisso</b>	<b>Sistema</b>	8 825	78,8		
<b>Total</b>		11 203	100		

Tabela A.18: Número de crianças de até sete anos presentes no domicílio

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
<b>Válido</b>	1,00	2 116	18,9	73,8	73,8
	2,00	602	5,4	21,0	94,8
	3,00	119	1,1	4,2	99,0
	4,00	26	0,2	0,9	99,9
	5,00	4	0,0	0,1	100
	<b>Total</b>		2 867	25,6	100
<b>Omisso</b>	<b>Sistema</b>	8 336	74,4		
<b>Total</b>		11 203	100		

Tabela A.19: Número de crianças de 7 a 12 anos presentes no domicílio

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa	
<b>Válido</b>	1,00	2 218	19,8	67,8	67,8	
	2,00	1 103	9,0	31,0	98,8	
	3,00	37	0,3	1,1	99,9	
	4,00	3	0,0	0,1	100	
	<b>Total</b>		3 271	29,2	100	
	<b>Omisso</b>	<b>Sistema</b>	7 932	70,8		
<b>Total</b>		11 203	100			

Tabela A.20: Número de idosos acima de 60 anos no domicílio

V0404 - Cor ou raça do responsável pelo domicílio (1 dígito): Consideraram-se cinco categorias para a pessoa se classificar quanto à característica cor ou raça: branca, preta, amarela (compreendendo-se nesta categoria a pessoa que se declarou de origem japonesa, chinesa, coreana etc.), parda (incluindo-se nesta categoria a pessoa que se declarou mulata, cabocla, cafuza, mameluca ou mestiça de preto com pessoa de outra cor ou raça), ou indígena (considerando-se nesta categoria a pessoa que se declarou indígena ou índia). Classificação da informação:

- 2 - Branca
- 4 - Preta
- 6 - Amarela
- 8 - Parda

0 - Indígena

9 - Sem declaração

V0404 - raça

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
<b>Válido</b>	0,00	35	0,3	0,3	0,3
	2,00	4 663	41,6	41,6	41,9
	4,00	1 329	11,9	11,9	53,8
	6,00	31	0,3	0,3	54,1
	8,00	5 144	45,9	45,9	100
	9,00	1	0,0	0,0	100
	<b>Total</b>	<b>11 203</b>	<b>100</b>	<b>100</b>	

Tabela A.21: Cor ou raça do responsável pelo domicílio

Estudante - Foi definida como a pessoa que frequentava curso de ensino regular (fundamental, médio, primeiro grau, segundo grau ou superior de graduação), de mestrado ou doutorado, pré-escolar, alfabetização de jovens e adultos, educação de jovens e adultos ou supletivo ministrado em escola, ou pré-vestibular. A pessoa que frequentava somente curso de especialização profissional, de extensão cultural (idioma, costura, datilografia etc.) ou de educação de jovens e adultos ou supletivo por meio de rádio, televisão ou correspondência não foi classificada como estudante.

V0602 - Frequência à escola ou creche do responsável pelo domicílio (1 dígito): Foi pesquisado se a pessoa era estudante, ou seja, se frequentava curso de ensino regular (fundamental, médio, primeiro grau, segundo grau ou superior de graduação), de mestrado ou doutorado, pré-escolar, alfabetização de jovens e adultos, educação de jovens e adultos ou supletivo ministrado em escola, ou pré-vestibular, ou se frequentava creche. Para a pessoa que não era estudante e nem frequentava creche, foi investigado se já havia frequentado escola (curso de ensino regular, de mestrado ou doutorado, pré-escolar, alfabetização de jovens e adultos, educação de jovens e adultos ou supletivo) ou creche. Para a pessoa que era estudante, foram pesquisados a série e o nível ou grau do ensino do curso que frequentava. Para a pessoa que não era estudante, mas já havia frequentado escola, foram investigados o nível ou grau do ensino do curso mais elevado que frequentou, a última série concluída e se o curso foi concluído. O sistema de ensino regular atualmente em vigor compreende: o ensino fundamental, o médio e o superior de graduação. O sistema de ensino regular anterior, mas que ainda pode ser encontrado em vigor, compreende: o primeiro grau, o segundo grau e o terceiro grau ou superior. O sistema de ensino regular anterior a estes dois compreendia: o elementar, o médio primeiro ciclo, o médio segundo ciclo e o superior. Considerou-se como creche o estabelecimento, juridicamente regulamentado ou não, destinado a dar assistência a crianças nas primeiras idades. Classificação da informação:

- 1 - Sim
- 0 - Não

	Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
<b>Válido</b>	0,00	10 862	97,0	97,0
	1,00	341	3,0	100
	<b>Total</b>	11 203	100	100

Tabela A.22: Número total de estudantes presentes no domicílio

V4706 - Posição na ocupação do responsável pelo domicílio (1 dígito): Foram definidas oito categorias de posição na ocupação: Classificação da informação:

1 - Empregado - Pessoa que trabalhava para um empregador (pessoa física ou jurídica), geralmente obrigando-se ao cumprimento de uma jornada de trabalho e recebendo em contrapartida uma remuneração em dinheiro, mercadorias, produtos ou benefícios (moradia, comida, roupa, etc.). Nesta categoria, incluiu-se a pessoa que prestava o serviço militar obrigatório e, também, o sacerdote, ministro de igreja, pastor, rabino, frade, freira e outros clérigos;

2 - Trabalhador doméstico - Pessoa que trabalhava prestando serviço doméstico remunerado em dinheiro ou benefícios, em uma ou mais unidades domiciliares;

3 - Conta própria - Pessoa que trabalhava explorando o seu próprio empreendimento, sozinha ou com sócio, sem ter empregado e contando, ou não, com a ajuda de trabalhador não remunerado;

4 - Empregador - Pessoa que trabalhava explorando o seu próprio empreendimento, com pelo menos um empregado;

5 - Trabalhador não remunerado membro da unidade domiciliar - Pessoa que trabalhava sem remuneração, durante pelo menos uma hora na semana, em ajuda a membro da unidade domiciliar que era: empregado na produção de bens primários (que compreende as atividades da agricultura, silvicultura, pecuária, extração vegetal ou mineral, caça, pesca e piscicultura), conta própria ou empregador;

6 - Outro trabalhador não remunerado - Pessoa que trabalhava sem remuneração, durante pelo menos uma hora na semana, como aprendiz ou estagiário ou em ajuda à instituição religiosa, beneficente ou de cooperativismo;

7 - Trabalhador na produção para o próprio consumo - Pessoa que trabalhava, durante pelo menos uma hora na semana, na produção de bens do ramo que compreende as atividades da agricultura, silvicultura, pecuária, extração vegetal, pesca e piscicultura, para a própria alimentação de pelo menos um membro da unidade domiciliar; e

8 - Trabalhador na construção para o próprio uso - Pessoa que trabalhava, durante pelo menos uma hora na semana, na construção de edificações, estradas privadas, poços e outras benfeitorias (exceto as obras destinadas unicamente à reforma) para o próprio uso de pelo menos um membro da unidade domiciliar. Para efeito de divulgação, em todas as tabelas que apresentam a classificação por posição na ocupação, as categorias trabalhador não remunerado membro da unidade domiciliar e outro trabalhador não remunerado foram reunidas em uma única, que recebeu a denominação de não remunerado. No banco de dados da PNAD 2014 quatro variáveis, que se adequam a esse estudo, foram chamadas de Posição na Ocupação. Sendo assim, foram lidas e abaixo encontra-se a frequência simples destas, a fim de auxiliar a escolha da variável de trabalho.

Família - Considerou-se como família o conjunto de pessoas ligadas por laços de parentesco, dependência doméstica ou normas de convivência, que residissem na mesma unidade domiciliar e, também, a pessoa que morasse só em uma unidade domiciliar. Entendeu-se por dependência doméstica a relação estabelecida entre a pessoa de referência e os empregados domésticos e agregados da família, e por normas de convivência as regras estabelecidas para o convívio de pessoas que morassem juntas sem estarem ligadas por laços de parentesco ou dependência doméstica. Definiram-se como famílias conviventes aquelas constituídas por, no mínimo, duas pessoas cada uma, que residissem na mesma unidade domiciliar.

tipofam - Tipo de Família do responsável pelo domicílio (1 dígito): Classificação da informação:

- 01 - Casal sem filhos
- 02 - Casal com todos os filhos menores de 14 anos
- 03 - Casal com todos os filhos de 14 anos ou mais
- 04 - Casal com filhos menores de 14 anos e de 14 anos ou mais
- 06 - Mãe com todos os filhos menores de 14 anos
- 07 - Mãe com todos os filhos de 14 anos ou mais
- 08 - Mãe com filhos menores de 14 anos e de 14 anos ou mais
- 10 - Outros tipos de família

		Frequência	Porcentagem	Porcentagem Válida	Porcentagem Cumulativa
Válido	1	2114	18,9	18,9	18,9
	2	2027	18,1	18,1	37,0
	3	2013	18,0	18,0	54,9
	4	847	7,6	7,6	62,5
	6	278	2,5	2,5	65,0
	7	1227	11,0	11,0	75,9
	8	213	1,9	1,9	77,8
	10	2484	22,2	22,2	100,0
	Total	11203	100,0	100,0	

Tabela A.23: Posição na ocupação do responsável pelo domicílio



# Apêndice B

## Resultados dos Modelos

### B.1 Modelo linear clássico - variável dependente renda

```
lm(formula = normalize((renda)) ~ n.comodos.1 + n.comodos.2 +  
n.dormitorios.1 + n.dormitorios.2 + n.dormitorios.3 + n.banheiros.1 +  
n.banheiros.2 + normalize(idade.anos) + cri_ate7 + cri_7a14 +  
n.idoso + tipofam.1 + tipofam.2 + sexo + raca.1 + raca.2 +  
raca.3 + posse.radio + posse.tv + posse.maqlavar + posse.geladeira +  
posse.computador + posse.auto + vive.conjugue + ocup.dom.1 +  
ocup.dom.2 + tipo.esgoto + freq.escola + pos.ocup.1 + pos.ocup.2 +  
pos.ocup.3 + pos.ocup.4 + pos.ocup.5 + pos.ocup.6 - 1, data = base.pnad.treino)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.288	-0.150	-0.025	0.094	69.737

Coefficients:

Estimate	Std. Error	t value	Pr(> t )
n.comodos.1	-0.201144	0.030745	-6.542 6.54e-11 ***
n.comodos.2	-0.166627	0.022704	-7.339 2.43e-13 ***
n.dormitorios.1	0.213765	0.042707	5.005 5.73e-07 ***
n.dormitorios.2	0.110157	0.039958	2.757 0.005854 **
n.dormitorios.3	0.060314	0.040424	1.492 0.135742
n.banheiros.1	-0.060217	0.112645	-0.535 0.592967
n.banheiros.2	-0.048278	0.016837	-2.867 0.004153 **
normalize(idade.anos)	0.028391	0.018173	1.562 0.118288
cri_ate7	-0.012893	0.016347	-0.789 0.430308

cri_7a14	-0.019393	0.015380	-1.261	0.207398
n.idoso	0.046589	0.021460	2.171	0.029971 *
tipofam.1	-0.178352	0.139986	-1.274	0.202684
tipofam.2	-0.158631	0.139517	-1.137	0.255583
sexo	0.033671	0.016424	2.050	0.040394 *
raca.1	0.025377	0.013531	1.875	0.060771 .
raca.2	-0.001788	0.019960	-0.090	0.928628
raca.3	-0.025075	0.103791	-0.242	0.809108
posse.radio	0.003923	0.015220	0.258	0.796630
posse.tv	0.032152	0.047890	0.671	0.502011
posse.maqlavar	0.027470	0.014635	1.877	0.060572 .
posse.geladeira	0.025837	0.060495	0.427	0.669334
posse.computador	0.057509	0.014996	3.835	0.000127 ***
posse.auto	-0.053207	0.015016	-3.543	0.000398 ***
vive.conjuge	0.082334	0.139592	0.590	0.555331
ocup.dom.1	-0.019526	0.017787	-1.098	0.272337
ocup.dom.2	0.009566	0.020863	0.459	0.646594
tipo.esgoto	0.020881	0.016620	1.256	0.209021
freq.escola	-0.006807	0.034729	-0.196	0.844617
pos.ocup.1	0.035344	0.030844	1.146	0.251876
pos.ocup.2	-0.087698	0.174418	-0.503	0.615120
pos.ocup.3	0.094073	0.037273	2.524	0.011631 *
pos.ocup.4	0.033584	0.032066	1.047	0.294984
pos.ocup.5	0.030172	0.030245	0.998	0.318512
pos.ocup.6	0.230769	0.037385	6.173	7.14e-10 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9653 on 6165 degrees of freedom

Multiple R-squared: 0.07311, Adjusted R-squared: 0.068

F-statistic: 14.3 on 34 and 6165 DF, p-value: < 2.2e-16

## B.2 Modelo linear clássico com logaritmo natural da renda como variável dependente

Call:

```
lm(formula = normalize(log(renda)) ~ n.comodos.1 + n.comodos.2 +
n.dormitorios.1 + n.dormitorios.2 + n.dormitorios.3 + n.banheiros.1 +
```

```
n.banheiros.2 + normalize(idade.anos) + cri_ate7 + cri_7a14 +
n.idoso + tipofam.1 + tipofam.2 + sexo + raca.1 + raca.2 +
raca.3 + posse.radio + posse.tv + posse.maqlavar + posse.geladeira +
posse.computador + posse.auto + vive.conjuge + ocup.dom.1 +
ocup.dom.2 + tipo.esgoto + freq.escola + pos.ocup.1 + pos.ocup.2 +
pos.ocup.3 + pos.ocup.4 + pos.ocup.5 + pos.ocup.6 - 1, data = base.pnad.treino)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.9499	-0.4219	0.0073	0.4017	4.2627

Coefficients:

Estimate	Std. Error	t value	Pr(> t )
n.comodos.1	-0.247915	0.021762	-11.392 < 2e-16 ***
n.comodos.2	-0.206840	0.016283	-12.703 < 2e-16 ***
n.dormitorios.1	0.318185	0.030903	10.296 < 2e-16 ***
n.dormitorios.2	0.138213	0.029051	4.758 2.00e-06 ***
n.dormitorios.3	0.044926	0.029431	1.527 0.126932
n.banheiros.1	-0.205762	0.087306	-2.357 0.018464 *
n.banheiros.2	-0.186308	0.012088	-15.413 < 2e-16 ***
normalize(idade.anos)	0.076675	0.013168	5.823 6.07e-09 ***
cri_ate7	-0.185153	0.011723	-15.793 < 2e-16 ***
cri_7a14	-0.213376	0.010957	-19.474 < 2e-16 ***
n.idoso	0.068041	0.015560	4.373 1.25e-05 ***
tipofam.1	-0.439560	0.099869	-4.401 1.09e-05 ***
tipofam.2	-0.448166	0.099665	-4.497 7.03e-06 ***
sexo	0.053070	0.011673	4.547 5.56e-06 ***
raca.1	0.032877	0.009649	3.407 0.000660 ***
raca.2	-0.005427	0.014140	-0.384 0.701139
raca.3	-0.128257	0.083875	-1.529 0.126278
posse.radio	0.002963	0.010842	0.273 0.784657
posse.tv	0.071872	0.032008	2.245 0.024774 *
posse.maqlavar	0.149308	0.010393	14.366 < 2e-16 ***
posse.geladeira	0.081559	0.041362	1.972 0.048671 *
posse.computador	0.164601	0.010658	15.444 < 2e-16 ***
posse.auto	-0.162573	0.010668	-15.240 < 2e-16 ***
vive.conjuge	0.296660	0.099653	2.977 0.002923 **
ocup.dom.1	-0.054629	0.012666	-4.313 1.64e-05 ***
ocup.dom.2	0.037493	0.014852	2.525 0.011610 *

tipo.esgoto	0.096124	0.011856	8.108	6.16e-16	***
freq.escola	0.022573	0.024687	0.914	0.360552	
pos.ocup.1	0.160426	0.022147	7.244	4.89e-13	***
pos.ocup.2	0.204434	0.126952	1.610	0.107380	
pos.ocup.3	0.297084	0.026645	11.150	< 2e-16	***
pos.ocup.4	0.082891	0.022933	3.614	0.000303	***
pos.ocup.5	0.128820	0.021723	5.930	3.19e-09	***
pos.ocup.6	0.410154	0.026665	15.382	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6863 on 6167 degrees of freedom

Multiple R-squared: 0.5315, Adjusted R-squared: 0.5289

F-statistic: 205.8 on 34 and 6167 DF, p-value: < 2.2e-16

### B.3 Modelo linear de superpopulação com renda como variável dependente

Call:

```
svyglm(formula = normalize((renda)) ~ n.comodos.1 + n.comodos.2 +
n.dormitorios.1 + n.dormitorios.2 + n.dormitorios.3 + n.banheiros.1 +
n.banheiros.2 + normalize(idade.anos) + cri_ate7 + cri_7a14 +
n.idoso + tipofam.1 + tipofam.2 + sexo + raca.1 + raca.2 +
raca.3 + posse.radio + posse.tv + posse.maqlavar + posse.geladeira +
posse.computador + posse.auto + vive.conjuge + ocup.dom.1 +
ocup.dom.2 + tipo.esgoto + freq.escola + pos.ocup.1 + pos.ocup.2 +
pos.ocup.3 + pos.ocup.4 + pos.ocup.5 + pos.ocup.6 - 1, design = d.pnad)
```

Survey design:

```
svydesign(ids = ~V4618 + dom, strata = ~V4617, weights = ~V4610,
nest = T, data = base.pnad.treino)
```

Coefficients:

Estimate Std. Error t value Pr(>|t|)

n.comodos.1	-0.345704	0.037428	-9.237	< 2e-16	***
n.comodos.2	-0.279944	0.032174	-8.701	< 2e-16	***
n.dormitorios.1	0.340721	0.039911	8.537	3.40e-16	***
n.dormitorios.2	0.166456	0.033286	5.001	8.76e-07	***

n.dormitorios.3	0.084106	0.031560	2.665	0.008031	**
n.banheiros.1	-0.131050	0.037051	-3.537	0.000455	***
n.banheiros.2	-0.113611	0.018045	-6.296	8.49e-10	***
normalize(idade.anos)	0.043372	0.012578	3.448	0.000627	***
cri_ate7	-0.050529	0.007565	-6.679	8.59e-11	***
cri_7a14	-0.055536	0.007707	-7.206	3.16e-12	***
n.idoso	0.039661	0.016595	2.390	0.017342	*
tipofam.1	-0.241294	0.069014	-3.496	0.000528	***
tipofam.2	-0.241491	0.064605	-3.738	0.000214	***
sexo	0.051236	0.016538	3.098	0.002093	**
raca.1	0.042589	0.012172	3.499	0.000523	***
raca.2	-0.002455	0.010716	-0.229	0.818947	
raca.3	-0.039058	0.055302	-0.706	0.480463	
posse.radio	-0.007083	0.017534	-0.404	0.686463	
posse.tv	0.026358	0.024260	1.086	0.277959	
posse.maqlavar	0.049570	0.014723	3.367	0.000839	***
posse.geladeira	0.053535	0.025759	2.078	0.038351	*
posse.computador	0.112701	0.012700	8.874	< 2e-16	***
posse.auto	-0.116905	0.012324	-9.486	< 2e-16	***
vive.conjugue	0.105343	0.060552	1.740	0.082722	.
ocup.dom.1	-0.044198	0.011867	-3.724	0.000226	***
ocup.dom.2	0.026030	0.021847	1.191	0.234216	
tipo.esgoto	0.037517	0.011388	3.294	0.001079	**
freq.escola	0.013955	0.029519	0.473	0.636679	
pos.ocup.1	0.055894	0.017036	3.281	0.001131	**
pos.ocup.2	-0.130249	0.072446	-1.798	0.072995	.
pos.ocup.3	0.176075	0.026206	6.719	6.76e-11	***
pos.ocup.4	0.056467	0.018301	3.086	0.002182	**
pos.ocup.5	0.061876	0.016221	3.815	0.000159	***
pos.ocup.6	0.351746	0.063653	5.526	6.11e-08	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.7258885)

Number of Fisher Scoring iterations: 2

## B.4 Modelo linear de superpopulação com logaritmo natural da renda como variável dependente

Call:

```
svyglm(formula = normalize(log(renda)) ~ n.comodos.1 + n.comodos.2 +
n.dormitorios.1 + n.dormitorios.2 + n.dormitorios.3 + n.banheiros.1 +
n.banheiros.2 + normalize(idade.anos) + cri_ate7 + cri_7a14 +
n.idoso + tipofam.1 + tipofam.2 + sexo + raca.1 + raca.2 +
raca.3 + posse.radio + posse.tv + posse.maqlavar + posse.geladeira +
posse.computador + posse.auto + vive.conjugue + ocup.dom.1 +
ocup.dom.2 + tipo.esgoto + freq.escola + pos.ocup.1 + pos.ocup.2 +
pos.ocup.3 + pos.ocup.4 + pos.ocup.5 + pos.ocup.6 - 1, design = d.pnad)
```

Survey design:

```
svydesign(ids = ~V4618 + dom, strata = ~V4617, weights = ~V4610,
nest = T, data = base.pnad.treino)
```

Coefficients:

Estimate	Std. Error	t value	Pr(> t )
n.comodos.1	-0.213323	0.027180	-7.849 4.37e-14 ***
n.comodos.2	-0.183873	0.021845	-8.417 8.08e-16 ***
n.dormitorios.1	0.289333	0.030664	9.436 < 2e-16 ***
n.dormitorios.2	0.107189	0.028641	3.742 0.000211 ***
n.dormitorios.3	0.003638	0.030169	0.121 0.904094
n.banheiros.1	-0.267284	0.082099	-3.256 0.001234 **
n.banheiros.2	-0.184899	0.014589	-12.674 < 2e-16 ***
normalize(idade.anos)	0.087438	0.013094	6.678 8.70e-11 ***
cri_ate7	-0.193295	0.011422	-16.923 < 2e-16 ***
cri_7a14	-0.207219	0.011155	-18.576 < 2e-16 ***
n.idoso	0.078307	0.015861	4.937 1.19e-06 ***
tipofam.1	-0.402752	0.131431	-3.064 0.002338 **
tipofam.2	-0.420665	0.131291	-3.204 0.001470 **
sexo	0.041985	0.011492	3.653 0.000295 ***
raca.1	0.036237	0.010693	3.389 0.000776 ***
raca.2	0.003278	0.010975	0.299 0.765370
raca.3	-0.063165	0.083371	-0.758 0.449137
posse.radio	0.007402	0.010654	0.695 0.487614

posse.tv	0.041415	0.029566	1.401	0.162113	
posse.maqlavar	0.138959	0.009774	14.218	< 2e-16	***
posse.geladeira	0.061822	0.057088	1.083	0.279528	
posse.computador	0.157726	0.009707	16.249	< 2e-16	***
posse.auto	-0.177801	0.010618	-16.746	< 2e-16	***
vive.conjugue	0.284722	0.130723	2.178	0.030019	*
ocup.dom.1	-0.059169	0.014126	-4.189	3.50e-05	***
ocup.dom.2	0.013441	0.016736	0.803	0.422413	
tipo.esgoto	0.111289	0.015115	7.363	1.14e-12	***
freq.escola	0.042002	0.024346	1.725	0.085314	.
pos.ocup.1	0.189765	0.027855	6.813	3.79e-11	***
pos.ocup.2	0.067961	0.122862	0.553	0.580488	
pos.ocup.3	0.317953	0.031634	10.051	< 2e-16	***
pos.ocup.4	0.119587	0.029529	4.050	6.22e-05	***
pos.ocup.5	0.163901	0.026581	6.166	1.80e-09	***
pos.ocup.6	0.443112	0.033118	13.380	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.4668861)

Number of Fisher Scoring iterations: 2

## B.5 Coeficientes da RNA-Clássica com variável dependente renda

Camada de Saída		Camada de Entrada							
B0	0,100113455	b11	0,213765451	b02	0,009953819	b03	-0,070416413	b04	0,03316851
B1	1	b12	0,110156538	b21	-0,018251379	b31	-0,070699936	b41	-0,017704721
B2	0,16069724	b13	0,060313989	b22	-0,10379455	b32	-0,03524903	b42	0,096653762
B3	-0,201144361	b14	-0,060216658	b23	0,027423937	b33	0,037111392	b43	0,020002172
B4	-0,166627312	b15	-0,048278361	b24	0,085649447	b34	0,089577502	b44	0,014192948
		b16	0,028390758	b25	-0,094532477	b35	0,042792884	b45	-0,010278729
		b17	-0,01289331	b26	-0,0259385	b36	0,076658161	b46	0,101270551
		b18	-0,019393018	b27	0,049473317	b37	-0,098453606	b47	0,057114419
		b19	0,04658851	b28	0,10801189	b38	-0,100401477	b48	0,074841349
		b110	-0,178352364	b29	0,027968619	b39	-0,082814992	b49	0,069713555
		b111	-0,158631111	b210	0,002121965	b310	0,090107303	b410	-0,035592525
		b112	0,033671096	b211	-0,007801002	b311	-0,036804941	b411	-0,065628909
		b113	0,025377313	b212	0,076722619	b312	-0,045803059	b412	0,029258678
		b114	-0,001787892	b213	0,014798046	b313	0,038757179	b413	-0,009727408
		b115	-0,025074646	b214	0,075638369	b314	0,018561528	b414	0,033661464
		b116	0,003922562	b215	0,08635594	b315	0,020110908	b415	0,012725528
		b117	0,032151538	b216	0,053523839	b316	0,030399015	b416	-0,03511248
		b118	0,027469953	b217	0,034660109	b317	-0,085844827	b417	0,006581009
		b119	0,025836543	b218	-0,102953546	b318	-0,037993074	b418	-0,049853269
		b120	0,057508771	b219	0,020812491	b319	-0,108578373	b419	-0,096958762
		b121	-0,053207062	b220	0,05192789	b320	0,055317668	b420	0,002375336
		b122	0,082334315	b221	-0,009893777	b321	0,090385521	b421	-0,107220895
		b123	-0,019526371	b222	0,095963053	b322	0,026235973	b422	-0,048128329
		b124	0,009565984	b223	0,062032017	b323	0,057396591	b423	0,072775549
		b125	0,020881197	b224	-0,041324608	b324	-0,008934375	b424	-0,023186694
		b126	-0,006806951	b225	-0,106364129	b325	0,09450749	b425	0,010838388
		b127	0,035344095	b226	0,067399321	b326	-0,095982212	b426	-0,00656467
		b128	-0,087698038	b227	-0,028794288	b327	0,090774382	b427	-0,000870285
		b129	0,094072595	b228	-0,082745633	b328	-0,05921297	b428	0,03470333
		b130	0,033583923	b229	0,004955254	b329	0,048485235	b429	0,08361373
		b131	0,030172	b230	0,065103243	b330	-0,036867864	b430	-0,074976346
		b132	0,230768625	b231	-0,036710316	b331	0,075341733	b431	0,085972839
		b133	-0,053193614	b232	-0,027067553	b332	-0,015126659	b432	-0,021206818
		b134	-0,016089844	b233	0,051538002	b333	-0,015126659	b433	0,043467684
				b234	-0,020289752	b334	0,050874185	b434	0,043467684

Tabela B.1: Coeficientes da RNA - Clássica da com variável dependente renda



## B.6 Coeficientes da RNA-Clássica com variável dependente o logaritmo natural da renda

Camada de Saída		Camada de Entrada							
B0	0,085523359	b11	0,318184783	b02	0,06572354	b03	0,049006615	b04	0,022899926
B1	1	b12	0,13821254	b21	0,07094163	b31	-0,103986535	b41	-0,035725799
B2	0,728574355	b13	0,044926272	b22	0,008977126	b32	0,015198067	b42	0,051204112
B3	-0,247914571	b14	-0,205761977	b23	0,108487626	b33	-0,050530937	b43	0,08731412
B4	-0,20683966	b15	-0,186308182	b24	0,119780216	b34	0,060493564	b44	-0,015957182
		b16	0,076674893	b25	-0,071246744	b35	0,034041885	b45	0,034362385
		b17	-0,185153127	b26	0,00605569	b36	-0,107920722	b46	0,023010972
		b18	-0,213375925	b27	0,029630545	b37	-0,087117764	b47	0,028458397
		b19	0,068040704	b28	0,094461544	b38	-0,095169471	b48	0,002343878
		b110	-0,439560417	b29	-0,000898393	b39	-0,081308635	b49	-0,048846836
		b111	-0,448166482	b210	-0,00469784	b310	0,034741114	b410	-0,001333872
		b112	0,053070407	b211	-0,037460406	b311	-0,05981033	b411	-0,014566692
		b113	0,032877416	b212	0,075571688	b312	0,070975241	b412	0,032560061
		b114	-0,005427037	b213	-0,040719487	b313	-0,075960591	b413	-0,031629476
		b115	-0,128257213	b214	0,074737989	b314	0,023582552	b414	0,032868245
		b116	0,002962853	b215	0,007140512	b315	-0,024677267	b415	-0,065721275
		b117	0,071871898	b216	-0,005129728	b316	-0,042667732	b416	0,029114043
		b118	0,14930769	b217	-0,094911571	b317	0,057332081	b417	0,012485234
		b119	0,081559046	b218	-0,041789376	b318	0,089673393	b418	-0,015174061
		b120	0,164601248	b219	0,024317143	b319	-0,010965146	b419	0,054465333
		b121	-0,162573082	b220	-0,001696693	b320	-0,087756719	b420	0,027522299
		b122	0,296659885	b221	-0,030000878	b321	0,106208211	b421	-0,036425619
		b123	-0,054628975	b222	-0,054253623	b322	-0,048006033	b422	-0,022272311
		b124	0,037493326	b223	-0,007438678	b323	0,031708308	b423	0,023307091
		b125	0,096124014	b224	0,082444285	b324	-0,037724443	b424	0,066500188
		b126	0,022573155	b225	0,054230497	b325	0,028694436	b425	0,029329357
		b127	0,16042647	b226	-0,049964905	b326	0,094654219	b426	-0,050249807
		b128	0,20443355	b227	-0,084335117	b327	0,075728413	b427	0,040077761
		b129	0,297084156	b228	-0,004064501	b328	-0,029703182	b428	0,002618073
		b130	0,082890969	b229	0,049021531	b329	-0,05196136	b429	0,004875554
		b131	0,128820404	b230	-0,041061764	b330	0,07556323	b430	-0,048521141
		b132	0,410154308	b231	0,006378932	b331	0,07297847	b431	-0,051552075
		b133	0,031452183	b232	0,087555262	b332	-0,069589661	b432	-0,084544586
		b134	0,008094329	b233	-0,010463326	b333	-0,069589661	b433	-0,049387304
				b234	0,018375093	b334	0,000770114	b434	-0,049387304

Tabela B.2: Coeficientes da RNA - Clássica da com variável dependente o logaritmo natural da renda

## B.7 Coeficientes da RNA-Superpopulação com variável dependente a renda

Camada de Saída		Camada de Entrada							
B0	0,054139335	b11	0,34072088	b02	0,005973561	b03	-0,057795911	b04	-0,020046818
B1	1	b12	0,166456097	b21	0,100851185	b31	-0,078406272	b41	0,048301949
B2	0,510168438	b13	0,084105851	b22	0,060692	b32	-0,007757711	b42	0,032433542
B2	-0,345704343	b14	-0,131049907	b23	-0,012255214	b33	0,080780345	b43	0,030770382
B3	-0,27994421	b15	-0,113611434	b24	-0,002529451	b34	0,107622634	b44	-0,040676614
		b16	0,043372261	b25	0,000491708	b35	0,072636368	b45	-0,014592534
		b17	-0,050529352	b26	-0,06964946	b36	-0,084255648	b46	-0,079432227
		b18	-0,055536168	b27	0,067647198	b37	-0,029910237	b47	0,058585546
		b19	0,039660583	b28	0,054315495	b38	-0,014758618	b48	0,016886739
		b110	-0,241293858	b29	-0,015225523	b39	-0,084450474	b49	-0,003368584
		b111	-0,241491225	b210	0,008417972	b310	0,100844056	b410	-0,018273392
		b112	0,051236045	b211	-0,016403332	b311	0,072593728	b411	0,05059719
		b113	0,042589333	b212	-0,03584446	b312	0,032752483	b412	0,056439106
		b114	-0,002454587	b213	-0,035351022	b313	-0,09581879	b413	0,036553063
		b115	-0,03905757	b214	-0,077203675	b314	-0,045615168	b414	0,054093795
		b116	-0,007083028	b215	-0,051423745	b315	0,012540186	b415	0,085840351
		b117	0,026357945	b216	-0,076032272	b316	0,011654242	b416	-0,047715656
		b118	0,049569899	b217	0,004698054	b317	0,032190761	b417	0,0356279
		b119	0,05353548	b218	-0,009588402	b318	-0,002665548	b418	-0,092118142
		b120	0,112700657	b219	-0,040321544	b319	0,063665441	b419	-0,007085889
		b121	-0,116904722	b220	-0,016474269	b320	0,033349583	b420	-0,034101687
		b122	0,105343044	b221	0,049738418	b321	-0,033029897	b421	0,04075693
		b123	-0,044198072	b222	-0,083000091	b322	0,001179817	b422	-0,023204832
		b124	0,026030489	b223	-0,0469482	b323	-0,081104438	b423	-0,033828851
		b125	0,037516655	b224	-0,007710778	b324	-0,013206415	b424	0,075422235
		b126	0,013954644	b225	0,097891497	b325	0,021405711	b425	-0,072195984
		b127	0,055893994	b226	0,009274376	b326	-0,09899264	b426	-0,107647545
		b128	-0,130249186	b227	-0,055062395	b327	0,081650825	b427	-0,032379351
		b129	0,176074986	b228	0,090929352	b328	0,071720043	b428	0,019922945
		b130	0,056467222	b229	-0,065990096	b329	0,082925748	b429	0,102966174
		b131	0,061876185	b230	0,035951679	b330	-0,070172626	b430	-0,077624096
		b132	0,351745513	b231	-0,049619791	b331	0,046170634	b431	0,08583173
		b133	-0,014682603	b232	0,025226631	b332	0,003080467	b432	0,044230451
		b134	-0,067999505	b233	-0,066690867	b333	0,003080467	b433	-0,01725656
				b234	-0,066063567	b334	0,07868104	b434	-0,01725656

Tabela B.3: Coeficientes da RNA-Superpopulação com variável dependente a renda

## B.8 Coeficientes da RNA-Superpopulação da variável dependente o logaritmo natural da renda

Camada de Saída		Camada de Entrada							
B0	0,030684491	b11	0,289332825	b02	-0,104126011	b03	0,108212192	b04	-0,081978717
B1	1	b12	0,107189314	b21	-0,060848834	b31	0,070625073	b41	-0,11173802
B2	0,392457259	b13	0,003637585	b22	-0,018934197	b32	-0,083203748	b42	-0,063069785
B3	-0,213323215	b14	-0,267283613	b23	-0,064252011	b33	-0,10616124	b43	-0,007959586
B4	-0,183872925	b15	-0,184898634	b24	0,031579787	b34	0,014593504	b44	0,085224041
		b16	0,087437877	b25	0,024600299	b35	0,082479631	b45	-0,078926121
		b17	-0,193294899	b26	-0,019077512	b36	-0,043014668	b46	0,052698104
		b18	-0,207218674	b27	-0,014013554	b37	-0,099785496	b47	0,018081063
		b19	0,078307144	b28	-0,042449878	b38	-0,008048996	b48	0,059694855
		b110	-0,4027524	b29	0,005445703	b39	-0,095055348	b49	-0,046536291
		b111	-0,42066468	b210	-0,011963479	b310	0,06262506	b410	-0,065536384
		b112	0,041984705	b211	-0,07062145	b311	0,026185304	b411	-0,058665319
		b113	0,036236718	b212	0,015023439	b312	-0,002355703	b412	-0,1040071
		b114	0,003277633	b213	0,096897688	b313	-0,033567367	b413	0,009581775
		b115	-0,063165309	b214	-0,01102884	b314	0,023721361	b414	-0,063820589
		b116	0,00740245	b215	0,048428278	b315	0,097655685	b415	-0,013881034
		b117	0,041414522	b216	0,022222327	b316	-0,032728595	b416	0,064563589
		b118	0,138959398	b217	-0,110912155	b317	0,062806317	b417	-0,025287522
		b119	0,061822232	b218	0,0972951	b318	0,08145605	b418	-0,022017498
		b120	0,15772617	b219	-0,095201116	b319	-0,022730428	b419	-0,100104078
		b121	-0,177800583	b220	0,024511286	b320	-0,028592894	b420	-0,056304601
		b122	0,284722402	b221	0,04727063	b321	0,006086384	b421	-0,099289545
		b123	-0,05916879	b222	0,062016886	b322	0,066275147	b422	-0,095838313
		b124	0,01344125	b223	-0,016584562	b323	0,083792489	b423	-0,05194836
		b125	0,111289342	b224	-0,073122552	b324	0,007432144	b424	0,046862272
		b126	0,042001679	b225	-0,011147841	b325	0,039385563	b425	0,081356531
		b127	0,189764505	b226	-0,031496469	b326	-0,021100321	b426	0,039572246
		b128	0,067960749	b227	0,098430808	b327	-0,010106001	b427	-0,033486161
		b129	0,317952939	b228	-0,02807784	b328	0,087203595	b428	0,050921254
		b130	0,119587289	b229	0,095710972	b329	-0,081498953	b429	-0,062499328
		b131	0,163901433	b230	0,004881435	b330	-0,074557212	b430	0,079182621
		b132	0,443112158	b231	0,100616888	b331	-0,015535644	b431	0,007534296
		b133	-0,000238417	b232	0,103054518	b332	0,001829541	b432	-0,049440373
		b134	-0,070924202	b233	0,004165969	b333	0,001829541	b433	-0,023510567
				b234	-0,078766352	b334	-0,122384261	b434	-0,023510567

Tabela B.4: Coeficientes da RNA-Superpopulação da variável dependente o logaritmo natural da renda

## B.9 Gráficos dos resíduos dos modelos ajustados

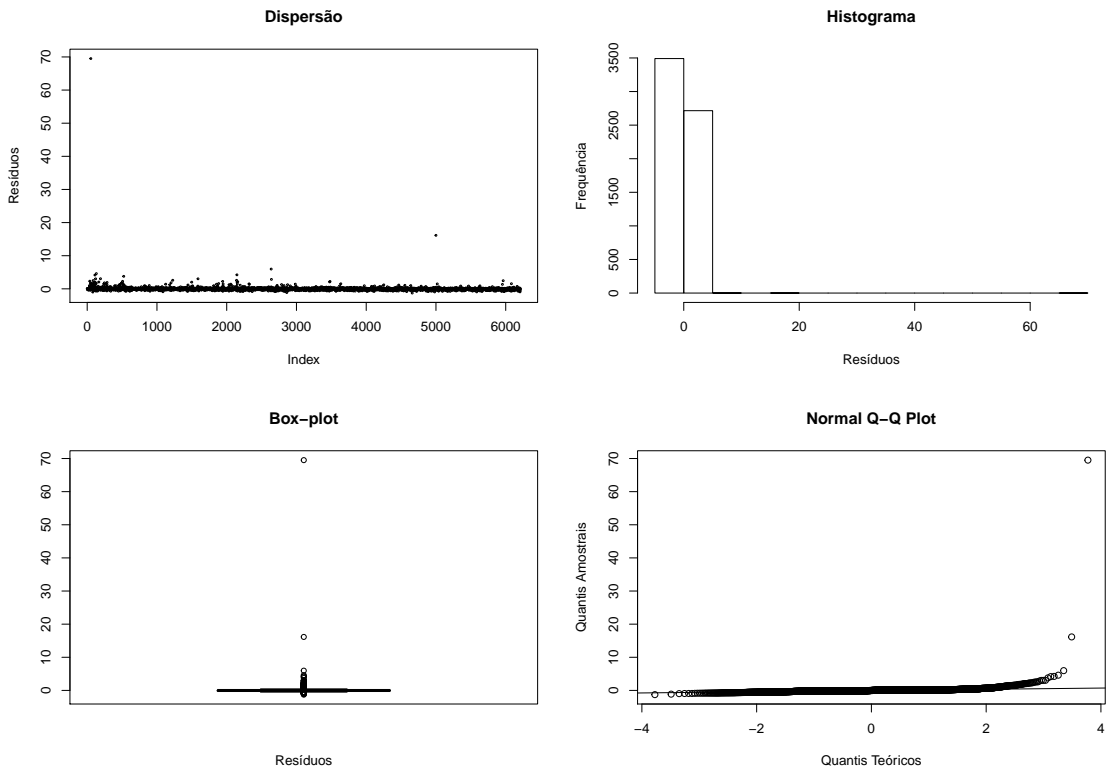


Figura B.1: Resíduos do modelo linear clássico da renda

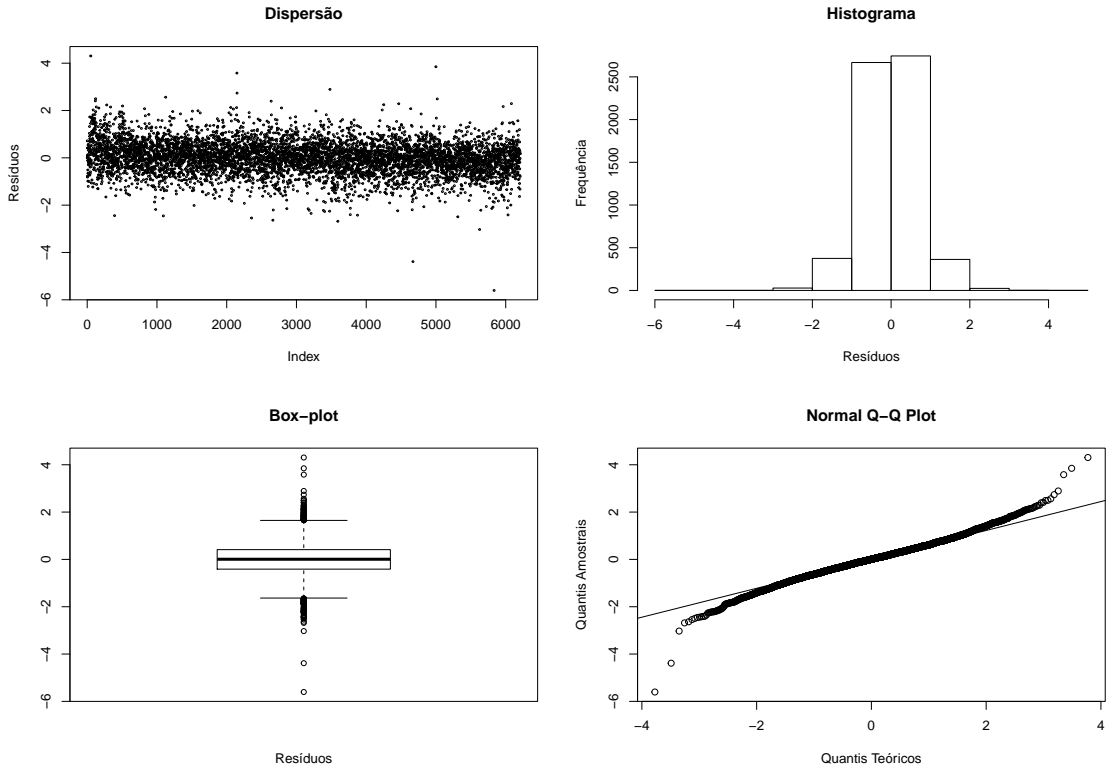


Figura B.2: Resíduos do modelo linear clássico do log. natural da renda

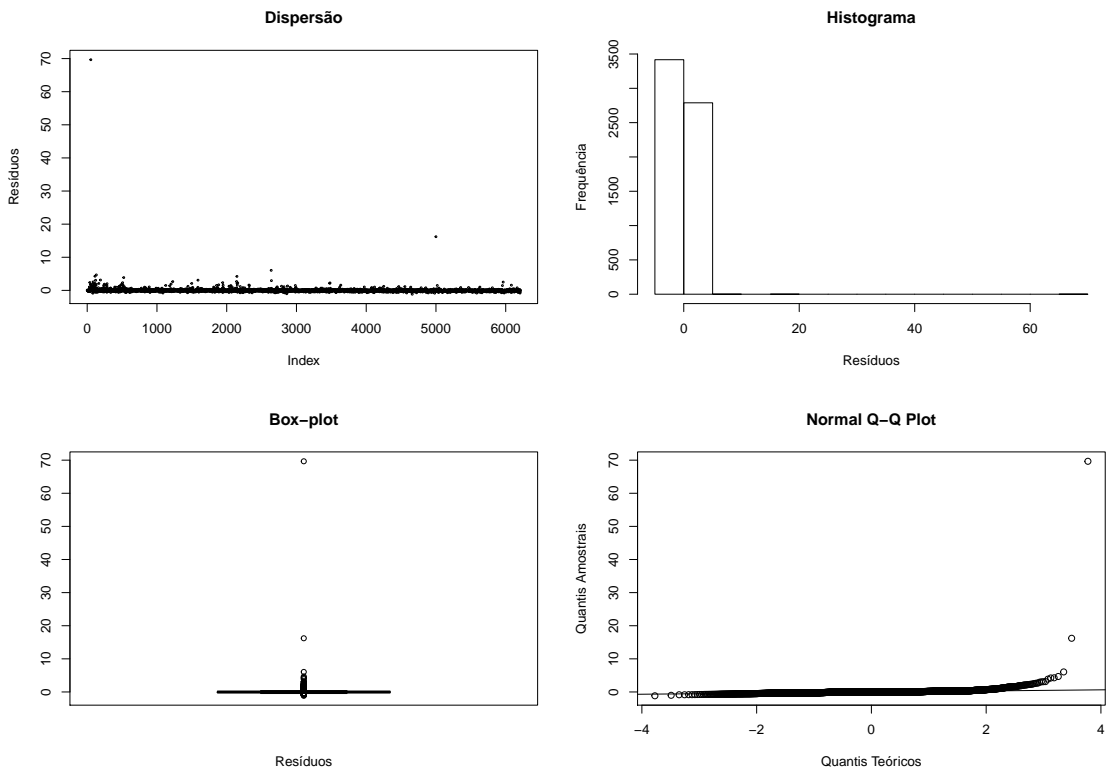


Figura B.3: Resíduos do modelo linear de superpopulação da renda

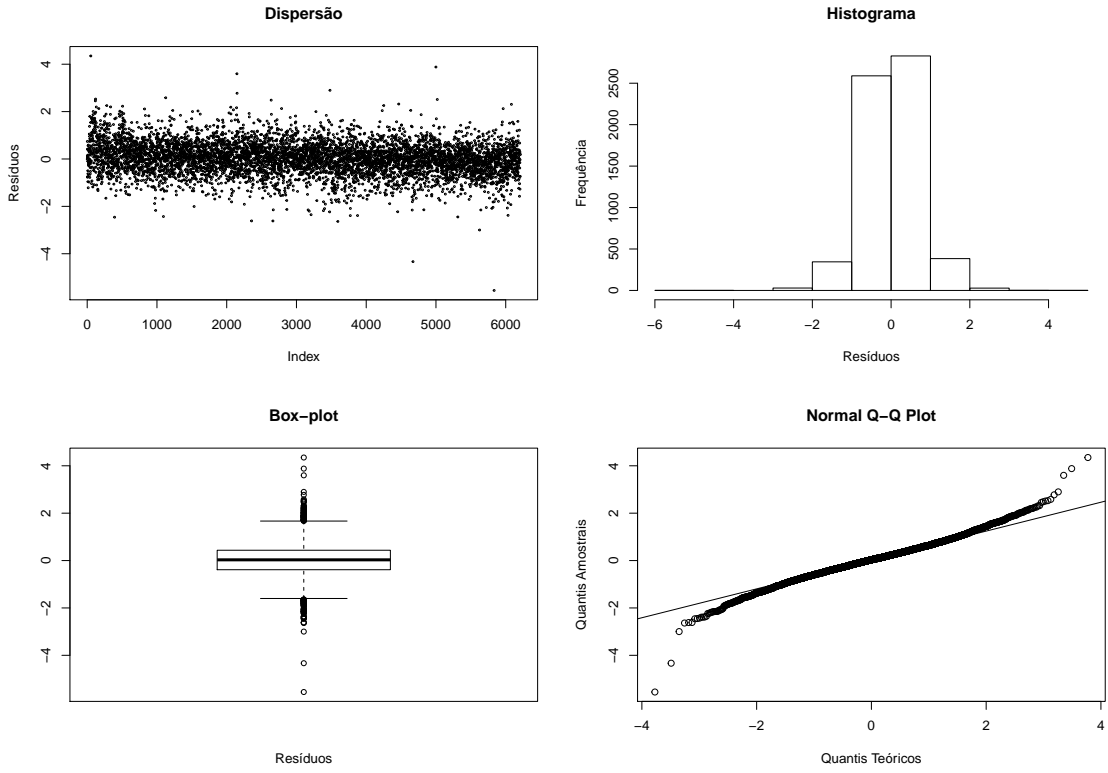


Figura B.4: Resíduos do modelo linear de superpopulação do log. natural da renda

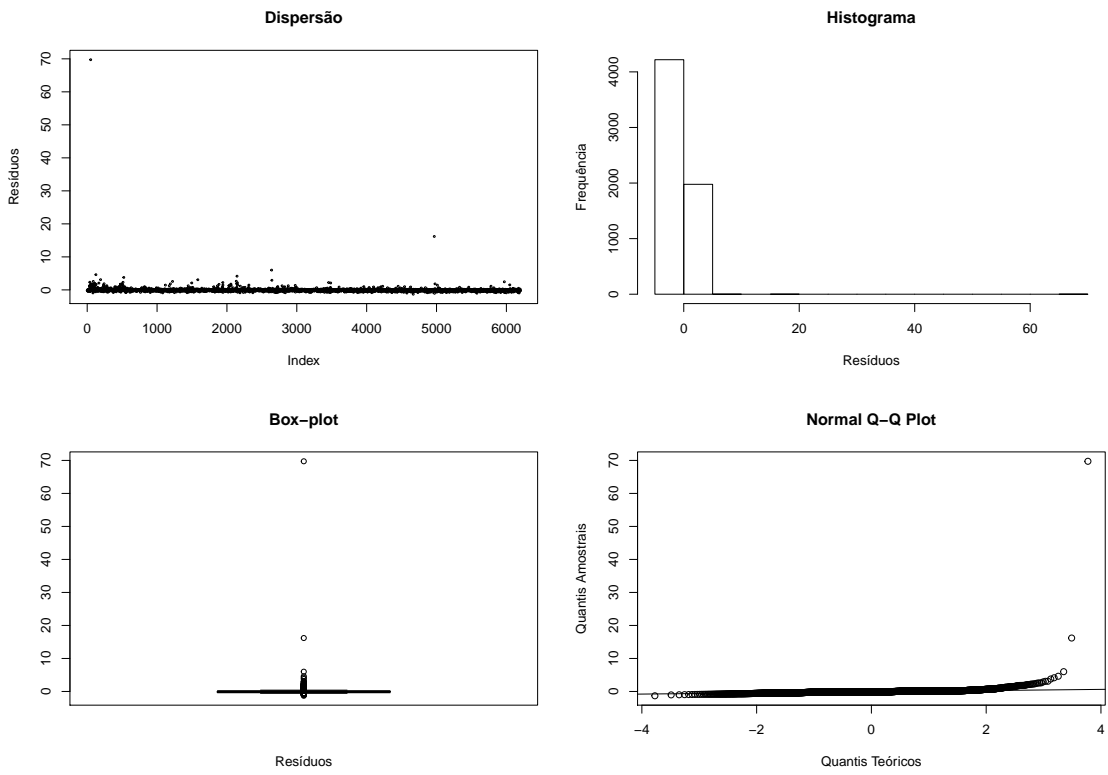


Figura B.5: Resíduos da RNA clássica da renda

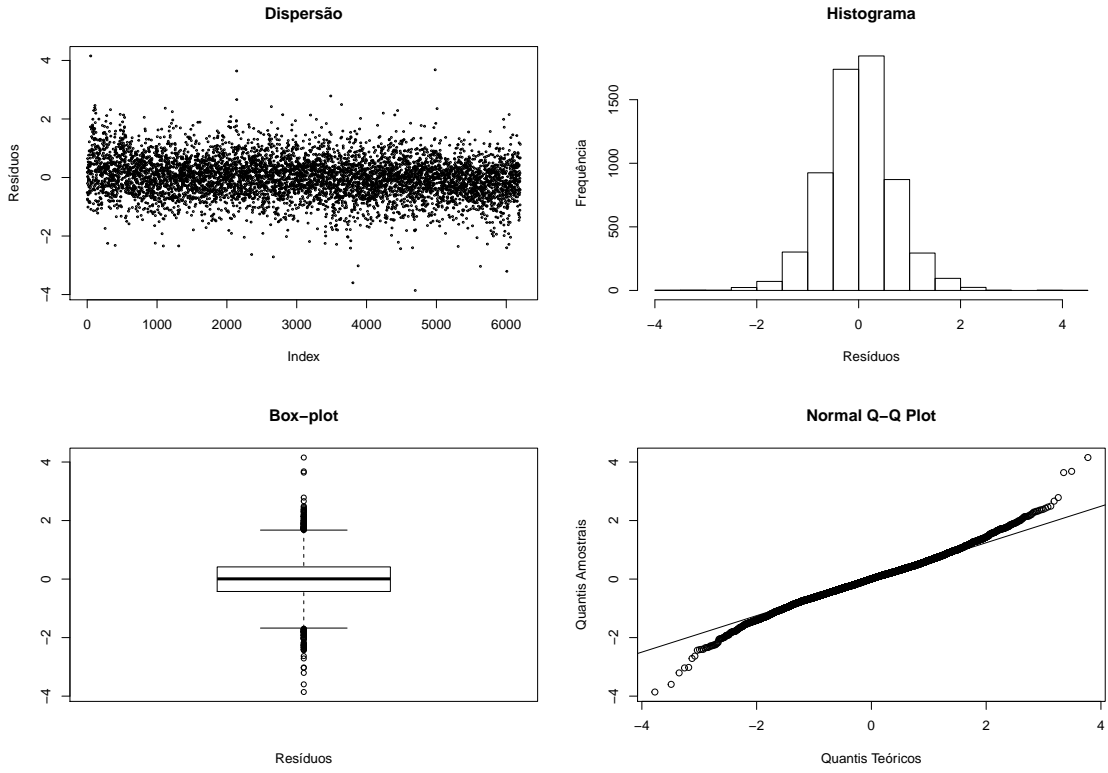


Figura B.6: Resíduos da RNA clássica do log. natural da renda

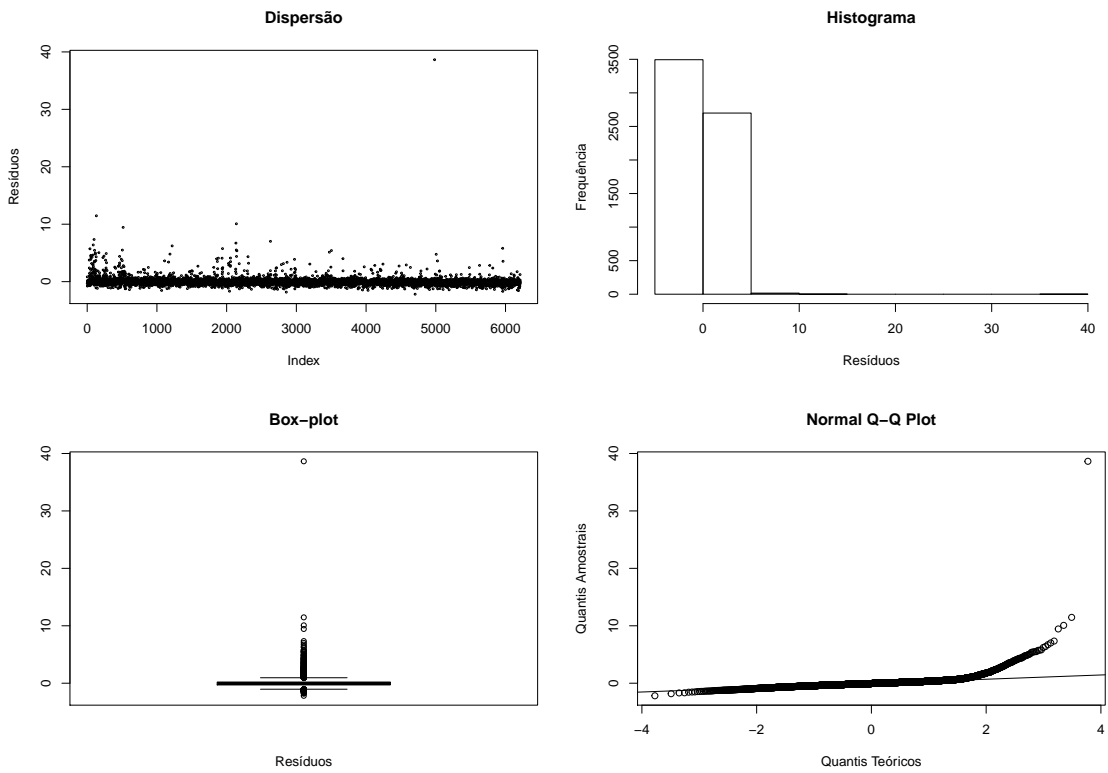


Figura B.7: Resíduos da RNA de superpopulação da renda

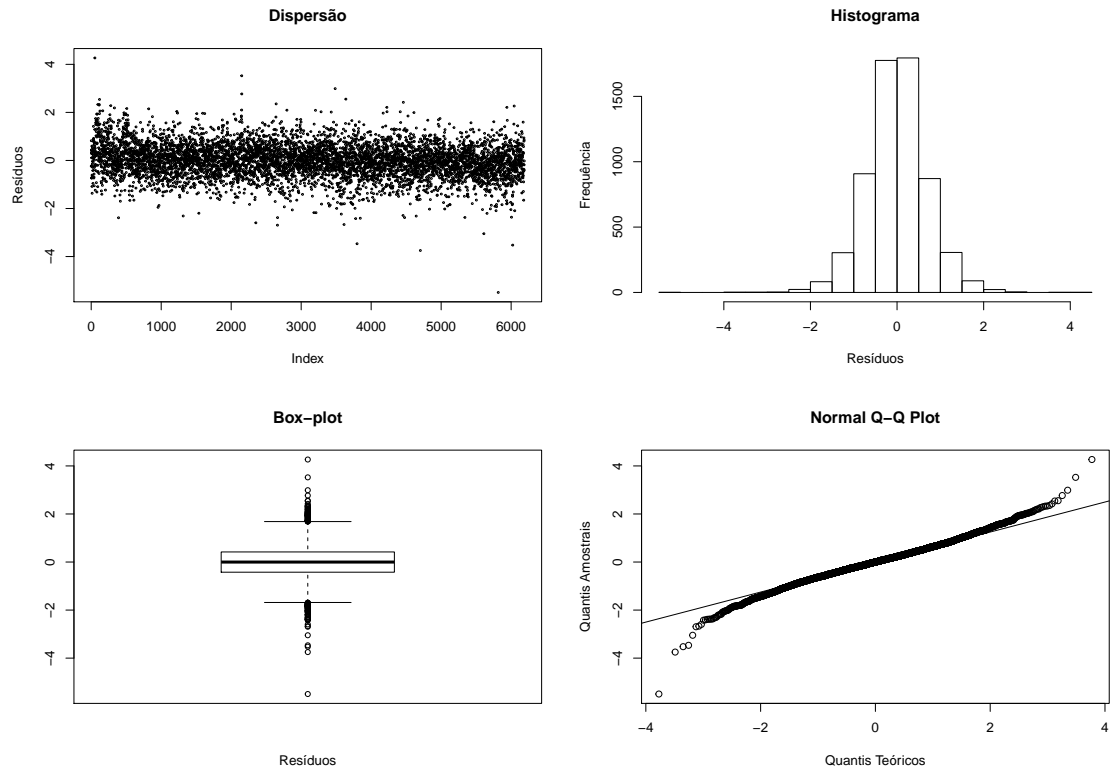


Figura B.8: Resíduos da RNA de superpopulação do log. natural da renda

## B.10 Gráficos da evolução do SQE dos modelos de RNA estimados



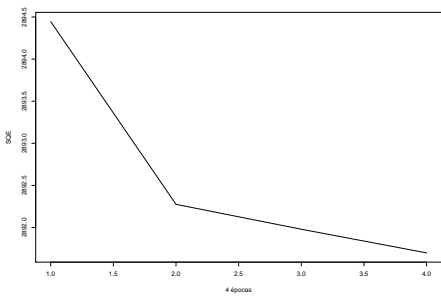


Figura B.9: Evolução do SQE da RNA clássica da renda

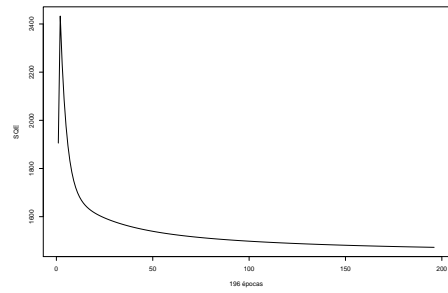


Figura B.10: Evolução do SQE da RNA clássica da log. natural da renda

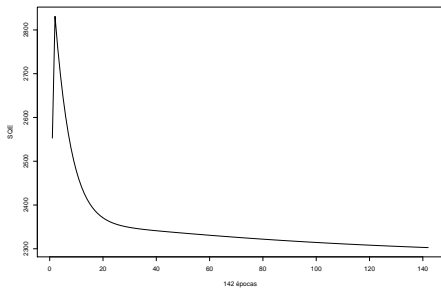


Figura B.11: Evolução do SQE da RNA de superpopulação da renda

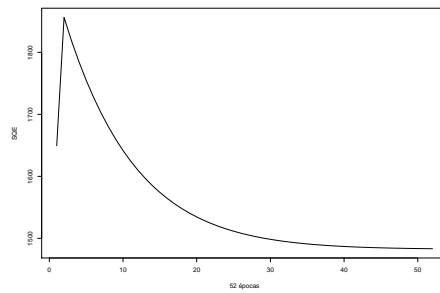


Figura B.12: Evolução do SQE da RNA de superpopulação do log. natural da renda

# Apêndice C

## Scripts R

### C.1 RNA sob abordagem clássica identificável

```
#####  
# Script R para ajuste de uma RNA identificável com dados amostrais simples #  
#####  
# filename: 004_tanh_2limiar_amostra2000_simples_identificavel.R #  
# Written by: Savano #  
# Date: Set,2016 #  
# #  
# O programa ajusta uma RNA-identificável com 1 neurônio na camada de entrada, #  
# 1 neurônio tanh com limiar na camada escondida e 1 neurônio linear com limiar #  
# na camada de saída. O algoritmo BP é utilizado. #  
# #  
# Input: amostra200_q.txt #  
# Output: rna_tanh1_2limiar_simples_amostra200_identificavel.txt #  
#####  
  
library(som)  
setwd('/home/savano/Documents/particular/amostra_novapop4/amostra2000/graficos')  
  
# Definição dos parâmetros da rede #  
replica <- 100  
pop.ref <- c(312,573)  
n.epocas <- 5000  
v.alfa <- 0.001  
c.sqe <- 0.0001  
c.coef <- 0.0001
```

```

m001 <- (matrix(nrow=replica,ncol=16))
m002 <- (matrix(nrow=0,ncol=16))

# Caminho e arquivos de entrada/saida
local.dados <- "/home/savano/Documents/particular/amostra_novapop4/amostra2000/"
arq.entrada <- "amostra2000_"
arq.saida <- "rna_tanh1_2limiar_simples_amostra2000_identificavel_log_batelada_ma

for(q in pop.ref){
arquivo <- paste(local.dados,arq.entrada,q, ".txt", sep="")

pop <- read.table(arquivo,sep="|",header=TRUE)

y <- as.matrix(normalize(log(as.numeric(pop$y))))
z <- as.matrix(normalize(log(as.numeric(pop$x))))

#b0 <- 3
#b1 <- 1

#while(b0>b1){
for(u in 1:replica){
b0 <- runif(1, min=0.1,max=0.22)
b1 <- runif(1, min=0.1,max=0.22)
B0 <- runif(1, min=0.01,max=0.022)
B1 <- runif(1, min=0.1,max=2)

b0.inicial <- b0
b1.inicial <- b1
B0.inicial <- B0
B1.inicial <- B1

v <- matrix(nrow=nrow(y),ncol=1)
yhat <- matrix(nrow=nrow(y),ncol=1)
erro <- matrix(nrow=nrow(y),ncol=1)
alfa <- v.alfa
epocas <- n.epocas
sqe <- matrix(nrow=epocas,ncol=1)

```

```

coef   <- matrix(nrow=epocas,ncol=4)
delta.B0 <- matrix(nrow=nrow(y),ncol=1)
delta.B1 <- matrix(nrow=nrow(y),ncol=1)
delta.b0 <- matrix(nrow=nrow(y),ncol=1)
delta.b1 <- matrix(nrow=nrow(y),ncol=1)

for(j in 1:epocas){
for(i in 1:nrow(y)){
v[i,]   <- tanh(b0+b1*z[i,])
yhat[i,] <- B0+B1*v[i,]
erro[i,] <- y[i,]-yhat[i,]

delta.B0[i,] <- alfa*erro[i,]
delta.B1[i,] <- alfa*erro[i,]*v[i,]
delta.b0[i,] <- alfa*erro[i,]*B1*(1-v[i,]^2)
delta.b1[i,] <- alfa*erro[i,]*z[i,]*(1-v[i,]^2)*B1

# B0 <- abs(B0 + alfa*erro[i,])
# B1 <- abs(B1 + alfa*erro[i,]*v[i,])
# b0 <- abs(b0 + alfa*erro[i,]*B1*(1-v[i,]^2))
# b1 <- abs(b1 + alfa*erro[i,]*z[i,]*(1-v[i,]^2)*B1)

}

B0 <- B0+colMeans(delta.B0)
B1 <- abs(B1+colMeans(delta.B1))
b0 <- abs(b0+colMeans(delta.b0))
b1 <- abs(b1+colMeans(delta.b1))

sqe[j,] <- sum(erro^2)/2
coef[j,1] <- b0
coef[j,2] <- b1
coef[j,3] <- B0
coef[j,4] <- B1

if(j>1000){
v.sqe   <- abs((sqe[j-1000,]-sqe[j,])/sqe[j-1000])
d.coef.b0 <- abs((coef[j,1]-coef[j-1,1]))
d.coef.b1 <- abs((coef[j,2]-coef[j-1,2]))
}

```

```

d.coef.B0 <- abs((coef[j,3]-coef[j-1,3]))
d.coef.B1 <- abs((coef[j,4]-coef[j-1,4]))
if((v.sqe<=c.sqe || d.coef.b0<=c.coef & d.coef.b1<=c.coef & d.coef.B0<=c.coef & d.c
break
}
}
}
m001[u,1] <- q
m001[u,2] <- u
m001[u,3] <- b0
m001[u,4] <- b1
m001[u,5] <- B0
m001[u,6] <- B1
m001[u,7] <- v.sqe
m001[u,8] <- j
m001[u,9] <- d.coef.b0
m001[u,10] <- d.coef.b1
m001[u,11] <- d.coef.B0
m001[u,12] <- d.coef.B1
m001[u,13] <- b0.inicial
m001[u,14] <- b1.inicial
m001[u,15] <- B0.inicial
m001[u,16] <- B1.inicial
}
#}
m002 <- rbind(m002,m001)

cat("população:",q,sep="\n")

converge <- as.data.frame(coef)

names(converge) <- c('b0','b1','B0','B1')
jpeg(file=paste('b0','populacao',q,'.jpeg',sep=''),
units="in",
width=20,
height=10,
pointsize=12,
res=72)
plot(converge$b0,type="l",xlab='Épocas',ylab='Coeficiente b0',main = paste('Amostra

```

```

dev.off()

jpeg(file=paste('b1', 'populacao', q, '.jpeg', sep=''),
units="in",
width=20,
height=10,
pointsize=12,
res=72)
plot(converge$b1,type="l",xlab='Épocas',ylab='Coeficiente b1',main = paste('Amostra
dev.off()

jpeg(file=paste('B0', 'populacao', q, '.jpeg', sep=''),
units="in",
width=20,
height=10,
pointsize=12,
res=72)
plot(converge$B0,type="l",xlab='Épocas',ylab='Coeficiente B0',main = paste('Amostra
dev.off()

jpeg(file=paste('B1', 'populacao', q, '.jpeg', sep=''),
units="in",
width=20,
height=10,
pointsize=12,
res=72)
plot(converge$B1,type="l",xlab='Épocas',ylab='Coeficiente B1',main = paste('Amostra
dev.off()

jpeg(file=paste('SQE_POP', q, '.jpeg', sep=''),
units="in",
width=20,
height=10,
pointsize=12,
res=72)
plot(na.exclude(sqe),type="l",xlab='Épocas',ylab='SQE',main=paste('SQE - Amostra 20
dev.off()
}

```

```
#plot(na.exclude(sqe),type="l")

pop.rna.tanh.2 <- as.data.frame(m002)
names(pop.rna.tanh.2) <- c("pop","replica","b0","b1","B0","B1","v.sqe","epoca",
"delta b0","delta b1","delta B0","delta B1",
"bo.inicial","b1.inicial","B0.inicial","B1.inicial")

write.table(pop.rna.tanh.2,
file=paste(local.dados,arq.saida,sep=""),
quote=FALSE,sep="|",row.names=FALSE)
```

## C.2 RNA sob abordagem de superpopulação identificável

```
#####
# Script R para ajuste de uma RNA identificável com dados amostrais complexos #
#####
# filename: 004_tanh_2limiar_amostra2000_complexo_identificavel.R #
# Written by: Savano #
# Date: Set,2016 #
# #
# O programa ajusta uma RNA-identificável com 1 neurônio na camada de entrada, #
# 1 neurônio tanh com limiar na camada escondida e 1 neurônio linear com limiar #
# na camada de saída. O algoritmo BP é utilizado. #
# #
# Input: amostra2000_q.txt #
# Output: rna_tanh1_2limiar_complexo_amostra2000_identificavel.txt #
#####

library(som)
setwd('/home/savano/Documents/particular/amostra_novapop4/amostra2000/graficos')

# Definição dos parâmetros da rede #
replica <- 100
pop.ref <- c(34)
n.epocas <- 5000
```

```

v.alfa <- 0.001
c.sqe <- 0.0001
c.coef <- 0.0001

N <- 14000
nh <- 2000

m001 <- (matrix(nrow=replica,ncol=16))
m002 <- (matrix(nrow=0,ncol=16))

# Caminho e arquivos de entrada/saida
local.dados <- "/home/savano/Documents/particular/amostra_novapop4/amostra2000/"
arq.entrada <- "amostra2000_"
arq.saida <- "rna_tanh1_2limiar_complexo_amostra2000_identificavel_log_batelada_m

for(q in pop.ref){
arquivo <- paste(local.dados,arq.entrada,q, ".txt", sep="")

pop <- read.table(arquivo,sep="|",header=TRUE)

y <- as.matrix(normalize(log(pop$y)))
z <- as.matrix(normalize(log(pop$x)))
w <- as.matrix(pop$Nh/pop$nh)/N*nh

# b0 <- 3
# b1 <- 1
#
#while(b0>b1){
for(u in 1:replica){
b0 <- runif(1, min=0.1,max=0.22)
b1 <- runif(1, min=0.1,max=0.22)
B0 <- runif(1, min=0.01,max=0.022)
B1 <- runif(1, min=0.1,max=2)

b0.inicial <- b0
b1.inicial <- b1
B0.inicial <- B0
B1.inicial <- B1

```



```

v      <- matrix(nrow=nrow(y),ncol=1)
yhat   <- matrix(nrow=nrow(y),ncol=1)
erro   <- matrix(nrow=nrow(y),ncol=1)
alfa   <- v.alfa
epocas <- n.epocas
sqe    <- matrix(nrow=epocas,ncol=1)
coef   <- matrix(nrow=epocas,ncol=4)
delta.B0 <- matrix(nrow=nrow(y),ncol=1)
delta.B1 <- matrix(nrow=nrow(y),ncol=1)
delta.b0 <- matrix(nrow=nrow(y),ncol=1)
delta.b1 <- matrix(nrow=nrow(y),ncol=1)

for(j in 1:epocas){
  for(i in 1:nrow(y)){
    v[i,]    <- tanh(b0+b1*z[i,])
    yhat[i,] <- B0+B1*v[i,]
    erro[i,] <- y[i,]-yhat[i,]

    delta.B0[i,] <- alfa*erro[i,]*w[i,]
    delta.B1[i,] <- alfa*erro[i,]*v[i,]*w[i,]
    delta.b0[i,] <- alfa*erro[i,]*B1*(1-v[i,]^2)*w[i,]
    delta.b1[i,] <- alfa*erro[i,]*z[i,]*(1-v[i,]^2)*B1*w[i,]

    # B0 <- abs(B0 + alfa*erro[i,])*w[i,]
    # B1 <- abs(B1 + alfa*erro[i,]*v[i,])*w[i,]
    # b0 <- abs(b0 + alfa*erro[i,]*B1*(1-v[i,]^2))*w[i,]
    # b1 <- abs(b1 + alfa*erro[i,]*z[i,]*(1-v[i,]^2)*B1)*w[i,]
  }

  B0 <- B0+colMeans(delta.B0)
  B1 <- abs(B1+colMeans(delta.B1))
  b0 <- abs(b0+colMeans(delta.b0))
  b1 <- abs(b1+colMeans(delta.b1))

  sqe[j,] <- sum(erro^2)/2
  coef[j,1] <- b0
  coef[j,2] <- b1
  coef[j,3] <- B0
  coef[j,4] <- B1
}

```

```

if(j>1000){
v.sqe      <- abs((sqe[j-1000,]-sqe[j,])/sqe[j-1000])
d.coef.b0  <- abs((coef[j,1]-coef[j-1,1]))
d.coef.b1  <- abs((coef[j,2]-coef[j-1,2]))
d.coef.B0  <- abs((coef[j,3]-coef[j-1,3]))
d.coef.B1  <- abs((coef[j,4]-coef[j-1,4]))
if((v.sqe<=c.sqe || d.coef.b0<=c.coef & d.coef.b1<=c.coef & d.coef.B0<=c.coef & d.c
break
}
}
}
m001[u,1] <- q
m001[u,2] <- u
m001[u,3] <- b0
m001[u,4] <- b1
m001[u,5] <- B0
m001[u,6] <- B1
m001[u,7] <- v.sqe
m001[u,8] <- j
m001[u,9]  <- d.coef.b0
m001[u,10] <- d.coef.b1
m001[u,11] <- d.coef.B0
m001[u,12] <- d.coef.B1
m001[u,13] <- b0.inicial
m001[u,14] <- b1.inicial
m001[u,15] <- B0.inicial
m001[u,16] <- B1.inicial
}
#}
m002 <- rbind(m002,m001)

cat("população:",q,sep="\n")

converge <- as.data.frame(coef)

names(converge) <- c('b0','b1','B0','B1')
jpeg(file=paste('b0','amostra',q,'.jpeg',sep=''),
units="in",

```

```

width=20,
height=10,
pointsize=12,
res=72)
plot(converge$b0,type="l",xlab='Épocas',ylab='Coeficiente b0',main = paste('Amostra',q),
dev.off()

jpeg(file=paste('b1','amostra',q,'.jpeg',sep=''),
units="in",
width=20,
height=10,
pointsize=12,
res=72)
plot(converge$b1,type="l",xlab='Épocas',ylab='Coeficiente b1',main = paste('Amostra',q),
dev.off()

jpeg(file=paste('B0','amostra',q,'.jpeg',sep=''),
units="in",
width=20,
height=10,
pointsize=12,
res=72)
plot(converge$B0,type="l",xlab='Épocas',ylab='Coeficiente B0',main = paste('Amostra',q),
dev.off()

jpeg(file=paste('B1','amostra',q,'.jpeg',sep=''),
units="in",
width=20,
height=10,
pointsize=12,
res=72)
plot(converge$B1,type="l",xlab='Épocas',ylab='Coeficiente B1',main = paste('Amostra',q),
dev.off()

jpeg(file=paste('SQE_amostra',q,'.jpeg',sep=''),
units="in",
width=20,
height=10,
pointsize=12,

```

```

res=72)
plot(na.exclude(sq),type="l",xlab='Épocas',ylab='SQE',main=paste('SQE - Amostra -
dev.off()
}

#plot(na.exclude(sq),type="l")

pop.rna.tanh.2 <- as.data.frame(m002)
names(pop.rna.tanh.2) <- c("pop","replica","b0","b1","B0","B1","v.sqe","epoca",
"delta b0","delta b1","delta B0","delta B1",
"bo.inicial","b1.inicial","B0.inicial","B1.inicial")

write.table(pop.rna.tanh.2,
file=paste(local.dados,arq.saida,sep=""),
quote=FALSE,sep="|",row.names=FALSE)

```

### C.3 Construção da base da PNAD-2014

```

# Modelagem PNAD 2014
library('dplyr')
library('RMySQL')
library('survey')
library('MASS')
options(survey.lonely.psu = "adjust")

con <-
dbConnect(
MySQL(),
user = 'root',
password = 'mariadb',
host = '127.0.0.1' ,
dbname = 'tese',
port = 3306
)

# PNAD 2014 #
rs.pnad <-
dbSendQuery(
con,

```

"select UF,  
V4611,  
V4618,  
V0102,  
V0103,  
V4617,  
V4618,  
V4105,  
V4107,  
V4600,  
V4601,  
V4602,  
V4604,  
V4605,  
V4606,  
V4607,  
V4608,  
V4609,  
V4610,  
V4614,  
V4621,  
V4620,  
V0105,  
V2032,  
V0215,  
V0231,  
V0228,  
V0230,  
V0225,  
TV,  
V4624,  
V0202,  
V0217,  
V0206,  
V0205,  
V0207,  
V0401,  
V0404,  
V0602,

```

V4111,
V8005,
V4706,
V0302,
tipofam,
cri_ate7,
cri_7a14,
idoso,
V2016
from pnad2014_mg;"
)
pnad.peso <- fetch(rs.pnad, n = -1)
pnad.rs <- dbHasCompleted(rs.pnad)
dbClearResult(rs.pnad)

# Plano amostral da PNAD com pesos do desenho e dois estagios
pnad.peso$dom <- pnad.peso$V0102 * 1000 + pnad.peso$V0103

pnad.peso$UF.reg[pnad.peso$UF == 31 & pnad.peso$V4107 == 1] <- 1
pnad.peso$UF.reg[pnad.peso$UF == 31 & pnad.peso$V4107 > 1] <- 2

# Cria objeto com informações da estrutura do plano amostral básico da PNAD 2014
d.pnad <-
svydesign(
ids = ~ V4618 + dom ,
strata = ~ V4617 ,
weights = ~ V4610 ,
nest = T,
data = pnad.peso
)

# Localiza valores das projeções de população das regiões do estado de MG
proj_UF.reg <- unique(pnad.peso$V4609)
tot_UF.reg_r <- as.data.frame(cbind(1:2 , proj_UF.reg))
UF.reg <- as.data.frame(pnad.peso$UF.reg)
names(tot_UF.reg_r) <- c('UF.reg', 'Freq')

## Cria objeto contendo dados e pesos calibrados da PNAD 2014
d.rak <- postStratify(d.pnad , ~ UF.reg, population = tot_UF.reg_r)

```

```

# Recupera pesos calibrados dos domicílios
# Tem que coincidir com valores da variável V4611
pesos.calib <- weights(d.rak)

# Número de comodos
pnad.v0205 <- pnad.peso$V0205

n.comodos.1 <- ifelse(pnad.v0205 %in% c(1,2,3,4), 1,-1)
n.comodos.2 <- ifelse(pnad.v0205 %in% c(5, 6, 7, 8), 1,-1)
n.comodos.3 <- ifelse(pnad.v0205 >= 9, 1,-1)

n.comodos <- cbind.data.frame(n.comodos.1, n.comodos.2, n.comodos.3)

# Número de comodos utilizados como dormitório
pnad.v0206 <- pnad.peso$V0206

n.dormitorios.1 <- ifelse(pnad.v0206 == 1, 1,-1)
n.dormitorios.2 <- ifelse(pnad.v0206 == 2, 1,-1)
n.dormitorios.3 <- ifelse(pnad.v0206 == 3, 1,-1)
n.dormitorios.4 <- ifelse(pnad.v0206 >= 4, 1,-1)

n.dormitorios <-
cbind.data.frame(n.dormitorios.1,
n.dormitorios.2,
n.dormitorios.3,
n.dormitorios.4)

# Número de banheiros
pnad.v2016 <- pnad.peso$V2016

n.banheiros.1 <- ifelse(pnad.v2016 == 0, 1,-1)
n.banheiros.2 <- ifelse(pnad.v2016 == 1, 1,-1)
n.banheiros.3 <- ifelse(pnad.v2016 >= 2, 1,-1)

n.banheiros <-
cbind.data.frame(n.banheiros.1,
n.banheiros.2,

```

```

n.banheiros.3)

# Idade em anos
idade.anos <- pnad.peso$V8005

# Crianças até 7 anos
pnad.cri_ate7 <- pnad.peso$cri_ate7

cri_ate7 <- ifelse(pnad.cri_ate7 >= 1, 1, -1)

# Crianças com idade entre 7 e 14 anos
pnad.cri_7a14 <- pnad.peso$cri_7a14

cri_7a14 <- ifelse(pnad.cri_7a14 >= 1, 1, -1)

# Número de idosos (pessoas com 60 anos ou mais)
pnad.idoso <- pnad.peso$idoso

n.idoso <- ifelse(pnad.peso$idoso >= 1, 1, -1)

# Tipo de família
pnad.tipofam <- pnad.peso$tipofam

tipofam.1 <- ifelse(pnad.tipofam == 1, 1, -1)
tipofam.2 <- ifelse(pnad.tipofam == 2, 1, -1)
tipofam.3 <- ifelse(pnad.tipofam == 3, 1, -1)

tipofam <-
cbind.data.frame(tipofam.1,
tipofam.2,
tipofam.3)

# Possui Banheiro
pnad.v0215 <- pnad.peso$V0215

possui.banheiro <- ifelse(pnad.v0215 == 0, -1, 1)

# Sexo do responsável pelo domicílio
pnad.v0302 <- pnad.peso$V0302

```



```

sexo <- ifelse(pnad.v0302 == 1, 1, -1)

# Cor raça do responsável pelo domicílio
pnad.v0404 <- pnad.peso$V0404
raca.1 <- ifelse(pnad.peso$V0404 == 1, 1, -1)
raca.2 <- ifelse(pnad.peso$V0404 == 2, 1, -1)
raca.3 <- ifelse(pnad.peso$V0404 == 3, 1, -1)
raca.4 <- ifelse(pnad.peso$V0404 == 4, 1, -1)
raca <- cbind.data.frame(raca.1, raca.2, raca.3, raca.4)

# Existência de rádio no domicílio
pnad.v0225 <- pnad.peso$V0225
posse.radio <- ifelse(pnad.v0225 == 1, 1, -1)

# Existência de televisão no domicílio
pnad.tv <- pnad.peso$TV
posse.tv <- ifelse(pnad.tv == 1, 1, -1)

# Existência de máquina de lavar roupa no domicílio
pnad.v0230 <- pnad.peso$V0230
posse.maqlavar <- ifelse(pnad.v0230 == 1, 1, -1)

# Existência de geladeira no domicílio
pnad.v0228 <- pnad.peso$V0228
posse.geladeira <- ifelse(pnad.v0228 == 1, 1, -1)

# Existência de microcomputador no domicílio
pnad.v0231 <- pnad.peso$V0231
posse.computador <- ifelse(pnad.v0231 == 1, 1, -1)

# Existência de automóvel para uso particular
pnad.v2032 <- pnad.peso$V2032
posse.auto <- ifelse(pnad.v2032 == 2, 1, -1)
posse.auto <- ifelse(pnad.v2032 == 4, 1, -1)
posse.auto <- ifelse(pnad.v2032 == 6, 1, -1)
posse.auto <- ifelse(pnad.v2032 == 8, 1, -1)

```

```

# Vive em companhia de c4njugue ou companheiro(a)
pnad.v4111 <- pnad.peso$V4111
vive.conjugue <- ifelse(pnad.v4111 == 1, 1, -1)
#cbind.data.frame(vive.conjugue.1, vive.conjugue.2, vive.conjugue.3)

# Condi4o4o de ocupa4o4o do domic4lio
pnad.v0207 <- pnad.peso$V0207
ocup.dom.1 <- ifelse(pnad.v0207 == 1, 1, -1)
ocup.dom.2 <- ifelse(pnad.v0207 == 3, 1, -1)
ocup.dom.3 <-
ifelse(pnad.v0207 == 2 |
pnad.v0207 == 4 | pnad.v0207 == 5 | pnad.v0207 == 6,
1,
-1)
ocup.dom <- cbind.data.frame(ocup.dom.1, ocup.dom.2, ocup.dom.3)

# Tipo de esgotamento sanit4rio
pnad.v0217 <- pnad.peso$V0217
pnad.v0217[pnad.v0217 == 0] <- NA

tipo.esgoto <- ifelse(pnad.v0217 == 1, 1, -1)

# Frequencia a escola
pnad.v0602 <- pnad.peso$V060
freq.escola <- ifelse(pnad.v0602 == 2, 1, -1)

# Posi4o4o na ocupa4o4o
pnad.v4706 <- pnad.peso$V4706
pnad.v4706[pnad.v4706 == 0] <- NA
pnad.v4706[pnad.v4706 == 6] <- 1
pnad.v4706[pnad.v4706 == 7] <- 4
pnad.v4706[pnad.v4706 == 9] <- 5
pnad.v4706[pnad.v4706 == 10] <- 6
pnad.v4706[pnad.v4706 == 11 | pnad.v4706 == 12 |
pnad.v4706 == 13] <- 7

pos.ocup.1 <- ifelse(pnad.v4706 == 1, 1, -1)

```

```

pos.ocup.2 <- ifelse(pnad.v4706 == 2, 1, -1)
pos.ocup.3 <- ifelse(pnad.v4706 == 3, 1, -1)
pos.ocup.4 <- ifelse(pnad.v4706 == 4, 1, -1)
pos.ocup.5 <- ifelse(pnad.v4706 == 5, 1, -1)
pos.ocup.6 <- ifelse(pnad.v4706 == 6, 1, -1)
pos.ocup.7 <- ifelse(pnad.v4706 == 7, 1, -1)

pos.ocup <-
cbind.data.frame(pos.ocup.1,
pos.ocup.2,
pos.ocup.3,
pos.ocup.4,
pos.ocup.5,
pos.ocup.6,
pos.ocup.7)

# Variáveis
var.pnad <- cbind(n.comodos,n.dormitorios,n.banheiros,idade.anos,cri_ate7,vive.conj

renda <- as.data.frame(pnad.peso$V4621)
names(renda) <- "renda"

# Banco PNAD

banco.pnad <-
cbind.data.frame(
renda,
pnad.peso$V4611,
pnad.peso$V0102,
pnad.peso$V0103,
pnad.peso$UF,
pnad.peso$V4107,
pnad.peso$V4618,
pnad.peso$V4617,
pnad.peso$V4610,
pnad.peso$V4609,
var.pnad)

# Renomeando Variáveis

```

```

names(banco.pnad)[names(banco.pnad) == 'pnad.peso$V4611'] <- 'V4611'
names(banco.pnad)[names(banco.pnad) == 'pnad.peso$V0102'] <- 'V0102'
names(banco.pnad)[names(banco.pnad) == 'pnad.peso$V0103'] <- 'V0103'
names(banco.pnad)[names(banco.pnad) == 'pnad.peso$UF'] <- 'UF'
names(banco.pnad)[names(banco.pnad) == 'pnad.peso$V4107'] <- 'V4107'
names(banco.pnad)[names(banco.pnad) == 'pnad.peso$V4618'] <- 'V4618'
names(banco.pnad)[names(banco.pnad) == 'pnad.peso$V4617'] <- 'V4617'
names(banco.pnad)[names(banco.pnad) == 'pnad.peso$V4610'] <- 'V4610'
names(banco.pnad)[names(banco.pnad) == 'pnad.peso$V4609'] <- 'V4609'

```

```

# Separando treino / teste

```

```

upa <- as.data.frame(table(banco.pnad$V4618))

```

```

names(upa) <- c('upa', 'freq')

```

```

upa$treino <- round(upa$freq*0.8)

```

```

upa$teste <- round(upa$freq*0.2)

```

```

treino <- as.data.frame(setNames(
replicate(57,numeric(0), simplify = F),
c(
"renda",
"V4611",
"V0102",
"V0103",
"UF",
"V4107",
"V4618",
"V4617",
"V4610",
"V4609",
"n.comodos.1",
"n.comodos.2",
"n.comodos.3",
"n.dormitorios.1",
"n.dormitorios.2",
"n.dormitorios.3",
"n.dormitorios.4",
"n.banheiros.1",
"n.banheiros.2",
"n.banheiros.3",

```

```

"idade.anos",
"cri_ate7",
"vive.conjugue",
"cri_7a14",
"n.idoso",
"tipofam.1",
"tipofam.2",
"tipofam.3",
"possui.banheiro",
"sexo",
"raca.1",
"raca.2",
"raca.3",
"raca.4",
"posse.radio",
"posse.tv",
"posse.maqlavar",
"posse.geladeira",
"posse.computador",
"posse.auto",
"ocup.dom.1",
"ocup.dom.2",
"ocup.dom.3",
"tipo.esgoto",
"freq.escola",
"pos.ocup.1",
"pos.ocup.2",
"pos.ocup.3",
"pos.ocup.4",
"pos.ocup.5",
"pos.ocup.6",
"pos.ocup.7"
)
))

for(i in 1:nrow(upa)){
  amostra <- sample_n(banco.pnad[banco.pnad$V4618==which(upa$upa==i)],size=upa$trein)
  treino <- rbind.data.frame(treino,amostra)
}

```

```

# sum(upa$treino)

teste <- banco.pnad[-as.numeric(row.names(treino)),]

# sum(upa$teste)
#base.pnad.treino <- treino
#base.pnad.teste <- teste
on.exit(dbDisconnect(con))

```

## C.4 Modelo Não Linear Fraco - Abordagem Clássica

```

library('RMySQL')
library('survey')
library('MASS')
library('som')
library('xtable')
library('Metrics')
library('neuralnet')

# source(
#   '~/Documents/particular/dados_censo/script_censo/construcao_base_modelagem_pnad
# )

boxplot(treino$renda,xlab='Variável Renda')
summary(treino$renda)
boxplot(log(treino$renda),xlab='Variável log(Renda)')
summary(log(treino$renda))

# Excluindo Valores faltantes das bases
base.pnad.treino <- na.exclude(treino[treino$renda<60000 & treino$renda>5000,])
base.pnad.teste <- na.exclude(teste[teste$renda<60000 & teste$renda>5000,])
basao <- rbind.data.frame(base.pnad.treino,base.pnad.teste)

boxplot(basao$renda,xlab='Variável Renda')
summary(basao$renda)

```

```

boxplot(log(basao$renda),xlab='Variável log(Renda)')
summary(log(basao$renda))

# Excluindo outlier
summary(log(basao$renda))
boxplot(log(basao$renda))

# Modelo de regressão linear clássica sem intercepto
lm.renda.c <- lm(
normalize(log(renda)) ~
n.comodos.1 +
n.comodos.2 +
n.dormitorios.1 +
n.dormitorios.2 +
n.dormitorios.3 +
#n.banheiros.1 +
#n.banheiros.2 +
normalize(idade.anos) +
cri_ate7 +
cri_7a14 +
n.idoso +
tipofam.1 +
tipofam.2 +
sexo +
raca.1 +
raca.2 +
raca.3 +
posse.radio +
posse.tv +
posse.maqlavar +
#posse.geladeira +
posse.computador +
posse.auto +
#vive.conjuge +
ocup.dom.1 +
ocup.dom.2 +
tipo.esgoto +
freq.escola - 1,

```

```

#pos.ocup.1 +
#pos.ocup.2 +
#pos.ocup.3 +
#pos.ocup.4 +
#pos.ocup.5 +
#pos.ocup.6,
data = basao
)

summary(lm.renda.c)

coef.lm <- lm.renda.c$coefficients

# Parte não-linear
# Definição dos parâmetros da rede #
replica <- 1
pop.ref <- 1
n.epocas <- 300
v.alfa <- 0.01
c.sqe <- 0.00001
#c.coef <- 0.0001

#w <- as.matrix(basao$V4611 / sum(basao$V4611) * nrow(basao))

w <- matrix(data = 1,
nrow = nrow(basao),
ncol = 1)

m001 <- (matrix(nrow = replica, ncol = 107))

for (q in 1:pop.ref) {
y <- as.matrix((normalize(log(basao$renda))))
z <- cbind.data.frame(
basao$n.comodos.1,
basao$n.comodos.2,
basao$n.dormitorios.1,
basao$n.dormitorios.2,
basao$n.dormitorios.3,
#basao$n.banheiros.1,

```



```

#basao$n.banheiros.2,
normalize(basao$idade.anos),
basao$cri_ate7,
basao$cri_7a14,
basao$n.idoso,
basao$tipofam.1,
basao$tipofam.2,
basao$sexo,
basao$raca.1,
basao$raca.2,
basao$raca.3,
basao$posse.radio,
basao$posse.tv,
basao$posse.maqlavar,
#basao$posse.geladeira,
basao$posse.computador,
basao$posse.auto,
#basao$vive.conjugue,
basao$ocup.dom.1,
basao$ocup.dom.2,
basao$tipo.esgoto,
basao$freq.escola
# basao$pos.ocup.1,
# basao$pos.ocup.2,
# basao$pos.ocup.3,
# basao$pos.ocup.4,
# basao$pos.ocup.5,
# basao$pos.ocup.6
)

for (u in 1:replica) {
r <- 0
rep <- 0
#repeat {
rep <- rep+1
B0 <- runif(1, min = 0.01, max = 0.011)
B1 <- 1
B2 <- runif(1, min = 0.01, max = B1)

```

```
B3 <- runif(1, min = -0.011, max = B2)
B4 <- runif(1, min = -0.011, max = B3)
```

```
b02 <- runif(1, min = -0.011, max = 0.011)
b21 <- runif(1, min = -0.011, max = 0.011)
b22 <- runif(1, min = -0.011, max = 0.011)
b23 <- runif(1, min = -0.011, max = 0.011)
b24 <- runif(1, min = -0.011, max = 0.011)
b25 <- runif(1, min = -0.011, max = 0.011)
b26 <- runif(1, min = -0.011, max = 0.011)
b27 <- runif(1, min = -0.011, max = 0.011)
b28 <- runif(1, min = -0.011, max = 0.011)
b29 <- runif(1, min = -0.011, max = 0.011)
b210 <- runif(1, min = -0.011, max = 0.011)
b211 <- runif(1, min = -0.011, max = 0.011)
b212 <- runif(1, min = -0.011, max = 0.011)
b213 <- runif(1, min = -0.011, max = 0.011)
b214 <- runif(1, min = -0.011, max = 0.011)
b215 <- runif(1, min = -0.011, max = 0.011)
b216 <- runif(1, min = -0.011, max = 0.011)
b217 <- runif(1, min = -0.011, max = 0.011)
b218 <- runif(1, min = -0.011, max = 0.011)
b219 <- runif(1, min = -0.011, max = 0.011)
b220 <- runif(1, min = -0.011, max = 0.011)
b221 <- runif(1, min = -0.011, max = 0.011)
b222 <- runif(1, min = -0.011, max = 0.011)
b223 <- runif(1, min = -0.011, max = 0.011)
b224 <- runif(1, min = -0.011, max = 0.011)
```

```
b03 <- runif(1, min = -0.011, max = 0.011)
b31 <- runif(1, min = -0.011, max = 0.011)
b32 <- runif(1, min = -0.011, max = 0.011)
b33 <- runif(1, min = -0.011, max = 0.011)
b34 <- runif(1, min = -0.011, max = 0.011)
b35 <- runif(1, min = -0.011, max = 0.011)
b36 <- runif(1, min = -0.011, max = 0.011)
b37 <- runif(1, min = -0.011, max = 0.011)
```

```
b38 <- runif(1, min = -0.011, max = 0.011)
b39 <- runif(1, min = -0.011, max = 0.011)
b310 <- runif(1, min = -0.011, max = 0.011)
b311 <- runif(1, min = -0.011, max = 0.011)
b312 <- runif(1, min = -0.011, max = 0.011)
b313 <- runif(1, min = -0.011, max = 0.011)
b314 <- runif(1, min = -0.011, max = 0.011)
b315 <- runif(1, min = -0.011, max = 0.011)
b316 <- runif(1, min = -0.011, max = 0.011)
b317 <- runif(1, min = -0.011, max = 0.011)
b318 <- runif(1, min = -0.011, max = 0.011)
b319 <- runif(1, min = -0.011, max = 0.011)
b320 <- runif(1, min = -0.011, max = 0.011)
b321 <- runif(1, min = -0.011, max = 0.011)
b322 <- runif(1, min = -0.011, max = 0.011)
b323 <- runif(1, min = -0.011, max = 0.011)
b324 <- runif(1, min = -0.011, max = 0.011)
```

```
b04 <- runif(1, min = -0.011, max = 0.011)
b41 <- runif(1, min = -0.011, max = 0.011)
b42 <- runif(1, min = -0.011, max = 0.011)
b43 <- runif(1, min = -0.011, max = 0.011)
b44 <- runif(1, min = -0.011, max = 0.011)
b45 <- runif(1, min = -0.011, max = 0.011)
b46 <- runif(1, min = -0.011, max = 0.011)
b47 <- runif(1, min = -0.011, max = 0.011)
b48 <- runif(1, min = -0.011, max = 0.011)
b49 <- runif(1, min = -0.011, max = 0.011)
b410 <- runif(1, min = -0.011, max = 0.011)
b411 <- runif(1, min = -0.011, max = 0.011)
b412 <- runif(1, min = -0.011, max = 0.011)
b413 <- runif(1, min = -0.011, max = 0.011)
b414 <- runif(1, min = -0.011, max = 0.011)
b415 <- runif(1, min = -0.011, max = 0.011)
b416 <- runif(1, min = -0.011, max = 0.011)
b417 <- runif(1, min = -0.011, max = 0.011)
b418 <- runif(1, min = -0.011, max = 0.011)
b419 <- runif(1, min = -0.011, max = 0.011)
```

```

b420 <- runif(1, min = -0.011, max = 0.011)
b421 <- runif(1, min = -0.011, max = 0.011)
b422 <- runif(1, min = -0.011, max = 0.011)
b423 <- runif(1, min = -0.011, max = 0.011)
b424 <- runif(1, min = -0.011, max = 0.011)

v      <- matrix(nrow = nrow(y), ncol = 4)
yhat   <- matrix(nrow = nrow(y), ncol = 1)
erro   <- matrix(nrow = nrow(y), ncol = 1)
alfa   <- v.alfa
epocas <- n.epocas
sqe    <- matrix(nrow = epocas, ncol = 1)
#coef  <- matrix(nrow = epocas, ncol = 4)

delta.B0 <- matrix(nrow = nrow(y), ncol = 1)
delta.B2 <- matrix(nrow = nrow(y), ncol = 1)
delta.B3 <- matrix(nrow = nrow(y), ncol = 1)
delta.B4 <- matrix(nrow = nrow(y), ncol = 1)

delta.b02 <- matrix(nrow = nrow(y), ncol = 1)
delta.b21 <- matrix(nrow = nrow(y), ncol = 1)
delta.b22 <- matrix(nrow = nrow(y), ncol = 1)
delta.b23 <- matrix(nrow = nrow(y), ncol = 1)
delta.b24 <- matrix(nrow = nrow(y), ncol = 1)
delta.b25 <- matrix(nrow = nrow(y), ncol = 1)
delta.b26 <- matrix(nrow = nrow(y), ncol = 1)
delta.b27 <- matrix(nrow = nrow(y), ncol = 1)
delta.b28 <- matrix(nrow = nrow(y), ncol = 1)
delta.b29 <- matrix(nrow = nrow(y), ncol = 1)
delta.b210 <- matrix(nrow = nrow(y), ncol = 1)
delta.b211 <- matrix(nrow = nrow(y), ncol = 1)
delta.b212 <- matrix(nrow = nrow(y), ncol = 1)
delta.b213 <- matrix(nrow = nrow(y), ncol = 1)
delta.b214 <- matrix(nrow = nrow(y), ncol = 1)
delta.b215 <- matrix(nrow = nrow(y), ncol = 1)
delta.b216 <- matrix(nrow = nrow(y), ncol = 1)
delta.b217 <- matrix(nrow = nrow(y), ncol = 1)

```

```
delta.b218 <- matrix(nrow = nrow(y), ncol = 1)
delta.b219 <- matrix(nrow = nrow(y), ncol = 1)
delta.b220 <- matrix(nrow = nrow(y), ncol = 1)
delta.b221 <- matrix(nrow = nrow(y), ncol = 1)
delta.b222 <- matrix(nrow = nrow(y), ncol = 1)
delta.b223 <- matrix(nrow = nrow(y), ncol = 1)
delta.b224 <- matrix(nrow = nrow(y), ncol = 1)
```

```
delta.b03 <- matrix(nrow = nrow(y), ncol = 1)
delta.b31 <- matrix(nrow = nrow(y), ncol = 1)
delta.b32 <- matrix(nrow = nrow(y), ncol = 1)
delta.b33 <- matrix(nrow = nrow(y), ncol = 1)
delta.b34 <- matrix(nrow = nrow(y), ncol = 1)
delta.b35 <- matrix(nrow = nrow(y), ncol = 1)
delta.b36 <- matrix(nrow = nrow(y), ncol = 1)
delta.b37 <- matrix(nrow = nrow(y), ncol = 1)
delta.b38 <- matrix(nrow = nrow(y), ncol = 1)
delta.b39 <- matrix(nrow = nrow(y), ncol = 1)
delta.b310 <- matrix(nrow = nrow(y), ncol = 1)
delta.b311 <- matrix(nrow = nrow(y), ncol = 1)
delta.b312 <- matrix(nrow = nrow(y), ncol = 1)
delta.b313 <- matrix(nrow = nrow(y), ncol = 1)
delta.b314 <- matrix(nrow = nrow(y), ncol = 1)
delta.b315 <- matrix(nrow = nrow(y), ncol = 1)
delta.b316 <- matrix(nrow = nrow(y), ncol = 1)
delta.b317 <- matrix(nrow = nrow(y), ncol = 1)
delta.b318 <- matrix(nrow = nrow(y), ncol = 1)
delta.b319 <- matrix(nrow = nrow(y), ncol = 1)
delta.b320 <- matrix(nrow = nrow(y), ncol = 1)
delta.b321 <- matrix(nrow = nrow(y), ncol = 1)
delta.b322 <- matrix(nrow = nrow(y), ncol = 1)
delta.b323 <- matrix(nrow = nrow(y), ncol = 1)
delta.b324 <- matrix(nrow = nrow(y), ncol = 1)
```

```
delta.b04 <- matrix(nrow = nrow(y), ncol = 1)
delta.b41 <- matrix(nrow = nrow(y), ncol = 1)
delta.b42 <- matrix(nrow = nrow(y), ncol = 1)
```

```

delta.b43 <- matrix(nrow = nrow(y), ncol = 1)
delta.b44 <- matrix(nrow = nrow(y), ncol = 1)
delta.b45 <- matrix(nrow = nrow(y), ncol = 1)
delta.b46 <- matrix(nrow = nrow(y), ncol = 1)
delta.b47 <- matrix(nrow = nrow(y), ncol = 1)
delta.b48 <- matrix(nrow = nrow(y), ncol = 1)
delta.b49 <- matrix(nrow = nrow(y), ncol = 1)
delta.b410 <- matrix(nrow = nrow(y), ncol = 1)
delta.b411 <- matrix(nrow = nrow(y), ncol = 1)
delta.b412 <- matrix(nrow = nrow(y), ncol = 1)
delta.b413 <- matrix(nrow = nrow(y), ncol = 1)
delta.b414 <- matrix(nrow = nrow(y), ncol = 1)
delta.b415 <- matrix(nrow = nrow(y), ncol = 1)
delta.b416 <- matrix(nrow = nrow(y), ncol = 1)
delta.b417 <- matrix(nrow = nrow(y), ncol = 1)
delta.b418 <- matrix(nrow = nrow(y), ncol = 1)
delta.b419 <- matrix(nrow = nrow(y), ncol = 1)
delta.b420 <- matrix(nrow = nrow(y), ncol = 1)
delta.b421 <- matrix(nrow = nrow(y), ncol = 1)
delta.b422 <- matrix(nrow = nrow(y), ncol = 1)
delta.b423 <- matrix(nrow = nrow(y), ncol = 1)
delta.b424 <- matrix(nrow = nrow(y), ncol = 1)

```

```

for (j in 1:epocas) {
  for (i in 1:nrow(y)) {
    v[i, 1] <- (
      #lm.renda.c$coefficients[1] +
      lm.renda.c$coefficients[1] * z[i, 1] +
      lm.renda.c$coefficients[2] * z[i, 2] +
      lm.renda.c$coefficients[3] * z[i, 3] +
      lm.renda.c$coefficients[4] * z[i, 4] +
      lm.renda.c$coefficients[5] * z[i, 5] +
      lm.renda.c$coefficients[6] * z[i, 6] +
      lm.renda.c$coefficients[7] * z[i, 7] +
      lm.renda.c$coefficients[8] * z[i, 8] +
      lm.renda.c$coefficients[9] * z[i, 9] +
      lm.renda.c$coefficients[10] * z[i, 10] +

```

```

lm.renda.c$coefficients[11] * z[i, 11] +
lm.renda.c$coefficients[12] * z[i, 12] +
lm.renda.c$coefficients[13] * z[i, 13] +
lm.renda.c$coefficients[14] * z[i, 14] +
lm.renda.c$coefficients[15] * z[i, 15] +
lm.renda.c$coefficients[16] * z[i, 16] +
lm.renda.c$coefficients[17] * z[i, 17] +
lm.renda.c$coefficients[18] * z[i, 18] +
lm.renda.c$coefficients[19] * z[i, 19] +
lm.renda.c$coefficients[20] * z[i, 20] +
lm.renda.c$coefficients[21] * z[i, 21] +
lm.renda.c$coefficients[22] * z[i, 22] +
lm.renda.c$coefficients[23] * z[i, 23] +
lm.renda.c$coefficients[24] * z[i, 24]
)

```

```

v[i, 2] <- tanh(
b02 +
b21 * z[i, 1] +
b22 * z[i, 2] + #comodos
b23 * z[i, 3] +
b24 * z[i, 4] +
b25 * z[i, 5] + #dormitorios
b26 * z[i, 6] + #idade anos
b27 * z[i, 7] + #cri_ate7
b28 * z[i, 8] + #cri_7a14
b29 * z[i, 9] + #n.idoso
b210 * z[i, 10] +
b211 * z[i, 11] + #tipofam
b212 * z[i, 12] + #sexo
b213 * z[i, 13] +
b214 * z[i, 14] +
b215 * z[i, 15] + #raça
b216 * z[i, 16] + #posse.radio
b217 * z[i, 17] + #posse.tv
b218 * z[i, 18] + #posse.maqlavar
b219 * z[i, 19] + #posse.computador
b220 * z[i, 20] + #posse.auto
b221 * z[i, 21] +

```

```

b222 * z[i, 22] + #ocup.dom
b223 * z[i, 23] + #tipo.esgoto
b224 * z[i, 24] #freq.escola
)

v[i, 3] <- tanh(
b03 +
b31 * z[i, 1] +
b32 * z[i, 2] + #comodos
b33 * z[i, 3] +
b34 * z[i, 4] +
b35 * z[i, 5] + #dormitorios
b36 * z[i, 6] + #idade anos
b37 * z[i, 7] + #cri_ate7
b38 * z[i, 8] + #cri_7a14
b39 * z[i, 9] + #n.idoso
b310 * z[i, 10] +
b311 * z[i, 11] + #tipofam
b312 * z[i, 12] + #sexo
b313 * z[i, 13] +
b314 * z[i, 14] +
b315 * z[i, 15] + #raça
b316 * z[i, 16] + #posse.radio
b317 * z[i, 17] + #posse.tv
b318 * z[i, 18] + #posse.maqlavar
b319 * z[i, 19] + #posse.computador
b320 * z[i, 20] + #posse.auto
b321 * z[i, 21] +
b322 * z[i, 22] + #ocup.dom
b323 * z[i, 23] + #tipo.esgoto
b324 * z[i, 24] #freq.escola
)

v[i, 4] <- tanh(
b04 +
b41 * z[i, 1] +
b42 * z[i, 2] + #comodos
b43 * z[i, 3] +
b44 * z[i, 4] +

```



```

b45 * z[i, 5] + #dormitorios
b46 * z[i, 6] + #idade anos
b47 * z[i, 7] + #cri_ate7
b48 * z[i, 8] + #cri_7a14
b49 * z[i, 9] + #n.idoso
b410 * z[i, 10] +
b411 * z[i, 11] + #tipofam
b412 * z[i, 12] + #sexo
b413 * z[i, 13] +
b414 * z[i, 14] +
b415 * z[i, 15] + #raça
b416 * z[i, 16] + #posse.radio
b417 * z[i, 17] + #posse.tv
b418 * z[i, 18] + #posse.maqlavar
b419 * z[i, 19] + #posse.computador
b420 * z[i, 20] + #posse.auto
b421 * z[i, 21] +
b422 * z[i, 22] + #ocup.dom
b423 * z[i, 23] + #tipo.esgoto
b424 * z[i, 24] #freq.escola
)

```

```

yhat[i, ] <-
(B0 + B1 * v[i, 1] + B2 * v[i, 2] + B3 * v[i, 3] + B4 * v[i, 4])
erro[i, ] <- y[i, ] - yhat[i, ]

```

```

delta.B0[i, ] <- alfa * erro[i, ] * w[i, ]
delta.B2[i, ] <- alfa * erro[i, ] * v[i, 2] * w[i, ]
delta.B3[i, ] <- alfa * erro[i, ] * v[i, 3] * w[i, ]
delta.B4[i, ] <- alfa * erro[i, ] * v[i, 4] * w[i, ]

```

```

delta.b02[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * w[i, ]
delta.b21[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 1] * w[i, ]
delta.b22[i, ] <-

```

```

alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 2] * w[i, ]
delta.b23[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 3] * w[i, ]
delta.b24[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 4] * w[i, ]
delta.b25[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 5] * w[i, ]
delta.b26[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 6] * w[i, ]
delta.b27[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 7] * w[i, ]
delta.b28[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 8] * w[i, ]
delta.b29[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 9] * w[i, ]
delta.b210[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 10] * w[i, ]
delta.b211[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 11] * w[i, ]
delta.b212[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 12] * w[i, ]
delta.b213[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 13] * w[i, ]
delta.b214[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 14] * w[i, ]
delta.b215[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 15] * w[i, ]
delta.b216[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 16] * w[i, ]
delta.b217[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 17] * w[i, ]
delta.b218[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 18] * w[i, ]
delta.b219[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 19] * w[i, ]
delta.b220[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 20] * w[i, ]
delta.b221[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 21] * w[i, ]

```

```

delta.b222[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 22] * w[i, ]
delta.b223[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 23] * w[i, ]
delta.b224[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 24] * w[i, ]

delta.b03[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * w[i, ]
delta.b31[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 1] * w[i, ]
delta.b32[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 2] * w[i, ]
delta.b33[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 3] * w[i, ]
delta.b34[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 4] * w[i, ]
delta.b35[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 5] * w[i, ]
delta.b36[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 6] * w[i, ]
delta.b37[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 7] * w[i, ]
delta.b38[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 8] * w[i, ]
delta.b39[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 9] * w[i, ]
delta.b310[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 10] * w[i, ]
delta.b311[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 11] * w[i, ]
delta.b312[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 12] * w[i, ]
delta.b313[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 13] * w[i, ]
delta.b314[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 14] * w[i, ]
delta.b315[i, ] <-

```

```

alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 15] * w[i, ]
delta.b316[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 16] * w[i, ]
delta.b317[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 17] * w[i, ]
delta.b318[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 18] * w[i, ]
delta.b319[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 19] * w[i, ]
delta.b320[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 20] * w[i, ]
delta.b321[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 21] * w[i, ]
delta.b322[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 22] * w[i, ]
delta.b323[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 23] * w[i, ]
delta.b324[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 24] * w[i, ]

delta.b04[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * w[i, ]
delta.b41[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 1] * w[i, ]
delta.b42[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 2] * w[i, ]
delta.b43[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 3] * w[i, ]
delta.b44[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 4] * w[i, ]
delta.b45[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 5] * w[i, ]
delta.b46[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 6] * w[i, ]
delta.b47[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 7] * w[i, ]
delta.b48[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 8] * w[i, ]
delta.b49[i, ] <-

```

```

alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 9] * w[i, ]
delta.b410[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 10] * w[i, ]
delta.b411[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 11] * w[i, ]
delta.b412[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 12] * w[i, ]
delta.b413[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 13] * w[i, ]
delta.b414[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 14] * w[i, ]
delta.b415[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 15] * w[i, ]
delta.b416[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 16] * w[i, ]
delta.b417[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 17] * w[i, ]
delta.b418[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 18] * w[i, ]
delta.b419[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 19] * w[i, ]
delta.b420[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 20] * w[i, ]
delta.b421[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 21] * w[i, ]
delta.b422[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 22] * w[i, ]
delta.b423[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 23] * w[i, ]
delta.b424[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 24] * w[i, ]
}

```

```
B0 <- abs(B0 + colMeans(delta.B0))
```

```
B2 <- abs(B2 + colMeans(delta.B2))
```

```
B3 <- abs(B3 + colMeans(delta.B3))
```

```
B4 <- abs(B2 + colMeans(delta.B4))
```

```
b02 <- abs(b02 + colMeans(delta.b02))
b21 <- abs(b21 + colMeans(delta.b21))
b22 <- abs(b22 + colMeans(delta.b22))
b23 <- abs(b23 + colMeans(delta.b23))
b24 <- abs(b24 + colMeans(delta.b24))
b25 <- abs(b25 + colMeans(delta.b25))
b26 <- abs(b26 + colMeans(delta.b26))
b27 <- abs(b27 + colMeans(delta.b27))
b28 <- abs(b28 + colMeans(delta.b28))
b29 <- abs(b29 + colMeans(delta.b29))
b210 <- abs(b210 + colMeans(delta.b210))
b211 <- abs(b211 + colMeans(delta.b211))
b212 <- abs(b212 + colMeans(delta.b212))
b213 <- abs(b213 + colMeans(delta.b213))
b214 <- abs(b214 + colMeans(delta.b214))
b215 <- abs(b215 + colMeans(delta.b215))
b216 <- abs(b216 + colMeans(delta.b216))
b217 <- abs(b217 + colMeans(delta.b217))
b218 <- abs(b218 + colMeans(delta.b218))
b219 <- abs(b219 + colMeans(delta.b219))
b220 <- abs(b220 + colMeans(delta.b220))
b221 <- abs(b221 + colMeans(delta.b221))
b222 <- abs(b222 + colMeans(delta.b222))
b223 <- abs(b223 + colMeans(delta.b223))
b224 <- abs(b224 + colMeans(delta.b224))
```

```
b03 <- abs(b03 + colMeans(delta.b03))
b31 <- abs(b31 + colMeans(delta.b31))
b32 <- abs(b32 + colMeans(delta.b32))
b33 <- abs(b33 + colMeans(delta.b33))
b34 <- abs(b34 + colMeans(delta.b34))
b35 <- abs(b35 + colMeans(delta.b35))
b36 <- abs(b36 + colMeans(delta.b36))
b37 <- abs(b37 + colMeans(delta.b37))
b38 <- abs(b38 + colMeans(delta.b38))
b39 <- abs(b39 + colMeans(delta.b39))
b310 <- abs(b310 + colMeans(delta.b310))
b311 <- abs(b311 + colMeans(delta.b311))
```

```
b312 <- abs(b312 + colMeans(delta.b312))
b313 <- abs(b313 + colMeans(delta.b313))
b314 <- abs(b314 + colMeans(delta.b314))
b315 <- abs(b315 + colMeans(delta.b315))
b316 <- abs(b316 + colMeans(delta.b316))
b317 <- abs(b317 + colMeans(delta.b317))
b318 <- abs(b318 + colMeans(delta.b318))
b319 <- abs(b319 + colMeans(delta.b319))
b320 <- abs(b320 + colMeans(delta.b320))
b321 <- abs(b321 + colMeans(delta.b321))
b322 <- abs(b322 + colMeans(delta.b322))
b323 <- abs(b323 + colMeans(delta.b323))
b324 <- abs(b324 + colMeans(delta.b324))
```

```
b04 <- abs(b04 + colMeans(delta.b04))
b41 <- abs(b41 + colMeans(delta.b41))
b42 <- abs(b42 + colMeans(delta.b42))
b43 <- abs(b43 + colMeans(delta.b43))
b44 <- abs(b44 + colMeans(delta.b44))
b45 <- abs(b45 + colMeans(delta.b45))
b46 <- abs(b46 + colMeans(delta.b46))
b47 <- abs(b47 + colMeans(delta.b47))
b48 <- abs(b48 + colMeans(delta.b48))
b49 <- abs(b49 + colMeans(delta.b49))
b410 <- abs(b410 + colMeans(delta.b410))
b411 <- abs(b411 + colMeans(delta.b411))
b412 <- abs(b412 + colMeans(delta.b412))
b413 <- abs(b413 + colMeans(delta.b413))
b414 <- abs(b414 + colMeans(delta.b414))
b415 <- abs(b415 + colMeans(delta.b415))
b416 <- abs(b416 + colMeans(delta.b416))
b417 <- abs(b417 + colMeans(delta.b417))
b418 <- abs(b418 + colMeans(delta.b418))
b419 <- abs(b419 + colMeans(delta.b419))
b420 <- abs(b420 + colMeans(delta.b420))
b421 <- abs(b421 + colMeans(delta.b421))
b422 <- abs(b422 + colMeans(delta.b422))
b423 <- abs(b423 + colMeans(delta.b423))
```

```

b424 <- abs(b424 + colMeans(delta.b424))

sqe[j, ] <- sum(erro ^ 2) / 2
# coef[j,1] <- b0
# coef[j,2] <- b1
# coef[j,3] <- B0
# coef[j,4] <- B1

if (j > 1) {
v.sqe <- abs((sqe[j - 1, ] - sqe[j, ]) / sqe[j - 1])
# d.coef.b0 <- abs((coef[j,1]-coef[j-1,1]))
# d.coef.b1 <- abs((coef[j,2]-coef[j-1,2]))
# d.coef.B0 <- abs((coef[j,3]-coef[j-1,3]))
# d.coef.B1 <- abs((coef[j,4]-coef[j-1,4]))
if ((v.sqe <= c.sqe) || j == n.epocas) {
break
}
}
if(j<=1) {
cat(paste("épocas:", j, "||", "sqe:", sum(erro ^ 2) / 2,"||", "repetição:",r), sep =
} else {cat(paste("épocas:", j, "||", "sqe:", sum(erro ^ 2) / 2,"||", "v.sqe:",v.sqe
}
}

m001[u, 1] <- q
m001[u, 2] <- u
m001[u, 3] <- B0
m001[u, 4] <- B1
m001[u, 5] <- B2

m001[u, 6] <- lm.renda.c$coefficients[1]
m001[u, 7] <- lm.renda.c$coefficients[2]
m001[u, 8] <- lm.renda.c$coefficients[3]
m001[u, 9] <- lm.renda.c$coefficients[4]
m001[u, 10] <- lm.renda.c$coefficients[5]
m001[u, 11] <- lm.renda.c$coefficients[6]
m001[u, 12] <- lm.renda.c$coefficients[7]
m001[u, 13] <- lm.renda.c$coefficients[8]
m001[u, 14] <- lm.renda.c$coefficients[9]

```



```
m001[u, 15] <- lm.renda.c$coefficients[10]
m001[u, 16] <- lm.renda.c$coefficients[11]
m001[u, 17] <- lm.renda.c$coefficients[12]
m001[u, 18] <- lm.renda.c$coefficients[13]
m001[u, 19] <- lm.renda.c$coefficients[14]
m001[u, 20] <- lm.renda.c$coefficients[15]
m001[u, 21] <- lm.renda.c$coefficients[16]
m001[u, 22] <- lm.renda.c$coefficients[17]
m001[u, 23] <- lm.renda.c$coefficients[18]
m001[u, 24] <- lm.renda.c$coefficients[19]
m001[u, 25] <- lm.renda.c$coefficients[20]
m001[u, 26] <- lm.renda.c$coefficients[21]
m001[u, 27] <- lm.renda.c$coefficients[22]
m001[u, 28] <- lm.renda.c$coefficients[23]
m001[u, 29] <- lm.renda.c$coefficients[24]
#m001[u, 30] <- lm.renda.c$coefficients[25]
```

```
m001[u, 31] <- b02
m001[u, 32] <- b21
m001[u, 33] <- b22
m001[u, 34] <- b23
m001[u, 35] <- b24
m001[u, 36] <- b25
m001[u, 37] <- b26
m001[u, 38] <- b27
m001[u, 39] <- b28
m001[u, 40] <- b29
m001[u, 41] <- b210
m001[u, 42] <- b211
m001[u, 43] <- b212
m001[u, 44] <- b213
m001[u, 45] <- b214
m001[u, 46] <- b215
m001[u, 47] <- b216
m001[u, 48] <- b217
m001[u, 49] <- b218
m001[u, 50] <- b219
m001[u, 51] <- b220
m001[u, 52] <- b221
```

```
m001[u, 53] <- b222
m001[u, 54] <- b223
m001[u, 55] <- b224

m001[u, 56] <- b03
m001[u, 57] <-      b31
m001[u, 58] <-      b32
m001[u, 59] <-      b33
m001[u, 60] <-      b34
m001[u, 61] <-      b35
m001[u, 62] <-      b36
m001[u, 63] <-      b37
m001[u, 64] <-      b38
m001[u, 65] <-      b39
m001[u, 66] <-      b310
m001[u, 67] <-      b311
m001[u, 68] <-      b312
m001[u, 69] <-      b313
m001[u, 70] <-      b314
m001[u, 71] <-      b315
m001[u, 72] <-      b316
m001[u, 73] <-      b317
m001[u, 74] <-      b318
m001[u, 75] <-      b319
m001[u, 76] <-      b320
m001[u, 77] <-      b321
m001[u, 78] <-      b322
m001[u, 79] <-      b323
m001[u, 80] <-      b324

m001[u, 81] <-      b04
m001[u, 82] <-      b41
m001[u, 83] <-      b42
m001[u, 84] <-      b43
m001[u, 85] <-      b44
m001[u, 86] <-      b45
m001[u, 87] <-      b46
m001[u, 88] <-      b47
```

```

m001[u, 89] <-      b48
m001[u, 90] <-      b49
m001[u, 91] <-      b410
m001[u, 92] <-      b411
m001[u, 93] <-      b412
m001[u, 94] <-      b413
m001[u, 95] <-      b414
m001[u, 96] <-      b415
m001[u, 97] <-      b416
m001[u, 98] <-      b417
m001[u, 99] <-      b418
m001[u, 100] <-     b419
m001[u, 101] <-     b420
m001[u, 102] <-     b421
m001[u, 103] <-     b422
m001[u, 104] <-     b423
m001[u, 105] <-     b424

```

```

m001[u, 106] <- v.sqe
m001[u, 107] <- j

```

```

cat(
paste(
"Replicação:",
u,
"Repetição:",
r <-
r + 1,
"époças:",
j,
"MSE:",
round(mean(na.exclude(sqe)), 2),
sep = " "
),
sep = "\n"
)
cat(paste(
"B0:",
round(B0, 4),

```

```

"B1:",
round(B1, 4),
"B2:",
round(B2, 4),
"B3:",
round(B3, 4),
"B4:",
round(B4, 4)
),
sep = '\n')

#if (B1 >= B2 & B2 >= B3 & B3 >= B4) {
#break
#}
#}

}
}

pop.rna.tanh.2 <- as.data.frame(m001)
names(pop.rna.tanh.2) <-
c(
'pop',
'replica',
'B0',
'B1',
'B2',
'b01',
'b11',
'b12',
'b13',
'b14',
'b15',
'b16',
'b17',
'b18',
'b19',
'b110',
'b111',

```

'b112',  
'b113',  
'b114',  
'b115',  
'b116',  
'b117',  
'b118',  
'b119',  
'b120',  
'b121',  
'b122',  
'b123',  
'b124',  
'b125',  
'b02',  
'b21',  
'b22',  
'b23',  
'b24',  
'b25',  
'b26',  
'b27',  
'b28',  
'b29',  
'b210',  
'b211',  
'b212',  
'b213',  
'b214',  
'b215',  
'b216',  
'b217',  
'b218',  
'b219',  
'b220',  
'b221',  
'b222',  
'b223',  
'b224',

'b03',  
'b31',  
'b32',  
'b33',  
'b34',  
'b35',  
'b36',  
'b37',  
'b38',  
'b39',  
'b310',  
'b311',  
'b312',  
'b313',  
'b314',  
'b315',  
'b316',  
'b317',  
'b318',  
'b319',  
'b320',  
'b321',  
'b322',  
'b323',  
'b324',  
'b04',  
'b41',  
'b42',  
'b43',  
'b44',  
'b45',  
'b46',  
'b47',  
'b48',  
'b49',  
'b410',  
'b411',  
'b412',  
'b413',

```

'b414',
'b415',
'b416',
'b417',
'b418',
'b419',
'b420',
'b421',
'b422',
'b423',
'b424',
'v.sqe',
'epoca'
)

write.table(
pop.rna.tanh.2,
file = '/home/savano/Documents/particular/dados_censo/rn_censo.txt',
quote = FALSE,
sep = "|",
row.names = FALSE
)

plot(
sqe,
type = 'l',
xlab = paste(j, 'épocas'),
xlim = c(1, j),
ylab = 'SQE'
)
plot(erro, ylab = 'Resíduos')

nf = layout(matrix(c(1, 2, 3, 4), 2, 2, byrow = TRUE))
plot(erro,
ylab = 'Resíduos',
cex = .3,
main = 'Dispersão')
hist(erro,

```

```

ylab = 'Frequência',
xlab = 'Resíduos',
main = 'Histograma')
boxplot(erro, xlab = 'Resíduos', main = 'Box-plot')
qqnorm(erro, ylab = 'Quantis Amostrais', xlab = 'Quantis Teóricos')
qqline(erro)

y.nn <- y*sd(log(basao$renda))+mean(log(basao$renda))
yhat.nn <- yhat*sd(log(basao$renda))+mean(log(basao$renda))

ssr.logrenda.treino.rna <- sum((exp(y.nn)-exp(yhat.nn))^2*basao$V4611)/sum(basao$V4

sqr.rna.treino <- sum((exp(y.nn)-exp(yhat.nn))^2*basao$V4611)
sqt.rna.treino <- sum((basao$renda-mean(basao$renda))^2*basao$V4611)

r.2.rna.treino <- (1-sqr.rna.treino/sqt.rna.treino)*100

lm.renda.c.linear <- lm(
log(renda) ~
n.comodos.1 +
n.comodos.2 +
n.dormitorios.1 +
n.dormitorios.2 +
n.dormitorios.3 +
#n.banheiros.1 +
#n.banheiros.2 +
idade.anos +
cri_ate7 +
cri_7a14 +
n.idoso +
tipofam.1 +
tipofam.2 +
sexo +
raca.1 +
raca.2 +
raca.3 +
posse.radio +
posse.tv +
posse.maqlavar +

```



```

#posse.geladeira +
posse.computador +
posse.auto +
#vive.conjugue +
ocup.dom.1 +
ocup.dom.2 +
tipo.esgoto +
freq.escola,
#pos.ocup.1 +
#pos.ocup.2 +
#pos.ocup.3 +
#pos.ocup.4 +
#pos.ocup.5 +
#pos.ocup.6,
data = basao
)

summary(lm.renda.c.linear)

ssr.logrenda.treino.lm.linear <- sum((basao$renda-exp(lm.renda.c.linear$fitted.val

sqr.linear.treino <- sum((basao$renda-exp(lm.renda.c.linear$fitted.values)) ^2*basao
sqt.linear.treino <- sum((basao$renda-mean(basao$renda)) ^2*basao$V4611)

r.2.linear.treino <- (1-sqr.linear.treino/sqt.linear.treino)*100

```

## C.5 Modelo Não Linear Fraco - Abordagem de Superpopulação

```

library('RMySQL')
library('survey')
library('MASS')
library('som')
library('xtable')
library('Metrics')
library('neuralnet')

```

```

# source(
#   '~/Documents/particular/dados_censo/script_censo/construcao_base_modelagem_pnad
# )

# Excluindo Valores faltantes das bases
base.pnad.treino <- na.exclui(trreino[treino$renda<60000 & treino$renda>5000,])
base.pnad.teste <- na.exclui(teste[teste$renda<60000 & teste$renda>5000,])

# Excluindo outlier
summary(log(base.pnad.treino$renda))
boxplot(log(base.pnad.treino$renda))

# Plano amostral da PNAD com pesos do desenho e dois estagios
base.pnad.treino$dom <- base.pnad.treino$V0102 * 1000 + base.pnad.treino$V0103

base.pnad.treino$UF.reg[base.pnad.treino$UF == 31 & base.pnad.treino$V4107 == 1] <-
base.pnad.treino$UF.reg[base.pnad.treino$UF == 31 & base.pnad.treino$V4107 > 1] <-

# Cria objeto com informações da estrutura do plano amostral básico da PNAD 2014
d.pnad <-
svydesign(
ids = ~ V4618 + dom ,
strata = ~ V4617 ,
weights = ~ V4610 ,
nest = T,
data = base.pnad.treino
)

# Localiza valores das projeções de população das regiões do estado de MG
proj_UF.reg <- unique(base.pnad.treino$V4609)
tot_UF.reg_r <- as.data.frame(cbind(1:2 , proj_UF.reg))
UF.reg <- as.data.frame(base.pnad.treino$UF.reg)
names(tot_UF.reg_r) <- c('UF.reg', 'Freq')

## Cria objeto contendo dados e pesos calibrados da PNAD 2014
d.rak <- postStratify(d.pnad , ~ UF.reg, population = tot_UF.reg_r)

# Recupera pesos calibrados dos domicílios
# Tem que coincidir com valores da variável V4611

```

```

pesos.calib <- weights(d.rak)

# Modelo de regressão linear clássica sem intercepto
lm.renda.c <- svyglm(
normalize(log(renda)) ~
n.comodos.1 +
n.comodos.2 +
n.dormitorios.1 +
n.dormitorios.2 +
n.dormitorios.3 +
#n.banheiros.1 +
#n.banheiros.2 +
normalize(idade.anos) +
cri_ate7 +
cri_7a14 +
n.idoso +
tipofam.1 +
tipofam.2 +
sexo +
raca.1 +
raca.2 +
raca.3 +
posse.radio +
posse.tv +
posse.maqlavar +
#posse.geladeira +
posse.computador +
posse.auto +
#vive.conjuge +
ocup.dom.1 +
ocup.dom.2 +
tipo.esgoto +
freq.escola - 1,
#pos.ocup.1 +
#pos.ocup.2 +
#pos.ocup.3 +
#pos.ocup.4 +
#pos.ocup.5 +
#pos.ocup.6,

```

```

design= d.pnad
)

summary(lm.renda.c)

coef.lm <- lm.renda.c$coefficients

# Parte não-linear
# Definição dos parâmetros da rede #
replica <- 1
pop.ref <- 1
n.epocas <- 300
v.alfa <- 0.01
c.sqe <- 0.00001
#c.coef <- 0.0001

w <- as.matrix(base.pnad.treino$V4611 / sum(base.pnad.treino$V4611) * nrow(base.pna

# w <- matrix(data = 1,
#             nrow = nrow(base.pnad.treino),
#             ncol = 1)

m001 <- (matrix(nrow = replica, ncol = 107))

for (q in 1:pop.ref) {
y <- as.matrix((normalize(log(base.pnad.treino$renda))))
z <- cbind.data.frame(
base.pnad.treino$n.comodos.1,
base.pnad.treino$n.comodos.2,
base.pnad.treino$n.dormitorios.1,
base.pnad.treino$n.dormitorios.2,
base.pnad.treino$n.dormitorios.3,
#base.pnad.treino$n.banheiros.1,
#base.pnad.treino$n.banheiros.2,
normalize(base.pnad.treino$idade.anos),
base.pnad.treino$cri_ate7,
base.pnad.treino$cri_7a14,
base.pnad.treino$n.idoso,
base.pnad.treino$tipofam.1,

```

```

base.pnad.treino$tipofam.2,
base.pnad.treino$sexo,
base.pnad.treino$raca.1,
base.pnad.treino$raca.2,
base.pnad.treino$raca.3,
base.pnad.treino$posse.radio,
base.pnad.treino$posse.tv,
base.pnad.treino$posse.maqlavar,
#base.pnad.treino$posse.geladeira,
base.pnad.treino$posse.computador,
base.pnad.treino$posse.auto,
#base.pnad.treino$vive.conjugue,
base.pnad.treino$ocup.dom.1,
base.pnad.treino$ocup.dom.2,
base.pnad.treino$tipo.esgoto,
base.pnad.treino$freq.escola
# base.pnad.treino$pos.ocup.1,
# base.pnad.treino$pos.ocup.2,
# base.pnad.treino$pos.ocup.3,
# base.pnad.treino$pos.ocup.4,
# base.pnad.treino$pos.ocup.5,
# base.pnad.treino$pos.ocup.6
)

for (u in 1:replica) {
r <- 0
rep <- 0
#repeat {
rep <- rep+1
B0 <- runif(1, min = 0.01, max = 0.011)
B1 <- 1
B2 <- runif(1, min = 0.01, max = B1)
B3 <- runif(1, min = -0.011, max = B2)
B4 <- runif(1, min = -0.011, max = B3)

b02 <- runif(1, min = -0.011, max = 0.011)
b21 <- runif(1, min = -0.011, max = 0.011)

```

```
b22 <- runif(1, min = -0.011, max = 0.011)
b23 <- runif(1, min = -0.011, max = 0.011)
b24 <- runif(1, min = -0.011, max = 0.011)
b25 <- runif(1, min = -0.011, max = 0.011)
b26 <- runif(1, min = -0.011, max = 0.011)
b27 <- runif(1, min = -0.011, max = 0.011)
b28 <- runif(1, min = -0.011, max = 0.011)
b29 <- runif(1, min = -0.011, max = 0.011)
b210 <- runif(1, min = -0.011, max = 0.011)
b211 <- runif(1, min = -0.011, max = 0.011)
b212 <- runif(1, min = -0.011, max = 0.011)
b213 <- runif(1, min = -0.011, max = 0.011)
b214 <- runif(1, min = -0.011, max = 0.011)
b215 <- runif(1, min = -0.011, max = 0.011)
b216 <- runif(1, min = -0.011, max = 0.011)
b217 <- runif(1, min = -0.011, max = 0.011)
b218 <- runif(1, min = -0.011, max = 0.011)
b219 <- runif(1, min = -0.011, max = 0.011)
b220 <- runif(1, min = -0.011, max = 0.011)
b221 <- runif(1, min = -0.011, max = 0.011)
b222 <- runif(1, min = -0.011, max = 0.011)
b223 <- runif(1, min = -0.011, max = 0.011)
b224 <- runif(1, min = -0.011, max = 0.011)
```

```
b03 <- runif(1, min = -0.011, max = 0.011)
b31 <- runif(1, min = -0.011, max = 0.011)
b32 <- runif(1, min = -0.011, max = 0.011)
b33 <- runif(1, min = -0.011, max = 0.011)
b34 <- runif(1, min = -0.011, max = 0.011)
b35 <- runif(1, min = -0.011, max = 0.011)
b36 <- runif(1, min = -0.011, max = 0.011)
b37 <- runif(1, min = -0.011, max = 0.011)
b38 <- runif(1, min = -0.011, max = 0.011)
b39 <- runif(1, min = -0.011, max = 0.011)
b310 <- runif(1, min = -0.011, max = 0.011)
b311 <- runif(1, min = -0.011, max = 0.011)
b312 <- runif(1, min = -0.011, max = 0.011)
b313 <- runif(1, min = -0.011, max = 0.011)
```

```
b314 <- runif(1, min = -0.011, max = 0.011)
b315 <- runif(1, min = -0.011, max = 0.011)
b316 <- runif(1, min = -0.011, max = 0.011)
b317 <- runif(1, min = -0.011, max = 0.011)
b318 <- runif(1, min = -0.011, max = 0.011)
b319 <- runif(1, min = -0.011, max = 0.011)
b320 <- runif(1, min = -0.011, max = 0.011)
b321 <- runif(1, min = -0.011, max = 0.011)
b322 <- runif(1, min = -0.011, max = 0.011)
b323 <- runif(1, min = -0.011, max = 0.011)
b324 <- runif(1, min = -0.011, max = 0.011)
```

```
b04 <- runif(1, min = -0.011, max = 0.011)
b41 <- runif(1, min = -0.011, max = 0.011)
b42 <- runif(1, min = -0.011, max = 0.011)
b43 <- runif(1, min = -0.011, max = 0.011)
b44 <- runif(1, min = -0.011, max = 0.011)
b45 <- runif(1, min = -0.011, max = 0.011)
b46 <- runif(1, min = -0.011, max = 0.011)
b47 <- runif(1, min = -0.011, max = 0.011)
b48 <- runif(1, min = -0.011, max = 0.011)
b49 <- runif(1, min = -0.011, max = 0.011)
b410 <- runif(1, min = -0.011, max = 0.011)
b411 <- runif(1, min = -0.011, max = 0.011)
b412 <- runif(1, min = -0.011, max = 0.011)
b413 <- runif(1, min = -0.011, max = 0.011)
b414 <- runif(1, min = -0.011, max = 0.011)
b415 <- runif(1, min = -0.011, max = 0.011)
b416 <- runif(1, min = -0.011, max = 0.011)
b417 <- runif(1, min = -0.011, max = 0.011)
b418 <- runif(1, min = -0.011, max = 0.011)
b419 <- runif(1, min = -0.011, max = 0.011)
b420 <- runif(1, min = -0.011, max = 0.011)
b421 <- runif(1, min = -0.011, max = 0.011)
b422 <- runif(1, min = -0.011, max = 0.011)
b423 <- runif(1, min = -0.011, max = 0.011)
b424 <- runif(1, min = -0.011, max = 0.011)
```

```

v          <- matrix(nrow = nrow(y), ncol = 4)
yhat      <- matrix(nrow = nrow(y), ncol = 1)
erro      <- matrix(nrow = nrow(y), ncol = 1)
alfa      <- v.alfa
epocas    <- n.epocas
sqe       <- matrix(nrow = epocas, ncol = 1)
#coef     <- matrix(nrow = epocas, ncol = 4)

```

```

delta.B0 <- matrix(nrow = nrow(y), ncol = 1)
delta.B2 <- matrix(nrow = nrow(y), ncol = 1)
delta.B3 <- matrix(nrow = nrow(y), ncol = 1)
delta.B4 <- matrix(nrow = nrow(y), ncol = 1)

```

```

delta.b02 <- matrix(nrow = nrow(y), ncol = 1)
delta.b21 <- matrix(nrow = nrow(y), ncol = 1)
delta.b22 <- matrix(nrow = nrow(y), ncol = 1)
delta.b23 <- matrix(nrow = nrow(y), ncol = 1)
delta.b24 <- matrix(nrow = nrow(y), ncol = 1)
delta.b25 <- matrix(nrow = nrow(y), ncol = 1)
delta.b26 <- matrix(nrow = nrow(y), ncol = 1)
delta.b27 <- matrix(nrow = nrow(y), ncol = 1)
delta.b28 <- matrix(nrow = nrow(y), ncol = 1)
delta.b29 <- matrix(nrow = nrow(y), ncol = 1)
delta.b210 <- matrix(nrow = nrow(y), ncol = 1)
delta.b211 <- matrix(nrow = nrow(y), ncol = 1)
delta.b212 <- matrix(nrow = nrow(y), ncol = 1)
delta.b213 <- matrix(nrow = nrow(y), ncol = 1)
delta.b214 <- matrix(nrow = nrow(y), ncol = 1)
delta.b215 <- matrix(nrow = nrow(y), ncol = 1)
delta.b216 <- matrix(nrow = nrow(y), ncol = 1)
delta.b217 <- matrix(nrow = nrow(y), ncol = 1)
delta.b218 <- matrix(nrow = nrow(y), ncol = 1)
delta.b219 <- matrix(nrow = nrow(y), ncol = 1)
delta.b220 <- matrix(nrow = nrow(y), ncol = 1)
delta.b221 <- matrix(nrow = nrow(y), ncol = 1)
delta.b222 <- matrix(nrow = nrow(y), ncol = 1)
delta.b223 <- matrix(nrow = nrow(y), ncol = 1)

```



```
delta.b224 <- matrix(nrow = nrow(y), ncol = 1)
```

```
delta.b03 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b31 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b32 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b33 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b34 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b35 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b36 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b37 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b38 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b39 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b310 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b311 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b312 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b313 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b314 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b315 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b316 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b317 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b318 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b319 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b320 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b321 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b322 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b323 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b324 <- matrix(nrow = nrow(y), ncol = 1)
```

```
delta.b04 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b41 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b42 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b43 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b44 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b45 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b46 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b47 <- matrix(nrow = nrow(y), ncol = 1)  
delta.b48 <- matrix(nrow = nrow(y), ncol = 1)
```

```

delta.b49 <- matrix(nrow = nrow(y), ncol = 1)
delta.b410 <- matrix(nrow = nrow(y), ncol = 1)
delta.b411 <- matrix(nrow = nrow(y), ncol = 1)
delta.b412 <- matrix(nrow = nrow(y), ncol = 1)
delta.b413 <- matrix(nrow = nrow(y), ncol = 1)
delta.b414 <- matrix(nrow = nrow(y), ncol = 1)
delta.b415 <- matrix(nrow = nrow(y), ncol = 1)
delta.b416 <- matrix(nrow = nrow(y), ncol = 1)
delta.b417 <- matrix(nrow = nrow(y), ncol = 1)
delta.b418 <- matrix(nrow = nrow(y), ncol = 1)
delta.b419 <- matrix(nrow = nrow(y), ncol = 1)
delta.b420 <- matrix(nrow = nrow(y), ncol = 1)
delta.b421 <- matrix(nrow = nrow(y), ncol = 1)
delta.b422 <- matrix(nrow = nrow(y), ncol = 1)
delta.b423 <- matrix(nrow = nrow(y), ncol = 1)
delta.b424 <- matrix(nrow = nrow(y), ncol = 1)

```

```

for (j in 1:epocas) {
  for (i in 1:nrow(y)) {
    v[i, 1] <- (
      #lm.renda.c$coefficients[1] +
      lm.renda.c$coefficients[1] * z[i, 1] +
      lm.renda.c$coefficients[2] * z[i, 2] +
      lm.renda.c$coefficients[3] * z[i, 3] +
      lm.renda.c$coefficients[4] * z[i, 4] +
      lm.renda.c$coefficients[5] * z[i, 5] +
      lm.renda.c$coefficients[6] * z[i, 6] +
      lm.renda.c$coefficients[7] * z[i, 7] +
      lm.renda.c$coefficients[8] * z[i, 8] +
      lm.renda.c$coefficients[9] * z[i, 9] +
      lm.renda.c$coefficients[10] * z[i, 10] +
      lm.renda.c$coefficients[11] * z[i, 11] +
      lm.renda.c$coefficients[12] * z[i, 12] +
      lm.renda.c$coefficients[13] * z[i, 13] +
      lm.renda.c$coefficients[14] * z[i, 14] +
      lm.renda.c$coefficients[15] * z[i, 15] +
      lm.renda.c$coefficients[16] * z[i, 16] +

```

```

lm.renda.c$coefficients[17] * z[i, 17] +
lm.renda.c$coefficients[18] * z[i, 18] +
lm.renda.c$coefficients[19] * z[i, 19] +
lm.renda.c$coefficients[20] * z[i, 20] +
lm.renda.c$coefficients[21] * z[i, 21] +
lm.renda.c$coefficients[22] * z[i, 22] +
lm.renda.c$coefficients[23] * z[i, 23] +
lm.renda.c$coefficients[24] * z[i, 24]
)

```

```

v[i, 2] <- tanh(
b02 +
b21 * z[i, 1] +
b22 * z[i, 2] + #comodos
b23 * z[i, 3] +
b24 * z[i, 4] +
b25 * z[i, 5] + #dormitorios
b26 * z[i, 6] + #idade anos
b27 * z[i, 7] + #cri_ate7
b28 * z[i, 8] + #cri_7a14
b29 * z[i, 9] + #n.idoso
b210 * z[i, 10] +
b211 * z[i, 11] + #tipofam
b212 * z[i, 12] + #sexo
b213 * z[i, 13] +
b214 * z[i, 14] +
b215 * z[i, 15] + #raça
b216 * z[i, 16] + #posse.radio
b217 * z[i, 17] + #posse.tv
b218 * z[i, 18] + #posse.maqlavar
b219 * z[i, 19] + #posse.computador
b220 * z[i, 20] + #posse.auto
b221 * z[i, 21] +
b222 * z[i, 22] + #ocup.dom
b223 * z[i, 23] + #tipo.esgoto
b224 * z[i, 24] #freq.escola
)

```

```

v[i, 3] <- tanh(

```

```

b03 +
b31 * z[i, 1] +
b32 * z[i, 2] + #comodos
b33 * z[i, 3] +
b34 * z[i, 4] +
b35 * z[i, 5] + #dormitorios
b36 * z[i, 6] + #idade anos
b37 * z[i, 7] + #cri_ate7
b38 * z[i, 8] + #cri_7a14
b39 * z[i, 9] + #n.idoso
b310 * z[i, 10] +
b311 * z[i, 11] + #tipofam
b312 * z[i, 12] + #sexo
b313 * z[i, 13] +
b314 * z[i, 14] +
b315 * z[i, 15] + #raça
b316 * z[i, 16] + #posse.radio
b317 * z[i, 17] + #posse.tv
b318 * z[i, 18] + #posse.maqlavar
b319 * z[i, 19] + #posse.computador
b320 * z[i, 20] + #posse.auto
b321 * z[i, 21] +
b322 * z[i, 22] + #ocup.dom
b323 * z[i, 23] + #tipo.esgoto
b324 * z[i, 24] #freq.escola
)

```

```

v[i, 4] <- tanh(
b04 +
b41 * z[i, 1] +
b42 * z[i, 2] + #comodos
b43 * z[i, 3] +
b44 * z[i, 4] +
b45 * z[i, 5] + #dormitorios
b46 * z[i, 6] + #idade anos
b47 * z[i, 7] + #cri_ate7
b48 * z[i, 8] + #cri_7a14
b49 * z[i, 9] + #n.idoso
b410 * z[i, 10] +

```

```

b411 * z[i, 11] + #tipofam
b412 * z[i, 12] + #sexo
b413 * z[i, 13] +
b414 * z[i, 14] +
b415 * z[i, 15] + #raça
b416 * z[i, 16] + #posse.radio
b417 * z[i, 17] + #posse.tv
b418 * z[i, 18] + #posse.maqlavar
b419 * z[i, 19] + #posse.computador
b420 * z[i, 20] + #posse.auto
b421 * z[i, 21] +
b422 * z[i, 22] + #ocup.dom
b423 * z[i, 23] + #tipo.esgoto
b424 * z[i, 24] #freq.escola
)

```

```

yhat[i, ] <-
(B0 + B1 * v[i, 1] + B2 * v[i, 2] + B3 * v[i, 3] + B4 * v[i, 4])
erro[i, ] <- y[i, ] - yhat[i, ]

```

```

delta.B0[i, ] <- alfa * erro[i, ] * w[i, ]
delta.B2[i, ] <- alfa * erro[i, ] * v[i, 2] * w[i, ]
delta.B3[i, ] <- alfa * erro[i, ] * v[i, 3] * w[i, ]
delta.B4[i, ] <- alfa * erro[i, ] * v[i, 4] * w[i, ]

```

```

delta.b02[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * w[i, ]
delta.b21[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 1] * w[i, ]
delta.b22[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 2] * w[i, ]
delta.b23[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 3] * w[i, ]
delta.b24[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 4] * w[i, ]
delta.b25[i, ] <-

```

```

alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 5] * w[i, ]
delta.b26[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 6] * w[i, ]
delta.b27[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 7] * w[i, ]
delta.b28[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 8] * w[i, ]
delta.b29[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 9] * w[i, ]
delta.b210[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 10] * w[i, ]
delta.b211[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 11] * w[i, ]
delta.b212[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 12] * w[i, ]
delta.b213[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 13] * w[i, ]
delta.b214[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 14] * w[i, ]
delta.b215[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 15] * w[i, ]
delta.b216[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 16] * w[i, ]
delta.b217[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 17] * w[i, ]
delta.b218[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 18] * w[i, ]
delta.b219[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 19] * w[i, ]
delta.b220[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 20] * w[i, ]
delta.b221[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 21] * w[i, ]
delta.b222[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 22] * w[i, ]
delta.b223[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 23] * w[i, ]
delta.b224[i, ] <-
alfa * erro[i, ] * B2 * (1 - v[i, 2] ^ 2) * z[i, 24] * w[i, ]

```

```

delta.b03[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * w[i, ]
delta.b31[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 1] * w[i, ]
delta.b32[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 2] * w[i, ]
delta.b33[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 3] * w[i, ]
delta.b34[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 4] * w[i, ]
delta.b35[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 5] * w[i, ]
delta.b36[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 6] * w[i, ]
delta.b37[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 7] * w[i, ]
delta.b38[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 8] * w[i, ]
delta.b39[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 9] * w[i, ]
delta.b310[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 10] * w[i, ]
delta.b311[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 11] * w[i, ]
delta.b312[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 12] * w[i, ]
delta.b313[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 13] * w[i, ]
delta.b314[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 14] * w[i, ]
delta.b315[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 15] * w[i, ]
delta.b316[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 16] * w[i, ]
delta.b317[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 17] * w[i, ]
delta.b318[i, ] <-

```

```

alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 18] * w[i, ]
delta.b319[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 19] * w[i, ]
delta.b320[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 20] * w[i, ]
delta.b321[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 21] * w[i, ]
delta.b322[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 22] * w[i, ]
delta.b323[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 23] * w[i, ]
delta.b324[i, ] <-
alfa * erro[i, ] * B3 * (1 - v[i, 2] ^ 2) * z[i, 24] * w[i, ]

delta.b04[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * w[i, ]
delta.b41[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 1] * w[i, ]
delta.b42[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 2] * w[i, ]
delta.b43[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 3] * w[i, ]
delta.b44[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 4] * w[i, ]
delta.b45[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 5] * w[i, ]
delta.b46[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 6] * w[i, ]
delta.b47[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 7] * w[i, ]
delta.b48[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 8] * w[i, ]
delta.b49[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 9] * w[i, ]
delta.b410[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 10] * w[i, ]
delta.b411[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 11] * w[i, ]
delta.b412[i, ] <-

```



```

alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 12] * w[i, ]
delta.b413[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 13] * w[i, ]
delta.b414[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 14] * w[i, ]
delta.b415[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 15] * w[i, ]
delta.b416[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 16] * w[i, ]
delta.b417[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 17] * w[i, ]
delta.b418[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 18] * w[i, ]
delta.b419[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 19] * w[i, ]
delta.b420[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 20] * w[i, ]
delta.b421[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 21] * w[i, ]
delta.b422[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 22] * w[i, ]
delta.b423[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 23] * w[i, ]
delta.b424[i, ] <-
alfa * erro[i, ] * B4 * (1 - v[i, 2] ^ 2) * z[i, 24] * w[i, ]
}

```

```

B0 <- abs(B0 + colMeans(delta.B0))
B2 <- abs(B2 + colMeans(delta.B2))
B3 <- abs(B3 + colMeans(delta.B3))
B4 <- abs(B2 + colMeans(delta.B4))

```

```

b02 <- abs(b02 + colMeans(delta.b02))
b21 <- abs(b21 + colMeans(delta.b21))
b22 <- abs(b22 + colMeans(delta.b22))
b23 <- abs(b23 + colMeans(delta.b23))
b24 <- abs(b24 + colMeans(delta.b24))
b25 <- abs(b25 + colMeans(delta.b25))

```

```
b26 <- abs(b26 + colMeans(delta.b26))
b27 <- abs(b27 + colMeans(delta.b27))
b28 <- abs(b28 + colMeans(delta.b28))
b29 <- abs(b29 + colMeans(delta.b29))
b210 <- abs(b210 + colMeans(delta.b210))
b211 <- abs(b211 + colMeans(delta.b211))
b212 <- abs(b212 + colMeans(delta.b212))
b213 <- abs(b213 + colMeans(delta.b213))
b214 <- abs(b214 + colMeans(delta.b214))
b215 <- abs(b215 + colMeans(delta.b215))
b216 <- abs(b216 + colMeans(delta.b216))
b217 <- abs(b217 + colMeans(delta.b217))
b218 <- abs(b218 + colMeans(delta.b218))
b219 <- abs(b219 + colMeans(delta.b219))
b220 <- abs(b220 + colMeans(delta.b220))
b221 <- abs(b221 + colMeans(delta.b221))
b222 <- abs(b222 + colMeans(delta.b222))
b223 <- abs(b223 + colMeans(delta.b223))
b224 <- abs(b224 + colMeans(delta.b224))
```

```
b03 <- abs(b03 + colMeans(delta.b03))
b31 <- abs(b31 + colMeans(delta.b31))
b32 <- abs(b32 + colMeans(delta.b32))
b33 <- abs(b33 + colMeans(delta.b33))
b34 <- abs(b34 + colMeans(delta.b34))
b35 <- abs(b35 + colMeans(delta.b35))
b36 <- abs(b36 + colMeans(delta.b36))
b37 <- abs(b37 + colMeans(delta.b37))
b38 <- abs(b38 + colMeans(delta.b38))
b39 <- abs(b39 + colMeans(delta.b39))
b310 <- abs(b310 + colMeans(delta.b310))
b311 <- abs(b311 + colMeans(delta.b311))
b312 <- abs(b312 + colMeans(delta.b312))
b313 <- abs(b313 + colMeans(delta.b313))
b314 <- abs(b314 + colMeans(delta.b314))
b315 <- abs(b315 + colMeans(delta.b315))
b316 <- abs(b316 + colMeans(delta.b316))
b317 <- abs(b317 + colMeans(delta.b317))
```

```
b318 <- abs(b318 + colMeans(delta.b318))
b319 <- abs(b319 + colMeans(delta.b319))
b320 <- abs(b320 + colMeans(delta.b320))
b321 <- abs(b321 + colMeans(delta.b321))
b322 <- abs(b322 + colMeans(delta.b322))
b323 <- abs(b323 + colMeans(delta.b323))
b324 <- abs(b324 + colMeans(delta.b324))
```

```
b04 <- abs(b04 + colMeans(delta.b04))
b41 <- abs(b41 + colMeans(delta.b41))
b42 <- abs(b42 + colMeans(delta.b42))
b43 <- abs(b43 + colMeans(delta.b43))
b44 <- abs(b44 + colMeans(delta.b44))
b45 <- abs(b45 + colMeans(delta.b45))
b46 <- abs(b46 + colMeans(delta.b46))
b47 <- abs(b47 + colMeans(delta.b47))
b48 <- abs(b48 + colMeans(delta.b48))
b49 <- abs(b49 + colMeans(delta.b49))
b410 <- abs(b410 + colMeans(delta.b410))
b411 <- abs(b411 + colMeans(delta.b411))
b412 <- abs(b412 + colMeans(delta.b412))
b413 <- abs(b413 + colMeans(delta.b413))
b414 <- abs(b414 + colMeans(delta.b414))
b415 <- abs(b415 + colMeans(delta.b415))
b416 <- abs(b416 + colMeans(delta.b416))
b417 <- abs(b417 + colMeans(delta.b417))
b418 <- abs(b418 + colMeans(delta.b418))
b419 <- abs(b419 + colMeans(delta.b419))
b420 <- abs(b420 + colMeans(delta.b420))
b421 <- abs(b421 + colMeans(delta.b421))
b422 <- abs(b422 + colMeans(delta.b422))
b423 <- abs(b423 + colMeans(delta.b423))
b424 <- abs(b424 + colMeans(delta.b424))
```

```
sqe[j, ] <- sum(erro ^ 2) / 2
# coef[j,1] <- b0
# coef[j,2] <- b1
```

```

# coef[j,3] <- B0
# coef[j,4] <- B1

if (j > 1) {
v.sqe      <- abs((sqe[j - 1, ] - sqe[j, ]) / sqe[j - 1])
# d.coef.b0 <- abs((coef[j,1]-coef[j-1,1]))
# d.coef.b1 <- abs((coef[j,2]-coef[j-1,2]))
# d.coef.B0 <- abs((coef[j,3]-coef[j-1,3]))
# d.coef.B1 <- abs((coef[j,4]-coef[j-1,4]))
if ((v.sqe <= c.sqe) || j == n.epocas) {
break
}
}
if(j<=1) {
cat(paste("épocas:", j, "||", "sqe:", sum(erro ^ 2) / 2,"||", "repetição:",r), sep =
} else {cat(paste("épocas:", j, "||", "sqe:", sum(erro ^ 2) / 2,"||", "v.sqe:",v.sqe
}
}

m001[u, 1] <- q
m001[u, 2] <- u
m001[u, 3] <- B0
m001[u, 4] <- B1
m001[u, 5] <- B2

m001[u, 6] <- lm.renda.c$coefficients[1]
m001[u, 7] <- lm.renda.c$coefficients[2]
m001[u, 8] <- lm.renda.c$coefficients[3]
m001[u, 9] <- lm.renda.c$coefficients[4]
m001[u, 10] <- lm.renda.c$coefficients[5]
m001[u, 11] <- lm.renda.c$coefficients[6]
m001[u, 12] <- lm.renda.c$coefficients[7]
m001[u, 13] <- lm.renda.c$coefficients[8]
m001[u, 14] <- lm.renda.c$coefficients[9]
m001[u, 15] <- lm.renda.c$coefficients[10]
m001[u, 16] <- lm.renda.c$coefficients[11]
m001[u, 17] <- lm.renda.c$coefficients[12]
m001[u, 18] <- lm.renda.c$coefficients[13]
m001[u, 19] <- lm.renda.c$coefficients[14]
m001[u, 20] <- lm.renda.c$coefficients[15]

```

```
m001[u, 21] <- lm.renda.c$coefficients[16]
m001[u, 22] <- lm.renda.c$coefficients[17]
m001[u, 23] <- lm.renda.c$coefficients[18]
m001[u, 24] <- lm.renda.c$coefficients[19]
m001[u, 25] <- lm.renda.c$coefficients[20]
m001[u, 26] <- lm.renda.c$coefficients[21]
m001[u, 27] <- lm.renda.c$coefficients[22]
m001[u, 28] <- lm.renda.c$coefficients[23]
m001[u, 29] <- lm.renda.c$coefficients[24]
#m001[u, 30] <- lm.renda.c$coefficients[25]
```

```
m001[u, 31] <- b02
m001[u, 32] <- b21
m001[u, 33] <- b22
m001[u, 34] <- b23
m001[u, 35] <- b24
m001[u, 36] <- b25
m001[u, 37] <- b26
m001[u, 38] <- b27
m001[u, 39] <- b28
m001[u, 40] <- b29
m001[u, 41] <- b210
m001[u, 42] <- b211
m001[u, 43] <- b212
m001[u, 44] <- b213
m001[u, 45] <- b214
m001[u, 46] <- b215
m001[u, 47] <- b216
m001[u, 48] <- b217
m001[u, 49] <- b218
m001[u, 50] <- b219
m001[u, 51] <- b220
m001[u, 52] <- b221
m001[u, 53] <- b222
m001[u, 54] <- b223
m001[u, 55] <- b224
```

```
m001[u, 56] <- b03
m001[u, 57] <- b31
```

m001[u, 58] <- b32  
m001[u, 59] <- b33  
m001[u, 60] <- b34  
m001[u, 61] <- b35  
m001[u, 62] <- b36  
m001[u, 63] <- b37  
m001[u, 64] <- b38  
m001[u, 65] <- b39  
m001[u, 66] <- b310  
m001[u, 67] <- b311  
m001[u, 68] <- b312  
m001[u, 69] <- b313  
m001[u, 70] <- b314  
m001[u, 71] <- b315  
m001[u, 72] <- b316  
m001[u, 73] <- b317  
m001[u, 74] <- b318  
m001[u, 75] <- b319  
m001[u, 76] <- b320  
m001[u, 77] <- b321  
m001[u, 78] <- b322  
m001[u, 79] <- b323  
m001[u, 80] <- b324

m001[u, 81] <- b04  
m001[u, 82] <- b41  
m001[u, 83] <- b42  
m001[u, 84] <- b43  
m001[u, 85] <- b44  
m001[u, 86] <- b45  
m001[u, 87] <- b46  
m001[u, 88] <- b47  
m001[u, 89] <- b48  
m001[u, 90] <- b49  
m001[u, 91] <- b410  
m001[u, 92] <- b411  
m001[u, 93] <- b412  
m001[u, 94] <- b413

```
m001[u, 95] <-      b414
m001[u, 96] <-      b415
m001[u, 97] <-      b416
m001[u, 98] <-      b417
m001[u, 99] <-      b418
m001[u, 100] <-     b419
m001[u, 101] <-     b420
m001[u, 102] <-     b421
m001[u, 103] <-     b422
m001[u, 104] <-     b423
m001[u, 105] <-     b424
```

```
m001[u, 106] <- v.sqe
m001[u, 107] <- j
```

```
cat(
paste(
"Replicação:",
u,
"Repetição:",
r <-
r + 1,
"épocas:",
j,
"MSE:",
round(mean(na.exclude(sqe)), 2),
sep = " "
),
sep = "\n"
)
cat(paste(
"B0:",
round(B0, 4),
"B1:",
round(B1, 4),
"B2:",
round(B2, 4),
"B3:",
round(B3, 4),
```

```

"B4:",
round(B4, 4)
),
sep = '\n')

#if (B1 >= B2 & B2 >= B3 & B3 >= B4) {
#break
#}
#}

}
}

pop.rna.tanh.2 <- as.data.frame(m001)
names(pop.rna.tanh.2) <-
c(
'pop',
'replica',
'B0',
'B1',
'B2',
'b01',
'b11',
'b12',
'b13',
'b14',
'b15',
'b16',
'b17',
'b18',
'b19',
'b110',
'b111',
'b112',
'b113',
'b114',
'b115',
'b116',
'b117',

```



'b118',  
'b119',  
'b120',  
'b121',  
'b122',  
'b123',  
'b124',  
'b125',  
'b02',  
'b21',  
'b22',  
'b23',  
'b24',  
'b25',  
'b26',  
'b27',  
'b28',  
'b29',  
'b210',  
'b211',  
'b212',  
'b213',  
'b214',  
'b215',  
'b216',  
'b217',  
'b218',  
'b219',  
'b220',  
'b221',  
'b222',  
'b223',  
'b224',  
'b03',  
'b31',  
'b32',  
'b33',  
'b34',  
'b35',

'b36',  
'b37',  
'b38',  
'b39',  
'b310',  
'b311',  
'b312',  
'b313',  
'b314',  
'b315',  
'b316',  
'b317',  
'b318',  
'b319',  
'b320',  
'b321',  
'b322',  
'b323',  
'b324',  
'b04',  
'b41',  
'b42',  
'b43',  
'b44',  
'b45',  
'b46',  
'b47',  
'b48',  
'b49',  
'b410',  
'b411',  
'b412',  
'b413',  
'b414',  
'b415',  
'b416',  
'b417',  
'b418',  
'b419',

```

'b420',
'b421',
'b422',
'b423',
'b424',
'v.sqe',
'epoca'
)

write.table(
pop.rna.tanh.2,
file = '/home/savano/Documents/particular/dados_censo/rn_censo.txt',
quote = FALSE,
sep = "|",
row.names = FALSE
)

plot(
sqe,
type = 'l',
xlab = paste(j, 'épocas'),
xlim = c(1, j),
ylab = 'SQE'
)
plot(erro, ylab = 'Resíduos')

nf = layout(matrix(c(1, 2, 3, 4), 2, 2, byrow = TRUE))
plot(erro,
ylab = 'Resíduos',
cex = .3,
main = 'Dispersão')
hist(erro,
ylab = 'Frequência',
xlab = 'Resíduos',
main = 'Histograma')
boxplot(erro, xlab = 'Resíduos', main = 'Box-plot')
qqnorm(erro, ylab = 'Quantis Amostrais', xlab = 'Quantis Teóricos')
qqline(erro)

```

```

y.nn <- y*sd(log(base.pnad.treino$renda))+mean(log(base.pnad.treino$renda))
yhat.nn <- yhat*sd(log(base.pnad.treino$renda))+mean(log(base.pnad.treino$renda))

ssr.logrenda.treino.rna <- sum((exp(y.nn)-exp(yhat.nn))^2*base.pnad.treino$V4611)/s

sqr.rna.treino <- sum((exp(y.nn)-exp(yhat.nn))^2*base.pnad.treino$V4611)
sqt.rna.treino <- sum((exp(y.nn)-mean(exp(yhat.nn)))^2*base.pnad.treino$V4611)

r.2.rna.treino <- (1-sqr.rna.treino/sqt.rna.treino)*100

lm.renda.c.linear <- svyglm(
(log(renda)) ~
n.comodos.1 +
n.comodos.2 +
n.dormitorios.1 +
n.dormitorios.2 +
n.dormitorios.3 +
#n.banheiros.1 +
#n.banheiros.2 +
(idade.anos) +
cri_ate7 +
cri_7a14 +
n.idoso +
tipofam.1 +
tipofam.2 +
sexo +
raca.1 +
raca.2 +
raca.3 +
posse.radio +
posse.tv +
posse.maqlavar +
#posse.geladeira +
posse.computador +
posse.auto +
#vive.conjuge +
ocup.dom.1 +
ocup.dom.2 +
tipo.esgoto +

```

```

freq.escola ,
#pos.ocup.1 +
#pos.ocup.2 +
#pos.ocup.3 +
#pos.ocup.4 +
#pos.ocup.5 +
#pos.ocup.6,
design= d.pnad
)

summary(lm.renda.c.linear)

ssr.logrenda.treino.lm.linear <- sum((base.pnad.treino$renda-exp(lm.renda.c.linear

sqr.linear.treino <- sum((base.pnad.treino$renda-exp(lm.renda.c.linear$fitted.value
sqt.linear.treino <- sum((base.pnad.treino$renda-mean(exp(lm.renda.c.linear$fitted.

r.2.linear.treino <- (1-sqr.linear.treino/sqt.linear.treino)*100

##### Teste #####

# RNA
y.teste <- as.matrix((normalize(log(base.pnad.teste$renda))))
z.teste <- cbind.data.frame(
base.pnad.teste$n.comodos.1,
base.pnad.teste$n.comodos.2,
base.pnad.teste$n.dormitorios.1,
base.pnad.teste$n.dormitorios.2,
base.pnad.teste$n.dormitorios.3,
#base.pnad.teste$n.banheiros.1,
#base.pnad.teste$n.banheiros.2,
normalize(base.pnad.teste$idade.anos),
base.pnad.teste$cri_ate7,
base.pnad.teste$cri_7a14,
base.pnad.teste$n.idoso,
base.pnad.teste$tipofam.1,
base.pnad.teste$tipofam.2,
base.pnad.teste$sexo,

```

```

base.pnad.teste$raca.1,
base.pnad.teste$raca.2,
base.pnad.teste$raca.3,
base.pnad.teste$posse.radio,
base.pnad.teste$posse.tv,
base.pnad.teste$posse.maqlavar,
#base.pnad.teste$posse.geladeira,
base.pnad.teste$posse.computador,
base.pnad.teste$posse.auto,
#base.pnad.teste$vive.conjugue,
base.pnad.teste$ocup.dom.1,
base.pnad.teste$ocup.dom.2,
base.pnad.teste$tipo.esgoto,
base.pnad.teste$freq.escola
# base.pnad.teste$pos.ocup.1,
# base.pnad.teste$pos.ocup.2,
# base.pnad.teste$pos.ocup.3,
# base.pnad.teste$pos.ocup.4,
# base.pnad.teste$pos.ocup.5,
# base.pnad.teste$pos.ocup.6
)

v.teste      <- matrix(nrow = nrow(y.teste), ncol = 4)
yhat.teste   <- matrix(nrow = nrow(y.teste), ncol = 1)
erro.teste   <- matrix(nrow = nrow(y.teste), ncol = 1)

for (i in 1:nrow(y.teste)) {
v.teste[i, 1] <- (
#lm.renda.c$coefficients[1] +
lm.renda.c$coefficients[1] * z.teste[i, 1] +
lm.renda.c$coefficients[2] * z.teste[i, 2] +
lm.renda.c$coefficients[3] * z.teste[i, 3] +
lm.renda.c$coefficients[4] * z.teste[i, 4] +
lm.renda.c$coefficients[5] * z.teste[i, 5] +
lm.renda.c$coefficients[6] * z.teste[i, 6] +
lm.renda.c$coefficients[7] * z.teste[i, 7] +
lm.renda.c$coefficients[8] * z.teste[i, 8] +
lm.renda.c$coefficients[9] * z.teste[i, 9] +
lm.renda.c$coefficients[10] * z.teste[i, 10] +

```

```

lm.renda.c$coefficients[11] * z.teste[i, 11] +
lm.renda.c$coefficients[12] * z.teste[i, 12] +
lm.renda.c$coefficients[13] * z.teste[i, 13] +
lm.renda.c$coefficients[14] * z.teste[i, 14] +
lm.renda.c$coefficients[15] * z.teste[i, 15] +
lm.renda.c$coefficients[16] * z.teste[i, 16] +
lm.renda.c$coefficients[17] * z.teste[i, 17] +
lm.renda.c$coefficients[18] * z.teste[i, 18] +
lm.renda.c$coefficients[19] * z.teste[i, 19] +
lm.renda.c$coefficients[20] * z.teste[i, 20] +
lm.renda.c$coefficients[21] * z.teste[i, 21] +
lm.renda.c$coefficients[22] * z.teste[i, 22] +
lm.renda.c$coefficients[23] * z.teste[i, 23] +
lm.renda.c$coefficients[24] * z.teste[i, 24]
)

```

```

v.teste[i, 2] <- tanh(
b02 +
b21 * z.teste[i, 1] +
b22 * z.teste[i, 2] + #comodos
b23 * z.teste[i, 3] +
b24 * z.teste[i, 4] +
b25 * z.teste[i, 5] + #dormitorios
b26 * z.teste[i, 6] + #idade anos
b27 * z.teste[i, 7] + #cri_ate7
b28 * z.teste[i, 8] + #cri_7a14
b29 * z.teste[i, 9] + #n.idoso
b210 * z.teste[i, 10] +
b211 * z.teste[i, 11] + #tipofam
b212 * z.teste[i, 12] + #sexo
b213 * z.teste[i, 13] +
b214 * z.teste[i, 14] +
b215 * z.teste[i, 15] + #raça
b216 * z.teste[i, 16] + #posse.radio
b217 * z.teste[i, 17] + #posse.tv
b218 * z.teste[i, 18] + #posse.maqlavar
b219 * z.teste[i, 19] + #posse.computador
b220 * z.teste[i, 20] + #posse.auto
b221 * z.teste[i, 21] +

```

```

b222 * z.teste[i, 22] + #ocup.dom
b223 * z.teste[i, 23] + #tipo.esgoto
b224 * z.teste[i, 24] #freq.escola
)

v.teste[i, 3] <- tanh(
b03 +
b31 * z.teste[i, 1] +
b32 * z.teste[i, 2] + #comodos
b33 * z.teste[i, 3] +
b34 * z.teste[i, 4] +
b35 * z.teste[i, 5] + #dormitorios
b36 * z.teste[i, 6] + #idade anos
b37 * z.teste[i, 7] + #cri_ate7
b38 * z.teste[i, 8] + #cri_7a14
b39 * z.teste[i, 9] + #n.idoso
b310 * z.teste[i, 10] +
b311 * z.teste[i, 11] + #tipofam
b312 * z.teste[i, 12] + #sexo
b313 * z.teste[i, 13] +
b314 * z.teste[i, 14] +
b315 * z.teste[i, 15] + #raça
b316 * z.teste[i, 16] + #posse.radio
b317 * z.teste[i, 17] + #posse.tv
b318 * z.teste[i, 18] + #posse.maqlavar
b319 * z.teste[i, 19] + #posse.computador
b320 * z.teste[i, 20] + #posse.auto
b321 * z.teste[i, 21] +
b322 * z.teste[i, 22] + #ocup.dom
b323 * z.teste[i, 23] + #tipo.esgoto
b324 * z.teste[i, 24] #freq.escola
)

v.teste[i, 4] <- tanh(
b04 +
b41 * z.teste[i, 1] +
b42 * z.teste[i, 2] + #comodos
b43 * z.teste[i, 3] +
b44 * z.teste[i, 4] +

```



```

b45 * z.teste[i, 5] + #dormitorios
b46 * z.teste[i, 6] + #idade anos
b47 * z.teste[i, 7] + #cri_ate7
b48 * z.teste[i, 8] + #cri_7a14
b49 * z.teste[i, 9] + #n.idoso
b410 * z.teste[i, 10] +
b411 * z.teste[i, 11] + #tipofam
b412 * z.teste[i, 12] + #sexo
b413 * z.teste[i, 13] +
b414 * z.teste[i, 14] +
b415 * z.teste[i, 15] + #raça
b416 * z.teste[i, 16] + #posse.radio
b417 * z.teste[i, 17] + #posse.tv
b418 * z.teste[i, 18] + #posse.maqlavar
b419 * z.teste[i, 19] + #posse.computador
b420 * z.teste[i, 20] + #posse.auto
b421 * z.teste[i, 21] +
b422 * z.teste[i, 22] + #ocup.dom
b423 * z.teste[i, 23] + #tipo.esgoto
b424 * z.teste[i, 24] #freq.escola
)

```

```

yhat.teste[i, ] <-

```

```

(B0 + B1 * v.teste[i, 1] + B2 * v.teste[i, 2] + B3 * v.teste[i, 3] + B4 * v.teste[i, 4] +
erro.teste[i, ] <- y.teste[i, ] - yhat.teste[i, ]
}

```

```

y.teste.nn <- y.teste*sd(log(base.pnad.teste$renda))+mean(log(base.pnad.teste$renda))

```

```

yhat.teste.nn <- yhat.teste*sd(log(base.pnad.teste$renda))+mean(log(base.pnad.teste$renda))

```

```

ssr.logrenda.teste.rna <- sum((exp(y.teste.nn)-exp(yhat.teste.nn))^2*base.pnad.teste$renda)

```

```

yhat.teste.linear <- predict.lm(lm.renda.c.linear,newdata = base.pnad.teste)

```

```

ssr.logrenda.treino.lm.linear <- sum((exp(yhat.teste.linear) -base.pnad.teste$renda)^2)

```