CLEIDE APARECIDA MOREIRA SILVA

EXPLORAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS PELA TÉCNICA DE REGRESSÃO LOGÍSTICA PARA ANÁLISE DE DADOS EPIDEMIOLÓGICOS

CAMPINAS

2006

CLEIDE APARECIDA MOREIRA SILVA

EXPLORAÇÃO DE MÉTODOS DE SELEÇÃO DE VARIÁVEIS PELA TÉCNICA DE REGRESSÃO LOGÍSTICA PARA ANÁLISE DE DADOS EPIDEMIOLÓGICOS

Dissertação de Mestrado apresentada à Pós-Graduação da Faculdade de Ciências Médicas da Universidade Estadual de Campinas para obtenção do título de Mestre em Saúde Coletiva

ORIENTADOR: *PROF. DR. DJALMA DE CARVALHO MOREIRA FILHO*

CAMPINAS

2006

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA FACULDADE DE CIÊNCIAS MÉDICAS DA UNICAMP
Bibliotecário: Sandra Lúcia Pereira – CRB-8ª / 6044

Si38e

Silva, Cleide Aparecida Moreira
Exploração de métodos de seleção de variáveis pela técnica de regressão logística para análise de dados epidemiológicos / Cleide Aparecida Moreira Silva. Campinas, SP : [s.n.], 2006.

Orientador : Djalma de Carvalho Moreira Filho
Dissertação (Mestrado) Universidade Estadual de Campinas. Faculdade de Ciências Médicas.

1. Modelos Estatísticos. 2. Mortalidade Infantil. 3. Modelos Logísticos. I. Moreira Filho, Djalma de Carvalho. II. Universidade Estadual de Campinas. Faculdade de Ciências Médicas. IV. Título.

Título em ingles : Exploration of variable selection methods by logistic regression techniques for epidemiologic data analysis

Keywords:
- Statistics models
- Infant mortality
- Logistic models

Área de concentração : Saúde coletiva

Titulação: Mestrado

Banca examinadora: Prof Dr Djalma de Carvalho Moreira Filho
Profa. Dra. Nívea da Silva Matuda
Prof Dr José Norberto Walter Dachs

Data da defesa: 23/02/2006

Banca Examinadora da Dissertação de Mestrado

Orientador: Prof. Dr. Djalma de Carvalho Moreira Filho

Membros:

1 – Prof.Dra.Nivea da Silva Matuda

2 – Prof.Dr.José Norberto Walter Dachs

Curso de Pós-Graduação em Saúde Coletiva da Faculdade de Ciências Médicas da Universidade Estadual de Campinas

Data: 23 / 02 / 2006

DEDICATÓRIA

Dedico este trabalho aos meus pais, Moacyr e Mercedes, início de tudo. Pelo exemplo de luta, honestidade e pela segurança.

Dedico também ao meu marido Tony, pelo amor, companheirismo e incentivo.

Ao Prof.Dr. José Guilherme Cecatti, pela oportunidade e reconhecimento.

Aos meus amigos Andréa Ferreira Semolini, Helymar da Costa Machado e Eduardo Luiz Hoehne, pela paciência, pelas dicas e ajuda prestada.

Ao Prof.Dr. Djalma de Carvalho Moreira Filho, por me aceitar como orientanda e permitir que realizasse mais esta etapa de minha carreira.

Ao Prof.Dr. José Butori Lopes de Faria pela compreensão e apoio.

Em especial à Dra.Solange Duarte de Mattos Almeida por me fornecer o banco de dados para realização deste trabalho.

À Monize Cocetti, pela amizade e incentivo.

À Prof.Dra.Sílvia Maria Santiago pelo apoio durante as aulas.

À Maria Aparecida Moreira Mendes pela amizade e ajuda.

À Prof.Dra. Marilisa Berti de Azevedo Barros, por despertar meu interesse pelo tema do trabalho e pela colaboração durante o curso.

À Leoci H.T.Santos pela serenidade e profissionalismo.

Enfim, a todos que direta ou indiretamente contribuíram para a realização deste trabalho.

SUMÁRIO

PÁG.

IC95%	Intervalo de 95% de confiança
OR	odds ratio (razão de chances)
SIM	Sistema de Informações sobre Mortalidade
SINASC	Sistema de Informações de Nascidos Vivos

PÁG.

PÁG.

PÁG.

RESUMO

Neste trabalho foi discutida a aplicação de dois métodos distintos de seleção de variáveis e modelos na análise de regressão logística múltipla: modelo hierarquizado e modelo selecionado pelo critério *stepwise*. Em um estudo caso-controle não-pareado realizado para identificar fatores de risco para o óbito neonatal em Campinas-SP foram analisadas variáveis sócio-econômicas, de morbidade materna e relacionadas à atenção à saúde. Foram selecionados 117 casos e 234 controles e as informações adicionais obtidas por meio de entrevista domiciliar. Pela análise de regressão logística múltipla com modelo hierarquizado foram identificados como fatores de risco para o óbito neonatal a renda familiar, a naturalidade da mãe, o número de moradores do domicílio, presença de sangramento vaginal, parto antecipado por problema de saúde, o número de orientações recebidas durante o pré-natal, a escolha do hospital para o parto, o tempo entre a internação e o parto, a idade gestacional, baixo peso ao nascer e Apgar do quinto minuto. As diferenças encontradas no modelo selecionado pelo critério *stepwise* foram: renda familiar que se mostrou associada à escolha do hospital, internação por problemas de saúde associada ao sangramento vaginal e naturalidade da mãe, contemplada apenas no modelo hierarquizado e associada ao parto precipitado. Houve também a inclusão de uma interação entre número de orientações recebidas e parto precipitado. A modelagem hierarquizada permitiu que variáveis associadas entre si ficassem no modelo final (colinearidade). A exploração das relações entre as variáveis foi realizada quando se empregou o procedimento *stepwise*. Independentemente da escolha do processo de seleção de variáveis ou modelo existem pontos que devem ser relevados: revisão exaustiva da literatura sobre o evento em estudo, análise univariada cuidadosa e avaliação das inter-relações ente as variáveis.

ABSTRACT

In this work it was discussed the application of two methods for selection of predictor variables and models in multiple logistic regression analysis: hierarchical model and a stepwise model. In a case-control study conduced to identify risk factors associated to neonatal mortality in Campinas, São Paulo, the effects of socio-economic, maternal morbidity and health care were studied. The study included 117 cases and 234 controls and the supplementary data were obtained from household interviews. The multiple logistic regression analysis, in a hierarchical model, identified as associated to neonatal death risk: income, immigration, number of dwellers, the choice of delivery hospital, vaginal bleeding, early delivery due to health problems, time elapsed between hospital admission and delivery, number of orientations received, gestational age, low birth weight and APGAR score at 5^{th} minute. The differences found for the stepwise model were: income was associated with the choice of delivery hospital, vaginal bleeding was associated with early delivery due to health problems and immigration. Immigration was selected only by the hierarchical model and was associated with early delivery due to health problems. A interaction effect between number of orientations received and early delivery due to health problems was included in model selected by the stepwise procedure. The hierarchical modeling allowed associated variables to be in the final model (collinearity). The inter-relation between variables was investigated with the stepwise procedure. Independently of the choice of the process of selection of variables or models there are points that are important: exhaustive literature review, a careful univariate analysis and the evaluation of the relationships between two or more independent variables.

1- INTRODUÇÃO

O grande marco da aplicabilidade e utilidade da técnica de regressão logística, principalmente na área da saúde, foi o estudo dos fatores de risco para doença coronariana em Framingham publicado por TRUETT, CORNFIELD e KANNEL (1967).

A partir daí, os modelos de regressão logística são amplamente utilizados em investigações epidemiológicas, principalmente quando envolvem grande número de variáveis consideradas como fatores de risco.

1.1- Regressão logística

Resumidamente, a regressão logística descreve a relação entre uma variável resposta discreta (dicotômica é o caso mais comum) e uma ou mais variáveis independentes, freqüentemente chamadas de variáveis preditoras, explicativas ou covariáveis.

Em qualquer problema envolvendo modelos de regressão simples procura-se quantificar o valor médio da variável resposta dado um determinado valor da variável independente. Esta quantidade é chamada média condicional e pode ser expressa como $E(Y|x)$, onde Y representa a variável resposta e x um valor da variável independente. Em regressão linear simples (variável resposta contínua) esta média pode ser expressa através de uma equação linear em x, como segue:

$E(Y|x)=\beta_0+\beta_1 x$, onde β_0 e β_1 são parâmetros desconhecidos e $E(Y|x)$ pode assumir qualquer valor entre $-\infty$ e $+\infty$, em função de x.

No caso de resposta dicotômica a média condicional deve ser maior ou igual a zero e menor ou igual a um, isto é, $0 \leq E(Y|x) \leq 1$. A curva desta média tem forma de "S" e assemelha-se a função de distribuição acumulada de uma variável aleatória com distribuição logística.

Para representar a média condicional de Y dado x quando a distribuição logística é usada, utiliza-se a quantidade $\pi(x)$. A forma específica do modelo de regressão logística é apresentada a seguir:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \qquad (1.1)$$

Aplicando-se a transformação logito a $\pi(x)$, a função resultante, $g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x$, é linear em seus parâmetros, pode ser contínua e variar entre $-\infty$ e $+\infty$, dependendo de x.

Os erros (quantidades que expressam os desvios entre cada observação e a média condicional) num modelo de regressão logística apresentam distribuição binomial.

Para o ajuste do modelo de regressão logística descrito pela função (1.1) a um conjunto de dados é preciso estimar os valores dos parâmetros β_0 e β_1, desconhecidos.

Um método geral de estimação é conhecido como máxima verossimilhança. Este método produz valores para os parâmetros que maximizam a probabilidade de se obter o conjunto de dados observado. Para sua aplicação constrói-se uma função chamada função de verossimilhança, que expressa a probabilidade dos valores observados como função dos parâmetros desconhecidos.

Se Y é codificada como 0 ou 1, então $\pi(x)$ expressa a probabilidade condicional de Y igual a 1 dado x, denotada como $P(Y=1|x)$ e, a quantidade $1-\pi(x)$ a probabilidade condicional de Y igual a 0 dado x, $P(Y=0|x)$. Assim, para os pares (x_i,y_i), onde $y_i=1$, a contribuição na função de verossimilhança é $\pi(x_i)$ e onde $y_i=0$ a contribuição é $1-\pi(x_i)$. A contribuição na função de verossimilhança para os pares (x_i,y_i), em termos gerais pode ser expressa por:

$$\zeta(x_i) = \pi(x_i)^{y_i}\,[1-\pi(x_i)]^{1-y_i}$$

Supondo-se que as observações sejam independentes, a função de verossimilhança é obtida pelo produto dos termos dados pela expressão acima, como segue:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \zeta(x_i)$$

Com a transformação logarítmica aplicada à expressão acima se obtém:

$$L(\boldsymbol{\beta}) = \ln[l(\boldsymbol{\beta})] = \sum_{i=1}^{n} \{ y_i \ln[\pi(x_i)] + (1 - y_i) \ln[1 - \pi(x_i)] \}$$

Determinam-se os valores que maximizam a função acima derivando $L(\boldsymbol{\beta})$ em relação à β_0 e β_1 e igualando as expressões resultantes a zero, como segue:

$$\sum_{i=1}^{n} [y_i - \pi(x_i)] = 0 \quad \text{e} \quad \sum_{i=1}^{n} x_i [y_i - \pi(x_i)] = 0$$

Métodos de cálculo numérico, por exemplo Newton-Raphson, são necessários para encontrar as estimativas de máxima verossimilhança denotadas por $\hat{\beta}_0$ e $\hat{\beta}_1$.

O modelo logístico pode ser generalizado para o caso de mais de uma variável independente. As variáveis independentes podem ser quantitativas ou qualitativas, neste caso, representadas por variáveis indicadoras (variáveis *dummy*).

Considerando um conjunto de *p* variáveis independentes, o logito do modelo de regressão logística múltiplo, para as *p* variáveis observadas é dado pela seguinte equação:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$$

Encontra-se ampla literatura sobre os aspectos teóricos do modelo de regressão logística e suas aplicações. KLEINBAUM (1994), HOSMER e LEMESHOW (1989), TABACHNICK e FIDELL (2001), entre outros, são de fácil compreensão.

1.2- Seleção de variáveis

Existem várias estratégias para a construção de modelos e o objetivo de qualquer uma delas, é selecionar as variáveis que resultem no "melhor" modelo dentro do contexto operacional do problema. O sucesso para a modelagem de um conjunto de dados complexo está relacionado à área específica, aos métodos estatísticos e à experiência e bom senso do pesquisador (HOSMER e LEMESHOW, 1989).

Os critérios para inclusão de uma variável num modelo variam de acordo com o problema e com a área de aplicação. O caminho tradicional da estatística é procurar por um modelo parcimonioso, estável e que descreva o fenômeno estudado.

Alguns passos devem ser seguidos para a construção do modelo:

1. realizar análises univariadas cuidadosas;

2. identificar variáveis com potencial impacto;

3. estudar as inter-relações entre as diferentes variáveis. É recomendado que somente as interações que tenham justificativas *a priori* para serem consideradas ou que sejam biologicamente importantes devam ser investigadas (HOSMER e LEMESHOW, 1989);

4. decidir qual, ou quais técnicas para seleção de variáveis serão empregadas.

O termo confundimento[1] é utilizado pelos epidemiologistas para descrever uma covariável que está associada tanto com o desfecho em estudo quanto com alguma outra variável independente. Quando estas associações acontecem, a relação entre a variável independente e a variável resposta é dita estar confundida.

Outro conceito empregado, e também objeto de atenção na modelagem, é a interação. Ela representa a interdependência entre duas ou mais variáveis independentes que altera a magnitude do efeito. Por exemplo, existe interação entre variáveis quando o coeficiente de incidência de uma doença na presença de dois ou mais fatores de risco difere

[1] Do inglês *counfunding* que tem o significado de "confusão de variáveis". "Confundimento" é termo dicionarizado no nosso idioma, estando registrado no vocabulário Ortográfico da Língua Portuguesa, publicação oficial da Academia Brasileira de Letras, datada de 1981 (PEREIRA, 1995).

do coeficiente de incidência que seria esperado pela combinação dos efeitos individuais destes fatores de risco (PEREIRA, 1995).

Alguns dos processos de seleção de variáveis mais utilizados, cujos algoritmos foram implementados em programas computacionais, são resumidamente descritos abaixo:

1. *forward* (seleção à frente) - este procedimento começa com a escolha da variável independente que melhor explica a variável dependente comparando-se, pelo teste da razão de verossimilhança, o modelo ajustado apenas com o intercepto e cada modelo univariado. Assim, a primeira variável escolhida é a que apresenta menor p-valor neste teste. O próximo passo é escolher uma segunda variável que produza o maior aumento na razão de verossimilhança quando adicionada ao modelo. Novamente aplica-se o teste da razão de verossimilhança para verificar se a contribuição desta nova variável é significativa. O processo continua até que nenhuma variável acrescida no modelo cause aumento significativo na razão de verossimilhança. Uma característica importante desse procedimento é que, uma vez que a variável foi selecionada e incluída por ser significativa, ela não deve mais ser excluída;

2. *backward* (seleção para trás) – este procedimento começa pelo ajuste de todas as variáveis independentes candidatas a ficar no modelo. Compara-se o desvio do modelo logístico contendo todas as variáveis com os desvios dos modelos que resultam da exclusão individual de cada variável. Se o nível descritivo do teste da razão de verossimilhança for significativo a variável fica no modelo e o procedimento se encerra; se não for, ela sai do modelo. Das variáveis que restaram no modelo, novamente se escolhe a que menos contribui e testa-se a sua significância. Se for significativa, ela fica no modelo e o processo se encerra, caso contrário ela sai do modelo e o processo continua. Neste procedimento uma vez que a variável sai do modelo ela não entra mais;

3. *stepwise* (seleção passo a passo) – o procedimento começa com o passo à frente, mas depois que a segunda variável entra no modelo, o teste da razão de verossimilhança é realizado para verificar se a primeira variável permanece no modelo. Caso permaneça, uma terceira variável é selecionada da mesma forma que no procedimento passo à frente. Se uma terceira variável entra no modelo, testa-se para verificar se as duas primeiras continuam no modelo. Pode acontecer que uma delas ou as duas sejam eliminadas. Tenta-se então a inclusão de uma nova variável. Caso entre, tenta-se a eliminação das que já estão no modelo. O procedimento acaba quando não se consegue nem adicionar, nem eliminar variáveis;

4. e, o menos utilizado, *best subsets* (melhores subgrupos) – este procedimento calcula a regressão para todos os possíveis subconjuntos de variáveis independentes. Modelos contendo uma, duas, três variáveis e assim por diante, até um único modelo com todas as variáveis. A avaliação dos subconjuntos é feita pela estatística C_q que mede a qualidade do subconjunto selecionado com q variáveis independentes. Para um subconjunto de q variáveis de um total de p variáveis, $C_q = \frac{X^2 + \lambda^*}{X^2/(n-p-1)} + 2(q+1) - n$, onde X^2 é a estatística Qui-quadrado de Pearson para o modelo com as p variáveis, λ^* é a estatística do teste de Wald sob a hipótese de que os coeficientes do modelo com as p-q variáveis são iguais a zero e n é o tamanho da amostra. Modelos com C_q próximos de (q+1) são candidatos à melhores modelos.

Os procedimentos de seleção de variáveis são baseados em algoritmos estatísticos que verificam a "importância" da variável e a inclui ou exclui do modelo baseados numa regra fixa de decisão. A "importância" da variável é definida em termos de uma medida da significância estatística do coeficiente estimado para a variável. A significância é obtida pelo teste da razão de verossimilhança que, basicamente, compara o modelo sem a variável com o modelo ajustado para a variável.

Estes procedimentos, em especial o *stepwise*, têm sido criticados, pois podem gerar modelos puramente matemáticos, ajustando variáveis que não são fatores de confundimento, uma vez que os pesquisadores freqüentemente não possuem conhecimento prévio sobre estes fatores (GREENLAND, 1989; MALDONADO e GREENLAND, 1993).

A seleção do modelo e variáveis apropriadas, assim como o modo como as variáveis entram no modelo, são tarefas complexas que devem ser executadas tomando-se o cuidado de explorar ao máximo as inter-relações entre as variáveis (HENNEKENS e BURING, 1987).

Embora a análise múltipla seja útil no controle de confundimentos, é importante que seja realizada e interpretada cuidadosamente.

Métodos de modelagem hierarquizada têm sido sugeridos como alternativa para lidar com as limitações dos métodos convencionais aplicados a estudos epidemiológicos com grande número de covariáveis. A hierarquia pode estar presente na unidade amostral, no modo de entrada das variáveis e entre as variáveis independentes.

A escolha de critérios para a seleção de variáveis de confundimento ultrapassa o aspecto puramente estatístico. A hierarquização das variáveis independentes é estabelecida no marco conceitual e mantida durante a análise dos dados, permitindo a seleção daquelas mais fortemente associadas com o desfecho de interesse. Ao estudarem-se múltiplas exposições, o modelo hierarquizado indica a ordem em que devem entrar as variáveis. Inicialmente determinam-se as variáveis mais fortemente associadas com o desfecho dentro de cada bloco. Para as análises subseqüentes mantém-se no modelo as variáveis que permanecem associadas após o ajuste para as variáveis dos blocos hierarquicamente superiores. A construção da estrutura conceitual requer conhecimento de determinantes sociais, biológicos e de outras naturezas para o evento estudado (FUCHS, 1993).

Assim, a seleção de um modelo logístico múltiplo deve ser um processo conjugado de seleção estatística e experiência do pesquisador.

1.3- Qualidade de ajuste do modelo

Além da seleção de variáveis de confundimento, outro ponto de interesse refere-se à verificação dos pressupostos do modelo. Uma vez que a validade das inferências depende de quão bem o modelo descreve os dados observados, torna-se necessário saber se o modelo foi bem ajustado.

Basicamente esta verificação é realizada através de uma medida da distância entre os valores observados e os valores preditos (ajustados) pelo modelo.

Na regressão logística existem várias medidas de diferença entre valores observados e ajustados (resíduos).

Os resíduos de Pearson e a Deviance são utilizados para identificar observações que não contribuem para a explicação do modelo.

A estatística Qui-quadrado de Pearson é a soma dos quadrados dos resíduos de Pearson.

A Deviance é a estatística do teste da razão de verossimilhança do modelo saturado com J (J>p+1) parâmetros em relação ao ajustado com (p+1) parâmetros. J representa o número de sub-populações (observações com valores iguais são consideradas como vindo de uma mesma sub-população) e p o número de variáveis independentes do modelo ajustado.

Estas estatísticas, sob a hipótese de que o modelo ajustado está correto em todos os aspectos, assumem distribuição Qui-quadrado com J-(p+1) graus de liberdade.

Os detalhes teóricos das medidas citadas aqui, bem como de outras medidas, são descritos no capítulo 5 de HOSMER e LEMESHOW (1989).

Os métodos empregados na criação de modelos são sujeitos a erros e não há um método único para se identificar o melhor modelo (GREENLAND, 1989).

Alguns trabalhos avaliam os modelos de regressão logística múltipla obtidos por diferentes métodos de seleção de variáveis.

MOSLEY e CHEN (1984) apresentaram uma proposta de análise da mortalidade infantil para países em desenvolvimento utilizando uma estrutura analítica que envolvia determinantes próximos (variáveis intermediárias), determinantes sociais e biológicos. Os determinantes próximos teriam um papel intermediário entre o nível que a influenciaria diretamente e os fatores sócio-econômicos, culturais, políticos e outros. Por sua vez, o impacto dos fatores sócio-econômicos sobre a saúde somente seria processado através de seus efeitos sobre os determinantes próximos. Indicaram o uso deste modelo para facilitar a especificação dos diferentes níveis de causalidade e possíveis interações entre os determinantes socioeconômicos, sugerindo que a mortalidade infantil deva ser estudada como um processo multifatorial.

GREENLAND (1994) comparou métodos de modelagem hierárquica com métodos convencionais de seleção de variáveis e sugere que os métodos hierárquicos devam ser utilizados em estudos de múltipla exposição.

FUCHS et al. (1996) concluem que a modelagem hierarquizada representa uma alternativa aos métodos tradicionais de análise, pois contempla os aspectos biológicos e estatísticos permitindo estruturar a investigação de fatores de risco e facilitar a interpretação.

WITTE e GREENLAND (1996) em um estudo de simulação mostraram que o modelo hierárquico geralmente consegue estimativas de efeito mais acuradas do que as técnicas convencionais.

VICTORA et al. (1997) questionam as estratégias apropriadas para análise de estudos epidemiológicos, especialmente na presença de uma estrutura complexa de inter-relação entre as variáveis. Têm aplicado a modelagem hierarquizada em diversos trabalhos e acreditam que esta técnica auxilia na interpretação dos resultados considerando o conhecimento social e biológico.

A contribuição do presente trabalho é apresentar empiricamente a aplicação de dois métodos de seleção de variáveis e modelos em regressão logística ao mesmo conjunto de dados.

Não se pretende identificar o melhor método de seleção de variáveis, apenas apontar as vantagens e desvantagens de um e de outro método.

Paralelamente são estudados os fatores de risco para a mortalidade infantil no município de Campinas - SP.

2- OBJETIVOS

2.1- Geral

Com base em um problema real, ajustar modelos de regressão logística múltipla utilizando um método convencional de seleção de variáveis, no caso, o procedimento *stepwise* e as recomendações apresentadas por HOSMER e LEMESHOW, 1989 e o método de modelagem hierarquizada segundo proposta de VICTORA et al. (1997).

2.2- Específico

Como exemplo de aplicação dos procedimentos de seleção e ajuste de modelos, identificar fatores associados à mortalidade neonatal.

3- MATERIAL E MÉTODOS

O banco de dados utilizado neste trabalho é o utilizado por ALMEIDA (2002). Trata-se de um estudo caso-controle desenvolvido para identificar fatores associados ao óbito neonatal. Este estudo foi realizado no município de Campinas (SP).

Foram considerados "casos" todos os nascidos vivos que morreram antes de completar 28 dias de vida no período compreendido entre primeiro de março de 2001 a 28 de fevereiro de 2002, cujo nascimento ocorreu em Campinas e que eram filhos de mães residentes neste município. A relação dos óbitos neonatais ocorridos foi obtida do Banco de Dados sobre Mortalidade (SIM) da Secretaria Municipal de Saúde. Os "controles" foram selecionados por meio de sorteio aleatório entre as crianças nascidas na mesma data que o caso e que sobreviveram ao vigésimo oitavo dia de vida. As listagens para sorteio foram obtidas do Banco de Dados de Nascidos Vivos (SINASC) da Secretaria Municipal de Saúde. Foram sorteados dois controles para cada caso.

Foram estudados 117 casos e 234 controles não-pareados. As informações adicionais foram obtidas através de entrevistas domiciliares com a mãe ou o responsável pela criança.

As variáveis analisadas são apresentadas no quadro 1.

Quadro 1- Nome, descrição, categorização utilizada e nível hierárquico para as variáveis investigadas.

Nome	Descrição	Categorias	Nível
CASOCONTRO	grupo de estudo	1=caso; 2=controle	5
TBENS_A	total de bens e equipamentos da família	0-2; 3-7; ≥8	1
REFAM	renda familiar em salários mínimos	contínua	1
SALARIOF	renda familiar em faixas de salário mínimo	0-2; >2	1
ESTUD_A	escolaridade da mãe em faixas de anos de estudo	0-4; 5-8; ≥9	1
NCOMOD	número de cômodos	até 2; ≥3	2
DOMIC	tipo de acabamento do domicílio	alvenaria com acabamento completo; outro	2
IMOV	tipo de propriedade	1=próprio/alugado; 2=outros (cedido/ invasão/ocupação)	2
TEMPOD	tempo no domicílio (anos)	≤1; 1-10; >10	2
SANEAM	condições de saneamento	adequadas; inadequadas	2
NPESSOA	número de pessoas no domicílio	até 3; ≥4	2
OCUPA	ocupação da mãe	empregada doméstica; outra ocupação; não trabalha	2
MESEST	quantos meses trabalhou durante a gestação	0; até 6 ; 7-9	2
ESFORCOF	esforço físico no trabalho	nenhum; leve/moderado grande	2
DESGASTE	desgaste emocional no trabalho	nenhum; leve/moderado grande	2
RACA_A	raça ou cor da mãe	branca; não branca	2
NAT_A	naturalidade da mãe	Campinas e região; outras	2
HABITODEFU	hábito de fumar	não; fumante; ex-fumante	2
FILHOSV	número de filhos vivos	0; ≥1	3
IDMAEAD	idade da mãe em faixas	≤19; ≥20	3
DOENCADURA	doença durante a gravidez	sim; não	3
TRATOUPRES	tratou de pressão alta	sim; não	3
TRATOUINFE	tratou de infecção urinária	sim; não	3
APRESENTOU	apresentou sangramento vaginal	sim; não	3
FOIINTERNA	foi internada durante a gestação	sim; não	3
NORI	número de orientações* recebidas no pré-natal	0-2; 3-5; 6-7	3
NEXROT	total de exames de rotina	0-7; ≥8	3
NPROC	total de procedimentos recebidos no pré-natal	0-4; ≥5	3
VISITPN	número de visitas ao pré-natal	0-7; ≥8	3

Quadro 1- (continuação). Nome, descrição, categorização utilizada e nível hierárquico para as variáveis investigadas.

Nome	Descrição	Categorias	Nível
CONVPRE	convênio do pré-natal	SUS; outro	3
ESCOLHEUM	escolheu o médico para o pré-natal	sim; não	3
MAIORPARTE	maior parte das consultas realizadas pelo mesmo médico	sim; não	3
ENCONTROUD	encontrou alguma dificuldade para iniciar o pré-natal	sim; não	3
FEZECOGRAF	fez ecografia	sim; não	3
TIPOPART	tipo de parto	vaginal; cesárea	3
SINAL_A	sinais apresentados antes do parto	usuais; não referiu; não usuais	3
PARTOPRECI	parto precipitado	sim; não	3
CONVPART	convênio do parto	SUS; outro	3
HOSPITALFO	hospital foi da sua escolha	sim; não	3
TE	tempo entre internação e o parto em horas	0-1; 1-10; >10	3
MEDPART	médico que realizou o parto	mesmo do pré-natal; outro	3
ESTN	estabelecimento do parto	SUS; hospital escola; outro	3
TIPODEGEST	tipo de gestação	única; gemelar	4
SEXO	sexo do recém-nascido	masculino; feminino	4
DGEST_A	idade gestacional em semanas	até 36; ≥37	4
BXPESO	peso ao nascer em faixas	<2500g; ≥2500g	4
APG5	Apgar do quinto minuto	0-8; 9-10	4

* orientações recebidas durante as consultas ou em atividades de grupo sobre: ganho de peso, uso de medicamentos, evolução da gestação, sinais de parto, amamentação e vacinação

O modelo hierarquizado, proposto na análise de regressão logística múltipla, obedeceu ao esquema descrito na figura 1.

Para caracterização da amostra estudada são apresentadas tabelas de freqüências para as variáveis categóricas e medidas de posição e dispersão para variáveis contínuas.

Para identificar fatores de risco para óbito neonatal foi utilizada a análise de regressão logística não condicional univariada e múltipla. Os modelos apresentados foram definidos segundo dois critérios de seleção de variáveis: *stepwise* e modelagem hierarquizada.

As análises foram realizadas utilizando-se o software estatístico SAS®. O programa para processamento das análises encontra-se no Anexo I.

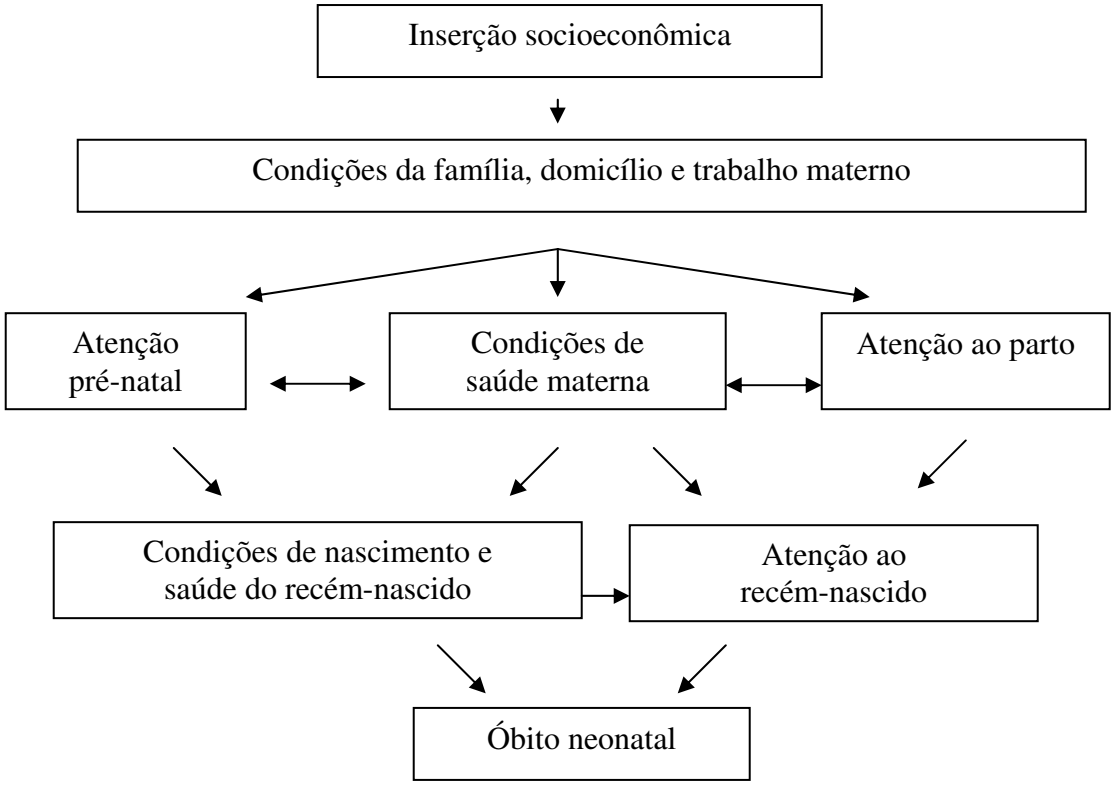

Figura 1- Modelo de análise hierarquizada para o óbito neonatal, segundo ALMEIDA (2002).

4- RESULTADOS

Parte dos resultados apresentados neste trabalho foi retirada da tese de ALMEIDA (2002). Foram mantidas as categorias de respostas para todas as variáveis apresentadas na referida tese.

A análise univariada é o primeiro passo para identificar possíveis problemas nas distribuições das variáveis.

A tabela 1 apresenta os resultados referentes às características sócio-econômicas da mãe e da família. A renda familiar esteve associada ao óbito neonatal. Para as famílias com renda inferior a 2 salários mínimos o risco foi aproximadamente 2 vezes maior que para as de maior renda. Não houve associação estatisticamente significativa com as demais variáveis sócio-econômicas, inclusive com a escolaridade da mãe.

Tabela 1- Distribuição dos casos (óbitos neonatais), controles, *odds ratio* bruta e respectivo intervalo de 95% de confiança, segundo variáveis sócio-econômicas pertencentes ao primeiro nível hierárquico. Campinas, SP, 2001.

Variável	**Casos n (%)**	**Controles n (%)**	**OR**	**IC95%**	**p-valor**
Número de bens e equipamentos					
0-2	12 (10,3)	25 (10,7)	1,15	0,51-2,59	0,7319
3-7	75 (64,1)	137 (58,5)	1,31	0,79-2,19	0,2947
>=8	30 (25,6)	72 (30,8)	1,00		
Renda familiar em salários mínimos*	6,3 (7,7)	6,6 (6,6)	1,00	0,96-1,03	0,7611
Remuneração familiar (em faixas de salários mínimos)					
0-2	32 (27,4)	38 (16,2)	1,94	1,13-3,31	0,0150
>2	85 (72,6)	196 (83,8)	1,00		
Escolaridade da mãe em faixas de anos de estudo					
0-4	19 (16,2)	28 (12,0)	1,47	0,76-2,85	0,2494
5-8	40 (34,2)	80 (34,2)	1,09	0,66-1,77	0,7412
>=9	58 (49,6)	126 (53,8)	1,00		

* média e desvio padrão

As variáveis de condições do domicílio, condições da família, condições de trabalho materno e hábitos maternos são apresentadas na tabela 2. Os recém-nascidos cujas mães residiam em domicílios com 1 a 2 cômodos, sem acabamento de alvenaria, localizados em áreas de invasão/ocupação, com saneamento inadequado ou com até 3 moradores apresentaram riscos maiores de óbito neonatal. O tempo de trabalho materno durante a gestação e o desgaste emocional no trabalho também estiveram associados ao óbito neonatal. O risco de óbito foi maior entre recém-nascidos de mães naturais de municípios fora de Campinas e região.

Tabela 2- Distribuição dos casos (óbitos neonatais), controles, *odds ratio* bruta e respectivo intervalo de 95% de confiança segundo variáveis de condições do domicílio, família, de trabalho e hábitos maternos pertencentes ao segundo nível hierárquico, Campinas, SP, 2001.

Variável	Casos n (%)	Controles n (%)	OR	IC95%	p-valor
Número de cômodos do domicílio					
1-2	10 (8,5)	8 (3,4)	2,64	1,01-6,88	0,0470
>=3	107 (91,5)	226 (96,6)	1,00		
Tipo de acabamento do domicílio					
Outros	62 (53,0)	82 (35,0)	2,09	1,33-3,28	0,0014
Alvenaria completa	55 (47,0)	152 (65,0)	1,00		
Local de moradia					
Áreas de invasão/ocupação	20 (17,1)	19 (8,1)	2,33	1,19-4,56	0,0135
Outros	97 (82,9)	215 (91,9)	1,00		
Tempo de residência no domicílio (anos)					
<=1	54 (46,1)	88 (37,6)	1,93	0,97-3,85	0,0623
1-10	49 (41,9)	102 (43,6)	1,51	0,76-3,01	0,2431
>10	14 (12,0)	44 (18,8)	1,00		
Tipo de saneamento (água/lixo/esgoto)					
Inadequado	30 (25,6)	39 (16,7)	1,72	1,01-2,96	0,0476
Adequado	87 (74,4)	195 (83,3)	1,00		
Número de moradores no domicílio					
<=3	65 (55,6)	66 (28,2)	3,18	2,00-5,05	<0,0001
>=4	52 (44,4)	168 (71,8)	1,00		
Ocupação da mãe					
Empregada doméstica	25 (21,4)	29 (12,4)	1,75	0,94-3,24	0,0740
Outras	30 (25,6)	79 (33,8)	0,77	0,46-1,30	0,3277
Não trabalha fora	62 (53,0)	126 (53,8)	1,00		
Tempo de trabalho durante a gestação					
Não trabalhou	62 (53,0)	126 (53,8)	1,00		
Até 6 meses	37 (31,6)	31 (13,2)	2,43	1,38-4,27	0,0021
Toda gestação	18 (15,4)	77 (32,9)	0,48	0,26-0,86	0,0145
Esforço físico trabalho					
Não trabalhou	62 (53,0)	126 (53,8)	1,00		
Moderado/leve	33 (28,2)	84 (35,9)	0,80	0,48-1,32	0,3818
grande	22 (18,8)	24 (10,3)	1,86	0,97-3,58	0,0621
Desgaste emocional no trabalho					
Não trabalhou	62 (53,9)	126 (53,8)	1,00		
Moderado/leve	32 (27,3)	84 (35,9)	0,77	0,47-1,29	0,3236
Grande	23 (19,7)	24 (10,3)	1,95	1,02-3,72	0,0437
Raça/cor					
Branca	54 (46,1)	126 (53,8)	1,00		
Não branca	63 (53,9)	108 (46,2)	1,36	0,87-2,12	0,1747
Naturalidade da mãe					
Campinas e região	48 (41,0)	126 (53,8)	1,00		
Outra	69 (59,0)	108 (46,2)	1,68	1,07-2,63	0,0241
Hábito de fumar					
Não fumante	77 (65,8)	149 (63,7)	1,00		
Ex-fumante	15 (12,8)	48 (20,5)	0,61	0,32-1,15	0,1248
fumante	25 (21,4)	37 (15,8)	1,31	0,73-2,33	0,3626

A tabela 3 mostra os efeitos das condições de saúde da mãe durante a gestação e da atenção ao pré-natal e parto. A presença de alguma doença durante a gestação, tratamento para hipertensão arterial, presença de sangramento vaginal e internações durante a gestação estiveram associadas ao óbito neonatal. Das variáveis de atenção e qualidade do pré-natal, o número de orientações recebidas, o número total de procedimentos recebidos, o número de consultas, a não escolha do médico, não fazer a maior parte das consultas com o mesmo médico, a dificuldade em realizar o pré-natal e não fazer ecografia durante o pré-natal foram fatores significativamente associados ao óbito neonatal. Das condições do parto apresentaram maior risco os recém-nascidos de mães que apresentaram sinais não usuais antes da internação, internação precipitada por problemas de saúde, não escolheram o hospital para o parto ou parto realizado pelo plantonista. O tempo decorrido entre a internação e o parto também esteve significativamente associado ao óbito neonatal.

Tabela 3- Distribuição dos casos (óbitos neonatais), controles, *odds ratio* bruta e respectivo intervalo de 95% de confiança segundo condições de saúde da mãe durante a gestação e de atenção ao pré-natal e parto, pertencentes ao terceiro nível hierárquico, Campinas, SP, 2001.

Variável	**Casos n (%)**	**Controles n (%)**	**OR**	**IC95%**	**p-valor**
Número de filhos vivos					
0	60 (51,3)	100 (42,7)	1,41	0,90-2,20	0,1302
≥1	57 (48,7)	134 (57,3)	1,00		
Idade da mãe					
≤19	25 (21,4)	38 (16,2)	1,40	0,80-2,46	0,2392
≥20	92 (78,6)	196 (83,8)	1,00		
Doença durante a gravidez					
Sim	72 (61,5)	104 (44,4)	2,00	1,27-3,15	0,0027
não	45 (38,5)	130 (55,6)	1,00		
Fez tratamento para pressão alta					
Sim	22 (18,8)	23 (9,8)	2,12	1,13-4,00	0,0196
não	95 (81,2)	211 (90,2)	1,00		
Fez tratamento para infecção urinária					
Sim	19 (16,2)	41 (17,5)	0,91	0,50-1,66	0,7636
não	98 (83,8)	193 (82,5)	1,00		
Sangramento vaginal durante a gestação					
Sim	26 (22,2)	19 (8,1)	3,23	1,70-6,13	<0,0001
Não	91 (77,8)	215 (91,9)	1,00		
Internação durante a gestação					
Sim	32 (27,4)	33 (14,1)	2,29	1,33-3,97	0,0030
não	85 (72,6)	201 (85,9)	1,00		
Número de orientações recebidas no pré-natal					
0-2	36 (30,8)	23 (9,8)	5,73	3,04-10,80	<0,0001
3-5	43 (36,8)	72 (30,8)	2,19	1,30-3,68	0,0033
6-7	38 (32,5)	139 (59,4)	1,00		
Número de exames de rotina					
0-7	15 (12,8)	16 (6,8)	2,00	0,95-4,21	0,0666
≥8	102 (87,2)	218 (93,2)	1,00		
Total de procedimentos recebidos no pré-natal					
0-4	11 (9,4)	3 (1,3)	7,99	2,18-29,21	0,0017
≥5	106 (90,6)	231 (98,7)	1,00		
Número de visitas ao pré-natal					
0-7	70 (59,8)	58 (24,8)	4,52	2,81-7,26	<0,0001
≥8	47 (40,2)	176 (75,2)	1,00		
Convênio do pré-natal					
SUS	66 (56,4)	125 (53,4)	1,13	0,72-1,76	0,5959
Outro	51 (43,6)	109 (46,6)	1,00		
Escolheu o médico que fez o pré-natal					
Sim	55 (47,0)	141 (60,3)	1,00		0,0002
não	62 (53,0)	93 (39,7)	1,71	1,09-2,67	

Tabela 3- (continuação). Distribuição dos casos (óbitos neonatais), controles, *odds ratio* bruta e respectivo intervalo de confiança de 95% segundo condições de saúde da mãe durante a gestação e de atenção ao pré-natal e parto, pertencentes ao terceiro nível hierárquico, Campinas, SP, 2001.

Variável	Casos n (%)	Controles n (%)	OR	IC95%	p-valor
Consultas realizadas pelo mesmo médico					
Sim	81 (69,2)	188 (80,3)	1,00		0,0213
não	36 (30,7)	46 (19,7)	1,82	1,09-3,02	
Encontrou dificuldade para fazer o pré-natal					
Sim	24 (20,5)	15 (6,4)	3,77	1,89-7,51	0,0002
Não	93 (79,5)	219 (93,6)	1,00		
Realizou ecografia durante o pré-natal					
Sim	19 (16,2)	6 (2,6)	1,00		<0,0001
Não	98 (83,8)	228 (97,4)	7,37	2,86-19,00	
Tipo de parto					
Cesárea	60 (51,3)	134 (57,3)	0,79	0,50-1,23	0,2883
vaginal	57 (48,7)	100 (42,7)	1,00		
Apresentou sinal antes da internação do parto					
Sinais não usuais	15 (12,8)	10 (4,3)	3,04	1,30-7,10	0,0101
Sinais usuais	73 (62,4)	148 (63,3)	1,00		
Não referiu sinal	29 (24,8)	76 (32,5)	0,77	0,46-1,29	0,3253
Internação do parto precipitada por problema de saúde					
Sim	72 (61,5)	66 (28,2)	4,07	2,55-6,51	<0,0001
Não	45 (38,5)	168 (71,8)	1,00		
Convênio do parto					
SUS	76 (65,0)	136 (58,1)	1,34	0,84-2,12	0,2175
outro	41 (35,0)	98 (41,9)	1,00		
Escolheu o hospital para realização do parto					
Sim	73 (62,4)	209 (89,3)	1,00		<0,0001
Não	44 (37,6)	25 (10,7)	5,04	2,88-8,81	
Tempo decorrido entre a internação e o parto (em horas)					
<1	35 (29,9)	41 (17,5)	2,66	1,54-4,59	0,0005
1-10	53 (45,3)	165 (70,5)	1,00		
>10	29 (24,8)	28 (12,0)	3,22	1,76-5,90	0,0001
Médico que realizou o parto					
Médico plantonista	93 (74,5)	161 (68,8)	1,76	1,04-2,98	0,0362
Médico que realizou o pré-natal	24 (20,5)	73 (31,2)	1,00		
Estabelecimento do parto					
Hospital-escola SUS	43 (36,7)	52 (22,2)	1,81	0,98-3,32	0,0568
Hospitais privados SUS	47 (40,2)	123 (52,6)	0,84	0,47-1,47	0,5324
Hospitais privados	27 (23,1)	59 (25,2)	1,00		

Das condições de nascimento e saúde do recém-nascido apresentadas na tabela 4, o risco de óbito neonatal foi maior entre os nascidos de gestação múltipla, com idade gestacional inferior a 37 semanas, peso menor que 2500 gramas ou com Apgar no quinto minuto menor ou igual a 8.

Tabela 4- Distribuição dos casos (óbitos neonatais), controles, *odds ratio* bruta e respectivo intervalo de 95% de confiança segundo variáveis de condições de nascimento e saúde do recém-nascido, pertencentes ao quarto nível hierárquico, Campinas, SP, 2001.

Variável	Casos n (%)	Controles n (%)	OR	IC95%	p-valor
Tipo de gestação					
Múltipla	20 (17,1)	3 (1,3)	15,88	4,61-54,66	<0,0001
única	97 (82,9)	231 (98,7)	1,00		
Sexo do recém-nascido					
Masculino	70 (59,8)	125 (53,4)	1,30	0,83-2,04	0,2550
Feminino	47 (40,2)	109 (46,6)	1,00		
Idade gestacional ao nascimento (em semanas)					
Até 36	92 (78,6)	26 (11,1)	29,44	16,14-53,72	<0,0001
≥37	25 (21,4)	208 (88,9)	1,00		
Peso ao nascimento (em gramas)					
<2500	85 (72,6)	23 (9,8)	24,37	13,48-44,05	<0,0001
≥2500	32 (27,4)	211 (90,2)	1,00		
Apgar no quinto minuto					
0-8	81 (69,8)	14 (6,0)	36,36	18,61-71,06	<0,0001
9-10	35 (30,2)	220 (94,0)	1,00		

A próxima etapa é selecionar as variáveis para a análise múltipla. HOSMER e LEMESHOW (1989) utilizam como critério escolher apenas àquelas que apresentaram p-valor<0,25 na análise univariada. Neste trabalho este critério não foi utilizado, sendo assim, partindo-se do modelo saturado (com todas as variáveis) aplica-se o método de seleção *stepwise* para identificar apenas os efeitos principais associados ao óbito neonatal. O resumo deste processo pode ser observado na tabela 5.

Tabela 5- Resumo do processo de seleção de variáveis *stepwise* para estudar o óbito neonatal.

Passo	Variável incluída	Variável removida	p-valor teste de Wald
1	Apg5		<0,0001
2	Dgest_a		<0,0001
3	Hospitalfo		<0,0001
4	Bxpeso		0,0004
5	Partopreci		0,0089
6	Nori		0,0335
7	Foiinterna		0,0448
8	Te		0,0182
9	Npessoa		0,0135
10	refam		0,0538
11		refam	0,0555

Os parâmetros estimados para o modelo com os efeitos principais selecionados são apresentados na tabela 6.

Conhecendo as variáveis do modelo reduzido deve-se investigar a necessidade da inclusão de interações e definir quais as principais.

Tabela 6- Resultados da regressão logística múltipla, modelando o risco de óbito neonatal pelo processo de seleção de variáveis *stepwise* – efeitos principais.

Parâmetro	Estimativa	Erro padrão	p-valor	OR	IC95%
Intercepto	- 6, 0967	0,8170	<0,0001	-	-
npessoa <=3 x >4	1,2206	0,5074	0,0161	3,389	1,254; 9,163
FOIINTERNA sim x não	1,6799	0,6195	0,0067	5,365	1,593; 18,067
nori 0-2 x 6-7	1,6401	0,6382	0,0102	5,156	1,476; 18,012
nori 3-5 x 6-7	1,2382	0,5489	0,0241	3,449	1,176; 10,115
PARTOPRECI sim x não	1,3286	0,4842	0,0061	3,776	1,462 9,753
HOSPITALFO não x sim	2,7073	0,6138	<0,0001	14,989	4,501; 49,918
te 0-1h x 1-10h	1,4411	0,5348	0,0070	4,225	1,481; 12,054
te >10h x 1-10h -	0,2113	0,6697	0,7524	0,810	0,218; 3,008
dgest_a ate 36s x >=37s	1,4081	0,5749	0,0143	4,088	1,325; 12,614
BXPESO <2500 x >=2500	1,4946	0,5953	0,0120	4,458	1,388; 14,315
apg5 0-8 x 9-10	3,9320	0,5953	<0,0001	51,007	15,883; 163,800

Critérios de adequação do ajuste: Deviance (p-valor=1,0000), Pearson (p-valor=0,7094)

Qui-quadrado dos resíduos (p-valor=0,4214)

OR= odds ratio (razão de chances)

IC95%= Intervalo de 95% de confiança para a odds

Os resultados da inclusão de cada interação no modelo controlando para as demais variáveis encontram-se no quadro 2, Anexo II. Dentre as interações testadas, somente NORI*PARTOPRECI (número de orientações recebidas na consulta de pré-natal e internação para o parto precipitada por problema de saúde) foi significativa. Acrescentado a interação no modelo com os efeitos principais os resultados são apresentados na tabela 7, a seguir.

Tabela 7- Resultados da regressão logística múltipla, modelando o risco de óbito neonatal pelo processo de seleção de variáveis *stepwise* – modelo final.

Parâmetro		**Estimativa**	**Erro padrão**	**p-valor**	**OR**	**IC95%**
Intercepto		-5,5669	0,7699	<,0001	**-**	**-**
npessoa	<=3 x >=4	1,2627	0,4930	0,0104	3,535	1,345; 9,290
FOIINTERNA	sim x não	1,4581	0,5995	0,0150	4,298	1,327; 13,917
nori	0-2 x 6-7	0,2352	0,8193	0,7740	-	-
nori	3-5 x 6-7	0,5069	0,6969	0,4670	-	-
PARTOPRECI	sim x não	-0,0684	0,6789	0,9197	-	-
HOSPITALFO	não x sim	2,4105	0,6137	<0,0001	11,139	3,346; 37,088
te	0-1h x 1-10h	1,4218	0,5490	0,0096	4,145	1,413; 12,156
te	>10h x 1-10h	-0,0382	0,6738	0,9547	0,962	0,257; 3,605
dgest_a	ate 36s x >=37s	1,3855	0,5862	0,0181	3,997	1,267; 12,609
BXPESO	<2500 x >=2500	1,6097	0,5978	0,0071	5,001	1,550; 6,141
apg5	0-8 x 9-10	3,9344	0,5689	<0,0001	51,130	6,767;155,919
nori*PARTOPRECI	0-2 sim	5,1065	1,6916	0,0025	-	-
nori*PARTOPRECI	3-5 sim	1,8649	1,0282	0,0697	-	-

Critérios de adequação do ajuste: Deviance (p-valor=1,0000), Pearson (p-valor=0,9994)

OR= odds ratio (razão de chances)

IC95%= Intervalo de 95% de confiança para a odds

Na presença da interação entre um fator de risco e outra variável independente, a estimativa da *odds ratio* para o fator de risco depende do nível da variável que está participando da interação. Fixando o parto precipitado as razões de chances calculadas para o número de orientações recebidas no pré-natal são detalhadas na tabela 8.

Tabela 8- *Odds ratio* estimada para o número de orientações recebidas durante o pré-natal, no modelo selecionado pelo *stepwise* para estudo do óbito neonatal, controlando para o parto precipitado.

PARTOPRECI	NORI	OR
Sim	0-2 x 6-7	154,18
	3-5 x 6-7	6,03
Não	0-2 x 6-7	0,01
	3-5 x 6-7	0,17

Utilizando a análise de regressão logística múltipla com modelo hierarquizado, ALMEIDA (2002) identificou como fatores de risco para o óbito neonatal a renda familiar (OR=2,02, IC95%:1,08-3,79), o número de pessoas no domicílio (OR=2,25, IC95%:1,34-3,77), a naturalidade da mãe (OR=2,39, IC95%:1,29-4,44), o sangramento vaginal (OR=3,36, IC95%:1,40-8,04), o parto precipitado por problema de saúde (OR=4,94, IC95%:2,64-9,24), o número de orientações recebidas durante o pré-natal (OR=5,22, IC95%:2,13-12,79 para 0 a 2 e OR=2,29, IC95%:1,16-4,54 para 3 a 5), a escolha do hospital do parto (OR=6,26, IC95%:2,86-13,68), o tempo decorrido entre a internação e o parto (OR=2,42, IC95%:1,17-4,99 para menos que 1 h e OR=1,63, IC95%:0,73-3,61 para mais que 10 h), a idade gestacional (OR=5,73, IC95%:1,83-17,98), o baixo peso ao nascer (OR=3,84, IC95%:1,18-12,50) e o Apgar de quinto minuto (OR=32,19, IC95%:11,35-91,25). As variáveis trabalho durante a gestação e número de visitas ao pré-natal estiveram significativamente associadas à duração da gestação e seus efeitos permanecem apenas para os pré-termos. A idade gestacional atuou como fator de confundimento para a associação do óbito neonatal com o número de visitas ao pré-natal.

Os detalhes das mudanças ocorridas nas razões de chances brutas e ajustadas, seguindo a hierarquia definida, podem ser encontrados em ALMEIDA e BARROS (2004).

5- DISCUSSÃO

Comparando os resultados obtidos, a diferença entre os modelos em relação à variável renda familiar deve-se ao fato de sua associação com a escolha do hospital (p-valor=0,0006, teste Qui-quadrado). Outra associação encontrada foi entre sangramento vaginal e internação por problemas de saúde (p-valor<0,0001, teste Qui-quadrado). A naturalidade da mãe foi um fator identificado pelo modelo hierarquizado, não contemplado pelo *stepwise*. Foi encontrada associação significativa entre esta variável e o parto precipitado por problemas de saúde (p-valor=0,0397, teste Qui-quadrado). Pelo procedimento *stepwise* permanecem no modelo as variáveis mais fortemente associadas ao desfecho, conforme descrito na introdução deste trabalho. As tabelas cruzadas para as associações aqui apresentadas podem ser consultadas no anexo III.

As interações entre variáveis não foram estudadas pelo modelo hierarquizado. O estudo das interações pelo processo *stepwise* identificou apenas uma como sendo significativa. Optou-se então por acrescentá-la ao modelo final, dado que sua presença modificou o risco de óbito neonatal para o número de orientações recebidas no pré-natal. Se o parto foi precipitado por problemas de saúde, o baixo número de orientações (0-2) aumenta a chance de óbito (OR=154,18), controlando para as demais variáveis do modelo.

Pesquisando não exaustivamente em literatura da área, utilizando ferramentas de busca pela Internet, não foram encontrados estudos que apresentassem este tipo de interação em modelos de risco para o óbito neonatal.

O método hieraquizado permitiu que variáveis associadas ficassem juntas no mesmo modelo. Este efeito, conhecido como colinearidade, pode gerar estimativas imprecisas para os parâmetros do modelo.

A escolha das variáveis que irão compor a estrutura teórica (marco teórico) é outro ponto que merece especial cuidado na opção por este tipo de modelo. As diferentes dimensões do processo em estudo, bem como a relação entre as variáveis que as compõem devem ser exaustivamente estudadas e bem definidas no início do estudo. Modelos imprecisos podem ser gerados caso variáveis de mesma natureza sejam alocadas em blocos distintos. Por exemplo, na estrutura apresentada por ALMEIDA (2002), a variável total de bens e equipamentos da família foi alocada no nível hierárquico 1 denominado inserção

sócio-econômica, porém não estaria incorreto se fosse alocada no nível 2: condições da família, domicílio e trabalho materno.

KAUFMAN et al. (2004) discutem esta estratégia de ajuste mediada por uma estrutura causal (hierarquizada) salientando que ainda necessita de fundamentos confiáveis. Admitem que existem conjuntos de hipóteses forçadas que permitem que esta estratégia seja válida, mas que é difícil saber quando estas hipóteses são satisfeitas.

FUCHS et al. (1996) apontam como vantagens da análise hierarquizada o estabelecimento prévio dos critérios de seleção e a visualização do desenvolvimento do processo em estudo.

REICHENHEIM e MORAES (1998) analisando a validade de estudos epidemiológicos especificam que se o processo é pouco ou nada conhecido, a obtenção de uma estrutura teórica fica prejudicada e assim o uso de procedimentos estatísticos de seleção de variáveis seriam cabíveis.

Quando se trabalha com grandes conjuntos de variáveis objetivando estudar a probabilidade de ocorrência de um determinado evento utilizando modelos de regressão logística, independente dos procedimentos para seleção de variáveis e modelos existem pontos comuns que devem ser observados: extensa revisão da literatura, análise univariada cuidadosa e avaliação das inter-relações entre as variáveis independentes.

A redução do número de variáveis, a identificação e controle dos confundimentos e a escolha das interações pertinentes à explicação da probabilidade de ocorrência do evento estudado são as maiores preocupações dos pesquisadores.

Segundo REICHENHEIM e MORAES (1998) o pouco cuidado com o processo de redução de variáveis pode fazer com que um "falso" representante do conceito seja erroneamente incorporado ao modelo prejudicando sua validade operacional.

PAULA (2004) enfatiza que a utilização de um processo puramente matemático de seleção pode levar a um modelo sem sentido e de difícil interpretação.

Vale salientar a importância de se deixar claro todo o processo utilizado na apresentação dos resultados quando se emprega a técnica de regressão logística múltipla. A omissão de detalhes da análise impossibilita sua reprodução por outros pesquisadores.

Num levantamento de literatura médica restrita a área de genética ligada à suscetibilidade ao câncer, BAGLEY et al. (2001) identificaram 15 artigos revisados que empregavam a técnica de regressão logística. Desses artigos, a maioria não informava o procedimento para seleção de variáveis e modelo utilizados e nenhum deles indicava a adequação do ajuste. É recomendável que pesquisadores, autores, revisores e editores fiquem mais atentos às normas de uso e apresentação de resultados que envolvem modelos de regressão logística.

6- CONCLUSÕES

Composição de modelos utilizando a regressão logística múltipla envolve diversidade e complexidade de métodos. A realização deste estudo permitiu avaliar diferenças em modelos selecionados por critérios distintos: *stepwise* e modelagem hierarquizada.

Dados de um estudo caso-controle não-pareado visando identificar fatores de risco para o óbito neonatal em Campinas, SP no ano de 2001 foram analisados segundo estes dois critérios. Os resultados apontaram diferenças nas variáveis selecionadas e na inclusão de interação no modelo.

Pela modelagem hierarquizada foram considerados como fatores de risco para o óbito neonatal a renda familiar, o número de pessoas no domicílio, a naturalidade da mãe, o sangramento vaginal, o parto precipitado por problema de saúde, o número de orientações recebidas durante o pré-natal, a escolha do hospital do parto, o tempo decorrido entre a internação e o parto, a idade gestacional, o baixo peso ao nascer e o Apgar de quinto minuto.

Pelo processo de seleção *stepwise* sugerido por HOSMER e LEMESHOW (1989) foram identificados como fatores de risco o número de pessoas no domicílio, o número de orientações recebidas durante o pré-natal, internação por problemas de saúde, parto precipitado, escolha do hospital, tempo decorrido entre a internação e o parto, idade gestacional, baixo peso ao nascer e Apgar no quinto minuto. A interação entre o número de orientações recebidas durante o pré-natal e o parto precipitado foi significativa aumentando a chance de óbito para o baixo número de orientações (0-2) quando controlada para o parto precipitado.

A modelagem hierarquizada permitiu que variáveis associadas ficassem no modelo, renda familiar está associada com a escolha do hospital para o parto e naturalidade da mãe com parto precipitado por problemas de saúde.

A exploração das relações entre as variáveis foi realizada quando se empregou o procedimento *stepwise*.

Os processos de seleção de variáveis implementados em computadores facilitam o emprego da técnica de regressão logística múltipla. As críticas atribuídas à estes processos por gerarem modelos puramente matemáticos devem ser reconsideradas, pois a falha pode estar no cuidado do pesquisador pela investigação das variáveis. Cabe ao pesquisador revisar e avaliar o modelo selecionado antes de apresentá-lo como sendo o melhor.

Independentemente da escolha do processo de seleção de variáveis ou modelos existem pontos que devem ser relevados: revisão exaustiva da literatura sobre o evento em estudo, análise univariada cuidadosa e avaliação das inter-relações entre as variáveis.

7- REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA,S.D.M. **Atenção à saúde da gestante e mortalidade neonatal**. Campinas, 2002 (Tese – Doutorado - Universidade Estadual de Campinas).

ALMEIDA, S.D.M.; BARROS, M.B.A. Atenção à saúde e mortalidade neonatal: estudo caso-controle realizado em Campinas, SP. **Rev. Bras. Epidemiol**., 7(1):22-35, 2004.

BAGLEY,S.C.; WHITE,H.; GOLOMB,B.A. Logistic regression in the medical literature: standards for use an reporting, with particular attention to one medical domain. **Journal of Clinical Epidemiology**, 54:979-985 (2001).

FUCHS, S.C. **Fatores de risco para diarréia complicada por desidratação moderada a grave: um estudo de casos e controles.** Porto alegre, 1993 (Tese – Doutorado – Faculdade de Medicina da UFRGS).

FUCHS,S.; VICTORA,C.G.; FACHEL,J. Modelo hierarquizado: uma proposta de modelagem aplicada à investigação de fatores de risco para diarréia grave**. Rev.Saúde Pública,** 30(2):168-78, 1996.

GREENLAND,S. Modeling and variable selection in epidemiologic analysis. **Am.J.Public Health**, 79:340-9, 1989.

GREENLAND,S. Hierarchical regression for epidemiologic analyses of multiple exposures. **Environmental Health Perspect**. 102 Supl 8:33-9, 1994.

HENNEKENS, C.H.; BURING, J.E. **Epidemiology in Medicine**. Boston: Little, Brown and Company, 1987.p.287-323.

HOSMER, D.W.; LEMESHOW,S. **Applied Logistic Regression.** Nova Iorque: John Wiley & Sons, 1989. 307p.

KAUFMAN, J.S.; MACLEHOSE, R.F.; KAUFMAN,S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. **Epidemiologic Perspectives & Innovations**. 1:4, 2004.

KLEINBAUM, D.G. **Logistic Regression: A self-Learning Text**. Nova Iorque: Springer-Verlag, 1994. 282p.

MALDONADO,G.; GREENLAND,S. Simulation study of confounder-selection strategies. **Am.J.Epidemiol.**, 138:923-26,1993.

MOSLEY,W.H.; CHEN,L.C. An analytical framework for the study of child survival in developing countries. **Population and Development Review**, 10:25-45, 1984.

PAULA, G. **Modelos de Regressão com Apoio Computacional.** São Paulo: Editora da Universidade de São Paulo, 2004. p.84-152.

PEREIRA, M.G. **Epidemiologia, Teoria e Prática.** Rio de Janeiro: Editora Guanabara Koogan, 1995. p.377-97.

REICHENHEIM, M.E.; MORAES, C.L. Pillars for assessing validity in epidemiological studies. ***Rev. bras. epidemiol***. [online], 1(2):131-148, 1998. [cited 24 January 2006]. Disponível em: <http://www.scielo.br/scielo.php?script=

sci_arttext&pid=S1415-790X1998000200004&lng=en&nrm=iso>. Acesso em: 24 jan 2006.

SAS INSTITUTE INC. **SAS System for Windows**, **versão 8.2.** Cary, North Carolina, USA, 1999-2001.

TABACHNICK, B.G.; FIDELL,L.S. **Using Multivariate Statistics**. 4ª ed. Needham Heights: Allyn&Bacon, 2001.p.517-82.

TRUETT,J.; CORNFIELD,J.; KANNEL,W. A multivariate analysis of the risk of coronary heart disease in Framingham. **Journal of Chronic Diseases**, 20**:**511-24, 1967.

VICTORA, C.G.; HUTTLY,S.R.; FUCHS,S.; OLINTO,M.T.A. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. **International Journal of Epidemiology**, 26(1):224-227, 1997.

WITTE,J.S.; GREENLAND,S. Simulation study of hierarchical regression. **Statistics in Medicine**, 15:1161-1170, 1996.

8- ANEXOS

ANEXO I

PROGRAMA UTILIZADO PARA AS ANÁLISES NO SAS

```
libname trab 'a:';
options ps=1000 ls=80 nodate nonumber nolabel;
proc format;
value gr            1='caso    '
                    2='controle';
value san           1='adeq'
                    2='inad';
value oc            1='emp.d'
                    2='outra'
                    9='n tra';
value ed 1='grd'
                    2='mod'
                    3='lev';
value fu 1='não'
                    2='ex'
                    3='fum';
value sn            1='sim'
                    2='não';
value sx            1='masc'
                    2='fem';
value ge            1='un'
                    2='ge';
run;
options             ls=120;

**********análise descritiva**********;

proc means data=trab.estati nonobs n mean std min median max maxdec=1;
var refam ;
class casocontro;
format casocontro gr.;
run;
options ls=80;

proc freq data=trab.estati formchar(1,2,7)='|-+';
tables (tbens_a salariof estud_a ncomod domic imov tempod saneam npessoa
        ocupa mesest esforcof desgaste raca_a nat_a habitodefu filhosv
        idmaead doencadura tratoupres tratouinfe apresentou foiinterna
        nori nexrot nproc visitpn convpre escolheuom maiorparte
        encontroud fezecograf tipopart sinal_a partopreci convpart
        hospitalfo te medpart estn ipodegest sexo dgest_a bxpeso
        apg5)*casocontro / nopercent norow;
format casocontro gr. saneam san. ocupa oc. habitodefu fu. doencadura
        tratoupres tratouinfe apresentou foiinterna escolheuom maiorparte
        encontroud fezecograf partopreci hospitalfo sn. tipodegest ge.
        sexo sx. bebenasceu sn.;
run;
```

**********regressão logística univariada**********;

```
proc logistic data=trab.estati ;
class tbens_a (ref='>=8')/param=ref;
model casocontro=tbens_a / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
class salariof (ref='>2 SAL')/param=ref;
model casocontro=salariof / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
model casocontro=refam / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
class estud_a (ref='>=9')/param=ref;
model casocontro=estud_a / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
class ncomod (ref='>=3')/param=ref;
model casocontro=ncomod / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
class domic (ref='alv c')/param=ref;
model casocontro=domic / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
class imov (ref='pp/al/ce')/param=ref;
model casocontro=imov / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
class tempod (ref='>10')/param=ref;
model casocontro=tempod / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
class saneam (ref='adeq')/param=ref;
model casocontro=saneam / scale=none aggregate risklimits ;
format casocontro gr. saneam san.;
run;
proc logistic data=trab.estati ;
class npessoa (ref='>=4')/param=ref;
model casocontro=npessoa / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
class ocupa (ref='n tra')/param=ref;
model casocontro=ocupa / scale=none aggregate risklimits ;
format casocontro gr. ocupa oc.;
run;
```

```
proc logistic data=trab.estati ;
class mesest (ref='0')/param=ref;
model casocontro=mesest / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class esforcof (ref='nt')/param=ref;
model casocontro=esforcof / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class desgaste (ref='nt')/param=ref;
model casocontro=desgaste / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class raca_a (ref='bca')/param=ref;
model casocontro=raca_a / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class nat_a (ref='c/r')/param=ref;
model casocontro=nat_a / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class habitodefu (ref='não')/param=ref;
model casocontro=habitodefu / scale=none aggregate risklimits ;
format casocontro gr. habitodefu fu.;
run;
proc logistic data=trab.estati ;
class filhosv (ref='>=1')/param=ref;
model casocontro=filhosv / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class idmaead (ref='>=20')/param=ref;
model casocontro=idmaead / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class doencadura (ref='não')/param=ref;
model casocontro=doencadura / scale=none aggregate risklimits ;
format casocontro gr. doencadura sn.;
run;
proc logistic data=trab.estati ;
class tratoupres (ref='não')/param=ref;
model casocontro=tratoupres / scale=none aggregate risklimits ;
format casocontro gr. tratoupres sn.;
run;
proc logistic data=trab.estati ;
class tratouinfe (ref='não')/param=ref;
model casocontro=tratouinfe / scale=none aggregate risklimits ;
format casocontro gr. tratouinfe sn.;
run;
```

```
proc logistic data=trab.estati ;
class apresentou (ref='não')/param=ref;
model casocontro=apresentou / scale=none aggregate risklimits ;
format casocontro gr. apresentou sn.;
run;
proc logistic data=trab.estati ;
class foiinterna (ref='não')/param=ref;
model casocontro=foiinterna / scale=none aggregate risklimits ;
format casocontro gr. foiinterna sn.;
run;
proc logistic data=trab.estati ;
class nori (ref='6-7')/param=ref;
model casocontro=nori / scale=none aggregate risklimits ;
format casocontro gr.;
run;
proc logistic data=trab.estati ;
class nexrot (ref='>=8')/param=ref;
model casocontro=nexrot / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class nproc (ref='>=5')/param=ref;
model casocontro=nproc / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class visitpn (ref='>=8')/param=ref;
model casocontro=visitpn / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class convpre (ref='NSUS')/param=ref;
model casocontro=convpre / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class escolheuom (ref='sim')/param=ref;
model casocontro=escolheuom / scale=none aggregate risklimits ;
format casocontro gr. escolheuom sn.;
run;
proc logistic data=trab.estati ;
class maiorparte (ref='sim')/param=ref;
model casocontro=maiorparte / scale=none aggregate risklimits ;
format casocontro gr. maiorparte sn.;
run;
proc logistic data=trab.estati ;
class encontroud (ref='não')/param=ref;
model casocontro=encontroud / scale=none aggregate risklimits ;
format casocontro gr. encontroud sn.;
run;
proc logistic data=trab.estati ;
class fezecograf (ref='sim')/param=ref;
model casocontro=fezecograf / scale=none aggregate risklimits ;
format casocontro gr. fezecograf sn.;
run;
```

```
proc logistic data=trab.estati ;
class tipopart (ref='VAGINAL')/param=ref;
model casocontro=tipopart / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class sinal_a (ref='usuais')/param=ref;
model casocontro=sinal_a / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class partopreci (ref='não')/param=ref;
model casocontro=partopreci / scale=none aggregate risklimits ;
format casocontro gr. partopreci sn.;
run;
proc logistic data=trab.estati ;
class convpart (ref='NSUS')/param=ref;
model casocontro=convpart / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class hospitalfo (ref='sim')/param=ref;
model casocontro=hospitalfo / scale=none aggregate risklimits ;
format casocontro gr. hospitalfo sn.;
run;
proc logistic data=trab.estati ;
class te (ref='1-10h') / param=ref;
model casocontro=te / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class medpart (ref='MEDPREN')/param=ref;
model casocontro=medpart / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class estn (ref='out')/param=ref;
model casocontro=estn / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class tipodegest (ref='un')/param=ref;
model casocontro=tipodegest / scale=none aggregate risklimits ;
format casocontro gr. tipodegest ge.;
run;
proc logistic data=trab.estati ;
class sexo (ref='fem')/param=ref;
model casocontro=sexo / scale=none aggregate risklimits ;
format casocontro gr. sexo sx.;
run;
proc logistic data=trab.estati ;
class dgest_a (ref='>=37s')/param=ref;
model casocontro=dgest_a / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
```

```
proc logistic data=trab.estati ;
class bxpeso (ref='>=2500')/param=ref;
model casocontro=bxpeso / scale=none aggregate risklimits ;
format casocontro gr. ;
run;
proc logistic data=trab.estati ;
class apg5 (ref='9e10')/param=ref;
model casocontro=apg5 / scale=none aggregate risklimits ;
format casocontro gr.;
run;

**********regressão logística múltipla - efeitos principais*****;

proc logistic data=trab.estati ;
class tbens_a (ref='>=8') salariof (ref='>2 SAL') estud_a (ref='>=9')
      ncomod (ref='>=3') domic (ref='alv c') imov (ref='pp/al/ce')
      tempod (ref='>10') saneam (ref='adeq') npessoa (ref='>=4')
      ocupa (ref='n tra') mesest (ref='0') esforcof (ref='nt')
      desgaste (ref='nt') raca_a (ref='bca') nat_a (ref='c/r')
      habitodefu (ref='não') filhosv (ref='>=1') idmaead (ref='>=20')
      doencadura (ref='não') tratoupres (ref='não') tratouinfe(ref='não')
      apresentou (ref='não') foiinterna (ref='não') nori (ref='6-7')
      nexrot (ref='>=8') nproc (ref='>=5') visitpn (ref='>=8')
      convpre (ref='NSUS') escolheuom (ref='sim') maiorparte (ref='sim')
      encontroud (ref='não') fezecograf (ref='sim') tipopart
      (ref='VAGINAL') sinal_a (ref='usuais') partopreci (ref='não')
      convpart (ref='NSUS') hospitalfo (ref='sim') te (ref='1-10h')
      medpart (ref='MEDPREN') estn (ref='out') tipodegest (ref='un')
      sexo (ref='fem') dgest_a (ref='>=37s') bxpeso (ref='>=2500')
      apg5 (ref='9e10')  / param=ref;
model casocontro=tbens_a salariof refam estud_a ncomod domic imov tempod
            saneam npessoa ocupa mesest esforcof desgaste raca_a
            nat_a habitodefu filhosv idmaead doencadura tratoupres
            tratouinfe apresentou foiinterna nori nexrot nproc
            visitpn convpre escolheuom maiorparte encontroud
            fezecograf tipopart sinal_a partopreci convpart
            hospitalfo te medpart estn tipodegest sexo dgest_a
            bxpeso apg5 / scale=none aggregate risklimits
                        selection=s /*details lackfit influence*/
                        rsquare;
format casocontro gr. saneam san. ocupa oc. habitodefu fu. doencadura sn.
        tratoupres tratouinfe apresentou foiinterna escolheuom maiorparte
        encontroud fezecograf partopreci hospitalfo sn. tipodegest ge.
        sexo sx. ;
run;

****verificando associações entre variáveis*****;

proc freq data=trab.estati;
tables npessoa*(tbens_a salariof refam estud_a ncomod domic
                imov tempod saneam npessoa)/chisq;
run;
proc means data=trab.estati;
class npessoa;
var refam;
run;
```

```
proc npar1way wilcoxon data=trab.estati;
class npessoa;
var refam;
run;

proc freq data=trab.estati;
tables idmaead*(casocontro tratoupres raca_a habitodefu bxpeso dgest_a
                habitodefu visitpn)/chisq;
run;
proc freq data=trab.estati;
tables salariof*npessoa / chisq;
run;
proc freq data=trab.estati;
tables salariof*hospitalfo / chisq;
run;
proc freq data=trab.estati;
tables dgest_a*casocontro*(nori bxpeso apg5)/*visitpn*//chisq;
run;
proc freq data=trab.estati;
tables partopreci*te/chisq;
run;
proc freq data=trab.estati;
tables partopreci*apresentou apresentou*foiinterna nat_a*foiinterna/chisq;
run;
proc freq data=trab.estati;
tables nat_a*te nat_a*salariof/chisq;
run;
proc freq data=trab.estati;
tables nat_a*te /*hospitalfo*casocontro*//chisq;
run;
proc freq data=trab.estati formchar(1,2,7)='|-';
tables npessoa*(FOIINTERNA nori PARTOPRECI HOSPITALFO te dgest_a BXPESO apg5)/chisq;
run;

*********testando interações ************+;

proc logistic data=trab.estati ;
class npessoa (ref='>=4') foiinterna (ref='não') nori (ref='6-7')
      partopreci (ref='não')hospitalfo (ref='sim') te (ref='1-10h')
      dgest_a (ref='>=37s') bxpeso (ref='>=2500') apg5 (ref='9e10')
      /param=ref;
model casocontro= npessoa foiinterna nori partopreci hospitalfo te
                  dgest_a bxpeso apg5 npessoa*foiinterna /*npessoa*nori
                  npessoa*partopreci npessoa*hospitalfo npessoa*te
                  npessoa*dgest_a npessoa*bxpeso npessoa*apg5
                  foiinterna*nori foiinterna*partopreci
                  foiinterna*hospitalfo foiinterna*te foiinterna*dgest_a
                  foiinterna*bxpeso foiinterna*apg5 nori*partopreci
                  nori*hospitalfo nori*te nori*dgest_a nori*bxpeso
                  nori*apg5 partopreci*hospitalfo partopreci*te
                  partopreci*dgest_a partopreci*bxpeso partopreci*apg5
                  hospitalfo*te hospitalfo*dgest_a hospitalfo*bxpeso
                  hospitalfo*apg5 te*dgest_a te*bxpeso te*apg5
                  dgest_a*bxpeso dgest_a*apg5 bxpeso*apg5 nori*partopreci
    / scale=none aggregate risklimits ;
```

```
format casocontro gr. saneam san. ocupa oc. habitodefu fu. doencadura sn.
          tratoupres tratouinfe apresentou foiinterna escolheuom maiorparte
          encontroud fezecograf partopreci hospitalfo sn. tipodegest ge.
          sexo sx. ;
run;



**********modelo final**********;

proc logistic data=trab.estati ;
class npessoa (ref='>=4') foiinterna (ref='não') nori (ref='6-7')
        partopreci (ref='não') hospitalfo (ref='sim') te (ref='1-10h')
        dgest_a (ref='>=37s') bxpeso (ref='>=2500') apg5 (ref='9e10')
        /param=ref;
model casocontro= npessoa foiinterna nori partopreci hospitalfo te
              dgest_a bxpeso apg5  nori*partopreci
     / scale=none aggregate risklimits details lackfit influence
       rsquare;
format casocontro gr. foiinterna partopreci hospitalfo sn. ;
run;
```

ANEXO II

Quadro 2- Valores de p para as possíveis interações entre variáveis a serem incluídas no modelo com os efeitos principais.

Interação	p-valor	Interação	p-valor
npessoa*foiinterna	0,4948	Nori*dgest_a	0,4420
npessoa*nori	0,1053	Nori*bxpeso	0,5539
Npessoa*partopreci	0,2153	Nori*apg5	0,2930
Npessoa*hospitalfo	0,7182	Partopreci*hospitalfo	0,1764
Npessoa*te	0,9188	Partopreci*te	0,2161
Npessoa*dgest_a	0,4407	Partopreci*dgest_a	0,2821
Npessoa*bxpeso	0,2230	Partopreci*bxpeso	0,8858
Npessoa*apg5	0,1697	Partopreci*apg5	0,9312
Foiinterna*nori	0,2553	Hospitalfo*te	0,6412
Foiinterna*partopreci	0,5737	Hospitalfo*dgest_a	0,8448
Foiinterna*hospitalfo	0,6650	Hospitalfo*bxpeso	0,5198
Foiinterna*te	0,4605	Hospitalfo*apg5	0,1675
Foiinterna*dgest_a	0,1407	Te*dgest_a	0,1117
Foiinterna*bxpeso	0,6363	Te*bxpeso	0,2189
Foiintena*apg5	0,1648	Te*apg5	0,5890
Nori*partopreci	0,0070	Dgest_a*bxpeso	0,9217
Nori*hospitalfo	0,9077	Dgest_a*apg5	0,8669
Nori*te	0,8788	Bxpeso*apg5	0,9251

ANEXO III

Quadro 3- Estudo das associações entre as variáveis independentes.

Tabela 9- Distribuição da renda familiar em faixas de salários mínimos e escolha do hospital para o parto. Campinas, SP, 2001.

SALARIOF	HOSPITALFO		Total
	Sim n(%)	Não n (%)	
0-2 sal	46 (65,7)	24 (34,3)	70 (100,0)
>2 sal	236 (84,0)	45 (16,0)	281 (100,0)
Total	282 (80,3)	69 (19,7)	351 (100,0)

p-valor=0,0006 (Qui-quadrado)

Tabela 10- Distribuição do sangramento vaginal e da internação por problema de saúde. Campinas, SP, 2001.

APRESENTOU	FOIINTERNA		Total
	Sim n(%)	Não n (%)	
sim	18 (40,0)	27 (60,0)	45 (100,0)
não	47 (15,4)	259 (84,6)	306 (100,0)
Total	65 (18,5)	286 (81,5)	351 (100,0)

p-valor<0,0001 (Qui-quadrado)

Tabela 11- Distribuição da naturalidade materna e do parto precipitado por problema de saúde. Campinas, SP, 2001.

NAT_A	PARTOPRECI		Total
	Sim n(%)	Não n (%)	
Campinas/região	59 (33,9)	115 (66,1)	174 (100,0)
outra	79 (44,6)	98 (55,4)	177 (100,0)
Total	138 (39,3)	213 (60,7)	351 (100,0)

p-valor=0,0397 (Qui-quadrado)