

Estimação em Pequenos Domínios

Fernando A. S. Moura

Instituto de Matemática
Universidade Federal do Rio de Janeiro
<http://www.dme.ufrj.br/~fmoura>

Minicurso 18^o SINAPE
28/07 a 01/08 de 2008
Estância de São Pedro - SP - Brasil

Para Gilda e Antonio (sempre presente), meus queridos pais,
por seus esforços incansáveis de me proverem uma boa formação.

Para minha esposa Inês pelo apoio e carinho.

Prefácio

A demanda por informações estatísticas em nível desagregado tem apresentado um crescimento considerável nos últimos anos. Este crescimento tem sido principalmente motivado pela necessidade das autoridades locais de obter um quadro mais preciso e atualizado da sua área de jurisdição. Além disso, os governos estaduais e federal também necessitam de informações em nível geográfico mais desagregado a fim de identificar subregiões menos desenvolvidas, auxiliando na elaboração de planos regionais de desenvolvimento. A dificuldade na obtenção de estimativas para pequenos domínios (ou comumente chamados de pequenas áreas) é que o tamanho da amostra resultante de uma pesquisa em âmbito nacional é muito pequeno para que estimativas baseadas somente no desenho amostral apresentem precisão aceitável. Em alguns casos constata-se a inexistência de unidades amostrais nos domínios de interesse. Vários métodos têm sido propostos e empregados com a finalidade de fornecer estimativas para pequenos domínios. Muitos destes métodos envolvem o uso de modelos de superpopulação na presença de informações auxiliares. Esta será a abordagem apresentada neste curso.

Nesta monografia apresentam-se os principais modelos encontrados na literatura, com aplicações bem sucedidas. Na primeira parte do curso abordamos os fundamentos teóricos da Teoria de Predição de População Finita, que servirão de base para a construção de modelos de previsão em pequenos domínios. Na segunda parte, apresentamos e ilustramos com aplicações alguns modelos bastante empregados em problemas encontrados na literatura. A abordagem clássica e a bayesiana para o problema serão apresentadas, baseadas na experiência de trabalhos realizados pelo autor e por outros pesquisadores. Vale ressaltar que procuramos discutir em detalhes aqueles modelos em que o autor esteve diretamente envolvido no seu desenvolvimento.

Embora o curso tenha como público alvo pesquisadores envolvidos com o problema de obter estimativas para pequenos domínios, será também de interesse para aqueles que possam estar interessados apenas em inferência paramétrica em modelos de superpopulação. Os fundamentos teóricos mais especializados e necessários ao acompanhamento e desenvolvimento do curso serão abordados com a devida profundidade. Contudo, para uma maior compreensão sobre o conteúdo do curso, é necessário que o aluno já tenha feito um curso de graduação em estatística ou tenha conhecimentos de inferência estatística, modelos de regressão e noções de amostragem em populações finitas.

Uma parte substancial deste livro relata alguns trabalhos realizados por mim e outros colaboradores, incluindo professores e alunos da pós-graduação em Estatística do Instituto de Matemática da UFRJ. Registro meus agradecimentos à eles. Agradeço à Comissão Organizadora do 18^o SINAPE e à Associação Brasileira de Estatística pela oportunidade de divulgação do nosso trabalho.

Rio de Janeiro, 06 de fevereiro de 2008.

Fernando A. S. Moura

Sumário

1	Introdução	1
1.1	Estrutura do Livro	4
2	Modelos de Superpopulação	7
2.1	Introdução	7
2.2	Notação e Terminologia	9
2.3	Inferência em Modelos de Superpopulação	11
2.3.1	Previsores Lineares Ótimos (BLUP)	13
2.3.2	Previsores Lineares Empíricos (EBLUP)	17
2.4	Modelos Lineares Mistos	20
2.4.1	Predição em Modelos Lineares Mistos: EBLUP	22
2.4.2	Erro Quadrático Médio do EBLUP	23
2.4.3	Estimador do Erro Quadrático Médio do EBLUP	28
2.5	Amostragem Informativa	29
2.5.1	Definição Formal de Ignorabilidade	30
2.5.2	Tratamento da Não-Resposta	31
2.6	Introdução à Inferência Bayesiana em Populações Finitas	34
2.6.1	Modelo Paramétrico na Família Exponencial	35

3	Modelos Lineares Mistos	41
3.1	Modelo no Nível de Pequenos Domínios	42
3.2	Modelo no Nível de Unidades Populacionais	44
3.3	Modelo Multinível de Moura e Holt	49
3.3.1	O BLUP da Média para o Pequeno Domínio	50
3.3.2	Aproximação do Erro Quadrático Médio	51
3.3.3	Estimação do Erro Quadrático Médio	53
3.3.4	Um Estudo de Simulação	54
4	Modelos Lineares Mistos Generalizados	59
4.1	Estimação	60
4.2	Modelo Hierárquico Logístico	62
4.3	Exemplo 4.1: Modelo Logístico para Proficiência Escolar	64
4.4	Exemplo 4.2: Modelo de Poisson para Incidências de Doenças	70
4.4.1	Modelo Poisson-Gamma	72
4.4.2	Modelo Log-Normal	74
5	Tópicos Avançados	77
5.1	Modelos com Erros Correlacionados	77
5.2	Modelos Espacialmente Estruturados	79
5.3	Exemplo 5.1: Modelo Espacial Logístico	81
5.4	Exemplo 5.2: Estimação de Populações Municipais	85
5.4.1	Conjunto de Dados	86
5.4.2	Descrição do Modelo	87
5.4.3	Modelagem da Variância das Observações	89
5.4.4	Aspectos Computacionais e Alguns Resultados	90
	Referências Bibliográficas	93

Capítulo 1

Introdução

O interesse em se obter informações sobre características populacionais em nível desagregado já vem de longa data. Brackstone (1987) menciona a existência de um livro na Inglaterra no século XI, denominado "Domesday" em que se registravam informações sobre proprietários de terra, com a finalidade principal de cobrança de impostos. Há também registro histórico de um Censo no Canadá em 1666.

Naquela época as estatísticas em nível desagregado eram baseadas numa investigação completa da população, ou de registros administrativos. Certamente, limitações de recursos e tecnologia somente permitiam que essa tarefa fosse realizada intermitentemente.

Nos últimos 50 anos, pesquisas por amostragem vêm se consolidando como uma forma efetiva de prover estimativas nacionais e subnacionais com maior periodicidade. Contudo, em geral, dada a inabilidade de pesquisas por amostragem produzirem estimativas com precisão aceitável para pequenos domínios de estudos, quando o estimador de expansão direta do desenho é empregado, pouca atenção foi dada ao problema. O desempenho do estimador de expansão

baseado no desenho amostral é insatisfatória porque geralmente o plano amostral é desenhado para prover estimativas acuradas num nível mais agregado do que os pequenos domínios de estudos, geralmente requeridos *a posteriori*. Consideramos que um domínio é pequeno se o tamanho da amostra no respectivo domínio não é suficientemente grande para que o uso do estimador de expansão direta produza estimativa de precisão aceitável. Outros autores preferem utilizar o termo "pequena área" para designar domínio ou área de estudo em que o tamanho da amostra é pequeno (ou mesmo inexistente) para produzir estimativa direta de precisão aceitável. Preferimos usar o termo pequeno domínio para enfatizar que o problema está relacionado ao tamanho da amostra e não ao tamanho da área.

Por outro lado, o interesse por estimativas confiáveis para pequenos domínios vem crescendo mais recentemente, tanto por parte do setor público quanto do setor privado. Este crescimento é motivado pela necessidade de se obter um quadro mais preciso de informação em nível mais desagregado do que é geralmente obtido por pesquisas periódicas por amostragem. Por exemplo, podem existir regiões ou subgrupos da população que estejam muito abaixo do índice de desenvolvimento médio do país. Portanto, há a necessidade de se indentificarem estas regiões, a partir de informação estatística confiável neste nível, para que se possa planejar uma política desenvolvimentista diferenciada. Recentemente, o Banco Mundial vem implantando, juntamente com, as agências de estatísticas de países em desenvolvimento, um projeto de mapeamento de indicadores de pobreza para subregiões. A metodologia está, em princípio, descrita em Elbers et al. (2002). Autoridades locais também precisam obter informações sobre as suas respectivas jurisdições, tais como atividade econômica, emprego e taxa de desemprego, características demográficas e etc, com o propósito de planejamento. Muitas políticas de decisões industriais

são baseadas ou sustentadas por informações sócio-econômicas locais, gerando-se demanda do setor privado por estatísticas desagregadas. Além disso, há uma forte demanda por dados desagregados por parte de analistas e usuários que desejam estender sua análise com maior detalhamento. Por exemplo, demógrafos podem estar interessados em obter estimativas de fertilidade por etnia e idade. Uma classificação refinada como esta resultaria em pequenos domínios de estudo.

No Brasil, como em todo o mundo, a demanda por estimação para pequenos domínios vem se tornando mais premente, pelas mesmas razões expostas acima. Um exemplo da necessidade de estimativas confiáveis é a distribuição de verbas federais pelo Fundo de Participação dos Municípios, instituída por legislação federal e definida a partir de critérios que levam em consideração os tamanhos das populações municipais. É preciso, então, obter estimativas que atendam às necessidades de pesquisa e de planejamento e que apresentem boa precisão. No caso específico da obtenção de estimativas das populações municipais, são realizadas projeções populacionais. Primeiramente é feita a projeção para uma área maior, considerando sua evolução populacional com base em hipóteses sobre a natalidade, a mortalidade e a migração, e depois o resultado é repartido de alguma forma entre os pequenos domínios. Contudo, algumas técnicas de projeção caracterizam-se por não apresentar o erro das estimativas, como aquela conhecida por método dos coeficientes, consultar IBGE (2002).

Dessa forma, destacam-se duas questões importantes ao considerarmos o problema de estimação em pequenos domínios: a primeira diz respeito a como produzir estimativas confiáveis das características de interesse investigadas, baseando-se na informação proveniente de amostras muito pequenas; a segunda está relacionada a como obter o erro de estimação. Com raras exceções,

o desenho e o tamanho da amostra são escolhidos de maneira que a amostra forneça estimativas com boa precisão para níveis geográficos maiores, formados pela agregação de pequenos domínios, ou para grupos demográficos mais amplos, optando-se, como exemplo, por não apresentar estimativas por faixas de idade, mas sim para o agregado. Existindo somente uma amostra pequena para uma particular área, uma solução possível para o problema de estimação é pedir emprestada informação de outros dados relacionados, tais como: dados obtidos para a característica de interesse em outros domínios considerados "similares" ao domínio em questão e/ou dados obtidos para a característica em ocasiões anteriores.

Os fatos apresentados fizeram com que crescessem na literatura as propostas de técnicas, com fundamentos teóricos sólidos, baseadas em estimadores indiretos onde ocorre a troca de informação entre os domínios. Desta forma, eleva-se o tamanho efetivo das amostras e, conseqüentemente, aumenta-se a precisão das estimativas. Por outro lado, a implantação destas técnicas não seria viável sem o progresso considerável de técnicas computacionais avançadas e sistemas computacionais cada vez mais eficientes. A estimação baseada em modelos de superpopulação destaca-se entre as alternativas propostas e é a adotada neste livro.

1.1 Estrutura do Livro

Este livro está estruturado em 5 capítulos. No Capítulo 2 introduz-se o conceito de modelos de superpopulação sob o enfoque de estimação clássica e *bayesiana*. O Capítulo 2 é fundamental para o entendimento de todo o conteúdo abordado nos capítulos posteriores.

O Capítulo 3 introduz os modelos lineares mistos, discutindo-se as suas

vantagens em relação a outros modelos. Obtêm-se os preditores das quantidades de interesse nos pequenos domínios, como também expressões para os seus respectivos erros quadráticos médios. Apresentam-se os principais modelos discutidos na literatura para cada tipo de informação disponível nos pequenos domínios, como também algumas aplicações dos mesmos a dados reais.

Uma vez introduzido o modelo linear misto, o Capítulo 4 apresenta modelos para o caso em que a variável resposta é discreta e/ou a função de ligação não é a identidade. Descreve-se o modelo linear geral misto para a família exponencial e apresentam-se com detalhes algumas aplicações à estimação em pequenos domínios, utilizando-se dados reais. Finalmente o Capítulo 5 apresenta alguns modelos com estrutura mais complexa e alguns exemplos de aplicações com dados reais.

Capítulo 2

Modelos de Superpopulação

2.1 Introdução

A Teoria Convencional da Amostragem para Populações Finitas, tal como em Bolfarine e Bussab (2005) ou em Cochran (1977), tem como objetivo principal estimar funções de uma ou mais características de interesse da população. Para alcançar este objetivo, um plano probabilístico de seleção da amostra é proposto. Este plano amostral é geralmente desenhado de forma a fornecer estimativas das características de interesse com precisões aceitáveis em domínios pré-fixados. Por exemplo, se uma pesquisa sobre as condições de saúde de uma população tem como domínio de interesse todos os municípios que compõem um determinado estado, então é desejável que o plano amostral seja estratificado por município. A escolha do plano amostral pode também estar sujeita a restrições de ordem prática. Uma amostra de alunos de uma determinada escola pode ser sorteada por meio de uma amostra de turmas (amostragem de conglomerados). A utilização de variáveis auxiliares, relacionadas com a variável de interesse pode também produzir estimadores mais precisos, como

por exemplo: o uso do estimador estratificado; de regressão e de razão.

As principais características da Teoria Convencional da Amostragem para Populações Finitas são:

- os parâmetros de interesse são funções das características de interesse das unidades populacionais, como por exemplo a média populacional:

$$\bar{Y}_i = \frac{1}{N} \sum_{i=1}^N y_i;$$

- a toda unidade amostral está associada uma probabilidade de inclusão na amostra, conhecida e diferente de zero;

- o único processo de aleatorização é devido à seleção da amostra;

- não há necessidade de se associar uma distribuição de probabilidade às características de interesse;

- as propriedades do estimador proposto são avaliadas sob a sua distribuição amostral, induzida pelo plano amostral empregado;

- via de regra, o interesse é obter apenas estimativas pontuais das quantidades de interesse e de suas variâncias;

- inferência intervalar para variáveis contínuas é apenas adequada para grandes amostras e, quando a fração de amostragem for pequena, via versão do Teorema Central do Limite para Populações Finitas, consultar Bolfarine e Bussab (2005).

Desde o início do século passado, a Teoria Convencional da Amostragem para Populações Finitas tem tido um papel importante como uma metodologia para se obterem informações acuradas de uma população para fins descritivos, como também analíticos. Desde então, governos necessitam obter estatísticas descritivas da população para fins de avaliar e implantar as suas políticas públicas e sociais. Por outro lado, é também de suma importância, entender e identificar mecanismos casuais que governam as relações entre variáveis de interesse e as razões pelas quais as características de interesse variam em função

de outras variáveis e/ou diferentes grupos da população. Neste caso, diz-se que o propósito da pesquisa é de uso analítico. Contudo, é importante notar que investigações sociais de uma população apresentam estruturas complexas, com diferenças sistemáticas entre subpopulações que podem ser devidas a localização geográfica, características dos indivíduos ou outros fatores relacionados com os indivíduos, ou grupos de indivíduos. Portanto, um desenho amostral apropriado, deveria levar em conta a estrutura da população, a fim de incorporar as complexidades da estratificação, os diferentes níveis de conglomerados e as medidas de tamanhos desses conglomerados. Como veremos mais adiante, a modelagem explícita de estruturas populacionais por meio de modelos de superpopulação resultam, para alguns casos especiais, nas mesmas inferências pontuais de parâmetros de interesse da população, realizadas sob a amostragem convencional.

2.2 Notação e Terminologia

Uma população finita é um conjunto de N unidades, com pelo menos uma característica em comum. Esta característica em comum permite determinar inequivocamente quais os elementos que pertencem à população. Esse conjunto é representado por U e os seus elementos são rotulados de 1 a N : $U = \{1, 2, \dots, N\}$. A cada unidade da população está associada uma (ou mais) característica de interesse, denotada por $y = (y_1, \dots, y_N)^T$. Um problema geral de interesse é inferir sobre uma função $h(y_1, \dots, y_N)$, por exemplo: a média populacional $\bar{Y} = N^{-1} \sum_{i=1}^N y_i$ por meio da observação de alguns valores da característica de interesse y obtidos de uma escolha de uma amostra s da população U .

A abordagem de modelos de superpopulação considera que os valores

y_1, \dots, y_N são realizações dos vetores aleatórios Y_1, \dots, Y_N . Denotamos a distribuição conjunta de Y_1, \dots, Y_N por ξ .

Exemplo 2.1: Considere modelo de superpopulação em que a nota y_{ik} do aluno $i = 1, \dots, N$ no teste $k = 1, 2$ sejam observações independentes e identicamente distribuídas de uma normal bivariada de média $\mu = (\mu_1, \mu_2)$ e matriz de covariância Σ . Pode-se estar interessado em inferir sobre os parâmetros do modelo de superpopulação μ e Σ ou o vetor das médias da população $\bar{Y} = \sum_{i=1}^N y_i$ por meio de uma observação de uma amostra s de U .

Contudo, se N for grande, pouca diferença numérica há entre o parâmetro do modelo populacional e o da função da característica populacional correspondente:

$$\bar{Y} = \mu + O_p(N^{-1/2})$$

Definição 2.1: Seja \mathfrak{S} o conjunto de todas as amostras s de U , tal que o tamanho da amostra $n(s) \leq N$, ou seja \mathfrak{S} é o conjunto das partes de U . Um plano amostral ou planejamento amostral é uma função de probabilidade $p(s) = Pr(S = s)$ definida para todo $s \in \mathfrak{S}$, logo $p(s)$ satisfaz:

$$p(s) \geq 0, \forall s \in \mathfrak{S}; \quad e \quad \sum_{s \in \mathfrak{S}} p(s) = 1$$

Podemos então considerar que qualquer amostra $s \in \mathfrak{S}$ é uma realização de uma variável aleatória S . A probabilidade de inclusão de uma unidade $k \in U$ é a probabilidade desta unidade ser selecionada, e é dada por $\pi_k = \sum_{s \ni k} p(s)$, onde $s \ni k$ denota que o somatório é sobre todas as amostras s que contém a unidade k .

A probabilidade de inclusão conjunta da unidade k e l é dada por:

$$\pi_{kl} = \sum_{s \ni k, l} p(s).$$

Exemplo 2.2: Plano Amostral de Bernoulli

Suponha que os N elementos da população U sejam selecionados seqüencialmente em ordem crescente de acordo com os N resultados de N experimentos de Bernoulli independentes com probabilidade constante $0 < \pi < 1$ de qualquer unidade $k \in U$ ser incluída na amostra. Uma forma de executar na prática este plano amostral é gerar N variáveis uniformes i.i.d no intervalo $(0, 1)$: $\varepsilon_k \sim U(0, 1)$; $k = 1, \dots, N$ e adotar a seguinte regra de inclusão da unidade k , $\forall k = 1, \dots, N$: k é selecionada se e somente se o valor gerado de ε_k for menor do que π_k .

Desta forma para qualquer amostra $s \in \mathfrak{S}$ de tamanho $n(s) = 0, 1, \dots, N$, o plano amostral é dado por $p(s) = \pi^{n(s)}(1 - \pi)^{N-n(s)}$.

2.3 Inferência em Modelos de Superpopulação

Conforme descrito na Seção 2.2 deste capítulo, na abordagem de modelos de superpopulação, a inferência sobre uma função de interesse $h(y_1, \dots, y_N)$ procede com respeito à distribuição amostral de uma estatística \hat{h} sob repetidas realizações y_1, \dots, y_N geradas pelo modelo ξ com a amostra selecionada s permanecendo fixa. Algumas propriedades pontuais da estatística \hat{h} são definidas abaixo:

a) Vício sob o modelo ξ (ξ -vício):

$$\text{vício}_\xi(\hat{h}) = E_\xi(\hat{h} - h)$$

b) Variância do error de predição sob o modelo ξ (ξ -variância): $V_\xi(\hat{h} - h)$

c) Erro Quadrático Médio de predição sob o modelo ξ (ξ -EQM):

$$E_{\xi}(\hat{h} - h)^2 = V_{\xi}(\hat{h} - h) + \left\{ \text{vício}_{\xi}(\hat{h}) \right\}^2$$

Exemplo 2.3: Um Modelo para Amostra Aleatória Simples

Suponha uma população em que não haja nenhuma estrutura relevante, e que o investigador acredite que a variável de interesse y apresente comportamento semelhante para todas as unidades da população. Um modelo de superpopulação adequado para esta situação é dado por:

$$\begin{aligned} E_{\xi}(Y_i) &= \mu; \quad \forall i = 1, \dots, N \\ Cov_{\xi}(Y_i, Y_j) &= \sigma^2; \quad \forall i = j = 1, \dots, N \\ Cov_{\xi}(Y_i, Y_j) &= 0; \quad \forall i \neq j = 1, \dots, N \end{aligned} \quad (2.1)$$

Considere o seguinte preditor para o total $T = \sum_{i=1}^N y_i$, baseado nos valores de y obtidos numa amostra s de tamanho n :

$\hat{T} = \sum_{i \in s} y_i + \sum_{i \notin s} \hat{y}_i = \sum_{i \in s} y_i + \sum_{i \notin s} \bar{y}_s = N\bar{y}_s$, onde $\bar{y}_s = n^{-1} \sum_{i \in s} Y_i$. Então é fácil verificar que \hat{T} é um preditor não viciado sob o modelo ξ em (2.1):

$E_{\xi}(\hat{T} - T) = NE_{\xi}(\bar{y}_s) - \sum_{i=1}^N E_{\xi}(Y_i) = N\mu - N\mu = 0$. Além disso, $V_{\xi}(\hat{T} - T) = V_{\xi} \left\{ (N - n)\bar{y}_s - \sum_{i \notin s} Y_i \right\} = N^2 \frac{\sigma^2}{n} (1 - f)$, onde $f = n/N$ é denominada de fração de amostragem. As fórmulas do preditor e de sua variância são semelhantes àquelas obtidas por meio da abordagem convencional da Teoria das Populações Finitas para um plano de amostragem aleatória simples sem reposição. Note que a diferença numérica é devida apenas ao parâmetro de variabilidade populacional σ^2 . No caso da abordagem convencional ele é substituído por $S^2 = (N - 1)^{-1} \sum_{i=1}^N (y_i - \bar{Y})^2$. Contudo, a diferença entre as duas quantidades tende a ser insignificante para uma população com tamanho N grande.

2.3.1 Previsores Lineares Ótimos (BLUP)

Nesta seção obteremos o melhor preditor linear não viciado, comumente denominado BLUP ("Best Linear Unbiased Predictor") para uma combinação linear dos valores da característica de interesse na população $Y = (Y_1, \dots, Y_N)^T$ sob um modelo linear geral.

Definição 2.2: Um preditor \hat{h} de uma combinação linear dos valores de Y na população $h = \gamma^T Y$ é dito linear, se puder ser escrito como: $\hat{h} = g_s^T Y_s$, onde: $g_s = (g_1, \dots, g_n)^T$ é um vetor de coeficientes conhecidos; $Y_s = (Y_1, \dots, Y_n)^T$ é o vetor aleatório associado aos valores de Y observados na amostra s .

Se, por exemplo, $\gamma_i = 1, \forall i$, então h é o total populacional, se $\gamma_i = 1/N$, tem-se que h é a média populacional.

Definição 2.3: Um preditor linear de uma quantidade $h = \gamma^T Y$ é denominado de "o melhor preditor linear não viciado" (BLUP), se for não viciado e tiver a menor variância do erro de predição na classe de todos os preditores lineares e não viciados de h .

Vimos no Exemplo 2.3 que o preditor linear $\hat{T} = N\bar{y}_s$ é não viciado para o total populacional $T = \sum_{i=1}^N Y_i$, sob o modelo descrito em (2.1). Além disso, não é difícil mostrar que possui a menor variância do erro de predição na classe de todos os preditores lineares não viciados. O Teorema a seguir devido a Royall (1976) fornece o BLUP para um modelo linear bastante geral.

Considere o seguinte modelo de superpopulação linear geral:

$$\begin{aligned} E_{\xi}(Y) &= X\beta \\ V_{\xi}(Y) &= V \end{aligned} \quad (2.2)$$

onde: X é uma matriz $N \times p$ de p covariáveis auxiliares, supostamente co-

nhecidas para todas as N unidades populacionais; β é um vetor $p \times 1$ de parâmetros desconhecidos e V é uma matriz de covariância positiva definida. Rearranjando-se os elementos da população de tal forma que os primeiros n elementos são os pertencentes a amostra s , podemos expressar X e V como:

$$X = \begin{pmatrix} X_s \\ X_{\bar{s}} \end{pmatrix} \text{ e } V = \begin{pmatrix} V_{ss} & V_{s\bar{s}} \\ V_{\bar{s}s} & V_{\bar{s}\bar{s}} \end{pmatrix}.$$

Note que por construção X_s é $n \times p$, $X_{\bar{s}}$ é $(N - n) \times p$, V_{ss} é $n \times n$, $V_{\bar{s}\bar{s}}$ é $(N - n) \times (N - n)$, $V_{s\bar{s}}$ é $n \times (N - n)$ e $V_{\bar{s}s} = V_{s\bar{s}}^T$. Assume-se também que V_{ss} é definida positiva.

Teorema 2.1: Sob o modelo (2.2), com V conhecida, o melhor preditor não viciado de $h = \gamma^T Y$ é dado por:

$$\hat{h}_B = \gamma_s^T Y_s + \gamma_{\bar{s}}^T \left\{ X_{\bar{s}} \tilde{\beta} + V_{\bar{s}s} V_{ss}^{-1} (Y_s - X_s \tilde{\beta}) \right\} \quad (2.3)$$

onde $\tilde{\beta} = (X_s^T V_{ss}^{-1} X_s)^{-1} X_s^T V_{ss}^{-1} Y_s$ é o estimador de mínimos quadrados de β . A variância do erro de predição de \hat{h}_B é dada por:

$$\begin{aligned} V_\xi(\hat{h}_B - h) &= \gamma_{\bar{s}}^T (V_{\bar{s}\bar{s}} - V_{\bar{s}s} V_{ss}^{-1} V_{s\bar{s}}^T) \gamma_{\bar{s}} \\ &+ \gamma_{\bar{s}}^T \left\{ (X_{\bar{s}} - V_{\bar{s}s} V_{ss}^{-1} X_s) (X_s^T V_{ss}^{-1} X_s)^{-1} (X_{\bar{s}} - V_{\bar{s}s} V_{ss}^{-1} X_s)^T \right\} \gamma_s \end{aligned} \quad (2.4)$$

Prova:

Seja $g_s^T Y_s$ um preditor de $\gamma^T Y$ não viciado, então a variância do seu erro de predição é dada por:

$$\begin{aligned} V_\varepsilon(g_s^T Y_s - \gamma^T Y) &= V_\varepsilon([g_s - \gamma_s]^T Y_s - \gamma_{\bar{s}}^T Y_{\bar{s}}) = V_\varepsilon(a_s^T Y_s - \gamma_{\bar{s}}^T Y_{\bar{s}}) = \\ &= E_\varepsilon(a_s^T Y_s - \gamma_{\bar{s}}^T Y_{\bar{s}})^2 + \left\{ E_\varepsilon(a_s^T Y_s - \gamma_{\bar{s}}^T Y_{\bar{s}}) \right\}^2 = \\ &= a_s^T V_{ss} a_s - 2a_s^T V_{s\bar{s}} \gamma_{\bar{s}} + \gamma_{\bar{s}}^T V_{\bar{s}\bar{s}} \gamma_{\bar{s}} + \left\{ (a_s^T X_s - \gamma_{\bar{s}}^T X_{\bar{s}}) \beta \right\}^2 \end{aligned} \quad (2.5)$$

onde $a_s = g_s - \gamma_s$.

Uma vez que o preditor em (2.5) deve ser não viciado, o último termo de (2.5) se anula. O valor de a_s que minimiza (2.5) pode ser encontrado, minimizando-se com respeito a a_s a seguinte função de Lagrange:

$$L(a_s) = a_s^T V_{ss} a_s - 2a_s^T V_{s\bar{s}} \gamma_{\bar{s}} + 2(a_s^T X_s - \gamma_{\bar{s}}^T X_{\bar{s}}) \lambda \quad (2.6)$$

onde λ é o vetor de multiplicadores de Langrange.

Diferenciando-se $L(a_s)$ com respeito a a_s e igualando-se a zero o resultado obtido, tem-se que o valor de a_s que minimiza (2.6) é dado por:

$$a_{opt} = V_{ss}^{-1} [V_{s\bar{s}} + X_s A_s^{-1} (X_{\bar{s}}^T - X_s^T V_{ss}^{-1} V_{s\bar{s}})] \gamma_{\bar{s}} \quad (2.7)$$

onde $A_s = X_s^T V_{ss}^{-1} X_s$.

Finalmente, $g_{opt} = a_{opt} + \gamma_s$ e portanto (2.3) está demonstrado. A variância do erro de predição em (2.4) é obtida por substituição de a_{opt} em (2.5).

Um caso particular importante, ocorre quando $V_{s\bar{s}} = 0$. Neste caso o BLUP e sua variância do erro de predição são dados respectivamente por: $\hat{h}_B = \gamma_s^T Y_s + \gamma_{\bar{s}}^T X_{\bar{s}} \tilde{\beta}$ e $V_{\xi}(\hat{h}_B - h) = \gamma_{\bar{s}}^T (V_{ss} + X_{\bar{s}} V_{ss}^{-1} X_{\bar{s}}^T) \gamma_{\bar{s}}$. É importante notar que o modelo linear Geral (2.2) contempla uma classe bastante ampla de modelos. Vejamos alguns exemplos.

Exemplo 2.4 : Regressão Linear Simples

Considere o seguinte modelo linear de regressão: $Y_i = a + bx_i + \epsilon_i$, onde os ϵ_i 's são supostamente não correlacionados de média zero e variância σ^2 . Aplicando-se o Teorema (2.1), tem-se que o BLUP do total $T = \sum_{i=1}^N Y_i$ é dado por $T_B = N[\bar{y}_s + \tilde{b}(\bar{X} - \bar{x}_s)]$ e sua variância do erro de predição é dada por $V_{\xi}(\hat{T}_B - T) = N^2(1-f)(\sigma^2/n) \{1 + (\bar{x}_s - \bar{X})^2\} / \{(1-f)c_s\}$, onde $c_s = n^{-1} \sum_{i \in s} (x_i - \bar{x}_s)^2$ e $\tilde{b} = \frac{\sum_{i \in s} (y_i - \bar{y}_s)(x_i - \bar{x}_s)}{\sum_{i \in s} (x_i - \bar{x}_s)^2}$ é o estimador de mínimos

quadrados de b .

Exemplo 2.5: Amostragem Estratificada

Considere o seguinte modelo ξ de superpopulação:

$$\begin{aligned} E_{\xi}(Y_{hi}) &= \mu_h, \quad i = 1, \dots, N_h; \quad h = 1, \dots, L \\ V_{\xi}(Y_{hi}) &= \sigma_h^2 \\ Cov(Y_{hi}, Y_{h'i'}) &= 0, \quad \forall (h, i) \neq (h', i'). \end{aligned}$$

Suponha que se queira prever $\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$. Pode-se mostrar que o BLUP da média acima é $\hat{Y}_B = \frac{1}{N} \sum_{h=1}^L N_h \bar{Y}_h$ e sua variância do erro de predição é dada por: $V_{\xi}(\hat{Y}_B - Y) = \frac{1}{N^2} \sum_{h=1}^L N_h^2 \sigma_h^2 (1 - f_h)/n_h$, onde $f_h = n_h/N_h$ é a fração de amostragem no estrato $h = 1, \dots, L$. O resultado é idêntico àquele obtido através de um plano amostral estratificado com seleção aleatória e sem reposição (abordagem probabilística baseada no plano amostral).

É importante notar que se a variável Y for binária, tal que $E_{\xi}(Y_{hi}) = P_h$ e $V_{\xi}(Y_{hi}) = P_h(1 - P_h)$ e os Y_{hi} 's não são correlacionados, tem-se que o BLUP da proporção populacional é dado por $\hat{Y}_B = \sum_{h=1}^L (N_h/N) p_h$, onde p_h é a proporção amostral no estrato h . Este preditor coincide com o estimador usual de expansão para a amostragem estratificada aleatória simples sob a abordagem probabilística. Além disso, a fórmula da variância do erro de predição de \hat{Y}_B :

$V_{\xi}(\hat{Y}_B - \bar{Y}) = \sum_{h=1}^L W_h^2 P_h(1 - P_h)(1 - f_h)/n_h$, onde $W_h = N_h/N$ e a da variância do estimador usual estratificado sob a abordagem probabilística são semelhantes. Institutos de pesquisas de opiniões ou de mercado frequentemente utilizam amostragem por cotas, onde os estratos são formados de acordo com certas características de interesse. Pesquisadores selecionam indivíduos propositalmente de cada estrato, até que uma determinada cota (n_h) de respondente seja atingida. Este esquema garante que a distribuição da variável

de estratificação da amostra seja igual ao da população. Contudo, a falta de controle da seleção das unidades amostrais (indivíduos) em cada estrato pode permitir que o mecanismo não aleatório de seleção esteja relacionado com a variável de resposta e logo não podemos supô-lo ignorável. O uso na amostragem por cotas das fórmulas da amostragem estratificada aleatória simples para se obter estimativas de proporção e o seus respectivos erros amostrais associados é viável somente se o mecanismo de seleção for ignorável. Podemos analisar os dados assumindo que o mecanismo de seleção seja ignorável, mas a possibilidade real de que não o seja pode comprometer a validade da nossa inferência. A amostragem probabilística evita este problema e portanto aumenta a plausibilidade da inferência, freqüentista ou bayesiana.

É importante notar que o melhor preditor \hat{h}_{BP} de uma quantidade h é dado pela esperança condicional $E(h | Y)$, uma vez que para qualquer outro preditor \hat{h}^* de h , não necessariamente viciado ou linear tem-se que:

$$E[\hat{h}^* - h]^2 \geq E[E(h | Y) - h]^2$$

Admitindo-se normalidade, o melhor preditor (BP) é dado por (2.3) com $\tilde{\beta}$ substituído por β , ou seja, depende do parâmetro desconhecido β . Este preditor também é o melhor preditor linear (BLP) sem assumir normalidade. Para uma demonstração, ver Rao (2003).

2.3.2 Previsores Lineares Empíricos (EBLUP)

Na Seção anterior apresentamos preditores lineares ótimos para uma classe geral de modelos lineares que incluem os modelos mistos, (vistos no Capítulo 3) e etc. Contudo, as expressões obtidas para os preditores e sua variância do erro de predição dependem de parâmetros desconhecidos da matriz de covariância V . Vários métodos de estimação destes parâmetros estão disponíveis na lite-

ratura e implantados em diversas rotinas de pacotes computacionais. Dentre eles se destacam os métodos de máxima verossimilhança (MLE) e o de máxima verossimilhança restrita (RMLE), utilizados quando se assume normalidade. Para uma excelente revisão destes métodos de estimação para modelos mistos, ver Jiang (2007). O preditor obtido ao se substituir as estimativas dos parâmetros associados à matriz de covariância V é denominado "melhor preditor empírico linear não viciado" (EBLUP). Porém, é importante frisar que o EBLUP não possui o menor erro quadrático, uma vez que se tem apenas uma estimativa de V .

Exemplo 2.6: Modelo de Intercepto Aleatório

Considere o seguinte modelo de intercepto aleatório sem covariáveis explicativas para a nota y_{ij} de um estudante j numa escola i , onde $i = 1, \dots, M$ e $j = 1, \dots, N_i$:

$$\begin{aligned} Y_{ij} &= \alpha + \nu_i + \epsilon_{ij} \\ \epsilon_{ij} &\sim N(0, \sigma_\epsilon^2), \quad i.i.d \\ \nu_i &\sim N(0, \sigma_\nu^2), \quad i, i, d \\ Cov(\epsilon_{ij}, \nu_i) &= 0 \quad \forall(i, j) \end{aligned} \tag{2.8}$$

Note que o modelo pode ser escrito como em (2.2) com:

$Y = (Y_{1,1}, Y_{1,2}, \dots, Y_{1,N_1}, \dots, Y_{M,1}, \dots, Y_{M,N_M})^T$; $X = \mathbf{1}_N$; $\beta = \alpha$ e V Bloco-diagonal com elementos $V_{N_i} = \sigma_\epsilon^2 I_{N_i} + \sigma_\nu^2 \mathbf{1}_{N_i} \mathbf{1}_{N_i}^T$. Onde $\mathbf{1}_{N_i}$ e I_{N_i} denotam, respectivamente, o vetor coluna de 1's de dimensão N_i e a matriz de 1's de dimensão igual a N_i . Suponha que uma amostra de tamanho n_i seja selecionada em cada escola $i = 1, \dots, M$ e o objetivo seja prever a média das notas em cada escola, i.é: $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} Y_{ij}$. Dada a estrutura bloco-diagonal

de V tem-se que:

$$\tilde{\alpha} = \left\{ \sum_{i=1}^M \mathbf{1}_{n_i}^T V_{n_i}^{-1} \mathbf{1}_{n_i} \right\}^{-1} \left\{ \sum_{i=1}^M \mathbf{1}_{n_i}^T V_{n_i}^{-1} y_i \right\} \quad (2.9)$$

onde $y_i = (y_{i1}, \dots, y_{i,n_i})^T$ representa o vetor de observações de y na amostra selecionada na i -ésima escola. O vetor $\mathbf{1}_{n_i}$, $n_i \times 1$, e a matriz V_{n_i} , $n_i \times n_i$, são respectivamente definidos de forma análoga a $\mathbf{1}_{N_i}$ e a $V_{N_i}^{-1}$.

Pode-se facilmente verificar, ver Graybill (1969), que a inversa da matriz V_{n_i} é dada por:

$$V_{n_i}^{-1} = (1/\sigma_\epsilon^2)I_{n_i} - (\sigma_\nu^2/\sigma_\epsilon^2) \{1 + n_i^{-1}(\sigma_\nu^2/\sigma_\epsilon^2)\}^{-1} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^T \quad (2.10)$$

Definindo-se $\gamma_i = \sigma_\nu^2/(\sigma_\nu^2 + n_i^{-1}\sigma_\epsilon^2)$ e substituindo-se (2.10) na expressão (2.9), tem-se:

$$\tilde{\alpha} = \frac{\sum_{i=1}^M \gamma_i \bar{y}_i}{\sum_{i=1}^M \gamma_i} \quad (2.11)$$

Além disso, podemos escrever $\bar{Y}_i = \gamma_i^T Y$ onde γ_i é o vetor $N \times 1$ com as N_i posições correspondentes aos alunos da i -ésima escola iguais a N_i^{-1} e as outras $N - N_i$ posições iguais a zero. Utilizando-se então o Teorema (2.1), após algumas simplificações, obtém-se que o BLUP de \bar{Y}_i é dado por:

$$\hat{\bar{Y}}_i^B = f_i \bar{y}_i + (1 - f_i) \{ \tilde{\alpha} + \gamma_i (\bar{y}_i - \tilde{\alpha}) \} \quad (2.12)$$

Fazendo-se $\tilde{\nu}_i = \gamma_i (\bar{y}_i - \tilde{\alpha})$, a equação (2.12) pode ser re-escrita como:

$$\hat{\bar{Y}}_i^B = f_i \bar{y}_i + (1 - f_i) \{ \tilde{\alpha} + \tilde{\nu}_i \}.$$

Conforme mencionado nesta seção, é importante notar que o preditor (2.12) depende das componentes de variância σ_ϵ^2 e σ_ν^2 que são geralmente desconhecidas. Substituindo-se os respectivos estimadores dessas componentes na expressão (2.12), obtém-se o EBLUP, cuja forma geral é dada por:

$$\hat{\bar{Y}}_i^{EB} = f_i \bar{y}_i + (1 - f_i) \{ \hat{\alpha} + \hat{\gamma}_i (\bar{y}_i - \hat{\alpha}) \} \quad (2.13)$$

Entretanto, a variância do erro de predição do EBLUP não pode ser obtido via equação (2.4), uma vez que no Teorema (2.1) assume-se que a matriz de covariância V é conhecida. Na próxima seção obteremos algumas propriedades do EBLUP para uma classe bastante geral de modelos lineares e muito utilizada na estimação de pequenos domínios, denominada modelos lineares mistos, onde o Exemplo 2.6 é um caso particular.

2.4 Modelos Lineares Mistos

A teoria e a aplicação de modelos mistos, também denominados de modelos hierárquicos, têm sido desenvolvidas consideravelmente nos últimos anos. Este desenvolvimento tem sido impulsionado pela disponibilidade de algoritmos computacionais eficientes, ver por exemplo Goldstein (1995), Longford (1987) e Gamerman e Lopes (2006), aliada à existência de computadores de uso doméstico cada vez mais velozes. Além disso, tem crescido a aplicação prática destes modelos, gerando-se também novos problemas teóricos relacionados a uma particular aplicação. A razão pela qual os modelos mistos são amplamente aplicados decorre do fato de que em muitas situações a população de interesse apresenta uma estrutura hierárquica e, portanto, modelos que relacionam variáveis da população devem levar em conta esta estrutura. Por exemplo, considere um estudo que tem por objetivo investigar o desempenho escolar dos alunos de uma determinada série do primeiro grau. Claramente, alunos de uma determinada escola tendem a apresentar desempenhos mais semelhantes do que alunos de escolas diferentes. Se considerarmos as turmas de uma mesma escola, podemos ainda obter um grau de homogeneidade mais refinado. Desta forma, podemos notar que a população está naturalmente estruturada em níveis hierárquicos: escolas, turmas e alunos e portanto de-

vemos considerar esta estrutura na modelagem. Vários outros exemplos de populações com estruturas hierárquicas podem ser encontrados, fazendo que seja vasta a aplicação dos mesmos em diversas áreas, tais como: dados longitudinais, modelos de credibilidade, medicina e etc, veja Goldstein (1995) para uma excelente coletânea de exemplos. Na Teoria da Amostragem para Populações Finitas os modelos hierárquicos têm sido aplicados na análise de relações entre variáveis investigadas na pesquisa, como também para predição em pequenos domínios utilizando-se informações auxiliares. Como exemplo, podemos citar a Pesquisa Nacional por Amostragem de Domicílios (PNAD) realizada pelo IBGE, onde os níveis de hierarquia coincidem com os vários estágios de seleção da amostra: municípios, setores censitários e domicílios. Desta forma, para estudar relações entre diversas variáveis investigadas na pesquisa, a modelagem hierárquica pode ser mais apropriada do que modelos de regressão que não levam em conta os efeitos de conglomeração em cada um destes níveis. Além disso, estes efeitos podem ser explorados por meio da modelagem hierárquica para emprestar informação dentro de um mesmo nível da hierarquia. Como veremos adiante, este empréstimo de informação é fundamental na estimação em pequenos domínios, devido aos mesmos apresentarem (por definição) pequenos tamanhos de amostras. O objetivo desta Seção é fornecer o embasamento teórico da predição linear sob modelos mistos.

2.4.1 Predição em Modelos Lineares Mistos: EBLUP

Um modelo linear é denominado modelo linear misto se puder ser escrito como:

$$Y = X\beta + Z\nu + \varepsilon \quad (2.14)$$

onde: Y é o vetor $N \times 1$ da variável resposta (característica de interesse na população); X e Z são matrizes conhecidas de dimensões $N \times p$ e $N \times M$, respectivamente, e ambas de postos completos; ν e ε são vetores aleatórios, independentes com médias zero e matriz de covariância G e R , respectivamente. Assume-se também que as matrizes G e R dependem de um vetor usualmente desconhecido $\theta = (\theta_1, \dots, \theta_q)^T$, cujas componentes são comumente denominadas de componentes de variância. É fácil ver que o modelo (2.14) é um caso particular do modelo (2.2) com $V = R + ZGZ^T$. Aqui também é suposto que o espaço paramétrico Θ de θ é tal que $\forall \theta \in \Theta$ tem-se que $V = V(\theta)$ é positiva definida. As componentes do parâmetro β são usualmente denominadas de parâmetros associados aos "efeitos fixos" e os associados ao vetor aleatório ν de "efeitos aleatórios".

Veremos mais adiante que em muitas situações práticas que envolvem estimação em pequenos domínios estaremos interessados em prever uma combinação linear de efeitos fixos e efeitos aleatórios da forma: $\mu = k^T\beta + l^T\nu$. Consideramos o BLUP de μ (com θ conhecido) e conseqüentemente o seu respectivo EBLUP e abordaremos as suas propriedades.

Aplicando-se o mesmo argumento utilizado na demonstração do Teorema (2.1), obtemos o seguinte BLUP para $\mu = k^T\beta + l^T\nu$:

$$\hat{\mu}^B = k^T\tilde{\beta} + l^T\tilde{\nu} \quad (2.15)$$

onde $\tilde{\beta} = (X^TV^{-1}X)^{-1}X^TV^{-1}Y$ é o BLUE ("Best Linear Unbiased Estimator") de β para θ conhecido e $\tilde{\nu} = \tilde{\nu}(\theta) = GZ^TV^{-1}(Y - X\tilde{\beta})$ é um pre-

ditor linear de ν . O preditor BLUP depende de θ , que na maioria das situações práticas, é desconhecido. Há na literatura vários métodos de estimação de θ , destacando-se os métodos iterativos de mínimos quadrados generalizados (IGLS) e de mínimos quadrados generalizados restritos (RIGLS). Assumindo-se normalidade de ν e de ε o IGLS e o RIGLS são respectivamente equivalentes aos métodos de estimação de máxima verossimilhança (MLE) e o de máxima verossimilhança restrita (RMLE), veja Goldstein, (1986, 1989) para maiores detalhes. Outros métodos não baseados na hipótese de normalidade também estão disponíveis em várias rotinas de pacotes computacionais, destacando-se o método dos momentos e o de ajuste pelo método das constantes. Seja, então, $\hat{\theta}$ um estimador de θ com as seguintes propriedades: i) $E[\hat{\mu}^B(\hat{\theta})]$ é finita; ii) $\hat{\theta}(y) = \hat{\theta}(-y)$; iii) $\hat{\theta}(y) = \hat{\theta}(y - Xb) \forall y \text{ e } \forall b$ e iv) assumem-se que as distribuições de ν e ε são simétricas, não necessariamente normais. Kachar e Harville (1981) mostraram que sob as condições acima o preditor $\hat{\mu}^{EB} = \hat{\mu}^B(\hat{\theta})$ é não viciado. Kachar e Harville (1981) também mostraram que vários métodos, amplamente empregados de estimação de componentes de variância, tais como: MLE, RMLE e o método do ajuste de constantes, satisfazem às quatro propriedades acima e, portanto, os respectivos preditores são não viciados.

2.4.2 Erro Quadrático Médio do EBLUP

Conforme já observado por diversos autores, expressões algébricas fechadas para o erro quadrático médio (EQM) de um preditor EBLUP de um modelo linear misto é apenas disponível para casos particulares, veja por exemplo, Peixoto e Harville (1986). Contudo, como ressaltado por Prasad e Rao (1990), o uso do estimador do EQM, que não leva em conta a incerteza adicional devido à estimação das componentes de variância de V , acarreta uma subestimação

do EQM, notadamente quando o número de áreas for pequeno. A fim de tratar deste problema, várias aproximações foram propostas. Destacando-se o artigos de Kachar e Harville (1984) e posteriormente o de Prasad e Rao (1990), que continuam sendo referências amplamente citadas em trabalhos posteriores sobre este assunto.

Sejam $EQM(\hat{\mu}^{EB})$ e $EQM(\hat{\mu}^B)$ os respectivos erros quadráticos médios de um preditor EBLUP e do BLUP de μ , então o $EQM(\hat{\mu}^{EB})$ pode ser decomposto como:

$$\begin{aligned} EQM(\hat{\mu}^{EB}) &= E_{\xi} [\{\hat{\mu}^{EB} - \mu\}^2] = E_{\xi} [(\{\hat{\mu}^B - \mu\} + \{\hat{\mu}^{EB} - \hat{\mu}^B\})^2] = \\ &= E_{\xi} [(\hat{\mu}^B - \mu)^2] + E_{\xi} [(\hat{\mu}^{EB} - \hat{\mu}^B)^2] + \\ &+ 2E_{\xi} [(\hat{\mu}^B - \mu)(\hat{\mu}^{EB} - \hat{\mu}^B)] \end{aligned} \quad (2.16)$$

Kachar e Harville (1984) mostraram (para uma prova alternativa ver Rao, 2003) que sob normalidade dos erros ε e ν e assumindo-se que $\hat{\theta}$ é um estimador invariante por translação de θ , o último termo da equação (2.16) se anula e tem-se:

$$\begin{aligned} EQM(\hat{\mu}^{EB}) &= E_{\xi} [(\hat{\mu}^B - \mu)^2] + E_{\xi} [(\hat{\mu}^{EB} - \hat{\mu}^B)^2] = \\ &= EQM(\hat{\mu}^B) + E_{\xi} [(\hat{\mu}^{EB} - \hat{\mu}^B)^2] \end{aligned} \quad (2.17)$$

Como o segundo termo de (2.17) é positivo tem-se, o resultado já esperado, de que o erro quadrático do EBLUP é sempre maior que o do BLUP correspondente. O segundo termo de (2.17) pode ser interpretado como a parcela extra do EQM devida à estimação das componentes de variância. O primeiro termo de (2.17) pode ser facilmente calculado, enquanto que o segundo é geralmente calculado por aproximação. Após alguma álgebra matricial e utilizando-se propriedades de esperanças de formas quadráticas tem-se:

$$EQM(\hat{\mu}^B) = g_1(\theta) + g_2(\theta) \quad (2.18)$$

onde:

$$\begin{aligned} g_1(\theta) &= l^T (G - GZ^T V^{-1} ZG) l \\ g_2(\theta) &= (k - X^T V^{-1} ZGl)^T (X^T V^{-1} X)^{-1} (k - X^T V^{-1} ZGl) \end{aligned}$$

Kachar e Harville (1984) obtiveram uma aproximação do segundo termo de (2.17) por meio de uma expansão de primeira ordem de $\hat{\mu}^{EB} - \hat{\mu}^B$ em série de Taylor em torno de θ , assumindo que os termos restantes são relativamente de ordem inferior ao primeiro: $\hat{\mu}^{EB} - \hat{\mu}^B \approx \frac{\partial \hat{\mu}^B(\theta)}{\partial \theta} (\hat{\theta} - \theta)$. Finalmente, Kachar e Harville (1984), baseados na expansão acima, propuseram a aproximação:

$$E_\xi [\hat{\mu}^{EB} - \hat{\mu}^B]^2 \approx \text{Trace} \left[\text{COV}(d(\theta)) E_\xi \left[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T \right] \right] \quad (2.19)$$

onde $\text{COV}(d(\theta))$ é a matriz de covariância de $d(\theta) = \partial \hat{\mu}^B(\theta) / \partial \theta$.

Prasad e Rao (1990) propuseram uma aproximação mais simples de se calcular do que a descrita em (2.19). Assumindo normalidade e algumas hipóteses de regularidade, Prasad e Rao (1990) obtiveram:

$$E_\xi [\hat{\mu}^{EB} - \hat{\mu}^B]^2 \approx \text{Trace} \left[(\partial b^T / \partial \theta) V (\partial b^T / \partial \theta)^T \text{COV}_A(\hat{\theta}) \right] \quad (2.20)$$

onde $b^T = l^T GZ^T V^{-1}$ e $\text{COV}_A(\hat{\theta})$ é a matriz de covariância assintótica de $\hat{\theta}$.

Denominando-se $g_3(\theta) = \text{Trace} \left[(\partial b^T / \partial \theta) V (\partial b^T / \partial \theta)^T \text{COV}_A(\hat{\theta}) \right]$, tem-se a seguinte aproximação para o erro quadrático médio de $\hat{\mu}^{EB}$:

$$EQM(\hat{\mu}^{EB}) \approx g_1(\theta) + g_2(\theta) + g_3(\theta) \quad (2.21)$$

O primeiro termo de (2.21) é a parcela de maior contribuição no EQM e pode ser interpretado como a variabilidade introduzida pelos erros aleatórios ε e ν . O segundo e o terceiro termos são, respectivamente, as contribuições devidas às estimações do parâmetro de efeitos fixos β e do vetor de componentes de variância θ . Prasad e Rao (1990) mostraram num estudo de simulação, para

casos particulares do modelo (2.14), que a contribuição do terceiro termo no EQM não pode ser ignorada, mormente quando o número de domínios e/ou o número de unidades amostrais em cada domínio for pequeno. Rao (2003) observou que a aproximação (2.21) pode ser apenas rigorosamente justificada para casos particulares do modelo (2.14). Prasad e Rao (1990) mostraram que sob o modelo de Fay e Herriot (1979) (ver Seção 3.1) a aproximação de Taylor para o segundo termo de (2.17) é de ordem $o(m^{-1})$. Entretanto, segundo Prasad e Rao (1990), parece ser difícil fornecer as condições gerais para que esta aproximação seja $o(m^{-1})$. Mais recentemente, assumindo uma estrutura Bloco Diagonal da matriz V no modelo (2.14) e algumas condições de regularidade, Datta e Lahiri (2000) mostraram que a aproximação em (2.21) é de ordem $o(m^{-1})$ para um número grande de domínios m na amostra.

Exemplo 2.7: Modelo de Intercepto Aleatório (cont.)

O modelo do intercepto aleatório é um caso particular do modelo (2.14)

com $X = 1_N^T$, $\nu = (\nu_1, \dots, \nu_M)^T$ e $Z =$

$$\begin{pmatrix} 1 & 0 & 0 & \dots & \dots & 0 \\ 1 & 0 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots & 1 \end{pmatrix}$$
 de di-

mensão $N \times M$.

Suponha que se esteja interessado em prever \bar{Y}_i e que a fração de amostragem f_i possa ser considerada desprezível. Então $\bar{Y}_i = \alpha + \nu_i + \bar{\epsilon}_i \approx \alpha + \nu_i$, uma vez que $\bar{\epsilon}_i \approx 0$ pela lei dos grandes números, já que N_i é grande e $E_\xi(\epsilon_{ij}) = 0 \quad \forall(i, j)$.

Por outro lado, $\mu_i = \alpha + \nu_i$ pode ser escrito como $\mu_i = K^T \beta + l^T \nu$ com $K = 1$, $\beta = \alpha$ e $l = (0, \dots, 1, 0, \dots, 0)^T$, o vetor $1 \times M$ com 1 na i -ésima posição e 0 nas demais. Fazendo-se $f_i = 0$ em (2.13) tem-se que o EBLUP de μ_i é dado por:

$$\hat{\mu}_i^{EB} = \hat{\alpha} + \hat{\gamma}_i(\bar{y}_i - \hat{\alpha}) \tag{2.22}$$

Aplicando-se a aproximação (2.21) ao modelo (2.8) obtém-se:

$$\begin{aligned}
 EQM(\hat{\mu}_i^{EB}) &= (1 - \hat{\gamma}_i)\sigma_\nu^2 + (1 - \hat{\gamma}_i)^2 \left\{ \sum_{i=1}^M \hat{\gamma}_i \right\}^{-1} + \\
 &+ n_i^{-2}(\sigma_\nu^2 + n_i^{-1}\sigma_\epsilon^2)^{-3}[\sigma_\epsilon^4 var(\hat{\sigma}_\nu^2) + \sigma_\nu^4 var(\hat{\sigma}_\epsilon^2) + \\
 &- 2\sigma_\nu^2\sigma_\epsilon^2 cov(\hat{\sigma}_\epsilon^2, \hat{\sigma}_\nu^2)]
 \end{aligned} \tag{2.23}$$

onde os símbolos $var(\hat{a})$ e $cov(\hat{a}, \hat{b})$, respectivamente, denotam a variância e a covariância dos estimadores das componentes de variância a e b . É importante notar que os valores de $var(\hat{a})$ e $cov(\hat{a}, \hat{b})$ dependem do método de estimação empregado.

2.4.3 Estimador do Erro Quadrático Médio do EBLUP

Como em geral, as componentes de variância são usualmente desconhecidas, há necessidade de se estimar o erro quadrático médio de um preditor EBLUP. Nesta seção, obteremos heurísticamente estimadores não viciados para o EQM de um preditor EBLUP.

Um estimador do $EQM(\hat{\mu}_i^{EB})$ pode ser obtido pela simples substituição das componentes desconhecidas por um estimador $\hat{\theta}$. No entanto, o estimador resultante do EQM é viciado. Conforme observado por Rao (2003), embora os dois últimos termos são aproximadamente não viciados, i. é:

$E_\xi(g_2(\hat{\theta})) \approx g_2(\theta)$ e $E_\xi(g_3(\hat{\theta})) \approx g_3(\theta)$, o vício do termo de maior contribuição $g_1(\hat{\theta})$ é geralmente da mesma ordem de $g_2(\theta)$ e $g_3(\theta)$. Portanto, a fim de se obter um estimador do EQM aproximadamente não viciado, devemos primeiramente obter um estimador aproximadamente não viciado de $g_1(\hat{\theta})$. Consideraremos inicialmente o caso em que $\hat{\theta}$ é um estimador não viciado de

θ . Expandindo-se $g_1(\hat{\theta})$ em torno de θ até segunda ordem tem-se:

$$g_1(\hat{\theta}) \approx g_1(\theta) + (\hat{\theta} - \theta)^T \nabla g_1(\theta) + \frac{1}{2}(\hat{\theta} - \theta)^T \nabla^2 g_1(\theta)(\hat{\theta} - \theta) \quad (2.24)$$

Tomando-se o valor esperado em ambos os termos de (2.24) e lembrando-se que, por hipótese, $E_\xi(\hat{\theta}) = \theta$ tem-se:

$$E_\xi(g_1(\hat{\theta})) \approx g_1(\theta) + \frac{1}{2} \text{Trace} \left[\nabla^2 g_1(\theta) \text{COV}(\hat{\theta}) \right] \quad (2.25)$$

Além disso, assumindo-se que V tenha uma estrutura linear em θ , a equação (2.25) reduz-se a:

$$E_\xi(g_1(\hat{\theta})) \approx g_1(\theta) - g_3(\theta) \quad (2.26)$$

Da aproximação (2.26) e do fato de que $E_\xi(g_2(\hat{\theta})) \approx g_2(\theta)$ e $E_\xi(g_3(\hat{\theta})) \approx g_3(\theta)$, tem-se que um estimador aproximadamente não viciado de $EQM(\hat{\mu}_i^{EB})$ é dado por:

$$eqm(\hat{\mu}_i^{EB}) = g_1(\hat{\theta}) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}) \quad (2.27)$$

No caso em que o estimador de θ seja viciado tal que $E_\xi(\theta) - \theta \approx v_\theta^T(\theta)$, então um estimador aproximadamente não viciado de $EQM(\hat{\mu}_i^{EB})$ é dado por:

$$eqm^*(\hat{\mu}_i^{EB}) = g_1(\hat{\theta}) - v_\theta^T(\theta) \nabla g_1(\theta) + g_2(\hat{\theta}) + 2g_3(\hat{\theta}) \quad (2.28)$$

2.5 Amostragem Informativa

Nas seções anteriores, a inferência sobre quantidades de interesse é baseada em modelos assumidos para a população, e o plano amostral não teve papel algum na inferência desses parâmetros. As propriedades dos preditores (vício, variância do erro de predição, normalidade assintótica e etc) foram obtidas sob o modelo proposto. Este fato poderia levar erroneamente à conclusão de que o plano amostral é irrelevante, desde que algum modelo seja assumido. Esta afirmação é verdadeira apenas sob certas condições, como veremos mais adiante nesta seção.

2.5.1 Definição Formal de Ignorabilidade

Neste seção damos as condições gerais para que um desenho amostral seja ignorável ou não informativo. Definimos o vetor indicador de pertinência de cada unidade populacional na amostra por $I_s = (I_1(s), \dots, I_N(s))^T$, onde $I_j(s) = 1$ se a unidade populacional j pertence à amostra s e $I_j(s) = 0$, em caso contrário. Um planejamento amostral é um sistema de regras que define uma distribuição conjunta do vetor I_s . No caso geral, esse sistema de regras pode depender de um conjunto de covariáveis X , da variável resposta Y e, possivelmente, de um vetor de parâmetros desconhecidos ϕ . Denotamos então por $f_{I_s}(I_s | X, Y, \phi)$ a função de probabilidade conjunta de I_s condicional a X, Y e ϕ . O modelo de superpopulação ξ é definido por uma função de densidade ou de probabilidade (no caso em que a característica de interesse y é discreta) que pode depender de covariáveis X e de um vetor desconhecido de parâmetros θ : $f_Y(Y | X, \theta)$.

O modelo conjunto do vetor Y e do vetor de pertinência I_s pode ser escrito como:

$$f_{Y, I_s}(Y, I_s | X, \theta, \phi) = f_Y(Y | X, \theta) \cdot f_{I_s}(I_s | X, Y, \phi) \quad (2.29)$$

Particionando-se o vetor Y na parte observada na amostra Y_s e na parte não observada $Y_{\bar{s}}$ tem-se que a função de densidade conjunta de Y_s e I_s é dada por:

$$f_{Y_s, I_s}(Y_s, I_s | X, \theta, \phi) = \int f_Y(Y_s, Y_{\bar{s}} | X, \theta) f_{I_s}(I_s | X, Y, \phi) dY_{\bar{s}} \quad (2.30)$$

Se I_s e Y forem condicionalmente independentes dado X e ϕ , i. é:

$$f_{I_s}(I_s | X, Y, \phi) = f_{I_s}(I_s | X, \phi) \quad (2.31)$$

Então da equação (2.30) tem-se:

$$\begin{aligned} f_{Y_s, I_s}(Y_s, I_s | X, \theta, \phi) &= f_{I_s}(I_s | X, Y, \phi) \int f_Y(Y_s, Y_{\bar{s}} | X, \theta) dY_{\bar{s}} \\ &= f_{I_s}(I_s | X, Y, \phi) f_{Y_s}(Y_s | X, \theta) \end{aligned} \quad (2.32)$$

Conclui-se, da equação (2.32), que se I_s e Y forem condicionalmente independentes dado X e ϕ , a informação adicional trazida por I_s pode ser descartada quando se deseja fazer inferência sobre θ . Em outras palavras, podemos considerar que o modelo de superpopulação assumido ξ pode ser empregado para se ajustar os dados y_s obtidos pela amostra selecionada s . Neste caso diz-se que o plano amostral é ignorável ou não informativo com respeito ao parâmetro de interesse θ .

Há muitos planos amostrais que satisfazem (2.5.1) e, portanto, não são informativos e, entre eles, todos os planos probabilísticos em que a probabilidade de inclusão não depende da variável de interesse y .

Exemplo 2.8: Amostragem Aleatória Simples sem Reposição

Neste caso

$$f_{I_s}(I_s | X, Y, \phi) = f_{I_s}(I_s) = \begin{cases} \binom{N}{n}^{-1} & \text{se } \sum_{j=1}^N I_j(s) = n \\ 0 & \text{caso contrário} \end{cases}$$

e, portanto, o plano amostral é ignorável, como já havíamos observado.

Muitos outros planos amostrais não probabilísticos também podem ser classificados como não informativos e; como exemplo, temos o plano amostral denominado balanceado, em que a média amostral da variável auxiliar x deve coincidir com sua média populacional, i. é: $\bar{x}_s = \bar{X}$, $\forall s \in \mathfrak{S}$.

2.5.2 Tratamento da Não-Resposta

Um problema quase sempre presente nas pesquisas por amostragem é a presença da não-resposta em algumas unidades amostrais ou para algumas características de interesse (no caso em que a pesquisa seja de múltiplos propósitos).

Em muitas situações práticas, os entrevistados recusam-se a responder certos quesitos do questionário ou todos eles. A não-resposta pode também decorrer da falta de empenho ou compromisso do entrevistador, como também a outros motivos. Caso o mecanismo de não-resposta depender da variável resposta, podemos estar numa situação em que a equação (2.5.1) não é válida, mesmo que a amostra inicial seja aleatoriamente selecionada. Neste caso, diz-se que a não-resposta é informativa. Para exemplificar esta situação, considere que uma secretaria de educação de um determinado município deseja fazer uma avaliação educacional dos alunos de uma determinada série escolar por meio da aplicação de um teste. Para simplificar, suponha que uma amostra aleatória de alunos seja selecionada. Se a taxa de não comparecimento ao teste for consideravelmente maior para alunos de baixo aproveitamento escolar, este auto-mecanismo de seleção (não-resposta) é não ignorável e portanto deve ser considerado na inferência de um modelo paramétrico, onde a variável resposta seja Y . Suponha que para algumas unidades amostradas não se obtém o valor de y . Então podemos repartir a amostra s em $s = s_o \cup s_{\bar{o}}$, onde os símbolos s_o e $s_{\bar{o}}$ denotam, respectivamente, as unidades observadas em s e aquelas em s em que não se obteve respostas. Conseqüentemente, o vetor de observações y_s pode ser repartido em $y_s = (y_{s_o}, y_{s_{\bar{o}}})$. Introduzimos, então, um vetor indicador R_s para o mecanismo de não-resposta para toda unidade amostrada $j \in 1, \dots, n$: $R_s = (R_1(s), \dots, R_n(s))^T$, onde $R_j(s) = 1$, se a unidade amostrada j é observada, e $R_j(s) = 0$, se não houver resposta da mesma. Seja φ um vetor de parâmetros que pode ser necessário para se modelar o mecanismo de não-resposta. De acordo com o modelo de superpopulação de Little (1982), veja também (Chambers e Skinner, 2003, Chap. 18), a distribuição conjunta

de Y , I_s e R_s pode ser escrita como:

$$f_{Y,I_s,R_s}(Y, I_s | X, \theta, \phi, \varphi) = f_Y(Y | X, \theta) \cdot f_{I_s}(I_s | X, Y, \phi) f_{R_s}(R_s | X, I_s, Y_s, \varphi) \quad (2.33)$$

Little em Chambers e Skinner (2003), a partir de (2.33) fornece condições de ignorabilidade na presença de não-resposta.

Exemplo 2.9: Modelo Normal com Não-Resposta Informativa

Considere o seguinte modelo de superpopulação: $Y_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, N$ e seja $\theta = (\mu, \sigma^2)^T$ o vetor de parâmetros de interesse. Suponha que uma amostra aleatória simples de tamanho n seja selecionada; contudo, algumas unidades amostrais não respondem. Seja então $n_r \leq n$ o número de respondentes da amostra n . Assume-se que a probabilidade de cada unidade responder dependa apenas do seu respectivo valor Y_j , e seja dada pelo seguinte modelo probit:

$$Pr(R_j(s) = 1 | Y_j, \varphi) = \Phi(\varphi_0 + \varphi_1 Y_j) \quad (2.34)$$

onde Φ em (2.34) denota a função de distribuição acumulada da normal padronizada. Portanto, $R_j(s) | Y_j, \varphi \sim \text{Bern}(\Phi(\varphi_0 + \varphi_1 Y_j))$ e condicionalmente independentes para $j = 1, \dots, N$.

Notando-se que o desenho amostral é ignorável, a função de verossimilhança de θ e φ pode ser facilmente obtida de (2.33):

$$\begin{aligned} L(\theta, \varphi | y_o, X, R_s) &= \int f_{Y,R_s}(Y, R_s | X, \theta, \phi, \varphi) dy_{\bar{s}} dy_{s_{\bar{o}}} = \\ &= \int f_Y(Y | X, \theta) f_{R_s}(R_s | Y_s, \varphi) dy_{\bar{s}} dy_{s_{\bar{o}}} \end{aligned} \quad (2.35)$$

Por outro lado:

$$\begin{aligned}
 f_Y(Y | X, \theta) &\propto (\sigma^2)^{-(N-n)/2} \prod_{i \in \tilde{s}} \exp \{-0.5\sigma^2(y_i - \mu)^2\} \cdot \\
 &\cdot (\sigma^2)^{-(n-n_r)/2} \prod_{i \in s_{\tilde{o}}} \exp \{-0.5\sigma^2(y_i - \mu)^2\} \cdot \\
 &\cdot (\sigma^2)^{-(n_r)/2} \prod_{i \in s_o} \exp \{-0.5\sigma^2(y_i - \mu)^2\} \quad (2.36)
 \end{aligned}$$

$$f_{R_s}(R_s | Y_o, \varphi) = \prod_{i \in s_{\tilde{o}}} \Phi(\varphi_0 + \varphi_1 y_i) \prod_{i \in s_o} \Phi(\varphi_0 + \varphi_1 y_i) \quad (2.37)$$

Substituindo-se (2.36) e (2.37) em (2.35) e resolvendo-se a integração tem-se:

$$\begin{aligned}
 L(\theta, \varphi | y_o, X, R_s) &\propto (\sigma^2)^{-n_r/2} \exp \left\{ -0.5\sigma^2 \left(\sum_{i \in s_{\tilde{o}}} (y_i - \mu)^2 \right) \right\} \cdot \\
 &\left\{ \prod_{i \in s_{\tilde{o}}} \Phi(\varphi_0 + \varphi_1 y_i) \right\} \Phi \left\{ \frac{\varphi_0 + \varphi_1 \mu}{\sqrt{1 + \varphi_1^2 \sigma^2}} \right\}^{n-n_r} \quad (2.38)
 \end{aligned}$$

Uma vez que o último termo multiplicativo da equação (2.38) envolve os parâmetros de interesse $\theta = (\mu, \sigma^2)^T$ e os parâmetros do modelo de não-resposta φ , tem-se que a estimação de verossimilhança de θ depende de φ . Métodos iterativos devem ser empregados para conjuntamente estimá-los por máxima verossimilhança.

2.6 Introdução à Inferência Bayesiana em Populações Finitas

Os primeiros trabalhos nesta área datam do final da década de 60. Destacando-se entre eles, o de Basu (1969) em que descreve como o princípio da verossimilhança deve ser aplicado para inferência em populações finitas e o de

Ericson (1969), no qual considera modelos paramétricos conjugados na família exponencial. Assumindo apenas a condição de linearidade da *posteriori*, Ericson (1969) obtém os estimadores de Bayes de quantidades de interesse na população sob perda quadrática. Mais recentemente, a partir da década de 90, com o advento de técnicas de estatística computacional mais avançadas, em particular com o surgimento do método de simulação estocástica via cadeias de Markov, a Estatística Bayesiana tem tido um grande impulso em todas as áreas. Em particular, tem sido também observado um número crescente de trabalhos teóricos e aplicados na área de predição em populações finitas sob o enfoque da Inferência Bayesiana. Nesta Seção, introduziremos alguns fundamentos básicos da Inferência Bayesiana para a predição em Populações Finitas.

2.6.1 Modelo Paramétrico na Família Exponencial

Em muitas situações práticas, eliciar uma distribuição *a priori*, quando a dimensão do espaço paramétrico for alta é uma tarefa bastante complicada, ou se não impossível. Este problema está particularmente presente na inferência em populações finitas, onde os parâmetros de interesse são funções dos N valores $Y_i; i = 1, \dots, N$ desconhecidos (parâmetros) da população. Neste caso, uma análise bayesiana parece ser impossível sem que sejam adotadas algumas hipóteses simplificadoras. Nesta Seção apresentaremos alguns resultados devidos a Ericson (1969) que podem ser aplicados quando for possível se atribuírem distribuições *a priori* para a média e a variância da população. Suponhamos que seja razoável se admitir que os rótulos (índices) tragam pouca ou nenhuma informação sobre as unidades populacionais. Gostaríamos também que os valores de y para a amostra observada s fornecessem informação para predição dos valores não observados de y . Uma forma de se realizar esta tarefa é se

atribuirm distribuições *a priori* permutáveis em relação aos rótulos das unidades populacionais. Uma técnica simples de se atingir permutabilidade é por meio da formulação paramétrica dada a seguir.

Seja θ um vetor de parâmetros desconhecidos de dimensão $p \ll N$ no qual se atribui uma distribuição *a priori* $p(\theta)$. Assumindo-se que condicional a θ , os valores Y_i 's sejam independentes e identicamente distribuídos com função de densidade $g(\cdot | \theta)$ (ou função de probabilidade, no caso em que Y for discreta), temos que a densidade de $Y = (Y_1, \dots, Y_N)^T$ é dada por:

$$f_Y(y) = \int_{-\infty}^{\infty} \prod_{i=1}^N g(y_i | \theta) p(\theta) d\theta \quad (2.39)$$

É importante notar que o parâmetro de mistura θ , também denominado hiperparâmetro, gera a distribuição permutável $f_Y(\cdot)$, a partir de distribuições independentes e identicamente distribuídas $g(\cdot | \theta)$.

A distribuição preditiva de $y_{\bar{s}}$, dados os valores y_s de y observados na amostra s é obtida por:

$$\begin{aligned} \pi(y_{\bar{s}} | y_s) &= \frac{\int_{-\infty}^{\infty} \prod_{i \notin s} g(y_i | \theta) \prod_{i \in s} g(y_i | \theta) p(\theta) d\theta}{\int \dots \int \left\{ \int_{-\infty}^{\infty} \prod_{i \notin s} g(y_i | \theta) \prod_{i \in s} g(y_i | \theta) p(\theta) d\theta \right\} d \prod_{j \notin s} y_j} = \\ &= \int_{-\infty}^{\infty} \prod_{i \notin s} g(y_i | \theta) \frac{\prod_{i \in s} g(y_i | \theta) p(\theta)}{f_{y_s}(y_s)} d\theta = \\ &= \int_{-\infty}^{\infty} \prod_{i \notin s} g(y_i | \theta) h(\theta | y_s) d\theta \end{aligned} \quad (2.40)$$

onde $f_{y_s}(y_s)$ e $h(\theta | y_s)$ são respectivamente a densidade conjunta do vetor y_s e a distribuição *a posteriori* do parâmetro θ dado o vetor y_s .

Exemplo 2:10: Modelo Normal Conjugado com Variância Conhecida

Considere o seguinte modelo normal: $Y_i | \theta \sim N(\theta, \sigma^2)$ com σ^2 conhecido. Assumindo-se uma distribuição *a priori* de θ normal com média μ e variância τ^2 , então não é difícil verificar que o vetor $Y = (Y_1, \dots, Y_N)^T$ tem distribuição multivariada normal com média $\mathbf{1}_N^T \mu$ e matriz de covariância $\sigma^2 I_N + \mathbf{1}_N \mathbf{1}_N^T \tau^2$. Além disso, $\theta | y_s \sim N(m, v)$ onde: $m = \frac{\tau^2 \bar{y}_s + m \{n(s)\}^{-1} \sigma^2}{\tau^2 + \{n(s)\}^{-1} \sigma^2}$ e $v = \frac{\{n(s)\}^{-1} \sigma^2}{\tau^2 + \{n(s)\}^{-1} \sigma^2} \tau^2$. Utilizando-se (2.40) obtém-se:

$$y_{\bar{s}} | y_s \sim MN \left(\mathbf{1}_{N-n(s)}^T m, \sigma^2 I_{N-n(s)} + \mathbf{1}_{N-n(s)} \mathbf{1}_{N-n(s)}^T v \right)$$

A partir do resultado acima, não é difícil mostrar que a distribuição *a posteriori* de $\bar{Y} = N^{-1} (\sum_{i \in s} y_i + \sum_{i \notin s} Y_i)$ é normal com média e variância dadas respectivamente por:

$$E(\bar{Y} | y_s) = N^{-1} \{n(s) \bar{y}_s + (N - n(s)) m\}$$

$$V(\bar{Y} | y_s) = \frac{N - n(s)}{N} \frac{\{n(s)\}^{-1} \sigma^2}{\tau^2 + \{n(s)\}^{-1} \sigma^2} (\tau^2 + \sigma^2 / N)$$

É importante observar que a medida que τ^2 cresce, ou seja a informação *a priori* torna-se mais vaga, os valores de $E(\bar{Y} | y_s)$ e $V(\bar{Y} | y_s)$ aproximam-se respectivamente de $N \bar{y}_s$ e $\frac{N - n(s)}{N} \sigma^2 / n(s)$. Portanto, neste caso limite, as respectivas estimativas pontuais clássicas (baseada no preditor BLUP) coincidem com as estimativas bayesianas.

O modelo conjugado acima pertence a uma classe de distribuições conhecidas como família exponencial. Os resultados abaixo generalizam aqueles obtidos no Exemplo 2.10 acima.

Teorema 2.2: Suponha que condicionalmente a um parâmetro θ , Y_i , para $i = 1, \dots, N$ sejam independentes e identicamente distribuídos com função de densidade comum pertencente à família exponencial de um parâmetro:

$$f(y | \theta) = \exp\{\theta y - a(\theta)\} v(y) \quad (2.41)$$

e o parâmetro θ tenha uma distribuição *a priori* conjugada da forma:

$$p(\theta) \propto \exp\{\alpha\theta - \nu a(\theta)\} \quad (2.42)$$

Então, a distribuição *a posteriori* de θ dado y_s também pertence à família exponencial:

$$h(\theta | y_s) \propto \exp\left\{\left(\sum_{i \in s} y_i + \alpha\right)\theta - (n(s) + \nu)a(\theta)\right\} \quad (2.43)$$

Além disso, é fácil ser verificado que:

$E(Y_i | \theta) = a'(\theta)$ e $V(Y_i | \theta) = a''(\theta)/n(s)$. Por integração por partes, tem-se que:

$$E\left[a'(\theta) | y_s\right] = \frac{n(s)\bar{y}_s + \alpha}{n(s) + \nu} \quad (2.44)$$

Da distribuição *a priori* (2.42), pode ser verificado por integração por partes que:

$$E\left[a'(\theta)\right] = \alpha/\nu; \quad E\left[a''(\theta)\right] = \nu V\left[a'(\theta)\right] \quad (2.45)$$

Das expressões (2.44) e (2.45) tem-se:

$$E\left[a'(\theta) | y_s\right] = \frac{\bar{y}_s \cdot V\left[a'(\theta)\right] + E\left[a'(\theta)\right] E\left[V(\bar{Y} | \theta)\right]}{V\left[a'(\theta)\right] + E\left[V(\bar{Y} | \theta)\right]} \quad (2.46)$$

O resultado em (2.46) foi obtido por Ericson (1969) apenas assumindo a condição de linearidade da distribuição *a posteriori* (veja a definição a seguir).

Definição 2.4: Suponha que condicionalmente a um vetor θ de p componentes, Y_i , $i = 1, \dots, N$ tenham uma média comum $\mu(\theta)$ e θ seja atribuída uma distribuição *a priori* tal que:

$$E[\mu(\theta) | y_s] = \alpha + \beta \bar{y}_s \quad (2.47)$$

onde α e β não dependem de y . Então (2.47) é referida como condição de linearidade da *posteriori*.

O Teorema 2.3 a seguir, devido a Ericson (1969), fornece uma expressão para $E[\mu(\theta) | y_s]$ sob a condição de linearidade da *posteriori*.

Teorema 2.3: Suponha que condicionalmente a um vetor de parâmetros θ , Y_i , $i = 1, \dots, N$ tenham uma média comum $\mu(\theta)$, $V(Y_i | \theta) < \infty \forall i = 1, \dots, N$, $0 < V[\mu(\theta)] < \infty$ e a condição (2.47) seja satisfeita. Então:

$$E[\mu(\theta) | y_s] = \frac{V[\mu(\theta)] \bar{y}_s + E[\mu(\theta)] E[V(\bar{y}_s | \theta)]}{V[\mu(\theta)] + E[V(\bar{y}_s | \theta)]} \quad (2.48)$$

Prova: Da condição de linearidade da *posteriori*, obtém-se:

$$E[\mu(\theta)] = E[E[\mu(\theta) | y_s]] = E[\alpha + \beta \bar{y}_s] = \alpha + \beta E(\bar{y}_s) \quad (2.49)$$

Por outro lado:

$$E(\bar{y}_s) = E[E(\bar{y}_s | \theta)] = E[\mu(\theta)] \quad (2.50)$$

Das equações (2.49) e (2.50) tem-se que:

$$\{E[\mu(\theta)]\}^2 = \alpha E[\bar{y}_s] + \beta \{E[\bar{y}_s]\}^2 \quad (2.51)$$

Multiplicando-se ambos os membros de (2.47) por \bar{y}_s e tomando-se o valor esperado, tem-se:

$$E[\mu(\theta) \bar{y}_s] = \alpha E[\bar{y}_s] + \beta E[\bar{y}_s^2] \quad (2.52)$$

Por outro lado:

$$E[\mu(\theta)\bar{y}_s] = E[E[\mu(\theta)\bar{y}_s | \theta]] = E[\mu(\theta)E[\bar{y}_s | \theta]] = E[\mu(\theta)^2] \quad (2.53)$$

Das equações (2.51) (2.52) e (2.53) tem-se:

$$V[\mu(\theta)] = \beta V(\bar{y}_s) \quad (2.54)$$

Utilizando-se a seguinte decomposição:

$V(\bar{y}_s) = E[V(\bar{y}_s | \theta)] + V[E(\bar{y}_s | \theta)] = E[V(\bar{y}_s | \theta)] + V[\mu(\theta)]$ e as equações (2.49) e (2.54) tem-se finalmente (2.48).

Concluimos, então, que o modelo conjugado na família exponencial é um caso particular de condição de linearidade da *posteriori* e portanto os resultados em (2.46) e (2.48) coincidem. Além disso, notando-se que:

$E(Y_i | y_s) = E[\mu(\theta) | y_s] \quad \forall i \notin s$, conclui-se que os resultados obtidos acima podem ser facilmente estendidos para estimação em populações finitas. Por exemplo, sob as condições do Teorema 2.3, a média da *posteriori* de \bar{Y} é dada por:

$$\begin{aligned} E[\bar{Y} | y_s] &= N^{-1} \left\{ n(s)\bar{y}_s + \sum_{i \notin s} E(Y_i | y_s) \right\} = \\ &= N^{-1} \{ n(s)\bar{y}_s + (N - n(s))E[\mu(\theta) | y_s] \} \quad (2.55) \end{aligned}$$

Em muitas situações práticas, não é razoável se trabalhar com modelos conjugados e portanto técnicas de estatística computacional devem ser empregadas para se obter uma aproximação numérica de quantidades de interesse, tais como $E[\bar{Y} | y_s]$ e $V[\bar{Y} | y_s]$. Muitas vezes, o interesse é em aproximar ou amostrar da distribuição *a posteriori* da quantidade de interesse, por exemplo $\bar{Y} | y_s$. Neste último caso, um método muito utilizado atualmente é o método de simulação estocástica via cadeia de Markov. Para exemplos no contexto de estimação em pequenos domínios veja os Capítulos 4 e 5.

Capítulo 3

Modelos Lineares Mistos

Neste capítulo, introduzimos os modelos lineares mistos de superpopulação, bastante empregados e amplamente discutidos na literatura sobre estimação em pequenos domínios. Estes modelos combinam dados provenientes de amostras e informações auxiliares, que podem ser providas por Censos, registros administrativos e etc. A vantagem do uso de modelos de efeitos mistos em relação aos modelos de regressão tem sido comprovada empiricamente e teoricamente por vários trabalhos na literatura, ver, por exemplo, Moura e Holt (1999) e Rao (2003). A principal desvantagem dos modelos de regressão em relação aos modelos mistos é assumir que as variáveis auxiliares sejam capazes de explicar toda fonte de variabilidade entre os domínios. Porém, já se tem como certo, de que mesmo após a introdução de covariáveis, há ainda diferenças não estruturadas entre os domínios. Estas diferenças residuais são então modeladas por efeitos aleatórios. O emprego de modelos mistos dependem do tipo de informação auxiliar disponível. Nas Seções abaixo apresentaremos os principais modelos discutidos na literatura para cada tipo de informação disponível.

3.1 Modelo no Nível de Pequenos Domínios

Em muitas situações práticas, as informações auxiliares não estão disponíveis no nível de unidades elementares da população, em que foi feita a medida da variável resposta (de interesse) y . Frequentemente, embora a informação no nível desagregado esteja disponível, não há possibilidade de se identificarem as unidades selecionadas (do cadastro próprio de seleção da amostra) no cadastro contendo as informações auxiliares. Contudo é bastante comum a possibilidade de se obter informação auxiliar em nível mais agregado, tal como no nível do pequeno domínio de interesse.

O primeiro trabalho objetivando prever renda per capita média para pequenos domínios com população até 1000 habitantes foi proposto por Fay e Herriot (1979) e, por esta razão, este modelo é denominado de modelo de Fay-Herriot. Suponha que o objetivo seja obter estimativas das médias de uma variável de interesse y nos m pequenos domínios amostrados de M domínios na população. Suponha que \hat{Y}_i , $i = 1, \dots, m$ seja um estimador não viciado (sob a aleatorização induzida pelo desenho amostral empregado) da verdadeira média desconhecida \bar{Y}_i no domínio de interesse i . Além disso, assuma que esteja apenas disponível um vetor de variáveis auxiliares $z_i = (z_{1i}, \dots, z_{pi})^T$ obtido para cada domínio i e relacionado com uma função $g(\cdot)$ dos parâmetros de interesse \bar{Y}_i , por meio do seguinte modelo linear misto:

$$\theta_i = z_i^T \beta + k_i \nu_i, \quad i = 1, \dots, m \quad (3.1)$$

onde $\theta_i = g(\bar{Y}_i)$ e k_i são constantes positivas para permitirem possíveis heterocedasticidade no nível de domínios, ν_i são efeitos aleatórios independentes e identicamente distribuídos no nível de domínios com $E_\xi(\nu_i) = 0$ e $V_{xi}(\nu_i) = \sigma_\nu^2$. Usualmente supõe-se normalidade dos efeitos aleatórios, embora seja possível relaxar a hipótese de normalidade a fim de se tornar o modelo mais robusto.

Supondo-se que $\hat{\theta}_i = g(\hat{Y}_i)$ seja aproximadamente não viciado para $\theta_i = g(\bar{Y}_i)$, $\forall i = 1, \dots, m$, podemos escrever:

$$\hat{\theta}_i = \theta_i + \epsilon_i, \quad i = 1, \dots, m \quad (3.2)$$

onde os erros amostrais ϵ_i são supostamente independentes com $E_p(\epsilon_i | \theta_i) = 0$ e $V_p(\epsilon_i | \theta_i) = \psi_i$. Os símbolos $E_p(\cdot)$ e $V_p(\cdot)$ denotam respectivamente que a expectativa e a variância são tomadas com respeito a aleatorização induzida pelo desenho amostral empregado. Usualmente assume-se que as variâncias ψ_i 's são conhecidas e dadas pelas suas respectivas estimativas. Conforme ressaltado por Rao(2003) as hipóteses assumidas sobre os erros amostrais ϵ_i 's são muito restritivas com relação a algumas aplicações. Por exemplo, $g(\hat{\theta}_i)$ pode ser viciado para $g(\theta_i)$ se $g(\cdot)$ for uma função não linear e o tamanho da amostra em cada área for pequeno. Os erros amostrais podem não ser independentes se as unidades de seleção da amostra contiver mais de um domínio. A hipótese de variância conhecida pode ser relaxada por meio da obtenção de estimativas pontuais das mesmas sob o desenho empregado e modelando-se estas variâncias, ver Neto et al. (2007) para maiores detalhes.

Combinando-se os modelos (3.1) e (3.2) tem-se finalmente o modelo de Fay-Heriot:

$$\hat{\theta}_i = z_i^T \beta + k_i \nu_i + \epsilon_i; \quad i = 1, \dots, m \quad (3.3)$$

É importante notar que o modelo (3.3) apresenta dois tipos de erros: os erros devidos ao processo de aleatorização amostral (ϵ_i); e os erros (ν_i) provenientes do modelo ξ assumido. Assume-se também que os erros ν_i e ϵ_i são independentes. Note que o modelo (3.3) torna-se não identificável se assumirmos que as variâncias ψ_i são desconhecidas. Portanto há duas possibilidades: considerá-las conhecidas e iguais às suas respectivas estimativas (hipótese simplificadora, mais não muito realista); ou propor um modelo adicional para as variâncias

ψ_i que emprestem informação entre os domínios.

3.2 Modelo no Nível de Unidades Populacionais

Em algumas situações práticas é possível que um vetor de $p + 1$ covariáveis $\mathbf{x}_{ij} = (x_{ij0}, \dots, x_{ijp})^T$ para toda unidade populacional j em cada pequeno domínio i esteja disponível. Geralmente $x_{ij0} = 1; \forall i, j$, indicando que o modelo possui o intercepto.

O seguinte modelo de superpopulação foi originalmente proposto por Battese e Fuller (1981):

$$Y_{ij} = \mathbf{x}_{ij}^T \beta + \nu_i + \epsilon_{ij}; \quad j = 1, \dots, N_i; \quad i = 1, \dots, M \quad (3.4)$$

No modelo (3.4) os efeitos aleatórios ν_i 's (não estruturados) são supostamente independentes com média zero e variância constante e denotada por σ_ν^2 . Os ϵ_{ij} 's são supostamente i.i.d com $E_\xi(\epsilon_{ij}) = 0$ e $V_\xi(\epsilon_{ij}) = \sigma_\epsilon^2$ e independentes dos ν_i 's. Além disso, normalidade é freqüentemente assumida para os ν_i 's e os ϵ_{ij} 's. Podemos introduzir heterocedasticidade no nível de unidades populacionais, fazendo $V_\xi(\epsilon_{ij}) = k_{ij}^2 \sigma_\epsilon^2$ com k_{ij} conhecido $\forall i, j$. Para maiores detalhes, ver Rao (2003). O modelo (3.4) pode ser escrito também em notação matricial na seguinte forma:

$$Y_i = X_i \beta + \nu_i \mathbf{1}_{N_i} + E_i, \quad i = 1, \dots, M \quad (3.5)$$

onde X_i é a matriz de desenho $N_i \times p$ e $E_i = (\epsilon_{i1}, \dots, \epsilon_{iN_i})^T$

Assumimos, inicialmente, que uma amostra s_i de tamanho n_i em cada pequeno domínio $i = 1, \dots, m$ investigado seja selecionada e que os valores amostrais Y_{ij} , $i = 1, \dots, m; \quad j = 1, \dots, n_i$ seguem o mesmo modelo de superpopulação em (3.5), ou seja o desenho amostral é não informativo, veja Seção 2.5. Desta forma a parte amostral do modelo em (3.5) pode ser escrita

como:

$$y_i = x_i^T \beta + \nu_i \mathbf{1}_{n_i} + e_i, \quad i = 1, \dots, m \quad (3.6)$$

Em geral, os parâmetros de interesse são as médias (ou totais) nos respectivos domínios. Escrevendo-se o modelo (3.5) como um caso particular do modelo linear misto em (2.14) ou por simples cálculo direto, encontra-se que o melhor preditor não viciado de \bar{Y}_i é dado por:

$$\bar{Y}_i^B = f_i \bar{y}_i + (\bar{X}_i - f_i \bar{x}_i)^T (\tilde{\beta} + \tilde{\nu}_i) \quad (3.7)$$

onde: \bar{x}_i é o vetor de médias amostrais das covariáveis do modelo (incluindo o intercepto); $\tilde{\beta}$ é o estimador de mínimos quadrados generalizados de β e $\tilde{\nu}_i = \gamma_i (\bar{y}_i - \bar{x}_i^T \tilde{\beta})$ com $\gamma_i = \sigma_\nu^2 / (\sigma_\nu^2 + n_i^{-1} \sigma_\epsilon^2)$. Quando a fração de amostragem no domínio i for desprezível então \bar{Y}_i sob o modelo (3.4) pode ser aproximado por $\mu_i = \bar{X}_i \beta + \nu_i$ e o seu BLUP é dado por $\hat{\mu}_i = \bar{X}_i \tilde{\beta} + \tilde{\nu}_i$. Os preditores EBLUP's de \bar{Y}_i e de μ_i podem ser facilmente obtidos pela substituição de estimadores apropriados das componentes de variância σ_ν^2 e σ_ϵ^2 nas correspondentes expressões dos seus BLUP's, ver Capítulo 2 para maiores detalhes.

Battese et al. (1988) propuseram e aplicaram um modelo misto de intercepto aleatório para estimar a produção média por segmento de soja e milho para 12 condados no estado de Iowa. Cada condado foi dividido em segmentos de áreas e segmentos com plantações de soja ou de milho foram amostrados. O tamanho da amostra de segmentos para cada condado, n_i é pequeno, variando de 1 a 5, com $\sum n_i = 37$. Porém o número total de segmentos em cada condado, N_i varia de 394 a 687. As variáveis auxiliares foram obtidas por dados de satélite:

x_{1ij} é número de pixels classificados como soja no j -ésimo segmento, $j = 1, \dots, N_i$ do i -ésimo condado, $i = 1, \dots, 12$;

x_{2ij} é número de pixels classificados como milho no j -ésimo segmento, $j = 1, \dots, N_i$ do i -ésimo condado, $i = 1, \dots, 12$.

Após uma análise exploratória preliminar dos dados, o segundo segmento do condado de Hardin foi excluído da análise e da aplicação subsequente, uma vez que a área de milho foi erroneamente reportada para aquele segmento.

O modelo proposto por Battese et al. (1988) para o valor reportado em hectares para a produção de soja (ou milho) no segmento j do condado i pode ser escrito como:

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \nu_i + \epsilon_{ij} \quad (3.8)$$

onde $\nu_i \sim N(0, \sigma_\nu^2)$, i.i.d. e $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, i.i.d. com ν_i e ϵ_{ij} independentes.

Primeiramente, Battese et al. (1988) estimaram as componentes de variância para cada modelo utilizando o método do ajuste por constantes e então obtiveram os estimadores de mínimos quadrados generalizados de β . A Tabela 3.1 abaixo apresenta respectivamente as estimativas dos parâmetros do modelo para a soja e para o milho.

Tabela 3.1: *Estimativas dos Parâmetros dos Modelos para Soja e Milho e seus respectivos desvios-padrão em parênteses*

Parâmetro	Soja	Milho
β_0	-16(29)	51(25)
β_1	0.028(0.058)	0.329(0.050)
β_2	0.494(0.065)	-0.134(0.056)
σ_ν^2	272(49)	140(89)
σ_ϵ^2	195(59)	150(45)

Da Tabela 3.1 pode-se concluir que as três estimativas das componentes do vetor β são estatisticamente significantes no modelo para o milho. Contudo,

somente o coeficiente para a variável número de pixels de soja é significativamente diferente de zero no modelo para a soja. É importante ressaltar que para o modelo de soja, a variância entre condados é aproximadamente 60% da variabilidade total, enquanto no modelo para o milho as duas componentes de variância são aproximadamente iguais. Battese et al. (1988) também deduziram que as distribuições dos estimadores das componentes de variância são múltiplas da distribuição qui-quadrado e concluíram que a variância entre condados são significantes ao nível de 10 % e 1 % para os modelos de previsão de milho e soja, respectivamente. Diagnósticos do ajuste para ambos os modelos baseados na análise de resíduos transformados, também são apresentados em Battese et al. (1988). Os testes aplicados sugerem que não há evidência contra a hipótese de normalidade.

Tabela 3.2: *Valores preditos para a produção média de milho em cada condado com os desvios-padrão correspondentes e de preditores alternativos*

Condado	n_i	EBLUP	D. EBLUP	D. Regressão	D. M. Amostral
Cero Gordo	1	122.2	9.6	13.7	30.5
Hamilton	1	126.3	9.5	12.9	30.5
Worth	1	106.2	9.3	12.4	30.5
Humboldt	2	108.0	8.1	9.7	21.5
Franklin	3	145.0	6.5	7.1	17.6
Pocahontas	3	112.6	6.6	7.2	17.6
Winnebago	3	112.4	6.6	7.2	17.6
Wright	3	122.1	6.7	7.3	17.6
Webstar	4	115.8	5.8	6.1	15.2
Hancock	5	124.3	5.3	5.7	13.6
Kossuth	5	106.3	5.2	5.5	13.6
Hardim	5	143.6	5.7	6.1	13.6

A Tabela 3.2 acima, reproduzida de Battese et al. (1988), fornece o EBLUP para a média da produção de milho em cada condado e seu respectivo desvio-padrão. Para fins de comparação são também apresentados os desvios-padrão dos estimadores de regressão, $\mu_i^R = \bar{y}_i + (\bar{X}_i - \bar{x}_i)$, e da média amostral, \bar{y}_i . Uma tabela semelhante para a produção de soja pode ser encontrada também no artigo citado. Como era esperado, o uso da média amostral produz estimativas bastante imprecisas, quando comparadas com os estimadores EBLUP e os de regressão. Contudo, à medida que o tamanho da amostra cresce, as discrepâncias entre as precisões vão diminuindo. É importante notar que há um considerável ganho de precisão do preditor EBLUP em relação ao estimador de regressão para pequenas amostras (até 2 segmentos). Entretanto, este

ganho passa a ser pequeno para condados com três ou mais segmentos.

3.3 Modelo Multinível de Moura e Holt

Moura e Holt (1999), baseados em Moura (1994), propuseram uma generalização do modelo do intercepto aleatório, no qual permite que as inclinações também variem com os pequenos domínios. Além disso todos os coeficientes do modelo podem ser escritos como uma soma de efeitos fixos, explicados por variáveis de pequenos domínios, e efeitos aleatórios não estruturados. O modelo de superpopulação de Moura e Holt para dois níveis pode ser escrito como:

$$\begin{aligned} Y_{ij} &= x_{ij}^T \beta_i + \epsilon_{ij}, \quad j = 1, \dots, N_i \\ \beta_i &= Z_i \gamma + \nu_i, \quad i = 1, \dots, M \end{aligned} \quad (3.9)$$

onde Z_i é uma matriz de desenho $p \times q$ de variáveis explicativas de pequenos domínios, γ é um vetor $q \times 1$ de parâmetros desconhecidos, $\nu_i \sim N(0, \Omega)$ e independentes, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ e independentes. Assume-se também que os erros de primeiro e segundo níveis sejam independentes.

O modelo de dois níveis (3.9) efetivamente integra o uso de covariáveis de pequenos domínios e de unidades populacionais num único modelo:

$$Y_{ij} = x_{ij}^T Z_i \gamma + x_{ij}^T \nu_i + \epsilon_{ij}, \quad i = 1, \dots, M \quad j = 1, \dots, N_i \quad (3.10)$$

É importante notar que o modelo do intercepto aleatório de Battese et al. (1988) pode ser considerado como um caso particular do modelo acima onde Z_i é a matriz identidade para cada pequeno domínio e $\Omega = \sigma_\nu$. Outros modelos intermediários existem; por exemplo, Ω é uma matriz diagonal,

ou seja, todos os coeficientes de regressão são aleatórios mas não correlacionados.

3.3.1 O BLUP da Média para o Pequeno Domínio

Suponha que uma amostra de m pequenos domínios seja obtida e que para cada pequeno domínio i selecionado, n_i unidades populacionais sejam investigadas.

Seja $\theta = (\text{vec}(\Omega), \sigma_\epsilon^2)^T = (\theta_1, \dots, \theta_s)^T$ as componentes de variância do modelo (3.9). Aplicando-se os resultados gerais obtidos no Capítulo 2, tem-se que o BLUP de $\mu_i = \bar{X}_i^T \beta_i = \bar{X}_i^T Z_i \gamma + \bar{X}_i^T \nu_i$ é dado por:

$$\tilde{\mu}_i = \bar{X}_i^T Z_i \tilde{\gamma} + \bar{X}_i^T \tilde{\nu}_i \quad (3.11)$$

onde $\tilde{\gamma}$ é o estimador de mínimos quadrados generalizados de γ :

$$\tilde{\gamma} = \left\{ \sum_{i=1}^m Z_i^T X_i^T V_i^{-1} X_i Z_i \right\}^{-1} \sum_{i=1}^m Z_i^T X_i^T V_i^{-1} Y_i \quad (3.12)$$

e $\tilde{\nu}_i = \Omega X_i^T V_i^{-1} (Y_i - X_i Z_i \tilde{\gamma})$ onde $V_i = \sigma_\epsilon^2 I_{n_i} + X_i \Omega X_i^T$ é a matriz de covariância de Y_i .

A equação (3.12) pode ser mais facilmente calculada pela seguinte forma alternativa:

$$\tilde{\gamma} = \left\{ \sum_{i=1}^m Z_i^T G_i^{-1} X_i^T X_i Z_i \right\}^{-1} \sum_{i=1}^m Z_i^T G_i^{-1} X_i^T Y_i \quad (3.13)$$

onde $G_i = I_{n_i} + \sigma_\epsilon^2 X_i^T X_i \Omega$. É importante notar que a equação (3.13) requer que se inverta uma matriz de dimensão $(p+1)$ ao invés de dimensão n_i , caso a equação (3.12) fosse usada. As matrizes V_i^{-1} também podem ser escritas em termos de G_i^{-1} : $V_i^{-1} = \sigma_\epsilon^{-2} I_{n_i} - \sigma_\epsilon^{-2} X_i \Omega G_i^{-1} X_i^T$. Quando θ for desconhecido, há vários métodos iterativos para se estimarem θ e γ . Veja por exemplo

Longford (1987) e Goldstein, (1986, 1989). Moura e Holt (1999) utilizaram o método denominado (RIGLS) como descrito em Goldstein (1989). Sejam $\hat{\theta}$ e $\hat{\gamma}$ os estimadores RIGLS de θ e γ , respectivamente. Substituindo-se em (3.11) os estimadores RIGLS nas equações acima obtém-se o EBLUP de μ_i :

$$\hat{\mu}_i = \bar{X}_i^T Z_i \hat{\gamma} + \bar{X}_i^T \hat{\nu}_i \quad (3.14)$$

onde $\hat{\nu}_i = \hat{\Omega} X_i^T \hat{V}_i^{-1} (Y_i - X_i Z_i \hat{\gamma})$.

3.3.2 Aproximação do Erro Quadrático Médio

O erro quadrático médio de $\hat{\mu}_i$ pode ser obtido utilizando-se a decomposição do erro quadrático médio fornecida na Seção 2.4.2. Supondo-se normalidade das variáveis aleatórias envolvidas no modelo (3.9) tem-se que $\hat{\theta}$ é equivalente ao RMLE (Estimador de Máxima Verossimilhança Restrita) que por sua vez é invariante por translação; logo, os resultados descritos na Seção 2.4.2 podem ser aplicados, obtendo-se:

$$E_{\xi}(\hat{\mu}_i - \mu_i)^2 = E_{\xi}(\tilde{\mu}_i - \mu_i)^2 + E_{\xi}(\hat{\mu}_i - \tilde{\mu}_i)^2 \quad (3.15)$$

Desenvolvendo-se o primeiro termo de (3.15), ver Moura (1994) para detalhes, obtém-se:

$$\begin{aligned} EQM_{\xi}(\tilde{\mu}_i) &= E_{\xi}(\tilde{\mu}_i - \mu_i)^2 = \bar{X}_i^T (G_i^{-1})^T \Omega \bar{X}_i + \\ &+ \sigma_{\epsilon}^2 \bar{X}_i^T (G_i^{-1})^T Z_i \left\{ \sum_{i=1}^m Z_i^T G_i^{-1} X_i^T X_i Z_i \right\}^{-1} Z_i^T G_i^{-1} \bar{X}_i \\ &= T_1 + T_2 \end{aligned} \quad (3.16)$$

Moura (1994) obteve uma aproximação para o segundo termo da equação (3.15), utilizando resultados desenvolvidos por Prasad e Rao (1990):

$$\begin{aligned}
E_{\xi}(\hat{\mu}_i - \tilde{\mu}_i)^2 &\approx \text{tr} \left[\left(I_{s-1} \otimes \bar{X}_i^T (G_i^{-1})^T \right) \Delta C_i \Delta^T \left(I_{s-1} \otimes G_i^{-1} \bar{X}_i \right) B_{11}^* \right] + \\
&\quad - 2\text{tr} \left[\left(I_{s-1} \otimes \bar{X}_i^T (G_i^{-1})^T \right) \Delta R_i \Omega \bar{X}_i B_{12}^* \right] + \bar{X}_i^T \Omega S_i \Omega \bar{X}_i B_{22}^* \\
&= T_3
\end{aligned} \tag{3.17}$$

onde $\text{tr}[\cdot]$ denota o traço da matriz entre colchetes, $C_i = \sigma_{\epsilon}^{-2} G_i^{-1} X_i^T X_i$, $R_i = \sigma_{\epsilon}^{-4} G_i^{-2} X_i^T X_i$, $S_i = \sigma_{\epsilon}^{-6} G_i^{-3} X_i^T X_i$, $B^* = B^{-1} = \begin{pmatrix} B_{11}^* & B_{12}^* \\ (B_{12}^*)^T & B_{22}^* \end{pmatrix}$ é a inversa da matriz de covariância assintótica de $\hat{\theta}$ e

$$\Delta = \begin{pmatrix} \Delta_1 \\ \cdot \\ \cdot \\ \cdot \\ \Delta_{s-1} \end{pmatrix} \text{ com } \Delta_k = \frac{\partial \Omega}{\partial \theta_k}, \quad k = 1, \dots, s-1$$

Das equações (3.15) (3.16) e (3.17) uma aproximação para o erro quadrático médio de $\hat{\mu}_i$ pode ser escrita como:

$$EQM_{\xi}(\hat{\mu}_i) \approx T_1 + T_2 + T_3 \tag{3.18}$$

Quando a fração amostral em cada pequeno domínio não for desprezível, pode-se construir um estimador para a média de y em cada pequeno domínio i , incorporando predições para apenas as unidades populacionais não pertencente à amostra:

$$\hat{\mu}_i^F = f_i \bar{y}_i + (\bar{X}_i - f_i \bar{x}_i)^T (Z_i \hat{\gamma} + \hat{\nu}_i) \tag{3.19}$$

onde f_i é a fração amostral na i -ésima pequena área; \bar{x}_i é o vetor $p+1$ de médias amostrais das variáveis auxiliares, incluindo o termo constante igual a 1. Denotando-se por $\bar{X}_i^c = (1 - f_i)^{-1} (\bar{X}_i - f_i \bar{x}_i)$ e notando-se que $\hat{\mu}_i^F - \bar{Y}_i = (1 - f_i) [Z_i (\hat{\gamma} - \gamma) + (\hat{\nu}_i - \nu_i) - \bar{\epsilon}_i^c]$, onde $\bar{\epsilon}_i^c$ é a média de ϵ_i para

as unidades populacionais na i -ésima pequena área que não estão na amostra, tem-se que uma aproximação para o erro quadrático médio de $\hat{\mu}_i^F$ é dada por:

$$EQM_{\xi}(\hat{\mu}_i^F) = (1 - f_i)^2 [EQM_{\xi}^*(\hat{\mu}_i) + N_i^{-1}(1 - f_i)^{-1}\sigma_{\epsilon}^2] \quad (3.20)$$

onde $EQM_{\xi}^*(\hat{\mu}_i)$ é utilizado para se representar a equação (3.18) quando se substitui \bar{X}_i por \bar{X}_i^c .

3.3.3 Estimação do Erro Quadrático Médio

Uma prática comum ao se estimar o erro quadrático de uma combinação linear de efeitos fixos e aleatórios em um modelo linear misto é substituir estimativas das componentes de variância na respectiva expressão do EQM. Vários estudos, veja por exemplo Prasad e Rao (1990) ou Harville e Jeske (1992), mostraram que este procedimento tende a subestimar o erro quadrático médio. Prasad e Rao (1990) mostraram através de um estudo de simulação que o uso deste tipo de estimador pode resultar numa subestimação severa do erro quadrático médio. Aplicando-se os resultados desenvolvidos na Seção 2.4.3, tem-se que um estimador aproximadamente não viciado para $EQM_{\xi}(\hat{\mu}_i)$ é dado por:

$$eqm_{\xi}(\hat{\mu}_i) = \hat{T}_1 + \hat{T}_2 + 2\hat{T}_3 \quad (3.21)$$

onde \hat{T}_k , $k = 1, 2, 3$ são obtidos, substituindo-se γ , σ_{ϵ}^2 e Ω em T_k , $k = 1, 2, 3$ pelos seus respectivos RIGLS estimadores.

Um estimador aproximadamente não viciado para (3.20) é dado por:

$$eqm_{\xi}(\hat{\mu}_i^F) = (1 - f_i)^2 [eqm_{\xi}^*(\hat{\mu}_i) + N_i^{-1}(1 - f_i)^{-1}\hat{\sigma}_{\epsilon}^2] \quad (3.22)$$

onde $eqm_{\xi}^*(\hat{\mu}_i)$ é dado por (3.21), substituindo-se \bar{X}_i por \bar{X}_i^c

3.3.4 Um Estudo de Simulação

Com o objetivo de avaliar as propriedades de vários estimadores para pequenos domínios foi realizado um estudo por simulação, utilizando-se os dados de um Censo Demográfico Experimental. Os setores urbanos foram considerados os pequenos domínios de interesse, totalizando 140 setores e a renda do chefe do domicílio como variável de interesse. Após uma análise sob o poder explicatório de um conjunto de variáveis, elegeu-se o grau de instrução do chefe do domicílio em escala ordinal (0 a 5) e o número de quartos do domicílio (1-11+) como variáveis auxiliares. A população em estudo é constituída de 38740 domicílios. O modelo geral (G) assumido é dado pelas equações abaixo:

$$\begin{aligned}
 Y_{ij} &= \beta_{0i} + \beta_{1i}(x_{1ij} - \bar{X}_1) + \beta_{2i}(x_{2ij} - \bar{X}_2) + \epsilon_{ij} \\
 \beta_{0i} &= \gamma_{00} + \nu_{0i} \\
 \beta_{1i} &= \gamma_{01} + \nu_{1i} \\
 \beta_{2i} &= \gamma_{02} + \nu_{2i}, \quad i = 1, \dots, m; j = 1, \dots, N_i
 \end{aligned} \tag{3.23}$$

onde x_1 e x_2 representam, respectivamente, o número de quartos do domicílio e o grau de instrução do chefe do domicílio. Com a finalidade de se acelerar o processo iterativo de obtenção das estimativas, as variáveis explicativas foram centralizadas em torno das suas respectivas médias gerais, \bar{X}_1 e \bar{X}_2 . As estimativas dos parâmetros do modelo (3.23) e seus respectivos desvios padrão são apresentados na Tabela 3.3. Como se pode perceber dos resultados da Tabela 3.3, todos os parâmetros acima são significativamente maiores que zero.

Tabela 3.3: *Estimativas dos Parâmetros do Modelo Geral*

Parâmetro	Estimativa	Erro Padrão
γ_{00}	8.456	0.108
γ_{01}	1.223	0.046
γ_{02}	2.596	0.086
σ_{00}	1.385	0.194
σ_{01}	0.354	0.066
σ_{02}	0.492	0.117
σ_{11}	0.234	0.035
σ_{12}	0.333	0.054
σ_{22}	0.926	0.124
σ_{ϵ}^2	47.74	0.345

Com o objetivo de investigar as propriedades dos estimadores sob repetidas realizações do modelo acima e seus casos particulares, foram mantidos fixos os pequenos domínios e o conjunto de variáveis auxiliares a elas associadas. Selecionou-se uma amostra de 10% em cada pequeno domínio. Este mesmo conjunto de variáveis auxiliares foi mantido constante para todas as simulações (denominado subconjunto de simulação). Os dados gerados através do processo de simulação foram obtidos em duas etapas, gerando-se 3 conjuntos de dados de acordo com o modelo geral (G), modelo diagonal (D) e o modelo do intercepto aleatório (I). Para o modelo (G) os parâmetros utilizados na simulação são aqueles da Tabela 3.1, para o modelo (D), os elementos fora da diagonal da matriz são iguados a zero e para o modelo (I) apenas a variância do intercepto, σ_{00} , é diferente de zero. A primeira etapa do processo de simulação consistiu na geração do vetor de efeitos aleatórios ν_i para cada pequeno domínio e para cada um dos três modelos considerados. Nesta etapa a esperança para a média no *i*-ésimo pequeno domínio condicionada a cada

realização de ν_i sob o modelo $m_1 = G, D, I$ na r -ésima simulação é dado por:

$$\mu_{i,m_1}^{(r)} = \beta_{0i}^{m_1,r} + \beta_{1i}^{m_1,r}(\bar{X}_{1i} - \bar{X}_1) + \beta_{2i}^{m_1,r}(\bar{X}_{2i} - \bar{X}_2), \quad m_1 = G, D, I; r = 1, \dots, 5000 \quad (3.24)$$

A segunda etapa consiste na geração dos valores de $y_{ij}^{m_1,r}$ para cada j -ésima unidade amostral, $j = 1, \dots, n_i$ pertencente ao i -ésimo pequeno domínio, $i = 1, \dots, m$ para cada modelo $m_1 = G, D, I$ considerado. Tendo sido gerados os valores de y , todos os 3 modelos são ajustados para cada uma simulação gerada de acordo com cada modelo, sendo, portanto, possível realizar 9 combinações diferentes de dados gerados e modelo ajustado, obtendo-se então estimativas dos parâmetros e dos respectivos preditores para $\mu_{i,m_1}^{(r)}$. Para cada pequeno domínio i e para cada combinação de dados gerados de acordo com o modelo $m_1 = G, D, I$ e ajustados de acordo com o modelo $m_2 = G, D, I$ tem-se para cada simulação $r = 1, \dots, R = 5000$ a esperança para a média no i -ésimo pequeno domínio condicionada a cada realização de ν_i^r e os respectivos preditores $\hat{\mu}_{i,m_1,m_2}^{(r)}$, onde o símbolo $m_2 = G, D, I$ é utilizado para denotar o modelo que está sendo ajustado. Pode-se então obter por simulação as seguintes estimativas de variabilidade dos preditores $\hat{\mu}_{i,m_1,m_2}^{(r)}$:

i) Erro Quadrático Médio (EQM)

$$EQM[\hat{\mu}_{i,m_1,m_2}] = R^{-1} \sum_{r=1}^R (\hat{\mu}_{i,m_1,m_2} - \hat{\mu}_{i,m_1})^2 \quad (3.25)$$

ii) Erro Absoluto Relativo (EAR)

$$EAR[\hat{\mu}_{i,m_1,m_2}] = R^{-1} \sum_{r=1}^R \frac{|\hat{\mu}_{i,m_1,m_2} - \hat{\mu}_{i,m_1}|}{|\hat{\mu}_{i,m_1}|} \quad (3.26)$$

Com propósito de comparação, podem-se definir medidas relativas de variabilidade média nos pequenos domínios de cada estimador em relação àquele em que o modelo gerado coincide com o ajustado:

i) As razões das médias dos erros quadráticos médios em (%)

$$RE\bar{Q}M_{m_1, m_2} = 100 \frac{\sum_{i=1}^m EQM[\hat{\mu}_{i, m_1, m_2}]}{\sum_{i=1}^m EQM[\hat{\mu}_{i, m_1}]} \quad (3.27)$$

ii) Razões das médias dos valores absolutos em (%)

$$RE\bar{A}R_{m_1, m_2} = 100 \frac{\sum_{i=1}^m EAR[\hat{\mu}_{i, m_1, m_2}]}{\sum_{i=1}^m EAR[\hat{\mu}_{i, m_1}]} \quad (3.28)$$

Como pode ser observado na Tabela 3.4 , quando o modelo no qual os dados são gerados for mais simples do que o modelo utilizado para ajustar os dados, os estimadores resultantes não sofrem considerável perda de eficiência e aumento do vício. Por outro lado, quando os dados são gerados por um modelo mais complexo (ex. G), os estimadores obtidos através de ajustes de modelos mais simples serão mais ineficientes. Contudo a diferença de desempenho entre os estimadores obtidos através dos ajustes (G) e (D) são pequenas quando comparadas com as respectivas diferenças em relação ao (I). Da Tabela 3.4 pode-se concluir que é vantajoso introduzir coeficientes aleatórios adicionais, além do modelo do intercepto aleatório, tal como (D), mas não necessariamente (G).

Tabela 3.4: Razões das Médias dos EQM e EAR (em parênteses) para os três estimadores e os três modelos de geração dos dados assumidos.

Estimador	Modelo		
	Geral	Diagonal	I. Aleatório
Geral	100.0 (100.0)	101.8 (100.9)	101.2 (100.6)
Diagonal	108.8 (82.6)	100.0 (100.0)	100.2 (100.1)
I. Aleatório	131.9 (176.9)	109.1 (105.6)	100.0 (100.0)

Dos resultados obtidos da simulação, investigaram-se também a acurácia da aproximação (3.18) e a participação relativa das suas componentes. Considerando-se o Estimador Geral e o modelo de geração (G), a aproximação para o EQM encontrada é bastante satisfatória. Em média a aproximação calculada pela equação (3.18) é 0,31% menor que o valor do EQM calculado por simulação. As diferenças relativas entre a aproximação e o valor obtido por simulação variaram de -5.4% a 4.8 %. Em média, a participação relativa da componente T_1 é de 94,6% em relação ao EQM (variando de 87.4% a 99.1%) e a componente T_3 é de 4.3% (variando de 0,7% a 10,5%). A componente T_2 nunca contribui mais do que 2.2% do total do EQM para qualquer pequeno domínio.

Capítulo 4

Modelos Lineares Mistos Generalizados

Neste capítulo, descreveremos alguns modelos apropriados para situações práticas em que a variável resposta não pode ser considerada contínua. Trataremos desta situação por meio da introdução dos modelos lineares mistos generalizados (MLMG). Descreveremos o modelo linear misto generalizado para a família exponencial e apresentaremos e discutiremos, com detalhes, algumas aplicações com dados reais à estimação em pequenos domínios.

Assume-se que a variável resposta y_{ij} é uma realização de uma variável aleatória Y_{ij} cuja distribuição pertence à família exponencial de parâmetro canônico θ_{ij} :

$$f_p(y_{ij}|x, \theta_{ij}) = a(y_{ij})exp(y_{ij}\theta_{ij} + b(\theta_{ij})) \quad (4.1)$$

com $E(Y_{ij}|\theta_{ij}) = \mu_{ij}$ e função de ligação dada por $g(\mu_{ij}) = x_{ij}^T\beta_i$, onde $\beta_i = Z_i\gamma + \nu_i$ é como no modelo de Moura e Holt.

4.1 Estimação

Infelizmente, em geral, a função de verossimilhança não possui uma fórmula fechada, dependendo de uma integração multidimensional. Por sua vez, esta integração é difícil de ser calculada, mesmo numericamente. Jiang (2007) ressalta que embora a estimação de máxima verossimilhança e de verossimilhança restrita sejam procedimentos comuns em modelos lineares mistos e facilmente implementados, os métodos de estimação baseados na função de verossimilhança para modelos lineares mistos generalizados possuem ainda desafios computacionais a serem superados. Portanto, há várias abordagens para se fazer inferência em modelos MLMG, tentando resolver ou evitar as dificuldades computacionais surgidas pela complexidade do modelo. Estas abordagens podem ser classificadas em:

- a) métodos computacionalmente intensivos para se obter iterativamente as estimativas de máxima verossimilhança, destacando-se o algoritmo EM: McCulloch (1997); Booth e Hobert (1999) e o método de estimação por partes (Song et al. (2003));
- b) métodos aproximados de inferência: Breslow e Clayton (1993); Lin e Breslow (1996) e Lee e Nelder (1996); entre outros;
- c) métodos GEE ("Generalized Estimation Equation"): Diggle et al. (1996), Jiang (1998) e Jiang e Zhang (2001);
- d) métodos Bayesianos, via MCMC, veja Gamerman e Lopes (2006).

Um método aproximado de inferência clássica muito empregado e de fácil implementação é conhecido pela sigla PQL ("Penalized Quasi-likelihood"). Este método baseia-se na maximização da distribuição conjunta do vetor de observações y e do vetor de efeitos aleatórios ν com respeito ao vetor de

parâmetros fixos γ e o vetor de efeitos aleatórios ν . Conforme demonstrado por Jiang (1998), a aproximação de Laplace da função de verossimilhança envolvida no método PQL introduz vício aos estimadores PQL. A aproximação de Laplace de segunda ordem tem vício consideravelmente menor quando comparada com a de primeira ordem. Contudo, a aproximação de segunda ordem não elimina o vício assintoticamente. (Jiang, 2007, chap. 3) ressalta que por mais elevada que seja a ordem da aproximação de Laplace os estimadores PQL são inconsistentes. Porém, sob o ponto de vista prático, a aproximação de segunda ordem apresenta vício relativamente pequeno para os principais modelos lineares mistos generalizados empregados.

Na abordagem clássica, após terem sido estimados os parâmetros do modelo, a estimação das médias em cada domínio pode ser obtido por:

$$\hat{Y}_i = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \notin s_i} \hat{y}_{ij} \right\} \quad (4.2)$$

onde $\hat{y}_{ij} = \hat{\mu}_{ij} = g^{-1}(X_i^T \hat{\beta}_i)$ é um preditor de y_{ij} para $j \notin s_i$.

Uma vez que o modelo não é linear misto, não podemos aplicar os resultados obtidos no Capítulo 3 para obter um estimador aproximadamente não viciado para o erro quadrático médio do preditor (4.2). Contudo, uma aproximação para o erro quadrático de \hat{Y}_i para alguns casos especiais de modelos lineares mistos generalizados pode ser encontrada em Ghosh e Maiti (2004).

Analogamente à abordagem clássica, na abordagem bayesiana, assume-se que os parâmetros β_i são condicionalmente independente dado γ e Ω . Comumente $\beta_i \sim N(Z_i \gamma, \Omega)$, embora outras distribuições possam ser atribuídas. No caso do modelo especificado em (4.1) os hiperparâmetros são dados pelo vetor γ e a componente de variância σ_ν^2 . Em geral, não se têm informações sobre os hiperparâmetros e assumem-se distribuições *a priori* relativamente vagas para os mesmos, como no Exemplo 4.1 da Seção 4.3. Comumente as seguintes

distribuições *a priori* independentes são atribuídas aos hiperparâmetros γ e Ω : $\gamma \sim N(0, \Sigma_\gamma)$ e $\Omega^{-1} \sim W(\alpha, R)$, onde $W(\alpha, R)$ denota a distribuição de Wishart com parâmetros α e R .

Infelizmente, para o modelo descrito acima, a distribuição *a posteriori* dos hiperparâmetros não pode ser obtida analiticamente. Uma forma apropriada de se fazer inferência Bayesiana em modelos complexos é obter amostras da distribuição *a posteriori* dos parâmetros do modelo via simulação estocástica de cadeias de Markov (MCMC). Após um período de "aquecimento" obtém-se uma amostra da distribuição *a posteriori* de $\hat{\beta}_i$ e, portanto, a partir da função de ligação empregada e da distribuição $f_p(y_{ij}|x, \theta_{ij})$ em (4.1) obtém-se uma amostra da distribuição preditiva para Y_{ij} , $\forall j \notin s_i$. Finalmente, de posse da amostra da distribuição preditiva, $y_{ij}^{(k)}$, $\forall j \notin s_i$, $k = 1, \dots, L$, onde L é o tamanho da amostra, obtém-se uma amostra a posteriori de \bar{Y}_i :

$$\bar{Y}_i^{(k)} = N_i^{-1} \left\{ \sum_{j \in s_i} y_{ij} + \sum_{j \notin s_i} y_{ij}^{(k)} \right\} \quad (4.3)$$

4.2 Modelo Hierárquico Logístico

Alguns autores têm considerado o problema de estimar proporções em pequenos domínios usando métodos bayesianos empíricos e completos. Nesta direção pode-se citar o artigo de MacGibbon e Tomberlim (1989), que propõe estimar proporções nas áreas locais, com base em modelos de regressão logística contendo efeitos fixos e aleatórios. Esta proposta foi desenvolvida por Farrell et al. (1997), que apresentaram um estudo de simulação para comparar o erro quadrático médio do previsor do modelo hierárquico e o modelo de regressão logística sem efeito aleatório, concluindo que o modelo hierárquico é mais eficiente. Malec et al. (1997) apresentam uma abordagem bayesiana completa para

estimar proporções usando Cadeias de Markov em integração de MonteCarlo (MCMC), em particular o amostrador de *Gibbs*, mostrando que a estimação hierárquica produz estimativas mais precisas para os parâmetros do modelo do que os estimadores de regressão logística convencional. Moura et al. (2000) propuseram uma abordagem Bayesiana para a escolha de modelos binários competitivos para o problema de predição em pequenos domínios. Moura e Migon (2002) também aplicaram modelos logísticos com efeitos espacialmente estruturados para estimação de proficiências escolares em pequenos domínios do Estado do Rio de Janeiro, ver Capítulo 5.

Consideremos um modelo com dois níveis de hierarquia para a variável resposta Y_{ij} , onde Y_{ij} é uma variável binária:

$$Y_{ij} = \begin{cases} 1; & \text{se } (i,j) \text{ possui o atributo de interesse} \\ 0; & \text{se } (i,j) \text{ não possui o atributo de interesse} \end{cases}$$

onde o índice i, j representa os dois níveis hierárquicos que são estabelecidos, j denota a unidade amostral no último nível (nível mais desagregado) e i representa o sub-índice correspondente aos pequenos domínios. Supõe-se que a variável resposta Y_{ij} tenha a seguinte distribuição de probabilidade:

$$Y_{ij} \sim \text{Bern}(\pi_{ij}) \text{ i.i.d} \quad (4.4)$$

A variável Y_{ij} pode ser explicada a partir da probabilidade de resposta π_{ij} . Para relacionar π_{ij} e o vetor de covariáveis $X_{ij}^T = (x_{1ij}, \dots, x_{pij})^T$, propõe-se o seguinte modelo hierárquico logístico:

$$\begin{aligned} \text{Log} \left\{ \frac{\pi_{ij}}{1 - \pi_{ij}} \right\} &= X_{ij}^T \beta_i \\ \beta_i &= Z_i \gamma + \nu_i \end{aligned} \quad (4.5)$$

onde: β_i é um vetor de parâmetros, no qual todas as componentes podem conter efeitos aleatórios; X_{ij} representa o vetor de variáveis auxiliares associadas

com os efeitos fixos; Z_i é um vetor de variáveis associadas com os pequenos domínios, γ é o vetor de parâmetros de efeitos fixos. O vetor aleatório ν_i tem média 0 e variância Ω . Cabe mencionar que variáveis auxiliares podem ser definidas em qualquer nível de hierarquia. Algumas destas variáveis podem ser medidas de uma forma natural. Por exemplo, se considerarmos como primeiro nível a escola, neste nível pode ser medido o tamanho da escola e a sua localização. Por outro lado, se considerarmos como segundo nível o aluno, poderíamos definir variáveis como média de aproveitamento escolar do aluno.

4.3 Exemplo 4.1: Modelo Logístico para Proficiência Escolar

Os objetivos principais do Sistema Nacional de Avaliação do Ensino Básico são avaliar e reportar o estado atual dos estudantes do ensino básico e fundamental e mudanças ocorridas na qualidade do ensino. Com a finalidade de alcançar esses objetivos, pesquisas periódicas vêm sendo realizadas na última década. Um entre vários importantes sistemas de avaliações, seleciona uma amostra de estudantes em âmbito nacional e lhes aplica um teste para avaliar seus conhecimentos, supostamente adquiridos durante seus respectivos cursos. Variáveis sócio-econômicas dos alunos também são pesquisadas com objetivo de se estudar suas relações com a proficiência escolar dos mesmos. As proficiências escolares são padronizadas numa escala comum a todas as séries escolares, permitindo-se comparar a aquisição e o grau de conhecimento de alunos de séries diferentes. Variáveis relacionadas às condições das escolas também são objeto de investigação .

Embora estas pesquisas sejam planejadas para fornecerem medidas de avaliação educacional no nível agregado (ex: estado), elas não provêem estimati-

vas com precisões aceitáveis no nível mais agregado. Contudo há uma grande demanda, principalmente pelas autoridades locais, por se produzirem estimativas no nível desagregado e de se analisarem diversos fatores relacionados à proficiência escolar. Por outro lado, já é amplamente reconhecido que o uso de modelos que emprestam informações entre os domínios de interesse, melhoram as precisões das estimativas nos mesmos.

Moura e Castaneda (2000) analisaram dados extraídos de uma avaliação escolar da rede municipal do Rio de Janeiro em 1996, utilizando-se as mesmas instruções do Sistema de Avaliação do Ensino Básico (SAEB) realizado em 1995. A população alvo era constituída de 15.288 alunos da oitava série do primeiro grau no município do Rio de Janeiro, que realizaram o teste de matemática. O Município foi dividido em 34 regiões de acordo com a localização e demarcação geográfica dos bairros. O número total de escolas consideradas foi de 392. Uma vez que a população de interesse era conhecida, foi possível obter os parâmetros da população. Estes parâmetros foram utilizados para se avaliarem as estimativas amostrais produzidas por diferentes métodos. A variável resposta Y é binária (1: aluno com proficiência dentro do primeiro quartil, 0: para os outros casos).

As covariáveis utilizadas foram: idade dos alunos (x_1 e x_2) categorizadas, onde $x_1 = 0$ se idade ≤ 14 ou idade ≥ 17 , $x_1 = 1$ se $15 \leq$ idade ≤ 16 ; $x_2 = 0$ se idade ≤ 14 ou $15 \leq$ idade ≤ 16 , $x_2 = 1$ se idade ≥ 17 ; $x_3 = 0$ se sexo for masculino, $x_3 = 1$ se sexo for feminino, escolaridade dos pais: $x_4 = 0$ se não for superior, $x_4 = 1$ se for superior; localização da escola: $z_1 = 0$ se for rural, $z_1 = 1$ se for urbana. Para cada uma das 34 regiões foram selecionadas 25% das escolas, e para cada escola selecionada foram selecionadas aleatoriamente 40% dos alunos, resultando em uma amostra de 1695 alunos. A partir da amostra selecionada, ajustou-se o modelo logístico com inter-

cepto aleatório. Os parâmetros do modelo foram estimados sob a abordagem clássica, utilizando-se o método PQL de segunda ordem e sob a abordagem bayesiana via simulação estocástica por meio de cadeia de Markov (MCMC).

Considerando-se a abordagem bayesiana assume-se que a distribuição condicional de ν_{0i} dado a precisão $\tau_\nu = \sigma_\nu^{-2}$ é normal com média zero e precisão τ_ν . Foram assumidas distribuições *a priori* normais e independentes para os parâmetros de regressão, todas relativamente vagas com média zero e variância 10^6 . Atribuiu-se ao hiperparâmetro τ_ν uma distribuição gama, relativamente vaga de parâmetros 10^{-3} e 10^{-3} . Consideramos 7.000 amostras geradas via MCMC e descartamos as primeiras 2.000 amostras (aquecimento da cadeia) para se obter uma amostra de 5.000 observações das distribuições *a posteriori* dos parâmetros de interesse. Para se verificar a convergência da cadeia, utilizamos o método descrito por Geweke (1996). Considerando como primeira amostra 1.000 unidades depois do aquecimento da cadeia, a segunda amostra considerada foi as 1.000 últimas amostras da seqüência da cadeia.

As equações abaixo sumarizam o processo de se obterem amostras das distribuições *a posteriori* dos parâmetros de interesse \bar{Y}_i para $i = 1, \dots, m$:

$$\begin{aligned} \text{Calcule } \pi_{ij}^{(k)} &= [1 + \exp(-x'_{ij}\beta_i^{(k)})]^{-1} \quad \text{para } j \notin s_i, \quad k = 1, \dots, 5000 \\ \text{Amostre } y_{ij}^{(k)} &\sim \text{Bern}(\pi_{ij}^{(k)}) \\ \text{Avalie } \theta_i^{(k)} &= N_i^{-1} \left(\sum_{j \in s_i} y_{ij} + \sum_{j \notin s_i} y_{ij}^{(k)} \right) \end{aligned}$$

Se o objetivo for apenas estimar a média e a variância da *posteriori* de \bar{Y}_i , é mais eficiente utilizar os estimadores de "Rao-Blackwellized", ver Gelfand e Smith (1990), dados respectivamente por:

$$\begin{aligned} \hat{E}(\bar{Y}_i | y(s)) &= \left(\sum_{j \in s_i} y_{ij} + \sum_{j \notin s_i} \bar{\pi}_{ij} \right) N_i^{-1} \\ \hat{V}(\bar{Y}_i | y(s)) &= \left(\sum_{j \notin s_i} \bar{\pi}_{ij} \cdot (1 - \bar{\pi}_{ij}) \right) N_i^{-2} \end{aligned}$$

onde $\bar{\pi}_{ij} = N_i^{-1} \sum_{k=1}^{5000} \pi_{ij}^{(k)}$ e $y(s) = \{y_{ij}, j \in s_i, i = 1, \dots, m\}$.

Na Tabela 4.1, observa-se que as estimativas dos parâmetros de regressão do modelo hierárquico aplicando o método PQL de segunda ordem são relativamente similares às estimativas obtidas sob a abordagem bayesiana via MCMC, dadas pelas respectivas médias e desvios-padrão das distribuições *a posteriori* dos hiperparâmetros de interesse.

Tabela 4.1: *Estimativas dos Parâmetros do Modelo Logístico Hierárquico sob a abordagem clássica e bayesiana*

Estimação	Parâmetro						
	γ_0	γ_1	γ_2	γ_3	γ_4	γ_5	σ_v^2
PQL	-2,024 (0,372)	0,682 (0,158)	1,902 (0,202)	0,591 (0,138)	-0,346 (0,192)	-0,331 (0,370)	0,336 (0,124)
MCMC	-1,988 (0,390)	0,679 (0,148)	1,898 (0,189)	0,590 (0,134)	-0,359 (0,182)	-0,370 (0,388)	0,321 (0,119)

Na Tabela 4.1, observa-se que as estimativas dos parâmetros γ_1 , γ_2 , γ_3 e γ_4 são significativamente maiores do que zero para os dois métodos de estimação empregados. Sendo γ_1 positivo, implica que alunos com idade menor ou igual a 14 anos têm menor probabilidade de apresentarem proficiência baixa comparado com alunos entre 15-16 anos. Analogamente a interpretação é similar para o parâmetro γ_2 . Desta forma pode-se afirmar que alunos acima da idade escolar adequada para a oitava série apresentam maior chance de obter proficiência baixa. Com relação aos parâmetros γ_3 e γ_4 as interpretações são análogas; assim alunos do sexo masculino têm menor probabilidade de apresentar uma proficiência baixa e alunos cujos pais não possuem escolaridade superior têm maior probabilidade de apresentarem baixa proficiência. Há

evidência estatística de que alunos que residem na área rural têm maior probabilidade de apresentarem proficiência baixa. Na Figura 4.1 apresentam-se as estimativas das proporções de alunos com baixa proficiência utilizando-se os métodos de estimação PQL e MCMC. É importante notar que as estimativas pontuais produzidas pelos dois métodos são bastante próximas. É importante ressaltar que as estimativas pontuais são razoavelmente próximas das respectivas proporções reais. Por outro lado, Moura e Castaneda (2000) observa que o uso da média amostral resulta em estimativas bastante imprecisas e que há um ganho razoável ao se introduzir o efeito aleatório no intercepto no modelo de regressão logística. Na Figura 4.2 apresenta-se o diagrama Box-Plot das últimas 1000 amostras, obtidas por MCMC após o período de aquecimento, da distribuição *a posteriori* da proporção de alunos com baixa proficiência em cada região. Gráficos deste tipo são bastante úteis para se avaliar as dispersões das distribuições *a posteriori* dos parâmetros de interesse e ter uma percepção de como se distribuem estes valores. Por exemplo, da Figura 4.2 pode-se notar que as regiões 15, 21 e 33 apresentaram as maiores proporções de alunos com baixa proficiência, enquanto a região 9 se destaca por apresentar a menor proporção de alunos com baixa proficiência.

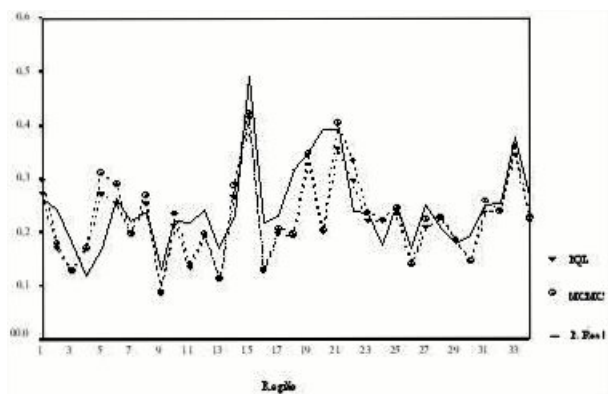


Figura 4.1: Comparação entre os métodos de estimação PQL e MCMC

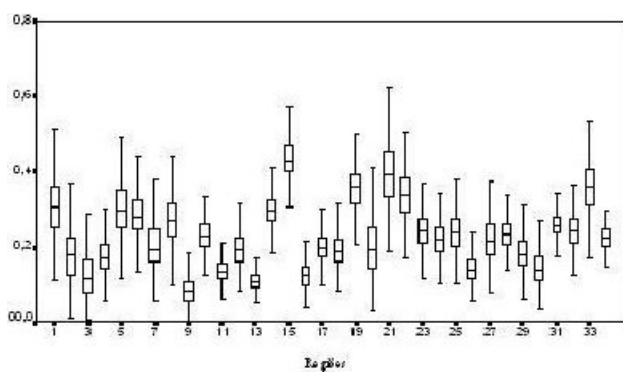


Figura 4.2: Box-Plot das últimas 1000 amostras da distribuição *a posteriori* de \bar{Y}_i para cada região, geradas via MCMC

4.4 Exemplo 4.2: Modelo de Poisson para Incidências de Doenças

O mapeamento da incidência de doenças como o câncer em pequenos domínios (resultantes da raridade das mesmas) é bastante importante para pesquisas médicas. Esses mapas permitem analisar as variações geográficas das taxas de incidência destas doenças, permitindo, por exemplo, serem indentificadas áreas de elevados riscos e que requerem intervenções urgentes. Rao (2003) cita vários estudos relativamente recentes sobre taxas de incidência de doenças e remete o leitor à edição de outubro de 2000 da revista "Statistics in Medicine" para recentes desenvolvimentos sobre mapeamento de doenças. Comumente, dados administrativos de contagem e possivelmente algumas variáveis auxiliares relacionadas estão disponíveis, e são utilizados na modelagem descrita a seguir. Dados amostrais não são empregados.

Suponha que uma determinada região (ex: estado) seja dividida em pequenas regiões (ex: municípios). Seja θ_i o risco relativo para cada pequena região $i = 1, \dots, m$. Um estimador direto de θ_i é dado por $\hat{\theta}_i = y_i/e_i$ onde y_i e e_i , respectivamente denotam o número observado de casos (mortes) da doença na pequena região $i = 1, \dots, m$ e o número esperado de casos (mortes) durante um determinado período. Os e_i 's são obtidos por meio da seguinte equação: $e_i = n_i (\sum_{i=1}^m y_i / \sum_{i=1}^m n_i)$, onde n_i é o número de pessoas-anos expostas ao risco na i -ésima pequena região. Assume-se que o termo $\sum_{i=1}^m y_i / \sum_{i=1}^m n_i$ é uma constante. O estimador $\hat{\theta}_i$ é denominado "Estimador de Razão Padronizado" (SMR). Sob o modelo $y_i | \theta_i \sim Po(e_i)$, o SMR coincide com o estimador de máxima verossimilhança de θ_i . Todavia, o mapeamento da doença, baseado nas estimativas $\hat{\theta}_i$ pode distorcer a distribuição geográfica da incidência da mesma, uma vez que o estimador θ_i é bastante impreciso para áreas em que e_i são pequenos. Note que sob o modelo de Poisson, e supondo-se que e_i são constantes, tem-se que $V_{\xi}(\theta_i) = \theta_i/e_i$. Assumindo-se uma estrutura

hierárquica para θ_i é possível se emprestar informação entre as pequenas áreas (pequenos domínios) e conseqüentemente aumentar a precisão das estimativas. Mollié (1996) relata e ilustra algumas dificuldades em se interpretar mapas de doenças utilizando os estimadores SMR. A tabela 4.2, extraída e adaptada de Mollié (1996), apresenta as estimativas das taxas relativas de risco de mortalidade de um determinado tipo de câncer em homens para algumas regiões da França no período 1971-1978 e o valor de prova para a hipótese de que $\theta_i = 1$. Conforme observado por Mollié (1996), valores extremos do SMR podem ser aqueles baseados numa amostra pequena. Por outro lado, os p-valores para se testar $\theta_i = 1$ podem ser influenciados pelos tamanhos da população, de tal forma que p-valores elevados podem apenas indentificar áreas mais populosas. Da Tabela 4.2 pode-se concluir que o valor do SMR é pequeno para a região de Lozère, mas θ_i não é significativamente diferente de 1 ao nível de 8 % de significância, enquanto para a região de Loire-Atlantique, o valor do SMR é próximo de 1, embora θ_i seja significativamente diferente de 1 (p-valor=0.002). O problema ocorre devido ao fato que a população de Loire-Atlantique é 12 vezes maior do que a de Lozère. O problema oposto ocorre com a região de Savoie (SMR alto, mais não significante ao nível de 0.092) e de Pas-de-Calais, cujo SMR é perto de 1, mas significante (p-valor=0.006). Uma forma de contornar estas dificuldades é assumir que os riscos relativos θ_i possuam uma estrutura hierárquica. Esta estrutura pode ser imposta por meio da introdução de efeitos aleatórios, também denominados de heterogeniedade não estruturada. Pode-se também supor uma heterogeniedade estruturada (ex: espaciamente estruturada) que será abordada no capítulo 5.

Tabela 4.2: Estimativas do número esperado de mortes, SMR e seus respectivos desvios-padrão, EB, e o valor de prova para a hipótese de que $\theta_i = 1$

Região	e_i	SMR	S.D. of SMR	EB	p-valor
Loire-Atlantique	48.34	0.58	0.11	0.72	0.002
Lozère	6.46	0.31	0.22	0.85	0.088
Meurthe-et-Moselle	37.66	1.75	0.22	1.44	$< 10^{-4}$
Moselle	46.40	1.72	0.19	1.46	$< 10^{-4}$
Pas-de-Calais	73.59	1.33	0.13	1.24	0.006
Savoie	18.05	1.38	0.28	1.15	0.092
Paris	158.35	0.92	0.08	0.93	0.348
Vendée	28.94	0.35	0.11	0.65	$< 10^{-4}$
Territoire-de-Belfort	6.74	1.19	0.42	1.03	0.476

4.4.1 Modelo Poisson-Gamma

Considere um vetor de observações $y = (y_1, \dots, y_m)^T$ onde os y_i 's são dados provenientes de uma contagem para cada pequeno domínio $i = 1, \dots, m$. Analogamente ao modelo logito de dois estágios descrito pelas equações (4.4) e (4.5) tem-se:

$$\begin{aligned} Y_i | \theta_i &\sim Po(e_i \theta_i); \quad i.i.d. \\ \theta_i | \alpha, \beta &\sim Ga(\alpha, \beta); \quad i.i.d. \end{aligned} \quad (4.6)$$

Na abordagem bayesiana, o modelo (4.6) é completado atribuindo-se uma distribuição *a priori* para os hiperparâmetros α e β . Por exemplo, $\alpha \sim Ga(a_\alpha, b_\alpha)$ e $\beta \sim Ga(a_\beta, b_\beta)$. Amostras da distribuição *a posteriori* de θ_i podem ser obtidas via MCMC.

Uma alternativa bastante utilizada na prática é o estimador empírico de Bayes (EB), que pode ser encontrado, calculando-se primeiramente:

$$E_\xi(\theta_i | y, \alpha, \beta) = (y_i + \alpha) / (e_i + \beta) \quad (4.7)$$

e substituindo-se as quantidades desconhecidas em (4.7) pelos seus respectivos estimadores, obtidos, por exemplo, pelo método de máxima verossimilhança ou dos momentos. Não existem formas fechadas para os estimadores de máxima verossimilhança de α e β . Contudo os seus respectivos estimadores de momentos, podem ser obtidos resolvendo-se as seguintes equações, ver Marshall (1991) para maiores detalhes:

$$\begin{aligned}\frac{\hat{\alpha}}{\hat{\beta}} &= \hat{\theta}_s \\ \frac{\hat{\alpha}}{\hat{\beta}^2} &= s_e^2 - \frac{\hat{\theta}_s}{e}.\end{aligned}\quad (4.8)$$

onde $\hat{\theta}_s = \frac{1}{m} \sum_{l=1}^m (e_l/e.) \hat{\theta}_l$, $s_e^2 = \frac{1}{m} \sum_{i=1}^m (e_i/e.) (\hat{\theta}_i - \hat{\theta}_s)^2$ e $e. = \frac{1}{m} \sum_{l=1}^m e_l$.

Rao (2003) sugere substituir α e β pelos seus respectivos estimadores de momentos, obtendo-se:

$$\hat{\theta}_i^{EB} = (y_i + \hat{\alpha}) / (e_i + \hat{\beta}) = \hat{\gamma}_i \hat{\theta}_i + (1 - \hat{\gamma}_i) \hat{\theta}_s \quad (4.9)$$

onde $\hat{\gamma}_i = e_i / (e_i + \hat{\beta})$. É importante notar que $\hat{\theta}_i^{EB}$ é uma média ponderada do estimador direto $\hat{\theta}_i$ e o estimador sintético $\hat{\theta}_s$. À medida que e_i aumenta, maior peso é dado ao estimador direto (SMR). Se $s_e^2 < \hat{\theta}_s/e.$, fixa-se $\hat{\theta}_i^{EB} = \hat{\theta}_s$, a fim de não serem obtidas estimativas negativas (ver a segunda equação em (4.8)). Rao (2003) ressalta que o estimador EB é aproximadamente não viciado, no sentido de que o vício é de ordem m^{-1} para m grande. Um estimador aproximadamente não viciado do erro quadrático médio de $\hat{\theta}_i^{EB}$ pode ser obtido pelo método de *Jackknife*, ver Rao (2003). Lahiri e Maiti (1999) obtiveram um estimador do erro quadrático médio por expansão em série de Taylor, utilizando *bootstrap* paramétrico para estimar a matriz de covariância de $\hat{\alpha}$ e $\hat{\beta}$.

Na Tabela 4.2, extraída de Mollié (1996) apresentam-se as estimativas EB das taxas relativas de risco de mortalidade. As estimativas de α e β foram

calculadas pelo método de máxima verossimilhança. Mollié (1996) observa que as estimativas EB apresentam menor variabilidade entre as regiões do que as SMR. Além disso, regiões com populações menores tendem a apresentar discrepâncias maiores entre as estimativas SMR e EB. Portanto, como já era esperado, as estimativas EB tendem a suavizar valores extremos de SMR, particularmente em regiões com populações pequenas. Mollié (1996) também emprega uma análise bayesiana completa via MCMC e obtém estimativas similares às obtidas pelo método EB. Contudo, argumenta que a análise bayesiana completa possui a vantagem adicional de produzir intervalos de credibilidade.

O modelo de Poisson-Gama pode ser estendido para possibilitar a introdução de covariáveis de regiões que possam explicar parte do risco de morte (doença). Clayton e Kaldor (1987) assumiram que o valor esperado do risco relativo em cada área varie com o parâmetro de escala e propuseram o seguinte modelo loglinear: $\text{Log}(E(\theta_i)) = \log(\alpha/\beta_i) = z_i^T \gamma$.

4.4.2 Modelo Log-Normal

Uma outra alternativa de modelagem em dois estágios para dados de contagem é assumir que os riscos relativos na escala logarítmica são normalmente distribuídos. Neste caso, covariáveis z_i de regiões são facilmente introduzidas:

$$\begin{aligned} Y_i | \theta_i &\sim \text{Po}(e_i \theta_i); \text{ i.i.d.} \\ \log(\theta_i) | \gamma, \sigma^2 &\sim N(z_i \gamma, \sigma^2); \text{ i.i.d.} \end{aligned} \quad (4.10)$$

A implantação do método de estimação EB torna-se mais complicada, uma vez que não existem expressões fechadas para $E_\xi(\theta_i | y, \mu, \sigma^2)$ e $V_\xi(\theta_i | y, \mu, \sigma^2)$. Clayton e Kaldor (1987) aproximaram a distribuição de $\log(\theta) | y, \mu, \sigma^2$ por uma distribuição normal multivariada, onde $\theta = (\theta_1, \dots, \theta_m)^T$. Aplicando o algoritmo EM, obtiveram as estimativas de máxima verossimilhança de μ e σ^2 .

A partir da aproximação normal multivariada, obtiveram aproximações para os respectivos EB's de $\log(\theta_i)$; $i = 1, \dots, m$, denotados por $\hat{\zeta}_i^{EB}$. Aplicando a transformação inversa, derivaram uma expressão para o EB de θ_i :

$\hat{\theta}_i^{EB} = \exp(\hat{\zeta}_i^{EB})$. No entanto, o estimador EB obtido por esta aproximação é aproximadamente viciado. Rao (2003) mostra como realizar ajuste em $\hat{\theta}_i^{EB}$ para se obter um estimador aproximadamente não viciado.

Jiang e Zhang (2001) sugere utilizar os estimadores de momentos de μ e σ^2 para simplificar o cálculo do estimador erro quadrático de $\hat{\theta}_i^{EB}$ por *jackknife*.

Capítulo 5

Tópicos Avançados

Neste capítulo trataremos de alguns tópicos usualmente não discutidos na literatura. Alguns modelos aqui apresentados são extensões daqueles apresentados nos capítulos anteriores. Discutiremos algumas aplicações com dados reais e simulados para ilustrar o uso da metodologia abordada.

5.1 Modelos com Erros Correlacionados

Várias extensões do modelo de Fay-Heriot, descrito no Capítulo 3 pelas equações (3.1) e (3.2), têm sido propostas na literatura para lidar com a situação em que os erros amostrais não podem ser considerados independentes. Nesta Seção descreveremos alguns destes modelos.

Rao e Yu (1994) propuseram uma extensão do Modelo de Fay-Heriot para lidar com observações correlacionadas no tempo. O modelo de Rao e Yu (1994) pode ser escrito como:

$$\begin{aligned}\hat{\theta}_{i,t} &= \theta_{i,t} + \epsilon_{i,t}, \quad t = 1, \dots, T; \quad i = 1, \dots, m \\ \theta_{i,t} &= z_{i,t}\gamma + \nu_i + u_{i,t}\end{aligned}\tag{5.1}$$

onde $\hat{\theta}_{i,t}$ é o estimador de $\theta_{i,t}$ obtido pelo desenho amostral, os $\epsilon_{i,t}$'s são normalmente distribuídos com médias iguais a zero e matriz Σ de covariância bloco-diagonal com blocos Σ_i ; $z_{i,t}$ é um vetor de variáveis auxiliares cujas componentes estão disponíveis para $t = 1, \dots, T$; $i = 1, \dots, m$. Além disso, $\nu_i \sim N(0, \sigma_\nu^2)$, $\forall i = 1, \dots, m$ e independentes. Assume-se que os $u_{i,t}$'s seguem um processo autoregressivo de primeira ordem (AR1):

$$u_{i,t} = \rho u_{i,t-1} + \varsigma_{i,t}, \quad |\rho| < 1 \quad (5.2)$$

com erros $\varsigma_{i,t} \sim N(0, \sigma^2)$ e independentes. Supõe-se também que os erros $\{\varsigma_{i,t}\}$, $\{\nu_{i,t}\}$ e $\{\epsilon_{i,t}\}$ sejam mutualmente independentes. Outros processos autoregressivos podem ser assumidos para u_{it} , como por exemplo o processo ARMA. Contudo, segundo Rao (2003), o ganho de eficiência em relação ao modelo (AR1) dificilmente será significativo. Outras referências importantes, envolvendo algumas modificações na estrutura de $u_{i,t}$ no modelo (5.1), podem ser encontradas em Datta et al. (1999), You (1999) e Datta et al. (2002).

Modelos dinâmicos também têm sido propostos na literatura para estimação em pequenos domínios, para se emprestar informações entre as séries de estimativas temporais. Ghosh e Nangia (1993) e Ghosh et al. (1996) propuseram o seguinte modelo com estrutura dinâmica no tempo:

$$\begin{aligned} \hat{\theta}_{i,t} | \theta_{i,t} &\sim N(\theta_{i,t}, \psi_{i,t}), \quad t = 1, \dots, T; \quad i = 1, \dots, m \\ \theta_{i,t} | \alpha_t &\sim N(z_{i,t}\beta + w_{it}\alpha_t, \sigma_t^2) \\ \alpha_t | \alpha_{t-1} &\sim N(H_t\alpha_{t-1}, \Delta) \end{aligned} \quad (5.3)$$

onde $z_{i,t}$ e w_{it} são vetores de variáveis auxiliares de pequenos domínios, as variâncias amostrais $\psi_{i,t}$'s são supostamente conhecidas, α_t é o vetor $r \times 1$ evoluindo dinamicamente no tempo e H_t é uma matriz conhecida de dimensão $r \times r$.

No modelo (5.3) assume-se independência no tempo dos estimadores θ_{it} para cada $i = 1, \dots, m$. Esta hipótese não é defensável no contexto de pesquisas longitudinais, em que as amostras de painéis tenham interseções. Outra deficiência do modelo (5.3) é a não inclusão de efeitos aleatórios de pequenos domínios e portanto não levar em consideração a variabilidade entre os mesmos não explicada pelas variáveis auxiliares. Pfeiffermann e Burck (1990) propuseram um modelo dinamicamente estruturado bastante geral, resultando em vários casos particulares encontrados na literatura.

5.2 Modelos Espacialmente Estruturados

Em muitas situações práticas é adequado assumir que as observações são espacialmente estruturadas. Um exemplo típico é a taxa de incidência de uma doença numa determinada região, onde se verifica que está relacionada com regiões vizinhas, devido à contaminação. Diferentemente dos modelos de efeitos aleatórios não estruturados, os modelos espaciais induzem correlações entre os efeitos aleatórios e portanto são denominados efeitos aleatórios espacialmente estruturados. Várias propostas têm sido apresentadas na literatura para tratar de dados espacialmente estruturados. Ao leitor interessado, sugerimos consultar Banerjee et al. (2004) e Schmidt e Sansó (2006). No contexto de estimação em pequenos domínios, alguns casos particulares do modelo MRF ("Markov Random Fields") têm sido bastante empregados e investigados na literatura, principalmente no contexto da Inferência Bayesiana. Nesta seção descrevemos o modelo CAR ("Conditional Autoregressive Model") na sua forma geral e fornecemos alguns exemplos de suas aplicações. O modelo é descrito no contexto da inferência Bayesiana.

Nos modelos de efeitos aleatórios não estruturados, geralmente assumimos

que condicionalmente, $\nu = (\nu_1, \dots, \nu_m)^T$ possui uma distribuição multivariada normal com média $\mathbf{0}$ e matriz de covariância diagonal, com elementos da diagonal iguais a (σ_ν^2) . Isto implica independência condicional entre estes efeitos. Contudo, conforme já observado no início desta seção, áreas vizinhas (pequenos domínios) podem ser semelhantes com relação à característica de interesse. Um modelo adequado que considera a estrutura de semelhança entre áreas vizinhas é o modelo CAR, cuja distribuição condicional conjunta de ν é dada por:

$$f(\underline{\nu}|\Phi) \propto |\Phi|^{m/2} \exp\left\{-\mathbf{0.5} \sum_{i,j=1}^m \mathbf{w}_{ij}(\nu_i - \nu_j)' \Phi (\nu_i - \nu_j)\right\} \quad (5.4)$$

onde w_{ij} são ponderadores associados a estrutura de vizinhança. A estrutura espacial pode ser representada por diferentes escolhas dos ponderadores. Por exemplo, fazendo-se $w_{ij} = 1$ se as regiões i e j são contíguas e $w_{ij} = 0$, caso contrário. Neste caso, o modelo (5.4) reduz-se à seguinte representação condicional equivalente:

$$\nu_i | (\nu_{\mathbf{k}}, \mathbf{k} \in \partial i, \nu) \sim \mathbf{N}(\bar{\nu}_i, \mathbf{k}_i^{-1} \Phi^{-1}) \quad (5.5)$$

onde o símbolo ∂i denota o conjunto de regiões vizinhas da área i , $\bar{\nu}_i = \mathbf{k}_i^{-1} \sum_{\mathbf{k} \in \partial i} \nu_{\mathbf{k}}$, sendo k_i o número de vizinhos da área i . Usualmente é atribuída *a priori* uma distribuição Wishart à matriz Φ , aqui denotada por $W(\alpha_\Phi, R_\Phi)$.

É importante observar que a distribuição em (5.4) é imprópria, uma vez que adicionando-se uma constante aos ν'_i s sua expressão não é alterada. Devemos impor uma restrição aos ν'_i s para que o modelo seja identificável. Bessag e Kooperberg (1995) sugere impor a restrição $\sum_{i=1}^m \nu_i = 0$ e especificar o intercepto separadamente com uma distribuição uniforme na reta. Esta metodologia está implantada no pacote computacional "Winbugs", ver Spiegelhalter et al. (2002).

5.3 Exemplo 5.1: Modelo Espacial Logístico

Moura e Migon (2002), entre outros modelos, propuseram e analisaram um modelo logístico com os coeficientes de regressão espacialmente estruturados segundo um modelo CAR. Os dados utilizados foram os de proficiência escolar analisados anteriormente por Moura e Castaneda (2000), ver Exemplo 4.1. De maneira análoga ao Exemplo 4.1 tem-se:

$$Y_{ij} \sim \text{Bern}(\pi_{ij}) \quad \text{i.i.d}$$

$$\text{Log} \left\{ \frac{\pi_{ij}}{1 - \pi_{ij}} \right\} = X_{ij}^T \beta_i$$

Moura e Migon (2002) assumiram que $\beta_i = \delta_i + \nu_i$, onde δ_i são efeitos aleatórios não estruturados, i.e: $\delta \sim N(0, \sigma_\delta^2)$; i.i.d e os ν_i 's seguem um modelo CAR tal como descrito na seção (5.2). Aqui consideramos o caso particular em que todos os coeficientes de regressão são fixos, exceto o intercepto. Portanto tem-se: $\beta_i = (\beta_{0i}, \dots, \beta_{pi})^T$ com $\beta_{0i} = \delta_{0i} + \nu_{0i}$ e $\beta_{ki} = \gamma_k$; $\forall k = 1, \dots, p$. Denominamos este caso especial de "Modelo Logístico de Intercepto Aleatório-Espacial". Distribuições *a priori* relativamente vagas foram atribuídas aos hiperparâmetros do modelo. Para maiores detalhes consultar Moura e Migon (2002). Similarmente ao método apresentado na seção 4, geraram-se 5000 amostras das distribuições *a posteriori* das quantidades de interesse θ_i , $i = 1, \dots, m$ via MCMC, após um período de aquecimento de 5000 amostras.

As Figuras 5.1 e 5.2 mostram como se distribuem espacialmente as médias das distribuições *a posteriori* das 34 regiões, consideradas como pequenos domínios, do estado do Rio de Janeiro. Analisando-se cuidadosamente as duas figuras, pode-se concluir que o modelo logístico sem efeito aleatório não captura corretamente diferenças entre as áreas com relação à proficiência escolar, enquanto que o modelo logístico de intercepto aleatório-espacial é mais

flexível, permitindo que os dados identifiquem semelhanças espacialmente estruturadas, como também as não estruturadas.

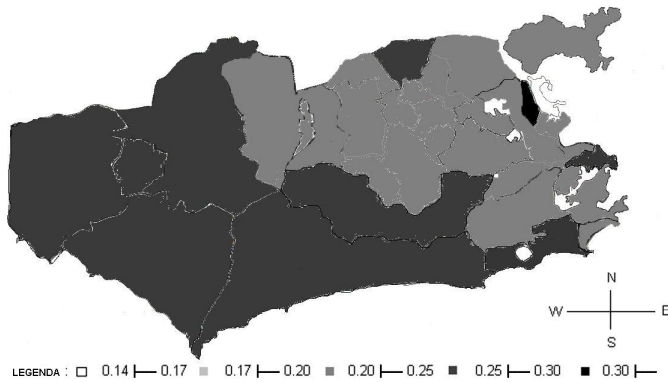


Figura 5.1: Mapa do Rio de Janeiro representando as médias das Distribuições *a posteriori* das proporções de alunos com baixa proficiência obtida com o modelo logístico sem efeito aleatório.

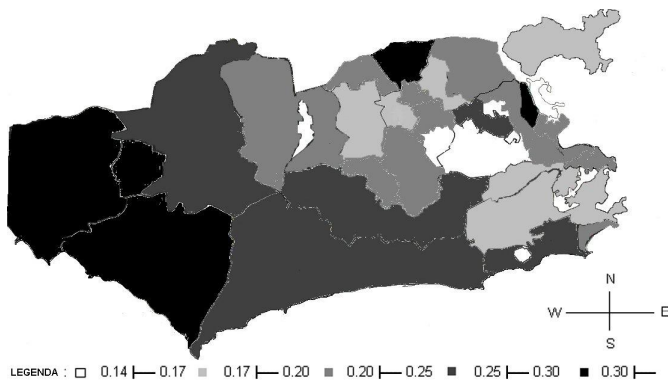


Figura 5.2: Mapa do Rio de Janeiro representando as médias das Distribuições *a posteriori* das proporções de alunos com baixa proficiência obtida com o modelo logístico de intercepto aleatório-espacial

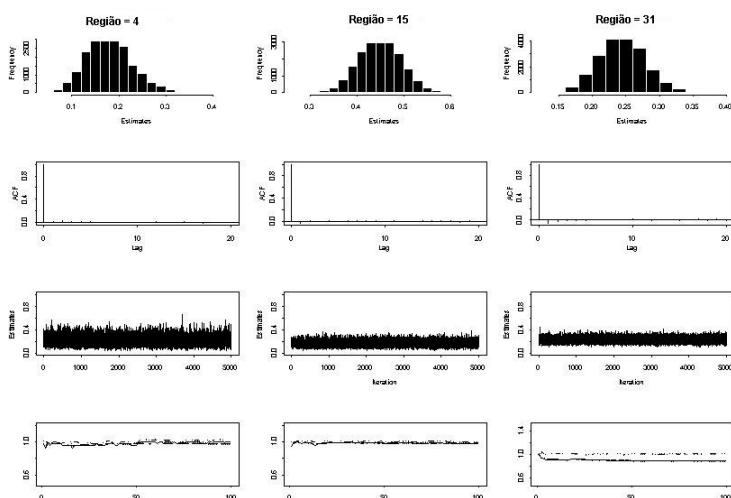


Figura 5.3: Histograma das 5000 amostras da distribuição *a posteriori* de θ_i , função de autocorrelação, os traços das 4 cadeias geradas e o gráfico para se checar convergência baseado em Brooks e Gelman (1998). Consideraram-se três regiões e o modelo logístico de intercepto aleatório-espacial

Diagnósticos informais, baseados em técnicas gráficas foram aplicados para checar a convergência das amostras geradas das proporções de interesse θ'_i s. Observaram-se o histograma das 5000 amostras geradas, a função de autocorrelação de primeira ordem, e o gráfico de série de 4 cadeias de 5000 amostras cada, sendo que para cada cadeia os hiperparâmetros do modelo são inicializados com diferentes valores. Diagnósticos formais de convergência baseados nos critérios de Geweke (1996) e Brooks e Gelman (1998) também foram realizados.

Na Figura 5.3 apresenta-se um resumo das técnicas gráficas empregadas para se checar convergência, considerando-se o modelo do intercepto aleatório-espacial e três regiões criteriosamente escolhidas. As três regiões, numeradas

4, 31 e 15 representam, respectivamente, regiões com baixo, médio e alto valor do parâmetro de interesse θ_i . O histograma apresentado na Figura 5.3 indica que, para alguns casos, a distribuição *a posteriori* de θ_i poder ser levemente assimétrica, como para a região 4. A função de autocorrelação decresce rapidamente. O traço das quatro cadeias indica que a convergência parece ser atingida rapidamente. O gráfico para checar convergência, baseado em Brooks e Gelman (1998), não revela nenhum problema de não convergência. O maior valor encontrado para a medida de diagnóstico de Geweke foi 1.895, correspondendo a um p-valor de 0.06 para hipótese de não convergência.

Moura e Migon (2002) propuseram um critério de seleção de modelos, apropriado para a estimação em pequenos domínios e baseado na medida de seleção de modelos, proposta por Gelfand e Ghosh (1998):

$$D(l) = \sum_{i=1}^m \left(\mu_i^{(l)} - \bar{y}_i \right)^2 + \sum_{i=1}^m \sigma_i^{2(l)} \quad (5.6)$$

onde l denota o modelo considerado, $\mu_i^{(l)}$ é a média a posteriori do parâmetro de interesse θ_i sob o modelo l , \bar{y}_i é a proporção amostral e $\sigma_i^{2(l)}$ é a variância *a posteriori* de θ_i sob o modelo l . Modelos com menores valores de $D(l)$ são preferíveis, segundo este critério. É importante notar que a medida $D(l)$ é composta pela soma de dois termos. O primeiro termo, denotado por $G(l)$, é uma medida de ajuste e geralmente decresce com o crescimento da complexidade do modelo. O segundo termo, $P(l)$, é uma medida de penalidade de modelos não parcimoniosos. Aplicando-se este critério a vários modelos alternativos, Moura e Migon (2002) concluíram que os modelos que envolviam alguma estrutura espacial pareciam ser mais adequados. Como as verdadeiras proporções são conhecidas, foi possível verificar se os modelos mais adequados correspondiam realmente àqueles que produziam a menor média das diferenças ao quadrado entre a média *a posteriori* de θ_i e o seu respectivo

verdadeiro valor P_i , i.e: $AMSE(l) = \frac{1}{34} \sum_{i=1}^{34} (\mu_i^{(l)} - P_i)^2$. Concluíram que a medida proposta é uma ferramenta útil para a seleção de modelos competitivos. Contudo, como qualquer outro critério de seleção de modelos, ela não consegue distinguir entre modelos com aproximadamente o mesmo poder médio de predição. Moura e Migon (2002) apresentam uma tabela com os valores das médias e dos desvios-padrão *a posteriori* para o modelo logístico sem efeitos aleatórios e o logístico com intercepto aleatório-espacial. Para fins de comparação, apresentam também as estimativas amostrais das proporções e os seus respectivos desvios-padrão. Os resultados encontrados mostram um ganho expressivo em termos de precisão do modelo logístico com intercepto aleatório-espacial em relação ao modelo logístico sem efeitos aleatórios. A razão entre o desvio-padrão da distribuição *a posteriori* do modelo logístico com intercepto aleatório-espacial e do modelo logístico sem efeitos aleatórios varia de 0.37 a 0.62. O ganho em precisão, ainda é mais expressivo quando os desvios-padrão das distribuições *a posteriori* do modelo de intercepto-aleatório são comparados com os respectivos desvios-padrão das proporções amostrais. Neste caso a razão varia apenas de 0.26 a 0.46.

5.4 Exemplo 5.2: Estimação de Populações Municipais

Nas últimas décadas, as pesquisas por amostragem com múltiplos propósitos tomaram grande impulso por apresentarem menor custo em relação aos censos e contagens, além de possibilitarem a obtenção de informações demográficas e sócio-econômicas da população com periodicidade menor que a dos censos. No entanto, essas pesquisas são, em sua maioria, de âmbito nacional e na maioria

das vezes os tamanhos de amostras são pequenos para que se possam produzir estimativas confiáveis no nível de detalhamento desejado. Um exemplo de pesquisa por amostragem de múltiplos propósitos é a Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). O procedimento comum adotado é a produção de estatísticas para níveis geográficos maiores dos que são desejáveis.

No caso específico da obtenção de estimativas das populações municipais, são realizadas projeções populacionais em que, primeiro, é feita a projeção para uma região maior, considerando sua evolução populacional com base em hipóteses sobre a natalidade, a mortalidade e a migração e, depois, o resultado é repartido de alguma forma entre os municípios. Isso é feito pelo método de relação de coortes (Duchesne (1987)), por exemplo. Contudo, algumas técnicas de projeção caracterizam-se por não apresentar o erro das estimativas, como aquela conhecida por método dos coeficientes (IBGE (2002)). Uma alternativa aos modelos determinísticos de projeção demográfica é assumir modelos probabilísticos, levando-se em consideração todas (ou grande parte) das fontes de incertezas das observações e conseqüentemente fornecendo-se uma medida do erro das estimativas. Adotando esta abordagem, Souza et al. (2007) propuseram um modelo de crescimento exponencial com componentes espacialmente estruturadas para se estimar as densidades demográficas por município e conseqüentemente as respectivas populações municipais. Nas seções seguintes faremos uma descrição sucinta do trabalho de Souza et al. (2007).

5.4.1 Conjunto de Dados

Os municípios do estado de São Paulo formam a base de dados do estudo. Nos anos censitários de 1991 e 1996, dispõe-se dos tamanhos e conseqüentemente das densidades populacionais de todos os seus m municípios. Nos anos in-

tercensitários é possível obter as estimativas dos tamanhos populacionais dos k municípios ($k < m$) selecionados pela Pesquisa Nacional por Amostragem de Domicílios (PNAD), adaptando-se o estimador apresentado em Klein e Moura (1998). Como o nível município não faz parte do domínio de divulgação da PNAD, as estimativas obtidas desta forma apresentam baixa precisão. Para detalhes sobre o plano amostral da PNAD, o leitor pode consultar Bianchini e Albieri (2003). No entanto, mostra-se que as informações da PNAD combinadas com os censos resultam em estimativas mais precisas, quando se utiliza modelos adequados.

Os dados da PNAD de 1992 a 1999 foram utilizados nesta aplicação, excetuando-se o ano de 1994, para o qual a pesquisa não foi realizada. Ao longo do período houve modificações na malha municipal do estado, ocorrendo criação de novos municípios a partir de alguns existentes. Este fato levou à decisão de seguir a base territorial vigente em 1991, descartando os municípios com diferenças significativas entre a área geográfica apurada no censo de 1991 e no censo de 2000. Assim, dentre os 111 municípios selecionados pela PNAD na década de 90, apenas 91 foram usados. Como censo, contagem e PNAD possuem datas de referência distintas foi preciso compatibilizar estas datas, ver Souza (2004) para detalhes.

5.4.2 Descrição do Modelo

Seja y_{it} a densidade observada (quando censo) ou estimada (quando PNAD) da pequena área i no tempo t e por π_{it} e μ_{it} os verdadeiros valores da densidade e da população, para $i = 1, \dots, m$ e $t = 1, \dots, n$. O interesse é fazer inferência sobre as populações μ_{it} de todos os m municípios a partir de π_{it} . Os modelos apresentados abaixo são hierarquicamente estruturados, onde os verdadeiros valores das densidades populacionais, π_{it} 's, são descritos via uma função não

linear de crescimento. Assume-se que as observações y_{it} são normalmente distribuídas e condicionalmente independentes dados a média π_{it} e a variância σ_{it}^2 . Nos censos demográficos, apesar do objetivo ser investigar inteiramente a população, ocorrem omissões ou inclusões indevidas que se constituem no erro de cobertura. Então é razoável, para anos censitários, considerar o valor observado como sendo igual à população verdadeira mais um erro, da forma que está descrito na primeira equação do modelo. Por sua vez, a PNAD é uma pesquisa por amostragem e, portanto, ϵ_{ij} representa a principal componente do erro de estimação, como descrito na primeira equação do modelo abaixo. Além disso, assume-se que a função média π_{it} seja dada por uma curva de crescimento exponencial. A curva de crescimento permite que ocorra tanto crescimento quanto decréscimo de densidade e portanto de população. As fontes de dados trabalhadas possuem datas de referência diferenciadas e não são igualmente espaçadas. Neste caso, a utilização da curva de crescimento exponencial constitui-se numa vantagem, já que é possível fazer uma simples codificação do tempo para compatibilizar as datas.

$$\begin{aligned}
 y_{it} &= \pi_{it} + \epsilon_{it}, & \epsilon_{it} &\sim N(0, \sigma_{it}^2) \\
 \pi_{it} &= a_i + \beta \exp(c_i t) \\
 a_i &= \alpha + \xi_{ai}, & \xi_{ai} &\sim N(0, \tau_{ai}^2) \\
 c_i &= \gamma + \xi_{ci}, & \xi_{ci} &\sim N(0, \tau_{ci}^2)
 \end{aligned} \tag{5.7}$$

Souza et al. (2007) também consideraram um modelo alternativo ao modelo (5.7) em que o parâmetro a_i é espacialmente estruturado segundo um modelo CAR, como descrito no Exemplo 5.1.

Supondo que a função média, π_{it} , é não explosiva, o parâmetro α pode ser interpretado como o ponto de estabilização populacional. Os parâmetros β e γ controlam o crescimento da densidade. As distribuições *a priori* de α ,

β e γ podem ser escolhidas aproveitando-se algum conhecimento da evolução populacional média dos municípios do estado de São Paulo. No modelo (5.7), quando $t = 0$, a média da população inicial do município i é dada por $a_i + \beta$. O crescimento do município i no tempo é dado por $\exp(c_i)$, que juntamente com β , constitui um fator de acréscimo ou de redução populacional. A hierarquia em $a = (a_1, \dots, a_m)^T$ mostra que, em média, a população inicial dos municípios é $\alpha + \beta$, sofrendo deslocamento devido ao efeito de município ξ_{ai} . Considerar c_i hierárquico equivale a dizer que os municípios experimentam crescimentos populacionais diferenciados, mas similares, dependendo de um efeito ξ_{ci} . Atribuíram-se distribuições relativamente vagas aos hiperparâmetros do modelo, ver Souza et al. (2007) para detalhes. No modelo espacialmente estruturado, a distribuição *a priori* de a_i é a mesma sugerida por Mollié (1996).

5.4.3 Modelagem da Variância das Observações

Como foram utilizados dados de duas pesquisas com diferentes variabilidades, assumiu-se que as variâncias das observações dependesse do ano. Foi proposto o seguinte modelo para as variâncias das observações: $\log(\sigma_{it}^2) = \eta_0 + \eta_1(1/n_i)$, se no ano t utiliza-se o dado da PNAD, onde n_i é o número de setores selecionados pela PNAD no município i . A expectativa era que a variância fosse reduzida com o aumento do tamanho da amostra de setores, o que foi confirmado. Para anos censitários tem-se: $\log(\sigma_{it}^2) = \log(\text{var}_{it})$ onde var_{it} é a variância do censo no município i no ano t , calculada em função do erro de cobertura do censo, considerado como sendo de 5% para todos os municípios. Então a verdadeira população (e também a densidade, sob a hipótese de que a extensão territorial do município permanece inalterada no período em estudo) estaria localizada no intervalo dado pela população observada no censo mais ou menos 5% de seu valor. Uma estimativa do desvio-padrão é dada

pela diferença entre o extremo superior do intervalo menos a população observada, dividindo-se a diferença por dois. A variância da densidade é então obtida, dividindo-se a variância da população pelo quadrado da área geográfica do município. Estabelecer variâncias conhecidas nos anos censitários é uma forma de se dar mais peso às suas observações, uma vez que se espera que um censo produza informações mais confiáveis para pequenos domínios que uma pesquisa por amostragem. Atribuíram-se distribuições *a priori* normais e relativamente vagas aos parâmetros η_0 e η_1 .

5.4.4 Aspectos Computacionais e Alguns Resultados

A distribuição *a posteriori* obtida sob o modelo (5.7) não é tratável, sendo necessário recorrer a métodos de simulação indireta, tais como os baseados em *Markov Chain Monte Carlo* (MCMC), para gerar uma amostra de tal distribuição e fazer inferências com base na mesma. Como as distribuições condicionais completas de cada parâmetro, dadas as observações e todos os demais parâmetros, eram conhecidas, com exceção do vetor $\mathbf{c} = (c_1, \dots, c_k)^T$, foi adotado o método de simulação conhecido por *Gibbs sampler*, introduzindo-se um passo de aceitação/rejeição para gerar da distribuição *a posteriori* de \mathbf{c} . Uma amostra de 25000 observações foi gerada, sendo as últimas 5000 utilizadas para inferência.

Seja μ_{it} a população do i -ésimo município no tempo t . O processo seguinte resume como se obtém uma observação da distribuição *a posteriori* de μ_{it} a partir de π_{it} :

Gerar $a_i^{(l)}$, $b_i^{(l)}$, $c_i^{(l)}$, $\alpha^{(l)}$, $\beta^{(l)}$, $\gamma^{(l)}$, $\tau_a^{-2(l)}$, $\tau_b^{-2(l)}$, $\tau_c^{-2(l)}$, $\eta_0^{(l)}$ e $\eta_1^{(l)}$, $l = 1, \dots, M$, com base nas distribuições condicionais completas, onde M é o número de iterações do *Gibbs sampler*;

Calcular $\pi_{it}^{(l)} = a_i^{(l)} + b_i^{(l)} \exp(c_i^{(l)} t)$;

Calcular $\mu_{it}^{(l)} = \pi_{it}^{(l)} A_i$

onde A_i é a área do município i . Como o objetivo final era inferir sobre as populações municipais em cada ano representada por μ_{it} , incluiu-se um passo adicional no algoritmo *Gibbs sampler* para se amostrar da sua distribuição *a posteriori*: $\mu_{it}^{(l)} = \pi_{it}^{(l)} A_i$. A média *a posteriori* de μ_{it} foi utilizada como estimativa pontual da população municipal.

Um aspecto importante no uso dos métodos MCMC para fazer inferências é garantir que as distribuições empíricas dos parâmetros tenham atingido equilíbrio. Foram utilizados os mesmos critérios apresentados no Exemplo 5.1 da Seção 5.3 para se verificar a convergência dos hiperparâmetros dos dois modelos considerados. Uma análise cuidadosa dos resultados mostrou que não havia razões para se suspeitar da convergência dos hiperparâmetros.

Como já foi dito anteriormente, o menor domínio de divulgação dos dados da PNAD é o nível de região metropolitana. Como forma de validar os resultados encontrados com o modelo (5.7), foram obtidas as estimativas da população da região metropolitana do estado de São Paulo a partir da aplicação desse modelo. O objetivo era compará-las aos resultados oficiais. A região metropolitana é formada pela agregação de um conjunto de municípios para os quais já obtivemos as estimativas dos tamanhos populacionais. Estamos interessados na distribuição *a posteriori* de $\sum_{i=1}^r \mu_{it} | D$, ou seja, $\sum_{i=1}^r \pi_{it} * A_i | D$, que é facilmente obtida, acrescentando-se no algoritmo do *Gibbs sampler* o passo

$$\theta_t^{(l)} = \sum_{i=1}^r \pi_{it}^{(l)} * A_i$$

onde θ_t representa a população da região metropolitana no tempo t , r é o

número de municípios pertencentes à região, e D é o conjunto de toda informação disponível.

Na Figura 5.4, tem-se o gráfico comparativo das estimativas da população da região metropolitana do Estado de São Paulo dadas pelo modelo (5.7) e pelo método de estimação utilizado pelo IBGE. A média *a posteriori* de θ_t é utilizada como estimativa pontual da população da região metropolitana. As linhas cheias representam os limites do intervalo de 95% de credibilidade para a população θ_t , enquanto a linha tracejada é a média *a posteriori*. O símbolo (+) representa a população projetada e o que foi observado nos censos em 1991, 1996 e 2000. Os resultados obtidos pelo modelo proposto são razoavelmente próximos daqueles obtidos pelo método atualmente empregado pelo IBGE. Souza et al. (2007) também mostram que as previsões de população municipais e as fornecidas pelos modelos propostos são razoavelmente próximas. Mais recentemente, Neto et al. (2007) propõem a modelagem conjunta de estimativas diretas, baseadas unicamente no desenho da amostra, e de suas estimativas da variância. Para tratar a evolução temporal, optaram pelo uso de Modelos Dinâmicos de Crescimento Exponencial e, para captar a dependência espacial desta evolução, utilizaram Processos Gaussianos.

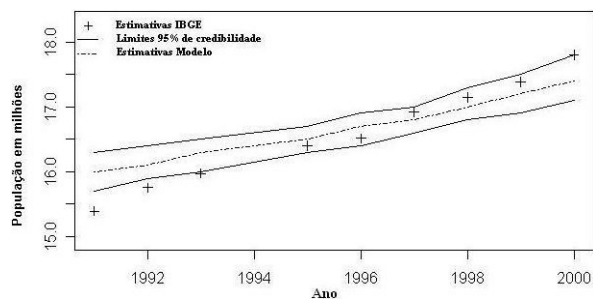


Figura 5.4: Comparação entre as populações estimadas pelo modelo e as estimativas oficiais do IBGE para a região metropolitana do Estado de São Paulo.

Referências Bibliográficas

- Banerjee, S., Carlin, B. P. e Gelfand, A. E. (2004) *Hierarchical Modeling and Analysis of Spatial Data*. New York: Chapman and Hall.
- Basu, D. (1969) Role of sufficiency and likelihood principles in sample survey theory. *Sankiā, Series B*, **31**, 441–454.
- Battese, G. E., Harter, R. M. e Fuller, W. A. (1988) An error component model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, **83**, 28–36.
- Bessag, J. e Kooperberg, C. L. (1995) On conditional and intrinsic autoregressions. *Biometrika*, **82**, 733–746.
- Bianchini, Z. M. e Albieri, S. (2003) Principais aspectos de amostragem das pesquisas domiciliares do ibge - revisão 2002. *Relatório técnico*, Texto para Discussão, 8, IBGE, Rio de Janeiro, RJ.
- Bolfarine, H. e Bussab, W. O. (2005) *Elementos de Amostragem*. São Paulo: Edgard Blücher.
- Booth, J. G. e Hobert, J. P. (1999) Maximum generalized linear mixed model likelihood with an automated Monte Carlo EM algorithm. *Journal of the Royal Statistical Society, Series B*, **61**, 265–285.

- Brackstone, G. J. (1987) *Small Area Data: Policy Issues and Technical Challenges*. New York: Wiley.
- Breslow, N. E. e Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, **88**, 9–25.
- Brooks, S. P. e Gelman, A. (1998) Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434–455.
- Chambers, R. L. e Skinner, C. J. (2003) *Analysis of Survey Data*. Great Britain, Uk: Wiley.
- Clayton, D. e Kaldor, J. (1987) Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**, 671–681.
- Cochran, W. (1977) *Sampling Techniques*. New York: John Wiley and Sons.
- Datta, G. S. e Lahiri, P. (2000) A unified measure of uncertainty of estimated best linear unbiased prediction in small area estimation problems. *Statistica Sinica*, **10**, 613–627.
- Datta, G. S., Lahiri, P. e Maiti, T. (2002) Empirical bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference*, **102**, 83–97.
- Datta, G. S., Lahiri, P., Maiti, T. e Lu, K. L. (1999) Hierarchical bayes estimation of unemployment rates for the u. s. states. *Journal of the American Statistical Association*, **94**, 1074–1082.
- Diggle, P. J., Liang, K. Y. e Zeger, S. L. (1996) *Analysis of Longitudinal Data*. Oxford University Press.

- Duchesne, L. (1987) Proyecciones de poblacion por sexo y edad para areas intermedias y menores. *Relatório técnico*, Centro Latinoamericano de Demografia, Santiago, Chile.
- Elbers, C., Lanjouw, J. O. e Lanjouw, P. (2002) Micro-level estimation of welfare. *Relatório Técnico 2911*, World Bank, Washington DC.
- Ericson, W. A. (1969) Subjective bayesian models in sampling finite population (with discussion). *Journal of the Royal Statistical Society, Series B*, **31**, 195–233.
- Farrell, P. J., MacGibbon, B. e Tomberlin, T. J. (1997) Empirical bayes estimators of small area proportions in multistage designs. *Statistica Sinica*, **7**, 1065–1083.
- Fay, R. E. e Herriot, R. A. (1979) Estimation of income from small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association*, **74**, 269–277.
- Gamerman, D. e Lopes, H. F. (2006) *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. London, UK: Chapman & Hall, 2nd. Edition.
- Gelfand, A. e Ghosh, S. (1998) Model choice: A minimum posterior predictive loss approach. *Biometrika*, **85**, 1–11.
- Gelfand, A. E. e Smith, A. F. M. (1990) Sample-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 972–985.
- Geweke, J. (1996) Evaluating the accuracy of sampler-based approaches to the

- calculation of posterior moments. Em *Bayesian Statistics 4*, 169–193. E. M. Bernardo et al. editors. Oxford University Press.
- Ghosh, M. e Maiti, T. (2004) Small-area estimation based on natural exponential family quadratic variance function models and survey weights. *Biometrika*, **91**, 95–112.
- Ghosh, M. e Nangia, N. (1993) Estimation of median income of four-person families: A bayesian time series approach. *Relatório técnico*, Department of Statistics, University of Florida, Gainesville.
- Ghosh, M., Nangia, N. e Kim, D. (1996) Estimation of median income of four-person families: A bayesian time series approach. *Journal of the American Statistical Association*, **91**, 1423–1431.
- Goldstein, H. (1986) Multilevel mixed linear model analysis using iterative least squares. *Biometrika*, **73**, 43–56.
- (1989) Restricted unbiased iterative generalized least squares estimation. *Biometrika*, **76**, 622–623.
- (1995) *Multilevel Statistical Models*. Edward Arnold, 2nd. Edition.
- Graybill, F. A. (1969) *Matrices with Applications in Statistics*. California: Wadsworth International Group, 2nd. Edition.
- Harville, D. A. e Jeske, D. R. (1992) Mean squared error of estimation on prediction under general linear model. *Journal of the American Statistical Association*, **87**, 724–731.
- IBGE (2002) Estimativas populacionais do brasil, grandes regiões, unidades da federação e municípios - metodologia. *Relatório técnico*, IBGE, Rio de Janeiro, RJ.

- Jiang, J. (1998) Consistent estimators in generalized linear mixed models. *Journal of the American Statistical Association*, **93**, 720–729.
- (2007) *Linear and Generalized Linear Mixed Models and their Applications*. New York: Springer.
- Jiang, J. e Zhang, W. (2001) Robust estimation in generalized linear mixed models. *Biometrika*, **88**, 753–765.
- Kachar, R. N. e Harville, D. A. (1981) Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communication in Statistics Series A*, **10**, 1249–1261.
- (1984) Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, **79**, 853–862.
- Klein, R. e Moura, F. A. S. (1998) Análise dos erros amostrais de indicadores educacionais obtidos através de dados da pesquisa nacional por amostra de domicílios. *Revista Brasileira de Estatística*, **59**, 47–58.
- Lahiri, P. e Maiti, T. (1999) Empirical bayes estimation of relative risks in disease mapping. *Relatório técnico*, Department of Statistics, University of Nebraska, Lincoln, USA.
- Lee, Y. e Nelder, J. A. (1996) Hierarchical generalized linear mixed model (with discussion). *Journal of the Royal Statistical Society, Series B*, **58**, 619–678.
- Lin, X. e Breslow, N. E. (1996) Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association*, **91**, 1007–1016.

- Little, R. J. A. (1982) Models for nonresponse in sample surveys. *Journal of the American Statistical Association*, **77**, 237–250.
- Longford, N. T. (1987) A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, **74**, 817–827.
- MacGibbon, B. e Tomberlim, T. J. (1989) Small area estimation of proportions via empirical bayes techniques. *Survey Methodology*, **15**, 237–252.
- Malec, D., Sedransk, J., Moriarity, C. L. e LeClere, F. B. (1997) Small area inference for binary variables in national health interview survey. *Journal of the American Statistical Association*, **92**, 815–826.
- Marshall, R. J. (1991) Mapping disease and mortality rates using empirical bayes estimators. *Applied Statistics*, **40**, 283–294.
- McCulloch, C. E. (1997) Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, **92**, 162–170.
- Mollié, A. (1996) Bayesian mapping of disease. Em *Markov Chain Monte Carlo in Practice*, 359–379. W. R. Gilks, S. Richardson and D. J. Spiegelhalter, editors. Chapman & Hall.
- Moura, F. A. S. (1994) *Small Area Estimation Using Multilevel Models*. Tese de Doutorado, Social Statistics Department, University of Southampton, Southampton, UK.
- Moura, F. A. S. e Castaneda, D. F. N. (2000) Estimação de Índices de proficiência escolar para pequenas Áreas do município do rio de janeiro via modelos logísticos hierárquicos. *Revista Brasileira de Estatística*, **61**, 35–49.

- Moura, F. A. S. e Holt, D. (1999) Small area estimation using multilevel models. *Survey Methodology*, **21**, 73–80.
- Moura, F. A. S. e Migon, H. S. (2002) Bayesian spatial models for small area estimation of proportions. *Statistical Modeling: an International Journal*, **2**, 183–201.
- Moura, F. A. S., Migon, H. S. e Ferreira, M. A. R. (2000) Small area estimation for binary data via hierarchical models. *Statistics in Transition*, **4**, 665–677.
- Neto, J. H. V., Schmidt, A. M. e Moura, F. A. S. (2007) Modelagem conjunta de média e variância: Uma aplicação ao problema de estimação em pequenos domínios. Em *12a Escola de Séries Temporais*.
- Peixoto, J. L. e Harville, D. A. (1986) Comparison of alternative predictors under the balanced one-way random model. *Journal of the American Statistical Association*, **81**, 431–436.
- Pfeffermann, D. e Burck, L. (1990) Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, **16**, 217–237.
- Prasad, N. G. N. e Rao, J. N. K. (1990) The estimation of the mean squared error of small area estimators. *Journal of the American Statistical Association*, **85**, 163–171.
- Rao, J. (2003) *Small Area Estimation*. New Jersey: Wiley Series in Survey Methodology.
- Rao, J. N. K. e Yu, M. (1994) Small area estimation by combining time series and cross-sectional data. *Canadian Journal of Statistics*, **22**, 511–528.
- Royall, R. M. (1976) The linear least-squares prediction approach to two-stage sampling. *Journal of the American Statistical Association*, **71**, 657–664.

- Schmidt, A. M. e Sansó, B. (2006) *Modelagem Bayesiana da Estrutura de Covariância de Processos Espaciais e Espaço-Temporais*. Minicurso 17o. SINAPE, ABE. Caxambu, MG, Brasil.
- Song, P. X. K., Fan, Y. e Kalbfleisch, J. D. (2003) Maximization by parts in likelihood inference (with discussion). *Journal of the American Statistical Association*, **100**, 1145–1158.
- Souza, D. F. (2004) *Estimação de População em Nível Municipal via Modelos Hierárquicos e Espaciais*. Dissertação de Mestrado, Instituto de Matemática - UFRJ, Rio de Janeiro, RJ.
- Souza, D. F., S., M. F. A. e S., M. H. (2007) Small area prediction via spatial growth models. Em *56th Session of the International Statistical Institute: Contributed Paper*.
- Spiegelhalter, D. J., Thomas, A. e Best, N. G. (2002) *WinBugs Version 1.4 User Manual*. *Relatório técnico*, Cambridge: Medical Research Council Biostatistics.
- You, Y. (1999) *Hierarchical Bayes and Related Methods for Model-Based Small Area Estimation*. Tese de Doutorado, Carleton University, Ottawa, Canada.