Mapeamento Genético
utilizando a teoria do
gráfico da variável adicionada
em modelos lineares mistos

Nubia Esteban Duarte

AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TITULO
DE
DOUTOR EM CIÊNCIAS

Programa: **Estatística**Orientadora: **Profa. Dra. Júlia Maria Pavan Soler**

Durante o desenvolvimento deste trabalho a autora recebeu apoio financeiro da CAPES e CNPq.

São Paulo, Maio de 2012

Mapeamento Genético utilizando a teoria do gráfico da variável adicionada em modelos lineares mistos

Esta tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defessa realizada por Nubia Esteban Duarte em 11/05/2012.

O original encontra-se disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Profa. Júlia Maria Pavan Soler (Orientadora) IME/USP
- Prof. Gilberto Alvarenga de Paula IME/USP
- Prof. Juvêncio Santos Nobre DEST/UFC
- Profa. Mariza de Andrade Division of Biostatistics, Mayo Clinic/MN-USA
- Profa. Suely Ruiz Giolo DEST/UFPR

"O temor do Senhor é o princípio da sabedoria, têm bom entendimento todos os que cumprem os seus preceitos."

Salmo 111:10

"O coração do homem dispõe o seu caminho, mas é o Senhor que dirige seus passos."

Provérbios 16:9

"O anjo do Senhor acampa-se ao redor dos que o temem, e os livra."

Salmo 34:7

À meu pai, minha mãe, minha tia, in memoriam,
à meus irmãos,
à meus tios,
à Eddy, Jorge e Marina.

Agradecimentos

Acima de tudo, agradeço a Deus pela saúde e oportunidades concedidas.

À minha mãe, a pessoa que mas me deu amor, obrigada por me apoiar durante todos estes anos de vida. Você sempre esteve presente nos momentos fáceis e difíceis que já deparei. Mesmo havendo partido para sempre ao finalizar esta etapa da minha vida, sua dedicação, seu extraordinário exemplo de vida e suas orações estarão sempre presentes em minha memoria e serão um incentivo para continuar...

À meus irmãos, meus tios e familiares em geral, agradeço a todos por estarem sempre próximos a minha pessoa, fazerem parte da minha vida e sempre me apoiarem em todo momento apesar da distância.

À minha orientadora, Profa. Dra. Júlia M. Paván Soler, pela oportunidade de ser sua orientanda, pelo apoio, confiança em mim depositada e pela força transmitida ao longo do desenvolvimento deste trabalho. Muito obrigada por fazer parte desta conquista.

Àos professores Gilberto de Paula, Juvêncio Santos Nobre, Suely Giolo e Mariza de Andrade, membros da comissão examinadora, pelas valiosas sugestões no desenvolvimento deste trabalho e por disponibilizarem seu tempo avaliando e dando sugestões para pesquisas futuras.

Aos meus professores do Departamento de Estatística do IME-USP pelo exemplo e pelos valiosos ensinamentos recebidos.

Aos Drs. José Eduardo Krieger e Alexandre C. Pereira do laboratório de Genética e Cardiologia Molecular do InCor/USP pelas sugestões e motivação para dar continuidade ao trabalho desenvolvido.

À Jacqueline pela amizade e confiança que depositou em mim desde que a conhecí, pelas discussões de nossa área comum de pesquisa. A Paulo Tadeu pela amizade e pelas sugestões e bibliografias precisas que recomendou para o encaminhamento do trabalho.

À todos os meus amigos Colombianos, estudantes do IME, pela amizade e ajuda que

sempre me ofereceram, são muitos, mas destacarei a alguns deles, começando por Marina pela magnífica amizade e pela ajuda no desenvolvimento deste trabalho, Sandra, Diana, Natalia, Paola, Mary Luz, Alex Valencia, José Domingo, Mauricio, Paula, Nelfi, Tabarez, Oscar Giraldo, Oscar Ocampo, Angélica, Gladys, Luz Mery, Guillermo, Deissy, Alex Holguin, Mauricio Mazo, Héctor, Faber, Germán, Elizabeth, Alejandro, ..., a todos muito obrigada pela amizade e os bons momentos compartilhados.

À Edwin, Ludy, primo Juan, Nataly, Damaris, Smith, minhas primas, entre muitas outras pessoas da Colômbia que desde lá sempre me acompanharam e torceram por mim.

À todos os meus amigos que conheci em São Paulo, brasileiros e de outros países, de novo nomearei aos meus amigos da turma de mestrado que me brindaram sua amizade incondicional e foi "bom demais" estar ao lado deles: Pedro, Augusto, Mariana, Juliana, Grazieli, Ela Cecilia, Fernando, Fernando Cordeiro, Joan, Iván, Jossivão, Marcelo, entre outros. À Betsabé, Alexandre, Artur, Rafael, Michel, Tiago, Lizandra, Marco, Gustavo, Jalmar, Eliani, Isabel, à Amanda e Mariana, que bom ter compartilhado com vocês muitos momentos no periodo de doutorado!!. À Verónica minha grande amiga por estar sempre presente en momentos que é necesário ter a alguém presente.

Também a Gleiciane, Adriana, Fernando, João, Luis, Rafael, Michel, Ana Maria e ultimamente Tatiana pela amizade e magnífica convivência.

Aos meus amigos da igreja: senhora Emilia, Edwiges, Yamily, Guido, Vania, Paul, Luis, Esther, Heriberto, Porfirio e família, Yury e família entre outros.

À todos meus amigos que assistiram a minha defessa.

Aos funcionários do IME-USP, pela colaboração em todo momento.

À CAPES e CNPq pelo apoio financeiro.

Não é possível citar todos os nomes que merecem agradecimentos, pois além dos já nomeados são muitas pessoas que conheci ao longo dos anos de estudo, saibam que sempre estarei agradecida pela oportunidade de ter lhes conhecido e sempre estarão no meu coração.

Em resumo, foi um grande privilegio ter estudado na USP!!!

À todos, MUITO OBRIGADA.

Resumo

Mapeamento Genético utilizando a teoria do gráfico da variável adicionada em modelos lineares mistos

Atualmente, um dos problemas mais importantes da Genética é a identificação de genes associados com doenças complexas. Um delineamento adequado para esta finalidade corresponde à coleta de dados de famílias e plataformas de marcadores moleculares do tipo SNP (do inglês, Single Nucleotide Polimorphism). Estas plataformas representam pontos de referência estrategicamente dispostos ao longo do genoma dos indivíduos e são de alta dimensão. A análise destes dados traz desafios analíticos como o problema de múltiplos testes e a seleção de variáveis preditoras.

Nesta tese, propõe-se um critério para discriminar as variáveis preditoras genéticas em efeitos devidos ao componente aleatório poligênico e ao componente aleatório residual, sob a estrutura de um modelo linear misto. Também, considerando que o efeito individual das variáveis preditoras é esperado ser pequeno, é sugerido um método para encontrar blocos ou subconjuntos ordenados destas variáveis e estudar o seu efeito simultâneo sobre a variável resposta. Neste contexto, utiliza-se a teoria associada ao gráfico da variável adicionada em modelos lineares mistos. As propostas são validadas por meio de um estudo de simulação, o qual é baseado em estruturas de famílias envolvidas no Projeto "Corações de Baependi" (InCor/USP), cujo objetivo é identificar genes associados a fatores de risco cardiovascular na população brasileira. Para a implementação dos procedimentos, usa-se o programa R e na geração das variáveis preditoras genéticas adota-se o aplicativo SimPed.

Palavras-Chave: Marcadores moleculares SNPs, Gráfico da variável adicionada, Modelo misto poligênico.

Abstract

Genetic mapping using the theory of the added variable plot in the mixed linear models

Recently, one of the most important problems in genetics is the identification of genes associated with complex diseases. A useful design for this proposal corresponds to collect data from extended families and molecular markers obtained from SNPs platforms (Single Nucleotide polymorphism). These platforms represent points of reference strategically placed along the genome of the individuals and are high dimensional. Analysis of these data brings analytical challenges as the problem of multiple testing and selection of predictive variables.

In this thesis, we propose a criterion for discriminating predictors of genetic effects due to random polygenic and the residual components, under the framework of a linear mixed model. By assuming, that the expected individual predictor variables effects is small, we suggested a method to identify ordered subsets of these variables and study their simultaneous effect on the response variable under study by using the theory of the added variable plot under a mixed model framework. The proposed methods are validated through a simulation studies, which are based on family structures from the Project "Baependi Heart Study" (FAPESP Process 2007/58150-7), whose objective is to identify genes associated with cardiovascular risk factors in the Brazilian population. The method was implemented using the R statistical environment and for the simulation of genetic predictors we adopted the SimPed program.

Keywords: Molecular markers SNPs, Added variable plot, Polygenic mixed model.

Índice

adecimentos	V
ımo	vii
tract	viii
a de Tabelas	xi
a de Figuras	xiii
ntrodução	1
1 Motivação	. 8
2 Proposta de tese	. 9
rincipais modelos estatísticos no mapeamento genético	12
1 Alguns Modelos Genéticos	. 13
2.1.1 Dados genéticos: Mapas de marcadores microsatélites	. 13
2.1.2 Dados genéticos: Marcadores microsatélites e refinamento de posições	
genéticas candidatas	. 16
2.1.3 Dados genéticos: Mapas de microsatélites e SNPs	. 18
2.1.4 Dados genéticos: Mapas de SNPs	. 18
Gráfico da variável adicionada em Modelos Lineares Mistos	21
1 Inclusão de efeitos fixos	. 26
2 Gráfico da variável adicionada para efeitos fixos	. 28
3 Decomposição do gráfico da variável adicionada	. 32
4 Índice para discriminar o efeito de uma variável preditora	. 38

	3.5	Índice para discriminar famílias influentes	40
	3.6	Geometria tridimensional do problema da variável adicionada	43
4	Es	tudo do efeito de um bloco de SNPs: Abordagem haplotípica	47
	4.1	Relação entre Genótipos e Haplótipos	48
		4.1.1 Ambiguidades na formação de haplótipos	49
	4.2	Configurações de haplótipos para um indivíduo	50
	4.3	Estudo de efeitos de variáveis haplotípicas	52
5	Es	tudo de simulação	57
	5.1	Geração das variáveis aleatórias	58
	5.2	Cenários de simulação	60
	5.3	Decomposição de efeitos de variáveis preditoras	64
	5.4	Análises no caso haplotípico	71
	5.5	Análises no nível de famílias	79
	5.6	Simulações complementares	82
6	Co	onsiderações finais	89
\mathbf{A}	Ro	otinas	93
	A.1	Código para a geração de marcadores SNPs no aplicativo SimPed $\ .\ .\ .\ .$.	93
		A.1.1 Codificação das saídas obtidas no aplicativo SimPed	95
В	An	nexos do estudo de simulação	99
	B.1	Decomposição do efeito das variáveis preditoras	99
	B.2	Análises no caso haplotípico	100
	В.3	Análises no nível de famílias	108
\mathbf{C}	Ge	enética-Alguns conceitos fundamentais	L11
	C.1	Marcadores Moleculares	111
	C.2	Desequilíbrio de ligação	114
	C.3	Equilíbrio de Hardy- Weinberg	116
\mathbf{R}	efer	ências Bibliográficas	118

Lista de Tabelas

1.1	Banco de dados
1.2	Herdabilidades de variáveis amostradas em famílias brasileiras
11	C. G. G. L. L. L. CND.
4.1	Genótipos e Haplótipos segundo o número de SNPs
4.2	Classes genotípicas para dois SNPs, SNP A e SNP B
4.3	Configurações haplotípicas para blocos de 2 SNPs
4.4	Configurações haplotípicas para blocos de 3 SNPs
5.1	Tamanho das famílias usadas nos cenários de simulação
5.2	Cenários de simulação
5.3	Resultados (em média) para variáveis resposta geradas com variância total
	$\sigma_y^2 = 7$ e covariável o SNP 7
5.4	Resultados (em média) para variáveis resposta geradas com variância total
	$\sigma_y^2 = 7$ e covariável v_3
5.5	Resultados quando os \mathbf{Y}_{kl} são gerados com herdabilidade de 8% 73
5.6	Resultados quando os \mathbf{Y}_{kl} são gerados com herdabilidade de 32% 75
5.7	Resultados quando os \mathbf{Y}_{kl} são gerados com herdabilidade de 70% 76
5.8	Efeitos de haplótipo
5.9	Estimativas para os cenários das variáveis $\mathbf{Y}_1,\mathbf{Y}_2,\mathbf{Y}_3$ e \mathbf{Y}_4,\ldots 86
B.1	Resultados (em média) para variáveis resposta geradas com variância total
	$\sigma_y^2=2$ e covariável o SNP 7
B.2	Resultados (em média) para variáveis resposta geradas com variância total
	$\sigma_y^2 = 15$ e covariável o SNP 7
В.3	Resultados (em média) para variáveis resposta geradas com variância total
	$\sigma_y^2 = 50$ e covariável o SNP 7

B.4	Resultados (em média) para variáveis resposta geradas com variância total
	$\sigma_y^2 = 100$ e covariável o SNP 7
B.5	Resultados quando os Y_{kl} são gerados com herdabilidade de 8% 105
B.6	Resultados quando os Y_{kl} são gerados com herdabilidade de 32% 106
B.7	Resultados quando os Y_{kl} são gerados com herdabilidade de 70% 107
C.1	Marcadores Microsatélites versus SNPs
C.2	Relação entre probabilidades de alelos e genótipos

Lista de Figuras

1.1	Padrões possíveis de regulação genética	2
3.1	Partição do efeito fixo em coeficientes devidos a efeitos aleatórios e residu-	
	ais	37
3.2	Decomposição do efeito aleatório por unidade amostral	42
3.3	Geometria da estatística F	44
3.4	Geometria do teste F e decomposição dos resíduos	46
4.1	Configurações Cis e Trans.	50
4.2	Haplótipos para 2 SNPs	51
4.3	Configurações haplotípicas para 1 indivíduo quando considerados 2 SNPs.	52
4.4	Haplótipos gerados a partir de 3 SNPs	53
5.1	Heredograma de uma família com 4 gerações.	57
5.2	Variáves preditoras para o estudo de simulação	60
5.3	Valores de \mathbf{Y}_1 para 51 famílias as quais estão ordenadas de acordo com a	
	respectiva média. A linha horizontal representa a média geral	61
5.4	Valores de \mathbf{Y}_2 para 51 famílias as quas estão ordenadas de acordo com a	
	respectiva média. A linha horizontal representa a média geral	61
5.5	Partição de $\hat{\beta}_2$ para variáveis resposta geradas com $\sigma_y^2=7.$	68
5.6	Indices d_g para o conjunto de variáveis preditoras	77
5.7	Indices d_g para o conjunto de variáveis preditoras	78
5.8	Representação da contribuição de cada família no componente poligênico.	82
5.9	Heredograma correspondente à família 6	83
5.10	Heredograma correspondente à família 7	83
5.11	Famílias caudais quando considerados os SNPs 6, 7 e 8 como um bloco	84

5.12	Heredograma correspondente à família 47
5.13	Indices d_g para as simulações complementares
5.14	Indices d_g para 300 SNPs
B.1	Representação do índice I_f de cada família no componente poligênico 108
B.2	Heredograma correspondente à família 6
B.3	Famílias caudais quando considerados os SNPs 6, 7 e 8 como um bloco 109
B.4	Heredograma correspondente à família 44
C.1	Representação de um mapa de marcadores moleculares
C.2	Ilustração do marcador molecular conhecido como SNP
C.3	Composição dos alelos em dois locos genéticos
C4	Ilustração do equilíbrio de Hardy-Weinberg 117

Capítulo 1

Introdução

O acelerado crescimento da área da Genômica na geração de dados de alta dimensão tem trazido inúmeros desafios analíticos aos estatísticos (Ziegler et al., 2008). Atualmente, um dos problemas mais importantes da Genética é a identificação de genes associados com doenças complexas (Altshuler et al., 2008). A maioria das doenças de interesse em saúde pública, tais como hipertensão, diabetes, é decorrente de um complicado mecanismo de regulação envolvendo componentes genéticos, ambientais e suas possíveis interações, sendo necessárias metodologias estatísticas específicas para o entendimento da complexidade dos dados envolvidos nas análises.

De maneira geral, é necessário estabelecer a relação

$$\mathbf{Y} = \mathbf{G} + \mathbf{E},\tag{1.1}$$

em que \mathbf{Y} é uma variável resposta (em termos genéticos denominada traço ou fenótipo) que representa medições de características dos indivíduos, tais como glicose, pressão arterial, colesterol entre outras; \mathbf{G} é o efeito genético exercido sobre \mathbf{Y} e \mathbf{E} é o componente de erro, envolvendo efeitos ambientais entre outros.

A Figura 1.1 apresenta padrões que explicam a variação de fenótipos em função de variáveis genéticas e ambientais, tal como foi indicado na expressão (1.1). Observa-se, em primeiro lugar, um modelo no qual apenas um gene controla um traço de interesse (flecha branca). Estes tipos de modelos são conhecidos na literatura como **modelos Mendelianos** de regulação monogênica. Em seguida, tem-se o caso em que um único

gene regula ou influencia mais de um fenótipo (flecha escura); esse fenômeno é conhecido como pleiotropia (Jiang & Zeng, 1995). Finalmente, observa-se o caso em que dois genes (ou mais) controlam o fenótipo, este modelo é denominado de regulação multilocos e, no caso em que há interação entre os genes tem-se epistasia (Cordell, 2002; Moore, 2005). O modelo de epistasia trata da situação na qual diferentes locos¹ estão envolvidos na regulação de um mesmo fenótipo, em que um gene de um determinado loco interfere na manifestação dos genes de outro loco, caracterizando um efeito de interação ou associação entre eles. No processo de regulação celular das características dos organismos é esperado que esteja envolvida uma complexa rede funcional em que todos estes modelos (multilocos, pleiotrópicos e epistáticos) estejam presentes (Balding et al., 2003).

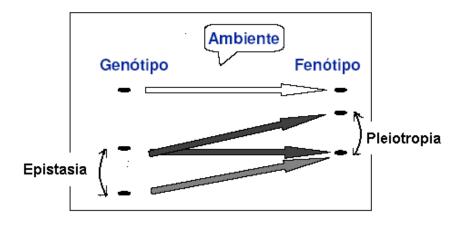


Figura 1.1 Padrões possíveis de regulação genética.

O procedimento que envolve amostragem de uma população alvo de estudo, identificação e localização de regiões cromossômicas do genoma bem como de estimar os seus efeitos sobre a variável resposta em estudo \mathbf{Y} é conhecido como **mapeamento genético**.

Em geral, existem dois tipos de populações envolvidas no mapeamento genético, a saber, populações experimentais cujos cruzamentos são controlados, comuns em experimentos

¹ Cada gene ocupa uma região estabelecida no cromossomo, tradicionalmente chamada **loco** (Farah, 1997).

envolvendo manejo de plantas e animais de laboratório (como ratos, por exemplo). Um segundo tipo refere-se a populações naturais, caracterizadas pelos cruzamentos aleatórios que nela ocorrem, a população humana é um exemplo de população natural. Em relação à amostragem do genoma, a base fundamental consiste em utilizar a informação observável de pontos de referência do genoma e, a partir disso, predizer genes e seus efeitos associados à variável resposta. Os pontos de referência são chamados **marcadores moleculares** cuja posição é conhecida e a constituição genotípica do indivíduo pode ser identificada; são estrategicamente dispostos ao longo do genoma e devidamente codificados como variáveis preditoras. A hipótese geral, empregada para a localização de genes, é que algum marcador estará próximo e, portanto, suficientemente ligado (do inglês *linkage*) ao gene de interesse de tal forma que os dados do marcador oferecem informação sobre o gene (Zeng, 1994; Lynch & Walsh, 1998).

Referente aos marcadores moleculares, as plataformas de Marcadores microsatélites. conhecidos também como SSR (do inglês, Simple Sequence Repeats), são pouco densas, seus marcadores são altamente polimórficos (muitas classes genotípicas) e, em geral, definem grandes regiões cromossômicas. Já as plataformas de Marcadores SNP (do inglês, Single Nucleotide Polimorphism) são altamente densas (cerca de um milhão de marcadores no caso da Affymetrics 6.0), seus marcadores são dialélicos (três classes genotípicas AA, Aa e aa) e correspondem a variações comuns de uma única base do genoma. Embora, mais de 99% da sequência do genoma humano seja idêntico para toda a população, as pequenas variações na sequência do DNA podem ter um impacto importante sobre a forma como os seres humanos respondem a diferentes condições, isto dá aos SNPs um grande valor para a investigação biomédica. Assim, atualmente, o entendimento da viabilidade de plataformas de SNPs para o mapeamento de genes é de grande interesse para os diferentes grupos de pesquisa devido ao acelerado crescimento e disponibilidade de mapas de marcadores deste tipo (Almasy & Blangero, 2000; Altshuler et al., 2008). No entanto, esta tarefa não é trivial devido ao grande número de marcadores moleculares (SNPs) que em um modelo estatístico, representam um problema de análise em um espaço de alta dimensão.

Associados a plataformas de marcadores moleculares, vários delineamentos amostrais

têm sido considerados no mapeamento de genes, por exemplo indivíduos não relacionados (Ziegler et al., 2008; Allen & Satten, 2009), pares de irmãos de diferentes famílias (Haseman & Elston, 1972), dados de famílias (Amos & Elston, 1989; Blangero et al., 2000; Schork, 1993; Blangero, 1993; Amos, 1994; Blangero & Almasy, 1997; Comuzzie et al., 1997; Aulchenko et al., 2007, Amos et al., 2008 entre outros) e trios (pai, mãe, filho) (Spielman et al., 1993; Spielman & Ewens, 1998).

Apesar do sucesso que alguns estudos têm trazido na identificação de genes (Arya et al., 2003; Aulchenko et al., 2007; Amos et al., 2008; Walsh, 2009) ainda existem na literatura muitos desafios analíticos a serem pesquisados. Neste contexto, um dos problemas é o de múltiplos testes em que se o mapa consiste de um milhão de SNPs, o mesmo número de modelos precisariam ser ajustados. Algumas estratégias têm sido propostas para correção do nível descritivo considerando os múltiplos testes (Cordell & Clayton, 2005; Skol et al., 2006). Um outro problema envolve a decisão de como modelar o efeito do SNP, como fixo ou aleatório. Também é conhecido da literatura que o efeito individual de marcadores SNP é pequeno (Spielman et al., 1993; Horvath & Baur, 2000; Yang et al., 2011) e, sozinho não gera covariância entre indivíduos como é o caso dos marcadores microsatélites (várias classes genotípicas) em que o efeito do marcador é considerado como aleatório e usado para modelar a estrutura de covariância entre indivíduos de uma mesma família (Amos, 1994; Almasy & Blangero, 1998; de Andrade et al., 1999). Uma visão geral dos avanços e desafios na área da Genética, em que alguns dos problemas existentes poderiam ser abordados com ferramentas estatísticas, são apresentados em Walsh (2009).

Focando a atenção em estudos baseados na utilização de delineamentos com famílias e plataformas de marcadores moleculares SNPs, a estrutura dos dados é apresentada na Tabela 1.1. Nesta tabela, as colunas 1 até 4 representam a estrutura familiar (família, indivíduo, pai, mãe), esta estrutura é utilizada para calcular a matriz de parentesco entre os indivíduos; a coluna 5 representa o sexo dos indivíduos; Y representa uma variável resposta, que pode ser, por exemplo, pressão arterial. Em todas as outras colunas apresentam-se variáveis com a informação genética, neste caso o genótipo de locos de SNP.

A informação genética está representada pela seguinte codificação:

$$SNP_{j} = \begin{cases} 2, & \text{se o indivíduo \'e homozigoto } AA, \\ 1, & \text{se o indivíduo \'e heterozigoto } Aa, \\ 0, & \text{se o indivíduo \'e homozigoto } aa, \end{cases}$$
 (1.2)

j=1,2,...K, em que K, em geral é da ordem de 1 milhão. Nota-se que o marcador SNP é sempre dialélico (ocorrendo em alelos A ou a, por exemplo). Em geral, para cada loco, o alelo A é aquele que ocorre em menor frequência. Assim, os SNPs representam o espaço das variáveis preditoras, que são categóricas, seguem uma ordenação e são obtidas de uma posição fixa do genoma humano.

Tabela 1.1 Banco de dados.

FAMID	ID	FA	MO	SEX	Y	SNP1	SNP2	SNP3	 SNPK
						rs103395	rs176588	rs131963	rs645804
1	4101	4501	4502	2	118,254	1	1	1	1
1	4201	0	0	1	122,144	0	0	2	1
1	4301	4201	4101	2	121,663	0	1	2	1
1	4302	4201	4101	2	124,024	0	1	2	1
1	4303	4201	4101	1	117,554	1	0	1	0
2	5101	5924	47604	2	113,150	2	0	2	1
2	5201	0	0	1	124,004	1	0	0	1
2	5302	5201	5101	2	115,699	2	0	1	1
2	5601	5924	47604	1	119,123	2	0	1	1
2	5603	5924	47604	1	124,375	2	0	2	1
2	5604	5924	47604	2	121,247	1	0	2	1
2	5704	5909	5604	1	121,039	0	0	1	2
	:						:		
51	117903	0	0	2	121,989	0	0	0	1
51	117904	117925	117924	2	122,049	2	0	1	0
51	117905	117902	117901	1	121,490	1	0	1	1
51	117906	117902	117901	1	123,933	0	0	0	2

Diferentes fases experimentais caracterizam a análise de dados genéticos oriundos de delineamentos com famílias. Inicialmente há interesse em avaliar se a variável resposta **Y**, tem um componente de regulação genético significante, o que é realizado mediante o estudo do coeficiente de correlação intraclasse, o qual em termos genéticos é amplamente

conhecido como coeficiente de herdabilidade ou simplesmente **herdabilidade**. Este coeficiente representa a proporção relativa da variância total de **Y** que é devido a estrutura familiar (variância entre famílias) e o seu valor varia entre 0 e 1. Quando o seu valor está próximo de zero, isto indica que as famílias são muito homogêneas entre si e que o valor da variável resposta de qualquer indivíduo independe de sua família; quando está próximo de 1, a maior parte da variabilidade na variável resposta é devido ao efeito de família. Verificar se a variável resposta em estudo, **Y**, tem um componente genético relevante não é trivial e corresponde à primeira etapa em uma investigação sobre o papel dos genes na regulação de uma variável resposta.

Em referência ao mapeamento de genes, com delineamentos de família e plataformas de SNPs, uma alternativa é proposta em Aulchenko et al. (2007). Os autores propõem uma análise multiestágios em que o modelo misto é estendido para incluir o efeito do SNP (variável preditora) como fixo. A análise proposta, divide-se em três passos: No primeiro passo, os dados são analisados sob o modelo de componentes de variância (sem incluir o efeito do SNP) para obter os resíduos condicionais; no segundo passo, os resíduos obtidos no passo 1 são usados como variável dependente em um modelo de regressão simples, independentemente, para cada SNP. Esta etapa é usada como um filtro já que são selecionados os SNPs significantes (adotando-se um nível corrigido para múltiplos testes). No passo 3, os SNPs selecionados no passo 2 são analisados como efeitos fixos, independentemente, por meio do ajuste de um modelo misto no intuito de encontrar SNPs com possíveis efeitos significantes que estejam regulando a variável resposta. Mediante esta proposta de análise, os SNPs discriminados na etapa final, são SNPs comuns à população geral mas, ao serem consideradas as plataformas de SNPs, é relevante não somente discriminar genes que sejam comuns a toda a população mas, também, discriminar aqueles que afetam algumas famílias em particular.

Também, sendo que o efeito individual de marcadores SNP é pequeno (Spielman et al., 1993; Horvath & Baur, 2000) e sozinho não gera covariância entre indivíduos, uma possível estratégia de análise dos dados é mensurar o efeito de regiões cromossômicas por

meio de conjuntos ordenados de SNPs ou seja blocos de SNPs² e não de SNPs individuais. Esta é uma proposta na literatura que tem sido abordada para análises com indivíduos provenientes de estudos observacionais caso-controle mas não no contexto de dados de família (Akey et al., 2001).

Partindo do fato que a variável resposta Y tenha um componente genético relevante e dado que o espaço das variáveis preditoras é de alta dimensão, neste trabalho propõe-se um critério de seleção de SNPs podendo serem classificados em duas categorias gerais, aqueles SNPs que tem efeito no componente aleatório residual (afetam a população geral) e aqueles SNPs que têm efeito no componente aleatório poligênico (afetam famílias específicas). Nesse intuito, em primerio lugar será obtida uma partição do coeficiente angular que representa o efeito total de um SNP em dois coeficientes angulares, um deles devido a efeitos residuais e o outro devido a efeitos genéticos. Por meio desta partição, é possível selecionar aqueles SNPs que explicam uma alta variabilidade genética, podendo-se obter blocos de SNPs adjacentes para serem estudados em conjunto e mensurar efeitos de haplótipos.

O desenvolvimento teórico para as propostas anteriores, é baseado na teoria associada ao grafico da variável adicionada no contexto de modelos lineares mistos. Esta técnica gráfica é sugerida para buscar evidências para a inclusão de uma variável preditora particular num modelo estatístico. Essa ferramenta visual foi proposta por Mosteller & Tuckey (1977) e subsequentemente discutida por vários autores, entre eles Cook & Weisberg (1982) e Atkinson (1985) por exemplo. Nos trabalhos de Belsley et al. (1980) e Drapper & Smith (1998) têm-se discutido o uso de gráficos de variáveis adicionadas em modelos de regressão linear ordinária. Wang (1985) estendeu o procedimento de diagnóstico para Modelos Lineares Generalizados. Hodges (1998) sugeriu o uso de gráficos da variável adicionada em análises de Modelos Lineares Mistos. Posteriormente, Hilden-Minton (1995) escreve o efeito da variável adicionada em função do resíduo do modelo e sugere uma

² Os blocos são conjuntos de SNPs adjacentes, ordenados de acordo com sua posição no genoma. Contêm duas sequências de informações genéticas, uma que provém do pai e a outra da mãe, tais sequências são denominadas **Haplótipos**.

1.1 Motivação

decomposição dos componentes de variância do modelo linear misto. Esta decomposição é estendida para o desenvolvimento teórico das propostas do presente trabalho.

1.1 Motivação

A motivação para o desenvolvimento do presente trabalho, decorre da necessidade na literatura de análises estatísticas específicas para dados genômicos de alta dimensão, particularmente quando utilizados delineamentos de famílias e plataformas de SNPs. No Brasil, poucas iniciativas têm sido feitas nessa direção, o fato de a população brasileira, ser considerada uma das mais heterogêneas do mundo (Pimenta et al., 2006) justifica ainda mais o interesse em caracterizar o padrão de regulação gênica de muitas doenças nesta população e é um desafio para diferentes grupos de pesquisa. Específicamente, os dados genômicos do Projeto "Corações de Baependi" (Processo Fapesp 2007/58150-7) do Laboratório de Genética e Cardiologia Molecular do Instituto do Coração de São Paulo (InCor/USP) fornecem uma excelente oportunidade para caracterizar as contribuições genéticas a fatores de risco cardiovascular nesta população.

Em referência à coleta dos dados do Projeto, entre dezembro de 2005 e janeiro de 2006, um total de 81 famílias (aproximadamente 1.700 indivíduos) foram amostradas no município de Baependi, MG. O tamanho dessas famílias varia de 3 a 156 membros com média de 21 indivíduos por família; a idade média dos indivíduos é de 44 anos, variando de 18 a 100 anos. A distribuição da amostra relativamente ao gênero é de 57% de mulheres (43% de homens).

Nessa primeira fase do Projeto, foram coletadas variáveis respostas dos indivíduos como glicose, colesterol total, fração de lipoproteínas e triglicérides, entre outras. Ainda, o DNA foi extraído das amostras de sangue de cada indivíduo e armazenado apropriadamente para posterior genotipagem.

Considerando os trabalhos de Oliveira et al. (2008), Giolo et al. (2009), Horimoto et al. (2011a) e Horimoto et al. (2011b), os resultados indicam efeitos poligênicos significantes em várias variáveis respostas avaliadas, o que justifica a busca por determinantes

genéticos específicos que expliquem tais efeitos. Algumas das variáveis resposta do estudo encontram-se na Tabela 1.2, assim como estimativas de herdabilidades e de componentes de variância (poligênico σ_g^2 e residual σ_e^2) obtidos do ajuste de modelos lineares mistos específicos. Os resultados justificam a pesquisa pelos determinantes genéticos que as regulam. Deste modo, nesta segunda fase do Projeto, a partir das amostras de DNA, a genotipagem dos indivíduos via a plataforma de SNPs da Affymetrics 6.0, está sendo realizada. Também, as mesmas famílias estão sendo revisitadas, além de algumas famílias adicionais, no intuito de coletar novos fenótipos dando lugar a estudos longitudinais.

Tabela 1.2 Herdabilidades de variáveis amostradas em famílias brasileiras.

Y	Covariáveis	Herdabilidade	σ_g^2	σ_e^2	
$\ln(\mathrm{SBP})$	Nenhuma	0.1532	0.1617	0.8941	
	Sexo	0.1563	0.1602	0.8649	
	Sexo, Idade	0,2582	0,2035	$0,\!5847$	
	Sexo, Idade, Med	0,1965	0,1320	0,5399	
ln(DBP)	Nenhuma	0,1492	0,1478	0,8428	
	Sexo	0,1489	$0,\!1472$	0,8414	
	Sexo, Idade	0,2052	$0,\!1867$	0,7231	
	Sexo, Idade, Med	0,1590	0,1371	0,7254	
ln(GLU)	Nenhuma	0,3264	0,2905	0,5992	
	Sexo	0,3264	0,2905	0,5992	
	Sexo, Idade	0,3397	$0,\!2794$	0,5430	
	Sexo, Idade, Med	0,3184	$0,\!2501$	0,5353	
$\ln(\text{LDL})$	Nenhuma	0,2693	0,2589	0,7024	
	Sexo	0,2682	$0,\!2560$	0,6984	
	Sexo, Idade	0,2638	0,2421	0,6758	
	Sexo, Idade, Med	0,2694	0,2480	0,6726	
ln(HDL)	Nenhuma	0,3132	0,4126	0,9046	
	Sexo	0,3123	$0,\!4062$	0,8945	
	Sexo, Idade	0,3217	$0,\!4174$	0,8800	
	Sexo, Idade, Med	0,3253	0,4211	0,8733	

1.2 Proposta de tese

Diante das limitações dos modelos estatísticos existentes quando as plataformas de

marcadores genéticos disponíveis são do tipo SNPs em dados de família, a seguir descrevese a proposta de pesquisa desenvolvida nesta tese:

- Propor um critério de seleção de variáveis preditoras em um modelo linear misto.
 O espaço destas variáveis é de altísssima dimensão. Essas variáveis preditoras são categorizadas (classificadas como 0, 1, 2 de acordo com a classe genotípica aa, Aa e AA, respectivamente), estão ordenadas segundo uma posição específica (no genoma) e ao serem testadas individualmente o efeito pode não ser significante.
- Estudo do efeito conjunto de um bloco de variáveis preditoras SNPs (abordagem haplotípica).

Como declarado, dado que os marcadores SNPs individualmente não modelam covariância entre indivíduos relacionados e o efeito individual é pequeno, uma proposta é "reconstruir" uma região informativa com um bloco de SNPs ordenados, denominados de **haplótipo**, que modele covariância entre indivíduos da mesma família. Uma vez encontradas essas regiões informativas, os SNPs envolvidos são codificados de forma adequada, segundo o interesse do estudo, para encontrar efeito conjunto sobre a variável resposta sob análise.

Dado que o interesse principal é conhecer quais variáveis preditoras contribuem em maior grau no efeito genético para se estabelecer a relação geral descrita em (1.1), a teoria associada ao gráfico da variável adicionada é considerada no intuito de selecionar variáveis preditoras (SNPs) que tenham uma contribuição considerável no componente aleatório poligênico. Além disso, com esta ferramenta estatística será possível encontrar blocos de variáveis preditoras (SNPs ordenados) que se destaquem do conjunto geral.

A implementação das propostas é realizada no programa R, que é um sofware livre, encontrado em http://cran.r-project.org. Neste programa utiliza-se a biblioteca kinship disponível em cran.r-project.org/package=kinship, a qual é específica para modelos lineares mistos no contexto de dados genéticos. Na simulação das variáveis preditoras genéticas é utilizado o programa SimPed (Leal et al., 2005). Mediante este programa é possível a simulação de marcadores moleculares com dois ou mais alelos e para diversas

estruturas familiares.

O trabalho está organizado da seguinte forma:

-Capítulo 2: É apresentada uma revisão das metodologias estatísticas para mapeamento de genes via dados de famílias e plataformas de marcadores moleculares. São discutidas as propostas de análise estatística que já existem na literatura, apresentando suas vantagens e desvantagens segundo o tipo de marcador utilizado.

-Capítulo 3: Apresentação da teoria geral de modelos lineares mistos dando ênfase à análise de sensibilidade no sentido da inclusão de efeitos fixos, gráfico da variável adicionada e sua decomposição em componentes associadas à parte aleatória residual e aleatória poligênica.

-Capítulo 4: Apresentação da teoria referente ao estudo de SNPs via configurações haplotípicas. São sugeridas algumas alternativas para a codificação dos blocos de SNPs e as possíveis abordagens de serem estudadas no intuito de encontrar o seu efeito sobre a variável resposta.

-Capítulo 5: Apresenta-se um estudo de simulação no intuito de ilustrar o comportamento das estatísticas envolvidas no estudo.

-Capítulo 6: Apresenta-se uma conclusão com os principais resultados e contribuições desta tese assim como algumas sugestões para pesquisas futuras.

Capítulo 2

Principais modelos estatísticos no mapeamento genético

A historia da Estatística com aplicações na área de Genética é bastante antiga. O artigo de Fisher (1918) foi seminal em teoria estatística de variáveis quantitativas e deu origem a muitos trabalhos futuros, assim como o livro publicado em 1925 (Fisher, 1925) trouxe contribuições à análise de componentes de variância que serviu de base para muitos estudos posteriores em Genética Quantitativa. Penrose (1938) foi o primeiro a desenvolver um método para análise de ligação genética em variáveis quantitativas para populações humanas.

Considerando as correlações geradas por indivíduos vindos da mesma família, Elston & Stewart (1971) e Lange et al. (1976) utilizaram o método de máxima verossimilhança para modelar dados agrupados em famílias as quais são consideradas independentes. Haseman & Elston (1972), assumindo o Equilíbrio de Hardy-Weinberg (independência de alelos de um mesmo loco) e de ligação (independência entre alelos de locos diferentes), aplicaram o modelo de regressão para o qual somente dados de pares de irmãos de diferentes famílias são considerados; o método utiliza como variáveis resposta a diferença entre fenótipos de pares de irmãos e como variável preditora a proporção de alelos IBD (idênticos por descendência) compartilhados entre os irmãos. Outros autores, tais como Amos & Elston (1989), Schork (1993); Blangero & Almasy (1997), estenderam o modelo de regressão de Haseman-Elston possibilitando a análise para todos os membros da família e no intuito de utilizar estruturas familiares mais complexas. Isto fornece covariâncias mais informativas podendo-se obter estimativas mais precisas nas respectivas análises. Amos (1994) propôs

um modelo de componentes de variância com efeitos mistos, possibilitando a análise dos efeitos de covariáveis e aumentado o poder dos testes resultantes.

Outros métodos foram desenvolvidos para se estudar famílias estendidas de tamanho e complexidade arbitrários (Comuzzie et al., 1997) e permitir análises que incluem interações entre genes e ambiente (Blangero, 1993), pleiotropia (Almasy et al., 1997; Kraft & de Andrade, 2003) bem como análises multivariadas (Schork, 1993; Almasy et al., 1997; Blangero & Almasy, 1997; Williams et al., 1997; de Andrade et al., 1997; Olswold & de Andrade, 2003).

2.1 Alguns Modelos Genéticos

Nesta seção será apresentada uma breve revisão sobre os modelos utilizados no mapeamento de genes em dados de família. Serão consideradas as diferentes abordagens respeitando-se o tipo de dados genotípicos coletados, os quais podem envolver somente marcadores microsatélites, ambos os marcadores, microsatélites e SNPs, ou somente marcadores SNPs.

2.1.1 Dados genéticos: Mapas de marcadores microsatélites

• Modelo Poligênico

Este modelo não incorpora a informação de marcadores moleculares, sendo utilizada a estrutura familiar para modelar o efeito genético.

Sob os mesmos pressupostos de um modelo misto linear (Verbeke & Molenberghs, 2000), para finalidade de mapeamento de genes, descreve-se o modelo poligênico (Amos, 1994, Almasy & Blangero, 1998):

$$\mathbf{Y}_f = \mathbf{X}_f \boldsymbol{\beta} + \mathbf{g}_f + \mathbf{e}_f, \ f = 1, ..., F,$$
 (2.1)

em que \mathbf{Y}_f representa um vetor $(n_f \times 1)$ de respostas da f-ésima família, $\boldsymbol{\beta}$ é um vetor $(p \times 1)$ de parâmetros (efeitos fixos), \mathbf{X}_f representa a matriz de planejamento $(n_f \times p)$ conhecida e de posto completo, \mathbf{g}_f indica o vetor de efeitos aleatórios poligênicos os

quais surgem de efeitos não mensuráveis de muitos genes que induzem à estrutura de correlação entre indivíduos parentes, em outras palavras, é o resumo da informação de muitos genes, \mathbf{e}_f representa o vetor de erros aleatórios. Note que em (2.1) a matriz de planejamento ou incidência associada aos efeitos aleatórios poligênicos é a identidade. Os efeitos aleatórios, \mathbf{g}_f e \mathbf{e}_f , são assumidos não correlacionados, com média zero e variância σ_g^2 e σ_e^2 , respectivamente.

Sob essa formulação as covariâncias entre as variáveis resposta para os indivíduos i e i' das famílias f e f' é dada por:

$$Cov(y_{if}, y_{i'f'}) = \begin{cases} \sigma_g^2 + \sigma_e^2, \text{ para } i = i', \\ 2\phi_{ii'}\sigma_g^2, \text{ para } i \neq i' \text{ e } f = f' \text{ (relacionados)}, \\ 0 \text{ para } i \neq i' \text{ e } f \neq f' \text{ (não-relacionados)}. \end{cases}$$
(2.2)

O parâmetro $2\phi_{ii'}$ é definido como o coeficiente de relacionamento entre os indivíduos i e i' sendo dado por $(\frac{1}{2})^r$, em que r representa o grau de relacionamento. Deste modo a variância de $\mathbf Y$ denotada por σ_y^2 é a soma $\sigma_y^2 = \sigma_g^2 + \sigma_e^2$.

Considerando todas as famílias, a matriz de covariâncias para o modelo poligênico dado em (2.1) pode ser definida como:

$$\mathbf{V} = 2\Phi\sigma_a^2 + \mathbf{I}\sigma_e^2,\tag{2.3}$$

com 2Φ uma matriz bloco diagonal (sob independência entre famílias) com elementos $2\phi_{ii'}$ correspondendo aos dados de cada família; **I** a matriz identidade de ordem $n \times n$. A estrutura familiar codificada como, por exemplo, aquela descrita nas primeiras 4 colunas da Tabela 1.1 é usada para calcular o coeficiente de relacionamento.

A herdabilidade, ou coeficiente de correlação intraclasse, é um conceito importante em Genética sendo definida como a proporção da variância total que é devida a componentes genéticos. Sob o modelo (2.1), a herdabilidade é dada por:

$$h_g^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}. (2.4)$$

Se h_g^2 for pequena pode-se inferir que a variável resposta não é regulada predominantemente por fatores genéticos. Alternativamente, pode-se definir herdabilidade residual

como:

$$h_e^2 = \frac{\sigma_e^2}{\sigma_g^2 + \sigma_e^2}. (2.5)$$

Estimadores dos componentes de variância σ_g^2 e σ_e^2 são obtidos pelos métodos de máxima verossimilhança restrita (Searle et al., 1992; Almasy & Blangero, 1998; de Andrade et al., 1999; Demidenko, 2004).

No modelo (2.1) o interesse está na hipótese $H_0: \sigma_g^2 = 0$ vs $H_1: \sigma_g^2 > 0$, o que corresponde a testar se o efeito poligênico é significante. Sob H_0 tem-se a estatística da razão de verossimilhanças generalizada que é assintoticamente distribuída como uma mistura (1/2):(1/2) de χ_1^2 e χ_0^2 , respectivamente (Self & Liang, 1987).

Modelo Oligogênico

Se no modelo (2.1) o coeficiente poligênico for significante de maneira geral, (herdabilidade poligênica significante) para a variável resposta \mathbf{Y} , justifica-se o interesse na identificação dos específicos genes que a regulam.

No modelo oligogênico, além do efeito poligênico (\mathbf{g}_f), assume-se que um gene de efeito maior (oligogene) influencie a variável resposta \mathbf{Y} . Este gene de "grande" efeito é amplamente conhecido na literatura como QTL (do inglês, Quantitative Trait Locus). Para se estudar o efeito de específicos QTLs sobre a resposta \mathbf{Y} , usa-se a informação de dados de marcadores moleculares do tipo microsatélites para se estimar a proporção de alelos compartilhados no QTL por dois indivíduos relacionados; esta proporção de alelos é conhecida como IBD (idêntico por descendência) e assume os valores 0, 1/2 e 1 se os dois indivíduos têm genótipos conhecidos e compartilham 0, 1 ou 2 alelos idênticos no marcador em consideração (Cockerham, 1971; Amos et al., 1990; Almasy & Blangero, 1998). Mais especificamente para indivíduos i e i' relacionados tem-se a variável $\pi_{ii'}$ definida como:

$$\pi_{ii'} = \begin{cases} 1, & \text{se } i = i', \\ 1, & \text{se } i \text{ e } i' \text{ compartilham 2 alelos}, \\ 1/2, & \text{se } i \text{ e } i' \text{ compartilham 1 alelo,} \\ 0, & \text{se } i \text{ e } i' \text{ não compartilham alelos,} \end{cases}$$

$$(2.6)$$

quando i e $i^{'}$ são não relacionados $\pi_{ii^{'}}=0$. Adicionando o efeito do QTL, o modelo (2.1) é estendido como:

$$Y_f = \mathbf{X}_f \beta + \mathbf{q}_f + \mathbf{g}_f + \mathbf{e}_f, \quad f = 1, ..., F, \tag{2.7}$$

em que \mathbf{q}_f é o vetor de efeitos aleatórios do QTL, este vetor é normalmente distribuído com média zero e com matriz de covariâncias $\mathbf{\Pi}_f \sigma_q^2$, $\mathbf{\Pi}_f = \{\pi_{ii'}\}$ sendo a matriz IBD. Se o QTL tem efeito sobre \mathbf{Y} , é assumido que quanto maior a proporção IBD compartilhada, mais similares serão os valores da variável resposta \mathbf{Y} de indivíduos relacionados.

Com esta nova especificação, temos que a covariância (2.2) fica dada por:

$$Cov(y_{if}, y_{i'f'}) = \begin{cases} \sigma_q^2 + \sigma_g^2 + \sigma_e^2 \text{ para } i = i', \\ \pi_{ii'}\sigma_q^2 + 2\phi_{ii'}\sigma_g^2 \text{ para } i \neq i', f = f' \text{ (relacionados)}, \\ 0 \text{ para } i \neq i', f \neq f' \text{ (não-relacionados)}. \end{cases}$$
(2.8)

Assim, a matriz dada em (2.3) pode ser escrita como,

$$\mathbf{V} = \mathbf{\Pi}\sigma_a^2 + 2\mathbf{\Phi}\sigma_a^2 + \mathbf{I}\sigma_e^2. \tag{2.9}$$

O interesse está em testar a hipótese $H_0: \sigma_q^2 = 0$ vs $H_1: \sigma_q^2 > 0$, o que pode ser feito via a estatística (1/2):(1/2) de χ_1^2 e χ_0^2 , respectivamente (Self & Liang, 1987).

Em muitas situações o QTL pode ser definido em posições intermediárias no mapa de marcadores, posições estas não genotipadas, e, neste caso, a matriz Π é estimada a partir dos dados de marcadores adjacentes a essa região, (Blangero & Almasy, 1997).

2.1.2 Dados genéticos: Marcadores microsatélites e refinamento de posições genéticas candidatas

O modelo oligogênico (2.7), usando dados de marcadores microsatélites, faz a pesquisa

por regiões muito grandes do genoma, sendo que o interesse é encontrar regiões mais refinadas, no nível de nucleotídeo. Uma proposta para refinamento de posições encontrase em Almasy et al. (1997) em que no modelo (2.7), além da informação IBD (idêntico por descendência) de indivíduos relacionados, é incluída a informação IBS (idêntico por estado) de indivíduos não relacionados, com o objetivo de refinar o mapeamento de QTLs. A informação IBS de indivíduos não relacionados é obtida da informação de marcadores moleculares de forma análoga à informação IBD.

A correspondente matriz de covariâncias para o modelo (2.7) que inclui dependência entre indivíduos não relacionados tem componentes dados por:

$$Cov(y_{if}, y_{i'f'}) = \begin{cases} \sigma_q^2 + \sigma_g^2 + \sigma_e^2 \text{ para } i = i', \\ \pi_{ii'}\sigma_q^2 + 2\phi_{ii'}\sigma_e^2 \text{ para } i \neq i', f = f' \text{ (relacionados)}, \\ \psi_{ii'}\sigma_q^2 \text{ para } i \neq i', f \neq f' \text{ (não-relacionados)}, \end{cases}$$
(2.10)

 $\psi_{ii'}$ corresponde à proporção de alelos compartilhados entre os indivíduos i e $i^{'}$ não relacionados. A matriz (2.9) pode ser escrita como:

$$\mathbf{V} = \mathbf{\Pi}^+ \sigma_q^2 + 2\mathbf{\Phi}\sigma_g^2 + \mathbf{I}\sigma_e^2, \tag{2.11}$$

em que $\Pi^+ = \{\pi_{ii'} + \psi_{ii'}\}$ representa a matriz estimada de probabilidades IBD e IBS obtida através da informação de marcadores microsatélites. Detalhes e aplicações do método podem ser encontradas em Almasy et al. (1999).

Esta proposta na prática não tem muita utilidade porque é esperado que $\psi_{ii'}=0$ para a maioria dos indivíduos não relacionados em dados de marcadores microsatélites (em que cada loco corresponde a uma grande região cromossômica). Contudo, a ideia que está sendo introduzida na modelagem é muito interessante, isto é, usar a informação de indivíduos não relacionados para mapear regiões finas (pequenas) do genoma, pois é esperado que estes indivíduos compartilhem pequenas regiões cromossômicas em comum, são os chamados genes comuns que ocorrem na população em geral e correspondem a variações antigas e estáveis do genoma. Já indivíduos relacionados, que compartilham grandes regiões em comum, são úteis para a identificação de genes raros associados a

doenças e que possivelmente, são variações recentes que ocorrem no genoma (de algumas famílias) e não estão segregando na população geral.

2.1.3 Dados genéticos: Mapas de microsatélites e SNPs

Esta situação é abordada em duas etapas de análises.

- Localizar uma região cromossômica candidata, um QTL. Essa região é obtida usando dados de marcadores moleculares microsatélites (em que a amostragem é realizada por "grandes regiões" do genoma) por meio do ajuste do modelo oligogênico descrito em (2.7).
- Determinada a região cromossômica obtida na primeira etapa, esta pode ser refinada saturando-a com marcadores genéticos adicionais do tipo SNPs. As análises são realizadas com o intuito de encontrar efeito e posição de variantes genéticas associadas com a variável resposta \mathbf{Y} como consequência de mapeamento fino de QTLs. A modelagem estatística é uma extensão do modelo oligogênico (2.7) em que a variável preditora \mathbf{X}_j com dados do SNP_j , é modelada como um efeito fixo:

$$\mathbf{Y}_f = \mathbf{X}_f \boldsymbol{\beta} + \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{q}_f + \mathbf{g}_f + \mathbf{e}_f, \ f = 1, ..., F,$$
 (2.12)

em que \mathbf{X}_j é um vetor que representa os dados do j-ésimo SNP e β_j o correspondente coeficiente de regressão. Os demais termos do modelo estão definidos como em (2.7). O interesse está em testar a hipótese $H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$, que corresponde ao teste do efeito de SNP_j condicionado à região candidata associada ao efeito aleatório \mathbf{q} .

A grande desvantagem desta alternativa é que o custo financeiro de fazer a genotipagem dos indivíduos para os dois tipos de marcadores (microsatélites e SNPs) é muito alto e, em geral, as duas plataformas não estão disponíveis para as respectivas análises.

2.1.4 Dados genéticos: Mapas de SNPs

Quando se dispõe somente de dados de marcadores SNP para informar sobre o genoma

de indivíduos em delineamentos com famílias, é questionável se o efeito do SNP deve ser modelado como um efeito fixo no modelo poligênico (2.1) ou como um efeito aleatório (q) associado a uma matriz de covariâncias IBD como no modelo oligogênico (2.7). Lembrando que este tipo de marcador SNP é formado pela variação em um único nucleotídeo (Apêndice C) na modelagem oligogênica (2.7) a informação IBD é esperada ser aproximadamente igual à informação IBS dado que estes marcadores não modelam correlação entre indivíduos da mesma família (matriz Π muito esparsa).

Uma alternativa de análise em multiestágios de mapeamento de genes é proposta por Aulchenko et al. (2007) em que o modelo poligênico dado em (2.1) é estendido para incluir o efeito das variáveis preditoras \mathbf{X}_j , j=1,2,...,K modeladas como efeitos fixos. A proposta inclui os seguintes passos:

Passo 1 Os dados são analisados sob o modelo poligênico descrito em (2.1) para se obter os resíduos condicionais que são dados por,

$$\hat{\mathbf{e}}_f = \mathbf{Y}_f - (\mathbf{X}_f \hat{\boldsymbol{\beta}} + \hat{\mathbf{g}}_f), \tag{2.13}$$

em que $\hat{\beta}$ é o estimador dos efeitos fixos e $\hat{\mathbf{g}}_f$ é o valor predito do efeito aleatório do poligene para os indivíduos da família f.

Passo 2 Os resíduos obtidos em (2.13) são usados como uma variável dependente em um modelo de regressão simples, independentemente, para cada \mathbf{X}_j (cada SNP) da seguinte forma,

$$\hat{\mathbf{e}}_{ij} = u + \mathbf{X}_{ij}\beta_j + \mathbf{e}_{ij}, \ i = 1, 2, ..., n; j = 1, 2, ..., K,$$
 (2.14)

em que u é a média geral, \mathbf{X}_{ij} representa o genótipo do marcador j para o indivíduo i, β_j é o correspondente efeito do SNP_j , os erros \mathbf{e}_{ij} são assumidos normalmente distribuídos com média 0 e variância σ^2 , K é o número de SNPs no mapa sob estudo (em geral K é da ordem de 1 milhão)

Passo 3 As variáveis preditoras \mathbf{X}_j (cada SNP_j) com efeitos significantes (adotando-se um nível descritivo para múltiplos testes) no Passo 2 são selecionadas para uma análise final como efeito fixo em um modelo poligênico descrito a seguir:

$$\mathbf{Y}_f = \mathbf{X}_f \boldsymbol{\beta} + \beta_j \mathbf{X}_j + \mathbf{g}_f + \mathbf{e}_f, \ f = 1, ..., F.$$
 (2.15)

Aulchenko et al. (2007) fazem as aplicações utilizando dados simulados em que o número de SNPs no estudo não ultrapassa 100.000, o qual é um número pequeno quando comparado com as plataformas que incluem aproximadamente um milhão de variáveis preditoras SNP_i para serem analisadas.

Uma outra proposta encontra-se em Amin et al. (2007). Os autores seguem os mesmos passos apresentados em Aulchenko et al. (2007) com a diferença que os elementos da matriz de relacionamento 2Φ são obtidos por meio da informação dos SNPs e não da informação da genealogia familiar declarada como nos outros casos. Os autores argumentam que muitas vezes a informação completa e correta da genealogia dos indivíduos não é conhecida ou traz erros de declaração dos indivíduos (Leutennegger et al., 2003).

No mapeamento de genes, os modelos poligênico (2.1) e oligogênico (2.7), apesar de suas limitações, têm sido os métodos padrão em estudos genéticos que envolvem dados de família quando os marcadores utilizados são marcadores microsatélites, já que este tipo de marcador é útil para modelar covariância entre indivíduos da mesma família. Note que vários esforços têm sido feitos no intuito de refinar posições genéticas candidatas (Almasy et al., 1998) sem se obter o sucesso esperado. Métodos envolvendo os dois tipos de marcadores (microsatélites e SNPs) talvez sejam os mais adequados para o mapeamento de genes mas trata-se de uma análise dirigida ao mapeamento fino de regiões previamente identificadas ou que não garantem uma pesquisa do genoma inteiro.

Ainda, referente ao uso exclusivo de plataformas de SNPs, a proposta de Aulchenko et al. (2007) tem limitações no sentido de que o efeito individual do SNP é esperado que seja muito "pequeno" e de difícil detecção.

Capítulo 3

Gráfico da variável adicionada em Modelos Lineares Mistos

Os modelos mistos são considerados uma ferramenta poderosa na modelagem estatística devido a sua flexibilidade para modelar a estrutura de variância-covariância entre e intra unidades amostrais, a sua capacidade de tratar com dados balanceados bem como desbalanceados e a disponibilidade de pacotes confiáveis para o ajuste desses modelos. Uma extensa teoria para o caso normal é apresentada em detalhe em Verbeke e Molenberghs (2000). Diversas aplicações dos modelos mistos podem ser encontradas em Robinson (1991); Verbeke e Molenberghs (2000); Diggle et al (2002); Demidenko (2004); Verbeke e Molenberghs (2008) entre outros.

Análise de sensibilidade e análises de resíduos dos supostos inerentes constituem ferramentas importantes para avaliar o ajuste de qualquer modelo estatístico para um conjunto de dados. Em modelos mistos estas técnicas são mais recentes, encontram-se por exemplo em Beckman et al. (1987), Hilden-Minton (1995); Demidenko (2004); Demidenko & Stukel (2005); Nobre & Singer (2007), Nobre & Singer (2011) entre outros. Procedimentos de seleção de efeitos fixos e aleatórios simultaneamente, em um modelo linear misto, podem ser encontrados em Bondell et al. (2010) e em Ibrahim et al. (2011).

Neste capítulo será apresentada a teoria geral usual de Modelos Lineares Mistos e, dado o enfoque do presente trabalho, dá-se ênfase à teoria associada ao gráfico da variável adicionada a qual é estendida no intuito de particionar um efeito fixo em componentes associados a efeitos aleatórios e residuais, no contexto da modelagem de dados genéticos.

O modelo linear misto pode ser escrito na forma:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i, \ i = 1, ..., c, \tag{3.1}$$

em que \mathbf{Y}_i representa um vetor $(n_i \times 1)$ de respostas da i-ésima unidade; $\boldsymbol{\beta}$ é um vetor $(p \times 1)$ de parâmetros dos efeitos fixos; \mathbf{X}_i é uma matriz $(n_i \times p)$ de especificação dos efeitos fixos, a qual é conhecida e de posto completo; $\boldsymbol{\gamma}_i$ é um vetor $(q \times 1)$ de variáveis latentes, comumente denominadas efeitos aleatórios, as quais refletem o efeito individual da i-ésima unidade; \mathbf{Z}_i é uma matriz $(n_i \times q)$ de especificação dos efeitos aleatórios, também conhecida e de posto completo e $\boldsymbol{\epsilon}_i$ é um vetor $(n_i \times 1)$ de erros aleatórios.

Fazendo, $\mathbf{Y} = (\mathbf{Y}_1^{\top}, ..., \mathbf{Y}_c^{\top})^{\top}; \mathbf{X} = (\mathbf{X}_1^{\top}, ..., \mathbf{X}_c^{\top})^{\top}; \mathbf{Z} = \operatorname{diag}(\mathbf{Z}_1, ..., \mathbf{Z}_c); \boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^{\top}, ..., \boldsymbol{\gamma}_c^{\top})^{\top}$ e $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^{\top}, ..., \boldsymbol{\epsilon}_c^{\top})^{\top}$, o modelo (3.1) pode ser reescrito matricialmente como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}. \tag{3.2}$$

Em geral, assume-se que $E[\gamma] = 0$, $E[\epsilon] = 0$, com matriz de covariâncias:

$$Cov \quad \begin{pmatrix} \gamma \\ \epsilon \end{pmatrix} = \begin{pmatrix} \Delta & \mathbf{0}_{cq \times n} \\ \mathbf{0}_{n \times cq} & \mathbf{\Sigma} \end{pmatrix}, \tag{3.3}$$

em que $\mathbf{0}_{k_1 \times k_2}$ representa uma matriz nula de ordem $k_1 \times k_2$, Δ e Σ são matrizes positivas definidas, quadradas cujas ordens são cq e $n = \sum_{i=1}^{c} n_i$ e correspondem às matrizes de covariâncias dos vetores aleatórios γ e ϵ , respectivamente.

No modelo (3.2), os efeitos fixos são usados para modelar o valor esperado da variável resposta \mathbf{Y} , isto é $E[\mathbf{Y}] = \mathbf{X}\boldsymbol{\beta}$ enquanto que os efeitos aleatórios são utilizados para modelar sua estrutura de covariância $cov[\mathbf{Y}] = \mathbf{V} = \mathbf{Z}\Delta\mathbf{Z} + \boldsymbol{\Sigma}$. Usualmente, assume-se que $\boldsymbol{\gamma}$ e $\boldsymbol{\epsilon}$ seguem uma distribuição normal cq e n-variada, respectivamente.

No modelo (3.2), fazendo $\boldsymbol{\xi} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$, obtém-se o modelo marginal

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi},\tag{3.4}$$

com $\boldsymbol{\xi}$ tendo distribuição normal n-variada com vetor de médias $\mathbf{0}_n$ e matriz de covariâncias $\mathbf{V} = \mathbf{Z} \boldsymbol{\Delta} \mathbf{Z}^\top + \boldsymbol{\Sigma}$.

Colocando um parâmetro de dispersão em evidência tem-se $\Delta = \sigma^2 \mathbf{D}$ e $\Sigma = \sigma^2 \mathbf{R}$, com \mathbf{D}

e R denotando matrizes positivas definidas. Assim, a matriz de covariâncias é dada por:

$$\mathbf{V} = (\mathbf{Z}\mathbf{D}\mathbf{Z}^{\top} + \mathbf{R})\sigma^{2}.\tag{3.5}$$

Diferentes estruturas para **D** e **R** podem ser encontradas na literatura (por exemplo, Searle et al., 1992; Verbeke e Molenberghs, 1997).

Quando \mathbf{R} é uma matriz diagonal o modelo (3.2) é denominado **modelo de independência condicional**; se $\mathbf{R} = \mathbf{I}_n$ e $\Delta = \mathbf{0}_{cq \times cq}$ o modelo (3.2) corresponde ao modelo homoscedástico usual.

Para o modelo linear misto o **melhor estimador linear não viesado** ("BLUE-best linear unbiased estimator") para β e o **melhor preditor linear não viesado** ("BLUP-best linear unbiased preditor") para o vetor de efeitos aleatórios γ podem ser obtidos por mínimos quadrados por meio da solução do seguinte sistema de equações, conhecidas na literatura como **Equações de Henderson**:

$$\begin{pmatrix} \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{Y} \\ \mathbf{Z}^{\top} \mathbf{R}^{-1} \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^{\top} \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^{\top} \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^{\top} \mathbf{R}^{-1} \mathbf{Z} + \mathbf{D}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix}. \tag{3.6}$$

O BLUE de β é dado por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\top} \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^{\top} \mathbf{V}^{-1} \mathbf{Y}. \tag{3.7}$$

O BLUP do vetor de efeitos aleatórios γ é dado por

$$\hat{\gamma} = (\mathbf{Z}^{\top} \mathbf{R}^{-1} \mathbf{Z} + \mathbf{D}^{-1})^{-1} \mathbf{Z}^{\top} \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}})$$

$$= \mathbf{C}^{-1} \mathbf{Z}^{\top} \mathbf{R}^{-1} (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}), \tag{3.8}$$

 $com \mathbf{C} = \mathbf{D}^{-1} + \mathbf{Z}^{\top} \mathbf{R}^{-1} \mathbf{Z}.$

Define-se a matriz M como

$$\mathbf{M} = \sigma^2 \mathbf{V}^{-1},\tag{3.9}$$

sendo a matriz V dada em (3.5). Assim temse que a matriz M pode ser escrita como $M = (\mathbf{Z}\mathbf{D}\mathbf{Z}^{\top} + \mathbf{R})^{-1}$ e portanto a matriz M^{-1} é dada por:

$$\mathbf{M}^{-1} = \sigma^{-2} \mathbf{V} = (\mathbf{Z} \mathbf{D} \mathbf{Z}^{\mathsf{T}} + \mathbf{R}). \tag{3.10}$$

Esta matriz será muito importante no desenvolvimento teórico que será apresentado nas seções a seguir, já que por meio desta será possível fazer uma partição conveniente dos efeitos aleatórios. O BLUE de β em função da matriz \mathbf{M} é dado por:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathsf{T}} \mathbf{M} \mathbf{X})^{-1} \mathbf{X}^{\mathsf{T}} \mathbf{M} \mathbf{Y}. \tag{3.11}$$

O BLUP de γ também pode ser escrito como:

$$\hat{\gamma} = \mathbf{D}\mathbf{Z}^{\mathsf{T}}\mathbf{M}\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \mathbf{D}\mathbf{Z}^{\mathsf{T}}\mathbf{Q}\mathbf{Y}.$$
(3.12)

Resíduos em Modelos Lineares Mistos

Análises de resíduos em modelos lineares mistos são tratados por vários autores. Hilden-Minton (1995), Verbeke & Lesafre (1996), Pinheiro & Bates (2000), por exemplo, estenderam as ideias dos modelos lineares normais para definir três tipos de resíduos que contêm a fonte extra de variabilidade dos modelos lineares. Sob o modelo (3.2) podem-se definir três tipos de erros:

- Erros condicionais: $\epsilon = \mathbf{Y} \mathbf{E}[\mathbf{Y}|\boldsymbol{\gamma}] = \mathbf{Y} \mathbf{X}\boldsymbol{\beta} \mathbf{Z}\boldsymbol{\gamma};$
- Efeitos aleatórios: $\mathbf{Z}\boldsymbol{\gamma} = E[\mathbf{Y}|\boldsymbol{\gamma}] E[\mathbf{Y}];$
- Erros marginais: $\boldsymbol{\xi} = \mathbf{Y} E[\mathbf{Y}] = \mathbf{Y} \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$;

em que os correspondentes valores preditos, denominados resíduos, são dados, respectivamente por, $\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{Z}\hat{\boldsymbol{\gamma}}$, $\mathbf{Z}\hat{\boldsymbol{\gamma}}$ e $\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ com $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$ representando, respectivamente, o BLUE de $\boldsymbol{\beta}$ e o BLUP de $\boldsymbol{\gamma}$. Cada tipo de resíduo é útil para avaliar alguma suposição do modelo descrito em (3.2).

No modelo descrito em (3.4), os erros marginais são escritos como:

$$\boldsymbol{\xi} = \mathbf{Y} - E[\mathbf{Y}] = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{3.13}$$

cujos correspondentes valores preditos, denominados resíduos, são dados por

$$\hat{\boldsymbol{\xi}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

$$= \mathbf{Y} - \mathbf{X}(\mathbf{X}^{\top}\mathbf{M}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{M}\mathbf{Y}$$

$$= [\mathbf{I} - \mathbf{X}(\mathbf{X}^{\top}\mathbf{M}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{M}]\mathbf{Y}.$$
(3.14)

No intuito de fazer a partição dos efeitos aleatórios, estes resíduos marginais podem ser escritos em função da matriz \mathbf{M}^{-1} . Multiplicando os dois lados da equação (3.14) por $\mathbf{I} = \mathbf{M}^{-1}\mathbf{M}$ tem-se:

$$\hat{\boldsymbol{\xi}} = \mathbf{M}^{-1}[\mathbf{M} - \mathbf{M}\mathbf{X}(\mathbf{X}^{\top}\mathbf{M}\mathbf{X})^{-1}\mathbf{X}^{\top}\mathbf{M}]\mathbf{Y}$$

$$= \mathbf{M}^{-1}\mathbf{Q}\mathbf{Y}, \qquad (3.15)$$

em que a matriz \mathbf{Q} é dada por:

$$\mathbf{Q} = \mathbf{M} - \mathbf{M} \mathbf{X} \left(\mathbf{X}^{\mathsf{T}} \mathbf{M} \mathbf{X} \right)^{-1} \mathbf{X}^{\mathsf{T}} \mathbf{M}. \tag{3.16}$$

A matriz \mathbf{Q} é simétrica de ordem n e posto n-p (pois a matriz \mathbf{M} é simétrica, positiva definida de posto n). Duas propriedades importantes verificadas pela matriz \mathbf{Q} são:

$$\mathbf{Q}\mathbf{M}^{-1}\mathbf{Q} = \mathbf{Q}, \quad \mathbf{Q}\mathbf{X} = \mathbf{0},\tag{3.17}$$

em decorrência, tem-se que $\mathbf{M}^{-1}\mathbf{Q}$ é idempotente $(\mathbf{M}^{-1}\mathbf{Q}\mathbf{M}^{-1}\mathbf{Q} = \mathbf{M}^{-1}\mathbf{Q})$ e ortogonal às colunas da matriz \mathbf{X} $(\mathbf{M}^{-1}\mathbf{Q}\mathbf{X} = \mathbf{0})$. $\hat{\mathbf{Y}}$ e $\hat{\boldsymbol{\xi}}$ são ortogonais e podem ser escritos em função da matriz \mathbf{Q} , respectivamente, $\hat{\mathbf{Y}} = [\mathbf{I} - \mathbf{M}^{-1}\mathbf{Q}]\mathbf{Y}$ e $\hat{\boldsymbol{\xi}} = \mathbf{M}^{-1}\mathbf{Q}\mathbf{Y}$.

Da definição da matriz \mathbf{Q} , cada um dos resíduos podem ser escritos como:

$$\hat{\epsilon} = \mathbf{RQY},$$
 (3.18)

$$\hat{\boldsymbol{\xi}} = \mathbf{M}^{-1}\mathbf{Q}\mathbf{Y},\tag{3.19}$$

$$\mathbf{Z}\hat{\boldsymbol{\gamma}} = \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathsf{T}}\mathbf{Q}\mathbf{Y},\tag{3.20}$$

em que as respectivas variâncias são dadas por:

$$Var[\hat{\epsilon}] = \mathbf{RQR}\sigma^2,$$

$$Var[\hat{\xi}] = \mathbf{M}^{-1}\mathbf{QM}^{-1}\sigma^2,$$

$$Var[\mathbf{Z}\hat{\boldsymbol{\gamma}}] = \mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathsf{T}}\mathbf{Q}\mathbf{Z}\mathbf{D}\mathbf{Z}^{\mathsf{T}}\sigma^{2}.$$

Hilden-Minton (1995) faz uma classificação do resíduo em *puro* e *confundido*, o resíduo *puro* é aquele que depende apenas das componentes fixas do modelo e do respectivo erro

do qual ele é preditor, enquanto que um resíduo que depende de dois ou mais erros é dito ser *confundido*.

Sob a validade do modelo (3.2), em Nobre (2004) é mostrado que os resíduos podem ser escritos também da seguinte forma:

$$\hat{\epsilon} = \mathrm{RQR} \epsilon + \mathrm{RQZ} \gamma$$

$$\hat{\xi} - \xi = -\mathrm{X}^{ op} \mathrm{M} \mathrm{X}^{-1} \mathrm{X}^{ op} \mathrm{M} \xi$$

$$\mathrm{Z} \hat{\gamma} = \mathrm{ZDZ}^{ op} \mathrm{QZ} \gamma + \mathrm{ZDZ}^{ op} \mathrm{Q} \epsilon.$$

Concluí-se portanto que $\hat{\epsilon}$ e $\mathbf{Z}\hat{\gamma}$ são resíduos confundidos pela presença de γ e ϵ , respectivamente. Sendo que no diagnóstico por meio de resíduos é necessário considerar as suposições de independência, normalidade e variância constante dos elementos de ξ , no que segue utiliza-se o resíduo $\mathbf{M}^{1/2}\xi$ o qual tem distribuição Normal com matriz de covariância $\mathbf{I}\sigma^2$. Da equação (3.15) infere-se que

$$\mathbf{M}^{1/2}\hat{\boldsymbol{\xi}} = \mathbf{M}^{-1/2}\mathbf{Q}\mathbf{Y}.\tag{3.21}$$

Dentre as abordagens mais utilizadas na prática para medir influência em modelos lineares mistos, destacam-se as análises baseadas em **influência local** (Cook, 1986) e aquelas obtidas via eliminação de observações, **influência global**. Nas seções a seguir, serão discutidas algumas propostas como inclusão de efeitos fixos e a abordagem da teoria associada ao gráfico da variável adicionada em modelos mistos o qual é o foco do presente trabalho.

3.1 Inclusão de efeitos fixos

Hilden Minton (1995) apresenta algumas fórmulas para a inclusão de novas covariáveis como efeitos fixos no modelo misto. Tais fórmulas são úteis para definir o gráfico da variável adicionada em modelos lineares mistos entre outras aplicações.

Considere o modelo (3.1), com as matrizes **D** e **R** conhecidas e contendo inicialmente

os efeitos fixos β_1 , ou seja,

$$\mathbf{Y} = \mathbf{X}_1 \boldsymbol{\beta}_1 + \boldsymbol{\xi},\tag{3.22}$$

em que $\boldsymbol{\xi} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$. Chama-se este modelo de reduzido. No caso, tem-se interesse em adicionar novos efeitos fixos $\boldsymbol{\beta}_2$ ao modelo (3.22) ajustando o modelo,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}$$
$$= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\xi}, \tag{3.23}$$

com $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\top}, \boldsymbol{\beta}_2^{\top})$ e $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2]$ de posto completo e tal que posto $(\mathbf{X}) = p = \text{posto}(\mathbf{X}_1) + \text{posto}(\mathbf{X}_2) = p_1 + p_2$. Este modelo será considerado como modelo completo, aumentado relativamente ao modelo reduzido (3.22).

Analogamente à equação (3.7) tem-se que o BLUE de β_1 , relativo ao modelo reduzido (3.22) oqual é dado por

$$\hat{\boldsymbol{\beta}}_{1}^{0} = \left(\mathbf{X}_{1}^{\top} \mathbf{M} \mathbf{X}_{1}\right)^{-1} \mathbf{X}_{1}^{\top} \mathbf{M} \mathbf{Y}, \tag{3.24}$$

enquanto que o BLUE de β_2 referente ao modelo completo (3.23), pode ser obtido da seguinte forma: considere a matriz \mathbf{Q}_1 dada por

$$\mathbf{Q}_1 = \mathbf{M} - \mathbf{M} \mathbf{X}_1 \left(\mathbf{X}_1^{\mathsf{T}} \mathbf{M} \mathbf{X}_1 \right)^{-1} \mathbf{X}_1^{\mathsf{T}} \mathbf{M}, \tag{3.25}$$

com,

$$\mathbf{Q}_1 \mathbf{M}^{-1} \mathbf{Q}_1 = \mathbf{Q}_1, \quad \mathbf{Q}_1 \mathbf{X}_1 = \mathbf{0}.$$
 (3.26)

Aplicando \mathbf{Q}_1 ao modelo dado em (3.23), tem-se

$$\mathbf{Q}_1 \mathbf{Y} = \mathbf{Q}_1 \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Q}_1 \boldsymbol{\xi},$$

em que $\text{Var}[\mathbf{Q}_1\boldsymbol{\xi}] = \sigma^2 \mathbf{Q}_1 \mathbf{M}^{-1} \mathbf{Q}_1 = \sigma^2 \mathbf{Q}_1$ (equação (3.26)). Assim no modelo (3.23) é obtido o BLUE para $\boldsymbol{\beta}_2$, escrito como:

$$\hat{\boldsymbol{\beta}}_2 = \left(\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2\right)^{-1} \mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{Y},\tag{3.27}$$

em que $E[b\hat{f}\beta_2] = \boldsymbol{\beta}_2$ e $Var[\hat{\boldsymbol{\beta}}_2] = (\mathbf{X}_2^{\top}\mathbf{Q}_1\mathbf{X}_2)^{-1}\sigma^2$. Portanto, $\hat{\boldsymbol{\beta}}_2$ é obtido ao se adicionar novas variáveis contidas na matriz \mathbf{X}_2 com efeitos fixos a um modelo que contém inicialmente só as variáveis contidas em \mathbf{X}_1 .

Hilden-Minton (1995) obtêm fórmulas de atualização para o BLUE e o BLUP. A fórmula de atualização para o BLUE de $\hat{\beta}_1$ é dado pela seguinte expressão:

$$\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_1^0 = \left(\mathbf{X}_1^{\mathsf{T}} \mathbf{M} \mathbf{X}_1\right)^{-1} \mathbf{X}_1^{\mathsf{T}} \mathbf{M} \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2. \tag{3.28}$$

Também, denotando por $\hat{\gamma^0}$ e $\hat{\gamma}$ os BLUP referentes ao vetor de efeitos aleatórios dos modelos (3.22) e (3.23), respectivamente, tem-se

$$\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}^0} = \mathbf{D} \mathbf{Z}^\top \mathbf{Q}_1^{-1} \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2. \tag{3.29}$$

A matriz \mathbf{Q} obtida no modelo completo pode ser escrita em termos da matriz \mathbf{Q}_1 como a seguir:

$$\mathbf{Q} = \mathbf{Q}_1 - \mathbf{Q}_1 \mathbf{X}_2 (\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2)^{-1} \mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1. \tag{3.30}$$

Em função destas matrizes, \mathbf{Q} e \mathbf{Q}_1 , sob a suposição de normalidade e \mathbf{D} e \mathbf{R} conhecidas, tem-se que $\mathbf{Y}^{\top}\mathbf{Q}\mathbf{Y}$; $\mathbf{Y}^{\top}\mathbf{Q}_1\mathbf{Y}$ e $\mathbf{Y}^{\top}\mathbf{Q}_1\mathbf{X}_2(\mathbf{X}_2^{\top}\mathbf{Q}_1\mathbf{X}_2)^{-1}\mathbf{X}_2^{\top}\mathbf{Q}_1\mathbf{Y}$ são $\sigma^2\chi^2$ distribuidas com $n-p,\ n-p_1$ e p_2 graus de liberdade, respectivamemente.

Uma estatística para testar $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$ (que corresponde a testar se o efeito da variável adicionada \mathbf{X}_2 é significante) é

$$F = \frac{(\mathbf{Y}^{\top} \mathbf{Q}_1 \mathbf{Y} - \mathbf{Y}^{\top} \mathbf{Q} \mathbf{Y})/p_2}{\mathbf{Y}^{\top} \mathbf{Q} \mathbf{Y}/(n-p)} = \frac{\mathbf{Y}^{\top} (\mathbf{Q}_1 - \mathbf{Q}) \mathbf{Y}/p_2}{\mathbf{Y}^{\top} \mathbf{Q} \mathbf{Y}/(n-p)},$$
(3.31)

que sob H_0 possui uma distribuição F com p_2 e n-p graus de liberdade.

O desenvolvimento teórico dos resultados anteriores encontra-se, por exemplo, em Hilden-Minton (1995) e em Nobre (2004).

3.2 Gráfico da variável adicionada para efeitos fixos

Esta técnica gráfica é introduzida na literatura para avaliar o efeito da inclusão de uma

variável (preditora) particular num modelo estatístico. Essa ferramenta visual é proposta por Mosteller & Tuckey (1977) e subsequentemente discutida por vários autores, entre eles Cook & Weisberg (1982); Atkinson (1985) e Johnson & McCulloch (1987). Drapper & Smith (1998) e Belsley et al. (1980) têm discutido o uso de gráficos de variáveis adicionadas em modelos de regressão linear ordinária. Wang (1985) estendeu o procedimento de diagnóstico para Modelos Lineares Generalizados. Cheng & Wang (1991) têm usado o metodo para diagnóstico em modelos de regressão de Cox. Rahmatullah Imon (2003) aperfeiçõou o gráfico da variável adicionada com remoção de resíduos para detectar dados fora do padrão em modelos de regressão normal. Chien (2011) estendeu as técnicas de verossimilhança robusta introduzidas por Royall & Sou (2003) com o propósito de melhorar o gráfico da variável adicionada. Hodges (1998) tem sugerido o uso de gráficos da variável adicionada em análises de Modelos Lineares Mistos. Posteriormente, Hilden-Minton (1995) escreve o efeito da variável adicionada em função do resíduo do modelo e sugere uma decomposição dos resíduos do modelo misto. Esta decomposição será tomada como base para decomposições adicionais no desenvolvimento teórico descrito nas seções seguintes.

Os gráficos da variável adicionada podem ser usados para identificar possíveis observações influentes nos coeficientes das variáveis explicativas. Também podem mostrar se a variável incluída tem um efeito linear sobre a variável resposta ou se alguma transformação deveria ser considerada, espera-se uma relação linear entre os resíduos considerados quando a respectiva variável preditora tem uma relação linear com a variável resposta. Assim, este tipo de gráficos podem fornecer ideias de grande valor na construção de um modelo.

Para se estudar o efeito que a inclusão de um efeito fixo exerce no componente aleatório total do modelo ($\boldsymbol{\xi} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}$), utiliza-se o gráfico da variável adicionada. Em primeiro lugar, considera-se o caso em que \mathbf{X}_2 no modelo (3.23) tem uma única coluna, ou seja em que há interesse em incluir um único efeito fixo. Em tudo o que segue, \mathbf{X}_2 representará a variável que está sendo adicionada e cujo interesse será estudar o seu efeito no modelo linear misto, especificamente o efeito causado nos componentes aleatórios.

A partir da variável que está sendo adicionada, descreve-se o seguinte modelo:

$$\mathbf{X}_2 = \mathbf{X}_1 \boldsymbol{\beta}_1^* + \boldsymbol{\xi}^*, \tag{3.32}$$

com $\boldsymbol{\xi}^* = \boldsymbol{\gamma}^* + \boldsymbol{\epsilon}^*$, \mathbf{X}_2 a variável resposta tendo \mathbf{X}_1 como variável preditora e $\boldsymbol{\gamma}^*$ e $\boldsymbol{\epsilon}^*$ definidos como no modelo (3.2). A partir dos resíduos obtidos do modelo reduzido dado em (3.22) e do modelo para a variável adicionada dado em (3.32) é obtido o gráfico da variável adicionada. De forma geral, se existir alguma relação linear entre estes resíduos, a variável \mathbf{X}_2 deve ser incluída ao modelo misto de forma linear.

O resíduo total (puro) no modelo (3.32), que será denotado por \mathbf{r}_{x_2} , é dado por:

$$\mathbf{r}_{x_2} = \hat{\boldsymbol{\xi}}^* = \mathbf{M}^{-1/2} \mathbf{Q}_1 \mathbf{X}_2. \tag{3.33}$$

Considerando o modelo reduzido (3.22), os resíduos, denotados por \mathbf{r}_y , podem ser escritos como:

$$\mathbf{r}_y = \hat{\boldsymbol{\xi}} = \mathbf{M}^{-1/2} \mathbf{Q}_1 \mathbf{Y},\tag{3.34}$$

em que \mathbf{r}_y representa a quantidade da variabilidade em \mathbf{Y} que não pode ser explicada por \mathbf{X}_1 . De forma mais explícita, se existir correlação entre \mathbf{r}_{x_2} e \mathbf{r}_y então existe informação em \mathbf{X}_2 , que não existe em \mathbf{X}_1 , para explicar a variável resposta \mathbf{Y} e, portanto, o modelo (3.23) em que é incluída \mathbf{X}_2 como covariável é o mais adequado.

Em Hilden-Minton (1995) e em Nobre (2004) é mostrado que os resíduos \mathbf{r}_{x_2} , sob o modelo (3.32), e \mathbf{r}_y , sob o modelo (3.22), podem ser escritos, respectivamente, como

$$\mathbf{r}_{r_2} = \mathbf{M}^{-1/2} \mathbf{Q}_1 \mathbf{X}_2 = \mathbf{M}^{-1/2} (\mathbf{X}_2 - \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1^*) = \mathbf{M}^{-1/2} (\mathbf{R})^{-1} (\mathbf{X}_2 - \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1^* - \mathbf{Z} \hat{\boldsymbol{\gamma}}^*), \quad (3.35)$$

e

$$\mathbf{r}_y = \mathbf{M}^{-1/2} \mathbf{Q}_1 \mathbf{Y} = \mathbf{M}^{-1/2} (\mathbf{Y} - \mathbf{X}_1 \hat{\beta}_1) = \mathbf{M}^{-1/2} (\mathbf{R})^{-1} (\mathbf{Y} - \mathbf{X}_1 \hat{\beta}_1^0 - \mathbf{Z} \hat{\gamma}^0).$$
 (3.36)

Ressalta-se o fato que cada resíduo descrito em (3.35) e (3.36) está em função dos seus respectivos BLUE e BLUP $(\hat{\beta}_1^*, \hat{\gamma}^* \in \hat{\beta}_1^0, \hat{\gamma}^0)$.

Hilden-Minton (1995) chama de gráficos de variáveis adicionadas marginais porque são

derivados do modelo marginal. São baseadas em resíduos puros, em que o termo de erro composto é usado.

De grande interesse é o fato que $\hat{\boldsymbol{\beta}}_2$, dado na equação (3.27), possa ser escrito em função dos resíduos \mathbf{r}_{x_2} e \mathbf{r}_y (Hilden-Minton, 1995; Nobre, 2004). Levando em consideração as propriedades da matriz \mathbf{Q}_1 dadas em (3.26) ($\mathbf{Q}_1\mathbf{M}^{-1}\mathbf{Q}_1 = \mathbf{Q}_1$), tem-se:

$$\hat{\boldsymbol{\beta}}_{2} = \left(\mathbf{X}_{2}^{\top}\mathbf{Q}_{1}\mathbf{X}_{2}\right)^{-1}\mathbf{X}_{2}^{\top}\mathbf{Q}_{1}\mathbf{Y}$$

$$= \left(\mathbf{X}_{2}^{\top}\mathbf{Q}_{1}\mathbf{M}^{-1}\mathbf{Q}_{1}\mathbf{X}_{2}\right)^{-1}\mathbf{X}_{2}^{\top}\mathbf{Q}_{1}\mathbf{M}^{-1}\mathbf{Q}_{1}\mathbf{Y}$$

$$= \left(\mathbf{X}_{2}^{\top}\mathbf{Q}_{1}\mathbf{M}^{-1/2}\mathbf{M}^{-1/2}\mathbf{Q}_{1}\mathbf{X}_{2}\right)^{-1}\mathbf{X}_{2}^{\top}\mathbf{Q}_{1}\mathbf{M}^{-1/2}\mathbf{M}^{-1/2}\mathbf{Q}_{1}\mathbf{Y}$$

$$= \left(\mathbf{r}_{x_{2}}^{\top}\mathbf{r}_{x_{2}}\right)^{-1}\mathbf{r}_{x_{2}}^{\top}\mathbf{r}_{y}.$$
(3.37)

Portanto $\hat{\boldsymbol{\beta}}_2$ pode ser interpretado como o estimador de mínimos quadrados (ordinários) do coeficiente angular de uma regressão linear sem intercepto tendo como variável resposta os resíduos \mathbf{r}_y obtidos do modelo reduzido (equação (3.22)) e como variável explicativa os resíduos \mathbf{r}_{x_2} obtidos do modelo de regressão dado na equação (3.32). Especificamente tem-se o modelo estrutural,

$$\mathbf{r}_{ui} = \beta_2 \mathbf{r}_{x_2 i} + e_i, i = 1, ..., n, \tag{3.38}$$

assim, o gráfico da variável adicionada é definido como o gráfico de dispersão dos resíduos $\mathbf{r}_{\mathbf{x_2}}$ (eixo horizontal) e \mathbf{r}_y (eixo vertical). No gráfico de dispersão é ajustada uma reta de regressão sem intercepto (passando pela origem) de \mathbf{r}_y como variável resposta e \mathbf{r}_{x_2} como preditora. Caso a inclinação dessa reta seja significante (estatisticamente diferente de zero), então existe evidência de que a covariável \mathbf{X}_2 deve ser incluída no modelo.

No modelo de regressão linear simples (3.38), pode-se usar a estatística t para testar $H_0: \beta_2 = 0$ vs $H_1: \beta_2 \neq 0$.

Considerando a estatística t^2 e substituindo os respectivos resíduos, $\mathbf{r}_{x_2}^{\top}$ e \mathbf{r}_y^{\top} , dados nas equações (3.35) e (3.36), respectivamente, chega-se à estátistica F descrita em (3.31) especificamente.

$$t^{2} = \frac{\mathbf{r}_{y}^{\mathsf{T}} \mathbf{r}_{x_{2}} (\mathbf{r}_{x_{2}}^{\mathsf{T}} \mathbf{r}_{x_{2}})^{-1} \mathbf{r}_{x_{2}}^{\mathsf{T}} \mathbf{r}_{y}}{\mathbf{r}_{y}^{\mathsf{T}} [\mathbf{I} - \mathbf{r}_{1} (\mathbf{r}_{1}^{\mathsf{T}} \mathbf{r}_{1})^{-1} \mathbf{r}_{1}^{\mathsf{T}}] \mathbf{r}_{y}} = F.$$

$$(3.39)$$

Lembrando que $\hat{\beta}_2$ no modelo misto depende da estrutura de covariância, pretende-se conhecer qual componente desta estrutura está sendo mais afetada pela adição de uma variável como efeito fixo ou, dito de outra maneira, qual estrutura, aleatória ou residual, está contribuindo mais para a possível significância do efeito em questão. Na seção seguinte aborda-se este tema.

3.3 Decomposição do gráfico da variável adicionada

Sendo o interesse estudar como a inclusão de uma covariável afeta o resíduo puro do modelo envolvido, a decomposição do gráfico da variável adicionada tem como objetivo estabelecer como essa nova covariável influencia cada um dos resíduos separadamente. Hilden-Minton (1995) propõe decompor o gráfico da variável adicionada para avaliar o efeito da inclusão de um efeito fixo na predição dos efeitos aleatórios. A ideia básica consiste em substituir a matriz $\mathbf{M}^{-1/2}$ por uma matriz \mathbf{A} de dimensão $(n+q) \times n$ tal que $\mathbf{M}^{-1} = \mathbf{A}^{\top} \mathbf{A}$ nas expressões (3.35) e (3.36). O autor sugere, em particular, utilizar a seguinte matriz

$$\mathbf{A} = \begin{pmatrix} \mathbf{R}^{1/2} \\ \mathbf{D}^{1/2} \mathbf{Z}^{\top} \end{pmatrix}. \tag{3.40}$$

Note que a matriz \mathbf{A} como definida em (3.40) também é particionada nos dois efeitos aleatório e residual. O gráfico da variável adicionada corresponde ao gráfico das componentes de $\mathbf{AQ}_1\mathbf{Y}$ versus $\mathbf{AQ}_1\mathbf{X}_2$, isto é:

$$\mathbf{r}_{y} = \mathbf{A}\mathbf{Q}_{1}\mathbf{Y} = \begin{pmatrix} \mathbf{R}^{1/2}\mathbf{Q}_{1}\mathbf{Y} \\ \mathbf{D}^{1/2}\mathbf{Z}^{\top}\mathbf{Q}_{1}\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \mathbf{R}^{-1/2}(\mathbf{Y} - \mathbf{X}_{1}\hat{\beta}_{1} - \mathbf{Z}\hat{\boldsymbol{\gamma}}) \\ \mathbf{D}^{-1/2}\hat{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_{y1} \\ \mathbf{r}_{y2} \end{pmatrix}, \tag{3.41}$$

 \mathbf{e}

$$\mathbf{r}_{x_2} = \mathbf{A}\mathbf{Q}_1\mathbf{X}_2 = \begin{pmatrix} \mathbf{R}^{1/2}\mathbf{Q}_1\mathbf{X}_2 \\ \mathbf{D}^{1/2}\mathbf{Z}^{\top}\mathbf{Q}_1\mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{R}^{-1/2}(\mathbf{X}_2 - \mathbf{X}_1\hat{\beta}_1^* - \mathbf{Z}\hat{\gamma}^*) \\ \mathbf{D}^{-1/2}\hat{\gamma}^* \end{pmatrix} = \begin{pmatrix} \mathbf{r}_{x_{2,1}} \\ \mathbf{r}_{x_{2,2}} \end{pmatrix}. \tag{3.42}$$

O gráfico de \mathbf{r}_{y1} versus $\mathbf{r}_{x2.1}$ é chamado por Hilden-Minton (1995) de **gráfico da** variável adicionada intra-unidades, pois leva em consideração as mudanças no BLUE $(\hat{\beta}_1 \in \hat{\beta}_1^*)$ e no BLUP $(\hat{\gamma} \in \hat{\gamma}^*)$. O segundo gráfico, de \mathbf{r}_{y2} versus $\mathbf{r}_{x2.2}$ é denominado de **gráfico da variável adicionada entre-unidades**, pois só leva em consideração a mudança existente entre os BLUP $(\hat{\gamma} \in \hat{\gamma}^*)$. A principal vantagem da decomposição proposta é que por meio desta é possível avaliar o efeito da inclusão de uma covariável no BLUP.

Neste trabalho, dando continuidade à teoria desenvolvida em Hilden-Minton (1995), levando em consideração os dois gráficos (intra-unidades e entre-unidades) de forma análoga ao modelo de regressão linear simples dado na equação (3.38) para o gráfico da variável adicionada no caso marginal, podem-se ter dois modelos de regressão linear simples para os dois gráficos (intra-unidades e entre-unidades). Definindo β_2^e e β_2^g como os coeficientes angulares das regressões lineares sem intercepto, os respectivos modelos podem ser escritos da seguinte forma:

$$\mathbf{r}_{v1i} = \beta_2^e \mathbf{r}_{x_{2,1}i} + e_{1i}, \ i = 1, ..., n, \tag{3.43}$$

е

$$\mathbf{r}_{y2i} = \beta_2^g \mathbf{r}_{x_{2,2}i} + e_{2i}, \ i = 1, ..., n.$$
(3.44)

O estimador de cada parâmetro envolvido no ajuste do modelo de regressão sem intercepto, sob as premissas clássicas, é dado por:

$$\hat{\beta}_{2}^{e} = (\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{x_{2.1}})^{-1} \mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{y_{1}}$$
(3.45)

$$= (\mathbf{X}_{2}^{\mathsf{T}} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{Q}_{1} \mathbf{X}_{2})^{-1} \mathbf{X}_{2} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{Y}, \tag{3.46}$$

e

$$\hat{\beta}_{2}^{g} = (\mathbf{r}_{x_{2.2}}^{\top} \mathbf{r}_{x_{2.2}})^{-1} \mathbf{r}_{x_{2.2}}^{\top} \mathbf{r}_{y_{2}}$$
(3.47)

$$= (\mathbf{X}_{2}^{\mathsf{T}} \mathbf{Q}_{1} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\mathsf{T}} \mathbf{Q}_{1} \mathbf{X}_{2})^{-1} \mathbf{X}_{2} \mathbf{Q}_{1} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\mathsf{T}} \mathbf{Q}_{1} \mathbf{Y}, \tag{3.48}$$

em que,

$$E[\hat{\beta}_{2}^{e}] = (\mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{X}_{2})^{-1} \mathbf{X}_{2} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} E[\mathbf{Y}]$$

$$= (\mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{X}_{2})^{-1} \mathbf{X}_{2} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{X}_{2} \beta_{2}$$

$$= \beta_{2}. \tag{3.49}$$

De forma análoga, tem-se,

$$E[\hat{\beta}_2^g] = \beta_2,\tag{3.50}$$

assim, $\hat{\beta}_2^e$ e $\hat{\beta}_2^g$ podem ser interpretados como estimadores de mínimos quadrados ordinários das respectivas regressões sem intercepto e ao mesmo tempo são estimadores do parâmetro de interesse β_2 . Os gráficos de dispersão entre os respectivos resíduos (\mathbf{r}_{y1} e $\mathbf{r}_{x_{2.1}}$; \mathbf{r}_{y2} e $\mathbf{r}_{x_{2.2}}$) representam os gráficos da variável adicionada intra-unidades e entre-unidades, respectivamente.

O estimador do parâmetro de interesse β_2 dado na expressão (3.27) pode ser escrito em termos dos coeficientes angulares $\hat{\beta}_2^e$ e $\hat{\beta}_2^g$, dados nas expressões (3.45) e (3.47), respectivamente. Em primeiro lugar, consideram-se os resíduos \mathbf{r}_y e $\mathbf{r}_{x_2}^{\top}$ obtidos da partição dada nas equações (3.41) e (3.42) para serem substituídos na expressão dada para $\hat{\beta}_2$ como segue:

$$\hat{\beta}_{2} = (\mathbf{r}_{x_{2}}^{\top} \mathbf{r}_{x_{2}})^{-1} \mathbf{r}_{x_{2}}^{\top} \mathbf{r}_{y}$$

$$= (\mathbf{r}_{x_{2}} \mathbf{r}_{x_{2}})^{-1} (\mathbf{r}_{x_{2,1}}^{\top} \mathbf{r}_{x_{2,2}}^{\top}) (\mathbf{r}_{y_{1}})$$

$$= (\mathbf{r}_{x_{2}}^{\top} \mathbf{r}_{x_{2}})^{-1} [\mathbf{r}_{x_{2,1}}^{\top} \mathbf{r}_{y_{1}} + \mathbf{r}_{x_{2,2}}^{\top} \mathbf{r}_{y_{2}}], \qquad (3.51)$$

tomando as duas expressões dentro do colchete, tem-se que, multiplicando a primeira expressão por $(\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{x_{2.1}})(\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{x_{2.1}})^{-1}$, pela equação (3.45) chega-se na seguinte igualdade

$$(\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{x_{2.1}})(\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{x_{2.1}})^{-1}\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{y1} = (\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{x_{2.1}})\hat{\boldsymbol{\beta}}_{2}^{e}.$$

De forma análoga, multiplicando a segunda expressão por $(\mathbf{r}_{x_{2.2}}^{\top}\mathbf{r}_{x_{2.2}})(\mathbf{r}_{x_{2.2}}^{\top}\mathbf{r}_{x_{2.2}})^{-1}$, pela equação (3.47) chega-se na seguinte expressão

$$(\mathbf{r}_{x_{2.2}}^{\top}\mathbf{r}_{x_{2.2}})(\mathbf{r}_{x_{2.2}}^{\top}\mathbf{r}_{x_{2.2}})^{-1}\mathbf{r}_{x_{2.2}}^{\top}\mathbf{r}_{y2} = (\mathbf{r}_{x_{2.2}}^{\top}\mathbf{r}_{x_{2.2}})\hat{\beta}_{2}^{g},$$

portanto substituindo as duas expressões dentro do colchete, $\hat{\beta}_2$ pode ser escrito como a seguir

$$\hat{\beta}_{2} = (\mathbf{r}_{x_{2}}^{\top} \mathbf{r}_{x_{2}})^{-1} \left[(\mathbf{r}_{x_{2,1}}^{\top} \mathbf{r}_{x_{2,1}}) \hat{\beta}_{2}^{e} + (\mathbf{r}_{x_{2,2}}^{\top} \mathbf{r}_{x_{2,2}}) \hat{\beta}_{2}^{g} \right]$$
(3.53)

$$= \frac{(\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{x_{2.1}})}{(\mathbf{r}_{x_{2}}^{\top} \mathbf{r}_{x_{2}})} \hat{\beta}_{2}^{e} + \frac{(\mathbf{r}_{x_{2.2}}^{\top} \mathbf{r}_{x_{2.2}})}{(\mathbf{r}_{x_{2}}^{\top} \mathbf{r}_{x_{2}})} \hat{\beta}_{2}^{g}, \tag{3.54}$$

e, sendo que $\mathbf{r}_{x_2}^{\intercal}\mathbf{r}_{x_2} = \mathbf{r}_{x_{2.1}}^{\intercal}\mathbf{r}_{x_{2.1}} + \mathbf{r}_{x_{2.2}}^{\intercal}\mathbf{r}_{x_{2.2}}$, chega-se na seguinte expressão para $\hat{\beta}_2$

$$\hat{\beta}_{2} = \frac{\mathbf{r}_{x_{2.2}}^{\top} \mathbf{r}_{x_{2.2}}}{\mathbf{r}_{x_{2.1}}^{\top} + \mathbf{r}_{x_{2.2}}^{\top} \mathbf{r}_{x_{2.2}}} \hat{\beta}_{2}^{g} + \frac{\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{x_{2.1}}}{\mathbf{r}_{x_{2.1}}^{\top} + \mathbf{r}_{x_{2.2}}^{\top} \mathbf{r}_{x_{2.2}}} \hat{\beta}_{2}^{e}$$

$$= w_{g} \hat{\beta}_{2}^{g} + w_{e} \hat{\beta}_{2}^{e}, \qquad (3.55)$$

com,

$$w_g = \frac{\mathbf{r}_{x_{2.2}}^{\top} \mathbf{r}_{x_{2.2}}}{\mathbf{r}_{x_{2.1}}^{\top} + \mathbf{r}_{x_{2.2}}^{\top} \mathbf{r}_{x_{2.2}}}, \quad w_e = \frac{\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{x_{2.1}}}{\mathbf{r}_{x_{2.1}}^{\top} + \mathbf{r}_{x_{2.2}}^{\top} \mathbf{r}_{x_{2.2}}},$$
(3.56)

em que $w_g + w_e = 1$.

Para uma análise mais detalhada de w_g e w_e , substituindo os resíduos pelas respectivas matrizes descritas nas equações (3.41) e (3.42), tem-se:

$$w_g = \frac{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{Z} \mathbf{D} \mathbf{Z}^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2}{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2 + \mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{Z} \mathbf{D} \mathbf{Z}^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2} = \frac{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{Z} \mathbf{D} \mathbf{Z}^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2}{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2},$$
(3.57)

e

$$w_e = \frac{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{R} \mathbf{Q}_1 \mathbf{X}_2}{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{R} \mathbf{Q}_1 \mathbf{X}_2 + \mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{Z} \mathbf{D} \mathbf{Z}^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2} = \frac{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{R} \mathbf{Q}_1 \mathbf{X}_2}{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2}.$$
 (3.58)

Note que tanto w_g quanto w_e estão associados com a variável \mathbf{X}_2 (que está sendo adicionada ao modelo linear misto); o peso w_g é associado especificamente com os efeitos aleatórios (γ) e o peso w_e está associado com os erros (ϵ) e refletem a proporção da contribuição de cada componente para a possível significância do efeito β_2 . Os valores máximos e mínimos que os pesos w_g e w_e podem assumir são obtidos mediante o Lema da maximização de formas quadráticas (Johnson & Wichern, 1998), da seguinte forma:

$$\max_{\mathbf{X}_{2} \neq 0} \frac{(\mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{X}_{2})}{(\mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{X}_{2})} = \max_{\mathbf{X}_{2} \neq 0} \frac{(\mathbf{X}_{2}^{\top} \mathbf{Q}_{1}^{1/2} \mathbf{Q}_{1}^{1/2} \mathbf{R} \mathbf{Q}_{1}^{1/2} \mathbf{Q}_{1}^{1/2} \mathbf{X}_{2})}{(\mathbf{X}_{2}^{\top} \mathbf{Q}_{1}^{1/2} \mathbf{Q}_{1}^{1/2} \mathbf{X}_{2})} = \max_{\mathbf{X}_{2} \neq 0} \frac{(\mathbf{X}^{*\top} \mathbf{Q}_{1}^{1/2} \mathbf{R} \mathbf{Q}^{1/2} \mathbf{X}^{*})}{(\mathbf{X}^{*\top} \mathbf{X}^{*})} = \lambda_{1}^{R},$$
 (3.59)

em que $\lambda_1^R, \lambda_2^R, ..., \lambda_n^R$ representam os n autovalores da matriz $\mathbf{Q}_1^{1/2}\mathbf{R}\mathbf{Q}_1^{1/2}$. Analogamente,

$$\max_{X_{2} \neq 0} \frac{(\mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\top} \mathbf{Q}_{1} \mathbf{X}_{2})}{(\mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{X}_{2})} = \max_{\mathbf{X}_{2} \neq 0} \frac{(\mathbf{X}_{2}^{\top} \mathbf{Q}_{1}^{1/2} \mathbf{Q}_{1}^{1/2} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\top} \mathbf{Q}_{1}^{1/2} \mathbf{X}_{2})}{(\mathbf{X}_{2}^{\top} \mathbf{Q}_{1}^{1/2} \mathbf{Q}_{2}^{1/2} \mathbf{X}_{2})} = \max_{\mathbf{X}_{2} \neq 0} \frac{(\mathbf{X}^{*\top} \mathbf{Q}_{1}^{1/2} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\top} \mathbf{Q}^{1/2} \mathbf{X}_{2})}{(\mathbf{X}^{*\top} \mathbf{X}^{*})} = \lambda_{1}^{D}$$
(3.60)

em que $\lambda_1^D, \lambda_2^D, ..., \lambda_n^D$ representam os n autovalores da matriz $\mathbf{Q}_1^{1/2}\mathbf{Z}\mathbf{D}\mathbf{Z}^{\top}\mathbf{Q}_1^{1/2}$ e, em geral, $\lambda_n^R \leq w_e \leq \lambda_1^R$ e $\lambda_n^D \leq w_g \leq \lambda_1^D$. $\mathbf{X}^* = \mathbf{Q}_1^{1/2}\mathbf{X}_2$.

Também, sendo que $E[\hat{\beta}_2] = E[\hat{\beta}_2^g] = E[\hat{\beta}_2^e] = \beta_2$, isto indica a decisiva importância dos pesos w_g e w_e na utilização da partição (3.55) para discriminar entre variáveis preditoras. No Capítulo 5 os pesos w_g e w_e serão estudados por meio de diferentes cenários de simulação.

Voltando à equação (3.55) em que $\hat{\beta}_2$ é particionado nas componentes $w_g \hat{\beta}_2^g$ e $w_e \hat{\beta}_2^e$, temse que cada componente (incluíndo o peso e o coeficiente angular) é um novo coeficiente angular obtido do ajuste de um modelo de regressão linear simples, especificamente:

$$w_{g}\hat{\beta}_{2}^{g} = w_{g}(\mathbf{r}_{x_{2,2}}^{\top}\mathbf{r}_{x_{2,2}})^{-1}\mathbf{r}_{x_{2,2}}^{\top}\mathbf{r}_{y_{2}}$$

$$= (\mathbf{r}_{x_{2,2}}^{\top}\mathbf{r}_{x_{2,2}})^{-1}\mathbf{r}_{x_{2,2}}^{\top}w_{g}\mathbf{r}_{y_{2}}$$

$$= (\mathbf{r}_{x_{2,2}}^{\top}\mathbf{r}_{x_{2,2}})^{-1}\mathbf{r}_{x_{2,2}}^{\top}\mathbf{r}_{y_{2}}^{*}$$

$$= \beta_{2}^{wg}$$
(3.61)

em que $\mathbf{r}_{y2}^* = w_g \mathbf{r}_{y2}$. De forma análoga,

$$w_{e}\hat{\beta}_{2}^{e} = w_{e}(\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{x_{2.1}})^{-1}\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{y_{1}}$$

$$= (\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{x_{2.1}})^{-1}\mathbf{r}_{x_{2.1}}^{\top}w_{e}\mathbf{r}_{y_{1}}$$

$$= (\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{x_{2.1}})^{-1}\mathbf{r}_{x_{2.1}}^{\top}\mathbf{r}_{y_{1}}^{*}$$

$$= \beta_{2}^{we}$$
(3.62)

em que $\mathbf{r}_{y1}^* = w_e \mathbf{r}_{y1}$.

Assim, $\hat{\beta}_2^{we}$ e $\hat{\beta}_2^{wg}$ são os coeficientes dos ajustes dos modelos de regressão simples dados por:

$$\mathbf{r}_{y1}^* i = \beta_2^{we} \mathbf{r}_{x_{2.1}} i + \mathbf{e}_1 i, \ i = 1, ..., n,$$
(3.63)

$$\mathbf{r}_{u2}^* i = \beta_2^{wg} \mathbf{r}_{x_{2,2}} i + \mathbf{e}_2 i, \ i = 1, ..., n,$$
(3.64)

respectivamente, obtidos sob premissas clássicas.

Para ilustrar a partição obtida em (3.55), a Figura 3.1 apresenta três linhas cuja inclinação representa os três coeficientes angulares envolvidos $(\hat{\beta}_2, \hat{\beta}_2^{we}, \hat{\beta}_2^{wg})$. O coeficiente angular da linha 1 (vermelha) representa $\hat{\beta}_2$, o efeito total da variável adicionada \mathbf{X}_2 . Este efeito total é então dado pela soma dos coeficientes angulares das retas em (3.63) e (3.64). Especificamente, o coeficiente angular $\hat{\beta}_2^{we}$ da linha 2 (azul) é devido a efeitos residuais e o outro coeficiente angular $\hat{\beta}_2^{wg}$ da linha 3 (verde) é devido a efeitos aleatórios. Por meio desta partição (3.55), é possível selecionar aquelas variáveis preditoras que explicam uma parte relevante da variabilidade associada ao efeito aleatório do modelo. Note contudo, que o uso de (3.55) como critério de seleção de variáveis preditoras está restrito a efeitos β_2 significantes.

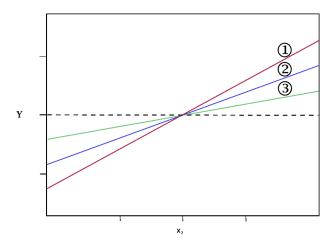


Figura 3.1 Partição do efeito fixo em coeficientes devidos a efeitos aleatórios e residuais.

Para testar a significância dos parâmetros envolvidos nos modelos de regressão dados em (3.63) e (3.64) tem-se a estatística t a qual ao quadrado é equivalente à estatística F como definida em (3.31).

Considerando a equação (3.17) tem-se que a estatística t^2 para se testar a hipótese $H_0:\beta_2^{we}=0$ é dada por:

$$t^{2} = \frac{\mathbf{r}_{y1}^{\top} \mathbf{r}_{x_{2.1}} (\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{x_{2.1}})^{-1} \mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{y1}}{\mathbf{r}_{y1}^{\top} [\mathbf{I} - \mathbf{r}_{x_{2.1}} (\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{x_{2.1}})^{-1} \mathbf{r}_{x_{2.1}}^{\top}] \mathbf{r}_{y1}},$$
(3.65)

e, substituindo os respectivos, resíduos chega-se na expressão

$$Fe = \frac{\mathbf{Y}^{\top} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{X}_{2} (\mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{X}_{2})^{-1} \mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{Y}}{\mathbf{Y}^{\top} [\mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} - \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{X}_{2} (\mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1} \mathbf{X}_{2})^{-1} \mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{R} \mathbf{Q}_{1}] \mathbf{Y}}.$$

$$(3.66)$$

De forma análoga, para testar a hipótese $H_0: \beta_2^{wg} = 0$, tem-se a estatística

$$t^{2} = \frac{\mathbf{r}_{y2}^{\top} \mathbf{r}_{x_{2,2}} (\mathbf{r}_{x_{2,2}}^{\top} \mathbf{r}_{x_{2,2}})^{-1} \mathbf{r}_{x_{2,2}}^{\top} \mathbf{r}_{y2}}{\mathbf{r}_{y2}^{\top} [\mathbf{I} - \mathbf{r}_{x_{2,2}} (\mathbf{r}_{x_{2,2}}^{\top} \mathbf{r}_{x_{2,2}})^{-1} \mathbf{r}_{x_{2,2}}^{\top}] \mathbf{r}_{y2}},$$
(3.67)

e substituindo os respectivos resíduos chega-se na expressão

$$Fg = \frac{\mathbf{Y}^{\top} \mathbf{Q}_{1} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\top} \mathbf{Q}_{1} \mathbf{X}_{2} (\mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\top} \mathbf{Q}_{1} \mathbf{X}_{2})^{-1} \mathbf{X}_{2}^{\top} \mathbf{Q}_{1} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\top} \mathbf{Q}_{1} \mathbf{Y}}{\mathbf{Y}^{\top} [\mathbf{Q}_{1} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\top} \mathbf{Q}_{1} \mathbf{Z} \mathbf{Q} \mathbf{Z}^{\top} \mathbf{$$

As estatísticas Fe e Fg são divididas pelos respectivos graus de liberdade, 1 grau de liberdade no numerador e (n-1) graus de liberdade no denominador. Observando a Figura 3.1, no caso de estudos Genéticos esta decomposição é relevante já que, por meio desta, é possível encontrar variáveis preditoras que além de poderem ser selecionadas para inclusão no modelo linear misto, podem ser escolhidas aquelas em que a parte poligênica (associada ao efeito aleatório) na estimativa do parâmetro β_2 é mais relevante. Este tema será discutido com mais detalhes no Capítulo 5 que trata sobre simulações realizadas.

Do que foi discutido anteriormente, conclui-se que a decomposição do gráfico da variável adicionada pode ter a vantagem de mostrar quais estruturas de covariância de um modelo lineal misto são mais influenciadas pela inclusão de um efeito fixo.

3.4 Índice para discriminar o efeito de uma variável preditora

No conjunto de variáveis preditoras a serem analizadas como efeitos fixos, precisa-se discriminar aquelas variáveis que afetam a parte aleatória de forma predominantemente, moderada ou baixa. As que afetam a parte aleatória de uma forma predominantemente moderada, são capturadas quando o parâmetro β_2 é significante e por meio da partição representada na Figura 3.1. No problema em estudo, é possível que algumas variáveis não tenham efeito significante sobre a variável resposta \mathbf{Y} e no entanto estas afetem o componente aleatório e seja possível, ao serem analisadas no nível de unidades amostrais

(vetor de respostas intraunidades), encontrar unidades amostrais que sejam caudais que possam ser analisadas em detalhes. Nesse sentido, a seguir é proposto um índice que torna mais operacional a discriminação das variáveis de interesse, sendo que em geral os bancos de dados a serem analisados são de altísima dimensão.

Em primeiro lugar, considere as estruturas dos pesos w_g e w_e descritas nas equações (3.57) e (3.58) especificamente dadas por:

$$w_g = \frac{\mathbf{X}_2^{\top} \mathbf{Q}_1 \mathbf{Z} \mathbf{D} \mathbf{Z}^{\top} \mathbf{Q}_1 \mathbf{X}_2}{\mathbf{X}_2^{\top} \mathbf{Q}_1 \mathbf{X}_2},$$
(3.69)

$$w_e = \frac{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{R} \mathbf{Q}_1 \mathbf{X}_2}{\mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2}.$$
 (3.70)

Os valores de w_g e w_e são obtidos a partir das estimativas dos parâmetros envolvidos no ajuste do modelo completo dado na equação (3.23). Também, nota-se que a matriz \mathbf{Q}_1 dada em (3.25) depende da matrix \mathbf{X}_1 .

Desta forma, analogamente, podemos obter medidas da parte aleatória e residual no modelo reduzido dado na equação (3.22). As respectivas medidas podem ser obtidas a partir das estimativas envolvidas no ajuste deste modelo. As expressões podem ser definidas por:

$$w_{g0} = \frac{\mathbf{X}_{2}^{\top} \mathbf{Q}_{0} \mathbf{Z} \mathbf{D}_{0} \mathbf{Z}^{\top} \mathbf{Q}_{0} \mathbf{X}_{2}}{\mathbf{X}_{2}^{\top} \mathbf{Q}_{0} \mathbf{X}_{2}},$$
(3.71)

$$w_{e0} = \frac{\mathbf{X}_2^{\top} \mathbf{Q}_0 \mathbf{R}_0 \mathbf{Q}_0 \mathbf{X}_2}{\mathbf{X}_2^{\top} \mathbf{Q}_0 \mathbf{X}_2},$$
 (3.72)

com,

$$\mathbf{Q}_0 = \mathbf{M}_0 - \mathbf{M}_0 \mathbf{X}_1 \left(\mathbf{X}_1^{\mathsf{T}} \mathbf{M}_0 \mathbf{X}_1 \right)^{-1} \mathbf{X}_1^{\mathsf{T}} \mathbf{M}_0, \tag{3.73}$$

em que a matriz \mathbf{M}_0 é definida de forma análoga à matriz \mathbf{M} dada na expressão (3.9), só que a partir das estimativas do ajuste do modelo reduzido.

Assim, os efeitos aleatórios e/ou residuais causados pela adição de uma determinada variável preditora sobre \mathbf{Y} são dados pelas respectivas diferenças

$$d_q = w_{q0} - w_q, (3.74)$$

е

$$d_e = w_{e0} - w_e. (3.75)$$

Considere o sistema de equações:

$$w_{q0} + w_{e0} = 1 (3.76)$$

$$w_q + w_e = 1. (3.77)$$

Dadas as restrições $w_g + w_e = 1$ e $w_{g0} + w_{e0} = 1$, tem-se que ao fazer a diferença, d_g é o inverso de d_e ($d_g + d_e = 0$), sendo suficiente fazer análises sobre um índice somente. Tomando como referência d_g , este índice pode ser positivo, zero ou negativo e cuja interpretação é descrita a seguir,

$$d_{gj} = \begin{cases} positivo, & \text{a } j - \acute{e}sima \text{ variável preditora teve efeito na parte aleatória,} \\ 0, & \text{a } j - \acute{e}sima \text{ variável preditora não teve nenhum efeito,} \\ negativo, & & \text{a } j - \acute{e}sima \text{ variável preditora teve efeito na parte residual.} \end{cases}$$
 (3.78)

j=1,2,...K, sendo K o número de variáveis preditoras em questão.

Nas análises do conjunto de variáveis preditoras, o índice d_g positivo discriminará aquelas variáveis de interesse, principalmente porque essas variáveis tem maior influência na parte aleatória do modelo liner misto.

Os coeficientes d_g e d_e serão analizados no Capítulo 5 por meio de vários cenários de simulação.

3.5 Índice para discriminar famílias influentes

É possível ainda decompor o gráfico da variável adicionada como uma combinação linear dos efeitos que cada unidade amostral (vetor de resposta intraunidades) exerce tanto no

componente aleatório quanto no componente residual. Nesse intuito, pode-se escrever os vetores obtidos na partição (equações (3.41) e (3.42)) de tal forma que os seus primeiros n_1 componentes correspondem as respostas da primeira unidade, os seguintes n_2 componentes correspondem às respostas da segunda unidade, e assim sucessivamente até os ultimos n_F componentes os quais correspondem à unidade F, ou seja,

$$\mathbf{r}_{v1}^{\top} = (\mathbf{r}_{y1,1}, \mathbf{r}_{y1,2}, \dots, \mathbf{r}_{y1,f}, \dots, \underline{\mathbf{r}}_{v1,F}), \tag{3.79}$$

$$\mathbf{r}_{y2}^{\top} = (\mathbf{r}_{y2,1}, \mathbf{r}_{y2,2}, \dots, \mathbf{r}_{y2,f}, \dots, \mathbf{r}_{y2,F}),$$
 (3.80)

$$\mathbf{r}_{x_{2.1}}^{\top} = (\mathbf{r}_{x_{2.1},1}, \mathbf{r}_{x_{2.1},2}, \dots, \mathbf{r}_{x_{2.1},f}, \dots \mathbf{r}_{x_{2.1},F}),$$
 (3.81)

$$\mathbf{r}_{x_{2,2}}^{\top} = (\mathbf{r}_{x_{2,2},1}, \mathbf{r}_{x_{2,2},2}, \dots \mathbf{r}_{x_{2,2},f}, \dots \mathbf{r}_{x_{2,2},F}).$$
 (3.82)

Das equações (3.79) e (3.81), tomando os resíduos da f-ésima unidade obtem-se o coeficiente angular ajustado gerado pela regressão dos correspondentes vetores, cuja expressão, sob as premissas clássicas, é dada por:

$$\hat{eta}_{2,f}^e = (\mathbf{r}_{x_{2.1},f}^ op \mathbf{r}_{x_{2.1},f})^{-1} \mathbf{r}_{x_{2.1},f}^ op \mathbf{r}_{y_{1},f} = rac{\mathbf{r}_{x_{2.1},f}^ op \mathbf{r}_{y_{1},f}}{\mathbf{r}_{x_{2.1},f}^ op \mathbf{r}_{x_{2.1},f}}.$$

De forma análoga, para o componente aleatório, tomando as equações (3.80) e (3.82), o coeficiente angular ajustado para cada unidade é dado por:

$$\hat{\beta}_{2,f}^g = (\mathbf{r}_{x_{2.2},f}^{\top} \mathbf{r}_{x_{2.2},f})^{-1} \mathbf{r}_{x_{2.2},f}^{\top} \mathbf{r}_{y_{2},f} = \frac{\mathbf{r}_{x_{2.2},f}^{\top} \mathbf{r}_{y_{2},f}}{\mathbf{r}_{x_{2.2},f}^{\top} \mathbf{r}_{x_{2.2},f}}.$$

Como ilustração, na Figura 3.2 para o componente aleatório, apresentam-se retas de regressão ajustadas a partir dos dados das unidades amostrais estudadas. Nota-se que com esta técnica é possível discriminar unidades amostrais caudais, a saber, unidades amostrais que apresentam altos coeficientes angulares (em módulo). Essas unidades amostrais caudais merecem um estudo mais detalhado já que podem apresentar alterações na variável

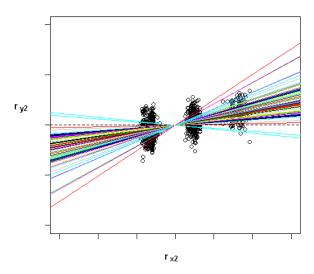


Figura 3.2 Decomposição do efeito aleatório por unidade amostral.

resposta observada que seguem um padrão de variação específico e diferente do restante da média da população.

Os estimadores dos coeficientes de regressão $\hat{\beta}_2^e$ e $\hat{\beta}_2^g$ podem ser escritos como uma combinação linear dos coeficientes para cada unidade amostral, especificamente da seguinte forma:

$$\hat{\beta}_{2}^{e} = \frac{\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{y_{1}}}{\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{x_{2.1}}} = \frac{\mathbf{r}_{x_{2.1,1}}^{\top} \mathbf{r}_{x_{2.1,1}}}{\mathbf{r}_{x_{2.1}}^{\top} \hat{\beta}_{2,1}^{e}} + \frac{\mathbf{r}_{x_{2.1,2}}^{\top} \mathbf{r}_{x_{2.1,2}}}{\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{x_{2.1}}} \hat{\beta}_{2,2}^{e} + \dots + \frac{\mathbf{r}_{x_{2.1,F}}^{\top} \mathbf{r}_{x_{2.1,F}}}{\mathbf{r}_{x_{2.1}}^{\top} \mathbf{r}_{x_{2.1}}} \hat{\beta}_{2,F}^{e}, \quad (3.83)$$

$$\hat{\beta}_{2}^{g} = \frac{\mathbf{r}_{x_{2,2}}^{\top} \mathbf{r}_{y_{2}}}{\mathbf{r}_{x_{2,2}}^{\top} \mathbf{r}_{x_{2,2}}} = \frac{\mathbf{r}_{x_{2,2,1}}^{\top} \mathbf{r}_{x_{2,2,1}}}{\mathbf{r}_{x_{2,2}}^{\top} \hat{\beta}_{2,1}^{g}} + \frac{\mathbf{r}_{x_{2,2,2}}^{\top} \mathbf{r}_{x_{2,2,2}}}{\mathbf{r}_{x_{2,2}}^{\top} \hat{\beta}_{2,2}^{g}} + \dots + \frac{\mathbf{r}_{x_{2,2,F}}^{\top} \mathbf{r}_{x_{2,2},F}}{\mathbf{r}_{x_{2,2}}^{\top} \hat{\beta}_{2,F}^{g}}.$$
(3.84)

Em consequência, o efeito total da variável adicionada em estudo, $\hat{\beta}_2$, pode ser escrito em função da contribuição de cada unidade amostral como segue:

$$\hat{\beta}_{2} = w_{g} \hat{\beta}_{2}^{g} + w_{e} \hat{\beta}_{2}^{e}$$

$$= w_{g} \sum_{f=1}^{F} w_{g,f} \hat{\beta}_{2,f}^{g} + w_{e} \sum_{f=1}^{F} w_{g,f} \hat{\beta}_{2,f}^{e}.$$
(3.85)

Note que o valor $w_{g,f}\hat{\beta}_{2,f}^g$ caracteriza cada família na sua respectiva contribuição para

o valor obtido de $\hat{\beta}_2^g$, podendo-se encontrar famílias caudais. Assim, é possível definir um índice que caracteriza as famílias ou unidades amostrais envolvidas no estudo,

$$I_f = w_{g,f} \hat{\beta}_{2,f}^g \; ; f = 1, ..., F.$$
 (3.86)

No Capítulo 5 do estudo de simulação será retomado este índice no intuito de encontrar famílias caudais.

A grande vantagem da partição (3.85) é que $\beta_{2,f}^g$ e $\beta_{2,f}^e$ podem ser significantes mesmo que β_2 (e, portanto β_2^g e β_2^e) não o sejam.

3.6 Geometria tridimensional do problema da variável adicionada

Nesta seção, apresenta-se um esquema da geometria tridimensional associada ao problema da variável adicionada no modelo misto, com dois componentes de variância. Em primeiro lugar, na Figura 3.3 apresenta-se geometricamente a teoria clássica que define a estatística F, dada na equação (3.31) para testar a possível significância do efeito da variável adicionada. Logo, na Figura 3.4 são descritas as partições dos resíduos obtidos do modelo reduzido (equação (3.22)) e do modelo completo (equação (3.23)) do ponto de vista geométrico.

Na Figura 3.3 observa-se que a projeção de \mathbf{Y} sobre o modelo completo é dado por $\mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2$ com vetor de resíduos $\boldsymbol{\xi}_C$.

A projeção de \mathbf{Y} sobre o modelo reduzido (o qual contém somente \mathbf{X}_1) é dado por $\mathbf{X}_1\hat{\boldsymbol{\beta}}_{1R}$ com vetor de resíduos $\boldsymbol{\xi}_R$.

Os vetores $\boldsymbol{\xi}_R$ e $\boldsymbol{\xi}_C$ são ortogonais às variáveis \mathbf{X}_1 pois, $\mathbf{Q}_1\mathbf{X}_1=0$ e $\mathbf{Q}\mathbf{X}_1=0$ e, portanto, também se verifica a ortoganalidade para a diferença $(\boldsymbol{\xi}_R-\boldsymbol{\xi}_C)$. Esta diferença é o resíduo que sobra depois da projeção de $\mathbf{X}_1\hat{\boldsymbol{\beta}}_1+\mathbf{X}_2\hat{\boldsymbol{\beta}}_2$ sobre o espaço de variáveis \mathbf{X}_1 , em que $\boldsymbol{\xi}_R-\boldsymbol{\xi}_C=\mathbf{M}^{-1/2}\mathbf{Q}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2$. Como o vetor $\boldsymbol{\xi}_C$ é ortogonal a \mathbf{X}_1 e \mathbf{X}_2 é também ortogonal a $\mathbf{M}^{-1/2}\mathbf{Q}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2$ então pelo Teorema de Pitágoras é obtida a equação:

$$\boldsymbol{\xi}_{B}^{\top}\boldsymbol{\xi}_{B} = \boldsymbol{\xi}_{C}^{\top}\boldsymbol{\xi}_{C} + \hat{\boldsymbol{\beta}}_{2}^{\top}\mathbf{X}_{2}^{\top}\mathbf{Q}_{1}\mathbf{M}^{-1}\mathbf{Q}_{1}\mathbf{X}_{2}\hat{\boldsymbol{\beta}}_{2}$$

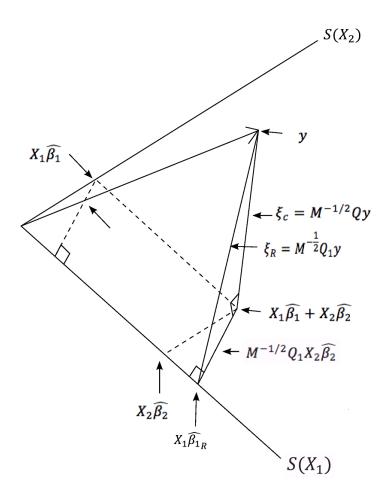


Figura 3.3 Geometria da estatística F.

e, dado que $\mathbf{Q}_1\mathbf{M}^{-1}\mathbf{Q}_1=\mathbf{Q}_1$, (equação (3.26)) tem-se:

$$\boldsymbol{\xi}_R^{\mathsf{T}} \boldsymbol{\xi}_R - \boldsymbol{\xi}_C^{\mathsf{T}} \boldsymbol{\xi}_C = \hat{\boldsymbol{\beta}}_2^{\mathsf{T}} \mathbf{X}_2^{\mathsf{T}} \mathbf{Q}_1 \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2. \tag{3.87}$$

A expressão (3.87) corresponde ao numerador da estatística F dada na expressão (3.31) para testar $\boldsymbol{\beta}_2 = 0$, o qual é equivalente a testar se a contribuição $\mathbf{M}^{-1/2}\mathbf{Q}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2$ é significante, isto é, se o comprimento de $\boldsymbol{\xi}_R$ é significativamente maior que o comprimento de $\boldsymbol{\xi}_C$ ou equivalentemente se a expressão $\boldsymbol{\xi}_R^{\top}\boldsymbol{\xi}_R - \boldsymbol{\xi}_C^{\top}\boldsymbol{\xi}_C$ difere significativamente de zero.

Na seção 3.3 foi utilizada a matriz \mathbf{A} definida na equação (3.40) no intuito de fazer a decomposição do gráfico da variável adicionada, substituindo a matriz $\mathbf{M}^{-1/2}$ pela matriz \mathbf{A} .

Na Figura 3.3, mediante a substituição da matriz $\mathbf{M}^{-1/2}$ pela matriz \mathbf{A} , é feita a decom-

posição dos resíduos obtidos do modelo reduzido e do modelo completo. Tal decomposição pode ser escrita como:

$$\boldsymbol{\xi}_{R} = \mathbf{A}\mathbf{Q}_{1}\mathbf{Y} = \begin{pmatrix} \mathbf{R}^{1/2}\mathbf{Q}_{1}\mathbf{Y} \\ \mathbf{D}^{1/2}\mathbf{Z}^{\top}\mathbf{Q}_{1}\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\xi}_{R1} \\ \boldsymbol{\xi}_{R2} \end{pmatrix}, \tag{3.88}$$

$$\boldsymbol{\xi}_C = \mathbf{AQY} = \begin{pmatrix} \mathbf{R}^{1/2} \mathbf{QY} \\ \mathbf{D}^{1/2} \mathbf{Z}^{\top} \mathbf{QY} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\xi}_{C1} \\ \boldsymbol{\xi}_{C2} \end{pmatrix}. \tag{3.89}$$

Nota-se na Figura 3.4 que o vetor $\boldsymbol{\xi}_C$ é decomposto nos vetores ortogonais $\boldsymbol{\xi}_{C1}$ e $\boldsymbol{\xi}_{C2}$ dados em (3.89) e, também o vetor $\boldsymbol{\xi}_R$ é decomposto nos vetores ortogonais $\boldsymbol{\xi}_{R1}$ e $\boldsymbol{\xi}_{R2}$ dados em (3.88). Fazendo a diferença das respectivas componentes pertencentes ao modelo reduzido e modelo completo tem-se:

$$\boldsymbol{\xi}_{R1} - \boldsymbol{\xi}_{C1} = \mathbf{R}^{1/2} (\mathbf{Q}_1 - \mathbf{Q}) \mathbf{Y} = \mathbf{R}^{1/2} \mathbf{Q}_1 \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2,$$

$$\boldsymbol{\xi}_{R2} - \boldsymbol{\xi}_{C2} = \mathbf{D}^{1/2} \mathbf{Z}^{\top} (\mathbf{Q}_1 - \mathbf{Q}) \mathbf{Y} = \mathbf{D}^{1/2} \mathbf{Z}^{\top} \mathbf{Q}_1 \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2.$$

Estes vetores (obtidos das respectivas diferenças) são paralelos aos vetores obtidos da decomposição de $\mathbf{M}^{-1/2}\mathbf{Q}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2$ em suas componentes ortogonais $(\mathbf{R}^{1/2}\mathbf{Q}_1\mathbf{X}_2\hat{\boldsymbol{\beta}}_2)$ e finalmente, pelo teorema de Pitágoras temos a igualdade:

$$\hat{\boldsymbol{\beta}}_2 \mathbf{X}_2^{\top} \mathbf{Q}_1 \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 = \hat{\boldsymbol{\beta}}_2 \mathbf{X}_2^{\top} \mathbf{Q}_1 \mathbf{R} \mathbf{Q}_1 \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{\beta}}_2 \mathbf{X}_2^{\top} \mathbf{Q} \mathbf{Z} \mathbf{D} \mathbf{Z}^{\top} \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2.$$

No caso em que \mathbf{X}_2 tenha somente uma coluna , $\hat{\beta}_2$ será uma constante, podendo chegar na equação

$$\mathbf{X}_2^{\top}\mathbf{Q}_1\mathbf{X}_2 = \mathbf{X}_2^{\top}\mathbf{Q}_1\mathbf{R}\mathbf{Q}_1\mathbf{X}_2 + \mathbf{X}_2^{\top}\mathbf{Q}_1\mathbf{Z}\mathbf{D}\mathbf{Z}^{\top}\mathbf{Q}_1\mathbf{X}_2$$

a qual é equivalente a ter a expressão $w_g + w_e = 1$ em que w_g e w_g , definidas em (3.57) e (3.58), respectivamente, são as contribuições aleatória e residual para a significância do parâmetro $\hat{\beta}_2$.

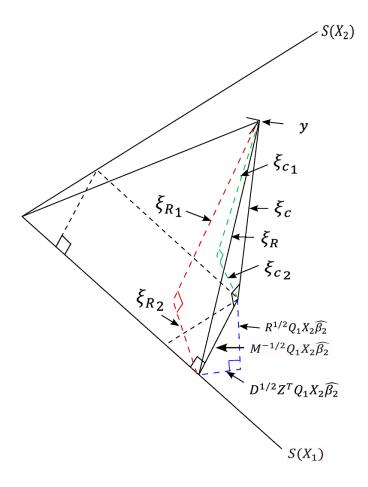


Figura 3.4 Geometria do teste F e decomposição dos resíduos.

Capítulo 4

Estudo do efeito de um bloco de SNPs: Abordagem haplotípica

Os seres humanos são considerados organismos diploides, pois os seus cromossomos homólogos estão pareados, sendo que um membro do par é vindo do pai e o outro vindo da mãe. Para uma específica posição ou região do genoma, a união das informações dos pais formam o genótipo do indivíduo naquela posição e a sequência de informações de um dos pais é denominada haplótipo (Ott, 1991). Portanto, cada indivíduo tem um par de haplótipos homólogos que podem ou não ser idênticos. Imagine os dados genotípicos de três SNPs ordenados (e localizados em um mesmo cromossomo): A_1A_1, A_2a_2, A_3A_3 . Neste caso, a partir dos dados genotípicos os dois haplótipos correspondentes são obtidos diretamente: $A_1A_2A_3$ e $A_1a_2A_3$ mas em uma sequência basta a ocorrência de dois ou mais genótipos heterozigotos para que a fase haplotípica seja desconhecida e não possa ser obtida diretamente. Procedimentos de genotipagem não podem separar os dois cromossomos homólogos para se obter diretamente os haplótipos (contendo diferentes locos) de modo que somente os genótipos do indivíduo são diretamente observados. Assim, a inferência de haplótipos busca extrair as informações do haplótipo a partir da observação de dados do genótipo do indivíduo. Para a codificação dos haplótipos de um indivíduo precisa-se de um conjunto de marcadores moleculares ordenados, neste caso conjuntos de SNPs que serão denominados "blocos".

Os blocos conformados pelo conjunto de SNPs têm um papel fundamental no risco de um indivíduo contrair ou responder a uma doença. É mais provável que o efeito de SNPs na regulação de doenças seja em blocos de SNPs adjacentes do que individualmente. Se

fosse possível identificar associações específicas de SNPs então seria possível detectar a evolução da população humana com maior precisão e aumentar a capacidade de mapear genes de doença. Por tal motivo o estudo de haplótipos é relevante em estudos genéticos quando são utilizados dados de família e plataformas de SNPs.

Neste capítulo, o efeito de um bloco de SNPs é estudado via a abordagem haplotípica, isto é, via modelos nos quais algumas colunas da matriz de planejamento **X** estão definidas segundo parametrizações específicas. Nota-se que, uma alternativa para se estudar o efeito de um bloco de SNPs pode ser definida em termos dos seus efeitos individuais principais e de interação (este não será o foco do presente trabalho). Além disso, na presente proposta de tese, um dos objetivos é estudar técnicas de busca por blocos de SNPs que gerem covariância entre os vetores correspondentes a indivíduos parentes. Uma forma é considerar os possíveis haplótipos que blocos formados por um número determinado de SNPs possam gerar.

4.1 Relação entre Genótipos e Haplótipos

Considerando locos dialélicos, segue na Tabela 4.1 a relação entre genótipos e haplótipos. Específicamente para K SNPs, nas colunas 2 e 3 tem-se 3^K genótipos e 2^K possíveis haplótipos, já na coluna 4, H^+ representa o número de possíveis configurações haplotípicas discriminando se determinado haplótipo foi doado pelo pai ou pela mãe, finalmente na coluna 5 tem-se H^- o número de configurações desconsiderando qual foi o progenitor doador.

Tabela 4.1 Genótipos e Haplótipos segundo o número de SNPs.

SNPs	Genótipos	Haplótipos	H^+	H^-
1	3^1	2^1	2^2	3
2	3^2	2^2	2^{4}	10
•				
k	3^k	2^k	2^{2k}	$2^k(2^k+1)/2$

Muitos experimentos em Genética são realizados para estudar os efeitos produzidos por dois ou mais genes sobre uma mesma variável resposta quantitativa e, o padrão de variação da resposta é investigado sob combinações dos níveis dos fatores genéticos. Por exemplo, se K=2 SNPs estão sendo estudados, SNP A e SNP B, cada um em três níveis genotípicos (AA, Aa, aa e BB, Bb, bb, respectivamente), então a variável resposta será avaliada em indivíduos pertencentes às 9 possíveis combinações dos níveis dos dois SNPs (ver Tabela 4.2). Note, sob esta formulação que o efeito dos dois SNPs poderia ser estudado via um modelo linear contendo os efeitos principais do SNP A (com 2 graus de liberdade), do SNP B (com 2 graus de liberdade) e os correspondentes efeitos de interação (com 4 graus de liberdade).

Tabela 4.2 Classes genotípicas para dois SNPs, SNP A e SNP B.

	SNP B			
SNP A	BB	Bb	bb	
$\mathbf{A}\mathbf{A}$	AA	AA	AA	
	BB	Bb	bb	
Aa	Aa	Aa	Aa	
	BB	Bb	bb	
aa	aa	aa	aa	
	BB	Bb	bb	

Contudo, na abordagem haplotípica, ao serem considerados os dois SNPs pode-se ter para cada indivíduo 2 sequências com informações provenientes dos dois progenitores, em que cada sequência é chamada de haplótipo.

4.1.1 Ambiguidades na formação de haplótipos

No estudo de haplótipos, quando um indivíduo é heterozigoto nos dois SNPs (genótipos Aa e Bb), existe ambiguidade na conformação dos haplótipos. Em geral, para K SNPs existirão 2^{K-1} pares ambíguos. Uma possível solução para o caso em que haja ambiguidade, é assumir uma configuração específica, Cis ou Trans, como verdadeira para formar

o conjunto de haplótipos (Conti & Gauderman, 2004). Especificamente, um indivíduo duplo heterozigoto pode ter os alelos arranjados de duas maneiras nos cromossomos como apresentado na Figura 4.1. Os alelos A e B se situam em um cromossomo, enquanto os alelos a e b se situam no homólogo correspondente. Esse tipo de arranjo é chamado de Cis. O alelo A e o alelo b se situam em um cromossomo, enquanto o alelo a e o alelo b, se situam no homólogo correspondente. Esse tipo de arranjo é chamado de Trans. Em geral, denomina-se A(B) o alelo menos frequente, o qual, possivelmente, estará envolvido com a doença (alelo de risco).

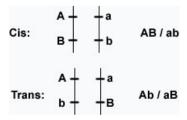


Figura 4.1 Configurações Cis e Trans.

Outra forma de contornar o problema de ambiguidades é dar uma ponderação para cada configuração haplotípica dependendo dos dados e da finalidade das análises. Para as análises realizadas neste trabalho, nos casos de ambiguidades, será assumida a configuração Cis.

4.2 Configurações de haplótipos para um indivíduo

Na Figura (4.2a) ilustra-se a sequência dos dois SNPs nos três níveis genotípicos, como descrito na Tabela 4.2. Na Figura 4.2b ilustram-se os 4 haplótipos que um progenitor pode doar para o filho, sem levar em consideração o progenitor, um indivíduo pode ter 10 possibilidades de configurações haplotípicas, essas possibilidades são específicamente ilustradas na Figura 4.3. Note que os genótipos associados às configurações 4 e 6 conduzem à mesma configuração haplotípica sob a estrutura Cis e são equivalentes sob a configuração Cis (a qual é adotada neste trabalho). Ao realizar um estudo via configurações haplotípicas,

um modelo bastante usado é considerar como haplótipo de risco o haplótipo que tem na sequência alelos de risco. O indivíduo que tem a configuração 1, carrega 2 haplótipos de risco; quem tiver as configurações 2, 3, 4 ou 6 carrega 1 haplótipo de risco; já quem tiver as configurações 5, 7, 8, 9 ou 10 não carrega haplótipos de risco.

Para o caso em que sejam estudados blocos de K=3 SNPs, ilustram-se as possíveis configurações na Figura 4.4. Neste caso, tem-se 8 haplótipos, em que cada indivíduo poderá ter uma das 36 possíveis configurações haplotípicas. Com 4 pares ambíguos e assumindo a configuração Cis, o total é de 34 possibilidades, em que cada indivíduo pode carregar 2, 1 ou nenhum haplótipo de risco.

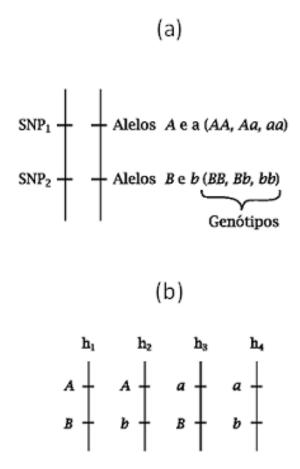


Figura 4.2 Haplótipos para 2 SNPs.

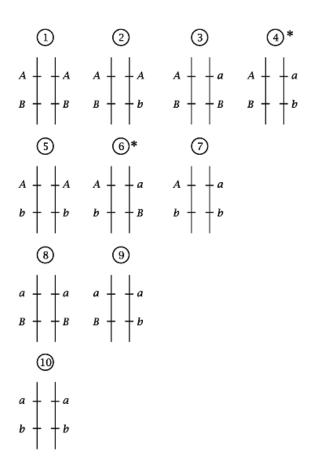


Figura 4.3 Configurações haplotípicas para 1 indivíduo quando considerados 2 SNPs.

4.3 Estudo de efeitos de variáveis haplotípicas

A abordagem via haplótipos para dados de famílias e marcadores moleculares SNPs não tem sido considerada na literatura. No entanto, vários autores que têm explorado esta abordagem haplotípica em modelos que consideram indivíduos não relacionados argumentam que pode haver um aumento de poder quando comparada com a análise de marcadores individuais (Akey et al., 2001).

Uma possível abordagem de estudo via haplótipos é estudar os efeitos produzidos por dois o mais genes sobre a variável resposta, em que esta é avaliada em indivíduos pertencentes às 10 possíveis configurações haplotípicas, como ilustrado na Figura 4.3 para o caso de blocos de 2 SNPs. Para o caso de três SNPs, a variável resposta será avaliada nas 36

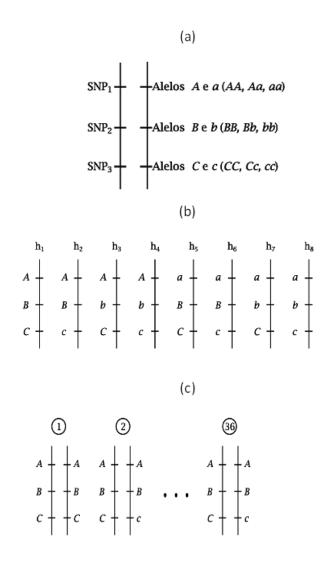


Figura 4.4 Haplótipos gerados a partir de 3 SNPs.

possíveis configuarações haplotípicas e assim sucessivamente, dependendo do número de SNPs no estudo. Neste caso, o efeito de haplótipos pode ser modelado por meio de uma variável preditora com 10 e 36 níveis, respectivamente.

Uma outra possível formulação em termos de haplótipos é a seguinte: quando analisado cada SNP de forma independente de acordo com a codificação dada no Capítulo 1 na equação (2.6) cada indivíduo carrega 2, 1 ou 0 alelos de risco correspondente aos genótipos (AA, Aa, aa). Do mesmo modo, na formulação haplotípica codificam-se os dados considerando se o indivíduo carrega 2, 1 ou 0 haplótipos de risco. Nas Tabelas 4.3 e 4.4

encontram-se as respectivas codificações, com \mathbf{X}_{hap} denotando a nova variável preditora em função de haplótipos. Esta nova variável é estudada como um efeito fixo em um modelo misto, mais específicamente é estudada como uma variável adicionada ao modelo misto poligênico por meio da teoria desenvolvida no Capítulo 3. Esta abordagem é considerada neste trabalho.

Num	SNP1	SNP2		\mathbf{X} hap
1	2	2	\rightarrow	2
2	2	1	\rightarrow	1
3	2	0	\rightarrow	0
4	1	2	\rightarrow	1
5	1	1	\rightarrow	1
6	1	0	\rightarrow	0
7	0	2	\rightarrow	0
8	0	1	\rightarrow	0
9	0	0	\rightarrow	0

Tabela 4.3 Configurações haplotípicas para blocos de 2 SNPs.

Assim como é possível estudar o efeito da variável \mathbf{X}_{hap} como efeito fixo, o efeito de haplótipo pode também ser estudado como efeito aleatório. Na seção 2.1 do Capítulo 2 foi apresentado o modelo Oligogênico em (2.7) para o qual foi definida a matriz $\mathbf{\Pi}$ com a informação IBD com valores 1, 1/2 ou 0 dependendo se dois indivíduos compartilham 2, 1, ou nenhum alelo de risco, respectivamente, para um determinado marcador (em geral, do tipo microsatélite). No caso de análises haplotípicas, ao se incluir o efeito de haplótipo como efeito aleatório, pode-se substituir a matriz $\mathbf{\Pi}$ pela matriz $\mathbf{\Pi}_h$ cujos elementos são descritos a seguir:

$$\Pi_{hij} = \begin{cases}
1, & \text{se } i = j, \\
1, & \text{se } i \in j \text{ compartilham 2 haplótipos em comum,} \\
1/2, & \text{se } i \in j \text{ compartilham 1 hapótipo em comum,} \\
0, & \text{se } i \in j \text{ não compartilham haplótipos comuns.}
\end{cases} (4.1)$$

Para o mapeamento de genes em termos de codificações haplotípicas uma questão relevante é como encontrar os "blocos" de SNPs para serem estudados no nível de haplótipos, sendo que o espaço é de alta dimensão e, além disso, o efeito individual do SNP é "pe-

Tabela 4.4 Configurações haplotípicas para blocos de 3 SNPs.

	SNP1	SNP2	SNP3		\mathbf{X} hap
1	2	2	2	\rightarrow	2
2	2	2	1	\rightarrow	1
3	2	2	0	\rightarrow	0
4	2	1	2	\rightarrow	1
5	2	1	1	\rightarrow	1
6	2	1	0	\rightarrow	0
7	2	0	2	\rightarrow	0
8	2	0	1	\rightarrow	0
9	2	0	0	\rightarrow	0
10	1	2	2	\rightarrow	1
11	1	2	1	$\begin{array}{c} \rightarrow \\ \rightarrow \end{array}$	1
12	1	2	0	\rightarrow	0
13	1	1	2	\rightarrow	1
14	1	1	1	\rightarrow	1
15	1	1	0	\rightarrow	0
16	1	0	2	\rightarrow	0
17	1	0	1	\rightarrow	0
18	1	0	0	\rightarrow	0
19	0	2	2	\rightarrow	0
20	0	2	1	\rightarrow	0
21	0	2	0	\rightarrow	0
22	0	1	2	$\begin{array}{c} \rightarrow \\ \rightarrow \end{array}$	0
23	0	1	1	\rightarrow	0
24	0	1	0	\rightarrow	0
25	0	0	2	\rightarrow	0
26	0	0	1	\rightarrow	0
27	0	0	0	\rightarrow	0

queno". Para que um conjunto de SNPs forme um haplótipo, estes têm que estar em Desequilíbrio de Ligação (conceito que é especificado no Apêndice C). Assim, para encontrar os blocos de interesse na pesquisa haplotípica, bastaria encontrar os "blocos" (subconjuntos de SNPs adjacentes) de variáveis correlacionadas. Mas para o foco do presente trabalho propõe-se encontrar especificamente aqueles "blocos" que tenham efeito sobre a doença em estudo, principalmente associados com o componente aleatório poligênico.

Da teoria clássica tem-se que se duas ou mais variáveis estão correlacionadas, estas manifestarão efeitos parecidos sobre a variável resposta em estudo. Assim, dentre o conjunto de SNPs a serem avaliados sob uma análise haplotípica, o interesse é encontrar SNPs consecutivos que tenham efeitos "parecidos" sobre a variável resposta **Y**. Isto é,

por meio da teoria da variável adicionada, apresentada no Capítulo 3, pode-se selecionar SNPs adjacentes que mostrem, por exemplo, β_f "significantes" para as mesmas famílias.

A decisão se o bloco de SNPs deve ser modelado como fator fixo (\mathbf{X}_{hap} , por exemplo) ou como fator aleatório ($\mathbf{\Pi}_h$) não é trivial, mas fica claro da teoria do gráfico da variável adicionada que se SNPs consecutivos mostram efeitos, principalmente associados ao componente residual ($\boldsymbol{\epsilon}$), o correspondente haplótipo deve ser modelado como fator fixo. Caso o bloco de SNPs mostre efeitos associados ao componente poligênico, então o efeito do haplótipo correspondente deve ser modelado como aleatório.

Capítulo 5

Estudo de simulação

Para o estudo de simulação foram consideradas 51 famílias de tamanhos diferentes, como apresentado na Tabela 5.1, com um total de 1486 indivíduos (média de 30 indivíduos por família). As estruturas foram escolhidas dos dados reais das famílias do Projeto Corações de Baependi (MG), do Laboratório de Genética e Cardiologia Molecular (InCor/USP). Como ilustração da estrutura familiar, apresenta-se na Figura 5.1 o heredograma correspondente à família 2 com o respectivo genótipo para cada indivíduo (codificado como 0, 1, 2). Todas as simulações estão baseadas na estrutura familiar destas 51 famílias como ilustrado nas primeiras 4 colunas da Tabela 1.1 no Capítulo de Introdução.

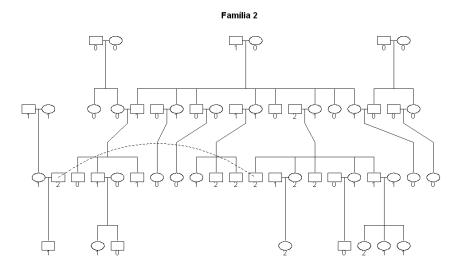


Figura 5.1 Heredograma de uma família com 4 gerações.

5.1 Geração das variáveis aleatórias

Para a geração das variáveis genéticas, ou seja os genótipos dos SNPs para os membros das 51 famílias, foi utilizado o pacote SimPed (Leal et al., 2005). O respectivo código utilizado encontra-se no Apêndice A. Inicialmente foram gerados 9 SNPs entre os quais, os SNPs 2 e 3 estão em LD¹ formando um bloco (haplótipos de 2 SNPs). Os SNPs 6, 7 e 8 também estão em LD formando blocos de 3 SNPs (haplótipos de 3 SNPs); já os SNPs 1, 4, 5 e 9 foram gerados independentemente verificando as condições genéticas requeridas para a execução do pacote SimPed. Também, no intuito de estabelecer algumas comparações relativamente ao efeito de SNPs, variáveis preditoras com agregado familiar², e variáveis preditoras aleatórias, sem agregado familiar, estas últimas foram geradas por meio de uma distribuição uniforme; estas variáveis são denotadas por v_1, v_2, v_3 e v_4 . Específicamente, $v_{ji} \sim U[0,1]$ em que $v_{ji} = 0,1,2$ se $v_{ji} \in [0,0,40], (0,40,0,75], (0,75,1]$ respectivamente, j = 1, 2, 3, 4 e i = 1, ..., 1486. Um esquema com a geração do banco de dados das variáveis preditoras é ilustrado na Figura 5.2. Assim, tem-se um conjunto de 13 variáveis preditoras a serem utilizadas nas diferentes simulações das variáveis resposta Y. A codificação corresponde a 0, 1, 2 de acordo com o genótipo dos indivíduos, como especificado na equação (2.6).

Na modelagem estatística, um desafio é explicar a variabilidade na variável resposta **Y** em termos da variabilidade nas covariáveis observadas, um modelo com efeitos aleatórios é uma forma de levar em consideração heterogeneidade não observada (Skrondal & Rabe-Hesketh, 2004).

No intuito de comparar o efeito sobre \mathbf{Y} de uma variável preditora com agregado familiar versus uma variável preditora sem agregado familiar, foi gerada a variável resposta \mathbf{Y}_1 com efeito do SNP 7 e a variável reposta \mathbf{Y}_2 com efeito da variável preditora aleatória v_2 . Especificamente, tem-se as seguintes distribuições multivariadas,

¹ LD, abreviação de Desequilíbrio de Ligação, do inglês *Linkage Desequilibrium*.

A agregação familiar numa dada característica acontece quando a semelhança dessa característica entre os seus membros é maior do que a esperada por mero acaso. Essa agregação reflete, na sua essência, a similaridade genética e cultural dentro de cada família.

Tabela 5.1 Tamanho das famílias usadas nos cenários de simulação.

Família	Tamanho	Família	Tamanho	Família	Tamanho
1	39	18	17	35	26
2	53	19	26	36	33
3	28	20	16	37	18
4	27	21	15	38	22
5	70	22	19	39	18
6	17	23	30	40	18
7	33	24	26	41	13
8	40	25	20	42	24
9	35	26	27	43	27
10	32	27	23	44	16
11	101	28	16	45	27
12	52	29	23	46	19
13	15	30	19	47	15
14	108	31	22	48	41
15	14	32	21	49	21
16	18	33	30	50	26
17	22	34	41	51	27

 $\mathbf{Y}_1 \sim N(\boldsymbol{\mu} + \beta_2(SNP7); 2\boldsymbol{\Phi}\sigma_g^2 + \mathbf{I}\sigma_e^2)$ e $\mathbf{Y}_2 \sim N(\boldsymbol{\mu} + \beta_2(v_2); 2\boldsymbol{\Phi}\sigma_g^2 + \mathbf{I}\sigma_e^2)$, com $\boldsymbol{\mu}$ o vetor de médias e $2\boldsymbol{\Phi}$ a matriz de parentesco descrita no Capítulo 2. Nas Figuras 5.3 e 5.4 os valores de \mathbf{Y}_1 e \mathbf{Y}_2 são graficados para cada família, as quais estão ordenadas de acordo a suas respectivas médias.

Na Figura 5.3, no caso em que a variável \mathbf{Y}_1 é gerada com efeito do SNP 7 (com agregado familiar), a heterogeneidade é um reflexo da variação das médias das famílias (linha tracejada) ao redor da média global (linha horizontal), resultante de correlações dentro das famílias porque, as suas respostas estão com tendência que direciona a maioria dos valores para o mesmo lado de sua média global. Observa-se que algumas famílias têm

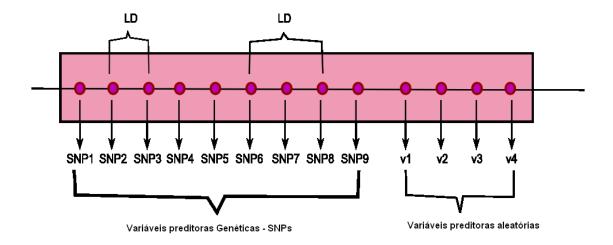


Figura 5.2 Variáves preditoras para o estudo de simulação.

as suas respectivas médias mais baixas, outras próximas e outras mais altas do que a média geral, o que sugere heterogeneidade entre famílias e correlação dentro da família.

Na Figura 5.4, em que \mathbf{Y}_2 foi gerada a partir da variável aleatória v_2 , percebe-se que as famílias não apresentam comportamentos muito diferenciados. Todas, em geral, têm o mesmo padrão de variação ao redor da média geral, percebendo-se que o valor da variável resposta de um indivíduo independe da sua família. Padrões de variação como os mostrados nas Figuras 5.3 e 5.4 são discutidos em (Skrondal & Rabe-Hesketh, 2004) para ilustrar o efeito de cluster em modelos mistos. Retoma-se esta discussão na Seção 5.5.

5.2 Cenários de simulação

A seguir, descrevem-se os cenários de simulação das variáveis resposta \mathbf{Y} que terão efeito das variáveis preditoras descritas na Figura 5.2. Os esquemas para os cenários de simulação são apresentados na Tabela 5.2. Inicialmente na linha 1, indicam-se as variâncias totais ($\sigma_y^2 = 2, 7, 15, 50, 100$) utilizadas nas simulações. Cada uma destas variâncias são subdivididas, por sua vez, em herdabilidades correspondentes a 8%, 32% e 70% como descrito na primeira coluna. Nas colunas 2, 7, 12, 17 e 22, a variância total está dividida nas variâncias poligênica (σ_g^2) e residual (σ_e^2). Para simular o efeito β_2 que uma variável

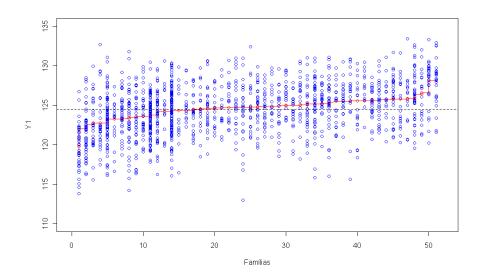


Figura 5.3 Valores de \mathbf{Y}_1 para 51 famílias as quais estão ordenadas de acordo com a respectiva média. A linha horizontal representa a média geral.

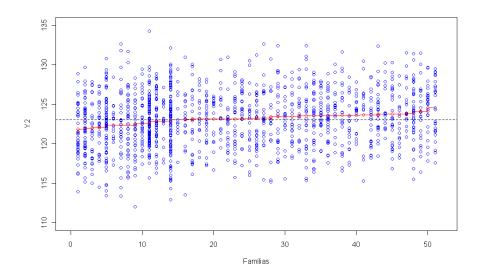


Figura 5.4 Valores de \mathbf{Y}_2 para 51 famílias as quas estão ordenadas de acordo com a respectiva média. A linha horizontal representa a média geral.

preditora tem sobre \mathbf{Y} , foram escolhidos os valores 0,1, 0,8, 1,5 e 3,0 descritos na linha 2 para cada variabilidade total σ_y^2 .

Tabela 5.2 Cenários de simulação.

	₀						
00	3,0						
$\sigma_y^2 = 100$	1,5						
	8,0						
	0,1						
		8	92	32	89	70	30
	3,0						
$\sigma_y^2 = 50$	2						
σ_{i}^{2}	8 1,5						
	0,1 0,8						
	cv 0,	4	46	16	34	35	15
	3,0	7	7				
15	••						
$\sigma_y^2 = 15$	1,5						
	0,1 0,8						
	0,1						
	cv	1,2	13,8	4,8	10,2	10,5	4,5
	3,0						
$\sigma_y^2 = 7$	1,5						
	8,0						
	0,1						
	CV	92,0	6,44	2,24	4,76	4,9	2,1
	3						
$\sigma_y^2 = 2$	1,5						
	0,1						
	cv $0,1$ $0,8$	0,16	0,84	32% 0,64	1,36	70% 1,4	9,0
1		%8					

cv:componentes de variância. Por exemplo para a herdabilidade de 32%, a variância total $\sigma_y^2 = 2$ é dividida nas variâncias, poligênica $\sigma_g^2 = 0,64 \text{ e residual } \sigma_e^2 = 1,36.$

De modo geral, \mathbf{Y}_{kl} k=1,2,3,4 ; l=1,2,...,15 representará a variável resposta que foi gerada a partir da l-ésima variância poligênica (σ_{gl}^2) , a l-ésima variância residual (σ_{el}^2) com efeito β_{2k} de uma determinada variável preditora, ou seja,

$$\mathbf{Y}_{kl} \sim N(\boldsymbol{\mu} + \boldsymbol{\beta}_{2k} \mathbf{X}_2; 2\boldsymbol{\Phi}\boldsymbol{\sigma}_{gl}^2 + \mathbf{I}\boldsymbol{\sigma}_{el}^2),$$

k=1,2,3,4; l=1,2,...,15, para um total de 60 cenários de simulação. \mathbf{X}_2 representa a variável preditora que tem efeito sobre \mathbf{Y}_{kl} . Nas simulações adota-se $\mu=120$.

Os parâmetros levados em consideração nas análises são: as componentes de variância, em que σ_y^2 representa a variância total (ou fenotípica) que varia com o modelo adotado, tal que $\sigma_y^2 = \sigma_g^2 + \sigma_e^2$; h_g^2 (proporção da variância total devida ao poligene) e h_e^2 (proporção da variância total devida ao erro). As seguintes estimativas são obtidas: $\hat{\mu}$, que representa a estimativa da média geral, $\hat{\beta}_2$, representa o efeito da covariável \mathbf{X}_2 (que no caso é o efeito do SNP 7 sobre \mathbf{Y}_{kl} , especificamente); w_g e w_e representam as proporções de $\hat{\beta}_2$ associadas à parte poligênica e residual, descritas nas equações (3.57) e (3.58) do Capítulo 3. Os valores lg e le correspondem aos máximos que w_g e w_e podem assumir bem como $\hat{\beta}_{wg}$ e $\hat{\beta}_{we}$ são os valores obtidos da partição de $\hat{\beta}_2$ especificamente ilustrados na expressão (3.55) do Capítulo 3 que correspondem à parte poligênica e residual, respectivamente. Ainda, w_{g0} e w_{e0} denotam os índices tomados no modelo reduzido, descrito nas expressões (3.71) e (3.72) do Capítulo 3; d_g e d_e são índices que medem o impacto na parte poligênica e residual quando é incluída uma variável preditora na análise (ver expressões (3.74) e (3.75) no Capítulo 3). Os valores entre parênteses, abaixo de algumas estimativas, representam seus erros padrão correspondentes.

Como uma primeira parte neste capítulo de simulação, foram geradas 100 réplicas para cada um dos 60 cenários e calculadas as médias das estimativas para cada parâmetro em questão. Em seguida, foram apresentados os resultados quando considerados na simulação os SNPs gerados em Desequilíbrio de Ligação (LD), os quais formam um bloco haplotípico. A Seção 5.5, corresponde às análises no nível de famílias e, finalmente, na Seção 5.6 são apresentadas algumas simulações complementares para o estudo do efeito conjunto de duas

variáveis preditoras sobre a variável resposta \mathbf{Y} no intuito de estudar o comportamento do índice d_g .

5.3 Decomposição de efeitos de variáveis preditoras

Nesta seção, na Tabela 5.3 apresentam-se os resultados para a média de 500 réplicas geradas quando os \mathbf{Y}_{kl} têm variância total $\sigma_y^2 = 7$. Os resultados (em média) dos outros cenários em que $\sigma_y^2 = 2$, $\sigma_y^2 = 15$, $\sigma_y^2 = 50$ e $\sigma_y^2 = 100$ são apresentados no Apêndice B nas Tabelas (B.1), (B.2), (B.3) e (B.4) respectivamente. Os resultados para cada tabela estão distribuídos da seguinte forma: na primeira coluna descrevem-se as respectivas herdabilidades poligênicas com que os \mathbf{Y}_{kl} foram gerados (8%, 32% e 70%) e na segunda coluna, estão os parâmetros de interesse os quais foram especificados na seção anterior. Na primeira linha, especificam-se os valores de β_2 (0,1, 0,8, 1,5 e 3,0) e, para cada valor deste parâmetro, s indica as estimativas dos parâmetros levados em consideração na geração dos \mathbf{Y}_{kl} (variância total, variância genética e variância residual e, em consequência, as respectivas herdabilidades); na coluna M_1 estão as respectivas estimativas dos parâmetros no ajuste do modelo reduzido, dado especificamente na equação (3.22); na coluna de M_2 estão as estimativas dos parâmetros sob ajuste do modelo completo (equação (3.23)). Ressalta-se que em todos estes cenários foi selecionado o SNP 7 como a covariável que tem efeito (segundo β_k) sobre os \mathbf{Y}_{kl} .

Para ilustrar o caso em que os \mathbf{Y}_{kl} foram gerados a partir de uma variável aleatória, na Tabela 5.4 apresentam-se os resultados em que o efeito é causado da variável aleatória v_3 e para uma variância geral também de $\sigma_y^2 = 7$.

Ao se incluir o SNP 7 no modelo é esperado que este absorva parte da variabilidade genética e, em consequência, a herdabilidade poligênica diminua. Já no caso em que seja gerada a variável \mathbf{Y}_{kl} a partir de uma variável aleatória sem agregado familiar, esperase que esta absorva parte da variabilidade residual e, em consequência, a herdabilidade residual diminua.

Na discussão dos resultados são abordados quatro enfoques descritos a seguir:

- 1. As estimativas das variâncias, poligênica e residual no modelo reduzido e no modelo completo, representados especificamente, por M1 e M2.
- 2. As porcentagens da estimativa do parâmetro β_2 correspondentes à parte genética e à parte residual, representadas por w_g e w_e , respectivamente, e os valores máximos que estes podem assumir, lg e le.
- 3. A partição de $\hat{\beta}_2$ em $\hat{\beta}_{wg}$ e $\hat{\beta}_{we}$, cujas equações foram obtidas nas expressões (3.61) e (3.62) do Capítulo 3.
- 4. As magnitudes e os sinais dos índices d_g e d_e .

Considerando a Tabela 5.3 para o caso em que os \mathbf{Y}_{kl} são gerados com variância total $\sigma_y^2 = 7$; de acordo ao primeiro item (focando a atenção nos modelos reduzido e completo, especificados por M1 e M2, respectivamente), para todos os cenários quando passamos de M1 para M2, a variância total σ_y^2 é sempre reduzida (em menor ou maior grau, dependendo do efeito de β_2). Isto indica que o SNP 7, quando adicionado ao modelo, absorve parte da variabilidade total, especificamente absorve parte da variabilidade poligênica (a herdabilidade poligênica é reduzida e, em consequência, a herdabilidade residual é aumentada). Como exemplo ilustrativo (Tabela 5.3), para uma herdabilidade de 32% e um efeito de $\beta_2=0,8,$ observa-se que ao passar do modelo sem covariáveis, M1, para o modelo M2, $\hat{\sigma}_y^2=6,984$ passa para $\hat{\sigma}_y^2=6,792$. Observando as estimativas das variâncias poligênica e residual temos que $\hat{\sigma}_g^2=2,197$ passa para $\hat{\sigma}_g^2=2,059$ enquanto $\hat{\sigma}_e^2=4,789$ passa para $\hat{\sigma}_e^2=4,733.$ Note que a diminuição (em média) foi mais alta na variabilidade poligênica do que na variabilidade residual (em consequência, a herdabilidade poligênica diminuiu de 0,315 para 0,303). Para a mesma herdabilidade e considerando $\beta_2=3,0$ tem-se em M1 que $\hat{\sigma}_y^2=9,908$ e, em M2, $\hat{\sigma}_y^2=7,041.$ Observando as estimativas das variâncias poligênica e residual temos que $\hat{\sigma}_g^2=5,067$ e $\hat{\sigma}_g^2=2,212$ enquanto $\hat{\sigma}_e^2=4,842$ e $\hat{\sigma}_e^2 = 4,830$. Note, neste caso que a redução também foi na variabilidade poligênica, a herdabilidade poligênica passou de 0,512 para 0,315 (em consequência, a herdabilidade residual aumentou, passando de 0,488 para 0,685).

Referente ao segundo item de interesse, nota-se que as estimativas do parâmetro β_2 são próximas ao verdadeiro valor e, com exceção do efeito de $\beta_2=0.1$ sobre \mathbf{Y} , todos os efeitos são significantes (pvalor $\leq 0,005$). Considerando a herdabilidade poligênica h_g^2 e w_g (a proporção do efeito do SNP correspondente à parte genética) existe uma relação interessante. Note que para a herdabilidade de 8% o valor de w_g é em média 0,13 (independente de β_2) podendo assumir o valor (teórico) máximo $l_g=0,48$. Para a herdabilidade de 32%, w_g é em média 0,34, podendo assumir o valor máximo $l_g=0,845$. Já para uma herdabilidade de 70%, w_g é em média de 0,63, podendo assumir um valor máximo $l_g=0,966$. Como exemplo ilustrativo, para uma variável resposta gerada com 32% de herdabilidade e um efeito β_2 de 0,8, as respectivas estimativas foram $\hat{\beta}_2=0,792$, $w_g=0,349$ e $w_e=0,651$. Isto sugere que, aproximadamente, 35% da contribuição para a significância do parâmetro β_2 é devido a efeitos genéticos.

Em relação ao item 3 referente à decomposição de $\hat{\beta}_2$ na parte poligênica e na parte residual ($\hat{\beta}_2 = \hat{\beta}_{wg} + \hat{\beta}_{we}$), na Figura 5.5 apresentam-se os gráficos com as respectivas partições de $\hat{\beta}_2$ para os \mathbf{Y}_{kl} gerados com as herdabilidades especificadas em cada caso. A linha sólida e preta representa $\hat{\beta}_2$, o coeficiente angular ajustado do efeito do SNP 7 sobre \mathbf{Y} ; a linha tracejada e azul representa $\hat{\beta}_{we}$ o coeficiente angular da parte residual e a linha pontilhada e verde representa $\hat{\beta}_{wg}$, o coeficiente angular estimado da parte genética. Note, no caso em que a herdabilidade é de 8% que a proporção correspondente à parte poligênica é "pequena" aumentando proporcionalmente assim que a herdabilidade e o efeito β_2 aumentam.

Em relação ao último item referente aos índices, d_g é o inverso de d_e . Note a partir da Tabela 5.3 para todos os casos, que d_g é positivo indicando que a covariável incluída no modelo absorveu parte da variabilidade poligênica; a magnitude do índice aumenta a medida que o efeito (do SNP 7) sobre a respectiva variável resposta aumenta.

\mathbf{Y}_{kl} gerados a partir da covariável sem agregação familiar

Na Tabela 5.4 apresentam-se os resultados em que o efeito é causado pela variável aleatória v_3 sem agregado familiar (resultados análogos aos dados na Tabela 5.3 em que o

Tabela 5.3 Resultados (em média) para variáveis resposta geradas com variância total $\sigma_y^2 = 7$ e covariável o SNP 7.

			$\beta_2 = 0$, 1		$\beta_2 = 0,$	8		$\boldsymbol{\beta}_2 = 1$, 5		$\beta_2 = 3$, 0
h_g^2	parâm	s	M1	M2	s	M1	M2	s	M1	M2	s	M1	M2
0,08	σ_{y}^{2}	7	7,053	7,042	7	7,203	6,967	7	7,908	7,169	7	9,764	6,874
	$egin{array}{c} \sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2 \end{array}$	0,56	0,609	0,604	0,56	0,7	0,498	0,56	1,1	0,495	0,56	3,552	0,658
	σ_e^2	6,44	6,442	6,437	6,44	6,502	6,468	6,44	6,807	6,673	6,44	6,212	6,215
	h_g^2	0,08	0,087	0,087	0,08	0,098	0,072	0,08	0,138	0,069	0,08	0,363	0,095
		0,92	0,913	0,913	0,92	0,902	0,928	0,92	0,862	0,931	0,92	0,637	0,905
	μ			119,968			119,973			119,997			119,998
	pv			0,309			0			0			0
	β_2			0,139 $(0,127)$			0,863 $(0,127)$			1,534 (0,128)			2,997 $(0,128)$
	wg			0,141			0,127)			0,117			0,128)
	we			0,859			0,88			0,883			0,843
	β_{wg}			0,118			0,762			1,36			2,52
	, wy			(0,114)			(0,114)			(0,115)			(0,111)
	β_{we}			0,021			0,099			0,173			0,473
				(0,014)			(0,011)			(0,012)			(0,014)
	lg			0,498			0,445			0,427			0,555
	le			0,994			0,995			0,995			0,992
	w_{go}			0,141			0,158			0,204			0,394
	w_{e0}			0,859			0,842			0,796			0,606
	d_g			0,001			0,036			0,088			0,238
0,32	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	7	7,033	7,029	7	6,984	6,792	7	7,883	7,072	7	9,908	7,041
	σ_g^2	2,24	2,156	2,15	2,24	2,197	2,059	2,24	3,038	2,154	2,24	5,067	2,212
	σ_e^2	4,76	4,878	4,879	4,76	4,789	4,733	4,76	4,845	4,92	4,76	4,842	4,83
	h_g^2	0,32	0,305	0,305	0,32	0,315	0,303	0,32	0,386	0,304	0,32	0,512	0,315
		0,68	0,695	0,695 $119,987$	0,68	0,685	0,697 $119,946$	0,68	0,614	0,696 119,988	0,68	0,488	0,685 $120,037$
	μ pv			0,552			0			0			0
	β_2			0,076			0,792			1,55			2,978
	P 2			(0,133)			(0,131)			(0,134)			(0,135)
	wg			0,35			0,349			0,349			0,357
	we			0,65			0,651			0,651			0,643
	β_{wg}			0,046			0,536			0,989			1,91
				(0,094)			(0,093)			(0,094)			(0,093)
	β_{we}			0,03			0,258			0,558			1,067
				(0,04)			(0,037)			(0,039)			(0,04)
	lg			0,846			0,847			0,844			0,852
	le			0,975			0,975			0,976			0,972
	w_{go}			0,351			0,357			0,409			0,495
	w_{e0}			0,649 0			0,643 0,008			0,591 0,057			0,505 0,139
	d_g												
0,7	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	7	6,934	6,921	7	7,021	6,828	7	7,765	6,971	7	9,881	6,951
	σ_g	4,9 2,1	4,904 2,029	4,888 2,033	4,9 2,1	4,959 2,06	4,777 $2,051$	$^{4,9}_{2,1}$	5,48 $2,285$	4,686 2,284	$^{4,9}_{2,1}$	7,895 $1,985$	4,85 2,1
	b^{σ}_{e}	0,7	0,708	0,707	0,7	0,707	0,699	0,7	0,704	0,669	0,7	0,798	0,697
	h_2^g	0,7	0,708	0,707	0,7	0,707	0,301	0,7	0,704	0,331	0,7	0,798	0,303
	μ	0,0	0,232	119,979	0,0	0,200	120,039	0,0	0,230	119,959	0,0	0,202	119,974
	pv			0,377			0			0			0
	β_2			0,106			0,77			1,554			2,954
				(0,13)			(0,131)			(0,132)			(0,132)
	wg			0,642			0,635			0,615			0,635
	we			0,358			0,365			0,385			0,365
	β_{wg}			0,04			0,283			0,599			1,066
	_			(0,052)			(0,051)			(0,055)			(0,054)
	β_{we}			0,066			0,487			0,955			1,886
	,			(0,08)			(0,08)			(0,077)			(0,079)
	lg le			0,968			0,967 $0,883$			0,964			0,967
	1. e.			0,876			U,083			0,893			0,883
				0.641			0.641			0.641			0.722
	w_{go} w_{e0}			0,641 0,359			0,641 0,359			0,641 0,359			0,722 $0,278$

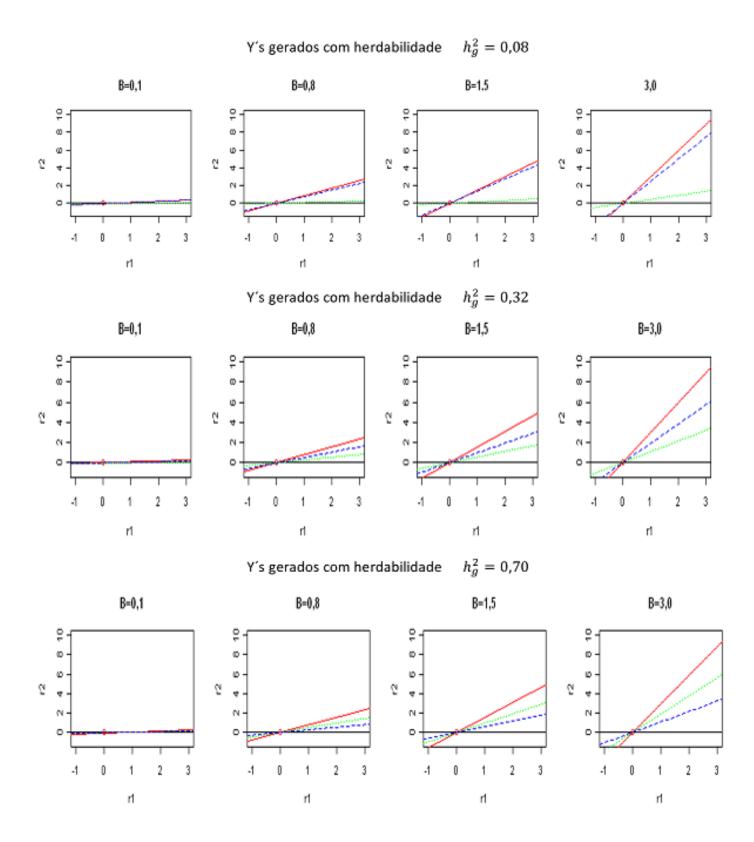


Figura 5.5 Partição de $\hat{\beta}_2$ para variáveis resposta geradas com $\sigma_y^2=7.$

efeito sobre a variável resposta é do SNP 7). O objetivo é estabelecer algumas diferenças nas estimativas dos parâmetros, quando o efeito em \mathbf{Y}_{kl} é causado por um SNP com agregado familiar versus uma variável aleatória, sem agregado familiar (as respectivas ilustrações gráficas estão nas Figuras 5.3 e 5.4).

Observando a Tabela 5.4, de acordo com o primeiro item das análises, quando passamos de M1 para M2 a variância total $\hat{\sigma}_y^2$ é reduzida mas, neste caso, é em consequência da redução na variabilidade residual, (a herdabilidade residual é diminuída e, portanto, a herdabilidade poligênica é aumentada), ocorrendo o caso contrário quando o efeito é causado pelo SNP7. Na Tabela 5.4 para uma herdabilidade de 32% e um efeito de $\beta_2=0,8$, observa-se que quando se passa de M1 para M2 (cuja covariável é v_3) a estimativa da variância total $\hat{\sigma}_y^2=7,363$ passa para $\hat{\sigma}_y^2=6,951$. Observando as estimativas das variâncias poligênica e residual temos que $\hat{\sigma}_g^2=2,123$ passa para $\hat{\sigma}_g^2=2,144$, enquanto $\hat{\sigma}_e^2=5,240$ passa para $\hat{\sigma}_e^2=4,806$. A diminuição foi muito mais alta na variabilidade residual (a herdabilidade residual passou de 0,712 para 0,692).

Para o efeito de β_2 igual a 3,0 em M1, $\hat{\sigma}_y^2 = 12,176$ e, em M2, $\hat{\sigma}_y^2 = 6,885$. Observase também que $\hat{\sigma}_g^2 = 2,472$ passa para $\hat{\sigma}_g^2 = 2,112$, enquanto $\hat{\sigma}_e^2 = 9,704$ passa para $\hat{\sigma}_e^2 = 4,773$. Note neste caso que, a maior diminuição foi também na variabilidade residual (a herdabilidade residual diminuiu, passando de 0,798 para 0,693).

Referente ao segundo item de interesse, nota-se que as estimativas do parâmetro β_2 são próximas ao verdadeiro valor e, com exceção do efeito de $\beta_2 = 0, 1$, todas as estimativas dos parâmetros são significantes. Nota-se uma grande diferença nas estimativas de w_g ao serem comparadas com as estimativas obtidas no caso em que o efeito é causado pelo SNP 7. Para as herdabilidades de 8%, 32% e 70%, w_g , para o SNP 7, é em média, 0,13, 0,34 e 0,63, respectivamente e, quando considerada a variável v_3 , as proporções de w_g são 0,071, 0,12 e 0,49, denotando uma baixa proporção de $\hat{\beta}_2$ na parte poligênica, como esperado.

Em relação ao item referente aos índices d_g e d_e , note para todos os casos que d_g é negativo e d_e é positivo, indicando que a covariável incluída no modelo absorveu parte da variabilidade residual.

Tabela 5.4 Resultados (em média) para variáveis resposta geradas com variância total $\sigma_y^2 = 7$ e covariável v_3 .

			$\beta_2 = 0,$	1		$\beta_2 = 0,$	8		$\beta_2 = 1$,	5		$\beta_2 = 3$,	0
h_g^2	parâm	s	M1	M2	s	M1	M2	s	M1	M2	s	M1	M2
0,08	σ_y^2	7,000	6,916	6,910	7,000	7,384	6,993	7,000	8,354	7,004	7,000	12,275	6,900
	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	0,560	0,472	0,475	0,560	0,630	0,676	0,560	0,771	0,666	0,560	0,874	0,561
	σ_e^2	6,440	6,444	6,434	6,440	6,754	6,316	6,440	7,584	6,339	6,440	11,402	6,339
	h_g^2	0,080	0,067	0,068	0,080	0,085	0,097	0,080	0,094	0,095	0,080	0,072	0,080
	h_e^2	0,920	0,933	0,932	0,920	0,915	0,903	0,920	0,906	0,905	0,920	0,928	0,920
	μ			120,020			120,011			120,020			120,040
	pv			0,421			0,000			0,000			0,000
	β_2			0,075			0,817			1,505			3,011
	an a			(0,090) 0,060			(0,091) $0,080$			(0,090) $0,078$			(0,090) 0,067
	wg we			0,000			0,920			0,922			0,933
	β_{wg}			0,069			0,755			1,386			2,802
				(0,082)			(0,081)			(0,081)			(0,081)
	β_{we}			0,005			0,060			0,122			0,211
				(0,006)			(0,010)			(0,009)			(0,007)
	lg			0,467			0,569			0,560			0,513
	le			0,996			0,993			0,992			0,993
	w_{go}			0,060			0,073			0,076			0,060
	w_{e0}			0,940			0,927			0,924			0,940
	d_g			0,000			-0,007			-0,002			-0,009
0,32	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	7,000	6,980	6,969	7,000	7,363	6,951	7,000	8,322	6,902	7,000	$12,\!176$	6,885
	σ_g^2	2,240	2,253	2,246	2,240	2,123	2,144	2,240	2,137	2,046	2,240	2,472	2,112
	σ_e^2	4,760	4,727	4,723	4,760	5,240	4,806	4,760	6,186	4,857	4,760	9,704	4,773
	h_g^2	0,320	0,323	0,322	0,320	0,288	0,308	0,320	0,258	0,296	0,320	0,202	0,307
	$\frac{n_e^-}{e}$	0,680	0,677	0,678 $119,938$	0,680	0,712	0,692 119,997	0,680	0,742	0,704 $119,984$	0,680	0,798	0,693 119,995
	pv			0,376			0,000			0,000			0,000
	β_2			0,093			0,845			1,542			2,978
	. 2			(0,086)			(0,085)			(0,087)			(0,085)
	wg			0,226			0,217			0,210			0,216
	we			0,774			0,783			0,790			0,784
	β_{wg}			0,065			0,669			1,217			2,328
				(0,064)			(0,065)			(0,068)			(0,065)
	β_{we}			0,025			0,176			0,326			0,649
	,			(0,019)			(0,021)			(0,019)			(0,019)
	lg le			0,853 0,972			0,845 $0,974$			0,838 $0,977$			0,844 $0,975$
				0,226			0,204			0,185			0,151
	w_{go} w_{e0}			0,774			0,796			0,815			0,849
	d_g			0,000			-0,011			-0,025			-0,064
0,7	σ^2	7,000	6,863	6,854	7,000	7,305	6,920	7,000	8,088	6,728	7,000	12,229	6,970
-,-	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	4,900	4,772	4,776	4,900	4,874	4,884	4,900	4,880	4,718	4,900	5,006	4,790
	σ_e^2	2,100	2,091	2,076	2,100	2,432	2,035	2,100	3,208	2,009	2,100	7,224	2,184
	h_{q}^{2}	0,700	0,694	0,696	0,700	0,667	0,705	0,700	0,601	0,699	0,700	0,409	0,687
	$h_e^{\stackrel{\circ}{2}}$	0,300	0,306	0,304	0,300	0,333	0,295	0,300	0,399	0,301	0,300	0,591	0,313
	μ			120,004			119,927			119,977			119,983
	pv			0,309			0,000			0,000			0,000
	β_2			0,109			0,817			1,491			2,975
				(0,076)			(0,076)			(0,072)			(0,078)
	wg			0,494			0,501			0,499			0,485
	we			0,506 $0,057$			0,499 $0,411$			0,501 $0,747$			0,515 $1,530$
	β_{wg}			(0,037)			(0,033)			(0,034)			(0,036)
	β_{we}			0,051			0,406			0,743			1,447
	·we			(0,040)			(0,040)			(0,039)			(0,040)
	lg			0,967			0,967			0,967			0,963
	le			0,884			0,876			0,881			0,887
	w_{go}			0,494			0,467			0,415			0,279
	w_{e0}			0,506			0,533			0,585			0,721
	d_g			-0,002			-0,033			-0,083			-0,205

Quando passamos do modelo 1 (modelo reduzido) para o modelo 2 (modelo completo) a variância total (σ_y^2) sempre diminui. Essa diminuição na variabilidade é devido ao fato de que a covariável em questão absorve uma parte da variabilidade total. De grande interesse para o estudo é saber se a variabilidade absorvida pela covariável corresponde à variabilidade poligênica ou à variabilidade residual.

5.4 Análises no caso haplotípico

. Chama-se a atenção para o banco de dados das variáveis preditoras ilustrado na Figura 5.2. Os SNPs 6, 7 e 8 foram gerados em Desequilíbrio de Ligação, especificamente estes SNPs foram gerados com probabilidades haplotípicas (ver no Apêndice A o código de simulação no pacote SimPed), em que um progenitor pode passar para um filho não só alelos de risco mas também um segmento do genoma, ou seja pode passar um haplótipo de risco (no caso, cada haplótipo é composto por três alelos, como descrito na Seção 4.2 do capítulo 4).

Para estas análises, cada \mathbf{Y}_{kl} foi gerada com efeito somente do SNP 7 e o intuito é ver, por meio dos diferentes cenários, se os efeitos dos SNPs adjacentes (SNP 6 e SNP 8) sobre a variável resposta também são capturados. Em geral, pretende-se ver as diferenças dos SNPs que formam o bloco haplotípico relativamente a todos os demais SNPs e variáveis aleatórias do banco de dados em questão (Figura 5.2) principalmente no índice d_g especificado na equação (3.74) do Capítulo 3.

Para este tipo de análise, são apresentados somente os cenários em que os \mathbf{Y}_{kl} foram gerados com variância total $\sigma_y^2 = 7$ e $\sigma_y^2 = 50$ e para cada uma das herdabilidades (8%, 32% e 70%). Nesta seção apresentam-se os resultados para o caso em que os \mathbf{Y}_{kl} foram gerados com variância total $\sigma_y^2 = 7$, os resultados do outro cenário são apresentados no Apêndice B. Cada variável resposta foi gerada somente com efeito do SNP 7 e o ajuste do modelo completo foi realizado, independentemente, para cada variável preditora. De forma geral os dados estão organizados em três tabelas; na Tabela 5.5 tem-se os resultados para os \mathbf{Y}_{kl} simulados com herdabilidade de 8% na Tabela 5.6 para os \mathbf{Y}_{kl} simulados com

herdabilidade de 32% e na Tabela 5.7 para os \mathbf{Y}_{kl} gerados com herdabilidade de 70%, variando sobre todos os efeitos β_2 (0, 1, 0, 8, 1, 5, 3, 0). Cada tabela está organizada da seguinte forma: na primeira linha descrevem-se os parâmetros envolvidos na geração dos \mathbf{Y}_{kl} com efeitos β_2 (do SNP 7) apresentados na coluna 1. Na linha onde está cada efeito β_2 , apresentam-se as estimativas dos componentes de variância obtidos ajustando o modelo reduzido. Na coluna 2, estão os parâmetros envolvidos nas análises para todos os 9 SNPS e três variáveis preditoras aleatórias (v_1, v_2, v_3) . As estimativas foram obtidas do ajuste de um modelo completo.

Resultados para os \mathbf{Y}_{kl} gerados com herdabilidade de 8%

Na Tabela 5.5, quando $\beta_2 = 0, 1,$ nenhuma variável preditora teve efeito significativo sobre Y (critério do p-valor), neste caso, o bloco de SNPs em observação (SNP 6, SNP 7, SNP 8) não é diferenciado das demais variáveis preditoras. Destaca-se o fato que a proporção da variabilidade poligênica (representada por w_g) explicada pelas variáveis aleatórias (v_1, v_2, v_3) é mais baixa do que a proporção explicada pelos SNPs, como é esperado. Para o caso em que o efeito representado por β_2 assume os valores 0,8, 1,5 e 3,0, o bloco em questão começa a ser destacado das outras variáveis preditoras. Note que as estimativas do parâmetro β_2 para os SNPs adjacentes ao SNP 7 em alguns casos são significantes. Isto é refletido no índice d_g , pois para os efeitos de 0,8, 1,5 e 3,0 o índice d_g é o mais alto para o SNP 7, seguido pelos índices no SNP 6 e no SNP 8. Como exemplo ilustrativo, toma-se o caso em que $\beta_2=3,0.$ Lembra-se que a variável resposta foi gerada só com efeito do SNP 7 e que o SNP 6, SNP7 e SNP 8 estão em LD. As únicas variáveis preditoras que tiveram efeito significante sobre a variável resposta foram os SNPs 6, 7 e 8. As estimativas de β_2 foram 1,466, 3,148 (verdadeiro valor é 3,0) e 1,098 e foram significantes. O índice d_g para o SNP 6 de 0,062, para o SNP 7 de 0,231 e para o SNP 8 de 0,028, os quais são os valores mais altos, quando comparados com os valores de d_g das outras variáveis preditoras cuja tendência é para zero, em geral.

Tabela 5.5 Resultados quando os \mathbf{Y}_{kl} são gerados com herdabilidade de 8%.

						$\sigma_y^2 = 7$	$h_g^2 = 0.08$	$\sigma_g^2 = 0.56$	$\sigma_e^2 = 6.44$				
β_2	parâm	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	v1	v2	v3
),1							$\sigma_{y0}^2 = 7,389$	$h_{g0}^2 = 0,113$	$h_{e0}^2 = 0,887$				
	σ_{ν}^{2}	7,371	7,383	7,384	7,388	7,354	7,386	7,389	7,389	7,389	7,388	7,384	7,371
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,111	0,111	0,111	0,113	0,111	0,111	0,113	0,113	0,114	0,113	0,113	0,115
	h_e^2	0,889	0,889	0,889	0,887	0,889	0,889	0,887	0,887	0,886	0,887	0,887	0,885
	pv	0,135	0,479	0,526	0,772	0,03	0,679	0,954	0,984	0,801	0,726	0,4	0,092
	β_2	0,209	0,165	-0,082	-0,041	0,314	0,054	0,008	-0,003	-0,038	-0,036	-0,089	-0,194
	-	(0,14)	(0,232)	(0,129)	(0,143)	(0,145)	(0,13)	(0,146)	(0,14)	(0,151)	(0,102)	(0,106)	(0,115)
	wg	0,188	0,182	0,18	0,207	0,193	0,201	0,21	0,217	0,186	0,084	0,086	0,088
	we	0,812	0,818	0,82	0,793	0,807	0,799	0,79	0,783	0,814	0,916	0,914	0,912
	β_{wg}	0,036	0,057	-0,031	-0,008	0,051	0,032	0,004	0,016	0,016	-0,004	-0,009	-0,013
	β_{we}	0,172	0,108	-0,051	-0,033	0,263	0,022	0,005	-0,019	-0,054	-0,032	-0,08	-0,181
	w_{g0}	0,19	0,184	0,182	0,207	0,195	0,203	0,21	0,217	0,185	0,084	0,086	0,087
	w_{e0}	0,81	0,816	0,818	0,793	0,805	0,797	0,79	0,783	0,815	0,916	0,914	0,913
	d_g	0,002	0,002	0,002	0	0,002	0,002	0	0	-0,001	0	0	-0,001
,8							$\sigma_{y0}^2 = 7,276$	$h_{g0}^2 = 0,104$	$h_{e0}^2 = 0,896$				
	σ_{u}^{2}	7,268	7,273	7,271	7,273	7,273	7,209	7,081	7,276	7,273	7,261	7,276	7,274
	$h_{\alpha}^{\stackrel{g}{2}}$	0,103	0,102	0,103	0,105	0,103	0,097	0,069	0,104	0,103	0,102	0,104	0,104
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,897	0,898	0,897	0,895	0,897	0,903	0,931	0,896	0,897	0,898	0,896	0,896
	pv	0,209	0,593	0,377	0,371	0,53	0	0	0,866	0,515	0,094	0,903	0,497
	β_2	0,143	0,098	-0,093	-0,105	-0,077	0,37	0,77	0,02	0,077	0,14	-0,011	0,062
	7- 2	(0,114)	(0,184)	(0,105)	(0,118)	(0,123)	(0,104)	(0,128)	(0,121)	(0,119)	(0,084)	(0,091)	(0,091)
	wg	0,18	0,165	0,18	0,17	0,168	0,168	0,126	0,185	0,168	0,08	0,087	0,086
	we	0,82	0,835	0,82	0,83	0,832	0,832	0,874	0,815	0,832	0,92	0,913	0,914
	β_{wg}	0,024	0,054	-0,023	-0,005	-0,021	0,06	0,113	0,007	0,019	0,017	-0,002	0,003
	β_{we}	0,119	0,044	-0,07	-0,101	-0,056	0,31	0,657	0,014	0,059	0,123	-0,009	0,059
	w_{g0}	0,181	0,167	0,182	0,168	0,169	0,177	0,17	0,185	0,168	0,081	0,087	0,086
	w_{e0}	0,819	0,833	0,818	0,832	0,831	0,823	0,83	0,815	0,832	0,919	0,913	0,914
	d_g	0,001	0,002	0,002	-0,001	0,001	0,009	0,044	0	0,001	0,001	0	0
1,5							$\sigma_{y0}^2 = 7,737$	$h_{g0}^2 = 0,24$	$h_{e0}^2 = 0,76$				
	$\begin{matrix} \sigma_y^2 \\ h_g^2 \\ h_e^2 \end{matrix}$	7,677	7,736	7,737	7,737	7,736	7,461	6,99	7,638	7,737	7,737	7,727	7,728
	$h_q^{\frac{5}{2}}$	0,229	0,243	0,24	0,239	0,239	0,195	0,12	0,224	0,239	0,24	0,243	0,236
	h_e^2	0,771	0,757	0,76	0,761	0,761	0,805	0,88	0,776	0,761	0,76	0,757	0,764
	pv	0,02	0,287	0,812	0,889	0,806	0	0	0,002	0,943	0,997	0,132	0,48
	β_2	0,342	0,261	0,032	-0,021	-0,038	0,714	1,353	0,467	-0,011	0	0,16	0,082
		(0,147)	(0,245)	(0,136)	(0,152)	(0,154)	(0,135)	(0,142)	(0,149)	(0,159)	(0,102)	(0,106)	(0,116)
	wg	0,289	0,291	0,289	0,322	0,305	0,284	0,218	0,325	0,294	0,163	0,168	0,165
	we	0,711	0,709	0,711	0,678	0,695	0,716	0,782	0,675	0,706	0,837	0,832	0,835
	β_{wg}	0,125	0,011	-0,002	-0,026	-0,024	0,231	0,323	0,164	-0,024	-0,001	0,014	0,047
	β_{we}	0,217	0,25	0,035	0,005	-0,014	0,483	1,029	0,302	0,013	0,001	0,146	0,035
	w_{g0}	0,297	0,289	0,289	0,322	0,305	0,319	0,323	0,337	0,294	0,163	0,166	0,167
	w_{e0}	0,703	0,711	0,711	0,678	0,695	0,681	0,677	0,663	0,706	0,837	0,834	0,833
	dg	0,008	-0,002	0	0	0	0,036	0,105	0,012	0	0	-0,002	0,002
3							$\sigma_{y0}^2 = 11,265$	$h_{g0}^2 = 0,417$	$h_{e0}^2 = 0,583$				
	$\sigma_y^2 \ h_g^2 \ h_e^2$	$11,\!273$	11,236	11,26	11,269	11,262	10,138	7,503	10,721	11,253	11,266	11,253	11,23
	h_g^2	0,419	0,42	0,416	0,418	0,417	0,323	0,111	0,373	0,415	0,417	0,416	0,414
	h_e^2	0,581	0,58	0,584	0,582	0,583	0,677	0,889	0,627	0,585	0,583	0,584	0,586
	pv	0,557	0,051	0,594	0,771	0,534	0	0	0	0,485	0,928	0,378	0,125
	β_2	-0,106	0,581	-0,088	0,055	-0,117	1,466	3,148	1,098	-0,135	-0,011	-0,108	0,205
		(0,181)	(0,298)	(0,166)	(0,188)	(0,188)	(0,161)	(0,147)	(0,182)	(0,194)	(0,118)	(0,123)	(0,134)
	wg	0,414	0,406	0,407	0,44	0,422	0,379	0,208	0,428	0,412	0,277	0,279	0,279
	we	0,586	0,594	0,593	0,56	0,578	0,621	0,792	0,572	0,588	0,723	0,721	0,721
	β_{wg}	0,025	0,179	-0,042	-0,039	-0,04	0,602	0,659	0,501	-0,086	0,017	-0,042	0,077
	β_{we}	-0,131	0,402	-0,046	0,094	-0,077	0,864	2,489	0,597	-0,05	-0,027	-0,067	0,128
	w_{g0}	0,412	0,404	0,407	0,439	0,421	0,441	0,439	0,456	0,413	0,277	0,28	0,28
	w_{e0}	0,588	0,596	0,593	0,561	0,579	0,559	0,561	0,544	0,587	0,723	0,72	0,72
		-0,002		0	-0,001	0	0,062	0,231	0,028	0,001	0	0,001	0,002

Resultados para os \mathbf{Y}_{kl} gerados com herdabilidade de 32%

Destaca-se o cenário em que $\beta_2 = 0, 8$. Neste caso, os efeitos dos SNPs 6 e 7 foram significantes o que não ocorreu para o efeito do SNP 8 (p-valor=0,031). No entanto o índice d_g para o SNP 8 é de 0,006 pois este SNP absorve parte da variabilidade poligênica. Note que a herdabilidade no modelo reduzido (M1) é de 0,277 e, no completo (M2) é de 0,269. Assim, este SNP 6 pode ser considerado em análises posteriores fazendo parte de um bloco haplotípico.

Como resumo da informação obtida a través do índice d_g , na Figura 5.6 apresentamse os índices cujos valores são dados nas Tabelas 5.5, 5.6 e 5.7 para as herdabilidades consideradas 8%, 32% e 70%. Esses gráficos permitem visualizar os efeitos do conjunto de variáveis preditoras, destacando o bloco dos três SNPs que formam o haplótipo. Note em alguns cenários que nem todos os três SNPs tem efeito significante. No entanto, o índice d_g mostra que esse SNP exerce influência genética sobre a variável resposta e deveria ser analisado como formando o haplótipo com os SNPs adjacentes que mostram algum efeito, mesmo que este seja "pequeno".

Da mesma forma, na Figura 5.7 apresenta-se o índice d_g para os valores obtidos nas Tabelas (B.5), (B.6), (B.7), dadas no Apêndice B, em que $\sigma_y^2 = 50$ e considerando todas as herdabilidades. Esses gráficos condensam a informação do efeito de todas as variáveis preditoras sobre a variável resposta \mathbf{Y}_{kl} quando analisadas independentemente. Nesta figura, percebe-se que para efeitos baixos do SNP 7 sobre a variável preditora, não é possivel diferenciar o bloco ou haplótipo, este é destacado das demais variáveis preditoras quando os efeitos de β_2 são de 1,5 e 3,0; em particular para uma herdabilidade de 70% o bloco só é destacado no caso em que $\beta_2 = 3,0$.

Para estudar o efeito conjunto (efeito de haplótipo) do bloco dos três SNPs gerados em DL, é obtida a variável \mathbf{X}_h por meio da codificação descrita na Tabela 4.4 no Capítulo 4. As estimativas dos parâmetros considerados encontram-se na Tabela 5.8. Nesta tabela, as estimativas dos parâmetros, quando incluídos independentemente os SNPs 6, 7 e 8 são as mesmas dos cenários considerados na Tabela 5.6 quando as variáveis resposta são geradas

Tabela 5.6 Resultados quando os \mathbf{Y}_{kl} são gerados com herdabilidade de 32%.

						$\sigma_y^2 = 7$	$h_q^2 = 32$	$\sigma_q^2 = 2.24$	$\sigma_e^2 = 4.76$				
β_2	parâm	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	v1	v2	v3
0,1							$\sigma_{y0}^2 = 6,993$	$h_{g0}^2 = 0,389$	$h_{e0}^2 = 0,611$				
	$\begin{matrix} \sigma_y^2 \\ h_g^2 \\ h_e^2 \end{matrix}$	6,98	6,983	6,993	6,971	6,993	6,882	6,871	6,97	6,995	6,992	6,993	6,992
	h_g^2	0,385	0,388	0,39	0,392	0,389	0,37	0,376	0,392	0,391	0,39	0,39	0,389
	h_e^2	0,615	0,612	0,61	0,608	0,611	0,63	0,624	0,608	0,609	0,61	0,61	0,611
	pv	0,528	0,374	0,738	0,039	0,991	0,002	0	0,04	0,817	0,678	0,884	0,934
	β_2	0,09	0,209	0,044	0,306	-0,002	0,415	0,548	0,302	0,035	-0,039	-0,014	0,009
		(0,142)	(0,235)	(0,131)	(0,148)	(0,148)	(0,133)	(0,15)	(0,147)	(0,153)	(0,093)	(0,098)	(0,106)
	wg	0,392	0,385	0,39	0,423	0,404	0,41	0,413	0,44	0,396	0,259	0,262	0,262
	we	0,608	0,615	0,61	0,577	0,596	0,59	0,587	0,56	0,604	0,741	0,738	0,738
	β_{wg}	0,084	0,098	0,005	0,1	0,005	0,193	0,226	0,104	-0,029	-0,007	0,006	0,008
	β_{we}	0,006	0,11	0,038	0,205	-0,007	0,221	0,321	0,198	0,065	-0,032	-0,021	0,001
	w_{g0}	0,395	0,386	0,389	0,422	0,404	0,423	0,422	0,439	0,395	0,259	0,262	0,262
	w_{e0}	0,605	0,614	0,611	0,578	0,596	0,577	0,578	0,561	0,605	0,741	0,738	0,738
	d_g	0,003	0,001	0	-0,002	0	0,013	0,009	-0,001	-0,001	0	0	0
0,8							$\sigma_{y0}^2 = 7,329$	$h_{g0}^2 = 0,277$	$h_{e0}^2 = 0,723$				
	$\sigma_y^2 \ h_g^2 \ h_e^2$	7,238	7,328	7,324	7,321	7,268	7,186	7,101	7,279	7,322	7,324	7,328	7,294
	h_g^2	0,26	0,277	0,277	0,277	0,267	0,267	0,261	0,269	0,274	0,277	0,277	0,275
	h_e^2	0,74	0,723	0,723	0,723	0,733	0,733	0,739	0,731	0,726	0,723	0,723	0,725
	pv	0,004	0,817	0,448	0,351	0,017	0	0	0,031	0,729	0,444	0,784	0,039
	β_2	0,412	0,055	-0,101	0,139	-0,36	0,564	0,803	0,319	-0,054	0,075	0,028	0,23
		(0,143)	(0,239)	(0,133)	(0,149)	(0,15)	(0,134)	(0,15)	(0,147)	(0,155)	(0,098)	(0,103)	(0,111)
	wg	0,311	0,314	0,314	0,348	0,324	0,34	0,338	0,358	0,319	0,186	0,189	0,189
	we	0,689	0,686	0,686	0,652	0,676	0,66	0,662	0,642	0,681	0,814	0,811	0,811
	β_{wg}	0,158	0,04	-0,034	0,046	-0,136	0,176	0,248	0,126	-0,093	0,021	0,011	0,055
	β_{we}	0,255	0,015	-0,067	0,093	-0,224	0,388	0,555	0,193	0,039	0,055	0,017	0,175
	w_{g0}	0,322	0,314	0,315	0,349	0,331	0,347	0,35	0,364	0,321	0,186	0,189	0,191
	w_{e0}	0,678	0,686	0,685	0,651	0,669	0,653	0,65	0,636	0,679	0,814	0,811	0,809
	d_g	0,012	0	0	0	0,007	0,007	0,011	0,006	0,002	0	0	0,002
1,5							$\sigma_{y0}^2 = 7,264$	$h_{g0}^2 = 0,334$	$h_{e0}^2 = 0, 0, 666$				
	$\sigma_y^2 \ h_g^2 \ h_e^2$	7,258	7,258	7,248	7,231	7,26	7,125	6,53	7,232	7,263	7,264	7,265	7,265
	h_g^2	0,333	0,333	0,335	0,327	0,333	0,33	0,271	0,33	0,334	0,337	0,339	0,335
	h_e^2	0,667	0,667	0,665	0,673	0,667	0,67	0,729	0,67	0,666	0,663	0,661	0,665
	pv	0,452	0,533	0,115	0,126	0,714	0	0	0,077	0,738	0,36	0,23	0,671
	β_2	0,109	-0,149	-0,21	-0,229	0,055	0,588	1,418	0,263	0,052	-0,088	0,121	-0,047
		(0,145)	(0,239)	(0,133)	(0,149)	(0,151)	(0,135)	(0,144)	(0,148)	(0,156)	(0,097)	(0,101)	(0,11)
	wg	0,359	0,35	0,353	0,382	0,368	0,384	0,345	0,4	0,359	0,224	0,229	0,227
	we	0,641	0,65	0,647	0,618	0,632	0,616	0,655	0,6	0,641	0,776	0,771	0,773
	β_{wg}	0,044	-0,08	-0,061	-0,124	0,06	0,199	0,471	0,114	0,022	-0,003	0	0,008
	β_{we}	0,065	-0,07	-0,149	-0,105	-0,005	0,389	0,947	0,149	0,031	-0,085	0,121	-0,055
	P_{g0}	0,359	0,351	0,353	0,386	0,368	0,386	0,387	0,403	0,359	0,222	0,226	0,227
	P_{e0} d_g	0,641 0	0,649 $0,001$	0,647	0,614 $0,005$	0,632 $0,001$	0,614 0,003	0,613 $0,042$	0,597 0,003	0,641 0	0,778 -0,002	0,774 -0,003	0,773 -0,001
3	y						$\sigma_{y0}^2 = 11,257$	$h_{g0}^2 = 0,605$	$h_{e0}^2 = 0,395$				
	2	11,207	11,255	11,296	11,252	11,251	9,929	7,586	10,889	11,245	11,236	11,253	11,251
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,6	0,604	0,62	0,604	0,604	0,523	0,382	0,578	0,603	0,602	0,604	0,604
	$\frac{n}{2}$	0,6 $0,4$	0,804	0,82	0,804	0,396	0,323	0,618	0,422	0,803	0,802	0,804	0,396
	$n_e^ pv$	0,4 $0,154$	0,396	0,38	0,396	0,396	0,477	0,618	0,422	0,397	0,398	0,396 $0,742$	0,396
		0,154 $0,254$	0,823	-0,355	-0,049	-0,1	1,577	3,07	0,852	-0,012	0,368	-0,038	0,829
	β_2	(0,234)	(0,293)	(0,163)	(0,187)	(0,187)	(0,161)	(0,157)	(0,185)	(0,192)	(0,11)	(0,116)	(0,126)
	wg	0,54	0,535	0,553	0,564	0,551	0,513	0,417	0,562	0,545	0,11 $0,417$	0,418	0,416
	we	0,34	0,335	0,333 $0,447$	0,436	0,331	0,487	0,583	0,438	0,345	0,583	0,582	0,584
		0,165	0,405	-0,122	-0,061	-0,063	0,813	1,245	0,496	-0,086	0,06	-0,029	0,047
	β_{wg} β	0,103	0,033	-0,122	0,012	-0,003	0,765	1,825	0,355	-0,030	0,04	-0,029	-0,02
	P_{g0}	0,544	0,535	0,541	0,565	0,551	0,571	0,566	0,581	0,546	0,419	0,418	0,417
	P_{e0}	0,456	0,335	0,459	0,303	0,331	0,429	0,434	0,419	0,340	0,581	0,582	0,583
	d_g	0,004	0,400	-0,012	0,001	0,449	0,058	0,149	0,019	0,001	0,002	0,001	0,001
	<i>g</i>	- ,		-,~-=	- ,		-,	-,0	-,0	- /	- ,	- ,	-,

Tabela 5.7 Resultados quando os \mathbf{Y}_{kl} são gerados com herdabilidade de 70%.

						$\sigma_y^2 = 7$	$h_g^2 = 0,70$	$\sigma_g^2 = 4, 9$	$\sigma_e^2=2,1$				
β_2	parâm	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	v1	v2	v3
,1							$\sigma_{y0}^2 = 6,58$	$h_{g0}^2 = 0,751$	$h_{e0}^2 = 0,249$				
	$\begin{matrix}\sigma_y^2\\h_q^2\\h_e^2\end{matrix}$	6,579	6,583	6,574	6,574	6,567	6,474	6,298	6,548	6,58	6,58	6,577	6,554
	h_{a}^{2}	0,751	0,753	0,751	0,75	0,749	0,741	0,721	0,75	0,752	0,751	0,751	0,748
	h_e^2	0,249	0,247	0,249	0,25	0,251	0,259	0,279	0,25	0,248	0,249	0,249	0,252
	pv	0,79	0,48	0,321	0,66	0,464	0,002	0	0,038	0,678	0,915	0,506	0,143
	β_2	0,035	-0,154	-0,121	0,062	-0,102	0,399	0,721	0,294	0,059	0,008	-0,055	0,131
		(0,133)	(0,218)	(0,122)	(0,14)	(0,139)	(0,127)	(0,141)	(0,141)	(0,143)	(0,079)	(0.083)	(0,09)
	wg	0,669	0,662	0,667	0,681	0,672	0,68	0,659	0,695	0,671	0,556	0,55	0,543
	we	0,331	0,338	0,333	0,319	0,328	0,32	0,341	0,305	0,329	0,444	0,45	0,457
	β_{wg}	0,032	-0,066	-0,079	0,066	-0,1	0,292	0,518	0,2	0,033	0,008	-0,033	0,091
	β_{we}	0,003	-0,088	-0,042	-0,005	-0,002	0,107	0,203	0,093	0,027	0	-0,022	0,04
	w_{g0}	0,669	0,66	0,667	0,682	0,674	0,69	0,686	0,696	0,671	0,556	0,55	0,546
	w_{e0}	0,331	0,34	0,333	0,318	0,326	0,31	0,314	0,304	0,329	0,444	0,45	0,454
	d_g	0	-0,001	0	0,001	0,002	0,009	0,027	0,001	0	0	0	0,003
,8							$\sigma_{y0}^2=7,107$	$h_{g0}^2 = 0,682$	$h_{e0}^2 = 0,318$				
_	σ_{u}^{2}	7,107	7,107	7,053	7,107	7,105	7,046	6,905	7,04	7,111	7,104	7,106	7,108
	h_{a}^{2}	0,682	0,682	0,678	0,683	0,682	0,681	0,668	0,676	0,684	0,682	0,682	0,682
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,318	0,318	0,322	0,317	0,318	0,319	0,332	0,324	0,316	0,318	0,318	0,318
	pv	0,989	0,982	0,024	0,444	0,821	0,004	0	0,013	0,404	0,748	0,92	0,936
	β_2	0,002	-0,005	-0,289	0,113	-0,033	0,386	0,691	0,366	0,126	-0,027	0,009	-0,008
	r 2	(0,14)	(0,23)	(0,128)	(0,147)	(0,147)	(0,134)	(0,149)	(0,148)	(0,151)	(0,085)	(0,089)	(0,096)
	wg	0,606	0,598	0,6	0,624	0,612	0,63	0,614	0,634	0,611	0,487	0,484	0,482
	we	0,394	0,402	0,4	0,376	0,388	0,37	0,386	0,366	0,389	0,513	0,516	0,518
	β_{wg}	0,005	-0,004	-0,185	0,054	-0,034	0,233	0,433	0,242	0,048	-0,023	0,01	0,012
	β_{we}	-0,004	-0,002	-0,105	0,059	0	0,154	0,259	0,124	0,078	-0,004	-0,001	-0,019
	w_{g0}	0,606	0,598	0,604	0,624	0,613	0,631	0,626	0,639	0,609	0,488	0,484	0,482
	w_{e0}	0,394	0,402	0,396	0,376	0,387	0,369	0,374	0,361	0,391	0,512	0,516	0,518
	d_g	0	0	0,004	-0,001	0	0,001	0,012	0,005	-0,002	0	0	0
,5							$\sigma_{y0}^2 = 8,273$	$h_{g0}^2 = 0,757$	$h_{e0}^2 = 0,243$				
	σ_{ν}^2	8,273	8,26	8,278	8,269	8,261	7,888	7,036	8,034	8,268	8,269	8,262	8,221
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,757	0,755	0,758	0,756	0,756	0,747	0,705	0,748	0,756	0,761	0,756	0,755
	h_{c}^{2}	0,243	0,245	0,242	0,244	0,244	0,253	0,295	0,252	0,244	0,239	0,244	0,245
	pv	0,847	0,438	0,492	0,516	0,358	0	0	0	0,528	0,064	0,386	0,017
	β_2	-0,029	0,189	-0,094	0,102	0,144	0,928	1,801	0,779	-0,101	0,164	-0,08	0,24
		(0,149)	(0,244)	(0,136)	(0,157)	(0,156)	(0,14)	(0,15)	(0,156)	(0,16)	(0,088)	(0,092)	(0,1)
	wg	0,674	0,664	0,674	0,687	0,678	0,685	0,645	0,693	0,676	0,566	0,554	0,55
	we	0,326	0,336	0,326	0,313	0,322	0,315	0,355	0,307	0,324	0,434	0,446	0,45
	β_{wg}	-0,022	0,155	-0,04	0,07	0,107	0,623	1,155	0,533	-0,074	0,085	-0,054	0,146
	β_{we}	-0,007	0,034	-0,053	0,031	0,037	0,305	0,647	0,246	-0,027	0,079	-0,026	0,094
	w_{g0}	0,674	0,665	0,672	0,687	0,679	0,694	0,691	0,701	0,676	0,562	0,555	0,551
	w_{e0}	0,326	0,335	0,328	0,313	0,321	0,306	0,309	0,299	0,324	0,438	0,445	0,449
	d_g	0	0,001	-0,001	0	0,001	0,009	0,046	0,008	0	-0,004	0,001	0,001
3							$\sigma_{y0}^2 = 9,255$	$h_{g0}^2 = 0,773$	$h_{e0}^2 = 0,227$				
	$\begin{matrix} \sigma_y^2 \\ h_g^2 \\ h_e^2 \end{matrix}$	9,254	9,255	9,208	9,255	9,244	8,439	6,489	8,951	9,257	9,256	9,256	9,261
	h_{a}^{2}	0,773	0,773	0,775	0,773	0,773	0,756	0,694	0,76	0,773	0,774	0,773	0,779
	$h_{\mathcal{L}}^{2}$	0,227	0,227	0,225	0,227	0,227	0,244	0,306	0,24	0,227	0,226	0,227	0,221
	pv	0,753	0,662	0,01	0,963	0,269	0	0	0	0,915	0,578	0,818	0,053
	β_2	-0,049	0,112	-0,368	-0,008	-0,182	1,357	2,784	0,838	-0,018	0,052	-0,022	-0,203
	. 2	(0,157)	(0,257)	(0,143)	(0,166)	(0,164)	(0,145)	(0,144)	(0,165)	(0,169)	(0,093)	(0,097)	(0,105)
		0,69	0,682	0,69	0,702	0,695	0,694	0,636	0,704	0,692	0,58	0,572	0,574
	wg		0,318	0,31	0,298	0,305	0,306	0,364	0,296	0,308	0,42	0,428	0,426
	$egin{array}{c} wg \ we \end{array}$	0,31					0,926	1,745	0,6	0,021	0,024	-0,006	-0,097
	we	0,31 -0,036	0,065	-0,239	-0,031	-0,125	0,920						
	we β_{wg}	-0,036	0,065 $0,048$	-0,239 -0,129	0,031	-0,125 -0,057			0,238		0,027	-0,017	
	$we \ eta_{wg} \ eta_{we}$	-0,036 -0,014	0,048	-0,129	0,023	-0,057	0,431	1,039	0,238	-0,039	0,027	-0,017	-0,106
	we β_{wg}	-0,036											

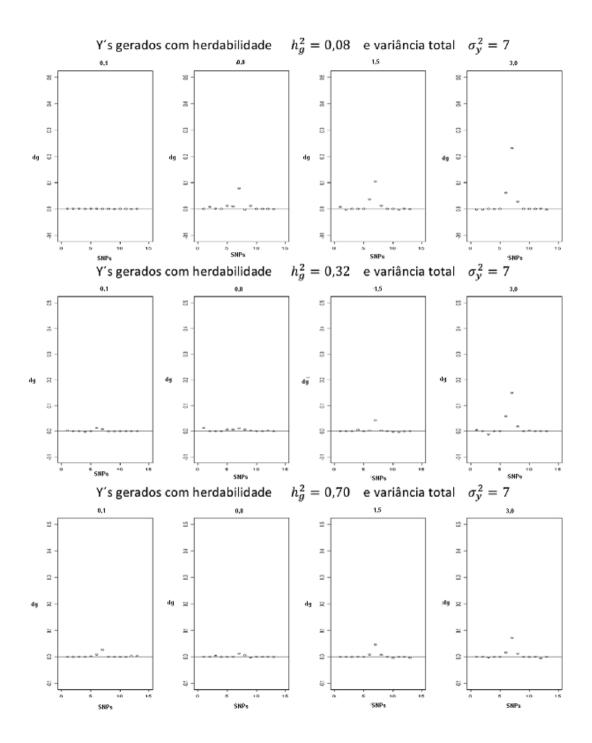


Figura 5.6 Indices d_g para o conjunto de variáveis preditoras.

com herdabilidade de 32%. Para este caso, percebe-se que a variável preditora \mathbf{X}_h absorve mais da variabilidade poligênica, pois note que a herdabilidade poligênica é menor do que aquelas dos modelos em que os SNPs em LD são incluídos individualmente, por exemplo

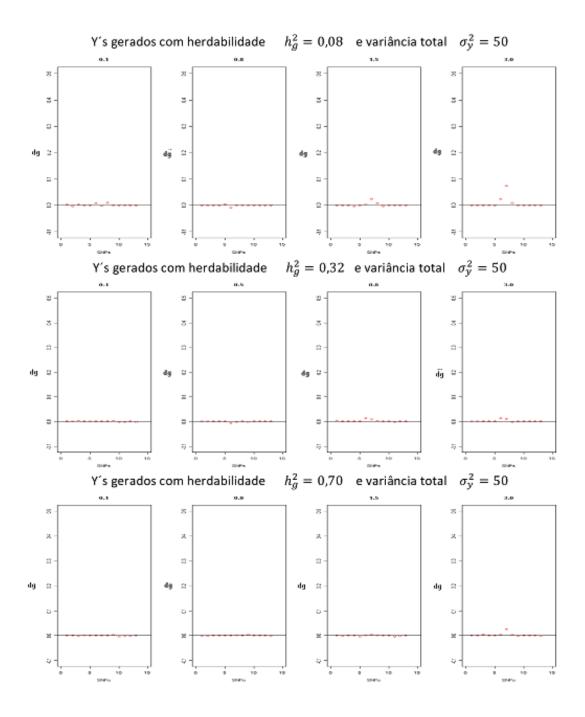


Figura 5.7 Indices d_g para o conjunto de variáveis preditoras.

para o cenário em que $\beta_2 = 1,5$ o valor da herdabilidade no modelo reduzido é $h_{g0}^2 = 0,334$ e quando incluida a variável \mathbf{X}_h no modelo completo a herdabilidae $h_{g0}^2 = 0,2691$. Também a estimativa do parâmetro β_2 é maior do que as respectivas estimativas obtidas para os vizinhos SNP 6 e o SNP 8. Na obtenção da variável \mathbf{X}_h , existem situações em que a

codificação desta variável pode ser igual ou estar altamente relacionada a algum dos SNPs envolvidos, podendo-se encontrar as mesmas estimativas ou estimativas muito próximas ao SNP em questão, mas a interpretação é diferente já que são considerados "pedaços" do genoma que os progenitores podem transmitir para o filho ou, mais especificamente, consideram-se haplótipos de risco. Note que para o exemplo dado, as estimativas estão próximas das estimativas obtidas para o SNP 7, que é o loco responsável pelo efeito.

Para os cenários descritos na Tabela 5.8 foi realizado um estudo de simulação, em que para cada efeito de β_2 sobre a variável resposta \mathbf{Y} foram realizadas 1000 réplicas no intuito de obter o número de vezes que o bloco de SNPs, simultaneamente, é discriminado dentre o conjunto geral das 13 de variáveis preditoras sob estudo. O bloco dos SNPs 6, 7 e 8 é discriminado dos demais se estes apresentam o maior valor do índice d_g (ver a Figura 5.6 por exemplo). Quando o efeito é "pequeno", $\beta_2 = 0, 1$, das mil réplicas, somente 28 discriminam o bloco; para $\beta_2 = 0, 8$, 715 discriminam o bloco; para $\beta_2 = 1, 5$, 815 discriminam o bloco e, finalmente para $\beta_2 = 3$, 985 discriminam o bloco.

5.5 Análises no nível de famílias

Para as análises no nível de famílias escolhem-se dois cenários. O primeiro deles corresponde à variável resposta \mathbf{Y} , gerada a partir de $\sigma_y^2 = 7$, com herdabilidade de 32% e cujo efeito é $\beta_2 = 1, 5$. O segundo cenário corresponde à variável resposta \mathbf{Y} , gerada a partir de $\sigma_y^2 = 50$, com herdabilidade de 32% e cujo efeito é $\beta_2 = 1, 5$ (resultados no Apêndice B).

Na Figura 5.8 apresentam-se os gráficos das contribuições totais de cada família para o parâmetro β_2^g dado na equação (3.84), por meio do índice I_f dado na equação (3.86) do Capítulo 3. Considera-se, de forma geral, uma família caudal se o valor desse índice I_f para essa família ultrapassa 1 desvio padrão à respectiva média do índice I_f . O primeiro gráfico apresenta as famílias quando incluído o SNP 6 no modelo completo representado por M2, resultando que as famílias caudais positivas são as famílias 1, 5, 7, 14 e 48 e uma única família caudal negativa (sinal do efeito dessa família oposto ao sinal do efeito

Tabela 5.8 Efeitos de haplótipo.

		$\sigma_y^2 = 7$	$h_g^2 = 32$	$\sigma_g^2 = 2.24$	$\sigma_e^2 = 4.76$
β_2	parâm	SNP6	SNP7	SNP8	X_h
0,1		$\sigma_{y0}^2 = 6,993$	$h_{g0}^2 = 0,389$	$h_{e0}^2 = 0,611$	
	$\sigma_y^2 \ h_g^2 \ h_e^2$	6,882	6,871	6,97	6,868
	h_q^2	0,37	0,376	0,392	0,368
	h_e^2	0,63	0,624	0,608	0,632
	pv	0,002	0	0,04	0,003
	β_2	0,415	0,548	0,302	0,225
		(-0,133)	(0,15)	(0,147)	(0,12)
	wg	0,41	0,413	0,44	0,409
	we	0,59	0,587	0,56	0,591
	β_{wg}	0,193	0,226	0,104	0,105
	β_{we}	0,221	0,321	0,198	0,120
0,8		$\sigma_{y0}^2 = 7,329$	$h_{g0}^2 = 0,277$	$h_{e0}^2 = 0,723$	
	$\sigma_y^2 \ h_g^2 \ h_e^2$	7,186	7,101	7,279	7,01
	$h_a^{\frac{g}{2}}$	0,267	0,261	0,269	0,258
	h_e^2	0,733	0,729	0,731	0,742
	pv	0	0	0,031	0,025
	β_2	0,564	1,418	0,319	0,612
		(0,134)	(0,144)	(0,147)	(0,13)
	wg	0,34	0,345	0,358	0,342
	we	0,66	0,655	0,642	0,658
	β_{wg}	0,176	0,471	0,126	0,208
	β_{we}	0,388	0,947	0,193	0,403
1,5		$\sigma_{y0}^2 = 7,264$	$h_{g0}^2 = 0, 334$	$h_{e0}^2 = 0,666$	
	σ_u^2	7,125	6,53	7,232	0,65
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,33	0,271	0,33	0,269
	h_{-}^{2}	0,67	0,729	0,67	0,731
	pv	0	0	0,077	0,002
	β_2	0,588	1,418	0,263	0,815
		(0,135)	(0,144)	(0,148)	(0,13)
	wg	0,384	0,345	0,4	0,36
	we	0,616	0,655	0,6	0,64
	β_{wg}	0,199	0,471	0,114	0,293
					0,522
	β_{we}	0,389	0,947	0,149	0,522
3	β_{we}	$\sigma_{y0}^2 = 11,257$	$h_{g0}^2 = 0,605$	$h_{e0}^2 = 0,395$	0,322
3					7,58
3		$\sigma_{y0}^2 = 11,257$	$h_{g0}^2 = 0,605$	$h_{e0}^2 = 0,395$	
3		$\sigma_{y0}^2 = 11,257$ $9,929$	$h_{g0}^2 = 0,605$ 7,586	$h_{e0}^2 = 0,395$ $10,889$	7,58
3	β_{we} σ_y^2 h_g^2 h_e^2 pv	$\sigma_{y0}^2 = 11,257$ $9,929$ $0,523$	$h_{g0}^2 = 0,605$ $7,586$ $0,382$	$h_{e0}^2 = 0,395$ $10,889$ $0,578$	7,58 0,378
3	$\sigma_y^2 \ h_g^2 \ h_e^2 \ pv$	$\sigma_{y0}^2 = 11,257$ $9,929$ $0,523$ $0,477$	$h_{g0}^2 = 0,605$ $7,586$ $0,382$ $0,618$	$h_{e0}^2 = 0,395$ $10,889$ $0,578$ $0,422$	7,58 0,378 0,622
3	$\sigma_y^2 \\ h_g^2 \\ h_e^2$	$\sigma_{y0}^2 = 11,257$ $9,929$ $0,523$ $0,477$ 0	$h_{g0}^2 = 0,605$ $7,586$ $0,382$ $0,618$ 0	$h_{e0}^2 = 0,395$ $10,889$ $0,578$ $0,422$ 0	7,58 0,378 0,622 0
3	$\sigma_y^2 \ h_g^2 \ h_e^2 \ pv \ eta_2$	$\sigma_{y0}^2 = 11,257$ $9,929$ $0,523$ $0,477$ 0 $1,577$	$h_{g0}^2 = 0,605$ $7,586$ $0,382$ $0,618$ 0 $3,07$	$h_{e0}^{2} = 0,395$ $10,889$ $0,578$ $0,422$ 0 $0,852$	7,58 0,378 0,622 0 2,01
3	$\sigma_y^2 \ h_g^2 \ h_e^2 \ pv$	$\sigma_{y0}^2 = 11,257$ $9,929$ $0,523$ $0,477$ 0 $1,577$ $(0,161)$	$h_{g0}^{2} = 0,605$ $7,586$ $0,382$ $0,618$ 0 $3,07$ $(0,157)$	$h_{e0}^{2} = 0,395$ $10,889$ $0,578$ $0,422$ 0 $0,852$ $(0,185)$	7,58 0,378 0,622 0 2,01 (0,135)
3	σ_y^2 h_g^2 h_e^2 pv β_2	$\sigma_{y0}^2 = 11,257$ $9,929$ $0,523$ $0,477$ 0 $1,577$ $(0,161)$ $0,513$	$h_{g0}^{2} = 0,605$ $7,586$ $0,382$ $0,618$ 0 $3,07$ $(0,157)$ $0,417$	$h_{e0}^2 = 0,395$ $10,889$ $0,578$ $0,422$ 0 $0,852$ $(0,185)$ $0,562$	7,58 0,378 0,622 0 2,01 (0,135) 0,418

do SNP) é identificada, a família 6. Para entender este resultado, nas Figuras 5.9 e 5.10 apresentam-se os heredogramas das famílias 6 e 7, respectivamente, indicando o valor do genótipo e o valor da variável resposta para cada indivíduo.

Observando a família 6 (caudal negativa), nota-se que esta é uma família com 17 indivíduos em que 10 indivíduos carregam 2 alelos de risco e os outros 7 indivíduos carregam um alelo de risco. A média geral da variável resposta em estudo (\mathbf{Y}) é 120,576. Dos 17 indivíduos, 15 destes têm valores da resposta abaixo da média. Como esperado, esta é uma família que tem um peso ($w_{g,f}$) baixo no cálculo de $\hat{\beta}_2^g$, mas seu coeficiente angular é alto (negativamente) que, em suma, é a família que tem menos contribuição para o valor de $\hat{\beta}_2^g$. Podemos dizer que é uma família muito diferente das demais, pois, ao contrário do esperado (devido ao efeito β_2 ser positivo), os indivíduos desta família com alelos de risco têm valores da variável resposta baixos. Uma explicação para um caso real seria que os indivíduos dessa família estejam tomando algum medicamento efetivo para controlar a pressão, ou que em regiões cromossômicas associadas a esta exista um possível gene protetor para o aumento da pressão arterial.

A Figura (5.10), correspondente ao heredograma da família 7, temos um total de 33 indivíduos, dos quais 11 carregam 2 alelos de risco, 12 carregam um alelo de risco e os restantes 19 carregam zero alelos de risco. Referente à variável resposta, quem carrega 2 alelos de risco está acima da média e, em geral, os que não carregam alelos estão abaixo da média dando uma alta contribuição para $\hat{\beta}_2^g$.

Também, na Figura 5.8 encontram-se os índices das famílias quando incluído o SNP 7 como variável preditora, tem-se várias famílias caudais positivas, a saber, as famílias 5, 11, 12, 14, 26 e 48, e famílias caudais negativas, 21, 22 e 47. Finalmente, na mesma Figura 5.8 são apresentados os índices das famílias quando incluído o SNP 8. Nota-se que são só duas famílias caudais positivas, famílias 5 e 6.

Quando analisados os SNPs 6, 7 e 8 por meio da codificação haplotípica (ver a codificação no Capítulo 4), na Figura 5.11 apresenta-se o índice para cada uma das famílias, encontrando-se as famílias 5 e 14 como caudais positivas e a família 47 como caudal nega-

tiva. Devido ao tamanho das famílias 5 (70 membros) e 14 (108 membros) não são apresentados os respectivos heredogramas, mas para a família 47, na Figura 5.12, apresentase seu respectivo heredograma. Ao lado de cada indivíduo está a seguinte codificação SNP6SNP7SNP8/hap/Y53. Por exemplo, para o indivíduo 1 tem-se 102/0/118, 6 indicando que esse indivíduo no SNP 6 carrega 1 alelo de risco, no SNP7 não carrega alelos de risco e no SNP 8 carrega 2 alelos de risco; no nível haplotípico, este indivíduo não carrega haplótipos de risco (codificando-se como 0); o valor da variável resposta para este indivíduo é 118,6, e assim sucessivamente para todos os indivíduos. Esta família é considerada caudal porque mesmo os indivíduos não carregando haplótipos de risco, a variável resposta está acima da média para a maioria de indivíduos.

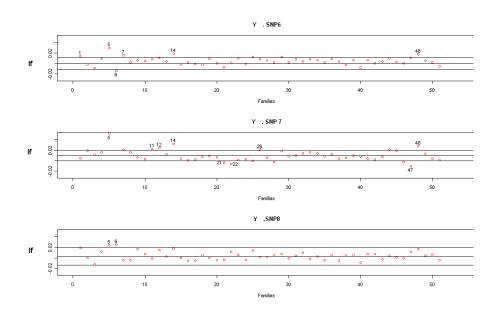


Figura 5.8 Representação da contribuição de cada família no componente poligênico.

5.6 Simulações complementares

Até agora ao serem geradas as variáveis resposta \mathbf{Y}_{kl} , o efeito sobre esta foi de uma única variável preditora (SNP 7 ou v_3). A seguir, apresentam-se resultados de 4 variáveis resposta $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$ e \mathbf{Y}_4 com efeitos de duas variáveis preditoras conjuntamente. As

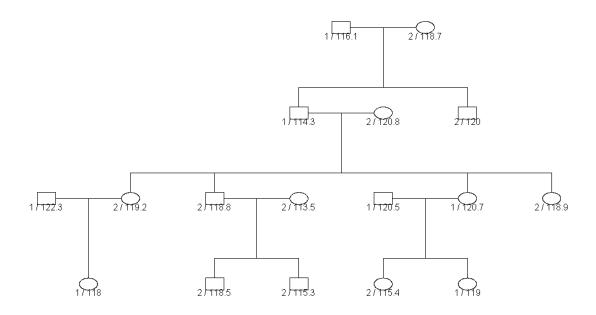


Figura 5.9 Heredograma correspondente à família 6.

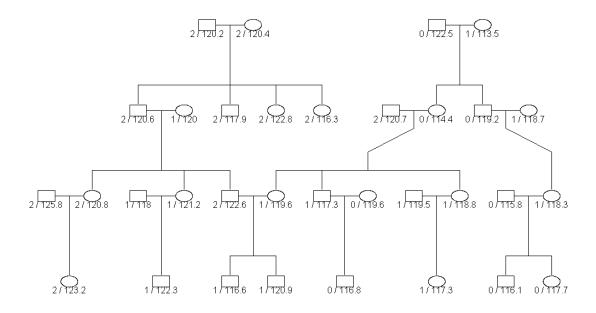


Figura 5.10 $Heredograma\ correspondente\ \grave{a}\ família\ 7.$

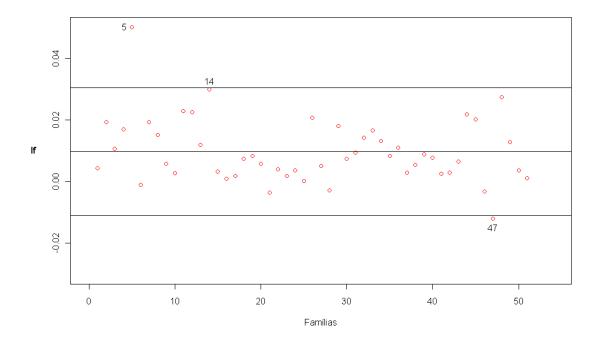


Figura 5.11 Famílias caudais quando considerados os SNPs 6, 7 e 8 como um bloco.

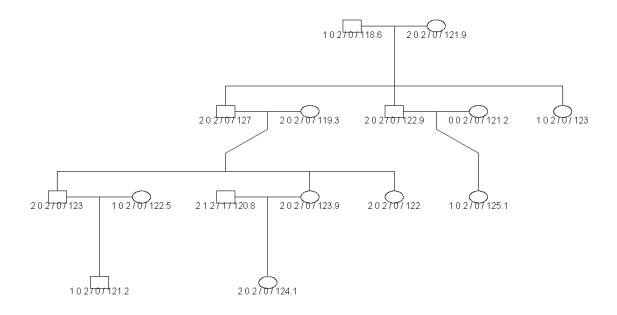


Figura 5.12 Heredograma correspondente à família 47.

análises focam-se principalmente nos valores do índice d_g , com o objetivo de entender seu comportamento nesta situação mais geral.

As 4 variáveis resposta foram geradas com herdabilidade de 32% (particularmente, o mesmo efeito das variáveis preditoras sobre \mathbf{Y}) da seguinte forma:

 \rightarrow efeito de dois SNPs independentes, SNP 4 e SNP 9:

$$\mathbf{Y}_1 \sim N(\boldsymbol{\mu} + 1, 5(SNP4 + SNP9); 2\boldsymbol{\Phi}(2, 24) + \mathbf{I}(4, 76))$$

 \rightarrow efeito de duas variáveis preditoras (aleatórias), v1 e v4:

$$\mathbf{Y}_2 \sim N(\boldsymbol{\mu} + 1, 5(v1 + v4); 2\boldsymbol{\Phi}(2, 24) + \mathbf{I}(4, 76))$$

 \rightarrow efeito do SNP 4 e uma variável preditora (aleatória) v_2 :

$$\mathbf{Y}_3 \sim N(\boldsymbol{\mu} + 1, 5(SNP4 + v2); 2\boldsymbol{\Phi}(2, 24) + \mathbf{I}(4, 76))$$

 \rightarrow efeito do SNP 2 e do SNP 6 o que implica em efeitos de dois haplótipos, pois o SNP 2 está em LD com o SNP 3 e os SNPs, 6, 7 e 8 também estão em LD:

$$\mathbf{Y}_4 \sim N(\boldsymbol{\mu} + 1, 5(SNP2 + SNP6); 2\boldsymbol{\Phi}(2, 24) + \mathbf{I}(4, 76))$$

ju Na Tabela 5.9 apresentam-se os resultados das simulações para as variáveis resposta $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$ e \mathbf{Y}_4 quando incluídas, independentemente, as variáveis preditoras no ajuste do modelo completo. Na mesma tabela, σ_{y0}^2 , h_{g0}^2 e h_{e0}^2 representam as estimativas obtidas sob o modelo reduzido e, em cada coluna, apresentam-se as estimativas de cada parâmetro para cada variável preditora no modelo completo. Na última linha, quando considerada cada variável resposta, apresentam-se os valores do índice d_g para cada uma das variáveis preditoras. Levando em consideração estes índices, na Figura 5.13 apresentam-se os respectivos gráficos.

No caso da variável \mathbf{Y}_1 em que os SNPs são variáveis preditoras com agregado familiar, o índice d_g apresenta valores positivos e altos quando comparados com os índices obtidos das outras preditoras. Para \mathbf{Y}_2 , os índices mais negativos são das variáveis preditoras

Tabela 5.9 Estimativas para os cenários das variáveis $\mathbf{Y}_1,\mathbf{Y}_2,\mathbf{Y}_3$ e $\mathbf{Y}_4.$

	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	v1	v2	v4
						Y_1						
					$\sigma_{y0}^2 = 8,663$	$h_{g0}^2 = 0,479$	$h_{e0}^2 = 0,521$					
σ_y^2	8,665	8,662	8,661	7,768	8,639	8,659	8,659	8,664	7,587	8,660	8,664	8,659
$\begin{matrix} \sigma_y^2 \\ h_g^2 \\ h_e^2 \end{matrix}$	0,482	0,480	0,479	0,431	0,475	0,480	0,478	0,479	0,412	0,478	0,480	0,478
h_e^2	0,518	0,520	0,521	0,569	0,525	0,520	0,522	0,521	0,588	0,522	0,520	0,522
pv	0,218	0,386	0,609	0,000	0,168	0,286	0,550	0,754	0,000	0,730	0,570	0,643
β_2	0,163	0,181	0,063	1,528	0,195	0,130	-0,090	-0,044	1,668	0,029	0,052	0,052
	(0,132)	(0,209)	(0,123)	(0,128)	(0,141)	(0,121)	(0,150)	(0,141)	(0,127)	(0,084)	(0,092)	(0,112)
wg	0,479	0,459	0,489	0,443	0,468	0,485	0,474	0,483	0,431	0,314	0,323	0,328
we	0,521	0,541	0,511	0,557	0,532	0,515	0,526	0,517	0,569	0,686	0,677	0,672
β_{wg}	0,035	0,051	0,029	0,657	0,123	0,047	-0,053	0,005	0,707	0,025	0,003	0,034
β_{we}	0,128	0,130	0,034	0,870	0,071	0,083	-0,037	-0,049	0,961	0,004	0,049	0,018
w_{g0}	0,477	0,459	0,490	0,475	0,470	0,484	0,474	0,482	0,478	0,315	0,323	0,328
w_{e0}	0,523	0,541	0,510	0,525	0,530	0,516	0,526	0,518	0,522	0,685	0,677	0,672
d_g	-0,002	-0,001	0,000	0,033	0,003	-0,001	0,000	0,000	0,047	0,000	-0,001	0,000
						Y_2						
~					$\sigma_{y0}^2 = 9,456$	$h_{g0}^2 = 0,186$	$h_{e0}^2 = 0,814$					
σ_y^2	9,452	9,456	9,455	9,456	9,453	9,454	9,454	9,456	9,455	7,892	9,443	8,592
$\sigma_y^2 \ h_g^2 \ h_e^2$	0,187	0,186	0,186	0,188	0,186	0,186	0,186	0,186	$0,\!187$	0,272	0,187	0,198
h_e^{2}	0,813	0,814	0,814	0,812	0,814	0,814	0,814	0,814	0,813	0,728	0,813	0,802
pv	0,309	0,920	$0,\!579$	0,526	0,503	0,508	0,569	0,834	0,494	0,000	0,123	0,000
β_2	-0,136	-0,022	0,068	0,087	-0,096	-0,081	-0,087	-0,030	-0,095	1,548	0,158	1,468
	(0,134)	(0,214)	(0,123)	(0,137)	(0,144)	(0,123)	(0,153)	(0,142)	(0,138)	(0,085)	(0,102)	(0,119)
wg	0,266	0,247	0,269	0,256	0,253	0,263	0,254	0,270	0,255	0,186	0,142	0,152
we	0,734	0,753	0,731	0,744	0,747	0,737	0,746	0,730	0,745	0,814	0,858	0,848
β_{wg}	-0,018	-0,071	0,010	-0,008	-0,027	-0,018	-0,023	-0,026	-0,009	0,268	0,017	0,223
β_{we}	-0,119	0,049	0,058	0,095	-0,069	-0,064	-0,064	-0,004	-0,085	1,280	0,141	1,246
w_{g0}	0,265	0,247	0,269	0,254	0,253	0,263	0,255	0,270	0,254	0,134	0,141	0,144
w_{e0}	0,735	0,753	0,731	0,746	0,747	0,737	0,745	0,730	0,746	0,866	0,859	0,856
d_g	-0,001	0,000	0,000	-0,001	0,000	0,000	0,000	0,000	-0,001	-0,052	-0,001	-0,008
						Y_3						
					$\sigma_{y0}^2 = 8,952$	$h_{q0}^2 = 0,296$	$h_{e0}^2 = 0,704$					
2	0.050	0.000	0.051	7.000				0.007	0.051	0.050	7.700	0.051
σ_y	8,953		8,951	7,920	8,946	8,938	8,942	8,937	8,951	8,952	7,708	8,951
		8,939	0.000				0.000					
h_g^2	0,297	0,297	0,296	0,196	0,295	0,295	0,293	0,296	0,297	0,297	0,333	0,296
$egin{array}{c} \sigma_y^2 \ h_g^2 \ h_e^2 \end{array}$	0,297 $0,703$	0,297 $0,703$	0,704	0,804	0,705	0,705	0,707	0,704	0,703	0,703	0,667	0,704
pv	0,297 0,703 0,873	0,297 0,703 0,120	0,704 $0,919$	0,804 $0,000$	$0,705 \\ 0,425$	$0,705 \\ 0,177$	0,707 $0,485$	0,704 $0,112$	0,703 $0,538$	0,703 0,540	0,667 $0,000$	0,704 $0,785$
	0,297 0,703 0,873 -0,021	0,297 0,703 0,120 0,330	0,704 0,919 -0,013	0,804 0,000 1,592	0,705 0,425 -0,114	0,705 0,177 -0,164	0,707 0,485 -0,106	0,704 0,112 -0,225	0,703 0,538 0,085	0,703 0,540 0,055	0,667 $0,000$ $1,445$	0,704 $0,785$ $0,032$
eta_2	0,297 0,703 0,873 -0,021 (0,133)	0,297 $0,703$ $0,120$ $0,330$ $(0,212)$	0,704 0,919 -0,013 (0,123)	0,804 0,000 1,592 (0,126)	0,705 0,425 -0,114 (0,142)	0,705 0,177 -0,164 (0,122)	0,707 0,485 -0,106 (0,151)	0,704 $0,112$ $-0,225$ $(0,141)$	0,703 0,538 0,085 (0,137)	0,703 0,540 0,055 (0,090)	0,667 0,000 1,445 (0,090)	0,704 0,785 0,032 (0,119)
pv eta_2 wg	0,297 0,703 0,873 -0,021 (0,133) 0,353	0,297 0,703 0,120 0,330 (0,212) 0,333	0,704 0,919 -0,013 (0,123) 0,360	0,804 0,000 1,592 (0,126) 0,263	0,705 0,425 -0,114 (0,142) 0,341	0,705 0,177 -0,164 (0,122) 0,353	0,707 0,485 -0,106 (0,151) 0,342	0,704 0,112 -0,225 (0,141) 0,358	0,703 0,538 0,085 (0,137) 0,346	0,703 0,540 0,055 (0,090) 0,201	0,667 0,000 1,445 (0,090) 0,231	0,704 0,785 0,032 (0,119) 0,213
pv eta_2 wg we	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667	0,704 0,919 -0,013 (0,123) 0,360 0,640	0,804 0,000 1,592 (0,126) 0,263 0,737	0,705 $0,425$ $-0,114$ $(0,142)$ $0,341$ $0,659$	0,705 0,177 -0,164 (0,122) 0,353 0,647	0,707 0,485 -0,106 (0,151) 0,342 0,658	0,704 0,112 -0,225 (0,141) 0,358 0,642	0,703 0,538 0,085 (0,137) 0,346 0,654	0,703 0,540 0,055 (0,090) 0,201 0,799	0,667 0,000 1,445 (0,090) 0,231 0,769	0,704 0,785 0,032 (0,119) 0,213 0,787
pv eta_2 wg we eta_{wg}	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422	0,705 0,425 -0,114 (0,142) 0,341 0,659 -0,045	0,705 0,177 -0,164 (0,122) 0,353 0,647 -0,057	0,707 0,485 -0,106 (0,151) 0,342 0,658 -0,086	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010
pv eta_2 wg we eta_{wg} eta_{we}	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170	0,705 0,425 -0,114 (0,142) 0,341 0,659 -0,045	0,705 0,177 -0,164 (0,122) 0,353 0,647 -0,057 -0,108	0,707 0,485 -0,106 (0,151) 0,342 0,658 -0,086	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022
pv eta_2 wg we eta_{wg} eta_{we} wg_0	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346	0,705 0,425 -0,114 (0,142) 0,341 0,659 -0,045 -0,069 0,342	0,705 0,177 -0,164 (0,122) 0,353 0,647 -0,057 -0,108 0,354	0,707 0,485 -0,106 (0,151) 0,342 0,658 -0,086 -0,020 0,345	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213
pv eta_2 wg we eta_{wg} eta_{we} w_{g0} w_{e0}	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654	0,705 0,425 -0,114 (0,142) 0,341 0,659 -0,045 -0,069 0,342 0,658	0,705 0,177 -0,164 (0,122) 0,353 0,647 -0,057 -0,108 0,354 0,646	0,707 0,485 -0,106 (0,151) 0,342 0,658 -0,086 -0,020 0,345 0,655	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787
pv eta_2 wg we eta_{wg} eta_{we} eta_{wg}	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346	0,705 0,425 -0,114 (0,142) 0,341 0,659 -0,045 -0,069 0,342	0,705 0,177 -0,164 (0,122) 0,353 0,647 -0,057 -0,108 0,354	0,707 0,485 -0,106 (0,151) 0,342 0,658 -0,086 -0,020 0,345	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213
pv eta_2 wg we eta_{wg} eta_{we} w_{g0} w_{e0}	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654	0,705 0,425 -0,114 (0,142) 0,341 0,659 -0,045 -0,069 0,342 0,658 0,001	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \end{array}$	0,707 0,485 -0,106 (0,151) 0,342 0,658 -0,086 -0,020 0,345 0,655 0,002	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787
pv eta_2 wg we eta_{wg} eta_{we} w_{g0} w_{e0} d_g	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654 0,000	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000
pv β_2 wg we β_{wg} β_{we} w_{g0} w_{e0} d_g	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083	$\begin{matrix} 0,705\\ 0,425\\ -0,114\\ (0,142)\\ 0,341\\ 0,659\\ -0,045\\ -0,069\\ 0,342\\ 0,658\\ 0,001\\ \end{matrix}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654 0,000	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000
pv β_2 wg we β_{wg} β_{we} w_{g0} w_{e0} d_g	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654 0,000	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000
pv β_2 wg we β_{wg} β_{we} w_{g0} w_{e0} d_g	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083	$\begin{matrix} 0,705\\ 0,425\\ -0,114\\ (0,142)\\ 0,341\\ 0,659\\ -0,045\\ -0,069\\ 0,342\\ 0,658\\ 0,001\\ \end{matrix}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654 0,000	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000
pv eta_2 wg we eta_{wg} eta_{we} w_{g0} w_{e0}	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083 8,022 0,421	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ 0,319 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \end{array}$ $\begin{array}{c} h_{e0}^2 = 0,579 \\ 7,703 \\ 0,383 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654 0,000	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000 8,023 0,421	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000
$\begin{array}{c} pv\\ \beta_2\\ \\ we\\ \\ \beta_wg\\ \\ \beta_we\\ \\ we0\\ \\ d_g\\ \\ \\ d_g\\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ $	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,332 0,668 -0,001	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000 7,913 0,408 0,592	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083 8,022 0,421 0,579	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ 0,319 \\ 0,681 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \end{array}$ $\begin{array}{c} h_{e0}^2 = 0,579 \\ 7,703 \\ 0,383 \\ 0,617 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000 7,755 0,393 0,607	0,703 0,538 0,085 (0,137) 0,346 0,654 0,071 0,346 0,654 0,000 8,022 0,421 0,579	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000 8,023 0,421 0,579	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000 8,024 0,422 0,578
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000 8,025 0,422 0,578 0,620	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001 7,733 0,390 0,610 0,000	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000 7,913 0,408 0,592 0,000	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083 8,022 0,421 0,579 0,944	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \end{array}$ $\sigma_{y0}^2 = 8,023 \\ 8,008 \\ 0,419 \\ 0,581 \\ 0,169 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ 0,319 \\ 0,681 \\ 0,000 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \end{array}$ $\begin{array}{c} h_{e0}^2 = 0,579 \\ 7,703 \\ 0,383 \\ 0,617 \\ 0,000 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000 7,755 0,393 0,607 0,000	0,703 0,538 0,085 (0,137) 0,346 0,654 0,071 0,346 0,654 0,000 8,022 0,421 0,579 0,635	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000 8,023 0,421 0,579 0,993	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000 8,024 0,422 0,578 0,727
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000 8,025 0,422 0,578 0,620 0,063	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001 7,733 0,390 0,610 0,000 1,251	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000 7,913 0,408 0,592 0,000 0,439	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083 8,022 0,421 0,579 0,944 -0,009	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \\ \\ \sigma_{y0}^2 = 8,023 \\ 8,008 \\ 0,419 \\ 0,581 \\ 0,169 \\ -0,187 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ 0,319 \\ 0,681 \\ 0,000 \\ 1,335 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \\ \end{array}$ $\begin{array}{c} h_{e0}^2 = 0,579 \\ 7,703 \\ 0,383 \\ 0,617 \\ 0,000 \\ 0,923 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000 7,755 0,393 0,607 0,000 0,806	0,703 0,538 0,085 (0,137) 0,346 0,654 0,071 0,346 0,654 0,000 8,022 0,421 0,579 0,635 -0,062	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000 8,023 0,421 0,579 0,993 -0,001	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022 8,022 0,421 0,579 0,957 0,005	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000 8,024 0,422 0,578 0,727 0,038
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000 8,025 0,422 0,578 0,620 0,063 (0,127)	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001 7,733 0,390 0,610 0,000 1,251 (0,198)	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000 7,913 0,408 0,592 0,000 0,439 (0,117)	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083 8,022 0,421 0,579 0,944 -0,009 (0,130)	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \\ \\ \sigma_{y0}^2 = 8,023 \\ 8,008 \\ 0,419 \\ 0,581 \\ 0,169 \\ -0,187 \\ (0,136) \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ 0,319 \\ 0,681 \\ 0,000 \\ 1,335 \\ (0,109) \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \\ \\ h_{e0}^2 = 0,579 \\ 7,703 \\ 0,383 \\ 0,617 \\ 0,000 \\ 0,923 \\ (0,142) \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000 7,755 0,393 0,607 0,000 0,806 (0,133)	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654 0,000 8,022 0,421 0,579 0,635 -0,062 (0,131)	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000 8,023 0,421 0,579 0,993 -0,001 (0,082)	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022 8,022 0,421 0,579 0,957 0,005 (0,090)	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000 8,024 0,422 0,578 0,727 0,038 (0,109)
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000 8,025 0,422 0,578 0,620 0,063 (0,127) 0,439	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001 7,733 0,390 0,610 0,000 1,251 (0,198) 0,398	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000 7,913 0,408 0,592 0,000 0,439 (0,117) 0,441	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083 8,022 0,421 0,579 0,944 -0,009 (0,130) 0,435	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \\ \\ \sigma_{y0}^2 = 8,023 \\ 8,008 \\ 0,419 \\ 0,581 \\ 0,169 \\ -0,187 \\ (0,136) \\ 0,429 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ 0,319 \\ 0,681 \\ 0,000 \\ 1,335 \\ (0,109) \\ 0,372 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \\ \\ h_{e0}^2 = 0,579 \\ 7,703 \\ 0,383 \\ 0,617 \\ 0,000 \\ 0,923 \\ (0,142) \\ 0,408 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000 7,755 0,393 0,607 0,000 0,806 (0,133) 0,425	0,703 0,538 0,085 (0,137) 0,346 0,654 0,071 0,346 0,654 0,000 8,022 0,421 0,579 0,635 -0,062 (0,131) 0,437	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000 8,023 0,421 0,579 0,993 -0,001 (0,082) 0,277	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022 8,022 0,421 0,579 0,957 0,005 (0,090) 0,285	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000 8,024 0,422 0,578 0,727 0,038 (0,109) 0,291
$\begin{array}{c} pv \\ \beta_2 \\ wg \\ we \\ \beta_{wg} \\ \beta_{we} \\ w_{g0} \\ d_g \\ \end{array}$ $\begin{array}{c} \sigma_{y}^2 \\ h_{g}^2 \\ h_{e}^2 \\ pv \\ \beta_2 \\ we \\ \beta_{wg} \end{array}$	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000 8,025 0,422 0,578 0,620 0,063 (0,127) 0,439 0,561	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001 7,733 0,390 0,610 0,000 1,251 (0,198) 0,398 0,602	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000 7,913 0,408 0,592 0,000 0,439 (0,117) 0,441 0,559	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083 8,022 0,421 0,579 0,944 -0,009 (0,130) 0,435 0,565	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \\ \\ \sigma_{y0}^2 = 8,023 \\ 8,008 \\ 0,419 \\ 0,581 \\ 0,169 \\ -0,187 \\ (0,136) \\ 0,429 \\ 0,571 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ 0,319 \\ 0,681 \\ 0,000 \\ 1,335 \\ (0,109) \\ 0,372 \\ 0,628 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \\ \\ h_{e0}^2 = 0,579 \\ 7,703 \\ 0,383 \\ 0,617 \\ 0,000 \\ 0,923 \\ (0,142) \\ 0,408 \\ 0,592 \\ 0,422 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000 7,755 0,393 0,607 0,000 0,806 (0,133) 0,425 0,575	0,703 0,538 0,085 (0,137) 0,346 0,654 0,071 0,346 0,654 0,000 8,022 0,421 0,579 0,635 -0,062 (0,131) 0,437 0,563	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000 8,023 0,421 0,579 0,993 -0,001 (0,082) 0,277 0,723	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022 8,022 0,421 0,579 0,957 0,005 (0,090) 0,285 0,715	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000 8,024 0,422 0,578 0,727 0,038 (0,109) 0,291 0,709
$\begin{array}{c} pv\\ \beta_2\\ wg\\ we\\ \beta_wg\\ \beta_we\\ wg0\\ dg\\ \end{array}$	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000 8,025 0,422 0,578 0,620 0,063 (0,127) 0,439 0,561 -0,014 0,077	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001 7,733 0,390 0,610 0,000 1,251 (0,198) 0,398 0,602 0,552	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000 7,913 0,408 0,592 0,000 0,439 (0,117) 0,441 0,559 0,211	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083 8,022 0,421 0,579 0,944 -0,009 (0,130) 0,435 0,565 -0,027	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \\ \\ \sigma_{y0}^2 = 8,023 \\ 8,008 \\ 0,419 \\ 0,581 \\ 0,169 \\ -0,187 \\ (0,136) \\ 0,429 \\ 0,571 \\ -0,088 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ 0,319 \\ 0,681 \\ 0,000 \\ 1,335 \\ (0,109) \\ 0,372 \\ 0,628 \\ 0,523 \\ \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \\ \\ \\ h_{e0}^2 = 0,579 \\ 7,703 \\ 0,383 \\ 0,617 \\ 0,000 \\ 0,923 \\ (0,142) \\ 0,408 \\ 0,592 \\ 0,422 \\ 0,501 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000 7,755 0,393 0,607 0,000 0,806 (0,133) 0,425 0,575 0,367	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654 0,000 8,022 0,421 0,579 0,635 -0,062 (0,131) 0,437 0,563 -0,018	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000 8,023 0,421 0,579 0,993 -0,001 (0,082) 0,277 0,723 0,012	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022 8,022 0,421 0,579 0,957 0,005 (0,090) 0,285 0,715 0,028	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000 8,024 0,422 0,578 0,727 0,038 (0,109) 0,291 0,709 -0,016
$\begin{array}{c} pv \\ \beta_2 \\ wg \\ we \\ \beta_{wg} \\ \beta_{we} \\ w_{g0} \\ d_g \\ \end{array}$ $\begin{array}{c} \sigma_{y}^2 \\ h_{g}^2 \\ h_{e}^2 \\ pv \\ \beta_2 \\ we \\ \beta_{wg} \end{array}$	0,297 0,703 0,873 -0,021 (0,133) 0,353 0,647 0,038 -0,059 0,352 0,648 0,000 8,025 0,422 0,578 0,620 0,063 (0,127) 0,439 0,561 -0,014	0,297 0,703 0,120 0,330 (0,212) 0,333 0,667 0,084 0,246 0,332 0,668 -0,001 7,733 0,390 0,610 0,000 1,251 (0,198) 0,398 0,602 0,552 0,699	0,704 0,919 -0,013 (0,123) 0,360 0,640 -0,025 0,012 0,361 0,639 0,000 7,913 0,408 0,592 0,000 0,439 (0,117) 0,441 0,559 0,211 0,228	0,804 0,000 1,592 (0,126) 0,263 0,737 0,422 1,170 0,346 0,654 0,083 8,022 0,421 0,579 0,944 -0,009 (0,130) 0,435 0,565 -0,027 0,018	$\begin{array}{c} 0,705 \\ 0,425 \\ -0,114 \\ (0,142) \\ 0,341 \\ 0,659 \\ -0,045 \\ -0,069 \\ 0,342 \\ 0,658 \\ 0,001 \\ \\ \\ \sigma_{y0}^2 = 8,023 \\ 8,008 \\ 0,419 \\ 0,581 \\ 0,169 \\ -0,187 \\ (0,136) \\ 0,429 \\ 0,571 \\ -0,088 \\ -0,098 \\ \end{array}$	$\begin{array}{c} 0,705 \\ 0,177 \\ -0,164 \\ (0,122) \\ 0,353 \\ 0,647 \\ -0,057 \\ -0,108 \\ 0,354 \\ 0,646 \\ 0,001 \\ \hline \\ Y_4 \\ h_{g0}^2 = 0,421 \\ 7,076 \\ 0,319 \\ 0,681 \\ 0,000 \\ 1,335 \\ (0,109) \\ 0,372 \\ 0,628 \\ 0,523 \\ 0,811 \\ \hline \end{array}$	$\begin{array}{c} 0,707 \\ 0,485 \\ -0,106 \\ (0,151) \\ 0,342 \\ 0,658 \\ -0,086 \\ -0,020 \\ 0,345 \\ 0,655 \\ 0,002 \\ \\ \\ h_{e0}^2 = 0,579 \\ 7,703 \\ 0,383 \\ 0,617 \\ 0,000 \\ 0,923 \\ (0,142) \\ 0,408 \\ 0,592 \\ 0,422 \\ \end{array}$	0,704 0,112 -0,225 (0,141) 0,358 0,642 -0,065 -0,159 0,358 0,642 0,000 7,755 0,393 0,607 0,000 0,806 (0,133) 0,425 0,575 0,367 0,439	0,703 0,538 0,085 (0,137) 0,346 0,654 0,014 0,071 0,346 0,654 0,000 8,022 0,421 0,579 0,635 -0,062 (0,131) 0,437 0,563 -0,018 -0,044	0,703 0,540 0,055 (0,090) 0,201 0,799 0,002 0,053 0,200 0,800 0,000 8,023 0,421 0,579 0,993 -0,001 (0,082) 0,277 0,723 0,012 -0,012	0,667 0,000 1,445 (0,090) 0,231 0,769 0,339 1,106 0,208 0,792 -0,022 8,022 0,421 0,579 0,957 0,005 (0,090) 0,285 0,715 0,028 -0,023	0,704 0,785 0,032 (0,119) 0,213 0,787 0,010 0,022 0,213 0,787 0,000 8,024 0,422 0,578 0,727 0,038 (0,109) 0,291 0,709 -0,016 0,054

(aleatórias) v_1 e v_4 correspondentes às posições 10 e 13. Para \mathbf{Y}_3 , percebe-se o contraste dos índices d_g encontrados da variável com agregado familiar, SNP 4 e a variável preditora v_2 (posição 11), em que o primeiro é positivo e o segundo negativo. Para \mathbf{Y}_4 , note que esta é uma situação relevante para o estudo de blocos haplotípicos, os efeitos sobre \mathbf{Y}_4 foram gerados a partir do SNP 2 e do SNP6 mas como o SNP 2 está em LD com o SNP 3, os dois SNPs mostram efeitos diferenciados; o mesmo acontece com o outro bloco, o efeito mais alto é do SNP 6 mas os outros que pertencem ao bloco, apresentam efeitos mais baixos do que o SNP 6 mas que merecem ser estudados como efeito de haplótipo.

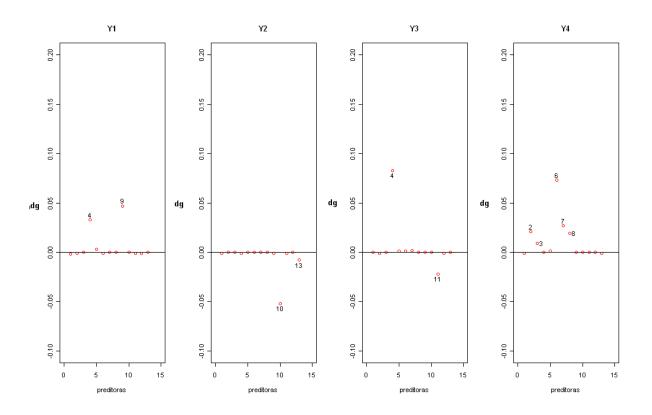


Figura 5.13 Indices d_g para as simulações complementares.

Finalmente, aumentando a dimensão do espaço de pesquisa das variáveis preditoras, uma simulação complementar foi realizada considerando uma amostra de 300 SNPs (gerados no pacote SimPed). Neste caso, foi gerada uma variável resposta \mathbf{Y} com herdabilidade de 32% e com efeito (1.5) do SNP 31 e do SNP 208 (o qual está em LD com os SNPs 206 e 207). Na Figura (5.14) apresentam-se os índices d_g para cada um dos 300 SNPs.

Percebe-se que o índice discrimina os efeitos de cada SNP sobre a variável resposta, indicando as possíveis regiões que merecem um estudo mais detalhado. Esta é a situação esperada quando analisado um segmento do cromossomo em que existam determinantes genéticos associados à regulação da variável resposta em estudo.

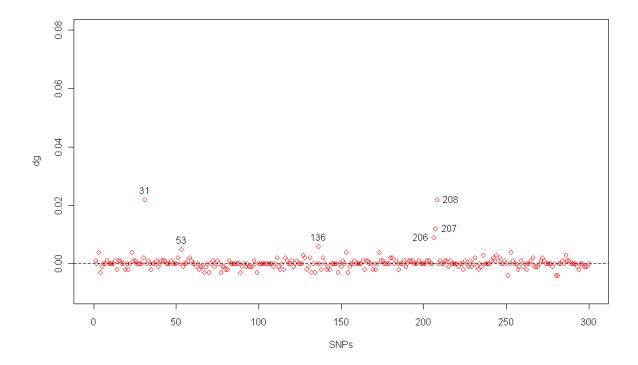


Figura 5.14 Indices d_g para 300 SNPs.

Capítulo 6

Considerações finais

No contexto da metodologia referente ao gráfico da variável adicionada na classe de modelos lineares mistos introduzida por Hodges (1994) e estendida em Hilden-Minton (1995), nesta tese é obtida uma decomposição do estimador do efeito da variável (preditora) adicionada em termos associados a componentes de variância. Em particular, tal decomposição foi considerada sob a estrutura de um modelo linear misto poligênico comumente adotado no mapeamento de genes em delineamentos que usam dados de famílias e plataformas de marcadores moleculares do tipo SNP. Neste caso, o modelo considera dois componentes de variância, um devido ao efeito aleatório poligênico (que está associado à correlação familiar) e outro devido ao componente de erro, sendo a variável preditora definida como o marcador SNP.

Na decomposição proposta a contribuição de cada componente ao efeito fixo é especificada por pesos w_g e w_e , dados nas equações (3.57) e (3.58), respectivamente. Com esta técnica é possível discriminar SNPs (variáveis preditoras) com contribuições relevantes especificamente no componente poligênico, o que tem um grande impacto em mapeamento genético, já que permite identificar regiões genômicas que influenciam variáveis resposta associadas à estrutura familar. Por outro lado, podem-se identificar SNPs com efeitos principalmente associados ao componente residual, sem contribuição poligênica, os quais estão segregando na população geral, independente de qualquer estrutura familiar. Este resultado estende a proposta de Aulchenko et al. (2007) que filtra SNPs principalmente associados ao componente residual e não leva em conta a possível associação com efeitos poligênicos.

Na prática, o mapeamento de genes considera plataformas de SNPs que incluem um número muito grande de variáveis preditoras a serem estudadas. No intuito de tornar mais operacional a discriminação de SNPs em espaços de alta dimensão via a técnica proposta nesta tese, foi apresentado um índice d_g obtido a partir de estimativas do componente de variância poligênico sob os modelos reduzido e completo (que inclui a variável adicionada). Os estudos de simulação feitos nesta tese ilustraram a utilidade deste índice na discriminação dos SNPs.

Ainda, um nível adicional de decomposição é proposto considerando o componente poligênico do efeito total da variável adicionada, o qual especifica a contribuição de cada família em termos de pesos $w_{g,f}$ e coeficientes angulares $\beta_{2,f}^g$. Nota-se que com esta técnica é possível discriminar famílias caudais, a saber, famílias que apresentam altos coeficientes angulares (em módulo). É interessante observar que famílias caudais com coeficientes $\beta_{2,f}^g$ altos e com sinal oposto a β_2 (o efeito da variável adicionada) recebem peso $w_{g,f}$ baixo. A composição destas duas medidas, o peso e o coeficiente angular, de cada família foi feita neste trabalho por meio do índice I_f , o qual permite especificar a contribuição de cada família diretamente ao efeito total da variável adicionada, permitindo assim identificar famílias caudais. Estas famílias merecem um estudo mais detalhado já que podem apresentar alelos de risco que causam alterações no fenótipo observado e que seguem um padrão de regulação gênica específico e diferente do restante da população.

Uma contribuição importante da aplicação desta decomposição refere-se à possibilidade de identificação de blocos de variáveis adicionadas (SNPs), posicionadas sequencialmente, com efeito poligênico relevante. Em mapeamento de genes isto caracteriza o efeito de um possível haplótipo, que pode ser formado por SNPs que, individualmente, não mostram efeito significante mas que se analisados como bloco mostram efeito. Vale ressaltar, que neste trabalho, quando um bloco é identificado foi proposto uma particular parametrização para se estudar o efeito da variável haplotípica dentre outras que podem ser consideradas.

Nesta tese vários estudos de simulação foram realizados considerando diferentes cenários para os parâmetros envolvidos no modelo linear misto. No caso das variáveis preditoras

(SNPs) adotou-se a estrutura de famílias sob a qual os descendentes recebem valores da variável condicionados aos valores dos pais. Ainda neste caso, adotou-se uma situação hipotética na qual os valores dos SNPs não atenderam à estrutura familiar. Os resultados mostraram que o efeito do SNP absorve variabilidade devida à agregação familiar.

No caso das variáveis resposta, a simulação foi feita adotando-se o modelo normal multivariado com diferentes valores para os parâmetros da média e dos componentes de variância. Especial atenção deve ser dada para o comportamento da decomposição proposta para diferentes valores de herdabilidade. Os resultados mostraram que o critério proposto (baseado no índice d_g) é útil em discriminar SNPs, tanto individualmente como em blocos, em situações de herdabilidades moderadas (em torno de 30%) a altas. É importante ressaltar que o índice d_g , apesar de fazer uma discriminação descritiva, ele é superior a uma análise baseada diretamente no teste de β_2 no sentido de identificar SNPs com algum efeito de agregação familiar, o que pode ser usado para análises haplotípicas e análises de famílias influentes.

Os comandos no R desenvolvidos para as análises usadas nesta tese estarão disponíveis para usuários interessados que solicitem ao autor, sendo que uma parte da programação está apresentada no Apêndice A. Em relação ao problema do tempo computacional requerido para as análises, alguns autores têm abordado este problema, por exemplo, Meyer & Tier (2012), em que os autores particionam a matriz de efeitos fixos em duas, uma correspondente às covariáveis gerais e a outra correspondente especificamente aos efeitos dos SNPs envolvidos, o que implica em uma particão no sistema de equações para se obter soluções em um modelo linear misto. A estratégia proposta tem sido implementada no pacote para modelos lineares mistos WOMBAT (Meyer 2007). Em situações de análise de dados reais em que o efeito do SNP deve ser avaliado na presença de outras covariáveis, a implementação computacional da decomposição proposta nesta tese pode ser feita incorporando estratégias como esta.

A seguir destacam-se alguns tópicos relevantes para pesquisas futuras.

- Aplicar as metodologias estatísticas desenvolvidas no trabalho a dados reais do Laboratório de Genética e Cardiologia Molecular obtidos do Projeto "Corações de Baependi".
- Os critérios obtidos nesta tese para a discriminação e seleção de variáveis preditoras (SNPs) foram apresentados sob um contexto descritivo, sendo necessário se obter cortes com a atribuição de significância estatística para melhor atender à sua utilização prática. Neste sentido, o trabalho de Nobre e Singer (2011), no contexto de modelos lineares mistos, apresentam uma proposta de identificação de unidades amostrais influentes que pode ser adaptada e estendida para a situação tratada nesta tese.
- Generalizar a teoria associada com o gráfico da variável adicionada para o caso em que sejam adicionadas conjuntamente duas ou mais variáveis preditoras ao modelo. Isto permitirá estudar o efeito de interações (epistasia) entre marcadores SNPs, sendo útil em situações em que apesar dos efeitos individuais não serem importantes o efeito de interação o é.
- Generalizar o critério proposto para situações multivariadas, envolvendo mais de uma variável resposta. Em Genética, isto permitirá selecionar SNPs com efeito pleiotrópico.
- Os resultados desenvolvidos nesta tese foram obtidos sob a suposição de normalidade
 e, neste caso, famílias caudais constribuem pouco (recebem peso baixo) para o efeito
 do SNP. Assim, há interesse em estender a análise do gráfico da variável adicionada
 em modelos mistos para classes robustas de distribuições, como a classe das elípticas,
 sob a qual o efeito de famílias caudais pode ser melhor absorvido pelo componente
 da média.

Apêndice A

Rotinas

A.1 Código para a geração de marcadores SNPs no aplicativo SimPed

Na geração de marcadores, são necesários dois arquivos, a saber **pedin.pre** e **in- put.dat**.

Arquivo **pedin.pre**: este arquivo contém a estrutura das famílias para as quais são gerados os marcadores SNPs. Como ilustração considere uma família com 13 indivíduos. As 4 colunas iniciais descrevem a estrutura familiar (família, indivíduo, pai, mãe). A coluna 5 descreve o género do indivíduo e, na coluna 6, o número 1 indica que para esse indivíduo todos os resultados serão gerados.

```
1
      1
                    0
                         1
                               1
      2
                    0
                         2
             0
                               1
      3
             1
                    2
                         1
                               1
1
      4
             0
                    0
                         1
                               1
      5
                         2
1
             0
                    0
                               1
      6
                         2
                    5
                               1
      7
1
             3
                    6
                         1
                               1
      8
             3
                    6
                         1
                               1
      9
             3
                    6
                         1
                               1
                         2
1
     10
             3
                    6
                               1
                         2
1
             3
                    6
     11
                               1
                         2
1
     12
             3
                    6
                               1
1
     13
             3
                    6
                         2
                               1
```

Neste arquivo qualquer estrutura familiar pode ser descrita. Também é possível gerar marcadores para indivíduos não relacionados, sendo suficiente deixar 0 nas colunas correspondentes ao pai e a mãe de cada indivíduo.

Arquivo input.dat: este arquivo contém a informação genética específica para a geração dos marcadores. Como ilustração, a seguir apresenta-se o arquivo utilizado na geração dos nove SNPs envolvidos no estudo de simulação apresentado no Capítulo 5. Por exemplo, a linha 1 contém o arquivo com a informação da estrutura familiar (pedin.pre) e o nome do arquivo de saída (file.out). Na linha 4, especifica-se uma réplica só, pois são consideradas unicamente as 51 famílias. Na linha 5 especifica-se o número total de marcadores. Na linha 9 pede-se para gerar o primeiro SNP com probabilidade de 0,70 para o alelo mais frequente. Nas linhas 10 a 14 especifica-se a geração do SNP 2 e 3 em LD (note que na linha 11 é possivel especificar as probabilidades de cada haplótipo). Nas linhas 15 a 17 especifica-se a geração dos SNPs 4 e 5 independentes e, nas linhas 18 a 23 pede-se para gerar os SNPs 6, 7 e 8 em LD e finalmente nas linhas 24 e 25 pede-se para gerar o SNP 9 independente dos outros SNPs.

```
pedin.pre file.out << name of pedigree file, name of output file
   23221 1601 21001
                       << three random seeds
   0 0 << # of columns for affection status/
            quantitative trait, autosomal data to be generated
   1<<number of replicates
   9 1 << Total # of marker loci, # of times pattern to be repeated
   1 << "1" recomb fraction, "2" Kosambi map distance & "3" Haldane map distance
   1 0,000 << # of ? values, ? values
7
   2 1 1 << genótipos a serem gerados
   2 0,7 << # of alleles, allele frequencies
   1 3 2 1 << "1" for haplotypes, # of haplotypes, # of marker loci,
10
              # of times pattern repeated
   0,60 0,30 0,10 << the frequency for each haplotype
12
   1 1
         < << observed alleles for each haplotype
   1 2
13
   2 2
14
15
   2 2 1 << genótipos a serem gerados, um marcador, modelo repetido uma vez
   2 0,7 << # of alleles, allele frequencies
   2 0,75 << # of alleles, allele frequencies
18
   1 4 3 1 << "1" for haplotypes, # of haplotypes, # of marker loci,
              # of times pattern repeated
   0,30 0,25 0,25 0,20 << the frequency for each haplotype
20
   1 1 1 < << observed alleles for each haplotype
```

```
21  2 1 2
22  1 1 2
23  2 2 2
24  2 1 1 << genótipos a serem gerados
25  2 0,75 << # of alleles, allele frequencies</pre>
```

O programa SimPed é flexível na geração de dados, podendo, por exemplo, especificar o número de alelos do marcador, especificar haplótipos com suas respectivas probabilidades, gerar bancos de dados combinando SNPs independentes com SNPs em Desequilíbrio de Ligação (haplótipos), como foi o caso na geração dos 9 SNPs utilizados no estudo de simulação. Para mais detalhes ver o Manual SimPed (Leal et al., 2005).

A.1.1 Codificação das saídas obtidas no aplicativo SimPed

As saídas do programa SimPed no arquivo **file.out** estão codificadas como 1 e 2, esta codificação é dada em duas colunas para cada SNP. O código a seguir classifica cada SNP da forma clássica, como 0, 1 e 2 segundo a descrição dada na equação (2.6) do Capítulo 1.

```
mat: saida do SimPed com a codificação dos marcadores.
#Aplica-se o critério
#2 2 recebe 2
#2 1 recebe 1
#1 2 recebe 1
#1 1 recebe 0
red<-function(a)
\{if (a[1]==2)
{
if(a[2]==2)
{
b<-2
}
if(a[2]==1)
{
b<-1
}
if (a[1]==1)
{
```

```
if(a[2]==2)
{
b<-1
if(a[2]==1)
b<-0
}
}
return(b)
}
#divide o vetor original nas duplas a avaliar (para o caso de marcadores com 2 alelos)
div<-function(c)</pre>
number1<-2 #numero de colunas a serem tomadas
cont<-1
aux1=length(c)
while(cont<=aux1)
{
aux2=reduccion(c[cont:(cont+(number1-1))])
if(cont==1){
d < -c(aux2)
}
else
{d<-c(d,aux2)}
}
cont<-cont+number1</pre>
return(d)
}
reduzida<-t(apply(mat,1,div))
```

Quando são obtidos os blocos de SNPs consecutivos (dois, três, quatro ou mais SNPs), o código a seguir faz a codificação para a obtenção das novas variáveis haplotípicas \mathbf{X}_h (de acordo à codificação sugerida no Capítulo 4) para serem estudadas mediante a metodologia respectiva.

```
hap: arquivo com os SNPs a serem codificados
# aplica-se o critério
```

```
#para todo 2 recebe 2
#se há algum 1 recebe 1
#se há algum 0 recebe 0
red<-function(a)</pre>
zero=any(a==0)
um=any(a==1)
dois=all(a==2)
if (zero==TRUE)
{
b<-0
}
else
{ if (um==TRUE)
{
b<-1
}
else
{ if(dois==TRUE){
b<-2
}
else
{
b<-NA
}
}
}
return(b)
#separa o vetor original no número de colunas a avaliar
separa<-function(c,nume=4)</pre>
number1<-nume #número de colunas a tomar
cont<-1
aux1=length(c)
while(cont<=aux1)
aux2=reduccion(c[cont:(cont+(number1-1))])
```

```
if(cont==1){
d<-c(aux2)
}
else
{d<-c(d,aux2)
}
cont<-cont+number1
}
return(d)
}
# le os datos da matriz
reducida<-t(apply(hap,1,separa,nume=3))
reducida</pre>
```

Apêndice B

Anexos do estudo de simulação

B.1 Decomposição do efeito das variáveis preditoras

Nesta seção, apresentam-se os resultados (para a média) de 500 réplicas para os diferentes cenários das variáveis resposta \mathbf{Y}_{kl} geradas com variância total $\sigma_y^2 = 2$, $\sigma_y^2 = 15$, $\sigma_y^2 = 50$ e $\sigma_y^2 = 100$. Os resultados são descritos nas Tabelas B.1, B.2, B.3 e B.4, respectivamente.

Para os \mathbf{Y}_{kl} gerados com variância total de $\sigma_y^2 = 2$ e $\sigma_y^2 = 15$, os resultados são próximos dos resultados apresentados no Capítulo 5 para a variância $\sigma_y^2 = 7$. Segue uma descrição mais detalhada dos resultados para as variáveis \mathbf{Y}_{kl} que foram geradas com $\sigma_y^2 = 50$. De acordo aos dados da Tabela B.3, em relação ao ítem 1, também ao ser incluido o SNP 7 no ajuste do modelo completo a variabilidade é absorvida justamente da variabilidade poligênica conduzindo a uma diminuição na herdabilidade poligênica. Como ilustração toma-se o caso em que a herdabilidade é de 32% e o efeito causado pelo SNP 7 foi de 3,0. Para este caso, a variabilidade total no modelo reduzido é de 53,853 e no modelo completo é de 50,528. A variabilidade poligênica passou de 19,558 para 15,913 e a variabilidade residual passou de 34,298 para 34,615 o qual implica específicamente em redução da herdabilidade poligênica (de 0,363 a 0,314).

Referente as estimativas do parâmetro β_2 , para todos os valores das herdabilidades, nota-se que para os efeitos mais próximos de zero, ($\beta_2=0,1$ e $\beta_2=0,8$) no foi significativo e as estimativas ficaram "longe" do verdadeiro valor do parâmetro; só foi significativo ($p-valor \leq 0,005$) para os efeitos de 1,5 e 3,0 em que as estimativas foram mais precisas.

Em relação ao índice d_g (última linha de cada cenário), este é positivo e as magnitudes em geral são baixas. Pode-se concluir que para a variabilidade alta da variável resposta e efeitos β_2 "pequenos", a variabilidade poligênica absorvida pelo SNP é baixa. Resultados

análogos para o caso en que a variável resposta foi gerada com variância total $\sigma_y^2 = 100$ são descritos na Tabela B.4.

De maneira geral, nota-se que para todos os cenários continua a relação entre a herdabilidade h_g^2 e o peso w_g descrita no Capítulo 5 para o caso em que $\sigma_y^2 = 7$, isto é, para a herdabilidade de 8%, w_g é em média de 0,13; para a herdabilidade de 32% w_g é em média de 0,34 e para a herdabilidade de 70% w_g é em média de 0,63.

Em relação ao índice d_g , este assume valores maiores ou iguais a zero, o qual indica que o SNP 7 absorve parte da variabilidade poligênica.

B.2 Análises no caso haplotípico

Os SNPs, 6, 7 e 8 foram gerados em Desequilíbrio de Ligação, formando um bloco haplotípico. O intuito é verificar, mediante simulações, o reconhecimento do bloco a través dos seus efeitos para uma posterior análise conjunta, como especificado no Capítulo 4. Para estas análises, cada variável resposta \mathbf{Y}_{kl} foi gerada com efeito somente do SNP 7 e o intuito é ver, por meio dos diferentes cenários, se os efeitos dos SNPs adjacentes (SNP 6 e SNP 8) sobre a variável resposta também são reconhecidos.

A seguir apresentam-se os cenários em que os \mathbf{Y}_{kl} foram gerados com variância total $\sigma_y^2 = 50$ e para cada uma das herdabilidades (8%, 32% e 70%). Os dados estão organizados em três tabelas; na Tabela B.5 tem-se os resultados para os \mathbf{Y}_{kl} simulados com herdabilidade de 8% na Tabela B.6 para os \mathbf{Y}_{kl} simulados com herdabilidade de 32% e na Tabela B.7 para os \mathbf{Y}_{kl} gerados com herdabilidade de 70%, variando sobre todos os diferentes efeitos do SNP 7, representado por β_2 .

Para estes cenários o bloco de SNPs é diferenciado somente para efeitos β_2 "altos", correspondendo a $\beta_2 = 1, 5$ e $\beta_2 = 3, 0$. Especificamente, quando o efeito do SNP 7 sobre a variável resposta é alto, os SNPs adjacentes (SNP 6 e SNP 8) também refletem o efeito. Uma forma de reconhecer dito efeito é por meio dos valores do índice d_g , nota-se que estes valores são mais altos do que os valores das outras variáveis envolvidas. Como ilustração, observa-se na Tabela B.6 em que o efeito para o SNP 7 foi de $\beta_2 = 1, 5$ que o valor do índice d_g é mais alto do que o índice para as otras variáveis, isto é, para o SNP 6 $d_g = 0,014$, para o SNP 7 tem-se $d_g = 0,010$ e para o SNP 8 $d_g = 0,003$.

Tabela B.1 Resultados (em média) para variáveis resposta geradas com variância total $\sigma_y^2=2$ e covariável o SNP 7.

		$\beta_2 = 0, 1$			$\beta_2 = 0, 8$			$\beta_2 = 1, 5$			$\beta_2 = 3, 0$			
h_g^2	parâm	s	M1	M2	s	M1	M2	s	M1	M2	s	M1	M2	
0,08	σ_u^2	2	1,981	1,974	2	2,182	1,983	2	2,684	1,971	2	4,945	2,008	
	σ_a^2	0,16	0,163	0,155	0,16	0,295	0,126	0,16	0,876	0,166	0,16	3,076	0,143	
	σ_{c}^{2}	1,84	1,818	1,818	1,84	1,889	1,856	1,84	1,808	1,805	1,84	1,872	1,866	
	h_{a}^{2}	0,08	0,082	0,08	0,08	0,135	0,063	0,08	0,326	0,083	0,08	0,623	0,07	
	$egin{array}{c} \sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2 \end{array}$	0,92	0,918	0,92	0,92	0,865	0,937	0,92	0,674	0,917	0,92	0,377	0,93	
	μ	- / -	- / -	120,008	- /-	- /	120,023	- / -	- /	120,002	- /-	-,	120,0	
	pv			0,103			0			0			0	
	β_2			0,137			0,799			1,489			3,01	
	F 2			(0,07)			(0,069)			(0,069)			(0,07	
	wg			0,135			0,114			0,139			0,12	
	we			0,865			0,886			0,861			0,87	
	β_{wg}			0,116			0,708			1,277			2,64	
	Pwg			0,061			0,061			0,06			0,06	
	В			0,001			0,001			0,211			0,36	
	β_{we}			0,008			0,000			0,006			0,00	
	10			0,508			0,007			0,516			0,45	
	lg le			0,995			0,998			0,994			0,43	
	w_{go}			0,141			0,203			0,365			0,57	
	w_{e0}			0,859			0,797 0,087			0,635			0,42	
	d_g			0,005			0,087			0,224			0,45	
),32	$egin{array}{c} \sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2 \end{array}$	2	1,994	1,989	2	2,246	2,04	2	2,721	2,018	2	4,853	1,95	
	σ_g^2	0,64	0,582	0,576	0,64	0,865	0,677	0,64	1,291	0,617	0,64	3,421	0,6	
	σ_e^2	1,36	1,412	1,411	1,36	1,383	1,36	1,36	1,429	1,403	1,36	1,432	1,32	
	h_g^2	0,32	0,293	0,291	0,32	0,382	0,332	0,32	0,474	0,304	0,32	0,705	0,32	
	h_e^2	0,68	0,707	0,709	0,68	0,618	0,668	0,68	0,526	0,696	0,68	0,295	0,67	
	μ			119,988			119,988			120			120,0	
	pv			0,199			0			0			0	
	β_2			0,114			0,811			1,481			3,00	
				0,07			0,072			0,07			0,0	
	wg			0,338			0,369			0,349			0,36	
	we			0,662			0,631			0,651			0,63	
	β_{wg}			0,075			0,515			0,963			1,92	
				0,05			0,05			0,05			0,0	
	β_{we}			0,039			0,295			0,517			1,08	
				0,021			0,023			0,02			0,02	
	lg			0,834			0,862			0,846			0,85	
	le			0,977			0,973			0,975			0,97	
	w_{go}			0,34			0,407			0,47			0,6	
	w_{e0}			0,66			0,593			0,53			0,3	
	d_g			0,001			0,037			0,119			0,27	
0,7	σ^2	2	1,963	1,957	2	2,225	2,015	2	2,717	1,968	2	4,901	1,96	
٠,٠	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h^2$	1,4	1,347	1,34	1,4	1,6	1,394	1,4	2,13	1,381	1,4	4,281	1,37	
	σ^{g}_{2}	0,6	0,615	0,617	0,6	0,624	0,621	0,6	0,588	0,588	0,6	0,619	0,59	
	$h^{\frac{e}{2}}$	0,7	0,686	0,686	0,7	0,719	0,693	0,7	0,784	0,698	0,7	0,872	0,69	
	h_e^g	0,3	0,314	0,314	0,3	0,281	0,307	0,3	0,216	0,302	0,3	0,128	0,30	
	μ	0,3	0,314	119,983	0,5	0,201	119,988	0,5	0,210	119,999	0,5	0,120	120,0	
				0,255			0			0			0	
	pv													
	β_2			0,1			0,807			1,513			3,01	
				0,07			0,07			0,07			0,0	
	wg			0,625			0,63			0,637			0,63	
	we			0,375			0,37			0,363			0,36	
	β_{wg}			0,03			0,301			0,547			1,10	
				0,029			0,029			0,028			0,0	
	β_{we}			0,07			0,509			0,968			1,90	
				0,041			0,041			0,042			0,04	
	lg			0,965			0,967			0,969			0,96	
	le			0,887			0,886			0,88			0,88	
	w_{go}			0,626			0,652			0,708			0,80	
	w_{e0}			0,374			0,348			0,292			0,19	
	d_g			0			0,021			0,072			0,16	

Tabela B.2 Resultados (em média) para variáveis resposta geradas com variância total $\sigma_y^2=15$ e covariável o SNP 7.

			$\beta_2 = 0, 1$	-		$\beta_2 = 0, 8$	1		$\beta_2 = 1, 5$;		$\beta_2 = 3,0$)
h_g^2	parâm	s	M1	M2	s	M1	M2	s	M1	M2	s	M1	M2
0,08	σ_{v}^{2}	15,000	15,607	15,591	15,000	15,199	15,019	15,000	15,518	14,789	15,000	17,811	14,978
	σ_q^2	1,200	1,359	1,359	1,200	1,188	0,961	1,200	1,686	1,106	1,200	3,694	1,028
	σ_e^2	13,800	14,250	14,233	13,800	14,014	14,056	13,800	13,832	13,685	13,800	14,116	13,953
	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	0,080	0,086	0,087	0,080	0,078	0,065	0,080	0,108	0,075	0,080	0,207	0,068
	$h_e^{\bar{2}}$	0,920	0,914	0,913	0,920	0,922	0,935	0,920	0,892	0,925	0,920	0,793	0,932
	μ			119,968			120,084			119,966			120,010
	pv			0,486			0,006			0,000			0,000
	β_2			0,138			0,722			1,516			2,978
				0,190			0,185			0,185			0,186
	wg we			0,149 $0,851$			0,114 0,886			0,129 $0,871$			0,123 $0,877$
	β_{wg}			0,126			0,625			1,331			2,612
	r wy			0,168			0,170			0,165			0,169
	β_{we}			0,013			0,096			0,185			0,367
				0,020			0,016			0,016			0,017
	lg			0,539			0,450			0,479			0,481
	le			0,994			0,997			0,994			0,998
	w_{go}			0,148			0,133			0,172			0,271
	w_{e0}			0,852			0,867			0,828			0,729
	d_g			0,000			0,020			0,046			0,148
0,32	σ_y^2	15,000	14,679	14,666	15,000	15,221	14,948	15,000	15,736	15,013	15,000	18,244	15,216
	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	4,800	4,693	4,678	4,800	4,945	4,722	4,800	5,381	4,667	4,800	8,218	5,099
	σ_e^2	10,200	9,987	9,986	10,200	10,276	10,226	10,200	10,353	10,350	10,200	10,025	10,118
	h_g^2	0,320	0,317	0,317	0,320	0,324	0,317	0,320	0,340	0,309	0,320	0,450	0,334
		0,680	0,683	0,683	0,680	0,676	0,683	0,680	0,660	0,691	0,680	0,550	0,666
	μ			120,034			119,963			120,010			120,059
	pv			0,422 $0,144$			0,001 $0,924$			0,000 1,486			0,000 $3,037$
	β_2			0,144			0,196			0,197			0,197
	wg			0,359			0,359			0,354			0,372
	we			0,641			0,641			0,646			0,628
	β_{wg}			0,091			0,606			0,959			1,901
	3			0,134			0,136			0,136			0,134
	β_{we}			0,055			0,319			0,529			1,137
				0,061			0,060			0,060			0,063
	lg			0,853			0,853			0,847			0,860
	le			0,974			0,976			0,977			0,972
	w_{g0}			0,360 0,640			0,366 $0,634$			0,376 $0,624$			0,454 0,546
	w_{e0}			0,040			0,007			0,024			0,082
	d_g			0,001			0,001			0,022			0,002
0,7	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	15,000	15,199	15,190	15,000	15,313	15,044	15,000	15,896	15,082	15,000	17,794	14,790
	σ_g^2	10,500	10,648	10,647	10,500	10,570	10,281	10,500	11,358	10,560	10,500	13,217	10,196
	σ_e^2	4,500	4,553	4,544	4,500	4,743	4,763	4,500	4,536	4,523	4,500	4,577	4,594
	h_g^-	0,700	0,698	0,700	0,700	0,689	0,683 0,317	0,700	0,713	0,701	0,700	0,739	0,689
		0,300	0,302	0,300 119,961	0,300	0,311	119,949	0,300	0,287	0,299 $120,001$	0,300	0,261	0,311 $119,990$
	$\mu \ pv$			0,461			0,002			0,000			0,000
	β_2			0,171			0,868			1,583			3,025
	. 2			0,197			0,197			0,195			0,193
	wg			0,638			0,621			0,637			0,627
	we			0,362			0,379			0,363			0,373
	β_{wg}			0,079			0,310			0,581			1,125
	_			0,076			0,079			0,076			0,077
	β_{we}			0,093			0,556			1,003			1,899
	1			0,120			0,115			0,118			0,114
	lg le			0,967 $0,880$			0,964 0,889			0,967 $0,881$			0,965 $0,886$
	w_{go}			0,637			0,628			0,648			0,673
	w_{e0}			0,363			0,372			0,352			0,327
	d_g			0,000			0,005			0,011			0,045

Tabela B.3 Resultados (em média) para variáveis resposta geradas com variância total $\sigma_y^2=50$ e covariável o SNP 7.

			$\beta_2 = 0, 1$			$\beta_2 = 0, 8$			$\beta_2 = 1, 5$	i		$\beta_2 = 3, 0$)
h_g^2	parâm	s	M1	M2									
0,08	σ_{u}^{2}	50,000	50,846	50,791	50,000	50,081	49,781	50,000	50,648	49,815	50,000	53,389	50,305
	σ_a^2	4,000	3,560	3,477	4,000	3,696	3,375	4,000	3,785	2,848	4,000	5,751	2,763
	σ_e^2	46,000	47,288	47,314	46,000	46,385	46,406	46,000	46,864	46,965	46,000	47,636	47,544
	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	0,080	0,070	0,067	0,080	0,074	0,068	0,080	0,073	0,057	0,080	0,106	0,054
	h_e^2	0,920	0,930	0,933	0,920	0,926	0,932	0,920	0,927	0,943	0,920	0,894	0,946
	μ			119,865			119,965			120,006			119,821
	pv			0,406			0,023			0,000			0,000
	β_2			0,303			0,935			1,591			3,100
				0,339			0,338			0,335			0,336
	wg			0,120			0,120			0,101			0,098
	we			0,880			0,880			0,899			0,902
	β_{wg}			0,259			0,828			1,416			2,796
	0			0,310			0,305			0,310			0,313
	β_{we}			0,044			0,108			0,176			0,304
	,			0,030			0,030			0,025			0,024
	lg			0,458			0,460			0,398			0,388
	le			0,998 $0,122$			0,997 $0,128$			0,997 $0,128$			0,996 $0,171$
	w_{go}			0,122			0,128			0,128			0,171
	w_{e0}			0,001			0,009			0,027			0,073
	$\frac{d_g}{d_g}$												
0,32	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	50,000	50,052	50,025	50,000	49,914	49,502	50,000	50,045	49,235	50,000	53,853	$50,\!528$
	σ_g^2	16,000	15,356	15,344	16,000	16,632	16,198	16,000	16,380	15,618	16,000	19,558	15,913
	σ_e^2	34,000	34,696	34,683	34,000	33,284	33,302	34,000	33,666	33,616	34,000	34,298	34,615
	h_{g}^{2}	0,320	0,305	0,305	0,320	0,335	0,329	0,320	0,325	0,318	0,320	0,363	0,314
		0,680	0,695	0,695	0,680	0,665	0,671	0,680	0,675	0,682	0,680	0,637	0,686
	μ			119,943			119,852			119,990			119,803
	pv			0,628			0,058			0,004			0,000
	β_2			0,116			0,935			1,545			3,129
				0,359 $0,351$			0,357 $0,367$			0,355 $0,359$			0,362 $0,358$
	wg			0,649			0,633			0,641			0,642
	we			0,049			0,661			0,996			1,956
	β_{wg}			0,055			0,243			0,330			0,252
	β_{we}			0,062			0,374			0,553			1,173
	r we			0,106			0,112			0,109			0,110
	lg			0,848			0,859			0,853			0,854
	le			0,976			0,973			0,975			0,974
	w_{go}			0,350			0,373			0,366			0,392
	w_{e0}			0,650			0,627			0,634			0,608
	d_g			0,000			0,005			0,006			0,034
0,7	σ^2	50,000	50,115	50,015	50,000	50,352	50,147	50,000	50,405	49,781	50,000	53,379	50,725
- , -	σ_{-}^{y}	35,000	35,685	35,580	35,000	35,188	34,937	35,000	34,830	34,228	35,000	39,144	36,458
	σ_{-}^{2}	15,000	14,432	14,435	15,000	15,163	15,210	15,000	15,577	15,554	15,000	14,232	14,268
	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	0,700	0,710	0,710	0,700	0,698	0,696	0,700	0,690	0,685	0,700	0,734	0,719
	h_{s}^{2}	0,300	0,290	0,290	0,300	0,302	0,304	0,300	0,310	0,315	0,300	0,266	0,281
	μ	•	,	120,009	,	,	120,055	,	*	120,000	,	*	120,071
	pv			0,411			0,186			0,000			0,000
	β_2			-0,078			0,627			1,385			2,818
	-			0,353			0,354			0,355			0,356
	wg			0,644			0,634			0,626			0,653
	we			0,356			0,366			0,374			0,347
	β_{wg}			-0,044			0,199			0,534			0,977
				0,136			0,143			0,143			0,134
	β_{we}			-0,038			0,430			0,851			1,841
				0,218			0,215			0,211			0,222
	lg			0,969			0,967			0,965			0,969
	le			0,876			0,883			0,886			0,869
	w_{go}			0,646			0,635			0,628			0,664
	w_{e0}			0,354			0,365			0,372			0,336
	d_g			0,000			0,001			0,003			0,012

Tabela B.4 Resultados (em média) para variáveis resposta geradas com variância total $\sigma_y^2=100$ e covariável o SNP 7.

			$\boldsymbol{\beta}_2 = 0, 1$			$\boldsymbol{\beta}_2 = 0, 8$			$\beta_2=1,5$			$\boldsymbol{\beta}_2 = 3, 0$	
h_g^2	parâm	s	M1	M2	s	M1	M2	s	M1	M2	s	M1	M2
0,08	σ_u^2	100,000	101,021	100,934	100,000	100,132	99,937	100,000	102,373	101,383	100,000	102,766	99,982
	σ_{a}^{2}	8,000	8,036	7,987	8,000	7,113	6,991	8,000	6,393	5,502	8,000	9,095	6,741
	σ_e^2	92,000	92,984	92,948	92,000	93,018	92,947	92,000	95,980	95,881	92,000	93,673	93,239
	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	0,080	0,077	0,077	0,080	0,071	0,071	0,080	0,063	0,054	0,080	0,087	0,067
	$h_e^{\overline{2}}$	0,920	0,923	0,923	0,920	0,929	0,929	0,920	0,937	0,946	0,920	0,913	0,933
	μ			119,963			120,149			119,962			119,862
	pv			0,458			0,295			0,003			0,000
	β_2			-0,093			0,673			1,740			2,949
				0,483			0,478			0,478			0,478
	wg			0,135			0,119			0,095			0,119
	we			0,865 -0,079			0,881 0,613			0,905 1,589			0,881 $2,610$
	β_{wg}			0,434			0,433			0,443			0,434
	β_{we}			-0,014			0,455			0,148			0,339
	rwe			0,047			0,042			0,032			0,041
	lg			0,504			0,437			0,371			0,450
	le			0,996			0,996			0,997			0,996
	w_{go}			0,135			0,121			0,110			0,147
	w_{e0}			0,865			0,879			0,890			0,853
	d_g			0,000			0,003			0,015			0,029
0,32	σ_y^2 σ_g^2 σ_e^2 h_g^2	100,000	103,005	102,888	100,000	97,587	97,348	100,000	101,438	100,780	100,000	104,203	100,999
	σ_g^2	32,000	35,662	35,567	32,000	28,609	28,465	32,000	34,601	34,114	32,000	36,657	33,208
	σ_e^2	68,000	67,342	67,319	68,000	68,979	68,879	68,000	66,837	66,664	68,000	67,546	67,790
	h_g^2	0,320	0,347	0,346	0,320	0,292	0,292	0,320	0,341	0,338	0,320	0,351	0,330
		0,680	0,653	0,654	0,680	0,708	0,708	0,680	0,659	0,662	0,680	0,649	0,670
	μ			120,186			120,002			120,157			119,907
	pv			0,408 -0,052			0,141 $0,864$			0,017 $1,445$			0,000 $3,074$
	β_2			0,515			0,504			0,510			0,510
	wg			0,382			0,339			0,375			0,369
	we			0,618			0,661			0,625			0,631
	β_{wg}			-0,058			0,592			0,934			1,924
	, w g			0,345			0,354			0,345			0,347
	β_{we}			0,006			0,271			0,513			1,149
				0,168			0,141			0,164			0,160
	lg			0,871			0,838			0,865			0,861
	le			0,971			0,977			0,972			0,973
	w_{go}			0,382			0,342			0,377			0,385
	w_{e0}			0,618			0,658			0,623			0,615
	d_g			0,000			0,000			0,002			0,017
0,7	$\sigma_y^2 \ \sigma_g^2 \ \sigma_e^2 \ h_g^2 \ h_e^2$	100,000	100,196	100,109	100,000	102,145	101,818	100,000	101,212	100,206	100,000	105,064	101,503
	σ_g^2	70,000	70,304	70,206	70,000	72,840	72,466	70,000	71,233	70,162	70,000	75,511	71,770
	$\frac{\sigma_e^2}{12}$	30,000	29,892	29,903	30,000	29,303	29,355	30,000	29,980	30,043	30,000	29,554	29,732
	n_g^-	0,700	0,701	0,700	0,700	0,712	0,712	0,700	0,702	0,698	0,700	0,718	0,705
		0,300	0,299	0,300 $120,013$	0,300	0,288	0,288 119,840	0,300	0,298	0,302 $119,649$	0,300	0,282	0,295 $119,912$
	$\frac{\mu}{pv}$			0,529			0,224			0,022			0,000
	β_2			-0,084			0,737			1,635			3,204
	P 2			0,501			0,507			0,501			0,505
	wg			0,637			0,644			0,634			0,643
	we			0,363			0,356			0,366			0,357
	β_{wg}			-0,031			0,264			0,579			1,108
	9			0,196			0,193			0,198			0,195
	β_{we}			-0,054			0,474			1,055			2,097
				0,305			0,313			0,304			0,308
	lg			0,968			0,969			0,967			0,972
	le			0,880			0,876			0,882			0,878
	w_{go}			0,637			0,646			0,637			0,652
	w_{e0}			0,363			0,354			0,363			0,348
	d_g			0,000			0,001			0,002			0,009

Tabela B.5 Resultados quando os Y_{kl} são gerados com herdabilidade de 8%.

						$\sigma_y^2 = 50$	$h_g^2 = 0,08$	$\sigma_g^2 = 4$	$\sigma_e^2 = 46$				
β_2	parâm	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	v1	v2	v3
),1							$\sigma_{y0}^2 = 49,560$	$h_{g0}^2 = 0,088$	$h_{e0}^2 = 0,912$				
	σ_{u}^{2}	49,545	49,528	49,540	49,559	49,559	49,432	49,501	49,410	49,430	49,561	49,559	49,558
	$egin{array}{c} \sigma_y^2 \ h_g^2 \ h_e^2 \end{array}$	0,087	0,091	0,087	0,088	0,088	0,082	0,089	0,081	0,089	0,088	0,089	0,089
	$h_e^{\frac{5}{2}}$	0,913	0,909	0,913	0,912	0,912	0,918	0,911	0,919	0,911	0,912	0,911	0,911
	pv	0,647	0,244	0,544	0,925	0,976	0,090	0,160	0,064	0,047	0,948	0,680	0,722
	β_2	0,136	0,557	-0,165	-0,029	0,010	0,459	0,478	0,578	0,612	-0,014	-0,098	-0,084
		0,296	0,478	0,272	0,305	0,320	0,271	0,340	0,312	0,308	0,219	0,237	0,237
	wg	0,159	0,151	0,159	0,148	0,150	0,148	0,152	0,155	0,149	0,070	0,076	0,075
	we	0,841	0,849	0,841	0,852	0,850	0,852	0,848	0,845	0,851	0,930	0,924	0,925
	β_{wg}	0,072	0,003	-0,057	-0,013	0,028	0,107	0,037	0,132	0,065	0,036	0,023	0,012
	β_{we}	0,063	0,555	-0,109	-0,016	-0,018	0,352	0,441	0,446	0,547	-0,051	-0,121	-0,097
	w_{g0}	0,161	0,148	0,161	0,149	0,150	0,156	0,150	0,165	0,149	0,070	0,076	0,075
	w_{e0}	0,839	0,852	0,839	0,851	0,850	0,844	0,850	0,835	0,851	0,930	0,924	0,925
	d_g	0,002	-0,003	0,002	0,000	0,000	0,008	-0,002	0,010	-0,001	0,000	-0,001	0,000
8,0							$\sigma_{y0}^2 = 51,153$	$h_{g0}^2 = 0,135$	$h_{e0}^2 = 0,865$				
	σ_{n}^{2}	51,067	51,074	51,133	51,146	51,092	51,120	50,962	51,158	51,157	51,152	51,002	51,120
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,134	0,136	0,134	0,135	0,132	0,144	0,135	0,136	0,136	0,135	0,134	0,133
	h^2	0,866	0,864	0,866	0,865	0,868	0,856	0,865	0,864	0,864	0,865	0,866	0,867
	pv	0,133	0,116	0,576	0,706	0,311	0,085	0,019	0,846	0,733	0,722	0,039	0,450
	β_2	0,461	0,774	0,158	0,119	-0,335	0,485	0,819	0,063	0,109	-0,079	-0,496	0,181
	P 2	0,306	0,492	0,282	0,315	0,330	0,282	0,350	0,325	0,318	0,221	0,240	0,240
	wg	0,215	0,200	0,216	0,203	0,200	0,221	0,204	0,221	0,205	0,102	0,108	0,106
	we	0,785	0,800	0,784	0,797	0,800	0,779	0,796	0,779	0,795	0,898	0,892	0,894
	β_{wg}	0,092	0,106	0,071	0,036	-0,132	-0,021	0,136	-0,090	-0,051	0,013	-0,060	0,059
	β_{we}	0,369	0,667	0,087	0,083	-0,203	0,507	0,684	0,153	0,160	-0,092	-0,435	0,122
	w_{g0}	0,216	0,200	0,218	0,204	0,204	0,212	0,205	0,220	0,203	0,101	0,108	0,107
	w_{e0}	0,784	0,800	0,782	0,796	0,796	0,788	0,795	0,780	0,797	0,899	0,892	0,893
	d_g	0,001	-0,001	0,001	0,000	0,004	-0,009	0,000	-0,001	-0,001	0,000	0,000	0,001
1,5							$\sigma_{y0}^2 = 48,911$	$h_{g0}^2 = 0, 0, 112$	$h_{e0}^2 = 0,888$				
	σ^2	48,911	48,908	48,895	48,891	48,913	48,695	47,981	48,677	48,841	48,876	48,853	48,895
	$egin{array}{c} \sigma_y^2 \ h_g^2 \ h_e^2 \end{array}$	0,112	0,112	0,112	0,115	0,113	0,110	0,091	0,105	0,114	0,111	0,111	0,113
	$h_g^{h_2}$	0,888	0,888	0,888	0,885	0,887	0,890	0,909	0,895	0,886	0,889	0,889	0,887
	$\frac{n_e}{pv}$	0,989	0,799	0,459	0,333	0,747	0,013	0,000	0,015	0,109	0,336	0,207	0,440
	β_2	-0,004	-0,122	-0,203	-0,337	-0,103	0,675	1,687	0,763	-0,495	-0,209	0,297	0,182
	$^{\rho}2$	0,297	0,478	0,274	0,306	0,321	0,272	0,335	0,313	0,309	0,217	0,235	0,132
							0,184				0,086		0,233
	wg	0,191	0,175 $0,825$	0,192 $0,808$	0,181 $0,819$	0,180 0,820	0,184	0,155 0,845	0,187 0,813	0,181 0,819	0,086	0,092 $0,908$	0,092
	we	0,809	-0,042	-0,020		0,820	0,816		0,813	-0,039	-0,029	0,908	0,908
	β_{wg}	-0,030 0,026	-0,042 -0,079	-0,020 -0,183	0,011	-0,149	0,111	0,286	0,163	-0,039 -0,456	-0,029 -0,180	0,039	0,005
	β_{we}		-0,079 0,176	-0,183 0,192	-0,348 0,178	0,149 0,179	0,564 0,186	1,401 0,179	0,600	-0,456 $0,178$	0,086	0,258	0,177
	w_{g0}	0,191 $0,809$	0,176	0,192	0,178	0,179 $0,821$	0,186	0,179	0,195	0,178 $0,822$	0,086 $0,914$	0,092	0,092
	w_{e0} d_g	0,000	0,000	0,000	-0,004	-0,001	0,003	0,821	0,803	-0,003	0,914	0,908	0,000
3							$\sigma_{y0}^2 = 51,558$	$h_{g0}^2 = 0,108$	$h_{e0}^2 = 0,892$				
	σ^2	51,559	51,543	51,557	51,553	51,541	50,665	48,409	51,162	51,554	51,508	51,531	51,481
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,108	0,109	0,108	0,108	0,107	0,089	0,052	0,100	0,107	0,107	0,109	0,107
	$^{\prime\prime}_{h^2}$	0,108	0,109	0,108	0,108	0,107	0,911	0,948	0,900	0,107	0,107	0,109	0,107
		0,892	0,891	0,892	0,582	0,893	0,000	0,948	0,900	0,857	0,893	0,891	0,893
	pv β_2	0,044	0,393	-0,067	-0,173	-0,209	1,325	3,149	1,022	0,057	0,262	0,255	0,131
	β_2	0,305	0,393	0,281		0,329	0,275	0,331		0,037	0,262		0,347
	****				0,314				0,321			0,241	
	wg	0,186	0,172	0,186	0,174	0,173	0,158	0,100	0,180	0,172	0,083	0,091	0,088
	we	0,814	0,828	0,814	0,826	0,827	0,842	0,900	0,820	0,828	0,917	0,909	0,912
	β_{wg}	-0,032	0,010	-0,003	0,006	-0,044	0,220	0,321	0,186	0,070	0,029	-0,003	0,041
	β_{we}	0,076	0,384	-0,063	-0,179	-0,165	1,105	2,828	0,837	-0,013	0,233	0,258	0,306
	w_{g0}	0,185	0,171	0,186	0,173	0,173	0,181	0,174	0,189	0,173	0,083	0,089	0,088
	w_{e0}	0,815	0,829	0,814	0,827	0,827	0,819	0,826	0,811	0,827	0,917	0,911	0,912
	d_g	0,000	-0,001	0,000	-0,001	0,001	0,023	0,074	0,009	0,001	0,000	-0,001	0,00

Tabela B.6 Resultados quando os Y_{kl} são gerados com herdabilidade de 32%.

						$\sigma_y^2 = 50$	$h_g^2 = 0,32$	$\sigma_g^2 = 16$	$\sigma_e^2 = 34$				
β_2	parâm	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	v1	v2	v3
),1							$\sigma_{y0}^2 = 52,020$	$h_{g0}^2 = 0,371$	$h_{e0}^2 = 0,629$				
	σ_y^2	52,005	51,929	51,896	52,013	51,969	52,008	51,964	52,012	51,892	51,939	52,025	51,962
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,370	0,368	0,367	0,370	0,368	0,370	0,368	0,371	0,367	0,371	0,372	0,369
	h_e^2	0,630	0,632	0,633	0,630	0,632	0,630	0,632	0,629	0,633	0,629	0,628	0,631
	pv	0,637	0,227	0,178	0,753	0,541	0,591	0,425	0,540	0,171	0,099	0,503	0,316
	β_2	0,152	0,619	0,402	0,104	-0,211	-0,159	0,294	-0,211	0,455	0,351	0,155	0,232
		0,323	0,512	0,298	0,330	0,345	0,296	0,367	0,344	0,333	0,212	0,232	0,231
	wg	0,404	0,383	0,412	0,400	0,394	0,409	0,397	0,410	0,398	0,246	0,255	0,255
	we	0,596	0,617	0,588	0,600	0,606	0,591	0,603	0,590	0,602	0,754	0,745	0,745
	β_{wg}	0,081	0,306	0,218	0,056	-0,194	-0,060	0,190	-0,058	0,244	0,085	0,002	0,093
	β_{we}	0,072	0,313	0,184	0,048	-0,017	-0,099	0,104	-0,153	0,212	0,266	0,153	0,139
	w_{g0}	0,404	0,385	0,415	0,400	0,396	0,409	0,399	0,410	0,401	0,245	0,254	0,256
	w_{e0}	0,596	0,615	0,585	0,600	0,604	0,591	0,601	0,590	0,599	0,755	0,746	0,744
	d_g	0,000	0,002	0,003	0,000	0,002	0,000	0,001	0,000	0,003	-0,001	-0,001	0,001
,8							$\sigma_{y0}^2 = 47,974$	$h_{g0}^2 = 0,270$	$h_{e0}^2 = 0,730$				
	$\sigma_y^2 \ h_g^2 \ h_e^2$	47,858	47,697	47,968	47,974	47,976	47,972	47,518	47,979	47,931	47,971	47,973	47,853
	$h_g^{\check{2}}$	0,267	0,267	0,270	0,270	0,270	0,277	0,272	0,270	0,272	0,270	0,270	0,269
	$h_e^{\overset{\sim}{2}}$	0,733	0,733	0,730	0,730	0,730	0,723	0,728	0,730	0,728	0,730	0,730	0,731
	pv	0,112	0,006	0,559	0,748	0,791	0,102	0,000	0,955	0,143	0,837	0,794	0,061
	β_2	-0,488	1,342	0,166	0,101	-0,087	0,460	1,347	0,019	0,463	-0,043	0,060	0,425
		0,306	0,488	0,283	0,314	0,329	0,281	0,348	0,326	0,316	0,209	0,227	0,227
	wg	0,331	0,311	0,340	0,325	0,322	0,339	0,326	0,338	0,327	0,184	0,192	0,192
	we	0,669	0,689	0,660	0,675	0,678	0,661	0,674	0,662	0,673	0,816	0,808	0,808
	β_{wg}	-0,187	0,403	0,026	0,004	0,016	0,014	0,353	-0,214	0,083	-0,019	0,005	0,096
	β_{we}	-0,301	0,939	0,140	0,097	-0,103	0,446	0,994	0,233	0,381	-0,024	0,055	0,329
	w_{g0}	0,333	0,313	0,340	0,325	0,322	0,334	0,324	0,338	0,325	0,184	0,192	0,193
	w_{e0}	0,667	0,687	0,660	0,675	0,678	0,666	0,676	0,662	0,675	0,816	0,808	0,807
	d_g	0,002	0,002	0,000	0,000	0,000	-0,005	-0,001	0,000	-0,002	0,000	0,000	0,000
,5							$\sigma_{y0}^2 = 51,119$	$h_{g0}^2 = 0,346$	$h_{e0}^2 = 0,654$				
	$\begin{matrix} \sigma_y^2 \\ h_g^2 \\ h_e^2 \end{matrix}$	50,891	51,085	51,124	51,110	51,040	50,528	49,698	50,997	51,068	51,098	51,073	50,954
	h_{q}^{2}	0,342	0,347	0,347	0,346	0,347	0,328	0,333	0,342	0,346	0,346	0,347	0,344
	$h_e^{\frac{5}{2}}$	0,658	0,653	0,653	0,654	0,653	0,672	0,667	0,658	0,654	0,654	0,653	0,656
	pv	0,030	0,262	0,713	0,657	0,118	0,002	0,000	0,225	0,218	0,575	0,190	0,045
	β_2	-0,691	0,570	-0,109	0,145	-0,535	0,889	2,165	0,412	-0,406	-0,119	0,303	0,462
		0,319	0,508	0,295	0,327	0,342	0,291	0,358	0,339	0,329	0,212	0,231	0,230
	wg	0,384	0,368	0,398	0,382	0,379	0,378	0,372	0,390	0,383	0,230	0,239	0,239
	we	0,616	0,632	0,602	0,618	0,621	0,622	0,628	0,610	0,617	0,770	0,761	0,761
	β_{wg}	-0,286	0,161	0,008	0,061	-0,169	0,443	0,766	0,256	-0,132	-0,053	0,065	0,138
	β_{we}	-0,405	0,409	-0,117	0,084	-0,366	0,445	1,399	0,156	-0,274	-0,066	0,237	0,324
	w_{g0}	0,387	0,368	0,398	0,383	0,378	0,391	0,381	0,393	0,383	0,230	0,239	0,240
	w_{e0}	0,613 $0,002$	0,632 $0,000$	0,602 $0,000$	0,617 $0,000$	0,622 $0,000$	0,609 0,014	0,619 0,010	0,607 0,003	0,617 $0,000$	0,770 $0,000$	0,761 -0,001	0,760 $0,001$
3	d_g	0,002	0,000	0,000	0,000	0,000	$\sigma_{y0}^2 = 55,279$	$h_{g0}^2 = 0,292$	$h_{e0}^2 = 0,708$	0,000	0,000	-0,001	0,001
	2	EE 000	EE 100	EE 000	EE 070	EE 070				EE 074	EE 074	EE 000	EE OFO
	$\sigma_y^2 \ h_g^2 \ h_e^2$	55,232	55,199	55,282	55,273	55,273	54,425	53,416	55,091	55,274	55,274	55,262	55,253
	n_{g}	0,290	0,289	0,293	0,292	0,293	0,274	0,278	0,294	0,293	0,293	0,292	0,291
		0,710	0,711	0,707	0,708	0,707	0,726	0,722	0,706	0,707	0,707	0,708	0,709
	pv	0,612	0,358	0,788	0,707	0,631	0,000	0,000	0,019	0,640	0,665	0,664	0,670
	eta_2	-0,167	-0,484	0,082	-0,127	0,170	1,213	2,513	0,826	-0,160	0,097	0,106	0,104
		0,330	0,526	0,305	0,338	0,354	0,300	0,369	0,351	0,341	0,223	0,243	0,243
	wg	0,348	0,327	0,358	0,343	0,339	0,337	0,331	0,356	0,343	0,198	0,206	0,206
	we	0,652	0,673	0,642	0,657	0,661	0,663	0,669	0,644	0,657	0,802	0,794	0,794
	β_{wg}	-0,196	-0,305	-0,020	-0,036	0,035	0,459	0,770	0,228	-0,028	0,013	0,055	0,089
	β_{we}	0,029	-0,179	0,102	-0,091	0,136	0,754	1,742	0,598	-0,131	0,084	0,051	0,015
	w_{g0}	0,349 $0,651$	0,330 $0,670$	0,358 0,642	0,343 0,657	0,339 0,661	0,351 0,649	0,342 $0,658$	0,355 0,645	0,343 $0,657$	0,198 $0,802$	0,206 0,794	0,207 0,793
	w_{e0}	0,001	0,670	0,042	0,007	0,001	0,049	0,008	0,045	0,007	0,802	0,794	0,793
		0,001	0,003	0,000	0,000	0,000	0,014	0,012	-0,001	0,000	0,000	0,000	-0,001

Tabela B.7 Resultados quando os Y_{kl} são gerados com herdabilidade de 70%.

						$\sigma_y^2 = 50$	$h_g^2 = 0,70$	$\sigma_g^2 = 35$	$\sigma_e^2 = 15$				
β_2	parâm	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	v1	v2	v3
,1							$\sigma_{y0}^2 = 52,122$	$h_{g0}^2 = 0,674$	$h_{e0}^2 = 0,326$				
	$\begin{matrix}\sigma_y^2\\h_g^2\\h_e^2\end{matrix}$	52,049	52,093	52,142	51,982	52,103	52,053	52,031	52,118	51,864	52,158	52,128	52,132
	h_q^{2}	0,673	0,674	0,675	0,672	0,674	0,673	0,673	0,674	0,670	0,679	0,676	0,675
	h_e^{2}	0,327	0,326	0,325	0,328	0,326	0,327	0,327	0,326	0,330	0,321	0,324	0,325
	pv	0,266	0,557	0,504	0,159	0,697	0,334	0,204	0,727	0,053	0,079	0,346	0,487
	β_2	-0,356	-0,296	0,200	0,461	0,133	0,285	0,463	0,119	0,640	-0,334	-0,197	-0,145
		0,320	0,503	0,299	0,327	0,341	0,295	0,364	0,342	0,330	0,190	0,209	0,209
	wg	0,614	0,601	0,628	0,615	0,610	0,624	0,615	0,618	0,619	0,465	0,470	0,474
	we	0,386	0,399	0,372	0,385	0,390	0,376	0,385	0,382	0,381	0,535	0,530	0,526
	β_{wg}	-0,242	-0,229	0,051	0,337	0,127	0,214	0,311	0,071	0,464	-0,105	-0,066	-0,040
	β_{we}	-0,114	-0,066	0,149	0,124	0,006	0,071	0,152	0,049	0,176	-0,229	-0,131	-0,105
	w_{g0}	0,615	0,601	0,627	0,617	0,610	0,625	0,616	0,618	0,622	0,461	0,469	0,473
	w_{e0}	0,385	0,399	0,373	0,383	0,390	0,375	0,384	0,382	0,378	0,539	0,531	0,527
	d_g	0,001	0,000	-0,001	0,002	0,000	0,001	0,001	0,000	0,003	-0,004	-0,001	-0,001
),8							$\sigma_{y0}^2=49,491$	$h_{g0}^2 = 0,695$	$h_{e0}^2 = 0,305$				
	σ^2	49,522	49,446	49,492	49,494	49,453	49,442	49,359	49,480	49,188	49,495	49,486	49,465
	$h^{\frac{y}{2}}$	0,697	0,696	0,695	0,696	0,696	0,695	0,693	0,695	0,691	0,696	0,695	0,695
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,303	0,304	0,305	0,304	0,304	0,305	0,307	0,305	0,309	0,304	0,305	0,305
	pv	0,568	0,128	0,972	0,839	0,242	0,400	0,171	0,732	0,024	0,768	0,758	0,487
	β_2	-0,178	-0,744	-0,010	0,065	0,388	0,241	0,484	0,114	-0,726	0,054	0,062	-0,140
	P 2	0,311	0,488	0,290	0,318	0,331	0,287	0,353	0,332	0,321	0,184	0,201	0,202
	wg	0,632	0,619	0,643	0,634	0,627	0,641	0,630	0,634	0,635	0,480	0,487	0,492
	we	0,368	0,381	0,357	0,366	0,373	0,359	0,370	0,366	0,365	0,520	0,513	0,508
		-0,002	-0,395	0,018	0,007	0,222	0,187	0,374	0,100	-0,524	0,010	0,040	-0,088
	β_{wg} β_{we}	-0,175	-0,349	-0,028	0,058	0,167	0,055	0,110	0,014	-0,202	0,045	0,022	-0,052
	w_{g0}	0,631	0,618	0,643	0,634	0,627	0,641	0,632	0,634	0,639	0,480	0,487	0,492
	_	0,369	0,382	0,357	0,366	0,373	0,359	0,368	0,366	0,361	0,520	0,513	0,508
	w_{e0} d_g	-0,001	-0,001	0,000	0,000	0,000	0,001	0,002	0,000	0,004	0,000	0,000	0,000
1,5	9						$\sigma_{y0}^2 = 53,157$	$h_{q0}^2 = 0,752$	$h_{e0}^2 = 0,248$				
	σ^2	53,148	53,170	52,931	53,156	53,222	52,877	52,374	52,910	53,157	53,162	53,214	53,157
	$egin{array}{c} \sigma_y^2 \ h_g^2 \ h_e^2 \end{array}$	0,752	0,753	0,750	0,752	0,756	0,750	0,748	0,750	0,752	0,752	0,755	0,754
	h_2^g	0,248	0,247	0,250	0,248	0,244	0,250	0,252	0,250	0,248	0,248	0,245	0,246
		0,795	0,347	0,032	0,802	0,244	0,021	0,000	0,029	0,899	0,926	0,350	0,240
	pv	0,793	0,347	0,638	-0,082	0,119	0,680	1,529	0,745	0,042	0,920	0,330	0,142
	β_2	0,083	0,500	0,038	0,326	0,331	0,294	0,361	0,340	0,330	0,017	0,189	0,298
	wg	0,678 $0,322$	0,668 $0,332$	0,687 $0,313$	0,681 $0,319$	0,678 0,322	0,686 $0,314$	0,676 $0,324$	0,678 0,322	0,687 $0,313$	0,534 $0,466$	0,544	0,548 0,452
	we		0,332 $0,221$	0,313 0,450			0,314	1,040		0,313	-0,027	0,456	
	β_{wg}	0,087		0,450 $0,188$	-0,050	0,230			0,529		0,044	0,043	0,138
	β_{we}	-0,004 0.678	0,250 $0,667$	0,188	-0.032 0.681	0,301 0,674	0,193 0,688	0,489 $0,680$	0,216 0,680	0,020 $0,687$	0,044	0,146 $0,541$	0,160 0,546
	w_{g0}	0,678 $0,322$	0,667	0,688 $0,312$	0,681	0,674 0.326	0,688	0,680	0,680	0,687	0,534 $0,466$	0,541 $0,459$	0,546 $0,454$
	w_{e0} d_g	0,000	-0,001	0,001	0,000	-0,004	0,002	0,004	0,320	0,000	0,466	-0,003	-0,002
3	<u>J</u>	-	*	*	*		$\sigma_{y0}^2 = 51,706$	$h_{g0}^2 = 0,728$	$h_{e0}^2 = 0,272$	*	•	*	*
	-2	51,687	51,679	51,369	51,696	51,534	51,178	48,332	51,371	51,672	51,696	51,657	51,703
	$\sigma_y^2 \ h_g^2 \ h_e^2$	0,728	0,728	0,722	0,728	0,728	0,723	0,697	0,724	0,729	0,729	0,727	0,728
	$\frac{n}{2}g$						0,723						
	n- 	0,272	0,272	0,278	0,272	0,272		0,303	0,276	0,271	0,271	0,273	0,272
	pv	0,671	0,331	0,04	0,679	0,02	0,001	0,000	0,013	0,192	0,381	0,459	0,821
	β_2	0,134	0,482	0,607	-0,134	-0,78	0,953	3,005	0,833	-0,426	0,162	0,15	-0,046
		0,316	0,496	0,295	0,323	0,336	0,291	0,35	0,337	0,327	0,185	0,202	0,203
	wg	0,658	0,646	0,664	0,66	0,654	0,664	0,633	0,657	0,667	0,51	0,516	0,522
	we	0,342	0,354	0,336	0,34	0,346	0,336	0,367	0,343	0,333	0,49	0,484	0,478
	β_{wg}	0,125	0,289	0,493	-0,099	-0,481	0,658	1,986	0,592	-0,246	0,076	0,117	-0,032
	β_{we}	0,01	0,193	0,114	-0,035	-0,299	0,295	1,019	0,242	-0,181	0,085	0,033	-0,014
	w_{g0}	0,658	0,646	0,669	0,66	0,653	0,668	0,659	0,66	0,666	0,51	0,517	0,522
	w_{e0}	0,342	0,354	0,331	0,34	0,347	0,332	0,341	0,34	0,334	0,49	0,483	0,478
		,	- ,										

B.3 Análises no nível de famílias

Para uma análise no nível de família, considera-se o cenário en que a variável \mathbf{Y} é gerada com herdabilidade de 32% e o efeito do SNP 7 é dado por $\beta_2=1,5$. As famílias caudais para esta variável são apresentadas na Figura B.1. Para o SNP 6 são influentes as famílias 5 e 6; para o SNP 7, as famílias 5, 14 e 47 e, para o SNP 8, as famílias 5 e 6. Quando os SNPs 6, 7 e 8 são analisados em conjunto (por meio da codificação haplotípica) tem-se várias famílias caudais que merecem uma análise mais detalhada, ver a Figura B.3.

Apresenta-se o heredograma da família 6 quando analisado o SNP 6 na Figura B.2. Note que os indivíduos apresentam alelos de risco (codificados como 2) e no entanto os valores da variável resposta na maioria deles está abaixo da média.

Quando analisados os SNPs 6, 7 e 8 em conjunto (Figura (B.3) encontram-se várias famílias caudais, entre elas a família 44 e cujo heredograma é dado na Figura B.4. Esta família é considerada caudal no nível haplotípico porque vários indivíduos carregam haplótipos de risco e têm valores da variável resposta acima da média geral.

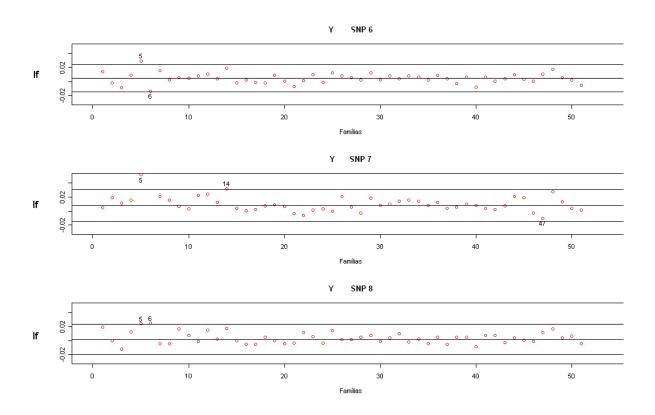


Figura B.1 Representação do índice I_f de cada família no componente poligênico.

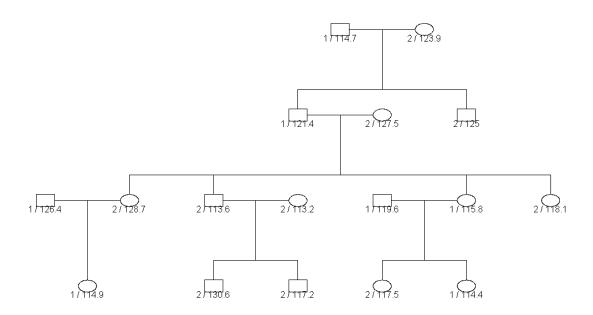


Figura B.2 Heredograma correspondente à família 6.

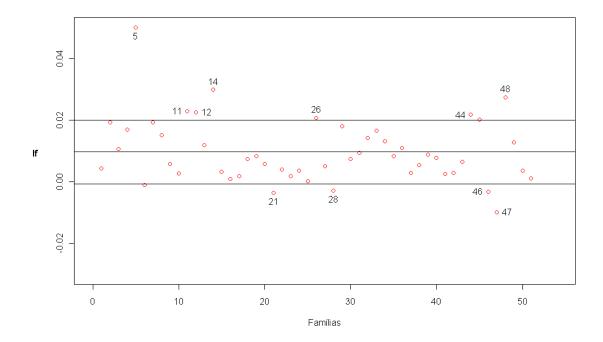


Figura B.3 Famílias caudais quando considerados os SNPs 6, 7 e 8 como um bloco.

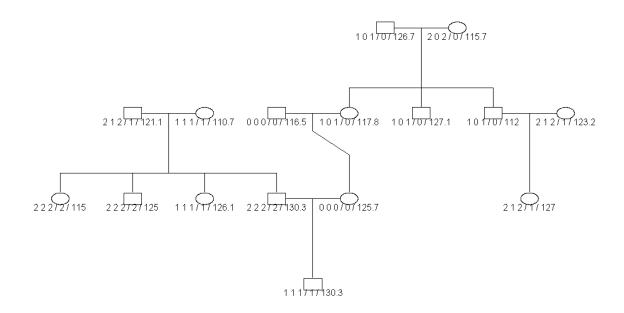


Figura B.4 Heredograma correspondente à família 44.

Apêndice C

Genética-Alguns conceitos fundamentais

Neste Apêndice são introduzidos alguns conceitos importantes de Genética, os quais ajudarão na compreensão e desenvolvimento da teoria abordada no trabalho.

O DNA ou ácido desoxirribonucleico, é a molécula que armazena a informação genética e está organizado em forma de Cromossomos no núcleo das células dos organismos. Gene é um termo geral que significa a entidade física transmitida de pai para filho, durante o processo de reprodução, que influencia características hereditárias (Falconer & Mackay, 1996). Formas alternativas de um gene que podem ocorrer em um determinado loco são chamados de alelos. O grupo de alelos de diferentes genes no mesmo cromossomo recebidos de um dos pais do indivíduo é chamado de haplótipo. O conjunto dos diferentes cromossomos de um indivíduo é conhecido como cariótipo e o conjunto formado por uma das cópias de todos os cromossomos é o Genoma. A constituição genética de um indivíduo definida pelo material recebido dos dois pais, é conhecida como o Genótipo. O genótipo não é observável, porém pode ser deduzido por meio de processos indiretos (de cruzamentos laboratoriais entre outros), é definido no momento da fecundação e não se altera, salvo alguma mutação, durante toda a vida do indivíduo. Já o conjunto de características observáveis nesse indivíduo no nível bioquímico, fisiológico ou morfológico é conhecido como o fenótipo ou variável resposta. Este é a "aparência" do indivíduo, ou seja, tudo aquilo que é visível ou detectável de alguma forma e que está associado a alguma característica do indivíduo, sendo que alguns fenótipos podem se alterar com o passar do tempo (Farah, 1997).

C.1 Marcadores Moleculares

Os marcadores moleculares são pontos de referência nos cromossomos do indivíduo

e correspondem a variações na sequência de DNA¹ cuja localização e genotipagem (determinar se um indivíduo é homozigoto ou heterozigoto) podem ser obtidos por meio de técnicas moleculares.

Na Figura 1 apresenta-se um mapa do genoma de uma espécie animal. Observa-se que cada cromossomo é representado por um arranjo linear de marcadores cuja distância entre eles é dada em unidades de distância citogenética (centimorgans, ilustrado na coluna da esquerda). Cada marcador ocupa uma posição fixa no cromossomo e tem um nome específico (ilustrado na coluna da direita).

A amostragem do genoma é diferenciada pela técnica molecular utilizada na obtenção dos marcadores moleculares. A seguir descrevem-se dois tipos de marcadores que têm auxiliado na procura de genes associados com doenças.

Marcadores microsatélites

Marcadores SNP (Single Nucleotide Polymorphism)

As áreas onde longas sequências de DNA diferem entre os indivíduos em apenas um nucleotídeo representam o tipo de marcador conhecido como SNP (do inglês, Single Nucleotide Polimorphism) ou polimorfismo de um único nucleotídeo (ver Figura C.2). Por exemplo, Maria tem em seu genoma a sequêcia AAATTTCGCCGGTA, essa mesma sequência em João pode ser AAATTTCGCCGTTA. Houve uma alteração em um único nucleotídeo o que pode ser a diferença entre Maria e João em ter um gene que produz

 $^{^{1}}$ As sequências são compostas pelas letras do conjunto $\{ACTG\}$ que correspondem aos quatro nucleotídeos e representam a estrutura de uma molécula ou cadeia de DNA com cerca de 3 bilhôes de nucleotídeos.

² Por exemplo se o marcador microsatélite tiver três alelos A_1 , A_2 , A_3 as possíveis respostas genotípicas são A_1A_1 , A_1A_2 , A_1A_3 , A_2A_2 , A_2A_3 , A_3A_3 . Em geral para n alelos tem-se $\frac{n(n+1)}{2}$ genótipos.

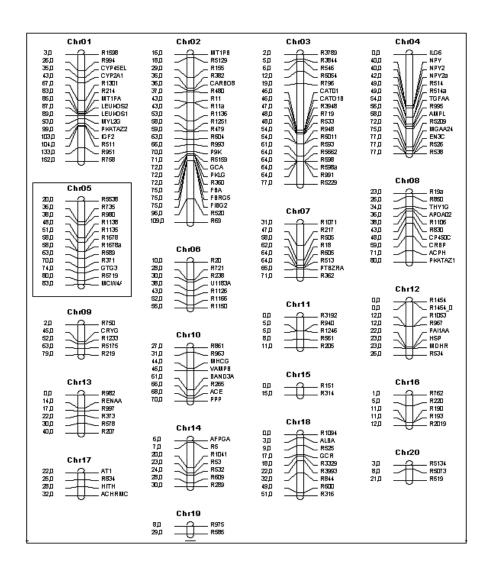


Figura C.1 Representação de um mapa de marcadores moleculares.

uma proteína defeituosa ou não. A Figura C.2 ilustra a estrutura de uma marcador SNP e está disponível no site:

httt://en.wikipedia.org/wiki/Single_nucleotide_polymorphism.

Em relação aos dois tipos de marcadores descritos ressaltam-se algumas caraterísticas que os diferenciam. Considerando marcadores microsatélites, com 300 ou 400 marcadores tem-se uma cobertura de aproximadamente 80% do genoma humano. Por exemplo, na Figura 1 o mapa apresentado contém 175 marcadores microsatélites. Sendo altamente polimórficos (possuem mais de 15 alelos) o tamanho físico permite que a amostragem seja realizada por grandes pedaços ou regiões do genoma o qual promove uma correlação entre

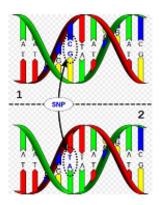


Figura C.2 Ilustração do marcador molecular conhecido como SNP.

indivíduos da mesma família pois a chance de dois indivíduos de famílias diferentes apresentarem o mesmo perfil de marcadores é extremamente pequena. No caso de marcadores do tipo SNPs, com aproximadamente um milhão de marcadores tem-se também uma cobertura de aproximadamente 80% do genoma humano (estima-se que existam aproximadamente 13 milhões de SNPs em todo o genoma humano). A amostragem é realizada por pequenas regiões (um único nucleotídeo) e dado que este tipo de marcador só tem dois alelos, é menos informativo do que outras alternativas multialélicas e não são úteis para identificar a estrutura familiar.

Finalmente, ressalta-se que a genotipagem dos microsatélites é realizada marcador por marcador o qual incrementa desmedidamente os custos financeiros. Enquanto que para marcadores SNPs têm sido desenvolvidas tecnologias de genotipagem em massa, reduzindo os custos de genoptipagem dos mesmos. Na Tabela C.1 apresenta-se um resumo ao dito anteriormente.

C.2 Desequilíbrio de ligação

A análise da distribuição dos alelos em um único loco pode ser estendida para mais de um loco. O desequilíbrio de ligação (D) é definido como qualquer desvio das probabilidades alélicas conjuntas em relação às probabilidades alélicas esperadas sob independência, indicando a existência de uma associação entre dois locos. Esses desvios podem ser repre-

	Microsatélites	SNPs
Cobertura do genoma	300 ou 400	aprox. um milhão
No de alelos	15 alelos ou mais	só tem dois alelos
Correlação	modela correlação familiar	não identificam a estrutura familiar
Quantidade		aprox. 13 milhões de SNPs
Genotipagem	marcador por marcador	em massa
Custo de genotipagem	muito alto	baixo

Tabela C.1 Marcadores Microsatélites versus SNPs.

sentados por:

$$D = P(A_i B_i) - P(A_i) P(B_i)$$

onde $P(A_iB_j)$ representa a probabilidade conjunta do haplótipo A_iB_j ; $P(A_i)$ e $P(B_j)$ representam as probabilidades marginais dos alelos A_i e B_j , respectivamente.

Se D = 0 então $P(A_iB_j) = P(A_i)P(B_j)$ e, nesse caso, a população é dita estar em equilíbrio de ligação. Este equilíbrio define independência probabilística na constituição dos alelos em locos do mesmo cromossomo chamado de independência vertical (entre locos) para diferenciar do EHW, uma independência horizontal (na constituição dos alelos dentro do loco).

Considerando o modelo de desequílíbrio baseado em dois locos, na Figura C.3 são representados os alelos A_1 e A_2 no loco L_1 , os alelos B_1 e B_2 no loco L_2 e os possíveis haplótipos A_1B_1 , A_1B_2 , A_2B_1 , A_2B_2 . Ao considerar o haplótipo A_1B_1 , a medida do desequilíbrio gamético é dada por:

$$D_{A_1B_1} = P(A_1B_1) - P(A_1)P(B_1). (C.1)$$

O parâmetro de desequilíbrio de ligação pode ser positivo ou negativo, sendo que sua amplitude depende das probabilidades alélicas nos dois locos:

$$-m\acute{a}x[P_{A_1}P_{B_1},(1-P_{A_1})(1-P_{B_1})] < D_{A_1B_1} < min[(1-P(B_1)P_{A_1},(1-P_{A_1})P_{B_1}].$$

Este parâmetro de desequilíbrio de ligação (D) é também uma medida de dependência alélica entre locos no mesmo cromossomo e tem sido explorado em estudos de ligação

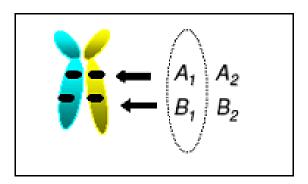


Figura C.3 Composição dos alelos em dois locos genéticos.

genéticos. Assim, quando o objetivo é encontrar genes que regulam um fenótipo (por exemplo, encontrar os genes que regulam doenças como a pressão arterial) é útil garantir que a população esteja em desequilíbrio de ligação pois, neste caso, a informação de um loco conhecido (marcador) ligado a um QTL poderá ser usada com maior precisão na predição do QTL. As populações derivadas de cruzamentos controlados são as mais adequadas para este tipo de estudo, pois nelas o desequilíbrio de ligação atinge o ponto máximo e, como conseqüência, a capacidade para detectar genes que estejam "muito próximos" também é máxima, permitindo o acompanhamento da segregação de um dado alelo de interesse (QTL) por meio da segregação de um gene (marcador) de manifestação conhecida (Tanksley, 1993 e Liu, 1998).

C.3 Equilíbrio de Hardy- Weinberg

Um conceito importante em Genética é o equilíbrio de Hardy-Weinberg, (EHW), que descreve as probabilidades genotípicas em termos das probabilidades alélicas, supondo independência na combinação dos alelos paternos que definem o genótipo (ver Figura C.4). Sob condições de cruzamentos aleatórios e ausência de processos como migração, mutação, seleção e oscilação, a população é esperada estar em equilíbrio de Hardy-Weinberg (Hartl & Clark, 1997).

Um sistema genético com r alelos em um loco ocorrendo com probabilidades p_i , i = 1, 2, ..., r, diz-se em equilíbrio de Hardy-Weinberg se as (r+1)r/2 probabilidades genotípicas p_{ij} , i < j são dadas por:

$$p_{ij} = p_i p_j [11(i=j) + 211(i < j)],$$
 (C.2)

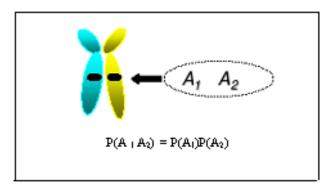


Figura C.4 Ilustração do equilíbrio de Hardy-Weinberg.

onde $\mathbb{1}C$ representa a função indicadora do conjunto C. Considerando locos dialélicos, segue na Tabela C.2 a relação entre probabilidades dos alelos e genótipos sob EHW (Falconer e Mackay, 1996). Por exemplo, a probabilidade que um indivíduo na população possua o genótipo A_1A_1 é p^2 . De maneira geral, pode-se dizer que o EHW é uma medida de associação entre alelos em um único loco. Existem possibilidades de desvios deste equilíbrio, uma delas é a estratificação da população, em que os cruzamentos são não aleatórios, ou seja, a união entre indivíduos de estratos diferentes são menos prováveis de ocorrer do que cruzamentos entre indivíduos de mesmo estrato, além disso a endogamía pode conduzir a desvios do equilíbrio (Sham, 1998).

Tabela C.2 Relação entre probabilidades de alelos e genótipos.

	Ale	elos	Génotipos				
	$\overline{A_1}$	A_2	A_1A_1	A_1A_2	A_2A_2		
Probabilidades	р	q	p^2	2pq	q^2		

Referências Bibliográficas

- [1] Akey, J., Jin, L. & Xiong, M. (2001). Haplotypes vs single marker linkage disequilibrium tests: What do we gain?. *European Journal of Human Genetics*. **9**, 291-300.
- [2] Almasy, L. & Blangero, J. (1998). Multipoint Quantitative-Trait Linkage Analysis in General Pedigrees. Am. J. Hum. Gent. 62, 1198-1211.
- [3] Almasy, L. & Blangero, J. (2000). Challenges for genetic analysis in the 21st century: localizing and characterizing genes for common complex diseases and their quantitative risk factors. *GeneScreen*. 1, 113-116.
- [4] Almasy, L., Dyer, T. D., & Blangero, J. (1997). Bivariate Quatitative Trait Linkage Analysis: Pleiotropy Versus Co-incident Linkages. *Genet. Epidem.* **14**, 953-958.
- [5] Almasy, L., Porjesz, B., Blangero, J. et al. (1999). Heritability of event-related brain potentials in families with a history of alcoholism. Am. J. Genet (Neuropsychiatry Genet). 88, 383-390.
- [6] Allen, A. S. & Satten, G. A. (2009). Genomewide association analysis of rheumatoid arthritis data via haplotype sharing. *BMC proceeding*. **3**, Suppl7: S30.
- [7] Altshuler, D., Daly, M. J. & Lander, E. S. (2008). Genetic mapping in human disease. Science. 322, 881-888.
- [8] Amin, N., Cornelia, S., Van Duijn, C. M., Aulchenko, Y. S. (2007). A Genomic Bacground Based Method for Association Analysis in Related Individuals. *PLoS One.* 2(12), e1 274.
- [9] Amos, C. I. (1994). Robust Variance-Components Approach for Assessing Genetic Linkage in Pedigrees. Am. J. Hum. Genet 54(3), 535-543.

- [10] Amos, C. I. & Elston R. C. (1989). Robust methods for the detection of genetic linkage for quantitative data from pedigrees. Genet. Epidemiol. 6, 349-360.
- [11] Amos, C. I., Dawson, D. V. & Elston, R. C. (1990). The probabilistic determination of identity-by-descent sharing for pairs of relatives from pedigrees. Am. J. Hum. Genet. 47(5), 842-853.
- [12] Amos, C.I., de Andrade M. & Zhu, D. (2001). Comparison of Multivariate Tests for Genetic Linkage. *Human Heredity*. **51**, 133-144.
- [13] Amos, C. I., Wu, X., Broderick, P., Gorlov, I. P., Gu, J., Eisen, T., Dong, Q., Zhang, Q., Gu, X., Vijayakrishnam, J., Sullivan, K., Matakidou, A., Wang, Y., Mills, S., Doheny, K., Tsai, Y., Chen, W. V., Shete, S., Spitz, M. R. & Houlston, R. S. (2008). Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15y25.1. Nat. Genet 40(5), 616-622.
- [14] Arya, R., Lehman, D., Hunt, K.J., Schneider, J., Almasy, L., Blangero, J., Stern, M.P. & Duggirala, R. (2003). Evidence for bivariate linkage of obesity and HDL-C levels in the Framingham Heart Study. BMC. Genet. [Suppl] 4, 52.
- [15] Atkinson, A. C. (1985). Plots, Transformations, and Regression. Clarendon Press, Oxford.
- [16] Aulchenko, Y. S., Koning, D. J. & Haley, C. (2007). Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigreebased quantitative trait loci association analysis. *Genetics.* 177, 577-585.
- [17] Balding, D. J., Bishop, M. & Cannings, C. (2003). Handbooks of statistical genetics. John Wiley & Sons Ltd, England.
- [18] Beckman, R. J., Nachtsheim, C. D. & Cook, R. D., (1987). Diagnostics for mixed-model analysis of variance. *Technometrics*. 29, 413-426.
- [19] Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). Regression Diagnostics: Identifying influential data and Sources of collinearity. John Wiley & Sons, New York.
- [20] Blangero, J. (1993). Statistical Genetic Approaches to Human Adaptability. Human Biology. 65, 941-966.

- [21] Blangero, J. & Almasy, L. (1997). Multipoint Oligogenic Linkage Analysis of Quantitative Traits. *Genet. Epidem. Biology.* **19**, (Suppl). S8-S14.
- [22] Blangero, J., Williams, J. & Almasy, L. (2000). Quantitative Trait Locus Mapping Using Human Pedigrees. *Human Biology.* **72(1)**, 35-62.
- [23] Bondell, H. D., Krishna, A. & Ghosh, S. K. (2010). Joint Variable Selection for Fixed and Random Effects in Linear Mixed-Effects Models. *Biometrics.* **66**, 1069-1077.
- [24] Breiman, L. (1996). Heuristics of instability and stabilization in model selection.

 Annals of Statistics. 24, 2350-2383.
- [25] Chatterjee, S. & Hadi, A. S. (1998). Sensitivity Analysis in Linear Regression. New York: Wiley.
- [26] Chen, C. H. & Wang, P. C. (1991). Diagnostic plots in Cox's regressão model. Biometrics. 47, 841-850.
- [27] Chien, L. C. (2011). A robust diagnostic plot for explanatory variables under model misspecification. *Journals of Applied Statistics*. **38(1)**, 113-126.
- [28] Cockerham, C. C. (1971). Higher order probability functions of identity of alleles by descent. *Genetics.* **69**, 235-246.
- [29] Comuzzie, A. G., Hixson, J. E., Mitchell, B. D., Mahaney, M. C., Dyer, T. D. & Stern, M. P. (1997). A Major Quantitative Trait Locus determining Serum Leptin Levels and Fat Mass is Located on Human Chromosome 2. Nat. Genet. 15, 273-275.
- [30] Conti, D. V. & Gauderman, J. (2004). SNPs, haplotypes and model selection in a candidate gene region: The SIMPLE analysis for multilocus data. *Genetic Epidemiology* **27**, 429-441.
- [31] Cook, R. D. (1986). Assessment of Local Influence (with discussion). Journal of the Statistical Society B 48, 133-169.
- [32] Cook, R. D. & Weisberg, S. (1982). Residuals and Influence in Regression. Chapman and Hall, New York.

- [33] Cordell, H. J. & Clayton, D. G. (2005). Genetic association studies. *Lancet* **366**, 1121-1131.
- [34] Cordell H. J. (2002). Epistasis: what it means, what it doesn't mean and methods to detected it in humans. *Human Molecular Genetics*. **11(20)**, 2463-2468.
- [35] Coulter K. L., Liu, D., Lin, X., Ghosh, D. & Michael, P. E. (2008). A Powerful and Flexible Multilocus Association Test for Quantitative Traits. The American Journal of Human Genetics. 82, 386-397.
- [36] de Andrade M., Thiel, T. J., Yu, L. & Amos, C. I. (1997). Assessing linkage in chromosome 5 using components of variance approach: univariate versus multivariate. *Genet Epidemiol.* 14, 773-778.
- [37] de Andrade, M., Amos, C. I. & Thiel, T. J. (1999). Methods to estimate genetic components of variance for quantitative traits in family studies. *Genetic. Epidemiol.* 17, 64-76.
- [38] Demidenko, E. (2004). Mixed Models: Theory and Applications. New York: Wiley.
- [39] Demidenko, E. & Stukel, T. A. (2005). Influence analysis for linear mixed-effects models. *Stat. Med.* **24**, 893-909.
- [40] Devlin, B. & Roeder, K. (1999). Genomic control for association studies. *Biometrics*. **55**, 997-1004.
- [41] Diggle, P. Heagerty, P., Liang, K. & Zeger, S. (2002). *Analysis of Longitudinal Data*. New York: Oxford University Press.
- [42] Draper, N. R. & Smith, H. (1998). Applied regression Analysis, 3rd Edition. John Wiley & Sons, New York.
- [43] Elston, R. C. & Stewart, J. (1971). A general Model for the Genetic Analysis of Pedigree Data. *Hum. Hered.* **21**, 523-542.
- [44] Falconer, D. S. & Mackay, T. F. C. (1996). Introduction to quantitative genetics. London: Prentice Hall.
- [45] Farah, S. B. (1997). DNA segredos e mistérios. São Paulo: Sarvier.

- [46] Fisher, R. T. (1918). The correlation between relatives on the supposition of Mendelian Inheritance. *Trans. R. Soc. Edinburg* **52**, 399-433.
- [47] Fisher, R. A. (1925). Statistical Methods for Research Workers. 1a ed. Oliver & Boyd, Edinburgh and London.
- [48] Giolo, S. R., Pereira, A. C., de Andrade, M., Oliveira, C. M., Krieger, J. E. & Soler, J. M. P. (2009). Genetic analysis of age-at-onset for cardiovascular risk factors in a Brazilian family study. *Human Heredity*. 68(2), 131-138.
- [49] Haseman, J. K. & Elston, R. C. (1972). The investigation of Linkage between a Quantitative Trait and a Marker Locus. *Behav. Genet.* **2**, 3-19.
- [50] Hartl, D. L. & Clark, A. G. (1997). Principles of Population Genetics. Sinauer.
- [51] Henderson, C. R. (1975). Best Linear Unbiased Estimation and Prediction under a Selection Model *Biometrics*. 31, 423-447.
- [52] Hilden-Minton, J. A. (1995). Multilevel Diagnostics for Mixed and Hierarchical Linear Models. PhD Thesis. University of California, Los Angeles. Ed Moderna.
- [53] Hodges, J. (1994). Some algebra and geometry for hierarchical models, applied to diagnostics. Research Report 94-0009, University of Minessota, Minneapolis, Minessota.
- [54] Hodges, J. (1998). Some algebra and geometry for hierarchical models, applied to diagnostics. J. R. Statis. Soc. B. **60**, 497-536.
- [55] Horimoto, A. R. V. R., Giolo, S. R., Oliveira, C. M., Alvim, R. O., Soler, J. M. P., de Andrade, M., Krieger, J. E. & Pereira, A. C. (2011a). Heritability of physical activity traits in Brazilian families: the Baependi Heart Study. BMC Medical Genetics (online). 2, 155.
- [56] Horimoto, A. R. V. R., Oliveira, C. M., Giolo, S. R., Soler, J. M. P., Krieger, J. E. & Pereira, A. C. (2011b). Genetic Analyses of Smoking Initiation, Persistence, Quantity and Age-At-Onset of Cigarette Use in Brazilian Families: Baependi Heart Study. Submitted BMC Medical Genetics.

- [57] Horvath, S. & Baur, M. P. (2000). Future directions of research in statistical genetics. Statis. Med 19, 3337-3343.
- [58] Ibrahim, J. G., Zhu, H., Garcia, R. I. & Guo, R. (2011). Fixed and Random Effects Selection in Mixed Effects Models. *Biometrics* 67, 495-503.
- [59] Jiang, C. & Zeng, Z. B. (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. *Genectics* **140**, 1111-1127.
- [60] Johnson, B. W. & McCulloch, R. E. (1987). Added-Variable Plots in Linear Regression. Technometrics 29, 427-433.
- [61] Kraft P. & de Andrade M. (2003). Group 6: Pleiotropy and multivariate analysis. Genet Epidemiol 25 (Suppl 1):S50-6.
- [62] Laird, N. M. & Ware, J. L. (1982). Random-effects models for longitudinal data. Biometrics 38, 963-974.
- [63] Lange, K.Westakle, J. & Spence, M. A. (1976). Extensions to pedigree analysis. III. Variance components by the Scoring Method. *Ann. Hum. Genet.* **39**, 485-491.
- [64] Leal, S. M., Yang, K. & Muller-Myhsok, B. (2005). SimPed: a simulation program to generate haplotype and genotype data for pedigree structures. *Hum Hered.* **60(2)**, 119-122.
- [65] Leutenegger, A. L., Prum, B., Genin, E., Verny C. & Lemainque, A. (2003). Estimação of the inbreeding coefficient through use of genomic data. Am J Hum Genet 73, 516-523.
- [66] Liu, B. H. (1998). Statistical Genomics: Linkage, Mapping, and QTL Analysis. Boca Raton FL: CRC Press.
- [67] Lynch, M. Walsh, B. (1998). Genetics and analysis of quantitative traits. Massachusetts: Sinauer Sunderland.
- [68] Meyer, K. (2007). WOMBAT a tool for mixed models analyses in quantitative genetics by REML. J. Zhejiang Uni. SCIENCE B 8, 815-821.

- [69] Meyer, K. & Tier, B. (2012). 'SNP Snappy": A Strategy for Fast Genome-Wide Association Studies Fitting a Full Mixed Model. Genetics 190, 275-277.
- [70] Moore, J. H. (2005). A global view of epistasis. S Nature Genetics. 37, 13-14.
- [71] Mosteller, F. & Tuckey, J. E. (1977). Data Analysis and Regression. Addison-Wesley, Reading MA.
- [72] Nobre, S. J. (2004). Métodos de Diagnóstico para Modelos Lineares Mistos. Dissertação para obtenção do grau de Mestre em Ciências. Área de concentração Estatística. Instituto de Matemática e Estatística da Universidade de São Paulo. São Paulo.
- [73] Nobre, J. S. & Singer, J. M. (2007). Residuals analysis for linear mixed models. Biom. J. 49, 863-875.
- [74] Nobre, J. S. & Singer, J. M. (2011). Leverage analysis for linear mixed models. Journal of Applied Statistics. 38(5), 1063-1072.
- [75] Oliveira, M. C., Pereira, A. C., de Andrade, M., Soler, J. M. P. & Krieger J. E. (2008). Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. BMC Medical Genetics. 9:32, 881-888.
- [76] Olswold, C. & de Andrade M. (2003). Localization of genes involved in the metabolic syndrome using multivariate linkage analysis. *BMC Genet.* **4** (Suppl) 1:S57.
- [77] Ott, J. (1991). Analysis of Human Genetic Linkage. London: Johns Hopkins.
- [78] Penrose, L. S. (1938). Genetic linkage in graded human characters. Ann. Eugenics 6, 133-138.
- [79] Pimenta, J. R., Zuccherato, L. W., Debes, A. A., Maselli, L., Soares, R. P., Moura-Neto, R. S., Rocha, J., Bydlowski, S. P. & Pena, S. D. J. (2006). Color and Genomic Ancestry in Brazilians: a study with forensic microsatelites. *Human Heredity* 62, 190-195.
- [80] Pinheiro, J. C. Bates, D. M. (2000). *Mixed-Effects in S and S-Plus*. New York: Springer.

- [81] Rahmatullah Imon, A. H. M. (2003). Simulation of errors in linear regression: an approach based on fixed percentage area. *Computational Statistics* 18.
- [82] Robinson, G. K. (1991). 'That blup is a good thing: The estimation of random effects'. Estatistical Science 6, 15-51.
- [83] Royall, R. M. & Sou, T. S. (2003). Interpreting statistical evidence by using imperfect models: Robust adjusted likelihood functions. J. R. Stat. Soc. Ser. B 65, 391-404.
- [84] Schork, N. J. (1993). Extended Multipoint Identity-by-Descent Analysis of Human Quantitative Traits: Efficiency, Power and Modeling Considerations. Am. J. Hum. Genet. 53(6), 1306-1319.
- [85] Searle, S. R., Cassela, G. McCullogh, C. E. (1992). Variance Components. New York: John Wiley & Sons.
- [86] Self, S. G. & Liang, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio test under nonstandard conditions. *Am. Stat. Assoc.* **82**, 605-610.
- [87] Sham, P. (1998). Statistics in human genetics. London: Arnold Applications of Statistics.
- [88] Skol A. D., et al. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209-213.
- [89] Skrondal, A. & Rabe-Hesketh, S. (2004). Generalized Latent Variable Modeling. Multilevel, Longitudinal, and Structural Equation Models Chapman & Hall.
- [90] Spielman, R. S., McGinnis, R. E. & Ewens, W. J. (1993). Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet 52, 506-516.
- [91] Spielman, R. S. & Ewens, W. J. (1998). A Sibship Test for linkage in the presence of Association: The Sib Transmission/Disequilibrium Test. Am. J. Hum. Genet 62, 450-458.
- [92] Tanksley, S. D. (1993). Mapping polygenes An. Rev. Genet 27, 205-233.

- [93] Verbeke, G. & Lessafre, E. (1996a). A linear mixed effects model with heterogeneity in the random - effects population. *Journal of the American Statistical Association*. 91, 217-221.
- [94] Verbeke, G. & Molenberghs, G. (1997). Linear Mixed Models In Practique: A SAS Oriented Approach. Lecture Notes in Statistics 126. New York: Springer-Verlag.
- [95] Verbeke, G. & Molenberghs, G. (2000). Linear Mixed Models for Longitudinal Data. New York: Springer-Verlag.
- [96] Walsh, B. (2009). Quantitative genetics, version 3.0: Where have we gone since 1987 and where are we headed?. *Genetica* **136**, 213-223.
- [97] Wang, P. C. (1985). Adding a Variable in generalized linear models. *Technometrics* **27**, 273-276.
- [98] Williams, J. T., Duggirala, R. & Blangero, J. (1997). Statistical Properties of Variance Components Method for Quantitative Trait Linkage Analysis in Nuclear Families and Extended Pedigrees. Genet. Epidem. 14, 1065-1070.
- [99] Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. M., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E. & Visscher, P. M. (2011). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*. 42, 565569.
- [100] Zeng, Z. B. (1994). Precision mapping quantitative trait loci. Genetics. 136, 1457-1468.
- [101] Ziegler, A., Konig, I. R. Thompson, J. R. (2008). Biostatistical aspects of Genomewide association studies. *Biometrical Journal.* **50**, 8-28.