



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Predição da Macrossomia Fetal Através da Regressão Logística e de Redes Neurais Artificiais

Mariana Teixeira Carballo

Orientador: Álvaro Vigo

Co-orientadores: Maria Inês Schmidt

Bruce B. Duncan

Monografia apresentada para obtenção do
grau em Bacharel em Estatística

Porto Alegre, Outubro de 2002

Agradecimentos

Este trabalho foi realizado com o apoio e incentivo de professores, amigos e colegas. A esses presto minha gratidão:

Ao meu orientador, Prof. Álvaro Vigo, agradeço o auxílio, a confiança e suas sugestões.

À professora Maria Inês Schmidt e ao Prof. Bruce B. Duncan, que acreditaram no meu trabalho, agradeço a confiança e a oportunidade de participação no Programa de Pós-Graduação em Epidemiologia, onde obtive valiosas experiências. Agradeço em especial a permissão da utilização do banco de dados para a realização desse trabalho.

Ao professor João Riboldi agradeço pelas recomendações e pelo empréstimo de material.

Desejo agradecer aos professores do Departamento de Estatística que no decorrer desses anos contribuíram para o meu crescimento.

À professora Elsa Cristina de Mundstock, que me apoiou desde o início do curso como orientadora em nossos trabalhos de iniciação científica, agradeço a orientação, preocupação, credibilidade, incentivo, apoio em todas as horas.

À professora Jandyra Maria G. Fachel, pela orientação e escolha da área de atuação, bioestatística.

Às professoras Liane Wemer e Dinara Fernandez pelo apoio em todos os momentos.

Ao Dr. Juarez Barbisan, pelo apoio e incentivo em todas as horas.

Aos meus familiares pais e irmãos pelo apoio e compreensão.

Aos meus amigos, em especial à Juliane e à Rose, aos meus colegas de curso, ao meu namorado, aos meus colegas da Pós-graduação em Epidemiologia, em especial à Iracema Santos, agradeço pelas horas de estudo, conselhos, incentivo, com momentos de descontração e diversão.

Sumário

1. Introdução	05
1.1 Objetivos	06
1.2 Estrutura da monografia	07
2. Regressão logística	08
2.1 Introdução	08
2.2 Estimação dos parâmetros	13
2.3 Seleção das variáveis	17
2.4 Ajuste do modelo	18
2.5 Validação do modelo	21
3. Redes neurais	22
3.1 Introdução	22
3.2 Definições básicas	25
3.3 Componentes de uma rede neural	29
3.4 Funcionamento da rede	30
3.5 Aprendizado	33
3.6 Aplicação de uma rede neural artificial	34
4. Diagnóstico clínico	36
4.1 Introdução	36
4.2 Conceitos probabilísticos básicos	37
4.3 Qualidade dos testes diagnósticos	39
4.4 Combinação de testes diagnósticos	42
4.5 Testes diagnósticos baseados em variáveis contínuas	45
4.6 Curva característica de operação (curva ROC)	46
5. Estudo de Caso: Predição da Macrossomia Fetal no EBDG	49
5.1 Características do EBDG	49
5.2 Delineamento e descrição da amostra	50
5.3 Métodos de coleta	51
5.4 Medidas	52

5.5 Predição da macrossomia fetal	53
5.5.1 Ajuste do modelo de regressão logística	54
5.5.2 Ajuste da rede neural	58
5.6 Comparação entre regressão logística e redes neurais	62
5.7 Vantagens e desvantagens da regressão logística e redes neurais	63
6. Considerações finais	67
7. Referências bibliográficas	69

Anexos

Anexo A – Separação dos bancos	73
A1 – Rotina do SAS para separar o banco de dados para o ajuste do modelo logístico	74
A2 – Rotina do SAS para reunificação do banco de dados	75
Anexo B – Ajuste do modelo de regressão logística	76
B1 – Rotina do SAS para ajuste do modelo logístico	77
B2 – Resultados do modelo final	80
B3 – Rotina do SAS para estimação da sensibilidade e especificidade	82
B4 – Resultados da estimação da sensibilidade e especificidade	84
Anexo C – Construção da rede neural artificial	91
C1 – Ajuste da rede neural	92
C2 – Resultados da rede neural	95
C3 – Estimação da sensibilidade e especificidade	100
C4 – Resultados da estimação da sensibilidade e especificidade	101
Anexo D – Construção da curva ROC no SPSS	106
D1 – Rotina do SPSS para construção da Curva Roc	107

1. Introdução

A gravidez provoca alterações metabólicas importantes - o organismo da mãe se prepara para fornecer ao feto toda a quantidade de energia necessária para garantir seu crescimento e seu desenvolvimento. Nas últimas semanas de gestação, as necessidades energéticas do feto aumentam, provocando uma resistência à ação da insulina nas células da gestante. Como consequência, mães com Diabete Melito durante a gestação não conseguem produzir insulina suficiente para controlar o metabolismo. Apesar da insulina não atravessar a placenta, o açúcar e macronutrientes circulando em excesso são transmitidos ao bebê e, assim, o pâncreas do bebê produz insulina extra para diminuir a glicemia. Como consequência, o bebê fica grande (macrossomia), porque a quantidade extra de açúcar faz com ele produza mais gordura.

A macrossomia é basicamente caracterizada por bebês com peso superior ao percentil 90 para sua idade gestacional (grande para a idade gestacional), podendo causar dificuldades durante o parto, como a fratura de clavículas (ossos dos ombros). Além disso, o bebê pode apresentar hipoglicemia logo após o parto, pois produzia muito mais insulina na gestação e não está mais recebendo o excesso de açúcar da mãe. Outro aspecto importante é o risco de apresentar problemas respiratórios devido à prematuridade. Estes aspectos reforçam a evidência de que recém-nascidos grandes têm maior risco de desenvolver obesidade.

A gravidade dos riscos associados à macrossomia e a severidade dos eventuais danos para o feto reforçam a necessidade de um procedimento para prever a ocorrência de macrossomia fetal ainda nos estágios iniciais da gravidez, permitindo intervenções para controlar o crescimento interuterino. Clinicamente, os médicos preveem se um bebê é macrossômico ou não baseados na combinação de sintomas observados nas gestantes e/ou em exames laboratoriais. Mediante uma ponderação dos sintomas observados nas gestantes, nos resultados dos exames e a devida importância de cada um, juntamente com a sua experiência, o médico elabora o diagnóstico mais provável para o bebê.

Na estatística, o diagnóstico médico pode ser comparado a uma previsão construída através de um modelo matemático, semelhante à ponderação dos sintomas. Muitas técnicas estatísticas podem ser usadas para fazer previsões, cuja escolha depende dos objetivos e da natureza dos dados.

Para uma variável com resposta dicotômica, o modelo de regressão logística é a técnica de análise comumente utilizada para descrever as relações entre o desfecho e os preditores, discretos ou contínuos. Na prática, entretanto, o modelo logístico parece relativamente limitado para identificar a existência de comportamento não linear entre o desfecho e os preditores. Em alguns casos, pode ser necessário fazer transformações nos regressores, mas muitas vezes elas não conhecidas *a priori*.

Um procedimento alternativo é utilizar a técnica denominada de redes neurais artificiais, cuja estrutura assemelha-se ao cérebro humano e permite modelar matematicamente as relações, lineares e não lineares, entre as variáveis independentes e dependentes.

1.1 Objetivos

O principal objetivo desse trabalho é construir procedimentos para prever a macrosomia fetal em gestantes de 24-28 semanas, através do modelo de regressão logística e de redes neurais artificiais.

A capacidade preditiva destes procedimentos usualmente pode ser comparada mediante as medidas de sensibilidade e de especificidade, determinando a área sob a curva característica de operação (*ROC - Receiver Operating Characteristic*, em inglês).

Em particular, deseja-se ilustrar empiricamente vantagens e desvantagens destas metodologias, incentivando a utilização de redes neurais no meio médico e epidemiológico.

1.2 Estrutura da monografia

Nesse trabalho serão apresentadas duas metodologias estatísticas que podem ser úteis para prever a macrosomia fetal, bem como outros desfechos dicotômicos. No Capítulo 2 são brevemente descritos aspectos do modelo de regressão logística, enquanto que o Capítulo 3 descreve resumidamente a metodologia das redes neurais artificiais.

No Capítulo 4 são apresentados aspectos básicos das medidas usualmente empregadas para comparar a capacidade preditiva dos modelos em estudo, em termos da sensibilidade, especificidade e da área sob a curva característica de operação.

O Capítulo 5 apresenta um estudo de caso, cujos dados foram gerados no estudo de coorte conduzido por pesquisadores do Programa de Pós-Graduação em Epidemiologia da UFRGS. Ele essencialmente trata da predição da ocorrência da macrosomia fetal, mediante as metodologias acima referidas. A descrição dos resultados obtidos com a regressão logística e com a rede neural é descrita nesse capítulo, bem como a construção das curvas características de operação.

No Capítulo 6 são discutidos os principais resultados gerados no trabalho, enquanto que o Capítulo 7 contém as referências utilizadas. Finalmente, as rotinas computacionais e os correspondentes resultados são mostrados nos anexos.

2. Regressão Logística

2.1 Introdução

O modelo de regressão logística pode ser extremamente útil para descrever a relação entre um desfecho binário e uma ou mais variáveis explicativas, pois permite estimar a magnitude e a direção dos efeitos dos preditores. O modelo ajustado também pode ser usado para prever o desfecho em um particular indivíduo, com base nos correspondentes valores observados das variáveis explanatórias. Se o modelo é parcimonioso e possui boa capacidade preditiva, então pode ser muito valioso na prática, em particular do ponto de vista clínico.

Aspectos teóricos do modelo de regressão logística estão amplamente discutidos na literatura, cabendo destacar Hosmer & Lemeshow (2000), Harrel (2001), Agresti (1990), Cox & Snell (1989) e Breslow & Day (1980), por exemplo.

É conveniente, entretanto, descrever resumidamente aspectos da metodologia. Para tanto, considere o desfecho binário Y que assume os valores genericamente representados por sucesso ($Y=1$) e fracasso ($Y=0$), enquanto as variáveis ou fatores explanatórios (também usualmente chamados de variáveis explicativas ou de covariáveis) são representadas pelo vetor $\mathbf{x} = (x_1, x_2, \dots, x_p)$. Assim, a probabilidade de sucesso da resposta binária, dado o vetor das variáveis explanatórias, será representada por $P(Y=1|\mathbf{x}) = \pi(\mathbf{x})$ e, naturalmente, a probabilidade de fracasso é $P(Y=0|\mathbf{x}) = 1 - \pi(\mathbf{x})$.

Presumivelmente, existe uma relação de dependência entre a probabilidade de sucesso e as variáveis explanatórias. Uma maneira tradicional de expressar matematicamente esta dependência é postular a existência de uma variável subjacente, latente e contínua, associada ao desfecho. Em outras palavras, os desfechos observados são oriundos da dicotomização de uma variável aleatória contínua subjacente, usualmente não observável diretamente.

Embora os valores das variáveis explanatórias sejam diferentes para cada indivíduo, usualmente podem estar arrançados em grupos, de tal forma que, dentro de cada grupo, todos os indivíduos têm os mesmos valores nas correspondentes variáveis explanatórias. Assim, sob o modelo ajustado os indivíduos de cada grupo apresentam a mesma estimativa de probabilidade de sucesso. Conseqüentemente, os aspectos relevantes a serem considerados sobre a variável resposta nos diferentes grupos são o número de indivíduos e o total de sucessos associados a cada grupo. Nesse caso, não se estará interessado na transformação logística da probabilidade de sucesso π , mas sim na transformação da variável aleatória que representa o número de sucessos do grupo. Essa transformação logística da variável aleatória é denominada de *transformação logito empírica*, sendo bastante discutida por Cox & Snell (1989, p.31,33).

O modelo logístico para resposta dicotômica foi primeiramente desenvolvido por Cox (1958) e Walker & Duncan (1967). O trabalho desenvolvido por Walker & Duncan (1967) teve uma motivação extremamente prática: identificar possíveis fatores de risco associados à ocorrência de doenças cardíacas coronarianas. O método é ilustrado para os registros dos 5209 participantes do estudo de Framingham, descrito por Dawber, Meadors e Moore (1951) – apud Walker & Duncan (1967).

Com base no conhecimento médico, os referidos autores argumentam que é razoável supor que a probabilidade de sucesso π pode ser descrita pela curva sigmóide. Assim, por exemplo, se Z representa a variável contínua subjacente, então a função

$$F(z) = \frac{\exp(z)}{1 + \exp(z)} = \frac{1}{1 + \exp(-z)}; \quad z \in \mathfrak{R},$$

é a função de distribuição acumulada de uma variável aleatória com distribuição logística. A Figura 2.1 ilustra o comportamento desta função.

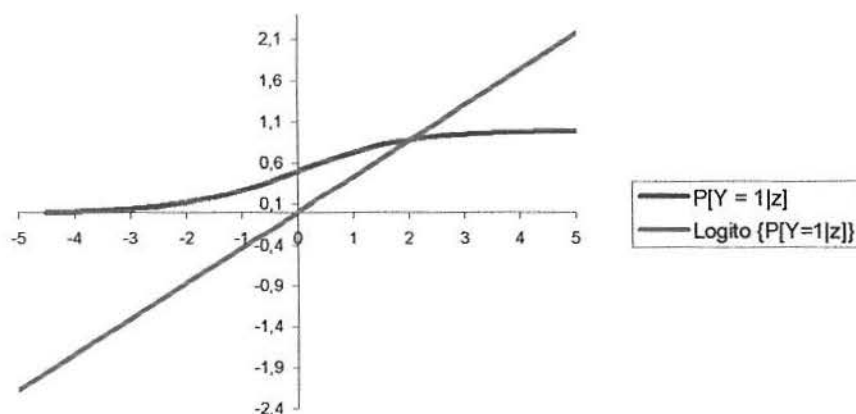


Figura 2.1: Comportamento da função de distribuição logística e da função logito.

→ A relação de dependência entre π e o vetor de variáveis explanatórias $\mathbf{x} = (x_1, x_2, \dots, x_p)$ pode ser expressada através da esperança condicional de Y , dado que as covariáveis assumem o valor \mathbf{x} , denotada por $E(Y|\mathbf{x})$. Como Y assume apenas os valores 0 e 1, sua esperança é dada por

$$E(Y|\mathbf{x}) = 0 \times P(Y = 0|\mathbf{x}) + 1 \times P(Y = 1|\mathbf{x}) = P(Y = 1|\mathbf{x}) = \pi(\mathbf{x}),$$

de tal forma que $0 \leq E(Y|\mathbf{x}) \leq 1$.

Admita, então, que para um dado valor do vetor de covariáveis \mathbf{x} , a função de distribuição condicional da variável latente e contínua Z é representada por $F(z|\mathbf{x})$. Como Z não pode ser observada, então, sem perda de generalidade, admita que a resposta binária assume valor $Y = 1$ se e somente se $Z > 0$, tal que

$$\pi(\mathbf{x}) = P(Y = 1|\mathbf{x}) = P(Z|\mathbf{x} > 0) = 1 - F(0|\mathbf{x}). \quad (2.1)$$

Assim, se $Z|\mathbf{x}$ é descrita pela densidade logística com parâmetro de locação $\alpha + \sum_{j=1}^p \beta_j x_j$, onde $\alpha, \beta_1, \beta_2, \dots, \beta_p$ são constantes desconhecidas que deseja-se estimar, e parâmetro de escala igual a 1, então

$$\begin{aligned} \pi(\mathbf{x}) &= P(Y = 1|\mathbf{x}) = 1 - F(0|\mathbf{x}) = 1 - \frac{\exp\left(0 - \alpha - \sum_{j=1}^p \beta_j x_j\right)}{1 + \exp\left(0 - \alpha - \sum_{j=1}^p \beta_j x_j\right)} \\ &= \frac{1 + \exp\left(-\left[\alpha + \sum_{j=1}^p \beta_j x_j\right]\right) - \exp\left(-\left[\alpha + \sum_{j=1}^p \beta_j x_j\right]\right)}{1 + \exp\left(-\left[\alpha + \sum_{j=1}^p \beta_j x_j\right]\right)} \\ &= \frac{1}{1 + \exp\left(-\left[\alpha + \sum_{j=1}^p \beta_j x_j\right]\right)} = \frac{\exp\left(\alpha + \sum_{j=1}^p \beta_j x_j\right)}{1 + \exp\left(\alpha + \sum_{j=1}^p \beta_j x_j\right)} \end{aligned} \quad (2.2)$$

e, naturalmente,

$$1 - \pi(\mathbf{x}) = P(Y = 0|\mathbf{x}) = \frac{1}{1 + \exp\left\{\alpha + \sum_{j=1}^p \beta_j x_j\right\}}, \quad (2.3)$$

estabelecendo uma relação matemática entre o desfecho e as variáveis explanatórias. Portanto, mediante a transformação inversa

$$\log \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} = \alpha + \sum_{j=1}^p \beta_j x_j, \quad (2.4)$$

o modelo é linear nos parâmetros. A transformação definida na equação (2.4) é denominada *logito* de $\pi(\mathbf{x}) = P(Y = 1|\mathbf{x})$, dando origem à denominação *modelo logístico* ou *de regressão logística*. A linearização da relação pode ser visualizada na Figura 2.1.

A medida da equação (2.5)

$$\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})} = \exp\left\{\alpha + \sum_{j=1}^p \beta_j x_j\right\}, \quad (2.5)$$

definida pela razão entre a probabilidade condicional de sucesso e seu complementar, é chamada de “odds” do evento $Y=1|\mathbf{x}$ e expressa a chance relativa de sua ocorrência. Assim, para indivíduos com valores distintos \mathbf{x}_1 e \mathbf{x}_2 para o vetor de variáveis explanatórias, a razão

$$\frac{\frac{\pi(\mathbf{x}_1)}{1-\pi(\mathbf{x}_1)}}{\frac{\pi(\mathbf{x}_2)}{1-\pi(\mathbf{x}_2)}} = \frac{\exp\left\{\alpha + \sum_{j=1}^p \beta_j x_{1j}\right\}}{\exp\left\{\alpha + \sum_{j=1}^p \beta_j x_{2j}\right\}} = \frac{\exp\{\boldsymbol{\beta}' \mathbf{x}_1\}}{\exp\{\boldsymbol{\beta}' \mathbf{x}_2\}} = \exp\{\boldsymbol{\beta}' (\mathbf{x}_1 - \mathbf{x}_2)\} \quad (2.6)$$

é chamada de razão de chances.

Cabe salientar, no entanto, que qualquer distribuição de probabilidade unimodal e simétrica pode ser postulada para a variável latente, geralmente produzindo resultados similares. Por exemplo, se variável latente Z tem distribuição normal com média

$\alpha + \sum_{j=1}^p \beta_j x_j$, então

$$\pi(\mathbf{x}) = P(Y=1|\mathbf{x}) = \Phi\left(\alpha + \sum_{j=1}^p \beta_j x_j\right), \quad (2.7)$$

onde $\Phi(\cdot)$ representa a função de distribuição acumulada de uma variável aleatória com *distribuição de probabilidade normal padrão*.

Esta relação é linearizada através da transformação inversa

$$\phi^{-1}(\pi(\mathbf{x})) = \alpha + \sum_{j=1}^p \beta_j x_j, \quad (2.8)$$

conduzindo ao modelo de probitos proposto por D.J. Finney em meados do século passado. Este modelo está extensamente descrito na literatura, em especial no contexto dos *bioensaios*. Veja Finney (1974) e Cox & Snell (1989, p.15).

Por outro lado, quando a variável latente apresenta uma acentuada assimetria, pode ser importante utilizar uma distribuição de probabilidade que contemple esta característica, tal como a *distribuição de valor extremo*, de *Gompert* ou *Gumbel*. Estes aspectos estão além dos objetivos propostos, mas ao leitor interessado recomenda-se a discussão apresentada por McCullagh & Nelder (1989, p.111), Agresti (1990, p.105) ou Cox & Snell (1989), entre outros.

Na próxima seção são apresentados aspectos básicos da estimação de parâmetros do modelo logístico.

2.2 Estimação dos parâmetros

Neste trabalho, o procedimento de estimação dos parâmetros do modelo logístico definido pela equação (2.4) é apresentado para o contexto em que os dados são gerados por uma amostra aleatória simples. Naturalmente, o modelo pode ser estendido para delineamentos amostrais mais complexos, mas estes aspectos estão além dos objetivos propostos. Ao leitor interessado, recomenda-se, por exemplo, Hosmer & Lemeshow (2000, p.203).

Assim, para uma amostra de tamanho n , os dados observados podem ser representados pelos pares (y_i, \mathbf{x}_i) , $\forall i = 1, 2, \dots, n$. O vetor p -dimensional $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^t$ contém os valores observados nas variáveis explanatórias para o i -ésimo indivíduo.

Embora existam outras alternativas, o procedimento da *máxima verossimilhança* tem sido utilizado com maior frequência para estimar os parâmetros do modelo logístico, o qual é brevemente descrito a seguir.

Quando mais de um valor de Y é observado para um determinado valor do vetor \mathbf{x} , fixo, é suficiente armazenar o número de observações n_i e o número de “1’s” gerados. Assim, considera-se Y_i como sendo o número de sucessos, ao invés de respostas binárias individuais. Estes $\{Y_i, i=1, \dots, I\}$ são variáveis aleatórias independentes com distribuição binomial, com esperança $E(Y_i) = n_i \pi_i$, onde $n_1 + n_2 + \dots + n_I = n$. Assim, a função massa de probabilidade conjunta do vetor (Y_1, Y_2, \dots, Y_I) é proporcional ao produto das I funções binomiais, isto é:

$$\begin{aligned} \prod_{i=1}^I \pi(\mathbf{x}_i)^{y_i} [1 - \pi(\mathbf{x}_i)]^{n_i - y_i} &= \left[\prod_{i=1}^I [1 - \pi(\mathbf{x}_i)]^{n_i} \right] \left[\prod_{i=1}^I \exp \left[\log \left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right)^{y_i} \right] \right] \\ &= \left[\prod_{i=1}^I [1 - \pi(\mathbf{x}_i)]^{n_i} \right] \exp \left[\sum_{i=1}^I y_i \log \left(\frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right) \right]. \end{aligned} \quad (2.9)$$

Fazendo $\alpha = \beta_0$, i -ésimo logito é $\sum_{j=1}^p \beta_j x_{ij}$, de tal forma que o termo exponencial da equação (2.9) é igual a

$$\exp \left\{ \sum_{i=1}^I y_i \left[\sum_{j=1}^p \beta_j x_{ij} \right] \right\} = \exp \left\{ \sum_j \left[\sum_i y_i x_{ij} \right] \beta_j \right\}.$$

Ainda, $(1 - \pi(\mathbf{x}_i)) = \left[1 + \exp \left\{ \sum_j \beta_j x_{ij} \right\} \right]^{-1}$, tal que o logaritmo da função de verossimilhança

é igual a

$$L(\boldsymbol{\beta}) = \sum_j \left(\sum_i y_i x_{ij} \right) \beta_j - \sum_i n_i \log \left[1 + \exp \left(\sum_j \beta_j x_{ij} \right) \right], \quad (2.10)$$

que depende da distribuição binomial apenas através da estatística suficiente $\left\{ \sum_i y_i x_{ij}; j = 0, 1, 2, \dots, p \right\}$. As equações de verossimilhança são obtidas derivando a equação (2.10) em relação aos elementos do vetor de parâmetros $(\beta_0, \beta_1, \dots, \beta_p)$ e igualando a zero estes resultados. Assim, como

$$\frac{\partial L}{\partial \beta_a} = \sum_i y_i x_{ia} - \sum_i n_i x_{ia} \left[\frac{\exp\left(\sum_j \beta_j x_{ij}\right)}{1 + \exp\left(\sum_j \beta_j x_{ij}\right)} \right], \quad (2.11)$$

e igualando em zero as derivadas parciais, as equações de verossimilhança são dadas por

$$\sum_i y_i x_{ia} - \sum_i n_i \hat{\pi}_i x_{ia} = 0, \quad a = 0, \dots, p, \quad (2.12)$$

onde

$$\hat{\pi}_i = \exp\left(\sum_j \hat{\beta}_j x_{ij}\right) / \left[1 + \exp\left(\sum_j \hat{\beta}_j x_{ij}\right)\right]$$

representa a estimativa de máxima verossimilhança para $\pi(\mathbf{x}_i)$.

Na forma matricial, as equações de verossimilhança podem ser escritas como

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\hat{\mathbf{m}} \quad (2.13)$$

onde $\hat{\mathbf{m}}_i = n_i \hat{\pi}_i$ e \mathbf{X} é uma matriz de dimensão $I \times (p+1)$ que contém os valores das variáveis explanatórias $\{x_{ij}\}$.

A matriz de informação é o valor esperado negativo da matriz de derivadas parciais de segunda ordem da função de verossimilhança. Sob condições de regularidade, os estimadores de máxima verossimilhança têm distribuição aproximadamente normal com matriz de covariância igual ao inverso da matriz de informação.

Para o modelo logístico, a derivada parcial de segunda ordem do logaritmo da função de verossimilhança especificado em (2.11) é

$$\frac{\partial^2 L}{\partial \beta_a \partial \beta_b} = - \sum_i \frac{x_{ia} x_{ib} n_i \exp\left(\sum_j \beta_j x_{ij}\right)}{\left[1 + \exp\left(\sum_j \beta_j x_{ij}\right)\right]^2} = - \sum_i x_{ia} x_{ib} n_i \pi_i (1 - \pi_i). \quad (2.14)$$

Como (2.14) não é função dos $\{y_i\}$, a matriz da derivada de segunda ordem observada e a esperada são idênticas. Assim, a matriz de covariâncias pode ser estimada mediante

$$\hat{Cov}(\hat{\beta}) = \{\mathbf{X}' \text{Diag}[n_i \hat{\pi}_i (1 - \hat{\pi}_i)] \mathbf{X}\}^{-1}, \quad (2.15)$$

onde $\text{Diag}[n_i \hat{\pi}_i (1 - \hat{\pi}_i)]$ representa a matriz diagonal com dimensão $I \times I$, contendo os elementos $\{n_i \hat{\pi}_i (1 - \hat{\pi}_i)\}$ na diagonal principal. A raiz quadrada dos elementos da diagonal principal da equação (2.15) produz as estimativas dos erros padrões associados às estimativas dos parâmetros do modelo.

Para um valor fixo \mathbf{x} , a variância estimada do logito predito $\hat{L} = \hat{\beta}' \mathbf{x}$ é $\hat{\sigma}^2(\hat{L}) = \mathbf{x}' \hat{Cov}(\hat{\beta}) \mathbf{x}$. Para grandes amostras, um intervalo assintótico com $(1 - \alpha)100\%$ de confiança para o verdadeiro logito é $\hat{L} \pm z_{\alpha/2} \hat{\sigma}(\hat{L})$. Para construir um intervalo de confiança assintótico para $\pi(\mathbf{x})$, basta inverter os extremos do intervalo de confiança acima, mediante a transformação $\hat{\pi} = \exp(\hat{L}) / [1 + \exp(\hat{L})]$.

Entretanto, as equações de verossimilhança (2.12) ou (2.13) são funções não lineares das componente do vetor de parâmetros β e, portanto, a solução do sistema requer um procedimento iterativo, como o método de Newton-Raphson, descrito em Agresti (1990, p.114-117).

A função de verossimilhança para o modelo de regressão logística é estritamente côncava, e as estimativas de máxima verossimilhança existem e são únicas, exceto em algumas situações extremas. Estes casos são discutidos, por exemplo, em Wedderburn (1976), Albert & Anderson (1984) e SAS Institute Inc. (1999).

Para amostras grandes, os estimadores de máxima verossimilhança apresentam boas propriedades, mas para amostras pequenas, podem gerar estimativas tendenciosas. Conseqüentemente, as estimativas das razões de chances podem não corresponder à realidade, sendo de pouca utilidade prática. Em tais casos, um procedimento alternativo é utilizar a metodologia da regressão logística exata, descrita, por exemplo, em Bressel (2001), Metha & Patel (1995) ou Cytel Software Corporation (2001).

Em situações práticas, na etapa de ajuste do modelo, usualmente são considerados diversos fatores explanatórios para descrever o comportamento do desfecho. Entretanto, um modelo com um número excessivo de covariáveis pode reduzir demasiadamente o número de graus de liberdade. Assim, sugere-se construir um modelo parcimonioso, mediante a rigorosa avaliação da importância relativa dos fatores explanatórios. Aspectos básicos da seleção das covariáveis do modelo são brevemente descritos na próxima seção.

2.3 Seleção de variáveis

Em muitos problemas epidemiológicos existem muitas variáveis que potencialmente podem ser incluídas no modelo, de tal forma que é necessário desenvolver uma estratégia para avaliar a importância relativa de cada variável. O objetivo básico é selecionar as variáveis que conduzam ao modelo que descreva da melhor forma possível, e com parcimônia, o fenômeno em estudo.

Após o ajuste do modelo, deve-se dar uma atenção cuidadosa à significância das covariáveis no modelo, podendo ser necessário utilizar testes de hipóteses para avaliar a significância dos fatores explanatórios. Um aspecto que deve ser observado é que manter covariáveis pouco relevantes no modelo pode prejudicar as estimativas dos outros parâmetros.

Testar a significância de um conjunto de covariáveis consiste essencialmente em verificar se o modelo que as inclui revela melhor ajuste (e interpretabilidade das relações com o desfecho) que o modelo que não inclui este conjunto de covariáveis. Isto pode ser realizado mediante a comparação dos valores observados com os preditos, para cada um dos dois modelos: um que inclui o conjunto de covariáveis sob avaliação e outro sem as referidas covariáveis. Se os valores preditos pelo modelo com as covariáveis são melhores do que aqueles preditos pelo modelo sem as variáveis, então elas são significativas.

A comparação entre os dois modelos é baseada no logaritmo da função de verossimilhança definido na equação (2.10), mediante a estatística

$$G = -2 \ln \left[\frac{\text{verossimilhança sem as covariáveis}}{\text{verossimilhança com as covariáveis}} \right]. \quad (2.16)$$

Sob a hipótese nula que especifica que os coeficientes do conjunto de fatores explanatórios em investigação são nulos, a estatística G tem distribuição assintótica de Qui-quadrado com graus de liberdade definidos pela diferença entre o número de parâmetros dos dois modelos. Naturalmente, em determinadas situações, a relevância prática de um determinado fator explanatório pode ser mais importante do que o resultado deste teste. Estes aspectos encontram-se extensamente detalhados na literatura e implementados em rotinas computacionais para o ajuste do modelo logístico. Maiores detalhes deste e outros procedimentos podem ser encontrados, por exemplo, em Hosmer & Lemeshow (2000, p.11-17) e SAS Institute Inc. (1999).

2.4 Ajuste do modelo

Até o momento foram considerado aspectos da estimação dos parâmetros e a avaliação da correspondente significância. Entretanto, os valores preditos, juntamente com as estimativas dos parâmetros, são as causas de variação do modelo e, para fazer inferências, essa variação precisa ser quantificada. Serão necessárias, então, algumas medidas do ajuste do modelo de regressão logística.

Testar o ajuste do modelo, significa verificar se o modelo contém todos os efeitos principais e interações que deveria conter para explicar adequadamente o comportamento da variável resposta.

O termo em inglês para esse tipo de verificação é “*goodness-of-fit*” e basicamente visa avaliar a discrepância entre os valores observados e os correspondentes valores preditos. Admita que os valores observados são (y_1, y_2, \dots, y_n) e os correspondentes preditos são $(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$. Em linhas gerais, diz-se que o modelo está bem ajustado se uma medida resumo das discrepâncias entre y_i e \hat{y}_i , $\forall i=1, 2, \dots, n$ assume um valor pequeno e a contribuição de cada par (y_i, \hat{y}_i) para essa medida resumo é assimétrica e é relativamente pequena em relação a estrutura de erro do modelo.

Uma avaliação completa do ajuste do modelo envolve o cálculo e a avaliação de uma medida global do ajuste, bem como o exame das componentes individuais dessa estatística, incluindo avaliações gráficas e exame de outras medidas das diferenças ou distâncias entre y_i e \hat{y}_i .

As estatísticas usuais para se testar o ajuste do modelo são a estatística Qui-quadrado de Pearson, a Deviance e os testes de Hosmer-Lemeshow, os quais estão detalhadamente descritos no Capítulo 5 de Hosmer & Lemeshow (2000, p.143).

Na regressão logística, os valores ajustados são calculados para cada padrão de covariáveis e dependem das correspondentes probabilidades estimadas. A estatística Qui-quadrado de Pearson é definida por

$$X^2 = \sum_{j=1}^J \left[\frac{(y_j - m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \right]^2, \quad (2.17)$$

e a deviance será definida por

$$D = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2, \quad (2.18)$$

onde $d(y_j, \hat{\pi}_j) = \pm \left\{ 2 \left[y_j \ln \left(\frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left(\frac{m_j - y_j}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{1/2}$. Sob a hipótese que o modelo

está bem ajustado em todos os aspectos, Ambas as estatísticas assumem distribuição qui-quadrado com $J - (p + 1)$ graus de liberdade. A deviance representará o teste da razão de verossimilhança do modelo saturado com J parâmetros em relação ao modelo ajustado com $p + 1$ parâmetros.

Para auxiliar na compreensão, considera-se a estatística X^2 como a estatística de Pearson e D como o log-verossimilhança da estatística Qui-quadrado de uma tabela de contingência de dimensão $2 \times J$. As linhas da tabela correspondem aos dois valores da variável de resposta, isto é, $y = 0$ ou $y = 1$, enquanto que as colunas representam as J possíveis covariáveis. A estimativa do valor esperado sob a hipótese de que o modelo logístico é o modelo correto para a célula de contingência correspondente a linha $y = 1$ e a j -ésima coluna é $m_j \hat{\pi}_j$.

Essa tabela de contingência $2 \times J$ terá valores esperados pequenos na medida que o número de colunas aumenta com o tamanho da amostra. Para solucionar esse problema, fixa-se um número de grupos, g , e então se calcula as frequências esperadas e observadas de cada célula, de tal forma que a estimativa do valor esperado aumenta com o aumento do tamanho da amostra.

Hosmer e Lemeshow (?) agruparam os valores das probabilidades estimadas considerando $J = n$ e construíram uma estatística para avaliar a qualidade do ajuste, denominada de estatística de Hosmer-Lemeshow, \hat{C} , que é calculada através dos valores da estatística de Qui-quadrado de Pearson associados à tabela de contingência $2 \times g$,

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}, \quad (2.19)$$

onde n'_k é o número da covariável padrão no k -ésimo grupo, $o_k = \sum_{j=1}^{n'_k} y_j$ representa o número de respostas sobre a covariável n'_k e $\bar{\pi}_k = \sum_{j=1}^{n'_k} m_j \hat{\pi}_j / n'_k$ representa a probabilidade média estimada. Quando $J = n$ e o modelo ajustado pela regressão logística é o modelo correto, a distribuição da estatística \hat{C} é muito bem aproximada pela distribuição Qui-quadrado com $g - 2$ graus de liberdade, $\chi^2 (g - 2)$. Para leitores mais interessados neste teste recomenda-se Hosmer & Lemeshow (1989, p. 140).

2.5 Validação do modelo

Acredita-se que a utilização de estimativas do risco global poderia ser mais relevante para guiar as decisões do que aquelas usando representações binárias dos fatores de risco. Entretanto, utilizando estimativas quantitativas do risco para um indivíduo que não pertence à população original estudada perde-se precisão e capacidade de generalização do risco estimado.

A precisão é normalmente representada pela variância ou pelos intervalos de confiança do risco estimado da população para a qual o modelo foi construído. A generalização pode ser visualizada através de um estudo externo da validação, ou seja, aplicando o modelo a alguns dados que não foram usados na sua construção, conforme Jaulent et. al. sugeriram.

A validação do modelo pode ser realizada através de diversos mecanismos. É possível que o procedimento mais usual consiste em dividir, ao acaso, o conjunto de observações em duas partes. Uma delas é usada para ajustar o modelo, cujas estimativas dos parâmetros são então usadas para fazer previsões para os indivíduos do segundo conjunto. Entretanto, nos últimos anos diversos melhoramentos deste procedimento têm sido implementados, mediante o emprego de técnicas de re-amostragem, tais como “*bootstrapping*” e “*jackknifing*”. Veja, por exemplo, Harrell (2001, p.90).

3. Redes Neurais

3.1 Introdução

O reconhecimento de padrões enfoca uma amplitude de problemas de processamento de informações de grande significado prático, como os diagnósticos clínicos. Frequentemente, os diagnósticos são solucionados pelos próprios médicos, tomando como base os exames solicitados, sua experiência e o relato dos sintomas.

Essa inovação tecnológica auxilia os clínicos naqueles diagnósticos mais difíceis, quando os pacientes não apresentam todos os sintomas da doença ou quando apresentam uma combinação de sintomas.

A primeira aplicação de uma rede neural artificial na clínica médica foi realizada para descrever um infarto no miocárdio. Segundo Baxt (1995), a rede neural construída identifica mais precisamente um infarto severo no miocárdio, avaliando não só as relações lineares clássicas mas também as relações multidimensionais.

As redes neurais artificiais têm uma grande aceitação no meio médico para realizar diagnósticos clínicos. Isso se deve ao fato de que se constituem de técnicas sofisticadas, com grande capacidade de modelar funções extremamente complexas, particularmente as não lineares, das variáveis de entrada em relação às variáveis de saída. O problema de dimensão das variáveis, que surge quando muitas variáveis são usadas em modelos não lineares, geralmente fica controlado quando as redes neurais são usadas.

Segundo Cross et al. (1995), um teste médico pode ser usado para discriminar os indivíduos que apresentam uma determinada doença ou condição, daqueles que não a apresentam. Entretanto, num procedimento deste tipo não poderá gerar a especificidade (capacidade do teste identificar indivíduos verdadeiramente livres da doença ou condição, não casos) e a sensibilidade (capacidade do teste detectar indivíduos verdadeiramente positivos, os casos).

Isto ocorre porque não existe uma maneira de se separar os pacientes com e sem doença, dado que os pacientes não estão agrupados em um determinado ponto, e sim estão alocados de forma aleatória.

Para que todos os pacientes sejam separados da mesma forma, um algoritmo deve ser gerado através de conhecimentos da população em estudo e de técnicas estatísticas avançadas.

Na etapa de discriminação, na entrada cada indivíduo será representado por x_i , que assume o valor 1 quando o indivíduo apresenta a doença, e o valor 0, caso contrário. Os valores da variável X podem ser representados pelo vetor $X = (x_1, \dots, x_d)$, onde d é o número total de indivíduos. Na saída, os sujeitos são discriminados mediante a variável y , que assume os valores 0 e 1, respectivamente para não doentes e doentes.

Assim, o sistema todo pode ser visualizado através do conjunto de entrada das variáveis x_1, \dots, x_d , representando os indivíduos, gerando a variável de saída y , representando a classificação dos indivíduos (doentes ou não). Em problemas mais complexos podem existir algumas variáveis de saída representadas por y_k , onde $k = 1, \dots, c$. Para aqueles leitores mais interessados recomenda-se Bishop (1995).

Em geral, não é possível determinar uma forma para o comportamento desejado das variáveis. O comportamento é modelado mediante alguma função matemática, que possui um número de parâmetros desconhecidos, os quais são determinados com o auxílio dos dados. Essa função pode ser escrita da seguinte forma

$$y_k = y_k(x; w), \quad (3.1)$$

onde w representa o vetor de parâmetros.

As redes neurais podem ser vistas como escolhas específicas para a forma funcional usada para representar a variável de saída na equação (3.1), juntamente com procedimentos particulares para otimizar os parâmetros que são chamados de pesos.

Elas constroem um modelo de seleção automática dos pacientes com e sem doença, e também podem ser usadas para selecionar aqueles itens que são mais importantes na performance da discriminação. O processo para determinar os valores dos parâmetros é baseado no conjunto de dados denominado de *aprendizado* (*training*, em inglês).

Os métodos básicos para utilizar as redes neurais são os mesmos da análise discriminante. A amostra coletada deve ser representativa para o número de casos e a saída gerada deve ser conhecida por algum método padrão, que no meio médico é comumente denominado padrão-ouro. No caso do infarto do miocárdio, por exemplo, a amostra em estudo consiste de um grande número de casos atendidos na emergência e o padrão-ouro utilizado foi o eletrocardiograma. Este estudo é bastante discutido por Baxt (1995).

Após a etapa de aprendizagem, o modelo construído deve ser testado sobre uma nova série de pacientes, para se verificar sua aplicabilidade na população em estudo. Usualmente, isto pode ser realizado mediante a divisão aleatória do conjunto de observações em duas partes: uma parte é usada para construir a rede e, a outra, para testá-la.

A rede que gera a discriminação de um conjunto de dados é conhecida como “*Perceptron*” e normalmente apresenta camadas simples, onde as variáveis de entrada são ligadas diretamente às de saída. Os “*Perceptron*” são combinações lineares das variáveis de entrada, as quais são transformadas por uma função de ativação não linear. Entretanto, quando se utilizam funções mais gerais, considerando redes com camadas sucessivas de processamento das unidades, as redes são chamadas de “*Perceptron*” com múltiplas camadas (*Multilayer Perceptron*, em inglês). Essas definições são encontradas em Bishop (1995), StatSoft Textbook (1998) e SAS Institute Inc (2000).

Para permitir uma visão mais geral, consideram-se transformações sucessivas, correspondendo a uma rede contendo algumas camadas com pesos adaptados. De fato, redes com somente duas camadas de pesos são capazes de aproximar qualquer função contínua.

Geralmente pode-se considerar um diagrama de rede arbitrário (não necessariamente tendo uma estrutura de camadas simples), desde que qualquer diagrama da rede possa ser convertido no seu mapa funcional correspondente. A única restrição feita é que o diagrama deve ser alimentado para frente, ao invés de para trás. Isso permite que a saída da rede seja calculada como uma função explícita das entradas e dos pesos.

Usualmente, na rede “*Perceptron*” com múltiplas camadas a função de ativação é a sigmóide ou a “*threshold*”. Essas funções serão apresentadas na Seção 3.4. Para as redes com outras funções de ativação, existe um método, denominado de erro de programação para trás (*error back propagation*, em inglês), para encontrar as derivadas da função do erro com respeito aos pesos e ao viés da rede.

A rede neural tem um funcionamento semelhante aos neurônios biológicos, os quais codificam a informação recebida e transmitem seus resultados aos demais neurônios. Nas próximas seções serão abordados aspectos básicos da metodologia de redes neurais artificiais. Aos leitores interessados em uma discussão mais detalhada, sugere-se consultar, por exemplo, Selau (2000) e Bishop (1995).

3.2 Definições básicas

Uma rede neural artificial consiste em um conjunto de unidades de processamento denominadas de nós, os quais simulam os neurônios humanos e estão interconectados através de um conjunto de pesos (de maneira similar às conexões sinápticas do sistema nervoso). Eles permitem que os sinais sejam transmitidos através da rede em paralelo. Os nós são elementos computacionais simples, baseados no comportamento de um neurônio: quando há neuro-transmissores suficientes acumulados no corpo da célula, uma ação é gerada. O mecanismo dos neurônios foi modelado matematicamente através da soma ponderada de todos os sinais que chegam a um determinado nó. Se o nível dos sinais exceder um certo limite, denominado de “*threshold*”, então é produzida uma resposta de saída. Nesse trabalho não se descreve o funcionamento dos neurônios biológicos: para leitores interessados, recomenda-se Selau (2000) e StatSoft Textbook (1998).

Segundo Bishop (1995), quando existe apenas uma camada, a rede neural pode ser introduzida no contexto de ajuste de uma curva polinomial. Aqui, o problema consiste em ajustar um polinômio ao conjunto de N observações, visando a minimização do erro. O polinômio considerado pode ser do tipo

$$y(x) = w_0 + w_1 x + \dots + w_m x^m = \sum_{j=0}^m w_j x^j, \quad (3.2)$$

onde x representa a entrada na rede e y é a saída gerada, cujo objetivo básico é representar o comportamento não linear. A forma precisa da função $y(x)$ será determinada pelos valores dos parâmetros w_0, \dots, w_m , os quais são análogos aos pesos da rede neural. É conveniente representar o conjunto de parâmetros (w_0, \dots, w_m) pelo vetor \vec{w} . O polinômio pode ser escrito na forma funcional como $y = y(x, w)$.

O valor desejado para a saída y é chamado de valor alvo (*target*, em inglês) no contexto da rede neural. Para se encontrar valores aceitáveis para os coeficientes do polinômio, é conveniente considerar o erro entre o valor desejado de saída t^n , para um particular valor de entrada x^n , e o correspondente valor predito pela função polinomial $y(x^n; w)$.

O procedimento de ajuste da curva envolve a minimização do erro quadrático que, quando somado sobre todas as observações, produz

$$E = \frac{1}{2} \sum_{n=1}^N \{y(x^n; w) - t^n\}^2. \quad (3.3)$$

Assim, a quantidade E definida na equação (3.3) é uma função de w , e o polinômio será ajustado aos dados considerando o valor estimado de w , que minimiza E , representado por \hat{w} . Para se encontrar o mínimo deve-se solucionar o sistema de equações lineares definido por (3.2) e (3.3).

A minimização da função do erro em (3.3), a qual envolve os valores alvo para a saída da rede, é chamada de aprendizagem supervisionada, desde que seja especificado o valor desejado de saída para cada padrão de entrada. Outro método de aprendizagem na rede neural, chamado de aprendizado sem supervisão, não envolve o uso dos dados alvos. Nesse tipo de rede não se está interessado no aprendizado dos valores de entrada e saída, mas sim na distribuição de probabilidade dos valores de entrada ou, então, em se encontrar agrupamentos nos dados. Esse tipo de rede não será abordado nesse trabalho. Veja Bishop (1995, p.33-76).

Quando são consideradas mais camadas, cada camada é conectada com cada uma das unidades da outra camada. Neste tipo de rede, unidades que não são de saída são denominadas unidades intermediárias. Assim, a rede é constituída por d camadas de entrada, m unidades intermediárias e c unidades de saída, e genericamente pode ser representada pela Figura 3.1.

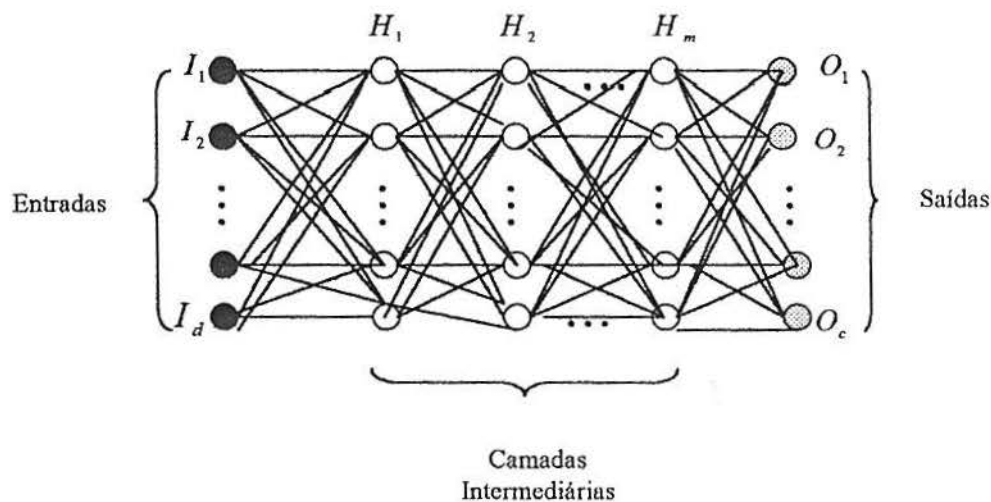


Figura 3.1: Representação Gráfica de uma rede neural com d camadas de entrada, m camadas intermediária e c camadas de saída

Conforme Bishop (1995), a função analítica dessa figura é representada como segue. A j -ésima saída da unidade intermediária é obtida pela primeira formação da combinação linear dos d valores de entrada e, adicionando-se um viés, gera

$$a_j = \sum_{i=0}^d w_{ij}(1) x_i . \quad (3.4)$$

Esta função é semelhante àquela definida em (3.2), onde $w_{ij}(1)$ representa o peso na primeira camada, variando da entrada i à unidade intermediária j , considerando camadas simples e uma variável de entrada extra x_0 que representa o viés, a qual pode assumir o valor $x_0 = 1$.

A ativação da unidade intermediária j é então obtida pela transformação da soma definida na equação (3.4), usando a função de ativação $g(\cdot)$ tal que

$$f_j = g(a_j), \quad (3.5)$$

que será definida na Seção 3.4. A saída da rede é obtida mediante transformações na ativação de cada unidade intermediária, usando os elementos de processamento da segunda camada. Assim, para cada unidade de saída k , constrói-se uma combinação linear das saídas das unidades intermediárias na forma

$$a_k = \sum_{i=1}^m w_{kj}(2) f_i , \quad (3.6)$$

onde o viés é representado por uma unidade intermediária extra, representada por $f_0 = 1$. A ativação dessa k -ésima unidade de saída é realizada pela transformação da combinação linear, usando uma função de ativação não linear

$$y_k = \tilde{g}(a_k). \quad (3.7)$$

A função de ativação da unidade de saída não precisa ser a mesma usada nas unidades intermediárias.

A combinação das funções definidas pelas equações (3.4), (3.5), (3.6) e (3.7) gera a expressão explícita para representar a função completa no diagrama, que assume a forma

$$y_k = \tilde{g} \left(\sum_{j=0}^m w_{kj}(2) g \left(\sum_{i=0}^d w_{ji}(1) x_i \right) \right). \quad (3.8)$$

Essa rede é denominada de rede com duas camadas intermediárias. Apesar de ser possível utilizar outras camadas intermediárias possíveis, elas não têm sido muito usadas, pois a utilização de muitas camadas pode gerar uma superestimava (*overfitting*, em inglês) na rede, conforme mencionado por Tu (1996).

Existem muitas maneiras para se representar o comportamento não linear entre os espaços multidimensionais. A importância das redes é a maneira pela qual elas tratam o problema de dimensão dos dados. Geralmente, esses modelos representam as funções não lineares entre muitas variáveis mediante a superposição das funções não lineares de uma única variável, a qual denomina-se de funções escondidas (*hidden function*, em inglês). O ponto chave é que as funções escondidas são adaptadas aos dados como parte do processo de aprendizagem e, então, o número dessas funções precisa somente crescer com a complexidade do problema, não simplesmente com a dimensão Selau (2000).

Portanto, a rede é constituída pelas unidades de entrada, pelas unidades intermediárias e pelas unidades de saída. Além disso, um viés pode ser introduzido nas camadas intermediárias da rede para auxiliar no efeito da função de ativação da rede.

3.3 Componentes de uma rede neural

Conforme Selau (2000), as redes neurais são constituídas por uma camada de entrada, uma de saída e camadas intermediárias (nós ocultos). Os nós de entrada obtêm a informação inicial que será transmitida a rede. Para as variáveis quantitativas existe somente um nó, enquanto que para as variáveis qualitativas existe um nó associado a cada uma de suas categorias.

Para as variáveis categorizadas, deve-se definir variáveis dicotômicas para cada uma de suas categorias, as quais identificam a ocorrência ou não do correspondente evento. Por sua vez, os nós de saída recebem entradas e produzem saídas, mas não transmitem as informações aos outros neurônios, pois são os resultados finais da rede. Os nós intermediários executam todas as operações internas, associadas às relações não lineares da rede. Isto, em geral, torna a rede mais poderosa.

Os componentes da rede neural podem ser descritos através de termos estatísticos cotidianos. Os neurônios de entrada, por exemplo, representam as variáveis independentes, enquanto que os de saída são os valores preditos. Já os neurônios escondidos, presentes nas camadas intermediárias, representam as interações entre as variáveis explicativas.

Ainda, os valores de treinamento representam as variáveis dependentes, o treinamento é o procedimento de estimação e os pesos são as estimativas dos parâmetros.

A rede é alimentada para frente (*feed-forward*). Isto significa que o sinal circula pelo nó de entrada (*input*), segue através dos nós intermediários (*hidden*) e, eventualmente, atinge os nós de saída (*output*).

Em muitas situações pode ser introduzido um viés nas conexões dos neurônios permitindo um aumento ou uma diminuição da função de ativação. Assim, pode-se aumentar os graus de liberdade da rede, facilitando sua adaptação ao conhecimento fornecido.

3.4 Funcionamento da rede

Quando existe uma correspondência direta entre o diagrama da rede e sua função matemática, é possível descrever o comportamento geral da rede. Nesse trabalho a atenção será voltada para o processo de alimentação para frente (*feed-forward*), o qual não volta para trás.

Uma rede alimentada dessa forma, quando possível, anexa números sucessivos aos valores de entrada e para todas as unidades intermediárias e de saída, de tal forma que cada unidade receba somente conexões das entradas ou de um pequeno número de unidades.

Conforme Selau (2000), um nó pode receber as informações de entrada X_1, X_2, \dots, X_p de diferentes meios: dos dados, da saída de outro nó ou das variáveis independentes. Cada informação de entrada recebe, através de suas conexões, um peso (w_i) que corresponde à eficiência sináptica do neurônio biológico. Cada neurônio tem um valor limite para o armazenamento de informações (denominado de *threshold*). Quando esse valor limite é atingido, o nó gera uma soma ponderada das entradas, ocorrendo sua ativação (Σ). O sinal de ativação é então passado através de uma função de ativação $f(a)$ para produzir a saída do neurônio y . Estes aspectos básicos do funcionamento de uma rede neural são ilustrados no diagrama da Figura 3.2.

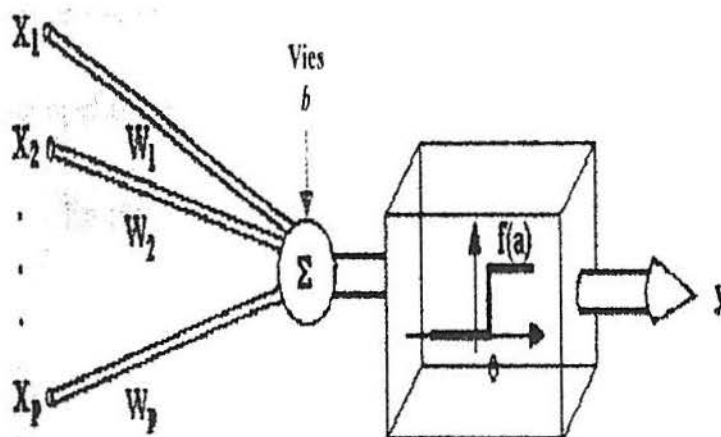


Figura 3.2: Descrição do funcionamento da rede. Fonte: Selau (2000)

Existem algumas maneiras, pelas quais a função linear das variáveis de entrada pode ser generalizada, produzindo a saída da rede. A principal maneira é considerar uma função não linear, denominada de função de ativação. Quando são consideradas diversas camadas, a saída da k -ésima unidade é realizada pela função de ativação representada por

$$f_k = g\left(\sum_j w_{kj} z_j\right), \quad (3.9)$$

onde a soma é realizada sobre todas as entradas e unidades que estão conectadas à unidade k . Um parâmetro de viés também está incluído no somatório. Para um dado conjunto de valores aplicado na entrada da rede, utilizações sucessivas da equação (3.9) permitem a ativação de todas as unidades da rede a serem incluídas nas unidades de saída. Esse processo pode ser visto como uma programação dos sinais para frente, através da rede. Na prática, existem poucas redes aleatórias, mas sempre existem vantagens consideráveis na construção de várias estruturas dentro da rede. Para leitores mais interessados, recomenda-se Bishop (1995).

Quando a função de ativação é usada, a saída do neurônio será igual a 0 se a entrada for menor do que zero, ou igual a 1, se a entrada for maior ou igual a zero. O nó agirá como o neurônio biológico, subtraindo o valor do *threshold* da soma ponderada e comparando com o valor zero, que é semelhante à comparação da soma ponderada pelo valor *threshold*.

Existem muitas escolhas para as funções de ativação não lineares, nas camadas intermediárias e de saída. Normalmente, a função de ativação é monótona, genericamente representada por

$$f(a) = g(w \cdot x + w_0). \quad (3.7)$$

A função de ativação $f(a)$ transforma a soma de ativação das entradas, em saídas. Esta função é o diferencial de cada nó, pois é ela que revela o valor a ser passado para os demais nós. Geralmente, são funções não lineares, sendo que as mais utilizadas são a função *threshold* e a função sigmóide, respectivamente definidas por

$$f(a) = \frac{1}{1 + e^{-a}} \quad \text{e} \quad f(a) = \tanh(a). \quad (3.8)$$

A função sigmóide, que pode ser comparada a função logística, é frequentemente a mais usada em redes neurais, enquanto que a função *threshold* é usada quando as entradas são binárias. As funções de ativação estão mais detalhadas em Bishop (1995; p. 122-128).

Nas unidades de entrada e nas camadas intermediárias, os valores computados são passados, através das conexões, para outros nós intermediários ou para o nó de saída.

Para as unidades de saída os valores computados são os termos preditos pelo modelo, os quais são comparados com os valores alvos para produzir os erros, que devem ser minimizados no treinamento.

A generalização da rede pode ser realizada através de diversos métodos, sendo que o mais simples consiste em minimizar a soma de quadrados dos erros.

3.5 Aprendizado

Quando a rede de percepção múltipla é inicializada, as conexões entre todos os nós são realizadas por pequenos pesos arbitrários, gerando uma predição errada para uma determinada entrada. Assim, a rede deve passar por um processo de aprendizado para melhorar seu desempenho. Esse aprendizado é feito através de um processo iterativo de ajustes aplicados aos seus pesos, chamado de treinamento. Para tanto, um grande número de casos cujas saídas são conhecidas é apresentado e os pesos dos conectores são alterados pelo treinamento do algoritmo, produzindo uma saída que melhor concorde com as saídas conhecidas. O número de casos necessários na etapa do treinamento depende essencialmente do número de variáveis de entradas, do número de decisões a serem tomadas e da complexidade do problema. O processo de treinamento é descrito em Bishop (1995), Cross et al.(1995), StatSoft Textbook (1998) e SAS Institute Inc.(2000).

Em termos estatísticos, durante o processo de aprendizagem ocorrem desvios entre a função alvo $f(x)$ e a função real $F(x,w)$ avaliada pela rede, onde x representa o sinal de entrada e w os pesos sinápticos. O vetor de entrada X é utilizado para prever a variável dependente Y , que é a saída da rede.

O aprendizado da rede inicia com a retro-propagação. O processo de retro-programação minimiza o erro quadrático médio da saída sobre o conjunto de entrada do treinamento.

O aprendizado termina quando um determinado número de iterações é atingido. Assim, o ponto ótimo de parada corresponde à situação em que a rede apresenta o menor erro quadrático médio e capacidade de generalização máxima.

Algumas vezes o treinamento da rede torna-se superespecialista, comprometendo a generalização dos dados (*overfitting*). Este problema usualmente pode ser evitado utilizando-se conjuntos de dados diferentes em cada treinamento da rede, reduzindo o número de camadas intermediárias ou adicionando um viés à função objetivo. Conforme Tu (1996), uma maneira alternativa de se aplicar um teste e um treinamento é utilizar técnicas de *bootstrapping*, nas quais todos os dados são usados para treinar a rede e avaliar sua performance.

3.6 Aplicação de uma rede neural artificial

Segundo Selau (2000) e Ambalavanan & Carlo (2001), o primeiro passo para a configuração de uma rede consiste na separação do conjunto de dados em duas partes: uma para o treinamento e a outra para o teste da rede. Para avaliar se a predição da variável de saída coincide com a realidade, os dados usados para o teste devem ser casos novos. Também é usual subdividir o conjunto de treinamento, utilizando-se uma parte para a validação. O processo de validação serve para verificar a eficiência da rede quanto a sua capacidade de generalização durante o treinamento, podendo ser empregado como critério de parada.

O próximo passo é a definição da configuração da rede. Nessa etapa, determina-se o número de camadas, o número de unidades em cada camada, a função de ativação, etc. Normalmente, essas escolhas são feitas de forma empírica.

O terceiro passo consiste no treinamento da rede, onde os pesos das conexões são ajustados. Nessa etapa, a inicialização da rede é fundamental, pois dependendo dos valores de inicialização atribuídos aos pesos, o tempo de processamento poderá ser longo. Geralmente, os valores iniciais dos pesos da rede são números aleatórios uniformemente distribuídos.

Um aspecto importante a ser analisado durante o treinamento é a ocorrência de valores extremos (*outliers*). Como eles interferem no desenvolvimento da rede, devem ser eliminados. (StatSoft Textbook, 1998).

O quarto passo é o teste da rede. O conjunto de teste é usado identificando a performance da rede com os dados que ainda não foram trabalhados. A performance da rede medida nessa fase será uma boa indicação da performance real da rede. Finalmente, a rede deve ser integrada no sistema do ambiente operacional da aplicação.

4. Diagnóstico clínico

4.1 Introdução

As decisões médicas são essencialmente baseadas na combinação de sintomas observados nos paciente e/ou em exames laboratoriais. Mediante uma ponderação entre os sintomas observados e sua importância relativa, bem como pela experiência adquirida, o médico elabora o diagnóstico mais provável.

Geralmente um teste diagnóstico é associado com um exame realizado em laboratório ou centro de imagem. Entretanto, um conjunto de achados clínicos pode servir como um teste diagnóstico. Os objetivos mais comuns destes testes são triagens de pacientes, diagnósticos de doenças e acompanhamento ou prognóstico da evolução de um paciente.

Conforme Soares & Siqueira (1999), seja qual for o tipo de resultado produzido pelo teste diagnóstico (dados em escalas nominais, ordinais ou intervalares) os médicos tendem a reduzi-lo a uma forma mais simples. Por razões práticas, freqüentemente as informações são reduzidas a uma forma dicotômica, tais como presença/ausência de uma doença ou condição, anormal/normal ou doente/sadio.

Uma linguagem comum é utilizar a expressão “teste positivo” para representar a presença de uma doença ou condição, e “teste negativo”, em caso contrário. Exceto em situações peculiares ou triviais, não existe um teste perfeito, ou seja, que revele com certeza a presença ou ausência de uma determinada doença ou condição. Assim, são necessários procedimentos para avaliar a validade de um teste de diagnóstico e, conseqüentemente, sua adequação. Para tanto, algumas medidas úteis são descritas a seguir. Inicialmente são apresentados alguns conceitos probabilísticos básicos.

4.2 Conceitos probabilísticos básicos

Os fenômenos aleatórios são essencialmente caracterizados pelo fato de que não é possível conhecer seus resultados de forma determinística (com certeza), mesmo quando o conjunto de possíveis realizações for conhecido. Nestas situações, modelos probabilísticos são postulados para descrevê-los, pois geralmente permitem estimar a probabilidade de ocorrência dos eventos de interesse.

Na clínica médica é muito comum este tipo de situação, haja vista que, exceto em circunstâncias extremas ou especiais, o médico pode não ter “certeza absoluta” sobre a verdadeira condição do paciente (doente/não doente) ou sobre a severidade da doença. Para auxiliar sua tomada de decisão, geralmente utiliza mecanismos auxiliares, tais como exames laboratoriais ou baseados em imagens. Mas o fato é que os resultados destes exames também podem não refletir (com certeza absoluta) a verdadeira condição ou severidade da doença, caracterizando-se, portanto, como um fenômeno aleatório. (Soares & Siqueira, 1999).

O conjunto de todos os resultados possíveis de um experimento aleatório é chamado de espaço amostral, enquanto um evento é um subconjunto particular deste espaço.

Para avaliar a validade de um teste diagnóstico (exame laboratorial ou de imagem, por exemplo), costuma-se utilizar um procedimento padrão que permita comparar o seu desempenho, usualmente chamado de padrão-ouro. O padrão-ouro pode ser um exame pré-existente, cujas propriedades são conhecidas, ou um exame anatomopatológico, etc. É usado para avaliar a verdadeira condição da doença (presente ou ausente, por exemplo). Mediante estas informações, os registros são divididos em dois grupos: um grupo com a doença e outro sem a doença. Por sua vez, o teste de diagnóstico sob investigação é utilizado nos mesmos indivíduos, produzindo os resultados chamados de teste positivo (reagente) e teste negativo (não reagente). Neste contexto, portanto, existem quatro resultados (eventos) possíveis, decorrentes das combinações dos resultados do teste de diagnóstico em investigação e do padrão-ouro.

Estes resultados são mostrados na Tabela 4.1, onde: (T_+) teste positivo, (T_-) teste negativo, (D_+) indivíduos doentes e (D_-) indivíduos não doentes.

Tabela 4.1: Resultados de um teste diagnóstico

Teste	Doença		Total
	Presente (D_+)	Ausente (D_-)	
Positivo (T_+)	a	b	$a + b$
Negativo (T_-)	c	d	$c + d$
Total	$a + c$	$b + d$	n

O resultado do teste está correto quando for positivo na presença da doença, ou negativo na ausência da doença. Em contraposição, é incorreto quando for negativo na presença da doença (denominado de falso-negativo) ou quando for positivo na ausência da mesma (falso-positivo).

Segundo Fletcher et al. (1996), o objetivo de todos os estudos clínicos que descrevem o valor dos testes diagnósticos deveria ser a obtenção de dados para as quatro células de contingência da Tabela 4.1. Sem todos esses dados, as classificações erradas tornam-se mais difíceis. Na prática, entretanto, isto nem sempre é possível, pois é necessário que haja meios objetivos e válidos para se estabelecer um diagnóstico. Em determinadas circunstâncias, razões éticas ou práticas impedem sua realização.

Quando informações sobre os testes diagnósticos são obtidas, para o bem estar do paciente, os médicos os usam somente quando os testes preliminares são positivos. Como consequência, este tipo de conduta pode gerar interpretações equivocadas sobre os testes, pois somente pacientes com a doença são freqüentemente avaliados.

Conforme Soares & Siqueira (1999), no contexto de teste de diagnóstico há um interesse especial em estimar a probabilidade do teste produzir resultado positivo quando o indivíduo verdadeiramente é portador da doença e de produzir resultado negativo, na ausência da doença. Estas medidas são chamadas de probabilidades condicionais e podem ser representadas, respectivamente, por

$$P(T_+|D_+) \quad \text{e} \quad P(T_-|D_-). \quad (4.1)$$

e serão definidas na próxima seção.

4.3 Qualidade dos testes diagnósticos

O uso de um teste diagnóstico, além das considerações clínicas, requer o conhecimento de medidas que caracterizam sua qualidade, tais como a *sensibilidade*, *especificidade*, os *valores de predição positiva* e de *predição negativa*. As duas últimas refletem a capacidade do teste produzir decisões clínicas corretas.

A sensibilidade e a especificidade correspondem às probabilidades condicionais definidas na equação (4.1). No contexto da Tabela 4.1, podem ser estimadas por

$$s = P(T_+|D_+) = \frac{a}{a+b} \quad \text{e} \quad e = P(T_-|D_-) = \frac{d}{c+d}, \quad (4.2)$$

onde s e e representam respectivamente a sensibilidade e a especificidade. A primeira representa a probabilidade do resultado do teste ser positivo, quando o indivíduo é doente, e mede a capacidade de reação do teste em um paciente doente. Por sua vez, a especificidade representa a probabilidade do teste ser negativo, sabendo-se que o paciente não apresenta a doença. Em outras palavras, a sensibilidade do teste mede sua capacidade de identificar os indivíduos verdadeiramente doentes, enquanto que a especificidade mede a capacidade do teste excluir os não doentes.

É importante observar que, subjacente a esses conceitos, assume-se a existência de um padrão-ouro, isto é, de um teste diagnóstico que sempre produz resultados corretos, o que pode ser difícil de se encontrar em muitas situações.

Segundo Fletcher et. al. (1996), ao selecionar um teste é necessário, assim, considerar sua sensibilidade e sua especificidade. Um teste sensível deve ser escolhido para os casos em que a penalidade por deixar de diagnosticar uma doença é grande. Testes sensíveis são úteis também nas fases iniciais de um processo diagnóstico, quando um grande número de possibilidades está sendo considerado e deseja-se reduzi-las. Os testes diagnósticos são usados nessas situações para excluir doenças, isto é, estabelecer que certas doenças são improváveis. Como por exemplo, poder-se-ia escolher precocemente um teste de anticorpo para HIV na avaliação de infiltrados pulmonares com perda de peso, para excluir uma infecção relacionada à AIDS. Assim, o teste sensível é mais útil ao clínico quando o resultado do teste é negativo. Testes específicos são úteis para confirmar um diagnóstico sugerido por outros dados. Isso porque um teste altamente específico é raramente positivo na ausência da doença, ou seja, há poucos falsos-positivos. Esse tipo de teste é adotado quando os resultados falsos-positivos podem lesar o paciente física, emocional ou financeiramente. Por exemplo, antes do paciente com câncer ser submetido à quimioterapia, que é um procedimento que acarreta riscos, trauma emocional e custos financeiros, é necessário que se faça diagnóstico tecidual, ao invés de confiar em exames menos específicos. Assim, é mais indicado quando o resultado do teste é positivo.

Embora a sensibilidade e a especificidade sejam índices ilustrativos e bons sintetizadores das qualidades gerais de um teste, apresentam uma limitação: não auxiliam a decisão da equipe médica, que recebendo um paciente com o resultado positivo do teste, precisa avaliar se o paciente está ou não doente. Isto não pode ser realizado apenas com as medidas de sensibilidade e especificidade, pois estes índices são resultantes de situações em que há certeza total sobre o diagnóstico. São necessários, portanto, medidas que reflitam de forma mais adequada à realidade prática.

Para tanto, podem ser úteis as medidas denominadas valor de predição positiva (VPP) e o valor de predição negativa (VPN), definidos respectivamente por

$$VPP = P[D_+|T_+] \quad \text{e} \quad VPN = P[D_-|T_-]. \quad (4.3)$$

O valor de predição positiva (VPP) é a probabilidade do paciente estar realmente doente quando o resultado do teste é positivo, enquanto que o valor de predição negativa (VPN) representa a probabilidade do paciente não estar doente quando o resultado do teste é negativo. Na prática, as probabilidades VPP e VPN podem ser determinadas mediante as informações dispostas na Tabela 4.2, descritas por Vecchio (1966) e reproduzidas em Soares & Siqueira (1999). Note que, aqui, p representa a prevalência da doença na população.

Tabela 4.2 – Cálculo do VPP e VPN

População	Proporção	Proporção com resultado	
		Positivo	Negativo
Doente	p	ps	$p(1-s)$
Sadia	$(1-p)$	$(1-p)(1-e)$	$(1-p)e$
Total	$p + (1-p)$	$ps + (1-p)(1-e)$	$p(1-s) + (1-p)e$

Assim, o valor de predição positiva é definido pela razão entre a frequência dos verdadeiros positivos (indivíduos doentes) e o total de positivos, conforme a expressão

$$VPP = \frac{ps}{ps + (1-p)(1-e)}. \quad (4.4)$$

De forma análoga, o valor de predição negativa é definida por

$$VPN = \frac{(1-p)e}{p(1-s) + (1-p)e}. \quad (4.5)$$

Como mencionado anteriormente, há também o risco do teste levar a decisões incorretas, as quais são denominadas de resultado falso-positivo e falso-negativo, cujas probabilidades são definidas pelas equações (4.6) e (4.7), respectivamente.

$$PFP = P[D_-|T_+] = 1 - P[D_+|T_+] = 1 - VPP \quad (4.6)$$

$$PFN = P[D_+|T_-] = 1 - P[D_-|T_-] = 1 - VPN \quad (4.7)$$

É conveniente observar que também é usual definir a probabilidade falso-positivo como o complementar da especificidade, ou seja, $PFP = 1 - e$ e, analogamente, $PFN = 1 - s$. Veja, por exemplo, Soares & Siqueira (1999) e Woodward (1999, p.95).

4.4 Combinação de testes diagnósticos

Muitas vezes, para o diagnóstico de algumas doenças, os testes disponíveis apresentam valores VPP ou VPN baixos ou, então, o melhor teste é muito caro ou oferece risco ao paciente. Nestas circunstâncias, uma opção freqüentemente adotada é o uso de uma combinação de testes mais simples ou baratos. A associação de testes eleva a qualidade do diagnóstico, diminuindo o número de resultados errados.

Quando dois ou mais testes são usados para se chegar a um diagnóstico, é preciso saber como são obtidos os índices de qualidade do teste múltiplo, composto pela agregação dos outros.

Segundo Soares & Siqueira (1999), as combinações mais fáceis são aquelas em série ou em paralelo. No caso do teste em paralelo, quando um dos dois testes for positivo, o teste conjunto também será. O teste em série é considerado positivo apenas se os dois testes individuais forem positivos.

A associação de testes em série é empregada em triagens (procedimento para classificar pessoas assintomáticas quanto à probabilidade de terem ou não determinada doença), diagnósticos individuais (testes caros e arriscados) e quando o custo é relevante. Os diagnósticos individuais também podem ser empregados em associações em paralelo, quando há casos de urgência ou quando os pacientes residem em lugares distantes.

Utilizando a linguagem probabilística, o teste em paralelo (T_p) e o teste em série (T_s) podem ser representados por

$$T_p = A_+ \cup B_+ \quad (4.8)$$

$$T_s = A_+ \cap B_+ \quad (4.9)$$

A sensibilidade e a especificidade dos testes T_p e T_s serão calculadas através de regras probabilísticas. Assim, para testes em paralelo, o resultado do teste será positivo se pelo menos um dos testes for positivo, como ilustrado na Tabela 4.3.

Tabela 4.3: Resultado de um teste em paralelo

Teste A	Teste B	Teste em Paralelo
-	-	-
-	+	+
+	-	+
+	+	+

Sua sensibilidade é determinada por

$$P[T_{p+} | D_+] = P[A_+ \cup B_+ | D_+] = P[A_+ | D_+] + P[B_+ | D_+] - P[A_+ \cap B_+ | D_+]$$

$$s_p = s_A + s_B - s_A \times s_B, \quad (4.10)$$

e a especificidade é tal que

$$P[T_{p-} | D_-] = P[A_- \cap B_- | D_-] = P[A_- | D_-] \times P[B_- | D_-]$$

$$e_p = e_A \times e_B. \quad (4.11)$$

Os parâmetros VPP e VPN são calculados através das equações (4.4) e (4.5), considerando a sensibilidade e a especificidade do teste em paralelo, bem como a prevalência da população de interesse.

Um teste é dito em série se os testes que o compõem são aplicados consecutivamente, de tal forma que o segundo teste só é aplicado se o primeiro apresentar resultado positivo. Esse procedimento é indicado quando o paciente pode ser acompanhado ao longo do tempo. A Tabela 4.4 resume o funcionamento de um teste em série com duas componentes.

Tabela 4.4: Resultado de um teste em série

Teste A	Teste B	Teste em série
-	desnecessário	-
+	-	-
+	+	+

Nesta situação, as medidas de sensibilidade e especificidade serão determinadas nas equações (4.12) e (4.13), respectivamente,

$$P[T_{s+} | D_+] = P[A_+ \cap B_+ | D_+] = P[A_+ | D_+] \times P[B_+ | D_+] \\ s_s = s_A \times s_B \quad (4.12)$$

$$e_s = e_A + e_B - e_A \times e_B . \quad (4.13)$$

Os valores VPP e VPN são calculados de maneira similar aos testes em paralelo. Para leitores mais interessados, recomenda-se Soares & Siqueira (1999) e Fletcher et. al. (1996).

4.5. Testes Diagnósticos baseados em variáveis contínuas

Até o momento, os índices construídos para determinar a qualidade de um teste eram baseados em resultados dicotômicos. Entretanto, variáveis contínuas são bastante freqüentes em testes de diagnóstico. Surge, então, a questão de como dicotomizar uma variável contínua, para que os resultados possam ser expressos em uma tabela 2x2, como anteriormente. Assim, será necessário determinar um ponto de corte para classificar os indivíduos como sadios ou doentes, ou seja, identificar o valor a partir do qual o teste é positivo ou negativo.

Cada teste clínico é caracterizado por um valor de referência e, variando-se este ponto de corte, a sensibilidade e a especificidade variam também. Uma alternativa para se determinar o ponto de corte é calcular a sensibilidade e a especificidade para vários valores de referência e adotar, então, aquele valor que produz a melhor combinação para as medidas da qualidade do teste. Para tanto, um procedimento bastante utilizado é a curva característica de operação (*ROC – Receiver Operator Characteristic*, em inglês), descrita na próxima seção.

4.6 Curva característica de operação (Curva ROC)

Segundo Fletcher et. al. (1996), a curva ROC mostra a intensidade do balanço desejável entre a sensibilidade e a especificidade de um teste. Usualmente, pode ser usada para auxiliar na decisão do melhor ponto de corte que, em geral, está localizado no “ombro” da curva, ou próximo a ele.

Testes de alto poder discriminatório concentram-se no canto superior esquerdo da curva ROC, pois à medida que a sensibilidade aumenta (diminuindo o ponto de corte) há pouca perda de especificidade. Já os teste com menor poder discriminatório aproximam-se da diagonal. Esta diagonal revela a relação entre as taxas de resultados da sensibilidade (verdadeiros-positivos) com o complementar da especificidade (falsos-positivos) que seria obtida pelo teste sem informação diagnóstica, por exemplo, se o clínico jogasse uma moeda.

As curvas ROC também podem ser usadas para avaliar a habilidade de discriminação de vários métodos estatísticos que geram previsões de diagnóstico. Muitas vezes, as curvas ROC são avaliadas qualitativamente, conforme a descrição anterior, pela sua facilidade. Entretanto, elas também podem ser avaliadas de forma quantitativa.

O método quantitativo mais utilizado é a área sob a curva, que avalia a acurácia global do diagnóstico. A área sob a curva corresponde à probabilidade de identificação correta dos doentes e não doentes e quanto maior a área sob a curva, melhor é o teste.

Hanley & McNeil (1982) estabelecem uma relação entre a área sob a curva ROC e a probabilidade do par de pacientes doente e não doente estar corretamente classificado, chamada de θ . Segundo eles, com uma amostra infinita de pacientes, a área sob a curva ROC é igual a essa probabilidade θ . A distribuição de probabilidade esperada dos doentes é maior que a dos não doentes. Sem considerar suposições sobre a forma das distribuições amostrais das populações de doentes e não doentes, a área sob a curva ROC pode ser quantificada pela probabilidade dos postos de doentes e não doentes tanto pela natureza com pela forma da distribuição exata. A estatística de Wilcoxon, permite estimar a probabilidade dos pares estarem corretamente ordenados conforme seus postos.

Assim, pode-se dizer que há uma equivalência entre a área sob a curva ROC, a probabilidade do par doente/não doente e a estatística de Wilcoxon. Entretanto há razões práticas para se considerar suposições sobre a distribuição amostral da população de doentes e não doentes. Os investigadores estão freqüentemente interessados nos outros aspectos da curva ROC, como pontos de cortes entre sensibilidade e especificidade.

No teste de Wilcoxon admite-se que as qualidades de uma certa variável x são maiores na população dos doentes do que na dos não doentes, sem admitir qualquer suposição para a variável x . Assim, x será um bom discriminador se sua probabilidade estiver próxima de 1 e classificará igualmente as duas populações se sua probabilidade for igual a 0,5. Assim, pode-se dizer que a estatística W de Wilcoxon, pode representar a probabilidade θ . Os escores médios de W variam de 0 a 1, refletindo quanto a probabilidade do discriminador x calculado para a população de doentes, é melhor que o mesmo discriminador calculado sob a população de não doentes. Obviamente que, em alguns casos, o discriminador para os doentes será dependente do discriminador para os não doentes. Essas correlações serão consideradas pelo erro padrão da estatística W .

O interesse principal em calcular o erro padrão da estatística W é avaliar o comportamento da área sob a curva ROC em amostras similares de diferentes tamanhos. Conforme mostrado por Hanley & McNeil (1982), o erro da estatística W poderá ser uma estimativa para θ .

Para se comparar duas áreas sob a curva ROC derivadas da mesma população de pacientes pode-se utilizar o método proposto por Hanley & McNeil (1983). Assim, calcula-se a razão crítica z , definida por

$$z = \frac{A_1 - A_2}{\sqrt{SE_1^2 + SE_2^2 - 2rSE_1SE_2}}, \quad (4.14)$$

onde A_1 e SE_1 referem-se à área observada e o erro padrão da curva ROC associada ao teste 1; A_2 e SE_2 referem-se às quantidades correspondentes para o teste 2; r representa a correlação estimada entre os dois testes. A distribuição da estatística z definida em (4.14) é assintoticamente normal.

A área sob a curva ROC e os erros padrões podem ser determinados pela estatística W de Wilcoxon, conforme mencionado anteriormente. A correlação poderá ser calculada pelo método de Pearson ou Tau de Kendall.

Outra maneira de se comparar dois testes através da sensibilidade e da especificidade é através do critério da razão de verossimilhança, definida por

$$\frac{s}{1-e} \quad (4.15)$$

onde s é a sensibilidade e e a especificidade. Essa razão aumenta se a sensibilidade ou a especificidade aumenta, mantendo a outra constante, ou se ambas aumentam. Assim, o teste com maior razão será considerado o melhor. Entretanto, se a sensibilidade e a especificidade se maximizam em direções contrárias, a razão decresce. Conseqüentemente, o uso da razão de verossimilhança definida em (4.15), tem limitações práticas. Veja Woodward (1999, p.97).

5. Estudo de caso: Predição da macrossomia fetal no EBDG

5.1 Caracterização do EBDG

O Estudo Brasileiro do Diabetes Gestacional – EBDG, é um estudo de coorte conduzido em seis capitais do país, com início em maio de 1991 e término em agosto de 1995. O objetivo geral é estudar o diabetes e a intolerância à glicose gestacional em grávidas com atendimento obstétrico junto ao SUS, em relação a alguns fatores de risco, quanto a prevalência e outros desfechos. Para esse trabalho o desfecho de interesse é a macrossomia fetal, caracterizada mais adiante.

O estudo foi constituído de 4 fases: na primeira fase, realizou-se o arrolamento das gestantes, entrevistas, medições antropométricas e um teste de tolerância à glicose (TTG). Na segunda fase, foi realizado o acompanhamento das gestantes, através da revisão dos prontuários, até o parto. A terceira fase foi composta pela obtenção das informações referentes ao parto e ao recém-nascido. Na quarta fase foi avaliada a morbi-mortalidade do recém nascido, através de um acompanhamento de até 28 dias após o parto.

O EBDG teve caráter multicêntrico com um conjunto de investigadores denominados de grupo de estudo. O grupo teve início no Ministério da Saúde, abrangendo a Secretaria Nacional de Ações de Saúde, o Departamento de Programas e a Coordenação de Doenças Crônicas. Após, formou-se um comitê de controle de qualidade, um centro de coordenação e grupos consultivos. A coordenação do estudo ficou sob responsabilidade do centro de Porto Alegre. O comitê e os grupos foram divididos nos demais centros: Salvador, Belo Horizonte, Rio de Janeiro, Fortaleza, São Paulo e Manaus.

Em cada centro de trabalho uma equipe foi formada por um coordenador, supervisores, entrevistadores e pessoal de apoio, treinados e supervisionados pela coordenação nacional do estudo.

5.2 Delineamento e descrição da amostra

A amostra foi constituída por gestantes com idade superior a 20 anos, que se encontravam entre a 21^a e a 28^a semana de gestação e sem histórico de Diabete Mellitus e se apresentavam para consulta em algum dos centros das seis capitais brasileiras

O estudo foi de natureza observacional. As decisões relativas ao controle de hiperglicemia foram deixadas para os clínicos que avaliaram as pacientes durante o atendimento obstétrico. As informações sobre a terapia e dieta de insulina foram obtidas através da revisão dos prontuários ou da carteira do pré-natal.

A idade gestacional das pacientes foi avaliada por quatro parâmetros: ecografia obstétrica, data da última menstruação, altura uterina, medida pelo obstetra no dia do arrolamento ou no dia imediatamente anterior ao arrolamento e estimativa da idade gestacional no parto (CAPURRO).

No início foram arroladas 5564 mulheres entre maio de 1991 até agosto de 1995, mas apenas 4998 (90%) concluíram o cronograma final. Para focalizar a avaliação de risco de escala da hiperglicemia, a qual fora da gravidez é considerada prejudicada pela tolerância a glicose, 21 mulheres foram excluídas da análise, pois apresentavam critérios de diabete (glicose de jejum $> 7,0$ mmol/l ou glicose de 2 horas $> 11,1$ mmol/l). Para investigar a associação da Diabete Mellitus Gestacional com a macrossomia estão disponíveis informações sobre 3925 mulheres (79% das mulheres que completaram o estudo). Devem ser excluídos, também, 546 devido a gemelaridade ou por falta de acompanhamento, 214 com informações incompletas nas covariáveis, 92 devido à ausência de informações sobre o peso de nascimento dos bebês, 198 deixadas a priori para completar as 35 semanas de gestação e 2 pela insulina. Considerando a idade, cor da pele, nível educacional, status nutricional e paridade (quantidade de filhos que tiveram), as mulheres que permaneceram na análise são similares às 1052 excluídas. Após essas exclusões adicionais, permaneceram na análise 3925 mulheres, tendo sido observados 379 (9,7%) casos de macrossomia fetal.

5.3 Métodos de coleta

Todas as mulheres responderam um questionário estruturado aplicado por entrevistadores especialmente treinados. As equipes de campo foram instruídas, em cada uma das etapas da coleta de dados, por um manual orientador, desenvolvido a partir de estudos pilotos realizados em Porto Alegre.

Juntamente com o arrolamento, foram realizadas as entrevistas e as medições antropométricas. Logo após, as pacientes foram convidadas a realizar um teste de tolerância à glicose (TTG) de 2 horas com 75 gramas entre a 24^a e a 28^a semanas de gestação. Então, elas foram acompanhadas até a chegada ao hospital e durante o pós-parto através da revisão do prontuário.

O teste de tolerância à glicose utilizou procedimentos padrões. Resumidamente, administrou-se uma carga de 75 gramas de glicose anidroxica após 12 a 14 horas de jejum e, obtendo o jejuado, amostras de 1 e 2 horas da veia antecubital. As amostras foram coletadas em tubos contendo fluido e mantidas a uma temperatura de 4° C até a centrifugação. Nas medidas de plasma, mediante métodos de glicose, utilizou-se um coeficiente de variação menor que 5%.

Diabete de Mellitus Gestacional foi definida de acordo com as novas recomendações da ADA (Associação Americana de Diabetes) para o teste de tolerância a glicose de 2 horas como tendo no mínimo dois valores maiores a uma glicemia de jejum de 5,3 mmol/l, a uma glicose de 1 hora maior que 10 mmol/l ou a uma glicose de 2 horas de 8,6 mmol/l. Isso foi adicionalmente definido pelo critério menos severo que o de WHO (jejum \geq 7,0 mmol/l ou 2 horas \geq 7,8 mmol/l).

5.4 Medidas

O desfecho principal do estudo é a ocorrência ou não de macrossomia fetal, definida como o peso de nascimento superior ao percentil 90 (grande para idade gestacional), para a correspondente idade gestacional.

Os valores do percentil 90, derivados da tabela de frequências do peso de nascimento, para as correspondentes idades gestacionais, são mostrados na Tabela 5.1.

Tabela 5.1 – Percentis da macrossomia

Semanas	Peso (g)
35	3425
36	3500
37	3590
38	3700
39	3790
40	3920
41	4040
42	4080
43	4200

As mortes perinatais foram definidas pela perda de peso fetal > 1 kg ou com idade gestacional estimada ≥ 28 semanas ou por uma morte neonatal cedo (acima de 7 dias). Pré-eclampsia (ou eclampsia), foi classificada de acordo com os critérios do Grupo Nacional do Funcionamento do Programa de Educação da Pressão Alta como sendo hipertensão depois de 20 semanas de gestação associada com proteinúria ou convulsões, considerando-se um novo começo ou superimposta sobre uma hipertensão crônica.

O índice de massa corporal foi calculado usando relatos do peso pré-gravídicos e os valores foram agrupados em categorias de acordo como estado nutricional em consonância com as recomendações da WHO 1997.

A idade gestacional foi definida de acordo com critérios hierárquicos baseados em quatro exames clínicos: ultra-sonografia antes de 26 semanas, para 52% da amostra, ultra-som para 15%; relatos da última menstruação conjuntamente com a estimação da idade neonatal ou altura uterina, para 23%, e idade gestacional estimada, ultra-som depois de 26 semanas, altura uterina ou último período menstrual para os outros 10%.

O hábito de fumar das gestantes foi registrado mediante três categorias: se nunca fumaram, fumavam mas tinham parado ou se ainda fumavam. Por sua vez, o ganho de peso na gravidez foi calculado através da diferença entre o peso no arrolamento e o peso no início da gestação.

Dentre os principais preditores, estão a idade gestacional do bebê, bem como a idade, a altura, a glicose e a obesidade materna, avaliada através das dobras cutâneas.

5.5 Predição da macrosomia fetal

Inicialmente, o banco de dados foi dividido em duas partes, denominadas de “Aprendizado” e “Teste”. A primeira, composta por 2964 observações (80%), foi utilizada para construir a rede neural e ajustar o modelo logístico, enquanto que a segunda, com 747 registros, foi usada para avaliar a capacidade preditiva dos métodos. Esse método de divisão foi sugerido por Ambalavanan (2001). A divisão do banco de dados se deu através de rotina computacional onde foi gerada uma variável aleatória uniformemente distribuída usada para alocar, ao acaso, os registros aos conjuntos mencionados. Esta rotina está disposta no Anexo A. Desta forma, a Tabela 5.2 apresenta a distribuição empírica de frequências da macrosomia fetal, para o conjunto de dados completo e, também, separadamente para as bases de aprendizado e teste”.

É importante salientar que 379 gestantes foram eliminadas da amostra por não apresentarem todas as informações nas variáveis analisadas.

Tabela 5.2 - Divisão do banco em aprendizado e teste

Conjunto de dados	Macrossomia Fetal		Total
	Sim	Não	
Teste	72 (9,64%)	675 (90,36%)	747
Aprendizado	292 (9,85%)	2672 (90,15%)	2964
Completo	364	3347	3711

5.5.1 Ajuste do modelo de regressão logística

O modelo de regressão logística foi construído através do procedimento PROC LOGISTIC do programa estatístico SAS. Inicialmente, para compreender melhor o comportamento dos fatores explanatórios, ainda com um enfoque descritivo, foram ajustados modelos univariados. Nessa etapa, para as variáveis quantitativas foram considerados os termos lineares e quadráticos, a fim de avaliar a natureza da relação com o desfecho. Para as variáveis categóricas foram geradas variáveis dicotômicas tais que, em conjunto, identificam as observações associadas a cada categoria. As variáveis são apresentadas na Tabela 5.3.

As informações geradas nos modelos univariados foram usadas para ajustar um primeiro modelo logístico múltiplo, no qual foram incorporados os fatores explanatórios que se mostraram significativos ao nível de significância 10%. Para tanto, o método automático de seleção de modelos stepwise foi usado nesta etapa, adotando-se o nível de significância 5% como critério para entrada ou saída das variáveis do modelo. O modelo final foi construído com a inclusão dos termos lineares que estavam ausentes, bem como com a inclusão das variáveis clinicamente importantes. Esse procedimento está apresentado no Anexo B1.

Tabela 5.3: Nome, descrição e tipo das variáveis

Nome da Variável	Descrição	Tipo
altura	Altura da paciente (em cm)	razão
altsq	Altura ao quadrado	razão
bmi	Índice de massa corporal	razão
bmisq	Índice de massa corporal ao quadrado	razão
circbra	Circunferência do braço	razão
cirqbrasq	Circunferência do braço ao quadrado	razão
escola	Escolaridade da gestante	nominal
gravi	Número de gestações anteriores	razão
idade	Idade da gestante	razão
idadesq	Idade da gestante ao quadrado	razão
pari	Outros partos	binária
pgabic	Pregas Bic	razão
pregas	Soma das pregas	razão
rescvjej	Glicemia de jejum	razão
rescv1h	Glicemia de 1 hora	razão
rescv2h	Glicemia de 2 horas	razão
wtgain	Ganho de peso	razão
wtgsq	Ganho de peso ao quadrado	razão
quadril	Circunferência do quadril	razão
quadrilsq	Circunferência do quadril ao quadrado	razão
cintura	Circunferência da cintura	razão
cinsq	Circunferência da cintura ao quadrado	razão
peso	Peso da gestante (em kg)	razão
hfdm2	Histórico familiar de macrossomia	binária
fumo	Pacientes fumantes, pacientes que já fumaram e que nunca fumaram cigarro	nominal
hxmacro	Histórico de Macrossomia	binário

O modelo é constituído pelas seguintes variáveis: altura, cintura, ganho de peso durante a gestação (diferença entre o peso no arrolamento pelo peso no parto), hábito de fumar e histórico de macrosomia fetal, e as estimativas dos parâmetros são mostradas na Tabela 5.4. Os resultados sugerem que esse modelo está bem ajustado, pois o valor observado para a estatística chamada de Deviance é igual 1779,8701, cuja probabilidade de significância associada é $p = 1,000$. Este resultado pode ser visualizado no Anexo B2.

Ainda, o valor observado da estatística de Wald foi igual a 124,43. Comparando-o com o valor tabelado da distribuição Qui-quadrado com seis graus de liberdade, rejeita-se a hipótese nula de que todos os coeficientes estimados sejam iguais a zero. A significância dos coeficientes individuais está representada na Tabela 5.4 e detalhes podem ser encontrados Anexo B2.

Tabela 5.4: Modelo logístico final

	Estimativa ($\hat{\beta}$)	Erro Padrão	P	Razão de Chances (IC95%)
Intercepto	-8,976	1,545	< 0,0001	
Altura	0,018	0,010	0,0702	1,018 (0,999; 1,038)
Cintura	0,044	0,008	< 0,0001	1,045 (1,030; 1,061)
Ganho de peso	0,044	0,012	0,0003	1,045 (1,020; 1,070)
Fumo				
Ex-Fumantes	-0,057	0,149	0,7027	0,945 (0,705; 1,266)
Fumantes	-0,889	0,215	< 0,0001	0,411 (0,270; 0,626)
Histórico de Macrossomia	1,007	0,178	< 0,0001	2,736 (1,930; 3,879)

Em seguida, as estimativas dos parâmetros ajustados foram usadas para avaliar a capacidade preditiva do modelo. Isto é realizado obtendo-se os valores preditos, pela equação ajustada, associados aos registros no conjunto de dados teste. A rotina computacional está disposta no Anexo B3.

Assim, as probabilidades preditas, calculadas para o conjunto de teste, foram ordenadas e agrupadas em 10 grupos de tamanhos iguais, os quais são brevemente descritos na Tabela 5.5. O grupo 1 é aquele em que se observa as menores estimativas da probabilidade de ocorrência de macrossomia, sendo que o valor médio é igual 0,030. No último grupo, por sua vez, o valor médio para esta probabilidade é 0,259. Na Tabela 5.6 são apresentadas as estimativas de sensibilidade e especificidade produzidas pelo modelo ajustado. Por sua vez, a curva característica de operação é ilustrada na Figura 5.1, cuja área sob a curva é 0,669 com erro padrão de 0,036. Anexo B4.

Tabela 5.5: Médias e desvios-padrão dos grupos.

	Valor predito pelo modelo	Função Logito
Grupo 1	-3,486 ± 0,214	0,030 ± 0,005
Grupo 2	-3,074 ± 0,090	0,044 ± 0,004
Grupo 3	-2,865 ± 0,050	0,054 ± 0,003
Grupo 4	-2,703 ± 0,049	0,063 ± 0,003
Grupo 5	-2,533 ± 0,037	0,074 ± 0,003
Grupo 6	-2,394 ± 0,043	0,084 ± 0,003
Grupo 7	-2,250 ± 0,044	0,095 ± 0,004
Grupo 8	-2,059 ± 0,068	0,113 ± 0,007
Grupo 9	-1,781 ± 0,098	0,145 ± 0,012
Grupo 10	-1,092 ± 0,417	0,259 ± 0,084

Média ± Desvio padrão

Tabela 5.6 - Estimativas da sensibilidade e especificidade no modelo logístico.

Grupo	Sensibilidade (%)	Especificidade (%)
10%	30,56%	95,01%
20%	40,28%	88,48%
30%	51,39%	82,05%
40%	59,72%	75,43%
50%	72,22%	69,10%
60%	77,78%	62,38%
70%	84,72%	55,66%
80%	88,89%	48,75%
90%	94,44%	41,94%
100%	100%	35,22%

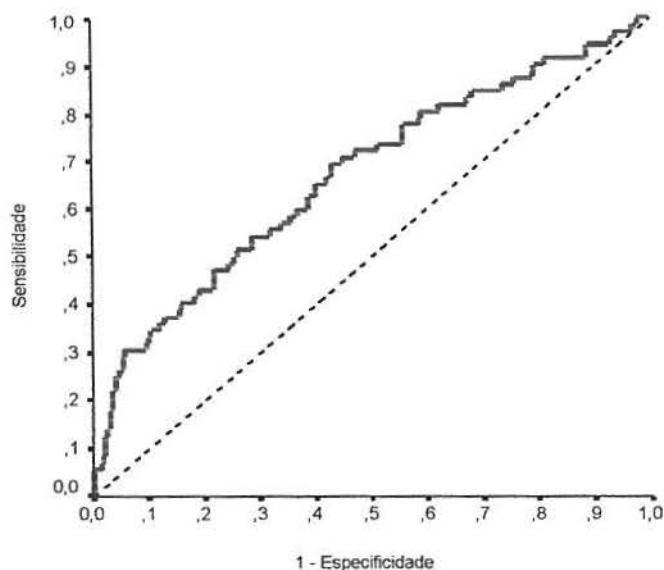


Figura 5.1: Curva característica de Operação para a regressão

5.5.2 Ajuste da rede neural

A rede neural foi construída através do procedimento Enterprise Miner do programa SAS, Versão 8.1, tendo como base as variáveis de entrada usadas no modelo logístico. Foram consideradas duas camadas intermediárias e um nó de saída. A camada de saída apresentava dois neurônios, um para cada categoria da variável de entrada. Foram permitidas conexões entre as camadas de entrada e as camadas intermediárias, das camadas intermediárias com a camada de saída e conexões diretas entre os valores de entrada com os valores de saída, representando as relações lineares das variáveis. Entre as camadas intermediárias e a camada de entrada existe um nó denominado de viés que auxilia a função de ativação, dando mais graus de liberdade ao modelo. Esse viés também foi introduzido nas camadas de saída, uma vez que conexões diretas entre as entradas e saídas foram aceitas.

A rede foi treinada pelo método de alimentação para frente “*feedforward*”. O método propagação para trás foi utilizado para o cálculo do gradiente do erro. O diagrama da rede foi construído de variáveis de razão (altura, cintura e ganho de peso) e nominais (Ex-Fumantes, fumantes, histórico de macrossomia), constituindo a camada de entradas ligadas às outras camadas (macrossomia). A função de ativação escolhida foi a função *threshold* para que as estimativas se assemelhem às obtidas pela regressão logística. A Figura 5.2 representa o diagrama da rede. Os comandos encontram-se no Anexo C1.

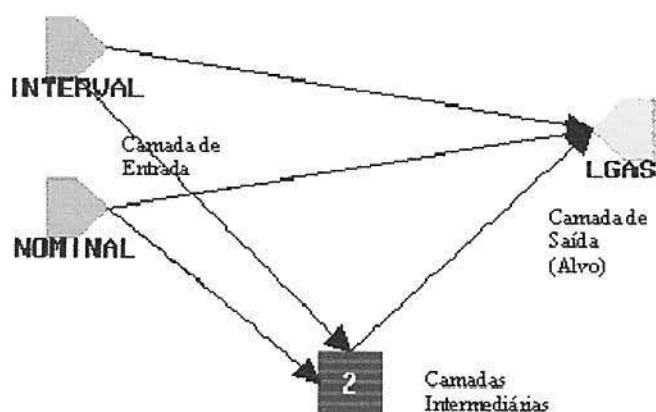


Figura 5.2: Diagrama da Rede Neural

O critério de seleção do modelo foi o erro médio, o qual estabelece o menor erro médio para a generalização da rede. Conforme o diagrama da rede, verifica-se que a arquitetura da rede considerada foi a de “*perceptron*” com várias camadas (*Multiple Perceptron*).

A função de ativação usada é a função *threshold*, enquanto que a função objetivo utilizada é a *default* do software, fazendo com que o nó escolha aquela que mais se adapta às outras funções de erros associados aos outros nós.

A rede neural foi otimizada na 32^a iteração, obtendo como resultado da função objetivo 0,2950, com 23 parâmetros estimados.

Os coeficientes estimados para as variáveis e seus respectivos pesos são mostrados na Tabela 5.7. Os Resultados encontra-se no Anexo C.2.

Tabela 5.7 - Estimativas dos parâmetros otimizados pela rede

Parâmetro	Estimativa	Função Gradiente Objetivo
Altura _ H11	-2,6932	0,000021
Cintura _ H11	0,9889	-0,000011
Wtgain _ H11	0,4250	-0,000007
Altura _ H12	3,2442	-0,000049
Cintura _ H12	-3,4486	0,000431
Ganho de peso _ H12	-0,8674	0,000120
Ex-Fumantes _ H11	-1,4039	0,000018
Fumantes _ H11	5,0705	-0,000033
Histórico de Macrossomia _ H11	4,1792	-0,000017
Ex-Fumantes _ H12	0,4644	-0,000118
Fumantes _ H12	1,0358	-0,000106
Histórico de Macrossomia _ H12	0,4623	-0,000092
Viés _ H11	-2,6599	0,000013
Viés _ H12	9,6964	0,000071
H11 _ Macrossomia	0,7838	0,0000001
H12 _ Macrossomia	0,9902	-0,000047
Altura _ Macrossomia	0,0950	-0,000023
Cintura _ Macrossomia	0,4986	-0,000062
Ganho de Peso _ Macrossomia	0,2331	0,000016
Ex-Fumantes _ Macrossomia	-0,0191	-0,000011
Fumantes _ Macrossomia	-0,9635	-0,000011
Histórico de Macrossomia _ Macrossomia	0,3106	-0,000012
Viés _ Macrossomia	-3,1309	0,000021

Legenda: H11 – Primeira camada intermediária; H12 – segunda camada intermediária

A rede foi testada e os valores preditos pelo modelo foram calculados. Com os valores preditos para os dados do teste, calculou-se a sensibilidade e a especificidade do modelo de forma análoga ao procedimento usado no para o modelo logístico.

A Tabela 5.8 apresenta as estimativas da sensibilidade e da especificidade. Detalhes são apresentados no Anexo C.4.

A curva característica de operação foi construída da mesma forma que no modelo logístico e está representada na Figura 5.3, cuja área sob a curva é igual a 0,623, com erro padrão de 0,039.

Tabela 5.8 - Estimativas da sensibilidade e especificidade pela rede neural

	Sensibilidade (%)	Especificidade (%)
10 %	30,56%	92,30%
20%	37,50%	81,93%
30%	48,61%	72,00%
40%	55,56%	61,63%
50%	61,11%	51,11%
60%	66,67%	40,74%
70%	80,76%	31,11%
80%	86,11%	20,59%
90%	91,67%	10,07%
100%	100%	0%

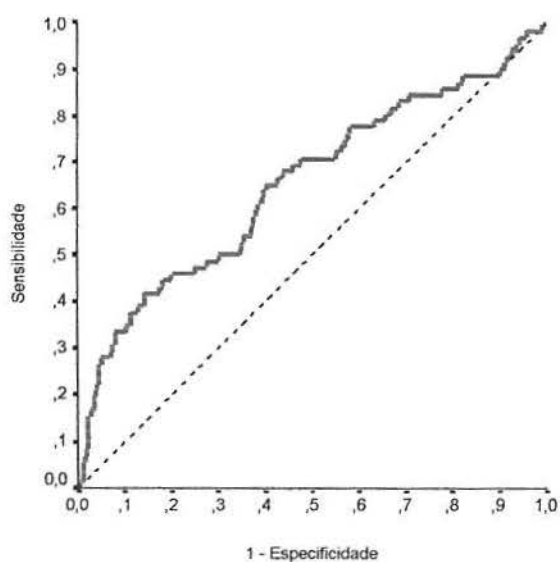


Figura 5.3: Área sob a curva ROC construída para a rede neural

5.6 Comparação entre regressão logística e redes neurais

O teste proposto por Hanley & McNeil (1983) para comparar duas áreas sob a curva ROC pode ser usado para comparar as capacidades preditivas do modelo logístico descrito na Tabela 5.4 e da rede neural artificial apresentada na Tabela 5.7. Da mesma maneira que no estudo de Duh (1998), as duas curvas ROC são derivadas da mesma população, e assim o erro padrão da diferença entre as duas áreas deve considerar a correlação entre os dois modelos. Assim, o valor observado da estatística de teste definida pela equação (4.14) é $z = 0,80$, cuja probabilidade associada é 0,4237, sugerindo que as áreas sob a curva ROC são iguais. O valor observado para correlação de Pearson entre as probabilidades preditas pelo modelo de regressão logística e pela rede neural é igual a 0,854.

O comportamento simultâneo das curvas ROC pode ser visualizado na Figura 5.4, onde claramente não se observa uma diferença significativa entre elas.

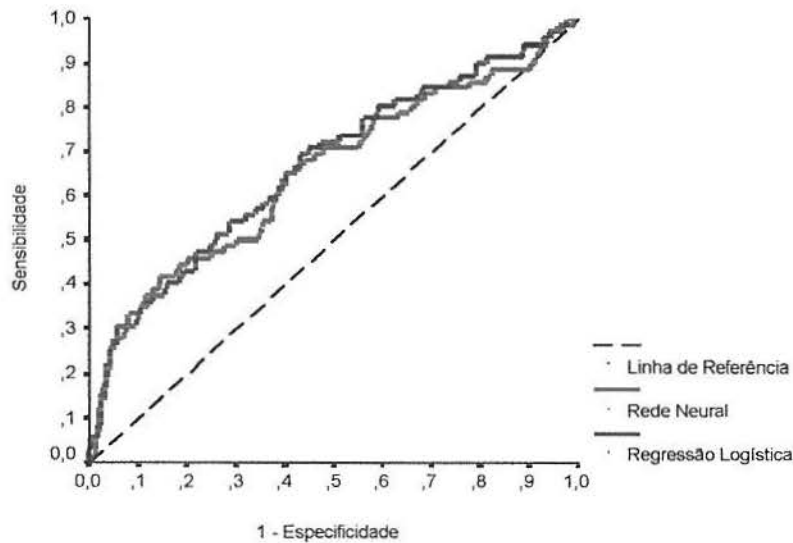


Figura 5.4: Curva característica de operação para os dois ajustes.

Uma maneira alternativa para comparar modelos é utilizar a medida de Akaike, sendo que o melhor modelo é aquele que apresenta o menor valor para esta estatística. Para o modelo logístico, o valor observado da estatística de Akaike foi igual a 1793,87, enquanto que para a rede neural foi igual a 1794,97. Claramente, estes resultados estão em consonância com aquele do teste de Hanley & McNeil (1983).

5.7 Vantagens e desvantagens da regressão logística e da rede neural

Pesquisadores da área médica estão bastante familiarizados com a utilização da regressão logística para fazer previsões, mas o mesmo não ocorre com a metodologia da rede neural. No entanto, estas metodologias possuem peculiaridades que, em determinadas situações, podem produzir resultados discrepantes. Assim, elas possuem vantagens e desvantagens, como foi salientado por Tu (1996) e Ohno-Machado (1999).

Uma grande vantagem das redes neurais é que elas possuem uma grande capacidade para descrever as relações não lineares das variáveis. Isso pode facilitar a solução de alguns problemas médicos, quando não estão claramente estabelecidas as relações entre as variáveis dependentes e as explanatórias.

Se existir uma quantidade substancial de relações não lineares entre as variáveis de entrada com as de saída, a rede ajustará automaticamente os pesos das conexões nessa estrutura para refletir as relações não lineares, através das camadas intermediárias. A regressão logística também pode ser usada para identificar as relações não lineares do modelo, mas usualmente é necessário realizar uma série de transformações nas variáveis. Em muitos casos, estas transformações podem ser desconhecidas e, além disso, perde-se graus de liberdade a cada variável acrescentada no modelo, podendo causar perturbações na estimação dos parâmetros.

As redes neurais não necessitam de muitos cuidados estatísticos para serem aplicadas. Todos os conjuntos de dados usados em modelos logísticos podem ser usados na rede, com variáveis categóricas e contínuas tanto nas camadas de entrada como nas camadas de saída. Já a regressão logística, precisa atender as suposições de linearidade e de que os dados foram gerados por uma distribuição multinomial. A rede apresenta uma grande capacidade de identificar as interações possíveis entre as covariáveis, o que é feito nas camadas intermediárias. Entretanto, na regressão logística, à medida que o número de covariáveis aumenta, aumenta também o número de interações possíveis, ficando cada mais complicado testar as relações com grau elevado. Além disso, muitas vezes não se dispõe de graus de liberdade suficiente para esse tipo de ajuste. Por outro lado, interações clinicamente irrelevantes ou inaceitáveis podem, eventualmente, ser consideradas pela rede. Este aspecto pode ser controlado com mais facilidade na regressão logística.

As redes neurais podem ser desenvolvidas a partir de algoritmos diferentes aplicados ao mesmo grupo de aprendizado da rede. Nesse trabalho o método desenvolvido foi o algoritmo da retro-propagação (*back-propagation*), mas existem outros métodos, como o algoritmo genético descrito por Dybowski et. al (1996), ou ainda a utilização métodos Bayesianos, descritos por Bishop (1995).

Segundo Freeman et. al. (2000), uma excelente característica das redes neurais é incorporar registros com informações ausentes em algumas variáveis (*missing values*), pois a rede trabalha com o reconhecimento de padrões das entradas relacionadas com as saídas.

Por outro lado, muitas vezes as redes neurais são chamadas de “caixa preta”, pois tendo caráter estritamente descritivo e pequena habilidade para explicar as possíveis relações causais. Sob este aspecto, a regressão logística é superior às redes neurais na identificação de relações causais das variáveis.

Na regressão logística, o modelo desenvolvido é capaz de determinar quais variáveis são mais importantes na predição do desfecho, através da magnitude das estimativas dos coeficientes de regressão e das correspondentes medidas de razão de chance (*odds ratio*). A rede simplesmente desenvolve um modelo e deixa que com o treinamento determine qual das variáveis é a mais importante.

Não se pode identificar qual a variável teve maior contribuição para a saída e provavelmente o modelo final conterá variáveis sem importância. Uma maneira adotada pelos pesquisadores para poder interpretar suas estimativas é treinar várias vezes a rede, retirando uma variável de entrada em cada ocasião e observando o resultado final. Outra forma é examinar os pesos das ligações entre as diversas variáveis de entrada e identificar qual delas pode ser eliminada sem alterar a performance da rede. Entretanto, nenhum desses métodos se compara a facilidade de interpretação das razões de chances da regressão logística.

As redes neurais são mais difíceis de serem determinadas. Por exemplo, para outra pessoa derivar a probabilidade de ocorrência do evento em um estudo na rede neural, são necessários o programa computacional usado e o conhecimento da matriz de pesos gerada no treinamento. Já na regressão logística, com a examinação dos coeficientes estimados pela regressão logística, qualquer pessoa obtém essa probabilidade.

As redes neurais necessitam de grandes recursos computacionais. Para se otimizar o tempo de treinamento da rede, deve-se utilizar computadores e rotinas computacionais potentes. Conforme a estrutura da rede, pode-se levar dias para se obter uma otimização do erro, com a propagação para trás. Além disso, as redes neurais apresentam uma grande capacidade de se superestimar, problema este chamado de *overfitting*. Isso se deve ao fato de que a rede é construída com todas as iterações possíveis do modelo. Elas são desenvolvidas por métodos empíricos e apresentam muitas questões metodológicas para serem resolvidas. Os modelos desenvolvidos necessitam percorrer através de processos empíricos a sensibilidade da análise nos parâmetros do treinamento como razões de aprendizado, termos momentâneos e números de nós intermediários. Poucos guias existem para se determinar se a rede treinada é a melhor solução para o problema dado.

Em comparação com a regressão logística, as vantagens e desvantagens da utilização de redes neurais estão resumidas no quadro 5.1.

Quadro 5.1: Desvantagens e Vantagens da Rede Neural comparada com a Regressão logística

Vantagens Rede Neural	<i>Desvantagens</i> <i>Regressão Logística</i>	Desvantagens Rede Neural	<i>Vantagens</i> <i>Regressão Logística</i>
Descrevem relações não lineares entre as variáveis	<i>Relações não lineares são realizadas mediante transformações nos dados</i>	Considerada uma “caixa preta”, pois é o treinamento que determina as variáveis importantes.	<i>Covariáveis identificadas pelo pesquisador, através das estimativas dos coeficientes e das razões de chances</i>
Não necessita suposições estatísticas	<i>Deve atender as suposições de linearidade e distribuição multinomial dos dados</i>	Dificuldades na interpretação das saídas	<i>Razões de chance são facilmente interpretadas</i>
Interações facilmente identificadas através das camadas intermediárias	<i>Como aumento do número de termos no modelo, aumenta o número e a ordem das possíveis interações</i>	É necessário o programa computacional para se obter a probabilidade de ocorrência	<i>Através das estimativas dos coeficientes, obtém-se a probabilidade de ocorrência</i>
Incorpora na rede resultados faltantes (“missing”) nas variáveis	<i>Em geral os registros com “missing” são excluídos da análise</i>	Necessitam de grandes recursos computacionais	<i>São apresentadas em qualquer software estatístico</i>

6. Considerações finais

Os resultados dos dois procedimentos utilizados para predizer a macrossomia fetal são consistentes com aqueles apresentados na literatura. As covariáveis incluídas nos modelos apresentam fortes evidências estatísticas e epidemiológicas para permanecerem no modelo.

A altura e o peso das pessoas descrevem a obesidade central através do índice de massa corporal. No modelo ajustado, a obesidade das gestantes pode ser descrita com as medidas da altura e cintura, podendo ser considerada como a obesidade central. Esse estudo comprovou que as gestantes mais pesadas apresentam maior chance de terem filhos macrossômicos.

Além disso, quanto maior o ganho de peso durante a gestação, maiores serão as chances dela ter um filho macrossômico. Nesse estudo muitas gestantes eram obesas, e assim, pode ter ocorrido uma subestimação da influência do ganho de peso na macrossomia fetal.

O fumo, como foi devidamente comprovado na literatura, diminui o peso dos bebês, sendo assim, um fator de proteção para a macrossomia. Para predizer a macrossomia fetal, não foram observadas diferenças estatísticas significativas entre os resultados gerados pela rede neural e pelo modelo logístico. Resultados semelhantes foram encontrados nos estudos reportados por Nguyen et al. (2002), Ottenbach et. al. (2001), Ambalavanan et.al. (2001), Duh et al. (1998), Freman et. al. (2000) e Zernikow et. al. (1998), usando outros desfechos.

A área sob a curva encontrada para os modelos não foi muito expressiva em termos estatísticos. Ambos os modelos obtiveram uma área de aproximadamente 0,7, enquanto que o esperado seria acima de 0,8. Entretanto este fato está coerente com as pesquisas epidemiológicas.

Mesmo com todas as vantagens da rede neural sob o modelo logístico, ainda defendo a utilização da regressão logística para a realização de predição, pois uma das grandes desvantagens da rede neural é sua pouca generalização.

Em outras palavras, para obtermos os resultados encontrados por um outro pesquisador, são necessários o programa utilizado e o banco de dados. Já com a regressão logística podemos obter essa generalização apenas com as estimativas dos coeficientes encontrados.

7. Referências bibliográficas

AGRESTI, A.(1990). *Categorical Data Analysis*. Wiley. Series in Probability and Mathematical Statistics. Florida

ALBERT, A. & ANDERSON, J.A. (1984). *On the existence of maximum likelihood estimates in the logistic regression models*. *Biometrika*,71, 1-10.

AMBALAVANAN, N. & CARLO, W.A.(2001). *Comparison of the prediction of extremely low birth weight neonatal mortality by regression analysis and by neural networks*. *Early Human development* 65: 123-137

American Diabetes Association: Gestational diabetes mellitus. *Diabetes Care* 23(Suppl. 1): S77-S79, 2000

BAGLEY, S.C, White, H. & Golomb, B. A. (2001). *Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain*. *Journal of Clinical Epidemiology* 54: 979-985.

BAXT, W.G. (1995). *Application of artificial neural networks to clinical medicine*. *Lancet* 346: 1135-38. October 28

BISHOP, C.M. (1995). *Neural Networks for Pattern Recognition*. New York: Oxford University Press.

BRESLOW, N.E. & DAY, N.E. (1980). *Statistical Methods in Cancer Research – The Analysis of Case-Control Studies*. Lyon, IARC Scientific Publications.

BRESSEL, M.A.(2001). *Introdução à Regressão Logística Exata*. Porto Alegre, R.S. (Monografia apresentada para obtenção do grau de Bacharel em Estatística/UFRGS), p.77.

COX, D.R. & SNELL, E.J. (1989). *Analysis of Binary Data*. Second Edition. Monographs on Statistics and Applied Probability 32.

CROSS, S.S., HARRISON, R.F. & KENNEDY, R.L. (1995). *Introduction to neural networks*. *Lancet* 346: 1075-79 October 21.

- CYTEL SOFTWARE CORPORATION (2001). *Proc-LogXact 4 for SAS Users*. Cambridge, MA, USA.
- DUH, M.S., WALKER, A.M., PAGANO, M. & KRONLUND, K. (1998). *Prediction and Cross-Validation of Neural Networks Versus Logistic Regression: Using Hepatic Disorders as an Example*. American Journal of Epidemiology 147, no. 4: 407 – 13
- DYBOWSKI, R., WELLER, P., CHANG, R. & GANT, V. (1996). *Prediction of outcome in critically ill pacientes using artificial neural network synthesized by genetic algorithm*. Lancet 347: 1146-50 April 27.
- FLETCHER, R.H., FLETCHER, S.W. & WAGNER, E.H (1996). *Epidemiologia Clínica: Elementos Essenciais*. 3^a ed. Artes Médicas. p 281
- FREEMAN, R.V., EAGLE, K.A., BATES, R.E., WERNS, S.W., ROGERS, E.K.R., KARAVITE, D. & MOSCUSSI, M. (2000). *Comparison of artificial neural network with logistic regression in prediction of in-hospital death after percutaneous transluminal coronary angioplasty*. American Heart Journal 140 (3): 511-20
- HANLEY, J.A & MCNEIL, B.J. (1982). *The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve*. Radiology 143: 29-36, April 1982
- HANLEY, J.A & MCNEIL, B.J. (1983). *A Method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases*. Radiology 148: 839-843, September de 1983.
- HARREL, JR. F. E. (2001). *Regression Modeling Strategies: with applications linear models, logistic regression, and survival analysis*. Springer-Verlag, New York.
- HOSMER, D.W. & LEMESHOW, S. (1989). *Applied Logistic Regression*. Nova York: Wiley.
- HOSMER, D.W. & LEMESHOW, S. (2001). *Applied Logistic Regression*. Nova York: Wiley.

JAULENT, M.C., COLOMBERT, I. & DEGOULET, P. *Logistic regression model: conditions required for stability of prediction.* Site: <http://www.amia.org/pubs/symposia/Doo5460.pdf>

METHA, C. R. & PATEL, N. R. (1995) *Exact Logistic Regression: Theory and Examples.* *Statistics in Medicine*, 14, 2143-2160.

McCULLAGH, P. & NELDER, J.A. (1989). *Generalized linear models.* Second Edition, Chapman and Hall.

NGUYEN, T., MALLEY, R., INKELIS, S.H. & KUPPERMANN, N. (2002). *Comparison of predictions models for adverse outcome in pediatric meningococcal disease using artificial neural network and logistic regression analyses.* *Journal of Clinical Epidemiology* 55: 687-695

OHNO-MACHADO, L., (1999). *Neural Network Applications in Physical Medicine and Rehabilitation.* *American Journal of Physical Medicine & Rehabil*, 78(4): 392-398.

OTTENBACHER, K.J., SMITH, P.M., ILLIG, S.B., LINN, R.T, FIEDLER, R.C & GRANGER, C.V. (2001). *Comparison of logistic regression and neural networks to predict rehospitalization in patients with stroke.* *Journal of Clinical Epidemiology* 54: 1159-1165.

SAS INSTITUTE INC. (1999). *SAS OnlineDoc®.* Version 8, Cary, NC: SAS Institute Inc.

SAS INSTITUTE INC.(2000). *What's New in SAS ® Software for Release 8.1.* Cary NC: SAS Institute Inc.,

SELAU, L.P.R (2000). *Redes Neurais Artificiais no Contexto Estatístico.* Porto Alegre, R.S. (Monografia apresentada para obtenção do grau de Bacharel em Estatística/UFRGS).

SOARES, J.F. & SIQUEIRA, A.L. (1999). *Introdução à estatística médica.* Belo Horizonte.

STATSOFT TEXTBOOK (1998). *Eletronic Textbook.* Copyright StatSoft, Inc.

TU, J.V. (1996). *Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes*. J Clin Epidemiol Vol 49, no. 11 p 1225-1231.

WALKER, S.H. & DUNCAN, D.B. (1967). *Estimation of the probability of an event as a function of several independent variables*. Biometrika 54 1 and 2, p.167-179.

WEDDERBURN, R.W.N. (1976). *On the existence and uniqueness of the maximum likelihood estimates for certain generalized linear models*. Biometrika, 63, 27-32.

WHO Consultation: Definition, Diagnosis and Classification of Diabetes Mellitus and Its Complications: Report of a WHO Consultation. Part 1: Diagnosis and Classification of Diabetes Mellitus. Geneva, WHO/NCD/NCS/99.2, World Health Org., 1999

WHO Expert Committee on Diabetes Mellitus: *Diabetes Mellitus*. Geneva, World Health Org., 1985 (Tech. Rep. Ser., no. 727)

WHO Consultation on Obesity: *Obesity: Preventing and Managing the Global Epidemic: Report of a WHO Consultation on Obesity*. Geneva World Health Org., 1998

WOODWARD, M., (1999). *Epidemiology: Study Design and Data analysis*. London CHAPMANN & HALL.

ZERNIKOW, B., HOLMANNPOETTER, K., MICHEL, E., PIELEMEIER, W., HORNSCHU, F., WESTERMANN, A. & HENNECKE, K.H. (1998). *Artificial Neural Network for risk assessment in preterm neonates*. Arch Dis Child Fetal Neonatal 79: F129-F134 (September).

ANEXO A

SEPARAÇÃO DOS BANCOS DE DADOS

ANEXO A1 – ROTINA DO SAS PARA SEPARAR O BANCO DE DADOS PARA O AJUSTE DO MODELO LOGÍSTICO

```

OPTIONS OBS=MAX;
LIBNAME DMP 'C:\Mariana\Monograf\dados\Pred01';
options formchar='|----|+|---+|=|/\<>*';
options linesize=78 ps=59;

```

```

data TEMP;
  set DMP.PRED03;
  x=ranuni(32);
run;

```

```

proc sort data=TEMP; by x; run;

```

```

data set1 set2;
  set TEMP;
  if lgas >.z;
  if _n_ <=4450
  then do;
    particao = 1; *training;
    output set1;
  end;
  else do;
    particao = 2; *testing;
    output set2;
  end;
run;

```

```

DATA DMP.TRAINING;
SET SET1 (DROP=X);
RUN;
DATA DMP.TESTING;
SET SET2 (DROP=X);
RUN;

```

```

PROC PRINT DATA=SET1;
  var registro particao;
RUN;

```

```

PROC PRINT DATA=SET2;
  var registro particao;
RUN;

```

```

proc freq data= set1;
tables lgas;
run;

```

```

proc freq data=set2;
tables lgas;
run;

```

ANEXO A2 – ROTINA DO SAS PARA REUNIFICAÇÃO DO BANCO DE DADOS

```

OPTIONS OBS=MAX;
LIBNAME DMP 'C:\Mariana\Monograf\dados\Pred01';
options formchar='|----|+|---+|=|/<>*';
options linesize=78 ps=59;

data NEURAL;
  length REGISTRO 8 CENTROA $ 4 CENTRON 8 IDADE 8 RACA 8 CINTURA 8 QUADRIL 8
    RCQ 8 ALTURA 8 RCA 8 PESO 8 IMCA 8 IMCP 8 GANHOPF1 8 PREGAS 8 PGASUBE 8
    PGASUPI 8 PGATRIC 8 PGABIC 8 CIRCBRA 8 CIRCCOX 8 RST 8 GLJ110 8 GLJ126 8
    GL2H140 8 FUMO 8 NCIGANT 8 NCIGAG 8 HFDM 8 HFDM2 8 ESCOLA 8 ESCCAT 8
    ANTOB2 8 TEMPMANH 8 TEMP5F 8 DMANT 8 DMGANT 8 DMGATUAL 8 PARI 8 GRAVI 8
    ALCO 8 AALCO 8 AUFIN 8 DUMAR $ 8 ECO12 8 DTECO12 $ 8 AUFIND 8 IGDUMAR 8
    IGECOAR 8 IGAR1 8 IGATTG 8 IGAPT 8 DTCONC $ 8 CRITIG2 8 CODATENO 8
    PESOPANT 8 RESCVJEJ 8 RESCV1H 8 RESCV2H 8 DTRESCV $ 8 DIETHIP 8
    INSULIN 8 GEMEO 8 PESORN 8 OBITO 8 DIASVIV 8 FMHOSP 8 PEES 8 LGAS 8
    IGAPTSEM 8 IGAPTS 8 WTGAIN 8 EXCLUDED 8 GIFG 8 BMI 8 BMISQ 8 CINSQ 8
    ALTSQ 8 GLJ105 8 BMIA 8 BMIASQ 8 BLACK 8 MIXED 8 PAR1 8 PAR2 8 PAR3P 8
    PARM 8 CHECK 8 OMSCOMB 8 COUNTADA 8 ADAGDM 8 ADAIGT 8 ADAGDMN 8 MPP 8
    MPPOMS 8 MPPADA 8 DMGANT2 8 WTGSQ 8 IDASQ 8 NENM4KGK 8 HXMACRO 8 HXPOS 8
    PARTICAO 8;
  set DMP.training
    DMP.testing
    ;
  keep REGISTRO CENTROA CENTRON IDADE RACA CINTURA QUADRIL RCQ ALTURA RCA
    PESO IMCA IMCP GANHOPF1 PREGAS PGASUBE PGASUPI PGATRIC PGABIC CIRCBRA
    CIRCCOX RST GLJ110 GLJ126 GL2H140 FUMO NCIGANT NCIGAG HFDM HFDM2 ESCOLA
    ESCCAT ANTOB2 TEMPMANH TEMP5F DMANT DMGANT DMGATUAL PARI GRAVI ALCO
    AALCO AUFIN DUMAR ECO12 DTECO12 AUFIND IGDUMAR IGECOAR IGAR1 IGATTG
    IGAPT DTCONC CRITIG2 CODATENO PESOPANT RESCVJEJ RESCV1H RESCV2H DTRESCV
    DIETHIP INSULIN GEMEO PESORN OBITO DIASVIV FMHOSP PEES LGAS IGAPTSEM
    IGAPTS WTGAIN EXCLUDED GIFG BMI BMISQ CINSQ ALTSQ GLJ105 BMIA BMIASQ
    BLACK MIXED PAR1 PAR2 PAR3P PARM CHECK OMSCOMB COUNTADA ADAGDM ADAIGT
    ADAGDMN MPP MPPOMS MPPADA DMGANT2 WTGSQ IDASQ NENM4KGK HXMACRO HXPOS
    PARTICAO;
run;

proc sort data=NEURAL;
  by PARTICAO;
run;

PROC PRINT DATA=NEURAL;
  var registro particao;
RUN;

* criação da dummy do fumo;
data fumo;
set neural;
IF FUMO > .Z
  THEN DO;
  FUM01=0; FUM02=0;
  IF FUMO=1 THEN FUM01=1;
  IF FUMO=2 THEN FUM02=1;
END;
ELSE DO;
  FUM01=-.;
  FUM02=-.;
END;
run;

DATA DMP.neural;
SET fumo;
RUN;

```

ANEXO B

AJUSTE DO MODELO DE REGRESSÃO LOGÍSTICA

ANEXO B1 – ROTINA DO SAS PARA O AJUSTE DO MODELO LOGÍSTICO

```

OPTIONS OBS=MAX;
LIBNAME DMP 'c:\mariana\monograf\dados\pred01';
options formchar='|---|+|---+|=|-\<*>';
options linesize=78 ps=59;

data training;
set DMP.pred03;
cirbrasq = (circbra*circbra);
quadrseq = quadril**2;
rescjesq = rescvjej**2;
resc1hsq = rescv1h**2;
resc2hsq = rescv2h**2;
imcasq = imca**2;
imcpsq = imcp**2;

* seleção do training, particao= 2;
if particao = 2;
* modelagem univariada com e sem termos quadraticos;
PROC LOGISTIC DESCENDING;
MODEL lgas = altura; * significativa no modelo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = altura altsq; * não significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = bmi; *significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = bmi bmisq; * significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = circbra; *significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = circbra cirbrasq; * não significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = escola; * significativo p = 0,08;
run;

PROC LOGISTIC DESCENDING;
MODEL lgas = gravi; * não significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = idade; * significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = idade idasq; * não significativo;
RUN;

```

```

PROC LOGISTIC DESCENDING;
MODEL lgas = pari; * não significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = pgabic; * significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = pregas; * significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = rescvjej; * significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = rescv1h; * significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = rescv2h; * significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = wtgain; * significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = wtgain wtgsq; *não significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = quadril; *significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = quadril quadrsq; *significativo p = 0,07;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = cintura; *significativo;
RUN;

PROC LOGISTIC DESCENDING;
MODEL lgas = cintura cinsq; *significativo p = 0,09;
RUN;

*analise múltipla;
*considerando soemnte as variáveis significativas anteriormente escolha pelo septwise;
PROC LOGISTIC DESCENDING;
class fumo/param = ref ref =first;
MODEL lgas = altura bmi bmisq circbra escola idade
pgabic pregas cintura cinsq quadril quadrsq rescvjej rescjesq rescv1h resc1hsq rescv2h
resc2hsq wtgain fumo/stepwise;
RUN;

* modelo final;
PROC LOGISTIC DESCENDING;
class fumo/param = ref ref =first;
MODEL lgas = altura cintura rescv1h wtgain fumo;
RUN;

```

```

* inclusão de mais alguma variável considerada clinicamente significativa a partir
do modelo final gerado;
* verificação se as gravides anteriores influenciam nas análises;
proc logistic descending;
model lgas = gravi; * não significativo p = 0,12;
run;

* verificação se o peso influencia na análise;
proc logistic descending;
model lgas = peso; * significativo;
run;
proc logistic descending;
class fumo/param = ref ref =first;
model lgas = altura cintura rescv1h wtgain peso fumo;
*peso nao significativo no modelo multiplo;
run;

*verificacao do historico familiar;
proc logistic descending;
class hfdm2/param = ref ref = first;
model lgas = hfdm2; * significativo;
run;
proc logistic descending;
class fumo hfdm2/param = ref ref =first;
model lgas = altura cintura rescv1h wtgain fumo hfdm2;
*hfdm2 nao significativo no modelo multiplo;
run;
* utilizacao do imc ao inves do bmi;
proc logistic descending;
model lgas = imca; * significativo;
run;
proc logistic descending;
model lgas = imcp; *significativo;
run;
proc logistic descending;
model lgas = imca imcasq; * significativo;
run;
proc logistic descending;
model lgas = imcp imcpsq; *significativo;
run;

proc logistic descending;
class fumo/param = ref ref =first;
model lgas = altura cintura imca imcasq imcp imcpsq rescv1h wtgain fumo /stepwise sle=0.05 sls=0.2;
* entraram no modelo variaveis com significancia maior que 5% e permaneceram com p <=0,20;
run;

* modelo final;
proc logistic descending;
class fumo/param = ref ref =first;
model lgas = altura cintura wtgain fumo;
run;

* acrescentando hxmacro;
proc logistic descending;
class fumo hxmacro/param = ref ref =first;
model lgas = altura cintura wtgain fumo hxmacro;
run;
* hxmacro deve ficar no modelo;

* teste do ajuste do modelo;
proc logistic descending;
class fumo hxmacro/param = ref ref =first;
model lgas = altura cintura wtgain fumo hxmacro/scale = deviance aggregate;
run;

```

ANEXO B2 – RESULTADOS DO MODELO FINAL

The LOGISTIC Procedure Model Information

Data Set	WORK.REFINAMENTO
Response Variable	LGAS
Number of Response Levels	2
Number of Observations	2964
Link Function	Logit
Optimization Technique	Fisher's scoring

Response Profile

Ordered Value	LGAS	Total Frequency
1	1	292
2	0	2672

NOTE: 1486 observations were deleted due to missing values for the response or explanatory variables.

Class Level Information

Class	Value	Design Variables	
		1	2
FUMO	0	0	0
	1	1	0
	2	0	1
HXMACRO	0	0	
	1	1	

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Deviance and Pearson Goodness-of-Fit Statistics

Criterion	DF	Value	Value/DF	Pr > ChiSq
Deviance	2954	1779.8701	0.6025	1.0000
Pearson	2954	2896.7843	0.9806	0.7705

Number of unique profiles: 2961

NOTE: The covariance matrix has been multiplied by the heterogeneity factor (Deviance / DF) 0.602529.

Model Fit Statistics

Criterion	Intercept and Covariates	
	Intercept Only	Intercept and Covariates
AIC	1909.683	1793.870
SC	1915.678	1835.830
-2 Log L	1907.683	1779.870

The LOGISTIC Procedure Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	127.8133	6	<.0001
Score	142.9560	6	<.0001
Wald	124.4321	6	<.0001

Type III Analysis of Effects

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
ALTURA	1	3.2777	0.0702
CINTURA	1	33.7247	<.0001
WTGAIN	1	12.9588	0.0003
FUMO	2	17.3616	0.0002
HXMACRO	1	31.9534	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard	Chi-Square	Pr > ChiSq
			Error		
Intercept	1	-8.9764	1.5445	33.7780	<.0001
ALTURA	1	0.0177	0.00978	3.2777	0.0702
CINTURA	1	0.0441	0.00760	33.7247	<.0001
WTGAIN	1	0.0436	0.0121	12.9588	0.0003
FUMO	1	-0.0570	0.1494	0.1457	0.7027
FUMO	2	-0.8889	0.2148	17.1300	<.0001
HXMACRO	1	1.0066	0.1781	31.9534	<.0001

Odds Ratio Estimates

Effect		Point	95% Wald	
		Estimate	Confidence	Limits
ALTURA		1.018	0.999	1.038
CINTURA		1.045	1.030	1.061
WTGAIN		1.045	1.020	1.070
FUMO	1 vs 0	0.945	0.705	1.266
FUMO	2 vs 0	0.411	0.270	0.626
HXMACRO	1 vs 0	2.736	1.930	3.879

The LOGISTIC Procedure

Association of Predicted Probabilities and Observed Responses

Percent Concordant	68.6	Somers' D	0.383
Percent Discordant	30.3	Gamma	0.387
Percent Tied	1.1	Tau-a	0.068
Pairs	780224	c	0.692

ANEXO B3 – ROTINA DO SAS PARA ESTIMAÇÃO DA SENSIBILIDADE E DA ESPECIFICIDADE

```

OPTIONS OBS=MAX;
LIBNAME DMP 'c:\mariana\monograf\dados\pred01';
options formchar='|----|+|---+|=|/\<*';
options linesize=78 ps=59;

data training;
*banco criado pelo dataset;
set DMP.TRAINING;
proc logistic descending outest=setlest;
class fumo hxmacro/param=ref ref=first;
model lgas = altura cintura wtgain fumo hxmacro;
output out=set1p predicted=LGASP;
run;

proc print data=setlest;
run;

* para poder testar construir as dummy das variaveis qualitativas;

data testing;
set DMP.TESTING;
IF FUMO > .Z
  THEN DO;
      FUMO1=0; FUMO2=0;
      IF FUMO=1 THEN FUMO1=1;
      IF FUMO=2 THEN FUMO2=1;
  END;
ELSE DO;
      FUMO1=.;
      FUMO2=.;
  END;
IF LGAS=1 THEN LGAS1="+";
ELSE LGAS1="-";
  if _n_=1 then set setlest
  (rename=(altura=ALTC cintura=CINTC wtgain=WTC fumo1=FUMOC1 fumo2=fumoc2
  hxmacro1=HXMC));
  xc=(INTERCEPT + altura*ALTC + cintura*CINTC + wtgain*WTC + fumo1*FUMOC1 + fumo2*fumoc2+
  + hxmacro*HXMC);
  pbia=1/(1+exp(-xc));
run;

proc print data=testing (obs =10);
var registro xc INTERCEPT altura ALTC cintura CINTC wtgain WTC
  fumo1 FUMOC1 fumo2 fumoc2 hxmacro HXMC pbia;
run;

proc means data=testing;
var xc pbia;
run;

proc FREQ;
TABLES pbia;
run;

proc rank groups=10 data=testing out=testia;
ranks pbiark;
var pbia;
run;

```

```
proc freq data=testia;
  tables pbiark*LGAS1;
run;
DATA SENS1;
SET TESTIA;

IF PBIARK>8 THEN POS10P="+";
ELSE POS10P="-";
IF PBIARK>7 THEN POS20P="+";
ELSE POS20P="-";
IF PBIARK>6 THEN POS30P="+";
ELSE POS30P="-";
IF PBIARK>5 THEN POS40P="+";
ELSE POS40P="-";
IF PBIARK>4 THEN POS50P="+";
ELSE POS50P="-";
IF PBIARK>3 THEN POS60P="+";
ELSE POS60P="-";
IF PBIARK>2 THEN POS70P="+";
ELSE POS70P="-";
IF PBIARK>1 THEN POS80P="+";
ELSE POS80P="-";
IF PBIARK>0 THEN POS90P="+";
ELSE POS90P="-";
IF PBIARK>=0 THEN POS100P="+";
ELSE POS100P="-";

*IF LGAS=1 THEN LGAS1="+";
*ELSE LGAS1="-";
RUN;
PROC FREQ;
TABLES (POS10P POS20P POS30P POS40P POS50P POS60P POS70P POS80P POS90P POS100P)*LGAS1;
RUN;

PROC SORT;
BY PBIARK;
RUN;

PROC MEANS;
VAR XC PBIARK;
BY PBIARK;

RUN;

PROC FREQ;
TABLES PBIARK;

RUN;
```

ANEXO B4 – RESULTADOS DA ESTIMAÇÃO DA SENSIBILIDADE E DA ESPECIFICIDADE

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	747	-2.4242303	0.6647272	-4.3805240	0.1767084
pbia	747	0.0959387	0.0682571	0.0123640	0.5440625

The FREQ Procedure

pbia	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0.0123640146	1	0.13	1	0.13
0.0149177122	1	0.13	2	0.27
0.0189802576	1	0.13	3	0.40
0.0200037948	1	0.13	4	0.54
0.0201543481	1	0.13	5	0.67
0.0233118041	1	0.13	6	0.80
0.0233516783	1	0.13	7	0.94
0.0237868709	1	0.13	8	1.07
0.0244301557	1	0.13	9	1.20
0.0246703749	1	0.13	10	1.34
0.0248210253	1	0.13	11	1.47
0.0249744729	1	0.13	12	1.61
0.0251912528	1	0.13	13	1.74
0.0253324059	1	0.13	14	1.87
0.025636895	1	0.13	15	2.01
0.0257406022	1	0.13	16	2.14
0.0257904884	1	0.13	17	2.28
0.0261015839	1	0.13	18	2.41
0.0265346104	1	0.13	19	2.54
0.0275120596	1	0.13	20	2.68
0.0277374079	1	0.13	21	2.81
0.0278086722	1	0.13	22	2.95
0.0278646287	1	0.13	23	3.08
0.02794381	1	0.13	24	3.21
0.0281385005	1	0.13	25	3.35
0.0281980115	1	0.13	26	3.48
0.0282731847	1	0.13	27	3.61
0.0287911451	1	0.13	28	3.75
0.0289959759	1	0.13	29	3.88
0.029276775	1	0.13	30	4.02
0.0295716156	1	0.13	31	4.15
0.0305194321	1	0.13	32	4.28
0.0305288672	1	0.13	33	4.42
0.0306245709	1	0.13	34	4.55
0.0308325307	1	0.13	35	4.69
0.030884615	1	0.13	36	4.82
0.0311489308	1	0.13	37	4.95
0.0312396716	1	0.13	38	5.09
0.0312895192	1	0.13	39	5.22
0.0317620811	1	0.13	40	5.35
0.0317832779	1	0.13	41	5.49
0.0320115798	1	0.13	42	5.62
0.0320649337	1	0.13	43	5.76
0.0321204948	1	0.13	44	5.89
0.032211323	1	0.13	45	6.02
.... itens removidos ...				

The FREQ Procedure

Table of pbiark by LGAS1

pbiark(Rank for Variable pbia)			
LGAS1			
Frequency			
Percent			
Row Pct			
Col Pct	+	-	Total
0	4	70	74
	0.54	9.37	9.91
	5.41	94.59	
	5.56	10.37	
1	4	71	75
	0.54	9.50	10.04
	5.33	94.67	
	5.56	10.52	
2	3	72	75
	0.40	9.64	10.04
	4.00	96.00	
	4.17	10.67	
3	5	70	75
	0.67	9.37	10.04
	6.67	93.33	
	6.94	10.37	
4	4	70	74
	0.54	9.37	9.91
	5.41	94.59	
	5.56	10.37	
5	9	66	75
	1.20	8.84	10.04
	12.00	88.00	
	12.50	9.78	
6	6	69	75
	0.80	9.24	10.04
	8.00	92.00	
	8.33	10.22	
7	8	67	75
	1.07	8.97	10.04
	10.67	89.33	
	11.11	9.93	
8	7	68	75
	0.94	9.10	10.04
	9.33	90.67	
	9.72	10.07	
9	22	52	74
	2.95	6.96	9.91
	29.73	70.27	
	30.56	7.70	
Total	72	675	747
	9.64	90.36	100.00

The FREQ Procedure

Table of POS10P by LGAS1

POS10P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	22	52		74
	1.97	4.67		6.64
	29.73	70.27		
	30.56	4.99		
-	50	990		1040
	4.49	88.87		93.36
	4.81	95.19		
	69.44	95.01		
Total	72	1042		1114
	6.46	93.54		100.00

POS20P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	29	120		149
	2.60	10.77		13.38
	19.46	80.54		
	40.28	11.52		
-	43	922		965
	3.86	82.76		86.62
	4.46	95.54		
	59.72	88.48		
Total	72	1042		1114
	6.46	93.54		100.00

POS30P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	37	187		224
	3.32	16.79		20.11
	16.52	83.48		
	51.39	17.95		
-	35	855		890
	3.14	76.75		79.89
	3.93	96.07		
	48.61	82.05		
Total	72	1042		1114
	6.46	93.54		100.00

Table of POS40P by LGAS1

POS40P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	43	256		299
	3.86	22.98		26.84
	14.38	85.62		
	59.72	24.57		
-	29	786		815
	2.60	70.56		73.16
	3.56	96.44		
	40.28	75.43		
Total	72	1042		1114
	6.46	93.54		100.00

POS50P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	52	322		374
	4.67	28.90		33.57
	13.90	86.10		
	72.22	30.90		
-	20	720		740
	1.80	64.63		66.43
	2.70	97.30		
	27.78	69.10		
Total	72	1042		1114
	6.46	93.54		100.00

POS60P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	56	392		448
	5.03	35.19		40.22
	12.50	87.50		
	77.78	37.62		
-	16	650		666
	1.44	58.35		59.78
	2.40	97.60		
	22.22	62.38		
Total	72	1042		1114
	6.46	93.54		100.00

The FREQ Procedure

Table of POS70P by LGAS1

POS70P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	61	462		523
	5.48	41.47		46.95
	11.66	88.34		
	84.72	44.34		
-	11	580		591
	0.99	52.06		53.05
	1.86	98.14		
	15.28	55.66		
Total	72	1042		1114
	6.46	93.54		100.00

POS80P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	64	534		598
	5.75	47.94		53.68
	10.70	89.30		
	88.89	51.25		
-	8	508		516
	0.72	45.60		46.32
	1.55	98.45		
	11.11	48.75		
Total	72	1042		1114
	6.46	93.54		100.00

POS90P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	68	605		673
	6.10	54.31		60.41
	10.10	89.90		
	94.44	58.06		
-	4	437		441
	0.36	39.23		39.59
	0.91	99.09		
	5.56	41.94		
Total	72	1042		1114
	6.46	93.54		100.00

Table of POS100P by LGAS1

POS100P	LGAS1		Total
Frequency	+	-	
Percent			
Row Pct			
Col Pct	+	-	Total
+	72	675	747
	6.46	60.59	67.06
	9.64	90.36	
	100.00	64.78	
-	0	367	367
	0.00	32.94	32.94
	0.00	100.00	
	0.00	35.22	
Total	72	1042	1114
	6.46	93.54	100.00

----- Rank for Variable pbia=- -----

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	0
pbia	0

----- Rank for Variable pbia=0 -----

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	74	-3.4864946	0.2144120	-4.3805240	-3.2399727
pbia	74	0.0302596	0.0054524	0.0123640	0.0376889

----- Rank for Variable pbia=1 -----

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	75	-3.0742226	0.0901020	-3.2312354	-2.9450738
pbia	75	0.0443368	0.0037854	0.0380071	0.0499699

----- Rank for Variable pbia=2 -----

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	75	-2.8656619	0.0509905	-2.9442892	-2.7905659
pbia	75	0.0539357	0.0025996	0.0500071	0.0578361

----- Rank for Variable pbia=3 -----

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	75	-2.7027083	0.0498203	-2.7859203	-2.6106514
pbia	75	0.0628769	0.0029431	0.0580898	0.0684561

----- Rank for Variable pbia=4 -----

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	74	-2.5333595	0.0371210	-2.5976949	-2.4776297
pbia	74	0.0735918	0.0025222	0.0692869	0.0774414

----- Rank for Variable pbia=5 -----

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	75	-2.3943304	0.0433838	-2.4775396	-2.3250196
pbia	75	0.0836651	0.0033081	0.0774478	0.0890719

----- Rank for Variable pbia=6 -----

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	75	-2.2506350	0.0440035	-2.3229863	-2.1756578
pbia	75	0.0953614	0.0038004	0.0892371	0.1019578

----- Rank for Variable pbia=7 -----

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	75	-2.0594345	0.0685501	-2.1720228	-1.9459074
pbia	75	0.1132823	0.0068715	0.1022911	0.1250003

----- Rank for Variable pbia=8 -----

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	75	-1.7812620	0.0979260	-1.9457907	-1.6005209
pbia	75	0.1445627	0.0121163	0.1250131	0.1679088

----- Rank for Variable pbia=9 -----

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
xc	74	-1.0920501	0.4173195	-1.5946619	0.1767084
pbia	74	0.2585085	0.0841915	0.1687290	0.5440625

ANEXO C

CONSTRUÇÃO DA REDE NEURAL ARTIFICIAL

ANEXO C1 – AJUSTE DA REDE NEURAL

A rede neural é construída dentro do procedimento Enterprise Miner do SAS. A configuração do procedimento Miner é um pouco diferente dos demais. A Figura C.1 mostra o modelo de procedimento executado.

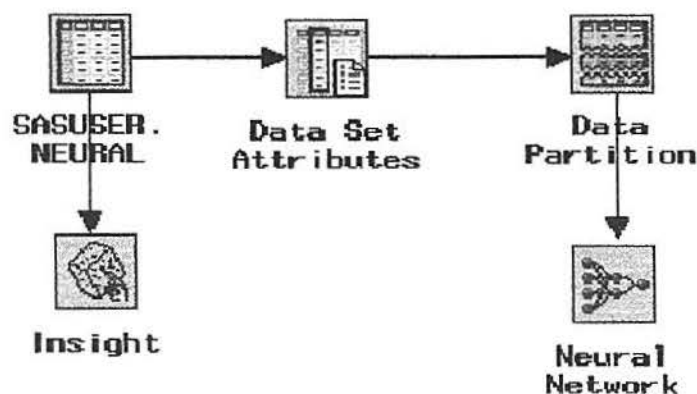


Figura C.1 - Janela do Enterprise Miner no SAS.

No menu **SASUSER** é definido a base de dados para a construção dos dados. O banco de dados utilizado é o **neural**, gerado pela rotina descrita no Anexo A2. No item **Data Set Attributes** indica-se quais variáveis devem ser consideradas na construção da rede neural, ou seja, identifica-se as variáveis de entrada e a variável alvo. Nesse item também são identificadas as variáveis indicadoras e as categorias que elas representam. Nesse trabalho, a variável fumo1 (representando as pacientes Ex-Fumantes) foi indicada como sendo ascendente, ou seja a rede irá considerar o valor 1 para estimativas dos pesos e não o valor 0, como o padrão do programa. O mesmo acontece com a variável, fumo 2(fumantes), e hxmacro (histórico familiar). No item **Data Partition**, é indicada a divisão do banco de dados, através da variável auxiliar chamada de PARTICAO.

Assim, os registros para os quais esta variável assume o valor igual a 1 correspondem à base do aprendizado (*training*), enquanto que aqueles com PARTICAO=2 formam a base de teste (*testing*). Finalmente, no item **Neural Network**, a rede é construída. Para isso, definiu-se a estrutura da rede, através de um modelo Perception multi-camada. Escolheu-se a opção de interface avançada do módulo de redes neurais para que fosse possível a criação da rede Perception com duas camadas escondidas. A seguir, apresenta-se todas as janelas de escolha da rede (Figura C.2).

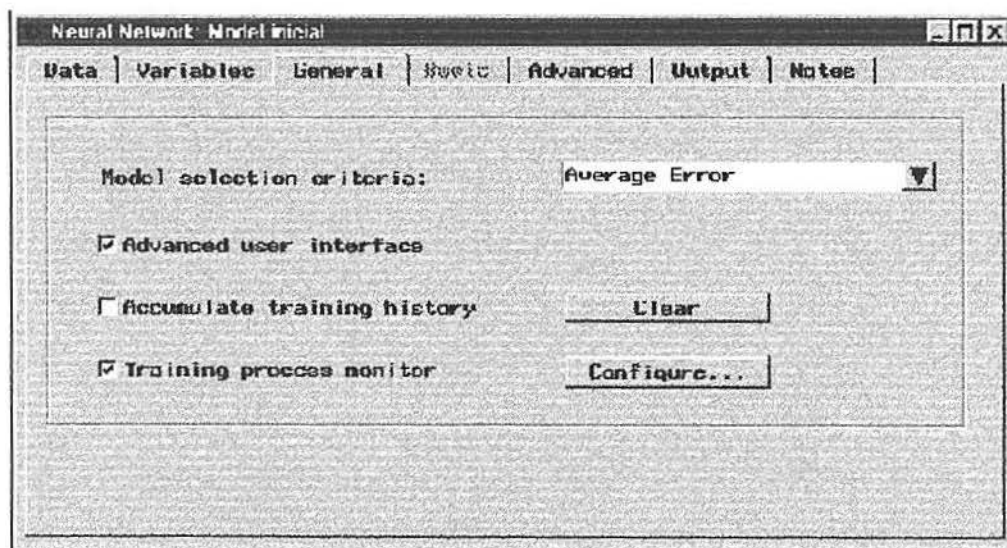


Figura C.2: Seleção do modelo de Redes Neurais

Logo após, começa-se a escolher o perfil da rede como o número de camadas ocultas bem como o número de nós escondidos. Na Figura C.3 pode-se observar a estrutura da rede que será criada. Os nós de entrada estão divididos pelo tipo de variável classificatória. O nó de saída é o resultado predito para a variável alvo, chamada de *Igas*.

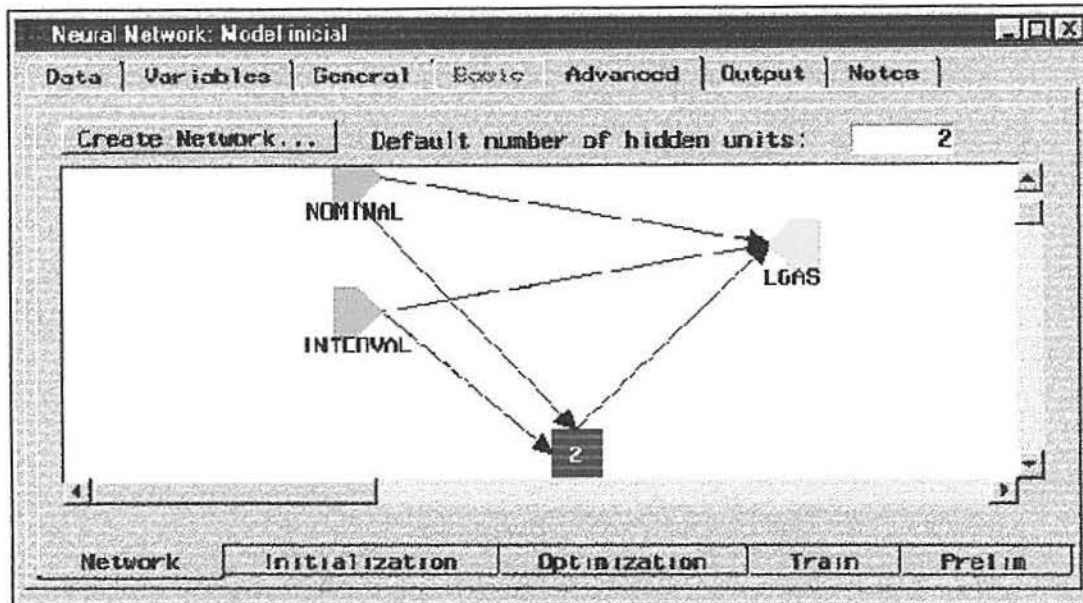


Figura C.3 –Arquitetura da rede com duas camadas escondidas composta por 2 nós.

Na opção **Create Network...**, tem-se a arquitetura da rede, onde se escolhe qual o modelo de rede neural a ser utilizado. Neste trabalho utilizou-se a rede perceptron multicamada, especificada na Figura C.4. O modelo gerado foi salvo e rodado para se obter os resultados das análises.

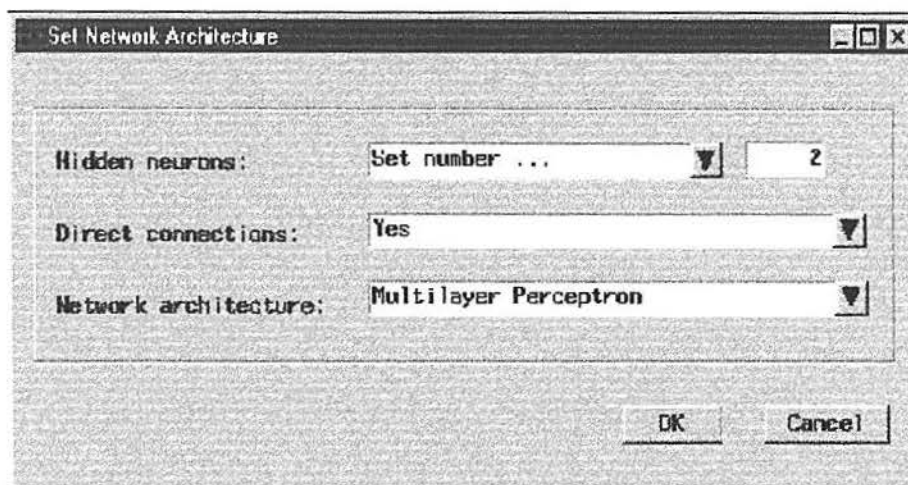


Figura C.4 – Escolha da arquitetura da rede neural.

ANEXO C2 – RESULTADOS DA REDE NEURAL

Pesos Sinápticos adotados pela rede.

From	To	Weight
FUMO11	H11	-1.403966897
FUMO21	H11	5.0705016263
HXMACRO1	H11	4.1792398338
FUMO11	H12	0.4644583857
FUMO21	H12	1.0358513663
HXMACRO1	H12	0.462318421
ALTURA	H11	-2.693245382
CINTURA	H11	0.9889143661
WTGAIN	H11	0.4249775052
ALTURA	H12	3.2442346855
CINTURA	H12	-3.448604088
WTGAIN	H12	-0.867435882
BIAS	H11	-2.659901013
BIAS	H12	9.6963747415
H11	LGAS1	0.783879658
H12	LGAS1	0.9901933179
FUMO11	LGAS1	-0.019138181
FUMO21	LGAS1	-0.963499944
HXMACRO1	LGAS1	0.3106471088
ALTURA	LGAS1	0.0950373008
CINTURA	LGAS1	0.4985971649
WTGAIN	LGAS1	0.2331252416
BIAS	LGAS1	-3.13088076

Rede Otimizada

The NEURAL Procedure

Optimization Start
Parameter Estimates

N	Parameter	Estimate	Gradient Objective Function
1	FUMO11_H11	-0.017547	0
2	FUMO21_H11	-0.040820	0
3	HXMACRO1_H11	-0.099406	0
4	FUMO11_H12	-0.090738	0
5	FUMO21_H12	0.030020	0
6	HXMACRO1_H12	0.203865	0
7	ALTURA_H11	-0.621022	0
8	CINTURA_H11	0.323252	0
9	WTGAIN_H11	0.233602	0
10	ALTURA_H12	0.071733	0
11	CINTURA_H12	-0.589350	0
12	WTGAIN_H12	0.183252	0
13	BIAS_H11	-0.644188	0
14	BIAS_H12	1.699142	0
15	H11_LGAS1	0	-0.006113
16	H12_LGAS1	0	0.004178
17	FUMO11_LGAS1	0	-0.006611
18	FUMO21_LGAS1	0	0.014610
19	HXMACRO1_LGAS1	0	-0.021347
20	ALTURA_LGAS1	0	-0.019300
21	CINTURA_LGAS1	0	-0.044581
22	WTGAIN_LGAS1	0	-0.032617
23	BIAS_LGAS1	-2.213829	-4.71114E-17

Value of Objective Function = 0.3218089396

The NEURAL Procedure

Levenberg-Marquardt Optimization

Minimum Iterations	-1
Maximum Iterations	100
Maximum Function Calls	2147483647
Maximum CPU Time	604800
ABSGCONV Gradient Criterion	0.00001
GCONV Gradient Criterion	1E-8
GCONV2 Gradient Criterion	0
ABSFCONV Function Criterion	0
FCONV Function Criterion	0.0001
FCONV2 Function Criterion	0
FSIZE Parameter	0
ABSXCONV Parameter Change Criterion	0
XCONV Parameter Change Criterion	0
XSIZE Parameter	0
ABSCONV Function Criterion	0.0024212536
Trust Region Initial Radius Factor	1
Singularity Tolerance (SINGULAR)	1E-8

Levenberg-Marquardt Optimization

Scaling Update of More (1978)

Parameter Estimates 23

Optimization Start

Active Constraints 0 Objective Function 0.3218089396
 Max Abs Gradient Element 0.0445810034 Radius 1

Iter	Restarts	Function Calls	Active Constraints	Objective Function	Objective Function Change	Max Abs Gradient Element	Lambda	Ratio Between Actual and Predicted
								Change
1*	0	2	0	0.30611	0.0157	0.0248	444E-16	0.498
2	0	3	0	0.30498	0.00113	0.0203	0.00008	0.116
3	0	4	0	0.30094	0.00404	0.0135	0.00052	0.410
4	0	5	0	0.29906	0.00188	0.00294	0.00018	0.512
5	0	7	0	0.29862	0.000433	0.00159	0.00914	0.735
6	0	8	0	0.29845	0.000177	0.000661	0.00567	0.690
7	0	9	0	0.29835	0.000095	0.000375	0.00230	0.850
8	0	10	0	0.29833	0.000022	0.000556	0.00029	0.247
9	0	12	0	0.29829	0.000038	0.000470	0.00403	0.369
10	0	13	0	0.29829	4.163E-6	0.000677	0.00262	0.0577
11	0	14	0	0.29825	0.000037	0.000226	0.0108	0.633
12	0	15	0	0.29823	0.000022	0.000166	0.00354	1.134
13	0	16	0	0.29818	0.000049	0.000228	0.00172	1.242
14	0	17	0	0.29808	0.000105	0.000458	0.00070	1.312
15	0	18	0	0.29786	0.000214	0.001000	0.00011	1.179
16	0	19	0	0.29741	0.000452	0.00199	0.00003	1.243
17	0	20	0	0.29662	0.000794	0.00171	0.00003	0.961
18	0	22	0	0.29626	0.000359	0.00206	0.00003	0.310
19	0	25	0	0.29586	0.000398	0.00113	0.00129	0.554
20	0	26	0	0.29572	0.000142	0.00107	0.00069	0.303
21	0	28	0	0.29551	0.000205	0.00120	0.00400	0.428
22	0	30	0	0.29545	0.000052	0.000779	0.00770	0.379
23	0	31	0	0.29544	0.000016	0.00103	0.00453	0.189
24	0	32	0	0.29543	9.925E-6	0.00110	0.00502	0.0987
25	0	33	0	0.29540	0.000039	0.000207	0.0166	0.553
26	0	34	0	0.29538	0.000015	0.000197	0.00472	0.912
27	0	35	0	0.29536	0.000023	0.000330	0.00199	0.864
28	0	36	0	0.29533	0.000028	0.000741	0.00077	0.528
29	0	39	0	0.29533	1.043E-6	0.000930	0.00496	0.0165
30	0	40	0	0.29530	0.000025	0.000172	0.0167	0.585
31	0	41	0	0.29530	8.12E-6	0.000112	0.00417	0.975
32	0	42	0	0.29528	0.000016	0.000177	0.00180	1.009
33	0	43	0	0.29525	0.000029	0.000329	0.00089	0.929
34	0	44	0	0.29520	0.000054	0.000913	0.00029	0.596
35	0	47	0	0.29517	0.000027	0.000280	0.0130	0.461
36	0	48	0	0.29516	7.461E-6	0.000344	0.00295	0.493
37	0	49	0	0.29516	2.168E-6	0.000568	0.00388	0.104
38	0	50	0	0.29516	1.982E-6	0.000580	0.00498	0.0656
39	0	51	0	0.29515	0.000012	0.000104	0.0167	0.566
40	0	52	0	0.29514	4.828E-6	0.000142	0.00462	0.961
41	0	53	0	0.29513	8.828E-6	0.000137	0.00230	0.951
42	0	54	0	0.29512	0.000016	0.000218	0.00114	0.868
43	0	55	0	0.29508	0.000033	0.000771	0.00037	0.570
44	0	58	0	0.29507	0.000017	0.000202	0.0123	0.491
45	0	59	0	0.29506	5.456E-6	0.000280	0.00341	0.628
46	0	60	0	0.29506	3.42E-6	0.000260	0.00399	0.343
47	0	61	0	0.29505	2.585E-6	0.000413	0.00449	0.201
48	0	62	0	0.29505	3.042E-6	0.000339	0.00546	0.193
49	0	63	0	0.29505	2.89E-6	0.000433	0.00542	0.177
50	0	64	0	0.29504	3.235E-6	0.000345	0.00571	0.193
51	0	65	0	0.29504	2.911E-6	0.000433	0.00549	0.175
52	0	66	0	0.29504	3.185E-6	0.000347	0.00569	0.190
53	0	67	0	0.29504	2.902E-6	0.000431	0.00551	0.174

The NEURAL Procedure
Optimization Results

Iterations	53	Function Calls	68
Jacobian Calls	54	Active Constraints	0
Objective Function	0.2950358875	Max Abs Gradient Element	0.0004310521
Lambda	0.0055083317	Actual Over Pred Change	0.1742448236
Radius	0.0444819105		

FCONV convergence criterion satisfied.

The NEURAL Procedure

Optimization Results
Parameter Estimates

N	Parameter	Estimate	Gradient Objective Function
1	FUMO11_H11	-1.403967	0.000017875
2	FUMO21_H11	5.070502	-0.000032626
3	HXMACRO1_H11	4.179240	-0.000016698
4	FUMO11_H12	0.464458	-0.000118
5	FUMO21_H12	1.035851	-0.000106
6	HXMACRO1_H12	0.462318	-0.000092012
7	ALTURA_H11	-2.693245	0.000021594
8	CINTURA_H11	0.988914	-0.000010867
9	WTGAIN_H11	0.424978	-0.000006680
10	ALTURA_H12	3.244235	-0.000049471
11	CINTURA_H12	-3.448604	0.000431
12	WTGAIN_H12	-0.867436	0.000120
13	BIAS_H11	-2.659901	0.000012710
14	BIAS_H12	9.696375	0.000070805
15	H11_LGAS1	0.783880	0.000000958
16	H12_LGAS1	0.990193	-0.000047639
17	FUMO11_LGAS1	-0.019138	-0.000011198
18	FUMO21_LGAS1	-0.963500	-0.000010544
19	HXMACRO1_LGAS1	0.310647	-0.000011751
20	ALTURA_LGAS1	0.095037	-0.000023610
21	CINTURA_LGAS1	0.498597	0.000061615
22	WTGAIN_LGAS1	0.233125	0.000015518
23	BIAS_LGAS1	-3.130881	0.000021259

Value of Objective Function = 0.2950358875

Ajuste do modelo da rede neural

Fit Statistic	Training	Test
[TARGET=LGAS]		
Average Profit	0.0985155196	0.0963855422
Misclassification Rate	0.0975033738	0.0977242303
Average Error	0.2950358875	0.3063764613
Average Squared Error	0.0833248276	0.0838823029
Sum of Squared Errors	493.94957828	125.32016053
Root Average Squared Error	0.2886604019	0.289624417
Root Final Prediction Error	0.2909091029	
Root Mean Squared Error	0.2897869336	0.289624417
Error Function	1748.9727411	457.72643316
Mean Squared Error	0.0839764669	0.0838823029
Maximum Absolute Error	0.9867643156	0.9912111896
Final Prediction Error	0.0846281061	
Divisor for ASE	5928	1494
Model Degrees of Freedom	23	
Degrees of Freedom for Error	2941	
Total Degrees of Freedom	2964	
Frequency of Classified Cases	2964	747
Sum of Frequencies	2964	747
Sum Case Weights * Frequencies	5928	1494
Akaike's Information Criterion	1794.9727411	
Schwarz's Bayesian Criterion	1932.8415258	

ANEXO C3 – ESTIMAÇÃO DA SENSIBILIDADE E ESPECIFICIDADE

```
LIBNAME DMP 'c:\mariana\monograf\dados';
```

```
data redetest;
set dmp.redetest;
IF LGAS=1 THEN LGAS1="+";
ELSE LGAS1="-";
proc freq;
tables pred1;
run;
```

```
proc rank groups=10 out=testern;
ranks pred1rk;
var pred1;
run;
```

```
proc freq data=testern;
tables pred1rk*LGAS1;
run;
```

```
DATA SENS1;
set testern;
IF pred1rk>8 THEN POS10P="+";
ELSE POS10P="-";
IF pred1rk>7 THEN POS20P="+";
ELSE POS20P="-";
IF pred1rk>6 THEN POS30P="+";
ELSE POS30P="-";
IF pred1rk>5 THEN POS40P="+";
ELSE POS40P="-";
IF pred1rk>4 THEN POS50P="+";
ELSE POS50P="-";
IF pred1rk>3 THEN POS60P="+";
ELSE POS60P="-";
IF pred1rk>2 THEN POS70P="+";
ELSE POS70P="-";
IF pred1rk>1 THEN POS80P="+";
ELSE POS80P="-";
IF pred1rk>0 THEN POS90P="+";
ELSE POS90P="-";
IF pred1rk>=0 THEN POS100P="+";
ELSE POS100P="-";
```

```
*IF LGAS=1 THEN LGAS1="+";
*ELSE LGAS1="-";
```

```
RUN;
```

```
PROC FREQ;
```

```
TABLES (POS10P POS20P POS30P POS40P POS50P POS60P POS70P POS80P POS90P POS100P)*LGAS1;
```

```
RUN;
```

ANEXO C4 – RESULTADOS DA ESTIMAÇÃO DA SENSIBILIDADE E ESPECIFICIDADE

The FREQ Procedure

Table of pred1rk by LGAS1

pred1rk(Rank for Variable pred1)			
LGAS1			
Frequency			
Percent			
Row Pct			
Col Pct	+	-	Total
0	7	104	111
	0.63	9.34	9.96
	6.31	93.69	
	9.72	9.98	
1	4	107	111
	0.36	9.61	9.96
	3.60	96.40	
	5.56	10.27	
2	10	102	112
	0.90	9.16	10.05
	8.93	91.07	
	13.89	9.79	
3	6	105	111
	0.54	9.43	9.96
	5.41	94.59	
	8.33	10.08	
4	6	65	71
	0.54	5.83	6.37
	8.45	91.55	
	8.33	6.24	
6	0	367	367
	0.00	32.94	32.94
	0.00	100.00	
	0.00	35.22	
7	2	6	8
	0.18	0.54	0.72
	25.00	75.00	
	2.78	0.58	
8	9	103	112
	0.81	9.25	10.05
	8.04	91.96	
	12.50	9.88	
9	28	83	111
	2.51	7.45	9.96
	25.23	74.77	
	38.89	7.97	
Total	72	1042	1114
	6.46	93.54	100.00

The FREQ Procedure

Table of POS10P by LGAS1

POS10P		LGAS1		
Frequency	Percent	Row Pct	Col Pct	Total
		+	-	
+	28	83	111	
	2.51	7.45	9.96	
	25.23	74.77		
	38.89	7.97		
-	44	959	1003	
	3.95	86.09	90.04	
	4.39	95.61		
	61.11	92.03		
Total	72	1042	1114	
	6.46	93.54	100.00	

POS20P		LGAS1		
Frequency	Percent	Row Pct	Col Pct	Total
		+	-	
+	37	186	223	
	3.32	16.70	20.02	
	16.59	83.41		
	51.39	17.85		
-	35	856	891	
	3.14	76.84	79.98	
	3.93	96.07		
	48.61	82.15		
Total	72	1042	1114	
	6.46	93.54	100.00	

POS30P		LGAS1		
Frequency	Percent	Row Pct	Col Pct	Total
		+	-	
+	39	192	231	
	3.50	17.24	20.74	
	16.88	83.12		
	54.17	18.43		
-	33	850	883	
	2.96	76.30	79.26	
	3.74	96.26		
	45.83	81.57		
Total	72	1042	1114	
	6.46	93.54	100.00	

Table of POS40P by LGAS1

POS40P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	39	559		598
	3.50	50.18		53.68
	6.52	93.48		
	54.17	53.65		
-	33	483		516
	2.96	43.36		46.32
	6.40	93.60		
	45.83	46.35		
Total	72	1042		1114
	6.46	93.54		100.00

POS50P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	39	559		598
	3.50	50.18		53.68
	6.52	93.48		
	54.17	53.65		
-	33	483		516
	2.96	43.36		46.32
	6.40	93.60		
	45.83	46.35		
Total	72	1042		1114
	6.46	93.54		100.00

POS60P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	45	624		669
	4.04	56.01		60.05
	6.73	93.27		
	62.50	59.88		
-	27	418		445
	2.42	37.52		39.95
	6.07	93.93		
	37.50	40.12		
Total	72	1042		1114
	6.46	93.54		100.00

The FREQ Procedure

Table of POS70P by LGAS1

POS70P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	51	729		780
	4.58	65.44		70.02
	6.54	93.46		
	70.83	69.96		
-	21	313		334
	1.89	28.10		29.98
	6.29	93.71		
	29.17	30.04		
Total	72	1042		1114
	6.46	93.54		100.00

POS80P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	61	831		892
	5.48	74.60		80.07
	6.84	93.16		
	84.72	79.75		
-	11	211		222
	0.99	18.94		19.93
	4.95	95.05		
	15.28	20.25		
Total	72	1042		1114
	6.46	93.54		100.00

POS90P		LGAS1		
Frequency				
Percent				
Row Pct				
Col Pct	+	-		Total
+	65	938		1003
	5.83	84.20		90.04
	6.48	93.52		
	90.28	90.02		
-	7	104		111
	0.63	9.34		9.96
	6.31	93.69		
	9.72	9.98		
Total	72	1042		1114
	6.46	93.54		100.00

Table of POS100P by LGAS1

POS100P	LGAS1		Total
Frequency			
Percent			
Row Pct			
Col Pct	+	-	
	72	1042	1114
	6.46	93.54	100.00
	6.46	93.54	
	100.00	100.00	
Total	72	1042	1114
	6.46	93.54	100.00

ANEXO D

CONSTRUÇÃO DA CURVA ROC NO SPSS

ANEXO D - CURVA ROC NO SPSS

ROC

```
pbia BY lgas (1)
/PLOT = CURVE(REFERENCE)
/PRINT = SE
/CRITERIA = CUTOFF(INCLUDE) TESTPOS(LARGE) DISTRIBUTION(FREE) CI(95)
/MISSING = EXCLUDE .
```

pbia = preditos da regressão logística

