

Universidade de São Paulo

Faculdade de Saúde Pública

**Estimação de sub-registro de nascidos vivos
pelo método de captura e recaptura**

Bianca Schmid

**Tese de Doutorado apresentada ao
Programa de Pós-Graduação em
Saúde Pública para obtenção do
título de Doutora em Saúde Pública**

**Área de Concentração:
Epidemiologia**

**Orientadora: Prof.^a Dr.^a Nilza Nunes
da Silva**

**São Paulo
2009**

Estimação de sub-registro de nascidos vivos pelo método de captura e recaptura

Bianca Schmid

**Tese de Doutorado apresentada ao
Programa de Pós-Graduação em
Saúde Pública para obtenção do
título de Doutora em Saúde Pública**

**Área de Concentração:
Epidemiologia**

**Orientadora: Prof.^a Dr.^a Nilza Nunes
da Silva**

**São Paulo
2009**

É expressamente proibida a comercialização deste documento, tanto na sua forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da tese.

Aos meus avós José Maria e Diva Lanes Maria, que finalmente poderão me chamar de “Dotôra” sem que eu tenha que corrigi-los (“Ainda não, Vô! Ainda não, Vó!”). Agora sim, Vô Zé! Agora sim, Vó Diva!

AGRADECIMENTOS

Ao DATASUS/Ministério da Saúde, em especial Dácio de Lyra Rabello Neto, Antony Stevens e Roberto Fernandes, pelo fornecimento dos dados e presteza na elucidação de dúvidas relativas ao material.

A COPIS e COADS/IBGE, especialmente Cláudio Dutra Crespo, Adauto Amadeu Bastos, Solange Pinto Ferreira, Fernanda Guedes, pelo acesso ao banco de dados do Registro Civil e Celso Cardoso Silva Simões, pela indicação de bibliografia deveras valiosa para a realização desta tarefa.

À Secretaria Estadual de Saúde de Sergipe, por prontamente informar os municípios que compõem as Regiões de Saúde desta Unidade da Federação.

À Professora Nilza Nunes da Silva, pelo incentivo à busca de novas fronteiras e aos meus colegas de orientação, em ordem alfabética para não demonstrar preferências, Carmen, Neussana, Regina e Silvio, pelas críticas sempre construtivas durante os seminários de grupo.

Ao pesquisador Filipe Monteiro, da Universidade de Évora, pelo esclarecimento das tão freqüentes dúvidas.

Aos meus pais, Sylvio e Branca, pois sem eles eu não estaria aqui!

À Ana Luíza Cordeiro, por acreditar em mim, algumas (muitas) vezes mais do que eu mesma.

Ao Prof. Dr. Cláudio Oliveira Ribeiro, pela revisão da tese.

Ao Bruno Dias, amigo, irmão, confidente, pela paciência com as minhas divagações acadêmico-filosófico-psico-sociais.

Por fim, aos familiares (“Rimão”, Tia Madrinha, primas amadas e todos os outros), amigos e colegas, de longe e de perto, que direta e indiretamente contribuíram para a realização deste trabalho.

Uma longa jornada começa com um único passo.

Pensamento chinês

RESUMO

Introdução - O método de captura-recaptura vem sendo empregado em Epidemiologia desde meados do século XX, e se consolidou a partir dos anos 1990, quando se nota grande número de publicações sobre sua aplicação e desenvolvimento nesta área. O sub-registro de eventos vitais ainda se revela um entrave para o cálculo direto de indicadores como os de fecundidade e mortalidade infantil, forçando seu cálculo indireto através de métodos demográficos, cujos procedimentos não permitem estimação em níveis geográficos menores do que unidade da federação, em períodos intercensitários. **Objetivo** – Estimar o sub-registro de nascidos vivos, aplicando o método de captura-recaptura para populações fechadas. **Métodos** - As bases de dados do Sistema Nacional de Informação sobre Nascidos Vivos (SINASC) e do Registro Civil do IBGE, nos segundo e terceiro trimestres de 2006 do estado de Sergipe, foram pareadas por relacionamento determinístico a partir do número da Declaração de Nascido Vivo. As desagregações geográficas adotadas foram as de microrregião e regional de saúde de residência da mãe. Os modelos de Huggins para populações fechadas foram aplicados para estimar as probabilidades de captura em cada uma das bases e o total de nascidos vivos ocorrido no período, dentro de cada desagregação geográfica. O aplicativo utilizado para as estimativas foi o *Software MARK*[®]. **Resultados** – A aplicação do método de captura e recaptura para estimar sub-registro de nascidos vivos é factível, inclusive para desagregações geográficas menores do que unidade da federação. O relacionamento determinístico foi prejudicado em quatro microrregiões e em uma regional de saúde, devido à falta de preenchimento do número da Declaração de Nascido Vivo na base do IBGE. O aplicativo *MARK*[®] apresenta interface amigável, o que facilitou a construção e seleção dos modelos estatísticos, permitindo identificar que a idade da mãe afeta a probabilidade de captura pelo Registro Civil, característica de heterogeneidade na população de nascidos vivos. **Conclusões** – O relacionamento determinístico destas duas bases de dados oficiais viabiliza ações localizadas, porque acaba por identificar onde e quantas vidas mantêm-se no anonimato jurídico, devido ao sub-registro e registro tardio. O método de captura e recaptura mostrou-se uma alternativa acessível e barata para a estimação de sub-registro de nascidos vivos.

Descritores: Captura-recaptura; Sub-registro; Estatísticas de saúde;

ABSTRACT

Introduction – Capture-recapture method has been used in Epidemiology since the middle of twentieth century and from the 1990's on, huge number of publications about applying and developing the method on this area have been noted. Underreporting of vital events is still a drawback for the direct calculation of health indicators like fertility and infant mortality, forcing indirect calculation with demographic methods, wich procedures do not allow estimation in geographic levels smaller than States, in intercensuses periods. **Objective** – To estimate underreporting of live births, using the capture-recapture method for closed populations. **Methods** – The National Live Birth Information System (SINASC) and the Civil Registry databases, in the second and third trimesters of 2006 of the State of Sergipe, Brazil, were linked through deterministic linkage using the Live Birth Declaration number as the connection key. Mother's micro-regions and health regions of residence were adopted as geographical subdivisions. Huggins closed capture models were used to estimate the capture probabilities for each database as well as the derived estimation of the total of live births in the selected time interval, at each of geographical subdivisions. MARK[®] Software was used to get all estimations. **Results** – The capture-recapture method is feasible for estimation of live birth underreporting, even for geographical subdivisions smaller than States. Deterministic linkage was damaged due to lack of filling of the Live Declaration number in some of the Civil Registry database's registries, mainly at four mother residence's micro-regions and at one health region. Program MARK[®] shows friendly interface, which favoured models building and selection, and allowed to point out mother age's influence in Civil Registry database capture probability, characterisc heterogeneity of live born population. **Conclusions** – Deterministic linkage of these two official databases makes possible local actions, because it allows do identify where and how many lives are neglected legal registration's duty, due to underreporting and to late reporting. Capture-recapture method appears as an accessible and cheap alternative for live birth underreporting estimation.

Descriptors: Capture-recapture; Underregistration; Health Statistics

ÍNDICE

1. INTRODUÇÃO.....	23
1.1. <i>APLICAÇÕES DO MÉTODO DE CAPTURA E RECAPTURA.....</i>	<i>25</i>
1.1.1. SUB-REGISTRO DE EVENTOS VITAIS E COBERTURA DE CENSOS DEMOGRÁFICOS.....	26
1.1.2. PREVALÊNCIA DE DOENÇAS	29
1.2. <i>FLUXO E CAPTAÇÃO DE NASCIDOS VIVOS NO BRASIL</i>	<i>38</i>
2. JUSTIFICATIVA	42
3. OBJETIVOS	44
3.1. <i>OBJETIVO GERAL</i>	<i>44</i>
3.2. <i>OBJETIVOS ESPECÍFICOS.....</i>	<i>44</i>
4. MATERIAL E MÉTODOS.....	45
4.1. <i>BASES METODOLÓGICAS E O PROGRAMA MARK[®]</i>	<i>45</i>
4.1.1. FUNDAMENTOS.....	45
4.1.2. O PROGRAMA MARK [®]	54
4.1.3. MODELOS PARA POPULAÇÕES FECHADAS NO PROGRAMA MARK [®]	57
4.1.4. MODELOS LINEARES	63
4.1.5. MODELOS COM VEROSSIMILHANÇA CONDICIONADA (HUGGINS)	68
4.1.6. SELEÇÃO DOS MODELOS E PONDERAÇÃO DOS PARÂMETROS.....	73
4.1.7. CRÍTICAS AO MÉTODO	80
4.2. <i>PROCEDIMENTOS METODOLÓGICOS</i>	<i>82</i>
4.2.1. LIMPEZA E PADRONIZAÇÃO DOS BANCOS.....	82
4.2.2. PAREAMENTO MENOS RESTRITO: NÚMERO DA DN	90
4.2.3. PAREAMENTO MAIS RESTRITO: DN + CRITÉRIOS	96
4.2.4. RELACIONAMENTO PROBABILÍSTICO.....	97
4.2.5. DADOS EMPÍRICOS DA UF SELECIONADA.....	99
4.2.6. DESAGREGAÇÕES GEOGRÁFICAS	116
4.2.7. BANCOS UTILIZADOS PARA A ESTIMAÇÃO.....	130
4.2.8. OS MODELOS DE ESTIMAÇÃO	133
4.2.9. ESTIMAÇÃO DO SUB-REGISTRO CIVIL DE NASCIDOS VIVOS E DA COBERTURA DO SINASC.....	158
5. RESULTADOS	160

5.1. PAREAMENTO MENOS RESTRITO.....	161
5.1.5. MICRORREGIÕES	163
5.1.6. REGIONAIS DE SAÚDE	172
5.2. PAREAMENTO MAIS RESTRITO	178
5.2.5. MICRORREGIÕES	181
5.2.6. REGIONAIS DE SAÚDE	189
6. DISCUSSÃO.....	194
7. CONSIDERAÇÕES FINAIS.....	219
8. REFERÊNCIAS	222
ANEXO I	236
ANEXO II	237
ANEXO III	238
ANEXO IV	239
ANEXO V	240
ANEXO VI	241
ANEXO VII	248
ANEXO VIII	252
ANEXO IX	259

LISTA DE TABELAS

TABELA	PÁGINA
Tabela 1 – Desenho, tipo de estudo, objetivo principal, local, período de realização e autores de alguns estudos epidemiológicos que aplicaram a metodologia de captura e recaptura, segundo a área de investigação.	30
Tabela 2 – Representação dos indivíduos presentes ou ausentes em duas amostras.	47
Tabela 3 – Número de registros de nascimentos vivos ocorridos em 2006 capturados pelas bases SINASC e Registro Civil (RC), segundo a unidade da federação de residência da mãe.	85
Tabela 4 – Total de registros de nascidos vivos ocorridos em 2006 e registros com Número da Declaração de Nascido Vivo (DN) em branco no Registro Civil. Brasil, Grandes Regiões e Unidades da Federação de residência da mãe. 2006	88
Tabela 5 – Número e proporção (%) de registros com variáveis ignoradas nos bancos Registro Civil e SINASC. Sergipe, 2006.	89
Tabela 6 – Pares de registros do SINASC e Registro Civil, relacionados pelo número da DN, classificados quanto à consistência do preenchimento das variáveis. Sergipe, segundo e terceiro trimestres de 2006	94
Tabela 7 – Número de nascidos vivos capturados pelo SINASC e Registro Civil (RC), no segundo e terceiro trimestres de 2006 em Sergipe, segundo o critério de pareamento adotado.	96
Tabela 8 – Estimativas da Fecundidade total, Mortalidade infantil e Mortalidade geral padronizada. Brasil, Região Nordeste e Sergipe, 1997 a 2005.	103
Tabela 9 – Total de registros de nascidos vivos da base do Registro Civil e registros com informações ignoradas nas variáveis número da DN, idade da mãe, sexo e local de nascimento, segundo a microrregião de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006	121

TABELA	PÁGINA
Tabela 10 – Total de registros de nascidos vivos da base do SINASC e registros com informações ignoradas nas variáveis tipo de gestação e local de nascimento, segundo a microrregião de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006	123
Tabela 11 – Número de pares relacionados pelos pareamentos menos e mais restritos e diferença percentual entre os pares encontrados, segundo a microrregião de residência da mãe constante na base do Registro Civil, Sergipe, segundo e terceiro trimestres de 2006	124
Tabela 12 – Total de registros de nascidos vivos da base do Registro Civil e registros com informações ignoradas nas variáveis número da DN, idade da mãe, sexo e local de nascimento, segundo a região de saúde de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006	127
Tabela 13 – Total de registros de nascidos vivos da base do SINASC e registros com informações ignoradas nas variáveis tipo de gestação e local de nascimento, segundo a região de saúde de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006	129
Tabela 14 – Número de pares relacionados pelos pareamentos menos e mais restritos, segundo a regional de saúde de residência da mãe constante na base do Registro Civil, Sergipe, segundo e terceiro trimestres de 2006	130
Tabela 15 – Número final de nascidos vivos capturados pelo SINASC e Registro Civil (RC), no segundo e terceiro trimestres de 2006 em Sergipe, segundo o critério de pareamento adotado.	131
Tabela 16 – Total de nascidos vivos capturados pelo SINASC, proporção de nascidos vivos do sexo feminino, estatísticas dos testes de diferença de proporções, média de idade das mães e estatísticas dos testes de diferença de médias, segundo a microrregião de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.	136
Tabela 17 – Total de nascidos vivos capturados pelo Registro Civil, proporção de nascidos vivos do sexo feminino, estatísticas dos testes de diferença de proporções, média de idade das mães e estatísticas dos testes de diferença de médias, segundo a microrregião de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.	137

TABELA	PÁGINA
Tabela 18 – Total de nascidos vivos capturados pelo SINASC, proporção de nascidos vivos do sexo feminino, estatísticas dos testes de diferença de proporções, média de idade das mães e estatísticas dos testes de diferença de médias, segundo a regional de saúde de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.	138
Tabela 19 – Total de nascidos vivos capturados pelo Registro Civil, proporção de nascidos vivos do sexo feminino, estatísticas dos testes de diferença de proporções, média de idade das mães e estatísticas dos testes de diferença de médias, segundo a regional de saúde de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.	139
Tabela 20 – Estimativa dos critérios de informação de Akaike (AIC), das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (ΔAIC), pesos normalizados de AIC, verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, segundo o pareamento menos restrito.	161
Tabela 21 – Modelos de Huggins para populações fechadas, com as respectivas estimativas de peso ponderado (w_i), probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) e estimativas ponderadas por w_i . Sergipe, pareamento menos restrito, segundo e terceiro trimestres de 2006.	162
Tabela 22 – Distribuição da captura de nascidos vivos pelo SINASC e Registro Civil no estado de Sergipe, segundo e terceiro trimestres de 2006, e estimativa do total de nascidos vivos, segundo o modelo $\{p(.)c(idmae)\}$ e o pareamento menos restrito.	162
Tabela 23 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (ΔAIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento menos restrito, segundo a microrregião de residência da mãe.	165

TABELA	PÁGINA
<p>Tabela 24 – Estimativas da probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) do modelo $\{p(g) c(g + idmae + g*idmae)\}$, distribuição da captura pelo SINASC (n_A), RC (n_B), sobreposição das bases (n_{11}) e total de nascidos vivos distintos capturados pelas duas bases ($r = n_A + n_B - n_{11}$). Sergipe, segundo a microrregião de residência da mãe, pareamento menos restrito, segundo e terceiro trimestres de 2006.</p>	169
<p>Tabela 25 – Estimativas da cobertura do SINASC (\hat{C}) e do sub-registro (\widehat{SR}) no Registro Civil (RC) de acordo com o modelo $\{p(g) c(g + idmae + g*idmae)\}$ e pareamento menos restrito. Sergipe, segundo a microrregião de residência da mãe, segundo e terceiro trimestres de 2006.</p>	170
<p>Tabela 26 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (ΔAIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento menos restrito, segundo a região de saúde de residência da mãe.</p>	174
<p>Tabela 27 – Estimativas da probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) do modelo $\{p(g + idmae + g*idmae) c(g + idmae + g*idmae)\}$, distribuição da captura pelo SINASC (n_A), RC (n_B), sobreposição das bases (n_{11}) e total de nascidos vivos distintos capturados pelas duas bases ($r = n_A + n_B - n_{11}$). Sergipe, pareamento menos restrito segundo a regional de saúde de residência da mãe, segundo e terceiro trimestres de 2006.</p>	176
<p>Tabela 28 – Estimativas da cobertura do SINASC (\hat{C}) e do sub-registro (\widehat{SR}) no Registro Civil (RC) de acordo com o modelo $\{p(g + idmae + g*idmae) c(g + idmae + g*idmae)\}$ e pareamento menos restrito. Sergipe, segundo a regional de saúde de residência da mãe, segundo e terceiro trimestres de 2006.</p>	177

TABELA	PÁGINA
Tabela 29 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (ΔAIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, segundo o pareamento mais restrito.	178
Tabela 30 – Modelos de Huggins para populações fechadas, com as respectivas estimativas de peso ponderado (w_i), probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) e estimativas ponderadas por w_i . Sergipe, pareamento mais restrito, segundo e terceiro trimestres de 2006.	179
Tabela 31 – Distribuição da captura de nascidos vivos pelo SINASC e Registro Civil no estado de Sergipe, segundo e terceiro trimestres de 2006, e estimativa do total de nascidos vivos, segundo o modelo $\{p(.) c(idmae)\}$ e o pareamento menos restrito.	180
Tabela 32 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (ΔAIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento mais restrito, segundo a microrregião de residência da mãe.	183
Tabela 33 – Estimativas da probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) do modelo $\{p(g) c(g + idmae + g*idmae)\}$, distribuição da captura pelo SINASC (n_A), RC (n_B), sobreposição das bases (n_{11}) e total de nascidos vivos distintos capturados pelas duas bases ($r = n_A + n_B - n_{11}$). Sergipe, pareamento mais restrito segundo a microrregião de residência da mãe, segundo e terceiro trimestres de 2006.	185
Tabela 34 – Estimativas da cobertura do SINASC (\hat{C}) e do sub-registro ($\hat{S}\hat{R}$) no Registro Civil (RC) de acordo com o modelo $\{p(g) c(g + idmae + g*idmae)\}$ e pareamento mais restrito. Sergipe, segundo a microrregião de residência da mãe, segundo e terceiro trimestres de 2006.	186

TABELA	PÁGINA
<p>Tabela 35 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (ΔAIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento mais restrito, segundo a regional de saúde de residência da mãe.</p>	190
<p>Tabela 36 – Estimativas da probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) do modelo $\{p(g) c(g + idmae + g*idmae)\}$, distribuição da captura pelo SINASC (n_A), RC (n_B), sobreposição das bases (n_{11}) e total de nascidos vivos distintos capturados pelas duas bases ($r = n_A + n_B - n_{11}$). Sergipe, pareamento mais restrito, segundo a regional de saúde de residência da mãe, segundo e terceiro trimestres de 2006.</p>	192
<p>Tabela 37 – Estimativas da cobertura do SINASC (\hat{C}) e do sub-registro (\hat{SR}) no Registro Civil (RC) de acordo com o modelo $\{p(g) c(g + idmae + g*idmae)\}$ e pareamento mais restrito. Sergipe, segundo a regional de saúde de residência da mãe, segundo e terceiro trimestres de 2006.</p>	193
<p>Tabela 38 – Número de nascidos vivos e número de registros civis constantes no Banco de 2006 do Registro Civil, segundo o trimestre de nascimento e trimestre de registro. Sergipe, 2006</p>	201

LISTA DE QUADROS

QUADRO	PÁGINA
Quadro 1 – Critérios para classificação dos pares reais, pares muito prováveis e pares prováveis entre as bases SINASC e Registro Civil. Os casos que não se enquadram nas situações abaixo foram classificados como pouco prováveis	95
Quadro 2 – Modelos de Huggins para populações fechadas e notação adotada	142
Quadro 3 – Matriz de Desenho do modelo linear completo [$p(g + id_{mae} + g \cdot id_{mae})$ $c(g + id_{mae} + g \cdot id_{mae})$] para as probabilidades de ser capturado pelo SINASC (p) e pelo Registro Civil (c), segundo a microrregião de residência da mãe	148
Quadro 4 – Matriz de Desenho do modelo linear completo [$p(g + id_{mae} + g \cdot id_{mae})$ $c(g + id_{mae} + g \cdot id_{mae})$] para as probabilidades de ser capturado pelo SINASC (p) e pelo Registro Civil (c), segundo a regional de saúde de residência da mãe	157

LISTA DE FIGURAS

FIGURA	PÁGINA
Figura 1 – Percentual de sub-registro de nascimentos vivos pelo método de Brass. Brasil e Unidades da Federação, 2006.	40
Figura 2 – Diagrama de Venn, representativo da história de captura de indivíduos em duas amostras	47
Figura 3 – Valores assumidos pela verossimilhança segundo a variação da probabilidade, em experimento de lançamento de moeda por 20 vezes, com resultado de 5 caras.	49
Figura 4 – Tela inicial do Programa MARK [®]	55
Figura 5 – Tela de importação dos dados e especificação do tipo de dados a serem analisados, do Programa MARK [®]	56
Figura 6 – Tela para escolha do tipo de modelo para populações fechadas, no programa MARK [®]	57
Figura 7 – Esquematização dos modelos e seus parâmetros para populações fechadas no Programa MARK [®]	59
Figura 8 – Gráficos dos Índices dos Parâmetros dos Modelos (PIM) ajustados no MARK [®] , para um estudo em populações fechadas, estimados a partir do modelo de Huggins, com duas amostras.	61
Figura 9 – Janela de Resultados dos modelos ajustados no Programa MARK [®]	74
Figura 10 – Variação do ajuste do modelo e da precisão das estimativas dos parâmetros, de acordo com o número de parâmetros incluídos no modelo	75
Figura 11 - Diferença percentual dos nascimentos vivos ocorridos em 2006 entre as bases SINASC e Registro Civil, segundo a unidade da federação de residência da mãe	86
Figura 12 – Localização geográfica, divisão política e classes de população dos municípios do Estado de Sergipe, 2007	101

FIGURA	PÁGINA
Figura 13 – Pirâmide populacional do estado de Sergipe, 2007	102
Figura 14 – Estimativas da mortalidade infantil. Brasil, Região Nordeste e Estado de Sergipe, 1997 a 2005.	104
Figura 15 - Mortalidade Proporcional por Causas Mal Definidas. Brasil, Região Nordeste e Sergipe, 1990 a 2005.	105
Figura 16 – Distribuição do número de municípios, segundo as classes de grau de urbanização. Sergipe, 2007	106
Figura 17 – Distribuição percentual do nível de escolaridade da população residente, em anos completos de estudo. Brasil, Sergipe, total e segundo o sexo, 2000.	108
Figura 18 – Percentual de mulheres de 10 a 49 anos sem instrução, segundo o município de residência. Sergipe, 2000.	110
Figura 19 - Percentual de sub-registro de Nascidos Vivos estimado pelo método de Brass, segundo a microrregião de residência da mãe. Sergipe, 1994, 2000 e 2005.	115
Figura 20 – Divisão territorial das microrregiões do Estado de Sergipe.	117
Figura 21 – Divisão territorial das Regiões de Saúde do Estado de Sergipe.	118
Figura 22 – Divisão territorial das Microrregiões e Regiões de Saúde do Estado de Sergipe.	119
Figura 23 – Percentual de registros com número da DN ignorado na Base do registro Civil, segundo a microrregião de residência .da mãe, Sergipe, segundo e terceiro trimestres de 2006.	122
Figura 24 – Percentual de registros com número da DN ignorado na Base do registro Civil, e diferença percentual do número de pares obtidos a partir de diferentes métodos de relacionamento entre as bases SINASC e Registro Civil, segundo a microrregião de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.	126

FIGURA	PÁGINA
Figura 25 – Percentual de registros com número da DN ignorado na Base do registro Civil, segundo a regional de saúde de residência .da mãe, Sergipe, segundo e terceiro trimestres de 2006	128
Figura 26 – Índices dos parâmetros do modelo completo $[p(g + idmae + g*idmae) c(g+ idmae + g*idmae)]$, para o agrupamento pelas microrregiões de residência da mãe.	144
Figura 27 – Índices dos parâmetros do modelo $[p(g) c(idmae)]$, para o agrupamento pelas microrregiões de residência da mãe.	147
Figura 28 – Índices dos parâmetros do modelo $[p(.) c(g + idmae)]$, para o agrupamento pelas microrregiões de residência da mãe.	151
Figura 29 – Índices dos parâmetros do modelo $[p = c(g + idmae)]$, para o agrupamento pelas microrregiões de residência da mãe.	153
Figura 30 – Índices dos parâmetros do modelo $[p(g + idmae + g*idmae) c(g+ idmae + g*idmae)]$, para o agrupamento pelas Regionais de Saúde de residência da mãe	155
Figura 31 – Estimativas do percentual de sub-registro civil de nascidos vivos pelos métodos de Brass, para o ano de 2005 e pela captura e recaptura segundo o pareamento menos restrito, para o segundo e terceiro trimestres de 2006. Sergipe e microrregião de residência da mãe	171
Figura 32 – Estimativas do percentual de sub-registro civil de nascidos vivos pelos métodos de Brass, para o ano de 2005 e pela captura e recaptura segundo o pareamento mais restrito, para o segundo e terceiro trimestres de 2006. Sergipe e microrregião de residência da mãe	188
Figura 33 – Estimativas pontuais do sub-registro civil, com o pareamento menos restrito, nas microrregiões de residência da mãe, segundo o grau de urbanização das microrregiões. Sergipe, segundo e terceiro trimestres de 2006.	210
Figura 34 – Estimativas pontuais da cobertura do SINASC, com o pareamento menos restrito, nas microrregiões de residência da mãe, segundo o grau de urbanização das microrregiões. Sergipe, segundo e terceiro trimestres de 2006.	211

FIGURA	PÁGINA
Figura 35 – Estimativas pontuais de sub-registro civil, com o pareamento menos restrito, nas microrregiões de residência da mãe, segundo o percentual de mulheres entre 10 e 49 anos sem instrução, residentes nas microrregiões. Sergipe, segundo e terceiro trimestres de 2006.	212
Figura 36 – Estimativas pontuais da cobertura do SINASC, com o pareamento menos restrito, nas microrregiões de residência da mãe, segundo o percentual de mulheres entre 10 e 49 anos sem instrução, residentes nas microrregiões. Sergipe, segundo e terceiro trimestres de 2006.	213
Figura 37 – Estimativas pontuais do sub-registro civil, com o pareamento menos restrito, segundo o grau de urbanização das regionais de saúde de residência da mãe. Sergipe, segundo e terceiro trimestres de 2006.	215
Figura 38 – Estimativas pontuais da cobertura do SINASC, com o pareamento menos restrito, segundo o grau de urbanização das regionais de saúde de residência da mãe. Sergipe, segundo e terceiro trimestres de 2006.	216
Figura 39 – Estimativas pontuais de sub-registro civil, com o pareamento menos restrito, segundo o percentual de mulheres entre 10 e 49 anos sem instrução, residentes regionais de saúde. Sergipe, segundo e terceiro trimestres de 2006.	217
Figura 40 – Estimativas pontuais da cobertura do SINASC, com o pareamento menos restrito, segundo o percentual de mulheres entre 10 e 49 anos sem instrução, residentes nas regionais de saúde. Sergipe, segundo e terceiro trimestres de 2006.	218

SIGLAS UTILIZADAS

AIC	– <i>Akaike Information Criteria</i>
AIDS	– Síndrome da Imunodeficiência Adquirida
CPF	– Cadastro de Pessoa Física
CPS	– Current Populatoin Survey
DN	– Declaração de Nascido Vivo
DNA	– Ácido Desoxirribonucleico
EUA	– Estados Unidos da América
GOF	– <i>Goodness of Fit</i>
IBGE	– Fundação Instituto Brasileiro de Geografia e Estatística
IC95%	– Estimação por intervalo com 95% de confiança
OMS	– Organização Mundial de Saúde
PEP	– Post-Enumeration Program
PIM	– <i>Parameters Index Model</i> (Índice de Parâmetros do Modelo)
PNAD	– Pesquisa Nacional por Amostra de Domicílios
PSF	– Programa de Saúde da Família
RC	– Registro Civil
SIAB	– Sistema de Informações sobre Atenção Básica
SIH	– Sistema de Informações Hospitalares
SIM	– Sistema de Informações sobre Mortalidade
SINAN	– Sistema de Nacional de Informação de Agravo de Notificação
SINASC	– Sistema de Informação sobre Nascidos Vivos
UF	– Unidade da Federação do Brasil

1. INTRODUÇÃO

Os eventos vitais são definidos como “todos os fatos relacionados com o começo e o fim da vida do indivíduo e as mudanças de estado civil que podem ocorrer durante a sua existência”. Os registros de tais eventos, como nascimentos vivos, nascimentos mortos, óbitos, casamentos, divórcios, adoções, legitimações, reconhecimentos, anulações e separações possuem especial importância jurídica, uma vez que caracterizam o conjunto de qualidades constitutivas que distinguem o indivíduo na sociedade e na família. O Registro de nascimentos e óbitos representa, também, papel fundamental para a construção de políticas públicas, inclusive de saúde, uma vez que tais registros permitem o cálculo de projeções populacionais, o que permite visualizar a evolução demográfica e epidemiológica da população. Com estes mesmos registros constroem-se vários indicadores fundamentais, como por exemplo, o coeficiente geral de mortalidade, os coeficientes específicos de mortalidade e o coeficiente de mortalidade infantil, ferramentas essenciais para conhecimento do estado de saúde de determinada população (Laurenti e colaboradores, 2005).

O sub-registro de eventos vitais é uma realidade em nosso país (Szwarcwald e colaboradores, 2002; Romero, 2002; IBGE, 2005). Segundo Simões (1999), a precariedade da cobertura dos sistemas de estatísticas vitais revela-se como um entrave para o cálculo direto do nível e padrão de fecundidade e mortalidade no Brasil.

De acordo com LaPorte (1994), historicamente, a principal abordagem para avaliação de populações, das diversas disciplinas das ciências, como a epidemiologia, a economia e a ciência política, é identificar os indivíduos de uma comunidade e conhecer seus atributos. Esta abordagem baseia-se na enumeração completa da população e das características que se desejam estudar. Entretanto, a enumeração completa é cara e ineficiente. Lança-se mão de técnicas de amostragens que também apresentam limitações e podem se

mostrar estranhas para aqueles que precisam dos dados para a formulação de políticas públicas, os governantes, por exemplo.

Para a estimação da prevalência de determinada doença, os sistemas de vigilância passiva não raro mostram-se subenumerados e, mesmo assim, estes dados são utilizados para os cálculos. Da mesma forma, o cálculo de fecundidade e mortalidade infantil através de métodos diretos, ou seja, sem correção para o sub-registro de nascimentos e óbitos, respectivamente, pode mascarar a realidade demográfica de uma população (Romero, 2002). Uma das técnicas utilizadas consiste na agregação das diferentes listas obtidas, eliminação dos registros comuns e obtenção de uma única lista. Este método fornece indicadores brutos, não-ajustados, onde verdadeiro número de ocorrências mantém-se desconhecido (McCarty e colaboradores, 1993). Para o cálculo de mortalidade infantil e fecundidade, também são empregadas técnicas indiretas para a estimação, tomando-se como fonte de informação censos demográficos e pesquisas amostrais. Muitas vezes, a quebra dos pressupostos implícitos, exigidos por tais técnicas, causam distorções nas estimativas. Ainda, quando a demanda por estimativas se dá para os níveis de desagregação geográfica menores do que unidade da federação, o problema torna-se mais complexo, dado o pequeno tamanho populacional de boa parte dos municípios brasileiros (Simões, 1999).

A estimação pelo método de captura-recaptura consiste em se aproveitar da superposição dos registros presentes em diversas fontes sabidamente incompletas, mas disponíveis, para formalmente medir o número de ocorrências e, por consequência, a subestimação destas, permitindo, assim, a correção das estatísticas, de maneira a produzir indicadores mais próximos da realidade. Estas fontes (listas) disponíveis podem ser registros de doenças de notificação compulsória, estatísticas hospitalares e de outros serviços de saúde e registros de óbitos, entre outros (LaPorte, 1994; McCarty e colaboradores, 1993; Coeli e colaboradores, 2000).

A partir do cenário exposto, deseja-se aprofundar o conhecimento sobre alternativas para a medida do tamanho das populações de interesse, de maneira a permitir a construção de indicadores válidos para a elaboração de políticas públicas que direcionem de modo mais racional as escassas verbas disponíveis, com o objetivo de melhorar as condições de vida destas populações, e conseqüentemente, da população como um todo.

O presente trabalho buscou aplicar a metodologia de captura e recaptura na estimação das ocorrências de nascidos vivos em Sergipe, nos níveis de microrregião e regionais de saúde de desagregação geográfica. A partir destas estimativas, calculou-se o sub-registro de nascidos vivos, de maneira a fornecer mais uma ferramenta para a construção de indicadores sociais.

1.1. APLICAÇÕES DO MÉTODO DE CAPTURA E RECAPTURA

Resumidamente, o método de captura-recaptura foi empregado na estimação da população da França, em 1793. A partir do século XIX começou a ser utilizado amplamente na estimação do tamanho de populações de animais selvagens (Dunn e Andreoli, 1994; Pradel, 1996), e em várias outras aplicações nas áreas da medicina, demografia e epidemiologia.

Em um experimento com duas amostras numa população de animais, faz-se uma amostra, os indivíduos são marcados e soltos no ambiente, por período suficiente para se misturarem aos demais. Nova amostra é feita e os indivíduos já marcados que foram capturados na segunda amostra são contados.

A idéia é relativamente simples: se, em determinado território a população é pequena, ao se realizar a segunda amostra haverá um grande número de indivíduos que foram marcados na primeira amostra. Por outro lado, se a população é grande, então, na segunda amostra haverá um número pequeno de indivíduos que foram marcados na primeira amostra.

Na epidemiologia, cada lista disponível é considerada uma amostra da

população, e “estar registrado em uma lista” é equivalente a “ser capturado” na amostra.

O International Working Group for Disease Monitoring and Forecasting (1995b) recomenda o uso do método para a avaliação dos sistemas de monitoramento de doenças, especialmente em países em desenvolvimento, pois esta técnica é barata e permite melhorar a informação disponível.

O método de captura e recaptura será detalhado no capítulo Material e Métodos. Nesta seção apresenta-se uma breve revisão bibliográfica da aplicação do método de captura e recaptura em epidemiologia e na demografia, no Brasil e em outros países.

1.1.1. SUB-REGISTRO DE EVENTOS VITAIS E COBERTURA DE CENSOS DEMOGRÁFICOS

O belo e pioneiro trabalho de Sekar e Demming, em 1949, introduziu o método na estimação de sub-registro de nascimentos e óbitos. Dado o reconhecimento de que os sistemas de registros eram ineficientes na Índia, os autores elaboraram um inquérito, em 1947, tendo como período de referência os anos de 1945 e 1946, na área coberta pelo Centro de Saúde de Singur, próximo a Calcutá, Índia. Este centro abrangia 4 subdivisões denominadas *Union Boards* num total de 68 vilas com uma população total em torno de 64.000 habitantes, distribuídos em 12.000 famílias que viviam em cerca de 8.300 domicílios. O inquérito fez o levantamento de todos os moradores desta área, investigando em especial a ocorrência de nascimentos e óbitos nos períodos de referência. Após a comparação com os registros oficiais, observou-se o sub-registro de nascimentos entre 40% e 53% dentro dos *Union Boards* no ano de 1945 e 23% e 39% para o ano de 1946. Já em relação aos óbitos, o sub-registro estimado variou, nos *Union Boards*, entre 23% e 39% em 1945 e entre 51% e 66% em 1946.

O Escritório Nacional de Estatísticas Vitais dos Estados Unidos (*National*

Center for Health Statistics – NCHS) aplicou o método para avaliar o registro de nascimentos naquele país, a partir de comparação dos dados do Censo Demográfico de 1940 com as estatísticas oficiais. A análise foi feita para o país e para os estados. Como resultado, detectou-se que entre 75 e 99% dos nascimentos foram registrados. Para o país como um todo, 92,5% dos nascimentos no ano de 1940 foram registrados (Shapiro, 1949).

No Egito, no período de 1974 a 1975, realizou-se inquérito, com amostra estratificada por área geográfica, com objetivo, dentre outros, de estimar o sub-registro de nascimentos vivos e óbitos infantis. O método empregado foi o de captura-recaptura, comparando os resultados do inquérito com os registros oficiais, e os resultados do estudo foram apresentados para 6 grandes áreas geográficas, onde se observou que entre 0 e 2,6% dos nascimentos e 0 e 5,9% dos óbitos não foram registrados (El-Shalakani, 1985).

Aslan e colaboradores (2003) estimaram os números de nascimentos e mortes ocorridos em uma área rural da Turquia, no período entre maio e outubro de 1999. Como fontes de dados, os pesquisadores utilizaram os registros oficiais e as informações obtidas a partir de um programa de planejamento familiar e saúde reprodutiva existente na área de estudo. Para o relacionamento entre as listas, os autores construíram cinco categorias diferentes de relacionamento, sendo a primeira a mais rígida e as demais progressivamente menos rígidas. Os autores discutem o quanto o método de captura-recaptura é influenciado pelo pareamento no resultado das estimativas. Este assunto é abordado com mais detalhes no capítulo Material e Métodos.

Para avaliar a cobertura do sistema de vigilância demográfica do oeste do Kênia, Eisele e colaboradores (2003) coordenaram um inquérito em 5.000 domicílios, selecionados através de amostragem probabilística, para a estimação do sub-registro de nascimentos e óbitos em menores de cinco anos. Os cálculos foram estratificados por sexo, para nascidos vivos, e sexo e idade em meses, para os óbitos. As estimativas dos totais foram obtidas a partir da

soma dos estratos. Os resultados foram apresentados para duas grandes áreas geográficas, Asembo e Gem. Os resultados encontrados revelaram cobertura total de nascimentos pelo sistema de vigilância demográfica de 77% em Gem, e de 88% em Asembo, para os nascimentos vivos. Nas duas áreas, o registro de meninos foi maior do que o de meninas recém-nascidas. Em relação aos óbitos em menores de cinco anos, a cobertura total foi de 75% em Gem, e de 71% em Asembo. O sub-registro de óbitos em recém-nascidos foi maior do que nas crianças entre 2 e 5 anos, nas duas áreas.

A avaliação de cobertura do Censo dos EUA em 1980 baseou-se em duas abordagens: o *Post-Enumeration Program* (PEP) e a análise demográfica (Fay e colaboradores, 1985; Ericken e Kadane, 1985; Wolter, 1991). No desenho do PEP, três fontes de dados foram utilizadas: dois inquéritos por amostragem de domicílios, o CPS (do inglês, *Current Population Survey*), em abril e agosto de 1980, com 84.000 domicílios cada e a chamada amostra E (*E-sample*), de aproximadamente 110.000 domicílios selecionados diretamente do censo. As CPSs são realizadas mensalmente para estimar o nível de desemprego nos EUA. A partir destas três amostras, obtiveram-se estimadores de captura-recaptura para a população, estratificada por sexo, idade, raça e localização geográfica (estado) e elaboraram-se “grupos” de estimadores, ponderando-se de várias maneiras de acordo com as três pesquisas e foram eleitos 12 destes grupos para comparação com o método de análise demográfica.

O IBGE utiliza a Pesquisa de Avaliação do Censo para a estimação do erro de cobertura do Censo demográfico, desde 1970. Trata-se de uma pesquisa amostral domiciliar que utiliza, também, a técnica de captura-recaptura para a estimação dos indicadores de cobertura de coleta (Oliveira e colaboradores, 2003; Costa, 1990).

1.1.2. PREVALÊNCIA DE DOENÇAS

O International Working Group for Disease Monitoring and Forecasting realizou revisão sobre as pesquisas envolvendo a metodologia de captura-recaptura e identificou sete grandes grupos de investigação, a saber: diabetes, uso de drogas ilícitas, câncer, anomalias congênitas, doenças infecciosas, lesões externas e outras (IWGDMF, 1995b). A Tabela 1 resume alguns trabalhos internacionais publicados que mencionam o método para a estimação das ocorrências dos eventos, de acordo com os grupos de investigação acima citados.

No Brasil, diversos trabalhos de pós-graduação estão catalogados no Banco de Teses da Capes empregando o método para estudar a prevalência de doenças crônicas e infecciosas, e também populações em risco [Coeli, 1998; Gonçalves, 1999; Oliveira, 2000; Côrtes, 2002; Mello, 2003; Brum, 2005; Elkhoury, 2005; Missiagia, 2005; Lanzieri, 2005; Gonçalves VF, 2006; Bezerra, 2007].

Coeli estudou a factibilidade de se implementar um programa de vigilância do *diabetes mellitus* na população idosa usuária do SUS, nas regiões administrativas da Tijuca e Vila Isabel, no município do Rio de Janeiro, RJ [Coeli, 1998].

Gonçalves IF buscou estimar a prevalência de diabetes mellitus do tipo 1 em crianças menores de 15 anos na Ilha do Governador, município do Rio de Janeiro, RJ [Gonçalves IF, 1999].

Oliveira buscou estimar a subnotificação dos casos de AIDS na população adulta de Belo Horizonte, Minas Gerais, no período 1995-1996. Como fonte de dados, utilizou as bases do Sistema de Informação de Agravos de Notificação (SINAN), do Sistema de Informação sobre Mortalidade (SIM), do Sistema de Informações Hospitalares (SIH), dos dados de notificação de tuberculose da Secretaria Municipal de Belo Horizonte e de fichas de notificações de consultórios médicos [Oliveira, 2000].

Tabela 1 – Desenho, tipo de estudo, objetivo principal, local, período de realização e autores de alguns estudos epidemiológicos que aplicaram a metodologia de captura e recaptura, segundo a área de investigação (continua).

DESENHO DO ESTUDO/TIPO DE PUBLICAÇÃO	OBJETIVO PRINCIPAL	LOCAL E PERÍODO DE REALIZAÇÃO DO ESTUDO	AUTORES E PUBLICAÇÃO
DIABETES			
Revisão	Apresentar as técnicas disponíveis para o conhecimento da epidemiologia do diabetes.	-	Ismail e Gill, <i>Baillière's Clin Endocrinol Metab</i> , 1999
Revisão	Apresentar as técnicas disponíveis para o conhecimento da epidemiologia do diabetes.	-	Gill e colaboradores, <i>Q J Med</i> , 2001
Inquérito e comparação com registros administrativos	Avaliar a cobertura do registro de diabetes.	Oguni, Japão, 1991	Sekikawa e colaboradores, <i>Int J Epidemiol</i> , 1999
Comparação entre listas	Verificar a dependência entre as listas.	Verona, Itália, 1986	Verlato e Muggeo, <i>Diabetes Care</i> , 2000
Comparação entre listas	Comparar as estimativas de prevalência de diabetes obtidas a partir do método de captura-recaptura e do método de agregação das listas.	Wales, Reino Unido, 1998	Harvey e colaboradores, <i>J Epidemiol Community Health</i> , 2002

Tabela 1 – Desenho, tipo de estudo, objetivo principal, local, período de realização e autores de alguns estudos epidemiológicos que aplicaram a metodologia de captura e recaptura, segundo a área de investigação (continua).

DESENHO DO ESTUDO/TIPO DE PUBLICAÇÃO	OBJETIVO PRINCIPAL	LOCAL E PERÍODO DE REALIZAÇÃO DO ESTUDO	AUTORES E PUBLICAÇÃO
DIABETES			
Comparação entre listas	Determinar o número ideal de listas a serem comparadas para estimação por captura e recaptura	Liverpool, Inglaterra, 1995-1996.	Ismail e colaboradores, <i>Int J Epidemiol</i> , 2000
Comparação entre listas	Corrigir a estimativa da prevalência de diabetes em crianças.	Madri, Espanha, 1985-1988.	McCarty e colaboradores, <i>Int J Epidemiol</i> , 1993
Comparação entre listas	Aplicar o método para monitorar o diabetes	República Dominicana, 1995-1996	Tull e colaboradores, <i>Rev Panam Salud Publica</i> , 1998
USUÁRIOS DE DROGAS ILÍCITAS/ ABUSO DE DROGAS LÍCITAS			
Revisão	Apresentar os métodos para a estimação do número de usuários de drogas ilícitas	-	Larson e Bammer, <i>Aust N Zealand J Pub Health</i> , 1996
Revisão	Rever os métodos epidemiológicos para pesquisa com usuários de drogas.	-	Dunn e Ferri, <i>Rev Saude Publica</i> , 1999

Tabela 1 – Desenho, tipo de estudo, objetivo principal, local, período de realização e autores de alguns estudos epidemiológicos que aplicaram a metodologia de captura e recaptura, segundo a área de investigação (continua).

DESENHO DO ESTUDO/TIPO DE PUBLICAÇÃO	OBJETIVO PRINCIPAL	LOCAL E PERÍODO DE REALIZAÇÃO DO ESTUDO	AUTORES E PUBLICAÇÃO
USUÁRIOS DE DROGAS ILÍCITAS/ ABUSO DE DROGAS LÍCITAS			
Revisão	Apresentar os métodos para estimar o número de usuários de drogas pesadas.	-	Maxwell, <i>Subst Use Misuse</i> , 2000
Revisão	Rever os métodos e achados empíricos para estimação do número de usuários de heroína nos Países Baixos	-	Korf e colaboradores, <i>Int J Addictions</i> , 1994
Comparação entre listas	Avaliar a utilidade do método para estimar o número de indivíduos com problemas relativos ao abuso de álcool	Voghera, Itália, 1997	Corrao e colaboradores, <i>J Epidemiol Community Health</i> , 2000
Comparação entre listas	Estimar prevalência do uso de ópio.	Barcelona, Espanha, 1993	Domingo-Salvany e colaboradores, <i>Am J Epidemiol</i> , 1998
Comparação entre listas	Estimar a prevalência de abuso de opióides e benzodiazepínicos	Dundee, Escócia, 1990-1994	Hay e McKeganey, <i>J Epidemiol Community Health</i> , 1996

Tabela 1 – Desenho, tipo de estudo, objetivo principal, local, período de realização e autores de alguns estudos epidemiológicos que aplicaram a metodologia de captura e recaptura, segundo a área de investigação (continua).

DESENHO DO ESTUDO/TIPO DE PUBLICAÇÃO	OBJETIVO PRINCIPAL	LOCAL E PERÍODO DE REALIZAÇÃO DO ESTUDO	AUTORES E PUBLICAÇÃO
USUÁRIOS DE DROGAS ILÍCITAS/ ABUSO DE DROGAS LÍCITAS			
Comparação entre listas	Estimar o número de usuários de drogas injetáveis infectados pelo vírus HIV	Bangkok, Tailândia, 1991	Mastro e colaboradores, <i>Am J Pub Health</i> , 1994
CÂNCER			
Comparação entre listas	Determinar o melhor sistema de vigilância de câncer de tireóide	Nouméa, Nova Caledônia (Território Francês no Pacífico), 1985-1992	Ballivet e colaboradores, <i>Eur J Epidemiol</i> , 2000
Comparação entre listas	Avaliar a cobertura dos registros de câncer	Saarland, Alemanha, 1970, 1975; 1980, 1985.	Brenner e colaboradores, <i>J Epidemiol Community Health</i> , 1995
Comparação entre listas	Avaliar a validade do método para o monitoramento da cobertura dos registros de câncer	Países Baixos, 1990	Schouten e colaboradores, <i>In J Epidemiol</i> , 1994

Tabela 1 – Desenho, tipo de estudo, objetivo principal, local, período de realização e autores de alguns estudos epidemiológicos que aplicaram a metodologia de captura e recaptura, segundo a área de investigação (continua).

DESENHO DO ESTUDO/TIPO DE PUBLICAÇÃO	OBJETIVO PRINCIPAL	LOCAL E PERÍODO DE REALIZAÇÃO DO ESTUDO	AUTORES E PUBLICAÇÃO
LESÕES EXTERNAS			
Comparação entre listas	Estimar a incidência de mordidas de cachorro	Pittsburg, EUA, 1993	Chang e colaboradores, <i>Am J Pub Health</i> , 1997
Comparação entre listas	Verificar a validade do método para estimar a incidência de acidentes de trânsito com crianças	Londres, Inglaterra, 1990-1995	Jarvis e colaboradores, <i>Inj Prev</i> , 2000
Comparação entre listas	Estimar a cobertura dos dados oficiais sobre acidentes de trânsito	Glasgow, Escócia, 1995	Morrison e Stone, <i>Inj Prev</i> , 2000
Comparação entre listas	Avaliar a validade do método para estimação de acidentes de trabalho na indústria	Maryland, EUA, 1980-1986	Cormack e colaboradores, <i>Int J Epidemiol</i> , 2000
DOENÇAS INFECTO-CONTAGIOSAS			
Comparação entre estudo prospectivo e registros administrativos	Estimar a transmissão de tuberculose	Madri, Espanha, 1997-1999	Iñigo e colaboradores, <i>Int J Epidemiol</i> , 2003

Tabela 1 – Desenho, tipo de estudo, objetivo principal, local, período de realização e autores de alguns estudos epidemiológicos que aplicaram a metodologia de captura e recaptura, segundo a área de investigação (continua).

DESENHO DO ESTUDO/TIPO DE PUBLICAÇÃO	OBJETIVO PRINCIPAL	LOCAL E PERÍODO DE REALIZAÇÃO DO ESTUDO	AUTORES E PUBLICAÇÃO
OUTROS			
Comparação entre listas	Avaliar a cobertura do registro de catarata congênita	Reino Unido, 1995-1996	Rahi e colaboradores, <i>Invest Ophthalm Vis Science</i> , 1999
Comparação entre inquérito domiciliar e listas	Estimar a prevalência de epilepsia	Benin, 1997	Debrock e colaboradores, <i>Int J Epidemiol</i> , 2000
Comparação entre listas	Estimar a prevalência de desordens mentais nos séculos XVI e XVII	Grã-Bretanha, 1540-1639	Neugebauer, <i>Am J Epidemiol</i> , 1984
Comparação entre listas	Estimar a prevalência de esclerose múltipla	EUA, 1950-1995	Nelson e Anderson, <i>Multiple Sclerosis</i> , 1995
Comparação entre listas	Estimar a prevalência de miastenia grave	Antioquia, Colômbia, 1995-2000	Sánchez e colaboradores, <i>Rev Neurol</i> , 2002

Tabela 1 – Desenho, tipo de estudo, objetivo principal, local, período de realização e autores de alguns estudos epidemiológicos que aplicaram a metodologia de captura e recaptura, segundo a área de investigação (conclusão).

DESENHO DO ESTUDO/TIPO DE PUBLICAÇÃO	OBJETIVO PRINCIPAL	LOCAL E PERÍODO DE REALIZAÇÃO DO ESTUDO	AUTORES E PUBLICAÇÃO
OUTROS			
Comparação entre listas	Estimar a cobertura dos programas de farmácias comunitárias	Países Baixos, 1999	Schrim e colaboradores, <i>Pharmacoepidemiol Drug Safety</i> , 2004
Comparação entre listas	Estimar prevalência de paralisia flácida aguda	Victoria, Austrália, 1998-2000	Whitfield e Kelly, <i>Bull WHO</i> , 2002
Comparação entre listas	Estimar a prevalência de mal de Parkinson	Nebraska, EUA, 2000	Strickland, <i>Mov Disord</i> , 2004
Comparação entre listas	Estimar a prevalência de demência	Carolina do Sul, EUA, 1999	Sanderson e colaboradores, <i>Ann Epidemiol</i> , 2003

Com objetivo de avaliar a vigilância epidemiológica das meningites na região metropolitana de Belo Horizonte, no ano de 1999, Côrtes utilizou três fontes de dados, o SINAN, o SIH e o SIM para estimar a incidência dos casos e a subnotificação, através do método de captura-recaptura [Côrtes, 2002].

Mello teve como objetivo estimar o número de casos de Acidente Vascular Encefálico na população coberta pelo SUS e identificar o perfil de mortalidade dos casos, no município do Rio de Janeiro no ano de 1998. Para isso, conduziu estudo de coorte para analisar a sobrevivência e, em estudo seccional para estimar a prevalência, utilizou a captura-recaptura tendo como fontes de dados as Autorizações de Internações Hospitalares e as declarações de óbito [Mello, 2003].

Brum estimou a leptospirose sintomática e seu subregistro na região central do Rio Grande do Sul, no período entre 2001 e 2002 e utilizou o método de captura-recaptura, a partir de três listas: fonte hospitalar, dados do Laboratório de Diagnóstico de Leptospirose do Departamento de Microbiologia e Parasitologia da Universidade Federal de Santa Maria e a 4.^a Coordenadoria Regional de saúde do Estado do Rio Grande do Sul [Brum, 2005].

Elkhoury (2005) utilizou o método para avaliar os Registros de Morbimortalidade da Leishmaniose Visceral em Sistemas de Informações do SUS, analisando os casos confirmados desta doença constantes no Sistema de Informação de Agravos de Notificação (SINAN), Sistema de Informação sobre Mortalidade (SIM) e Sistema de Informação Hospitalar (SIH).

Missiagia abordou a estimação do tamanho de uma população de diabéticos. Através de listas de pacientes e dados simulados, utilizando métodos de inferência bayesiana, foi feita uma comparação dos desempenhos dos estimadores clássicos e Bayesianos [Missiagia, 2005].

Lanzieri avaliou o Sistema de Vigilância de Rubéola Congênita em Recife, empregando a metodologia para estimar a sensibilidade do sistema, utilizando dados do SINAN e estudos sobre a síndrome da rubéola congênita. [Lanzieri, 2005].

Gonçalves VF utilizou dados secundários do SINAN, Sistema de Controle de Exames Laboratoriais e SIM para estimar a subnotificação de casos de AIDS em Fortaleza, empregando a captura-recaptura [Gonçalves VF, 2006]

Bezerra aplicou o método para estimar o número de crianças e adolescentes em situação de rua em Maceió, utilizando três listas geradas por uma Organização Não Governamental [Bezerra, 2007].

Outros estudos realizados no Brasil foram publicados, como de Argollo e Lessa (1999) que estimou a prevalência de neoplasia cerebral na faixa pediátrica no Estado da Bahia. Spichler e colaboradores (2001) estimaram o coeficiente de amputação de membros inferiores no Rio de Janeiro. Ainda, dentro do Projeto AJUDE Brasil II, o número de usuários de drogas injetáveis foi estimado com o emprego do método de captura e recaptura [Caiaffa WT e colaboradores, 2003; Mingoti e colaboradores, 2006]

1.2. FLUXO E CAPTAÇÃO DE NASCIDOS VIVOS NO BRASIL

O nascimento vivo é, segundo a OMS, "a expulsão ou extração completa do corpo da mãe, independente da duração da gestação, de um produto de concepção que, depois dessa separação, respira ou manifesta qualquer outro sinal de vida, tal como batimento cardíaco, pulsação do cordão umbilical ou contração efetiva de algum músculo de contração voluntária, haja ou não sido cortado o cordão umbilical e esteja ou não desprendida a placenta" (Laurenti e colaboradores, 2005).

No Brasil, todo indivíduo que nasce com vida deve ser inscrito em Cartório, de acordo com a Lei 6.015, de 31/12/73, que regulamenta este ato:

“Art. 50. Todo nascimento que ocorrer no território nacional deverá ser dado a registro no lugar em que tiver ocorrido o parto ou no lugar da residência dos pais, dentro do prazo de quinze dias, podendo ser ampliado até três meses, para os lugares distantes mais de trinta quilômetros da sede do cartório.” (Redação dada pela Lei 9.053, de 1995).

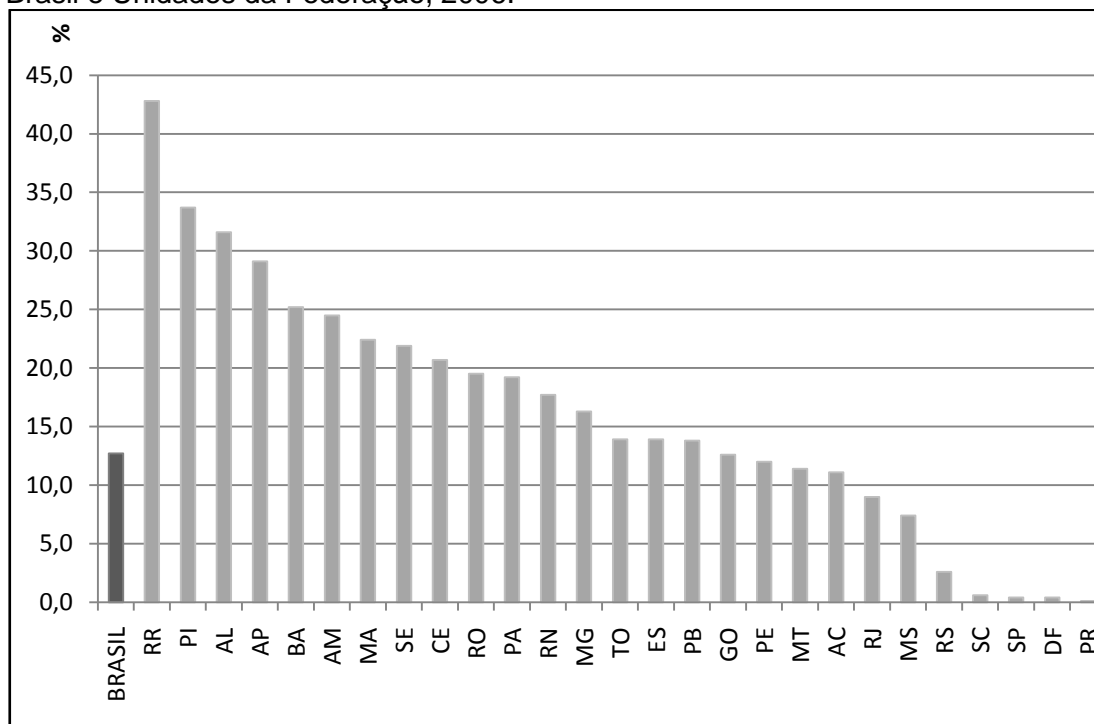
O registro do nascimento vivo no Brasil é gratuito desde 1997, como incentivo ao registro em cartório destes eventos (Brasil, 1997). Outras iniciativas também são tomadas, por parte das autoridades federais, para o estímulo ao registro dos nascimentos dentro do prazo legal, como a Campanha Nacional do Registro Civil, o estabelecimento de gratificações para as unidades de assistência à saúde que incentivam o registro do nascimento antes da alta hospitalar da mãe e da instalação de postos de cartórios de registro civil nas maternidades (IBGE, 2005).

Apesar destes esforços e da gratuidade do registro, no Brasil ainda há o sub-registro dos nascimentos, que se distribui de maneira diferente ao longo das regiões e estados brasileiros (Simões, 1999; IBGE, 2005; Laurenti e colaboradores, 2005). A cobertura dos registros estimada pelo IBGE teve um incremento substancial, passando de 80,6% em 2000 para 88,5% em 2005, no Brasil. Entretanto, no ano de 2006, os estados do Norte e Nordeste ainda apresentam as menores coberturas, chegando a 42,8% o sub-registro em Roraima, como ilustra a Figura 1 (IBGE, 2006). Oliveira e Simões (2005) destacam que a cobertura de nascimentos vivos do Registro Civil aumenta sensivelmente quando se analisam os registros ocorridos até um ano após o nascimento.

Os motivos para o não registro dos nascimentos vivos se revelam em diversas dimensões: fatores sócio-econômicos, culturais, geográficos e político-institucionais. No campo dos fatores sócio-econômicos, os baixos níveis educacionais e de rendimento familiar estão relacionados com o sub-registro. A falta de conhecimento sobre cidadania, da importância do registro civil do nascimento e casos onde a paternidade não é reconhecida constituem a dimensão cultural da baixa cobertura dos registros. Os fatores geográficos que afetam o registro do nascimento são a distância a ser percorrida até o cartório e as dificuldades de transporte, em determinadas áreas do país. Já no campo político-institucional, a falta de fiscalização sobre a aplicabilidade da lei que obriga os registros, a inexistência de rede de proteção à criança e a ausência de Cartórios de Registro Civil em mais de 400 municípios brasileiros contribuem para a baixa cobertura dos registros

de nascimentos nestas áreas (IBGE, 2005).

Figura 1 – Percentual de sub-registro de nascimentos vivos pelo método de Brass. Brasil e Unidades da Federação, 2006.



Fonte: IBGE, Estatísticas do Registro Civil, 2006

Com os dados dos Censos Demográficos e das Pesquisas Nacionais por Amostra de Domicílios (PNADs), emprega-se o método de Brass, que utiliza os quesitos sobre o total de filhos nascidos vivos e nascidos mortos nos últimos 12 meses para o cálculo da fecundidade por idade da mãe, a partir do qual se estima o número total de nascimentos. Uma vez calculada esta estimativa, faz-se a comparação com as informações do Sistema de Informação sobre Nascidos Vivos (SINASC) e com as Estatísticas do Registro Civil, estimando-se, assim, o sub-registro civil e a cobertura do SINASC.

O Sistema de Informação sobre Nascidos Vivos, SINASC, foi implantado pelo Ministério da Saúde em 1990, com objetivo de analisar as estatísticas e fornecer informações que direcionem ações básicas em saúde. Junto com a criação deste sistema, criou-se, também, o documento-padrão Declaração de Nascido Vivo, a DN. A implementação deste sistema ocorreu

de forma lenta e gradual, ao longo da década de 1990 e hoje o SINASC está presente em todas as unidades da federação do Brasil (FUNASA, 2001).

As DN's são impressas pelo Ministério da Saúde, cada declaração em três vias, por intermédio da Secretaria de Vigilância em Saúde (MS/SVS, 2009), e distribuídas gratuitamente às secretarias estaduais de saúde, que as repassam aos estabelecimentos de saúde e cartórios. O fluxo das três vias da DN dependerá do local de ocorrência do parto, se em estabelecimento de saúde ou se no domicílio ou outro local (FUNASA, 2001). O modelo da DN é apresentado no Anexo I.

Quando o nascimento vivo ocorre em um estabelecimento de saúde, este deve fornecer a DN, preenchida nas três vias, sendo que a primeira via, de cor branca, permanece no estabelecimento até que o órgão estadual ou municipal responsável a retire, por busca ativa. A segunda via, de cor amarela, é entregue à família para a efetivação do registro do nascimento em cartório, que reterá esta via para seus procedimentos legais. A terceira via, de cor rosa, fica arquivada no estabelecimento de saúde (FUNASA, 2001). O fluxo da informação, sugerido pelo Ministério da Saúde, está representado no Anexo II.

No caso de partos domiciliares, ou em outro local, a primeira via, de cor branca, deve ser encaminhada à Secretaria de Saúde responsável para o processamento. A segunda via, de cor amarela, fica retida no cartório. Já a terceira via, de cor rosa, deve ser entregue ao responsável do recém-nascido, para que a leve ao estabelecimento de saúde que fará o acompanhamento médico de rotina da criança (FUNASA, 2001). O Anexo III apresenta o fluxo da informação, para o caso de nascimentos vivos fora do estabelecimento de saúde.

As informações do Registro Civil são publicadas desde 1974 pelo IBGE. Os resultados apresentados refletem a totalidade dos assentos de nascidos vivos, casamentos, óbitos e óbitos fetais informados pelos Cartórios de Registro Civil de Pessoas Naturais, e de separações e divórcios declarados pelas Varas de Família, Foros ou Varas Cíveis. O Anexo IV mostra o modelo do questionário de coleta das informações sobre nascidos

vivos nos cartórios, no ano de 2006 (IBGE, 2006).

2. JUSTIFICATIVA

Para estimativa de sub-registro de nascimentos vivos, além do método de Brass, utilizado pelo órgão oficial de estatística do Brasil, IBGE, inquéritos e estudos prospectivos vem sendo realizados em determinadas áreas do país, com o objetivo de quantificar o sub-registro destes eventos vitais. Alguns destes são resumidos a seguir.

Em inquérito domiciliar por amostra na área urbana de Salvador, em 1977, Almeida observou que 31,9% das crianças com menos de um ano de idade não tinham registro de nascimento (Almeida, 1979).

Barros e colaboradores (1985) realizaram estudo de coorte em todas as crianças nascidas em hospitais no município de Pelotas, RS, no ano de 1982. Até o ano de 1984, foi feito acompanhamento nos domicílios destas famílias e, a partir deste estudo, calculou-se o sub-registro de nascimentos vivos (7,8%) e óbitos infantis (24%) na Secretaria de Saúde e Meio-Ambiente do município.

Com o objetivo de estimar o sub-registro de nascimentos vivos no município de Piriá, PI, entre 1.º de julho de 1983 e 30 de junho de 1984, Portela (1989) realizou pareamento entre três fontes de informação: o cartório de Registro Civil, a Fundação Serviço Especial de Saúde Pública e livros de registro de batismo da Igreja Católica e agregou as três diferentes listas, de maneira que observou que cerca de 70% dos nascimentos ocorridos (e registrados em pelo menos uma das fontes) no período não foram registrados no cartório.

Em análise das vias hospitalares e cartoriais das Declarações de Nascidos Vivos em Londrina, PR, entre abril e setembro de 1994, Mello Jorge e colaboradores (1997) observaram sub-registro de nascimentos em torno de 5% e encontraram associação estatística significativa entre o sub-registro e tipo de parto (cesáreo x vaginal/fórceps), escolaridade e renda da mãe.

Através de inquérito domiciliar em Centro Novo do Maranhão, MA,

realizado a partir de amostra por conglomerados, Cardoso e colaboradores (2003) estimaram o sub-registro de nascimentos vivos de 8,6%, no ano de 2002. Os autores atribuem o relativamente baixo sub-registro a três campanhas de registro civil que ocorreram no município, nos anos de 2001 e 2002.

Como se pode perceber, o sub-registro dos eventos vitais no Brasil tem sido preocupação em todas as regiões do país e muito ainda há a ser feito para o seu completo conhecimento e, finalmente, sua eliminação. O método oficial de estimativa do sub-registro de nascimentos vivos tem a grande limitação de não permitir o nível de desagregação microrregiões ou municípios em períodos intercensitários, pois toma como referência a PNAD, o que permite inferência somente no nível de regiões metropolitanas e de unidades da federação. Cálculos de sub-registro para outras desagregações geográficas, utilizando o método de Brass, são aproximações. Os inquéritos e estudos prospectivos realizados permitem o diagnóstico do sub-registro nas áreas estudadas, entretanto, são estudos que demandam recursos humanos e financeiros raramente disponíveis para sua implementação em todos os municípios brasileiros. A técnica de captura e recaptura mostra-se uma alternativa em potencial, barata de ser implementada, permitindo estimativas em níveis de maior desagregação geográfica.

3. OBJETIVOS

3.1. OBJETIVO GERAL

Estimar o sub-registro de nascimentos vivos mediante o uso do método de captura e recaptura e indicar as vantagens e desvantagens dos modelos utilizados.

3.2. OBJETIVOS ESPECIFICOS

- Obter estimativas do sub-registro de nascidos vivos nos níveis geográficos de microrregiões e regionais de saúde.
- Investigar a existência de associação entre fatores sócio-econômicos, geográficos e institucionais e o sub-registro de nascimentos vivos.

4. MATERIAL E MÉTODOS

4.1. BASES METODOLÓGICAS E O PROGRAMA MARK[®]

Esta seção tem como objetivo introduzir o método de captura-recaptura para populações fechadas e apresentar o Programa MARK[®], elaborado inicialmente para aplicar o método em experimentos em ecologia.

4.1.1. FUNDAMENTOS

O método de captura-recaptura, também conhecido como *dual system estimation*, atribuído inicialmente a La Place, em 1783, foi utilizado para estimar a população da França. Desde então, tem sido aplicado amplamente na área da biologia, para estimação de tamanho e de taxa de crescimento de populações de animais selvagens, e em várias outras aplicações nas áreas da medicina, demografia e epidemiologia (Dunn e Andreoli, 1994; Pradel, 1996). Em investigações clínicas, utiliza-se, por exemplo, para estimação do volume e sobrevivência de glóbulos vermelhos, através de marcação por rádio-isótopos; verificação de especificidade em procedimentos de rastreamento médico e, também, em análise de proteínas e DNA. Na área da demografia, o método de captura e recaptura é utilizado para estimação da cobertura dos censos demográficos e estimação de sub-registro de nascimentos vivos e óbitos. Na epidemiologia, usa-se para estimar a prevalência de doenças pouco freqüentes e de populações móveis, como usuários de drogas injetáveis (Dunn e Andreoli, 1994).

O modelo mais simples de estimação pelo método de captura-recaptura, o de duas amostras, foi empregado por Petersen, em 1894 e a partir daí o método passou a ser utilizado na área da ecologia. (IWGDMF, 1995a). Sekar e Demming (1949) aplicaram o método para estimar o sub-registro de nascimentos vivos e mortes em uma comunidade na Índia. Shapiro utilizou a técnica para o cálculo do sub-registro de nascimento vivo nos EUA, a partir de dados censitários (Shapiro, 1949). No final da década de 1960, Wittes e Sidel (1968) apresentaram a generalização do método de

captura-recaptura para as aplicações em epidemiologia, utilizando-se duas ou mais listas. Desde então, o método foi ganhando visibilidade e a partir da década de 1990 houve um crescimento considerável no interesse do tema nas aplicações em pesquisas epidemiológicas (Coeli, 1998).

Em experimentos ecológicos, o procedimento mais simples consiste em realizar uma amostra da população, marcar os indivíduos, esperar período suficiente para que se misturem ao restante da população e, em seguida, realizar nova amostra. Os indivíduos marcados e recapturados são contados, e a partir do número de indivíduos presentes na primeira amostra, na segunda amostra e daqueles presentes nas duas, estima-se o tamanho total da população. Os seguintes pressupostos devem ser assumidos (IWGDMF, 1995a; Hook e Regal, 1995):

1. A população é fechada, ou seja, não há nascimentos, óbitos e migrações no período compreendido entre as amostras;
2. A marcação é unívoca e inequívoca, isto é, cada indivíduo é identificado através da marcação e não há possibilidade de “perda” desta marca;
3. Em cada amostra, cada indivíduo tem a mesma probabilidade de pertencer à amostra (equiprobabilidade);
4. As duas amostras são independentes, ou seja, os eventos “um indivíduo é capturado em uma amostra” e “um indivíduo é capturado em outra amostra” são estocasticamente independentes; e
5. Em cada amostra, qualquer indivíduo é capturado (recapturado) independentemente dos demais.

Para apresentar o método, considere as notações apresentadas a seguir. Sejam duas amostras de indivíduos, A e B. Tem-se o número de indivíduos incluídos na amostra A, os incluídos na amostra B, os que pertencem às duas amostras e aqueles que não pertencem a nenhuma das duas. O diagrama de Venn (Figura 2) e a Tabela 2 apresentam as representações e notações para a situação descrita.

Figura 2 – Diagrama de Venn, representativo da história de captura de indivíduos em duas amostras.

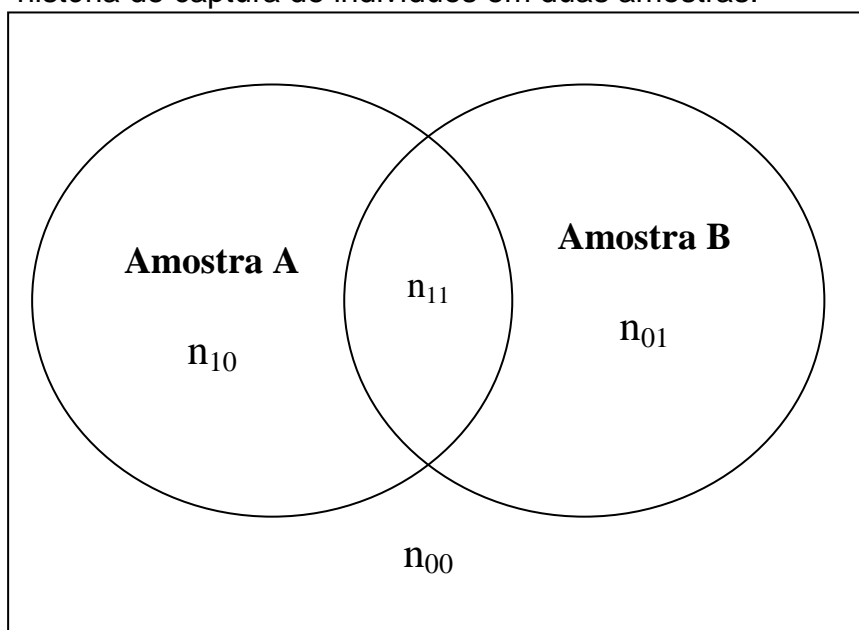


Tabela 2 – Representação dos indivíduos presentes ou ausentes em duas amostras

		Amostra B		
		Presente	Ausente	
Amostra A	Presente	n_{11}	n_{10}	n_A
	Ausente	n_{01}	n_{00}	
		n_B		N

Onde,

$N = n_{11} + n_{10} + n_{01} + n_{00}$ é o tamanho total da população;

$n_A = n_{11} + n_{10}$ é o número de indivíduos presentes na amostra A;

$n_B = n_{11} + n_{01}$ é o número de indivíduos presentes na amostra B;

n_{11} = número de indivíduos presentes nas duas amostras;

n_{10} = número de indivíduos presentes somente na amostra A;

n_{01} = número de indivíduos presentes somente na amostra B;

n_{00} = número de indivíduos ausentes nas duas amostras;

Define-se

p_A = probabilidade de estar presente na amostra A;

p_B = probabilidade de estar presente na amostra B;

Existem diversas técnicas para estimação em Estatística, como o método dos mínimos quadrados e o método dos momentos, por exemplo. Para as estimações em captura e recaptura, utiliza-se a estimação pelo método de máxima verossimilhança. Os estimadores obtidos pela máxima verossimilhança possuem propriedades que os tornam extremamente úteis, a saber: a) são consistentes; b) são assintoticamente não viciados; c) são assintoticamente eficientes, ou seja, a variância é pequena; e d) assintoticamente possuem distribuição normal, o que permite testes estatísticos baseados na distribuição normal (COOCH e WHITE, 2008).

Para rapidamente introduzir a estimação por verossimilhança, tome-se o exemplo de lançamento de moedas. Considere a probabilidade de se obter 5 caras em 20 lançamentos de uma moeda não viciada. Em outras palavras, a probabilidade de se obter $y = 5$ caras dados $N = 20$ lançamentos e probabilidade $p = 0,5$ de a cara da moeda cair voltada para cima, será:

$$f(y | N, p) = \binom{N}{y} p^y (1 - p)^{(N-y)}$$

Onde,

$$\binom{N}{y} = \frac{N!}{y!(N-y)!}$$

Assim,

$$f(5 | 20, 0.5) = \frac{20!}{5!(20-5)!} 0.5^5 (1 - 0.5)^{(20-5)} = 0,0148$$

A probabilidade de obterem 5 caras em 20 lançamentos de uma moeda não viciada é de 1,48%.

Neste exemplo, obtém-se a probabilidade de $y = 5$ caras sabendo-se

qual é a probabilidade do evento cara ($p = 0,5$) acontecer em um lançamento. Quando se deseja estimar a probabilidade de um evento acontecer a partir dos dados conhecidos, encontra-se o valor de p que maximiza a probabilidade (verossimilhança) de se observarem os dados disponíveis. Mantendo o exemplo acima, se não se soubesse que a moeda é honesta, poder-se-ia perguntar qual a probabilidade de ocorrer um evento cara, dado que em 20 lançamentos obtiveram-se 5 caras. Desta maneira, a verossimilhança da probabilidade de se obter cara, dado que em 20 lançamentos a cara da moeda caiu 5 vezes voltada para cima será:

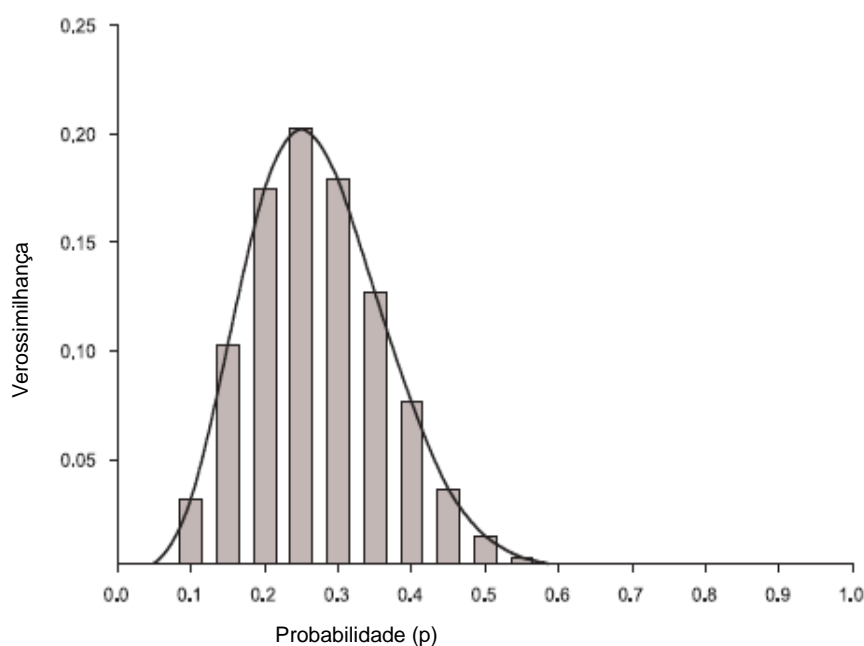
$$f(p | N, y) = \binom{N}{y} p^y (1 - p)^{(N-y)}$$



$$f(p | 20, 0.5) = \frac{20!}{5!(20 - 5)!} p^5 (1 - p)^{(20-5)}$$

Atribuindo vários valores para p na função acima, pode-se obter o gráfico (Figura 3):

Figura 3 – Valores assumidos pela verossimilhança segundo a variação da probabilidade, em experimento de lançamento de moeda por 20 vezes, com resultado de 5 caras.



Percebe-se que a verossimilhança é maximizada quando a probabilidade é igual a 0,25. Em outras palavras, se o p , desconhecido, for 0,25, então corresponderá à máxima probabilidade de se observar 5 caras em 20 lançamentos da moeda. O gráfico mostra que alguns valores para p são relativamente “inverossímeis”, dadas as informações disponíveis. O valor do parâmetro p que maximiza a função da verossimilhança é o valor mais provável de p , dados os resultados dos lançamentos (COOCH e WHITE, 2009).

Retornando à captura e recaptura, a probabilidade p_B de estar presente na segunda amostra segue uma distribuição binomial e a verossimilhança L (do Inglês, *likelihood*) desta probabilidade, dado que n_B indivíduos foram amostrados e n_{11} estavam presentes na primeira amostra será:

$$L(p_B | n_B, n_{11}) = p_B^{n_{11}} (1 - p_B)^{(n_B - n_{11})}$$

Para simplificar os cálculos, aplica-se a transformação logarítmica da verossimilhança da probabilidade p_B :

$$\ln[L(p_B | n_B, n_{11})] = n_{11} \ln p_B + (n_B - n_{11}) \ln(1 - p_B)$$

A verossimilhança tem seu valor maximizado quando se iguala a primeira derivada a zero e resolve-se a equação. Ao se diferenciar para o parâmetro desconhecido, p_B , igualando a zero a equação, encontra-se o estimador de p_B que maximiza a verossimilhança da probabilidade de se observar n_{11} indivíduos dada a amostra n_B :

$$\frac{\partial \ln[L(p_B | n_B, n_{11})]}{\partial p_B} = \frac{n_{11}}{p_B} - \frac{(n_B - n_{11})}{(1 - p_B)} = 0$$

Resolvendo para p_B , tem-se:

$$\hat{p}_B = \left(\frac{n_{11}}{n_B} \right)$$

O estimador para a variância de p_B é obtido a partir do inverso negativo da segunda derivada da verossimilhança log-transformada:

$$\text{var}(\hat{p}_B) = \left[- \frac{\partial^2 \ln(p_B | n_B n_{11})}{\partial^2 p_B} \right]_{p_B = \hat{p}_B}^{-1}$$

Resolvendo a equação, tem-se:

$$\text{var}(\hat{p}_B) = \frac{p_B(1-p_B)}{n_B}$$

Intuitivamente é possível perceber a relação entre os indivíduos presentes na primeira amostra e na segunda. Considerando os pressupostos, pode-se afirmar que a razão entre o número de indivíduos capturados na primeira amostra e o número total de indivíduos na população é igual à razão entre os indivíduos marcados na segunda amostra e o número de indivíduos presentes nesta segunda amostra. Utilizando a notação da Tabela 2, tem-se:

$$\frac{n_A}{N} = \frac{n_{11}}{n_B}$$

Resolvendo para N , tem-se o estimador de Petersen:

$$\hat{N} = \frac{n_A n_B}{n_{11}}$$

A premissa da independência entre as amostras é essencial para que

os estimadores não sejam viciados. Em situações onde há dependência positiva entre as amostras, isto é, a presença do indivíduo em uma aumenta a probabilidade da sua presença na outra, a sobreposição de indivíduos (n_{11}) será grande, e o \hat{N} subestimar N . Nas situações de dependência negativa, ou seja, a presença de um indivíduo em uma amostra diminui sua probabilidade de figurar na outra, então n_{11} será pequeno e \hat{N} superestimar N .

Além da independência entre as amostras, a heterogeneidade dos indivíduos que as compõem também vicia a estimativa da população, pois afeta a equiprobabilidade. Esta heterogeneidade se expressa por características sócio-econômicas, demográficas ou epidemiológicas, tais como sexo, idade, presença de doença, entre outras. Esta seria a origem de uma “aparente dependência”, que é similar ao efeito de uma variável de confusão sobre a associação entre exposição e desfecho, conceito já conhecido em epidemiologia (Coeli, 1998).

Ao comparar as proporções relativas entre indivíduos marcados e não marcados em sucessivas amostras, vários estimadores podem ser derivados. Grande número de modelos e métodos foi desenvolvido, buscando identificar a heterogeneidade entre os indivíduos e quebra das demais premissas do método de captura e recaptura (Lucacks, 2008). Aqui, será dada ênfase nos modelos baseados em verossimilhança disponíveis no programa MARK[®], para populações fechadas, em especial os modelos com a verossimilhança condicionada (Huggins).

A chave para o entendimento dos modelos de populações fechadas consiste em compreender que as histórias de captura são guiadas pela “taxa de visibilidade”, isto é, a probabilidade de um indivíduo ser identificado, dado que ele sobreviveu desde a amostra anterior e está presente na amostra seguinte. De fato, o processo de estimar o tamanho da população, nestes tipos de modelo, é, com efeito, o processo de se estimarem as probabilidades de “detecção”. Isto quer dizer que se deseja estimar a probabilidade de um indivíduo ser amostrado pela primeira vez (se o for) e a

probabilidade de, dado que foi amostrado pela primeira vez, o ser novamente. Os diferentes modelos para populações fechadas se diferenciam tão somente pela maneira como a “taxa de visibilidade” é abordada (Lucacks, 2008).

Considerando a notação “11” para expressar a história de captura do indivíduo presente nas duas amostras, “10” expressa a história do indivíduo presente somente na primeira amostra e assim por diante, é possível modelar as probabilidades de cada história de captura como no quadro a seguir:

História de Captura	Probabilidade
11	pc
10	$p(1 - c)$
01	$(1 - p)p$
00	$(1 - p)(1 - p)$

Onde p é a probabilidade do indivíduo ser amostrado pela primeira vez e c é a probabilidade de ser recapturado.

A função das probabilidades em um estudo com duas amostras é dada por:

$$P(n_A, n_B, n_{11}|N, p, c) = \frac{N!}{n_{11}!(n_A - n_{11})!(n_B - n_{11})!(N - r)!} \\ * \{pc^{n_{11}} * [p(1 - c)]^{n_A - n_{11}} * [(1 - p)p]^{n_B - n_{11}} * [(1 - p)(1 - p)]^{N - r}\}$$

Onde $r = n_A + n_B - n_{11}$, o número de indivíduos distintos observados nas duas amostras.

Cabe ressaltar que esta função, a partir da qual são derivados os estimadores de máxima verossimilhança dos parâmetros desconhecidos, inclui a probabilidade dos indivíduos não observados em nenhuma das duas amostras, representados na equação acima por $(N - r)$. É possível construir a função das probabilidades condicionada somente a partir dos indivíduos

observados no estudo (r). A probabilidade de um indivíduo ser amostrado pelo menos uma vez no estudo de duas amostras será:

$$p' = 1 - (1 - p)(1 - c)$$

Reescrevendo a probabilidade condicional, tem-se:

$$P(\{n_{ij}\} | r, p, c) = \frac{r!}{n_{11}! n_{01}! n_{10}!} * \left(\frac{pc}{p'}\right)^{n_{11}} * \left(\frac{(1-p)p}{p'}\right)^{n_{01}} * \left(\frac{p(1-c)}{p'}\right)^{n_{10}}$$

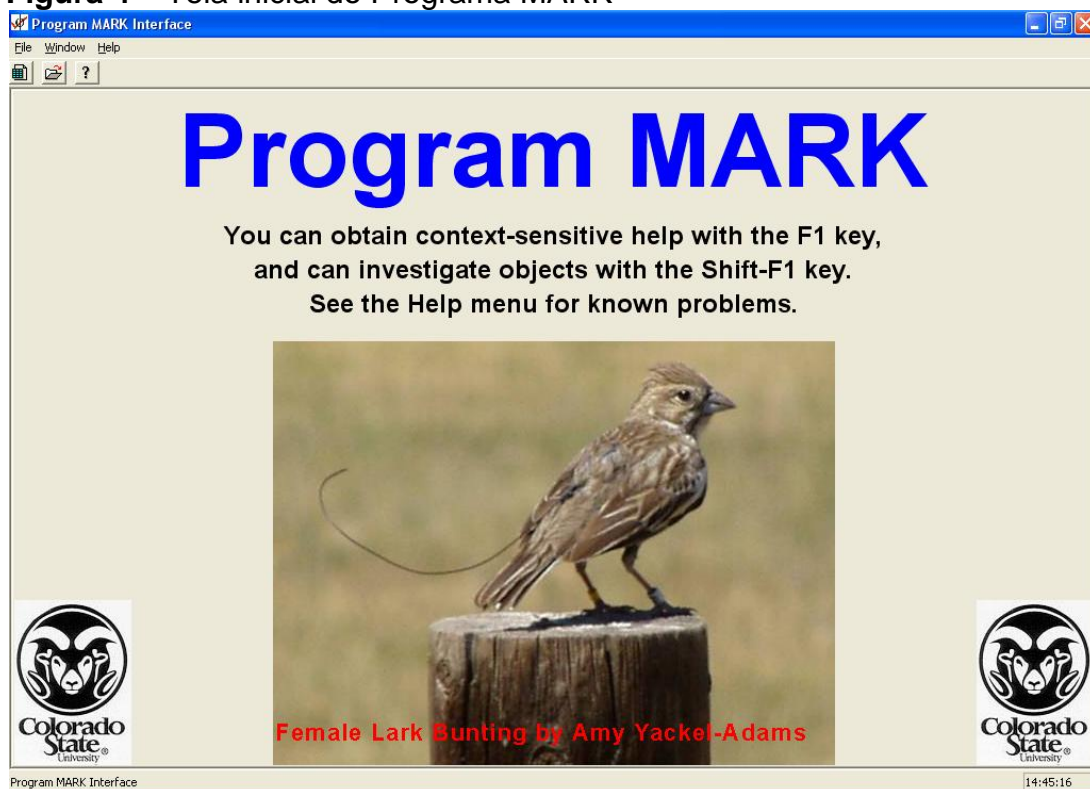
Onde n_{ij} é o indivíduo i capturado na amostra j .

Desta probabilidade são derivados os estimadores de máxima verossimilhança (Lucacks, 2008). A principal vantagem de se condicionar a probabilidade aos indivíduos presentes no estudo (modelos de Huggins) é que se permite ajuste de modelos com co-variáveis, de uma maneira direta.

4.1.2. O PROGRAMA MARK[®]

O programa MARK[®], desenvolvido por Gary C. White, da Universidade do Estado de Colorado, EUA, apresenta uma interface amigável para a estimação dos parâmetros em experimentos com animais identificados e reencontrados em outro momento posterior.

Estes reencontros podem ser recaptura de animais mortos ou vivos, monitoramento por rádio, identificação visual, ou combinação destas formas de reencontro. Os intervalos de tempo entre os reencontros podem ou não ser uniformes. É possível ajustar modelos com mais de um grupo de animais, assim como incluir variáveis individuais. Os parâmetros dos modelos podem ser fixados ou podem variar entre os reencontros e o programa também disponibiliza alguns modelos pré-definidos.

Figura 4 – Tela inicial do Programa MARK[®]

Fonte: MARK – Mark and Recapture Survival Rate Estimation[®], version 5.1 (software na internet). Fort Collins: Department of Fishery and Wildlife, Colorado State University; 2007 (acesso em 02/01/2008). Disponível em: <http://welcome.warnercnr.colostate.edu/~gwhite/mark/mark.htm#Downloading>.

O programa ajusta modelos para populações geográfica e demograficamente abertas e fechadas (WHITE, 2009). O tipo de modelo que se deseja utilizar é selecionado logo na tela de importação dos dados (Figura 5)

Figura 5 – Tela de importação e especificação do tipo de dados a serem analisados, do Programa MARK®

The screenshot shows the 'Enter Specifications for MARK Analysis' dialog box. On the left, under 'Select Data Type', the 'Closed Captures' option is selected. The right side of the dialog contains several input fields and controls:

- Title for this set of data:** A text box containing 'EXEMPLO'.
- Encounter Histories File Name:** A text box containing 'C:\Biká\Doutorado\projeto\EXEMPLO.inp' with a 'Click to Select File' button next to it and a 'View File' button to the right.
- Results File Name:** A text box containing 'C:\Biká\Doutorado\projeto\EXEMPLO.DBF'.
- Encounter occasions:** A spinner box set to '5' with a 'Set Time Intervals' button and the text 'Default Time Intervals Used'.
- Attribute groups:** A spinner box set to '4' with an 'Enter Group Labels' button and the text 'Default Group Labels Used'.
- Individual covariates:** A spinner box set to '2' with an 'Enter Ind. Cov. Names' button and the text 'Default Ind. Cov. Names Used'.
- Strata:** A spinner box set to '2' with an 'Enter Strata Names' button and the text 'Default Strata Names Used'.
- Mixtures:** A spinner box set to '2'.

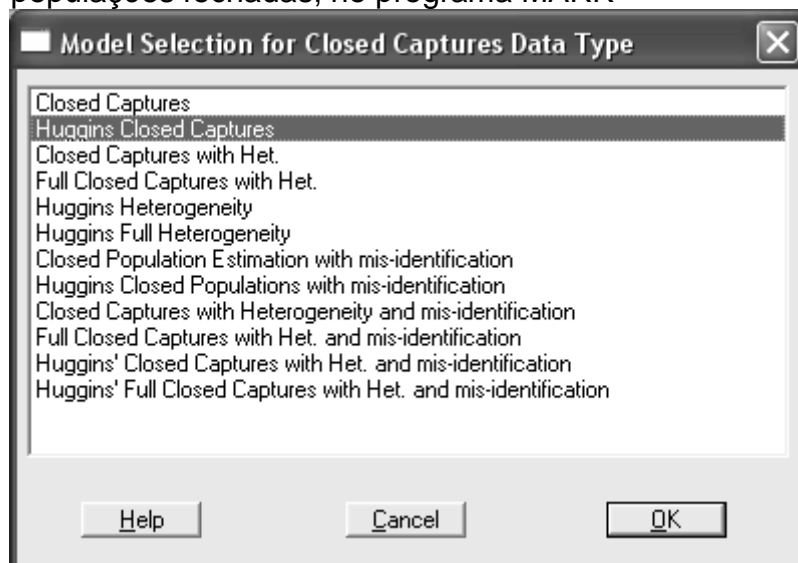
At the bottom right, there are three buttons: 'Help', 'Cancel', and 'OK'.

Ao selecionar a opção populações fechadas (*Closed Captures*), uma nova janela se abre, para se determinar qual tipo de modelos se quer ajustar (Figura 6). Os detalhes sobre os modelos para populações fechadas são apresentados nas seções a seguir.

Inúmeros recursos e funções estão disponíveis neste programa, porém, como já salientado, nesta seção apresentam-se somente aqueles utilizados em estimações para populações fechadas, pela verossimilhança condicionada aos dados observados (modelos de Huggins).

Este programa foi escolhido para a pesquisa devido à sua interatividade e interface amigável. Ainda, os modelos de Huggins podem ser aplicados aos dados disponíveis, a saber as bases de dados sobre nascidos vivos do Registro Civil e do SINASC.

Figura 6 – Tela para escolha do tipo de modelo para populações fechadas, no programa MARK[®]



Tilling e Sterne (1999) empregaram a abordagem da verossimilhança condicionada aos dados de Huggins para ajustarem modelos logísticos e log-lineares em simulações com o objetivo de medir efeitos da dependência atribuída a variáveis individuais, em epidemiologia. Tilling e colaboradores (2001) ajustaram modelos logísticos incluindo variáveis demográficas e de severidade da doença para estimar a probabilidade de estar presente em registros administrativos sobre derrame cerebral no sul de Londres, e por consequência, o sub-registro desta doença na área estudada. Assim, acredita-se que o *software* e as estatísticas por ele calculadas sejam adequados ao objetivo geral da pesquisa, que é estimar o sub-registro de nascidos vivos em Sergipe utilizando o método de captura e recaptura.

4.1.3. MODELOS PARA POPULAÇÕES FECHADAS NO PROGRAMA MARK[®]

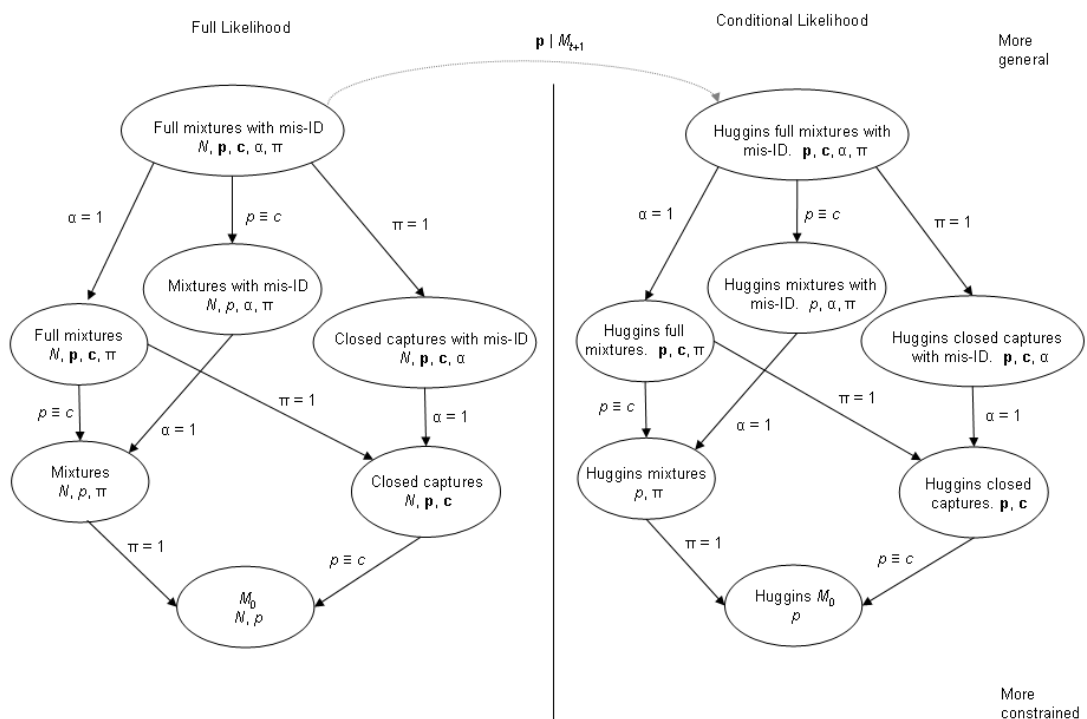
O Programa MARK[®] oferece uma gama de modelos para populações fechadas, 12 no total. A principal diferença entre eles é a forma de se abordar a verossimilhança. Nos modelos considerados “cheios”, o número total de indivíduos na população, N , é incluído na verossimilhança, ao passo que os modelos de Huggins incluem na verossimilhança somente

parâmetros dos indivíduos efetivamente observados nas amostras. De uma maneira geral, os modelos são classificados dentro de uma hierarquia de divisões dicotômicas (Figura 7). Na parte esquerda do diagrama da Figura 7 são apresentados os modelos com os parâmetros incluídos na verossimilhança “cheia”, ao passo que na parte direita encontram-se os modelos com os parâmetros incluídos na verossimilhança condicionada aos indivíduos observados nas amostras (Lucacks, 2008).

À esquerda do diagrama (Figura 7), os modelos *Closed Captures* baseiam-se na parametrização da verossimilhança “cheia”, com três tipos de parâmetros, onde p_i é a probabilidade da primeira captura (o indivíduo é capturado pela primeira vez) na amostra i , c_i é a probabilidade de recaptura na amostra i , dado que o indivíduo já havia sido amostrado pelo menos uma vez anteriormente e N é o número total de indivíduos na população. Nestes modelos, tanto p_i quanto c_i podem variar ao longo do tempo, quando se acredita que o tempo percorrido entre uma amostra e outra pode afetar as probabilidades em questão, ou seja, p_{i-1} e c_{i-1} na amostra $i - 1$ são diferentes dos p_i e c_i , respectivamente, na amostra i .

Como tentativa de se medir o efeito da heterogeneidade na probabilidade de captura, algumas abordagens podem ser feitas. Uma delas é considerar que na população há “composições” ou “grupos” (*mixtures*), onde nestes “grupos” os indivíduos são homogêneos em relação às probabilidades de captura e recaptura. Este recurso é útil quando não se sabe exatamente quantas “composições” existem na população. Considera-se que existe um número A , finito, de grupos e que dentro de cada grupo a , os indivíduos apresentam equi-probabilidade. Não se sabe qual indivíduo está em qual grupo. Assume-se que cada indivíduo é oriundo independentemente do grupo a com probabilidade π_a ($a = 1, \dots, A; \sum \pi_a = 1$). Esta probabilidade é incorporada na probabilidade de captura, como uma aproximação da heterogeneidade individual de captura (Pledger, 2000). Estes modelos são os denominados *Full Mixtures* na Figura 7, que contêm, além dos parâmetros p_i , c_i e N , o parâmetro π .

Figura 7 – Esquematização dos modelos e seus parâmetros para populações fechadas no Programa MARK®



Fonte: LUCKACKS P. Closed Population Capture-Recapture Models. In: COOCH E, WHITE G (Organizadores). **Program MARK: A Gentle Introduction**. Ithaca: Cornell University, 2008 (Acesso em 02/03/2009). Disponível em: <http://www.phidot.org/software/mark/docs/book/>

Já os modelos nomeados *Full Mixtures with Mis-Identification* são aplicáveis a experimentos onde a identificação dos indivíduos é feita através do DNA. Para estes modelos assume-se que: i) o indivíduo capturado mais de uma vez foi corretamente identificado e ii) indivíduos capturados somente uma vez podem ou não ter sido corretamente identificados. Estes modelos incluem na parametrização da verossimilhança, além de p_i , c_i , N e π , o parâmetro α , que é a probabilidade de um genótipo ser identificado corretamente, dado que foi observado pela primeira vez (Lucacks & Burnham, 2005).

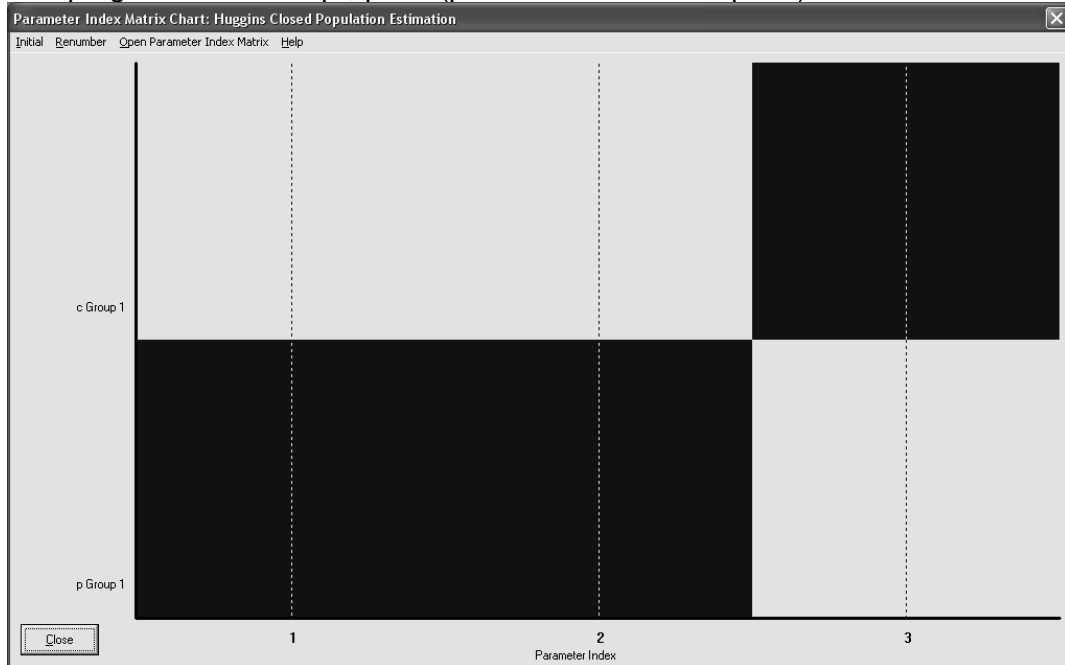
Como se pode perceber na Figura 7, os demais modelos são “casos especiais” dos três principais acima comentados. Os conceitos e parâmetros apresentados também se aplicam aos modelos de Huggins, localizados na

parte direita, que se diferenciam por não conter o parâmetro N na verossimilhança (Lucacks, 2008).

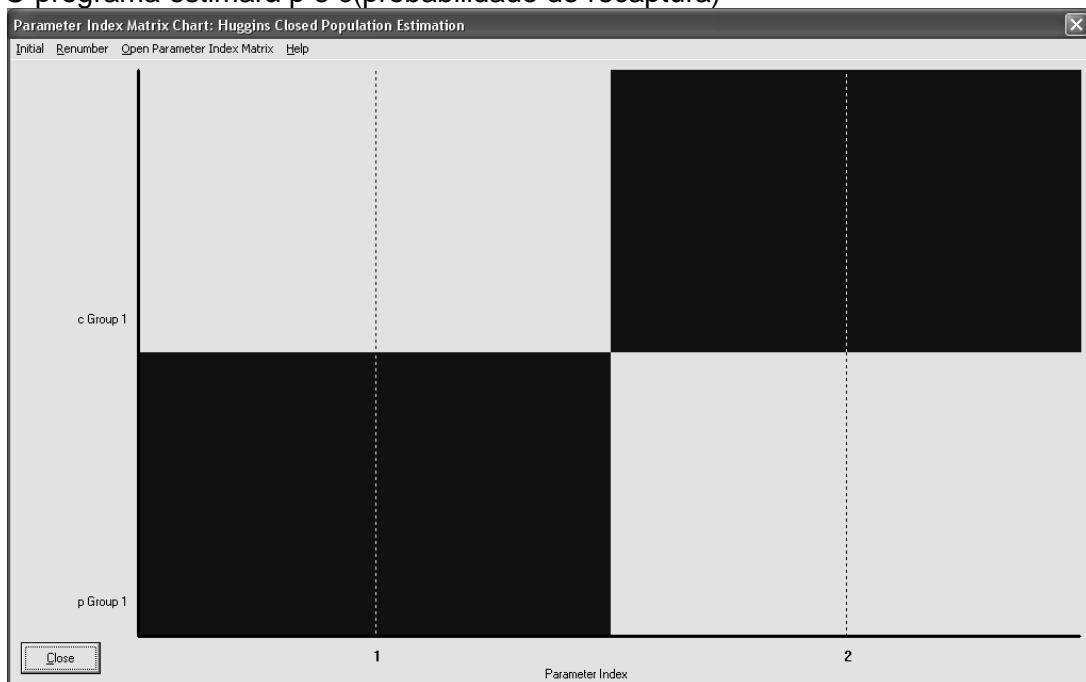
A parametrização do modelo é ajustada através do gráfico de Índices de Parâmetros do Modelo (*PIM chart*). Neste gráfico, é possível manter, por exemplo, a probabilidade de captura constante entre as amostras e entre os grupos, e a probabilidade de recaptura variando entre as amostras, mas constante entre os grupos; ou manter tanto a captura quanto recaptura constantes entre as amostras, mas variando entre os grupos; ou, ainda, as probabilidades são iguais em todos os grupos, mas variam entre as amostras; entre outras tantas possibilidades. Considere um estudo com duas amostras, onde se deseja ajustar dois modelos: um onde a probabilidade da primeira captura varia entre as duas amostras e outro onde a probabilidade da primeira captura é igual nas duas amostras. O gráfico dos parâmetros a ser ajustado para o primeiro modelo é apresentado na Figura 8A, enquanto que o gráfico dos parâmetros para o segundo modelo é mostrado na figura 8B.

Figura 8 – Gráficos dos Índices dos Parâmetros dos Modelos (PIM) ajustados no MARK®, para um estudo em populações fechadas, estimados a partir do modelo de Huggins, com duas amostras:

- A.** A probabilidade da primeira captura (p) é diferente na amostra 1 e na amostra 2.
2. O programa estimará p_1 , p_2 e c (probabilidade de recaptura)



- B.** A probabilidade da primeira captura (p) é igual na amostra 1 e na amostra 2.
0 programa estimará p e c (probabilidade de recaptura)



O ajuste de modelos com variáveis individuais é realizado a partir da Matriz de Desenho (*Design Matrix*). De acordo com a parametrização escolhida no gráfico PIM, incluem-se as variáveis individuais no modelo na Matriz de desenho. O número de linhas na Matriz de Desenho é o número de parâmetros determinado no gráfico do PIM, e o número de colunas é o número de parâmetros individuais que se deseja estimar. Assim, as Matrizes de Desenho para o exemplo cujos gráficos PIM estão representados na Figura 8 são:

- 8.A: A probabilidade da primeira captura é diferente nas amostras (p_1, p_2, c):

$$\begin{bmatrix} \text{Parâmetro} & \beta_{\text{intercepto}} & \beta_{\text{amostra}} & \beta_{\text{intercepto}} \\ p_1 & 1 & 1 & 0 \\ p_2 & 1 & 0 & 0 \\ c & 0 & 0 & 1 \end{bmatrix}$$

- 8.B: A probabilidade da primeira captura é igual nas amostras (p, c):

$$\begin{bmatrix} \text{Parâmetro} & \beta_{\text{intercepto}} & \beta_{\text{intercepto}} \\ p_1 & 1 & 0 \\ c & 0 & 1 \end{bmatrix}$$

Suponha, agora, que se deseja ajustar modelos com as parametrizações acima, porém incluindo o efeito do sexo na recaptura dos indivíduos. As Matrizes de desenho seriam, então:

- A probabilidade da primeira captura é diferente nas amostras, e a recaptura está associada ao sexo ($p_1, p_2, c(\text{sexo})$):

$$\begin{bmatrix} \text{Parâmetro} & \beta_{\text{intercepto}} & \beta_{\text{amostra}} & \beta_{\text{intercepto}} & \beta_{\text{sexo}} \\ p_1 & 1 & 1 & 0 & 0 \\ p_2 & 1 & 0 & 0 & 0 \\ c & 0 & 0 & 1 & 1 \end{bmatrix}$$

- A probabilidade da primeira captura é igual nas amostras, e a recaptura está associada ao sexo (p , $c(\text{sexo})$):

$$\begin{bmatrix} \text{Parâmetro} & \beta_{\text{intercepto}} & \beta_{\text{intercepto}} & \beta_{\text{sexo}} \\ p_1 & 1 & 0 & 0 \\ c & 0 & 1 & 1 \end{bmatrix}$$

O Programa MARK utiliza funções de ligação para utilizar modelos lineares para estimar as probabilidades de captura e recaptura, o que permite a inclusão de variáveis individuais nos modelos, através das matrizes de desenho. A transformação que o MARK faz a partir das funções de ligação será apresentada adiante.

4.1.4. MODELOS LINEARES

Há situações em que se pode supor que determinadas características dos indivíduos, ou mesmo características do ambiente, afetam as probabilidades de captura e recaptura. Com isso, é conveniente modelar estas características, com o objetivo de conhecer suas influências nas probabilidades. O Programa MARK[®] emprega modelos lineares, o que fornece flexibilidade para se lidar com um grande número de questões e hipóteses com os dados disponíveis.

A idéia básica dos modelos lineares é simples: a variável resposta em qualquer análise estatística pode ser expressa em função de um ou mais fatores. De uma maneira geral, um modelo linear é expresso na forma matricial como:

$$y = \beta X + \varepsilon$$

Onde y é o vetor da variável resposta, β é o vetor de parâmetros, X é a matriz com os valores das variáveis independentes e ε é o vetor de termos dos erros aleatórios. Estes modelos são chamados lineares porque a parte não aleatória da expressão (βX) é uma combinação linear dos parâmetros.

Como exemplo, considere que se deseja investigar a diferença entre a

medida da circunferência do crânio de meninos e meninas em idade escolar, onde se supõe que esta diferença está relacionada somente ao sexo, se estiver. Realiza-se a medida em k meninas e j meninos. O modelo linear que representa esta relação será:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Onde cada valor de Y_i é a soma do intercepto (β_0), do produto do coeficiente angular (β_1) e da variável x , e do termo de erro aleatório (ε). A variável x representa a codificação da variável que especifica o sexo e assume valor 1 para meninas e 0 para os meninos. A hipótese a ser testada é se o valor de β_1 é estatisticamente diferente de zero ($H_0: \beta_1 = 0$). Se o valor de β_1 não difere de zero, diz-se que não há diferença na circunferência craniana entre meninos e meninas.

Na notação matricial acima, o modelo de regressão apresentado no exemplo é:

$$y = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1j} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2k} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} * \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1j} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{2k} \end{bmatrix} = X\beta + \varepsilon$$

Onde k é o número de meninas e j , o número de meninos, a matriz X , formada por duas colunas, é a parte não aleatória do modelo. A primeira coluna, composta por "1s", corresponde ao intercepto (β_0) e a segunda, composta por "0s" e "1s", corresponde ao sexo (β_1).

A estrutura da Matriz de Desenho, no programa MARK[®], está relacionada com o número de parâmetros do modelo que se deseja ajustar. Utilizando o exemplo acima, pode-se decompor a equação que modela a circunferência craniana em relação ao sexo em duas:

$$y_i = \beta_0 + \beta_1 x_i$$

Sexo	Equação
Masculino (x = 0)	$y_1 = \beta_0(\mathbf{1}) + \beta_1(\mathbf{0})$
Feminino (x = 1)	$y_2 = \beta_0(\mathbf{1}) + \beta_1(\mathbf{1})$

A matriz de desenho corresponde aos coeficientes que multiplicam os “ β s”:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Suponha, agora, que se investiga a redução no nível de colesterol frente à exposição a determinado medicamento entre os sexos, com dois níveis de exposição ao tratamento: os submetidos ao tratamento e os controles. A equação para este modelo seria:

$$y_{ij} = \beta_0 + \beta_1(\text{sexo}) + \beta_2(\text{tratamento}) + \beta_3(\text{sexo} * \text{tratamento}) + \varepsilon$$

Onde o termo $\beta_3(\text{sexo} * \text{tratamento})$ representa a interação entre o sexo e o tratamento. Decompondo esta equação, tem-se:

Grupo	Equação
Masculino tratamento (sexo = 0; tratamento = 1)	$y_{11} = \beta_0(\mathbf{1}) + \beta_1(\mathbf{0}) + \beta_2(\mathbf{1}) + \beta_3(\mathbf{0})$
Masculino controle (sexo = 0; tratamento = 0)	$y_{12} = \beta_0(\mathbf{1}) + \beta_1(\mathbf{0}) + \beta_2(\mathbf{0}) + \beta_3(\mathbf{0})$
Feminino tratamento (sexo = 1; tratamento = 1)	$y_{21} = \beta_0(\mathbf{1}) + \beta_1(\mathbf{1}) + \beta_2(\mathbf{1}) + \beta_3(\mathbf{1})$
Feminino controle (sexo = 1; tratamento = 0)	$y_{22} = \beta_0(\mathbf{1}) + \beta_1(\mathbf{1}) + \beta_2(\mathbf{0}) + \beta_3(\mathbf{0})$

A matriz de desenho, construída com base nos coeficientes, será:

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Para ajustar um modelo sem o termo de interação, ou:

$$y_{ij} = \beta_0 + \beta_1(\text{sexo}) + \beta_2(\text{tratamento}) + \varepsilon$$

Decompondo a equação:

Grupo	Equação
Masculino tratamento	$y_{11} = \beta_0(\mathbf{1}) + \beta_1(\mathbf{0}) + \beta_2(\mathbf{1})$
Masculino controle	$y_{12} = \beta_0(\mathbf{1}) + \beta_1(\mathbf{0}) + \beta_2(\mathbf{0})$
Feminino tratamento	$y_{21} = \beta_0(\mathbf{1}) + \beta_1(\mathbf{1}) + \beta_2(\mathbf{1})$
Feminino controle	$y_{22} = \beta_0(\mathbf{1}) + \beta_1(\mathbf{1}) + \beta_2(\mathbf{0})$

A matriz de desenho será:

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

Observa-se que a matriz deste último modelo é a mesma do modelo anterior, sem a última coluna.

Nos dois exemplos acima, a variável resposta é contínua, isto é, podem assumir valores no intervalo $[-\infty, +\infty]$. Entretanto, nos modelos de captura e recaptura, as variáveis-resposta são dicotômicas, ou seja, 1 = capturado e 0 = não-capturado. Torna-se necessário, então, realizar uma “transformação” da probabilidade, de maneira que seja possível ajustar os modelos lineares com as variáveis que representam as características dos indivíduos e/ou do ambiente. Esta transformação é feita através da função de ligação (*link function*). O Programa MARK[®] disponibiliza uma gama de funções de ligação, a saber: seno, logística, log-log, log-log complementar, logarítmica, Identidade, parâmetro-específica e absoluta. Os modelos de captura para populações fechadas, condicionados às observações (modelos de Huggins) utilizam a função logística como função de ligação (Huggins, 1989, 1991).

Para explicitar a transformação da probabilidade de captura através da função logística, considere a variável resposta Y , dicotômica, como uma função de uma ou mais variáveis explicativas. Seja $Y = 1$ se o indivíduo foi capturado e $Y = 0$ se não o foi. Admite-se x como o vetor das variáveis explicativas e $p = \Pr(Y = 1|x)$ como a probabilidade da variável resposta que se deseja ajustar. A função linear desta probabilidade será obtida utilizando-se a transformação logística da probabilidade p :

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

Onde β_0 é o intercepto e β_1 é o vetor dos coeficientes dos parâmetros. Uma vez que $\theta = \ln[p/(1-p)]$ tem o inverso $p = e^\theta/(1 + e^\theta) = 1/(1 + e^{-\theta})$, então a estimativa retro-transformada de \hat{p} será:

$$\hat{p} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}}$$

Em outras palavras, é possível expressar a probabilidade do evento (captura ou recaptura) como uma função linear de um vetor de variáveis explicativas. O modelo logístico é um caso especial de uma classe mais generalizada de modelos lineares onde uma função $f = f(m)$ permite que seja assumido que qualquer variável está relacionada linearmente com o vetor de variáveis explicativas. A função f é a ligação entre o componente aleatório do modelo (variável resposta) e o componente fixo (variáveis explicativas). Por esta razão, a função $f(m)$ é referida como função de ligação. O Programa MARK[®] estima o intercepto e o vetor dos coeficientes dos parâmetros e então reconstitui o valor da captura a partir dos valores das variáveis explicativas, x . Este processo é feito em duas etapas: i) reconstituição das estimativas do parâmetro de captura a partir de β_0 , β_1 e x ; e ii) computação do valor do parâmetro de captura de f usando a retro-transformação f^{-1} (Cooch e White, 2008).

4.1.5. MODELOS COM VEROSSIMILHANÇA CONDICIONADA (HUGGINS)

Huggins (1989, 1991) introduziu um procedimento para a estimação do tamanho de uma população fechada, quando as probabilidades de captura são heterogêneas, modelando as probabilidades de captura e recaptura em termos de variáveis observáveis, como sexo, peso, história de captura, etc. A modelagem é realizada construindo a verossimilhança condicionada aos indivíduos capturados para estimar os parâmetros.

Seja p_{ij} a probabilidade do indivíduo i ser capturado na amostra j , onde $i = 1, 2, 3, \dots, N$ são os indivíduos da população (N é o tamanho da população) e $j = 1, 2, \dots, t$ são as amostras. A verossimilhança condicionada aos indivíduos capturados pode ser escrita em termos de:

$$\gamma_{ij} = \frac{p_{ij}}{1 - (1 - z_{ij}) \prod_{l=j}^t (1 - p_{ij}^*)}$$

Onde p_{ij}^* é p_{ij} quando $z_{ij} = 0$, com z_{ij} sendo o indicador da história

anterior de captura do indivíduo i . Ou seja:

$$z_{ij} = \begin{cases} 1, & \text{se o indivíduo } i \text{ foi capturado antes da amostra } j \\ 0, & \text{se o indivíduo } i \text{ não foi capturado antes da amostra } j \end{cases}$$

Então, γ_{ij} é a probabilidade do indivíduo i ser capturado na amostra j dado sua história anterior de captura e dado que foi capturado pelo menos uma vez durante o estudo.

Admitindo-se agora que $x_{ij} = 1$ se o indivíduo i é capturado na amostra j e $x_{ij} = 0$ se não o for, e renomeando os indivíduos capturados como 1, 2, 3, ..., n e os não capturados como $n+1$, $n+2$, $n+3$, ..., N , a verossimilhança condicional é proporcional a:

$$L = \prod_{i=1}^n \prod_{j=1}^t \gamma_{ij}^{x_{ij}} (1 - \gamma_{ij})^{(1-x_{ij})}$$

Que depende somente dos indivíduos presentes nas amostras. A função de ligação para o ajuste segundo características dos indivíduos e/ou ambiente é a logística, $\ln[p_{ij}/(1 - p_{ij})]$ (Huggins, 1991). Segundo o autor, os estimadores são assintoticamente normais e suas variâncias podem ser estimadas a partir da matriz das segundas derivadas. Utilizando dados de um experimento com informações sobre sexo; idade categorizada em jovem, semi-adulto e adulto; peso em gramas e histórico de captura da espécie *peromyscus maniculatus*, Huggins ajustou diversos modelos, a partir das variáveis observáveis e a história de captura. Considerando $i = 1, 2, 3, \dots, N$ os indivíduos da população e $j = 1, 2, \dots, t$ o número de amostras, expõem-se os modelos:

MODELO 0 (M_0)

$$\ln \frac{p_{ij}}{(1 - p_{ij})} = \beta_0$$

Este modelo assume que todos os indivíduos têm a mesma probabilidade de captura e recaptura, que não variam entre as amostras, nem de acordo com variáveis individuais e/ou ambientais, nem com a história anterior de captura.

MODELO 1 (M_1)

$$\ln \frac{p_{ij}}{(1 - p_{ij})} = \beta_0 + \beta_b z_{ij}$$

Esta parametrização permite que a probabilidade de captura varie de acordo com a história anterior individual de captura (o subscrito “b” vem do Inglês *behavioral*)

MODELO 2 (M_2)

$$\ln \frac{p_{ij}}{(1 - p_{ij})} = \beta_0 + \beta_j$$

Neste modelo a probabilidade de captura varia somente entre as amostras.

MODELO 3 (M_3)

$$\ln \frac{p_{ij}}{(1 - p_{ij})} = \beta_0 + \beta_j + \beta_b z_{ij}$$

Aqui a parametrização permite que a probabilidade de captura varie entre as amostras e de acordo com a história individual anterior de captura.

MODELO 4 (M₄)

$$\ln \frac{p_{ij}}{(1 - p_{ij})} = \beta_0 + \beta_{\text{sexo}} + \beta_{\text{idade}} + \beta_{\text{peso}} * \text{peso}_i$$

Onde β_{sexo} é o efeito do sexo; β_{idade} , o efeito da idade, β_{peso} é o efeito de uma unidade do peso e peso_i é o peso de cada indivíduo i . Neste modelo a probabilidade de captura varia somente em função das características individuais.

MODELO 5 (M₅)

$$\ln \frac{p_{ij}}{(1 - p_{ij})} = \beta_0 + \beta_{\text{sexo}} + \beta_{\text{idade}} + \beta_{\text{peso}} * \text{peso}_i + \beta_b z_{ij}$$

Com esta parametrização, a probabilidade de captura depende tanto das variáveis individuais quanto da história individual anterior de captura.

MODELO 6 (M₆)

$$\ln \frac{p_{ij}}{(1 - p_{ij})} = \beta_0 + \beta_{\text{sexo}} + \beta_{\text{idade}} + \beta_{\text{peso}} * \text{peso}_i + \beta_j + \beta_b z_{ij}$$

A probabilidade de captura varia de acordo com as variáveis individuais, entre as amostras e de acordo com o histórico anterior individual de captura.

MODELO 7 (M₇)

$$\ln \frac{p_{ij}}{(1 - p_{ij})} = \beta_0 + \beta_{\text{sexo}} + \beta_{\text{idade}} + \beta_{\text{peso}} * \text{peso}_i + \beta_b z_{ij} + \theta_i$$

Onde θ_i representa características não observadas do indivíduo i . Aqui, além das variáveis observadas, da história anterior de captura e entre as amostras, a probabilidade de captura pode variar também em função

destas características não observadas.

Como já mencionado anteriormente e visto neste trabalho de Huggins, uma gama de modelos podem ser ajustados quando se dispõe de informações sobre a história de captura e características do indivíduo ou do ambiente. O autor, em seu estudo, utilizou os critérios AIC (*Akaike Information Criteria*) para escolher o modelo que melhor se ajustou aos dados, que foi o Modelo 5 (Huggins, 1991).

Para estimar o tamanho da população, a probabilidade de o indivíduo i ser capturado pelo menos uma vez durante o estudo será:

$$p_i(\beta) = 1 - \prod_{j=1}^t (1 - p_{ij}^*) \quad (1)$$

Onde β é o vetor dos parâmetros associados ao modelo ajustado. Um estimador não-viciado do tamanho da população será:

$$\hat{N}(\beta) = \sum_{i=1}^n p_i(\beta)^{-1} \quad (2)$$

E a variância (Huggins, 1991):

$$\text{var}[\hat{N}(\beta)] = \sum_{i=1}^n p_i(\beta)^{-2} [1 - p_i(\beta)] \quad (3)$$

O erro-padrão de $\hat{N}(\beta)$ é a raiz quadrada de sua variância. O intervalo com 95% de confiança é estimado adicionando-se ao/diminuindo-se do valor da estimativa o produto $\{1,96 \times \text{ep}[\hat{N}(\beta)]\}$:

$$\text{IC95\%} = \hat{N}(\beta) \pm 1,96 * \text{ep}[\hat{N}(\beta)] \quad (4)$$

Cabe ressaltar que o Programa MARK[®] calcula as variâncias dos parâmetros e do estimador derivado da população antes de realizar a transformação logística, de sorte que os limites do intervalo, apresentados na janela de resultados, não são simétricos.

Os modelos de Huggins foram os ajustados para os dados sobre nascidos vivos em Sergipe, no segundo e terceiro trimestres de 2006, que têm como fontes o Registro Civil do IBGE e o Sistema de Informação sobre Nascidos Vivos, SINASC. Cada fonte de dados foi considerada uma amostra, ou ocasião, sendo o SINASC a primeira ocasião (primeira captura) e o Registro Civil a segunda ocasião (recaptura). Para a desagregação geográfica, duas abordagens foram tomadas: microrregiões e regionais de saúde. Estas desagregações foram consideradas grupos de indivíduos e a inclusão de variáveis individuais é discutida nos procedimentos metodológicos.

4.1.6. SELEÇÃO DOS MODELOS E PONDERAÇÃO DOS PARÂMETROS

As saídas dos modelos ajustados são armazenadas em um banco de dados, que também mantém os dados importados para o programa, o que viabiliza uma completa descrição do processo de construção do modelo. Uma janela de resultados resume e apresenta estas informações (Figura 9).

Através desta janela, recuperaram-se os gráficos de Índices de Parâmetros dos Modelos e as matrizes de desenho, que são as informações sobre o processo de construção do modelo. A janela de resultados também exibe os critérios de informação AIC (*Akaike Information Criteria*), conceito que é explanado adiante; a diferença dos critérios AIC entre o modelo em questão e o modelo com menor critério AIC; o peso com base nos critérios AIC; a verossimilhança do modelo; o número de parâmetros estimados e os resíduos de deviance, dados que permitem avaliação dos modelos ajustados.

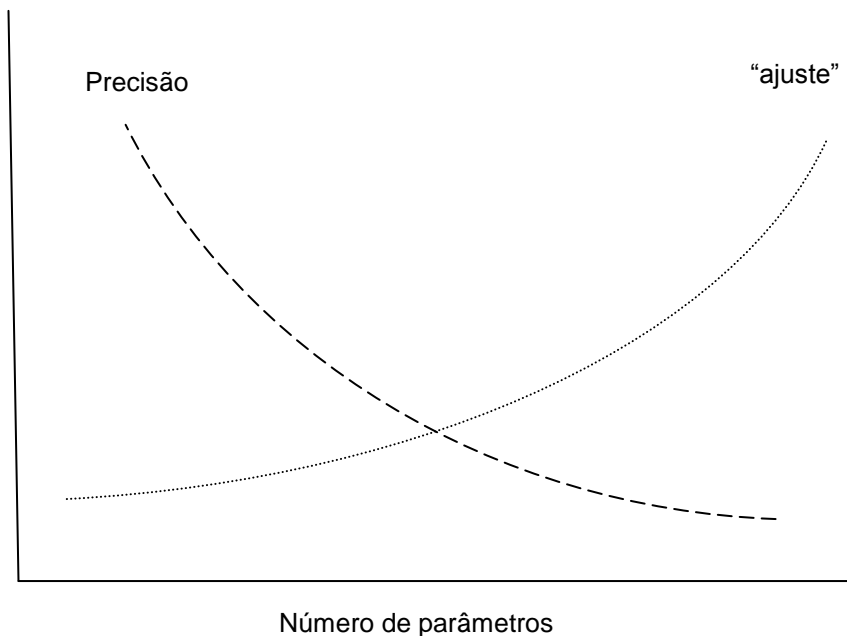
Cada um dos tipos de modelos disponíveis no MARK[®] permite o ajuste de “sub-modelos”, onde é possível estimar o número total de indivíduos considerando os p_i 's e c_i 's variando ou não ao entre as amostras e entre grupos definidos. É possível, também, considerar que p_i e c_i são iguais e constantes em todas as amostras e grupos, e assim por diante.

Figura 9 – Janela de Resultados dos modelos ajustados no Programa MARK[®]

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{p(idmae) c(.)}	2906,7378	0,0000	0,51087	1,0000	3	2900,7313
{p(idmae) c(idmae)}	2908,6925	1,9547	0,19224	0,3763	4	2900,6816
{p(idmae) c(sexo)}	2908,6925	1,9547	0,19224	0,3763	4	2900,6816
{p(idmae) c(idmae+sexo)}	2910,6143	3,8765	0,07354	0,1440	5	2900,5980
{p(idmae) c(idmae+sexo+sexo*idmae)}	2912,5313	5,7935	0,02820	0,0552	6	2900,5084
{p(.) c(.)}	2918,4248	11,6870	0,00148	0,0029	2	2914,4215
{p(.) c(idmae)}	2920,3420	13,6042	0,00057	0,0011	3	2914,3355
{p(.) c(sexo)}	2920,3784	13,6406	0,00056	0,0011	3	2914,3719
{p(.) c(idmae+sexo)}	2922,2951	15,5613	0,00021	0,0004	4	2914,2882
{p(.) c(idmae+sexo+sexo*idmae)}	2924,2149	17,4771	0,00008	0,0002	5	2914,1986

Esta especificação é feita pelo gráfico PIM (Índices de Parâmetros do Modelo). Desta maneira, é possível ajustar vários modelos com os dados disponíveis, o que permite a comparação entre eles e a seleção daquele que melhor se ajusta aos dados. Cada “sub-modelo”, de acordo com sua especificação (um com p_i variando ao longo do tempo e outro com p_i constante, por exemplo) estimará um número determinado de parâmetros. Quanto maior o número de parâmetros de um modelo, melhor este se ajustará aos dados, porém a precisão das estimativas dos parâmetros diminui (Figura 10).

Figura 10 – Variação do ajuste do modelo e da precisão das estimativas dos parâmetros, de acordo com o número de parâmetros incluídos no modelo



Uma alternativa para se realizar um “balanço” entre ajuste e precisão é avaliar os modelos segundo critérios de informação. Um destes métodos é o AIC (*Akaike Information Criteria*), que consiste em se relacionar a verossimilhança do modelo (representando o seu ajuste) com o número de parâmetros que ele estima:

$$\text{AIC} = -2 \ln(L) + 2k$$

Onde L é a verossimilhança do modelo e k o número de parâmetros. Quanto maior o ajuste do modelo, maior será sua verossimilhança, diminuindo o valor $[-2 \ln(L)]$. O termo aditivo $[+ 2k]$ “penaliza” o valor do AIC. Quanto maior o número de parâmetros, menor será o AIC, pois a verossimilhança aumentará. Porém, a soma do termo $[2k]$ balanceia o valor do AIC. Assim, o modelo com menor AIC será aquele mais parcimonioso, em relação à sua verossimilhança e número de parâmetros.

Apesar de simples a interpretação e escolha do modelo que melhor se ajusta aos dados, por vezes pode haver modelos com valores AIC muito próximos, o que dificulta a escolha. Os modelos podem, então, ser calibrados de maneira a fornecer um índice de relativa “plausibilidade”, empregando-se os pesos normalizados de Akaike. Estes pesos, w_i , são calculados para cada modelo no conjunto I de modelos “candidatos” de acordo com a fórmula:

$$w_i = \frac{\exp\left(\frac{-\Delta AIC_i}{2}\right)}{\sum_{i=1}^I \left\{\exp\left(\frac{-\Delta AIC_i}{2}\right)\right\}}$$

Onde ΔAIC_i é a diferença entre o valor de AIC do modelo i e o valor do modelo com menor AIC. O peso w_i é considerado como uma evidência de que o modelo i é o melhor modelo no conjunto dos candidatos. Quanto maior o peso do modelo, pode-se interpretar que ele melhor “suporta” os dados (Cooch e White, 2008).

Ainda na análise dos pesos, é possível também atribuir o “quanto” determinado modelo é melhor do que outro, dentro do conjunto de modelos ajustados. Como exemplo, suponha que em um estudo com três amostras foram ajustados 5 modelos para populações fechadas (*Closed Captures*, ver Figura 7) e os resultados do ΔAIC_i e w_i são:

MODELO	PARAMETRIZAÇÃO	NÚMERO DE PARÂMETROS	SIGNIFICADO	ΔAIC_i	w_i
1	p_i, c_i, N	6 ($p_1, p_2, p_3, c_1, c_2, N$)	As probabilidades de captura e recaptura variam entre as amostras	1,6	0,287
2	p, c_i, N	4 (p, c_1, c_2, N)	A probabilidade de ser capturado pela primeira vez é constante e a probabilidade de recaptura varia entre as amostras	0,0	0,639
3	p_i, c, N	5 (p_1, p_2, p_3, c, N)	A probabilidade de ser capturado pela primeira vez varia entre as amostras e a probabilidade de recaptura é constante	7,0	0,055
4	p, c, N	3 (p, c, N)	As probabilidades de captura e recaptura são constantes nas duas amostras	8,5	0,019
5	$p = c, N$	2 (p, N)	As probabilidades de captura e recaptura são iguais e constantes entre as amostras	13,5	0,000
TOTAL					1,000

Verifica-se que o modelo 2 é o que melhor se ajusta aos dados, pois tem o menor valor AIC ($\Delta AIC_2 = 0$) e o maior peso. Entretanto, o modelo 1 também apresenta relativo “suporte” aos dados, com 1/3 do peso total. Diz-se, aqui, que o ajuste do modelo 2 é 2,23 vezes o do modelo 1 ($0,639/0,287 = 2,23$), enquanto que os demais modelos essencialmente não são “suportados” (Cooch e White, 2008).

Mesmo sendo possível identificar o modelo que melhor se ajusta aos dados, sempre resta a dúvida: o modelo representa a realidade? Esta “incerteza” de se saber qual modelo é o correto, o que mede a realidade, pode ser acomodada nas estimativas através da ponderação destas, em um conceito similar ao de média ponderada. Para esta ponderação, utilizam-se os pesos normalizados de Akaike (w_i) (Cooch e White, 2008).

Calcula-se a média de um parâmetro θ a partir dos valores de $\hat{\theta}_i$ de cada modelo i do conjunto de I modelos, ponderando pelo w_i :

$$\hat{\theta} = \sum_{i=1}^I \hat{w}_i \hat{\theta}_i$$

A variância desta ponderação possui dois componentes: i) a variância condicional da amostra, dado o modelo e ii) a variação associada à incerteza da seleção do modelo:

$$\widehat{\text{var}}(\hat{\theta}) = \left[\sum_{i=1}^I \hat{w}_i * \sqrt{\widehat{\text{var}}(\hat{\theta}_i | M_i) + (\hat{\theta}_i - \hat{\theta})^2} \right]^2$$

Onde $\widehat{\text{var}}(\hat{\theta}_i | M_i)$ é a variância do parâmetro θ_i dado o modelo i , e $\hat{\theta}$ é a estimativa do parâmetro ponderada por w_i ($\sum_{i=1}^I w_i \hat{\theta}_i$). Esta estimativa da variância do parâmetro ponderado leva em consideração tanto a variância dos modelos, quanto a variação resultante da incerteza da seleção dos modelos (Cooch e White, 2008).

Além da ponderação dos parâmetros, no Programa MARK[®] também é possível ponderar a estimativa derivada do tamanho total da população, (\hat{N}), e calcular sua variância. O cálculo do Intervalo de Confiança é efetuado a partir da abordagem dos indivíduos ausentes nas duas amostras, n_{00} , que é a diferença entre o total de indivíduos na população e o número de indivíduos distintos observados nas duas amostras ($n_{00} = N - r$). Sob o pressuposto de que este número segue uma distribuição log-normal, os limites do intervalo de confiança de \hat{N} são (Lucacks, 2008):

$$\left[r + \frac{\hat{n}_{00}}{C}, \quad r + \hat{n}_{00} * C \right]$$

$$\text{Onde } \hat{n}_{00} = \hat{N} - r, \text{ e } C = \exp \left\{ 1,96 \left[\ln \left(1 + \frac{\widehat{\text{var}}(\hat{N})}{\hat{n}_{00}^2} \right) \right]^{1/2} \right\}$$

Uma vez que $\widehat{N} = \widehat{n}_{00} + r$, então $\widehat{\text{var}}(\widehat{N})$ é a mesma de \widehat{n}_{00} , porque r é um número conhecido e constante. Assim,

$$\widehat{\text{var}}(\widehat{N}) = \left[\sum_{i=1}^I \widehat{w}_i * \sqrt{\widehat{\text{var}}(\widehat{N}_i | M_i) + (\widehat{N}_i - \widehat{N})^2} \right]^2$$

Onde

$$\widehat{N} = \sum_{i=1}^I \widehat{w}_i \widehat{N}_i$$

A variância dos estimadores ponderados fornecidos pelo Programa MARK[®] são meras ponderações das variâncias de cada estimativa e não levam em consideração os componentes de variação de incerteza associada ao modelo e de variação condicional da amostra, dado o modelo (Cooch e White, 2008). Assim, o cálculo da variância dos parâmetros ponderados deve ver “manual”.

Ao se ajustar um conjunto de modelos candidatos, é necessário saber o quanto o modelo saturado, o mais parametrizado, se ajusta aos dados, isto é, fazer o diagnóstico deste modelo. Este diagnóstico pode ser realizado de várias maneiras, e uma delas é através dos testes GOF (do Inglês, *Goodness Of Fit*). Resumidamente, o teste GOF é um procedimento para testar os pressupostos relacionados ao modelo que se deseja ajustar aos dados. Para se corrigir a falta de ajuste, mede-se o quanto de variação extra binomial existe (*deviance*). A magnitude desta “extra-dispersão” é conhecida como fator de inflação da variância, \hat{c} . A partir desta medida, corrigem-se os modelos do conjunto de candidatos. Este procedimento não será detalhado, pois não há testes GOF disponíveis no Programa MARK[®] para modelos com co-variáveis para populações fechadas (Cooch e White, 2008).

Esta seção apresentou resumidamente o método de captura e recaptura que foi empregado na pesquisa. Acredita-se que o modelo de

Huggins para populações fechadas é adequado para os dados disponíveis, que permitirão a aplicação deste método para o conhecimento do número estimado de nascidos vivos ocorridos nas microrregiões e regionais de saúde do estado de Sergipe, e, conseqüentemente, do sub-registro destes eventos vitais nas bases de dados oficiais.

4.1.7. CRÍTICAS AO MÉTODO

Algumas considerações devem ser feitas em relação ao método de captura-recaptura, visto que seu emprego indiscriminado pode levar a resultados completamente errôneos. As principais críticas tecidas residem na necessidade de se assumirem pressupostos tão difíceis de serem contemplados nos dados epidemiológicos. Na área da ecologia, é possível garantir a independência entre as amostras, uma vez que estas são desenhadas especificamente para este fim. Na epidemiologia, por sua vez, os pesquisadores “aproveitam” listas já existentes, elaboradas por vezes com objetivos distintos (Papoz e colaboradores, 1996). A impossibilidade de verificação da dependência quando só há duas listas disponíveis é o grande “Calcanhar de Aquiles” ou o “escândalo silencioso” do método (Hook e Regal, 1995).

Listas não identificadas, ou cuja identificação é proibida por questões éticas, também dificultam o relacionamento entre os registros, obrigando os pesquisadores a lançarem mão de técnicas probabilísticas de relacionamento, que adicionam mais erros não aleatórios às estimativas e cujos modelos disponíveis não abrangem este tipo de erro (Desenclos e Hubert, 1994; Papoz e colaboradores, 1996).

Hook e Regal (1999), grandes entusiastas desta metodologia, dada a vasta literatura por eles produzida sobre o tema, publicaram um guia de “recomendações para apresentação e avaliação das estimativas por captura-recaptura na epidemiologia”:

- i) Identificar no artigo não somente os objetivos do estudo, mas

também o uso que se quer fazer com as estimativas, comentando sobre as conseqüências que podem ter a utilização da sub ou superestimação que pode ocorrer. Considerar *quem precisa com qual nível de confiança para qual propósito*;

- ii) Explicitar todas as fontes utilizadas e antecipar seus padrões e características;
- iii) Explicitar o critério de pareamento dos dados, seus possíveis erros e critérios utilizados;
- iv) Relatar as prováveis dependências entre as listas. Nos resultados e discussão observar tendências específicas que aparentemente contradizem as expectativas, mesmo que não significantes, e discutir as possíveis explicações para estas contradições, além de flutuação amostral;
- v) Especificar as características da população estudada, no que tange à área geográfica, local de residência, migração, morte, nascimentos e as prováveis incertezas frente ao fechamento geográfico e demográfico desta população;
- vi) Apresentar os dados numéricos da maneira mais clara e completa possível, exibindo a intersecção entre as listas, estratificando pelas co-variáveis importantes para o estudo;
- vii) Explorar a estrutura dos dados e examinar sua consistência interna, suas implicações e sua coerência com o conhecimento sobre o tema;
- viii) Considerar as conseqüências sobre os modelos frente à ocorrência de caselas com números pequenos ou nulos;
- ix) Discutir a sensibilidade das estimativas em relação à acurácia do pareamento e outros pressupostos implícitos;
- x) Apresentar o conhecimento anterior disponível que pode ser utilizado para interpretar os resultados e pesquisar esta informação para estudo *ad hoc*, se necessário;
- xi) Interpretar os resultados exclusivamente à luz dos objetivos do estudo.

Frente à quase constante ocorrência de heterogeneidade, os autores sugerem que, para modelagem log-linear, sejam adotadas pelo menos uma das seguintes abordagens:

- a) Ajustar modelos log-lineares em cada estrato obtido através da combinação das co-variáveis suspeitas de gerar heterogeneidade;
- b) Ajustar um único modelo log-linear com inclusão das co-variáveis geradoras da heterogeneidade.
- c) Ponderar as estimativas pelos métodos bayesianos (AIC).

Após estas breves apresentações da problemática do método de captura e recaptura, com o conhecimento de suas potencialidades e limitações, descreve-se, a seguir, o protocolo da pesquisa que foi implementado.

4.2. PROCEDIMENTOS METODOLÓGICOS

4.2.1. LIMPEZA E PADRONIZAÇÃO DOS BANCOS

Independente das definições de outras áreas de conhecimento, neste trabalho, doravante, são aplicados como sinônimos os termos “bancos de dados” e “bases de dados”, e referem-se aos arquivos onde os dados foram alocados. Em tais arquivos, cada linha (registro) representa um nascido vivo e cada coluna (campo) representa um atributo deste nascido vivo, como local de nascimento, município de residência da mãe, sexo, etc. Os bancos foram exportados para o programa Access[®], onde foram realizadas as rotinas de limpeza, padronização e pareamento descritas a seguir.

O Ministério da Saúde disponibilizou os dados preliminares de 2006 da base de dados do SINASC e o IBGE disponibilizou sua base do Registro Civil de Nascidos Vivos do ano de 2006, para o presente estudo. O SINASC totalizou 2.803.287 nascimentos vivos em 2006, ao passo que o Registro Civil identificou 2.625.454 registros civis deste evento vital, ocorridos em 2006. De uma maneira geral, o SINASC identifica mais nascidos vivos do

que o Registro Civil, o que caracteriza o já conhecido sub-registro civil de nascimento vivo, com exceção dos estados de São Paulo e Goiás. Estes números maiores de registro em cartório do que os captados pelo SINASC podem ser devidos ao fato dos dados do Ministério da Saúde serem preliminares, e sofrerem acréscimos posteriores a esta análise, diminuindo ou até mesmo eliminando esta diferença (Tabela 3 e Figura 11).

A maior diferença entre o número de nascimentos vivos capturados pelo SINASC e os registrados em cartório está nas unidades da federação das Regiões Norte e Nordeste, diminuindo consideravelmente nas demais regiões (Tabela 3 e Figura 11).

A partir de 2006, o IBGE incluiu o número da Declaração de Nascido Vivo em seu formulário de coleta. A partir deste número é possível realizar o pareamento entre as duas bases e assim identificar os nascimentos vivos que não foram capturados por uma delas.

Descreve-se, em seguida, a fase de limpeza dos bancos, padronização das variáveis comuns às duas bases, o relacionamento dos registros, a avaliação deste pareamento e a busca por alternativas para o melhor pareamento possível.

Na base do Registro Civil, há campos para o código da unidade da federação de residência da mãe e ocorrência do nascimento vivo e os campos do município de residência da mãe e município de nascimento apresentavam o dígito verificador, tendo, então, tamanho de 5 dígitos. Já na base do SINASC, o município de residência da mãe e o município de nascimento são codificados de maneira que os dois primeiros dígitos correspondem à unidade da federação e os quatro dígitos seguintes são o código do município, sem o dígito verificador. Para a padronização das variáveis que representam o município de residência da mãe e município de nascimento, os campos de unidade da federação e município foram concatenados e o dígito verificador eliminado na base do IBGE. Desta maneira, os códigos dos municípios de ocorrência do nascimento vivo e residência da mãe obedecerão ao padrão do SINASC, ou seja, tem seis dígitos, sendo os dois primeiros indicativos da unidade da federação e os

quatro seguintes apontam o código do município.

A fim de facilitar a realização de filtros por unidade da federação, foram mantidos os campos de código das UFs de ocorrência e residência, assim como foram criados estes campos na base do SINASC.

Em relação ao sexo do recém-nascido, as codificações tanto no Registro Civil quanto no SINASC são idênticas, sendo o código 1 atribuído a meninos e 2 para meninas. Estes códigos foram modificados para 0 para meninos e 1 para meninas.

A idade da mãe, nas duas bases, é coletada em números inteiros da idade em anos completos e não houve modificações nesta variável.

A data de nascimento em ambas as bases é apresentada no formato texto DDMMAAAA e realizou-se transformação para o formato data dd/mm/aaaa.

Tabela 3 – Número de registros de nascimentos vivos ocorridos em 2006 capturados pelas bases SINASC e Registro Civil (RC), segundo a unidade da federação de residência da mãe.

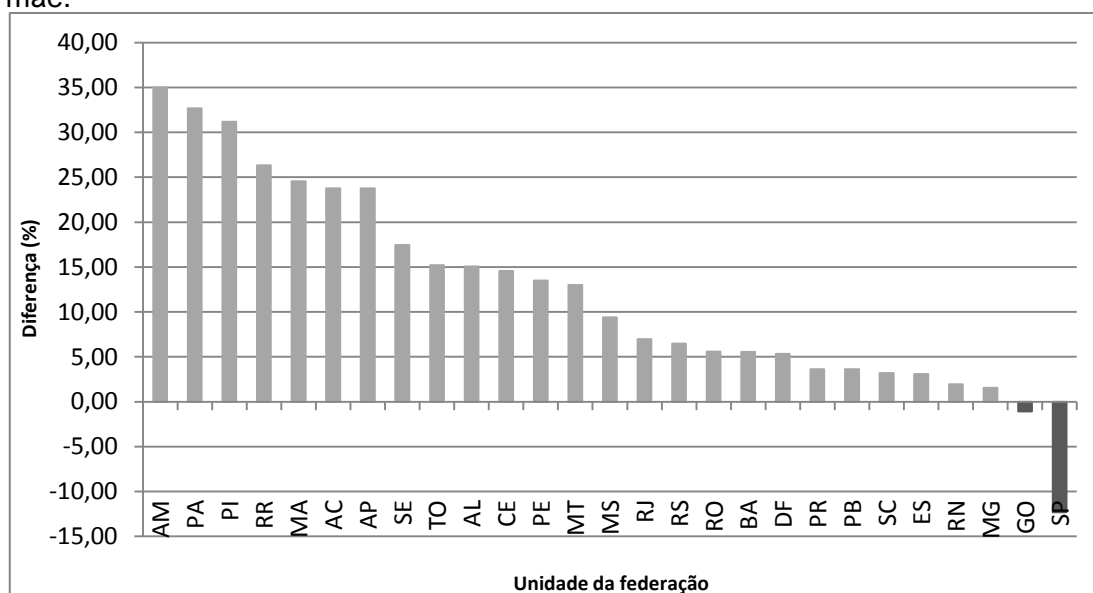
REGIÃO/UF	SINASC	RC	DIFERENÇA (%)
TOTAL	2.803.287	2.625.454	6,34
<i>ESTRANGEIRO</i>	-	285	-
Brasil	2.803.287	2.625.169	6,35
<i>UF DE RESIDÊNCIA DESCONHECIDA</i>	-	4.310	-
Região Norte	311.157	221.945	28,67
AC	16.067	12.247	23,78
AM	74.350	48.325	35,00
AP	14.573	11.110	23,76
PA	148.749	100.126	32,69
RO	24.895	23.503	5,59
RR	8.484	6.249	26,34
TO	24.039	20.385	15,20
Região Nordeste	863.136	747.452	13,40
AL	56.563	48.033	15,08
BA	210.078	198.397	5,56
CE	134.053	114.518	14,57
MA	121.192	91.452	24,54
PB	59.021	56.884	3,62
PE	144.643	125.135	13,49
PI	55.257	38.032	31,17
RN	45.413	44.526	1,95
SE	36.916	30.475	17,45
Região Sudeste	1.035.681	1.087.164	(4,74)*
ES	50.380	48.820	3,10
MG	262.095	258.011	1,56
RJ	211.960	197.213	6,96
SP	511.246	583.120	(12,33)*
Região Sul	378.347	360.956	4,60
PR	153.016	147.450	3,64
RS	141.362	132.229	6,46
SC	83.969	81.277	3,21
Região Centro-Oeste	214.966	203.342	5,41
DF	45.072	42.664	5,34
GO	81.315	82.195	(1,07)*
MS	39.491	35.783	9,39
MT	49.088	42.700	13,01

Fontes: MS. Sistema de Informação sobre Nascidos Vivos 2006. Dados Preliminares. Brasília, 2006, CD; e

IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

* Em São Paulo e Goiás o número nascidos vivos de registrados em cartório foi maior do que o captado pelo SINASC.

Figura 11 - Diferença percentual dos nascimentos vivos ocorridos em 2006 entre as bases SINASC e Registro Civil, segundo a unidade da federação de residência da mãe.



Fontes: MS. Sistema de Informação sobre Nascidos Vivos 2006. Dados Preliminares. Brasília, 2006, CD; e
 IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

Os códigos para o preenchimento do local de nascimento são parecidos nos dois sistemas, porém não são iguais. Os códigos para “local ignorado” e “outros locais” são diferentes no Registro Civil e no SINASC. Por outro lado, os códigos para hospital, outro estabelecimento de saúde e domicílios são idênticos nos dois sistemas. Realizou-se a padronização dos códigos.

O dicionário dos bancos finais do Registro Civil e do SINASC podem ser conferidos no Anexo V.

Antes do pareamento, realizou-se verificação de entradas duplicadas, onde foram identificados os registros com o mesmo número da DN e todas as demais variáveis do banco foram comparadas. Havendo uma simples repetição das entradas, estes registros foram classificados como ‘duplicados convergentes’ e um dos registros foi eliminado do banco. Havendo diferença em qualquer um dos demais campos, estes registros foram classificados como ‘duplicados divergentes’, retirados do banco e adicionados a um banco adicional, chamado ‘DUPLICADOS’. O banco SINASC para Sergipe

apresentou 4 entradas duplicadas, consideradas “divergentes” e estas foram mantidas à parte. No banco do Registro Civil não foram encontrados registros com número da DN duplicado para o estado de Sergipe.

A intenção de se realizar o pareamento através do número da DN esbarrou no fato de grande quantidade de registros do Registro Civil estar com esta variável em branco, mais de 40% (Tabela 4). O preenchimento desta variável tornou-se mais completo ao longo do ano (dados não mostrados). Entretanto, em nenhuma das unidades da federação o preenchimento foi completo. O Estado de São Paulo foi o que se mostrou com menor percentual de número da DN ignorada, porém, o sub-registro de nascidos vivos neste estado é considerado baixo, ou praticamente nulo (IBGE 2006), sendo, inclusive os cálculos de natalidade, fecundidade e mortalidade infantil realizados diretamente, sem necessidade de cálculos por métodos demográficos indiretos (RIPSA 2008). Dentre as unidades da federação do Norte e Nordeste, Acre e Sergipe tiveram os menores percentuais de número da DN em branco. A estimativa de sub-registro pelo método de Brass para o estado do Acre, no ano de 2006, não está entre as maiores do Brasil (ver Figura 1, página 40). Por esta razão, optou-se por analisar o estado de Sergipe, a fim de se verificar a possibilidade do pareamento através do número da DN e explorar alternativas para o pareamento.

Em seguida à montagem dos bancos, o relacionamento entre os registros foi realizado através de consultas de identificação de registros coincidentes e não coincidentes do programa MS-Access[®], a partir da variável DN, de maneira a permitir a construção da tabela de contingência 2 x 2 incompleta utilizada para estimação.

A fim de se capturarem registros inicialmente excluídos das bases de dados, uma vez identificados os registros não coincidentes, buscaram-se na tabela de duplicados e também foram procurados nas bases originais, da fonte onde não houve correspondência, de maneira a não haver perdas.

Tabela 4 – Total de registros de nascidos vivos ocorridos em 2006 e registros com Número da Declaração de Nascido Vivo (DN) em branco no Registro Civil. Brasil, Grandes Regiões e Unidades da Federação de residência da mãe, 2006

REGISTRO CIVIL: NASCIDOS VIVOS 2006			
UF/REGIÃO	NASCIDOS EM 2006	DN EM BRANCO	%
TOTAL	2.625.454	1.076.570	41,01
<i>ESTRANGEIRO</i>	285	152	53,33
BRASIL	2.625.169	1.076.418	41,00
<i>UF DESCONHECIDA</i>	4.310	3.027	70,23
Região Norte	221.945	128.886	58,07
AC	12.247	1.946	15,89
AM	48.325	28.533	59,04
AP	11.110	7.040	63,37
PA	100.126	63.264	63,18
RO	23.503	15.632	66,51
RR	6.249	3.606	57,71
TO	20.385	8.865	43,49
Região Nordeste	747.452	432.364	57,85
AL	48.033	45.017	93,72
BA	198.397	131.631	66,35
CE	114.518	56.342	49,20
MA	91.452	42.886	46,89
PB	56.884	44.625	78,45
PE	125.135	57.892	46,26
PI	38.032	17.090	44,94
RN	44.526	30.277	68,00
SE	30.475	6.638	21,78
Região Sudeste	1.087.164	290.033	26,68
ES	48.820	38.493	78,85
MG	258.011	141.699	54,92
RJ	197.213	109.594	55,57
SP	583.120	247	0,04
Região Sul	360.956	148.109	41,03
PR	147.450	67.867	46,03
RS	132.229	32.285	24,42
SC	81.277	47.957	59,00
Região Centro-Oeste	203.342	73.999	36,39
DF	42.664	12.526	29,36
GO	82.195	34.199	41,61
MS	35.783	9.468	26,46
MT	42.700	17.806	41,70

Fonte: IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

Assim, os registros que estavam no Registro Civil e não foram inicialmente relacionados no SINASC foram novamente procurados nas tabelas de duplicados, e nos arquivos dos outros estados do SINASC. Da mesma forma, os registros do SINASC não relacionados num primeiro momento com o Registro Civil foram buscados na base do IBGE. De fato, 222 registros do SINASC foram identificados na base de outros estados do Registro Civil, através do número da DN, onde constava município de residência da mãe fora de Sergipe. Entretanto, dos novos registros identificados no Registro Civil, em 36 constava a data de nascimento anterior ao ano de 2006 e por esta razão estes foram excluídos. Por outro lado, 205 registros do Registro Civil que inicialmente não foram pareados com o SINASC foram identificados nas bases de outros estados deste sistema.

Após esta fase de inclusão e exclusão de registros, a base do Registro Civil apresentou 30.661 registros de nascimentos vivos ocorridos de mães ali residentes, no ano de 2006. Já o SINASC resultou em 37.121 nascimentos vivos de mães residentes neste estado, no mesmo ano (Tabela 7).

Tabela 5 - Número e proporção (%) de registros com variáveis ignoradas nos bancos Registro Civil e SINASC. Sergipe, 2006.

VARIÁVEL	RC	SINASC
Número da DN	6.638 (21,65)	-
Município de residência da mãe	22 (0,07)	-
Município de nascimento	1 (0,00)	-
Sexo	1 (0,00)	4 (0,01)
Idade da mãe	201 (0,66)	-
Data de Nascimento	-	-
Tipo de gestação	-	59 (0,16)
Local de Nascimento	4 (0,01)	2 (0,00)
<i>TOTAL</i>	<i>30.661 (100)</i>	<i>37.121 (100)</i>

Fontes: MS. Sistema de Informação sobre Nascidos Vivos 2006. Dados Preliminares. Brasília, 2006, CD; e IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

Em análise da completude das variáveis que foram utilizadas nos dois bancos, verificou-se que ambas as bases mostram-se com excelente valor, mais de 99%, com exceção do número da DN no banco do Registro Civil (Tabela 5). Destaca-se, no SINASC, o preenchimento de 100% das variáveis município de residência da mãe, município de nascimento e a idade da mãe.

4.2.2. PAREAMENTO MENOS RESTRITO: NÚMERO DA DN

O pareamento pelo número da DN, batizado de Menos Restrito, gerou 22.785 pares, totalizando 44.997 ocorrências de nascimentos vivos em Sergipe, no ano de 2006. No Registro Civil, 7.876 registros não foram relacionados com o SINASC, em parte devido ao fato de que 6.638 registros da base do Registro Civil (21,64%) estavam com a variável número da DN em branco, o que prejudicou o pareamento. Na base do SINASC, 14.336 não foram relacionados com o Registro Civil, em parte também devido a esta ausência de informação na variável número da DN na base do IBGE. Cabe notar que a falta do número da DN na base do Registro Civil pode levar à uma superestimação do número total de ocorrências de nascidos vivos. O preenchimento desta variável na base do IBGE aumentou com a prática de coleta desta, que pôde ser verificada ainda no ano de 2006, com as informações dos terceiro e quarto trimestres bem mais completas do que as dos trimestres anteriores. No primeiro trimestre, 99,94% dos registros estavam com a DN em branco, proporção reduzida para 11,87% no segundo trimestre e 1,04% e 1,74% no terceiro e quarto trimestres, respectivamente.

A Lei 6.015, de 31/12/1973 (Brasil, 1973), determina que o registro civil do nascimento vivo deve ocorrer no máximo até três meses após o evento. Por esta razão, escolheu-se o período de nascimento no segundo e terceiro trimestres para eliminar a possibilidade de um nascido vivo ter sido registrado ainda dentro do prazo legal, porém fora da data englobada pelas bases, que é até 31/12/2006. Assim, ao se obter a estimativa do número de

nascidos vivos no segundo e terceiro trimestres de 2006, e compará-los com o Registro Civil e o SINASC, será possível estimar o sub-registro nestas bases.

Com o banco obtido na fase de limpeza e organização, realizou-se filtro para datas de nascimento entre 01/04/2006 e 30/09/2006. No período compreendido entre estas datas, o SINASC capturou 19.593 nascidos vivos e o Registro Civil, 17.344. O relacionamento entre as bases através do número da DN gerou 15.612 pares. A partir deste pareamento, conclui-se que as duas bases somam 21.325 registros, para o segundo e terceiro trimestres de 2006.

O Registro Civil, no período analisado, apresenta 808 registros com o número da DN ignorado, representando 4,65% dos registros nesta base. Quando se comparam os registros com DN preenchida e os que tem a DN ignorada, não há diferença estatística na média de idade da mãe ($p = 0,992$) nem na proporção do sexo do nascido vivo ($p = 0,510$). Entretanto, pergunta-se: destes 808 nascimentos vivos, sem o identificador único, quantos foram também capturados pelo SINASC?

Esta questão é de grande importância, uma vez que o número total de nascidos vivos capturados pelas duas fontes depende da correta identificação dos pares nelas presentes. Seriam estes 808 registros considerados como “pertencentes somente ao Registro Civil” realmente não pareáveis com o SINASC? Este número poderá ser estimado a partir dos resultados dos modelos selecionados, que são tratados mais adiante. Se, por um lado, é possível estimar quantos dos nascimentos vivos sem número da DN também foram capturados pelo SINASC, por outro, é necessário investigar a consistência entre os registros pareados: Os pares coincidem no preenchimento das variáveis colhidas pelos dois sistemas? Para responder a esta questão, os registros pareados foram comparados e a seção a seguir relata esta comparação.

4.2.2.1 OS PARES SÃO REAIS?

Além da falta do número da DN em uma parte da base do IBGE, resta dúvida sobre a exatidão deste pareamento, isto é, mesmo com número de DN igual, os registros indicam o mesmo evento? Os dois sistemas de informação coletam algumas variáveis em comum, o que permite avaliar a consistência do pareamento. São elas: município de residência da mãe, município de nascimento, idade da mãe, data de nascimento, sexo do nascido vivo, tipo de gestação e local de nascimento, além do número da DN.

Foram criadas quatro classes de pares, a saber: pares reais, pares muito prováveis, pares prováveis e pares pouco prováveis. Dentre os registros pareados através da DN, foram considerados pares reais aqueles em que todas as demais variáveis coincidiam.

Acredita-se haver um vício de informação do município de residência da mãe, devido às diferenças nas ofertas de saúde nos municípios, tendendo as gestantes a mentirem o município de residência, com medo de não serem atendidas no município onde procuram atendimento médico (Waldvogel e Ferreira, 2003; Oliveira e Simões, 2005). No período analisado, segundo e terceiro trimestres de 2006, a idade da mãe é ignorada no Registro Civil em 71 dos registros pareados com o SINASC e isso gera diferença nesta variável entre as duas bases, quando é feita a pesquisa de comparação das idades da mãe. Na base do SINASC, o tipo de gravidez é ignorado em 28 registros pareados com o Registro Civil, e também gera diferença na comparação. Não houve ocorrências de local de nascimento ignorado dentre os registros pareados. Desta maneira, os critérios para determinar que os pares muito provavelmente são corretos (pares muito prováveis) foram:

- Município de residência diferente e demais variáveis iguais; ou
- Município de nascimento diferente e demais variáveis iguais; ou

- Idade da mãe ignorada em uma das bases e demais variáveis iguais; ou
- Tipo de gestação ignorada em uma das bases e demais variáveis iguais.

Por vezes, pode haver diferença na idade da mãe, desde o momento do nascimento vivo até o seu registro civil. Como a data de nascimento da mãe não é coletada em nenhuma das duas bases, não é possível afirmar que a diferença de um ano caracterize registros diferentes. A diferença de um dia na data de nascimento também pode não caracterizar esta distinção. O sexo da criança, por ser uma variável dicotômica, pode ter a ordem da entrada de dados diferente em cada um dos sistemas. Pares onde tanto o município de nascimento e o município de residência da mãe são diferentes, mas as demais variáveis são iguais também podem significar erro na informação por parte das mães ou responsáveis. Pares com estas pequenas diferenças, tendo as demais variáveis iguais, podem ser considerados como pares prováveis. Os critérios para os pares prováveis, por conseguinte, são:

- Local de nascimento diferente e demais variáveis iguais; ou
- Diferença na idade da mãe de um ano e demais variáveis iguais; ou
- Diferença de um dia na data de nascimento e demais variáveis iguais; ou
- Município de residência da mãe e município de nascimento diferentes e demais variáveis iguais; ou
- Sexo diferente e demais variáveis iguais; ou
- Tipo de gestação diferente e demais variáveis iguais

Os demais casos foram considerados como pares pouco prováveis, isto é, aqueles que apresentavam diferença em mais de uma variável, com exceção do município de nascimento e município de residência da mãe; ou a diferença de idade da mãe fosse maior do que um ano; ou, ainda, a data de

nascimento diferisse em mais de um dia. Os critérios para a definição dos pares estão resumidos no Quadro 1.

Dos 15.612 pares encontrados, 11.710 foram considerados pares reais, isto é, todas as variáveis coincidem nas duas bases. O que se enquadraram como pares muito prováveis somam 1.433 registros, ao passo que os pares prováveis são 1.137. Os pares pouco prováveis são 1.332 (Tabela 6).

Tabela 6 – Pares de registros do SINASC e Registro Civil, relacionados pelo número da DN, classificados quanto à consistência do preenchimento das variáveis. Sergipe, segundo e terceiro trimestres de 2006

CLASSIFICAÇÃO	NÚMERO DE PARES	PROPORÇÃO (%)	PROPORÇÃO ACUMULADA (%)
PAR REAL	11.710	75,01	75,01
PAR MUITO PROVÁVEL	1.433	9,18	84,19
PAR PROVÁVEL	1.137	7,28	91,47
PAR POUCO PROVÁVEL	1.332	8,53	100,00
TOTAL	15.612	100,00	...

Fontes: MS. **Sistema de Informação sobre Nascidos Vivos 2006**. Dados Preliminares. Brasília, 2006, CD; e IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

A maior parte dos registros pareados (75%) cumpriu os quesitos para serem considerados pares reais. Ao considerar que os critérios para definir um par muito provável permitem inferir que são pares reais, este número salta para mais de 84%. Por outro lado, em números absolutos, 2.469 pares foram considerados como prováveis ou pouco prováveis, o que levanta a dúvida da identificação destes eventos como sendo referentes aos mesmos nascidos vivos. A partir desta dúvida, um pareamento alternativo é proposto e será tratado na seção a seguir.

Quadro 1 – Critérios para classificação dos pares reais, pares muito prováveis e pares prováveis entre as bases SINASC e Registro Civil. Os casos que não se enquadram nas situações abaixo foram classificados como pouco prováveis.

SITUAÇÃO								CLASSIFICAÇÃO
DN	CODMUNRES	CODMUNNASC	DTNASC	IDMAE	SEXO	GEST	LOCNASC	
=	=	=	=	=	=	=	=	PAR REAL
=	≠	=	=	=	=	=	=	PAR MUITO PROVAVEL
=	=	≠	=	=	=	=	=	PAR MUITO PROVAVEL
=	=	=	=	IGNORADA	=	=	=	PAR MUITO PROVAVEL
=	=	=	=	=	=	IGNORADA	=	PAR MUITO PROVAVEL
=	≠	≠	=	=	=	=	=	PAR PROVÁVEL
=	=	=	=	≠ 1 ANO	=	=	=	PAR PROVÁVEL
=	=	=	≠ 1 DIA	=	=	=	=	PAR PROVÁVEL
=	=	=	=	=	≠	=	=	PAR PROVÁVEL
=	=	=	=	=	=	≠	=	PAR PROVÁVEL
=	=	=	=	=	=	=	≠	PAR PROVÁVEL

DN = Número da Declaração de Nascido Vivo;
 CODMUNRES = Município de residência da mãe;
 CODMUNNASC = Município de nascimento;
 DTNASC = Data de nascimento;
 IDMAE = Idade da mãe;
 SEXO = Sexo do recém-nascido;
 GEST = Tipo de gestação;
 LOCNASC = Local de nascimento.

4.2.3. PAREAMENTO MAIS RESTRITO: DN + CRITÉRIOS

Além do pareamento utilizando somente o número da DN como identificador único, um novo pareamento foi realizado, a saber: foram considerados pares os registros que se encaixam nos critérios de “par real” e “par muito provável”, já descritos. Neste novo procedimento, as bases do SINASC e do Registro Civil, no segundo e terceiro trimestres de 2006 em Sergipe, com 19.953 e 17.344 registros, respectivamente, foram novamente utilizadas para identificação dos pares.

Neste pareamento alternativo, batizado de Mais Restrito, 13.143 pares foram identificados, totalizando 23.794 nascimentos vivos no segundo e terceiro trimestre de 2006, em Sergipe (Tabela 7).

Tabela 7 – Número de nascidos vivos capturados pelo SINASC e Registro Civil (RC), no segundo e terceiro trimestres de 2006 em Sergipe, segundo os critérios de pareamento adotados.

Pareamento Menos restrito				
RC				
		Sim	Não	
SINASC	Sim	15.612	3.981	19.593
	Não	1.732		
		17.344		
<i>TOTAL DE NASCIDOS VIVOS: 21.325</i>				
Pareamento Mais restrito				
RC				
		Sim	Não	
SINASC	Sim	13.143	6.450	19.593
	Não	4.201		
		17.344		
<i>TOTAL DE NASCIDOS VIVOS: 23.794</i>				

Fontes: MS. **Sistema de Informação sobre Nascidos Vivos 2006**. Dados Preliminares. Brasília, 2006, CD; e IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

Como era de se esperar, diferentes critérios de pareamento irão gerar diferentes número totais de nascimento vivos. O pareamento menos restrito foi o que identificou o maior número de pares entre as duas bases, e por conseqüência, o menor número total de nascidos vivos diferentes.

4.2.4. RELACIONAMENTO PROBABILÍSTICO

A correta identificação dos registros correspondentes dentre as diferentes listas é essencial para o método de estimação de captura-recaptura, sendo esta um dos seus pressupostos. A violação deste pressuposto pode levar a estimações viciadas.

Se as listas disponíveis contem campos identificadores unívocos, como CPF, por exemplo, então o relacionamento entre as listas não terá grandes dificuldades. Entretanto, as listas epidemiológicas disponíveis dispõem somente de campos mais genéricos, como nome, data de nascimento, endereço, data do diagnóstico da doença. O preenchimento incorreto, ou ausente, nos campos identificadores torna a tarefa mais complicada. Uma alternativa é o relacionamento probabilístico das listas, que tem tomado espaço na literatura especializada nos últimos anos, onde são propostos vários métodos para esta tarefa (Seber e colaboradores, 2000; Lee e colaboradores, 2001; Laska e colaboradores, 2003; Machado, 2004; Cadwell e colaboradores, 2005; Drummond e colaboradores, 2006; Silva e colaboradores, 2006; Coutinho e Coeli, 2006).

Coeli (1998), em sua tese de doutoramento, apresenta boa explicação do funcionamento do relacionamento probabilístico entre listas. A autora descreve os passos necessários para um bom relacionamento, sendo o primeiro a organização e padronização das listas. Nesta etapa, sugere que os registros que possuem entradas livres, como nome e logradouro, por exemplo, sejam “limpos”, trocando todas as letras para maiúsculas, eliminando-se os espaços entre as palavras, exclusão do tipo de logradouro (rua, avenida, estrada, etc). Para otimizar o pareamento, o segundo passo é

a organização de blocos lógicos, onde os arquivos são divididos em blocos mutuamente exclusivos (bairro de residência, por exemplo), e as listas são comparadas dentro destes blocos.

A última etapa consiste no pareamento dos registros, onde para esta tarefa são construídos escores para os possíveis pares a serem obtidos a partir da blocagem empregada. Estes escores são construídos a partir de pesos atribuídos às variáveis utilizadas para o pareamento, e estes pesos, por sua vez, são formados de acordo com a concordância ou discordância do registro para a variável avaliada. Um registro, dentro de um bloco de uma lista será comparado com um registro na outra lista, dentro do mesmo bloco lógico. Se uma variável do registro concorda com a variável do outro registro da outra lista, então o fator de ponderação desta variável será de concordância. Se discordar, o fator de ponderação será de discordância. Este fator de ponderação é obtido com base nos conceitos já sedimentados na epidemiologia de sensibilidade e especificidade, onde a sensibilidade é a capacidade do campo identificar um par como verdadeiro, quando ele realmente é verdadeiro, e a especificidade é a probabilidade do campo identificar um par como falso, dado que este par é realmente falso (Coeli, 1998).

Os procedimentos para o pareamento probabilístico podem ser executados no aplicativo RECLINK III[®], desenvolvido pelos pesquisadores Cláudia Medina Coeli e Kenneth Camargo Jr. Este programa permite executar funções que padronizam nomes, endereços e datas; calculam escores para o pareamento; executam rotinas de blocagem e relacionamento e verificam duplicidades (Coeli e Camargo Jr, 2007).

No caso das duas bases em estudo, o número da DN deveria tomar vezes de identificador único, entretanto, a falta de preenchimento deste campo na base do Registro Civil prejudica o pareamento com o SINASC. Como visto, mesmo entre os pares identificados pela DN, divergência entre as variáveis deixam dúvidas acerca da qualidade dos pares. Desta maneira, é útil e prudente utilizar mais de um método para a identificação dos pares entre as duas bases.

Em virtude destas falhas no relacionamento determinístico, buscou-se realizar o relacionamento probabilístico, com auxílio do programa RECLINK III[®]. Entretanto, as variáveis disponíveis para o pareamento são pouco discriminatórias o que demanda várias estratégias de blocagem e pareamento, o que foge ao escopo deste trabalho, uma vez que o foco é o relacionamento determinístico. De fato, em uma tentativa de fazer o relacionamento através das variáveis CODMUNRES, CODMUNNASC, IDMAE, DTNASC, SEXO, LOCNASC e GEST no Access geraram muitos valores duplicados, significando muitos recém-nascidos nascidos no mesmo dia, mesmo município, de mães de mesma idade, residentes no mesmo município, mesmo tipo de gestação e mesmo local de nascimento. Na base do Registro Civil, há 1.212 ocorrências destes valores duplicados, ao passo que no SINASC há 1.351 destas ocorrências.

Machado (2002), em sua tese de doutoramento, associou as bases do SIM e SINASC para o município de São Paulo através do relacionamento probabilístico utilizando variáveis também pouco discriminatórias, porém a pesquisadora tinha disponível o peso ao nascer nas duas bases, variável um pouco mais discriminatória, o que facilitou o pareamento, ainda que o trabalho para a execução da tarefa também tenha sido extremamente extenso.

Por fim, apesar das limitações do relacionamento determinístico, as estimações foram realizadas a partir dos dois bancos obtidos e, seus resultados, comparados.

4.2.5. DADOS EMPÍRICOS DA UF SELECIONADA.

Nesta seção apresentam-se alguns indicadores sócio-demográficos e de mortalidade do estado de Sergipe. Buscou-se obter as informações com referências temporais mais próximas do período analisado neste trabalho, uma vez que nem todos os dados estão disponíveis para o ano de 2006. Ainda, em situações onde havia informações oriundas do Censo

Demográfico 2000 e da Contagem Populacional 2007, deu-se preferência para o emprego dos dados de 2007.

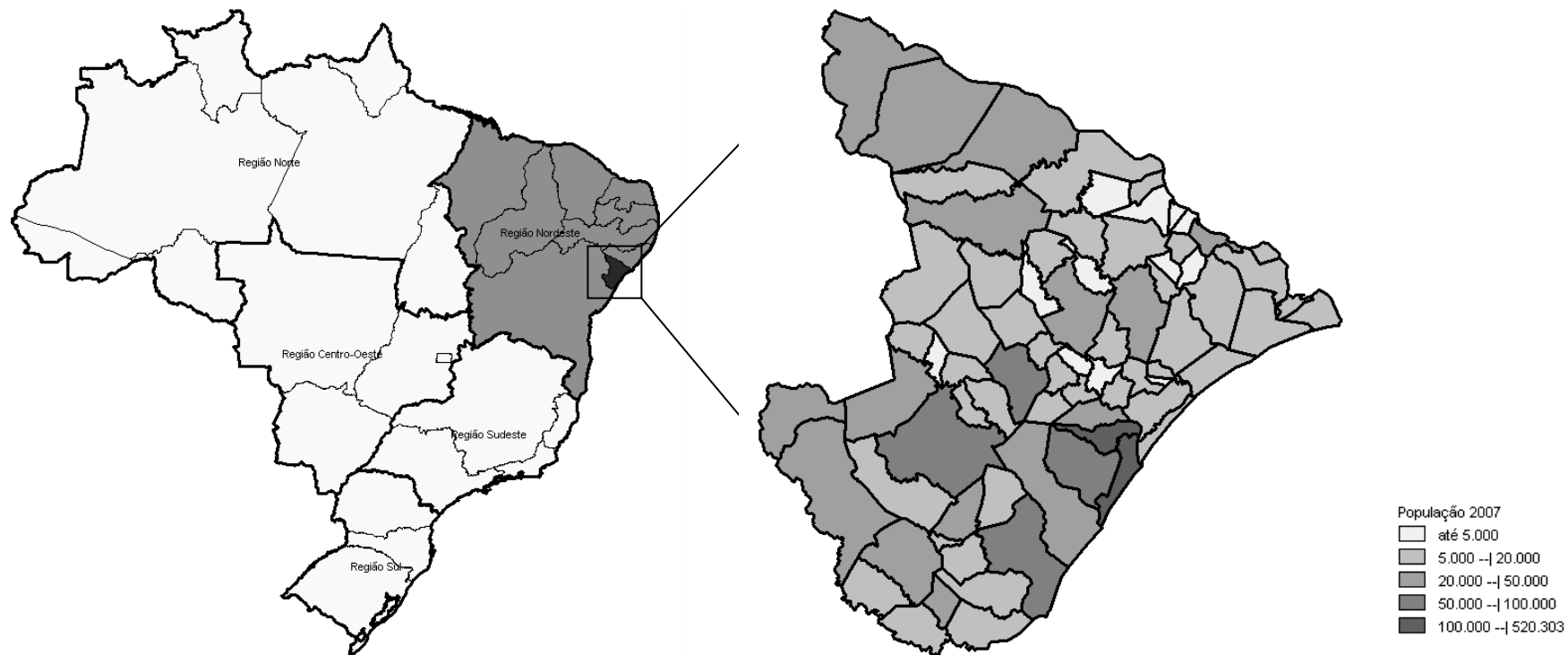
Os gráficos foram construídos no MS-Excel[®] 2007 e Stata[®] versão 9, os cartogramas foram elaborados no aplicativo Tabwin[®] e os testes estatísticos foram realizados no programa Stata[®], versão 9.

4.2.5.1 INDICADORES SÓCIO-DEMOGRÁFICOS E DE MORTALIDADE

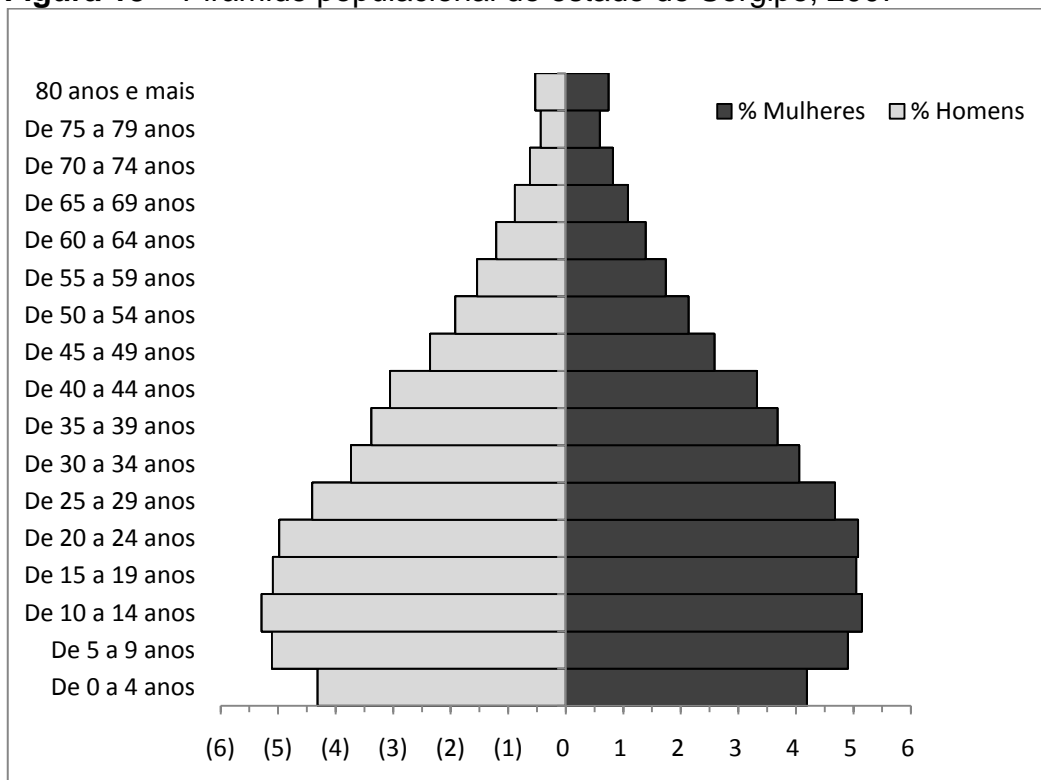
O Estado de Sergipe localiza-se na Região Nordeste do Brasil, possui 75 municípios instalados e a Contagem Populacional de 2007 (IBGE, 2007), apurou que o estado de Sergipe tinha 1.939.426 pessoas ali residentes, representando 1,05% da população residente no Brasil. A maioria dos municípios de Sergipe possui menos de 50.000 habitantes. (Figura 12). Por outro lado, a capital Aracaju, com 520.303 habitantes, concentra cerca de 27% da população do Estado. Ao considerar os municípios do entorno da capital, Aracaju, Barra do Coqueiro, Nossa Senhora do Socorro e São Cristóvão, a população soma 759.998 habitantes, ou 39% da população total de Sergipe.

A estrutura etária da população residente sugere que Sergipe inicia sua transição demográfica, onde se observa a pirâmide populacional com tendência à forma de colméia (Figura 13), decorrente da queda da fecundidade e não necessariamente da queda da mortalidade geral (Tabela 8). Crê-se que a tendência ao envelhecimento seja função da queda de fecundidade (Laurenti et al, 2005).

Figura 12 – Localização geográfica, divisão política e classes de população dos municípios do Estado de Sergipe, 2007



Fonte: IBGE, Contagem Populacional 2007

Figura 13 – Pirâmide populacional do estado de Sergipe, 2007

Fonte: IBGE. Contagem Populacional 2007

Assim como no Brasil, a taxa de fecundidade total na Região Nordeste mostra-se em declínio entre os anos de 1997 e 2005, porém, ainda, é sensivelmente maior do que a medida nacional. O estado de Sergipe, por sua vez, tem fecundidade total ainda maior do que da Região Nordeste, apesar de também apresentar queda no período. Nos anos de 2000 a 2003, as estimativas de fecundidade de Sergipe foram menores do que as da Região Nordeste, mas voltaram a se revelar maiores em 2004 e 2005 (Tabela 8).

Tabela 8 – Estimativas da Fecundidade total, Mortalidade infantil e Mortalidade geral padronizada. Brasil, Região Nordeste e Sergipe, 1997 a 2005.

Ano	1997	1998	1999	2000	2001	2002	2003	2004	2005
Fecundidade total (N.º filhos/mulher)									
Brasil	2,40	2,36	2,33	2,36	2,23	2,14	2,06	2,04	2,01
Reg. Nordeste	2,72	2,65	2,59	2,73	2,65	2,52	2,43	2,39	2,23
Sergipe	2,85	2,80	2,75	2,45	2,33	2,15	2,09	2,48	2,42
Mortalidade Geral Padronizada* (N.º óbitos/1000 habitantes)									
Brasil	-	-	5,32	6,54	6,46	6,17	6,11	6,01	5,93
Reg. Nordeste	-	-	6,31	7,53	7,43	7,16	7,03	6,90	6,90
Sergipe	-	-	5,39	7,10	6,88	6,66	6,56	6,47	6,51
Mortalidade Infantil (N.º óbitos < 1 ano/1000 nascidos vivos)									
Brasil	31,90	30,43	28,37	26,77	25,63	24,34	23,56	22,58	21,17
Reg. Nordeste	50,36	47,11	44,31	41,40	39,22	37,24	35,48	33,94	31,61
Sergipe	45,03	44,70	43,90	39,00	37,57	36,32	35,22	34,27	32,18

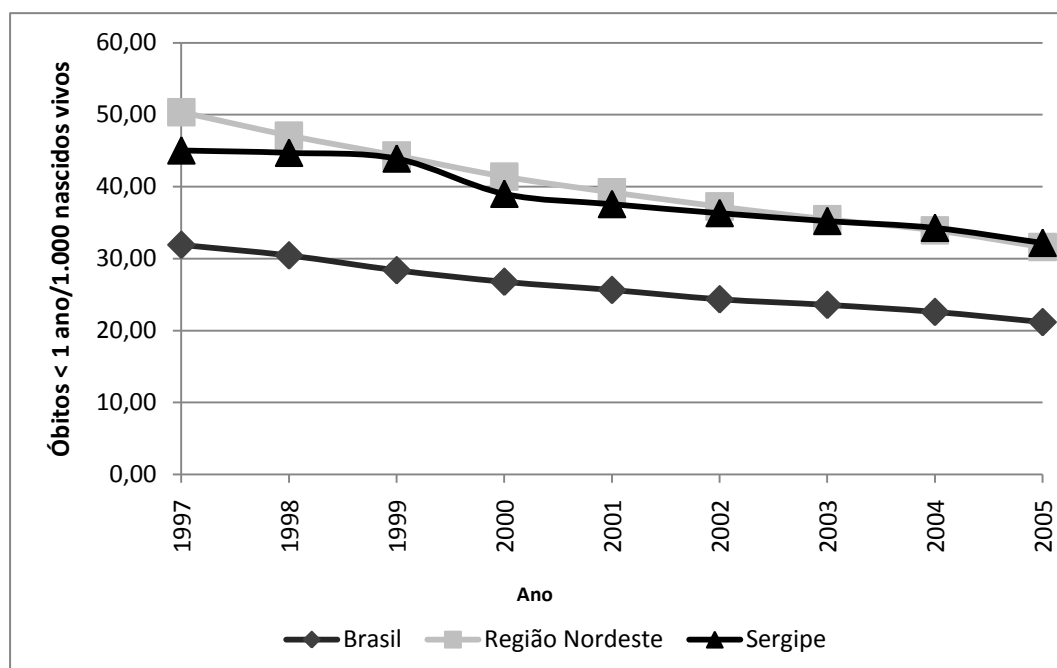
* Mortalidade Geral Padronizada não disponível para os anos 1997 e 1998.

Fonte: RIPSAs – Indicadores e Dados Básicos, Brasil, 2007.

A mortalidade infantil em Sergipe apresentou queda consistente nos últimos anos, saltando de mais de 45 óbitos infantis por mil nascidos vivos em 1997 para cerca de 32 óbitos por mil nascidos vivos em 2005 (Tabela 8).

As taxas de mortalidade infantil neste estado mostraram-se abaixo da média da Região Nordeste, como um todo, até o ano de 2003, porém, ainda muito acima da média nacional. A Região Nordeste experimenta queda gradual, porém contínua, ao longo dos anos do período analisado, passando de 50 óbitos por mil nascidos vivos, em 1997, para pouco mais de 31 óbitos por mil nascidos vivos, em 2005 (Tabela 8 e Figura 14).

Figura 14 – Estimativas da mortalidade infantil. Brasil, Região Nordeste e Estado de Sergipe, 1997 a 2005

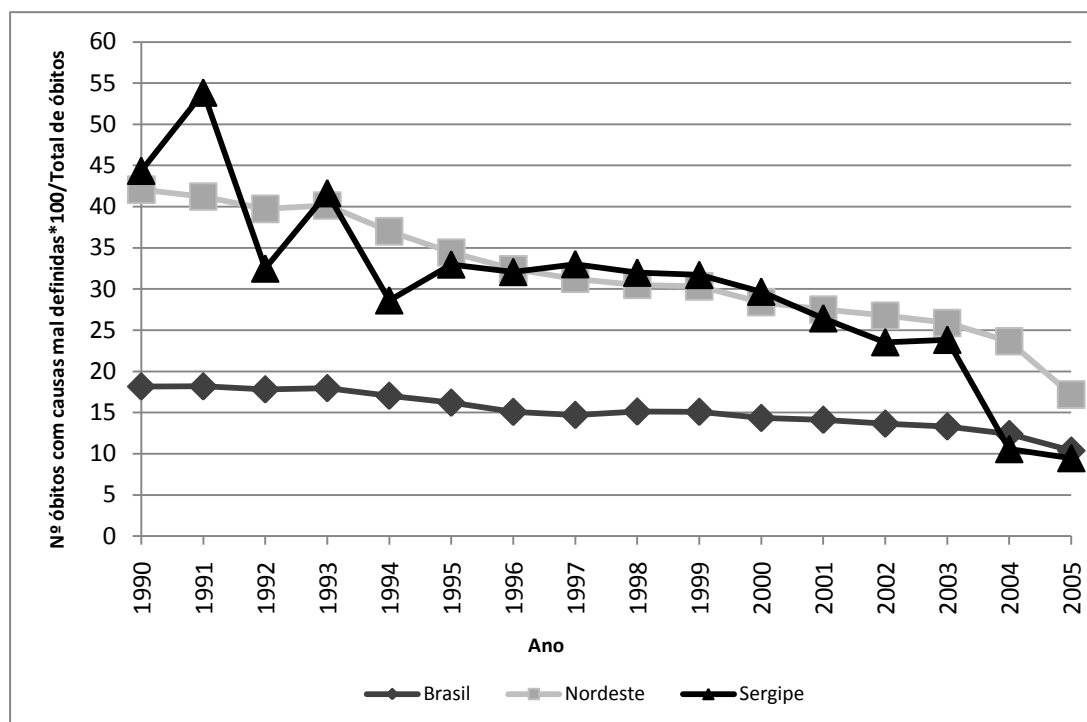


Fonte: RIPSA – Indicadores e Dados Básicos, Brasil, 2007

Pode-se perceber um grande avanço na identificação da causa da morte tanto no Nordeste quanto no estado de Sergipe, exibindo incremento na qualidade dos registros administrativos sobre mortalidade, ao longo da década de 1990 e início dos anos 2000 (Figura 15). Esta melhora permite sinalizar, também, melhora na qualidade dos serviços de saúde disponíveis para a população, evidenciando condições para o diagnóstico correto de doenças e capacitação profissional para o preenchimento adequado da Declaração de Óbito (RIPSA, 2008).

Em Sergipe, nos anos entre 1990 e 1995 a proporção de óbitos com causas mal definidas exibe comportamento irregular, chegando a mais de 50% em 1991. Entre 1996 e 2000, esta proporção estabilizou-se em torno dos 32% e, a partir de 2001, apresenta significativa queda, ficando em 2004 e 2005 abaixo, inclusive, da proporção nacional. A Região Nordeste, por sua vez, apresentou queda constante ao longo do período, mas ainda manteve-se mais de 60% maior do que a proporção do Brasil como um todo, em 2005 (Figura 15).

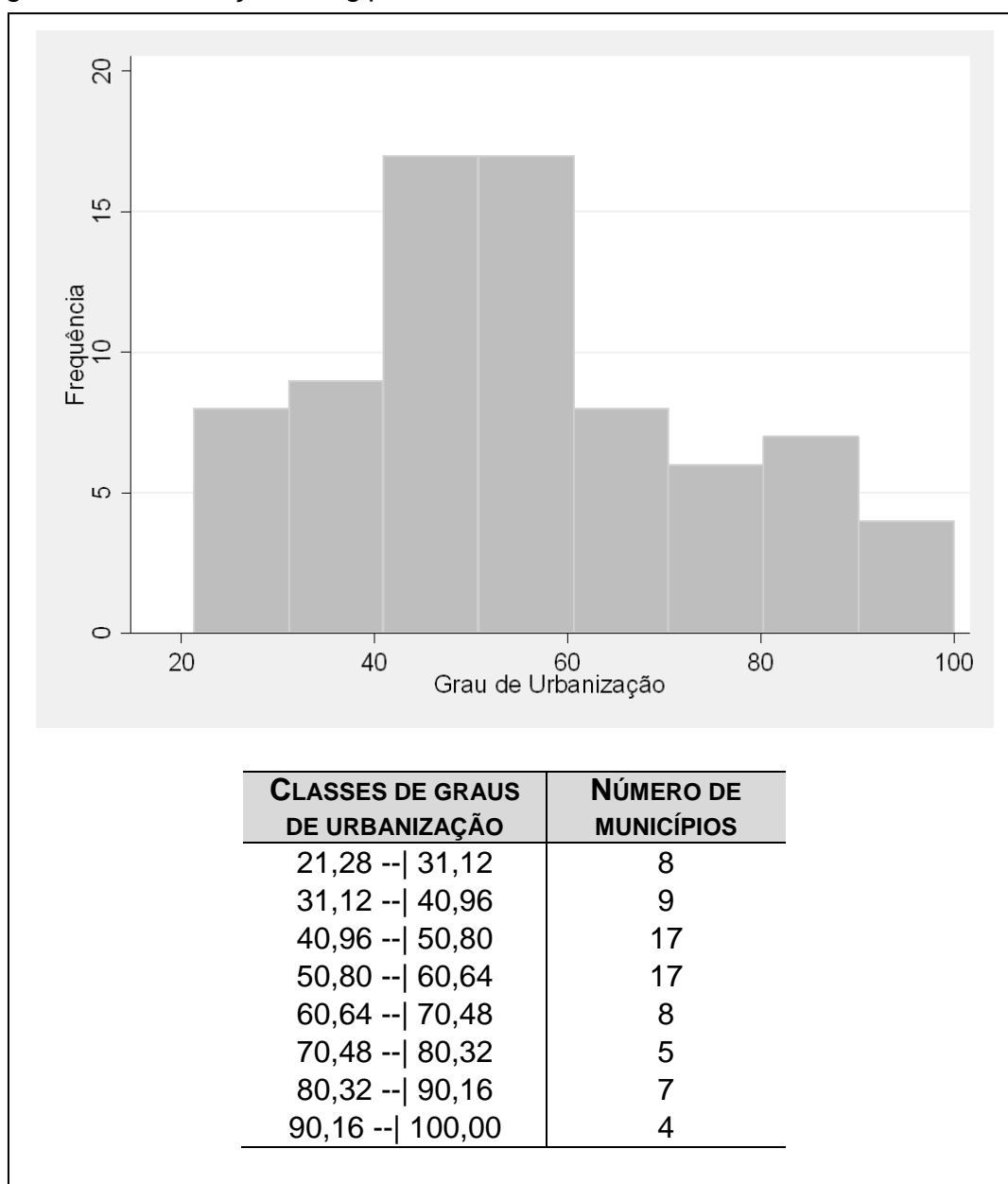
Figura 15 - Mortalidade Proporcional por Causas Mal Definidas. Brasil, Região Nordeste e Sergipe, 1990 a 2005.



Fonte: RIPSAs – Indicadores e Dados Básicos, Brasil, 2007

O grau de urbanização de uma divisão territorial define a proporção da população total que reside em áreas urbanas (RIPSA, 2008). Em Sergipe, no ano de 2007, o grau de urbanização era de 72,34%. Esta proporção é alta em relação à da Região Nordeste (64,92%), pois tem as menores proporções de urbanização, embora apresentem aumento ao longo das últimas décadas (RIPSA, 2008). O grau de urbanização do Estado de Sergipe é muito próximo à proporção do Brasil como um todo, e apesar de ser um estado relativamente urbanizado, quando se analisa este grau no nível de municípios, percebe-se grande amplitude na proporção de pessoas residentes em áreas urbanas (Figura 16). Dos 75 municípios instalados, 51 (68%) tem grau de urbanização igual ou menor que 60,64% e 16 (21,33%) apresentam proporção acima dos 70,48%, mais próximas à proporção do Estado.

Figura 16 – Distribuição do número de municípios, segundo as classes de grau de urbanização. Sergipe, 2007



Fonte: IBGE. Contagem Populacional, 2007

Percebe-se que há municípios predominantemente rurais, onde mais de 70% da população vive em áreas rurais e em apenas 4 mais de 90% das pessoas vivem em áreas urbanas (Figura 16). A partir desta distribuição, conclui-se que na maioria dos municípios grande parte da população reside em áreas rurais.

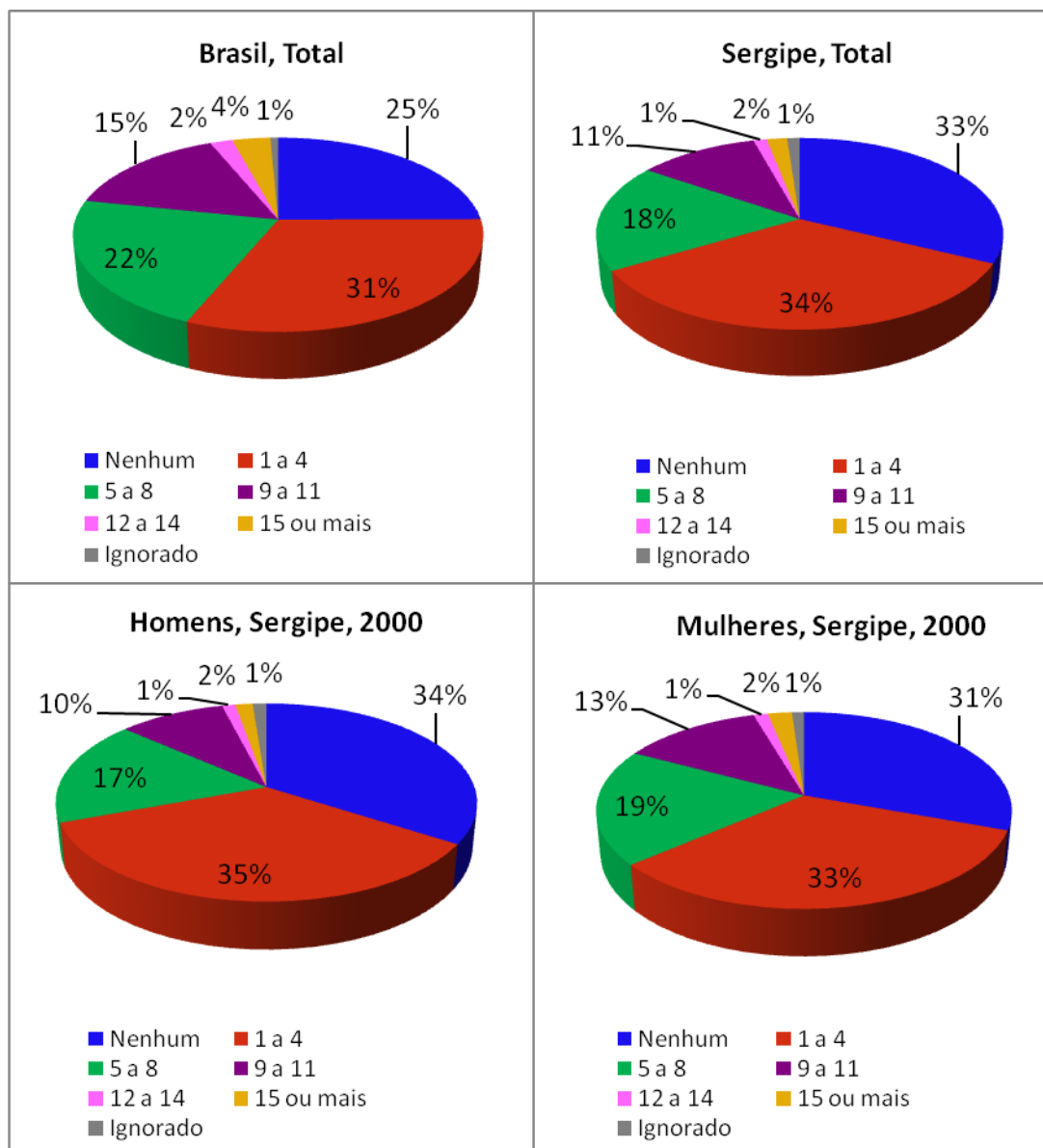
Em relação à escolaridade, o Censo Demográfico 2000 (IBGE, 2000)

constatou que, no Brasil, um quarto da população não tinha nenhum ano de estudo, 31% frequentaram a escola por 1 a 4 anos, 22% tinha entre 5 e 8 anos de estudo, 15% estudou entre 9 e 11 anos, 2% entre 12 e 14 anos de estudo e 4% da população tinha 15 anos de estudo ou mais. O estado de Sergipe apresentou proporção maior de pessoas sem nenhum ano de estudo, 33%, e apenas 2% tinham 15 anos ou mais de estudo, refletindo um nível de escolaridade menor do que o padrão nacional. Ao se analisar a escolaridade da população de Sergipe segundo o gênero, observa-se que, em 2000, as mulheres freqüentaram mais a escola do que os homens, pois 31% delas não tinham estudo algum, contra 34% dos homens sem estudo. Da mesma maneira, enquanto que 19% das mulheres tinham de 5 a 8 anos de estudo, a proporção nesta faixa entre os homens era de 17%. Das mulheres residentes em Sergipe, 13% freqüentaram a escola por 9 a 11 anos, ao passo que 10% dos homens informaram esta faixa de escolaridade. A partir dos 12 anos ou mais de estudo, a proporção de homens e mulheres residentes em Sergipe, no ano 2000, é a mesma (Figura 17).

Estudos em diferentes localidades no Brasil identificaram associação entre a baixa escolaridade da mãe e o sub-registro e registro civil tardio de nascidos vivos (Almeida, 1979; Mello Jorge e colaboradores, 1997; Crespo e colaboradores, 2006). Em estudo sobre o registro civil tardio nas microrregiões de Sergipe, no terceiro e quarto trimestres de 2006, Schmid (2009) verificou associação entre a baixa escolaridade da mãe e o registro civil após o período legal, na maioria das microrregiões do estado.

As bases de dados utilizadas no emprego do método de captura e recaptura são o SINASC e Registro Civil do IBGE (vide seção de Limpeza dos bancos e Relacionamento dos registros). A escolaridade da mãe é informada na Declaração de Nascido Vivo, porém, esta variável está presente somente na base do SINASC, tornando a inclusão desta variável no modelo a ser utilizado para estimação inviável. Entretanto, é pertinente exibir a distribuição da proporção de mulheres com idade entre 10 e 49 anos sem instrução, ao longo dos municípios de Sergipe, com o intuito de realizar comparação *a posteriori* dos resultados das estimações e esta distribuição.

Figura 17 – Distribuição percentual do nível de escolaridade da população residente, em anos completos de estudo. Brasil, Sergipe, total e segundo o sexo, 2000



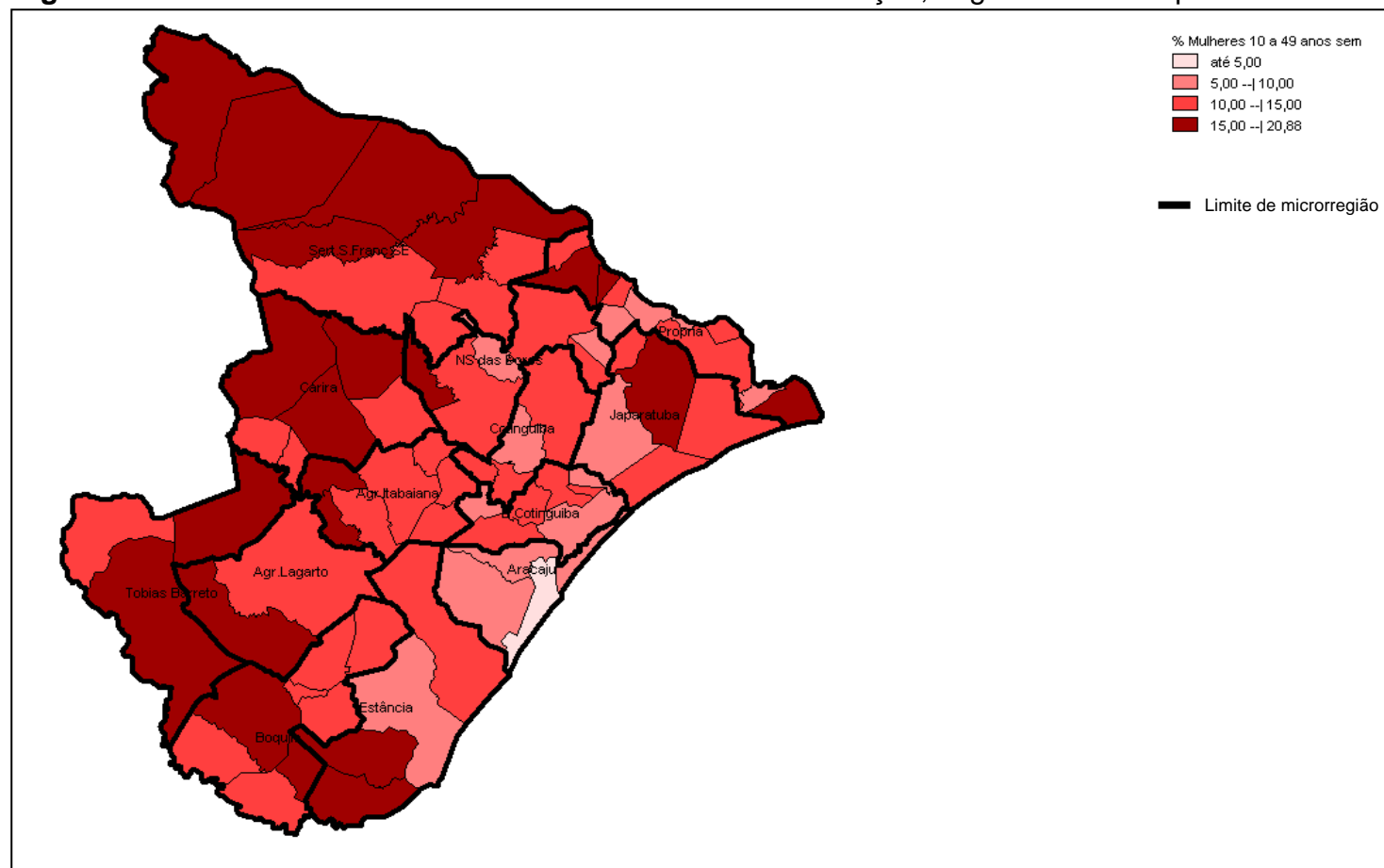
Fonte: IBGE. Censo Demográfico 2000

Segundo o Censo Demográfico 2000 (IBGE, 2000), em Sergipe havia 585.269 mulheres com idade entre 10 e 49 anos. Destas, 60.003 (10,25%) não tinham nenhum ano completo de estudo. Ao longo dos municípios, esta proporção variou de 4,44% em Aracaju a 20,88% em Riachão do Dantas, localizado na microrregião de Agreste de Lagarto, ao sudoeste do estado

(Figura 18). Os municípios de agreste e de sertão, ao sudoeste, norte e nordeste de Sergipe são os que apresentam maior percentual de mulheres na faixa etária analisada sem instrução. É interessante notar, também, que nos municípios mais próximos da capital as proporções de mulheres entre 10 e 49 anos sem nenhum ano de estudo são menores do que as dos mais afastados.

Ao observar os percentuais dos municípios dentro das microrregiões, percebe-se que, de uma maneira geral, as proporções de mulheres nesta faixa etária sem instrução são próximos, com exceção da microrregião de Propriá, que possui municípios onde os percentuais estão entre 5 e 10%, mas também engloba municípios com as demais faixas (Figura 18).

Figura 18 – Percentual de mulheres de 10 a 49 anos sem instrução, segundo o município de residência. Sergipe, 2000



Fonte: IBGE. Censo Demográfico 2000

4.2.5.2. ESTIMATIVAS DE SUB-REGISTRO CIVIL DE NASCIDOS VIVOS PELO MÉTODO DE BRASS

A estimação demográfica consiste na tentativa de mensurar valores de parâmetros demográficos básicos, como taxas de natalidade, de mortalidade ou o nível total de fecundidade. Estes parâmetros permitem identificar de que maneira determinada população irá evoluir, ao longo do tempo. O termo “indireto” é empregado para qualificar métodos usados em demografia, os quais produzem estimativas de determinado parâmetro com base em informações que se relacionam somente de maneira indireta com o seu valor. Um exemplo clássico é a utilização da proporção de crianças mortas, filhos de mulheres na faixa de 20 a 24 anos, para estimar a probabilidade de morte antes dos dois anos de idade. Em países com baixas coberturas censitárias e onde os sistemas de registro de eventos vitais são falhos, os métodos demográficos indiretos são empregados, de maneira a permitir estimativas dos parâmetros demográficos (UN, 1983).

Brass propôs, em 1975, um método de correção de diferentes parâmetros, baseado na comparação de taxas de fecundidade específicas de determinado período com a parturição média relatada. Variações do método foram elaboradas, tanto para suprir ausência de cumprimento dos pressupostos quanto para se ajustar à maior ou menor disponibilidade de dados. O IBGE utiliza o método original de Brass, que emprega a razão entre a parturição média, P , e a fecundidade média acumulada, F , com base nos dados de todas as crianças (Simões, 1999).

O método original (P/F) busca ajustar o nível das taxas de fecundidade específicas por grupos de idade, que são assumidas como representativas do verdadeiro padrão de fertilidade, para que sejam coerentes com o nível de fecundidade indicado pelas parturições médias de mulheres de grupos etários menores do que 30 ou 35 anos, tidos como acurados. As medidas das fecundidades médias acumuladas, F , e das parturições médias relatadas, P , em cada grupo etário com intervalo de 5 anos, são obtidas das taxas de fecundidade de cada grupo por acúmulo e

interpolação. Razões P/F são calculadas grupo por grupo e razão estimada para as mulheres mais jovens é usada como fator de correção, pelo qual são multiplicadas todas as taxas específicas de fecundidade correntes. Uma vez feita esta correção, as taxas específicas de fecundidade são multiplicadas pelo número de mulheres em cada grupo, obtendo-se, assim, o número estimado de nascidos vivos no período (UN, 1983).

O método de Brass pode ser interpretado como uma “padronização indireta, onde se aceita o padrão das taxas de fecundidade específicas correntes como verdadeiro, estimando-se o nível a partir da parturição observada pelas mulheres mais jovens” (Simões, 1999).

Resumidamente apresentando o método, três informações são necessárias: número de mulheres em idade fértil (entre 15 e 49 anos), divididas em grupos quinquenais de idade; número de filhos tidos desde o início do período fértil, classificados em grupos quinquenais de idade das mães e número de filhos tidos em período recente, geralmente nos 12 meses anteriores ao cálculo, também divididos em grupos quinquenais de idade das mães (UN, 1983).

Primeiro, calculam-se as parturições médias, P_i , em cada grupo quinquenal:

$$P_i = \frac{NFT_i}{M_i}$$

Onde

P_i = parturição média no grupo etário quinquenal i ;

NFT_i = número total de filhos tidos por mulheres do grupo quinquenal i ;

M_i = número total de mulheres do grupo quinquenal i ;

Em seguida, calculam-se as fecundidades correntes, preliminares, f_i :

$$f_i = \frac{NFPA_i}{M_i}$$

Em que

f_i = fecundidade corrente do grupo quinquenal i ;

$NFPA_i$ = Número de filhos tidos no período anterior, no grupo quinquenal i ;

M_i é definido como na expressão anterior.

A soma das fecundidades correntes preliminares, multiplicada por cinco, estima a fecundidade corrente acumulada até o limite de idade do grupo i (Φ_i):

$$\Phi(i) = 5 \left[\sum_{j=1}^i f(j) \right]$$

A estimação da parturição média acumulada, F_i , é feita por interpolação, empregando-se os valores de f_i e $\Phi(i)$:

$$F_i = \Phi(i - 1) + a * f_i + b * f_{i+1}$$

Onde a e b são constantes disponíveis em tabelas de interpolação para o cálculo das parturidades médias acumuladas. Uma vez obtidos os valores de P_i e F_i , calculam-se as razões destas estimativas (P_i/F_i). Idealmente, estas razões devem apresentar valores similares em cada grupo quinquenal i . Geralmente observa-se que entre as mulheres mais velhas, há vício de omissão do número de filhos tidos, diminuindo o valor da razão nos grupos de idades mais avançadas. Entretanto, pode haver diferenças nas razões até mesmo nos grupos de idades mais jovens. Na prática, tem-se como satisfatório se as razões P/F são próximas, nos grupos etários de 20-24 e 25-29 anos. Recomenda-se empregar a razão P/F do grupo de mulheres entre 20 e 24 anos como fator de correção (K) para a fecundidade corrente, porém, a razão de outro grupo etário, ou mesmo uma média entre duas ou mais razões podem ser empregadas:

$$f_i^* = K * f_i$$

Em que,

f_i^* = fecundidade corrente ajustada do grupo quinquenal i ;

K = fator de correção escolhido (P/F do grupo 20 a 24 anos, ou outra).

As estimativas do número de nascidos vivos em cada grupo etário (\widehat{NV}_i) são feitas multiplicando-se a fecundidade corrente ajustada pelo número de mulheres em cada grupo:

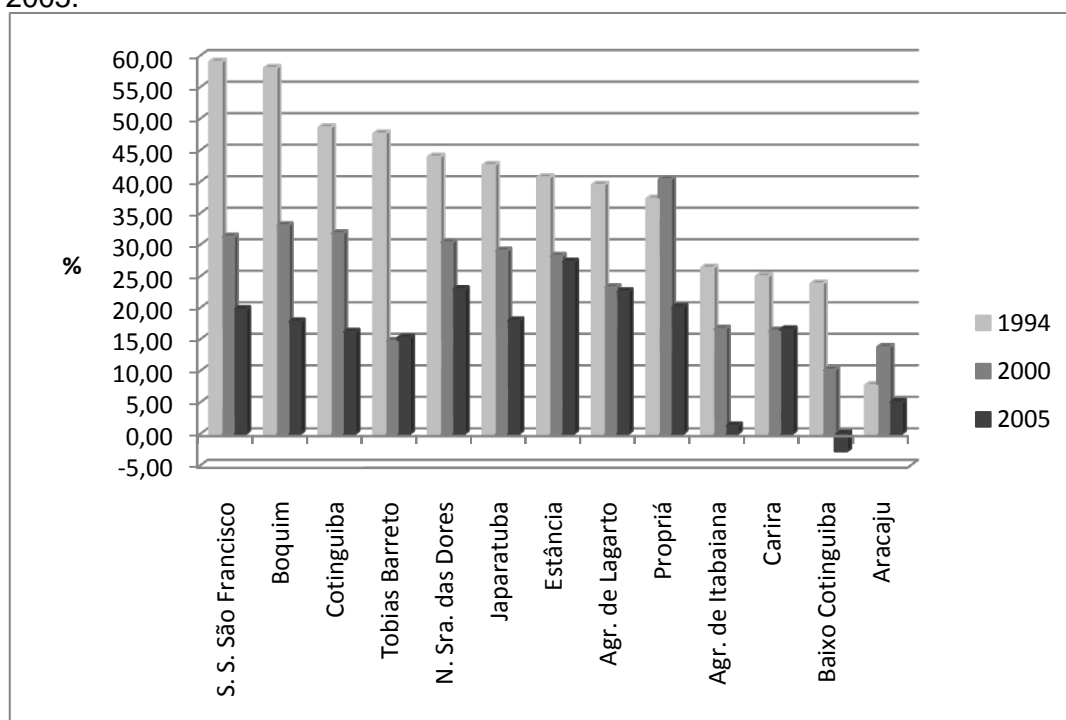
$$\widehat{NV}_i = f_i^* * M_i$$

E a soma dos nascidos vivos estimados, em cada grupo, dá a estimativa do total de nascidos vivos no período (UN, 1983).

O IBGE calcula anualmente, pelo método demográfico de Brass, o sub-registro civil de nascidos vivos, segundo a unidade da federação de residência da mãe e eventualmente é calculado o sub-registro na base do Registro Civil segundo a microrregião de residência da mãe (Figura 19).

Observa-se queda no sub-registro civil ao longo do decênio 1994-2005, em todas as microrregiões do Sergipe, indicando avanços, quer na percepção de cidadania, quer na facilidade ao acesso ao registro civil de nascimento vivo. A partir de 1997, o registro de nascimento vivo passou a ser gratuito em todo o território nacional (Brasil, 1997), o que pode ter estimulado a prática do registro civil (Figura 19).

Figura 19 - Percentual de sub-registro de Nascidos Vivos estimado pelo método de Brass, segundo a microrregião de residência da mãe. Sergipe, 1994, 2000 e 2005.



Fontes: SIMÕES CCS, 1999; e
IBGE. Censo Demográfico 2000 e PNAD 2005. Estatísticas do Registro Civil de 2000 e 2005

4.2.6. DESAGREGAÇÕES GEOGRÁFICAS

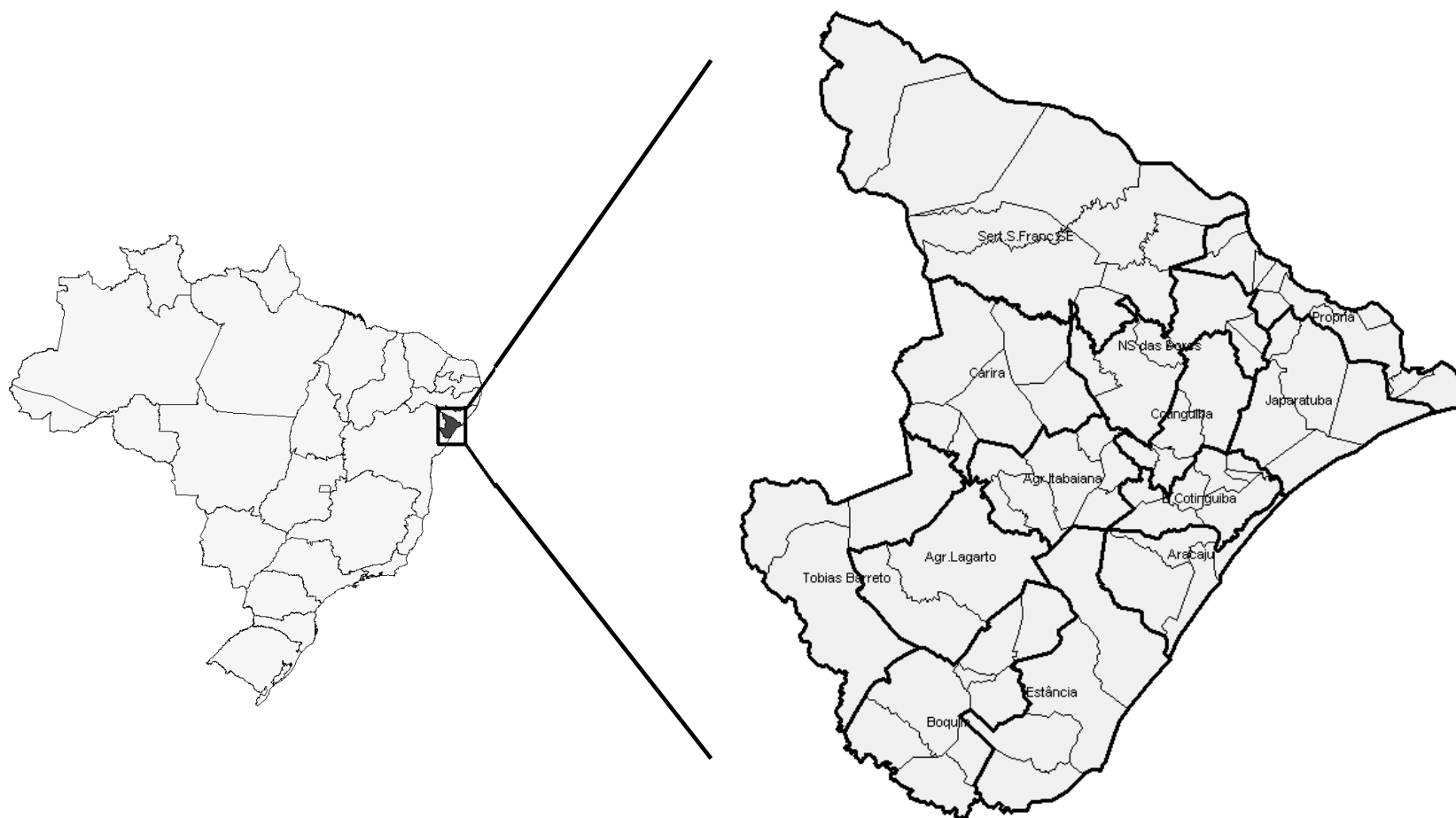
As microrregiões geográficas, definidas pelo IBGE, constituem um conjunto de municípios contíguos, agrupados segundo especificidades de organização e ocupação do espaço, de acordo com as estruturas de produção, relações sociais e elementos do quadro natural destes municípios (Guimarães, 2005). O Estado de Sergipe contabiliza treze microrregiões geográficas (Figura 20).

Na segunda edição da Norma Operacional de Assistência à Saúde – NOAS/SUS 01/2002 – a Portaria MS/GM n.º 373, que regulamentou a Norma, deliberou a região de saúde como:

“base territorial de planejamento da atenção à saúde, não necessariamente coincidente com a divisão administrativa do estado, a ser definida pela Secretaria Estadual de Saúde, de acordo com as especificidades e estratégias de regionalização da saúde em cada estado, considerando-se as características demográficas, sócio-econômicas, geográficas, sanitárias, epidemiológicas, oferta de serviços, relações entre municípios, entre outras.” (Portaria MS/GM n.º 373)

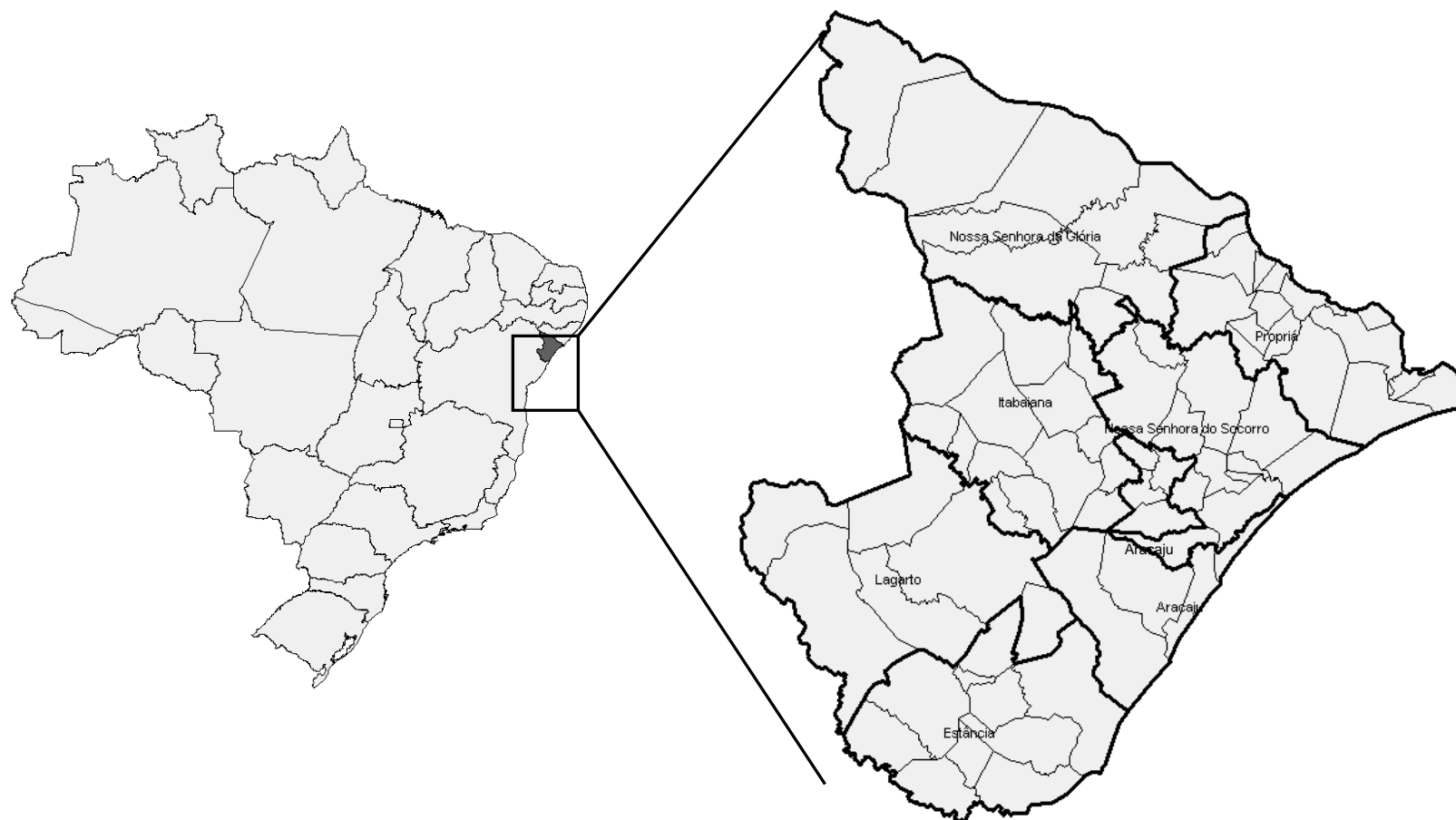
A Regional de saúde é definida como um “conjunto de municípios, entre os quais há um município-sede” (Portaria MS/GM n.º 373). As unidades da federação têm autonomia para a criação e divisão de suas regionais de saúde (Guimarães, 2005). O estado de Sergipe possui 7 regionais de saúde, que são visualizadas na Figura 21, e a relação entre as duas divisões são apresentadas na Figura 22.

Figura 20 – Divisão territorial das microrregiões do Estado de Sergipe.



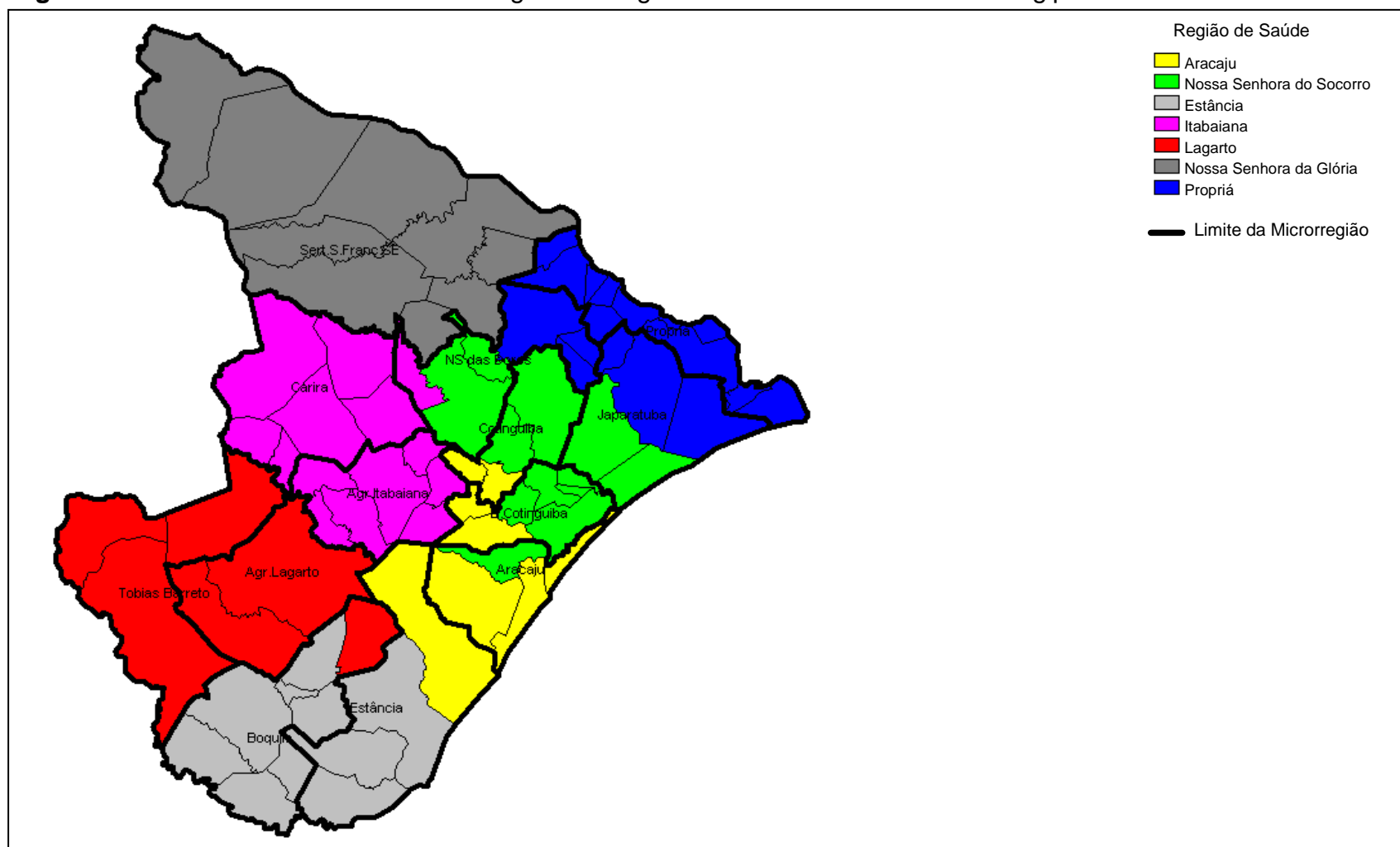
Fonte: Ministério da Saúde/DATASUS/Tabwin. Disponível em
[http://w3.datasus.gov.br/datasus/datasus.php?area=361A3B373C8D478E5F373G17H011Jd3L1M0N&VInclui=../site/din_sist.php&VSis=1&VAb=0&VCoit=478]

Figura 21 – Divisão territorial das Regiões de Saúde do Estado de Sergipe.



Fontes: Secretaria Estadual de Saúde de Sergipe e Ministério da Saúde/DATASUS/Tabwin. Disponível em [http://w3.datasus.gov.br/datasus/datasus.php?area=361A3B373C8D478E5F373G17H011Jd3L1M0N&VInclude=../site/din_sist.php&VSis=1&VAb=0&VCoit=478]

Figura 22 – Divisão territorial das Microrregiões e Regiões de Saúde do Estado de Sergipe.



Fontes: Secretaria Estadual de Saúde de Sergipe e Ministério da Saúde/DATASUS/Tabwin. Disponível em [http://w3.datasus.gov.br/datasus/datasus.php?area=361A3B373C8D478E5F373G17H011Jd3L1M0N&VInclude=../site/din_sist.php&VSis=1&VAb=0&VCoit=478]

4.2.6.1 MICRORREGIÕES

No Registro Civil, a proporção de registros com o número da DN em branco, ao longo das microrregiões, variou entre 0,41% em Senhora das Dores e quase 32% em Agreste do Lagarto. A idade da mãe é ignorada em 98 registros, ou 0,56%, sendo a microrregião Sergipana do Sertão do São Francisco a que apresentou maior percentual (2,43%). O sexo é ignorado em apenas um registro de Propriá e o local de nascimento é ignorado também em um registro de Japaratuba. No período analisado, em nenhum registro o município de nascimento nem o tipo de gestação foi ignorado (Tabela 9).

Tabela 9 – Total de registros de nascidos vivos da base do Registro Civil e registros com informações ignoradas nas variáveis número da DN, idade da mãe, sexo e local de nascimento, segundo a microrregião de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006

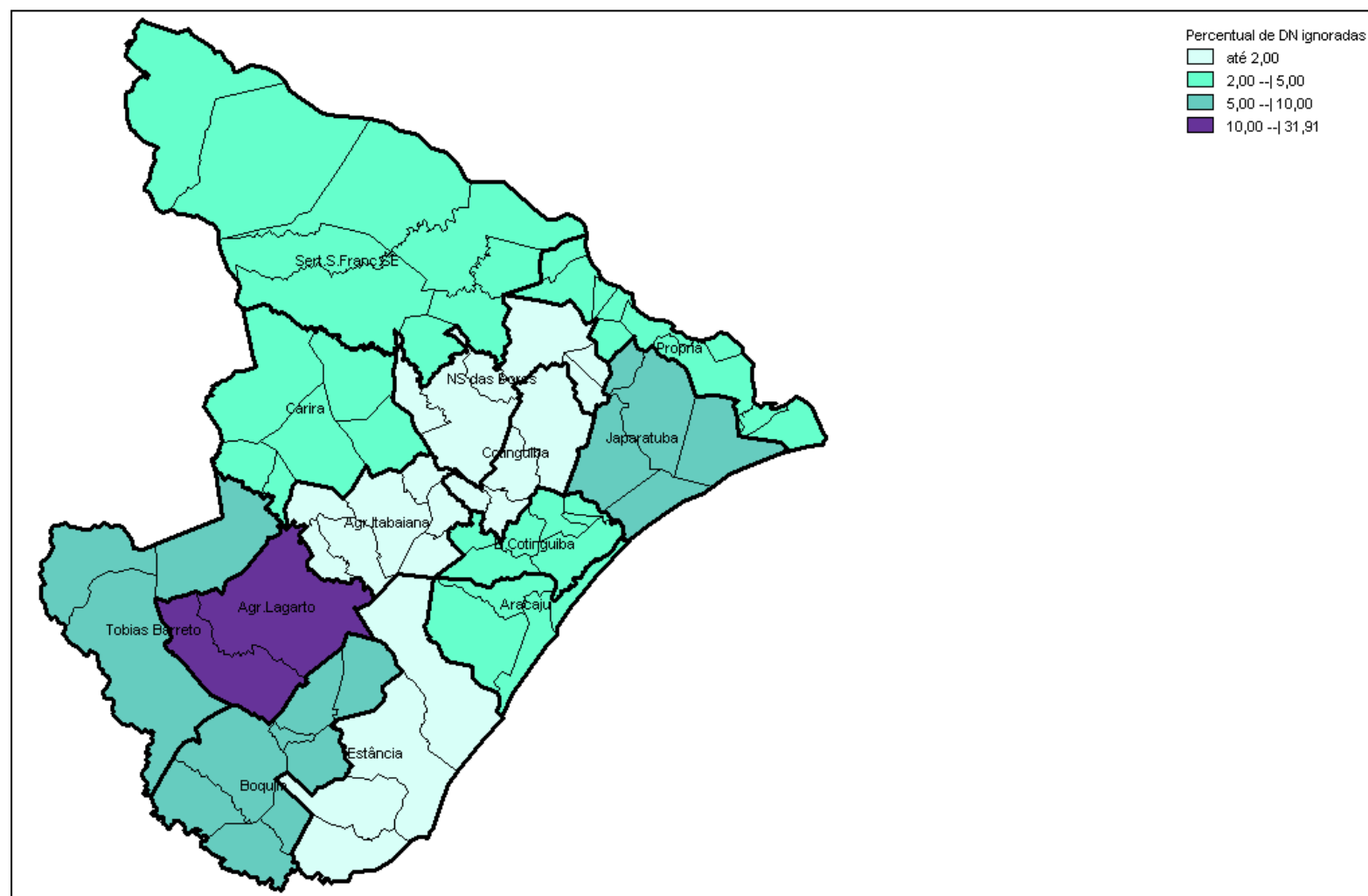
MICRORREGIÃO	TOTAL	DN IGNORADA (%)	IDADE MÃE IGNORADA (%)	SEXO IGNORADO (%)	LOC. NASC. IGNORADO (%)
S.S.Francisco-SE	1.359	46 (3,38)	33 (2,43)	-	-
Carira	479	19 (3,97)	1 (0,21)	-	-
N.Sra. Dores	493	2 (0,41)	-	-	-
Agr. Itabaiana	1.399	23 (1,64)	-	-	-
Tobias Barreto	833	77 (9,24)	3 (0,36)	-	-
Agr. Lagarto	868	277 (31,91)	2 (0,23)	-	-
Propriá	821	26 (3,17)	1 (0,12)	1 (0,12)	-
Cotinguiba	448	3 (0,67)	-	-	-
Japaratuba	493	39 (7,91)	-	-	1 (0,20)
Baixo Cotinguiba	901	31 (3,44)	1 (0,11)	-	-
Aracaju	6.590	138 (2,09)	43 (0,65)	-	-
Boquim	1.434	117 (8,16)	9 (0,63)	-	-
Estância	1.136	10 (0,88)	3 (0,26)	-	-
Munic. Res. ignorado	2	-	-	-	-
Outra UF ou UF desconhecida*	88	-	2 (2,27)	-	-
TOTAL	17.344	808 (4,66)	98 (0,56)	1 (0,01)	1 (0,01)

*Registros inicialmente não pareados com o SINASC e identificados pelo número da DN em outras unidades da federação.

Não houve registros com município de nascimento ou tipo de gestação ignorados.

Fonte: IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

Figura 23 – Percentual de registros com número da DN ignorado na Base do registro Civil, segundo a microrregião de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.



Fonte: IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

Há microrregiões onde se deve ter mais atenção na análise do total de nascidos vivos estimados, uma vez que são as microrregiões onde há o maior percentual de registros com número da DN em branco, acarretando, assim, em um pareamento menos fidedigno da realidade (Figura 23).

Na base do SINASC, não há registros sem o número da DN, e o número e percentual de registros com informações faltantes nas demais variáveis são apresentados na Tabela 10. Esta base mostra-se com excelente completude, havendo registros com informações faltantes somente no tipo de gestação e local de nascimento.

Tabela 10 – Total de registros de nascidos vivos da base do SINASC e registros com informações ignoradas nas variáveis tipo de gestação e local de nascimento, segundo a microrregião de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006

MICRORREGIÃO	TOTAL	TIPO DE GESTAÇÃO IGNORADA (%)	LOC. NASC. IGNORADO (%)
S.S Francisco-SE	1.643	19 (1,16)	-
Carira	571	2 (0,35)	-
Nossa Senhora das Dores	564	4 (0,71)	-
Agreste de Itabaiana	1.476	5 (0,34)	-
Tobias Barreto	955	1 (0,10)	-
Agreste de Lagarto	1.091	1 (0,09)	-
Propriá	823	-	1 (0,12)
Cotinguiba	468	-	-
Japaratuba	541	-	-
Baixo Cotinguiba	947	-	-
Aracaju	7.430	1 (0,01)	-
Boquim	1.617	1 (0,06)	-
Estância	1.334	3 (0,22)	-
Outras UFs*	133	-	-
TOTAL	19.593	37 (0,19)	1 (0,01)

*Registros inicialmente não pareados com o Registro Civil e identificados pelo número da DN em outras unidades da federação.

Não houve registros onde a idade da mãe, o município de nascimento ou o sexo do recém nascido eram ignorados.

Fonte: MS. **Sistema de Informação sobre Nascidos Vivos 2006**. Dados Preliminares. Brasília, 2006, CD

Em relação à qualidade do pareamento, a diferença entre os pares encontrados, de acordo com o critério de relacionamento proposto, foi menor na microrregião de Estância (10,77%), ao passo que a maior foi no Agreste de Lagarto (36,33%).

Tabela 11 – Número de pares relacionados pelos pareamentos menos e mais restritos e diferença percentual entre os pares encontrados, segundo a microrregião de residência da mãe constante na base do Registro Civil, Sergipe, segundo e terceiro trimestres de 2006

MICRORREGIÃO	MENOS RESTRITO	MAIS RESTRITO	DIFERENÇA %
S.S Francisco-SE	1.210	958	20,83
Carira	445	371	16,63
Nossa Senhora das Dores	464	380	18,10
Agreste de Itabaiana	1.296	1.135	12,42
Tobias Barreto	709	546	22,99
Agreste de Lagarto	545	347	36,33
Propriá	711	510	28,27
Cotinguiba	406	325	19,95
Japaratusba	424	319	24,76
Baixo Cotinguiba	826	704	14,77
Aracaju	6.183	5.473	11,48
Boquim	1.245	1.073	13,82
Estância	1.068	953	10,77
UF de residência desconhecida*	15	0	100,00
Outras UFs*	65	49	24,62
TOTAL	15.612	13.143	15,81

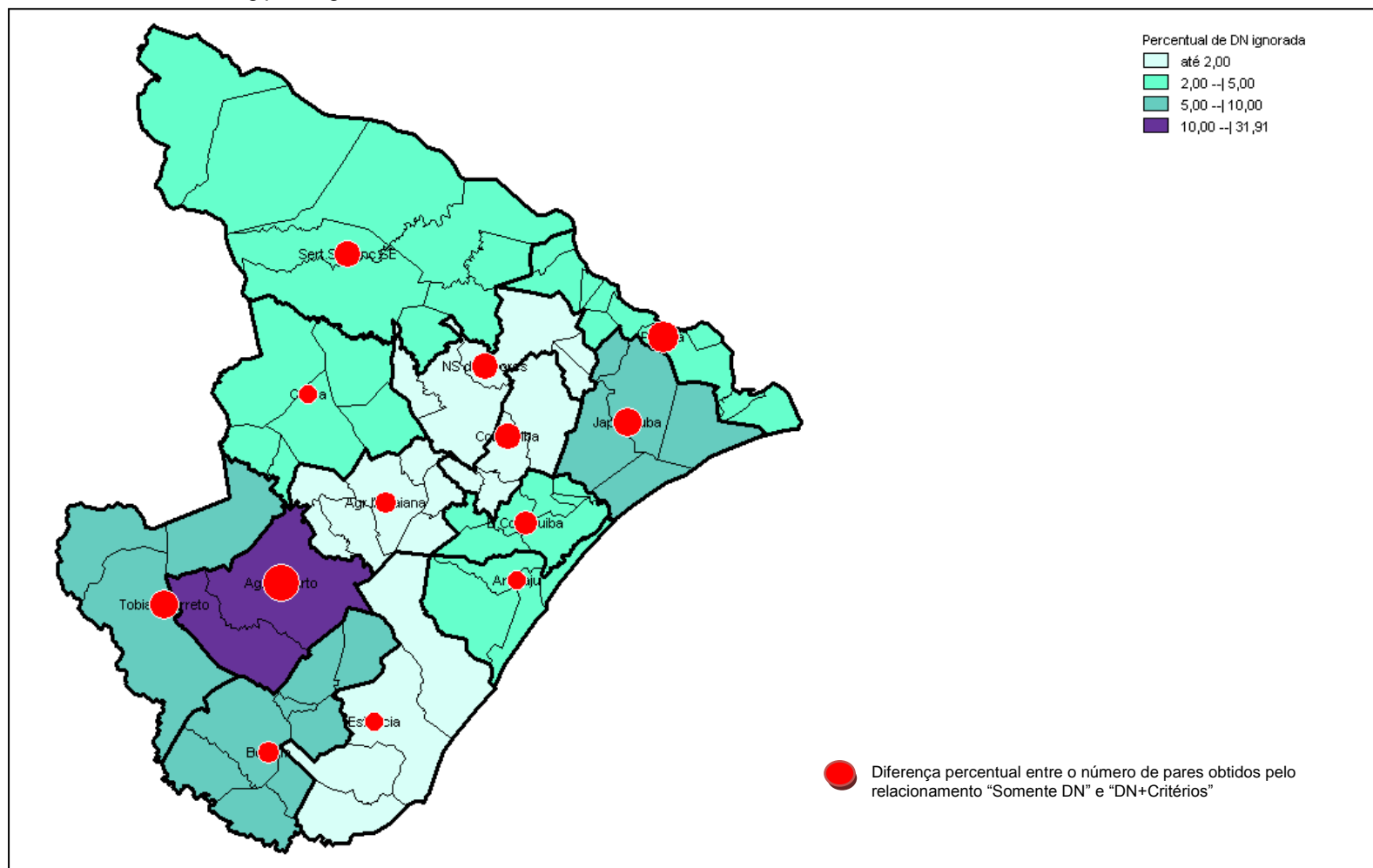
*Registros inicialmente não pareados com o SINASC e identificados pelo número da DN em outras unidades da federação.

Fontes: MS. **Sistema de Informação sobre Nascidos Vivos 2006**. Dados Preliminares. Brasília, 2006, CD; e IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

Percebe-se que o Agreste de Lagarto teve o maior percentual de registros sem número da DN preenchido e também é a microrregião onde houve maior diferença no número de pares encontrados, de acordo com o critério de relacionamento utilizado (Tabelas 9 e 11). Entretanto, este padrão não se repete nas demais microrregiões (Figura 24). Naquelas onde houve menor número de registros com número da DN em branco, na base do

Registro Civil, menos do que 5%, a diferença entre os pares não necessariamente é menor.

Figura 24 – Percentual de registros com número da DN ignorado na Base do registro Civil, e diferença percentual do número de pares obtidos a partir de diferentes métodos de relacionamento entre as bases SINASC e Registro Civil, segundo a microrregião de residência da mãe, Sergipe, segundo o primeiro e terceiro trimestres de 2006.



Fontes: MS. **Sistema de Informação sobre Nascidos Vivos 2006**. Dados Preliminares. Brasília, 2006, CD; e IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

4.2.6.1 REGIONAIS DE SAÚDE

Nesta divisão territorial, a proporção de ocorrências do Registro Civil sem o número da DN preenchido é melhor distribuída, tendo somente a regional de Lagarto com proporção elevada (Tabela 12 e Figura 25). Em relação à idade da mãe, somente a regional de Nossa Senhora da Glória apresenta mais de 1% dos registros com esta variável ignorada.

Tabela 12 – Total de registros de nascidos vivos da base do Registro Civil e registros com informações ignoradas nas variáveis número da DN, idade da mãe, sexo e local de nascimento, segundo a região de saúde de residência da mãe, Sergipe, segundo o primeiro e terceiro trimestres de 2006

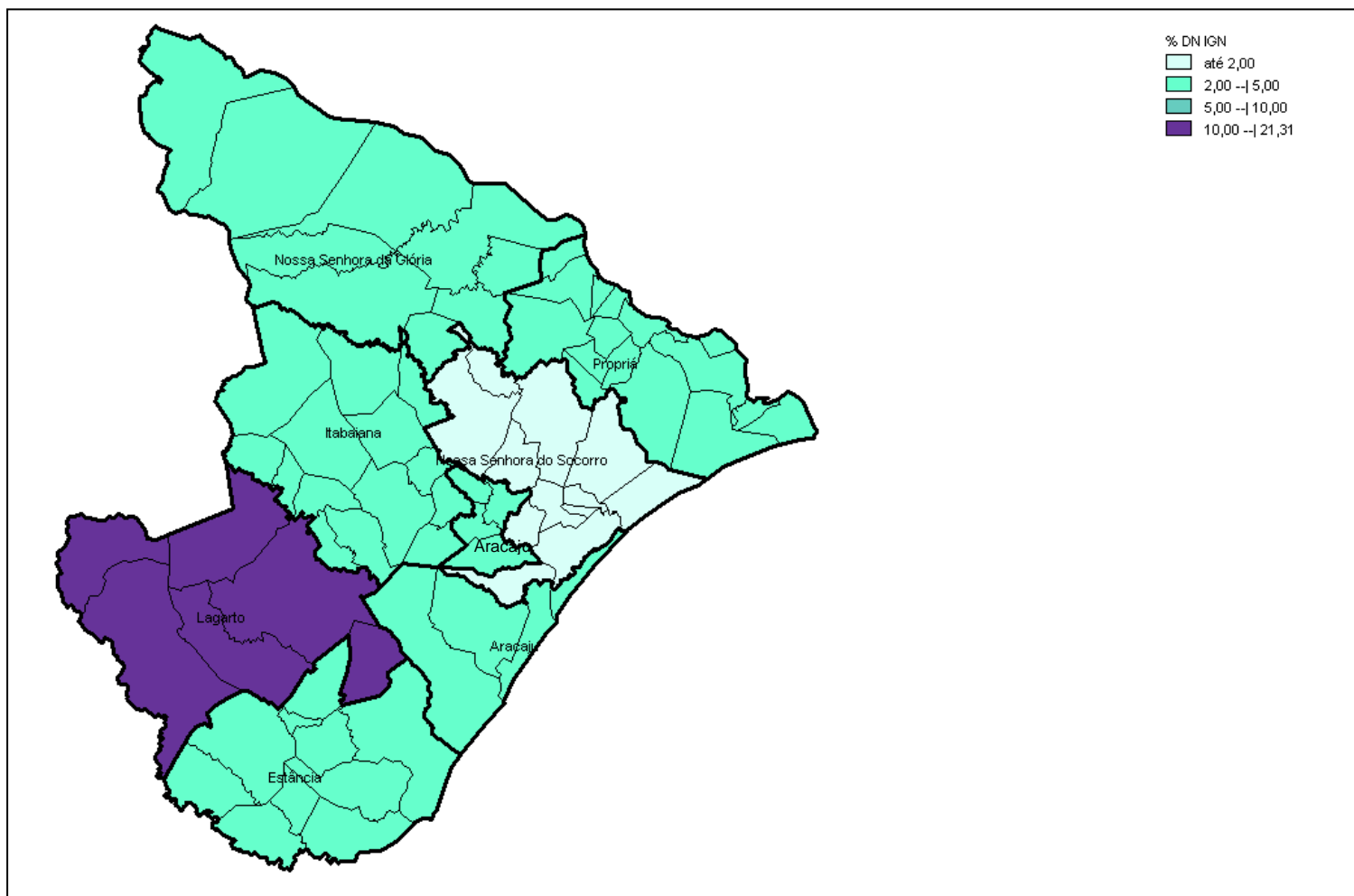
REGIONAL DE SAÚDE	TOTAL	DN IGNORADA (%)	IDADE MÃE IGNORADA (%)	SEXO IGNORADO (%)	LOC. NASC. IGNORADO (%)
Aracaju	6.332	165 (2,61)	37 (0,58)	-	-
Estância	2.088	65 (3,11)	6 (0,29)	-	-
Itabaiana	1.818	40 (2,20)	1 (0,06)	-	-
Lagarto	1.952	416 (21,31)	10 (0,51)	-	-
N.Sra. Glória	1.359	46 (3,38)	33 (2,43)	-	-
N.Sra. Socorro	2.369	14 (0,59)	8 (0,34)	-	-
Propriá	1.336	62 (4,64)	1 (0,08)	1 (0,08)	1 (0,08)
Mun. res. ignorado	2	-	-	-	-
Outra UF ou UF desconhecida*	88	-	2 (2,27)	-	-
TOTAL	17.344	808 (4,66)	98 (0,57)	1 (0,01)	1 (0,01)

*Registros inicialmente não pareados com o SINASC e identificados pelo número da DN em outras unidades da federação.

Não houve registros onde o município de nascimento ou o tipo de gestação eram ignorados.

Fonte: IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

Figura 25 – Percentual de registros com número da DN ignorado na Base do registro Civil, segundo a regional de saúde de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.



Fonte: IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

Na base do SINASC, como já visto anteriormente, somente o tipo de gestação e local de nascimento apresentam dados ignorados, em poucos registros. O tipo de gestação é ignorado em mais de 1% dos registros do SINASC somente na regional de Nossa Senhora da Glória (Tabela 13).

Tabela 13 – Total de registros de nascidos vivos da base do SINASC e registros com informações ignoradas nas variáveis tipo de gestação e local de nascimento, segundo a região de saúde de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006

REGIONAL DE SAÚDE	TOTAL	TIPO DE GESTAÇÃO IGNORADA (%)	LOCAL DE NASCIMENTO IGNORADO (%)
Aracaju	6.615	4 (0,06)	-
Estância	2.448	1 (0,04)	-
Itabaiana	1.972	5 (0,25)	-
Lagarto	2.319	5 (0,22)	-
Nossa Senhora da Glória	1.643	19 (1,16)	-
Nossa Senhora do Socorro	3.060	3 (0,10)	-
Propriá	1.403	-	1 (0,07)
Outra UF*	133	-	-
TOTAL	19.593	37 (0,19)	1 (0,01)

*Registros inicialmente não pareados com o Registro Civil e identificados pelo número da DN em outras unidades da federação.

Não houve registros onde a idade da mãe, o sexo da criança, ou o município de nascimento eram ignorados.

Fonte: MS. **Sistema de Informação sobre Nascidos Vivos 2006**. Dados Preliminares. Brasília, 2006, CD

Como nas microrregiões, na divisão territorial das regionais de saúde, a regional de Lagarto, onde há a maior proporção de registros com número da DN em branco no Registro Civil (Tabela 13), é a regional onde a diferença no número de pares é grande, a segunda maior (Tabela 14). Por outro lado, em Propriá, a diferença entre os pares, de acordo com o critério de pareamento adotado, é a maior (Tabela 14), mas esta regional de saúde não apresenta grande proporção de registros sem número da DN no Registro Civil (Tabela 12).

Tabela 14 – Número de pares relacionados pelos pareamentos menos e mais restritos, segundo a regional de saúde de residência da mãe constante na base do Registro Civil, Sergipe, segundo e terceiro trimestres de 2006

REGIONAL DE SAÚDE	MENOS RESTRITO	MAIS RESTRITO	DIFERENÇA %
Aracaju	5.908	5207	11,87
Estância	1.910	1664	12,88
Itabaiana	1.683	1451	13,78
Lagarto	1.433	1044	27,15
Nossa Senhora da Glória	1.210	958	20,83
Nossa Senhora do Socorro	2.226	1928	13,39
Propriá	1.162	842	27,54
UF de residência desconhecida*	15	0	100,00
Outras UFs*	65	49	24,62
TOTAL	15.612	13.143	15,81

*Registros inicialmente não pareados com o SINASC e identificados pelo número da DN em outras unidades da federação.

Fontes: MS. **Sistema de Informação sobre Nascidos Vivos 2006.**

Dados Preliminares. Brasília, 2006, CD; e

IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

4.2.7. BANCOS UTILIZADOS PARA A ESTIMAÇÃO

Como visto nas seções anteriores, há registros que foram pareados onde há discordância entre algumas das variáveis comuns aos dois bancos. Como padronização de qual informação utilizar, optou-se, em caso de divergência no pareamento, pelo município de residência da mãe que consta na base do Registro Civil. Já para o sexo do nascido vivo e a idade da mãe, em caso de divergência, foram considerados os da base do SINASC.

Durante o processo de pareamento, alguns registros foram identificados, em ambas as bases, como ocorrências de nascimento vivo de mães residentes fora de Sergipe. Estes registros foram excluídos, seguindo o critério de se utilizar o município de residência da mãe do Registro Civil. No total, 90 registros do Registro Civil foram identificados com a residência da mãe fora do estado de Sergipe ou município ignorado. Assim, o banco do Registro Civil conta com 17.254 nascimentos vivos. Cabe ressaltar que esta

escolha arbitrária modificou o número de registros do SINASC, para cada critério de pareamento adotado. Esta mudança se deve ao fato de, quando o pareamento é o menos restrito, há 15.532 pares (Tabela 15), cujo município de residência da mãe considerado foi o do Registro Civil. Dos 4.061 (19.593 – 15.532) nascidos vivos capturados somente pelo SINASC, 91 são de mães residentes fora de Sergipe, e por esta razão, foram excluídos. Quando o pareamento é pelo critério mais restrito, houve 13.094 pares (Tabela 15) e, dos 6.499 (19.593 – 13.094) nascidos vivos capturados somente pelo SINASC, 118 referem-se a mães residentes fora de Sergipe.

Com estes novos bancos, a distribuição nas caselas da tabela 2 x 2 é (Tabela 15):

Tabela 15 – Número final de nascidos vivos capturados pelo SINASC e Registro Civil (RC), no segundo e terceiro trimestres de 2006 em Sergipe, segundo o critério de pareamento adotado.

Pareamento menos restrito				
		RC		
		Sim	Não	
SINASC	Sim	15.532	3.970	19.502
	Não	1.722		
		17.254		
<i>TOTAL DE NASCIDOS VIVOS CAPTURADOS: 21.224</i>				
Pareamento mais restrito				
		RC		
		Sim	Não	
SINASC	Sim	13.094	6.381	19.475
	Não	4.160		
		17.254		
<i>TOTAL DE NASCIDOS VIVOS CAPTURADOS: 23.635</i>				

Fontes: MS. **Sistema de Informação sobre Nascidos Vivos 2006**. Dados Preliminares. Brasília, 2006, CD; e IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007.

O pareamento pelo número da DN foi prejudicado pela falta de preenchimento desta variável, no Registro Civil, em mais de 4% dos registros, no período considerado para a análise. Entretanto, ao se realizar a desagregação geográfica, tanto por microrregiões, quanto por regionais de saúde, esta proporção pode ser desprezível em algumas microrregiões e regionais de saúde, como as microrregiões de Nossa Senhora das Dores, Estância e Contiguiba, e a regional de saúde de Nossa Senhora do Socorro, com menos de 1% dos registros sem o número da DN preenchida. Por outro lado, a microrregião de Agreste de Lagarto e a regional de saúde de Lagarto merecem cuidado na análise dos resultados das estimações vindouras, por apresentarem proporções muito altas de registros sem o número da DN, o que afeta a qualidade do pareamento.

Pode-se questionar se a falta do preenchimento da variável número da DN é realmente variável incompleta, ou trata-se de nascimentos vivos ocorridos e sem a emissão da respectiva Declaração por parte do estabelecimento de saúde. É possível afirmar que a causa da incompletude desta variável é mesmo o não preenchimento do número da DN nos questionários do IBGE pelos cartórios, uma vez que o número de nascimentos vivos capturados pelo SINASC é maior do que o Registro Civil, e o campo para o número da DN foi incluído no questionário do IBGE em 2006. O preenchimento deste campo pelos cartórios, no questionário do IBGE, tornou-se mais freqüente ao longo dos trimestres de coleta, não somente em Sergipe, mas em todas as unidades da federação.

Dentre os registros pareados pelo número da DN, a consistência entre as duas bases com este pareamento mostrou-se regular, onde em torno de 75% dos pares coincidiram em todas as demais variáveis comparáveis. Ao se levar em consideração o possível vício de informação do endereço de residência da mãe, e desconsiderar o município de residência como uma variável-requisito para o pareamento ser considerado correto, o número de pares considerados “reais” aumenta para quase 85% do pares, o que ainda não é excelente.

A dificuldade em se realizar o pareamento probabilístico, frente ao

baixo poder discriminatório das variáveis disponíveis, forçou a empregar apenas os dois critérios de relacionamento determinísticos aqui descritos.

Por fim, apesar do problema da falta de preenchimento do número da DN no Registro Civil, e da qualidade do pareamento poder ser considerada apenas regular, as duas bases, de uma maneira geral, apresentam boa completude de seus registros e o uso do método de captura e recaptura para estimar o número total de nascidos vivos é viável com os dados disponíveis.

4.2.8. OS MODELOS DE ESTIMAÇÃO

Descreve-se nesta seção o procedimento utilizado para a construção dos modelos lineares aplicados na estimação, pelo método de captura e recaptura, no programa MARK[®], para estimar o total de nascidos vivos de Sergipe, no segundo e terceiro trimestres de 2006. Foram avaliadas as variáveis que podem afetar as probabilidades de captura de nascidos vivos pelo Registro Civil e SINASC, de maneira a incluí-las no modelo completo de estimação.

Vários fatores podem ser determinantes para o sub-registro de nascidos vivos e algumas características poderiam ser utilizadas no modelo completo, como a escolaridade da mãe, por exemplo. Entretanto, para o emprego deste método, é necessário que as co-variáveis individuais incluídas no modelo linear estejam disponíveis em todas as ocasiões de captura (Cooch & White, 2008), que no caso do presente trabalho, são as duas fontes de dados. Como visto na seção de Limpeza dos bancos e Relacionamento dos registros, somente o sexo do nascido vivo e a idade da mãe estão disponíveis nas duas bases. Fatores institucionais, como inexistência de cartório de registro de pessoas naturais no município também pode dificultar o registro civil do nascido vivo. Em Sergipe, somente dois municípios não possuíam este tipo de cartório, em 2006 (IBGE, 2006b). Desta maneira, este fator não foi considerado, uma vez que os níveis de

desagregação geográfica aplicados foram microrregiões e regionais de saúde, e nestes níveis, os cartórios se fazem presentes em todas as subdivisões geográficas.

Na seção de Limpeza dos bancos e Relacionamento dos registros, viu-se que o Banco do Registro Civil utilizado tem 17.254 registros, referentes a nascidos vivos de mães residentes em Sergipe, no segundo e terceiro trimestres de 2006, e que dependendo do critério de relacionamento adotado, o número de registros do SINASC foi 19.502 (menos restrito) ou 19.475 (mais restrito). Para a verificação da potencialidade da inclusão do sexo do recém-nascido e da idade da mãe no modelo completo, utilizou-se a base do SINASC com 19.593 registros.

Para o sexo do nascido vivo, testou-se a hipótese nula de que a proporção de meninas é a mesma na microrregião/regional de saúde avaliada, em relação à proporção do estado como um todo. Já para a idade da mãe, a hipótese nula é de que a média de idade na microrregião/regional de saúde é a mesma de Sergipe.

No Registro Civil, dentre os nascidos vivos de mães residentes em Sergipe, no segundo e terceiro trimestres de 2006, 96 registros tinham a idade da mãe ignorada e a estes foi atribuído o valor médio da idade da mãe, 25,16 anos.

A proporção de nascidos vivos do sexo feminino, em Sergipe, no período analisado, foi de 49%, nas duas bases. Não há diferença significativa na proporção de meninas capturadas tanto pelo Registro Civil, quanto pelo SINASC, em ambas as desagregações geográficas (Tabelas 16, 17, 18 e 19).

A idade média das mães foi de 25,11 anos no SINASC, sendo a idade mínima de 10 anos, e a máxima, 54 anos. No Registro Civil, a média de idade da mãe foi de 25,16 anos, variando entre 12 e 52 anos. Em relação à idade da mãe, percebe-se que há microrregiões onde a média de idade difere estatisticamente entre a microrregião em questão a média total. No SINASC, não há diferença entre a média de idade nas microrregiões de Carira, Nossa Senhora das Dores, Tobias Barreto, Agreste de Lagarto,

Contiguiba, Boquim e Estância, em relação à idade média total (Tabela 16). No Registro Civil, também não há diferença significativa em Carira, Nossa Senhora das Dores, Tobias Barreto, Agreste de Lagarto, Contiguiba, Japaratuba e Estância (Tabela 17).

Quando se agregam os registros segundo a regional de saúde de residência da mãe, a idade média difere da média total em praticamente em todas as regionais, com exceção de Lagarto, tanto no SINASC quanto no Registro Civil (Tabelas 18 e 19).

Frente a estes resultados, pode-se concluir que o sexo do nascido vivo não influencia na captura pelas duas bases de dados. Já a idade da mãe, por mostrar média estatisticamente diferente em algumas microrregiões e regionais de saúde, foi incluída no modelo final, cujos parâmetros e matrizes de desenhos são apresentados na subseção a seguir.

Tabela 16 – Total de nascidos vivos capturados pelo SINASC, proporção de nascidos vivos do sexo feminino, estatísticas dos testes de diferença de proporções, média de idade das mães e estatísticas dos testes de diferença de médias, segundo a microrregião de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.

MICRORREGIÃO	TOTAL DE REGISTROS	%	SEXO FEMININO			IDADE DA MÃE		
			PROPORÇÃO (IC95%)	H0: A PROPORÇÃO DE SEXO FEMININO NA MICRORREGIÃO É IGUAL À DA POPULAÇÃO		MÉDIA (IC95%)	H0: A IDADE MÉDIA DA MÃE NA MICRORREGIÃO É IGUAL À DA POPULAÇÃO	
				z	P(Z > z)		t	P(T > t)
Sertão do São Francisco- SE	1.643	8,44	0,49 (0,47 - 0,52)	0,34	0,732	24,35 (24,03 - 24,67)	-4,66	0,000
Carira	571	2,93	0,52 (0,48 - 0,56)	1,41	0,158	25,57 (25,03 - 26,11)	1,68	0,093
Nossa Senhora das Dores	564	2,90	0,48 (0,43 - 0,52)	-0,73	0,464	24,78 (24,23 - 25,33)	-1,16	0,248
Agreste de Itabaiana	1.476	7,58	0,50 (0,47 - 0,52)	0,46	0,644	25,52 (25,19 - 25,86)	2,44	0,015
Tobias Barreto	955	4,91	0,50 (0,47 - 0,53)	0,55	0,583	25,14 (24,71 - 25,57)	0,17	0,867
Agreste de Lagarto	1.091	5,61	0,49 (0,46 - 0,52)	0,11	0,915	25,46 (25,06 - 25,86)	1,72	0,085
Propriá	823	4,23	0,48 (0,45 - 0,52)	-0,47	0,637	23,81 (23,38 - 24,25)	-5,79	0,000
Cotinguiba	468	2,40	0,47 (0,42 - 0,51)	-0,98	0,327	24,68 (24,08 - 25,27)	-1,41	0,150
Japaratuba	541	2,78	0,50 (0,46 - 0,54)	0,57	0,571	24,20 (23,65 - 24,75)	-3,26	0,001
Baixo Cotinguiba	947	4,87	0,49 (0,46 - 0,53)	0,22	0,825	24,27 (23,88 - 24,67)	-4,12	0,000
Aracaju	7.430	38,18	0,49 (0,48 - 0,50)	-0,24	0,814	25,56 (25,42 - 25,71)	6,14	0,000
Boquim	1.617	8,31	0,50 (0,48 - 0,52)	0,78	0,435	24,86 (24,54 - 25,18)	-1,49	0,137
Estância	1.334	6,86	0,47 (0,44 - 0,54)	-1,78	0,076	24,85 (24,50 - 25,20)	-1,42	0,157
<i>TOTAL</i>	<i>19.460*</i>	<i>100,00</i>	<i>0,49 (0,48 - 0,50)</i>	<i>-</i>	<i>-</i>	<i>25,11 (25,01 - 25,20)</i>	<i>-</i>	<i>-</i>

* Dos 19.593 registros de nascidos vivos capturados, em 133 a UF de residência informada não é Sergipe
 Fonte: MS. **Sistema de Informação sobre Nascidos Vivos 2006**. Dados Preliminares. Brasília, 2006, CD

Tabela 17 – Total de nascidos vivos capturados pelo Registro Civil, proporção de nascidos vivos do sexo feminino, estatísticas dos testes de diferença de proporções, média de idade das mães e estatísticas dos testes de diferença de médias, segundo a microregião de residência da mãe, Sergipe, segundo o terceiro trimestre de 2006.

MICRORREGIÃO	TOTAL DE REGISTROS	%	SEXO FEMININO			IDADE DA MÃE		
			PROPORÇÃO (IC95%)	H0: A PROPORÇÃO DE SEXO FEMININO NA MICRORREGIÃO É IGUAL À DA POPULAÇÃO		MÉDIA (IC95%)	H0: A IDADE MÉDIA DA MÃE NA MICRORREGIÃO É IGUAL À DA POPULAÇÃO	
				z	P(Z > z)		t	P(T > t)
Sertão do São Francisco-SE	1.359	7,88	0,48 (0,46 - 0,51)	0,06	0,950	24,37 (24,03 - 24,71)	-4,56	0,000
Carira	479	2,78	0,51 (0,46 - 0,55)	0,86	0,392	25,58 (24,99 - 26,16)	1,42	0,157
Nossa Senhora das Dores	493	2,86	0,49 (0,45 - 0,53)	0,23	0,819	24,99 (24,40 - 25,58)	-0,56	0,577
Agreste de Itabaiana	1.399	8,11	0,50 (0,48 - 0,54)	1,27	0,206	25,42 (25,08 - 25,77)	1,52	0,128
Tobias Barreto	833	4,83	0,50 (0,47 - 0,54)	0,88	0,378	25,44 (24,98 - 25,90)	1,20	0,230
Agreste de Lagarto	868	5,03	0,50 (0,46 - 0,53)	0,65	0,513	25,31 (24,87 - 25,76)	0,71	0,480
Propriá	821	4,76	0,48 (0,45 - 0,51)	-0,42	0,678	24,14 (23,70 - 24,57)	-4,58	0,000
Cotinguiba	448	2,60	0,46 (0,41 - 0,51)	-1,18	0,237	24,79 (24,19 - 25,38)	-1,22	0,225
Japaratuba	493	2,86	0,50 (0,45 - 0,54)	0,32	0,750	24,44 (23,85 - 25,03)	-2,40	0,017
Baixo Cotinguiba	901	5,22	0,48 (0,45 - 0,51)	-0,36	0,716	24,23 (23,82 - 24,63)	-4,49	0,000
Aracaju	6.590	38,19	0,49 (0,47 - 0,50)	-0,72	0,472	25,71 (25,06 - 25,86)	-7,09	0,000
Boquim	1.434	8,31	0,49 (0,47 - 0,52)	0,45	0,650	24,57 (24,25 - 24,89)	-3,56	0,000
Estância	1.136	6,58	0,47 (0,44 - 0,50)	-1,07	0,283	24,78 (24,41 - 25,15)	-1,99	0,046
<i>TOTAL</i>	17.254	100,00	0,49 (0,48 - 0,50)	-	-	25,16 (25,06 - 25,25)	-	-

Fonte: IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007

Tabela 18 – Total de nascidos vivos capturados pelo SINASC, proporção de nascidos vivos do sexo feminino, estatísticas dos testes de diferença de proporções, média de idade das mães e estatísticas dos testes de diferença de médias, segundo a regional de saúde de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.

REGIONAL DE SAÚDE	TOTAL DE REGISTROS	%	SEXO FEMININO			IDADE DA MÃE		
			PROPORÇÃO (IC95%)	H0: A PROPORÇÃO DE SEXO FEMININO NA MICRORREGIÃO É IGUAL À DA POPULAÇÃO		MÉDIA (IC95%)	H0: A IDADE MÉDIA DA MÃE NA MICRORREGIÃO É IGUAL À DA POPULAÇÃO	
				z	P(Z > z)		t	P(T > t)
Aracaju	6.615	33,99	0,49 (0,48 - 0,50)	0,07	0,945	25,60 (25,45 - 25,76)	6,30	0,000
N.Sra. Socorro	3.060	15,72	0,48 (0,47 - 0,50)	-0,80	0,422	24,79 (24,56 - 25,02)	-2,73	0,006
Estância	2.448	12,58	0,49 (0,47 - 0,51)	-0,40	0,687	24,82 (24,56 - 25,08)	-2,14	0,033
Itabaiana	1.972	10,13	0,50 (0,58 - 0,52)	0,70	0,484	25,51 (25,22 - 25,80)	2,75	0,006
Lagarto	2.319	11,92	0,50 (0,48 - 0,52)	1,01	0,313	25,34 (25,01 - 25,61)	1,67	0,095
N.Sra. Glória	1.643	8,44	0,49 (0,47 - 0,52)	0,34	0,732	24,35 (24,03 - 24,67)	-4,66	0,000
Propriá	1.403	7,21	0,48 (0,45 - 0,50)	-0,92	0,355	23,88 (23,54 - 24,21)	-7,18	0,000
<i>TOTAL</i>	<i>19.460*</i>	100,00	<i>0,49 (0,48 - 0,50)</i>	-	-	<i>25,11 (25,01 - 25,20)</i>	-	-

* Dos 19.593 registros de nascidos vivos capturados, em 133 a UF de residência informada não é Sergipe

Fonte: MS. **Sistema de Informação sobre Nascidos Vivos 2006**. Dados Preliminares. Brasília, 2006, CD

Tabela 19 – Total de nascidos vivos capturados pelo Registro Civil, proporção de nascidos vivos do sexo feminino, estatísticas dos testes de diferença de proporções, média de idade das mães e estatísticas dos testes de diferença de médias, segundo a regional de saúde de residência da mãe, Sergipe, segundo e terceiro trimestres de 2006.

REGIONAL DE SAÚDE	TOTAL DE REGISTROS	%	SEXO FEMININO			IDADE DA MÃE		
			PROPORÇÃO (IC95%)	H0: A PROPORÇÃO DE SEXO FEMININO NA MICRORREGIÃO É IGUAL À DA POPULAÇÃO		MÉDIA (IC95%)	H0: A IDADE MÉDIA DA MÃE NA MICRORREGIÃO É IGUAL À DA POPULAÇÃO	
				z	P(Z > z)		t	P(T > t)
Aracaju	6.332	36,70	0,48 (0,47 - 0,49)	-0,84	0,401	25,65 (25,49 - 25,81)	6,13	0,000
Nossa Senhora do Socorro	2.369	13,73	0,48 (0,46 - 0,50)	-0,39	0,697	24,86 (24,61 - 25,12)	-2,23	0,026
Estância	2.088	12,10	0,48 (0,46 - 0,50)	-,037	0,713	24,61 (25,06 - 25,25)	-4,00	0,000
Itabaiana	1.818	10,54	0,50 (0,48 - 0,53)	1,33	0,184	25,45 (25,15 - 25,75)	1,91	0,056
Lagarto	1.952	11,31	0,50 (0,48 - 0,52)	1,36	0,175	25,39 (25,09 - 25,68)	1,53	0,125
Nossa Senhora da Glória	1.359	7,88	0,48 (0,46 - 0,51)	0,06	0,950	24,37 (24,02 - 24,71)	-4,56	0,000
Propriá	1.336	7,74	0,48 (0,45 - 0,51)	-1,45	0,656	24,25 (23,90 - 24,60)	-5,08	0,000
<i>TOTAL</i>	17.254	100,00	0,49 (0,48 - 0,50)	-	-	25,16 (25,06 - 25,25)	-	-

Fonte: IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007

Relembra-se aqui a distribuição da captura em uma tabela 2 x 2, e são definidas as denominações das caselas. Para efeitos de estimação, o SINASC foi considerado a “Amostra A” e o Registro Civil, a “amostra B”. Desta maneira, n_A é o número de nascidos vivos capturados pelo SINASC e n_B , o número de nascidos vivos capturados pelo Registro Civil:

		Registro Civil		
		Sim	Não	
SINASC	Sim	n_{11}	n_{10}	n_A
	Não	n_{01}	n_{00}	
		n_B		

Analogamente, n_{11} são os nascidos vivos capturados pelas duas bases, n_{10} são aqueles capturados somente pelo SINASC e n_{01} , os que foram somente capturados pelo Registro Civil. Os índices das caselas também representam a história de captura que alimenta o Programa MARK[®]. Por exemplo, um nascido vivo que foi capturado pelas duas bases, filho de mãe de 24 anos e outro que foi capturado somente pelo SINASC, de mãe de 16 anos, ambos na microrregião Sergipana do Sertão de São Francisco, foram importados da seguinte maneira para o Programa:

```
/*HIST  MICRO1  MICRO2  (...)  MICRO13  IDMAE  INDIV*/
  11      1      0      (...)      0      24      1;
  10      1      0      (...)      0      16      1;
```

Em que “11” e “10” são as histórias de captura; as sequências de “1” seguidos de 12 “0s” representam a microrregião 1; “24” e “16” são as idades das mães e “1” indica que há um indivíduo com cada uma destas características.

Os modelos construídos são similares ao modelo de número 4, de Huggins (1991), mais detalhado na seção de Apresentação do Método.

Neste modelo, assume-se que a probabilidade de captura é influenciada por grupos de indivíduos e por variáveis individuais. As microrregiões geográficas e as regionais de saúde são os agrupamentos de indivíduos e a idade da mãe na ocasião do parto, a variável individual. Desta forma, quatro bancos de dados foram gerados para imputação no programa MARK[®]: dois para o agrupamento por microrregiões, sendo um com o pareamento menos restrito, um com o pareamento mais restrito; e dois para as regionais de saúde. Os modelos em cada banco foram estimados de maneira independente, isto é, como resultado, houve quatro modelos finais, dois para as microrregiões e dois para as regionais de saúde.

O modelo completo estimou um parâmetro de primeira captura e um de recaptura para cada um dos grupos considerados. Não foi considerado, para o modelo inicial, aquele onde a primeira captura varia entre as ocasiões de captura, uma vez que no presente estudo, a primeira captura (representada pelo parâmetro p) é definida como a presença na base do SINASC, e a segunda captura (c) é a presença na base do Registro Civil. A segunda captura (presença no Registro Civil) não é uma segunda ocorrência de nascimento vivo. Desta maneira, considera-se que a primeira captura é constante ao longo das ocasiões.

Como notação para os nomes dos modelos, adotou-se o seguinte padrão:

$$p(g + idmae + g*idmae) c(g+ idmae + g*idmae)$$

Onde:

p = probabilidade da primeira captura;

c = probabilidade de recaptura;

g = grupo;

$idmae$ = idade da mãe; e

$g*idmae$ = termo de interação entre o grupo e a idade da mãe.

Esta é a notação para o modelo completo e estima p como função do

grupo no qual o indivíduo está inserido, da idade da mãe e da interação entre o grupo e a idade da mãe, assim como c também é função do grupo no qual o indivíduo está inserido, da idade da mãe e da interação entre o grupo e a idade da mãe.

A partir deste modelo completo, 29 outros sub-modelos são derivados (Quadro 2), e foram ajustados em cada um dos bancos gerados.

Quadro 2 – Modelos de Huggins para populações fechadas e notação adotada (continua)

N.º	MODELO	NOTAÇÃO
1	Captura e recaptura dependem do grupo, idade da mãe e da interação	$p(g + idmae + g*idmae) c(g+ idmae + g*idmae)$
2	Captura depende do grupo, idade da mãe e da interação. Recaptura depende do grupo e da idade da mãe	$p(g + idmae + g*idmae) c(g+ idmae)$
3	Captura depende do grupo, idade da mãe e da interação. Recaptura depende do grupo	$p(g + idmae + g*idmae) c(g)$
4	Captura depende do grupo, idade da mãe e da interação. Recaptura depende da idade da mãe	$p(g + idmae + g*idmae) c(idmae)$
5	Captura depende do grupo, idade da mãe e da interação. Recaptura não depende de nenhum destes fatores	$p(g + idmae + g*idmae) c(.)$
6	Captura depende do grupo e da idade da mãe. Recaptura depende do grupo, da idade da mãe e da interação	$p(g + idmae) c(g+ idmae + g*idmae)$
7	Captura e recaptura dependem do grupo e da idade da mãe	$p(g + idmae) c(g+ idmae)$
8	Captura depende do grupo e da idade da mãe. Recaptura depende do grupo.	$p(g + idmae) c(g)$
9	Captura depende do grupo e da idade da mãe. Recaptura depende da idade da mãe	$p(g + idmae) c(idmae)$
10	Captura depende do grupo e da idade da mãe. Recaptura não depende de nenhum destes fatores	$p(g + idmae) c(.)$
11	Captura depende do grupo. Recaptura depende do grupo, da idade da mãe e da interação	$p(g) c(g+ idmae + g*idmae)$
12	Captura depende do grupo. Recaptura depende do grupo e da idade da mãe	$p(g) c(g+ idmae)$
13	Captura e recaptura dependem do grupo	$p(g) c(g)$

Quadro 2 – Modelos de Huggins para populações fechadas e notação adotada (conclusão)

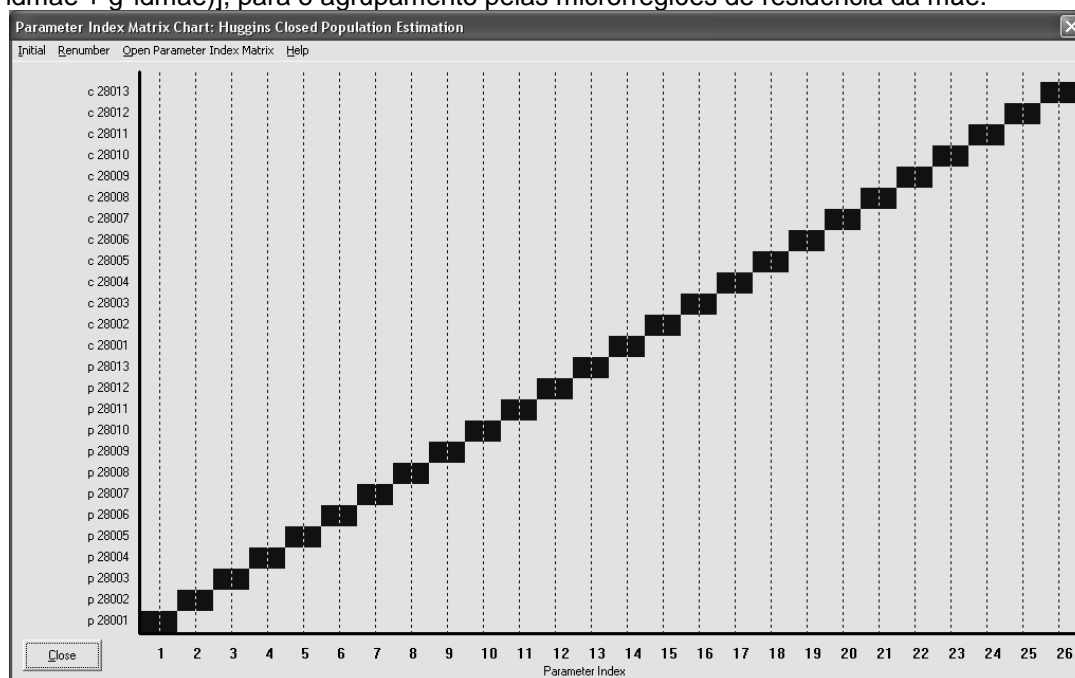
N.º	MODELO	NOTAÇÃO
14	Captura depende do grupo. Recaptura depende da idade da mãe	$p(g) c(idmae)$
15	Captura depende do grupo. Recaptura não depende de nenhum destes fatores	$p(g) c(.)$
16	Captura depende da idade da mãe. Recaptura depende do grupo, da idade da mãe e da interação	$p(idmae) c(g+ idmae + g*idmae)$
17	Captura depende da idade da mãe. Recaptura depende do grupo e da idade da mãe	$p(idmae) c(g+ idmae)$
18	Captura depende da idade da mãe. Recaptura depende do grupo.	$p(idmae) c(g)$
19	Captura e recaptura dependem da idade da mãe.	$p(idmae) c(idmae)$
20	Captura depende da idade da mãe. Recaptura não depende de nenhum destes fatores	$p(idmae) c(.)$
21	Captura não depende de nenhum destes fatores. Recaptura depende do grupo, da idade da mãe e da interação	$p(.) c(g+ idmae + g*idmae)$
22	Captura não depende de nenhum destes fatores. Recaptura depende do grupo e da idade da mãe.	$p(.) c(g+ idmae)$
23	Captura não depende de nenhum destes fatores. Recaptura depende do grupo.	$p(.) c(g)$
24	Captura não depende de nenhum destes fatores. Recaptura depende da idade da mãe.	$p(.) c(idmae)$
25	Captura e recaptura não dependem de nenhum dos fatores	$p(.) c(.)$
26	As probabilidades de captura e recaptura são iguais e dependem do grupo, da idade da mãe e da interação	$p = c (g + idmae + g*idmae)$
27	As probabilidades de captura e recaptura são iguais e dependem do grupo e da idade da mãe	$p = c (g + idmae)$
28	As probabilidades de captura e recaptura são iguais e dependem do grupo	$p = c (g)$
29	As probabilidades de captura e recaptura são iguais e dependem da idade da mãe	$p = c (idmae)$
30	As probabilidades de captura e recaptura são iguais e não dependem de nenhum dos fatores	$p = c (.)$

O modelo completo estimou um parâmetro de primeira captura e um de recaptura para cada um dos grupos considerados, além de um parâmetro para a idade da mãe e um parâmetro de interação entre o grupo e a idade da mãe, para cada grupo. O número de parâmetros de cada modelo variou de acordo com o agrupamento utilizado, pois há 13 microrregiões e 7 regionais de saúde. Os modelos, para cada agrupamento, são detalhados nas subseções a seguir.

4.2.8.1 MICRORREGIÕES

Como visto anteriormente, foi considerada a probabilidade da primeira captura constante ao longo das ocasiões e o Gráfico de Índices dos Parâmetros, para o modelo completo, é (Figura 26):

Figura 26 – Índices dos parâmetros do modelo completo $[p(g + idmae + g*idmae) c(g + idmae + g*idmae)]$, para o agrupamento pelas microrregiões de residência da mãe.



Considerando a idade da mãe uma variável individual e cada microrregião como um grupo de nascidos vivos, ajustam-se modelos lineares completos para a probabilidade de captura pelo SINASC:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 g_1 + \beta_2 g_2 + \dots + \beta_{12} g_{12} + \beta_{13} \text{idmae}_i + \beta_{14}(g_1 * \text{idmae}_i) + \beta_{15}(g_2 * \text{idmae}_i) + \dots + \beta_{25}(g_{12} * \text{idmae}_i) \quad (5)$$

Onde,

p_i é a probabilidade do indivíduo i estar na base do SINASC;

β_0 é o parâmetro do intercepto;

β_k é o parâmetro estimado para o grupo k (k é a microrregião);

g_k são os indivíduos pertencem ao grupo k ;

β_{13} é o parâmetro estimado para a idade da mãe;

idmae_i é a idade da mãe do indivíduo i , em anos; e

β_{k+13} é o parâmetro estimado para a interação entre o grupo k e a idade da mãe;

Similarmente, para a recaptura, ou seja, presença na base do Registro Civil, o modelo completo, adotando subscritos seqüenciais, é:

$$\ln\left(\frac{c_i}{1-c_i}\right) = \beta_{26} + \beta_{27} g_1 + \beta_{28} g_2 + \dots + \beta_{38} g_{12} + \beta_{39} \text{idmae}_i + \beta_{40}(g_1 * \text{idmae}_i) + \beta_{41}(g_2 * \text{idmae}_i) + \dots + \beta_{51}(g_{12} * \text{idmae}_i) \quad (6)$$

Onde,

c_i é a probabilidade do indivíduo i estar na base do Registro Civil;

β_{26} é o parâmetro do intercepto;

β_{k+26} é o parâmetro estimado para o grupo k (k é a microrregião);

g_k são os indivíduos pertencem ao grupo k ;

β_{39} é o parâmetro estimado para a idade da mãe;

idmae_i é a idade da mãe do indivíduo i , em anos; e

β_{k+39} é o parâmetro estimado para a interação entre o grupo k e a idade da mãe;

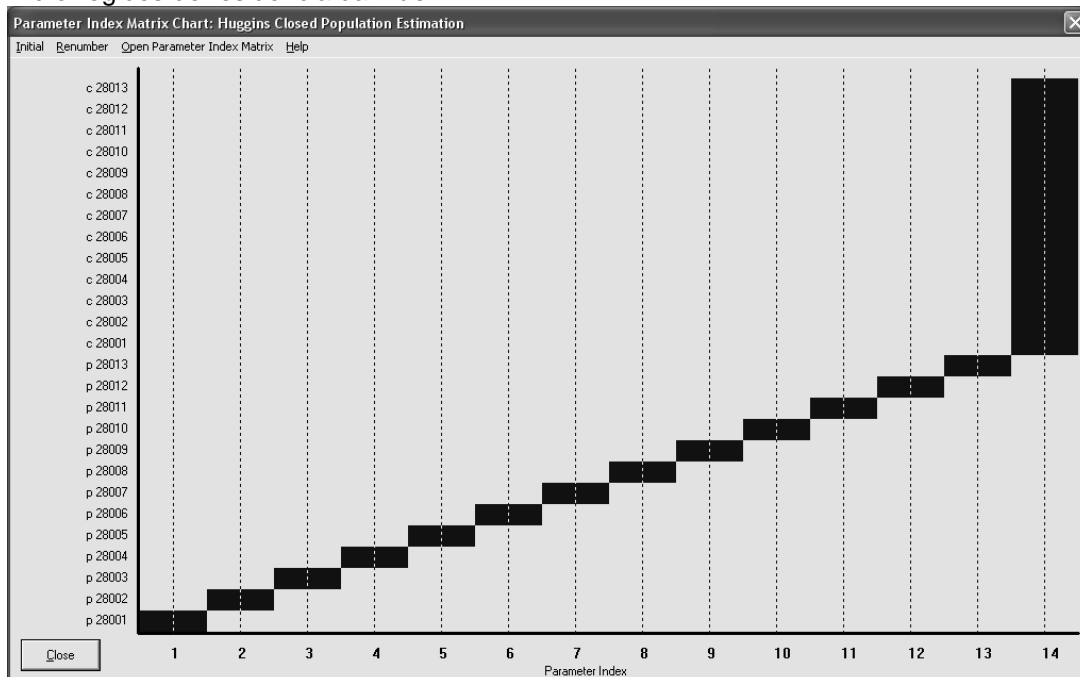
Com esta parametrização, construiu-se a matriz de desenho do modelo linear, onde cada coluna representa um parâmetro das equações (5) e (6). A Matriz de Desenho para p e c , do modelo completo, no programa MARK[®], então, é como especificada no Quadro 3.

A partir do modelo completo, os demais 29 sub-modelos, cuja notação é exibida no Quadro 2, foram derivados. Os Gráficos de Índices dos Parâmetros e Matrizes de Desenho de todos os sub-modelos não são mostrados aqui. Para exemplificar, exibem-se, entretanto, os Gráficos e Matrizes de Desenho de mais três deles, de maneira a ficar claro ao leitor o procedimento adotado para suas construções. Considere as notações dos sub-modelos de n.ºs 14, 22 e 27 apresentadas no Quadro 2:

N.º	MODELO	NOTAÇÃO
14	Captura depende do Grupo. Recaptura depende da idade da mãe	$p(g) c(idmae)$
22	Captura não depende de nenhuma destas variáveis. Recaptura depende do grupo e da idade da mãe.	$p(.) c(g+ idmae)$
27	As probabilidades de captura e recaptura são iguais e dependem do grupo	$p = c (g + idmae)$

Para o modelo 14, onde a captura pelo SINASC depende da microrregião de residência da mãe e a captura pelo Registro Civil depende somente da idade da mãe, o Gráfico de Índices de Parâmetros é (Figura 27):

Figura 27 – Índices dos parâmetros do modelo [p(g) c(idmae)], para o agrupamento pelas microrregiões de residência da mãe.



Para caracterizar que a captura pelo Registro Civil (c) não varia ao longo das microrregiões (grupos), atribui-se o mesmo índice para todos os grupos. O MARK[®] estimará a população total de nascidos vivos em cada grupo, porém usará a mesma estimativa de c em cada um dos grupos. O modelo para a captura pelo SINASC (p) construído foi:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 g_1 + \beta_2 g_2 + \dots + \beta_{12} g_{12}$$

Onde,

p_i é a probabilidade do indivíduo i estar na base do SINASC;

β_0 é o parâmetro do intercepto;

β_k é o parâmetro estimado para o grupo k (k é a microrregião); e

g_k são os indivíduos pertencem ao grupo k ;

Para a probabilidade de captura pelo Registro Civil, o modelo, adotando subscritos seqüenciais, é:

$$\ln\left(\frac{c_i}{1 - c_i}\right) = \beta_{13} + \beta_{14}idmae_i$$

Onde,

c_i é a probabilidade do indivíduo i estar na base do Registro Civil;

β_{13} é o parâmetro do intercepto;

β_{14} é o parâmetro estimado para a idade da mãe; e

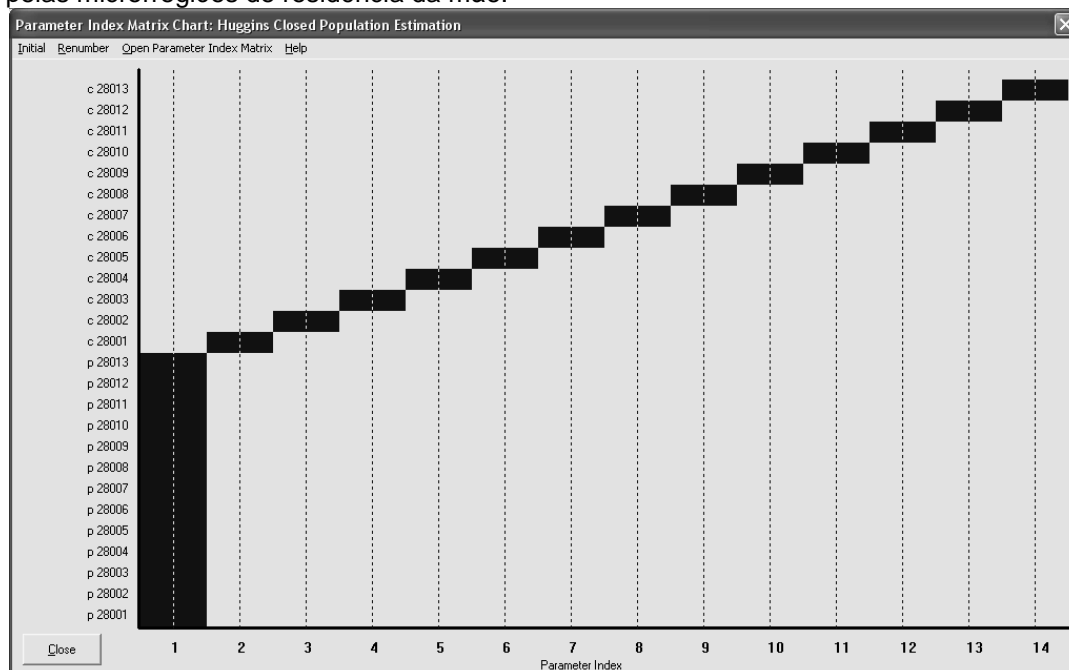
$idmae_i$ é a idade da mãe do indivíduo i , em anos;

A Matriz de desenho, então, é dada por:

p/c	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}
p ₁	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
p ₂	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
p ₃	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
p ₄	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
p ₅	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
p ₆	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
p ₇	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
p ₈	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
p ₉	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
p ₁₀	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
p ₁₁	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
p ₁₂	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
p ₁₃	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c ₁	0	0	0	0	0	0	0	0	0	0	0	0	0	1	IDMAE

O modelo 23 especifica que a captura pelo SINASC (p) não depende nem da microrregião de residência, nem da idade da mãe, e que a captura pelo Registro Civil (c) depende da microrregião de residência e da idade da mãe, mas não da interação entre estes fatores. O Gráfico dos Índices dos Parâmetros é (Figura 28):

Figura 28 – Índices dos parâmetros do modelo $[p(.) c(g + idmae)]$, para o agrupamento pelas microrregiões de residência da mãe.



O modelo para p é:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0$$

Onde,

p_i é a probabilidade do indivíduo i estar na base do SINASC; e

β_0 é o parâmetro do intercepto;

E o modelo para c :

$$\ln\left(\frac{c_i}{1 - c_i}\right) = \beta_1 + \beta_2 g_1 + \beta_3 g_2 + \dots + \beta_{13} g_{12} + \beta_{14} idmae_i$$

Onde,

c_i é a probabilidade do indivíduo i estar na base do Registro Civil;

β_1 é o parâmetro do intercepto;

β_{k+1} é o parâmetro estimado para o grupo k (k é a microrregião);

g_k são os indivíduos pertencem ao grupo k ;

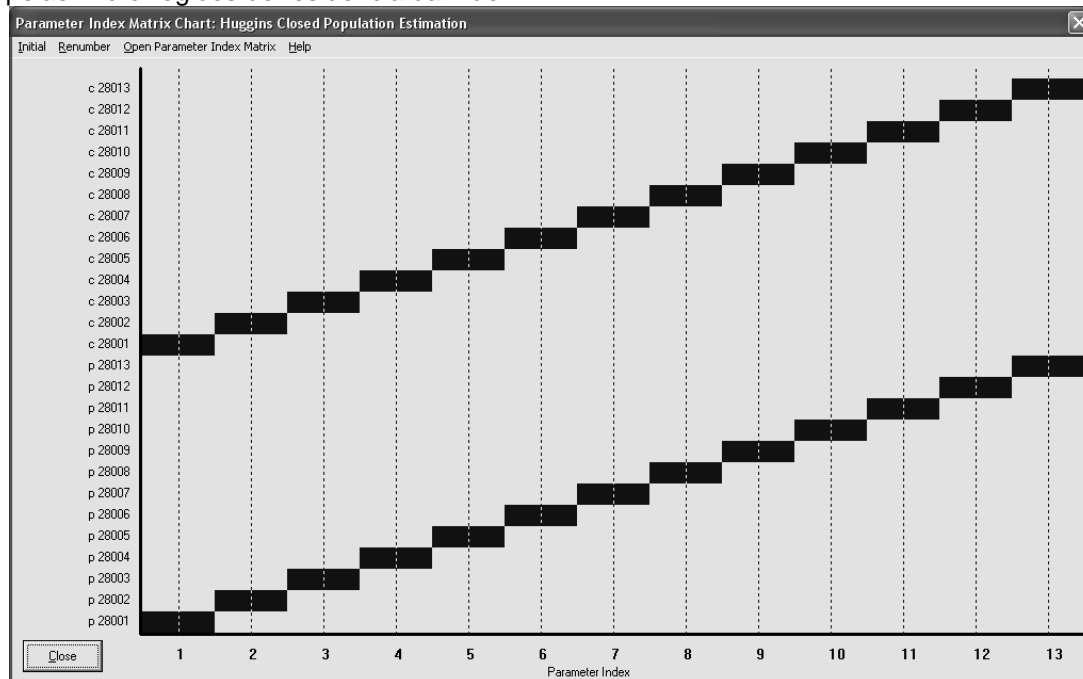
β_{14} é o parâmetro estimado para a idade da mãe; e
 $idmae_i$ é a idade da mãe do indivíduo i , em anos;

Para este modelo, a Matriz de Desenho é:

p/c	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}
p_1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c_1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	IDMAE
c_2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	IDMAE
c_3	0	1	0	0	1	0	0	0	0	0	0	0	0	0	IDMAE
c_4	0	1	0	0	0	1	0	0	0	0	0	0	0	0	IDMAE
c_5	0	1	0	0	0	0	1	0	0	0	0	0	0	0	IDMAE
c_6	0	1	0	0	0	0	0	1	0	0	0	0	0	0	IDMAE
c_7	0	1	0	0	0	0	0	0	1	0	0	0	0	0	IDMAE
c_8	0	1	0	0	0	0	0	0	0	1	0	0	0	0	IDMAE
c_9	0	1	0	0	0	0	0	0	0	0	1	0	0	0	IDMAE
c_{10}	0	1	0	0	0	0	0	0	0	0	0	1	0	0	IDMAE
c_{11}	0	1	0	0	0	0	0	0	0	0	0	0	1	0	IDMAE
c_{12}	0	1	0	0	0	0	0	0	0	0	0	0	0	1	IDMAE
c_{13}	0	1	0	0	0	0	0	0	0	0	0	0	0	0	IDMAE

No modelo 27, onde a captura pelo SINASC e pelo Registro Civil são consideradas iguais e variam ao longo das microrregiões, atribui-se um índice para cada grupo, e estes índices são atribuídos tanto para p quanto para c (Figura 29):

Figura 29 – Índices dos parâmetros do modelo [$p = c(g + idmae)$], para o agrupamento pelas microrregiões de residência da mãe.



O modelo para p é:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 g_1 + \beta_2 g_2 + \dots + \beta_{12} g_{12} + \beta_{13} idmae_i$$

Onde,

p_i é a probabilidade do indivíduo i estar na base do SINASC; e

β_0 é o parâmetro do intercepto;

β_k é o parâmetro estimado para o grupo k (k é a microrregião);

g_k são os indivíduos pertencem ao grupo k ;

β_{13} é o parâmetro estimado para a idade da mãe; e

$idmae_i$ é a idade da mãe do indivíduo i , em anos;

E o modelo para c é o mesmo:

$$\ln\left(\frac{c_i}{1 - c_i}\right) = \beta_0 + \beta_1 g_1 + \beta_2 g_2 + \dots + \beta_{12} g_{12} + \beta_{13} idmae_i$$

Onde,

c_i é a probabilidade do indivíduo i estar na base do Registro Civil

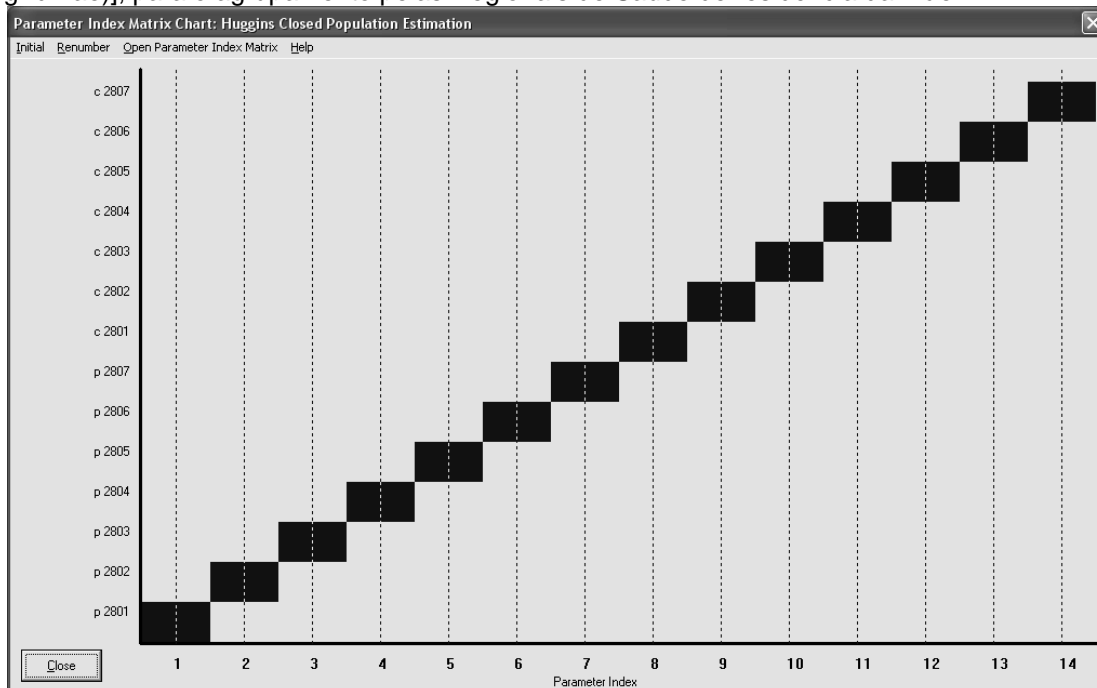
A Matriz de Desenho é:

c = p	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}
$c_1 = p_1$	1	1	0	0	0	0	0	0	0	0	0	0	0	IDMAE
$c_2 = p_2$	1	0	1	0	0	0	0	0	0	0	0	0	0	IDMAE
$c_3 = p_3$	1	0	0	1	0	0	0	0	0	0	0	0	0	IDMAE
$c_4 = p_4$	1	0	0	0	1	0	0	0	0	0	0	0	0	IDMAE
$c_5 = p_5$	1	0	0	0	0	1	0	0	0	0	0	0	0	IDMAE
$c_6 = p_6$	1	0	0	0	0	0	1	0	0	0	0	0	0	IDMAE
$c_7 = p_7$	1	0	0	0	0	0	0	1	0	0	0	0	0	IDMAE
$c_8 = p_8$	1	0	0	0	0	0	0	0	1	0	0	0	0	IDMAE
$c_9 = p_9$	1	0	0	0	0	0	0	0	0	1	0	0	0	IDMAE
$c_{10} = p_{10}$	1	0	0	0	0	0	0	0	0	0	1	0	0	IDMAE
$c_{11} = p_{11}$	1	0	0	0	0	0	0	0	0	0	0	1	0	IDMAE
$c_{12} = p_{12}$	1	0	0	0	0	0	0	0	0	0	0	0	1	IDMAE
$c_{13} = p_{13}$	1	0	0	0	0	0	0	0	0	0	0	0	0	IDMAE

4.2.8.2 REGIONAIS DE SAÚDE

Quando o agrupamento considerado foi a Regional de Saúde de residência da mãe, houve 7 grupos. O modelo completo, $[p(g + idmae + g*idmae) c(g+ idmae + g*idmae)]$, tem o Gráfico de Índices de Parâmetros da Figura 30:

Figura 30 – Índices dos parâmetros do modelo $[p(g + idmae + g*idmae) c(g+ idmae + g*idmae)]$, para o agrupamento pelas Regionais de Saúde de residência da mãe.



A construção do modelo é similar àquela feita pelo agrupamento pelas microrregiões. O modelo completo para a captura pelo SINASC (p), é:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 g_1 + \beta_2 g_2 + \dots + \beta_6 g_6 + \beta_7 idmae_i + \beta_8 (g_1 * idmae_i) + \beta_9 (g_2 * idmae_i) + \dots + \beta_{13} (g_6 * idmae_i) \quad (7)$$

Onde,

p_i é a probabilidade do indivíduo i estar na base do SINASC;

β_0 é o parâmetro do intercepto;

β_k é o parâmetro estimado para o grupo k (k é a regional de saúde);

g_k são os indivíduos pertencem ao grupo k ;

β_7 é o parâmetro estimado para a idade da mãe;

$idmae_i$ é a idade da mãe do indivíduo i , em anos; e

β_{k+7} é o parâmetro estimado para a interação entre o grupo k e a idade da mãe;

Similarmente, para a recaptura, ou seja, presença na base do Registro Civil, o modelo completo, adotando subscritos seqüenciais, é:

$$\ln\left(\frac{c_i}{1 - c_i}\right) = \beta_{14} + \beta_{15}g_1 + \beta_{16}g_2 + \dots + \beta_{20}g_6 + \beta_{21}idmae_i + \beta_{22}(g_1 * idmae_i) + \beta_{23}(g_2 * idmae_i) + \dots + \beta_{27}(g_6 * idmae_i) \quad (8)$$

Onde,

c_i é a probabilidade do indivíduo i estar na base do Registro Civil;

β_{14} é o parâmetro do intercepto;

β_{k+14} é o parâmetro estimado para o grupo k (k é a regional de saúde);

g_k são os indivíduos pertencem ao grupo k ;

β_{21} é o parâmetro estimado para a idade da mãe;

$idmae_i$ é a idade da mãe do indivíduo i , em anos; e

β_{k+21} é o parâmetro estimado para a interação entre o grupo k e a idade da mãe;

E a Matriz de Desenho é exibida no Quadro 4.

Os procedimentos para a construção dos demais 29 sub-modelos lineares propostos foram os mesmos adotados para as microrregiões.

Quadro 4 – Matriz de Desenho do modelo linear completo $[p(g + idmae + g*idmae) c(g + idmae + g*idmae)]$ para as probabilidades de ser capturado pelo SINASC (p) e pelo Registro Civil (c), segundo a regional de saúde de residência da mãe

p/c	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}	β_{17}	β_{18}	β_{19}	β_{20}	β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	β_{26}	β_{27}	
p₁	1	1	0	0	0	0	0	idmae	idmae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p₂	1	0	1	0	0	0	0	idmae	0	idmae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p₃	1	0	0	1	0	0	0	idmae	0	0	idmae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p₄	1	0	0	0	1	0	0	idmae	0	0	0	idmae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p₅	1	0	0	0	0	1	0	idmae	0	0	0	0	idmae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p₆	1	0	0	0	0	0	1	idmae	0	0	0	0	0	idmae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p₇	1	0	0	0	0	0	0	idmae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	idmae	idmae	0	0	0	0	0	0
c₂	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	idmae	0	idmae	0	0	0	0	0
c₃	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	idmae	0	0	idmae	0	0	0	0
c₄	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	idmae	0	0	0	idmae	0	0	0
c₅	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	idmae	0	0	0	0	idmae	0	0
c₆	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	idmae	0	0	0	0	0	0	idmae
c₇	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	idmae	0	0	0	0	0	0	0

Apesar de inúmeros fatores poderem interferir na probabilidade da captura do nascimento vivo pelo SINASC e pelo Registro Civil, os modelos de Huggins para captura e recaptura em populações fechadas tem como requisito a disponibilidade de informações para todos os indivíduos presentes no estudo. Por esta razão, somente a idade da mãe e o sexo do recém-nascido foram incluídos como potenciais co-variáveis dos modelos. Uma vez que somente a idade da mãe mostrou-se com diferenças estatisticamente significantes em ambas as bases, ao se fazer a desagregação geográfica, apenas a idade da mãe foi aplicada como co-variável individual.

A partir do modelo completo, vários outros podem ser gerados, além dos 30 eleitos para o estudo. Estes foram escolhidos por serem os mais gerais e que permitirão traçar um perfil do comportamento da probabilidade de captura pelas bases de dados, de acordo com os grupos e idade da mãe.

4.2.9. ESTIMAÇÃO DO SUB-REGISTRO CIVIL DE NASCIDOS VIVOS E DA COBERTURA DO SINASC

Uma vez obtidas as estimativas do total de nascidos vivos, partiu-se para o cálculo do sub-registro civil e da cobertura do SINASC, no segundo e terceiro trimestres de 2006, em Sergipe, assim como as estimativas para as subdivisões territoriais de microrregião e regional de saúde de residência da mãe. Os estimadores são dados de maneira bem simples, na forma de percentual:

Sub-registro civil:

$$\widehat{SR}_s = \frac{(\widehat{N}_s - n_{BS})}{\widehat{N}_s} \times 100 \quad (9)$$

Onde,

\widehat{SR}_s = Percentual de sub-registro na subdivisão s (microrregião, regional de saúde).

\widehat{N}_s = Total de nascidos vivos estimados para a subdivisão s .

n_{BS} = Número de nascidos vivos capturados pelo Registro Civil, na

subdivisão s .

A variação do sub-registro, aproximada pelos limites dos Intervalos de Confiança obtidos para NV, conforme expressão de número (4) apresentada na página 75, é:

$$\text{Valor mínimo } \widehat{SR}_s = \frac{(Li\widehat{N}_s - n_{BS})}{Li\widehat{N}_s} \times 100 \quad (10)$$

Onde,

Valor mínimo \widehat{SR}_s = Valor mínimo do percentual de sub-registro na subdivisão s (microrregião, regional de saúde), obtido a partir de $Li\widehat{N}_s$.

$Li\widehat{N}_s$ = Limite inferior do intervalo com 95% de confiança do total de nascidos vivos estimados para a subdivisão s .

n_{BS} = Número de nascidos vivos capturados pelo Registro Civil, na subdivisão s .

$$\text{Valor máximo } \widehat{SR}_s = \frac{(Ls\widehat{N}_s - n_{BS})}{Ls\widehat{N}_s} \times 100 \quad (11)$$

Onde,

Valor máximo \widehat{SR}_s = Valor máximo do percentual de sub-registro na subdivisão s (microrregião, regional de saúde), obtido a partir de $Ls\widehat{N}_s$.

$Ls\widehat{N}_s$ = Limite superior do intervalo com 95% de confiança do total de nascidos vivos estimados para a subdivisão s .

n_{BS} = Número de nascidos vivos capturados pelo Registro Civil, na subdivisão s .

E a cobertura do SINASC:

$$\widehat{C}_s = \frac{n_{AS}}{\widehat{N}_s} \times 100 \quad (12)$$

Onde,

\widehat{C}_s = Percentual de cobertura pelo SINASC na subdivisão s (microrregião, regional de saúde).

\widehat{N}_s = Total de nascidos vivos estimados para a subdivisão s .

n_{AS} = Número de nascidos vivos capturados pelo SINASC, na subdivisão s .

E a variação:

$$\text{Valor mínimo } \hat{C}_s = \frac{n_{As}}{Li\hat{N}_s} \times 100 \quad (13)$$

Onde,

Valor mínimo \hat{C}_s = Valor mínimo do percentual de cobertura pelo SINASC na subdivisão s (microrregião, regional de saúde), obtido a partir de $Li\hat{N}_s$.

$Li\hat{N}_s$ = Limite inferior do intervalo com 95% de confiança do total de nascidos vivos estimados para a subdivisão s .

n_{As} = Número de nascidos vivos capturados pelo SINASC, na subdivisão s .

$$\text{Valor máximo } \hat{C}_s = \frac{n_{As}}{Ls\hat{N}_s} \times 100 \quad (14)$$

Onde,

Valor máximo \hat{C}_s = Valor máximo do percentual de cobertura pelo SINASC na subdivisão s (microrregião, regional de saúde), obtido a partir de $Ls\hat{N}_s$.

$Ls\hat{N}_s$ = Limite superior do intervalo com 95% de confiança do total de nascidos vivos estimados para a subdivisão s .

n_{As} = Número de nascidos vivos capturados pelo SINASC, na subdivisão s .

5. RESULTADOS

A partir dos dois critérios de pareamento dos dados, ajustaram-se os modelos para o estado de Sergipe como um todo e para cada tipo de desagregação geográfica adotada, microrregiões e regionais de saúde de residência da mãe, o que gerou quatro conjuntos de resultados. A seguir expõem-se os resultados dos modelos, as estimativas das probabilidades de captura no SINASC e no RC, além das estimativas derivadas do sub-registro civil de nascidos vivos e da cobertura do SINASC.

5.1. PAREAMENTO MENOS RESTRITO

Conforme visto na seção que descreve os bancos utilizados para a estimação, o pareamento somente pelo número da DN levou à identificação de 21.224 nascidos vivos em Sergipe, no segundo e terceiro trimestres de 2006. A fim de se estimar o total de nascidos vivos para o estado de Sergipe como um todo, sem desagregação geográfica, foram ajustados quatro modelos de Huggins, onde se investigou a influência da idade da mãe na captura pelo SINASC e pelo RC. O modelo com maior peso, o que “melhor suporta os dados” é aquele onde a idade da mãe interfere na captura de nascidos vivos pelo Registro Civil, com cerca de 66% do peso total entre os modelos. O modelo que inclui a idade da mãe na captura tanto pelo SINASC, quanto na captura pelo Registro Civil foi o segundo modelo com maior peso, em torno de 33% (Tabela 20).

Tabela 20 – Estimativa dos critérios de informação de Akaike (AIC), das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (Δ AIC), pesos normalizados de AIC, verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, segundo o pareamento menos restrito.

MODELO	AIC	Δ AIC	w_i	VEROSSIMILHANÇA DO MODELO	N.º PARÂMETROS
{p(.) c(idmae)}	31652,237	0,00	0,66138	1,0000	3
{p(idmae) c(idmae)}	31653,596	1,36	0,33519	0,5068	4
{p(.) c(.)}	31663,579	11,34	0,00228	0,0034	2
{p(idmae) c(.)}	31664,939	12,70	0,00115	0,0017	3

O protocolo da pesquisa previa a ponderação dos parâmetros, de maneira a acomodar todos os modelos nas estimativas finais. Entretanto, como se vê na Tabela 21, a ponderação não alterou as estimativas do modelo com maior peso, {p(.) c(idmae)}. Desta maneira, foram consideradas somente as estimativas do referido modelo, para o cálculo do sub-registro no Estado de Sergipe. Este comportamento das estimativas ponderadas se repetiu nas desagregações geográficas, de modo que também nestas somente o modelo com maior peso ponderado de Akaike foi selecionado. As tabelas com as estimativas ponderadas das probabilidades de captura pelo

SINASC e pelo Registro Civil e do total de nascidos vivos, de acordo com o pareamento menos restrito, podem ser conferidas no Anexo VI.

Tabela 21 – Modelos de Huggins para populações fechadas, com as respectivas estimativas de peso ponderado (w_i), probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) e estimativas ponderadas por w_i . Sergipe, pareamento menos restrito, segundo e terceiro trimestres de 2006.

MODELO	w_i	\hat{p} (SINASC)	ERRO- PADRÃO \hat{p}	\hat{c} (RC)	ERRO- PADRÃO \hat{c}	\hat{N}	ERRO- PADRÃO \hat{N}
{p(.) c(idmae)}	0,661	0,912	0,002	0,804	0,004	21.391	15,474
{p(idmae) c(idmae)}	0,335	0,913	0,003	0,804	0,004	21.391	15,483
{p(.) c(.)}	0,002	0,912	0,002	0,796	0,003	21.391	15,474
{p(idmae) c(.)}	0,001	0,913	0,003	0,796	0,003	21.391	15,483
Ponderação (MARK®)*		0,912	0,002	0,804	0,004	21.391	15,477

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

Com a estimativa \hat{N} , completa-se a tabela 2 x 2 de distribuição de captura de nascidos vivos pelas duas fontes analisadas (Tabela 22).

Tabela 22 – Distribuição da captura de nascidos vivos pelo SINASC e Registro Civil no estado de Sergipe, segundo e terceiro trimestres de 2006, e estimativa do total de nascidos vivos, segundo o modelo {p(.) c(idmae)} e o pareamento menos restrito.

		Pareamento menos restrito		
		RC (B)		
		Sim	Não	
SINASC (A)	Sim	15.532 (n_{11})	3.970 (n_{10})	19.502 (n_A)
	Não	1.722 (n_{01})	167 (\hat{n}_{00})	
		17.254 (n_B)		

TOTAL DE NASCIDOS VIVOS CAPTURADOS ($r = n_A + n_B - n_{11}$): 21.224

TOTAL DE NASCIDOS VIVOS ESTIMADOS (\hat{N}): **21.391**

O modelo {p(idmae) c(.)} estimou 21.391 nascidos vivos no segundo e terceiro trimestre de 2006 em Sergipe, e um intervalo com 95% de confiança de [21.363 – 21.423]. O sub-registro civil e a cobertura do SINASC neste período, de acordo com o pareamento menos restrito, então, foram:

Sub-registro no Registro Civil

$$\widehat{SR}_{\text{Sergipe}} = \frac{(21.391 - 17.254)}{21.391} * 100 = 19,34\%$$

$$\text{Valor mínimo } \widehat{SR}_{\text{Sergipe}} = \frac{(21.363 - 17.254)}{21.363} * 100 = 19,23\%$$

$$\text{Valor máximo } \widehat{SR}_{\text{Sergipe}} = \frac{(21.423 - 17.254)}{21.423} * 100 = 19,46\%$$

Cobertura do SINASC

$$\widehat{C}_{\text{Sergipe}} = \left[\frac{(19.502)}{21.391} \right] * 100 = 91,17\%$$

$$\text{Valor mínimo } \widehat{C}_{\text{Sergipe}} = \left[\frac{(19.502)}{21.423} \right] * 100 = 91,03\%$$

$$\text{Valor máximo } \widehat{C}_{\text{Sergipe}} = \left[\frac{(19.502)}{21.363} \right] * 100 = 91,29\%$$

Resumindo, no período analisado, em Sergipe, o sub-registro na base do Registro Civil de nascidos vivos foi 19,34% (19,23 – 19,46) e a cobertura na base do SINASC, 91,17% (91,03 – 91,29), segundo o pareamento com menos restrito.

5.1.5. MICRORREGIÕES

Ao se incluírem as microrregiões de residência da mãe como

agrupamentos de nascidos vivos, dos 30 modelos ajustados aos dados, apenas 5 mostraram algum peso relativo, quando se avaliam os critérios AIC e a verossimilhança (Tabela 23). Os modelos são exibidos em ordem decrescente dos pesos normalizados de Akaike (w_i).

Como salientado anteriormente, as estimativas ponderadas não alteraram o resultado da estimativa do modelo com maior peso ponderado. Então será utilizado, para fins de estimação do sub-registro, o modelo $\{p(g) c(g + idmae + g*idmae)\}$ (Tabela 24).

A probabilidade de captura pelo SINASC é alta em todas as microrregiões, variando de 0,69 no Agreste de Lagarto até 0,95 em Estância e Nossa Senhora das Dores. Destaca-se que em Agreste de Lagarto há um grande número de registros no RC sem o número da DN preenchido (mais de 31%, Tabela 9), o que prejudicou o pareamento das bases. Também em Tobias Barreto, Japarutuba e Boquim, o percentual de registros no RC com a DN em branco é maior do que 5% (Tabela 9), o que pode explicar a probabilidade de captura pelo SINASC menor do que 0,90. Além destas microrregiões, onde o pareamento foi afetado pela falta do número da DN, somente em Propriá a probabilidade de captura pelo SINASC foi menor do que 0,90 (Tabela 24).

Tabela 23 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (Δ AIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento menos restrito, segundo a microrregião de residência da mãe (continua).

Modelo	AIC	Δ AIC	w_i	Verossimilhança do modelo	Nº Parâmetros
{p(g) c(g + idmae + g*idmae)}	30457,151	0,00	0,66837	1,0000	39
{p(g + idmae) c(g + idmae + g*idmae)}	30458,914	1,76	0,27672	0,4140	40
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	30462,151	5,00	0,05485	0,0821	52
{p(g) c(g + idmae)}	30476,520	19,37	0,00004	0,0001	27
{p(g + idmae) c(g + idmae)}	30478,282	21,13	0,00002	0,0000	28
{p(g + idmae + g*idmae) c(g + idmae)}	30481,507	24,36	0,00000	0,0000	40
{p(g) c(g)}	30484,221	27,07	0,00000	0,0000	26
{p(g + idmae) c(g)}	30485,983	28,83	0,00000	0,0000	27
{p(g + idmae + g*idmae) c(g)}	30489,208	32,06	0,00000	0,0000	39
{p(.) c(g + idmae + g*idmae)}	30910,791	453,64	0,00000	0,0000	27
{p(idmae) c(g + idmae + g*idmae)}	30912,153	455,00	0,00000	0,0000	28
{p(.) c(g + idmae)}	30930,174	473,02	0,00000	0,0000	15
{p(idmae) c(g + idmae)}	30931,535	474,38	0,00000	0,0000	16
{p(.) c(g)}	30937,876	480,73	0,00000	0,0000	14
{p(idmae) c(g)}	30939,236	482,09	0,00000	0,0000	15
{p(g) c(idmae)}	31198,568	741,42	0,00000	0,0000	15
{p(g + idmae) c(idmae)}	31200,330	743,18	0,00000	0,0000	16
{p(g + idmae + g*idmae) c(idmae)}	31203,540	746,39	0,00000	0,0000	28
{p(g) c(.)}	31209,910	752,76	0,00000	0,0000	14
{p(g + idmae) c(.)}	31211,671	754,52	0,00000	0,0000	15
{p(g + idmae + g*idmae) c(.)}	31214,882	757,73	0,00000	0,0000	27
{p = c(g + idmae + g*idmae)}	31423,536	966,39	0,00000	0,0000	26

Tabela 23 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (Δ AIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento menos restrito, segundo a microrregião de residência da mãe (conclusão).

Modelo	AIC	ΔAIC	w_i	Verossimilhança do modelo	Nº Parâmetros
{p = c(g + idmae)}	31453,780	996,63	0,00000	0,0000	14
{p = c(g)}	31460,969	1003,8	0,00000	0,0000	13
{p(.) c(idmae)}	31652,237	1195,0	0,00000	0,0000	3
{p(idmae) c(idmae)}	31653,597	1196,4	0,00000	0,0000	4
{p(.) c(.)}	31663,579	1206,4	0,00000	0,0000	2
{p(idmae) c(.)}	31664,939	1207,7	0,00000	0,0000	3
{p = c(idmae)}	32563,270	2106,1	0,00000	0,0000	2
{p = c(.)}	32574,058	2116,9	0,00000	0,0000	1

Em relação às probabilidades de captura pelo RC, nota-se que são sensivelmente menores do que as do SINASC, o que já era de se esperar, uma vez que o número de registros no RC é menor do que no SINASC. No RC, a maior probabilidade de captura estimada foi na microrregião de Aracaju (0,85) e a menor foi na Sergipana do Sertão do São Francisco (0,71), excetuando-se as microrregiões com pareamento problemático, já mencionadas (Tabela 24).

O total de nascidos vivos estimados (\hat{N}) foi bem próximo ao total medido (r) em todas as microrregiões, devido à alta sobreposição das listas. Novamente no Agreste de Lagarto, o alto percentual de registros com DN em branco, no RC, levou a uma sobreposição menor (n_{11} relativamente baixo), o que acabou por inflar a estimativa do total de nascidos vivos. A diferença absoluta entre os nascidos vivos estimados e capturados é menor do que 20, em quase todas as microrregiões onde o percentual de DNs em branco no RC é abaixo de 5%, sendo apenas 2 em Carira e Nossa Senhora das Dores (Tabela 24).

O sub-registro civil de nascidos no estado de Sergipe gira em torno de 19%. Entretanto, ao longo das microrregiões, variou de pouco mais de 12% no Baixo Contiguiba até quase 27% na Sergipana do Sertão do São Francisco. O sub-registro estimado no Agreste de Lagarto superou os 40%, porém este resultado deve ser avaliado com cautela, pelas razões anteriormente discutidas. Nota-se que o sub-registro civil é menor nas microrregiões localizadas na parte central do Estado, Aracaju, Baixo Contiguiba e Agreste de Itabaiana, todas com sub-registro civil estimado em menos de 15% dos nascidos vivos. À medida que as microrregiões de residência da mãe se afastam destas partes centrais, o sub-registro civil de nascidos vivos aumenta (Tabela 25).

A cobertura do SINASC, por sua vez, mostra-se relativamente alta, sendo menor do que 90% apenas em Propriá, Japaratuba, Agreste de Lagarto, Tobias Barreto e Boquim. Ainda, nestas microrregiões, somente Propriá não apresentou percentual de registros com DN em branco maior do que 5%, no Registro Civil. Em Aracaju, Estância e Nossa Senhora das Dores

a cobertura do SINASC chega a quase 95% (Tabela 25).

Ao se comparar o sub-registro civil estimado em 2005 pelo método de Brass com o sub-registro no segundo e terceiro trimestres de 2006, estimado pelo método de captura e recaptura, nota-se que a diferença entre as estimativas é bem pequena nas microrregiões de Carira, Contiguiba e Propriá. As maiores diferenças estão nas microrregiões de Agreste de Lagarto, Agreste de Itabaiana, Baixo Contiguiba e Aracaju, onde os valores de sub-registro obtidos pelo método de captura e recaptura superam em duas vezes ou mais os estimados pelo método de Brass (Figura 31).

Percebe-se, também, que apesar das distintas metodologias, de uma maneira geral, as menores estimativas do sub-registro pelo método de Brass também são as menores pela captura e recaptura, assim como as maiores estimativas também coincidem. As exceções são as microrregiões de Tobias Barreto, que pelo método da captura e recaptura é uma das microrregiões com maior sub-registro, mas não pelo método de Brass, e em contrapartida, em Estância e Nossa Senhora das Dores, os sub-registros estimados pelo método de Brass estão entre os mais altos, o que não acontece na estimação pela captura e recaptura (Figura 31)

Tabela 24 – Estimativas da probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) do modelo $\{p(g) c(g + idmae + g*idmae)\}$, distribuição da captura pelo SINASC (n_A), RC (n_B), sobreposição das bases (n_{11}) e total de nascidos vivos distintos capturados pelas duas bases ($r = n_A + n_B - n_{11}$). Sergipe, segundo a microrregião de residência da mãe, pareamento menos restrito, segundo e terceiro trimestres de 2006.

MICRORREGIÃO	P (PROBABILIDADE DE CAPTURA PELO SINASC)			C (PROBABILIDADE DE CAPTURA PELO RC)			TOTAL DE NASCIDOS VIVOS						
	\hat{p}	ERRO-PADRÃO	IC 95%	\hat{c}	ERRO-PADRÃO	IC 95%	n_A (SINASC)	n_B (RC)	n_{11}	r	\hat{N}	ERRO-PADRÃO	IC 95%
S.São Francisco-SE	0,912	0,008	0,896 - 0,926	0,714	0,011	0,692 - 0,735	1.692	1.359	1.210	1.841	1.855	4,544	1.849 - 1.867
Carira	0,937	0,011	0,912 - 0,956	0,822	0,016	0,787 - 0,852	541	479	445	575	577	1,716	576 - 583
N.Sra. Dores	0,949	0,010	0,926 - 0,965	0,823	0,016	0,789 - 0,852	565	493	464	594	596	1,388	594 - 601
Agr. Itabaiana	0,932	0,007	0,917 - 0,944	0,861	0,009	0,843 - 0,878	1.505	1.399	1.296	1.608	1.616	3,161	1.611 - 1.625
Tobias Barreto*	0,874	0,012	0,849 - 0,896	0,719	0,014	0,690 - 0,746	986	833	709	1.110	1.128	5,480	1.120 - 1.142
Agr. Lagarto*	0,694	0,019	0,655 - 0,731	0,516	0,015	0,486 - 0,546	1.056	868	545	1.379	1.521	23,574	1.482 - 1.575
Propriá	0,876	0,013	0,849 - 0,898	0,807	0,014	0,779 - 0,832	885	821	711	995	1.011	5,113	1.003 - 1.024
Cotinguiba	0,914	0,014	0,883 - 0,938	0,829	0,017	0,793 - 0,860	491	448	406	533	537	2,361	534 - 545
Japaratuba*	0,874	0,016	0,839 - 0,903	0,775	0,018	0,738 - 0,809	549	493	424	618	628	4,087	623 - 640
Baixo Cotinguiba	0,921	0,009	0,900 - 0,938	0,870	0,011	0,847 - 0,890	949	901	826	1.024	1.030	2,980	1.027 - 1.039
Aracaju	0,944	0,003	0,938 - 0,949	0,848	0,004	0,840 - 0,856	7.286	6.590	6.183	7.693	7.717	5,491	7.708 - 7.730
Boquim*	0,886	0,009	0,868 - 0,902	0,749	0,011	0,728 - 0,770	1.661	1.434	1.245	1.850	1.874	6,231	1.865 - 1.890
Estância	0,949	0,006	0,935 - 0,960	0,799	0,011	0,777 - 0,820	1.336	1.136	1.068	1.404	1.408	2,117	1.405 - 1.414

* As microrregiões sombreadas são as que tiveram mais de 5% dos registros no RC com o número da DN em branco, o que prejudicou o pareamento

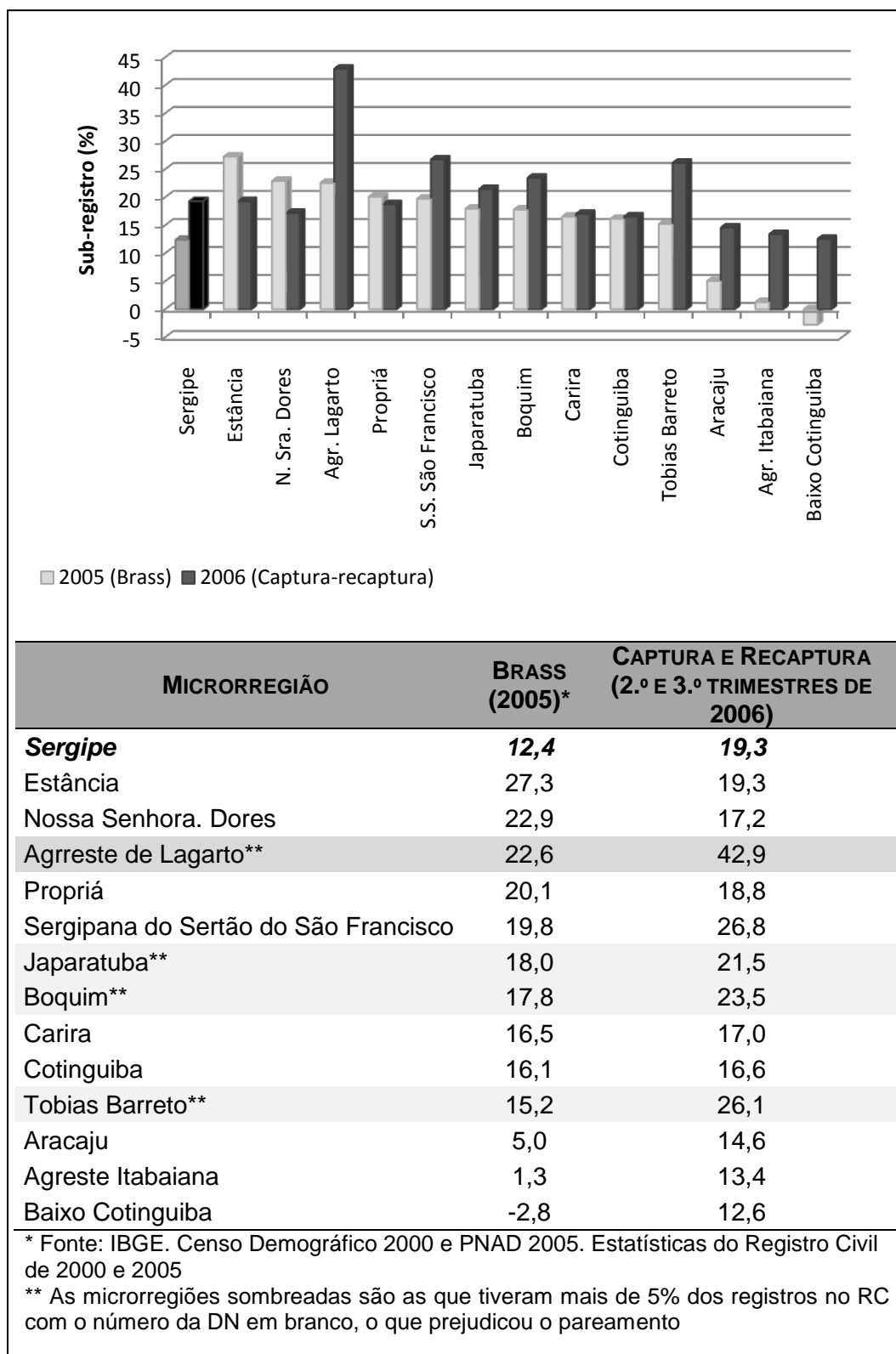
Tabela 25 – Estimativas da cobertura do SINASC (\hat{C}) e do sub-registro (\widehat{SR}) no Registro Civil (RC) de acordo com o modelo $\{p(g) c(g + idmae + g*idmae)\}$ e pareamento menos restrito. Sergipe, segundo a microrregião de residência da mãe, segundo e terceiro trimestres de 2006.

MICRORREGIÃO	\hat{N}	SINASC			RC		
		n_A	\hat{C}	VARIAÇÃO*	n_B	\widehat{SR}	VARIAÇÃO*
Sergipana do Sertão do São Francisco	1.855	1.692	91,19	90,63 - 91,51	1.359	26,75	26,50 - 27,21
Carira	577	541	93,72	92,80 - 93,92	479	17,02	16,84 - 17,84
Nossa Senhora das Dores	596	565	94,87	94,01 - 95,12	493	17,22	17,00 - 17,97
Agreste de Itabaiana	1.616	1.505	93,16	92,62 - 93,42	1.399	13,41	13,16 - 13,91
Tobias Barreto**	1.128	986	87,42	86,34 - 88,04	833	26,14	25,63 - 27,06
Agreste de Lagarto**	1.521	1.056	69,41	67,05 - 71,26	868	42,94	41,43 - 44,89
Propriá	1.011	885	87,57	86,43 - 88,24	821	18,76	18,15 - 19,82
Cotinguiba	537	491	91,45	90,09 - 91,95	448	16,56	16,10 - 17,80
Japaratuba**	628	549	87,43	85,78 - 88,12	493	21,49	20,87 - 22,97
Baixo Cotinguiba	1.030	949	92,10	91,34 - 92,41	901	12,56	12,27 - 13,28
Aracaju	7.717	7.286	94,41	94,26 - 94,53	6.590	14,61	14,50 - 14,75
Boquim**	1.874	1.661	88,62	87,88 - 89,06	1.434	23,49	23,11 - 24,13
Estância	1.408	1.336	94,91	94,48 - 95,09	1.136	19,30	19,15 - 19,66

* As variações apresentadas são aproximações, obtidas a partir dos limites do intervalo de confiança de \hat{N} .

** As microrregiões sombreadas são as que tiveram mais de 5% dos registros no RC com o número da DN em branco, o que prejudicou o pareamento

Figura 31 – Estimativas do percentual de sub-registro civil de nascidos vivos pelos métodos de Brass, para o ano de 2005 e pela captura e recaptura segundo o pareamento menos restrito, para o segundo e terceiro trimestres de 2006. Sergipe e microrregião de residência da mãe



5.1.6. REGIONAIS DE SAÚDE

Ao se empregar a regional de saúde de residência da mãe como o agrupamento de nascidos vivos, sete modelos apresentaram peso ponderado diferente de zero (Tabela 26). Novamente, as estimativas dos parâmetros ponderadas por w_i não foram diferentes das estimativas do modelo com maior peso, $\{p(g + idmae + g*idmae) c(g + idmae + g*idmae)\}$, o modelo completo. As estimações do sub-registro e cobertura foram feitas a partir dos dados do modelo selecionado (com maior peso). Vide o anexo VII para as estimativas ponderadas dos parâmetros no agrupamento pelas regionais de saúde e pareamento menos restrito.

Na Tabela 27 são apresentados os resultados do modelo $\{p(g + idmae + g*idmae) c(g + idmae + g*idmae)\}$. As probabilidades de captura pelo SINASC (\hat{p}) estimadas em quase todas as regionais de saúde foram superiores a 0,9, salvas as regionais de Propriá, 0,88 e Lagarto, 0,78. Nesta última regional, o percentual de registros no RC com o número da DN em branco foi bastante alto, cerca de 21% (Tabela 12), levando a um n_{11} relativamente pequeno e conseqüente menores estimativas de \hat{p} e \hat{c} , quando comparado às demais regionais. Já as estimativas da probabilidade de captura pelo Registro Civil, (\hat{c}) são menores, em todas as regionais, sendo a mais alta em Aracaju, 0,86. Em Estância e Nossa Senhora da Glória a probabilidade de captura pelo RC é menor do que 0,80.

O número total de nascidos vivos estimado pelo modelo completo é ligeiramente maior do que o número total de nascidos vivos distintos capturados pelas duas bases. A diferença absoluta entre estas duas medidas é de apenas 8 nascidos vivos em Nossa Senhora do Socorro e a maior diferença, de 29 nascidos vivos, foi verificada em Aracaju, excluindo a regional de Lagarto (diferença de 150 nascidos vivos), que teve o pareamento prejudicado (Tabela 27).

O sub-registro civil no período analisado, estimado a partir do modelo selecionado, mostrou-se, como nas microrregiões, com razoável amplitude, variando de menos de 15% em Aracaju e Itabaiana até mais de 25% em

Nossa Senhora da Glória. Em Lagarto, o sub-registro estimado superou os 30%, entretanto este número deve ser analisado com cautela (Tabela 28).

A cobertura do SINASC, por sua vez, também é bastante alta, superando os 90% em quase todas as regionais de saúde, exceto em Propriá e Lagarto, com 88% e 78%, respectivamente. As maiores coberturas estimadas foram em nossa Senhora do Socorro e Aracaju, quase 95% em cada (Tabela 28).

Tabela 26 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (Δ AIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento menos restrito, segundo a região de saúde de residência da mãe (continua).

Modelo	AIC	Δ AIC	w_i	Verossimilhança do modelo	Nº Parâmetros
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	30526,363	0,00	0,62377	1,0000	28
{p(g) c(g + idmae + g*idmae)}	30528,251	1,89	0,24275	0,3892	21
{p(g + idmae) c(g + idmae + g*idmae)}	30529,450	3,09	0,13328	0,2137	22
{p(g + idmae + g*idmae) c(g + idmae)}	30543,567	17,20	0,00011	0,0002	22
{p(g) c(g + idmae)}	30545,458	19,10	0,00004	0,0001	15
{p(g + idmae) c(g + idmae)}	30546,657	20,29	0,00002	0,0000	16
{p(g + idmae + g*idmae) c(g)}	30549,807	23,44	0,00001	0,0000	21
{p(g) c(g)}	30551,699	25,34	0,00000	0,0000	14
{p(g + idmae) c(g)}	30552,898	26,53	0,00000	0,0000	15
{p(.) c(g + idmae + g*idmae)}	30946,416	420,05	0,00000	0,0000	15
{p(idmae) c(g + idmae + g*idmae)}	30947,777	421,41	0,00000	0,0000	16
{p(.) c(g + idmae)}	30963,627	437,26	0,00000	0,0000	9
{p(idmae) c(g + idmae)}	30964,987	438,62	0,00000	0,0000	10
{p(.) c(g)}	30969,868	443,51	0,00000	0,0000	8
{p(idmae) c(g)}	30971,228	444,86	0,00000	0,0000	9
{p(g + idmae + g*idmae) c(idmae)}	31232,169	705,81	0,00000	0,0000	16
{p(g) c(idmae)}	31234,064	707,70	0,00000	0,0000	9
{p(g + idmae) c(idmae)}	31235,262	708,90	0,00000	0,0000	10
{p(g + idmae + g*idmae) c(.)}	31243,510	717,15	0,00000	0,0000	15
{p(g) c(.)}	31245,406	719,04	0,00000	0,0000	8
{p(g + idmae) c(.)}	31246,604	720,24	0,00000	0,0000	9
{p = c (g + idmae + g*idmae)}	31512,429	986,07	0,00000	0,0000	14

Tabela 26 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (ΔAIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento menos restrito, segundo a região de saúde de residência da mãe (conclusão).

Modelo	AIC	ΔAIC	w_i	Verossimilhança do modelo	Nº Parâmetros
{p = c (g + idmae)}	31537,167	1010,8	0,00000	0,0000	8
{p = c (g)}	31544,366	1018,0	0,00000	0,0000	7
{p(.) c(idmae)}	31652,237	1125,8	0,00000	0,0000	3
{p(idmae) c(idmae)}	31653,597	1127,2	0,00000	0,0000	4
{p(.) c(.)}	31663,579	1137,2	0,00000	0,0000	2
{p(idmae) c(.)}	31664,939	1138,5	0,00000	0,0000	3
{p = c (idmae)}	32563,270	2036,9	0,00000	0,0000	2
{p = c (.)}	32574,058	2047,6	0,00000	0,0000	1

Tabela 27 – Estimativas da probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) do modelo $\{p(g + idmae + g*idmae) c(g + idmae + g*idmae)\}$, distribuição da captura pelo SINASC (n_A), RC (n_B), sobreposição das bases (n_{11}) e total de nascidos vivos distintos capturados pelas duas bases ($r = n_A + n_B - n_{11}$). Sergipe, segundo a regional de saúde de residência da mãe, pareamento menos restrito, segundo e terceiro trimestres de 2006.

REGIONAL DE SAÚDE	P (PROBABILIDADE DE CAPTURA PELO SINASC)			C (PROBABILIDADE DE CAPTURA PELO RC)			TOTAL DE NASCIDOS VIVOS						
	\hat{p}	ERRO-PADRÃO	IC 95%	\hat{c}	ERRO-PADRÃO	IC 95%	n_A (SINASC)	n_B (RC)	n_{11}	r	\hat{N}	ERRO-PADRÃO	IC 95%
Aracaju	0,938	0,003	0,932 - 0,944	0,864	0,004	0,855 - 0,872	6.839	6.332	5.908	7.263	7.292	6,111	7.282 - 7.306
N.Sra. Socorro	0,949	0,004	0,939 - 0,957	0,809	0,008	0,794 - 0,824	2.754	2.369	2.226	2.897	2.905	3,144	2.901 - 2.914
Estância	0,928	0,006	0,916 - 0,938	0,774	0,008	0,757 - 0,790	2.466	2.088	1.910	2.644	2.658	4,314	2.652 - 2.669
Itabaiana	0,932	0,006	0,919 - 0,943	0,851	0,008	0,835 - 0,866	1.977	1.818	1.683	2.112	2.122	3,618	2.117 - 2.132
Lagarto	0,776	0,011	0,754 - 0,797	0,619	0,010	0,599 - 0,639	2.314	1.952	1.433	2.833	2.983	19,923	2.949 - 3.028
N.Sra. Glória	0,911	0,008	0,895 - 0,925	0,714	0,011	0,692 - 0,735	1.692	1.359	1.210	1.841	1.856	4,687	1.849 - 1.868
Propriá	0,880	0,010	0,859 - 0,897	0,800	0,011	0,778 - 0,820	1.460	1.336	1.162	995	1.011	5,113	1.003 - 1.024

*A regional sombreada teve 21,31% dos registros no RC com o número da DN em branco, o que prejudicou o pareamento.

Tabela 28 – Estimativas da cobertura do SINASC (\hat{C}) e do sub-registro (\widehat{SR}) no Registro Civil (RC) de acordo com o modelo $\{p(g + idmae + g*idmae) c(g + idmae + g*idmae)\}$ e pareamento menos restrito. Sergipe, segundo a regional de saúde de residência da mãe, segundo e terceiro trimestres de 2006.

REGIONAL DE SAÚDE	\hat{N}	SINASC			RC		
		n_A	\hat{C}	VARIAÇÃO*	n_B	\widehat{SR}	VARIAÇÃO*
Aracaju	7.292	6.839	93,79	93,61 - 93,92	6.332	13,16	13,05 - 13,33
Nossa Senhora do Socorro	2.905	2.754	94,80	94,51 - 94,93	2.369	18,45	18,34 - 18,70
Estância	2.658	2.466	92,78	92,39 - 92,99	2.088	21,44	21,27 - 21,77
Itabaiana	2.122	1.977	93,17	92,73 - 93,39	1.818	14,32	14,12 - 14,73
Lagarto**	2.983	2.314	77,56	76,42 - 78,47	1.952	34,57	33,81 - 35,54
Nossa Senhora da Glória	1.856	1.692	91,17	90,58 - 91,51	1.359	26,77	26,50 - 27,25
Propriá	1.658	1.460	88,07	87,27 - 88,59	1.336	19,41	18,93 - 20,14

* As variações apresentadas são aproximações, obtidas a partir dos limites do intervalo de confiança de \hat{N} .

**A regional sombreada teve 21,31% dos registros no RC com o número da DN em branco, o que prejudicou o pareamento.

5.2. PAREAMENTO MAIS RESTRITO

Ao se restringirem os critérios para considerar um par real, a casela n_{11} é menor do que no pareamento menos restrito, e por conseqüência, o total de nascidos vivos estimados é maior. Com este pareamento, foram contabilizados 23.635 nascidos vivos distintos em Sergipe, no segundo e terceiro trimestres de 2006. Para todo o estado, foram ajustados quatro modelos, para investigar a influência da idade da mãe na captura pelas bases de dados. Também neste critério de pareamento, o modelo com maior peso foi o $\{p(.) c(idmae)\}$, ou seja, a idade da mãe interfere na probabilidade de captura pelo Registro Civil, mas não no SINASC (Tabela 29).

Tabela 29 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (ΔAIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, segundo o pareamento mais restrito.

MODELO	AIC	ΔAIC	w_i	VEROSSIMILHANÇA DO MODELO	N.º PARÂMETROS
$\{P(.) C(IDMAE)\}$	46631,9	0,00	0,5553	1,0000	3
$\{P(IDMAE)C(IDMAE)\}$	46633,9	2,00	0,2043	0,3678	4
$\{P(.) C(.)\}$	46634,2	2,30	0,1758	0,3165	2
$\{P(IDMAE) C(.)\}$	46636,2	4,30	0,0647	0,1164	3

A ponderação dos parâmetros não alterou as estimativas para p , c e N do modelo com maior peso (Tabela 30). Desta forma, o modelo $\{p(.) c(idmae)\}$ foi utilizado para estimar o sub-registro civil e a cobertura do SINASC no estado de Sergipe, no segundo e terceiro trimestres de 2006.

Tabela 30 – Modelos de Huggins para populações fechadas, com as respectivas estimativas de peso ponderado (w_i), probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) e estimativas ponderadas por w_i . Sergipe, pareamento mais restrito, segundo e terceiro trimestres de 2006.

MODELO	w_i	\hat{p} (SINASC)	ERRO- PADRÃO \hat{p}	\hat{c} (RC)	ERRO- PADRÃO \hat{c}	\hat{N}	ERRO- PADRÃO \hat{N}
{P(.) C(IDMAE)}	0,55532	0,786	0,004	0,677	0,004	24.765	53,101
{P(IDMAE) C(IDMAE)}	0,20426	0,786	0,004	0,677	0,004	24.765	53,103
{P(.) C(.)}	0,17576	0,786	0,004	0,672	0,003	24.765	53,101
{P(IDMAE) C(.)}	0,06465	0,786	0,004	0,672	0,003	24.765	53,102
Ponderação (MARK®)*		0,786	0,004	0,676	0,004	24.765	53,101

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

Monta-se, então, a tabela 2 x 2 que permite visualizar a distribuição da captura pelos dois sistemas de informação utilizados (Tabela 31). Chama a atenção a diferença na casela n_{00} , estimada pela diferença entre \hat{N} e r , obtida pelos pareamentos menos e mais restritos. Enquanto que no pareamento menos restrito, estimou-se que apenas 167 nascidos vivos não foram capturados por nenhuma das duas bases (Tabela 22), com o critério mais restrito estimou-se que 1.130 nascidos vivos não foram capturados nem pelo SINASC, nem pelo Registro Civil. Este é o reflexo da importância do pressuposto de marcação unívoca e inequívoca da captura e recaptura: mudanças na “chave de identificação” dos indivíduos levam a resultados discrepantes.

Tabela 31 – Distribuição da captura de nascidos vivos pelo SINASC e Registro Civil no estado de Sergipe, segundo e terceiro trimestres de 2006, e estimativa do total de nascidos vivos, segundo o modelo $\{p(.) c(idmae)\}$ e o pareamento mais restrito.

		Pareamento mais restrito		
		RC (B)		
		Sim	Não	
SINASC (A)	Sim	13.094 (n_{11})	6.381 (n_{10})	19.475 (n_A)
	Não	4.160 (n_{01})	1.130 (\hat{n}_{00})	
		17.254 (n_B)		
TOTAL DE NASCIDOS VIVOS CAPTURADOS ($r = n_A + n_B - n_{11}$): 23.635				
TOTAL DE NASCIDOS VIVOS ESTIMADOS (\hat{N}): 24.765				

O sub-registro civil e a cobertura do SINASC, calculados com os resultados do modelo $\{p(.) c(idmae)\}$, $\hat{N} = 24.765$ e IC95% [24.666 – 24.874], são os seguintes.

Sub-registro no Registro Civil

$$\widehat{SR}_{Sergipe} = \frac{(24.765 - 17.254)}{24.765} * 100 = 30,33\%$$

$$\text{Valor mínimo } \widehat{SR}_{Sergipe} = \frac{(24.666 - 17.254)}{24.666} * 100 = 30,05\%$$

$$\text{Valor máximo } \widehat{SR}_{Sergipe} = \frac{(24.874 - 17.254)}{24.874} * 100 = 30,63\%$$

Cobertura do SINASC

$$\widehat{C}_{Sergipe} = \left[1 - \frac{(24.765 - 19.502)}{24.765} \right] * 100 = 78,75\%$$

$$\text{Valor mínimo } \widehat{C}_{Sergipe} = \left[1 - \frac{(24.874 - 19.502)}{24.874} \right] * 100 = 78,40\%$$

$$\text{Valor máximo } \hat{C}_{\text{Sergipe}} = \left[1 - \frac{(24.666 - 19.502)}{24.666} \right] * 100 = 79,06\%$$

Resumindo, o sub-registro civil em Sergipe, no período considerado, pelo pareamento mais restrito, foi de 30,33% [30,05% - 30,63%]. Já a cobertura do SINASC em Sergipe, no mesmo período, foi de 78,75% [78,40% - 79,06%].

5.2.5. MICRORREGIÕES

Dos 30 modelos ajustados, apenas 6 mostraram algum peso ponderado por w_i (Tabela 32). A ponderação dos parâmetros não alterou os resultados do modelo com maior peso, $\{p(g) c(g + idmae + g*idmae)\}$, e as estimativas deste foram utilizadas para o cálculo do sub-registro no Registro Civil e cobertura do SINASC em cada uma das microrregiões. As estimativas ponderadas podem ser conferidas no Anexo VIII.

Como era de se esperar, as estimativas das probabilidades de captura, tanto pelo SINASC quanto pelo RC, foram mais baixas do que aquelas estimadas com o pareamento menos restrito. As probabilidades de captura pelo SINASC mais altas são registradas em Estância, Aracaju e no Agreste de Itabaiana. No Agreste de Lagarto a probabilidade de captura pelo SINASC foi de apenas 0,5, devido tanto ao alto percentual de registros no RC sem o número da DN, quanto aos critérios mais rígidos de pareamento. De fato, as microrregiões com menores probabilidades de captura pelo SINASC são aquelas onde o percentual de registros no RC sem o número da DN supera os 5%. Na Sergipana do Sertão do São Francisco e em Propriá, a aplicação dos critérios adicionais para o pareamento gerou uma diferença considerável nos pares, de 20,83% e 28,27%, respectivamente (Tabela 12), e por esta razão também aqui a probabilidade de captura pelo SINASC acabou menor do que 0,80 (Tabela 33).

O Registro Civil, por sua vez, apresentou ainda menores estimativas de probabilidade de captura, sendo somente Aracaju, Baixo Contiguiba, Agreste de Itabaiana e Estância com probabilidades maiores do que 0,7. A estimativa no Agreste de Lagarto foi de apenas 0,3, pelas razões já expostas (Tabela 33). Adicionalmente, a diferença absoluta entre o total de nascidos vivos estimados e os nascidos vivos capturados pelas duas bases mostrou-se bem mais elevada, quando comparada com a diferença obtida quando o pareamento adotado foi menos restrito.

Tabela 32 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (ΔAIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento mais restrito, segundo a microrregião de residência da mãe (continua).

Modelo	AIC	ΔAIC	w_i	Verossimilhança do modelo	Nº Parâmetros
{p(g) c(g + idmae + g*idmae)}	45147,719	0,00	0,65520	1,0000	39
{p(g + idmae) c(g + idmae + g*idmae)}	45149,007	1,29	0,34405	0,5251	40
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	45161,365	13,65	0,00071	0,0011	52
{p(g) c(g)}	45169,385	21,67	0,00001	0,0000	26
{p(g) c(g + idmae)}	45170,133	22,41	0,00001	0,0000	27
{p(g + idmae) c(g)}	45170,671	22,95	0,00001	0,0000	27
{p(g + idmae) c(g + idmae)}	45171,419	23,70	0,00000	0,0000	28
{p(g + idmae + g*idmae) c(g)}	45183,014	35,30	0,00000	0,0000	39
{p(g + idmae + g*idmae) c(g + idmae)}	45183,765	36,05	0,00000	0,0000	40
{p(.) c(g + idmae + g*idmae)}	45594,473	446,75	0,00000	0,0000	27
{p(idmae) c(g + idmae + g*idmae)}	45596,476	448,76	0,00000	0,0000	28
{p(.) c(g)}	45616,152	468,43	0,00000	0,0000	14
{p(.) c(g + idmae)}	45616,899	469,18	0,00000	0,0000	15
{p(idmae) c(g)}	45618,153	470,43	0,00000	0,0000	15
{p(idmae) c(g + idmae)}	45618,901	471,18	0,00000	0,0000	16
{p = c(g + idmae + g*idmae)}	45641,106	493,39	0,00000	0,0000	26
{p = c(g)}	45663,796	516,08	0,00000	0,0000	13
{p = c(g + idmae)}	45665,531	517,81	0,00000	0,0000	14
{p(g) c(idmae)}	46185,116	1037,4	0,00000	0,0000	15
{p(g + idmae) c(idmae)}	46186,403	1038,6	0,00000	0,0000	16
{p(g) c(.)}	46187,416	1039,7	0,00000	0,0000	14
{p(g + idmae) c(.)}	46188,702	1040,9	0,00000	0,0000	15

Tabela 32 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (Δ AIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento mais restrito, segundo a microrregião de residência da mãe (conclusão).

Modelo	AIC	ΔAIC	w_i	Verossimilhança do modelo	Nº Parâmetros
{p(g + idmae + g*idmae) c(idmae)}	46198,735	1051,0	0,00000	0,0000	28
{p(g + idmae + g*idmae) c(.)}	46201,032	1053,3	0,00000	0,0000	27
{p(.) c(idmae)}	46631,895	1484,1	0,00000	0,0000	3
{p(idmae) c(idmae)}	46633,895	1486,1	0,00000	0,0000	4
{p(.) c(.)}	46634,195	1486,4	0,00000	0,0000	2
{p(idmae) c(.)}	46636,196	1488,4	0,00000	0,0000	3
{p = c(idmae)}	47102,538	1954,8	0,00000	0,0000	2
{p = c(.)}	47103,688	1955,9	0,00000	0,0000	1

Tabela 33 – Estimativas da probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) do modelo $\{p(g) c(g + idmae + g*idmae)\}$, distribuição da captura pelo SINASC (n_A), RC (n_B), sobreposição das bases (n_{11}) e total de nascidos vivos distintos capturados pelas duas bases ($r = n_A + n_B - n_{11}$). Sergipe, segundo a microrregião de residência da mãe, pareamento mais restrito, segundo e terceiro trimestres de 2006.

MICRORREGIÃO	P (PROBABILIDADE DE CAPTURA PELO SINASC)			C (PROBABILIDADE DE CAPTURA PELO RC)			TOTAL DE NASCIDOS VIVOS						
	\hat{p}	ERRO-PADRÃO	IC 95%	\hat{c}	ERRO-PADRÃO	IC 95%	n_A (SINASC)	n_B (RC)	n_{11}	r	\hat{N}	ERRO-PADRÃO	IC 95%
S.S Francisco - SE	0,761	0,013	0,734 - 0,786	0,569	0,012	0,545 - 0,593	1.678	1.359	958	2.079	2.205	18,814	2.173 - 2.247
Carira	0,803	0,021	0,760 - 0,841	0,675	0,020	0,634 - 0,713	549	479	371	657	683	7,816	672 - 704
N.Sra. Dores	0,800	0,021	0,756 - 0,837	0,675	0,020	0,635 - 0,713	564	493	380	677	705	8,153	693 - 726
Agr. Itabaiana	0,824	0,012	0,800 - 0,846	0,759	0,011	0,737 - 0,780	1.498	1.399	1.135	1.762	1.818	10,903	1.801 - 1844
Tobias Barreto*	0,706	0,020	0,666 - 0,743	0,559	0,016	0,528 - 0,590	976	833	546	1.263	1.383	20,966	1.348 - 1.431
Agr. Lagarto*	0,520	0,026	0,470 - 0,570	0,319	0,014	0,292 - 0,347	1.086	868	347	1.607	2.087	71,034	1.967 - 2.248
Propriá	0,641	0,024	0,594 - 0,686	0,591	0,017	0,557 - 0,624	867	821	510	1.178	1.352	29,936	1.302 - 1.421
Cotinguiba	0,745	0,026	0,692 - 0,795	0,673	0,021	0,630 - 0,713	483	448	325	606	648	11,283	631 - 676
Japaratuba*	0,684	0,028	0,627 - 0,735	0,587	0,021	0,545 - 0,628	550	493	319	724	805	18,210	776 - 849
Baixo Cotinguiba	0,792	0,016	0,758 - 0,822	0,742	0,014	0,713 - 0,769	945	901	704	1.142	1.194	11,242	1.176 - 1.221
Aracaju	0,847	0,005	0,837 - 0,856	0,749	0,005	0,739 - 0,759	7.295	6.590	5.473	8.412	8.614	19,586	8.579 - 8.656
Boquim*	0,781	0,013	0,756 - 0,805	0,649	0,012	0,626 - 0,672	1.652	1.434	1.073	2.013	2.114	16,056	2.087 - 2.151
Estância	0,863	0,011	0,840 - 0,882	0,715	0,012	0,690 - 0,739	1.332	1.136	953	1.515	1.544	7,188	1.533 - 1.562

* As microrregiões sombreadas são as que tiveram mais de 5% dos registros no RC com o número da DN em branco, o que prejudicou o pareamento

Tabela 34 – Estimativas da cobertura do SINASC (\hat{C}) e do sub-registro (\widehat{SR}) no Registro Civil (RC) de acordo com o modelo $\{p(g) c(g + idmae + g*idmae)\}$ e pareamento mais restrito. Sergipe, segundo a microrregião de residência da mãe, segundo e terceiro trimestres de 2006.

MICRORREGIÃO	\hat{N}	SINASC			RC		
		n_A	\hat{C}	VARIAÇÃO*	n_B	\widehat{SR}	VARIAÇÃO*
Sergipana do Sertão do São Francisco	2.205	1.678	76,10	74,68 - 77,22	1359	38,37	37,46 - 39,52
Carira	683	549	80,33	77,98 - 81,70	479	29,91	28,72 - 31,96
Nossa Senhora das Dores	705	564	79,96	77,69 - 81,39	493	30,10	28,86 - 32,09
Agreste de Itabaiana	1.818	1.498	82,38	81,24 - 83,18	1399	23,07	22,32 - 24,13
Tobias Barreto**	1.383	976	70,60	68,20 - 72,40	833	39,75	38,20 - 41,79
Agreste de Lagarto**	2.087	1.086	52,03	48,31 - 55,21	868	58,42	55,87 - 61,39
Propriá	1.352	867	64,13	61,01 - 66,59	821	39,27	36,94 - 42,22
Cotinguiba	648	483	74,53	71,45 - 76,55	448	30,87	29,00 - 33,73
Japaratuba**	805	550	68,36	64,78 - 70,88	493	38,72	36,47 - 41,93
Baixo Cotinguiba	1.194	945	79,15	77,40 - 80,36	901	24,53	23,38 - 26,21
Aracaju	8.614	7.295	84,69	84,28 - 85,03	6590	23,50	23,18 - 23,87
Boquim**	2.114	1.652	78,15	76,80 - 79,16	1434	32,16	31,29 - 33,33
Estância	1.544	1.332	86,26	82,28 - 86,89	1136	26,43	25,90 - 27,27

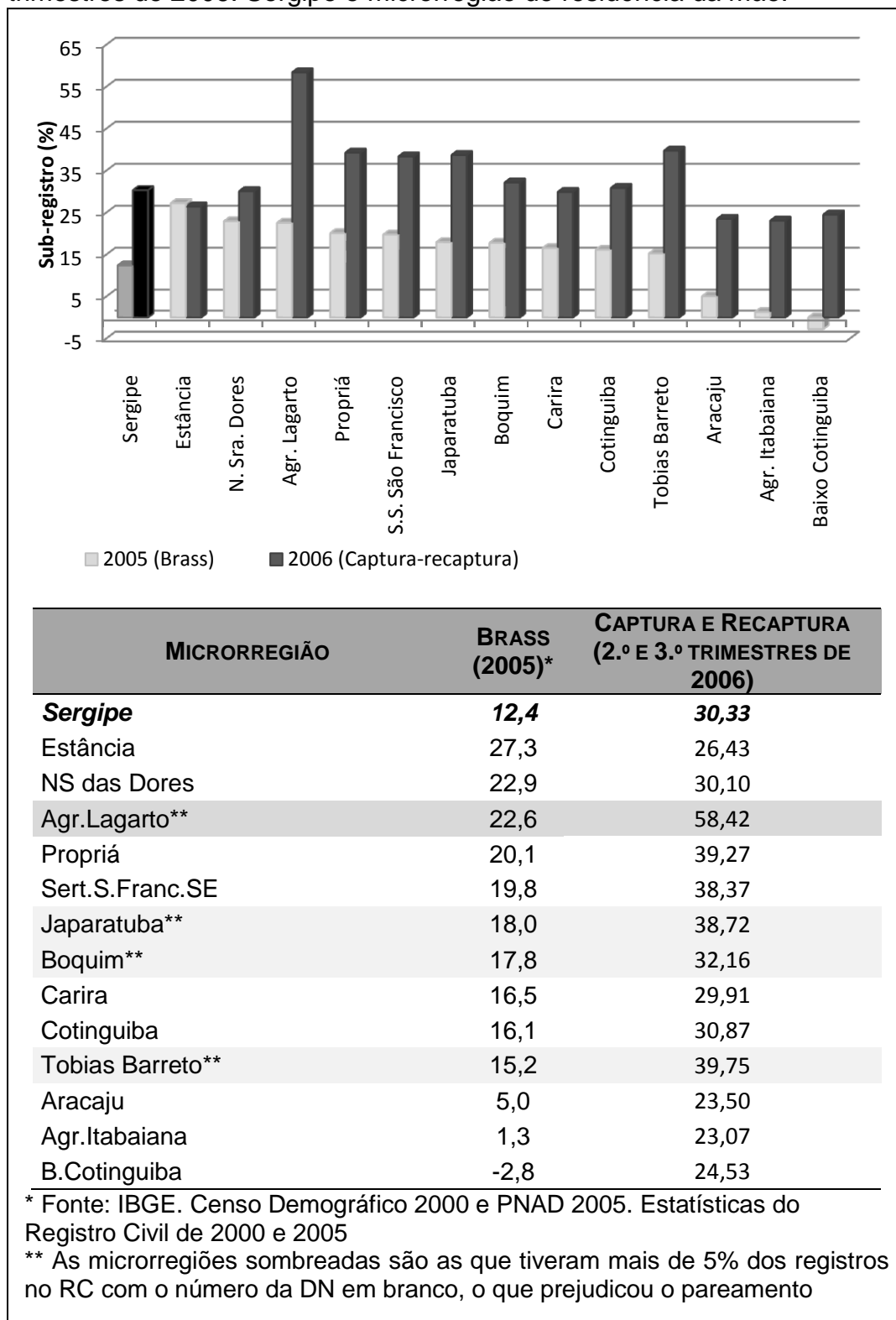
* As variações apresentadas são aproximações, obtidas a partir dos limites do intervalo de confiança de \hat{N} .

** As microrregiões sombreadas são as que tiveram mais de 5% dos registros no RC com o número da DN em branco, o que prejudicou o pareamento

Como conseqüência das relativas baixas probabilidades estimadas de captura pelo SINASC, a cobertura deste sistema também surge mais limitada. Apenas em Aracaju, Estância, Baixo Contiguiba e Agreste de Itabaiana a cobertura supera os 80%. Inversamente, no Registro Civil, o sub-registro estimado é consideravelmente alto, maior que 20% em todas as microrregiões. Em ambos os sistemas de informação, as microrregiões mais problemáticas são as que tem alto percentual de registros com DN em branco no Registro Civil, Agreste de Lagarto, com cobertura do SINASC de apenas 52% e sub-registro civil de 58%, além de Tobias Barreto, Japaratuba e Boquim (Tabela 34).

As diferenças entre as estimativas para as microrregiões obtidas pelo método de captura e recaptura com o pareamento mais restrito e o método de Brass são ainda mais dilatadas. Com exceção de Estância, em todas as microrregiões, o sub-registro calculado pelo método de captura-recaptura é maior do que o obtido pelo método de Brass. Inclusive, no Agreste de Lagarto, Japaratuba, Tobias Barreto, Aracaju, Agreste de Itabaiana e Baixo Contiguiba os valores estimados pela captura-recaptura chegam a ser mais do que dobro do que aqueles calculados pelo método de Brass (Figura 32).

Figura 32 – Estimativas do percentual de sub-registro civil de nascidos vivos pelos métodos de Brass, para o ano de 2005 e pela captura e recaptura segundo o pareamento mais restrito, para o segundo e terceiro trimestres de 2006. Sergipe e microrregião de residência da mãe.



5.2.6. REGIONAIS DE SAÚDE

Nove dos trinta modelos ajustados tiveram algum peso ponderado por w_i diferente de zero (Tabela 35). A ponderação dos parâmetros, que pode ser vista no Anexo IX, não alterou os valores das estimativas do modelo com maior peso, $\{p(g) c(g + idmae + g*idmae)\}$. Este modelo foi utilizado para estimação do sub-registro civil e da cobertura pelo SINASC.

A probabilidade de captura pelo SINASC nas regionais de saúde de residência da mãe foi em torno de 0,80, sendo menores em Lagarto, Nossa Senhora da Glória e Propriá (Tabela 36). Lagarto, além ter problemas no pareamento, devido ao alto percentual de registros com DN em branco, no RC, apresentou alta discrepância no número de pares, ao se adicionarem os critérios para o pareamento, além do número da DN. Nossa Senhora da Glória e Propriá também mostraram diferença maior do que 20% no número de pares, quando se comparam os critérios de pareamento (Tabela 14). Em relação ao Registro Civil, as probabilidades de captura são ainda menores, não havendo nenhuma regional de saúde que tenha atingido probabilidade de 0,80. Novamente, as menores estimativas estão em Lagarto, Nossa Senhora da Glória e Propriá, 0,45, 0,57 e 0,59, respectivamente. A maior estimativa de probabilidade foi em Aracaju, 0,77 (Tabela 36).

As estimativas de cobertura do SINASC nas regionais de saúde variaram entre 61% em Lagarto e 84% em Nossa Senhora do Socorro. Propriá e Nossa Senhora da Glória apresentaram valores de cobertura menores do que 80%, como consequência da baixa estimativa da probabilidade de captura. Já o Registro Civil teve estimativas de sub-registro mais elevadas, com quase 50% em Lagarto e quase 40% em Nossa Senhora da Glória e Propriá, devido tanto ao alto percentual de registros sem o número da DN, no caso de Lagarto, quanto à grande diferença no número de pares, segundo o tipo de pareamento. Em Aracaju, o sub-registro estimado foi cerca de 22%, enquanto que nas demais regionais de saúde (Nossa Senhora do Socorro, Estância e Itabaiana) o sub-registro estimado variou de pouco mais de 24% até quase 30% (Tabela 37).

Tabela 35 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (Δ AIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento mais restrito, segundo a regional de saúde de residência da mãe (continua).

Modelo	AIC	ΔAIC	w_i	Verossimilhança do modelo	Nº Parâmetros
{p(g) c(g + idmae + g*idmae)}	45268,754	0,00	0,67644	1,0000	21
{p(g + idmae) c(g + idmae + g*idmae)}	45270,381	1,63	0,29975	0,4431	22
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	45275,588	6,83	0,02219	0,0328	28
{p(g) c(g)}	45282,447	13,69	0,00072	0,0011	14
{p(g) c(g + idmae)}	45283,706	14,95	0,00038	0,0006	15
{p(g + idmae) c(g)}	45284,075	15,32	0,00032	0,0005	15
{p(g + idmae) c(g + idmae)}	45285,335	16,58	0,00017	0,0003	16
{p(g + idmae + g*idmae) c(g)}	45289,281	20,53	0,00002	0,0000	21
{p(g + idmae + g*idmae) c(g + idmae)}	45290,538	21,78	0,00001	0,0000	22
{p(.) c(g + idmae + g*idmae)}	45675,320	406,57	0,00000	0,0000	15
{p(idmae) c(g + idmae + g*idmae)}	45677,322	408,57	0,00000	0,0000	16
{p(.) c(g)}	45689,018	420,26	0,00000	0,0000	8
{p(.) c(g + idmae)}	45690,277	421,52	0,00000	0,0000	9
{p(idmae) c(g)}	45691,019	422,27	0,00000	0,0000	9
{p(idmae) c(g + idmae)}	45692,278	423,52	0,00000	0,0000	10
{p = c(g+ idmae+ g*idmae)}	45784,424	515,67	0,00000	0,0000	14
{p = c(g)}	45797,402	528,65	0,00000	0,0000	7
{p = c(g+ idmae)}	45799,158	530,40	0,00000	0,0000	8
{p(g) c(idmae)}	46225,322	956,57	0,00000	0,0000	9
{p(g + idmae) c(idmae)}	46226,950	958,20	0,00000	0,0000	10
{p(g) c(.)}	46227,622	958,87	0,00000	0,0000	8
{p(g + idmae) c(.)}	46229,249	960,50	0,00000	0,0000	9

Tabela 35 – Estimativa dos critérios de informação de Akaike (AIC); das diferenças entre o AIC do modelo, em relação ao modelo com menor critério de informação (Δ AIC); pesos normalizados de AIC; verossimilhança e número de parâmetros dos modelos de Huggins para populações fechadas para os nascidos vivos no segundo e terceiro trimestres de 2006 de Sergipe, pareamento mais restrito, segundo a regional de saúde de residência da mãe (conclusão).

Modelo	AIC	ΔAIC	w_i	Verossimilhança do modelo	Nº Parâmetros
{p(g + idmae + g*idmae) c(idmae)}	46232,151	963,40	0,00000	0,0000	16
{p(g + idmae + g*idmae) c(.)}	46234,450	965,70	0,00000	0,0000	15
{p(.) c(idmae)}	46631,895	1363,1	0,00000	0,0000	3
{p(idmae) c(idmae)}	46633,895	1365,1	0,00000	0,0000	4
{p(.) c(.)}	46634,195	1365,4	0,00000	0,0000	2
{p(idmae) c(.)}	46636,196	1367,4	0,00000	0,0000	3
{p = c(idmae)}	47102,538	1833,7	0,00000	0,0000	2
{p = c(.)}	47103,688	1834,9	0,00000	0,0000	1

Tabela 36 – Estimativas da probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) do modelo $\{p(g) c(g + idmae + g*idmae)\}$, distribuição da captura pelo SINASC (n_A), RC (n_B), sobreposição das bases (n_{11}) e total de nascidos vivos distintos capturados pelas duas bases ($r = n_A + n_B - n_{11}$). Sergipe, segundo a regional de saúde de residência da mãe, pareamento mais restrito, segundo e terceiro trimestres de 2006.

REGIONAL DE SAÚDE	P (PROBABILIDADE DE CAPTURA PELO SINASC)			C (PROBABILIDADE DE CAPTURA PELO RC)			TOTAL DE NASCIDOS VIVOS						
	\hat{p}	ERRO-PADRÃO	IC 95%	\hat{c}	ERRO-PADRÃO	IC 95%	n_A (SINASC)	n_B (RC)	n_{11}	r	\hat{N}	ERRO-PADRÃO	IC 95%
Aracaju	0,834	0,005	0,824 - 0,845	0,766	0,005	0,756 - 0,776	6.793	6.332	5.207	7.918	8.141	21,166	8.104 - 8.187
N.Sra. Socorro	0,842	0,008	0,825 - 0,857	0,691	0,009	0,674 - 0,708	2.789	2.369	1.928	3.230	3.313	12,677	3.291 - 3.342
Estância	0,827	0,009	0,809 - 0,844	0,677	0,009	0,658 - 0,695	2.457	2.088	1.664	2.881	2.969	13,531	2.947 - 3.000
Itabaiana	0,814	0,011	0,793 - 0,834	0,735	0,010	0,715 - 0,754	1.976	1.818	1.451	2.343	2.427	13,562	2.404 - 2.458
Lagarto	0,611	0,015	0,581 - 0,641	0,446	0,010	0,426 - 0,467	2.336	1.952	1.044	3.244	3.821	59,240	3.716 - 3.949
N.Sra. Glória	0,761	0,013	0,734 - 0,786	0,569	0,012	0,545 - 0,593	1.678	1.359	958	2.079	2.205	18,811	2.173 - 2.247
Propriá	0,659	0,018	0,623 - 0,693	0,589	0,013	0,563 - 0,615	1.446	1.336	842	1.940	2.196	34,683	2.137 - 2.273

*A regional sombreada teve 21,31% dos registros no RC com o número da DN em branco, o que prejudicou o pareamento.

Tabela 37 – Estimativas da cobertura do SINASC (\hat{C}) e do sub-registro (\widehat{SR}) no Registro Civil (RC) de acordo com o modelo $\{p(g) c(g + idmae + g*idmae)\}$ e pareamento mais restrito. Sergipe, segundo a regional de saúde de residência da mãe, segundo e terceiro trimestres de 2006.

REGIONAL DE SAÚDE	\hat{N}	SINASC			RC		
		n_A	\hat{C}	VARIAÇÃO*	n_B	\widehat{SR}	VARIAÇÃO*
Aracaju	8.141	6.793	83,44	82,97 - 83,82	6.332	22,22	21,87 - 22,66
Nossa Senhora do Socorro	3.313	2.789	84,19	83,45 - 84,75	2.369	28,49	28,02 - 29,11
Estância	2.969	2.457	82,74	81,90 - 83,37	2.088	29,68	29,15 - 30,40
Itabaiana	2.427	1.976	81,43	80,39 - 82,20	1.818	25,08	24,38 - 26,04
Lagarto**	3.821	2.336	61,13	59,15 - 62,86	1.952	48,92	47,47 - 50,57
Nossa Senhora da Glória	2.205	1.678	76,10	74,68 - 77,22	1.359	38,37	37,46 - 39,52
Propriá	2.196	1.446	65,85	63,62 - 67,66	1.336	39,16	37,48 - 41,22

* As variações apresentadas são aproximações, obtidas a partir dos limites do intervalo de confiança de \hat{N} .

** A regional sombreada teve 21,31% dos registros no RC com o número da DN em branco, o que prejudicou o pareamento.

6. DISCUSSÃO

A implantação do Sistema Nacional de Informação sobre Nascidos Vivos, SINASC, buscou padronizar e compilar os dados exigidos por lei e outros importantes para o conhecimento do perfil epidemiológico dos nascidos vivos. Com este objetivo, elaborou-se um documento básico e padronizado para todo o país, a Declaração de Nascido Vivo, DN, de cujas informações são possíveis se obter coeficientes de mortalidade infantil específicos, uma vez que é possível analisar em conjunto variáveis da DN e as variáveis da Declaração de Óbito. O SINASC também propicia o conhecimento do perfil do recém-nascido de alto risco, permitindo a identificação deste problema e fornecendo ferramentas para o seu estudo e elaboração de políticas públicas para minimizar a ocorrência de óbitos neonatais, aumentando, assim, a sobrevivência dos recém-nascidos e por consequência a esperança de vida ao nascer (Mello Jorge e colaboradores, 1992). Ainda, como mostra este trabalho, este sistema permite, inclusive, estimar o sub-registro civil.

O banco do SINASC é, de uma maneira geral, consistente, completo e de fácil acesso. Tabulações customizadas assim como o *download* de todo o banco podem ser feitos a partir do sítio na internet do DATASUS, do Ministério da Saúde (www.datasus.gov.br). O que se verificou neste trabalho vai ao encontro de outros estudos realizados sobre o SINASC na Região Nordeste. Szwarcwald (2008) recentemente avaliou a qualidade do SINASC e do SIM, verificando que, na primeira base, perto de 85% dos municípios do Nordeste apresentam informações adequadas, no que tange à cobertura de nascidos vivos, abrangendo pouco mais de 86% da população residente na Região. Costa e Frias (2009) avaliaram a completitude das variáveis da DN do estado de Pernambuco, no período 1996-2005, e verificaram grande incremento no preenchimento das variáveis relativas ao recém-nascido, à mãe e à gravidez.

As Estatísticas do Registro Civil, consolidadas pelo IBGE desde 1974, dentre tantas outras pesquisas desta Fundação, visam cumprir o Princípio

de Relevância, Imparcialidade e Igualdade de Acesso, listado como o primeiro dos Princípios Fundamentais das Estatísticas Oficiais da Comissão de Estatísticas da ONU, estabelecidos em Sessão Especial em abril de 1994 (UNSD, 2006). As Estatísticas do Registro Civil “constituem um importante instrumento no acompanhamento da evolução populacional no País, proporcionando, além de estudos demográficos, subsídios para a implementação de políticas públicas e o monitoramento do exercício da cidadania” (IBGE, 2006).

Dentre os usos destas Estatísticas, destaca-se seu emprego para o cálculo das Projeções e Estimativas Populacionais realizadas pelo IBGE, cujos valores são entregues ao Tribunal de Contas da União, responsável pela definição da distribuição do Fundo de Participação dos Municípios (IBGE, 2008).

Vários indicadores podem ser calculados a partir das Estatísticas do Registro Civil, como taxas de fecundidade, coeficientes de mortalidade e esperança de vida ao nascer, por exemplo. Entretanto, o sub-registro dos eventos vitais, especialmente o de nascidos vivos, em determinadas unidades da federação, é um obstáculo aos cálculos diretos destes indicadores. A busca do diagnóstico desta realidade de baixa cobertura, objetivo do presente trabalho, visa justamente subsidiar ações que permitam reverter esta situação, de maneira a colaborar no cumprimento dos Princípios de Padrões Profissionais e de Eficiência, dentre os Fundamentais das Estatísticas Oficiais da ONU.

O banco do Registro Civil, apesar do preenchimento falho do número da DN nos primeiros trimestres de implantação deste campo nos questionários do IBGE, também se mostrou com boa completude nas demais variáveis, e por tratar-se de um banco nacional, a melhora na sua qualidade viabiliza estudos válidos sobre o sub-registro civil.

O relacionamento entre bases de dados oficiais tem despertado interesse dos pesquisadores na última década. Relacionamentos determinísticos e probabilísticos foram desenvolvidos com o intuito de se agregarem informações presentes em bases de dados distintas,

aumentando as potencialidades de análise da realidade experimentada pela população (Coutinho e Coeli, 2006; Ferrari e colaboradores, 2006; Drummond e colaboradores, 2008; Nascimento e colaboradores, 2008). Adicionar o banco do Registro Civil a estes relacionamentos pode agregar ainda mais para o conhecimento desta realidade.

Destaca-se aqui a potencial factibilidade da associação entre as bases do IBGE e do SINASC, desde que resolvidas as falhas no preenchimento do número da DN, no Registro Civil. O IBGE também incluiu no questionário de óbitos o campo para o número da Declaração de Óbito (IBGE, 2006), o que implica na possível associação com o SIM. Estas viáveis ligações entre grandes bases nacionais acabam por potencializar o avanço na qualidade dos dados oficiais no Brasil, pois inúmeros estudos podem ser desenvolvidos para a identificação não só do sub-registro de eventos vitais, como também na investigação de suas causas, com vistas à sua, quiçá, erradicação. O relacionamento entre o RC e o SINASC permite ações localizadas, porque acaba por identificar onde e quantas vidas mantêm-se no anonimato jurídico, devido ao sub-registro e registro tardio.

Um exemplo de potencial estudo sobre os fatores associados ao registro tardio do nascimento vivo foi desenvolvido pela autora da presente tese. Schmid (2009) identificou o registro tardio de nascidos vivos de mães residentes em Sergipe, no segundo semestre de 2006, através da diferença entre a data do registro e a data de nascimento dos nascidos vivos na base do IBGE, o que permitiu a determinação dos registros civis dentro e fora do período legal. Com os pares formados pelo relacionamento entre o SINASC e o Registro Civil através do número da DN, foi possível investigar, por regressão logística, a associação entre o registro tardio e o sexo do nascido vivo, baixo peso ao nascer, escolaridade da mãe, idade da mãe e número de filhos tidos anteriormente, segundo a microrregião de residência da mãe.

Como resultado, verificou-se que a alta escolaridade da mãe é potencial fator de proteção para o registro tardio em quase todas as microrregiões, e que em algumas microrregiões, crianças com baixo peso ao nascer apresentam maiores chances para serem registrados após o período

legal, assim como filhos de mulheres com 2 ou mais filhos tidos anteriormente são registrados tardiamente (Schmid, 2009). Outro tipo de estudo que pode ser desenvolvido é a análise de sobrevivência, onde o tempo para o registro civil é a variável de desfecho.

A aplicação do método de captura e recaptura, empregando as duas bases de dados consideradas, revelou-se viável e de simples execução. Entretanto, algumas considerações sobre os pressupostos assumidos para a estimação pela captura e recaptura merecem ser tecidas. Toma-se como uma população fechada aquela onde não há migração nem nascimentos ou mortes durante o período de estudo. Neste trabalho, a população estimada foi o número total de nascidos vivos. O evento nascer é único e número de nascidos vivos é fixo no período e na delimitação geográfica adotada. Claramente, óbitos neonatais e/ou infantis ocorreram no período analisado, assim como as famílias das crianças podem ter mudado a residência para outro município ou unidade da federação, após o parto. Porém, estas ocorrências não alteram o tamanho da “população de nascidos vivos de mães residentes em Sergipe”, uma vez que óbito e mudança de endereço não podem modificar o fato de que, naquele momento, a criança nasceu viva e a mãe residia naquela localidade.

Em relação à marcação unívoca e inequívoca, as falhas no preenchimento do número da DN prejudicaram o relacionamento determinístico adotado. Dos 17.254 nascidos vivos presentes no Registro Civil de Sergipe, no segundo e terceiro trimestres de 2006, 808, ou 4,7%, tinham o campo DN em branco. Como dito na seção de Material e Métodos, o primeiro trimestre tinha quase a totalidade dos registros da base do IBGE com a DN não preenchida. Já o segundo trimestre tinha cerca de 12% de registros com DN em branco, ao passo que nos dois últimos trimestres, estas proporções eram menores do que 2%. O pareamento menos restrito foi realizado entre os nascidos vivos do segundo e terceiro trimestre e dos 17.254 constantes no Registro Civil, 15.532 foram pareados com o SINASC e 1.722 figuravam somente na base do IBGE (Tabela 15). Destes, 1.212 referem-se a nascidos vivos no segundo trimestre, e 719 dos registros não

tinham o número da DN. Estes resultados sugerem, então, que se os números da DN estivessem preenchidos, o número de pares seria maior, especialmente para os nascidos no segundo trimestre. Entretanto, resta a pergunta: Quantos, dos 808 registros no RC sem o número da DN são pareáveis com os registros do SINASC?

Tomando como ponto de partida os resultados do modelo para todo o Estado de Sergipe, segundo o pareamento menos restrito, da Tabela 21, tem-se que a probabilidade estimada de captura pelo SINASC é de 0,912 e a do Registro Civil é de 0,804. A probabilidade de um nascido vivo figurar nas duas bases, então, será, $0,912 * 0,804 = 0,733$. Logo, dos 808 registros sem número da DN, no Registro Civil, $0,733 * 808 = 592$ também teriam sido capturados pelo SINASC. A partir desta conta, sugere-se que o número correto de pares formados pelo relacionamento entre as duas bases seria $15.532 + 592 = 16.124$ nascidos vivos. Apesar de ser possível estimar o provável número de pares, realizar tal relacionamento pós-estimação torna-se tarefa árdua e imprecisa, pois como argumentado na subseção Relacionamento Probabilístico, as variáveis disponíveis tem pouco poder discriminatório, o que implica em se encontrar mais de um registro com as mesmas características.

Outro ponto a ser abordado é o estreitamento dos critérios para se considerar um par. Dentre os pareados pelo número da DN, 75% coincidiram em todas as variáveis possíveis de comparação. Noventa e nove pares resultantes do relacionamento pelo número da DN tinham alguma variável tida como ignorada em uma das bases, 71 com idade da mãe ignorada no RC e 28 do SINASC sem informação no tipo de gestação, o que impediu de classificá-los como “pares reais” (Quadro 1). Estes 99 pares de registros foram classificados como “muito prováveis”, juntamente com 1.334 onde somente o município de residência da mãe ou o município de nascimento eram discordantes nas duas bases (Tabela 6). Estes 1.334 registros representam 8,54% dos 15.612 pares formados pelo relacionamento pelo número da DN. É interessante sublinhar que este tipo de divergência já fora abordado anteriormente, por Waldvogel e Ferreira (2003) e por Oliveira e

Simões (2005), e aqui é possível identificar e quantificar tal divergência, no Estado de Sergipe. Waldvogel e Ferreira (2003) destacam que o vício da informação do endereço tende a ser menor no Registro Civil, pois, em caso de divergência, o cartório exige explicação por escrito. Por esta razão, nos casos de discordância no município de residência da mãe, considerou-se a informação do Registro Civil, para fins de estimação segundo as desagregações geográficas. Este procedimento explica, também, a variação do número de nascidos vivos utilizados na base do SINASC, de acordo com o critério de pareamento. Ao se restringir as condições para determinado par ser considerado “real” ou “muito provável”, alguns registros classificados como “não par” do SINASC acabaram excluídos da análise, pois o município de residência da mãe constava como fora do Estado do Sergipe.

Em sua tese de doutoramento, Machado (2002) realizou relacionamento probabilístico entre o SIM, para os óbitos infantis daqueles nascidos na coorte de 1998, com seus respectivos registros no SINASC. As variáveis utilizadas para o relacionamento probabilístico foram sexo, peso ao nascer, data de nascimento, idade da mãe, município de residência da mãe, tipo de parto e tipo de gestação. Como método de relacionamento, os registros foram comparados em cada variável e códigos foram atribuídos, de acordo com a concordância: total, parcial ou discordância. Em etapas seguintes, a comparação por blocos e regras de relacionamento foram adotadas e não serão detalhadas. O que se destaca, aqui, são os critérios para codificação adotados pela autora. Para Machado, a divergência de um dia na data de nascimento foi considerada “datas muito próximas” e divergência de dois dias nas datas foi tida como “datas próximas”. No presente trabalho, pares onde havia apenas a diferença de um dia na data de nascimento, aliada à concordância nas demais variáveis, foram avaliados como “pares prováveis” e não foram levados em consideração no relacionamento mais restrito. O mesmo ocorreu com a idade da mãe. Para Machado, a divergência de um ano na idade da mãe foi dada como “idades muito próximas”, ao passo que divergência de dois anos resultou em “idades próximas”. Novamente, aqui a disparidade de um ano na idade da mãe

definiu o par como “provável” e este não foi considerado no relacionamento mais restrito. Esta diferença nos critérios se justifica pela presença do número da DN como identificador único, o que não ocorreu no trabalho de Machado. Estes critérios mais restritos para o relacionamento se baseiam na hipotética e desejável concordância total das variáveis entre os registros com mesmo número da DN, em consonância com o relacionamento determinístico dos bancos.

Mais uma vez ressalta-se, aqui, que o ano de 2006 foi o primeiro de coleta do número da DN nos questionários do IBGE, e espera-se que nos anos subseqüentes haja melhora na qualidade dos dados, com menores discrepâncias entre a base do SINASC e do Registro Civil. De uma maneira geral, nota-se que pouco mais de 84% dos pares podem ser considerados efetivamente como pares. Em torno de 75% coincidem em todas as variáveis e os demais quase 9% referem-se basicamente ao vício de informação do local de residência da mãe ou do local de ocorrência do nascimento. Este vício sempre estará presente, e não pode ser atribuído como erro de preenchimento do questionário do IBGE, ao se lembrar dos argumentos de Waldvogel e Ferreira (2003) e Oliveira e Simões (2005).

No que concerne à equiprobabilidade, isto é, cada indivíduo tem a mesma probabilidade de ser capturado em uma dada amostra, enfrentam-se duas questões. A primeira refere-se à influência da idade da mãe na captura pelo Registro Civil. Entretanto, esta “fonte” de heterogeneidade pode ser incluída nos modelos e, de fato, mostrou-se significativa. A segunda remete ao intervalo de tempo tomado para o pareamento das bases. O banco de dados do Registro Civil, no IBGE, é organizado por ano de vigência da coleta dos dados. Os registros e suas respectivas datas de nascimento não são perdidos, porém constam, no banco do IBGE, no ano em que foram coletados. Desta maneira, o número do total de nascidos vivos em 2006 registrados irá variar à medida que os anos de coleta dos dados avançam. É evidente que esta variação não será muito grande, pois os registros tardios serão residuais com o passar dos anos. Como salientam Oliveira e Simões (2005), a cobertura do Registro Civil é sensivelmente incrementada quando

se analisam os registros ocorridos até um ano após o nascimento vivo. Este comportamento pode ser observado na Tabela 38, para os nascidos vivos de mães residentes em Sergipe no primeiro trimestre de 2006. Quase 92% daqueles que nasceram no primeiro trimestre foram registrados até o segundo trimestre, passando para quase 98% quando se verificam os registros do terceiro trimestre. Para aqueles que nasceram no segundo trimestre de 2006, este padrão ainda pode ser observado. A partir do terceiro trimestre, o padrão artificialmente muda, dando a equivocada impressão de que a proporção de nascidos vivos e registrados no mesmo trimestre é maior nos últimos quartos de 2006. Como o acesso aos dados ficou restrito ao ano de coleta de 2006, os nascidos nos dois últimos trimestres de 2006 e registrados a partir de 2007 em diante não foram detectados neste trabalho. Acredita-se, então, que não houve equi-probabilidade de captura pelo Registro Civil entre os indivíduos que nasceram no terceiro trimestre de 2006, em relação aos que nasceram no segundo trimestre daquele ano.

Tabela 38 – Número de nascidos vivos e número de registros civis constantes no Banco de 2006 do Registro Civil, segundo o trimestre de nascimento e trimestre de registro. Sergipe, 2006

TRIM.	NASCIDOS VIVOS (%)	REGISTROS CIVIS (%)			
		1.º TRIM.	2.º TRIM.	3.º TRIM.	4.º TRIM.
1.º	8.350 (100,00)	5.549 (66,46)	2.106 (25,22)	562 (6,73)	233 (2,79)
2.º	9.383 (100,00)	-	5.931 (63,21)	2.928 (31,21)	524 (5,58)
3.º	7.871 (100,00)	-	-	5.706 (72,49)	2.165 (27,51)
4.º	4.871 (100,00)	-	-	-	4.833 (99,22)

TOTAL DE NASCIDOS VIVOS NO BANCO DO REGISTRO CIVIL DE 2006: 30.475

Fonte: IBGE/DPE/COPIS. Acesso direto ao banco de dados do Registro Civil, junho, 2007

Ao se considerar que o padrão de registro civil dos nascidos no terceiro trimestre de 2006 será similar aos padrões dos nascidos no primeiro e segundo trimestres, pode-se afirmar uma parte dos nascidos vivos no terceiro trimestre de 2006 foram registrados em cartório no primeiro e segundo trimestres de 2007. Uma vez que o acesso ao banco do IBGE foi restrito ao ano de coleta de 2006, ressalta-se que o pareamento entre o SINASC e o RC para os nascidos vivos no terceiro trimestre apresenta mais

um revés, além da falta do número da DN. Deduz-se, então, que o número de pares, representado na tabela 2 x 2 de captura por n_{11} , será maior do que o medido, evidenciando ainda mais a grande sobreposição das duas bases.

Em relação ao SINASC, os dados são consolidados a partir do envio das informações das Secretarias Estaduais de Saúde para o Ministério da Saúde, e eventualmente também são atualizados (MS, 2009), de sorte que os números de nascidos vivos ora disponíveis no sítio da internet (www.datasus.gov.br) também diferem dos dados preliminares disponibilizados para este trabalho, em 2007. Deve-se levar em consideração o dinamismo dos dois bancos de dados, sendo sempre abordada esta contínua mudança nos totais de registros.

Recomenda-se, então, para associações futuras entre o SINASC e o Registro Civil, que as buscas por nascidos vivos sejam feitas na base do IBGE até seis meses depois da ocorrência, de maneira a minimizar o vício do registro tardio. Este vício sempre será um problema, pois, como no caso do presente trabalho, quem nasceu no primeiro dia do segundo trimestre teve mais tempo para ser capturado do que quem nasceu no último dia do terceiro trimestre. Ou se faz o pareamento dia a dia, ou acrescenta-se um termo na equação do estimador da probabilidade de captura pelo Registro Civil, que leve em consideração este tempo relativo ao registro tardio. Tanto parear dia-a-dia, quanto introduzir um fator de correção no modelo exige aprofundamento das técnicas, seja de pareamento, seja na modelagem estatística. No primeiro caso, faz-se necessária a implementação de rotinas automatizadas de associação entre as bases, pois da maneira como foi feita neste trabalho, implicaria na repetição manual das consultas de associação, uma para cada dia do período analisado. No caso de introduzir mais um termo ao modelo de probabilidade, é requisitado estudo de várias séries dos registros, de maneira a determinar os padrões mais próximos do real comportamento do registro tardio de nascidos vivos. Esse estudo permitiria estimar de maneira mais precisa e válida a distribuição dos registros ao longo do tempo e assim se desenvolver um modelo específico para caracterizar a probabilidade de captura de nascidos vivos pelo Registro Civil.

Quanto à independência entre as amostras, aqui bases de dados, a grande sobreposição entre elas (n_{11} grande) sugere que há dependência positiva. Isto significa que o número de nascidos vivos estimados não será muito maior do que aqueles indivíduos distintos identificados nas duas bases. Com o objetivo de quantificar o vício gerado pela dependência entre duas fontes (listas) epidemiológicas, Brenner (1995) incluiu um fator de correção para a dependência, obtido a partir da covariância entre as probabilidades, aplicado à probabilidade de o indivíduo ser capturado nas duas listas. O autor simula situações em que variam as probabilidades de captura em cada fonte (n_a e n_b) e os valores dos fatores correção das probabilidades de inclusão nas duas fontes (n_{11}). Estes fatores geraram dependência positiva e negativa, de maneira a observar o comportamento do fator de sub e superestimação, respectivamente, das estimativas obtidas pelo método de captura-recaptura. Para o caso da dependência negativa, o investigador conclui que o vício de superestimação do tamanho total da população será mais grave quando as listas têm baixa cobertura de indivíduos, ou seja, tem pequena probabilidade de inclusão destes. Com os achados, o autor sugere cautela para a utilização do método de captura-recaptura com apenas duas listas quando é sabido que estas têm dependência negativa. Por outro lado, o autor afirma que no caso da dependência positiva, listas com probabilidade alta de inclusão têm fatores de subestimação menores, e que mesmo conscientes deste tipo de dependência, as estimativas do tamanho da população ainda serão mais próximas da realidade do que a simples agregação das fontes (Brenner, 1995).

Neste estudo, as probabilidades estimadas de captura em ambas as bases, no estado de Sergipe, são relativamente altas, sendo o SINASC com maior probabilidade de captura, 0,91 para o pareamento menos restrito e 0,78 para o pareamento mais restrito. Já o Registro Civil teve probabilidades de captura estimadas de 0,80 e 0,67 (Tabelas 21 e 30). Os resultados aqui apresentados sugeririam valores mínimos de sub-registro civil e de cobertura do SINASC, se os demais pressupostos fossem atingidos. Não há motivos

para se suspeitar de dependência negativa entre o SINASC e o Registro Civil, tanto pela alta sobreposição entre os bancos, quanto pelas políticas adotadas na última década, de estímulo ao registro civil especialmente para partos ocorridos em hospitais (IBGE, 2005).

Frente a estas exposições, é coerente supor que os valores de sub-registro encontrados pela estimação com o pareamento menos restrito sejam os mais próximos da realidade, ainda que haja 25% de divergência entre os pares encontrados. Tal suposição fundamenta-se na grande sobreposição das bases, mesmo que prejudicada pelos registros no RC sem o número da DN e pela “censura” artificial do ano de coleta dos dados do IBGE. Ainda, a adoção de critérios muito rígidos no pareamento alternativo pode ter produzido redução artificial exagerada do número de pares, levando a uma superestimação do \hat{N} .

Por último, pode-se aventar que o pressuposto de que as probabilidades de captura e recaptura de um indivíduo não afetam as probabilidades dos demais não foi violado, porque um nascimento não deve interferir nem na identificação de outro no SINASC, tampouco no registro civil de outrem. Cabe refletir sobre os partos gemelares, onde os nascidos vivos serão registrados, ou não, em conjunto, o que não atenderia ao pressuposto em questão. Entretanto, a ocorrência de gestações gemelares é muito rara, e o número de gêmeos nascidos vivos não chega a prejudicar tal assunção.

Um pouco de palavras sobre o aplicativo empregado neste estudo devem ser mencionadas. O Programa MARK[®] foi desenvolvido para uso em estimativas advindas de experimentos com populações de animais, que geralmente envolvem amostras pequenas. Por consequência, sua velocidade para os cálculos limita-se a bancos de dados pequenos. O aplicativo realizou todas as estimativas solicitadas, porém o tempo para a execução das tarefas foi demasiado longo. Por exemplo, para o modelo completo nas microrregiões, $[p(g + idmae + g*idmae) c(g + idmae + g*idmae)]$, que contava com 13 grupos, o tempo de execução foi de 91,54 minutos. O cálculo em modelos mais simples era relativamente mais rápido,

porém, deve-se recordar que, para cada um dos quatro bancos de dados gerados, 30 modelos foram ajustados. Desta maneira, o tempo de execução, em cada banco, foi em média de 4 a 5 horas. Soma-se a esta demora o fato de que não era possível interromper a estimação para retomá-la depois, pois o arquivo interno gerado pelo aplicativo tornava-se demasiado grande para visualização posterior, ficando a capacidade de recuperação do MARK[®] aquém do arquivo gerado.

No presente trabalho, estudou-se a associação entre duas bases relativamente pequenas, quando comparadas às do Estado de São Paulo, Minas Gerais ou Bahia, por exemplo. Desta forma, para que o emprego do aplicativo MARK[®] seja viável em futuras estimações, faz-se mister que seja estabelecida colaboração entre os desenvolvedores deste programa com os interessados na implementação do método em nível nacional, com o objetivo de transformar o MARK[®] em um programa que suporte bancos de dados grandes. A grande vantagem deste aplicativo reside na sua interface amigável, além de sua gratuidade e manual de instrução acessível.

Este aplicativo oferece uma gama de modelos do método de captura, tanto para populações fechadas quanto para populações abertas e, apesar de ter sido elaborado para dar suporte a estudos em ecologia, o *software* merece atenção por parte dos estudiosos em outras áreas da saúde, como a Epidemiologia.

No tocante ao modelo escolhido, Tilling e Sterne (1999) e Tilling e colaboradores (2001) aplicaram o modelo de Huggins para estimação em dados epidemiológicos, mostrando a viabilidade do modelo para estes tipos de dados. Ainda, a abordagem da verossimilhança condicionada aos indivíduos observados dá a flexibilidade da inclusão de co-variáveis para modelar as probabilidades de captura, através de ajustes de modelos lineares. Em associações futuras do SINASC e do Registro Civil, uma vez resolvido o problema da equi-probabilidade de captura pela base do IBGE, no que tange ao período empregado para a identificação dos registros, enfrentado neste trabalho, acredita-se que o modelo de Huggins pode continuar a ser empregado para a estimação do total de nascidos vivos. Por

outro lado, ao optar-se por incluir um termo que aborde o registro tardio na probabilidade de captura pelo Registro Civil, novos estudos são necessários para se verificar se o modelo de Huggins ainda será adequado para a estimação.

Neste estudo, não se desenvolveu um estimador para a variância da Cobertura do SINASC, tampouco para o Sub-registro civil, de sorte que os intervalos derivados a partir do número total de nascidos vivos representam uma variação aproximada das estimativas em questão.

Para o Estado de Sergipe, a estimativa de sub-registro civil é maior pela captura e recaptura do que pelo método de Brass (Figuras 31 e 32). Cabe salientar, entretanto, que o presente estudo estimou o sub-registro em um período de seis meses do ano de 2006, ao passo que a estimativa pelo método de Brass foi feita para todo o ano de 2005. Ainda, a identificação de registros na base do IBGE restringiu-se ao ano de coleta de 2006, já discutida, o que prejudicou a captura pelo RC de parte dos nascidos vivos no terceiro trimestre de 2006, de maneira que o número total de nascidos vivos estimados pode ser menor, o que diminuiria o sub-registro e, por conseqüência, a diferença entre as estimativas dos dois métodos. Por outro lado, após esta correção do número de nascidos vivos capturados pelo RC, que geraria uma estimativa de sub-registro menor, é coerente se esperar que o comportamento do sub-registro civil estimado pela captura e recaptura não se modifique de maneira significativa ao se calcular para todo o ano, ao invés do período de seis meses.

Apesar de serem necessárias análises de mais séries de registros civis, a Tabela 38 indica que o comportamento destes tende a seguir o padrão de que a maioria dos registros civis ocorre até o fim do primeiro trimestre de nascimento, isto é, dentro do prazo legal, e não há implicações de que o sub-registro civil seja maior ou menor em determinada época do ano, de tal modo que venha a afetar significativamente a estimativa de sub-registro civil. Assim, pode-se afirmar que, para o Estado de Sergipe, a diferença entre as estimativas de sub-registro pelo método de Brass e de captura e recaptura não é grande.

A partir dos resultados ora disponíveis, é possível aludir que o método de captura e recaptura apresenta vantagem sobre o método de Brass, uma vez que não gera sub-registro negativo (Figura 19). Mesmo os limites inferiores dos intervalos de confiança para o tamanho da população não são menores do que o número total de nascidos vivos distintos observados (Tabelas 24, 27, 33 e 36). Ainda, o cálculo do sub-registro para desagregações geográficas menores do que unidade da federação pelo método de Brass representa uma aproximação, onde os pressupostos exigidos por esta técnica também não podem ser assegurados, como salientam Oliveira e Simões (2005).

Szwarcwald (2008) recorda que o cálculo das taxas de mortalidade infantil para as microrregiões do Brasil, realizado em 1998 a pedido do Ministério da Saúde (Simões, 1999), empregando métodos demográficos indiretos geraram valores superestimados, o que causou polêmica entre os gestores locais, que endereçaram Manifesto ao Ministério. No referido trabalho, Simões (1999) aplicou os métodos demográficos indiretos tanto para o cálculo do sub-registro de nascidos vivos quanto para a mortalidade infantil. As estimativas do sub-registro de nascidos vivos, comparadas com as estimativas obtidas pelo método de captura e recaptura (Figuras 31 e 32), também foram calculadas por um método demográfico indireto, Brass. O autor ressalva que as estimativas devem ser encaradas como *proxy* da real situação enfrentada nas localidades, e que “as metodologias utilizadas, ao ajustarem tendências históricas observadas e a seguir as projetarem no tempo, por certo deixam de considerar possíveis ações conjunturais” (Simões, 1999), de sorte que permite-se inferir que não há, ainda, uma metodologia correta para a estimação de sub-registro de nascidos vivos e de óbitos infantis. Não cabe aqui apontar qual é o método mais adequado para tais estimativas, exercita-se, somente, a investigação de alternativas para a mensuração da realidade de maneira válida e precisa.

Sobre as desagregações geográficas adotadas, como reza a Portaria MS/GM n.º 373, as unidades da federação tem autonomia para definir suas regionais de saúde, levando em conta as diretrizes sugeridas. As regionais

de saúde são determinadas de maneira a incrementar a eficiência da descentralização dos serviços de saúde preconizada pelo SUS, que no estado de Sergipe foram definidas em sete. Em termos de precisão das estimativas, os erros-padrão estimados tanto para as probabilidades de captura e recaptura quanto para o tamanho total da população são proporcionalmente menores nas regionais de saúde do que nas microrregiões, e podem ser conferidas nas Tabelas 24 e 27, para o pareamento menos restrito e Tabelas 33 e 36, para o pareamento mais restrito. Este ganho na precisão está associado ao maior número de observações na desagregação geográfica de regionais de saúde, em contraposição com as microrregiões. Quanto à validade das estimativas, acredita-se que o tipo de desagregação não interfere nesta propriedade, pois o que se modificou foi tão somente o arranjo dos agrupamentos, sendo aplicados os mesmos modelos e estimadores, em ambas as desagregações geográficas.

Por fim, é interessante notar o comportamento do sub-registro civil e da cobertura do SINASC, segundo algumas características das microrregiões e regionais de saúde, como o grau de urbanização e o percentual de mulheres com idade entre 10 e 49 anos, sem instrução formal.

Há nítida tendência de o sub-registro civil decrescer à medida que aumenta o grau de urbanização das microrregiões. As microrregiões do Baixo Contiguiba e de Aracaju, onde maior proporção da população reside em área urbana são as com menores estimativas de sub-registro civil, ao passo que a Sergipana do Sertão de São Francisco, com mais de 50% da população residindo em áreas rurais, apresenta alta estimativa de sub-registro civil (Figura 33). A cobertura do SINASC, apesar de revelar-se com menor amplitude ao longo das microrregiões, com um comportamento quase constante, mostra uma leve tendência a aumentar juntamente com o aumento do grau de urbanização (Figura 34).

Quando se analisam as estimativas a partir do percentual de mulheres sem instrução, com idade entre 10 e 49 anos, o sub-registro exhibe tendência visível de aumento ao se aumentar esta proporção de mulheres residentes

nas microrregiões. As microrregiões onde há menor percentual de mulheres nesta faixa etária, sem instrução, são também aquelas onde as estimativas do sub-registro civil são menores, Aracaju e Baixo Contiguiba. E a Sergipana do Sertão do São Francisco, com mais de 18% das mulheres nestas idades, sem educação formal, tem sub-registro civil alto (Figura 35). Quanto à cobertura do SINASC, não há tendência clara de aumento ou diminuição ao se variar o percentual de mulheres de 10 a 49 anos sem instrução. O SINASC representa o acesso da população aos serviços de saúde, sejam hospitais ou estabelecimentos sem internação e este achado sugere que não há seleção no atendimento quanto ao nível educacional e pode ir ao encontro do princípio de universalidade do SUS (Figura 36).

Nas regionais de saúde, o comportamento das estimativas do sub-registro civil mantém a tendência inversamente proporcional ao grau de urbanização. As regionais de Aracaju e Nossa Senhora do Socorro, com maiores graus de urbanização, tem as estimativas mais baixas de sub-registro civil. Nossa Senhora da Glória, com grau de urbanização menor do que 50%, tem estimativa de sub-registro entre as mais altas (Figura 37). Por outro lado, a leve tendência de aumento da cobertura do SINASC à medida que o grau de urbanização aumenta, observada nas microrregiões, não se repete nas regionais de saúde (Figura 38).

Figura 33 – Estimativas pontuais do sub-registro civil, com o pareamento menos restrito, nas microrregiões de residência da mãe, segundo o grau de urbanização das microrregiões. Sergipe, segundo e terceiro trimestres de 2006.

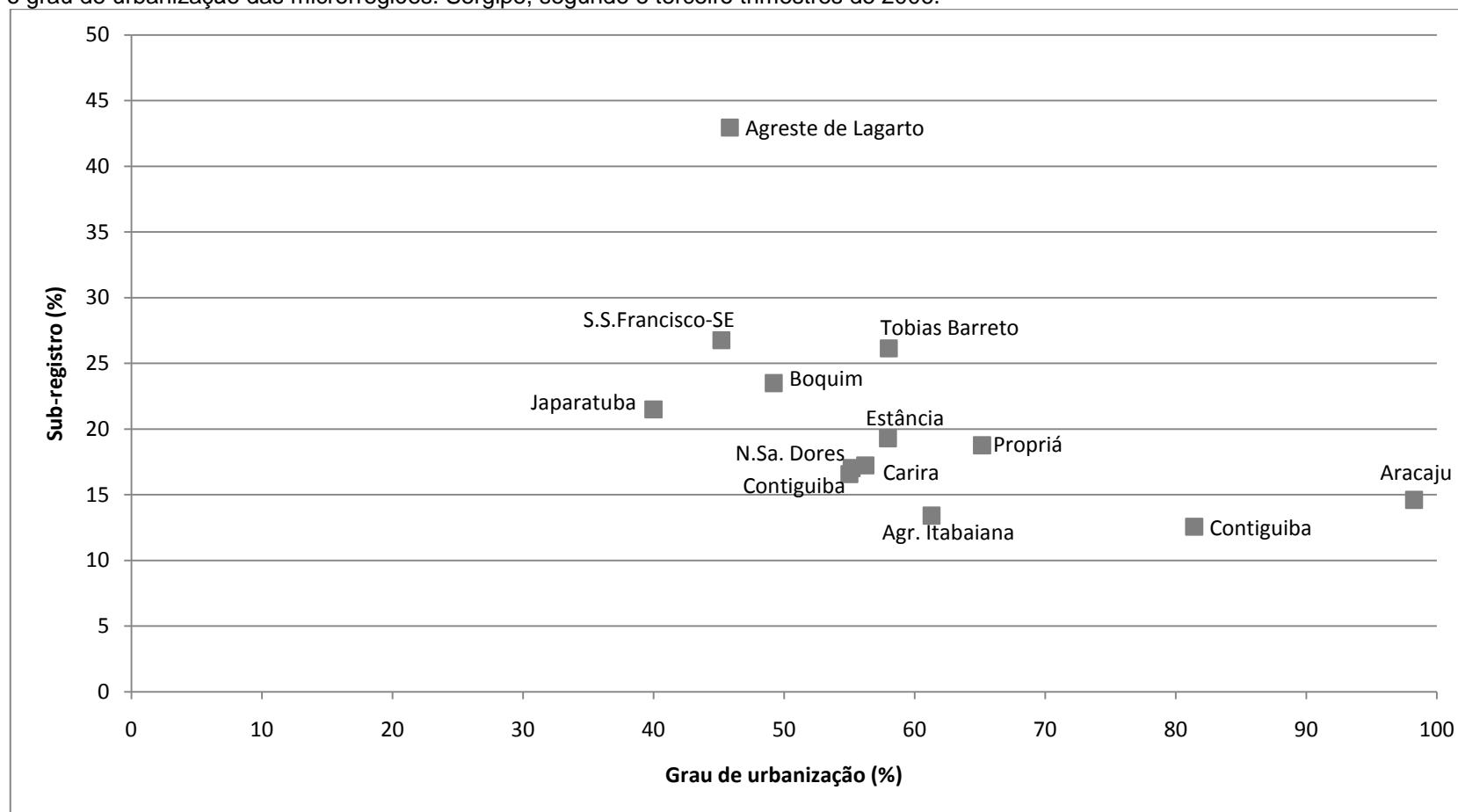


Figura 34 – Estimativas pontuais da cobertura do SINASC, com o pareamento menos restrito, nas microrregiões de residência da mãe, segundo o grau de urbanização das microrregiões. Sergipe, segundo e terceiro trimestres de 2006.

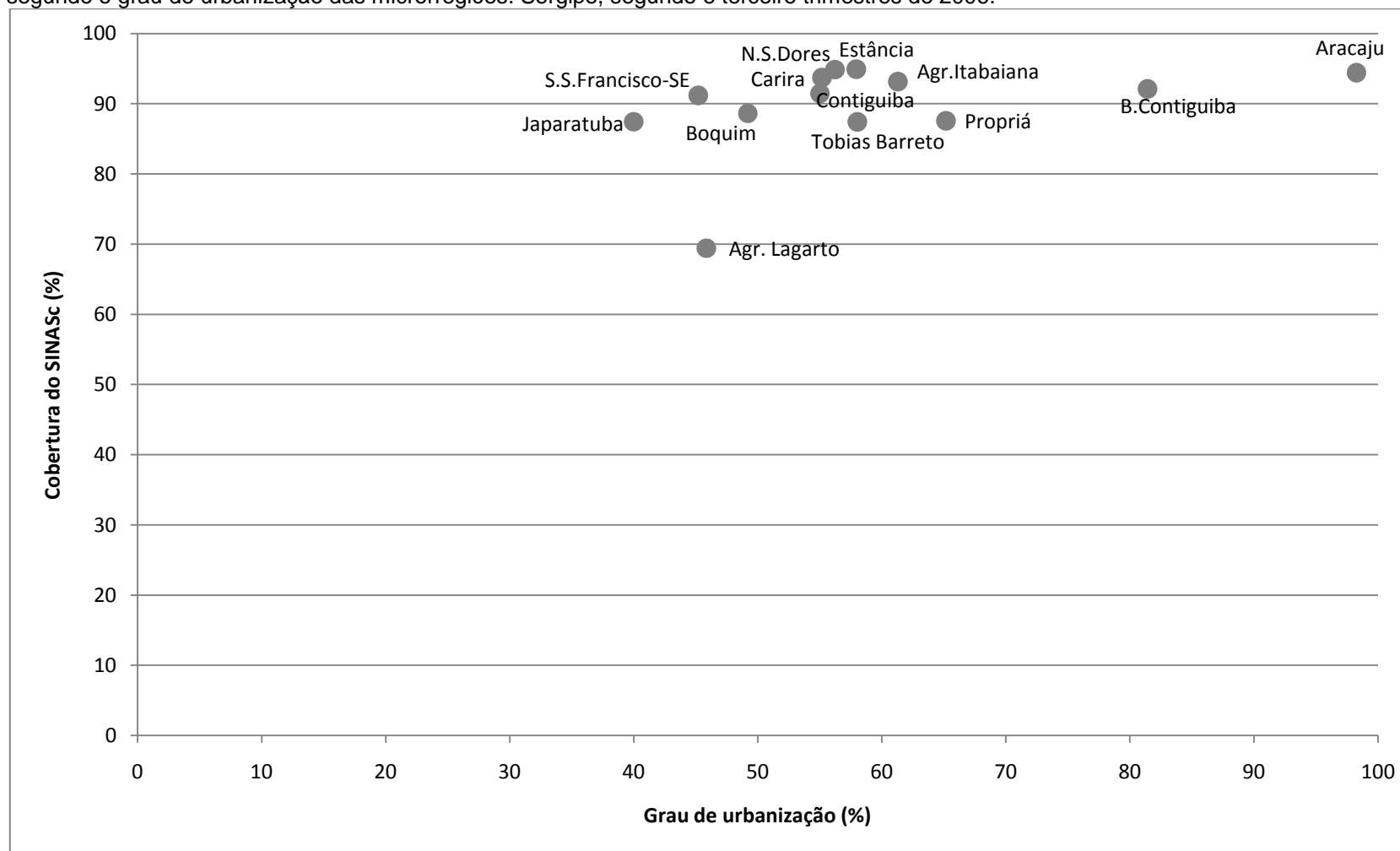


Figura 35 – Estimativas pontuais de sub-registro civil, com o pareamento menos restrito, nas microrregiões de residência da mãe, segundo o percentual de mulheres entre 10 e 49 anos sem instrução, residentes nas microrregiões. Sergipe, segundo e terceiro trimestres de 2006.

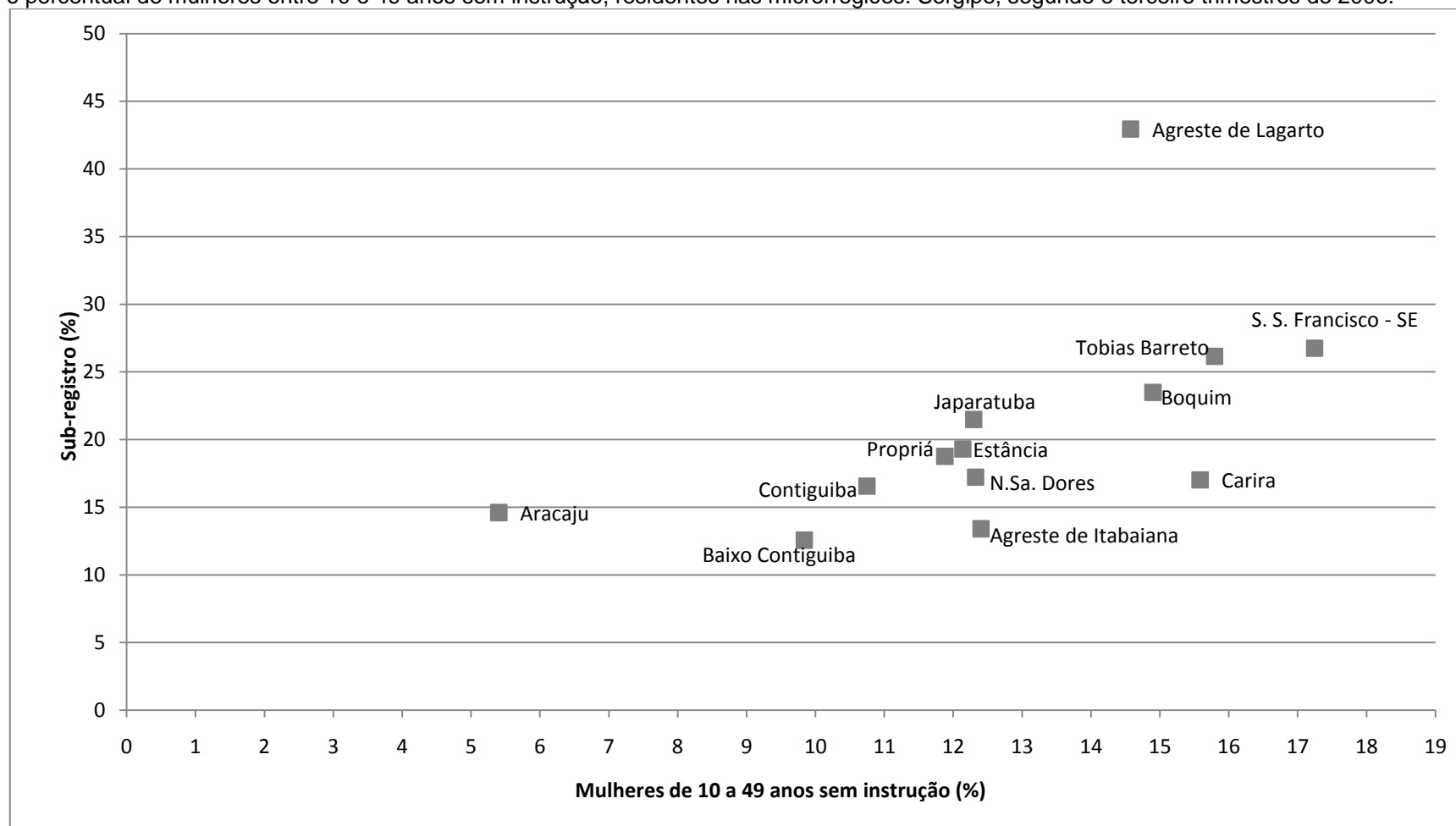
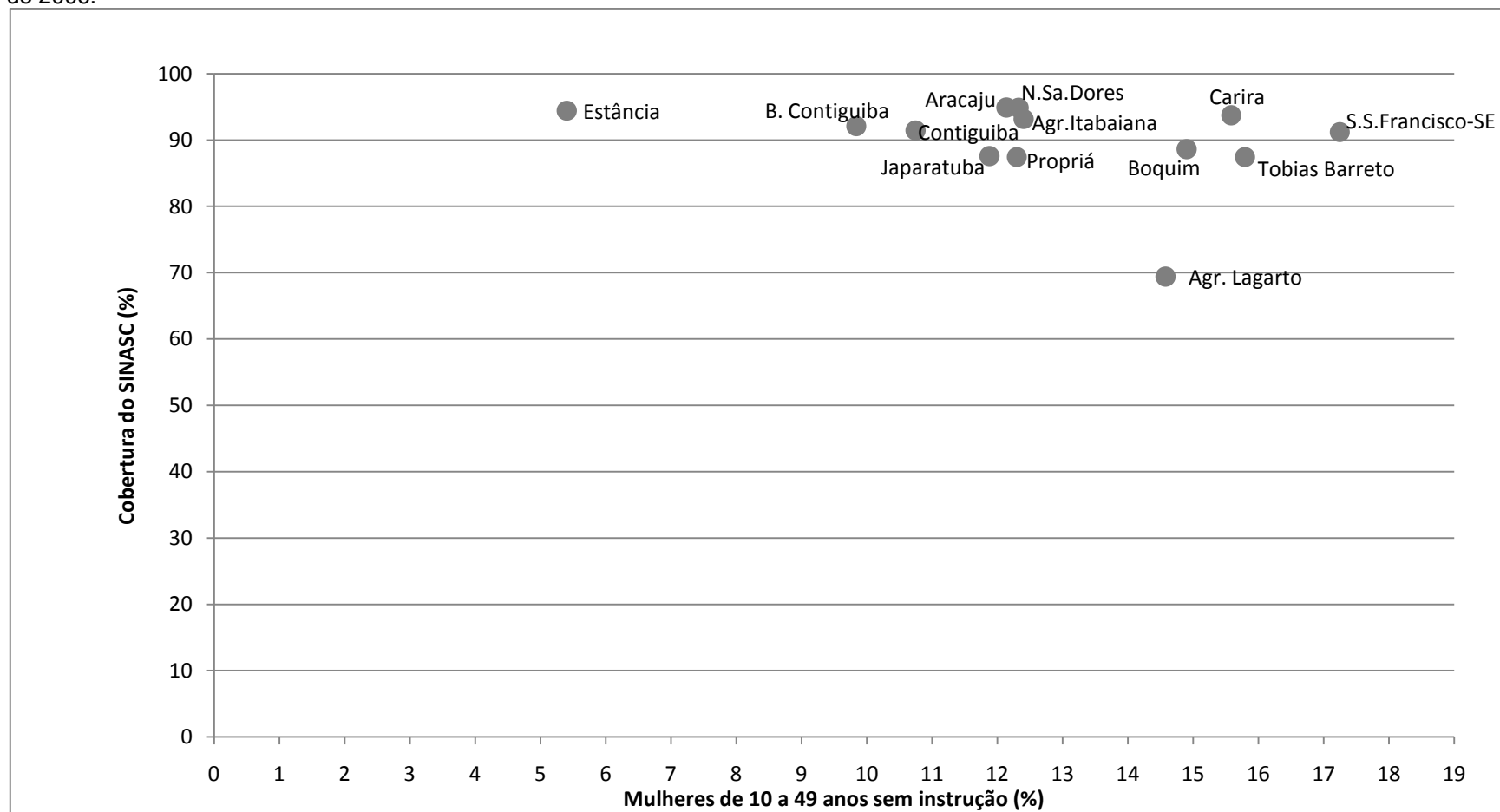


Figura 36 – Estimativas pontuais da cobertura do SINASC, com o pareamento menos restrito, nas microrregiões de residência da mãe, segundo o percentual de mulheres entre 10 e 49 anos sem instrução, residentes nas microrregiões. Sergipe, segundo e terceiro trimestres de 2006.



Também nestas desagregações geográficas, o sub-registro civil revela tendência de aumentar proporcionalmente com o percentual de mulheres sem instrução, com idade entre 10 e 49 anos, residentes nas regionais (Figura 39). A cobertura do SINASC, mais uma vez não apresenta nenhuma tendência (Figura 40).

Estas constatações apontam para a suscetibilidade do ato “registrar o nascido vivo em cartório” a fatores sócio-econômicos, como idade da mãe, escolaridade da mãe, grau de urbanização da localidade e outros não abordados aqui, revelando que não há equiprobabilidade de captura (heterogeneidade), de modo que para futuras estimações da probabilidade de captura pelo RC, na medida do possível, tais fatores devem ser levados em consideração, seja no modelo de estimação, seja por estratificação. Aparentemente, o SINASC recebe menos influências de características sócio-econômicas, pois, com exceção da desagregação geográfica de regionais de saúde, no pareamento menos restrito, cujo modelo completo foi selecionado, a idade da mãe não se mostrou influência na probabilidade de captura pelo SINASC, ao passo que para o Registro Civil, em todos os modelos selecionados a variável idade da mãe se fez presente. Em adição, na análise pós-estimação, a cobertura do SINASC aparentemente não é sensível a outros fatores sócio-econômicos locais, como o grau de urbanização e percentual de mulheres de 10 a 49 anos sem instrução, ao contrário do Registro Civil.

Figura 37– Estimativas pontuais do sub-registro civil, com o pareamento menos restrito, segundo o grau de urbanização das regionais de saúde de residência da mãe. Sergipe, segundo e terceiro trimestres de 2006.

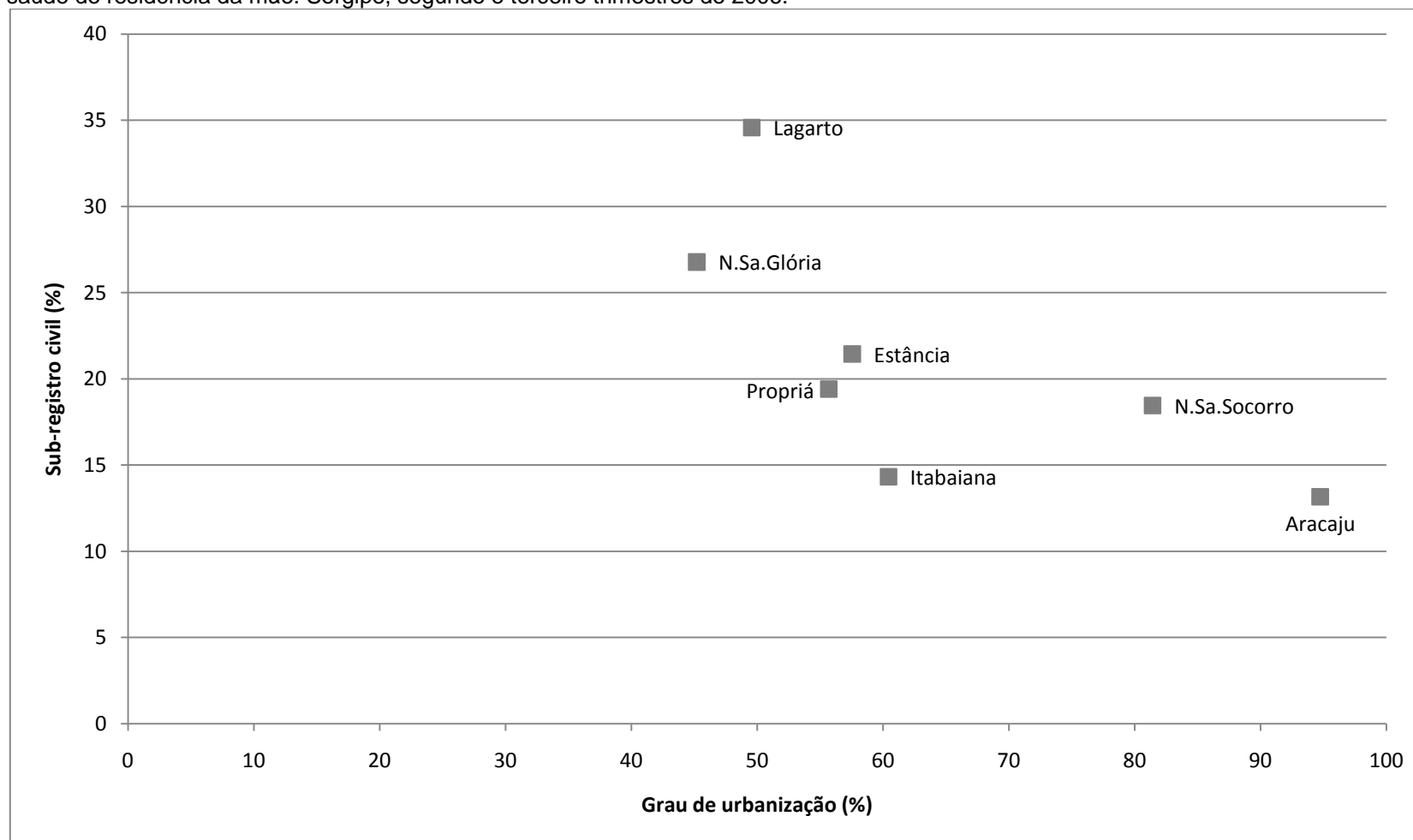


Figura 38 – Estimativas pontuais da cobertura do SINASC, com o pareamento menos restrito, segundo o grau de urbanização das regionais de saúde de residência da mãe. Sergipe, segundo e terceiro trimestres de 2006.

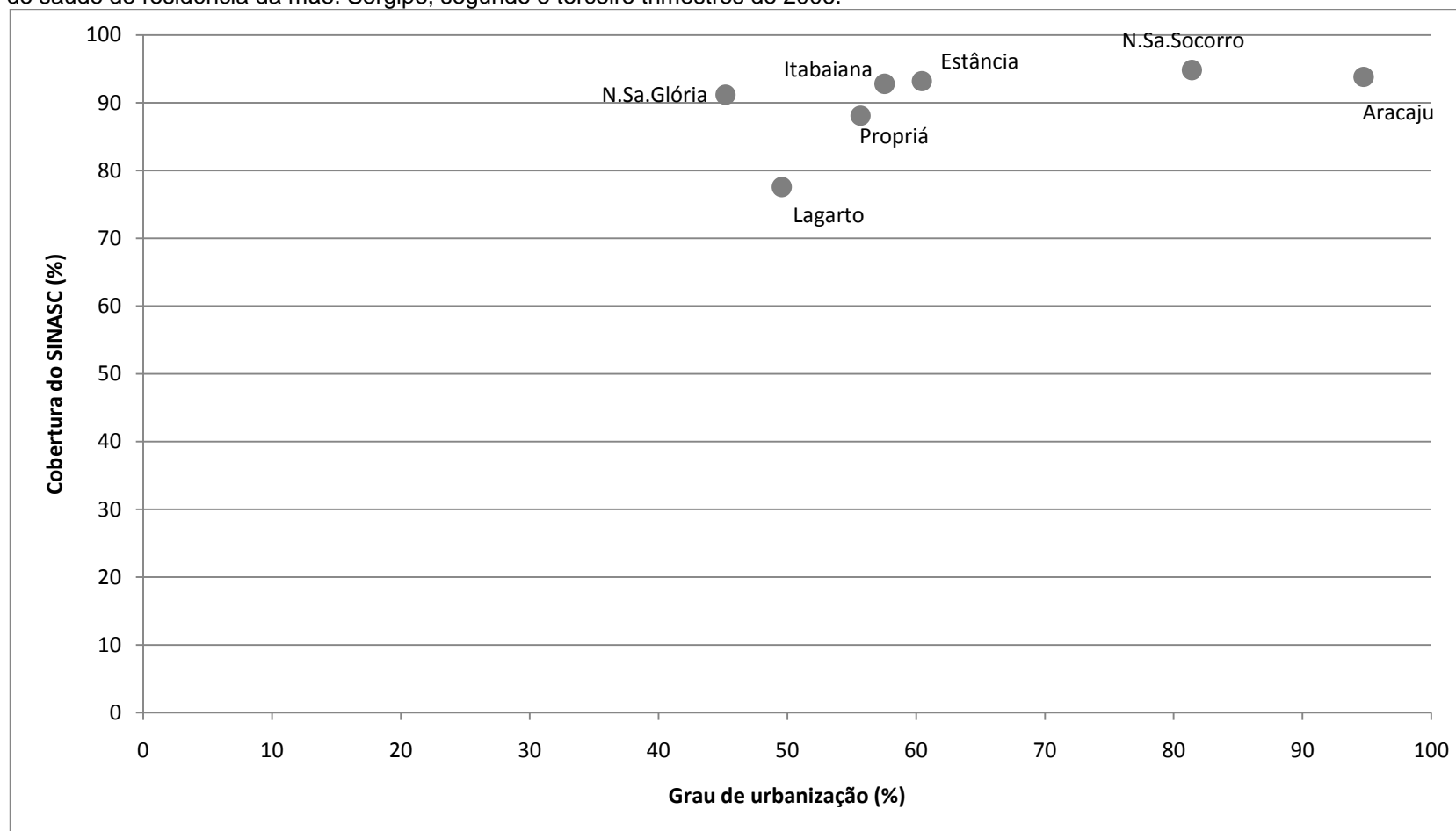


Figura 39 – Estimativas pontuais do sub-registro civil, com o pareamento menos restrito, segundo o percentual de mulheres de 10 a 49 anos sem instrução, residentes nas regionais de saúde. Sergipe, segundo e terceiro trimestres de 2006.

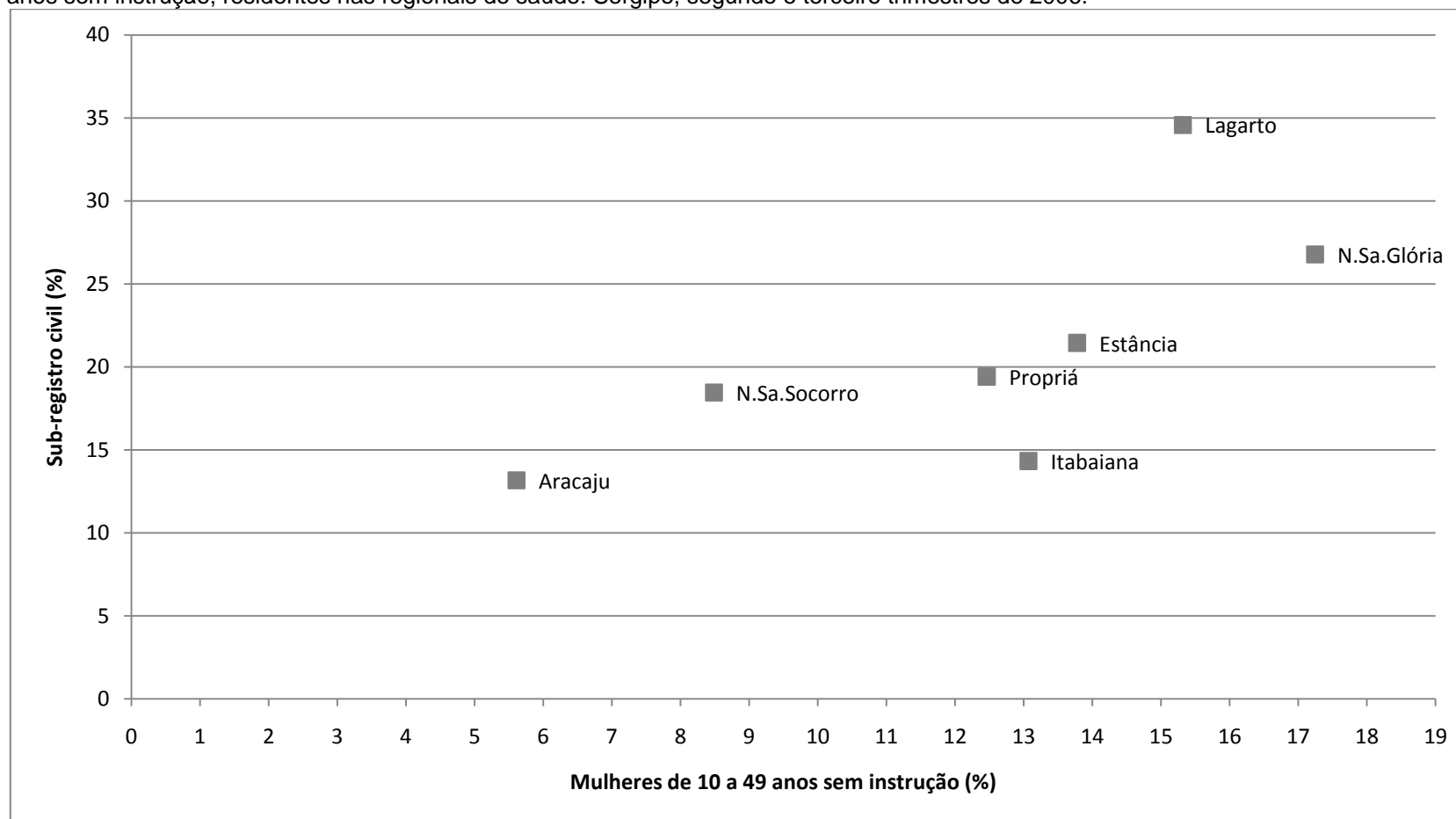
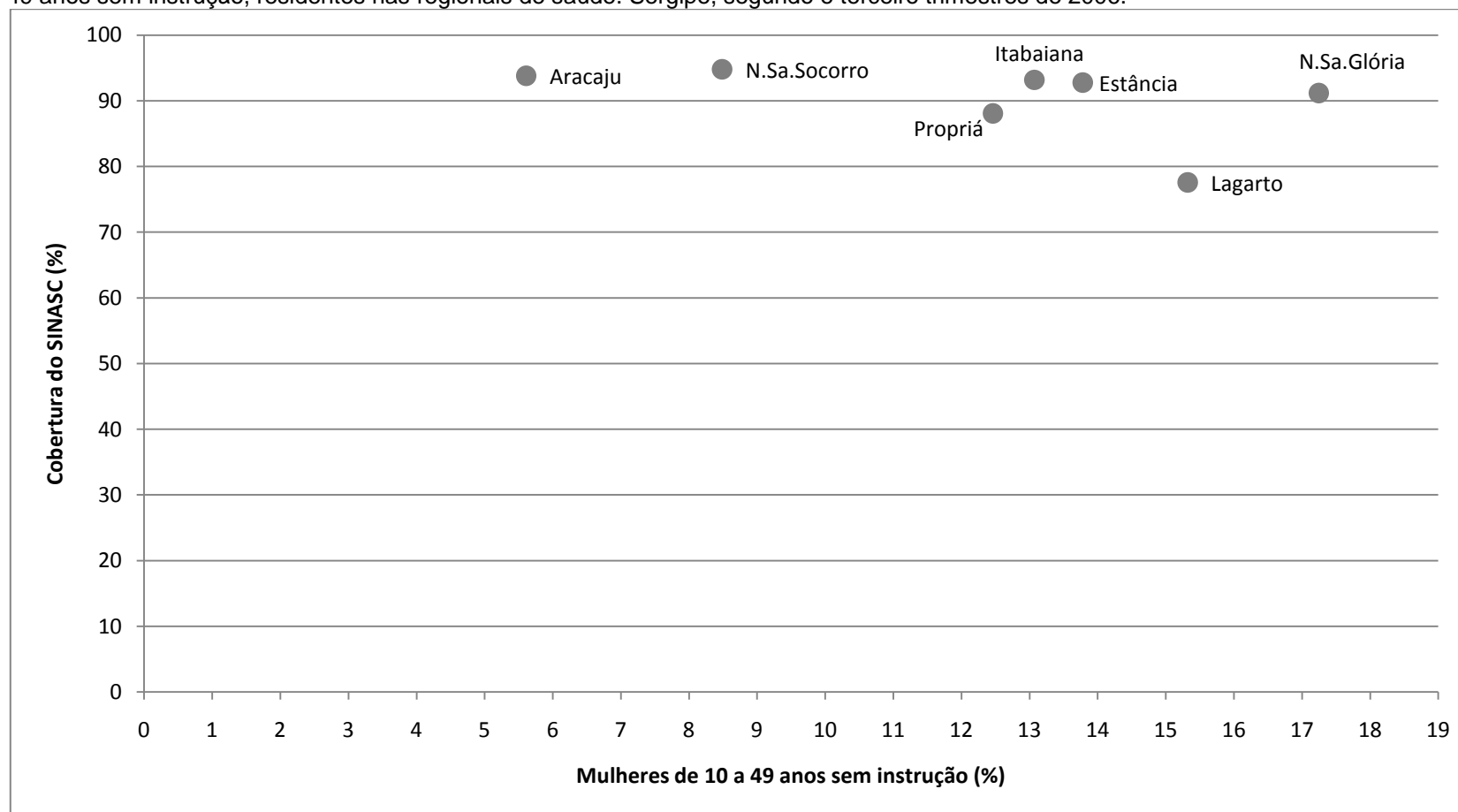


Figura 40 – Estimativas pontuais da cobertura do SINASC, com o pareamento menos restrito, segundo o percentual de mulheres entre 10 e 49 anos sem instrução, residentes nas regionais de saúde. Sergipe, segundo e terceiro trimestres de 2006.



7. CONSIDERAÇÕES FINAIS

Buscou-se nortear o presente trabalho pelas recomendações de Hook e Regal (1999), tanto no que tange ao ajuste de modelos que incluam variáveis representativas dos fatores que levam à heterogeneidade e na tentativa de ponderação dos parâmetros, quanto à apresentação das bases utilizadas; das falhas no pareamento e da sensibilidade das estimativas a diferentes métodos de relacionamento; além da possível dependência entre as fontes.

Considerou-se que a população de nascidos vivos em determinada área geográfica é fechada, pois óbitos infantis e migração da família após o parto não afetam o tamanho desta população.

O não preenchimento do número da DN em 4,7% dos registros da base do IBGE prejudicou o pareamento determinístico, especialmente nas microrregiões de Agreste de Lagarto, Tobias Barreto, Japaratuba e Boquim, onde a proporção de registros sem o número da DN superou os 5%. Nas regionais de Saúde, apenas Lagarto teve alta proporção de registros sem o número da DN (21,31%). Por outro lado, apesar das falhas no relacionamento, nota-se grande sobreposição das bases. Outro ponto importante a ser destacado é o vício de informação do município de residência, responsável por divergência de cerca de 8% nos pares.

Em relação à equi-probabilidade, os modelos selecionados sugerem que a idade da mãe afeta a probabilidade de captura da base do Registro Civil, indicando heterogeneidade relacionada a esta característica. O registro tardio deve ser levado em conta para se evitar a geração de heterogeneidade artificial, devido à censura dos dados. Acredita-se que estipular o prazo de seis meses após o nascimento vivo, na seleção dos dados do IBGE, pode minimizar o efeito do registro tardio.

Ressalta-se a continuidade que este trabalho demanda, no que tange aos estimadores da variância do sub-registro civil e da cobertura do SINASC. É importante desenvolver estimadores válidos, que permitam a estimação por intervalo destes indicadores.

Cabe chamar atenção para as potencialidades do uso do Programa MARK[®] em estudos epidemiológicos. Apesar do suporte limitado de grandes bancos de dados, a oferta de modelos deste aplicativo pode torná-lo ferramenta de grande utilidade na área da saúde.

Quanto aos objetivos específicos deste estudo, o método de captura e recaptura revelou-se viável para estimações em níveis geográficos menores do que o de unidade da federação. A presença ou ausência de cartório no município de residência da mãe representaria um fator institucional a ser pesquisado. Porém, em Sergipe, em 2006, apenas dois municípios não dispunham de cartório de registro civil e por esta razão não se averiguou associação entre este fator e o sub-registro civil. Por restrições do modelo empregado, a falta de disponibilidade de informações sobre escolaridade, renda da mãe, entre outras características, na base do IBGE, impediu a investigação de outros fatores sócio-demográficos diretamente no modelo, além da idade da mãe. Por outro lado, a observação *a posteriori* do comportamento das estimativas do sub-registro civil e da cobertura do SINASC permitiram apontar a aparente suscetibilidade do sub-registro civil à baixa instrução da mãe e ao baixo grau de urbanização das microrregiões e regionais de saúde. Ainda, esta análise sugeriu que a cobertura do SINASC não é afetada por estes fatores sócio-demográficos.

Apesar de todos os reveses enfrentados, os resultados do presente trabalho permitem concluir que é possível a aplicação da metodologia de captura e recaptura para estimar sub-registro de nascidos vivos. Este método pode ser estendido, também, para outros eventos vitais, como óbitos. No caso de grande sobreposição entre duas bases, o *International Working Group for Disease Monitoring and Forecasting* (1995b) recomenda que estas sejam agregadas e transformadas em uma só fonte. A saída seria o relacionamento determinístico entre o SINASC e o Registro Civil e o relacionamento probabilístico entre esta nova base de dados e outras que podem ser empregadas para a aplicação da captura e recaptura, como os cadastros dos Programas de Saúde da Família e o SIH, por exemplo.

Por fim, acredita-se que o presente trabalho abriu as portas para a ampliação das fronteiras do conhecimento acerca da realidade do registro dos eventos vitais no Brasil; permitiu identificar as potencialidades da metodologia de captura e recaptura para além da sua aplicação em ecologia e apontou as lacunas a serem preenchidas para a plena utilização deste método. Muito ainda há a ser feito, inúmeras são as dificuldades, mas, retomando a epígrafe, uma longa jornada não se inicia antes do primeiro passo.

8. REFERÊNCIAS

- ALMEIDA MMG. Sub-registro de nascimentos em Salvador BA (Brasil). *Rev Saude Publica*. 1979; 13:208-19.
- ARGOLLO N, LESSA I. [Estimated prevalence of cerebral neoplasm of childhood from the capture-recapture method] *Arq Neuropsiquiatr*. 1999 Jun;57(2B):435-41
- ASLAN D, AZCEBE H, BERTAN M, KARAAAOLU E. Capture-recapture methods for estimation of fertility and mortality in a rural district of Turkey. *East Mediterr Health Journal*. 2004 Jan-Mar;10(1-2):56-63.
- BALLIVET S, SALMI LR, DUBOURDIEU D. Capture-recapture method to determine the best design of a surveillance system. Application to a thyroid cancer registry. *Eur J Epidemiol*. 2000 Feb;16(2):147-53.
- BARROS FC, VITORIA CG, TEIXEIRA AMB, PUERTO FILHO M. Mortalidade Infantil em Pelotas, RS: nossas estatísticas são confiáveis? *Cad Saúde Publica*. 1985; 1(3): 348-58.
- BEZERRA KF. Estimativa do número de crianças e adolescentes em situação de rua em dois municípios do Brasil através do método de captura - recaptura. [Dissertação de Mestrado]. Aracaju: Universidade Federal de Sergipe; 2007. 65p.
- BRASIL. Lei no 6.015, de 31 de dezembro de 1973. Dispõe sobre os registros públicos e dá outras providências. *Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 31 dez. 1973. P. 13528, col. 1. Retificada no Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 30 out. 1975, p. 14337, col. 4.*
- BRASIL. Lei no 9.534, de 10 de dezembro de 1997. Dá nova redação ao artigo 30 da Lei no 6.015, de 31 de dezembro de 1973, que dispõe sobre os registros públicos; acrescenta inciso ao artigo 1 da Lei no 9.265, de 12 de fevereiro de 1996, que trata da gratuidade dos atos necessários ao exercício de cidadania; e altera os artigos 30 e 45 da Lei no 8.935, de 18 de novembro de 1994, que dispõe sobre os serviços notariais e de

- registro. Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 11 dez. 1977. p. 29440, col. 1.
- BRENNER H. Use and limitations of the capture-recapture method in disease monitoring with two dependent sources. *Epidemiology*. 1995 Jan;6(1):42-8.
- BRENNER H, STEGMAIER C, ZIEGLER H. Estimating completeness of cancer registration: an empirical evaluation of the two source capture-recapture approach in Germany. *J Epidemiol Community Health*. 1995 Aug;49(4):426-30.
- BRUM LM. Estimativa do número de casos de leptospirose sintomática na região central do Rio Grande do sul, Brasil, a partir do método de captura-recaptura [Dissertação de Mestrado]. Florianópolis: Universidade Federal de Santa Catarina; 2005. 93p.
- CADWELL BL, SMITH PJ, BAUGHMAN AL. Methods for capture-recapture analysis when cases lack personal identifiers. *Stat Med*. 2005 Jul 15;24(13):2041-51.
- CAIAFFA WT, ET AL. Estimation of the number of injecting drug users attending an outreach syringe-exchange program and infection with human immunodeficiency virus (HIV) and hepatitis C virus: the AjUDE-Brasil project. *J Urban Health*. 2003 Mar;80(1):106-14.
- CARDOSO AL, FERNANDES MIM, SILVA AAM, ARAGÃO VMF, SILVA RA. Sub-registro de nascimentos no município de Centro Novo do Maranhão, 2002. *Rev Bras Epidemiol*. 2003; 6(3): 237-44.
- CHANG YF, MCMAHON JE, HENNON DL, LAPORTE RE, COBEN JH. Dog bite incidence in the city of Pittsburgh: a capture-recapture approach. *Am J Public Health*. 1997 Oct;87(10):1703-5.
- COELI CM, CAMARGO JR, KR. Tutorial RECLINK III. Rio de Janeiro, 2007. Disponível em <http://www.mediafire.com/?sharekey=bc99288de55df934d2db6fb9a8902bda>. [Acesso em 15/03/2009].

- COELI CM, VERAS RP, COUTINHO ESF. Capture-recapture methodology: an option for surveillance of non-communicable diseases in the elderly. *Cad Saude Publica*, 2000 Oct-Dec; 16(4):1071-8230.
- COELI CM. Vigilância do diabetes mellitus em uma população idosa: aplicação da metodologia de captura-recaptura [Tese de Doutorado]. Rio de Janeiro: Universidade do Estado do Rio de Janeiro; 1998. 183p.
- COOCH E, WHITE G (Organizadores). Program MARK: A Gentle Introduction. Ithaca: Cornell University, 2008 [Acesso em 02/03/2009]. Disponível em: <http://www.phidot.org/software/mark/docs/book/>.
- CORMACK RM, CHANG YF, SMITH GS. Estimating deaths from industrial injury by capture-recapture: a cautionary tale. *Int J Epidemiol*. 2000 Dec;29(6):1053-9.
- CORRAO G, BAGNARDI V, VITTADINI G, FAVILLI S. Capture-recapture methods to size alcohol related problems in a population. *J Epidemiol Community Health*. 2000 Aug;54(8):603-10.
- CÔRTEZ MCJW. Vigilância das meningites na região metropolitana de Belo Horizonte, MG, 1999: o uso dos sistemas de informação em saúde e o método de captura-recaptura na estimação da incidência e da subnotificação [Tese de Doutorado]. Belo Horizonte: Universidade Federal de Minas Gerais; 2002. 219p.
- COSTA LN. Estudos e Pesquisas de Avaliação dos Censos Demográficos 1970 a 1990. Rio de Janeiro: IBGE [Texto para Discussão n.º 34]; 1990.
- COSTA JMBS, FRIAS PG. Avaliação da completude das variáveis da Declaração de Nascido Vivo de residentes em Pernambuco, Brasil, 1996 a 2005. *Cad S Pub*, 2009; 25(3): 613-24.
- COUTINHO ESF, COELI CM. Acurácia da metodologia de relacionamento probabilístico de registros para identificação de óbitos em estudos de sobrevivência. *Cad S Pub*, 2006; 22(10): 2249-52.
- CRESPO CD, BASTOS AA, CAVALCANTI WA. A Pesquisa do Registro Civil: condicionantes do sub-registro de nascimentos e perspectivas de melhorias da cobertura. In: Anais do XV Encontro Nacional de Estudos Populacionais; 2006; Caxambu, BR. São Paulo: Associação Brasileira de

- Estudos Populacionais; 2006. Publicação online. Disponível em www.abep.org.br. [Acesso em 20/05/2009].
- DEBROCK C, PREUX PM, HOUINATO D, DRUET-CABANAC M, KASSA F, ADJEN C, AVODE G, DENIS F, BOUTROS-TONI F, DUMAS M. Estimation of the prevalence of epilepsy in the Benin region of Zinvie using the capture-recapture method. *Int J Epidemiol*. 2000 Apr;29(2):330-5.
- DESENCLOS JC, HUBERT B. Limitations to the universal use of capture-recapture methods. *Int J Epidemiol*. 1994 Dec;23(6):1322-3
- DOMINGO-SALVANY A, HARTNOLL RL, MAGUIRE A, BRUGAL MT, ALBERTIN P, CAYLA JA, CASABONA J, SUELVES JM. Analytical considerations in the use of capture-recapture to estimate prevalence: case studies of the estimation of opiate use in the metropolitan area of Barcelona, Spain. *Am J Epidemiol*. 1998 Oct 15;148(8):732-40.
- DRUMMOND EF, FRANÇA EB, MACHADO CJ. SIH-SUS e SINASC: Utilização do método probabilístico para relacionamento de dados. *Cad S Col*. 2006; 14(2) :251-64.
- DRUMMOND EF, MACHADO CJ, FRANÇA E. Subnotificação de nascidos vivos: procedimentos de mensuração a partir do Sistema de Informação Hospitalar. *Revista de Saúde Pública* 2008; 42(1): 55-63.
- DUNN J, ANDREOLI SB. Método de captura e recaptura: nova metodologia para pesquisas epidemiológicas. *Rev Saude Publica*, 28 (6): 449-53, 1994.
- DUNN J, FERRI CP. Epidemiological methods for research with drug misusers: review of methods for studying prevalence and morbidity. *Rev Saude Publica*. 1999 Apr;33(2):206-15
- EISELE TP, LINDBLADE KA, ROSEN DH, ODHIAMBO F, VULULE JM, SLUTSKER L. Evaluating the completeness of demographic surveillance of children less than five years old in western Kenya: a capture-recapture approach. *Am J Trop Med Hyg*. 2003 Jul;69(1):92-7.
- ELKHOURY ANSM. Avaliação dos Registros de Morbimortalidade da Leishmaniose Visceral em Sistemas de Informações do SUS

- [Dissertação de Mestrado]. Salvador: Universidade Federal da Bahia; 2005. 97p.
- EL-SHALAKANI M. Estimating the completeness of births and deaths registration in Egypt through dual record systems. *Genus*. 1985 Jan-Jun;41(1-2):119-32.
- ERICKEN EP, KADANE JB. Estimating the population in a Census year, 1980 and beyond. *J Am Stat Assoc*, 80 (389): 98-109, 1985.
- FAY R, PASSEL JS, ROBINSON JG, COWAN CD. The coverage of the population in the 1980 Census. 1980 Census of Population and Housing - Evaluation and Research Reports. US Bureau of Census, 1985.
- FERRARI LS, et al. Mortalidade neonatal no Município de Londrina, Paraná, Brasil, nos anos 1994, 1999 e 2002. *Cadernos de Saúde Pública* 2006; 22(6): 1063-71.
- FUNASA – FUNDAÇÃO NACIONAL DE SAÚDE. Manual de procedimentos do sistema de informações de nascidos vivos. Brasília: Ministério da Saúde, 2001.
- GILL GV, ISMAIL AA, BEENCHING NJ. The use of capture-recapture techniques in determining the prevalence of type 2 diabetes. *QJM*. 2001 Jul;94(7):341-6.
- GONÇALVES, IF. Prevalência do diabetes mellitus do tipo 1 (insulino dependente) em crianças menores de quinze anos de idade na Ilha do Governador – Rio de Janeiro [Dissertação de Mestrado]. Rio de Janeiro: Universidade Federal do Rio de Janeiro; 1999. 61p.
- GONÇALVES VF. Estimativa de subnotificação de casos de AIDS em Fortaleza, Ceará – 2002 e 2003: Uma aplicação do método de captura – recaptura. [Dissertação de Mestrado]. Fortaleza: Universidade Federal do Ceará; 2006
- GUIMARÃES RB. Regiões de Saúde e escalas geográficas. *Cadernos de Saúde Pública* 2005; 21(4): 1017-25.
- HARVEY JN, CRANEY L, KELLY D. Estimation of the prevalence of diagnosed diabetes from primary care and secondary care source data:

- comparison of record linkage with capture-recapture analysis. *J Epidemiol Community Health*. 2002 Jan;56(1):18-23.
- HAY G, MCKEGANEY N. Estimating the prevalence of drug misuse in Dundee, Scotland: an application of capture-recapture methods. *J Epidemiol Community Health*. 1996 Aug;50(4):469-72.
- HOOK EB, REGAL RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev*. 1995; 17(2):243-64. Erratum in: *Am J Epidemiol* 1998 Dec 15;148(12):1219.
- HOOK EB, REGAL RR. Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology. *J Clin Epidemiol*. 1999 Oct;52(10):917-26; discussion 929-33.
- HUGGINS RM. On the statistical analysis of capture experiments. *Biometrika*, 1989; 76(1): 133-40
- HUGGINS RM. Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, 1991; 47: 725-32
- IBGE – Instituto Brasileiro de Geografia e Estatística. Estatísticas do Registro Civil. Rio de Janeiro; 2005. v: 32.
- IBGE – Instituto Brasileiro de Geografia e Estatística. Estatísticas do Registro Civil. Rio de Janeiro; 2006. v: 33.
- IBGE – Instituto Brasileiro de Geografia e Estatística. Projeção da População do Brasil por sexo e idade. 1980 a 2050. Revisão 2008. Rio de Janeiro: 2008. (Série Estudos e Pesquisas, Informação Demográfica e Socioeconômica, n.º 24)
- IBGE, BANCO MULTIDIMENSIONAL DE ESTATÍSTICAS. Censo Demográfico 2000. Dados da Amostra. Rio de Janeiro: 2000. Disponível em www.bme.ibge.gov.br. [Acesso em 20/05/2009].
- IBGE, BANCO MULTIDIMENSIONAL DE ESTATÍSTICAS. Estimativa Populacional para os Municípios, 2006. Rio de Janeiro; 2006a. Disponível em www.bme.ibge.gov.br. [Acesso em 20/05/2009].
- IBGE. Contagem Populacional 2007. Rio de Janeiro; 2007. Disponível em www.sidra.ibge.gov.br. [Acesso em 20/05/2009].

- IÑIGO J, ARCE A, MARTIN-MORENO JM, HERRUZO R, PALENQUE E, CHAVES F. Recent transmission of tuberculosis in Madrid: application of capture-recapture analysis to conventional and molecular epidemiology. *Int J Epidemiol.* 2003 Oct;32(5):763-9.
- ISMAIL AA, BEECHING NJ, GILL GV, BELLIS MA. How many data sources are needed to determine diabetes prevalence by capture-recapture? *Int J Epidemiol.* 2000 Jun;29(3):536-41.
- ISMAIL AA, GILL GV. The epidemiology of Type 2 diabetes and its current measurement. *Baillieres Best Pract Res Clin Endocrinol Metab.* 1999 Jul;13(2):197-220.
- IWGDMF - International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation I: History and theoretical development. *Am J Epidemiol.* 1995a Nov 15;142(10):1047-58.
- IWGDMF - International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation II: Applications in human diseases. *Am J Epidemiol.* 1995b Nov 15;142(10):1059-68.
- JARVIS SN, LOWE PJ, AVERY A, LEVENE S, CORMACK RM. Children are not goldfish--mark/recapture techniques and their application to injury data. *Inj Prev.* 2000 Mar;6(1):46-50.
- KORF DJ, REIJNEVELD SA, TOET J. Estimating the number of heroin users: a review of methods and empirical findings from The Netherlands. *Int J Addict.* 1994 Sep;29(11):1393-417
- LANZIERI TM. Avaliação do sistema de vigilância da síndrome da rubéola congênita no Brasil. [Dissertação de Mestrado Profissionalizante]. Salvador: Universidade Federal da Bahia; 2005. 49p
- LAPORTE RE. Assessing the human condition: capture-recapture techniques. *BMJ,* 308:5-6, 1994.
- LARSON A, BAMMER G. Why? Who? How? Estimating numbers of illicit drug users: lessons from a case study from the Australian Capital Territory. *Aust N Z J Public Health.* 1996 Oct;20(5):493-9.

- LASKA EM, MEISNER M, WANDERLING J, SIEGEL C. Estimating population size and duplicates rates when records cannot be linked. *Stat Med*. 2003 Nov 15;22(21):3403-17.
- LAURENTI R, MELLO JORGE, MHP, LEBRÃO, ML, GOTLIEB, SLD. *Estatísticas de Saúde*. São Paulo: EPU, 2005. 2ª. Edição.
- LEE AJ, SEBER GA, HOLDEN JK, HUAKAU JT. Capture-recapture, epidemiology, and list mismatches: several lists. *Biometrics*. 2001 Sep;57(3):707-13.
- LUCKACKS P. Closed Population Capture-Recapture Models. In: COOCH E, WHITE G (Organizadores). *Program MARK: A Gentle Introduction*. Ithaca: Cornell University, 2008 [Acesso em 02/03/2009]. Disponível em: <http://www.phidot.org/software/mark/docs/book/>
- LUKACS PM, BURNHAM KP. Estimating population size from DNA-based closed capture-recapture data incorporating genotyping error. *J Wildlife Manag*, 2005; 69: 396–403
- MACHADO CJ. A literature review of record linkage procedures focusing on infant health outcomes. *Cad S Pub*, 2004; 20(2): 362-71
- MACHADO, CJ. Early infant morbidity and infant mortality in the city of São Paulo, Brazil: A probabilistic record linkage approach [Tese de Doutorado]. Baltimore: Johns Hopkins University; 2002. 335p
- MARK – Mark and Recapture Survival Rate Estimation[®], version 5.1 [software na internet]. Fort Collins: Department of Fishery and Wildlife, Colorado State University; 2007 [acesso em 02/01/2008]. Disponível em :<http://welcome.warnercnr.colostate.edu/~gwhite/mark/mark.htm#Downloading>.
- MASTRO TD, KITAYAPORN D, WENIGER BG, VANICHSENI S, LAOSUNTHORN V, UNEKLABH T, UNEKLABH C, CHOOPANYA K, LIMPAKARNJANARAT K. Estimating the number of HIV-infected injection drug users in Bangkok: a capture--recapture method. *Am J Public Health*. 1994 Jul;84(7):1094-9.
- MAXWELL JC. Methods for estimating the number of "hard-core" drug users. *Subst Use Misuse*. 2000 Feb;35(3):399-420.

- MCCARTY DJ, TULL ES, MOY CS, KWOH CK, LAPORTE RE. Ascertainment corrected rates: applications of the capture-recapture methods. *Int J Epidemiol*, 22 (3): 559-65, 1993.
- MELLO ALLC. Prevalência e sobrevida de casos de acidente vascular encefálico, no município do Rio de Janeiro, no ano de 1998 [Dissertação de Mestrado]. Rio de Janeiro: Escola Nacional de Saúde Pública; 2003. 73p.
- MELLO JORGE MHP, GOTLIEB SLD, ANDRADE SM. Análise dos registros de nascimentos vivos em localidade urbana no Sul do Brasil. *Rev. Saúde Pública*, 31 (1), 1997
- MELLO JORGE MHP, GOTLIEB SLD, SOBOLL MLMS, BALDIJÃO MFA, LATORRE MRDO. O Sistema de Informação Sobre Nascidos Vivos - SINASC. São Paulo: Centro Brasileiro de Classificação de Doenças; 1992, Reimpressão 1999. (Série Divulgação, n.º 7)
- MINGOTI SA, CAIAFFA WT; PROJETO AJUDE-BRASIL II. A capture-recapture technique to estimate the size of the injecting drug user population attending syringe exchange programs: AjUDE-Brasil II Project. *Cad Saude Publica*. 2006 Apr;22(4):783-9. Epub 2006 Apr 5.
- MISSIAGIA JG. Estimativa Bayesiana do tamanho de uma população de diabéticos através de listas de pacientes [Dissertação de Mestrado]. São Carlos: Universidade Federal de São Carlos; 2005. 103p
- MORRISON A, STONE DH. Capture-recapture: a useful methodological tool for counting traffic related injuries? *Inj Prev*. 2000 Dec;6(4):299-304.
- MS – Departamento de Atenção Básica – DAB. Atenção Básica e Saúde da Família. Brasília: 2009. Disponível em <http://dtr2004.saude.gov.br/dab/atencaobasica.php#diretriz>. [Acesso em 20/05/2009]
- MS – Departamento de Informática do SUS – DATASUS. Informações em Saúde. Brasília: 2009. Disponível em <http://w3.datasus.gov.br/datasus/datasus.php>. [Acesso em 20/05/2009]
- MS – Secretaria de Vigilância em Saúde – SVS. Portaria n.º 116, de 11 de fevereiro de 2009.

- NASCIMENTO EM, et al. Estudo de fatores de risco para óbitos de menores de um ano mediante compartilhamento de bancos de dados. *Cadernos de Saúde Pública* 2008; 24(11): 2593-602.
- NELSON LM, ANDERSON DW. Case finding for epidemiological surveys of multiple sclerosis in United States communities. *Mult Scler*. 1995 Apr;1(1):48-55
- NEUGEBAUER R. Application of a capture-recapture method (the Bernoulli census) to historical epidemiology. *Am J Epidemiol*. 1984 Oct;120(4):626-34.
- OLIVEIRA ATR, SIMÕES CCS. Perfil dos municípios com informações precárias sobre eventos vitais. Brasília, DF: Rede Interagencial de Informações para a Saúde; Organização Pan-Americana da Saúde, 2005. 91 p. Relatório.
- OLIVEIRA LCS, DIAS MRML, NASCIMENTO CMF; MATTOS ES, RAMALHO JR JJA. Censo Demográfico 2000 - Resultados da Pesquisa de Avaliação da Cobertura da Coleta. [Texto para Discussão n.º 09]. Rio de Janeiro: IBGE, 2003.
- OLIVEIRA MTC. A subnotificação de casos de AIDS em Belo Horizonte, Minas Gerais: uma aplicação do método de captura-recaptura [Tese de Doutorado]. Belo Horizonte: Universidade Federal de Minas Gerais; 2000. 167p.
- PAPOZ L, BALKAU B, LELLOUCH J. Case counting in epidemiology: limitations of methods based on multiple data sources. *Int J Epidemiol*. 1996 Jun;25(3):474-8.
- PLEDGER S. Unified Maximum Likelihood Estimates for Closed Capture-Recapture Models Using Mixtures. *Biometrics*, 2000; 56: 434-52.
- PORTELA MHRB. Sub-registro de nascimentos vivos em localidade do Estado do Piauí, Brasil. *Rev Saude Publica*. 1989; 23: 493-501.
- PRADEL, R. Utilization of capture-recapture for the study of recruitment and population growth rate. *Biometrics*, 1996; 52 (2): 703-9.

- RAHI JS, DEZATEUX C. Capture-recapture analysis of ascertainment by active surveillance in the British Congenital Cataract Study. *Invest Ophthalmol Vis Sci.* 1999 Jan;40(1):236-9.
- RIPSA – Rede Interagencial de Informação para a Saúde. Indicadores básicos para a Saúde no Brasil: Conceitos e Aplicações. 2.^a Ed. Brasília; 2008 Organização Pan-Americana de Saúde.
- RIPSA – Rede Interagencial de Informação para a Saúde. Indicadores e Dados Básicos (IDB), Brasil, 2007. Disponível em <http://tabnet.datasus.gov.br/cgi/idb2007/matriz.htm>. [Acesso em 20/05/2009].
- ROMERO DEM. Vantagem e limitações do Método Demográfico Indireto e dos dados da PNAD'98 para a estimativa da Mortalidade Infantil. In: XIII Encontro Nacional de Estudos Populacionais, Ouro Preto, MG, Brasil. Anais ... Associação Brasileira de Estudos Populacionais (ABEP). 2002. Disponível em <http://www.abep.org.br/usuario/GerenciaNavegacao.php?caderno_id=134enivel=2> [Acesso em 03/01/2007].
- SANCHEZ JL, URIBE CS, FRANCO AF, JIMENEZ ME, ARCOS-BURGOS OM, PALACIO LG. [Prevalence of myasthenia gravis in Antioquia, Colombia] *Rev Neurol.* 2002 Jun 1-15;34(11):1010-2.
- SANDERSON M, BENJAMIN JT, LANE MJ, CORNMAN CB, DAVIS DR. Application of capture-recapture methodology to estimate the prevalence of dementia in South Carolina. *Ann Epidemiol.* 2003 Aug;13(7):518-24
- SCHIRM E, MONSTER TB, DE VRIES R, VAN DEN BERG PB, DE JONG-VAN DEN BERG LT, TOBI H. How to estimate the population that is covered by community pharmacies? Na evaluation of two methods using drug utilisation information. *Pharmacoepidemiol Drug Saf.* 2004 Mar;13(3):173-9.
- SCHMID B. Fatores associados ao registro tardio de nascidos vivos em Sergipe, 2006: Análise de dados administrativos. São Paulo: 2009 [Digitado]

- SCHOUTEN LJ, STRAATMAN H, KIEMENEY LA, GIMBRERE CH, VERBEEK AL. The capture-recapture method for estimation of cancer registry completeness: a useful tool? *Int J Epidemiol*. 1994 Dec;23(6):1111-6.
- SEBER GA, HUAKAU JT, SIMMONS D. Capture-recapture, epidemiology, and list mismatches: two lists. *Biometrics*. 2000 Dec;56(4):1227-32.
- SEKAR CC, DEMING WE. On a method of estimating birth and death rates and the extent of registration. *J Am Med Assoc*. 1949 Mar; 44: 101-15.
- SEKIKAWA A, EGUCHI H, TOMINAGA M, MANAKA H, SASAKI H, CHANG YF, KATO T. Evaluating the reported prevalence of type 2 diabetes mellitus by the Oguni diabetes registry using a two-sample method of capture-recapture. *Int J Epidemiol*. 1999 Jun;28(3):498-501.
- SHAPIRO S. Estimating birth registration completeness. *J Am Stat Assoc*. 1949;45:261-4.
- SILVA JPL, TRAVASSOS C, VASCONCELLOS MM, CAMPOS LM. Revisão sistemática sobre encadeamento ou linkage de bases de dados secundários para uso em pesquisa em saúde no Brasil. *Cad S Col*, 2006; 14 (2): 197-224.
- SIMÕES CCS. Brasil: Estimativas da Mortalidade Infantil por Microrregiões e Municípios. Brasília: Ministério da Saúde, 1999.
- SPICHLER ER, SPICHLER D, LESSA I, COSTA E FORTI A, FRANCO LJ, LAPORTE RE. Capture-recapture method to estimate lower extremity amputation rates in Rio de Janeiro, Brazil. *Rev Panam Salud Publica*. 2001 Nov; 10(5): 334-40.
- STRICKLAND D. Parkinson's prevalence estimated by a state registry. *Mov Disord* 2004; 19(3): 318-23
- SZWARCWALD CL, LEAL MC, ANDRADE CLT, SOUZA JR PRB. Estimación da mortalidade infantil no Brasil: o que dizem as informações sobre óbitos do Ministério da Saúde? *Cad Saude Publica*. 2002; 18(6): 1725-36.
- SZWARCWALD CL. Strategies for improving the monitoring of vital events in Brazil. *Int J Epidemiology*. 2008; 37: 738-44.

- TILLING K, STERNE JAC. Capture-recapture models including covariate effects. *Am J epidemiol.* 1999; 149(4): 392-400.
- TILLING K, STERNE JAC, WOLFE CDA. Estimation of the incidence of stroke using a capture-recapture model including covariates. *Int J Epidemiol.* 2001; 30: 1351-9.
- TULL ES, BUTLER C, GUMBS L, WILLIAMS S. The use of capture-recapture methods to monitor diabetes in Dominica, West Indies. *Rev Panam Salud Publica.* 1998 May;3(5):303-7.
- UNITED NATIONS – UN. DEPARTMENT OF INTERNATIONAL ECONOMIC AND SOCIAL AFFAIRS. Manual X. Indirect techniques for demographic estimation. New York: 1983 (Populations Studies Series, number 81).
- UNITED NATIONS STATISTICAL DIVISION – UNSD. Official statistics: principles and practices, organization and management. New York; c2006. [acesso em 26/10/2009]. Disponível em: <http://unstats.un.org/unsd/methods/statorg/default.htm>.
- VERLATO G, MUGGEO M. Capture-recapture method in the epidemiology of type 2 diabetes: a contribution from the Verona Diabetes Study. *Diabetes Care.* 2000 Jun; 23(6):759-64.
- WALDVOGEL BC, FERREIRA CEC. Estatísticas da Vida. São Paulo em Perspectiva, 2003 Jul-Dez; 17(3-4): 55-66.
- WHITE CG. Program MARK, Introduction. Colorado State University, 2009 [Acesso em 14/03/2009]. Disponível em: <http://welcome.warnercnr.colostate.edu/~gwhite/mark/mark.htm#Introduction>
- WHITFIELD K, KELLY H. Using the two-source capture-recapture method to estimate the incidence of acute flaccid paralysis in Victoria, Australia. *Bull World Health Organ.* 2002; 80(11):846-51. Epub 2002 Dec 3.
- WITTES JT, SIDEL VW. A generalization of the simple capture-recapture model with applications to epidemiological research. *J Chronic Dis.* 1968 Aug;21(5):287-301.
- WOLTER KM. Accounting for America's miscounting. *Science*, 253: 12-8, 1991.

ANEXO I

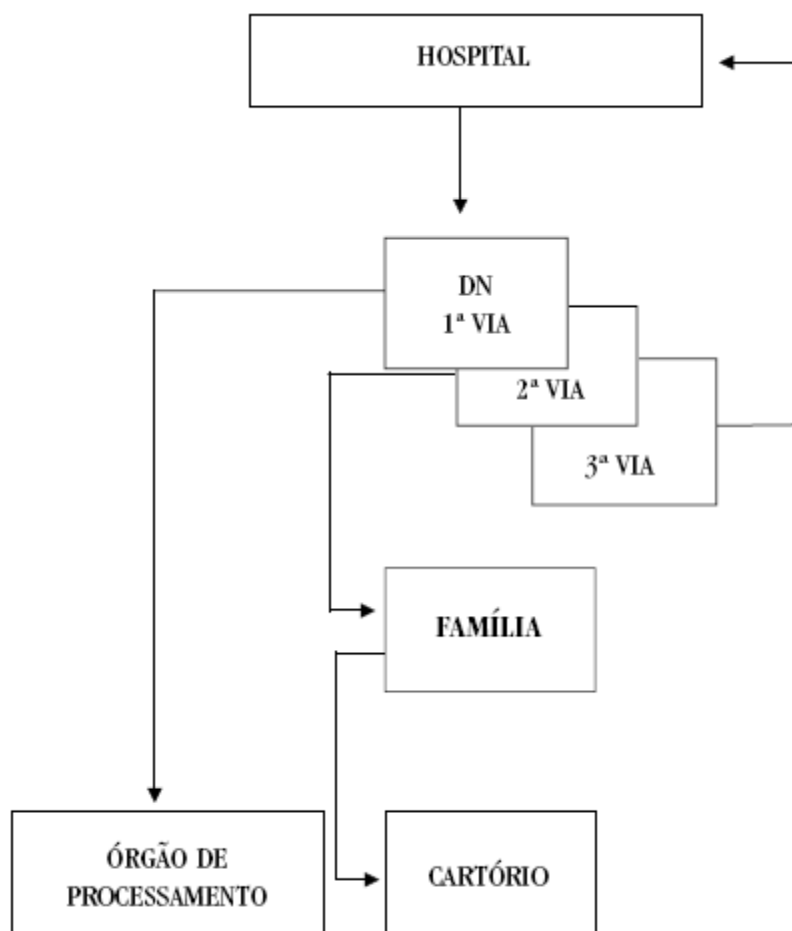
Modelo da Declaração de Nascido Vivo vigente em 2006.

República Federativa do Brasil Ministério da Saúde 1º Via - Secretaria de Saúde		Declaração de Nascido Vivo Nº		
I Centros	1) Cartório	Código	2) Registro	
	3) Data			
II Local da Ocorrência	4) Município	5) UF		
	6) Local da ocorrência 1 - Hospital 2 - Outros estabelecimentos 3 - Domicílio 4 - Outros 5 - Ignorado	7) Estabelecimento	Código	
	8) Endereço da ocorrência, ao fora do estabelecimento, ou da residência da mãe (Rua, praça, avenida, etc)		Número	Complemento
			9) CEP	
III Mãe	10) Bairro/Distrito	Código	11) Município de ocorrência	
			Código	
			12) UF	
	13) Nome da mãe		14) SBC	
	15) Idade (anos)	16) Estado civil 1 - Solteira 2 - Casada 3 - Viúva 4 - Sep. Judicial 5 - Ignorado	17) Escolaridade (anos de estudo concluídos) 1 - Menos de 1 2 - De 1 a 3 3 - De 4 a 7 4 - De 8 a 11 5 - 12 e mais 6 - Ignorado	18) Ocupação habitual e ramo de atividade Código
19) Número filhasidas em gestações anteriores (obs: incluir 99 se ignorado) Nascidos vivos Nascidos mortos				
IV Gestação e parto	20) Logradouro		Número	
			Complemento	
			21) CEP	
	22) Bairro/Distrito	Código	23) Município	
		Código	24) UF	
V Recém-nascido	25) Duração da gestação em semanas	26) Tipo de gravidez 1 - Única 2 - Dupla 3 - Tripla e mais 4 - Ignorado	27) Tipo de parto 1 - Vaginal 2 - Cesáreo 3 - Ignorado	
	28) Nascimento Data		29) Sexo M - Masculino F - Feminino Ignorado	
	30) Peso ao nascer em gramas		31) Índice de Apgar 1º minuto 5º minuto	
VI Identificação	32) Raça/cor 1 - Branca 2 - Preta 3 - Amarela 4 - Parda 5 - Indígena		33) Detectada alguma malformação congênita e/ou anomalia cromossômica? 1 - Sim 2 - Não 3 - Ignorado	
	34) Polígama da mãe		35) Pê direito da criança	
	36) Responsável pelo preenchimento Nome		37) Função	38) Identidade
		39) Órgão Emissor	40) Data	
ATENÇÃO : ESTE DOCUMENTO NÃO SUBSTITUI A CERTIDÃO DE NASCIMENTO O Registro de Nascimento é obrigatório por lei. Para registrar esta criança, o pai ou responsável deverá levar este documento ao cartório de registro civil.				

Fonte: FUNASA. *Manual de procedimentos do sistema de informações de nascidos vivos*. Brasília: Ministério da Saúde, 2001.

ANEXO II

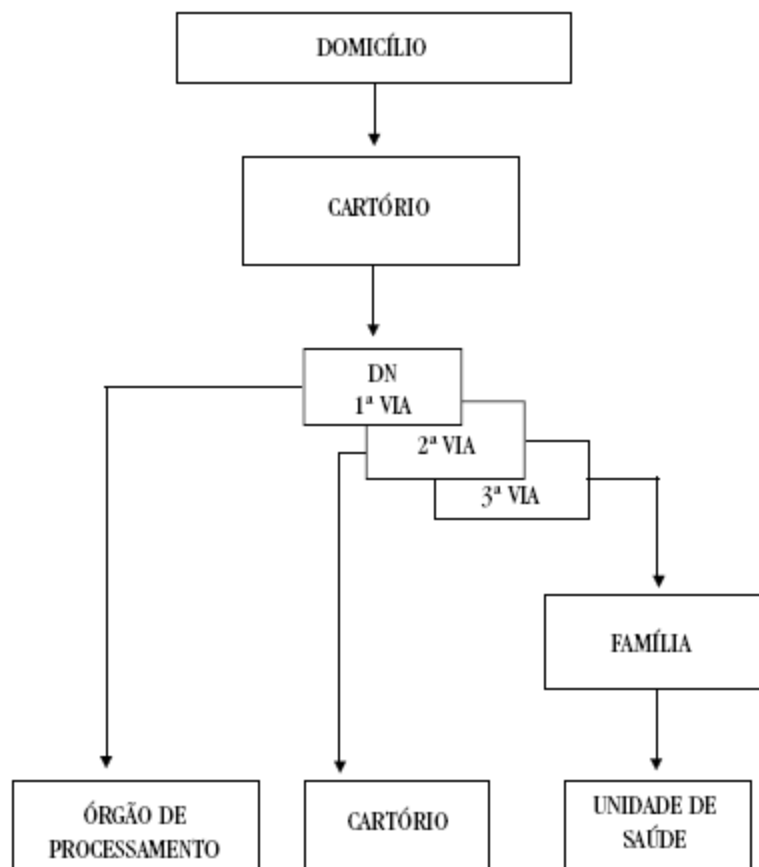
Fluxo da informação de nascidos vivos capturados pelo SINASC, para partos em estabelecimentos de saúde



Fonte: FUNASA. *Manual de procedimentos do sistema de informações de nascidos vivos*. Brasília: Ministério da Saúde, 2001.

ANEXO III

Fluxo da informação dos nascidos vivos capturados pelo SINASC, para partos ocorridos em domicílio e com DN preenchida pelo cartório.



Fonte: FUNASA. *Manual de procedimentos do sistema de informações de nascidos vivos*. Brasília: Ministério da Saúde, 2001.

ANEXO IV

Modelo de questionário sobre Nascidos Vivos. IBGE, 2006

 Instituto Brasileiro de Geografia e Estatística Diretoria de Pesquisas Coordenação de População e Indicadores Sociais REGISTRO CIVIL NASCIDOS VIVOS - RC.1		1 IDENTIFICAÇÃO								PARA USO DO IBGE									
		UF _____ MUNICÍPIO _____ DISTRITO _____ TRM _____ ANO _____ LIVRO Nº _____ Nº DE QUESTIONÁRIO POR LIVRO _____								CAMPO-CHAVE									
										MOD	TRM	UF	MUNICÍPIO - DV	DISTR	CART	ARROL			
		2								1	1	1	1	1	1	1			
<p>INSTRUÇÕES: Preencher o questionário de forma legível, de acordo com as instruções, com esferográficas azul ou preta. Destina-se este questionário aos arrolamentos dos Nascidos Vivos registrados no Cartório durante o trimestre considerado. Os arrolamentos deverão ser feitos em questionários distintos para cada livro. Não anotar registro que pertença a mês não abrangido pelo trimestre considerado. Não há necessidade de mudar de questionário para separar cada mês do trimestre. Quando houver registro anulado, o Nº DO REGISTRO NO LIVRO (coluna 2) deverá ser anotado e, a seguir, na mesma linha, escrever: ANULADO PELO CARTÓRIO. Não usar aspas para qualquer tipo de registro. Qualquer observação ou informação complementar deverá ser feita no verso deste modelo.</p> <p>PREENCHIMENTO: Coluna 2 Registrar o número da declaração de nascidos vivos do Ministério da Saúde. Colunas 4 e 6 Registrar com 2 (dois) algarismos o dia 01-02-03...31. Colunas 5 e 7 Registrar com 2 (dois) algarismos o mês: 01 = jan.; 2 = fev.;...12 = dez.; Coluna 8 Registrar com 2 (dois) algarismos o ano: 1999 = 99; 2000 = 00; 2001 = 01. Coluna 9 Registrar com 1 (um) algarismo: 1 = Hospital (Casa de saúde, maternidade); 2 = Outros estabelecimentos de saúde, sem internação (Posto de Saúde, Centro de Saúde, etc.); 3 = Domicílio; 4 = Via pública; 5 = Outros; 9 = Ignorado. Colunas 10 e 16 Registrar a sigla da Unidade da Federação. Coluna 11 Registrar o nome do Município, não o abreviando, de tal forma que torne impossível a sua identificação. Coluna 12 Registrar com 1 (um) algarismo: 1 = Não; 2 = Sim, gêmeo; 3 = Triplêto ou mais; 9 = Ignorado. Coluna 13 Registrar com 1 (um) algarismo: 1 = Masculino; 2 = Feminino; 9 = Ignorado. Colunas 14 e 15 Registrar a sigla da Unidade da Federação de nascimento dos genitores quando se tratar de brasileiros natos, ou o País de nascimento se forem estrangeiros ou naturalizados. Coluna 17 Registrar o nome do Município ou País, não o abreviando, de tal forma que torne impossível a sua identificação. Coluna 18 Registrar a idade da genitora com 2 (dois) algarismos. Quando for ignorado registrar (99).</p>																			
2		CARACTERÍSTICAS DO REGISTRO								3					CARACTERÍSTICAS INDIVIDUAIS DOS GENITORES				
Nº DE ORDEM DO ARROLAMENTO	DECLARAÇÃO DE NASCIMENTO Nº	Nº DO REGISTRO NO LIVRO	DATA DO REGISTRO		DATA DO NASCIMENTO			LOCAL DO NASCIMENTO	LUGAR DO NASCIMENTO		E GÊMEO?	SEXO	NATURALIDADE DOS GENITORES		NA OCASIÃO DO PARTO			Idade da genitora em anos completos	Nº DE ORDEM DO ARROLAMENTO
			Dia	Mês	Dia	Mês	Ano	1 = Hospital 2 = Outros estab. de saúde sem internação 3 = Domicílio 5 = Outros 9 = Ignorado	Sigla da UF	Município			1 = Mãe 2 = Sim (gêmeo) 3 = Triplêto ou mais 9 = Ignorado	1 = Masculino 2 = Feminino 9 = Ignorado	País	Mãe	Lugar de domicílio ou residência da genitora		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
01																		01	
02																		02	
03																		03	
04																		04	
05																		05	
06																		06	
07																		07	
08																		08	
09																		09	
10																		10	
11																		11	
12																		12	
13																		13	
14																		14	
15																		15	
2			2	2	2	2	2	1			1	1						2	2

02.08.1.3336/01

O Cartório deverá preencher o questionário em uma só via.

(continua verso)

ANEXO V

Dicionário das variáveis (campos) dos bancos finais SINASC e Registro Civil

CAMPO	DESCRIÇÃO	TAMANHO (DÍGITOS)	VALORES ASSUMIDOS	SIGNIFICADO
DN	Número da DN	8	-	-
CODUFRES	Código da Unidade da Federação de residência da mãe	2	11 a 17 21 a 29 31 a 33 e 35 41 a 43 50 a 53 99	UFs da Região Norte UFs da Região Nordeste UFs da Região Sudeste UFs da Região Sul UFs da Região Centro-Oeste e Distrito federal UF ignorada
CODMUNRES	Município de residência da mãe	6	- XX9999 999999	- UFs conhecida, mas município ignorado UFs e município ignorados
CODUFNASC	Código da Unidade da Federação de ocorrência do nascimento vivo	2	11 a 17 21 a 29 31 a 33 e 35 41 a 43 50 a 53 99	UFs da Região Norte UFs da Região Nordeste UFs da Região Sudeste UFs da Região Sul UFs da Região Centro-Oeste e Distrito federal UF ignorada
CODMUNNASC	Município de ocorrência do nascimento vivo	6	- XX9999 999999	- UFs conhecida, mas município ignorado UFs e município ignorados
SEXO	Sexo do recém-nascido	1	0 1 Branco	Masculino Feminino Ignorado
IDMAE	Idade da mãe em anos	2	10-55 Branco	- Ignorado
DTNASC	Data de nascimento	8	dd/mm/aaaa	-
GEST	Tipo de gestação	1	1 2 3 Branco	Única Dupla Tripla ou mais Ignorada
LOCNASC	Local de nascimento	1	1 2 3 4 Branco	Hospital Outro estabelecimento de saúde Domicílio Outro Ignorado

ANEXO VI

Modelos de Huggins para populações fechadas que apresentaram algum peso ponderado, com as respectivas estimativas de peso ponderado (w_i), probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) e estimativas ponderadas por w_i . Microrregiões de residência da mãe em Sergipe, pareamento menos restrito, segundo e terceiro trimestres de 2006.

SERGIPANA DO SERTÃO DE SÃO FRANCISCO (28001)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,912	0,008	0,714	0,011	1.855	4,544
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,912	0,008	0,714	0,011	1.855	4,545
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,911	0,008	0,714	0,011	1.856	4,687
{p(g) c(g + idmae)}	0,00004	0,912	0,008	0,717	0,011	1.855	4,543
{p(g + idmae) c(g + idmae)}	0,00002	0,912	0,008	0,717	0,011	1.855	4,546
Ponderação (MARK®)*		0,912	0,008	0,714	0,011	1.855	4,552

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

CARIRA (28002)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,937	0,011	0,822	0,016	577	1,716
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,937	0,011	0,822	0,016	577	1,716
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,937	0,011	0,822	0,016	577	1,717
{p(g) c(g + idmae)}	0,00004	0,937	0,011	0,822	0,016	577	1,716
{p(g + idmae) c(g + idmae)}	0,00002	0,937	0,011	0,822	0,016	577	1,716
Ponderação (MARK®)*		0,937	0,011	0,822	0,016	577	1,716

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

NOSSA SENHORA DAS DORES (28003)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,949	0,010	0,823	0,016	596	1,388
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,949	0,010	0,823	0,016	596	1,390
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,949	0,010	0,823	0,016	596	1,396
{p(g) c(g + idmae)}	0,00004	0,949	0,010	0,822	0,016	596	1,388
{p(g + idmae) c(g + idmae)}	0,00002	0,949	0,010	0,822	0,016	596	1,390
Ponderação (MARK®)*		0,949	0,010	0,823	0,016	596	1,389

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

AGRESTE DE ITABAIANA (28004)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,932	0,007	0,861	0,009	1.616	3,161
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,932	0,007	0,861	0,009	1.616	3,162
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,931	0,007	0,861	0,009	1.616	3,171
{p(g) c(g + idmae)}	0,00004	0,932	0,007	0,861	0,009	1.616	3,164
{p(g + idmae) c(g + idmae)}	0,00002	0,932	0,007	0,861	0,009	1.616	3,163
Ponderação (MARK®)*		0,932	0,007	0,861	0,009	1.616	3,162

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

TOBIAS BARRETO (28005)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,874	0,012	0,719	0,014	1.128	5,480
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,874	0,012	0,719	0,014	1.128	5,487
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,877	0,012	0,719	0,014	1.129	5,943
{p(g) c(g + idmae)}	0,00004	0,874	0,012	0,719	0,014	1.128	5,482
{p(g + idmae) c(g + idmae)}	0,00002	0,874	0,012	0,719	0,014	1.128	5,487
Ponderação (MARK®)*		0,874	0,012	0,719	0,014	1.128	5,508

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

AGRESTE DE LAGARTO (28006)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,694	0,019	0,516	0,015	1.521	23,574
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,694	0,019	0,516	0,015	1.521	23,571
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,694	0,019	0,516	0,015	1.522	23,722
{p(g) c(g + idmae)}	0,00004	0,694	0,019	0,516	0,015	1.521	23,573
{p(g + idmae) c(g + idmae)}	0,00002	0,694	0,019	0,516	0,015	1.521	23,573
Ponderação (MARK®)*		0,694	0,019	0,516	0,015	1.521	23,581

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

PROPRIÁ (28007)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,876	0,013	0,807	0,014	1.011	5,113
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,876	0,013	0,807	0,014	1.011	5,115
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,874	0,013	0,807	0,014	1.011	5,164
{p(g) c(g + idmae)}	0,00004	0,876	0,013	0,805	0,013	1.011	5,113
{p(g + idmae) c(g + idmae)}	0,00002	0,876	0,013	0,805	0,013	1.011	5,115
Ponderação (MARK®)*		0,876	0,013	0,807	0,014	1.011	5,117

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

CONTIGUIBA (28008)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,914	0,014	0,829	0,017	537	2,361
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,915	0,014	0,829	0,017	537	2,362
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,915	0,014	0,829	0,017	537	2,373
{p(g) c(g + idmae)}	0,00004	0,914	0,014	0,828	0,017	537	2,360
{p(g + idmae) c(g + idmae)}	0,00002	0,915	0,014	0,828	0,017	537	2,362
Ponderação (MARK®)*		0,915	0,014	0,829	0,017	537	2,362

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

JAPARATUBA (28009)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,874	0,016	0,775	0,018	628	4,087
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,875	0,016	0,775	0,018	628	4,088
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,874	0,016	0,775	0,018	628	4,089
{p(g) c(g + idmae)}	0,00004	0,874	0,016	0,774	0,018	628	4,088
{p(g + idmae) c(g + idmae)}	0,00002	0,875	0,016	0,774	0,018	628	4,088
Ponderação (MARK®)*		0,874	0,016	0,775	0,018	628	4,087

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

BAIXO CONTIGUIBA (28010)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,921	0,009	0,870	0,011	1.030	2,980
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,921	0,009	0,870	0,011	1.030	2,982
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,921	0,010	0,870	0,011	1.030	2,983
{p(g) c(g + idmae)}	0,00004	0,921	0,009	0,871	0,011	1.030	2,982
{p(g + idmae) c(g + idmae)}	0,00002	0,921	0,009	0,871	0,011	1.030	2,982
Ponderação (MARK®)*		0,921	0,009	0,870	0,011	1.030	2,980

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ARACAJU (28011)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,944	0,003	0,848	0,004	7.717	5,491
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,944	0,003	0,848	0,004	7.717	5,494
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,944	0,003	0,848	0,004	7.718	5,602
{p(g) c(g + idmae)}	0,00004	0,944	0,003	0,848	0,004	7.717	5,501
{p(g + idmae) c(g + idmae)}	0,00002	0,944	0,003	0,848	0,004	7.717	5,495
Ponderação (MARK®)*		0,944	0,003	0,848	0,004	7.717	5,498

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

BOQUIM (28012)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,886	0,009	0,749	0,011	1.874	6,231
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,886	0,009	0,749	0,011	1.874	6,233
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,886	0,009	0,749	0,011	1.874	6,238
{p(g) c(g + idmae)}	0,00004	0,886	0,009	0,750	0,011	1.874	6,230
{p(g + idmae) c(g + idmae)}	0,00002	0,886	0,009	0,750	0,011	1.874	6,233
Ponderação (MARK®)*		0,886	0,009	0,749	0,011	1.874	6,232

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ESTÂNCIA (28013)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,66837	0,949	0,006	0,799	0,011	1.408	2,117
{p(g + idmae) c(g + idmae + g*idmae)}	0,27672	0,949	0,006	0,799	0,011	1.408	2,118
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,05485	0,949	0,006	0,799	0,011	1.408	2,121
{p(g) c(g + idmae)}	0,00004	0,949	0,006	0,800	0,011	1.408	2,117
{p(g + idmae) c(g + idmae)}	0,00002	0,949	0,006	0,800	0,011	1.408	2,118
Ponderação (MARK®)*		0,949	0,006	0,799	0,011	1.408	2,118

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ANEXO VII

Modelos de Huggins para populações fechadas que apresentaram algum peso ponderado, com as respectivas estimativas de peso ponderado (w_i), probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) e estimativas ponderadas por w_i . Regionais de saúde de residência da mãe em Sergipe, pareamento menos restrito, segundo e terceiro trimestres de 2006.

ARACAJU (2801)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,62377	0,938	0,003	0,864	0,004	7.292	6,111
{p(g) c(g + idmae + g*idmae)}	0,24275	0,938	0,003	0,864	0,004	7.291	6,008
{p(g + idmae) c(g + idmae + g*idmae)}	0,13328	0,938	0,003	0,864	0,004	7.291	6,007
{p(g + idmae + g*idmae) c(g + idmae)}	0,00011	0,938	0,003	0,864	0,004	7.292	6,110
{p(g) c(g + idmae)}	0,00004	0,938	0,003	0,864	0,004	7.291	6,007
{p(g + idmae) c(g + idmae)}	0,00002	0,938	0,003	0,864	0,004	7.291	6,007
{p(g + idmae + g*idmae) c(g)}	0,00001	0,938	0,003	0,864	0,004	7.292	6,111
Ponderação (MARK[®])*		0,938	0,003	0,864	0,004	7.291	6,072

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK[®].

ESTÂNCIA (2802)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,62377	0,949	0,004	0,809	0,008	2.905	3,144
{p(g) c(g + idmae + g*idmae)}	0,24275	0,948	0,004	0,809	0,008	2.905	3,108
{p(g + idmae) c(g + idmae + g*idmae)}	0,13328	0,948	0,004	0,809	0,008	2.905	3,110
{p(g + idmae + g*idmae) c(g + idmae)}	0,00011	0,949	0,004	0,809	0,007	2.905	3,143
{p(g) c(g + idmae)}	0,00004	0,948	0,004	0,809	0,007	2.905	3,108
{p(g + idmae) c(g + idmae)}	0,00002	0,948	0,004	0,809	0,007	2.905	3,110
{p(g + idmae + g*idmae) c(g)}	0,00001	0,949	0,004	0,808	0,008	2.905	3,141
Ponderação (MARK®)*		0,948	0,004	0,809	0,008	2.905	3,131

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ITABAIANA (2803)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,62377	0,928	0,006	0,774	0,008	2.658	4,314
{p(g) c(g + idmae + g*idmae)}	0,24275	0,928	0,006	0,774	0,008	2.658	4,312
{p(g + idmae) c(g + idmae + g*idmae)}	0,13328	0,928	0,006	0,774	0,008	2.658	4,311
{p(g + idmae + g*idmae) c(g + idmae)}	0,00011	0,928	0,006	0,775	0,008	2.658	4,312
{p(g) c(g + idmae)}	0,00004	0,928	0,006	0,775	0,008	2.658	4,312
{p(g + idmae) c(g + idmae)}	0,00002	0,928	0,006	0,775	0,008	2.658	4,313
{p(g + idmae + g*idmae) c(g)}	0,00001	0,928	0,006	0,775	0,008	2.658	4,314
Ponderação (MARK®)*		0,928	0,006	0,774	0,008	2.658	4,313

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

LAGARTO (2804)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,62377	0,932	0,006	0,851	0,008	2.122	3,618
{p(g) c(g + idmae + g*idmae)}	0,24275	0,932	0,006	0,851	0,008	2.122	3,614
{p(g + idmae) c(g + idmae + g*idmae)}	0,13328	0,932	0,006	0,851	0,008	2.122	3,617
{p(g + idmae + g*idmae) c(g + idmae)}	0,00011	0,932	0,006	0,851	0,008	2.122	3,618
{p(g) c(g + idmae)}	0,00004	0,932	0,006	0,851	0,008	2.122	3,618
{p(g + idmae) c(g + idmae)}	0,00002	0,932	0,006	0,851	0,008	2.122	3,617
{p(g + idmae + g*idmae) c(g)}	0,00001	0,932	0,006	0,851	0,008	2.122	3,618
Ponderação (MARK®)*		0,932	0,006	0,851	0,008	2.122	3,617

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

NOSSA SENHORA DA GLÓRIA (2805)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,62377	0,776	0,011	0,619	0,010	2.983	19,923
{p(g) c(g + idmae + g*idmae)}	0,24275	0,776	0,011	0,619	0,010	2.983	19,840
{p(g + idmae) c(g + idmae + g*idmae)}	0,13328	0,775	0,011	0,619	0,010	2.983	19,867
{p(g + idmae + g*idmae) c(g + idmae)}	0,00011	0,776	0,011	0,619	0,010	2.983	19,920
{p(g) c(g + idmae)}	0,00004	0,776	0,011	0,619	0,010	2.983	19,836
{p(g + idmae) c(g + idmae)}	0,00002	0,775	0,011	0,619	0,010	2.983	19,867
{p(g + idmae + g*idmae) c(g)}	0,00001	0,776	0,011	0,619	0,010	2.983	19,921
Ponderação (MARK®)*		0,776	0,011	0,619	0,010	2.983	19,895

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

NOSSA SENHORA DO SOCORRO (2806)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,62377	0,911	0,008	0,714	0,011	1.856	4,687
{p(g) c(g + idmae + g*idmae)}	0,24275	0,912	0,008	0,714	0,011	1.855	4,542
{p(g + idmae) c(g + idmae + g*idmae)}	0,13328	0,912	0,008	0,714	0,011	1.855	4,547
{p(g + idmae + g*idmae) c(g + idmae)}	0,00011	0,911	0,008	0,716	0,011	1.856	4,688
{p(g) c(g + idmae)}	0,00004	0,912	0,008	0,716	0,011	1.855	4,546
{p(g + idmae) c(g + idmae)}	0,00002	0,912	0,008	0,716	0,011	1.855	4,548
{p(g + idmae + g*idmae) c(g)}	0,00001	0,911	0,008	0,715	0,011	1.856	4,687
Ponderação (MARK®)*		0,912	0,008	0,714	0,011	1.856	4,633

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

PROPRIÁ (2807)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,62377	0,880	0,010	0,800	0,011	1.658	6,256
{p(g) c(g + idmae + g*idmae)}	0,24275	0,881	0,010	0,800	0,011	1.658	6,208
{p(g + idmae) c(g + idmae + g*idmae)}	0,13328	0,881	0,010	0,800	0,011	1.658	6,215
{p(g + idmae + g*idmae) c(g + idmae)}	0,00011	0,880	0,010	0,797	0,011	1.658	6,256
{p(g) c(g + idmae)}	0,00004	0,881	0,010	0,797	0,011	1.658	6,208
{p(g + idmae) c(g + idmae)}	0,00002	0,881	0,010	0,797	0,011	1.658	6,215
{p(g + idmae + g*idmae) c(g)}	0,00001	0,880	0,010	0,796	0,011	1.658	6,255
Ponderação (MARK®)*		0,880	0,010	0,800	0,011	1.658	6,239

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ANEXO VIII

Modelos de Huggins para populações fechadas que apresentaram algum peso ponderado, com as respectivas estimativas de peso ponderado (w_i), probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) e estimativas ponderadas por w_i . Microrregiões de residência da mãe em Sergipe, pareamento mais restrito, segundo e terceiro trimestres de 2006.

SERGIPANA DO SERTÃO DE SÃO FRANCISCO (28001)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,761	0,013	0,569	0,012	2.205	18,814
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,761	0,013	0,569	0,012	2.205	18,807
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,759	0,013	0,569	0,012	2.207	19,268
{p(g) c(g)}	0,00001	0,761	0,013	0,571	0,012	2.205	18,809
{p(g) c(g + idmae)}	0,00001	0,761	0,013	0,571	0,012	2.205	18,813
{p(g + idmae) c(g)}	0,00001	0,761	0,013	0,571	0,012	2.205	18,807
Ponderação (MARK[®])*		0,761	0,013	0,569	0,012	2.205	18,812

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK[®].

CARIRA (28002)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,803	0,021	0,675	0,020	683	7,816
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,803	0,021	0,675	0,020	683	7,823
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,803	0,021	0,675	0,020	684	7,939
{p(g) c(g)}	0,00001	0,803	0,021	0,676	0,020	683	7,813
{p(g) c(g + idmae)}	0,00001	0,803	0,021	0,676	0,020	683	7,815
{p(g + idmae) c(g)}	0,00001	0,803	0,021	0,676	0,020	683	7,823
Ponderação (MARK®)*		0,803	0,021	0,675	0,020	683	7,818

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

NOSSA SENHORA DAS DORES (28003)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,800	0,021	0,675	0,020	705	8,153
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,800	0,021	0,675	0,020	705	8,155
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,800	0,021	0,675	0,020	705	8,154
{p(g) c(g)}	0,00001	0,800	0,021	0,674	0,020	705	8,161
{p(g) c(g + idmae)}	0,00001	0,800	0,021	0,674	0,020	705	8,153
{p(g + idmae) c(g)}	0,00001	0,800	0,021	0,674	0,020	705	8,156
Ponderação (MARK®)*		0,800	0,021	0,675	0,020	705	8,154

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

AGRESTE DE ITABAIANA (28004)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,824	0,012	0,759	0,011	1.818	10,903
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,824	0,012	0,759	0,011	1.819	10,906
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,824	0,012	0,759	0,011	1.818	10,902
{p(g) c(g)}	0,00001	0,824	0,012	0,758	0,011	1.818	10,899
{p(g) c(g + idmae)}	0,00001	0,824	0,012	0,757	0,011	1.818	10,902
{p(g + idmae) c(g)}	0,00001	0,824	0,012	0,758	0,011	1.819	10,906
Ponderação (MARK®)*		0,824	0,012	0,759	0,011	1.818	10,904

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

TOBIAS BARRETO (28005)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,706	0,020	0,559	0,016	1.383	20,966
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,706	0,020	0,559	0,016	1.382	20,957
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,707	0,020	0,559	0,016	1.384	21,572
{p(g) c(g)}	0,00001	0,706	0,020	0,559	0,016	1.383	20,975
{p(g) c(g + idmae)}	0,00001	0,706	0,020	0,559	0,016	1.383	20,968
{p(g + idmae) c(g)}	0,00001	0,706	0,020	0,559	0,016	1.382	20,957
Ponderação (MARK®)*		0,706	0,020	0,559	0,016	1.383	20,963

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

AGRESTE DE LAGARTO (28006)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,520	0,026	0,319	0,014	2.087	71,034
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,520	0,026	0,319	0,014	2.088	71,227
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,519	0,026	0,319	0,014	2.093	73,111
{p(g) c(g)}	0,00001	0,520	0,026	0,320	0,014	2.087	71,033
{p(g) c(g + idmae)}	0,00001	0,520	0,026	0,319	0,014	2.087	71,057
{p(g + idmae) c(g)}	0,00001	0,520	0,026	0,320	0,014	2.088	71,231
Ponderação (MARK®)*		0,520	0,026	0,319	0,014	2.088	71,101

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

PROPRIÁ (28007)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,641	0,024	0,591	0,017	1.352	29,936
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,641	0,024	0,591	0,017	1.352	29,908
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,636	0,025	0,591	0,017	1.355	30,987
{p(g) c(g)}	0,00001	0,641	0,024	0,588	0,017	1.352	29,932
{p(g) c(g + idmae)}	0,00001	0,641	0,024	0,589	0,017	1.352	29,938
{p(g + idmae) c(g)}	0,00001	0,641	0,024	0,588	0,017	1.352	29,911
Ponderação (MARK®)*		0,641	0,024	0,591	0,017	1.352	29,928

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

CONTIGUIBA (28008)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,745	0,026	0,673	0,021	648	11,283
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,745	0,026	0,673	0,021	648	11,288
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,745	0,026	0,673	0,021	648	11,360
{p(g) c(g)}	0,00001	0,745	0,026	0,673	0,021	648	11,284
{p(g) c(g + idmae)}	0,00001	0,745	0,026	0,673	0,021	648	11,283
{p(g + idmae) c(g)}	0,00001	0,745	0,026	0,673	0,021	648	11,285
Ponderação (MARK®)*		0,745	0,026	0,673	0,021	648	11,285

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

JAPARATUBA (28009)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,684	0,028	0,587	0,021	805	18,210
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,683	0,028	0,587	0,021	805	18,237
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,686	0,028	0,587	0,021	805	18,432
{p(g) c(g)}	0,00001	0,684	0,028	0,580	0,021	805	18,216
{p(g) c(g + idmae)}	0,00001	0,684	0,028	0,581	0,021	805	18,216
{p(g + idmae) c(g)}	0,00001	0,683	0,028	0,580	0,021	805	18,240
Ponderação (MARK®)*		0,683	0,028	0,587	0,021	805	18,219

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

BAIXO CONTIGUIBA (28010)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,792	0,016	0,742	0,014	1.194	11,242
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,791	0,016	0,742	0,014	1.194	11,237
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,789	0,017	0,742	0,014	1.196	11,897
{p(g) c(g)}	0,00001	0,792	0,016	0,745	0,014	1.194	11,240
{p(g) c(g + idmae)}	0,00001	0,792	0,016	0,745	0,014	1.194	11,245
{p(g + idmae) c(g)}	0,00001	0,791	0,016	0,745	0,014	1.194	11,237
Ponderação (MARK®)*		0,791	0,016	0,742	0,014	1.194	11,241

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ARACAJU (28011)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,847	0,005	0,749	0,005	8.614	19,586
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,847	0,005	0,749	0,005	8.614	19,593
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,847	0,005	0,749	0,005	8.614	19,594
{p(g) c(g)}	0,00001	0,847	0,005	0,750	0,005	8.614	19,582
{p(g) c(g + idmae)}	0,00001	0,847	0,005	0,750	0,005	8.614	19,585
{p(g + idmae) c(g)}	0,00001	0,847	0,005	0,750	0,005	8.614	19,592
Ponderação (MARK®)*		0,847	0,005	0,749	0,005	8.614	19,588

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

BOQUIM (28012)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,781	0,013	0,649	0,012	2.114	16,056
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,781	0,013	0,649	0,012	2.114	16,059
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,781	0,013	0,649	0,012	2.114	16,068
{p(g) c(g)}	0,00001	0,781	0,013	0,650	0,012	2.114	16,058
{p(g) c(g + idmae)}	0,00001	0,781	0,013	0,650	0,012	2.114	16,054
{p(g + idmae) c(g)}	0,00001	0,781	0,013	0,650	0,012	2.114	16,060
Ponderação (MARK®)*		0,781	0,013	0,649	0,012	2.114	16,057

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ESTÂNCIA (28013)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,65520	0,863	0,011	0,715	0,012	1.544	7,188
{p(g + idmae) c(g + idmae + g*idmae)}	0,34405	0,863	0,011	0,715	0,012	1.544	7,189
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,00071	0,862	0,011	0,715	0,012	1.544	7,193
{p(g) c(g)}	0,00001	0,863	0,011	0,716	0,012	1.544	7,166
{p(g) c(g + idmae)}	0,00001	0,863	0,011	0,716	0,012	1.544	7,186
{p(g + idmae) c(g)}	0,00001	0,862	0,011	0,715	0,012	1.544	7,189
Ponderação (MARK®)*		0,863	0,011	0,715	0,012	1.544	7,188

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ANEXO IX

Modelos de Huggins para populações fechadas que apresentaram algum peso ponderado, com as respectivas estimativas de peso ponderado (w_i), probabilidade de captura pelo SINASC (\hat{p}), probabilidade de captura pelo Registro Civil (\hat{c}), total de nascidos vivos (\hat{N}) e estimativas ponderadas por w_i . Regionais de saúde de residência da mãe em Sergipe, pareamento mais restrito, segundo e terceiro trimestres de 2006.

ARACAJU (2801)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,67644	0,834	0,005	0,766	0,005	8.141	21,166
{p(g + idmae) c(g + idmae + g*idmae)}	0,29975	0,835	0,005	0,766	0,005	8.141	21,175
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,02219	0,834	0,005	0,766	0,005	8.142	21,207
{p(g) c(g)}	0,00072	0,834	0,005	0,767	0,005	8.141	21,167
{p(g) c(g + idmae)}	0,00038	0,834	0,005	0,766	0,005	8.141	21,169
{p(g + idmae) c(g)}	0,00032	0,835	0,005	0,767	0,005	8.141	21,178
{p(g + idmae) c(g + idmae)}	0,00017	0,835	0,005	0,766	0,005	8.141	21,176
{p(g + idmae + g*idmae) c(g)}	0,00002	0,834	0,005	0,767	0,005	8.142	21,210
{p(g + idmae + g*idmae) c(g + idmae)}	0,00001	0,834	0,005	0,766	0,005	8.142	21,207
Ponderação (MARK®)*		0,834	0,005	0,766	0,005	8.141	21,170

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ESTÂNCIA (2802)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,67644	0,842	0,008	0,691	0,009	3.313	12,677
{p(g + idmae) c(g + idmae + g*idmae)}	0,29975	0,842	0,008	0,691	0,009	3.313	12,678
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,02219	0,842	0,008	0,691	0,009	3.314	12,841
{p(g) c(g)}	0,00072	0,842	0,008	0,691	0,009	3.313	12,678
{p(g) c(g + idmae)}	0,00038	0,842	0,008	0,691	0,009	3.313	12,679
{p(g + idmae) c(g)}	0,00032	0,842	0,008	0,691	0,009	3.313	12,677
{p(g + idmae) c(g + idmae)}	0,00017	0,842	0,008	0,691	0,009	3.313	12,678
{p(g + idmae + g*idmae) c(g)}	0,00002	0,842	0,008	0,691	0,009	3.314	12,841
{p(g + idmae + g*idmae) c(g + idmae)}	0,00001	0,842	0,008	0,691	0,009	3.314	12,839
Ponderação (MARK®)*		0,842	0,008	0,691	0,009	3.313	12,681

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

ITABAIANA (2803)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,67644	0,827	0,009	0,677	0,009	2.969	13,531
{p(g + idmae) c(g + idmae + g*idmae)}	0,29975	0,827	0,009	0,677	0,009	2.969	13,530
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,02219	0,827	0,009	0,677	0,009	2.970	13,558
{p(g) c(g)}	0,00072	0,827	0,009	0,677	0,009	2.969	13,531
{p(g) c(g + idmae)}	0,00038	0,827	0,009	0,677	0,009	2.969	13,527
{p(g + idmae) c(g)}	0,00032	0,827	0,009	0,677	0,009	2.969	13,532
{p(g + idmae) c(g + idmae)}	0,00017	0,827	0,009	0,677	0,009	2.969	13,530
{p(g + idmae + g*idmae) c(g)}	0,00002	0,827	0,009	0,677	0,009	2.970	13,558
{p(g + idmae + g*idmae) c(g + idmae)}	0,00001	0,827	0,009	0,677	0,009	2.970	13,558
Ponderação (MARK®)*		0,827	0,009	0,677	0,009	2.969	13,531

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

LAGARTO (2804)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,67644	0,814	0,011	0,735	0,010	2.427	13,562
{p(g + idmae) c(g + idmae + g*idmae)}	0,29975	0,814	0,011	0,735	0,010	2.427	13,564
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,02219	0,814	0,011	0,735	0,010	2.427	13,568
{p(g) c(g)}	0,00072	0,814	0,011	0,734	0,010	2.427	13,557
{p(g) c(g + idmae)}	0,00038	0,814	0,011	0,734	0,010	2.427	13,559
{p(g + idmae) c(g)}	0,00032	0,814	0,011	0,734	0,010	2.427	13,563
{p(g + idmae) c(g + idmae)}	0,00017	0,814	0,011	0,734	0,010	2.427	13,563
{p(g + idmae + g*idmae) c(g)}	0,00002	0,814	0,011	0,734	0,010	2.427	13,573
{p(g + idmae + g*idmae) c(g + idmae)}	0,00001	0,814	0,011	0,734	0,010	2.427	13,572
Ponderação (MARK®)*		0,814	0,011	0,735	0,010	2.427	13,563

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

NOSSA SENHORA DA GLÓRIA (2805)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,67644	0,611	0,015	0,446	0,010	3.821	59,240
{p(g + idmae) c(g + idmae + g*idmae)}	0,29975	0,611	0,015	0,446	0,010	3.822	59,274
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,02219	0,611	0,015	0,446	0,010	3.822	59,281
{p(g) c(g)}	0,00072	0,611	0,015	0,447	0,010	3.821	59,223
{p(g) c(g + idmae)}	0,00038	0,611	0,015	0,447	0,010	3.822	59,254
{p(g + idmae) c(g)}	0,00032	0,611	0,015	0,447	0,010	3.822	59,280
{p(g + idmae) c(g + idmae)}	0,00017	0,611	0,015	0,447	0,010	3.822	59,279
{p(g + idmae + g*idmae) c(g)}	0,00002	0,611	0,015	0,447	0,010	3.822	59,306
{p(g + idmae + g*idmae) c(g + idmae)}	0,00001	0,611	0,015	0,447	0,010	3.822	59,288
Ponderação (MARK®)*		0,611	0,015	0,446	0,010	3.821	59,251

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

NOSSA SENHORA DO SOCORRO (2806)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,67644	0,761	0,013	0,569	0,012	2.205	18,811
{p(g + idmae) c(g + idmae + g*idmae)}	0,29975	0,761	0,013	0,569	0,012	2.205	18,806
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,02219	0,759	0,013	0,569	0,012	2.207	19,267
{p(g) c(g)}	0,00072	0,761	0,013	0,571	0,012	2.205	18,817
{p(g) c(g + idmae)}	0,00038	0,761	0,013	0,571	0,012	2.205	18,816
{p(g + idmae) c(g)}	0,00032	0,761	0,013	0,571	0,012	2.205	18,807
{p(g + idmae) c(g + idmae)}	0,00017	0,761	0,013	0,571	0,012	2.205	18,808
{p(g + idmae + g*idmae) c(g)}	0,00002	0,759	0,013	0,571	0,012	2.207	19,271
{p(g + idmae + g*idmae) c(g + idmae)}	0,00001	0,759	0,013	0,571	0,012	2.207	19,269
Ponderação (MARK®)*		0,761	0,013	0,569	0,012	2.205	18,819

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.

PRÓPRIÁ (2807)

MODELO	w_i	\hat{p} (SINASC)	ERRO-PADRÃO \hat{p}	\hat{c} (RC)	ERRO-PADRÃO \hat{c}	\hat{N}	ERRO-PADRÃO \hat{N}
{p(g) c(g + idmae + g*idmae)}	0,67644	0,659	0,018	0,735	0,010	2.196	34,683
{p(g + idmae) c(g + idmae + g*idmae)}	0,29975	0,658	0,018	0,735	0,010	2.196	34,708
{p(g + idmae + g*idmae) c(g + idmae + g*idmae)}	0,02219	0,657	0,018	0,735	0,010	2.197	34,791
{p(g) c(g)}	0,00072	0,658	0,018	0,734	0,010	2.196	34,703
{p(g) c(g + idmae)}	0,00038	0,658	0,018	0,734	0,010	2.196	34,692
{p(g + idmae) c(g)}	0,00032	0,658	0,018	0,734	0,010	2.196	34,719
{p(g + idmae) c(g + idmae)}	0,00017	0,658	0,018	0,734	0,010	2.196	34,727
{p(g + idmae + g*idmae) c(g)}	0,00002	0,657	0,018	0,734	0,010	2.197	34,790
{p(g + idmae + g*idmae) c(g + idmae)}	0,00001	0,657	0,018	0,734	0,010	2.197	34,793
Ponderação (MARK®)*		0,658	0,018	0,735	0,010	2.196	34,693

* As estimativas do erro-padrão dos parâmetros ponderados são as simples ponderações por w_i fornecidas pelo Programa MARK®.