

COMPARAÇÃO DE DESEMPENHO DE TÉCNICAS
PARA DETECÇÃO DE BAIXO PESO EM RÉCEM-
NASCIDOS USANDO REGRESSÃO LINEAR MUL-
TIPLA E LOGÍSTICA.

LUIS APARECIDO MILAN



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO

89C

34/BC

CAMPINAS - SÃO PAULO
BRASIL

Impl. 8.10.92

TÍTULO DA TESE: COMPARAÇÃO DE DESEMPENHO DE TÉCNICAS PARA DETECÇÃO DE BAIXO PESO EM RECÉM-NASCIDOS USANDO REGRESSÃO LINEAR MÚLTIPLA E LOGÍSTICA.

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Luís Aparecido Milan e aprovada pela Comissão Julgadora.

Campinas, 21 de maio 1987


Prof. Dr. J. Norberto W. Dachs *ten*
(Orientador)

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do título de mestre em estatística.

INDICE

I - INTRODUÇÃO	1
II - DESCRIÇÃO DO PROBLEMA	4
III - INFORMAÇÕES DISPONÍVEIS	10
III.1 - Introdução	10
III.2 - As Variáveis	11
III.3 - Recodificação das Variáveis	18
IV - METODOLOGIA	22
IV.1 - Introdução	22
IV.2 - Regressão Linear	23
IV.3 - Seleção de Variáveis para o Modelo Linear	30
IV.4 - Modelo Logístico	32
V - REGRESSÃO LINEAR E VARIANTES	37
V.1 - Modelo Provisório	37
V.2 - Detecção de "Outliers"	38
V.3 - Modelo 0	41
V.4 - Sensibilidade e Especificidade	43
V.5 - Modelo 1	46
V.6 - Comparação de Modelos	48
V.7 - Modelo 2	51
V.8 - Modelo 3	54
V.9 - Modelos 4, 5 e 6	57
V.10 - Modelos 7, 8, 9, 10 e 11	66
VI - REGRESSÃO LOGÍSTICA	81
VI.1 - Introdução	81
VI.2 - Modelos 12, 13, 14 e 15	82
VII - COMPARAÇÃO GERAL DOS MODELOS	94
VIII - DISCUSSÃO E CONCLUSÕES	101
IX - ANEXO 1 - FICHA OBSTÉTRICA	105
X - REFERÊNCIAS	110

1 - INTRODUÇÃO

Grande parte do desenvolvimento de novos métodos em Estatística, ocorrido nos últimos vinte anos, foi motivado pela busca de soluções para problemas práticos ocorridos nas áreas de ciências biológicas em geral e médicas em particular. Boa parte desse desenvolvimento ocorreu no estudo de problemas de doenças degenerativas como câncer e enfermidades cardiovasculares. No caso específico de problemas mais frequentes nos países do terceiro mundo nem mesmo as metodologias existentes tem sido provadas com muita frequência.

Neste trabalho é feito um estudo de um dos problemas sérios de saúde, que é mais grave em regiões e países de menor desenvolvimento, especificamente, o de tentar prever peso de recém-nascidos.

O baixo peso ao nascer de uma criança se constitui num dos problemas mais sérios de saúde porque é diretamente responsável por altas taxas de mortes e também por consequências invalidantes para os que não morrem. Por outro lado as estatísticas mostram que quando este peso é adequado e se chegou ao final da gestação são poucas as chances de acontecer um desenlace negativo para a mesma.

A possibilidade de se prever antes do nascimento o peso deste futuro recém-nascido e consequentemente saber se o mesmo estaria dentro dos níveis considerados adequados, pode ser de muita utilidade ao pessoal da área de saúde para que possam ser tomadas as providências necessárias com a devida antecedência. Estas providências podem variar desde mudanças na atenção até o traslado da gestante a regiões ou estabelecimentos com tecnologias adequadas para diminuir o risco da criança quando o peso for insuficiente.

Neste trabalho serão desenvolvidos diversos modelos de ajuste para PESO, usando variáveis da gravidez atual e sobre antecedentes da mãe. O objetivo principal, do ponto de vista da Estatística é o de confrontar várias maneiras diferentes de desenvolver essa modelagem, numa tentativa de estabelecer se existem vantagens marcantes de uma ou mais dessas técnicas sobre as demais.

Os resultados dos modelos aqui apresentados, como será discutido mais adiante não deverão ser utilizados para fins preditivos em nenhuma população específica, pois os dados são de partos ocorridos em um hospital de referência e não de uma amostra de uma população.

Existem várias metodologias que podem ser empregadas para obtenção de modelos de ajuste para uma variável como PESO. Neste caso, como será argumentado no Capítulo II, na maioria das vezes,

não é importante prever ou ajustar o próprio valor de peso do recém-nascido, mas sim prever se esse peso estará abaixo de um determinado limiar mínimo. Por essa razão, além das técnicas de regressão linear múltipla, pode-se usar também as de regressão logística, e outras. Entre essas últimas, uma que parece promissora em outras áreas, e que poderá vir a ser investigada no contexto do problema abordado neste trabalho, é a de análise de classificação e árvores de regressão, como proposta por Breiman et al. (1984). Aqui serão usadas técnicas de regressão linear múltipla, com várias alterações e modificações e também regressão logística. O objetivo central não é obter modelos visando interpretar seus coeficientes e o possível relacionamento das variáveis preditoras com a resposta (PESO). O que se pretende é averiguar a adequação dos ajustes em termos de capacidade preditiva, ou seja, considerando uma resposta binária (do tipo baixo peso - peso normal), pesquisar se existem modelos e ajustes que fornecem melhores valores de sensibilidade e especificidade. Essas duas medidas são bastante adequadas neste caso para medir o desempenho de métodos de classificação pois ambas tem interpretações simples e úteis do ponto de vista médico e de saúde.

Considerando que a resposta de interesse é binária, se obtém a Tabela 1.1.

		Observado	
		Baixo Peso	Peso Normal
Previsão	Baixo Peso	a	b
	Peso Normal	c	d

Tabela 1.1 - Tabela de classificação 2 X 2 para baixo peso e peso normal.

Nessa tabela a sensibilidade é dada por $a / (a+c)$ e a especificidade por $d / (b+d)$. Ou seja, a sensibilidade mede a capacidade do processo preditivo de classificar corretamente os recém-nascidos de baixo peso, isto é, qual a fração dos de baixo pe-

so que o processo classifica como tal. A especificidade, por outro lado, mede qual a fração dos de peso normal que o método de classificação classifica como normais.

Em qualquer dos métodos de predição que se use ao tentar aumentar o valor da sensibilidade perde-se necessariamente em especificidade e vice-versa. O ponto "ideal" de corte na classificação depende de muitos fatores, incluindo recursos disponíveis, gravidade do problema tratado, incidência do mesmo na população em que se for usar o processo de classificação, e outros. Por esta razão é conveniente dispor de uma maneira de fazer comparações globais de duas metodologias classificatórias quanto a sua sensibilidade e especificidade. Uma maneira de se fazer comparações globais é usada ao longo dos Capítulos V e VI, através de gráficos dos logitos da especificidade versus logitos da sensibilidade para diferentes "pontos de corte" no ajuste.

O esquema geral do trabalho é o seguinte:

- 1 - É feita uma descrição sumária sobre a importância do problema a ser abordado, ou seja o de previsão de baixo peso ao nascer, seguida de uma apresentação dos dados disponíveis para as análises, as variáveis, suas codificações e as transformações feitas. (Capítulos II e III).
- 2 - No Capítulo IV são apresentados resumos sobre as metodologias de regressão linear múltipla e regressão logística. Servem apenas para situar o leitor e não devem ser tomados como fontes de referência sobre os dois assuntos, para o que são citados alguns textos, como referências, sem nenhuma tentativa de se fazer uma resenha bibliográfica sobre os dois temas.
- 3 - O Capítulo V apresenta os resultados obtidos com diversas variantes de uso de regressão linear múltipla.
- 4 - O Capítulo VI apresenta resultados obtidos com alguns ajustes usando regressão logística.
- 5 - No Capítulo VII é apresentado um resumo de todos os resultados e ajustes obtidos no processo descrito em detalhe nos Capítulos IV e V.
- 6 - O Capítulo VIII traz as conclusões com algumas observações de natureza metodológica, e não sobre aspectos de interpretação biológica dos ajustes obtidos.

Os dados utilizados na construção de todos os gráficos e na obtenção dos resultados numéricos apresentados neste trabalho foram obtidos do banco de dados existente no computador do Centro de Apoio Integral à Saúde da Mulher (CAISM) formado a partir de partos realizados na Maternidade da Faculdade de Ciências Médicas (FCM) da UNICAMP e aqui utilizados com a autorização dos docentes responsáveis pelos mesmos.

II - DESCRIÇÃO DO PROBLEMA

Já é de conhecimento médico há vários anos que, à medida que aumenta o peso de recém-nascidos, até um certo limite, diminuem o risco de morte dos mesmos, diminuindo também a probabilidade de contrair doenças e de ocorrerem outras complicações. A partir deste limite "ideal" (na verdade uma faixa) de peso, à medida que o peso aumenta, novamente voltam a aumentar a mortalidade, as doenças e as complicações. Este fato está representado de forma esquemática, para o coeficiente de mortalidade perinatal, na Figura 2.1.

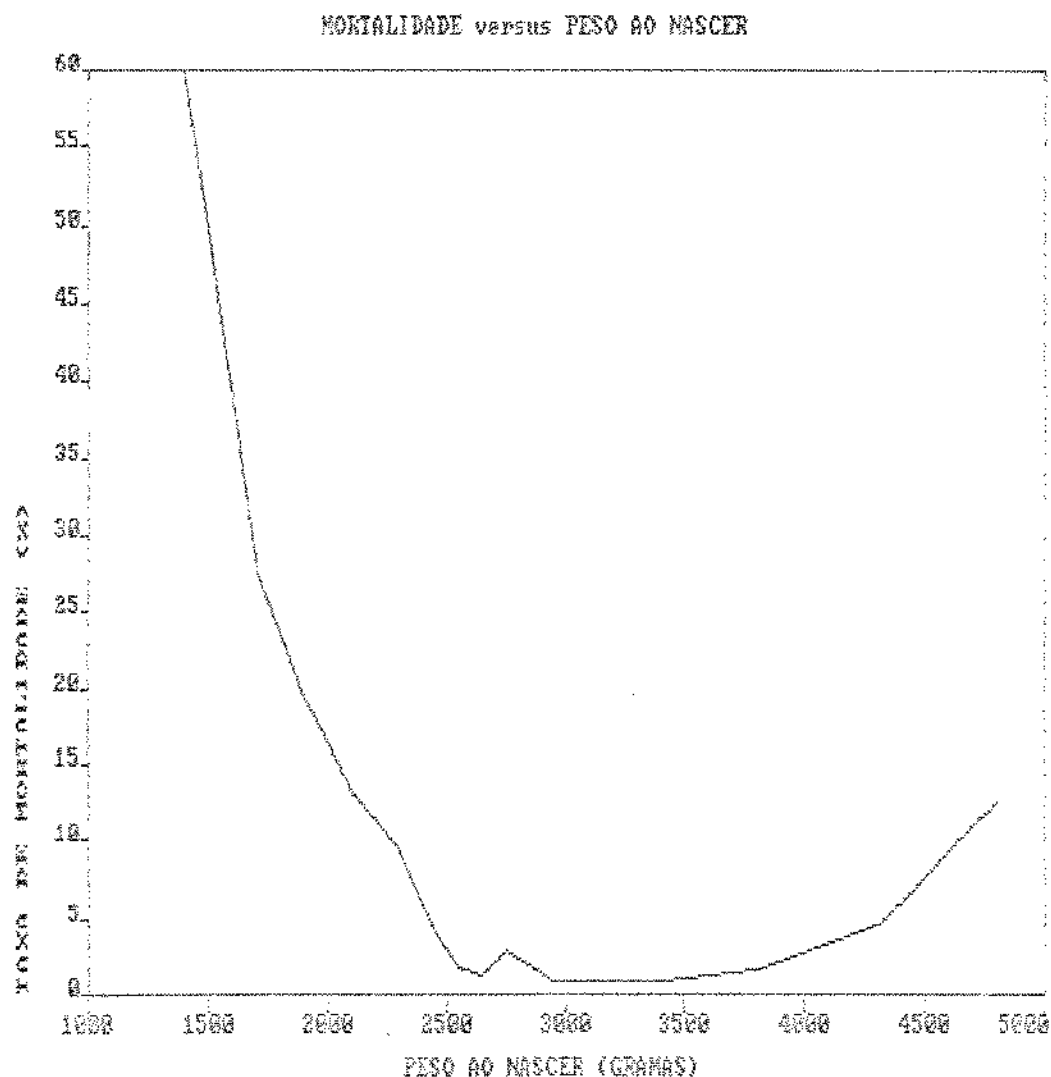


Figura 2.1 - Variação do coeficiente de mortalidade perinatal com o peso de recém-nascidos. (Fonte: FCM UNICAMP)

Sabe-se também que este limite de peso "ideal" do recém-nascido varia de região para região, e até dentro de uma região específica há diferenças entre diferentes grupos populacionais determinados. De um modo geral podemos dizer que este peso ideal está entre 2500 e 3500 gramas.

A partir dos trabalhos de Yerushalmy et al (1965) e Battaglia e Lubchenco (1967) ficou demonstrada a importância de outra variável, também relacionada com a mortalidade, que é o tempo de gestação. O período ou tempo de gestação que aqui denominaremos idade gestacional corresponde ao número de semanas que dura uma gestação, contadas a partir de 10 dias após a última menstruação até o momento do parto.

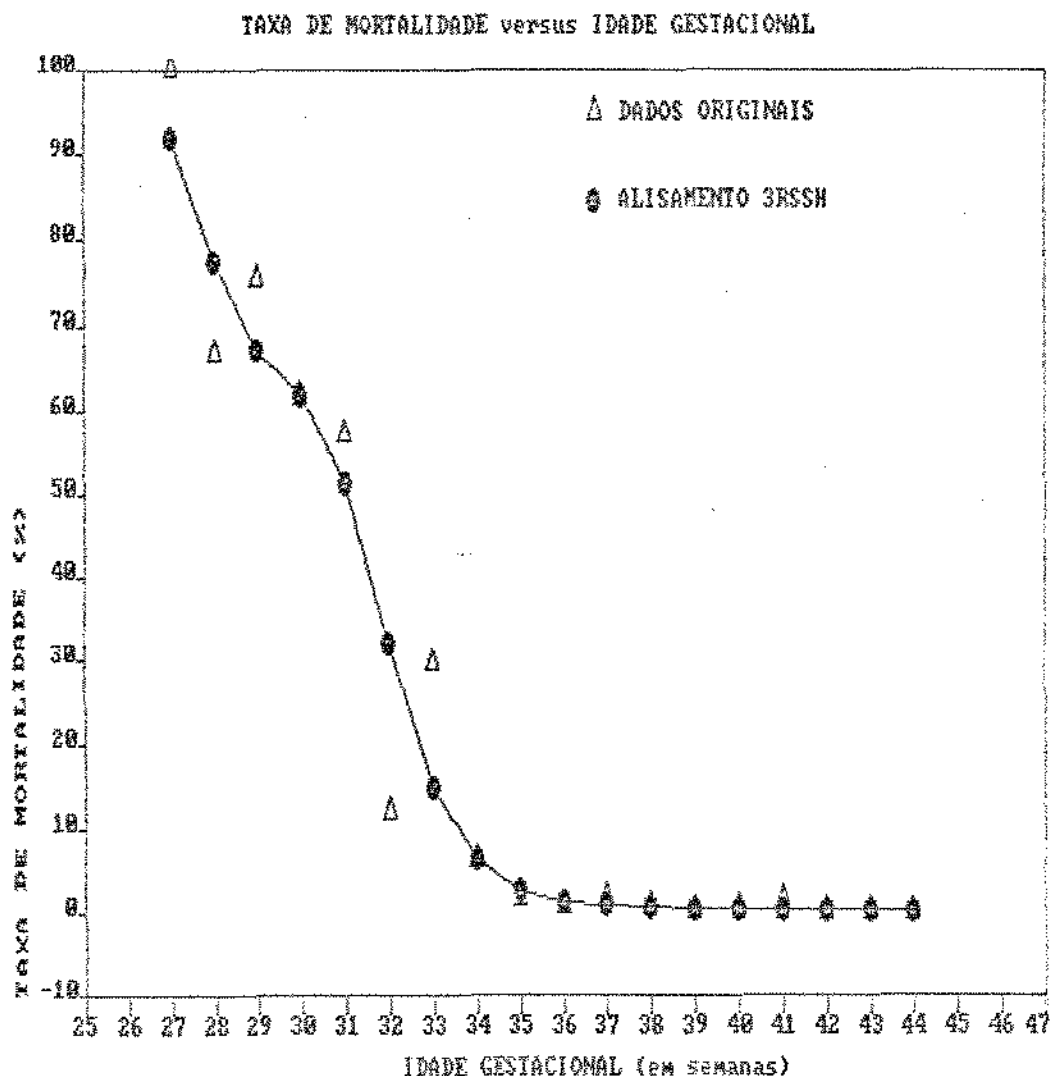


Figura 2.2 - Variação do coeficiente de mortalidade perinatal com a idade gestacional. (Fonte: FCM UNICAMP)

Da literatura disponível temos que a idade gestacional "ideal" está em torno de 37 a 42 semanas, Yerushalmy et al (1965). Isto pode ser constatado na figura 2.2, principalmente para valores menores que 37 semanas de gestação onde taxa de mortalidade aumenta à medida que a idade gestacional diminui. Para os casos com mais que 42 semanas de gestação temos poucas amostras, onze ao todo, e nenhuma comparação poderia ser feita.

A figura 2.2 mostra dois tipos de informação: A primeira representada por triângulos, apresentando os dados como foram obtidos originalmente, sem nenhum tratamento. A segunda representada por figuras arredondadas que apresentam o resultado obtido da aplicação de um alisamento 3RSSH. Trata-se de uma técnica exploratória destinada a evidenciar a tendência em um conjunto de dados em que esta não se manifesta de forma evidente. Para maiores detalhes sobre esta técnica exploratória ver Tukey (1977).

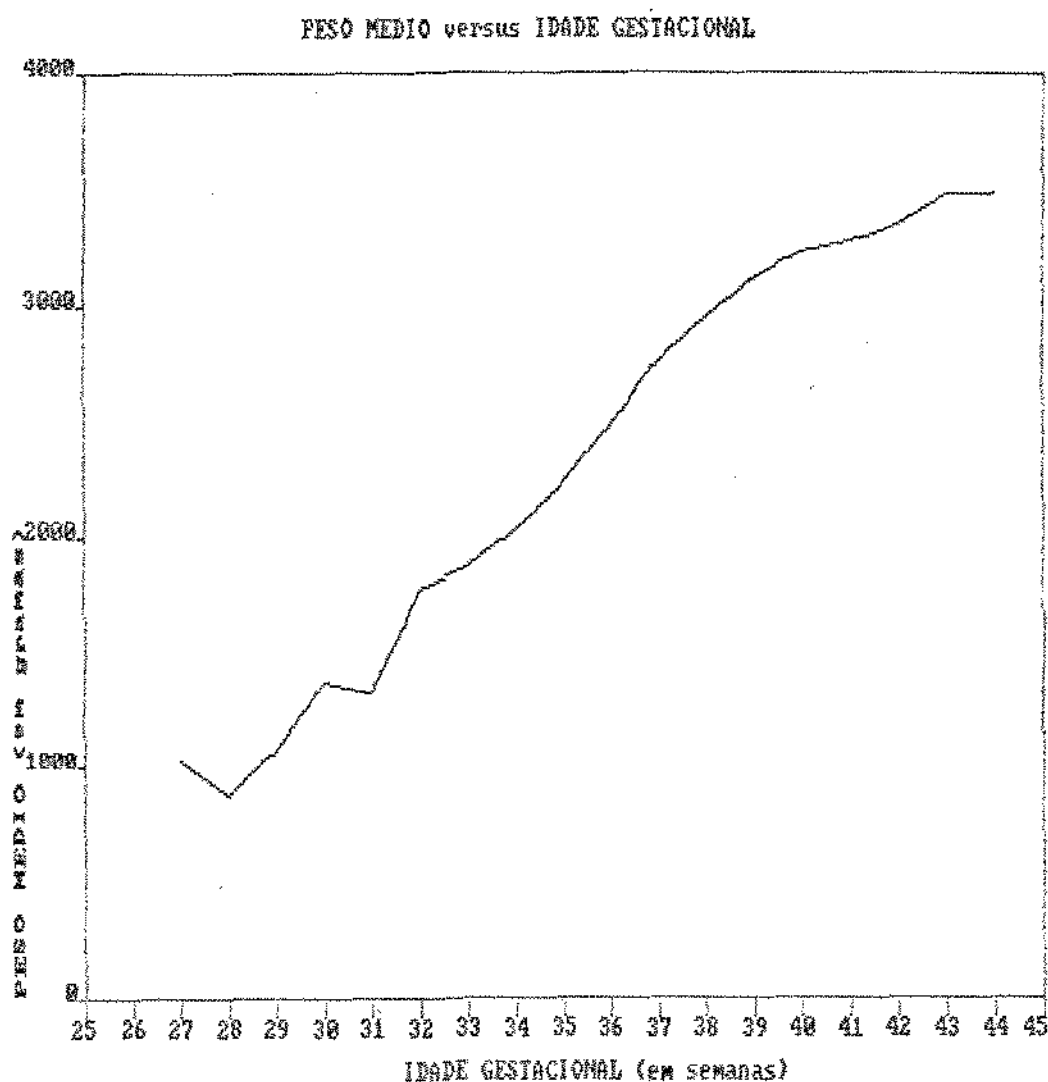


Figura 2.3 - Variação do peso de recém-nascidos com a idade gestacional. (Fonte: FCM UNICAMP)

A idade gestacional tem influência direta sobre o peso do recém-nascido, como se vê na Figura 2.3. À medida que aumenta a idade gestacional aumenta também o peso do recém-nascido, como pode ser observado na figura 2.3, mas isto não ocorre de maneira linear, sendo que o ganho de peso tem velocidades diferentes em etapas diferentes da gestação. A primeira vista podemos identificar três tipos diferentes de velocidade em três etapas diferentes da gestação: A primeira etapa compreende os três primeiros meses da gestação quando o ganho de peso do feto ocorre com uma velocidade pequena. A segunda etapa acontece aproximadamente entre o terceiro e o sexto mês de gestação quando o ganho de peso pelo feto se dá de forma bastante rápida. A terceira etapa ocorre nos três últimos meses da gestação quando a velocidade diminui novamente, principalmente nas últimas semanas, da 36ª a 42ª semana de gestação.

O conhecimento das relações entre peso do recém-nascido, idade gestacional e mortalidade, permitiram uma classificação das mortes perinatais de forma a considerar a influência simultânea de peso e idade gestacional na taxa de mortalidade perinatal, levando Battaglia e Lubchenco a formularem nove combinações diferentes de idade gestacional e peso do recém-nascido ao nascer.

A idade gestacional foi classificada em três fases:

- 1) Pré-termo: para nascimentos ocorridos antes das 37 semanas.
- 2) Termo: para nascimentos ocorridos entre 37 e 42 semanas.
- 3) Pós-termo: para nascimentos ocorridos após as 42 semanas.

O peso foi classificado em três grupos também:

- 1) Menor que 2500 gramas.
- 2) Entre 2500 e 3500 gramas.
- 3) Maior que 3500 gramas.

Esta classificação conduziu a elaboração da tabela cruzada mostrada na Tabela 2.1.

Os resultados apresentados por Battaglia e Lubchenco (1967) foram confirmados por outros autores, entre eles pode-se citar Jurado Garcia et al (1970), e indicam que, quando o peso do recém-nascido é adequado à idade gestacional, a mortalidade é menor em cada faixa etária. Quando o recém-nascido não tem o peso que deveria ter aumenta o risco de vida, não apenas no período intrauterino, mas também na vida extrauterina. Já existem também algumas evidências, ainda bastante fracas de que o peso de um recém-nascido está relacionado com o peso ao nascer de sua mãe. Isto estaria relacionado a fatores genéticos e ambientais.

Muitas das prováveis causas que levam a aumentar este risco de vida do recém-nascido ainda não são conhecidas. Por outro lado estas causas podem ser diferentes em diferentes grupos populacionais com diferenças regionais, de costumes, etc. Elas podem atuar só ou em combinação de duas ou mais variáveis, modificando assim os níveis de risco de um recém-nascido.

ta de apresentar com maior probabilidade problemas, tanto no período gestacional como na hora do parto.

Por outro lado, dada a limitação dos leitos existentes, é realizada uma triagem na hora da internação, dando preferência aos casos que apresentam riscos e sendo derivados para outras maternidades os casos que não apresentam riscos evidentes.

Estes dois fatores fazem com que os dados analisados neste trabalho não sejam representativos de Campinas ou região. Por esta razão, no estágio atual do estudo, não se pode utilizar os resultados para prever baixo peso para outras populações. Para isto serão necessários estudos similares futuros com um número maior de casos e com amostras representativas da população sobre a qual se quer inferir.

III - INFORMAÇÕES DISPONÍVEIS

III.1 - Introdução

Todas as gestantes que procuram a Maternidade da Faculdade de Ciências Médicas da UNICAMP recebem atendimento médico ou paramédico, segundo as necessidades de cada caso. As gestantes que estão no final de seu período gestacional (trabalho de parto), são internadas para se providenciarem os cuidados médicos necessários.

Em todos os casos de internação para atendimento do parto é preenchida uma ficha obstétrica pré-codificada (ver anexo 1). Nesta ficha estão quantificadas 83 variáveis com informações sobre a mãe, o feto, o parto e o recém-nascido.

Esta ficha é composta de variáveis quantitativas como anos de idade, altura uterina, duração do período de dilatação e nota de Apgar e variáveis qualitativas tais como estado civil, tipo de rotura de membranas, malformações congênitas, são preenchidas na forma de respostas com múltiplas escolhas mutuamente exclusivas. O preenchimento da ficha obstétrica está detalhadamente explicado no Manual de Instruções para o Preenchimento da Ficha Obstétrica do Departamento de Tocoginecologia da Faculdade de Ciências Médicas da UNICAMP, que é lido por todas as pessoas encarregadas do preenchimento da ficha, o que ajuda a assegurar uma uniformidade de critérios e a boa qualidade das informações.

A ficha é preenchida à medida que ocorrem os fatos, sendo que a primeira página é preenchida no instante da internação e a última é completada no momento da alta, quando a ficha é revisada pelo residente responsável pela paciente. A ficha é então numerada e enviada para ser digitada. Após a digitação os dados ainda passam por dois programas de consistência, um para detectar valores absurdos e outro para detectar contradições. Estes programas fazem uma lista dos erros encontrados, as fichas originais são localizadas e providenciam-se então as correções necessárias. Após as correções estes dados são incorporados ao arquivo geral de dados. Os dados utilizados neste trabalho correspondem aos primeiros 2902 casos desse arquivo, para partos únicos (eliminam-se os partos que resultaram gêmeos). Para maiores detalhes ver Sabatino e Araki (1979), palestra sobre o sistema Gênese.

Dentre as 83 variáveis existentes na ficha obstétrica nem todas são capazes de fornecer subsídios para a predição de peso. Algumas destas variáveis com pouca ou nenhuma relação com o peso são de fácil identificação como hora em que aconteceu o parto ou

número de toques vaginais efetuados, etc. Outras variáveis são mais difíceis de identificar como tendo influência no peso.

Dado ao interesse em se trabalhar com variáveis relevantes para a previsão do peso do recém-nascido foi solicitada a colaboração de pessoal médico, docentes da FCM da UNICAMP, que segundo suas experiências pessoais e de literatura puderam dizer quais dentre estas 83 variáveis seriam as mais adequadas e com maior relação com o peso.

Estes docentes identificaram 26 variáveis consideradas relevantes, com alguma relação com o peso. Estas variáveis são as que foram utilizadas neste trabalho e são descritas a seguir.

III.2 - AS VARIÁVEIS

As variáveis seleccionadas são apresentadas com suas descrições e, quando for o caso, com suas respectivas codificações, e o significado de cada código, segundo informações do Manual de Instruções para o Preenchimento da Ficha Obstétrica, Sabatino et al. (1978).

- 1 - IDM Idade da gestante em anos, número de anos completados no último aniversário.
- 2 - ESTCIV Estado civil da gestante.
 - 1 - Solteira, a mulher que nunca esteve casada legalmente, nem vivendo em união estável e que não vive com um companheiro atualmente.
 - 2 - Casada, é a mulher que está legalmente unida a um homem e que vive com ele.
 - 3 - Amasiada, é a mulher que vive em união estável com um homem que não é legalmente seu esposo, ou a mulher que está legalmente casada com um homem, mas vive em união estável com outro.
 - 4 - Desquitada, é a mulher que esteve unida por matrimônio legal ou era amasiada e que, atualmente, não vive com seu companheiro. Pode ser divorciada, desquitada ou simplesmente separada.
 - 5 - Viúva, é a mulher cujo marido ou companheiro faleceu e que não está amasiada.
 - 9 - Ignorado
- 3 - UP Número de meses que transcorreram entre a data do último parto e a data do parto atual. Se o tempo transcorrido for igual ou maior do que 8 anos e dois meses assume o valor 98.
- 4 - NTM Número de natimortos anteriores, onde natimorto significa o parto de um feto que pesa mais de 500 gra-

mas ou que tem 25 cm de altura ou mais e que não apresenta sinais vitais quando ocorre a expulsão.

- 5 - CSR Número de cesáreas a que foi submetida a gestante.
- 6 - ABOR Número de abortos espontâneos e provocados somados.
- 7 - FILMOR Número de filhos que nasceram vivos e que faleceram.
- 8 - TOTPAR Número de gestações que terminaram por via vaginal ou cesárea, independente de que o recém-nascido fosse vivo ou morto. As gestações gemelares contam como dois partos.
- 9 - ANTM Patologias anteriores a esta gestação. No caso de haver mais do que uma é codificada somente a que é considerada mais importante.
- 1 - Sem, a gestante nega haver tido qualquer patologia prévia a esta gestação.
 - 2 - Tuberculose, a gestante relata um diagnóstico de laboratório de tuberculose de qualquer localização, antes desta gestação.
 - 3 - Sífilis, a gestante relata um diagnóstico médico ou há sorologia ou exame microscópico positivo antes desta gravidez.
 - 4 - Diabetes, a gestante relata diagnóstico médico ou há diagnóstico de laboratório antes desta gravidez.
 - 5 - Cardiopatia, a gestante relata um diagnóstico médico ou apresenta evidências de quadro clínico ou há diagnóstico de especialista.
 - 6 - Infecção urinária, a gestante relata um diagnóstico médico ou há diagnóstico de laboratório, anterior a esta gravidez.
 - 7 - Hipertensão, a gestante relata diagnóstico médico de níveis tensionais superiores aos considerados normais ou apresenta evidência do quadro clínico ou há diagnóstico de especialista.
 - 8 - Sequela óssea de pélvis ou inferiores, a gestante apresenta algum impedimento ósseo para a realização do parto por via vaginal.
 - 9 - Cirurgia abdominal, a gestante relata ter sido submetida a operações que deixaram cicatriz uterina, como de miomectomia, sutura de perfurações, correção de anomalias congênitas, conização, etc., não se deve incluir neste grupo as pexias ou operações corretoras da posição uterina.
- Y - Ignorado.
- 10 - ANTOBS Patologias obstétricas acontecidas em gestações anteriores.

- 1 - Sem, a gestante é primigesta ou em gestações anteriores não apresentou nenhum dos problemas especificados a seguir.
- 2 - Toxemia, a gestante relata diagnóstico de toxemia em qualquer das gestações anteriores ou relata haver apresentado sintomas de edema, hipertensão e albuminúria.
- 3 - Metrorragia durante a gravidez, a gestante relata perdas de sangue por via vaginal durante qualquer uma das gestações anteriores.
- 4 - Metrorragia pós parto, a gestante relata perdas de sangue aumentadas após os partos anteriores.
- 5 - Gravidez prolongada, a gestante relata ter ultrapassado 42 semanas de amenorréia durante qualquer das gestações anteriores, considerando-se as dificuldades de precisão existentes.
- 6 - Malformação congênita, a gestante relata antecedentes de malformações congênitas no produto de qualquer das gestações anteriores.
- 7 - Infecções puerperais, a gestante relata antecedentes de infecções puerperais (nos 40 dias depois do parto) após qualquer das gestações anteriores.
- 8 - Polihidramnio, a gestante relata aumentos anormais de líquido amniótico em qualquer das gestações anteriores.
- 9 - Partos prematuros, a gestante relata parto prematuro em qualquer das gestações anteriores, considerando-se as dificuldades de precisão que as pacientes tenham.
- 0 - Outras patologias obstétricas, a gestante relata outra patologia não descrita acima e apresentada em qualquer das gestações anteriores.
- Y - Ignorado.

11 - UAAG

Uso e tipo de anticoncepcional que a gestante usava antes desta gestação.

- 1 - Não usava, a gestante nega o uso de qualquer tipo de anticoncepcional.
- 2 - Gestágeno suspenso, gravidez planejada, a gestante relata ter suspenso o uso de gestágeno porque queria engravidar.
- 3 - Gestágeno suspenso por intolerância, a gestante relata ter suspenso o uso de gestágeno por intolerância e, logo após, engravidou.
- 4 - Gestágeno suspenso por gravidez não planejada, a gestante relata ter engravidado enquanto usava gestágeno e suspendeu o uso.
- 5 - DIU retirado, gravidez planejada, a gestante relata ter retirado o DIU, porque queria ficar grávida.
- 6 - DIU retirado por não adaptação, a gestante relata ter retirado o DIU por não adaptação (hemorragias, dor, etc.). Considera-se também os casos de expulsão.

- 7 - DIU fracassado, a gestante relata ter engravidado com o DIU in situ.
- 8 - Outro anticoncepcional, gravidez planejada, a gestante relata ter suspenso o uso de qualquer outro método anticonceptivo (preservativo, diafragma, Ogino-Knaus, coitus interruptus, etc.) porque queria engravidar.
- 9 - Outro anticoncepcional, intolerância, a gestante relata ter suspenso o uso de qualquer outro anticoncepcional por intolerância e logo após engravidou.
- 0 - Outro anticoncepcional, gravidez não planejada, a gestante relata ter engravidado enquanto usava qualquer outro método anticoncepcional.
- Y - Ignorado.

12 - ES Grau de instrução da gestante, considerando o último ano que a mulher completou, tendo sido aprovada.

- 1 - Nenhum
- 2 - 1 a 2 anos
- 3 - 3 a 4 anos
- 4 - 5 anos
- 5 - 6 a 7 anos
- 6 - 8 a 9 anos
- 7 - 10 a 11 anos
- 8 - 12 anos ou mais
- Y - Ignorado

13 - TOTCON Número total de consultas pré-natais realizadas pela gestante.

14 - MIPN Número do mês de gestação em que a mulher se encontrava quando realizou sua primeira visita pré-natal. No caso de não ter feito pré-natal assume valor 0 e se a primeira visita ocorreu no nono mês assume valor 8.

15 - ALTM Altura da gestante descalça, medida segundo técnica habitual, em centímetros.

16 - SEMAM Semanas de amenorréia, calculada segundo a data da última menstruação pela regra de Nagele.

17 - DIFPON Diferença ponderal ou variação de peso da gestante obtido pela diferença entre o peso no ingresso ao pré-natal menos o peso habitual aproximado antes de engravidar informado pela gestante.

18 - EDEMA Presença de edemas.

- 1 - Sem, a gestante não apresenta edema.
- 2 - Maleolar, apresenta sinal de Godet nas apófises maleolares.
- 3 - Tibial, apresenta sinal de Godet no osso tibial.

- 4 - Abdômem, apresenta sinal de Codet no abdômem.
 5 - Anasarca, a gestante apresenta edema generalizado.
 9 - Ignorado.
- 19 - ALT Altura do recém-nascido em centímetros.
- 20 - PESO Peso do recém-nascido em gramas no momento do nascimento ou até 12 horas depois.
- 21 - CAPU Idade gestacional clínica estimada pelo neonatólogo no momento do parto, segundo método descrito por Capurro et al (1978).
- 22 - ERNA Estado do recém-nascido no momento da alta.
- 1 - Vivo sã, o recém-nascido estava sadio no momento da alta.
 2 - Vivo enfermo, o recém-nascido estava enfermo no momento da alta, mas sua enfermidade não requeria mais internação.
 3 - Natimorto, ver definição na variável NTM.
 4 - Morte neo-natal até 24 horas, a criança morreu antes de completar 24 horas de vida.
 5 - Morte neo-natal de 24 horas até o 7º dia, a criança morreu entre o 2º e o 7º dia de vida, inclusive.
 6 - Morte neo-natal do 8º dia até 28º dia, a criança morreu entre o 8º e o 28º dia de vida inclusive.
 Y - Ignorado.
- 23 - PRART Evolução da pressão arterial da gestante na gravidez e no parto.
- 1 - Normal, a paciente apresentou níveis tensionais inferiores a 140 mmHg sistólica e inferiores a 90 mmHg de diastólica, ou valores de pressão arterial média, (PAM), inferiores a 105 mmHg. A PAM se obtém da seguinte forma:
- $$PAM = \frac{PAS + 2 \times PAD}{3}$$
- onde,
- PAS = pressão arterial sistólica e
 PAD = pressão arterial diastólica.
- 2 - Caídas tensionais sem sinais de choque < 100/60 mmHg, a paciente apresentou SPONÍVEIS

- arterial menores que 100 mmHg de sistólica e menores de 60 mmHg de diastólica, sem sinais de choque (taquicardia, sudorese, palidez, etc.)
- 3 - Choque clínico reversível de qualquer origem, a paciente apresentou quedas tensionais e choque, mas recuperou os níveis normais logo após tratamento efetuado.
 - 4 - Choque clínico irreversível de qualquer origem, a paciente apresentou quedas tensionais e choque, mas não recuperou os níveis normais logo após tratamento efetuado.
 - 5 - Hipertensão leve, a paciente apresentou níveis tensionais diastólicos iguais ou maiores que 90 e menores que 100 mmHg ou a PAM iguais ou maiores que 105 mmHg e menores que 115 mmHg ou ainda quando houve um aumento de 30 mmHg na pressão arterial sistólica ou 15 mmHg na pressão arterial diastólica sobre os níveis tensionais anteriores.
 - 6 - Hipertensão moderada, a paciente apresentou níveis tensionais diastólicos iguais ou maiores que 100 e menores que 120 mmHg ou a PAM apresentou valores iguais ou maiores que 115 e menores que 130 mmHg.
 - 7 - Hipertensão grave, a paciente apresentou níveis tensionais diastólicos iguais ou maiores que 120 mmHg ou a PAM apresentou valores iguais ou maiores que 130 mmHg. Inclui também os casos em que além dos níveis assinalados em hipertensão moderada houve sinais clínicos de irritação do sistema nervoso central (alterações visuais, cefaléia fronto-occipital, náuseas ou vômitos, dor epigástrica, confusão mental, hiper-reflexia, febre, etc.).
 - 8 - Acidente vascular ou visceral hipertensivo, quando, além da hipertensão arterial e como consequência da mesma, acontece qualquer acidente vascular ou visceral (derrame) independente dos níveis tensionais observados.
 - Y - Ignorado

24 - TOX

Ocorrência de toxemia e a gravidade da mesma na gestação atual.

- 1 - Sem, a paciente apresentou níveis normais de tensão arterial, sem sinais de toxemia (edema + albuminúria).
- 2 - Hipertensão leve, provável toxemia, a paciente apresentou hipertensão leve, bem como sinais de provável toxemia (retenção de líquido intensa, aumento abrupto de peso, edema e/ou proteinúria).
- 3 - Hipertensão moderada, provável toxemia, a paciente apresentou hipertensão moderada bem como sinais de provável toxemia.

- 4 - Hipertensão severa, provável toxemia, a paciente apresentou hipertensão grave bem como sinais de provável toxemia.
- 5 - Pré-eclâmpsia moderada, a paciente apresentou sinais de provável toxemia, bem como sintomas de pré-eclâmpsia (cefaléia, tontura, náuseas, perturbações visuais, diarreia, etc.).
- 6 - Pré-eclâmpsia severa, a paciente apresentou os sintomas descritos em pré-eclâmpsia moderada e também quadro de obnublação mental, modificações na capacidade sensorial e de memória, agitação psicomotora, hiper-reflexia, etc.).
- 7 - Hipertensão crônica essencial, a paciente refere pressão alta ou existe diagnóstico da mesma, independente de gravidez.
- 8 - Hipertensão outras origens, não pré-eclâmpsia, a paciente apresentou pressão alta independente da gravidez, que apareceu agudamente (hipertensão endocraniana por tumor, etc.).
- 9 - Eclâmpsia gravídica, a paciente apresentou hipertensão com sinais de provável toxemia e também um quadro convulsivo e/ou comatoso (neurológico) durante a gravidez.
- 0 - Eclâmpsia puerperal, a paciente apresentou as mesmas condições descritas em eclâmpsia gravídica mas que ocorreram até 72 horas após o parto.
- Y - Ignorado.

25 - DIAB Diabetes na gestante.

- 0 - Sem, a paciente não apresenta diabetes.
- 1 - Com, a paciente apresenta diabetes.
- Y - Ignorada, não se tem a informação.

26 - INFEC Doenças infecciosas acontecidas durante a gravidez.

- 1 - Sem, a paciente não apresentou qualquer doença infecciosa durante a gravidez.
- 2 - Toxoplasmose, a paciente apresentou níveis elevados no teste de toxoplasmose ou sinais de reagudização de uma toxoplasmose antiga.
- 3 - Tuberculose, foi diagnosticado TBC ou a paciente apresentou sinais de reagudização de uma TBC antiga.
- 4 - Infecções urinárias sem febre, a paciente apresentou sinais de infecção urinária (disúria, poliúria, punho percussão lombar positiva, bacterioscopia positiva etc), sem febre.
- 5 - Infecções urinárias com febre, a paciente apresentou características de infecção urinária mais febre.
- 6 - Gonorréia, foi diagnosticada gonorréia ou reagudização de gonorréia antiga.
- 7 - Virose, a paciente apresentou sinais de gripe, catapora ou outras viroses.

- 8 - Estafilococos ou estreptococos, foi diagnosticada estafilococcia e/ou estreptococcia (anginas, otitis, piodermitis, etc).
- 9 - Sífilis, foi diagnosticada sífilis ou a paciente apresentou reagudização de sífilis antiga.
- 0 - Outras, a paciente apresentou alguma doença infecciosa não especificada acima.
- Y - Ignorada.

Um detalhamento maior sobre o significado preciso de cada um dos termos e sobre todas as classes e forma de medição ou levantamento de cada uma das variáveis pode ser encontrado no Manual de Instruções para o Preenchimento da Ficha Obstétrica, do Departamento de Tocoginecologia da Faculdade de Ciências Médicas da UNICAMP. Em Sabatino et al. (1980) é apresentado um resumo e são discutidos detalhes sobre o uso desse manual, dos programas e sistemas de crítica e consistência de dados para essa ficha.

Para algumas dessas variáveis, infelizmente, não havia quantidade de casos suficiente para que a variável fosse mantida no estudo, como no caso de diabetes, devido a um número excessivo de faltas de informação. Em outros, a codificação não é a melhor possível, como no caso de antecedentes obstétricos. É importante ressaltar que apesar desses problemas a qualidade dos dados é bastante boa, principalmente se considerarmos que para variáveis de natureza biológica existem, muitas vezes grandes imprecisões nas determinações e, frequentemente, surgem erros do tipo classificado como "grosseiros".

Um outro aspecto a se chamar a atenção é sobre a natureza pioneira do trabalho que se realiza na Maternidade do Departamento de Tocoginecologia da FCM da UNICAMP. Trata-se da primeira experiência dessa escala a ser feita no país.

III.3 - RECODIFICAÇÃO DAS VARIÁVEIS

Para que fosse possível usar essas variáveis nas análises que se contemplava, regressão linear múltipla e regressão logística, foi necessário recodificar diversas delas. As variáveis categóricas foram quase sempre transformadas em variáveis indicadoras, às vezes mais de uma usando a mesma variável original. Em outras situações foram combinadas várias variáveis para formar uma ou mais variáveis indicadoras. Todas essas recodificações foram feitas com base em informações, algumas vezes de natureza imprecisa, mas sempre por sugestão de membros do Departamento de Tocoginecologia da FCM, com vivência do problema e com base na pouca literatura existente sobre o tema de previsão de peso do recém-nascido.

As faltas de informação (FI) foram recodificadas como "." objetivando usar os recursos de análise do Statistical Analysis

System (SAS), instalado no Centro de Computação da Universidade Federal de São Carlos.

A seguir é feita uma breve apresentação de todas as recodificações que foram feitas nos dados, e em alguns casos é dada uma razão para a recodificação escolhida. Houve, além de razões de simplificação, oportunidades em que a recodificação foi feita visando recuperar a natureza contínua de uma variável, como no caso de escolaridade da mãe. A quase totalidade das recodificações feitas foi no sentido de transformar variáveis com mais de duas categorias em variáveis indicadoras, tanto para simplicidade de uso como para permitir maiores facilidades na interpretação de resultados.

A variável ESTCIV, estado civil da mãe, que especificava como possibilidades solteira, casada, amasiada, desquitada e viúva, foi transformada de forma a indicar ou não a presença do cônjuge, sendo atribuídos os valores:

ESTCIV=1 para solteira, desquitada ou viúva (sem a presença do cônjuge) e

ESTCIV=0 para casada ou amasiada (com a presença do cônjuge).

A variável ES, escolaridade, da maneira como estava codificada era uma variável classificatória mas, ainda que apresentasse uma ordem natural, não podia ser tratada como variável contínua. Ela foi transformada na variável ESC de forma a representar, de maneira aproximada, o número de anos de escolaridade da mãe. Os erros mais importantes ocorrem na classe 8, por ser aberta na parte superior. No caso dos dados desses 2902 partos essa classe representa, no entanto, menos de 0.5% dos casos.

ES	NÚMERO DE ANOS DE ESCOLARIDADE	ESC
1	0	0.0
2	1 a 2	1.5
3	3 a 4	3.5
4	5	5.0
5	6 a 7	6.5
6	8 a 9	8.5
7	10 a 11	10.5
8	12 ou mais	13.0
Y	F1	"."

A variável EDEMA foi transformada de forma a indicar ou não a presença de edemas na gestante, sendo atribuídos os valores:

EDEMA=0 se não apresentar edemas e

EDEMA=1 se apresentar.

A variável ANTOBS1 foi criada para indicar a presença de prematuros no histórico da gestante (partos anteriores ao atual), sendo:

ANTOBS1=1 em caso de ter havido prematuros e
ANTOBS1=0 caso contrário.

A variável ANTOBS2 foi criada para indicar todos os antecedentes obstétricos (partos anteriores ao presente) indicados por ANTOBS menos prematuros. Desta forma:

ANTOBS2=1 indica pelo menos uma das seguintes patologias, toxemia, metrorragia durante a gravidez, metrorragia pós parto, gravidez prolongada, malformação congênita, infecções puerperais, polihidramnio ou outras patologias obstétricas e
ANTOBS2=0 indica a não ocorrência das patologias acima.

A variável ANTM, antecedentes mórbitos, foi transformada para ANTMOR de forma que:

ANTMOR=1 indica a ocorrência de pelo menos um dos antecedentes: tuberculose, sífilis, diabetes, cardiopatia ou hipertensão e
ANTMOR=0 indica a não ocorrência de qualquer das patologias citadas acima.

A variável UAAG, relativa ao uso de anticoncepcionais antecedendo a esta gestação, foi transformada para ANTCON, indicadora do uso de anticoncepcionais à base de gestágeno. Desta forma:

ANTCON=1 indica o uso de anticoncepcionais à base de gestágeno e
ANTCON=0 indica o não uso dos mesmos.

A variável PRART, indicadora de nível de pressão arterial na gravidez atual foi dividida em duas variáveis binárias, PRALT que indica situações de pressão arterial diferente da normal e PRGR que indica situações de pressão arterial elevadas, maiores que 180/120, de maneira que

PRALT=0 se evolução da pressão arterial é normal e
PRALT=1 se diferente da normal,

PRGR=1 se a pressão esteve acima de 180/120 e
PRGR=0 caso contrário.

A variável TOX, indicadora de níveis de toxemia, foi transformada em binária, de forma que a nova variável

TOXE = 0 indica ausência de toxemia ou hipertensão leve,

TOXE = 1 indica todos os outros níveis da variável TOX, hipertensão moderada ou severa, pré-eclâmpsia e eclâmpsia.

A variável INFEC, indicadora de tipos de infecções foi recodificada para INF de forma a indicar algumas doenças infecciosas na gestante, de forma que,

INF=1 se a gestante apresentar pelo menos uma das seguintes doenças, toxoplasmose, tuberculose, infecções urinárias com ou sem febre, gonorréia, virose, estafilococos, estreptococos ou outras, exceto sífilis que é tratada separadamente,
INF=0 caso não se enquadre na descrição acima.

A variável SIFI foi criada para indicar as gestantes portadoras de sífilis sendo que

SIFI=1 se a gestante é portadora de sífilis e
SIFI=0 caso contrário.

As transformações, como mencionado anteriormente, foram feitas para transformar variáveis classificatórias em variáveis indicadoras, e sustentadas por experiência médica anterior.

Uma das variáveis mais importantes para o estudo, como descrito no Capítulo II, é a idade gestacional que será chamada de IDG e terá como unidade a semana. Existem, na ficha obstétrica, duas fontes de informação sobre a idade gestacional. A primeira, por anamnese, usando informações das parturientes quanto ao número de semanas de amenorréia (SEMAM) e a segunda obtida através do exame clínico do recém-nascido (CAPU). Tanto numa como na outra existem faltas de informação.

A variável IDG foi construída da seguinte maneira:

Se CAPU não for FI (falta de informação) então IDG é igual a CAPU. Caso contrário IDG é igual a SEMAM.

Neste caso CAPU, a variável que contém os valores da idade gestacional clínica, determinada pelo neonatólogo no momento do parto, foi considerada primeiro por ser determinada com métodos mais precisos que a idade gestacional obtida através de informação verbal da gestante sobre a data da última menstruação, SEMAM, variável em que também se observam faltas de informação mais frequentemente que em CAPU.

IV - METODOLOGIA

IV.1 - INTRODUÇÃO

Existem várias técnicas estatísticas disponíveis para se tentar construir um modelo de previsão de baixo peso ao nascer (peso menor do que 2500 gramas). Entre elas destacam-se regressão linear e regressão logística. Neste capítulo são descritas brevemente essas duas metodologias, visando estabelecer notação e terminologia.

A primeira técnica utilizada é o ajuste de um modelo de regressão linear, com a variável peso como dependente e as demais variáveis disponíveis, descritas no capítulo III, como possíveis variáveis independentes. Uma vez feito o ajuste utiliza-se o peso ajustado para prever se o parto terá ou não como resultado um recém-nascido de baixo peso. Este ajuste foi feito inicialmente através da aplicação de regressão linear dando origem ao MODELO 0. A seguir foram tentadas algumas variantes deste ajuste inicial, visando melhorar as previsões, o que foi feito nos MODELOS 1 a 11. Foram feitas tentativas de usar a altura prevista do recém-nascido para melhorar a previsão do peso, bem como dar pesos maiores que um às observações cujas previsões erraram para maior, previram um peso acima do verdadeiro e peso um nas demais, isto considerando que uma previsão de peso normal quando este é baixo é mais grave que a situação inversa. Os resultados obtidos através dessas técnicas estão apresentados no capítulo V.

A outra técnica utilizada é a regressão logística. Com esta metodologia é estimada a probabilidade de que uma dada observação esteja associada a um parto com recém-nascido com peso normal, onde aqui considera-se parto com peso normal aquele em que resulta um recém-nascido com peso maior ou igual a 2500 gramas. Para trabalhar com regressão logística foi criada uma variável indicadora, IPESO, igual a um quando o peso do recém-nascido é maior ou igual a 2500 gramas e igual a zero caso contrário. Isto está explicado com maiores detalhes no capítulo VI.

Existem muitos livros que tratam do tema de regressão linear, como Draper e Smith (1981) e Montgomery e Peck (1982). Para regressão logística as referências são mais escassas podendo ser mencionados os livros de Cox (1970) e de Anderson et al. (1980), capítulo 09.

IV.2 - REGRESSÃO LINEAR

O modelo a ser ajustado é da forma

$$y(i) = A_1 x(i,1) + A_2 x(i,2) + \dots + A_p x(i,p) + e(i) \quad (4.1)$$

onde,

$y(i)$ é a i-ésima observação da variável y .

$x(i,k)$ é a i-ésima observação da variável X_k .

A_1, A_2, \dots, A_p são os parâmetros do modelo.

$e(i)$ é o erro cometido.

A variável y é usualmente chamada de variável dependente e as variáveis X_1, X_2, \dots, X_k são chamadas de variáveis independentes, ou explicativas. Quase sempre inclui-se um termo constante no modelo e, neste caso, uma das variáveis independentes, X_1 , por exemplo, tem todos os valores iguais a um para representar esse termo constante do ajuste. Todos os ajustes que serão feitos no Capítulo V incluem um termo constante.

O modelo (4.1) pode ser escrito com notação matricial na forma

$$Y = XA + E \quad (4.2)$$

onde,

Y é um vetor contendo as observações da variável y .

$$Y = \begin{bmatrix} y(1) \\ y(2) \\ . \\ . \\ y(n) \end{bmatrix}$$

X é uma matriz composta pelos valores das variáveis explicativas, da forma:

$$X = \begin{bmatrix} x(1,1) & x(1,2) & \dots & x(1,p) \\ x(2,1) & x(2,2) & \dots & x(2,p) \\ \vdots & \vdots & \ddots & \vdots \\ x(n,1) & x(n,2) & \dots & x(n,p) \end{bmatrix}$$

E é o vetor de erros.

$$E = \begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(n) \end{bmatrix}$$

A é um vetor de parâmetros.

$$A = \begin{bmatrix} A1 \\ A2 \\ \vdots \\ Ap \end{bmatrix}$$

No modelo linear tem-se os parâmetros $A1, A2, \dots, Ap$ que são valores desconhecidos, mas que podem ser aproximados através de técnicas estatísticas. Ajustar um modelo a um conjunto de dados significa encontrar valores que se aproximam bastante dos valores desconhecidos dos parâmetros $A1, A2, \dots, Ap$, usando os dados para obtê-los. Estes valores são chamados "estimativas" dos parâmetros e são denotados por $a1, a2, \dots, ap$. Na notação matricial, $a = [a1, a2, \dots, ap]'$ é um vetor que contém as estimativas dos parâmetros.

Este ajuste pode ser feito através de vários critérios, sendo que o mais usual é o método de mínimos quadrados.

No método de mínimos quadrados as estimativas dos parâmetros são aquelas que tornam a soma de quadrados de resíduos mínima. Utilizando notação matricial as estimativas dos parâmetros são os valores de "a" que substituídos em lugar de "A", minimizam a forma quadrática da equação (4.3):

$$E'E = (Y - XA)'(Y - XA) \quad (4.3)$$

Obter as estimativas dos parâmetros, o vetor "a", no método de mínimos quadrados equivale a encontrar o ponto de mínimo da equação (4.3). Assim sendo, para obter os estimadores de A deve-se derivar a equação (4.3) com relação a A e igualar a zero. O resultado deste procedimento é apresentado a seguir:

$$-2X'Y + 2X'Xa = 0$$

Passando o termo negativo para o lado direito da igualdade e dividindo ambos os lados da igualdade por dois, tem-se como resultado a seguinte igualdade:

$$X'Xa = X'Y$$

Premultiplicando ambos os lados da igualdade pela inversa de $(X'X)$ tem-se a equação (4.4).

$$a = (X'X)^{-1} X'Y \quad (4.4)$$

Através dos parâmetros estimados pode-se obter os valores previstos pelo modelo para a variável y através da equação (4.5), denotar-se-a por Y_P o vetor com as previsões.

$$Y_P = Xa = X(X'X)^{-1} X'Y \quad (4.5)$$

Uma vez ajustado o modelo é indispensável averiguar se ele é satisfatório. Existem várias técnicas para esse fim, incluindo testes de hipóteses e, principalmente, análises dos resíduos, incluindo gráficos de resíduos versus valores ajustados.

Para testar a validade do modelo ajustado, $Y_P = Xa$, é necessário fazer algumas suposições a respeito da distribuição de probabilidade do erro E:

- S1 - Os erros são variáveis aleatórias com esperança zero e variância constante, qualquer que seja i.

$$E(e(i)) = 0$$

$$\text{Var}(e(i)) = \text{constante}$$

para $i=1, 2, \dots, n$.

S2 - Os erros de observações diferentes são não correlacionados.

$$\text{COV}(e(i), e(j)) = 0 \quad \text{para } i \neq j.$$

S3 - Os erros tem distribuição normal com média zero e variâncias iguais.

$$e(i) \sim N(0, S^2) \quad \text{para } i=1, 2, \dots, n.$$

Onde S^2 é a variância do resíduo, variância que deve ser constante para todos os resíduos segundo S1.

Estas suposições conjuntamente equivalem a dizer que os erros tem distribuições normais independentes com médias iguais a zero e variâncias iguais entre si.

Tendo ajustado o modelo através de (4.4) ou (4.5) a pergunta que surge é a seguinte: Será que este modelo ajustado é um bom modelo? Para respondê-la deve-se testar estatisticamente a validade do modelo. Uma forma de fazê-lo pode ser testando-se se os parâmetros são conjuntamente diferentes de zero. O modelo seria considerado válido se pelo menos um dos parâmetros testados fosse significativamente diferente de zero, não sendo considerado válido se em nenhum dos parâmetros estimados houvesse evidências suficientes para ser rejeitada a hipótese de que seja igual a zero.

Para testar se os parâmetros são conjuntamente diferentes de zero formulam-se duas hipóteses:

Uma hipótese conservativa a qual denomina-se "hipótese nula" que é aqui representada pela letra "H" e cujo significado é: "todos os parâmetros são iguais a zero".

E outra hipótese chamada "hipótese alternativa", aqui representada pela letra "K" e cujo significado é: "pelo menos um dos parâmetros é diferente de zero".

A hipótese "H", todos os parâmetros são iguais a zero, é aquela que será testada. Se for rejeitada fica-se com a hipótese alternativa, "K", pelo menos um dos parâmetros é diferente de zero.

Neste caso então, testar o modelo se resume a testar H, se rejeitada H o modelo é considerado válido, se não rejeitada o modelo não é considerado válido.

O teste pode ser representado esquematicamente da seguinte maneira:

$$H: A = 0 \quad \text{versus} \quad K: A \neq 0$$

Este teste é baseado na teoria das probabilidades e está detalhadamente explicado em Draper-Smith (1981) e Montgomery-Peck (1982). A seguir descreve-se sucintamente o procedimento de teste.

Inicialmente definimos algumas estatísticas que serão utilizadas nos testes. São elas o quadrado médio de regressão que será representado por QMReg, o quadrado médio de resíduo será representado por QMRes, e a estatística de teste do modelo representada por f . Estas estatísticas são definidas por:

$$\begin{aligned} \text{QMReg} &= \frac{a'X'Xa}{p} = \frac{Y'X(X'X)^{-1}X'Y}{p} \\ \text{QMRes} &= \frac{(Y - Xa)'(Y - Xa)}{n - p} = \frac{Y'(I - X(X'X)^{-1}X')Y}{n - p} \\ f &= \frac{\text{QMReg}}{\text{QMRes}} \end{aligned} \quad (4.6)$$

Utilizando as suposições S1, S2 e S3 e os resultados dos teoremas de Gauss-Markov e de Fisher-Cochran tem-se que a estatística f calculada em (4.6) segue a distribuição de probabilidade F com p e $(n-p)$ graus de liberdade, quando a hipótese nula, H , é a verdadeira.

Isto permite considerar válido o modelo quando o valor da estatística f for maior que F_c , onde F_c é o valor da abscissa da função de distribuição acumulada F que fornece uma probabilidade acumulada de $(1-\alpha)$.

Em outras palavras, consideramos válido o modelo se

$$f > F_c$$

onde,

$$F_c = F(p, (n-p); (1-\alpha)).$$

Neste caso alfa é a probabilidade de se cometer o erro de validar o modelo fundamentando-se somente em variações aleatórias, quando o mesmo não é significativo. Alfa é chamado de nível de significância do teste e geralmente escolhe-se para ele valores pequenos, sendo mais frequentemente utilizado 0.05, valores menores podem ser escolhidos quando se desejar que a probabilidade de se cometer este erro seja menor ainda.

Além de testar-se o modelo como um todo, onde se testa se todos os parâmetros são conjuntamente iguais a zero, também pode-se testar se os parâmetros das variáveis são individualmente iguais a zero ou não, o que equivale a testar se a variável correspondente ao parâmetro testado deveria ou não estar incluída no conjunto de variáveis explicativas do modelo.

Para testar a validade da inclusão de uma variável em particular no modelo, deve-se testar se o parâmetro correspondente a ela é igual a zero. Se por exemplo há interesse em testar a inclusão da variável X_j no modelo, pode-se fazê-lo testando se o coeficiente de X_j , neste caso A_j , é igual a zero. Neste caso pode ser representado esquematicamente nas forma:

$$H: A_j = 0 \quad \text{versus} \quad K: A_j \neq 0.$$

Para testar hipóteses deste tipo calculado-se o acréscimo na soma de quadrados de resíduos ocorrido com a saída do modelo da variável explicativa que se está querendo testar. Por exemplo se a variável a ser testada é X_j , calcula-se o acréscimo na soma de quadrados de resíduos quando $A_j = 0$, denotado por $ASQ (A_j = 0)$.

Isto é feito da seguinte maneira:

$$ASQ (A_j = 0) = a_j^2 (X'X)^{-1}_{j,j}$$

onde, $(X'X)^{-1}_{j,j}$ é j-ésimo elemento da diagonal principal da inversa da matriz $X'X$ e

a_j é o parâmetro estimado para a variável X_j .

$$f = \frac{ASQ (A_j = 0)}{QMRes} \quad (4.7)$$

Neste caso a estatística f calculada em (4.7) segue a distribuição de probabilidade F com 1 e $(n-p)$ graus de liberdade, sob a hipótese nula, e o procedimento de teste é semelhante ao descrito anteriormente para o modelo como um todo.

Considera-se a variável que está sendo testada válida no modelo quando

$$f > F_c$$

sendo,

$$F_c = F(1, (n-p); (1-\alpha)). \quad (4.8)$$

Estas ferramentas de teste de variáveis serão utilizadas nos procedimentos de seleção de variáveis na seção seguinte.

IV.3 - SELEÇÃO DE VARIÁVEIS PARA O MODELO LINEAR.

Quando se tem uma variável dependente e uma variável explicativa, ajusta-se o modelo, testa-se e está pronto. Mas quando se tem várias variáveis explicativas, o resultado não é tão direto.

O primeiro problema que surge é saber quais são as variáveis que devem integrar o conjunto das variáveis explicativas do modelo e quais devem ficar fora dele. Para resolver este problema existem diversas técnicas para seleção de variáveis.

As técnicas de seleção de variáveis mais comuns e também as que são utilizadas neste trabalho são a "FORWARD", a "BACKWARD" e a "STEPWISE".

Segue uma breve descrição das três técnicas.

A técnica "FORWARD" de seleção começa com um modelo sem variáveis explicativas, tendo somente a constante ou média.

Calcula-se então a estatística f , definida em (4.7), que reflete a contribuição de cada variável para a explicação da variável dependente se ela fosse incluída. Isto é feito para todas as variáveis candidatas a entrar no modelo.

A seguir a estatística f de cada variável é comparada com um valor de referência F_i , obtido da mesma forma que F_c em (4.8). Dentre estas são consideradas aptas a entrar no modelo as variáveis cuja estatística f for maior que F_i .

Neste estudo o valor de referência F_i permite a inclusão de variáveis com níveis de significância $p < 0.50$, este valor de referência permite a inclusão de variáveis com pouca importância mas posteriormente permitirá a comparação das variáveis escolhidas neste método e em outros para níveis de significância menores que 0.50.

Se nenhuma estatística f for maior que F_i o processo pára. Caso contrário inclui-se no modelo a variável que tiver maior valor na estatística f e continua-se o processo.

A seguir com a variável já incluída calcula-se o valor da estatística f para as variáveis restantes fora do modelo, redefine-se o valor de F_i como em (4.8). Convém notar a cada nova variável incluída no modelo a distribuição F usada na definição de F_i perde um grau de liberdade.

Repete-se este processo até que nenhuma das variáveis fora do modelo produza uma estatística f suficientemente alta, que permita a sua inclusão no modelo.

A técnica "BACKWARD" inicia com um modelo com todas as variáveis independentes possíveis.

Então as variáveis são excluídas do modelo até que todas as variáveis remanescentes no modelo produzam estatísticas f significantes a um nível alfa pré-estabelecido.

Neste caso utiliza-se um nível de significância $p < 0.10$ para determinar o limite mínimo aceitável da estatística f para que uma variável permaneça no conjunto das variáveis explicativas do modelo. Este limite é determinado da mesma forma que F_c em (4.8) e é representado como F_o .

A cada passo a variável com menor contribuição para o modelo, ou menor estatística f , é excluída se $f < F_o$. Esta exclusão é feita uma a uma e a cada passo as estatísticas f são recalculadas e F_o também é redefinido.

A técnica "STEPWISE" é uma modificação da técnica "FORWARD" e a diferença entre elas é que nesta técnica uma variável que já está no modelo pode ser excluída enquanto que na "FORWARD" isso não acontece.

Na técnica "STEPWISE" as variáveis são incluídas no modelo uma a uma da mesma maneira que na "FORWARD".

Após uma variável ser incluída, a técnica "STEPWISE" testa todas as variáveis que estão no modelo e exclui as variáveis que não tenham a estatística f maior que F_o , uma a uma, iniciando pela menor estatística f recalculando-as e redefinindo F_o a cada passo.

Após serem excluídas as variáveis com pequena contribuição ao modelo então é que se busca uma nova variável para ser incluída no mesmo.

Termina-se a inclusão de variáveis quando nenhuma das variáveis que estão fora do modelo puder ser incluída.

As estatísticas f usadas na regressão STEPWISE são calculadas como em (4.7) e os limites F_i e F_o são obtidos como F_c em (4.8) para níveis de significância alfa igual a 0.50 em F_i e alfa igual a 0.10 em F_o .

Os procedimentos computacionais utilizados foram os do SAS implantado na Universidade Federal de São Carlos.

Em nosso caso, o modelo foi escolhido com base na comparação dos resultados das três técnicas, resultados estes que foram bastante semelhantes.

IV.4 - MODELO LOGÍSTICO.

Como visto no início deste capítulo, quando se tem uma variável dependente e um conjunto de variáveis explicativas pode-se ajustar um modelo linear através do critério de mínimos quadrados também chamado de regressão linear.

Esta técnica é muito utilizada já que do ponto de vista teórico é simples, sendo muito interessante também do ponto de vista computacional, existindo métodos rápidos e precisos para se calcularem as estatísticas necessárias.

Contudo quando se tem um modelo cuja resposta é binária (a variável dependente assumindo somente dois valores), a situação não é tão simples assim. Neste caso haverão problemas com o modelo de regressão linear no que se refere às suposições feitas.

Quando se tem resposta binária na variável dependente, pode-se representar as respostas possíveis por 0 (zero) e 1 (um).

Seja o modelo linear apresentado em (4.1),

$$y(i) = A_1 x(i,1) + A_2 x(i,2) + \dots + A_p x(i,p) + e(i)$$

logo por S1 a esperança de $y(i)$ é

$$E[y(i)] = A_1 x(i,1) + A_2 x(i,2) + \dots + A_p x(i,p) \quad (4.9)$$

mas como $y(i)$ assume somente os valores 0 e 1, tem-se

$$E[y(i)] = P[y(i) = 1] \quad (4.10)$$

Em (4.10) constata-se que $E[y(i)]$ é uma probabilidade, já em (4.9) tem-se para $E[y(i)]$ uma expressão que pode assumir valores fora do intervalo $[0,1]$ à medida que variam os $x(i,k)$'s, o que é um absurdo, já que se trata de uma probabilidade.

Em segundo lugar a suposição de normalidade nos erros, necessária para os testes de hipóteses, não se justifica já que a distribuição do erro é binária e assume os valores

$$- E(y(i) / x(i,1), x(i,2), \dots, x(i,p)) \text{ se } y(i) = 0$$

e

$$1 - E(y(i) / x(i,1), x(i,2), \dots, x(i,p)) \text{ se } y(i) = 1.$$

Uma alternativa à regressão linear é dizer que alguma função da média de y é uma combinação linear das variáveis explicativas. Chamando de $P(i)$ a probabilidade de $\{ y(i) = 1, \text{ dados os valores de } x(i,1), x(i,2), \dots, x(i,p) \}$, então

$$P(i) = E[y(i) / x(i,1), x(i,2), \dots, x(i,p)]$$

Usando agora uma função de $P(i)$,

$$\ln[P(i) / (1 - P(i))] = x(i,1) A_1 + x(i,2) A_2 + \dots + x(i,p) A_p \quad (4.11)$$

A quantidade no lado esquerdo da equação (4.11) é chamada logito de $P(i)$. O logito tem a propriedade de transformar pontos em um intervalo $(0,1)$ para pontos em um intervalo $(-\infty, +\infty)$. Isolando $P(i)$ do lado esquerdo da expressão obtém-se a transformação inversa, qualquer ponto no intervalo $(-\infty, +\infty)$ é transformado para um ponto no intervalo $(0,1)$. Isto faz com que qualquer valor previsto em (4.12) esteja no intervalo $(0,1)$ enquadrando-o na condição de ser uma probabilidade.

Isolando $P(i)$ no lado esquerdo de (4.11) obtém-se

$$P(i) = 1 / (1 + \exp(- X_i' A)) \quad (4.12)$$

A equação mostrada em (4.11) e (4.12) é chamada de modelo logístico.

Ajustar um modelo logístico significa encontrar estimativas de " A ", denotadas por " a ". Isto pode ser feito através dos estimadores de máxima verossimilhança. O método de máxima verossimilhança encontra-se descrito em Bickel & Docksum (1977), a seguir é apresentado um resumo da aplicação do método ao problema em questão.

No método de máxima verossimilhança as estimativas dos parâmetros assumem os valores que maximizam a função de verossimilhança mostrada em (4.13)

$$L = \prod_{i=1}^n \exp (X_i' A y(i)) / (1 + \exp (X_i' A)) \quad (4.13)$$

onde

X_i é um vetor com os valores das variáveis explicativas relativas à i -ésima ocorrência da variável y , $y(i)$.

Sendo (4.13) uma função não linear, para se encontrar o vetor " a " que substituindo " A " torna (4.13) máxima deve-se utilizar um método numérico. O método geralmente usado é o método de Newton-Raphson.

O método de Newton-Raphson é um método iterativo para se encontrar raízes de funções. Portanto, para encontrar o máximo da função de verossimilhança L , deve ser aplicado às derivadas de L . Como o logaritmo de uma função qualquer não altera a abscissa de seu ponto de máximo, pode-se obter o máximo da função L calculando o máximo da função $\ln [L]$, com a vantagem de $\ln [L]$ ter uma derivada com expressão mais simples.

$$\ln (L) = \sum_{i=1}^n X_i' A y(i) - \ln (1 + \exp(X_i' A)) \quad (4.14)$$

O método de Newton-Raphson aplicado ao logaritmo da função de verossimilhança do modelo logístico (4.14) leva ao método iterativo descrito na equação (4.15).

$$a_{t+1} = a_t + [I(a_t)]^{-1} S(a_t) \quad (4.15)$$

onde,

$S(a_t)$ é a matriz "score" na iteração t .

$I(a_t)$ é a matriz informação na iteração t .

Os elementos da matriz "score" são dados por:

$$S_{(j)}(a) = \sum_{i=1}^n \left[x(i,j) \left(y(i) - \frac{\exp(Xi'a)}{(1 + \exp(Xi'a))} \right) \right]$$

para $j = 1, \dots, p$

Os elementos da matriz informação são dados por:

$$I_{(j,k)}(a) = \sum_{i=1}^n x(i,j) x(i,k) \exp(Xi'a) / (1 + \exp(Xi'a))^2$$

para $j = 1, \dots, p$ e $k = 1, \dots, p$.

Como no caso linear pode-se testar o modelo como um todo ou as variáveis isoladamente. Isto pode ser feito através do teste da razão de verossimilhança. Para maiores detalhes sobre o teste da razão de verossimilhança ver Bickel & Doksum (1977).

Para testar o modelo como um todo, ou seja todas as variáveis conjuntamente, ou um subconjunto das variáveis do modelo, que pode ser de uma única variável, pode ser usado o teste da razão de verossimilhança. Neste teste utiliza-se a estatística mostrada na equação (4.16).

$$X_q = 2 (\ln (L) - \ln (L_1)) \quad (4.16)$$

Onde,

L é a função de verossimilhança para o modelo com todas as variáveis e

L_1 é a função de verossimilhança para o modelo com as q variáveis que permanecem no modelo quando as variáveis que se deseja testar estão fora do mesmo.

Quando se testa se um subconjunto de q variáveis é conjuntamente igual a zero, a estatística X_q tem uma distribuição qui-quadrado com q graus de liberdade, sob a hipótese de validade de H .

Neste caso testa-se:

H : Os coeficientes das q variáveis são conjuntamente iguais a zero.

K : Pelo menos um dos parâmetros é diferente de zero.

Se a estatística X_q é maior que um limite X_c rejeita-se a hipótese H . O limite X_c é definido de forma que a função acumulada da distribuição qui-quadrado com q graus de liberdade no ponto X_c assume o valor $(1-\alpha)$, onde α é o nível de significância do teste.

O teste da razão de verossimilhança pode ser usado para testar subconjuntos das variáveis explicativas com apenas uma variável. Testar se um certo parâmetro é igual a zero equivale a testar a validade da inclusão da respectiva variável no modelo. Desta forma pode-se saber se a contribuição desta variável é ou não significativa.

Outro modo de testar esta contribuição é através da construção de intervalos de confiança para os parâmetros estimados.

Para se construir os intervalos de confiança utiliza-se a inversa da matriz de informação que na última iteração é assintoticamente a matriz de variâncias e covariâncias. Assim sendo, para grandes amostras, o erro padrão do parâmetro estimado para a i -ésima variável, $a(i)$, é a raiz quadrada do i -ésimo elemento da diagonal principal da inversa da matriz de informação na última iteração.

$$EP(a(i)) = \sqrt{\frac{-1}{[I(a)]_{k,i,i}}}$$

sendo k a última iteração.

Para aplicação desta técnica de regressão logística aos dados foram utilizados os recursos computacionais do Statistical Analysis System (SAS) instalado no Centro de Atendimento Integral à Mulher (CAISM) na Faculdade de Ciências Médicas da UNICAMP.

O manual de utilização do SAS, no capítulo referente ao "LOGIST PROCEDURE" tem informações detalhadas sobre os procedimentos computacionais utilizados.

V - REGRESSÃO LINEAR E VARIANTES

Neste capítulo está apresentado o ajuste do modelo linear utilizando as técnicas descritas em IV.2, passando pela detecção de outliers até chegar a um modelo inicial que é chamado de MODELO 0. Seguindo o trabalho então são testadas variações deste. A cada variação experimentada os resultados são comparados com os resultados obtidos no modelo inicial.

V.1 - MODELO PROVISÓRIO

O ajuste tem início com a modelagem através das técnicas de regressão "stepwise", "backward" e "forward" para seleção das variáveis explicativas para o peso do recém-nascido.

As variáveis que concorreram nesta seleção foram as seguintes: IDM, ESTCIV, NTM, ABOR, FILMOR, TOTPAR, ANTHOR, ANTOBS1, ANTOBS2, DIAB, INF, ANTCON, ESC, TOTCON, ALTM, EDEMA, PRALT, PRGR, TOXE, SIFI, DIFPON e IDG.

Pelo exposto anteriormente e a luz dos resultados obtidos durante o trabalho, a variável IDG se manifestou como sendo de grande importância na previsão de PESO. Por outro lado ela frequentemente não está disponível no momento da avaliação do caso (geralmente pela ignorância da mulher sobre a data da última menstruação). Este problema ainda não tem solução (substituí-la por outra com similar poder preditivo). Isto forçaria o uso do modelo somente nas gestantes que conhecem e informam a data da sua última menstruação e através desta informação calcular aproximadamente a idade gestacional segundo a regra de Nagele, conforme Sabatino (1987).

As técnicas de regressão "stepwise", "forward" e "backward" foram aplicadas aos dados e os conjuntos de variáveis selecionadas nas diferentes técnicas foram comparados. O resultado desta comparação foi que todas as técnicas apontam para conjuntos de variáveis muito semelhantes, esta concordância é um resultado bastante desejado pois indica a ausência de colinearidades. A indicação de ausência de colinearidade foi confirmada posteriormente por um teste de multicolinearidade.

Para estabelecer as variáveis que comporiam o conjunto de variáveis do modelo provisório foram selecionadas as variáveis obtidas na regressão stepwise e que apresentavam um nível de significância $p < 0.05$. Estabelecido este modelo provisório passou-se à detecção de outliers.

V.2 - DETECÇÃO DE "OUTLIERS"

Quando é ajustado um modelo pelo método de mínimos quadrados, um único ponto ou um conjunto de pontos estranhos, devidos a erros de digitação ou anotação, a falhas em aparelhos de medição ou mesmo ao comportamento estranho da variável podem causar alterações drásticas nos resultados obtidos no ajuste. Tais pontos são chamados "outliers". Em função destas alterações drásticas causadas por estes outliers torna-se importante detectá-los e identificar a sua origem. Reunindo as informações disponíveis sobre o(s) ponto(s) pode-se, quando possível, corrigir os erros que houverem, quando não for possível corrigir tais erros nos restará a opção de eliminar a observação quando ela estiver comprovadamente incorreta.

Convém ressaltar que a eliminação de observações é uma prática que gera polémica entre os especialistas quanto à validade da sua aplicação. Neste trabalho ela será aplicada somente em casos extremos.

Quando se tem um "outlier" e se constata a não existência de erros, decide-se sobre a permanência ou não da observação no ajuste utilizando critérios relacionados com a análise dos fenómenos que as variáveis representam.

Partindo do ajuste de um modelo provisório passou-se ao estudo para detectar "outliers" utilizando as técnicas conhecidas, D de Cooke, DFFITS, gráficos de RESÍDUOS versus PESO PREVISTO, gráficos de resíduos parciais e diagonal da matriz de projeção como explicado, por exemplo em Belsley et al (1980) e Carvalho & Dachs (1985).

Observando na figura 5.1 o gráfico de RESÍDUOS versus PESO PREVISTO, já à primeira vista pode-se notar um ponto estranho que requer uma atenção especial. Analisando a observação correspondente ao ponto isolado no gráfico da figura 5.1 constatou-se que se tratava de uma gestante com idade gestacional de 52 semanas, o que equivale a uma gravidez de doze meses. Não se conseguiu determinar a origem deste número, se devido a erro de digitação ou preenchimento de questionário ou tratou-se de um caso de prolongamento exagerado da gravidez. A decisão então foi pela exclusão desta observação no ajuste do modelo, dado que a literatura especializada nos informa que uma gestação durar doze meses é um evento quase impossível.

Os gráficos das figuras 5.1 e 5.2 são compostos de letras onde normalmente estão os pontos. Estas letras são colocadas de forma que a letra A representa 1 (uma) observação, a letra B 2 (duas) observações superpostas, a letra C 3 (tres) observações e assim por diante até a letra Z que representa 26 ou mais observações superpostas.

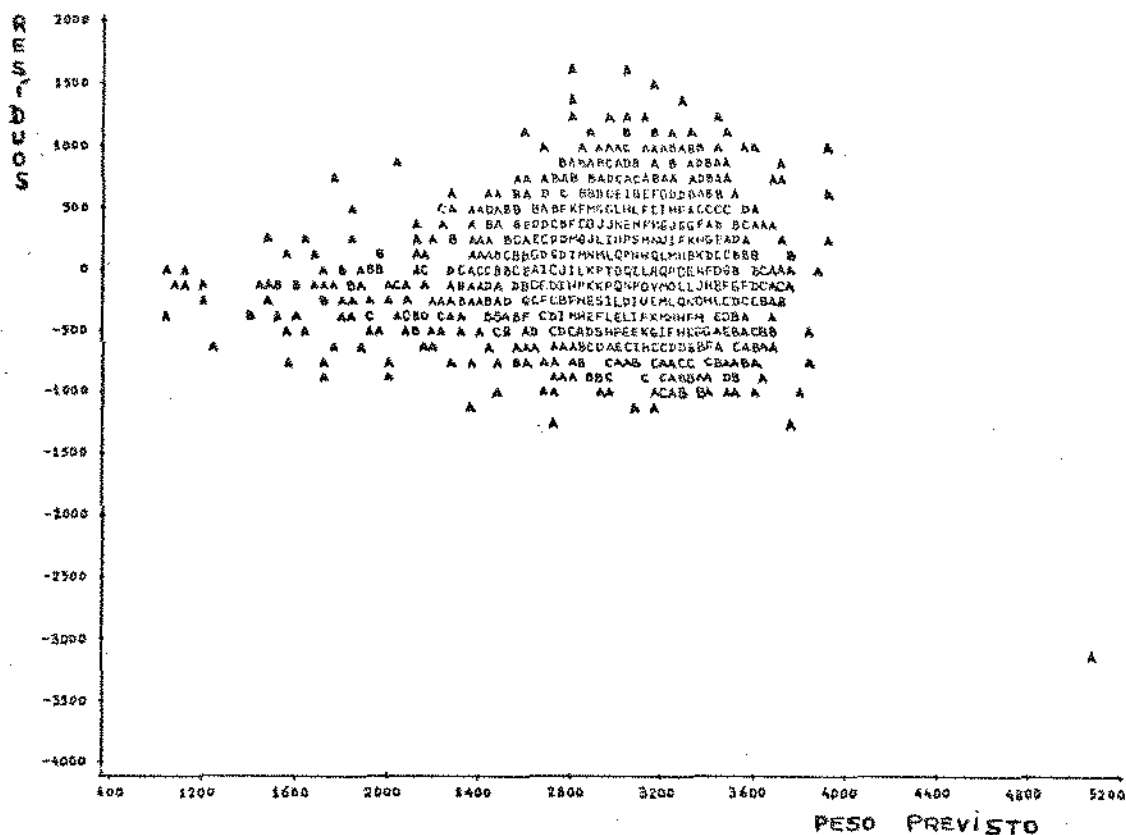


Figura 5.1 - Gráfico de RESÍDUOS versus PESO PREVISTO antes da
----- eliminação de outliers. (Fonte: FCM UNICAMP).

Através da variável ERNA, estado do recém-nascido no momento da alta, foram identificados os natimortos e foram eliminados aqueles com menos de 1000 gramas. Estas observações foram eliminadas porque os natimortos tem comportamento diferenciado dos recém-nascidos ou natimortos com mais de 1000 gramas com relação à variável que se está estudando, o peso do recém nascido. Segundo este critério foram eliminadas três observações.

O gráfico de RESÍDUOS versus PESO ESTIMADO com ajuste realizado após a eliminação dos outliers pode ser visto na figura 5.2. Comparando visualmente este gráfico de resíduos com o anterior pode-se dizer que este se assemelha mais ao que se espera dos resíduos numa regressão linear, ou seja, resíduos com distribuição normal.

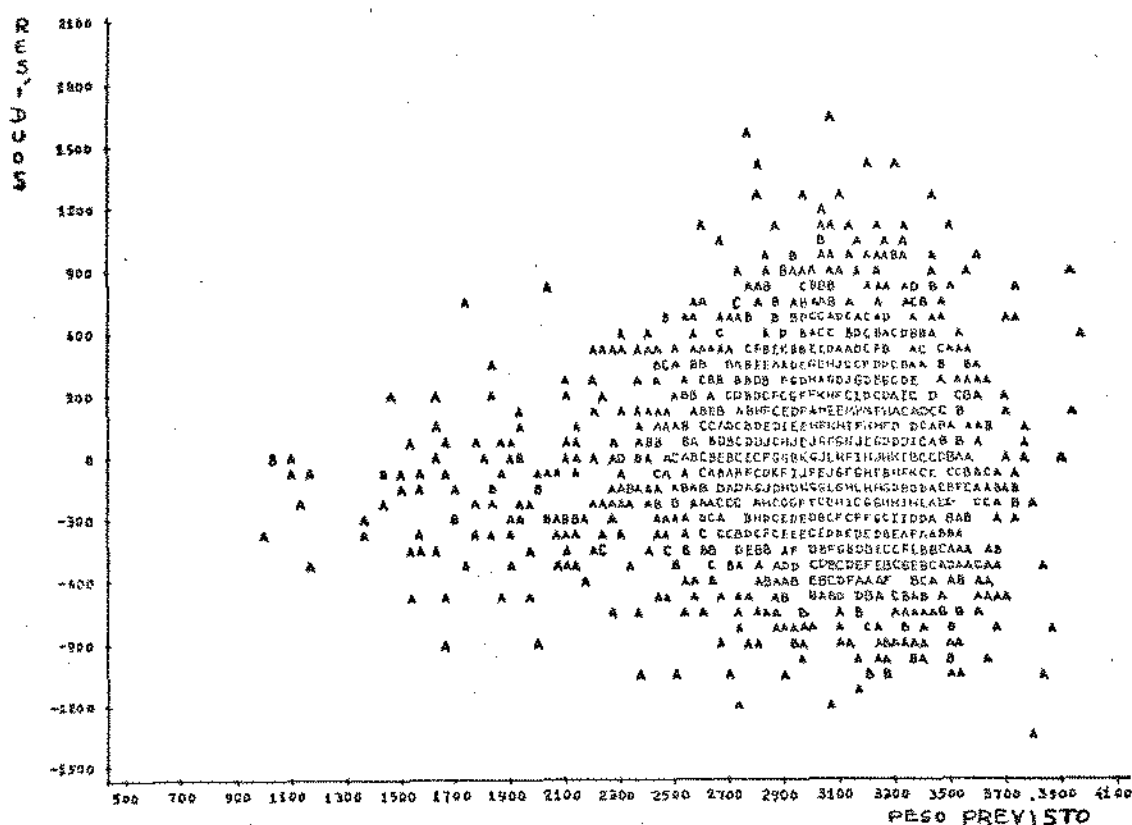


Figura 5.2 - Gráfico de RESÍDUOS versus PESO PREVISTO após a eliminação de outliers. (Fonte: FCM UNICAMP)

Nota-se também uma aparente dependência da variância dos resíduos com relação ao peso do recém-nascido de forma que a variância aumentaria à medida que o peso aumentasse. Isto não é verdade e esta aparência no gráfico deve-se ao fato de haver um número maior de observações para pesos entre 2700 e 3500 gramas e o número de observações diminui à medida que o peso decresce. Permanecendo válida ainda a suposição de que os resíduos tem variância constante.

Além destes foram detectados outros onze pontos, geralmente em crianças prematuras com baixo peso, que foram mantidos no ajuste após a verificação de que sua saída não alterava significativamente as estimativas obtidas.

Eliminados os outliers continuou-se a trabalhar as variáveis de forma a obter aquele que será o nosso primeiro modelo, o MODELO 0.

V.3 - MODELO 0

A seguir foi criada uma variável indicadora de natimortos, NTM1, a partir da variável NTM, número de natimortos, ou seja, $NTM1 = 0$ se $NTM = 0$ e $NTM1 = 1$ se $NTM \geq 1$. Construída desta forma a nova variável NTM1 será igual a 1 (um) se o histórico da gestante registrar algum natimorto e será igual a 0 (zero) se não houver nenhum indício de natimorto anterior.

Quando postas a concorrer num ajuste usando regressão "stepwise" as variáveis NTM e NTM1 juntamente com as demais, a variável indicadora de natimortos, NTM1 foi incluída no modelo em detrimento de NTM. Isto é uma indicação de que o fato de haver ou não natimorto(s) anterior(es) é mais importante na determinação do peso do recém-nascido do que o número de natimortos havidos anteriormente. Assim sendo, passou-se a trabalhar com NTM1 e não mais com NTM.

O mesmo procedimento foi adotado com relação às variáveis ABOR, número de abortos anteriores e FILMOR, número de filhos mortos anteriores; e em ambos os casos as variáveis selecionadas foram ABOR1, indicadora de abortos anteriores e FILMOR1, indicadora de filhos mortos anteriores e abandonou-se ABOR e FILMOR número de abortos e número de filhos mortos anteriores.

Durante os ajustes que foram feitos até aqui uma variável veio se destacando como de grande importância. Trata-se da variável DIAB, que é a indicadora de diabetes nas gestantes, sempre acrescentando algumas gramas à previsão de peso do recém-nascido no caso da gestante ser diabética, mas ela teve de ser retirada do ajuste por ser uma variável em que houve muita falta de informação, o que eliminaria do ajuste um número grande de observações.

Finalmente foi estabelecido o primeiro modelo para prever o peso de recém-nascidos, somente com o inconveniente de se ter um intercepto negativo. Causa estranheza interpretar um peso de -3000 gramas. Para eliminar este desconforto foram feitas algumas transformações que permitiram obter um intercepto positivo e aceitável em termos de ordem de grandeza.

A primeira delas é a seguinte:

$$IDM = IDM - 20$$

O uso desta transformação implica que a nova variável IDM valerá a quantidade de anos que a idade da gestante exceder a 20 anos, podendo assumir valores negativos caso esta tenha menos que 20 anos. Por exemplo, se a gestante tem 22 anos a variável IDM assume o valor 2, se tiver 19 assume o valor -1 e assim por diante.

Tratamento semelhante foi dado às variáveis ALTM, altura da gestante e IDG, idade gestacional de forma que:

$$ALTM = ALTM - 145$$

$$IDG = IDG - 30$$

Com o uso destas transformações conseguiu-se um intercepto em torno de 1400 gramas.

Após a detecção de "outliers" e todas estas transformações obteve-se um ajuste para o PESO denominado de MODELO 0 e cujos parâmetros estimados encontram-se na tabela 5.1.

Neste ajuste foram utilizadas 2382 observações válidas, obteve-se um R-quadrado de 0.4643 e no teste de significância do modelo um nível de significância p menor que 0.0001.

MODELO 0

VARIAVEIS	PARAMETROS ESTIMADOS	ERRO PADRÃO	t PARA TESTAR (Bi=0)	PROB > t
INTECEPTO	1399	40.2	34.8	0.0001
NTM1	-95.1	35.2	-2.7	0.0069
FILMO1	-64.4	24.5	-2.6	0.0086
TOTPAR	30.0	3.9	7.7	0.0001
ANTOBS1	-182	53.0	-3.4	0.0006
TOTCON	13.1	2.7	4.8	0.0001
ALTM	9.15	1.3	6.8	0.0001
EDEMA	125	17.9	7.0	0.0001
PRGR	-279	59.4	-4.7	0.0001
TOXE	-115	26.4	-4.4	0.0001
INF	-117	35.9	-3.3	0.0011
IDG	163	4.2	39.1	0.0001

Tabela 5.1 - Estimativas obtidas em regressão linear para o
MODELO 0 (Fonte: FCM UNICAMP)

Já tendo um instrumento para prever peso, pode-se agora analisá-lo em sua capacidade de detectar recém-nascidos com peso menor que 2500 gramas, ou baixo peso. Isto pode ser feito através das curvas de sensibilidade e especificidade.

V.4 - SENSIBILIDADE e ESPECIFICIDADE

Sendo um dos objetivos deste trabalho detectar "baixo peso", ou peso menor que 2500 gramas, em recém-nascidos e já tendo o MODELO 0 para prever o peso do recém-nascido. Pode-se tentar detectar os casos de baixo peso estabelecendo uma regra, por exemplo, considerando previsão de baixo peso todos os casos em que o peso previsto pelo modelo estiver abaixo de 2500 gramas. Este valor limite que no caso é 2500 gramas será chamado de "ponto de corte". Em função desta regra pode-se classificar cada observação como prevista como baixo peso ou peso normal e sabendo a sua real condição pode-se cruzar estas duas informações e montar a tabela 5.2.

		PREVISÃO		
		+-----+	+-----+	+-----+
		< 2500	>= 2500	TOTAL
REAL	< 2500	152	187	339
	>= 2500	33	2022	2055
	TOTAL	185	2209	2394

Tabela 5.2 - MODELO 0, ponto de corte 2500 gramas.
----- (Fonte: FCM UNICAMP)

Na tabela 5.2 observa-se 339 casos de baixo peso, 2055 de peso normal dando um total de 2394 observações. Destas 152 casos de baixo peso sendo corretamente previstos e 2022 pesos normais também sendo corretamente previstos, 33 pesos normais sendo previsto como baixo peso e 187 casos de baixo peso sendo previstos como normais.

Quando se deseja que o modelo detecte um número maior de casos de baixos pesos, basta alterar para maior o ponto de corte, considerando-se, por exemplo, como previsão de baixo peso as observações que apresentarem previsão de peso menor que 2600 gramas e não mais 2500 gramas como no caso anterior. O resultado do estabelecimento deste novo ponto de corte pode ser visto na tabela 5.3.

Comparando as duas tabelas nota-se que à medida que se aumenta o ponto de corte observa-se um aumento no número de previsões corretas de baixo peso e também um aumento nas previsões de baixo peso falsas, pesos normais que antes eram previstos corretamente agora são previstos como baixo peso.

		PREVISÃO		
		< 2600	>= 2600	TOTAL
REAL	< 2500	179	160	339
	>= 2500	64	1991	2055
	TOTAL	243	2151	2394

Tabela 5.3 - MODELO 0, ponto de corte 2600 gramas.
 ----- (Fonte: FCM UNICAMP)

Partindo de pontos de corte diferentes podem-se montar tabelas diferentes. A comparação destas tabelas é uma tarefa difícil. Por este motivo para se compararem diferentes pontos de corte e o desempenho de diferentes modelos definimos duas medidas que são a sensibilidade e a especificidade:

SENSIBILIDADE: fração de observações com baixo peso que foram previstos corretamente

$$\text{SENSIBILIDADE} = \frac{\text{nº de obs. com baixo peso previstas corretamente}}{\text{número total de observações com baixo peso}}$$

ESPECIFICIDADE: fração de observações com peso normal que foram previstas corretamente.

$$\text{ESPECIFICIDADE} = \frac{\text{nº de obs. com peso normal previstas corretamente}}{\text{número total de obs. com peso normal}}$$

Pode-se então calcular estas medidas para a tabela 5.2:

SENSIBILIDADE = 0.448 (baixo peso previsto corretamente)

ESPECIFICIDADE = 0.984 (peso normal previsto corretamente)

Uma alternativa de apresentação gráfica que pode ser utilizada é escolher vários pontos de corte, calcular a sensibilidade e especificidade nestes pontos e colocar os valores obtidos num gráfico. Como a sensibilidade e especificidade são funções aproximadamente contínuas do ponto de corte, escolhendo estrategicamente

os pontos e ligando-os por uma linha reta pode-se ter uma boa aproximação das mesmas.

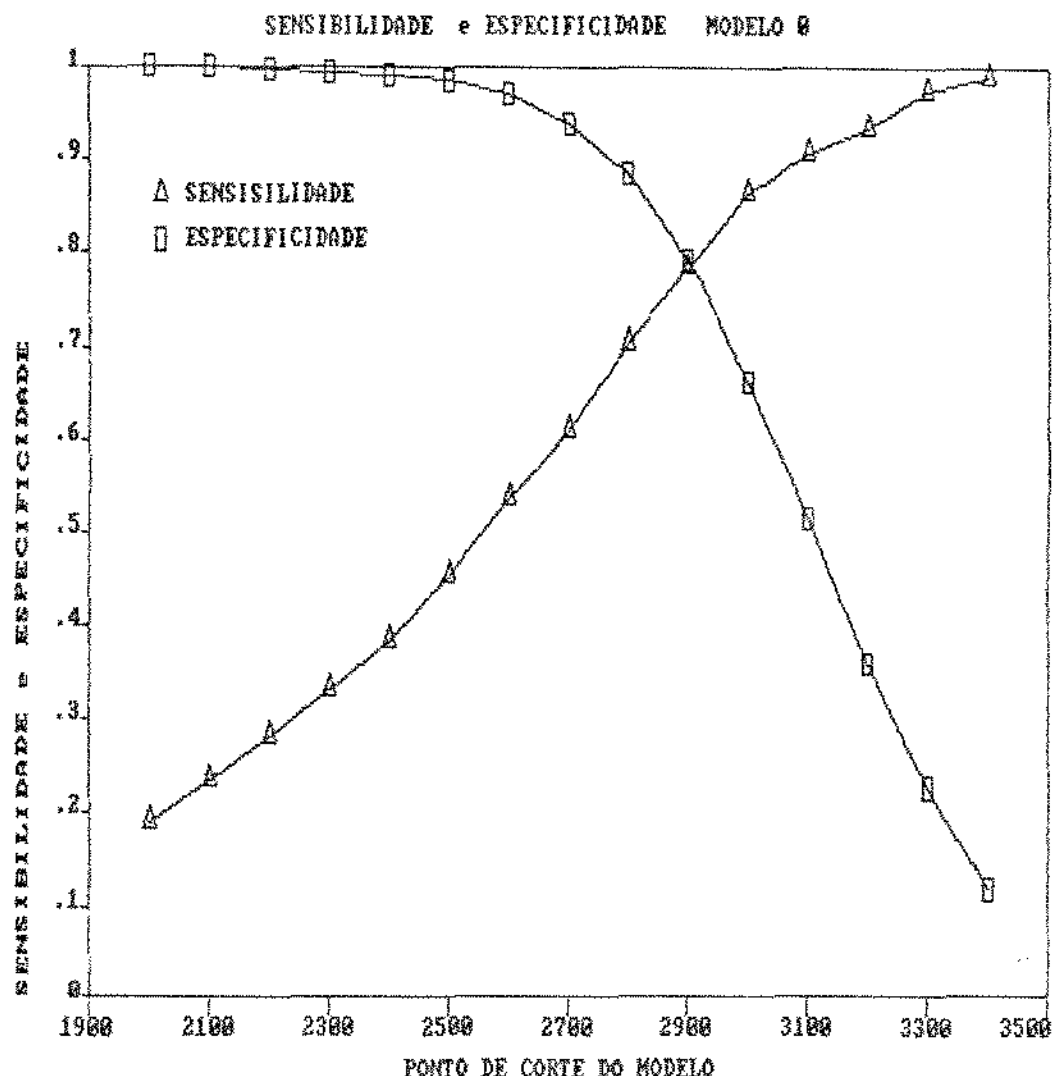


Figura 5.3 - Sensibilidade e especificidade para o MODELO 0.
----- (Fonte: FCM UNICAMP)

Como se observa no gráfico da figura 5.3 a medida que se aumenta o ponto de corte, também aumenta a sensibilidade ou o percentual de baixo peso detectados e em consequência diminui a especificidade, o que equivale a aumentar o número de previsões de baixo peso falsas.

Neste trabalho os pontos de corte foram escolhidos entre 2000 e 3400 gramas com intervalos de 100 gramas para todos os gráficos de sensibilidade e especificidade.

V.5 - MODELO 1

Uma variável altamente correlacionada com peso do recém-nascido é a sua altura. Mas esta não pode ser usada diretamente no modelo já que esta informação somente está disponível no momento do parto, quando o peso já é conhecido. Como a variável altura está disponível no arquivo, ajusta-se um modelo para prever a altura utilizando as variáveis disponíveis antes do parto. As previsões fornecidas por este modelo geram uma nova variável, a altura prevista, que poderia integrar o conjunto de variáveis do modelo. Esta variável é então anexada ao conjunto de variáveis explicativas do MODELO 0 e os resultados obtidos com este novo modelo comparados aos do MODELO 0.

O ajuste para a altura do recém-nascido, seguiu o mesmo esquema usado no ajuste do peso. As variáveis que concorreram na formação do modelo para a altura foram as mesmas que concorreram na modelagem do peso no MODELO 0. Utilizando regressão "stepwise" para seleção de variáveis, o conjunto de variáveis escolhido para explicar a altura foi o seguinte: ANTOBS1, TOTCON, ALTM, EDEMA, PRGR, INF, IDG, ESTCIV e ANTMOR. Apesar do conjunto de variáveis explicativas da altura ter variáveis comuns ao conjunto das explicativas do MODELO 0, não é um subconjunto do mesmo. Isto é importante pois se fosse um subconjunto, ALTP seria uma combinação linear das variáveis do MODELO 0 e as previsões do PESO no novo modelo seriam idênticas às do MODELO 0.

MODELO 1

VARIÁVEIS	PAR. EST. PESO (INICIAL)	PAR. EST. ALTURA ALTP1	PAR. ESTIMADOS PESO COM ALTURA PREVISTA
INTECEPTO	1399	41.6	-2905
NTM1	-95.1	x	-89.0
FILMO1	-64.4	x	-62.4
TOTPAR	30.0	x	27.8
ANTOBS1	-182	-0.915	-87.7
TOTCON	13.1	0.038	8.07
ALTM	9.15	0.050	4.14
EDEMA	125	0.372	90.3
PRGR	-279	-0.996	-170
TOXE	-115	x	-101
INF	-117	-0.576	-58.2
IDG	163	0.712	89.4
ESTCIV	x	-0.269	x
ANTMOR	x	-0.320	x
ALTP1	x	x	104

Tabela 5.4 - Parâmetros dos ajustes para peso, altura, e peso com altura prevista, para o MODELO 1. (Fonte: FCM UNICAMP)

A tabela 5.4 mostra os parâmetros estimados obtidos nos ajustes para o PESO, para a altura e para o PESO com a variável ALTP1.

No ajuste da altura foram utilizadas 2267 observações válidas, foi obtido um R-quadrado de 0.3423 e no teste de significância do modelo um nível de significância p menor que 0.0001. No ajuste para o PESO com ALTP1 foram utilizadas 2340 observações válidas, foi obtido um R-quadrado de 0.4670 e no teste de significância do modelo um nível de significância p menor que 0.0001.

A figura 5.4 apresenta o gráfico da sensibilidade e especificidade para alguns pontos de corte no MODELO 1. Observando o gráfico nota-se uma grande semelhança com o mesmo gráfico para o MODELO 0.

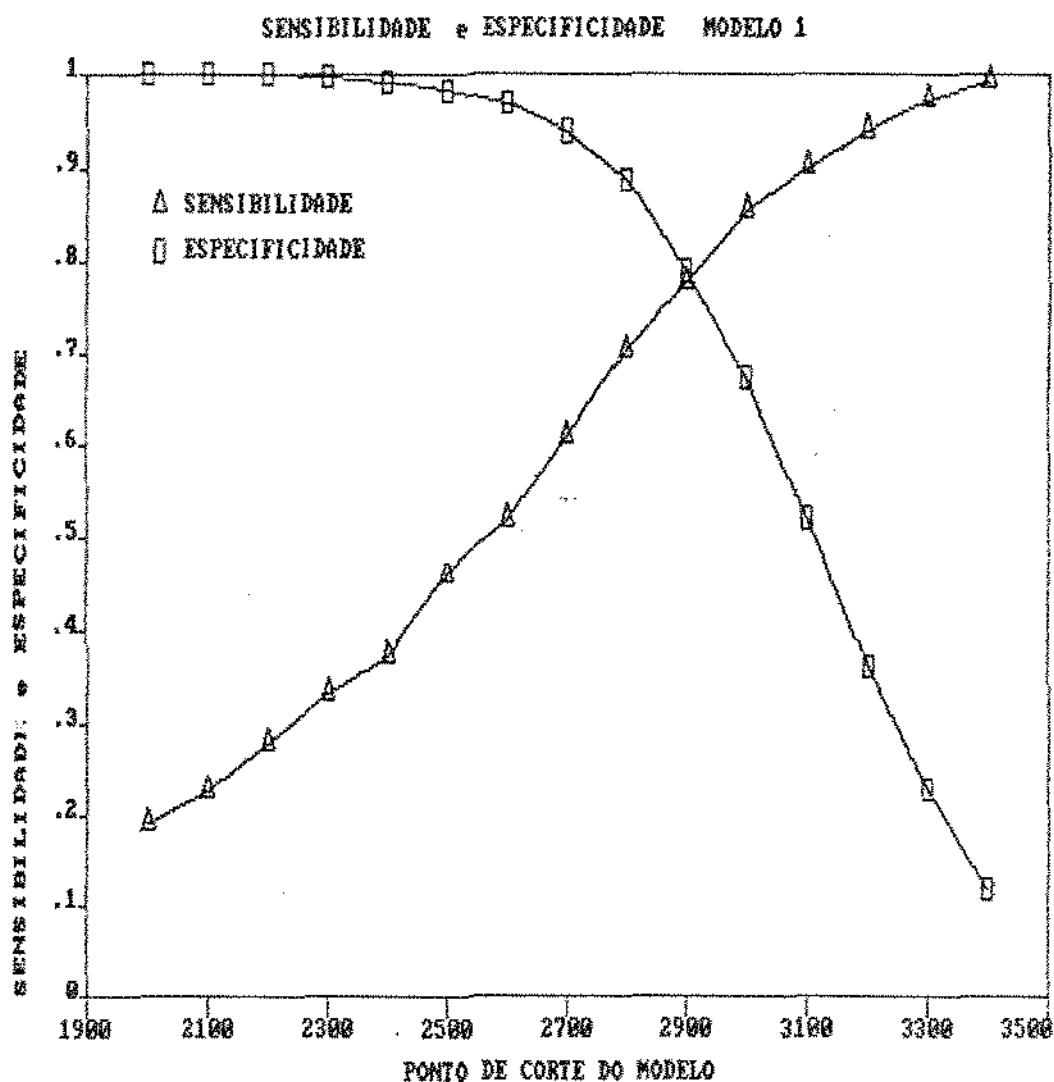


Figura 5.4 - Sensibilidade e especificidade para o MODELO 1.
 ----- (Fonte: FCM UNICAMP)

V.6 - COMPARAÇÃO DE MODELOS

Tendo disponíveis dois modelos, MODELO 0 e MODELO 1, e em cada um deles a possibilidade de escolher diferentes pontos de corte, torna-se necessário então de um método que permita compará-los em sua capacidade para detectar recém-nascidos com baixo peso.

Uma forma de se comparar modelos com relação à capacidade de detectar baixo peso em recém-nascidos é através de gráficos.

Utilizando as medidas definidas anteriormente, sensibilidade e especificidade, pode-se afirmar que o melhor modelo é aquele que maximiza conjuntamente as duas. Como os valores destas medidas se alteram a medida que varia o ponto de corte, a comparação dos modelos deverá levar em conta este fator, já que é possível que em alguns pontos de corte um determinado modelo possa ser o melhor e não ser em outros.

Uma alternativa para compará-los é colocar em um gráfico a sensibilidade e a especificidade de um mesmo ponto de corte como ordenada e abscissa de um mesmo ponto no gráfico e ligar os pontos na sequência crescente dos pontos de corte através de linhas retas para o modelo de referência. Posteriormente sobre o mesmo gráfico repete-se o processo para o modelo que se quer comparar. O melhor entre dois modelos será aquele cuja curva resultante da união dos pontos estiver mais acima e a direita no gráfico.

Nesta primeira comparação e em todas as outras que venham a ser feitas neste trabalho o MODELO 0 será considerado como referência para comparação.

Observando a figura 5.5 nota-se alguma dificuldade em comparar as curvas referentes ao MODELO 0 e ao MODELO 1, isto se deve principalmente a forma da curva resultante, a de um L invertido. Numa tentativa de linearizar as curvas obtidas e facilitar a comparação delas aplicou-se aos pontos de cada curva a seguinte transformação:

$$\text{SENSIBILIDADE} = P$$

$$\text{ESPECIFICIDADE} = Q$$

$$\text{LOGITO } P = \ln (P / (1 - P))$$

$$\text{LOGITO } Q = \ln (Q / (1 - Q))$$

Trata-se da transformação logística.

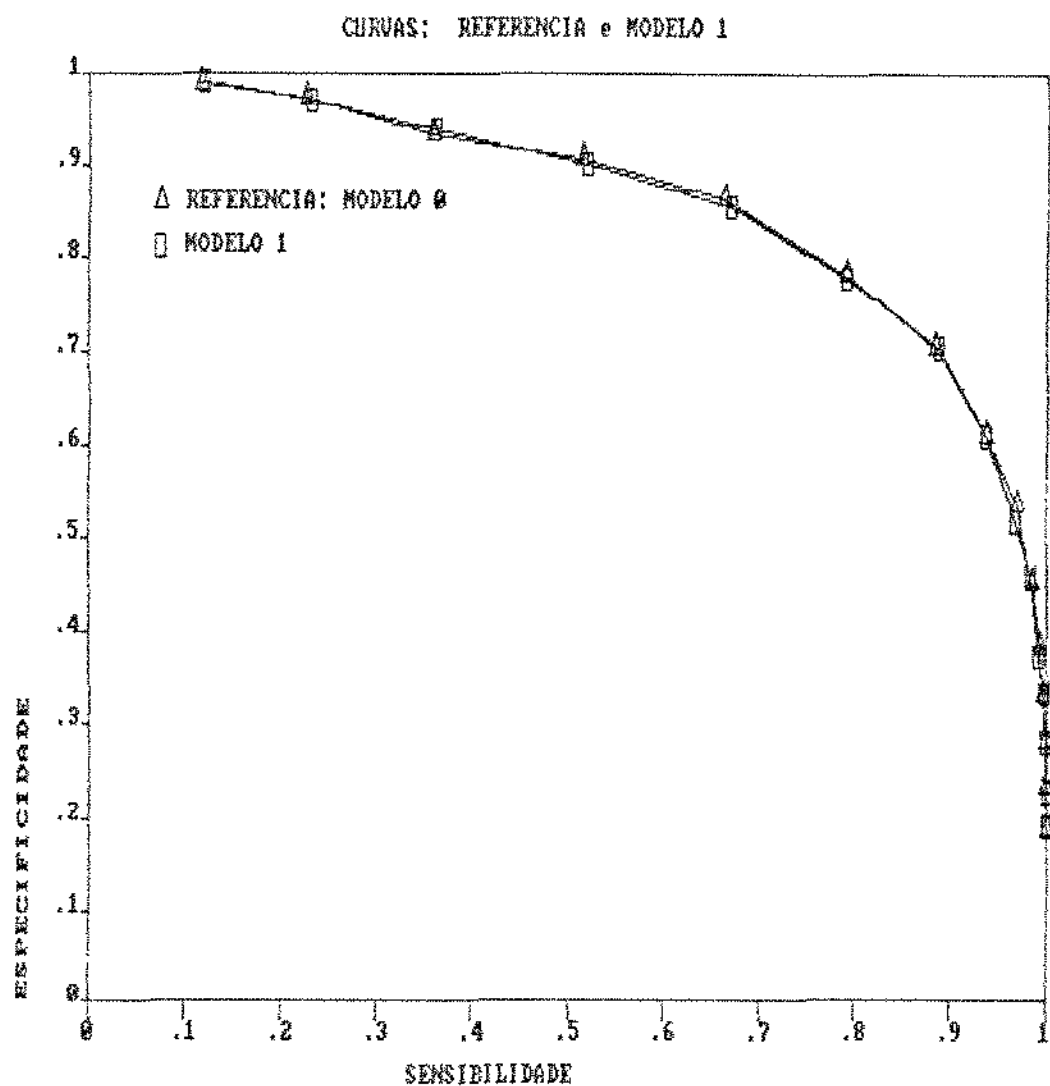


Figura 5.5 - Sensibilidade versus especificidade para MODELO 0 e
----- MODELO 1. (Fonte: FCM UNICAMP)

Na figura 5.6 nota-se uma forte linearização em relação as curvas obtidas antes da transformação na figura 5.5, isto permite comparar mais confortavelmente os dois modelos.

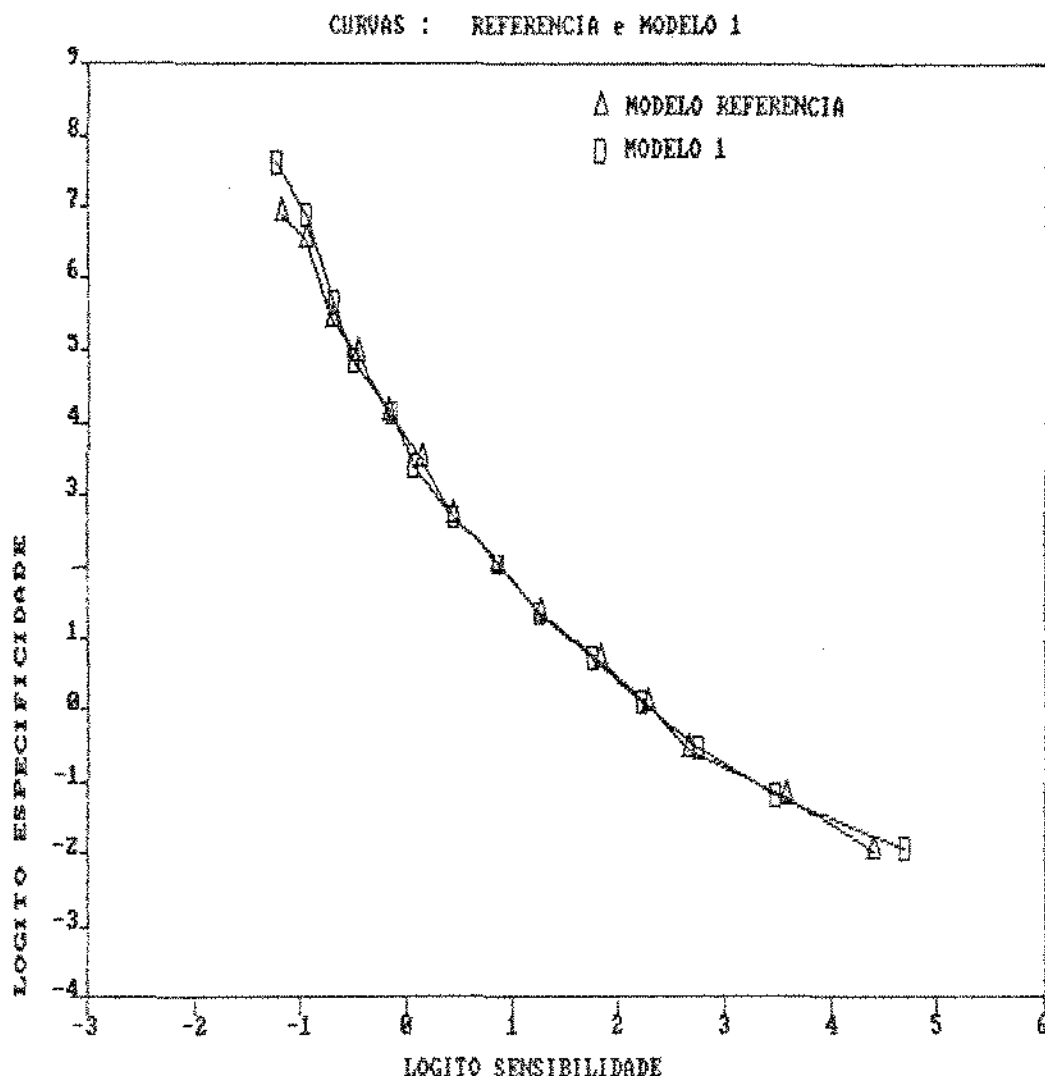


Figura 5.6 - LOGITO P versus LOGITO Q para o MODELO 0 e
 ----- MODELO 1. (Fonte: FCM UNICAMP)

Comparando o desempenho dos dois modelos o que se pode notar é que as duas curvas estão praticamente superpostas, o que indica que praticamente não houve melhora no MODELO 1 com relação ao MODELO 0 no que se refere a detecção de baixo peso. Para que houvesse diferença entre os modelos as curvas deveriam estar separadas pelo menos parcialmente e modelo melhor seria aquele cuja curva estivesse mais acima e mais a direita.

V.7 - MODELO 2

O ajuste do MODELO 2 é similar ao MODELO 1. Neste como no anterior o conjunto das variáveis explicativas é formado pelo conjunto das variáveis explicativas do MODELO 0 mais a altura prevista, neste caso a variável ALTP2.

A diferença está no ajuste que vai fornecer a altura prevista. No MODELO 1 foram usadas todas as observações disponíveis para o ajuste de ALTP1. Já no MODELO 2 usou-se somente as observações que apresentaram resíduos altos negativos no ajuste do peso, aquelas cujo resíduo "studentizado" externamente é menor que -1, para a regressão que fornece um ajuste para ALTP2.

Restringindo desta forma as observações, utilizaria-se uma possível contribuição da variável ALTP2 para os piores erros de previsão cometidos no ajuste inicial para o peso.

Isso é razoável considerando-se que quando um resíduo "studentizado" externamente é menor que -1, o erro na previsão é alto em relação aos demais erros e sendo negativo este erro, a previsão errará para mais, o que neste trabalho é o pior caso. Quer dizer que pode-se estar prevendo peso normal e obtendo ao final da gravidez um caso de baixo peso.

MODELO 2

VARIÁVEIS	PAR. EST. PESO (INICIAL)	PAR. EST. ALTURA ALTP2	PAR. ESTIMADOS PESO COM ALTURA PREVISTA
INTERCEPTO	1399	38.0	3217
NTM1	-95.1	x	-97.4
FILMO1	-64.4	x	-65.4
TOTPAR	30.0	x	32.8
ANTOBS1	-182	x	-180
TOTCON	13.1	0.085	17.7
ALTM	9.15	0.026	10.6
EDEMA	125	0.553	152
PRGR	-279	x	-279
TOXE	-115	-0.815	-152
INF	-117	-0.846	-160
IDG	163	0.841	203
IDM	x	0.031	x
SIF1	x	1.478	x
ALTP2	x	x	-47.8

Tabela 5.5 - Parâmetros dos ajustes para peso, altura e peso com altura prevista, MODELO 2. (Fonte: FCM UNICAMP)

Além disso é de se esperar um comportamento diferente deste modelo para a altura com relação ao anterior já que trabalha-se com as observações problemáticas com relação ao peso.

Este comportamento diferente é o que pode ser constatado na tabela 5.5 se comparada com a tabela 5.4. O conjunto de variáveis explicativas para ALTP2 é diferente do obtido no MODELO 1.

Neste modelo, no ajuste da altura foram utilizadas 399 observações válidas, o R-quadrado foi de 0.4903 e no teste do modelo o nível de significância p obtido foi menor que 0.0001. No ajuste do PESO com ALTP2 foram utilizadas 2393 observações válidas, o R-quadrado foi de 0.4655 e no teste do modelo o nível de significância p foi menor que 0.0001.

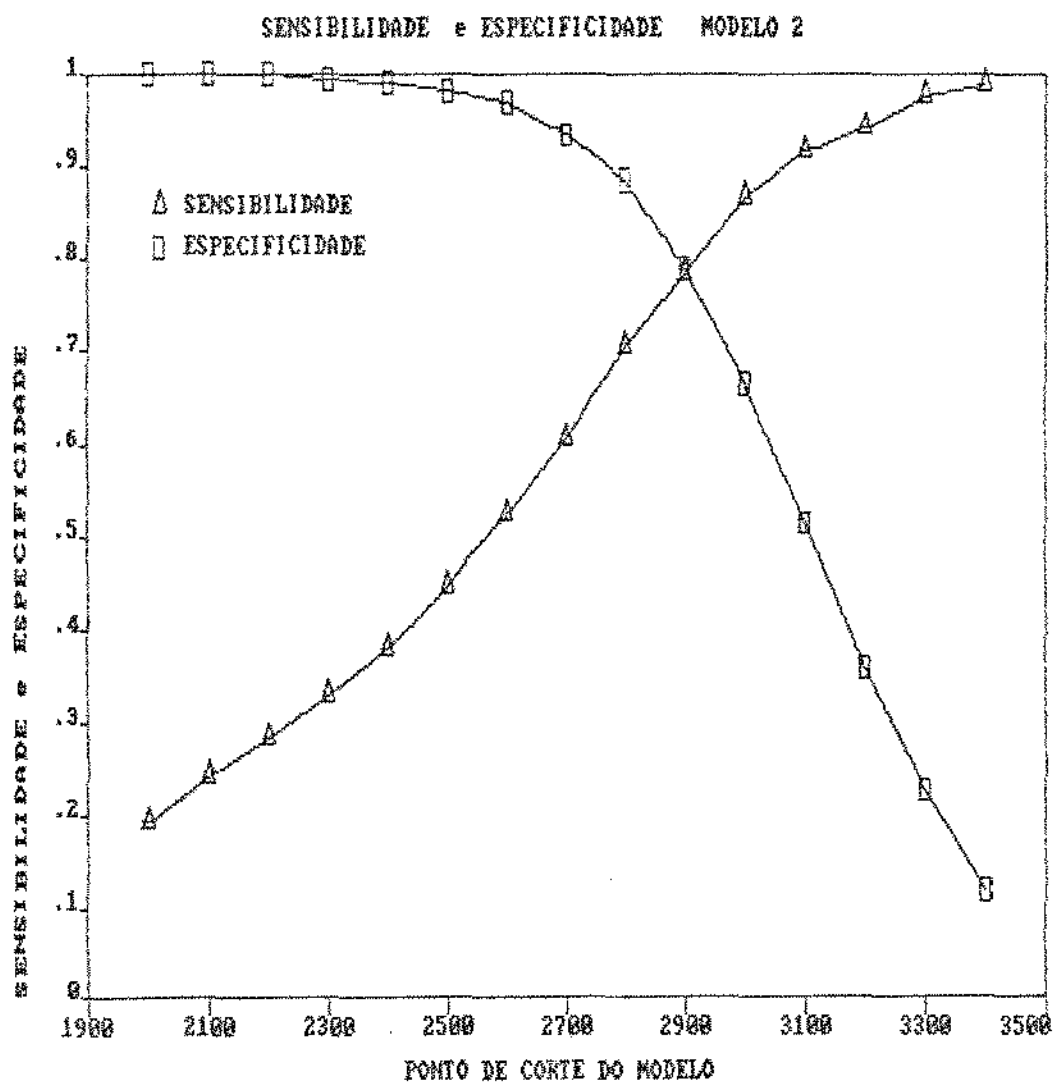


Figura 5.7 - Sensibilidade e especificidade para o MODELO 2.
(Fonte: FCM UNICAMP)

A figura 5.7 mostra um gráfico com as curvas da sensibilidade e da especificidade que são bastante semelhantes aos mesmos gráficos do MODELO 0.

A comparação dos MODELOS 0 e 2 pode ser feita através da figura 5.8, onde se pode notar que praticamente não há diferença de desempenho entre ambos.

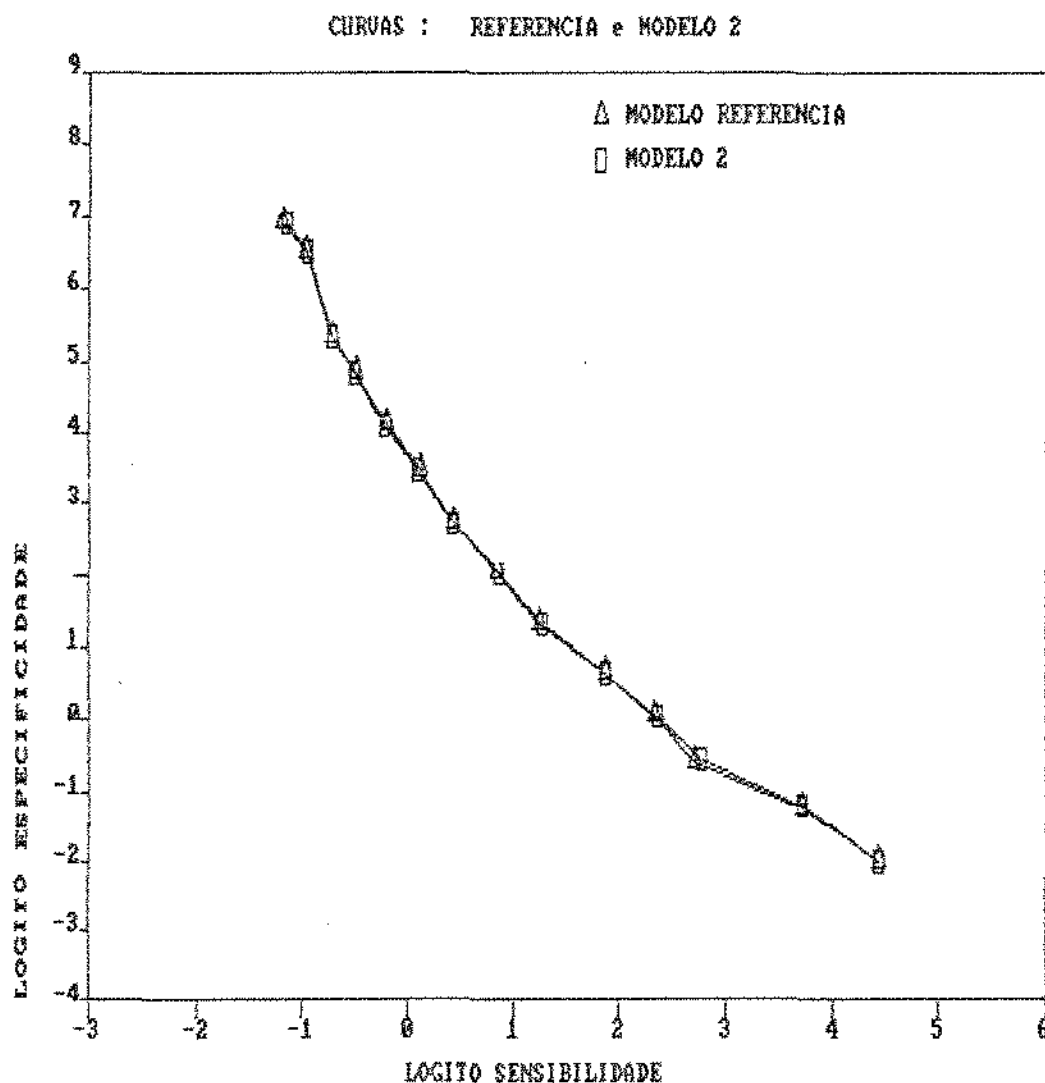


Figura 5.8 - LOGITO P versus LOGITO Q para MODELOS 0 e 2.
----- (Fonte: FCM UNICAMP)

V.8 - MODELO 3

No MODELO 3 foi utilizado inicialmente o ajuste feito para o MODELO 0 para selecionar as observações com resíduos altos negativos. A seguir ajustou-se a altura prevista usando apenas estas observações, como no MODELO 2.

A diferença deste com relação ao anterior é que a nova variável, ALTP2, não foi apenas anexada ao conjunto de variáveis explicativas do MODELO 0, mas a regressão stepwise foi repetida com todas as variáveis do início mais esta nova variável.

Mais uma vez o conjunto de variáveis explicativas do peso do recém-nascido foi diferente dos obtidos até agora, como pode ser constatado na tabela 5.6.

MODELO 3

VARIÁVEIS	PAR. EST. PESO (INICIAL)	PAR. EST. ALTURA ALTP2	PAR. ESTIMADOS PESO COM ALTURA PREVISTA
INTERCEPTO	1399	38.0	-5923
NTM1	-95.1	x	-98.4
FILMO1	-64.4	x	-60.9
TOTPAR	30.0	x	32.9
ANTOBS1	-182	x	-177
TOTCON	13.1	0.085	x
ALTM	9.15	0.026	4.31
ABOR1	x	x	-45.7
EDEMA	125	0.553	x
PRGR	-279	x	-281
TOXE	-115	-0.815	47.8
INF	-117	-0.846	x
IDG	163	0.841	x
IDM	x	-0.031	-6.37
SIFI	x	1.478	-374
ALTP2	x	x	193

Tabela 5.6 - Parâmetros dos ajustes para peso, altura e peso com altura prevista, MODELO 3. (Fonte: FCM UNICAMP)

No MODELO 3, no ajuste de PESO com ALTP2 foram utilizadas 2393 observações válidas, o R-quadrado foi de 0.4659 e no teste do modelo o nível de significância p foi menor que 0.0001.

O gráfico da sensibilidade e da especificidade versus ponto de corte, na figura 5.9, é bastante semelhante ao mesmo gráfico do MODELO 0.

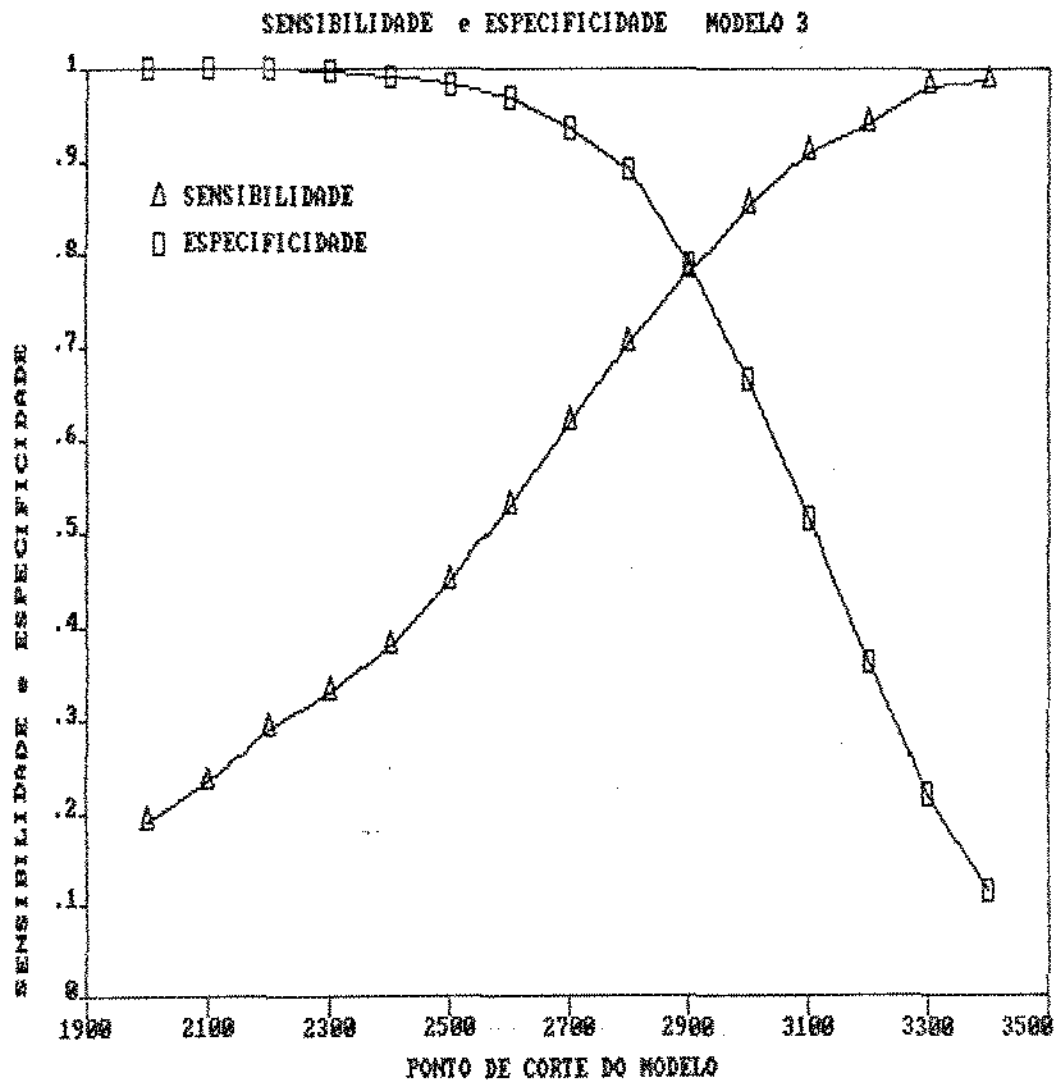


Figura 5.9 - Sensibilidade e especificidade para o MODELO 3.
----- (Fonte: FCM UNICAMP)

A figura 5.10 apresenta a comparação do desempenho dos MODELOS 0 e 3, onde pode ser constatado que as previsões de ambos também são bastante semelhantes.

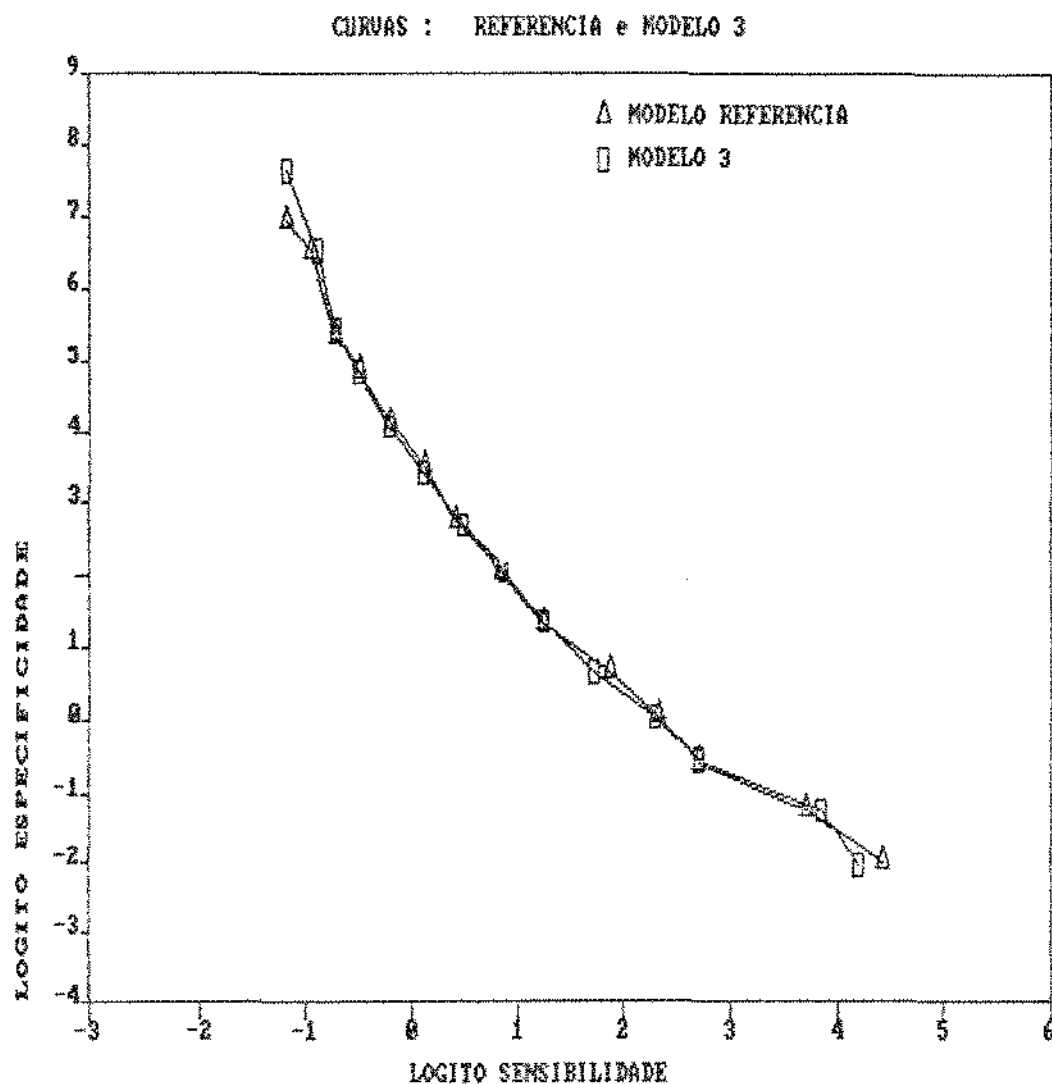


Figura 5.10 - LOGITO P versus LOGITO Q para MODELOS 0 e 3.
 ----- (Fonte: FCM UNICAMP)

V.9 - MODELOS 4 5 e 6

Nos MODELOS 2 e 3 utilizou-se a altura para auxiliar a previsão de baixo peso nos casos em que os erros eram grandes em relação aos demais erros de forma que a previsão era maior que o valor real que se desejava prever.

Seguindo a linha de interesse que é minimizar o número de previsões "falsas negativas", as que prevêm peso normal quando na realidade trata-se de um caso de baixo peso. Alterou-se o critério utilizado nos MODELOS 2 e 3 para a obtenção da altura prevista, de forma que as observações selecionadas para o ajuste da altura sejam somente aquelas em que houveram previsões falsas negativas no MODELO 0, não importando a magnitude dos erros.

Repetiu-se então os ajustes para a altura, utilizando este novo critério de seleção de observações e usando-se os mesmos conjuntos de variáveis usados nos MODELOS 1, 2 e 3. Com estas novas variáveis, alturas previstas, anexadas ao conjunto de variáveis explicativas do MODELO 0, uma por vez, foram obtidos os MODELOS 4, 5 e 6.

MODELO 4

VARIÁVEIS	PAR. EST. PESO (INICIAL)	PAR. EST. ALTURA ALTP4	PAR. ESTIMADOS PESO COM ALTURA PREVISTA
INTERCEPTO	1399	42.9	-3304
NTM1	-95.1	x	-88.9
FILMO1	-64.4	x	-63.2
TOTPAR	30.0	x	29.4
ANTOBS1	-182	0.23	-206
TOTCON	13.1	0.057	6.32
ALTM	9.15	0.020	7.17
EDEMA	125	0.059	122
PRGR	-279	-0.0044	-271
TOXE	-115	x	-98.8
INF	-117	-0.33	-82.6
IDG	163	0.28	132
ESTCIV	x	0.013	x
ANTHOR	x	-0.36	x
ALTP4	x	x	110

Tabela 5.7 - Parâmetros dos ajustes para peso, altura e peso com altura prevista, MODELO 4.(F; FCM UNICAMP)

No MODELO 4 usou-se o mesmo conjunto de variáveis usado no ajuste do MODELO 1, diferenciando-se apenas o ajuste da altura onde foram utilizadas somente as observações que no ajuste inicial do PESO apresentaram uma previsão de peso maior que 2500 gramas quando na realidade o peso era menor que este valor.

No ajuste da altura foram utilizadas 197 observações válidas, o R-quadrado foi de 0.0713 e no teste do modelo o nível de significância p obtido foi 0.1189. No ajuste do PESO com a variável ALTP4 foram utilizadas 2341 observações válidas, o R-quadrado foi de 0.4667 e no teste do modelo o nível de significância p foi menor que 0.0001.

Na figura 5.11 observa-se um gráfico com as curvas da sensibilidade e da especificidade.

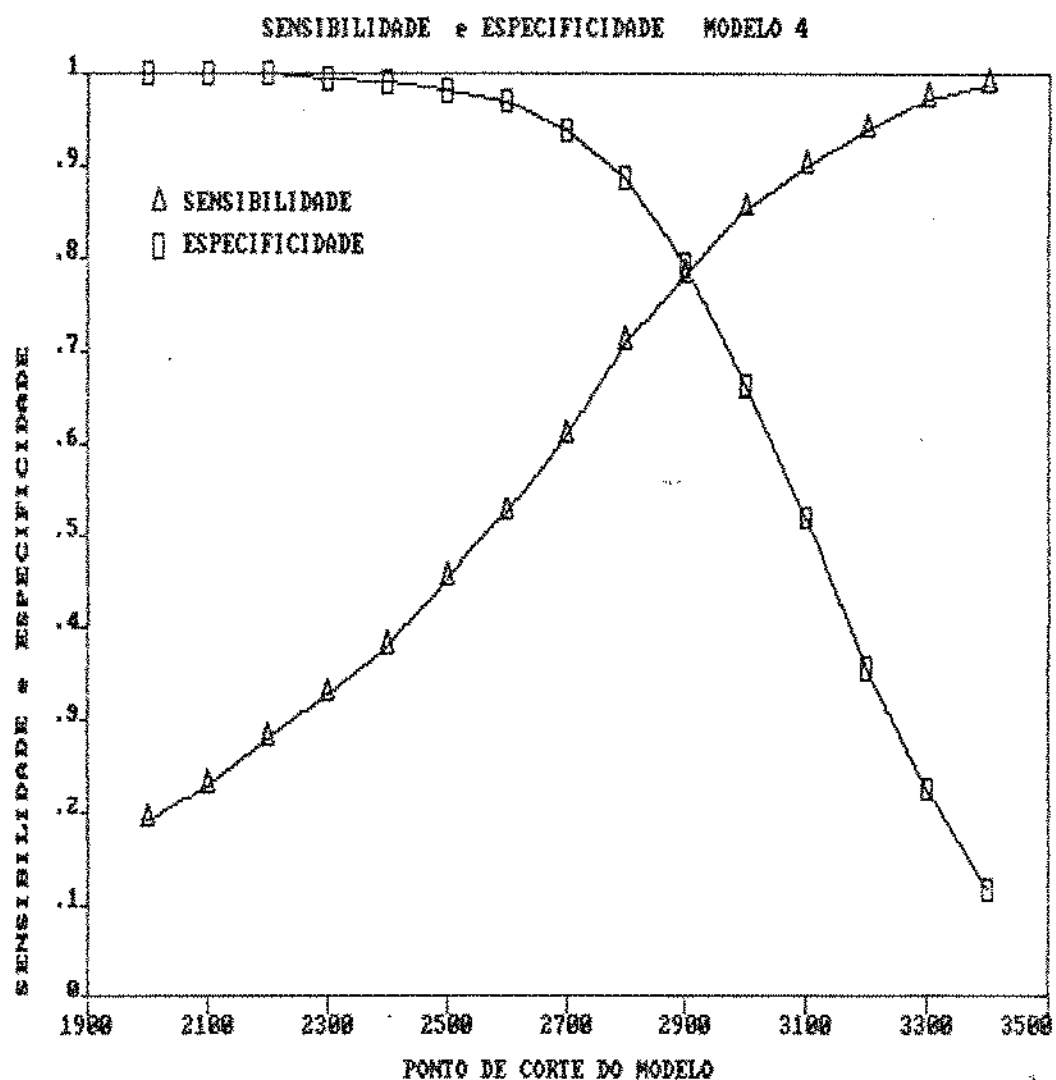


Figura 5.11 - Sensibilidade e especificidade para o MODELO 4.
 ----- (Fonte: FCM UNICAMP)

A comparação dos MODELO 0 e 4 pode ser feita através da figura 5.12, onde se pode notar que praticamente não há diferença de desempenho entre ambos.

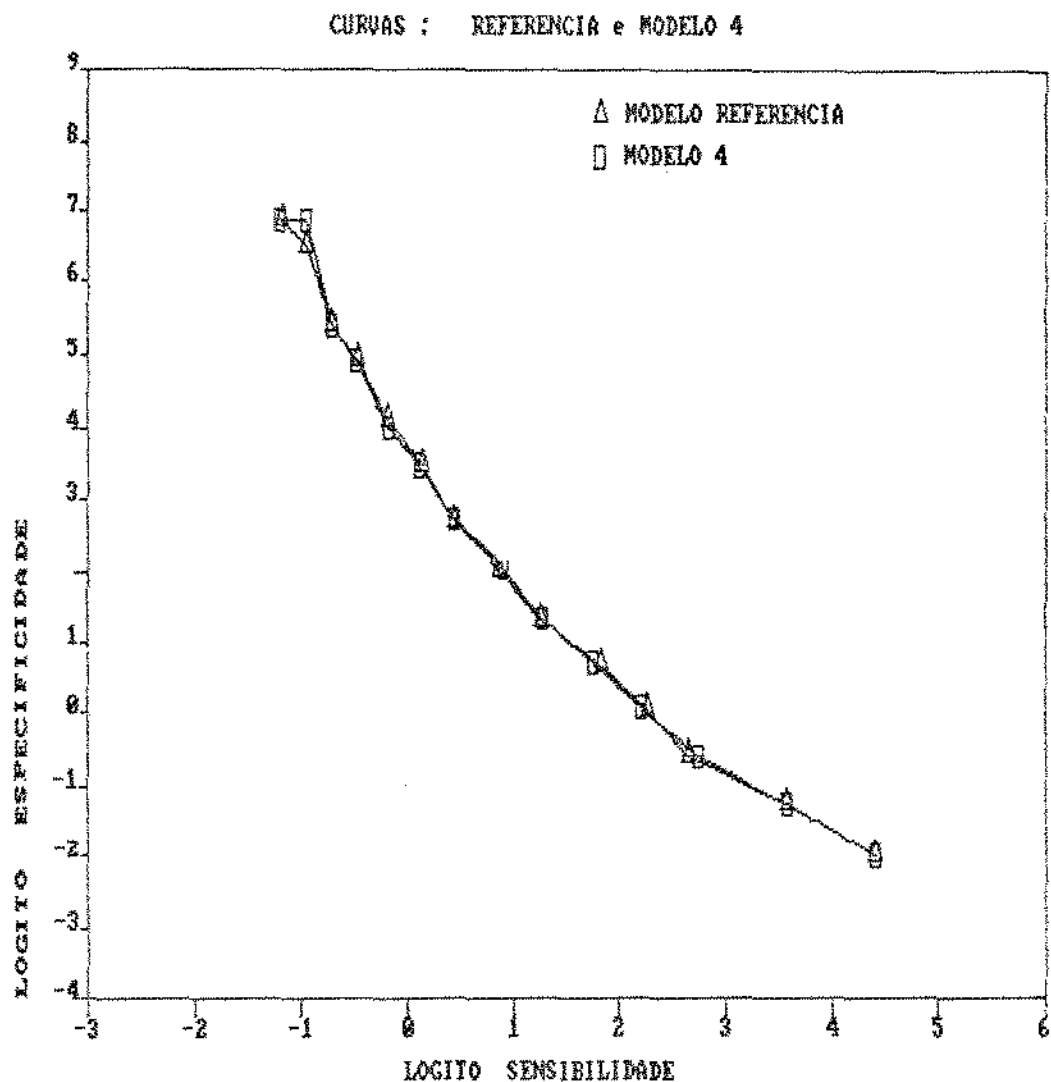


Figura 5.12 - LOGITO P versus LOGITO Q para os MODELOS 0 e 4.
 ----- (Fonte: FCM UNICAMP)

No MODELO 5 repete-se o conjunto de variáveis do MODELO 2, somente alterando-se o critério de seleção das observações que estarão presentes no ajuste da altura, ou seja, somente aquelas com previsões falsas negativas no modelo inicial serão selecionadas.

MODELO 5

VARIÁVEIS	PAR. EST. PESO (INICIAL)	PAR. EST. ALTURA ALTP5	PAR. ESTIMADOS PESO COM ALTURA PREVISTA
INTERCEPTO	1399	42.7	6547
NTM1	-95.1	x	-95.9
FILMO1	-64.4	x	-63.5
TOTPAR	30.0	x	30.7
ANTOBS1	-182	x	-180
TOTCON	13.1	0.035	17.5
ALTM	9.15	0.019	11.6
EDEMA	125	0.242	153
PRGR	-279	x	-281
TOXE	-115	-0.714	-200
INF	-117	-0.275	-153
IDG	163	0.319	202
IDM	x	0.00374	x
SIFI	x	0.743	x
ALTP5	x	x	-120

Tabela 5.8 - Parâmetros dos ajustes para peso, altura e peso com altura prevista, MODELO 5.(F: FCM UNICAMP)

Neste modelo, no ajuste da altura foram utilizadas 205 observações válidas, o R-quadrado foi de 0.0925 e no teste do modelo o nível de significância p obtido foi 0.0133. No ajuste do PESO com ALTP5 foram utilizadas 2393 observações válidas, o R-quadrado foi de 0.4656 e no teste do modelo o nível de significância p foi menor que 0.0001.

Na figura 5.13 observa-se no gráfico as curvas da sensibilidade e da especificidade que são bastante semelhantes aos mesmos gráficos do MODELO 0.

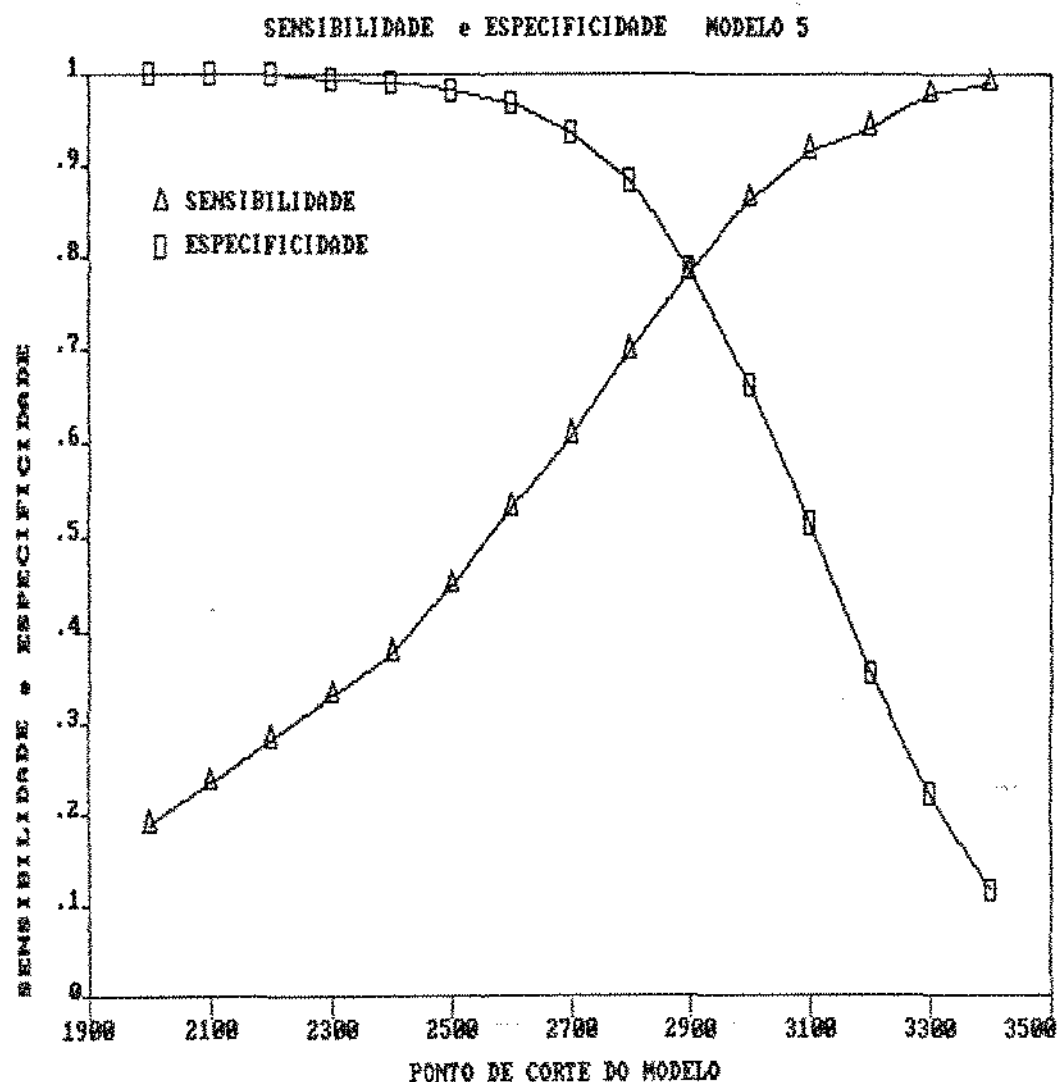


Figura 5.13 - Sensibilidade e especificidade para o MODELO 5
 ----- (Fonte: FCM UNICAMP)

A comparação de desempenho dos MODELOS 0 e 5 pode ser feita através da figura 5.14, onde se pode notar que praticamente não há diferença entre ambos.

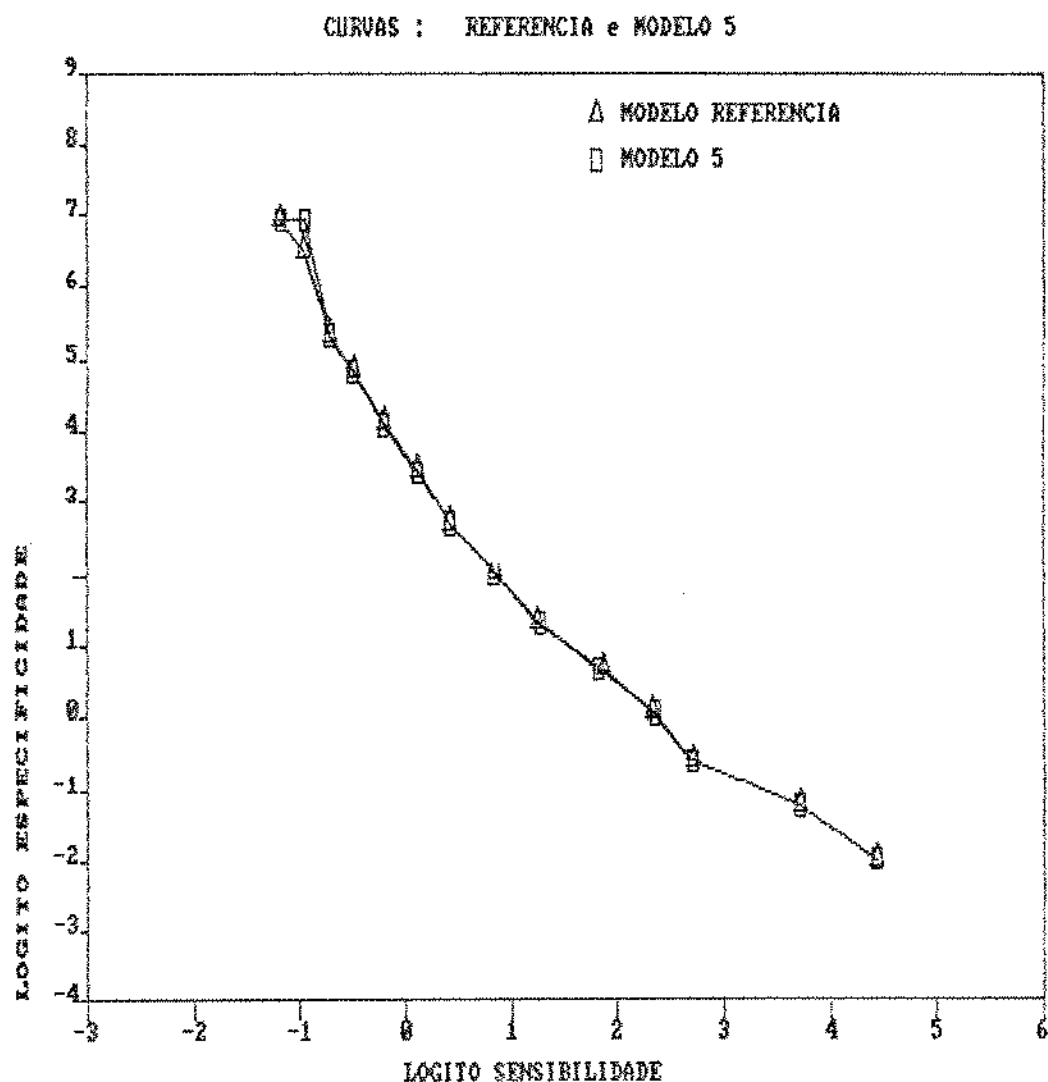


Figura 5.14 - LOGITO P versus LOGITO Q para MODELOS 0 e 5.
 ----- (Fonte: FCM UNICAMP)

No MODELO 6 ocorre a repetição do conjunto de variáveis do MODELO 3, sendo que o ajuste da altura foi feito usando somente as observações cujas previsões no MODELO 0 foram falsas negativas.

MODELO 6

VARIÁVEIS	PAR. EST. PESO (INICIAL)	PAR. EST. ALTURA ALTP5	PAR. ESTIMADOS PESO COM ALTURA PREVISTA
INTERCEPTO	1399	42.7	-20181
NTM1	-95.1	x	-99.3
FILMO1	-64.4	x	-61.2
TOTPAR	30.0	x	33.4
ANTOBS1	-182	x	-179
TOTCON	13.1	0.035	x
ALTM	9.15	0.019	-0.221
ABOR1	x	x	-47.8
EDEMA	125	0.242	x
PRGR	-279	x	-284
TOXE	-115	-0.714	248
INF	-117	-0.275	x
IDG	163	0.319	x
IDM	x	0.0037	-2.5
SIFI	x	0.743	-463
ALTP5	x	x	506

Tabela 5.9 - Parâmetros dos ajustes para peso, altura e peso com altura prevista, MODELO 6.(F: FCM UNICAMP)

Neste MODELO 6, no ajuste da altura foram utilizadas 205 observações válidas, o R-quadrado foi de 0.0925 e no teste do modelo o nível de significância p obtido foi 0.0133. No ajuste do PESO com ALTP5 foram utilizadas 2393 observações válidas, o R-quadrado foi de 0.4662 e no teste do modelo o nível de significância p foi menor que 0.0001.

Na figura 5.15 observa-se no gráfico as curvas da sensibilidade e da especificidade que são bastante semelhantes aos mesmos gráficos do MODELO 0.

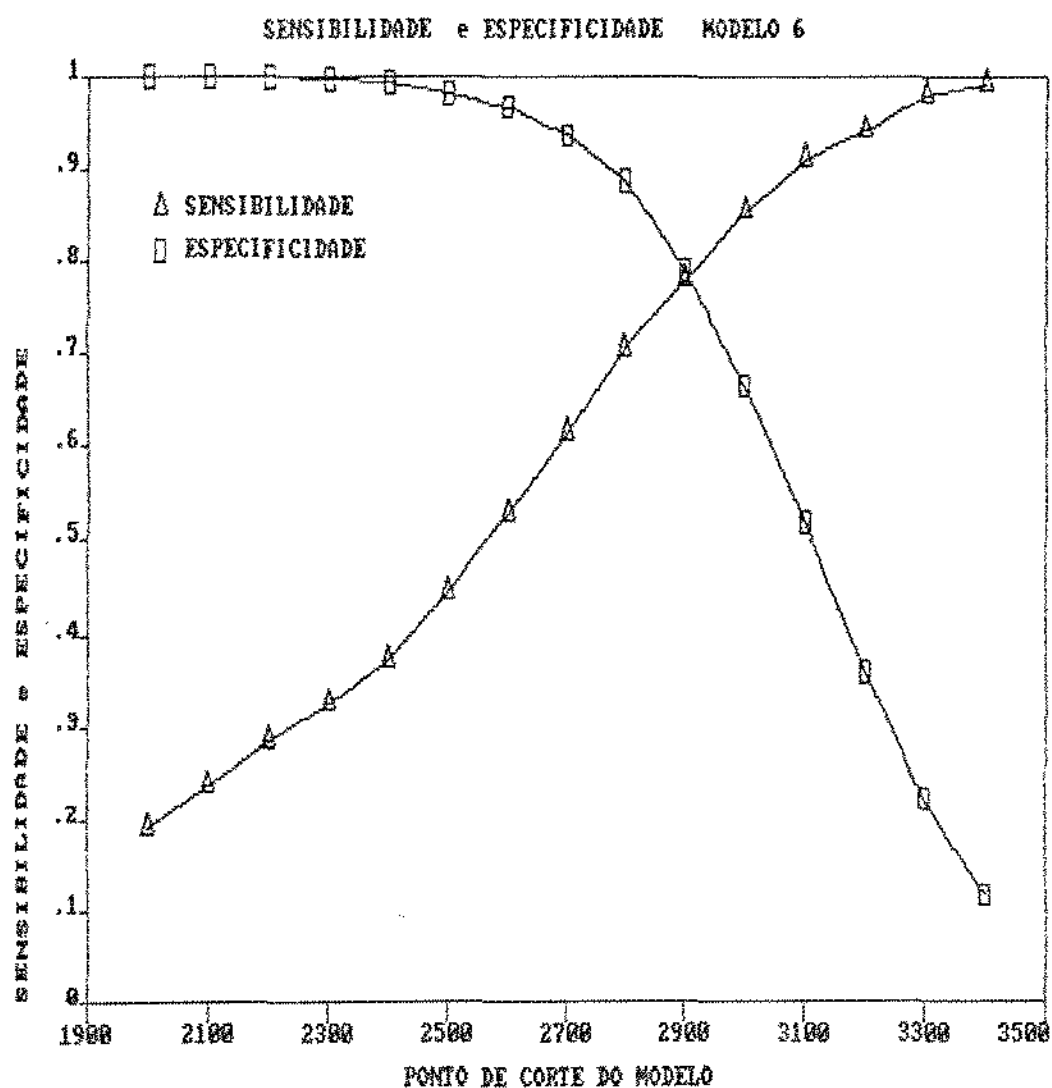


Figura 5.15 - Sensibilidade e especificidade para o MODELO 6.
----- (Fonte: FCM UNICAMP)

A comparação dos MODELO 0 e 6 pode ser feita através da figura 5.16, onde se pode notar que as curvas estão quase sobrepostas, indicando que praticamente não há diferença de desempenho entre ambos.

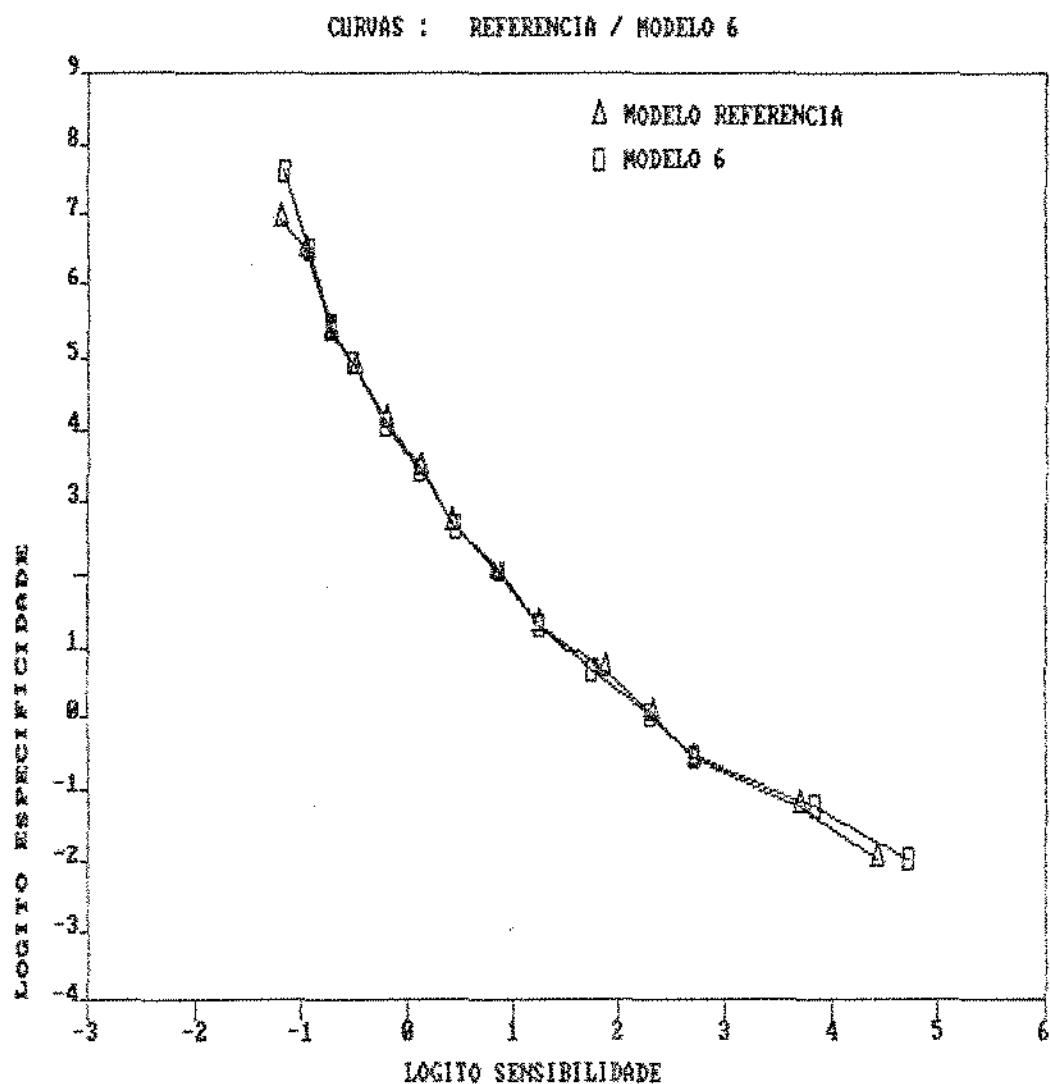


Figura 5.16 - LOGITO P versus LOGITO Q para os MODELOS 0 e 6.
 ----- (Fonte: FCM UNICAMP)

10) MODELOS 7 8 9 10 e 11

Outra tentativa de melhorar a capacidade de detecção de baixo peso pelos modelos foi a utilização de um método iterativo, onde pesos são atribuídos às observações.

Estes pesos são atribuídos às observações de forma que aqueles em que ocorreram previsões falsas negativas nas iterações anteriores recebem um peso $W > 1$ e se não ocorreram, um peso $W = 1$.

O conjunto de variáveis explicativas que utilizamos é o mesmo obtido no ajuste inicial para o MODELO 0 e valores iniciais dos parâmetros no método iterativo são também os valores obtidos no MODELO 0.

O número de iterações usadas é de três. Nas tabelas 5.10 a 5.14 que mostram os valores dos parâmetros nas diversas iterações pode-se observar através da progressão dos valores dos mesmos que este número de iterações é suficiente para a convergência do método.

Nos MODELOS 7 e 8 foram usados pesos fixos independentes da magnitude dos resíduos falsos negativos ocorridos nas observações. Já nos MODELOS 9, 10 e 11 os pesos utilizados são proporcionais às mesmas.

MODELO 7

VARIÁVEIS	PAR. EST. MÍNIMOS QUADRADOS COMUNS	PARÂMETROS ESTIMADOS		
		ITERAÇÃO		
		1	2	3
INTERCEPTO	1399	1323	1336	1334
NTM1	-95.1	-78.7	-80.7	-80.6
FILMO1	-64.4	-82.0	-81.0	-82.1
TOTPAR	30.0	29.8	30.1	30.2
ANTOBS1	-182	-191	-194	-194
TOTCON	13.1	14.5	14.5	14.5
ALTM	9.15	8.71	8.81	8.78
EDEMA	125	121	119	119
PRGR	-279	-262	-264	-264
TOXE	-115	-128	-130	-130
INF	-117	-110	-111	-111
IDG	163	167	166	166

Tabela 5.10 - Parâmetros iniciais e estimados nas iterações no MODELO 7. (Fonte: FCM UNICAMP)

No MODELO 7 foi atribuído peso 2 (dois) às observações onde houveram previsões falsas negativas na iteração anterior e peso 1 (um) às demais. Desta forma, sendo $W(i)$ o peso da i -ésima observação numa certa iteração, eles assumem os seguintes valores:

$W(i) = 2$ se na iteração anterior à i -ésima observação produziu uma previsão falsa negativa.

$W(i) = 1$ nas demais observações.

Aplicado este tratamento podemos observar na tabela 5.10 os valores iniciais e os valores dos parâmetros nas três iterações.

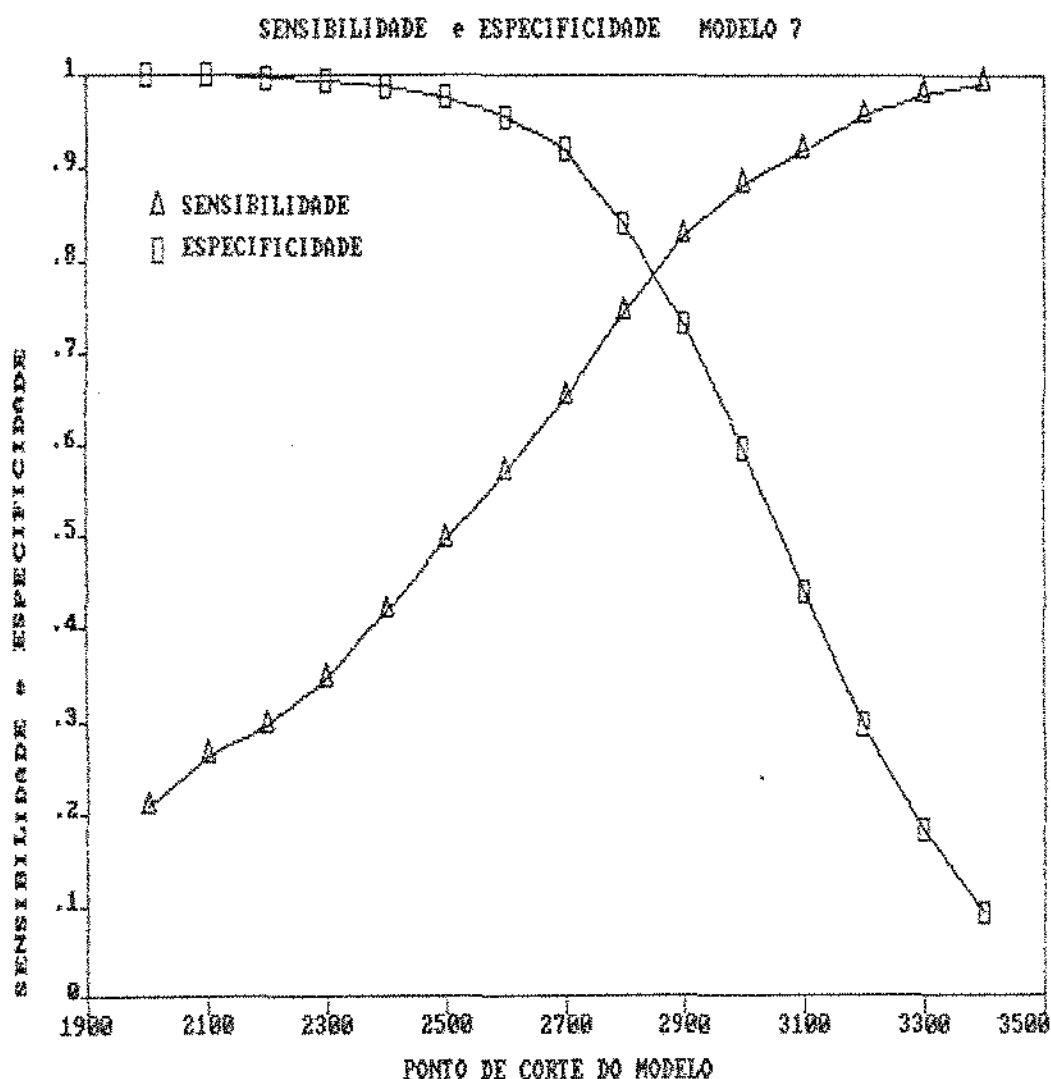


Figura 5.17 - Sensibilidade e especificidade para o MODELO 7.
 ----- (Fonte: FCM UNICAMP)

As curvas da sensibilidade e da especificidade para o MODELO 7 podem ser vistas no gráfico da figura 5.17.

A comparação de desempenho entre o MODELOS 7 e o MODELO 0 pode ser feita através das curvas da figura 5.18.

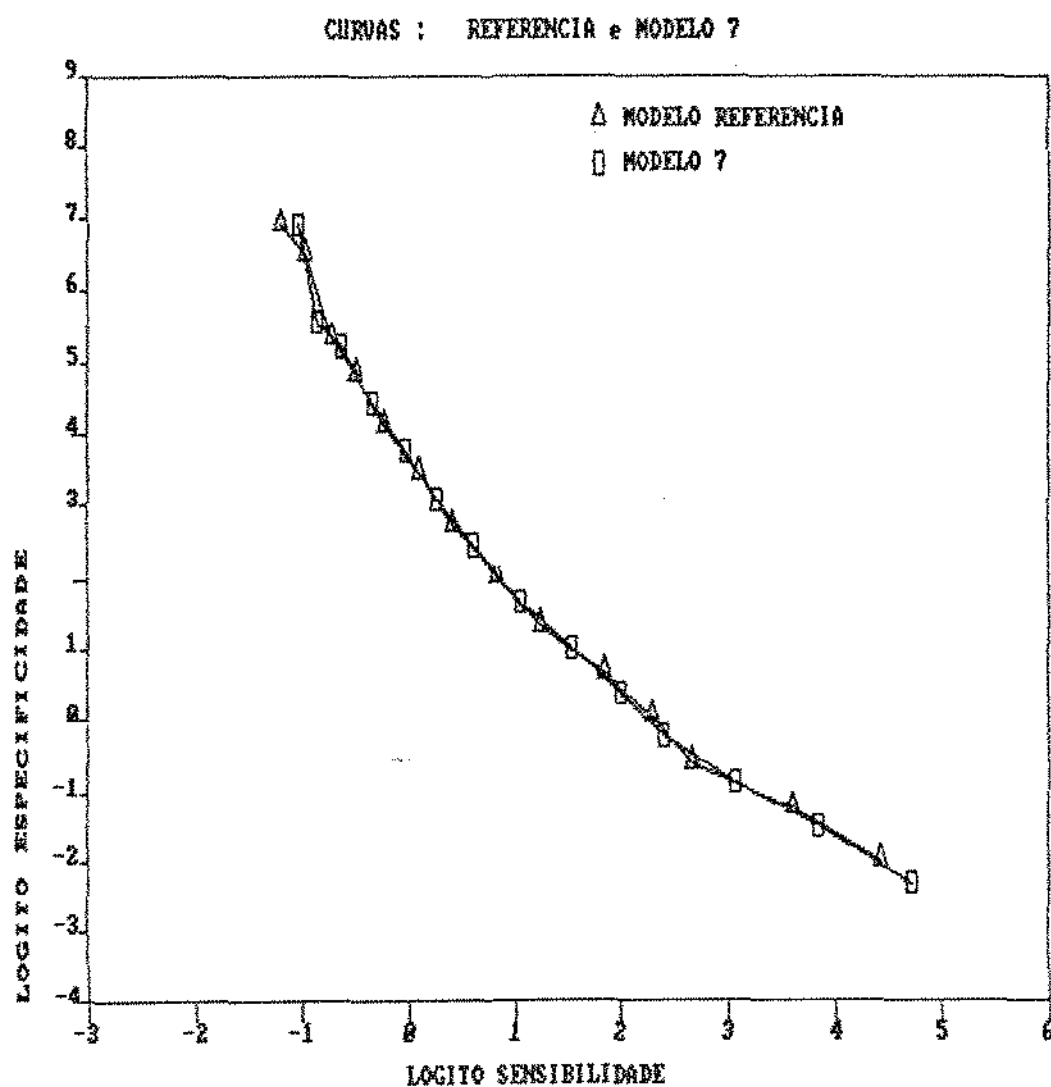


Figura 5.18 - LOGITO P versus LOGITO Q para MODELOS 0 e 7
 ----- (Fonte: FCM UNICAMP)

No MODELO 8 foi atribuído peso 5 (cinco) às observações onde houveram previsões falsas negativas na iteração anterior e peso 1 (um) às demais. Desta forma, sendo $W(i)$ o peso da i -ésima observação numa certa iteração, ele assume os seguintes valores:

$W(i) = 5$ se na iteração anterior a i -ésima observação produziu uma previsão falsa negativa.

$W(i) = 1$ nas demais observações.

Os resultados das iterações é o que pode ser observado na tabela 5.11.

MODELO 8

VARIÁVEIS	PAR. EST. MÍNIMOS QUADRADOS COMUNS	PARAMETROS ESTIMADOS		
		1	2	3
INTERCEPTO	1399	1197	1273	1255
NTM1	-95.1	-52.4	-65.2	-62.4
FILMO1	-64.4	-109	-112	-110
TOTPAR	30.0	28.7	30.0	29.9
ANTOBS1	-182	-200	-208	-205
TOTCON	13.1	16.8	16.8	16.3
ALTM	9.15	7.84	7.75	7.92
EDEMA	125	112	105	106
PRGR	-279	-236	-229	-222
TOXE	-115	-149	-142	-146
INF	-117	-98.3	-106	-103
IDG	163	172	164	166

Tabela 5.11 - Parâmetros iniciais e estimados nas iterações no
MODELO 8. (Fonte: FCM UNICAMP)

As curvas da sensibilidade e da especificidade para o MODELO 8 podem ser observadas no gráfico da figura 5.19.

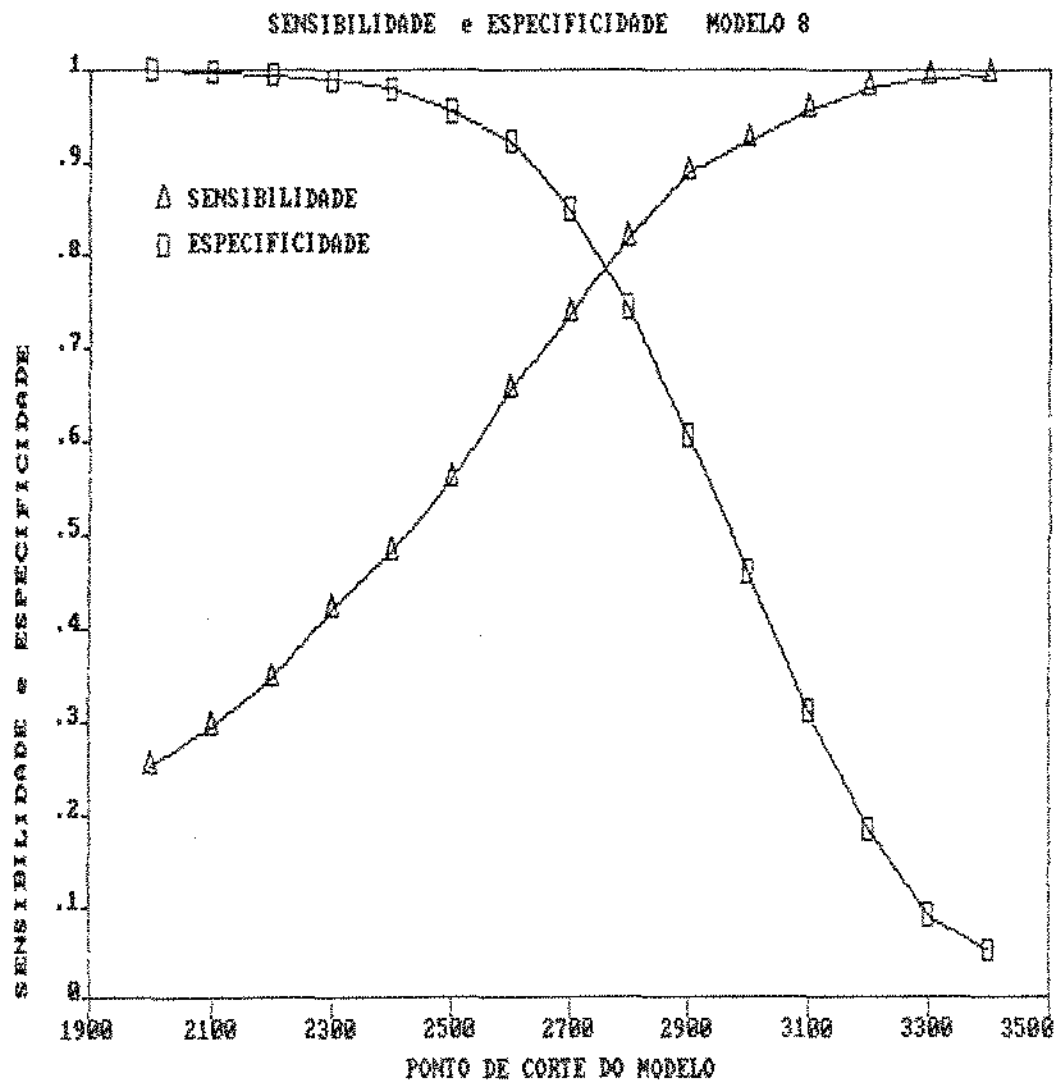


Figura 5.19 - Sensibilidade e especificidade para o MODELO 8.
----- (Fonte: FCM UNICAMP)

A comparação de desempenho do MODELO 8 com o MODELO 0 pode ser feita através da figura 5.20.

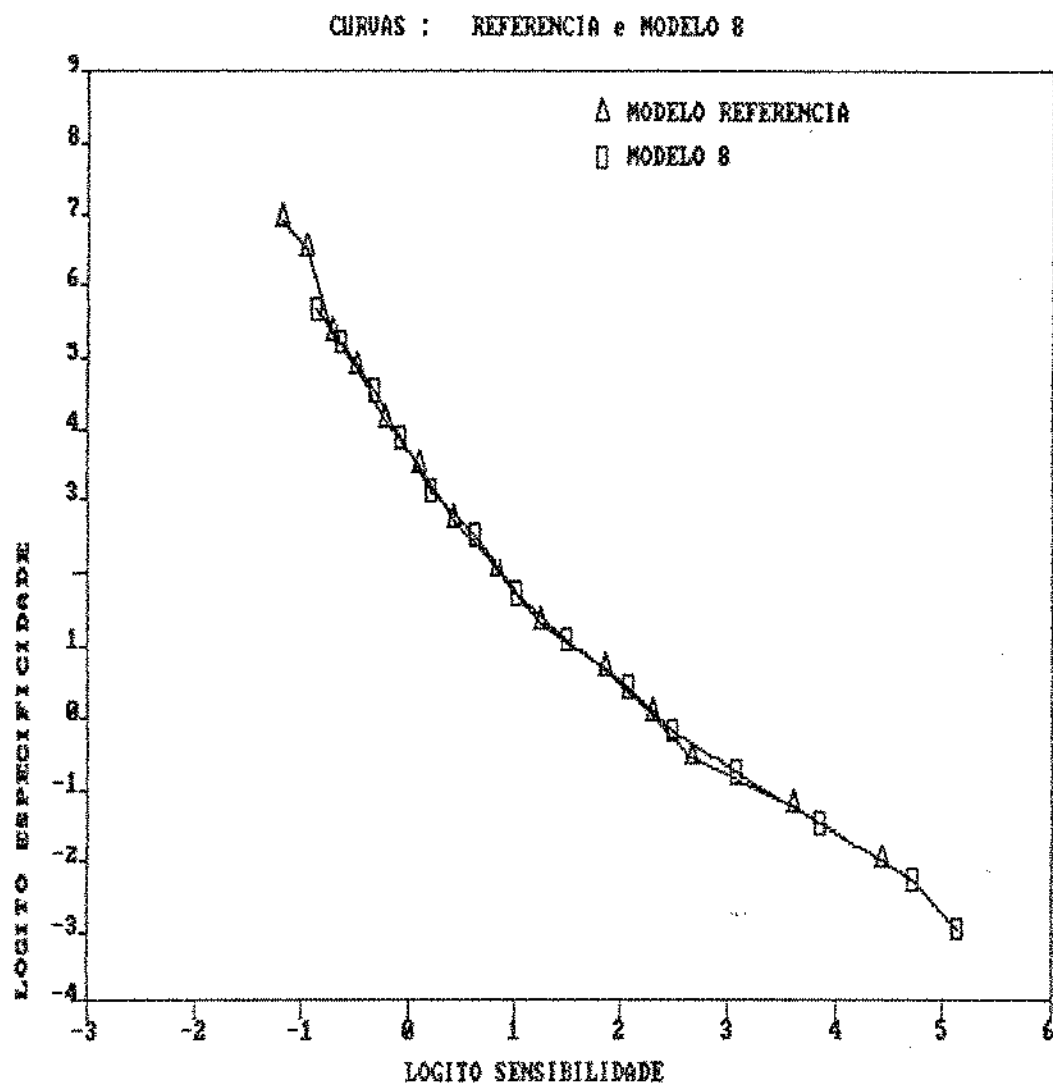


Figura 5.20 - LOGITO P versus LOGITO Q para MODELOS 0 e 8.
 (Fonte: FCM UNICAMP)

No MODELO 9 foram atribuídos pesos às observações cujas previsões na iteração anterior tenham sido falsas negativas, pesos proporcionais aos módulos dos resíduos. Sendo que o valor mínimo estabelecido para cada peso é 1 (um) e o peso é igual a 1 (um) nas demais observações.

Sendo $W(i)$ o peso da i -ésima observação, este assume o seguinte valor:

$$W(i) = \frac{|y(i) - y_p(i)|}{s} \quad \begin{array}{l} \text{se na iteração anterior na } i\text{-ésima} \\ \text{observação houve uma previsão} \\ \text{falsa negativa e nesta } W(i) \geq 1. \end{array}$$

e

$$W(i) = 1 \quad \text{nas demais observações.}$$

Onde,

s - é o desvio padrão dos resíduos no ajuste inicial,

$y(i)$ - é o peso real de cada recém-nascido e

$y_p(i)$ - é o peso previsto da i -ésima observação na iteração anterior.

Os parâmetros estimados neste método iterativo estão na tabela 5.12.

MODELO 9

VARIÁVEIS	PAR. EST. MÍNIMOS QUADRADOS COMUNS	PARÂMETROS ESTIMADOS		
		ITERAÇÃO		
		1	2	3
INTERCEPTO	1399	1402	1408	1405
NTM1	-95.1	-86.2	-88.7	-87.9
FILMO1	-64.4	-78.0	-74.9	-75.1
TOTPAR	30.0	28.8	29.0	28.9
ANTOBS1	-182	-175	-176	-175
TOTCON	13.1	14.9	14.5	14.6
ALTM	9.15	7.91	8.19	8.07
EDEMA	125	114	115	115
PRGR	-279	-251	-256	-255
TOXE	-115	-131	-128	-128
INF	-117	-112	-113	-113
IDG	163	161	161	161

Tabela 5.12 - Parâmetros iniciais e estimados nas iterações no MODELO 9. (Fonte: FCM UNICAMP)

As curvas da sensibilidade e da especificidade para o MODELO 9 estão no gráfico da figura 5.21.

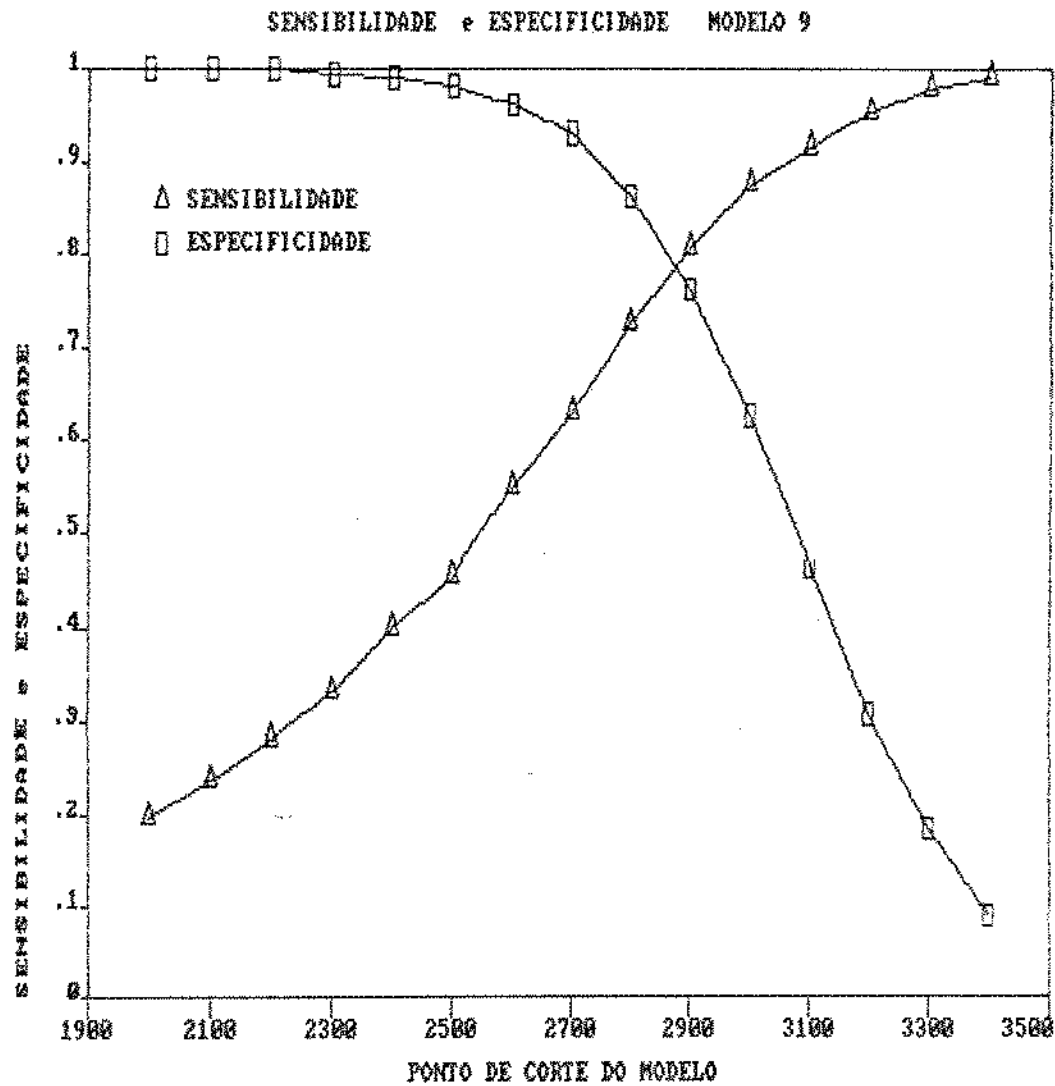


Figura 5.21 - Sensibilidade e especificidade para o MODELO 9.
----- (Fonte: FCM UNICAMP)

A comparação de desempenho entre o MODELO 9 e o MODELO 0 pode ser feita através da figura 5.22.

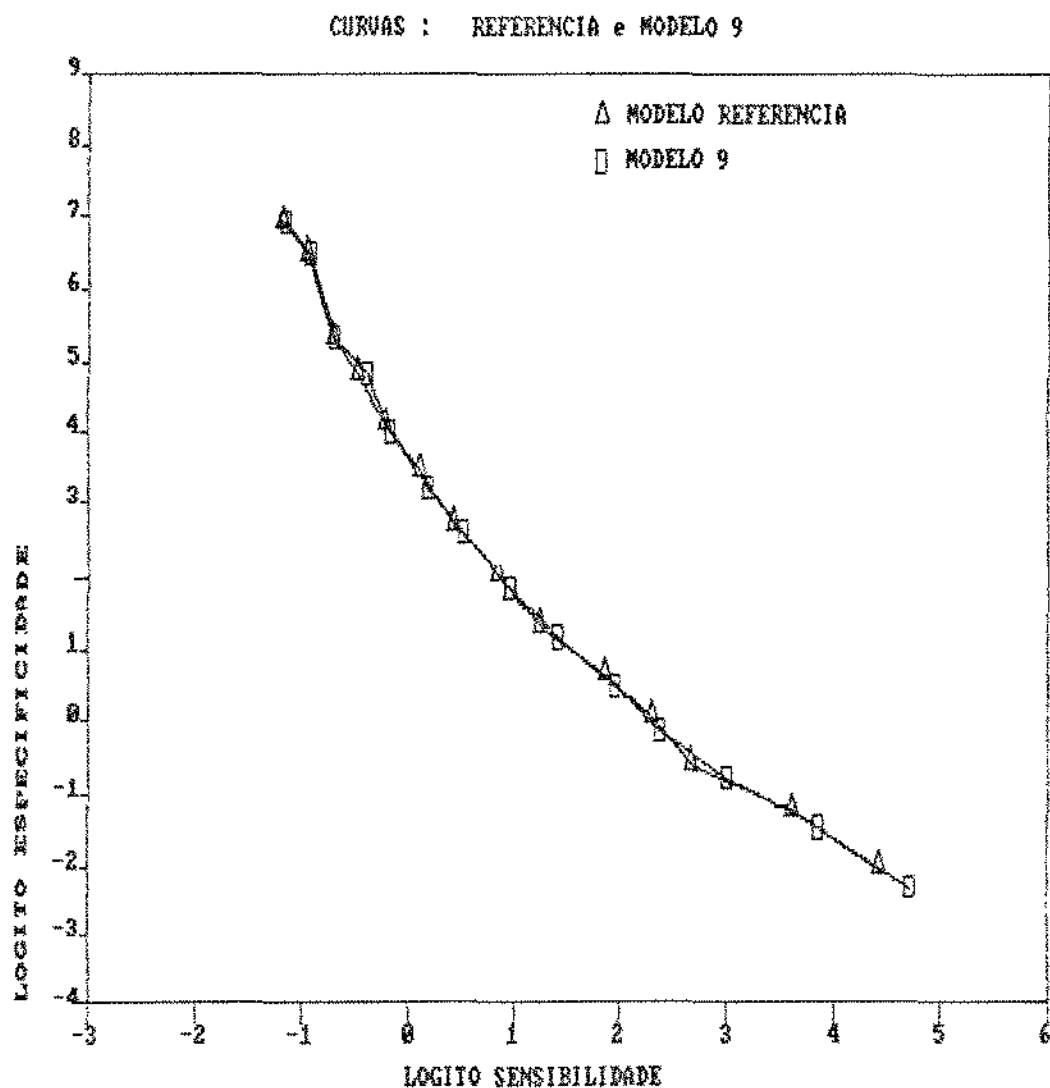


Figura 5.22 - LOGITO P versus LOGITO Q para os MODELO 0 e 9.
----- (Fonte: FCM UNICAMP)

No MODELO 10, como no MODELO 9, atribuiu-se pesos às observações cujas previsões na iteração anterior tenham sido falsas negativas, pesos proporcionais aos módulos dos resíduos. Sendo que o valor mínimo estabelecido para cada peso é 1 (um) e o peso é igual a 1 (um) nas demais observações. A diferença entre ambos que neste os pesos são maiores (duas vezes) que no MODELO 9 para observações com diferenças iguais.

Sendo $W(i)$ o peso da i -ésima observação, este assume os seguintes valores:

$$W(i) = 2 \frac{|y(i) - yp(i)|}{s}$$

se na iteração anterior na i -ésima observação houve uma previsão falsa negativa e nesta $W(i) \geq 1$.

$$W(i) = 1$$

nas demais observações.

- Onde,
- s - é o desvio padrão dos resíduos no ajuste inicial,
 - $y(i)$ - é o peso real de cada recém-nascido e
 - $yp(i)$ - é o peso previsto da i -ésima observação na iteração anterior.

Na tabela 5.13 podem ser observados os parâmetros estimados segundo este critério:

MODELO 10

VARIÁVEIS	PAR. EST. MÍNIMOS QUADRADOS COMUNS	PARÂMETROS ESTIMADOS		
		ITERAÇÃO		
		1	2	3
INTERCEPTO	1399	1331	1375	1363
NTM1	-95.1	-66.3	-79.4	-74.3
FILMO1	-64.4	-101	-87.7	-92.7
TOTPAR	30.0	27.9	28.5	28.3
ANTOBS1	-182	-177	-178	-178
TOTCON	13.1	17.4	15.8	16.5
ALTM	9.15	6.69	7.64	7.35
EDEMA	125	102	104	103
PRGR	-279	-218	-234	-231
TOXE	-115	-152	-140	-145
INF	-117	-103	-110	-107
IDG	163	163	159	160

Tabela 5.13 - Parâmetros iniciais e estimados nas iterações no MODELO 10. (Fonte: FCM UNICAMP)

As curvas da sensibilidade e da especificidade para o MODELO 10 podem ser vistas no gráfico da figura 5.23.

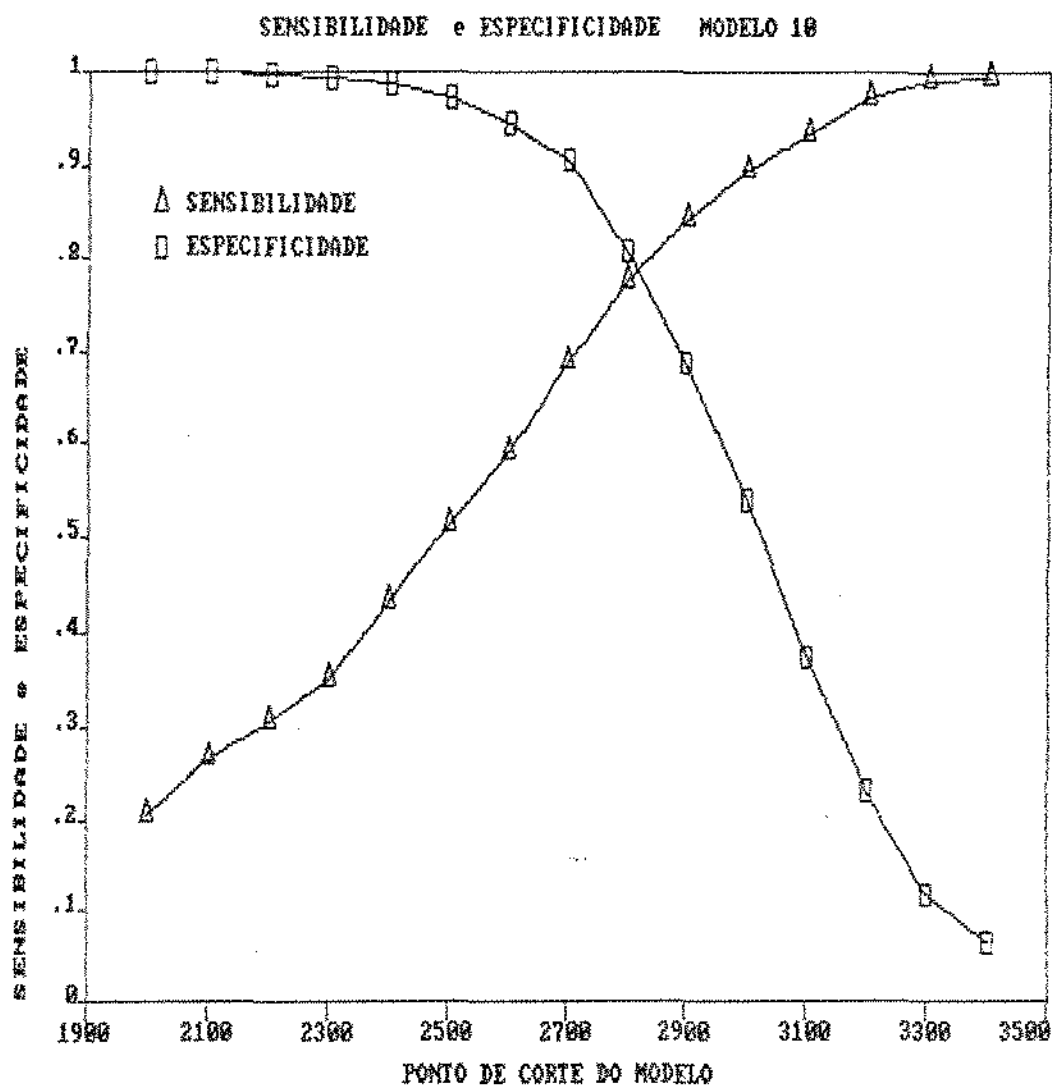


Figura 5.23 - Sensibilidade e especificidade para o MODELO 10.
(Fonte: FCM UNICAMP)

A comparação de desempenho dos MODELOS 10 e 0 pode ser feita através das curvas no gráfico da figura 5.24.

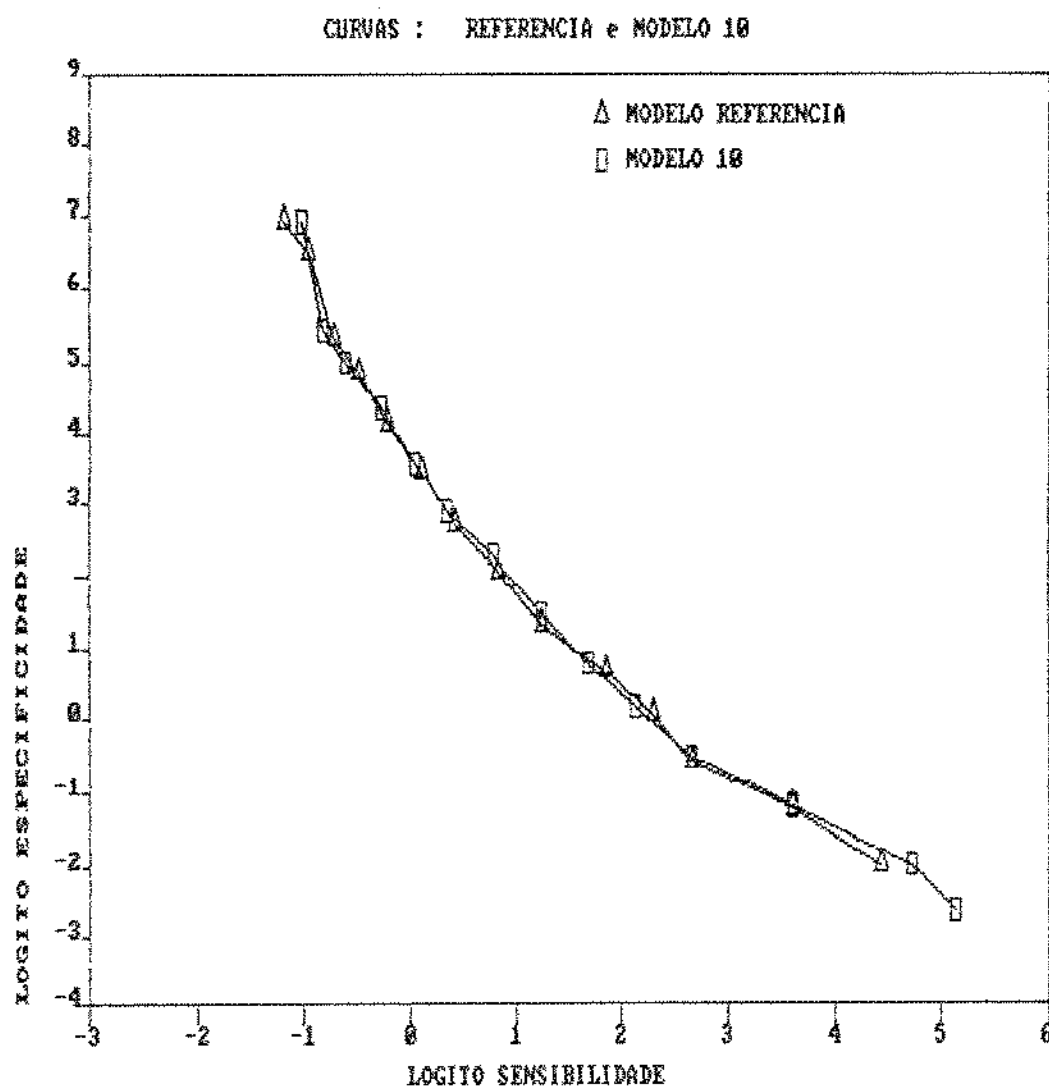


Figura 5.24 - LOGITO P versus LOGITO Q para os MODELOS 0 e 10.
----- (Fonte: FCM UNICAMP)

No MODELO 11, como nos MODELOS 9 e 10, foram atribuídos pesos às observações cujas previsões na iteração anterior tenham sido falsas negativas, pesos proporcionais aos módulos dos resíduos. Sendo que o valor mínimo estabelecido para este peso é 1 (um) e o peso é igual a 1 (um) nas demais observações. A diferença entre ambos é que neste último atribuímos pesos maiores que nos MODELOS 9 e 10 (duas vezes o MODELO 10 e quatro vezes o MODELO 9) às observações com diferenças iguais.

Sendo $W(i)$ o peso da i -ésima observação, este assume os seguintes valores:

$$W(i) = 4 \frac{|y(i) - y_p(i)|}{s} \quad \begin{array}{l} \text{se na iteração anterior na } i\text{-ésima} \\ \text{observação houve uma previsão} \\ \text{falsa negativa e nesta } W(i) \geq 1. \end{array}$$

$$W(i) = 1 \quad \text{nas demais observações.}$$

Onde,

s - é o desvio padrão dos resíduos no ajuste inicial,

$y(i)$ - é o peso real de cada recém-nascido e

$y_p(i)$ - é o peso previsto da i -ésima observação na iteração anterior.

Na tabela 5.14 estão os resultados das iterações com estes pesos.

MODELO 11

VARIÁVEIS	PAR. EST. MÍNIMOS QUADRADOS COMUNS	PARAMETROS ESTIMADOS		
		ITERAÇÃO		
		1	2	3
INTERCEPTO	1399	1257	1412	1314
NTM1	-95.1	-47.6	-86.8	-55.5
FILMO1	-64.4	-122	-101	-109
TOTPAR	30.0	26.2	30.0	27.1
ANTOBS1	-182	-173	-185	-168
TOTCON	13.1	20.1	15.4	18.7
ALTM	9.15	5.30	7.76	6.12
EDEMA	125	87.0	91.4	89.2
PRGR	-279	-183	-281	-165
TOXE	-115	-170	-89.0	-179
INF	-117	-92.5	-81.6	-107
IDG	163	162	149	159

Tabela 5.14 - Parâmetros iniciais e estimados nas iterações no MODELO 11. (Fonte: FCM UNICAMP)

As curvas da sensibilidade e da especificidade para o MODELO 11 podem ser observadas no gráfico da figura 5.25.

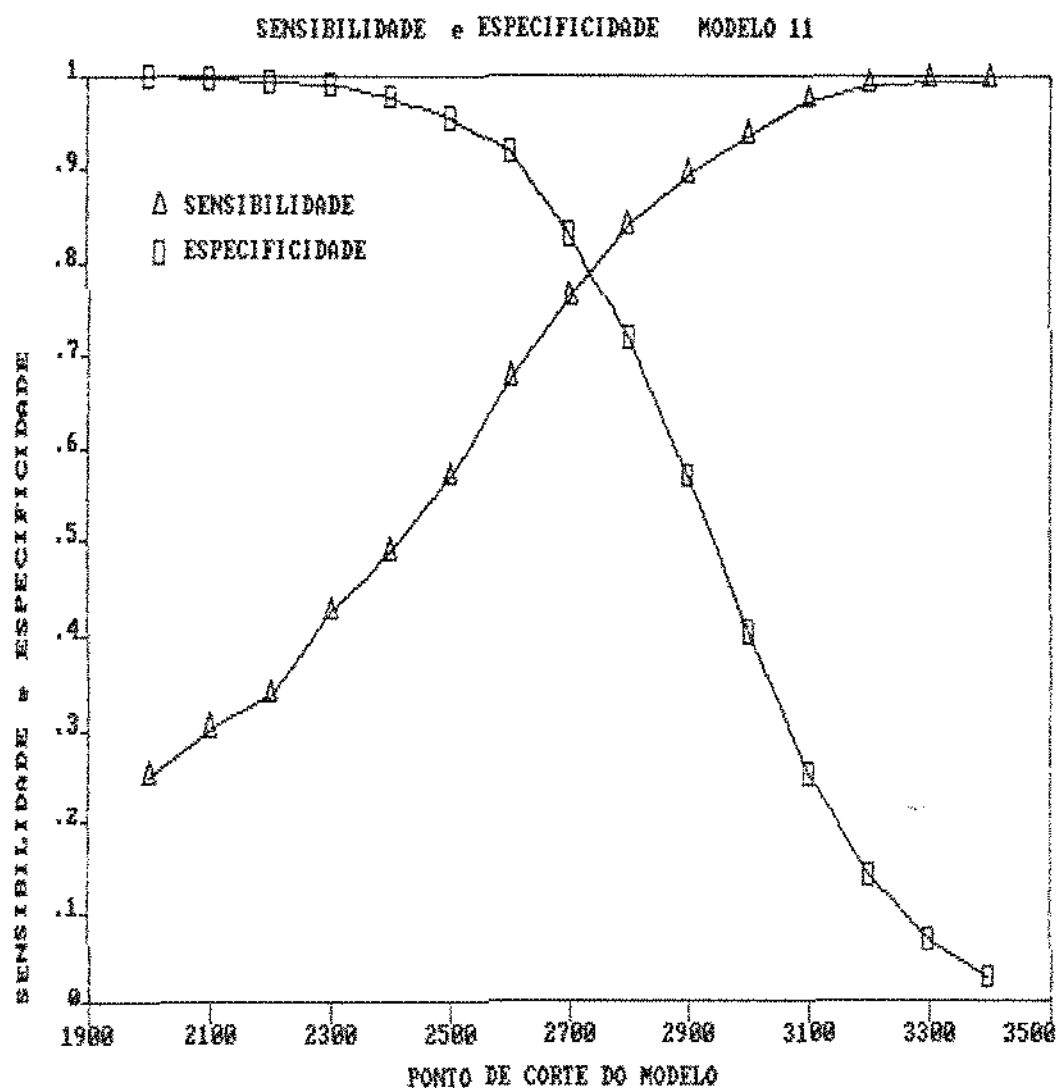


Figura 5.25 - Sensibilidade e especificidade para o MODELO 11.
 ----- (Fonte: FCM UNICAMP)

A comparação dos MODELOS 11 e 0 quanto ao desempenho pode ser feita através das curvas no gráfica da figura 5.26.

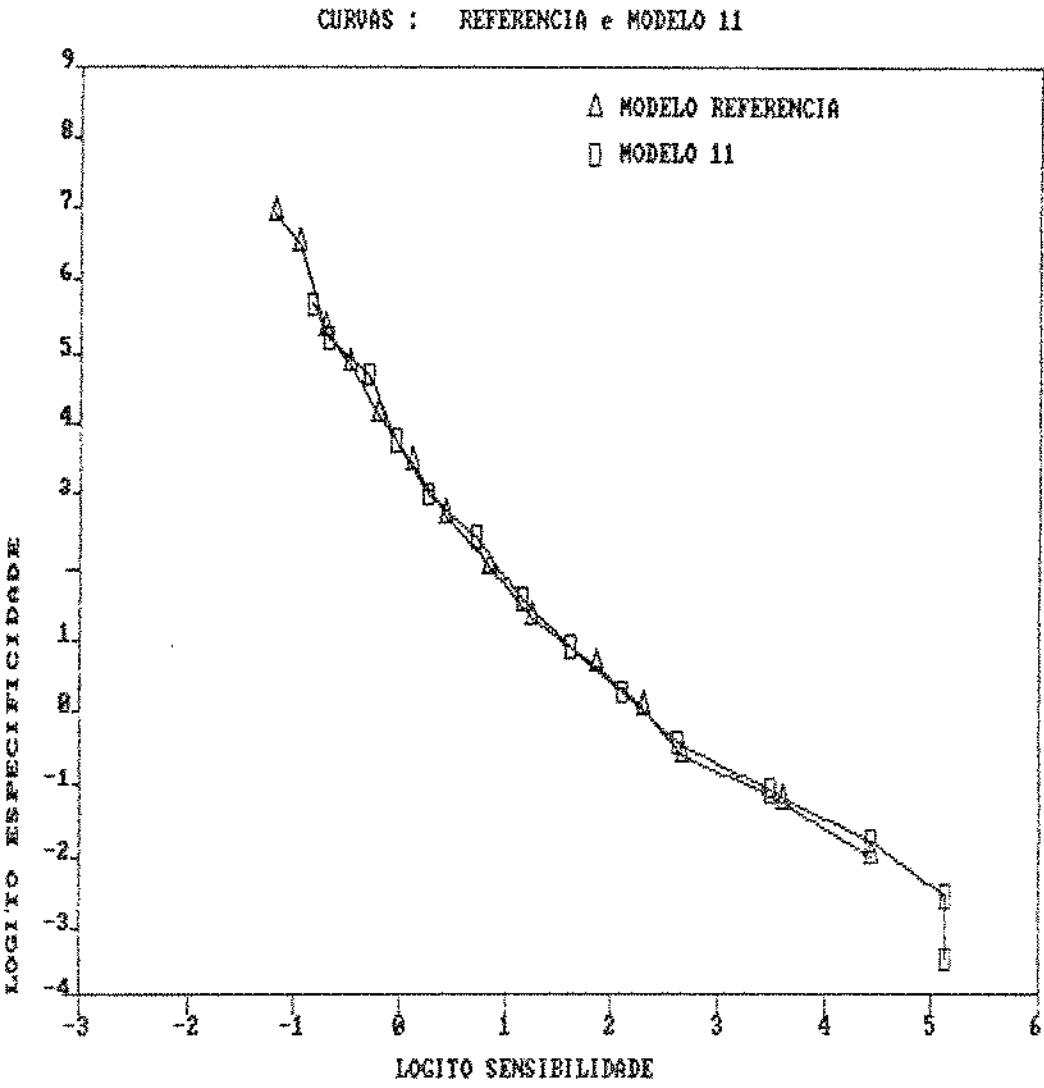


Figura 5.26 - LOGITO P versus LOGITO Q para os MODELOS 0 e 11.
----- (Fonte: FCM UNICAMP)

VI - REGRESSÃO LOGÍSTICA

VI.1 - INTRODUÇÃO

Dando continuidade à busca de um método melhor para prever baixo peso em recém-nascidos testa-se neste capítulo a aplicação da técnica de regressão logística ao problema em questão. Como foi visto na seção IV.4 esta técnica não se aplica na previsão de variáveis contínuas, sendo portanto inadequada para o ajuste do peso de recém-nascidos.

A transformação da variável PESO de forma a obter uma nova variável viabilizaria a aplicação de regressão logística neste caso se esta nova variável fosse binária. Uma forma de realizar esta transformação é fazer com que o resultado da transformação seja uma variável indicadora de peso normal em recém-nascidos.

Considerando que 2500 gramas é uma boa fronteira para detectar partos de risco, a variável PESO poderia ser recodificada, criando a variável IPESO, indicadora de "peso normal", da seguinte forma:

$IPESO=1$ se $PESO \geq 2500$ e

$IPESO=0$ se $PESO < 2500$.

Esta recodificação permitirá ajustar um modelo para IPESO através de regressão logística, estimando a probabilidade de que esta variável assuma o valor 1 (um), o que corresponde a se estimar a probabilidade um peso normal no recém-nascido, dados os valores das variáveis independentes ou explicativas.

Através da estimativa da probabilidade da variável dependente IPESO assumir o valor 1 (um), nos modelos que seguiremos tentar-se-á detectar os casos de baixo peso entre os recém-nascidos. Isto será feito de forma semelhante ao que foi feito no capítulo V para a regressão linear múltipla, utilizando a sensibilidade e especificidade para este fim. A diferença neste caso é que os pontos de corte serão estabelecidos em probabilidades e não mais em gramas como antes.

Como foi feito no capítulo V apresentaremos os gráficos com as curvas da sensibilidade e especificidade. Estas curvas não podem ser comparadas diretamente com as suas similares do capítulo

anterior por terem escalas diferentes, muito embora tenham sido construídas de forma semelhante.

Quanto às curvas de LOGITO P versus LOGITO Q não haverá o problema na escala, o que permitirá a comparação dos resultados aqui obtidos com os resultados do MODELO 0, que desta forma continuará a ser o modelo referência.

VI.2 - MODELOS 12 13 14 e 15.

O MODELO 12 consiste na aplicação de regressão logística ao mesmo conjunto de variáveis explicativas e ao mesmo conjunto de observações utilizados no MODELO 0, tendo como variável dependente IPESO.

MODELO 12

VARIÁVEIS	PARAMETROS ESTIMADOS	ERRO PADRÃO	QUI-QUADRADO	P
INTERCEPTO	-5.3549845	0.4106	170.04	.
NTM1	-0.2375522	0.3015	0.62	0.4308
FILMO1	-0.5367280	0.2113	6.45	0.0111
TOTPAR	0.0668864	0.0369	3.28	0.0701
ANTOBS1	-0.7389108	0.3945	3.51	0.0611
TOTCON	0.0706417	0.0251	7.93	0.0049
ALTM	0.0123181	0.0120	1.05	0.3053
EDEMA	0.2368660	0.1666	2.02	0.1552
PRGR	-1.2085875	0.4151	8.48	0.0036
TOXE	-0.5026649	0.2242	5.02	0.0250
INF	-0.6498526	0.2869	5.13	0.0235
IDG	0.8713833	0.0488	319.14	.

Tabela 6.1 - Estatísticas calculadas em regressão logística para o
----- MODELO 12. (Fonte: FCM UNICAMP)

As curvas da sensibilidade e especificidade do MODELO 12 são apresentadas no gráfico da figura 6.1.

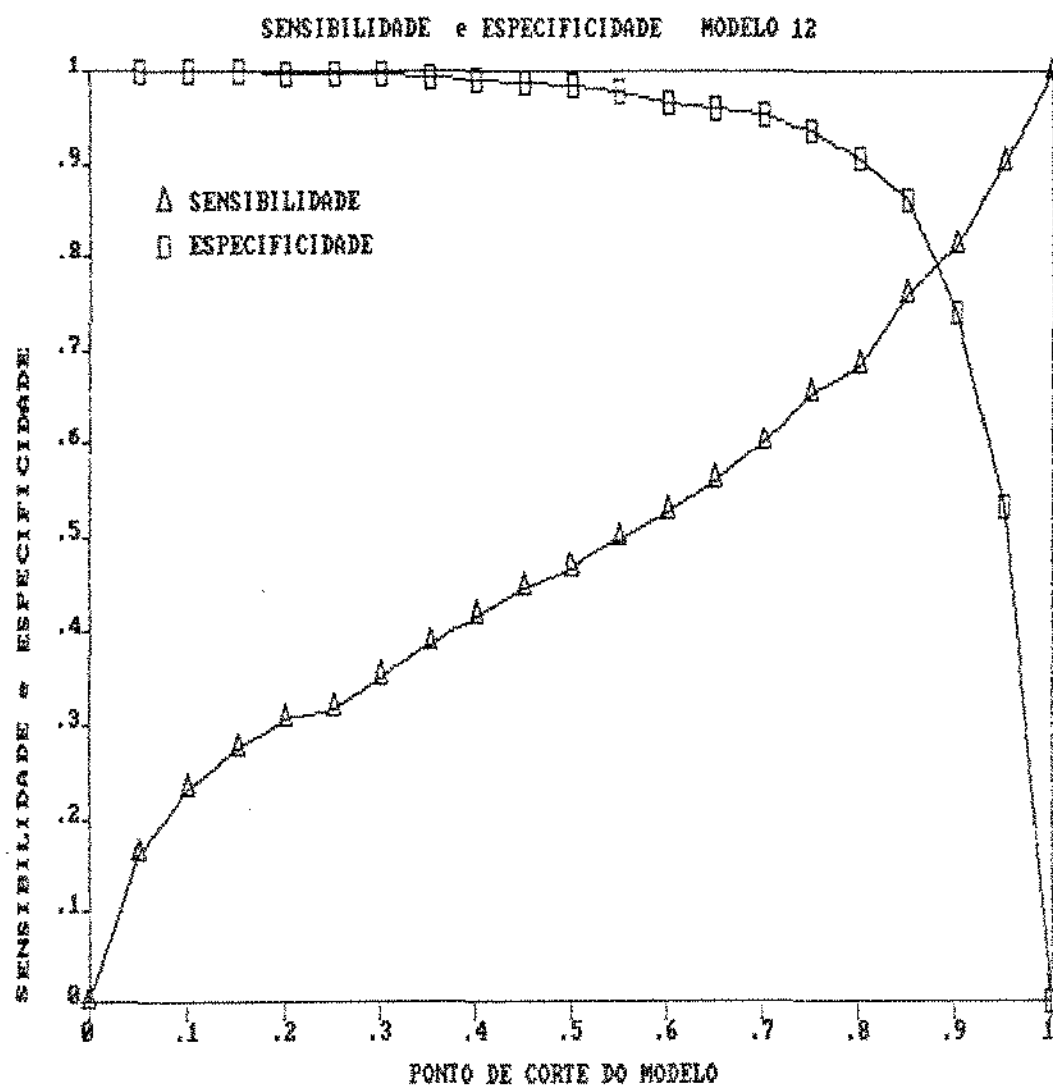


Figura 6.1 - Sensibilidade e especificidade para o MODELO 12.
----- (Fonte: FCM UNICAMP)

A comparação dos MODELOS 12 e 0 quanto aos seus desempenhos pode ser feita através das curvas no gráfico da figura 6.2.

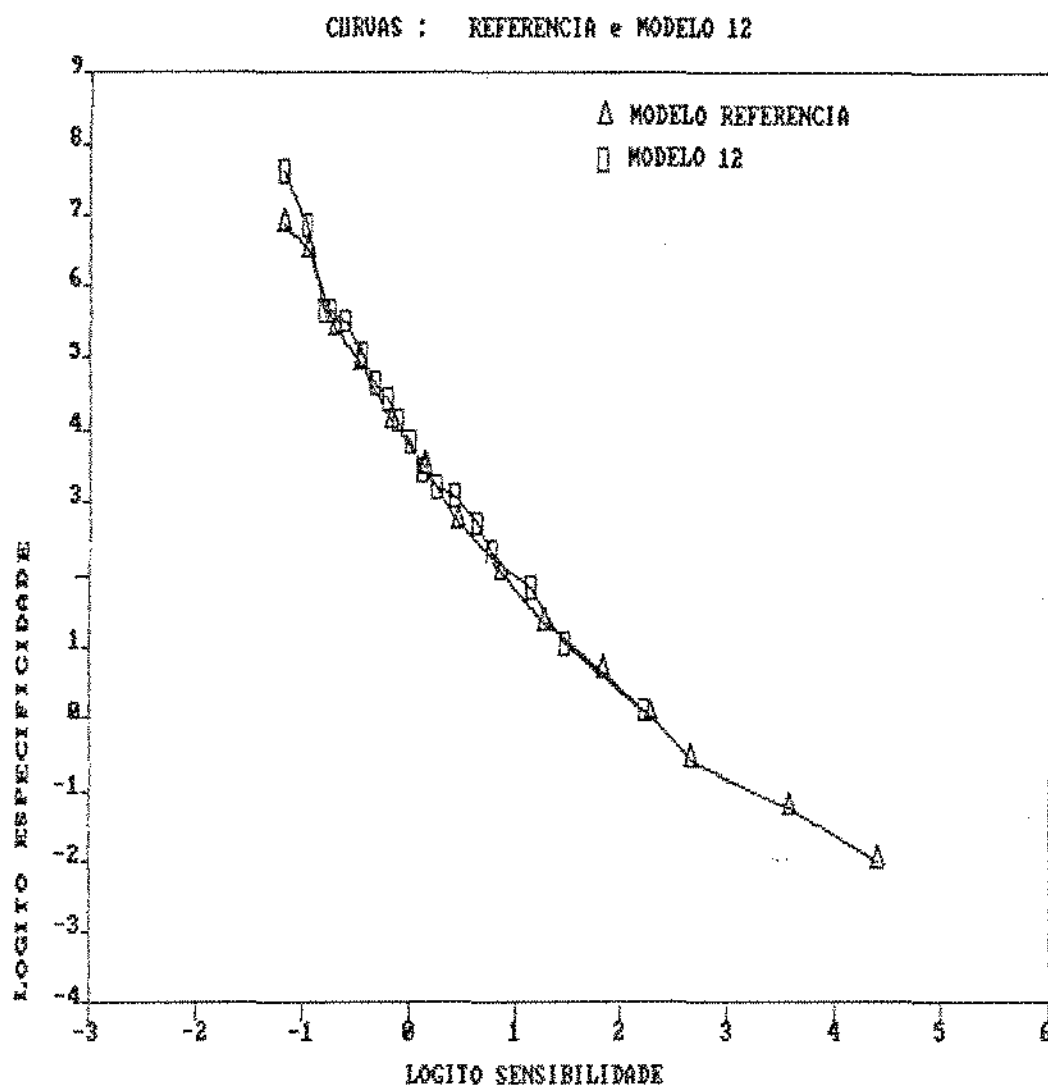


Figura 6.2 - LOGITO P versus LOGITO Q para os MODELOS 0 e 12.
 ----- (Fonte: FCM UNICAMP)

O MODELO 13 consiste na aplicação de regressão logística ao mesmo conjunto de variáveis explicativas e ao mesmo conjunto de observações utilizados no MODELO 1, tendo como variável dependente IPESO.

MODELO 13

VARIAVEIS	PARAMETROS ESTIMADOS	ERRO PADRAO	QUI-QUADRADO	P
INTERCEPT0	-23.7738364	30.8859	0.59	0.4415
NTM1	-0.1725065	0.3080	0.31	0.5754
FILMD1	-0.5406605	0.2161	6.26	0.0123
TOTPAR	0.0685477	0.0379	3.27	0.0705
ANTOBS1	-0.6265992	0.4493	1.95	0.1631
TOTCON	0.0220545	0.0799	0.08	0.7827
ALTM	-0.0066380	0.0344	0.04	0.8474
EDEMA	-0.0637419	0.5367	0.01	0.9055
PRGR	-0.8029353	0.8286	0.94	0.3326
TOXE	-0.4664426	0.2322	4.04	0.0446
INF	-0.0695484	1.0089	0.00	0.9450
IDG	0.4742782	0.6752	0.49	0.4824
ALTP1	0.4857154	0.8159	0.35	.

Tabela 6.2 - Estatísticas calculadas em regressão logística para o
MODELO 13. (Fonte: FCM UNICAMP)

As curvas da sensibilidade e especificidade do MODELO 13 estão apresentadas no gráfico da figura 6.3.

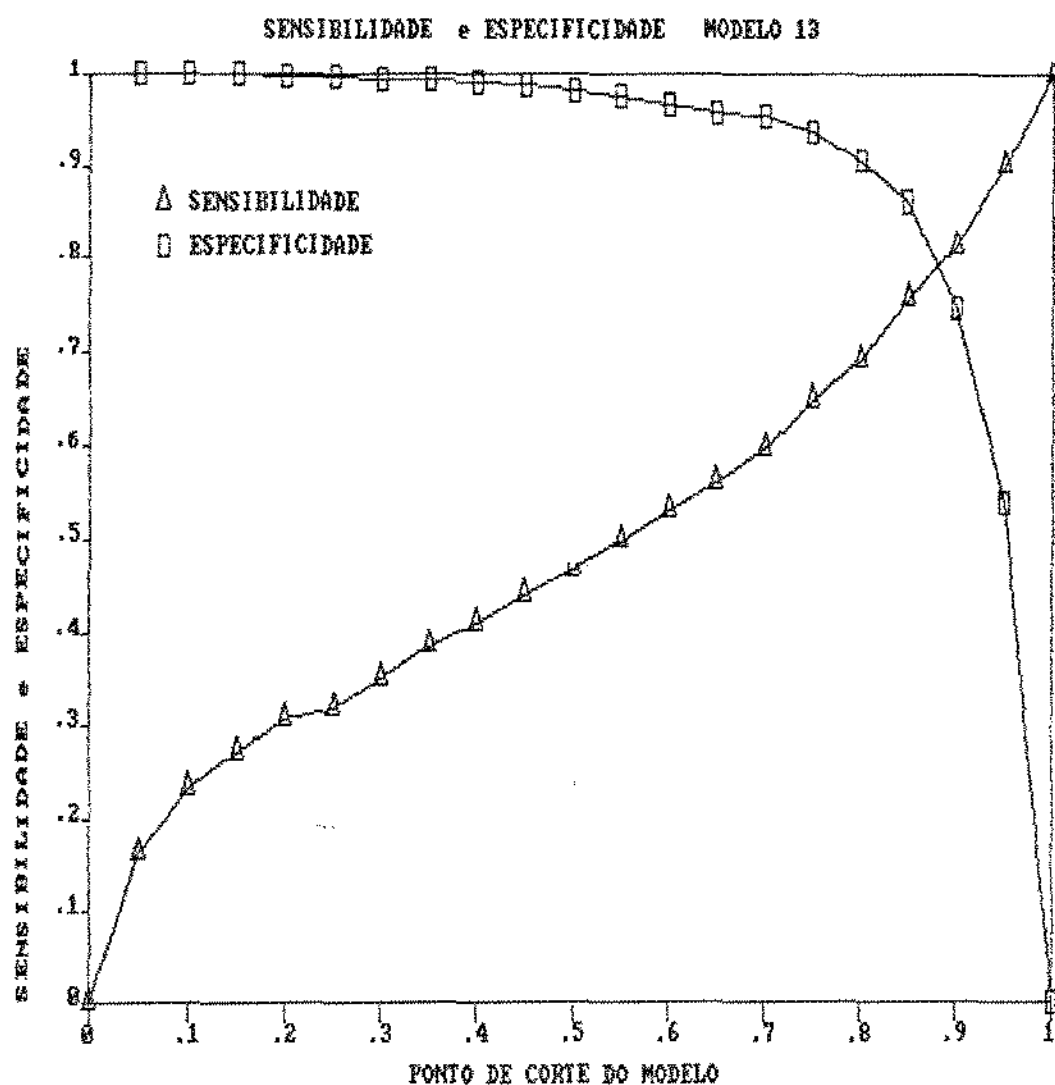


Figura 6.3 - Sensibilidade e especificidade para o MODELO 13.
 ----- (Fonte: FCM UNICAMP)

A comparação dos MODELOS 13 e 0 quanto aos seus desempenhos pode ser feita através do gráfico na figura 6.4.

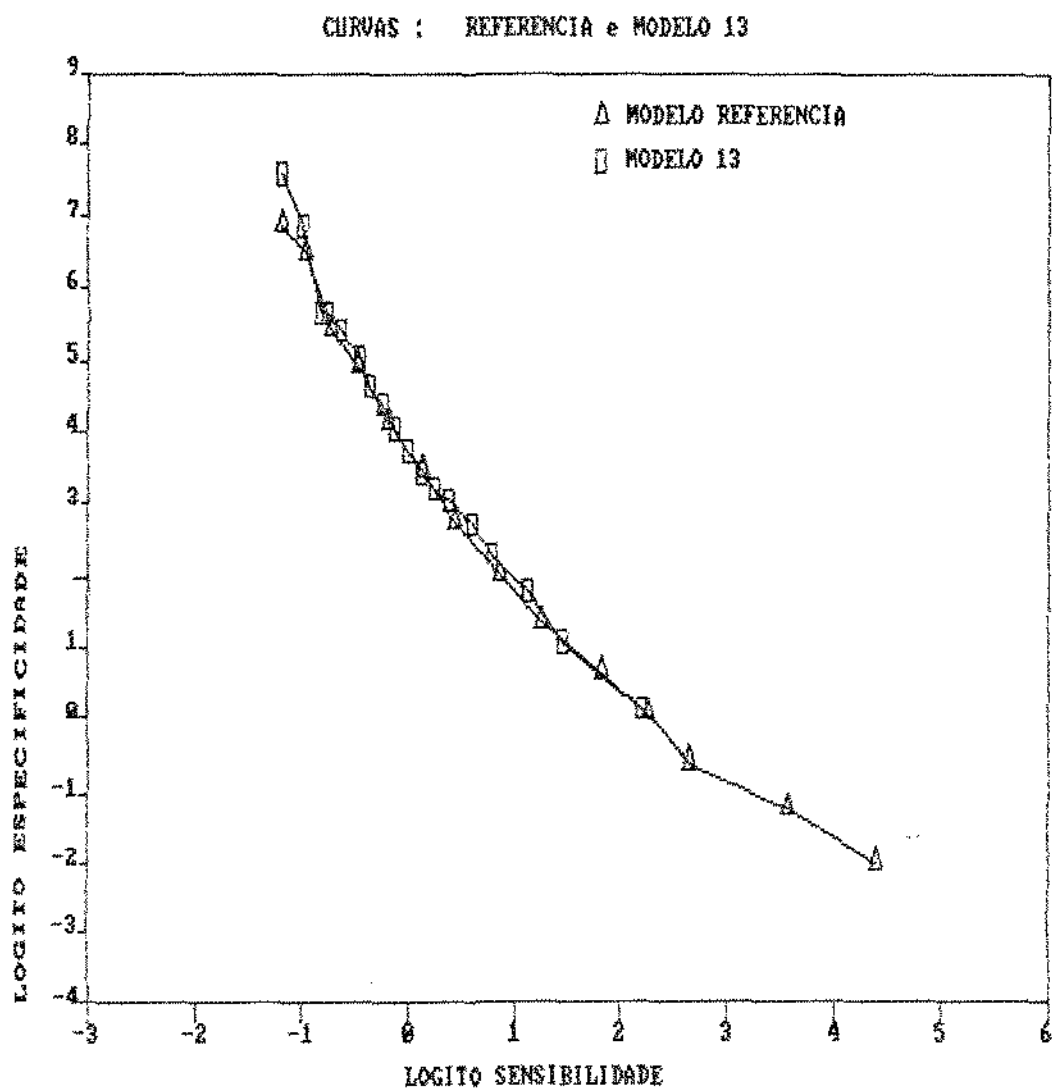


Figura 6.4 - LOGITO P versus LOGITO Q para MODELOS 0 e 13.
 ----- (Fonte: FCM UNICAMP)

O MODELO 14 consiste na aplicação de regressão logística ao mesmo conjunto de variáveis explicativas e ao mesmo conjunto de observações utilizados no MODELO 2, tendo como variável dependente IPESO.

MODELO 14

VARIÁVEIS	PARÂMETROS ESTIMADOS	ERRO PADRÃO	QUI-QUADRADO	P
INTERCEPT0	10.5338093	10.3752	1.03	0.3100
NTM1	-0.2578807	0.3035	0.72	0.3956
FILMO1	-0.5511106	0.2118	6.77	0.0093
TOTPAR	0.0934760	0.0407	5.28	0.0216
ANTOBS1	-0.7293676	0.3941	3.43	0.0642
TOTCON	0.1110439	0.0366	9.22	0.0024
ALTM	0.0258594	0.0144	3.21	0.0733
EDEMA	0.4706405	0.2298	4.19	0.0406
PRGR	-1.2208521	0.4180	8.53	0.0035
TOXE	-0.8324720	0.3082	7.29	0.0069
INF	-1.0349925	0.3810	7.38	0.0066
IDG	1.2295576	0.2358	27.19	0.0000
ALTP2	-0.4194225	0.2731	2.36	0.1245

Tabela 6.3 - Estatísticas calculadas em regressão logística para o MODELO 14. (Fonte: FCM UNICAMP)

As curvas da sensibilidade e especificidade do MODELO 14 podem ser vistas no gráfico da figura 6.5.

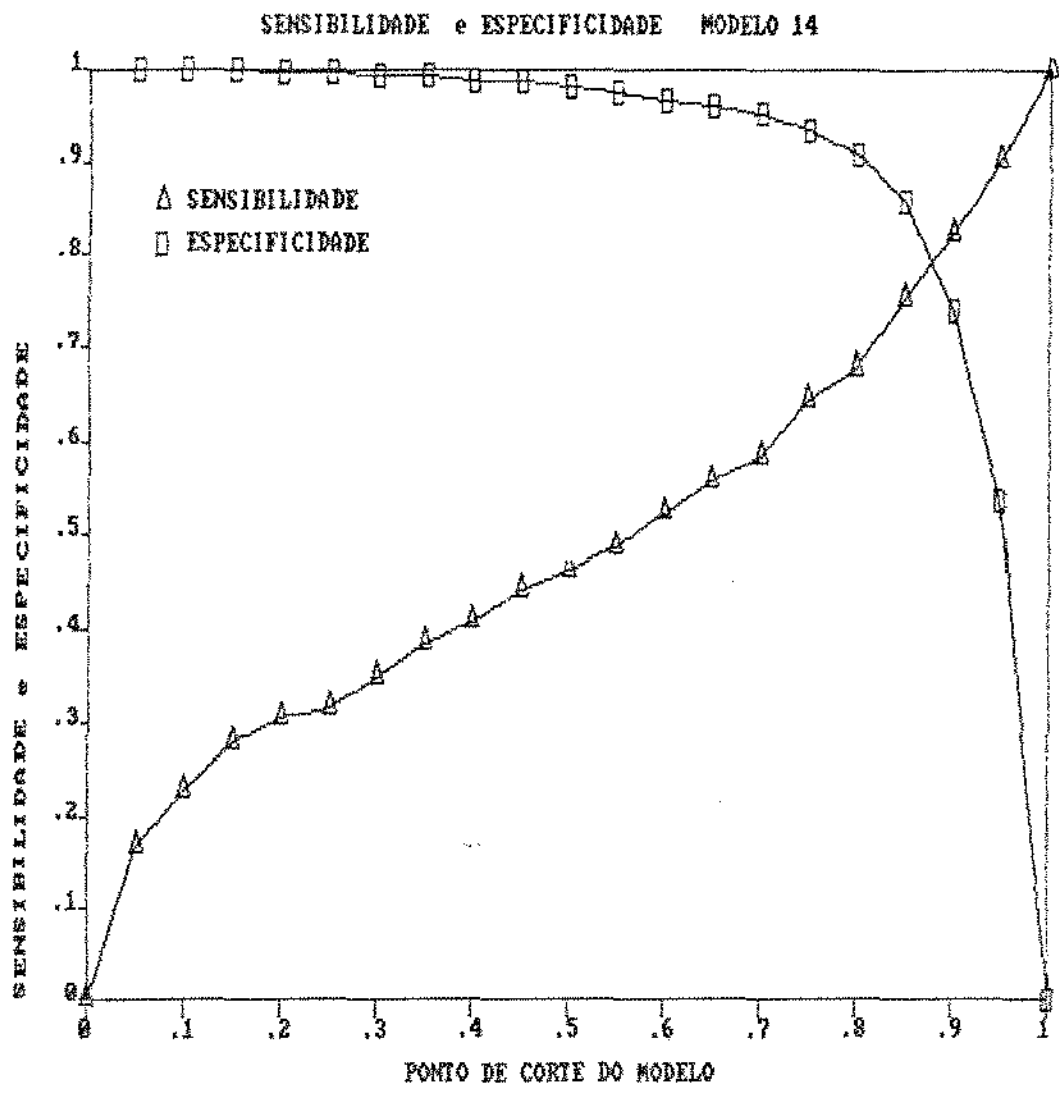


Figura 6.5 - Sensibilidade e especificidade para o MODELO 14.
----- (Fonte: FCM UNICAMP)

A comparação dos MODELOS 14 e 0 quanto aos seus desempenhos pode ser feita através das curvas na figura 6.6.

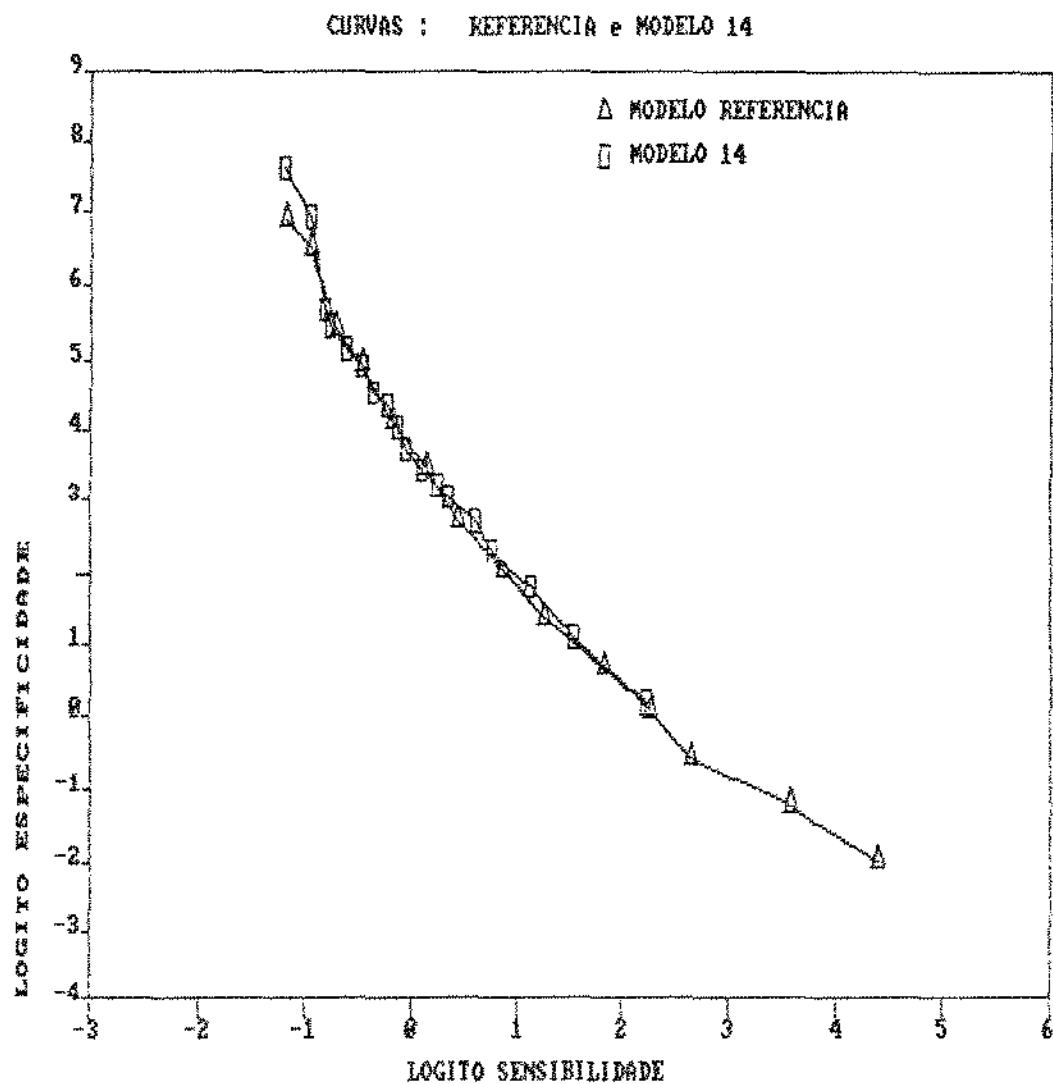


Figura 6.6 - LOGITO P versus LOGITO Q para os MODELOS 0 e 14.
 ----- (Fonte: FCM UNICAMP)

O MODELO 15 consiste na aplicação de regressão logística ao mesmo conjunto de variáveis explicativas e mesmo conjunto de observações utilizados no MODELO 3, tendo como variável dependente IPESO.

MODELO 15

VARIÁVEIS	PARAMETROS ESTIMADOS	ERRO PADRÃO	QUI-QUADRADO	P
INTERCEPT0	-43.9727018	2.5169	305.24	.
NTM1	-0.2703989	0.3056	0.78	0.3763
FILMD1	-0.5491640	0.2137	6.61	.
TOTPAR	0.1233080	0.0515	5.73	0.0166
ANTOBS1	-0.7711726	0.3918	3.87	0.0490
TOTCON	-0.0051763	0.0260	0.04	0.8421
ALTM	-0.0115164	0.0122	0.89	0.3452
ABOR1	-0.1872947	0.1851	1.02	0.3116
SIFI	-1.9337905	0.4832	16.01	0.0001
PRGR	-1.1972746	0.4171	8.24	0.0041
TOXE	0.2643800	0.2226	1.41	0.2350
IDM	-0.0576515	0.0169	11.67	0.0006
ALTP2	1.0171121	0.0564	324.72	.

Tabela 6.6 - Estatísticas calculadas em regressão logística para o
MODELO 15. (Fonte: FCM UNICAMP)

As curvas da sensibilidade e especificidade do MODELO 15 podem ser vistas no gráfico da figura 6.7.

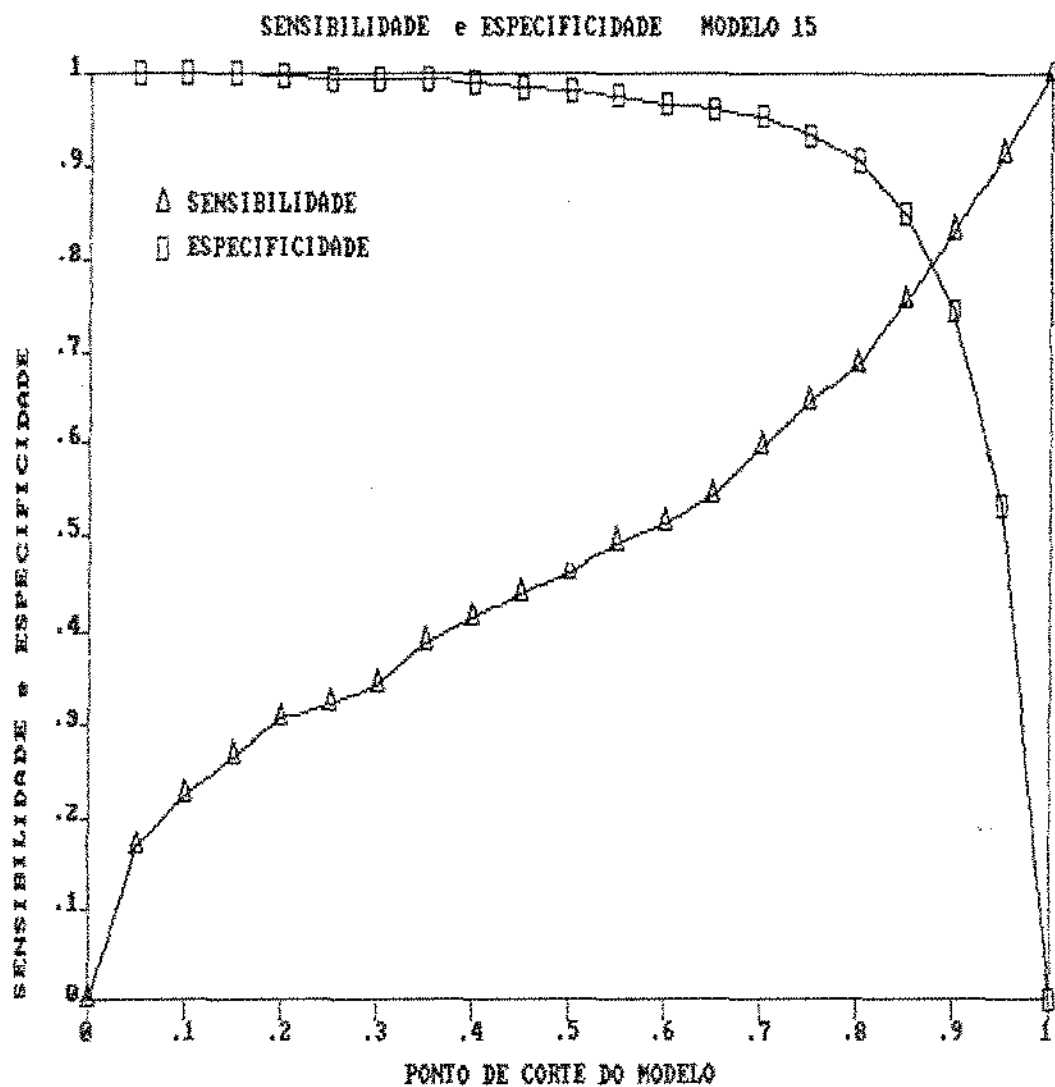


Figura 6.7 - Sensibilidade e especificidade para o MODELO 15.
----- (Fonte: FCM UNICAMP)

A comparação dos MODELOS 15 e 0 quanto aos seus desempenhos pode ser feita através das curvas na figura 6.8.

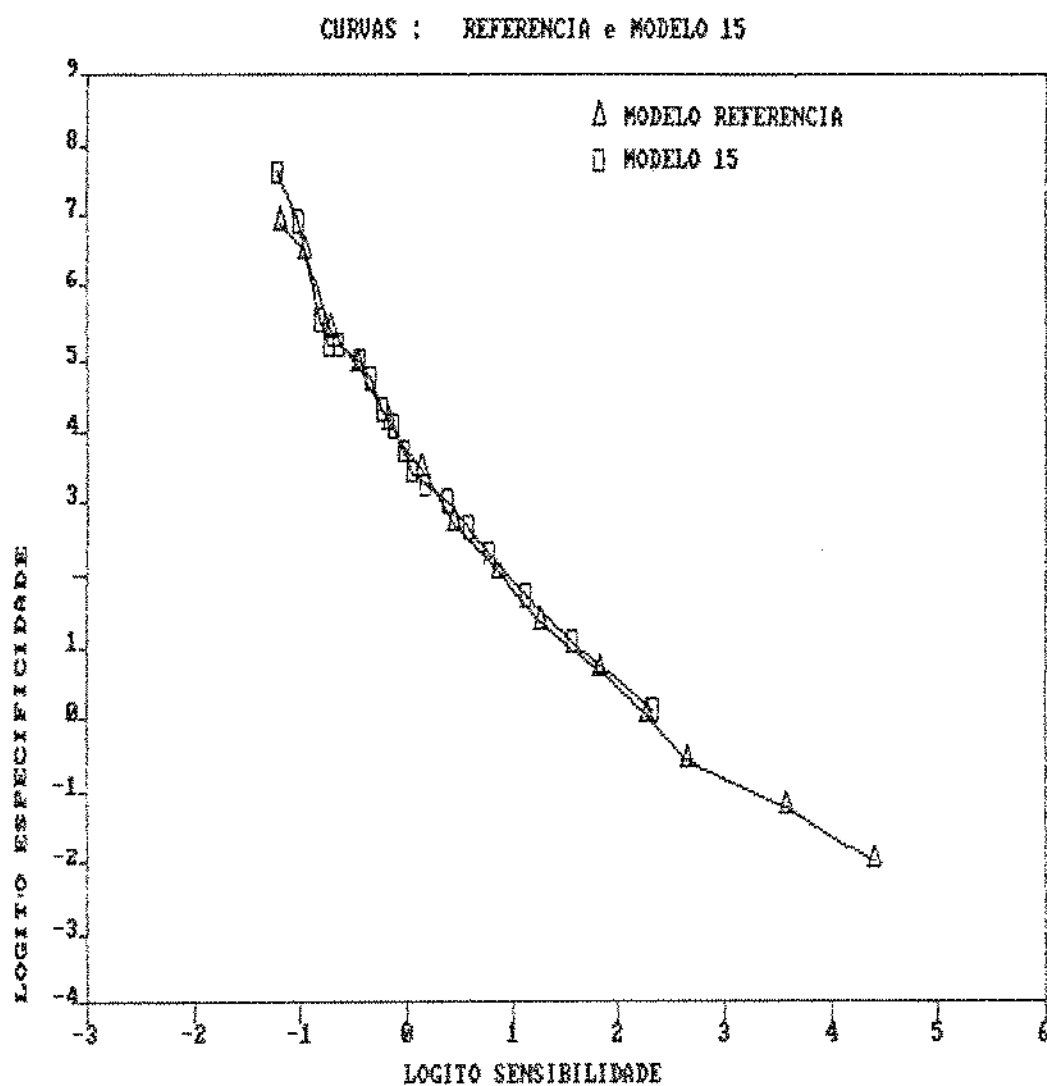


Figura 6.8 - LOGITO P versus LOGITO Q para os MODELOS 0 e 15.
 ----- (Fonte: FCM UNICAMP)

VII - COMPARAÇÃO GERAL DOS MODELOS

Neste capítulo procura-se comparar os modelos do ponto de vista construtivo, em grande parte repetindo comentários feitos nos capítulos V e VI mas aqui de forma mais sucinta, explicitando as razões que levaram à adoção dos procedimentos adotados e que levaram a construí-los na forma em que estão apresentados.

O MODELO 0 é resultado da aplicação dos métodos tradicionais de seleção de variáveis. Aplicando as técnicas "forward", "backward" e "stepwise" ao conjunto de variáveis previamente selecionadas por docentes da Faculdade de Medicina da UNICAMP. Comparando os resultados obtidos em cada técnica, observam-se resultados semelhantes quando é fixado o nível de significância máximo para que uma variável permaneça no modelo. Isto propiciou uma situação confortável para a escolha do modelo. Ao final de todo o processo de detecção de "outliers" e recodificações de variáveis optou-se pelo modelo resultante do ajuste obtido na regressão "stepwise" considerando somente as variáveis com nível de significância $p < 0.05$. Desta forma foi obtido o conjunto de variáveis explicativas para o MODELO 0.

MODELO 0:

$$\begin{aligned} \text{PESO PREVISTO} = & 1399 + 9.15 * \text{ALTM} + 13.1 * \text{TOTCON} + 125 * \text{EDEMA} \\ & - 95.1 * \text{NTM1} + 30.0 * \text{TOTPAR} + 163 * \text{IDG} \\ & - 279 * \text{PRGR} - 64.4 * \text{FILMO1} - 117 * \text{INF} \\ & - 115 * \text{TOXE} - 182 * \text{ANTOBS1} \end{aligned}$$

Tendo estabelecido o MODELO 0 procurou-se melhorar os resultados relativos à detecção das observações com baixo peso, usando para isto as informações disponíveis no arquivo de dados. A primeira tentativa nesse sentido foi a utilização da altura do recém-nascido na previsão de peso do MODELO 0. A utilização da altura diretamente no modelo seria ótimo já que esta variável é altamente correlacionada com o peso. Mas isto não traz vantagem já que a altura do recém-nascido somente está disponível no momento do parto, quando já se tem acesso ao peso do recém-nascido. Uma alternativa, portanto é a utilização da altura prevista por algum modelo cujas variáveis explicativas estejam disponíveis antes do parto. O que se fez então foi ajustar um modelo para a altura utilizando as mesmas técnicas usadas no ajuste do peso. O resultado obtido nesse ajuste foi um conjunto de variáveis explicativas diferente daquele obtido no ajuste do MODELO 0, com variáveis que não constam do mesmo, de forma que as variáveis explicativas da altura não são subconjunto das variáveis explicativas obtidas para o peso no MODELO 0. Isto é um resultado interessante pois possibi-

lita que a nova variável, a altura prevista ALTP1, seja anexada ao conjunto de variáveis explicativas do MODELO 0, formando assim o MODELO 1. Se o conjunto de variáveis explicativas da altura fosse subconjunto do conjunto de variáveis explicativas do peso no MODELO 0, de nada adiantaria anexar a altura prevista ao conjunto de variáveis explicativas do peso pois a altura prevista seria combinação linear destas variáveis e sua inclusão não alteraria em nada o peso previsto, devido ao problema de colinearidade exata.

MODELO 1:

$$\begin{aligned} \text{ALTP1} = & 41.6 + 0.050 * \text{ALTM} + 0.038 * \text{TOTCON} + 0.372 * \text{EDEMA} \\ & - 0.996 * \text{PRGR} - 0.269 * \text{ESTCIV} + 0.712 * \text{IDG} \\ & - 0.576 * \text{INF} - 0.320 * \text{ANTMOR} - 0.915 * \text{ANTOBS1} \end{aligned}$$

$$\begin{aligned} \text{PESO PREVISTO} = & -2905 + 4.14 * \text{ALTM} + 8.07 * \text{TOTCON} + 90.3 * \text{EDEMA} \\ & - 89.0 * \text{NTM1} + 27.8 * \text{TOTPAR} + 89.4 * \text{IDG} \\ & - 170 * \text{PRGR} - 62.4 * \text{FILMO1} - 58.2 * \text{INF} \\ & - 101 * \text{TOXE} - 87.7 * \text{ANTOBS1} + 104 * \text{ALTP1} \end{aligned}$$

Outra alternativa utilizada na tentativa de obter melhoria na detecção de baixo peso em recém-nascidos foi ajustar um modelo para a altura somente nos casos onde o modelo inicial previa o peso muito abaixo do peso real e posteriormente anexar esta altura prevista ao conjunto de variáveis do MODELO 0. Esta alternativa se justifica quando se admite a possibilidade de que o modelo para a altura nestes casos seja diferente daquele em que são consideradas todas as observações e que esta diferença pode ajudar a explicar o peso do recém-nascido. Esta suposição de diferença nos modelos para a altura é reforçada quando através de uma regressão "stepwise" ajusta-se um modelo para a altura nos casos em que o resíduo studentizado é menor que -1, e se obtém um conjunto de variáveis diferente daquele em que se utiliza todas as observações. Anexando a variável altura prevista obtida neste modelo ao conjunto de variáveis do MODELO 0 forma-se então o MODELO 2.

MODELO 2:

$$\begin{aligned} \text{ALTP2} = & 38.0 + 0.026 * \text{ALTM} + 0.085 * \text{TOTCON} + 0.553 * \text{EDEMA} \\ & - 0.815 * \text{TOXE} - 0.031 * \text{IDM} + 0.841 * \text{IDG} \\ & + 1.478 * \text{SIFI} - 0.846 * \text{INF} \end{aligned}$$

$$\begin{aligned} \text{PESO PREVISTO} = & 3217 + 10.6 * \text{ALTM} + 17.7 * \text{TOTCON} + 152 * \text{EDEMA} \\ & - 97.4 * \text{NTM1} + 32.8 * \text{TOTPAR} + 203 * \text{IDG} \\ & - 279 * \text{PRGR} - 65.4 * \text{FILMO1} - 160 * \text{INF} \\ & - 152 * \text{TOXE} - 180 * \text{ANTOBS1} - 47.8 * \text{ALTP2} \end{aligned}$$

Uma constante que se observa nos MODELOS 1 e 2 é que a inclusão da variável altura prevista, anexada ao conjunto de variáveis explicativas do MODELO 0 faz com que a presença de algumas variáveis no modelo não faça sentido, já que o seu nível de significância atinge valores altos, chegando até 0.80. Isto sugere a experiência de se fazer uma regressão "stepwise" após a obtenção da altura prevista no MODELO 2. Nesta regressão concorreram a altura prevista juntamente com as variáveis selecionadas do arquivo de dados no início dos trabalhos, resultando no MODELO 3.

MODELO 3:

$$\begin{aligned} \text{PESO PREVISTO} = & -5923 + 4.31 * \text{ALTM} - 45.7 * \text{ABOR1} - 374 * \text{SIF1} \\ & - 98.4 * \text{NTM1} + 32.9 * \text{TOTPAR} - 6.37 * \text{IDM} \\ & - 281 * \text{PRGR} - 60.9 * \text{FILMO1} + 193 * \text{ALTP2} \\ & + 47.8 * \text{TOXE} - 177 * \text{ANTOBS1} \end{aligned}$$

Outra forma possível para se usar a altura do recém-nascido para auxiliar na previsão de seu peso foi fazer o ajuste da altura nos casos em que houve previsões de peso falsas negativas. Isto significa usar somente as observações onde a previsão de peso foi acima de 2500 gramas e o valor verdadeiro esteve abaixo de 2500 gramas no ajuste da altura. Neste caso repete-se o conjunto de variáveis explicativas da altura e do peso dos MODELOS 1, 2 e 3 somente trocando as observações usadas no ajuste da altura e dando origem aos MODELOS 4, 5 e 6 respectivamente.

No MODELO 4 utilizou-se o mesmo conjunto de variáveis do MODELO 1, com o ajuste para a previsão da altura usando somente as observações com previsões falsas negativas na iteração anterior.

MODELO 4:

$$\begin{aligned} \text{ALTP4} = & 42.9 + 0.020 * \text{ALTM} + 0.057 * \text{TOTCON} + 0.059 * \text{EDEMA} \\ & - 0.0044 * \text{PRGR} + 0.013 * \text{ESTCIV} + 0.280 * \text{IDG} \\ & - 0.330 * \text{INF} - 0.360 * \text{ANTMOR} + 0.230 * \text{ANTOBS1} \end{aligned}$$

$$\begin{aligned} \text{PESO PREVISTO} = & -3304 + 7.17 * \text{ALTM} + 6.32 * \text{TOTCON} + 122 * \text{EDEMA} \\ & - 88.9 * \text{NTM1} + 29.4 * \text{TOTPAR} + 132 * \text{IDG} \\ & - 271 * \text{PRGR} - 63.2 * \text{FILMO1} - 82.6 * \text{INF} \\ & - 98.8 * \text{TOXE} - 206 * \text{ANTOBS1} + 110 * \text{ALTP4} \end{aligned}$$

No MODELO 5 utilizou-se o mesmo conjunto de variáveis do MODELO 2, com o ajuste para a previsão da altura usando somente as observações com previsões falsas negativas na iteração anterior.

MODELO 5:

$$\begin{aligned} \text{ALTP5} = & 42.7 + 0.019 * \text{ALTM} + 0.035 * \text{TOTCON} + 0.242 * \text{EDEMA} \\ & - 0.714 * \text{TOXE} + .00374 * \text{IDM} + 0.319 * \text{IDG} \\ & + 0.743 * \text{SIFI} - 0.275 * \text{INF} \end{aligned}$$

$$\begin{aligned} \text{PESO PREVISTO} = & 6547 + 11.6 * \text{ALTM} + 17.5 * \text{TOTCON} + 153 * \text{EDEMA} \\ & - 95.9 * \text{NTM1} + 30.7 * \text{TOTPAR} + 202 * \text{IDG} \\ & - 281 * \text{PRGR} - 63.5 * \text{FILMO1} - 153 * \text{INF} \\ & - 200 * \text{TOXE} - 180 * \text{ANTOBS1} - 120 * \text{ALTP5} \end{aligned}$$

No MODELO 6 utilizou-se o mesmo conjunto de variáveis do MODELO 3, com o ajuste para a previsão da altura usando somente as observações com previsões falsas negativas na iteração anterior.

MODELO 6:

$$\begin{aligned} \text{PESO PREVISTO} = & -20181 - 0.221 * \text{ALTM} - 47.8 * \text{ABOR1} - 463 * \text{SIFI} \\ & - 99.3 * \text{NTM1} + 33.4 * \text{TOTPAR} - 2.50 * \text{IDM} \\ & - 284 * \text{PRGR} - 61.2 * \text{FILMO1} + 506 * \text{ALTP5} \\ & + 248 * \text{TOXE} - 179 * \text{ANTOBS1}. \end{aligned}$$

Uma outra tentativa de se reduzir o número de previsões falsas negativas foi através de um método iterativo utilizando ponderação. Isto é feito de forma que são atribuídos pesos maiores que um às observações onde ocorreram previsões falsas negativas na iteração anterior, enquanto que nas demais os pesos atribuídos são iguais a um. Este procedimento faz com que o modelo seja mais sensível às observações onde ocorrem previsões falsas negativas na iteração anterior forçando a que haja menores erros nelas. Este método iterativo foi aplicado tendo como variáveis e parâmetros iniciais aqueles obtidos no MODELO 0. Utilizaram-se pesos fixos nos MODELOS 7 e 8 e pesos proporcionais aos erros cometidos na iteração anterior nos MODELOS 9, 10 e 11. Nos ajustes realizados através de métodos iterativos, MODELOS 7 a 11, os parâmetros estimados foram obtidos em três iterações.

No MODELO 7 utilizou-se peso 2 (dois) nas observações com previsão falsa negativa na iteração anterior e peso 1 (um) nas demais observações.

MODELO 7:

$$\begin{aligned} \text{PESO PREVISTO} = & 1334 + 8.78 * \text{ALTM} + 14.5 * \text{TOTCON} + 119 * \text{EDEMA} \\ & - 80.6 * \text{NTM1} + 30.2 * \text{TOTPAR} + 166 * \text{IDG} \\ & - 264 * \text{PRGR} - 82.1 * \text{FILMO1} - 111 * \text{INF} \\ & - 130 * \text{TOXE} - 194 * \text{ANTOBS1} \end{aligned}$$

No MODELO 8 utilizou-se peso 5 (cinco) nas observações com previsão falsa negativa na iteração anterior e peso 1 (um) nas demais observações.

MODELO 8:

$$\begin{aligned}\text{PESO PREVISTO} = & 1255 + 7.92 * \text{ALTM} + 16.3 * \text{TOTCON} + 106 * \text{EDEMA} \\ & - 62.4 * \text{NTM1} + 29.9 * \text{TOTPAR} + 166 * \text{IDG} \\ & - 222 * \text{PRGR} - 110 * \text{FILMO1} - 103 * \text{INF} \\ & - 146 * \text{TOXE} - 205 * \text{ANTOBS1}\end{aligned}$$

No MODELO 9 foram usados pesos proporcionais ao módulo do resíduo nas observações com previsão falsa negativa na iteração anterior e peso 1 (um) nas demais observações.

MODELO 9:

$$\begin{aligned}\text{PESO PREVISTO} = & 1405 + 8.07 * \text{ALTM} + 14.6 * \text{TOTCON} + 115 * \text{EDEMA} \\ & - 87.9 * \text{NTM1} + 28.9 * \text{TOTPAR} + 161 * \text{IDG} \\ & - 255 * \text{PRGR} - 75.1 * \text{FILMO1} - 113 * \text{INF} \\ & - 128 * \text{TOXE} - 175 * \text{ANTOBS1}\end{aligned}$$

No MODELO 10, como no MODELO 9, foram utilizados pesos proporcionais aos módulos dos resíduos nas observações com previsões falsas negativas na iteração anterior e peso 1 (um) nas demais observações.

MODELO 10:

$$\begin{aligned}\text{PESO PREVISTO} = & 1363 + 7.35 * \text{ALTM} + 16.5 * \text{TOTCON} + 103 * \text{EDEMA} \\ & - 74.3 * \text{NTM1} + 28.3 * \text{TOTPAR} + 160 * \text{IDG} \\ & - 231 * \text{PRGR} - 92.7 * \text{FILMO1} - 107 * \text{INF} \\ & - 145 * \text{TOXE} - 178 * \text{ANTOBS1}\end{aligned}$$

No MODELO 11, como nos MODELOS 9 e 10, foram usados pesos proporcionais aos módulos dos resíduos nas observações com previsões falsas negativas na iteração anterior e peso 1 (um) nas demais observações.

MODELO 11:

$$\begin{aligned}\text{PESO PREVISTO} = & 1314 + 6.12 * \text{ALTM} + 18.7 * \text{TOTCON} + 89.2 * \text{EDEMA} \\ & - 55.5 * \text{NTM1} + 27.1 * \text{TOTPAR} + 159 * \text{IDG} \\ & - 165 * \text{PRGR} - 109 * \text{FILMO1} - 107 * \text{INF} \\ & - 179 * \text{TOXE} - 168 * \text{ANTOBS1}\end{aligned}$$

Nos MODELOS 12 a 15 aplicou-se a técnica de regressão logística aos conjuntos de variáveis dos MODELOS 0 a 3. Neste caso estimou-se a probabilidade de que as observações sejam de peso normal, maiores que 2500 gramas.

Os ajustes obtidos nestes modelos são apresentados a seguir, mas a probabilidade, não é obtida diretamente, devendo-se fazer a seguinte transformação para obtê-la:

$$P = \frac{1}{1 + \exp(-L)}$$

Onde P é a probabilidade de se ter um peso normal e L é dado pelos ajustes apresentados abaixo.

No MODELO 12 aplica-se regressão logística ao conjunto de variáveis do MODELO 0.

MODELO 12:

$$\begin{aligned}L = & -5.3550 + 0.01232 * \text{ALTM} + 0.07064 * \text{TOTCON} + 0.23686 * \text{EDEMA} \\ & - 0.23755 * \text{NTM1} + 0.06588 * \text{TOTPAR} + 0.87138 * \text{IDG} \\ & - 1.20859 * \text{PRGR} - 0.53673 * \text{FILMO1} - 0.64985 * \text{INF} \\ & - 0.50266 * \text{TOXE} - 0.73891 * \text{ANTOBS1}\end{aligned}$$

No MODELO 13 aplica-se regressão logística ao conjunto de variáveis do MODELO 1.

MODELO 13:

$$\begin{aligned}L = & -23.773 - 0.00664 * \text{ALTM} + 0.02205 * \text{TOTCON} - 0.06374 * \text{EDEMA} \\ & - 0.17251 * \text{NTM1} + 0.06855 * \text{TOTPAR} + 0.47428 * \text{IDG} \\ & - 0.80294 * \text{PRGR} - 0.54066 * \text{FILMO1} - 0.06955 * \text{INF} \\ & - 0.46644 * \text{TOXE} - 0.62660 * \text{ANTOBS1} + 0.48572 * \text{ALTP1}\end{aligned}$$

No MODELO 14 aplica-se regressão logística ao conjunto de variáveis do MODELO 2.

MODELO 14:

$L = 10.533 + 0.02586 * \text{ALTM}$ ificância $p < 0.05$. Desta forma foi obtido o conjunto de variáveis explicativas para o MODELO 0.

- $0.25788 * \text{NTM1}$ e as variáveis com nível de significância $p < 0.05$. Desta forma foi obtido o conjunto de variáveis explicativas para o MODELO 0.

- $1.22085 * \text{PRGR}$ o "stepwise" considerando somente as variáveis com nível de significância $p < 0.05$. Desta forma foi obtido o conjunto de variáveis explicativas para o MODELO 0.

- $0.83247 * \text{TOXE} \quad \text{NTM1} + 30.0 * \text{TOTPAR} + 163 * \text{IDG}$

No MODELO 15 aplica- $\text{ALTM} + 13.1 * \text{TOTCON} + 125 * \text{EDEMA}$

MODELO 15:

$L = -43.973 - 0.01152 * \text{ALTM}$ citando as razões que levaram à adoção dos procedimentos adotados e que levaram a construí-los na forma em que estão apresentados.

- $0.27040 * \text{NTM1}$ icionais de seleção de variáveis. Aplicando as técnicas "forward", "backward" e "stepwise" ao conjunto de variáveis previamente selecionadas por docentes da Faculdade de Medicina da UNICAMP. Comparando os resultados obtidos em cada técnica, observam-se resultados semelhantes quando é fixado o nível de significância máximo para que uma variável permaneça no modelo. Isto propiciou uma situação confortável para a escolha do modelo. Ao final de todo o processo de detecção de "outliers" e recodificações de variáveis optou-se pelo modelo resultante do ajuste obtido na regressão "stepwise" considerando somente as variáveis com nível de significância $p < 0.05$. Desta forma foi obtido o conjunto de variáveis explicativas para o MODELO 0.

- $1.19727 * \text{PRGR}$ o processo de detecção de "outliers" e recodificações de variáveis optou-se pelo modelo resultante do ajuste obtido na regressão "stepwise" considerando somente as variáveis com nível de significância $p < 0.05$. Desta forma foi obtido o conjunto de variáveis explicativas para o MODELO 0.

+ $0.26438 * \text{TOXE}$ na forma em que estão apresentados.

VIII - DISCUSSÃO E CONCLUSÕES

Como foi indicado na introdução, este trabalho não objetiva discutir e interpretar os modelos e ajustes obtidos para peso de recém-nascidos ou para a probabilidade de baixo peso.

O objetivo principal era investigar diferentes maneiras de obter ajustes e modelos, visando detectar se existem alguns que são marcadamente melhores do que outros, no sentido de que, se usados para triagem, permitiriam a obtenção de níveis adequados de sensibilidade e especificidade.

A chave para o entendimento das conclusões está, em primeiro lugar, no entendimento das curvas de sensibilidade e especificidade versus ponto de corte, como as da Figura 5.3. Na abscissa estão os valores de peso previsto pelo modelo (MODELO 0, no caso). Para cada valor de peso previsto pode ser construída uma tabela 2 X 2, como a da Figura 5.3, na qual determinam-se os valores para sensibilidade e especificidade. Qual o ponto de corte a ser usado numa situação real depende de diversos fatores, entre os quais os mais importantes são o coeficiente de falsos negativos que se está disposto a tolerar e o montante de recursos disponíveis para dispensar atenção diferencial e especial aos casos classificados como sendo previstos de baixo peso. Por esta razão, ao comparar o desempenho de dois modelos para o fim proposto é necessário comparar o par de curvas de um deles com o par de curvas do outro.

A Figura 5.5 apresenta uma forma de se fazer a comparação. Neste gráfico, para diferentes valores de sensibilidade, são apresentados os valores de especificidade para dois modelos, no caso os MODELOS 0 e 1. Se uma das curvas estiver, por exemplo, sempre acima da outra, isto significa que o modelo correspondente apresenta características de desempenho melhor que o outro. Uma forma mais adequada de se fazer a comparação é através de outro gráfico, com valores transformados de sensibilidade e especificidade, como mostrado, por exemplo, na Figura 5.6. Neste gráfico são apresentados os logitos das especificidades de cada um dos modelos versus logitos da sensibilidade. A grande vantagem de se fazer a comparação usando estes gráficos é que as curvas são mais suaves, e muitas vezes próximas de retas. A idéia de se fazer as comparações dessa forma foi obtida de Mardones e Jones (1985).

Pode ocorrer também que um dos modelos se apresente mais adequado apenas para certas faixas de valores de sensibilidade, sem ser melhor sempre. Foi usado como referência o MODELO 0, por ser o primeiro. Poderia ter sido usado qualquer um deles. Na realidade o problema global é um problema complexo de múltiplas comparações, que, por necessidade foi reduzido a comparar o desempenho de cada modelo com um deles, usado como "referência".

A seguir é apresentado um resumo de desempenho dos modelos de 1 a 15, comparado ao do modelo 0.

No MODELO 0 aplicou-se regressão linear múltipla para se obter uma previsão de peso e através desta tentou-se detectar os casos de baixo peso existentes. Este MODELO 0 é considerado então como referência para os outros 15 modelos testados.

No MODELO 1 usou-se a variável ALTP1, altura prevista através de um ajuste utilizando todas as observações disponíveis, como variável auxiliar na previsão do peso, anexada ao conjunto de variáveis explicativas do MODELO 0. Na comparação dos MODELOS 0 e 1 feitas através da figura 5.6 notou-se as curvas referentes aos dois modelos praticamente superpostas, o que indica que os desempenhos de ambos são muito semelhantes.

No MODELO 2 usou-se ALTP2, altura prevista por ajuste que utilizou somente as observações que no MODELO 0 apresentaram resíduos "studentizados" externamente menores que -1, anexada ao conjunto de variáveis explicativas do MODELO 0. Neste caso também, como se pode observar na figura 5.8 os resultados foram bastante semelhantes aos do MODELO 0.

A seguir a variável ALTP2 juntamente com as variáveis disponíveis no início do trabalho foi submetida a uma regressão "stepwise" e foi obtido um novo conjunto de variáveis explicativas que formou o MODELO 3, cujo desempenho foi bastante semelhante ao do MODELO 0 como pode ser visto na figura 5.10.

Nos MODELOS 4, 5 e 6 foram utilizados os mesmos conjuntos de variáveis explicativas dos MODELOS 1, 2 e 3, com a diferença que no ajuste da altura prevista, utilizou-se somente as observações que no ajuste do MODELO 0 apresentaram previsões falsas negativas. Os resultados na detecção de baixo peso foram praticamente idênticos aqueles obtidos no MODELO 0, como pode ser observado nas figuras 5.12, 5.14 e 5.16.

Nos MODELOS 7 e 8 utilizou-se o mesmo conjunto de variáveis obtido para o MODELO 0 num método iterativo onde foram aplicados pesos fixos, 2 e 5 respectivamente, às observações que no MODELO 0 apresentaram previsões falsas negativas. Nestes dois casos os resultados foram equivalentes aos obtidos no MODELO 0 como se observa nas figuras 5.18 e 5.20.

Nos MODELOS 9 a 11 foram aplicadas ponderações proporcionais aos desvios nas previsões de peso quando estas eram falsas negativas, quanto maiores os desvios, maiores os pesos atribuídos às observações que apresentavam estes resíduos. Novamente os resultados obtidos na detecção de baixo peso foram semelhantes aos do MODELO 0, como se pode observar nas figuras 5.22, 5.24 e 5.26.

Nos MODELOS 12 a 15 utilizou-se regressão logística nos mesmos conjuntos de variáveis dos MODELOS 0 a 3. Nos gráficos das figuras 6.2, 6.4, 6.6 e 6.8 pode-se notar a equivalência destes

modelos com o MODELO 0 em termos de detetar baixo peso em recém-nascidos, muito embora na curva referente ao MODELO 15 tenha havido uma pequena diferença em favor do mesmo na metade inferior direita do gráfico, esta diferença é muito pequena e não foi considerada.

Do ponto de vista estatístico o que se tem é que, dentre as técnicas classificatórias testadas, não há evidências para considerar que elas não são equivalentes, sob as condições em que foram empregadas, ou seja, tendo uma variável dependente contínua deseja-se detectar quais os casos que estão acima de um certo limite e quais estão abaixo deste, utilizando para isto um conjunto de variáveis explicativas. Este resultado é bastante surpreendente. O que se esperava era encontrar modelos, possivelmente usando regressão múltipla com pesos, em que o desempenho fosse melhor. Há casos de mesmo tipo em que o desempenho obtido com regressão logística é, por exemplo, marcadamente melhor do que o que se obtém usando técnicas classificatórias simples baseadas em risco relativo, como em Mardones e Jones (1985), Figuras 7 e 8.

Ainda sob o ponto de vista estatístico temos que a regressão múltipla e a regressão logística apresentaram desempenhos semelhantes na detecção de casos de baixos pesos. Este é um resultado não esperado já que se tratam de técnicas bastante diferentes, na primeira temos como variável dependente uma variável contínua e na outra tentamos explicar uma variável binária.

Desta maneira a escolha deve ser feita, então usando outros critérios, como simplicidade de uso e adequação da interpretação do modelo do ponto de vista médico, ou de saúde em geral.

Alguns comentários adicionais que podem ser feitos sobre os resultados são os seguintes:

1 - Em todos os ajustes as medidas tradicionais de adequação são baixas. No caso das regressões lineares, por exemplo, os valores do R quadrado são sempre inferiores a 50%.

2 - Apesar do que se menciona em 1 acima, o desempenho dos modelos, com o fim de triagem, parece ser bastante adequado. Com o MODELO 0, por exemplo, usando um ponto de corte de aproximadamente 2.900 gramas pode-se trabalhar com um sistema em que a sensibilidade e a especificidade são, ambas, próximas de 80%.

Os modelos, no entanto, são complexos ao se considerar o uso dos mesmos na prática. Certamente um ajuste com mais de três ou quatro variáveis, com coeficientes fracionários, só poderá vir a ser usado com fins de triagem e classificação em um hospital de referência, como um hospital universitário, provavelmente empregando microcomputadores para fazer os cálculos.

Convém lembrar que os resultados aqui apresentados não podem ser generalizados para nenhuma população, já que os dados utilizados referem-se somente às gestantes que procuraram e foram atendidas na Maternidade da FCM da UNICAMP, não se tratando por-

tanto de uma amostra de população, e sim de partos desproporcionalmente de alto risco, por tratar-se de uma maternidade de referência.

IX - ANEXO 1 - FICHA OBSTÉTRICA

Neste anexo é apresentada uma cópia da FICHA OBSTÉTRICA do Departamento de Tocoginecologia da Faculdade de Ciências Médicas da UNICAMP que é preenchida em todos os casos de internação para parto.

FICHA OBSTÉTRICA
HOSPITAL DAS CLÍNICAS - UNICAMP

NONE

RESIDÊNCIA _____

LOCAL DE NASCIMENTO:	PROCEDÊNCIA:
----------------------	--------------

OBSERVAÇÕES_____

	1	2	3	4	5	6	7
ESTADO CIVIL	SOLTEIRO	CASADO	ANULADO	DESOLTO	VIVO	MORTO	

DUM D.P.P.

CONVULSIVA	SEX		IDADE DA MENARCA		11 12	RITMO MENSTRUAL	1	2	3	4	5
	RAÇA			REGULAR			IRREGULAR	SALENTE	IGN		

ALCOOLISMO

ANTES DE ENGRAVIDAR	DURANTE GRAVIDEZ
<input type="checkbox"/> NUNCA BEBEU	<input type="checkbox"/> NÃO BEBEU
<input type="checkbox"/> < 1231 PINGA OU EQUIV. X DIA	<input type="checkbox"/> < 1231 PINGA OU EQUIV. X DIA
<input type="checkbox"/> > 1231 PINGA OU EQUIV. X DIA	<input type="checkbox"/> > 1231 PINGA OU EQUIV. X DIA

GESTACE

ULTIMO PARTO NA MESSE:			14 18	FILHOS VIVOS ATUAL //			18 17	NATIMORTOS			NR
GRAV GEMELAR			10	CEÁREAS			30	TOTAL GRAVÍDEZ			21 22
TOTAL ABORTOS			23	FILHOS MORTOS			34	TOTAL PARTOS			75 76
ABORTOS ESPONTÂNEOS _____			ABORTOS PROVOCADOS _____								

	1	2	3	4	5	6	7	8	9	10	11
27. GRUPO SANGUÍNEO	A	B	AB	O	-	-	-	-	-	-	IGN
28. ANTÍGENS RH ENF. SENSIBILIZADO	RH +	1-1 NÃO SENS	1-1 SENS S/ ANT DE ENF. HEMOL DO RH	1-1 SENS O ANT DE ENF. HEMOL DO RH	1-1/1/1/1 SENSIBILIZADO	-	-	-	-	-	IGN
29. ANTECEDENTES MÓRBIDOS	SEM	TBC	SÍFILIS	DIABETE	CARDIOPATIA	INFECÇÃO UTERINA	NÃO FETADO	SEQUELA DO 2º DE PR. VIS. OU INF. RIMES	CIRURGIA ABDOMINAL	OUTROS	IGN
30. ANTECEDENTES GINECOLÓGICOS	SEM	TORÇÃO	METFORMIA DIA DUR A GRAVIDEZ	METFORMIA DIA POS PARTO	GRAVIDEZ PROLONGADA	MAIOREM CONGÊNITA	INFECÇÃO PUERPERAL	POLÍMORFO AMNIO	PARTOS PREMAT	OUTROS PATOL. UTERO	IGN
31. USO ANTECEDENTES ANTERIORES ESTA VERIFICAÇÃO	NÃO USAR	GASTAGENO SUSPENSO GRAVIDEZ PLANEJADA	GASTAGENO SUSPENSO POR INTOX. LERANÇIA	GESTAGENO INTERROMPIDO GRAVIDEZ NÃO PLANEJ.	OU RETIRO DO GRAVIDEZ PLANEJADA	OU RETIRO DO PM NÃO ADOPTADO	OU FETADO	OUTRO ANTICÓMP. GRAVIDEZ PLANEJADA	OU OUTRO ANTICÓMP. GRAVIDEZ PLANEJADA	OUTRO ANTICÓMP. GRAVIDEZ NÃO PLANEJ.	IGN
32. ESPÉCIE SANGUÍNEA	U	1-2	3-4	5	6-7	8-9	10-11	12 ou Mais	-	-	IGN

OBSERVAÇÕES

[illegible]

IN TOTAL COMPARAÇÃO	30	INÍCIO DE GESTAÇÃO	30	RISCO OBSTÉTRICO	30
		RISCO PRENATAL	30		30

DE NÃO TER PRENATAL PORQUE NÃO FEZ)

ADMITIDA PARA PARTO

DATA / / HORA : :

EXAME FÍSICO GERAL

Motivo de consulta

EST. GERAL <u> </u>	<table border="1"> <tr><td>PESO HABITUAL</td></tr> <tr><td>PESO NO PRESENTE</td></tr> <tr><td>DIFERENÇA PONDERAL</td></tr> </table>	PESO HABITUAL	PESO NO PRESENTE	DIFERENÇA PONDERAL	ALTURA <table border="1"><tr><td> </td><td> </td><td> </td><td> </td></tr></table>								
PESO HABITUAL													
PESO NO PRESENTE													
DIFERENÇA PONDERAL													
PULSO <u> </u> P.A. <u> </u> T. AXILAR <u> </u>	SEM. DE AMENORRÉIA <table border="1"><tr><td> </td><td> </td><td> </td><td> </td></tr></table>												
CABEÇA E PESCOÇO <u> </u>	DIFERENÇA PONDERAL <table border="1"><tr><td>1. IGUAL</td><td> </td><td> </td><td> </td></tr><tr><td>2. AUMENTOU</td><td> </td><td> </td><td> </td></tr><tr><td>3. DIMINUIU</td><td> </td><td> </td><td> </td></tr></table>	1. IGUAL				2. AUMENTOU				3. DIMINUIU			
1. IGUAL													
2. AUMENTOU													
3. DIMINUIU													
TÓRAX (mamas, coração, pulmões) <u> </u>	DIFERENÇA PONDERAL : Kg <table border="1"><tr><td> </td><td> </td><td> </td><td> </td></tr></table>												

ABDOMEM

EXTREMIDADES

	1	2	3	4	5	6	7
EDEMA	SEM	MAIORES	TIBIAIS	ABDOMEN	ANASARCA	IGUALADO	AD

OUTROS DADOS

INÍCIO DO TRABALHO DE PARTO DATA / / HORA : : RUTURA DA BOLSA DATA / / HORA : :

EXAME OBSTÉTRICO Nº B.C.F. Xmin DINÂMICA

T	ESTADO DO COLO	DILAT	MPS E LIG OVULAR	APRESENTAÇÃO	CAVID FOS C	PLANO	PROMONTÓRIO	C. DIAG	C. VERA	ESP. CÍATICAS	SEXO
V											
R											

OUTROS DADOS

DIAGNÓSTICOS DE ADMISSÃO	INDICAÇÕES NA ADMISSÃO E PLANO DE ASSISTÊNCIA	NOME RESPONSÁVEL

PARTO: HORA : : DIA / / MES / ANO

--	--	--	--

180 DURAÇÃO F. DILATAÇÃO 180 T. ADMISSÃO AO PARTO 180 T. R. M. PARTO 180 DURAÇÃO F. EXPULSÃO

OBSERVAÇÕES (DISTÓCIAS, ACIDENTES, INTERVENÇÕES, etc.)

PARTO ATENDIDO POR NOME LEGÍVEL (LETRA DE FÓRMAL)

CODIFICAÇÃO DE DADOS SOBRE EVOLUÇÃO DO TRABALHO DE PARTO

	1	2	3	4	5	6	7	8	9	10	11
30	TIPO DE RUPTURA DE MEMBRANAS	SEM PREMATURA	SEM ATÉ 5 cm	SEM 5-9 cm	SEM > 9 cm	RUPURA ALTA	SEM ATÉ 5 cm	SEM 5-9 cm	SEM > 9 cm	CEÁREA C/MEIO INTERIORES	PARTO COM DO SERVIÇO
31	ESTADO DO LÍQUIDO OVULAR	CLARO	LIGERAMENTE TINGIDO DE AMARELO	TINGIDO DE AMARELO	MACONTO ESPERMO	MACONTO ANTO	NECROCORADO	PURULENTO	COM MANTO CHIRRO	AMORFO	INFECÇÃO
32	QUANT. LÍQUIDO OVULAR	NORMAL	OLIGO ANGIO	POLIO GRAMING	-	-	-	-	-	-	1 CM
33	S.C.F. SINTOMAS GRAVES	NORMAL 170-180	DM 1	TRADUCAR. DIA (HEV ou H)	DM UNILATERAL	DM 11	BRADICAR. DIAUS TENTADA	DESAP. FOCO SEM SINAIS ROP.UMENTO	DESAP. FOCO COM SINAIS ROP.UMENTO	AUSENTE NA ADMISSÃO	DESAP. FOCO DUA MAMMARI
34	MT. TUBEROS DEADE PLUTURA DE MEMBRANAS	1 DO NENHUM	2	3	4	5	6	7	8	9 cm +	COM MEMBR. INTERIORES DO PARTO FORA DO SERVIÇO
35	ANALGESIA NO PER. DILATACÃO	SEM	MEKRO DO BENZODIAZ	BLOO EPID SIMPLES OU COMIN	BLOO RAO RAO OU ALTO	MEKRO DO BENZOD + BLOO EPID	MEKRO DO BENZOD + BLOO RAO	BLOO EPID + BLOO RAO	MEKRO DO BENZOD + BLOO EPID + BLOO RAO	MEKRO DO BENZOD + BLOO RAO + OUTRAS	OUTRAS
36	HORAS ANULO ANTES PULSIVO	SEM	1	2	3	4	5	6	7	8	9 cm +
37	INDICAÇÃO DE OCTOGONIA	SEM OCTOGONIA	GRAVIDEZ PROLONGADA	AMNIOCORRE PREMATURA	INFECÇÃO OVULAR	DIABETE	CONDUÇÃO DE PARTO	TOXEMIA	UNTO FETAL	SEM 1	OUTRAS

CODIFICAÇÃO DE DADOS SOBRE O PARTO

	1	2	3	4	5	6	7	8	9	10	11
38	HORA PARTO	0000-0530	0600-0130	0200-1130	1200-1330	1400-1930	2000-2130	PARTO FORA DO SERVIÇO	-	-	1 CM
39	FORMA DE INÍCIO DE PARTO	ESPONTÂNEA	INDUÇÃO DISTOLÓGICA PARA	CEÁREA C/MEIO INTERIORES	AMORFO	ABUREL	INDUÇÃO DISTOLÓGICA + ABUREL	PROSTAGLANDINA	-	OUTRAS	1 CM
40	FORMA DE TERMINO DE PARTO	ESPONTÂNEO	ASSISTIDO EM PÉLVICA	EXTRAÇÃO PÉLVICA	GRANDE EXTRAÇÃO PÉLVICA	FÓRCIPE	VACUO	VACUO + FÓRCIPE	FÓRCIPE PRAG. + CEÁREA	VACUO + CEÁREA	CEÁREA
41	APRESENTAÇÃO	VÉRTECE	FACE	FRONTE	DEFALICA INDETERMINADA	PÉLVICA COMPLETA	PÉLVICA INDETERMINADA	PÉLVICA INDETERMINADA	COMINICA	OBELIGUA	OUTRA
42	VARIANTE DE POSICÃO	LA	RY	EP	DR	DT	OP	PUBICA	SACRA	CEÁREA SEM DIAU	OUTRAS
43	DURAÇÃO PERÍODO DILATACÃO	SEM PERÍODO DILAT	ATE 3 HORAS	ATE 4 HORAS	ATE 5 HORAS	ATE 6 HORAS	ATE 7 HORAS	ATE 8 HORAS	ATE 9 HORAS	ATE 10 HORAS	ATE 11 HORAS
44	TEMPO DESDE ADMISSÃO ATÉ AO PARTO	ATE 30 min	ATE 1 HORA	ATE 2 HORAS	ATE 3 HORAS	ATE 4 HORAS	ATE 5 HORAS	ATE 6 HORAS	ATE 7 HORAS	ATE 8 HORAS	ATE 9 HORAS
45	TEMPO DESDE R.M. AO PARTO	ATE 3 HORAS	ATE 4 HORAS	ATE 5 HORAS	ATE 6 HORAS	ATE 7 HORAS	ATE 8 HORAS	ATE 9 HORAS	ATE 10 HORAS	ATE 11 HORAS	ATE 12 HORAS
46	DURAÇÃO PERÍODO EXPULSIVO (min)	CEÁREA ATÉ EXPULSIVO	ATE 5 min	ATE 10 min	ATE 15 min	ATE 20 min	ATE 25 min	ATE 30 min	ATE 35 min	ATE 40 min	ATE 45 min
47	PRINCIPAL	DI PARTO VAGINAL	DI SÓFAMENTO FETAL	DI DESP. CEF. PÉLVICA	DI COMPLETO MATO	DI PÉLVICA OU COMINICA	DI OUTRAS DIST. APRES	DI PLACENTA PRÉVIA	DI GRANDEZ PROLONG.	DI INFECÇÃO OVULAR	DI SINAIS FATOR RH
48	INDICAÇÃO DE CEÁREA	11 TOXEMIA	12 AM. CES	13 ACIDENTE COMATO	14 PRIMP. PARA TARDIA	15 ANT. INFER. OUTROTER	16 DIABETE	17 DIST. DE PARTE MOLE	18 OPF	19 SEM INDICAÇÃO	20 OUTRAS
49	ANESTESIA PARA EXP. OU INTERVENÇÕES	SEM	GERAL	ANQUIDEA	SIMPLES	COMINICA	FLUORENA	LOCAL	CAUDAL	CEÁREA DESDE PREPARTO	OUTRAS
50	INÍCIO OU VACUO	SEM	ALPHO	DISTOGIA ROTACÃO	DISTOGIA PROGRESSÃO	ANNEVIACÃO PERÍODO EXPULSIVO	CASECA DEGRADADA	FÓRCIPE NA CEÁREA	OUTRAS	OUTRAS	OUTRAS
51	OPERAÇÕES COMPLEMENTARES MENORES	SEM	EPISIOR. RAPIA	SUTURA DO COLO	EPISIOR. FIA + SUT. DO COLO	SUTURA DESGARR. VAGINAL	SUT. DESG. VAG. + SUT. DO COLO + EPISIOR	SUTURA DESG. VAG. + SUT. DO COLO + EPISIOR	SUTURA DESG. VAG. + SUT. DO COLO + EPISIOR	SUTURA DESG. VAG. + SUT. DO COLO + EPISIOR	SUTURA DESG. VAG. + SUT. DO COLO + EPISIOR
52	OPERAÇÕES COMPLEMENTARES MAIORES	SEM	ESTERELIZ. INTRA. PARTO	ESTERELIZ. POS. PARTO IMEDIATO	SUTURA RUTURA UTERINA	SUTURA DESG. TOTAL	HISTEREC. TOTAL	LIGADURA HIPÓCAS. TRICA	SUT. RUTURA UTERINA + ESTERELIZ	OUTRAS	OUTRAS

OBSERVAÇÕES

DEQUITAÇÃO

(1) DURAÇÃO

(2) PESO

COMPLETA	ADER. ANORMAIS	INCOMPLETA	ACRETA
----------	----------------	------------	--------

SEM REVISÃO	CURAGEM	CURETAGEM
-------------	---------	-----------

OBSERVAÇÕES (ACIDENTES, ANOMALIAS PLACENTÁRIAS, INTERVENÇÕES, etc.)

MÊS

CODIFICAÇÃO DE DADOS DA DEQUITAÇÃO

	1	2	3	4	5	6	7	8	9	10	11
53	DURAÇÃO (min)	1-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
54	PESO PLACENTA (g)	700-799	800-899	900-999	1000-1099	1100-1199	1200-1299	1300-1399	1400-1499	1500-1599	1600-1699
55	TIPO DEQUITAÇÃO	NATURAL	DIRIGIDO	CRESE	EXTRAÇÃO MANUA	CURAGEM	EXTRA MAN. + CURAGEM	EXTRA. C/ MAN. UTER	HISTEREC. POR PLAC. ACNE	HISTEREC. POR PLAC. ACNE	OUTRAS
56	CORDÃO	NORMAL	CIRCULAR SIMPLES	CIRCULAR DUPLA	MATE DE DUAS	NÓ	PROLAP. CORDÃO	PROCIDENCIA	PREZIADADA	CIRCULAR + NO	OUTRAS

F	O	2	1	77	60
---	---	---	---	----	----

SAME Nº _____ Nº _____

RECÉM-NASCIDO

RESUSCITAÇÃO	SEM	COM	IGN	ENV. ANAL. PATOL.	VIVO	SIM	NÃO	IGN
	1	2	3	4	5	6	7	8

ALTURA **PESO**

OBSERVAÇÕES (LESÕES, MALFORMAÇÕES, MORBILIDADE, etc.) _____

A 1 mm 16
P 17
B 5 mm 18
A 19
B 10 mm 20
USAR DESTAÇÃO CLÍNICA (CAPURRO) 21

MÉDICO BERÇÁRIO _____

DADOS DO RN	1	2	3	4	5	6	7	8	9	10	11
22 SEXO E TIPO DO PARTO	MASCULINO SIMPLES	FEMININO SIMPLES	MASCULINO GÊMELO	FEMININO GÊMELO	MASCULINO TRÍPLIO	FEMININO TRÍPLIO					IGN
23 MALFORMAÇÃO CONGÊNITA	SEM	ACRORIA DO ANENCEFALIA	HIDROCE SÁLIA	ANENGO CÉL	CARDIOPATIA CONGÊNITA	MONGOLISMO	UNILATERAL	LABIO LEPORINO	SINUS DOBROS	OUTRAS	IGN
24 MORBIDADE NEONATAL OU CAUSA DE MORTE	01 SEM	02 SÍNDROME DE CHOKKY	03 SÍNDROME DE CHOKKY	04 SÍNDROME DE CHOKKY	05 SÍNDROME DE CHOKKY	06 SÍNDROME DE CHOKKY	07 SÍNDROME DE CHOKKY	08 SÍNDROME DE CHOKKY	09 SÍNDROME DE CHOKKY	10 SÍNDROME DE CHOKKY	IGN
25 ESTADO DO RN NA ALTA	VIVO	ENFERMO	MORTO	MORTO	MORTO	MORTO	MORTO	MORTO	MORTO	MORTO	IGN

OBSERVAÇÕES (TRATAMENTO, ANÁLISE, etc.) _____

COMPLICAÇÃO DO PARTO	1	2	3	4	5	6	7	8	9	10	11
27 EVOLUÇÃO DA PRESSÃO ARTERIAL	NORMAL	CARDIOPATIA	CHOC	CHOC	HIPERTEN	HIPERTEN	HIPERTEN	HIPERTEN	ACIDENTE VASCULAR		IGN
28 TOXEMIA E HEMORRAGIA	SEM	HIPERTEN	HIPERTEN	HIPERTEN	HIPERTEN	HIPERTEN	HIPERTEN	HIPERTEN	HIPERTEN	HIPERTEN	IGN
29 DIABETES	SEM	PRÉ-DIABETES	DIABETES	DIABETES	DIABETES	DIABETES	DIABETES	DIABETES	DIABETES	DIABETES	IGN
30 INFECÇÕES	SEM	TOXOPLASMOSE	TUBERCULOSE	INF. UTR. SÍNDROME	INF. UTR. SÍNDROME	CONDILOMATA	VÍRUS	ESTAFILOCOCO	SÍFISIS	OUTRAS	IGN
31 HEMORRAGIAS E IMPORTÂNCIA	SEM	DO 1º TRIMESTRE	DO 2º TRIMESTRE	PLACENTA PRÉVIA	DO 3º TRIMESTRE	DO 4º TRIMESTRE	DO 5º TRIMESTRE	DO 6º TRIMESTRE	DO 7º TRIMESTRE	DO 8º TRIMESTRE	IGN
32 CIRURGIA DURANTE GRAVIDEZ	SEM	CERCLA	CERCLA	CERCLA	CERCLA	CERCLA	CERCLA	CERCLA	CERCLA	CERCLA	IGN

COMPLICAÇÃO DO PARTO	1	2	3	4	5	6	7	8	9	10	11
33 HEMORRAGIAS	SEM	RETENÇÃO PLACENTAL	ADHESÃO PLACENTAL	RETENÇÃO PLACENTAL	INFECÇÃO UTERINA	INFECÇÃO UTERINA	INFECÇÃO UTERINA	INFECÇÃO UTERINA	INFECÇÃO UTERINA	INFECÇÃO UTERINA	IGN
34 INFECÇÕES	SEM	ENDOMETRITE	INFEÇÃO SÉRIA	MASTITE	TRONCO FLETITE	ANEXITE	ABCESSO DE DOUGLAS	PERITONITE	SEPSIS	OUTRAS	IGN
35 CIRCUNSIÇÃO	SEM	DESCONF. PERÍDIA	EVISC. RACÃO	COMO EXTRANEO	HEMATOMA	HEMATOMA	HEMATOMA	HEMATOMA	HEMATOMA	HEMATOMA	IGN
36 ESTADO DA MÃE NA ALTA	SADA	VIVA EM COMPLEXO	VIVA EM COMPLEXO	VIVA EM COMPLEXO	VIVA EM COMPLEXO	VIVA EM COMPLEXO	VIVA EM COMPLEXO	VIVA EM COMPLEXO	VIVA EM COMPLEXO	VIVA EM COMPLEXO	IGN
37 CURETAGEM	SEM	POSSÍVEL	PURPÚRA	NEINGRÊSO	ITENS	ITENS	ITENS	ITENS	ITENS	ITENS	IGN
38 ESTUDO PLACENTAR	SEM	ANÁLISE	ANÁLISE	VÍDEO	MONIT. RIZACÃO	R.R.	MADE DE UN	FOROS	QUALQUER ANTERIOR	80 OUTROS	IGN
39 TABAGISMO	NUNCA FUMOU	NÃO FUMOU DURANTE A GRAVIDEZ	DO 1º A 3º MÊS	DO 4º A 6º MÊS	DO 7º A 9º MÊS	DO 10º MÊS	DO 11º MÊS	DO 12º MÊS	DO 13º MÊS	DO 14º MÊS	IGN

ALTA DATA _____

DIAS DE INTERNAÇÃO DA MÃE

ANTES DO PARTO	40	41	DEPOIS DO PARTO	42	43	REINTER. NACAO	44	45
----------------	----	----	-----------------	----	----	----------------	----	----

CONDIÇÃO _____ **CAUSA** _____

MÉDICO _____

DIAS INTERNAÇÃO DO RN **IDENTIFICAÇÃO**

X - REFERÊNCIAS

- Anderson, S.; Auquier, A.; Hauck, W.W.; Oakes, D.; Vandaele, W.; Weisberg, H.I. (1980). Statistical Methods for Comparative Studies. John Wiley & Sons, Nova Iorque.
- Battaglia, F.C. e Lubchenco, L.A. (1967). A practical classification of newborn infants by weight and gestational age. *Journal of Pediatrics*, 21, 159-163.
- Bickel, P.J. & Doksum, K.A. (1977). *Mathematical Statistics: Basic Ideas and Select Topics*. Holden-Day, San Francisco.
- Belsley, D.A. ; Kuh, E. e Welsch, R.E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Colinearity*. John Wiley & Sons, Nova Iorque
- Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J (1984). *Classification and regression trees*. Wadsworth International Group, Belmont, California.
- Capurro, H.; Konichezky, S.; Fonseca, D.; Caldeiro-Barcia, R. (1978). A simplified method for diagnosis of gestational age in the newborn infant. *J. Pediatrics* 93, 120-122.
- Carvalho, J.F. & Dachs, J.N.W. (1985). Diagnóstico em Regressão. Notas de Minicurso do VI SINAPE, UFRJ.
- Cox, D.R. (1970). *The Analysis of Binary Data*. Methuen, Londres.
- Draper, N.R. & Smith, H. (1981). *Applied Regression Analysis*. 2a. Edição. John Wiley & Sons, Nova Iorque.
- Jurado Garcia, E.; Abarca Arroyo, A.; Osorio, R.C.; Campos, O.R.; Saavedra, M.A.; Alvarez de los Cobos, J. e Parra Jimenez, S. (1970) El crecimiento intrauterino. Evaluación del peso y la longitud corporal fetal en la ciudad de México. Análisis estadístico de 16807 nacimientos consecutivos de producto único, vivo. *Boletín Médico del Hospital Infantil (México)*, 22, 163-195.

- Marcondes-Restat, F. e Jones, G. (1985). Prediccion de Fallas en el crecimiento infantil. Rev. Chil. Nutr. Vol.13, No 3, 156-169.
- Montgomery, D.C. & Peck, E.A. (1982). Introduction to Linear Regression Analysis. John Wiley & Sons, Nova Iorque.
- Organização Panamericana de Saúde. (1987). Analisis Dela Investigacion de Riesgo Materno Infantil: Resultados Preliminares. Cuba 1984.
- Sabatino, H.; Araki, M. (1979). Computadorized perinatal evaluation an experience at the state university of Campinas. Conferência do 10º aniversário da associação brasileira de físicos em medicina - São Paulo.
- Sabatino, H.; Silva, J. C. G. da; Silva, J. L. C. P.; Faundes, A. (1980). Experiência com o uso de uma ficha clínica obstétrica pré-codificada. Revista Brasileira de Ginecologia e Obstetrícia, jan/fev/mar de 1980, 5-10.
- Sabatino, H.; Vichi, A.; Hardy, E.; Silva, J. L. C. P.; Silva, J. C. G. da (1978). Manual de Instruções para o Preenchimento da Ficha Obstrétrica. FCM - UNICAMP.
- Sabatino, H. (1987)
Informação Verbal.
- Statistical Analysis System. (1980) Supplemental Library User's Guide. 1980 ed. SAS Institute Inc.
- Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley Publishing Company.
- Villar, J. e Belizan, J.M. (1986). The evaluation of the methods used in the Diagnosis of intrauterine growth retardation. Obstetrical and Gynecological Survey, Vol 41, No 4, 187-199.
- Yerushalmy, J., Van Den Bergh, B.J., Erhardt, C.L. e Jacobziner, H. (1965). Birth weight and gestation as indexes of "immaturity". Neonatal mortality and congenital anomalies of the "immature". American Journal of Diseases Children, 109, 43-57.