

Aplicações em Meta-análise sob um enfoque Bayesiano usando dados médicos.

Carla Fernanda Pissini

Orientador: Prof. Dr. Jorge Alberto Achcar

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

São Carlos
março de 2006

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

P678am

Pissini, Carla Fernanda.

Aplicações em meta-análise sob um enfoque bayesiano usando dados médicos / Carla Fernanda Pissini. -- São Carlos : UFSCar, 2006.

107 p.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2006.

1. Teoria bayesiana de decisão estatística. 2. Análise bayesiana hierárquica. 3. MCMC. 4. Modelo bayesiano hierárquico. 5. Meta-análises. I. Título.

CDD: 519.542 (20^a)

Dedicatória

Dedico este trabalho aos meus pais José Carlos e Velda e ao meu irmão Fernando pelo amor, compreensão e incentivo ao longo de toda minha vida.

Agradecimentos

Agradeço,

à Deus, pela força durante esses anos de estudo,

aos meus pais, José Carlos e Velda, a meu irmão Fernando e a todos os meus familiares pelo amor, pela paciência, pela compreensão, pelo afeto, pelo incentivo, sempre presentes ao longo desta jornada,

ao professor Dr. Jorge Alberto Achcar pela orientação, pelos ensinamentos no desenvolvimento deste trabalho, pelas idéias e principalmente pelo exemplo de dedicação e disciplina ao trabalho,

à todos os professores do Departamento de Estatística da UFSCar, em especial aqueles aos quais tive o privilégio de ser aluna e aos funcionários, pelos ensinamentos e pelos momentos de convívio,

à professora Dra. Teresa Cristina Dias e ao professor Dr. Francisco Antonio Rojas Rojas pelas correções em meu exame de qualificação,

à todos os colegas de pós-graduação e aos amigos da minha turma do mestrado, pelos momentos de amizade, companheirismo, pela troca de experiências e, sobretudo, pela possibilidade de me proporcionar momentos de grande felicidade,

à todos os meus amigos e professores da UNESP de São José do Rio Preto, pela convivência durante os anos de graduação, pelo companheirismo, pelos momentos inesquecíveis que vivemos durante os quatro anos, pelos diversos ensinamentos e pela grande amizade que se formou,

à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pelo apoio, sob a forma de bolsa de estudo, durante o período de realização deste trabalho,

à todos os amigos que, de alguma forma, contribuíram para a realização deste trabalho.

Resumo

Neste trabalho, consideramos o uso de Meta-análise sob um enfoque Bayesiano. Meta-análise é uma técnica estatística que combina resultados de diversos estudos independentes, com o propósito de descrever conclusões gerais. Este termo foi introduzido por *Glass* (1976) usado quando o número de estudos sobre alguma pesquisa científica é pequeno. Os modelos propostos para Meta-análise usualmente assumem muitos parâmetros e o enfoque Bayesiano com MCMC (Monte Carlo em Cadeias de Markov) é uma alternativa apropriada para combinar informações de estudos independentes. O uso de modelos Bayesianos hierárquicos permite combinações de vários estudos independentes, para a obtenção de inferências precisas sobre um determinado tratamento. Como ilustração numérica consideramos conjuntos de dados médicos de diferentes estudos e, na análise, utilizamos modelos de efeitos fixos e aleatórios e mistura de distribuições normais para o erro do modelo de regressão. Em uma outra aplicação relacionamos Meta-análise e Educação, através do efeito da expectativa do professor associada ao QI dos estudantes.

Palavras-chave: Análise Bayesiana, modelos Bayesianos hierárquicos, Meta-análise, modelos de efeitos aleatórios, mistura de distriuições, MCMC.

Abstract

In this work, we consider the use of Meta-analysis with a Bayesian approach. Meta-analysis is a statistical technique that combines the results of different independent studies with purpose to find general conclusions. This term was introduced by Glass (1976) and it has been used when the number of studies about some research project is small. Usually, the models for Meta-analysis assume a large number of parameters and the Bayesian approach using MCMC (Markov Chain Monte Carlo) methods is a good alternative to combine information of independent studies, to obtain accurate inferences about a specified treatment. As illustration, we consider real medical data sets on different studies, in which, we consider fixed and random effects models. We also assume mixture of normal distributions for the error of the models. Another application is considered with educational data.

keywords: Bayesian analysis, Bayesian hierarchical models, Meta-analysis, effects random models, mixture of distributions, MCMC.

Sumário

1	Introdução	1
2	Uma Introdução à Inferência Bayesiana.	6
3	Meta-análise Bayesiana em um estudo médico.	10
3.1	Uma análise Bayesiana do efeito de uma nova droga antiepilética.	10
3.2	Análise Clássica	12
3.3	Análise Bayesiana	14
3.3.1	Análise gráfica.	25
4	Relação entre o uso de estrógeno e a incidência de câncer no endométrio.	30
4.1	O uso de estrógeno e a incidência de câncer no endométrio.	33
4.2	Análise Bayesiana	35
5	Modelo com mistura finita de distribuições normais para o erro.	43
5.1	Modelo com mistura de distribuições para o erro.	44
5.1.1	Análise gráfica.	58
5.2	Critério de seleção de Modelos.	60
6	Meta-Análise de efeitos de aspirina em ataques cardíacos.	62
6.1	Análise Clássica	64
6.2	Análise Bayesiana	67
6.2.1	Modelo com efeitos aleatórios	68
6.2.2	Análise gráfica.	72
6.3	Modelo clássico de efeitos aleatórios.	74

7	Meta-análise em Educação.	79
7.1	Análise Bayesiana	81
7.1.1	Análise gráfica.	88
8	Conclusão final.	93
	Apêndices	93
A	Algoritmo Gibbs Sampling	94
B	Algoritmo Metropolis-Hastings	95
C	Verificação de convergência	97
D	Alguns programas desenvolvidos no WinBugs	99
	Referências Bibliográficas	103

Capítulo 1

Introdução

É comum encontrar relatórios de estudos médicos (ou de outras áreas) sobre um assunto, mas realizados sob diferentes condições.

A Meta-análise pode ser definida como uma técnica estatística que compara e combina resultados de diversos estudos independentes, buscando uma estimativa mais precisa do efeito de tratamento, e conclusões gerais sobre o aspecto/assunto considerado.

O objetivo da técnica Meta-análise é combinar informações de dados coletados sob diferentes condições e com diferentes níveis de precisão para produzir conclusões mais amplas do que aquelas disponíveis de cada fonte de informação. Esta técnica é muito usada quando o número de estudos sobre algum assunto é pequeno (*Fagard et al., 1996*).

Avaliar separadamente os estudos similares, um por vez, pode deixar uma impressão confusa, o que significa que um efeito real do tratamento não esteja sendo detectado. Tal problema pode ser reduzido usando Meta-análise, pela combinação dos dados de todos os estudos relevantes.

As etapas envolvidas na Meta-análise são similares às de qualquer outro empreendimento em pesquisa: formulação dos problemas, coleção e análise dos dados, e relatório dos resultados. Previamente, é proposto um protocolo detalhado da pesquisa que indique claramente os objetivos, as hipóteses a serem testadas, os subgrupos de interesse, os métodos e os critérios propostos, buscando identificar e selecionar estudos relevantes.

Entre os anos de 1945 e 1970 existiam poucos artigos publicados, relacionados ao tema de combinar resultados a partir de estudos distintos. Técnicas para combinar dados de vários estudos sobre um mesmo tópico, especialmente na área de ciências sociais, surgiram

em meados de 1970, onde houve um aumento no número de artigos publicados sobre Meta-análise. A aplicação da Meta-análise originou-se na pesquisa educacional, teve aceitação imediata em medicina e o seu emprego tem se estendido em várias áreas.

O termo Meta-análise foi criado por *G. V. Glass*, em 1976, no seu artigo "Primary, se-condary and meta-analysis of research", publicado na revista *Educational Researcher*. *Glass*, pesquisador da área da ciências sociais, mais precisamente da pesquisa educacional, Considera a Meta-análise como uma técnica específica, estimulada pelo uso de métodos de síntese quantitativa por cientistas sociais, particularmente por psicólogos.

G. V. Glass define Meta-análise como "a análise de uma grande coleção de resultados de análises provenientes de estudos individuais, tendo como propósito completar o que foi encontrado". Ou seja, como uma análise de análises.

Segundo *Egger et al.* (1997), há algumas considerações a fazer:

► A Meta-análise exprime um plano cuidadoso, como qualquer projeto de pesquisa científica, como um protocolo escrito e detalhado que está sendo preparado com antecedência;

► A definição a *priori* de critérios qualificados para os estudos são incluídos em uma busca detalhada, de modo que estudos estão centrados para uma alta qualidade de Meta-análise;

► Diferentes métodos estatísticos existem para combinar os dados, mas não existe um único "método correto";

Embora os estudos a serem combinados envolvam o mesmo assunto, há, diferenças entre eles, o que é denominado heterogeneidade entre estudos. A heterogeneidade é um dos problemas usuais em Meta-análise, uma solução encontrada para resolve-la foi incorporar na análise a variabilidade existente entre os estudos (ver *Biggerstaff e Tweedie* (1997), mas devemos ter cuidado em combinar os resultados. Se a análise mostrar resultados homogêneos, então as diferenças entre os estudos são conseqüências da variação da amostragem e um modelo de efeitos fixos é apropriado. Se a análise mostrar resultados heterogêneos, isso não deve ser considerado como um problema para a Meta-análise, mas como indicação de que os efeitos dos tratamentos diferem em circunstâncias diferentes.

Em medicina, Meta-análise levará a resultados finais que podem ser estatisticamente significantes - a favor ou contra determinado tratamento - ou serem inconclusivos (não

significantes). Se os resultados forem inconclusivos, ou seja, não forem constatadas diferenças estatísticas entre os tratamentos, existem duas possibilidades:

(a) um tratamento realmente não é melhor do que o outro. Neste caso, o número total de pacientes estudados deve constituir uma amostragem suficiente, dando poder estatístico que permita a detecção de efeitos clinicamente relevantes;

(b) o número de casos ou de eventos nos grupos estudados é insuficiente para responder a questão, nesta circunstância, mais ensaios clínicos são necessários.

Os modelos de Meta-análise que incorporam o componente de variância entre os estudos, são amplamente discutidos tanto na teoria clássica quanto na Bayesiana. Autores como *Gelman* (1995) e *Larose e Dey* (1997), consideram que a formulação de um modelo Bayesiano hierárquico oferece um mecanismo natural para descrever e explicar a heterogeneidade existente entre os estudos individuais, enquanto que a teoria clássica para condução da Meta-análise envolve a combinação da magnitude de efeitos. Desta forma, vamos considerar o uso de modelos Bayesianos hierárquicos (ver *Sillman*, 1997) pois, oferecem um sistema muito flexível para combinar informações de vários estudos como fonte entre grupos de estudos.

Algumas das vantagens dos métodos Bayesianos sobre os clássicos, na aplicação da Meta-análise, são: que os primeiros oferecem uma estrutura conveniente para a incorporação de informações disponíveis *a priori*, e são aplicáveis a problemas para combinar estudos com tamanhos amostrais pequenos, quando os estimadores Bayesianos devem ser mais precisos do que os obtidos via máxima verossimilhança. Uma outra vantagem, é o grande número de parâmetros que geralmente estão presentes em uma pesquisa, fazendo com que o uso de métodos clássicos seja intratável. Assim, as inferências a serem conduzidas, observando-se os resultados da análise, serão, possivelmente, mais precisas do que aquelas obtidas pelo método clássico.

As aplicações de Meta-análise se estendem a vários campos do conhecimento científico. A primeira aplicação ocorreu na área das ciências sociais, como citado anteriormente, mais precisamente em educação. Somente mais tarde a Meta-análise foi descoberta pela Medicina, Saúde Pública e campos de pesquisa para serviços de saúde, como estudos epidemiológicos. Na área médica, a Meta-análise conseguiu aceitação imediata em virtude, principalmente, da possibilidade de obtenção de um tamanho amostral grande e pela

combinação de estudos menores, proporcionando uma resposta mais precisa (ver *Egger et al.*, 1997).

Existem inúmeros artigos de aplicação da Meta-análise em medicina. Um exemplo clássico é o trabalho de *Draper* (ver *Draper*, 2000), que compara seis estudos clínicos multicêntricos, nos Estados Unidos e Europa, sobre a utilização e efeito de aspirina e placebo no tratamento de pacientes após um ataque cardíaco. Esse trabalho foi abordado na avaliação de diferentes metodologias e enfoques em Meta-análise.

O uso comum de Meta-análise em medicina tem coincidido com o aumento de projetos de pesquisa médica em experimentos clínicos aleatorizados e tem, evidentemente, beneficiado o aumento no nível de interesse sobre a interpretação de pequenos experimentos clínicos e individualmente inconclusivos (não significantes). Inegavelmente a medicina é a área onde houve maiores avanços no uso da Meta-análise, até o presente momento. Talvez, os motivos disso sejam as dificuldades práticas, os riscos, os custos e ainda as implicações éticas que envolvem os experimentos com seres humanos, e mesmo a obtenção de dados de pessoas submetidas a tratamentos médicos. Daí que cada vez mais pesquisadores das áreas biomédicas buscam no uso da Meta-análise a solução mais rápida, segura e econômica de problemas, principalmente os que dizem respeito ao estabelecimento de critérios que subsidiem a tomada de decisão por parte dos responsáveis pelas políticas de saúde pública.

Além de medicina e ciências sociais, o emprego da Meta-análise se estende a outras áreas da pesquisa, com aplicações bastante diversas, como por exemplo em agronomia, na qual há estudos sobre a influência da aplicação de diferentes dosagens de vinhaça na produção de cana-de-açúcar, em solos arenosos e argilosos. Em ecologia há estudos sobre o aumento do número e do tamanho em diferentes espécies de microorganismos de ambientes terrestres, marinhos e lacustres e também há estudos sobre os efeitos da elevação do CO_2 em plantas lenhosas. Em genética e melhoramento foi conduzida uma Meta-análise Bayesiana sobre a herdabilidade do quociente de inteligência em humanos e também estudos sobre os componentes genéticos maternos para características de importância econômica em raças cruzadas de bovinos de corte; podemos ver em *Giannotti et al.* (2002) as estimativas de correlação genética entre pesos ao nascer e à desmama em bovinos de corte. Em zootecnia foi estudada a influência que diferentes composições nutricionais de

dietas para gado leiteiro promovem na redução da febre do leite nos animais.

A proposta do trabalho é analisar modelos Bayesianos para combinar informações de estudos independentes usando a técnica de Meta-análise. Para tanto, utilizamos os algoritmos de simulação Metropolis-Hastings e Gibbs Sampling (ver *Casella e George*, 1992).

No capítulo 2, apresentamos uma breve introdução à inferência Bayesiana. No capítulo 3 consideramos um modelo de efeitos aleatórios agrupados e Meta-análise Bayesiana, aplicada a um conjunto de dados introduzido por *Van Belle* (1992), que descreve a eficácia de uma nova droga antiepilética. No capítulo 4 apresentamos um conjunto de dados introduzido por *Larose e Dey* (1997) e o analisamos através da Meta-análise Bayesiana, utilizando um modelo de regressão. No capítulo 5 vamos utilizar o mesmo conjunto de dados do capítulo 4, considerando o modelo de regressão com uma mistura de duas distribuições normais para o erro.

No capítulo 6 apresentamos métodos clássicos e Bayesianos de Meta-análise, considerando um conjunto de dados introduzido por *Draper et al.* (1993), onde compara-se o uso de aspirina e placebo para pacientes, após um ataque cardíaco: no capítulo 7 apresentamos um conjunto de dados de Meta-análise em Educação, analisado sob o enfoque Bayesiano; e no capítulo 8 apresentamos as conclusões finais.

Capítulo 2

Uma Introdução à Inferência Bayesiana.

Os métodos Bayesianos são métodos alternativos aos métodos clássicos de inferência, que podem melhorar a precisão e a interpretação dos resultados. Entre as diferenças importantes que existem entre os métodos clássicos e os Bayesianos salientamos: os métodos Bayesianos permitem a incorporação de informações *a priori* sobre os parâmetros do modelo, e ao contrário dos métodos clássicos, os métodos Bayesianos consideram o parâmetro da distribuição como uma variável aleatória, associando à ele uma distribuição de probabilidade.

Axiomas básicos da probabilidade.

Sejam A e B dois eventos e $\{A_i, i = 1, 2, 3, \dots\}$ um conjunto de eventos. A probabilidade $P(A)$ é um número que satisfaz,

axioma 1: $0 \leq P(A) \leq 1$ e $P(A) = 1$ se A é o evento certo.

axioma 2: se os eventos A_i são mutuamente exclusivos, $P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$.

axioma 3: $P(A \cap B) = P(B/A)P(A)$.

Através da fórmula de Bayes, que combina a informação vinda dos dados através da verossimilhança com a incorporação de conhecimentos *a priori*, ou seja, com as informações obtidas antes dos dados, obtém-se uma distribuição denominada distribuição *a posteriori*.

Fórmula de Bayes.

A probabilidade condicional do evento A dado outro evento B é dado por,

$$P(A/B) = \frac{P(B/A)P(A)}{P(B)}$$

Se A_i é um conjunto de eventos mutuamente exclusivos e exaustivos (formam um partição do espaço amostral Ω , isto é, $\bigcup_i A_i = \Omega$ e $P\left(\bigcup_i A_i\right) = \sum_i P(A_i) = 1$), temos

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_i P(B/A_i)P(A_i)}, \quad (2.1)$$

para todo i .

Assim,

$$P(A_i/B) \propto P(B/A_i)P(A_i).$$

Seja $\mathbf{y} = (y_1, \dots, y_n)'$ um vetor de dados ou observações de uma variável aleatória y com distribuição dada pela função densidade de probabilidade $f(\mathbf{y}/\theta)$ na qual θ é um parâmetro. Assumir que θ seja uma variável aleatória com uma distribuição dada pela densidade $\pi(\theta)$, a distribuição a *priori* para θ . Na inferência Bayesiana, toda inferência sobre θ é determinada pela distribuição a *posteriori* para θ dada por $\pi(\theta/\mathbf{y})$ obtida a partir da fórmula de Bayes (ver *Box e Tião, 1992*).

Os métodos estatísticos usuais sugerem uma função de verossimilhança $f(\mathbf{y}/\theta)$ e em adição vamos supor uma distribuição a *priori* $\pi(\theta)$ que expressa nossa incerteza sobre θ antes de observarmos os dados e a distribuição a *posteriori* $\pi(\theta/\mathbf{y})$ expressa nossa incerteza sobre θ após observarmos os dados.

Pela fórmula de Bayes (2.1), temos,

$$\pi(\theta/\mathbf{y}) = \frac{\pi(\theta)f(\mathbf{y}/\theta)}{\int \pi(\theta)f(\mathbf{y}/\theta)d\theta} \propto \pi(\theta)f(\mathbf{y}/\theta).$$

Densidades a *priori* e a *posteriori*.

Uma das limitações dos métodos Bayesianos nas aplicações, em geral está relacionada à integração da densidade conjunta a *posteriori* para obtenção das distribuições marginais, que na maioria das vezes é muito complexa. Neste sentido, uma alternativa é a utilização de técnicas de simulação, como os métodos Gibbs Sampling e Metropolis-Hastings (ver *Gelfand e Smith (1990)*).

O algoritmo Gibbs Sampling é um método de simulação via cadeias de Markov que fornece uma forma de obter uma amostra da distribuição de interesse, baseada em sucessivas gerações das distribuições condicionais (ver apêndice A). Esse algoritmo é utilizado no caso em que as distribuições condicionais marginais são identificadas com uma distribuição conhecida. Em modelos mais complexos, geralmente não temos formas conhecidas nas distribuições.

Componentes da inferência Bayesiana

Distribuição a *priori*: usa a probabilidade como um meio de quantificar incerteza sobre quantidades desconhecidas (variáveis).

Função de verossimilhança: relaciona todas as variáveis num modelo de probabilidade completa.

Distribuição a *posteriori*: quando observamos algumas variáveis (os dados), usar a fórmula de Bayes para encontrar distribuições de probabilidades condicionais para quantidades não observadas de interesse.

Utilidades: se desejar quantificar o valor de consequências diferentes, é possível usar as probabilidades a *posteriori* como base para tomada de decisões.

Modelo de efeito fixo e aleatório

No modelo de efeito fixo, os valores (populacionais) dos efeitos dos tratamentos são constantes fixas, mas desconhecidas. Na modelagem para Meta-análise, supondo estudos homogêneos, medindo aproximadamente a mesma quantidade sem variação entre estudos, isto é, a resposta y_i é modelada como $y_i = \mu + \varepsilon_i$, $\varepsilon_i \sim N(0, S_i^2)$, onde S_i^2 é a variância da resposta, sendo que a variabilidade assumida é a variação dentro do estudo. No caso

do modelo de efeito aleatório, os efeitos dos tratamentos variam aleatoriamente de estudo para estudo, como se fossem amostras de um universo de possíveis efeitos de tratamentos, com média e variância próprias; assim, é possível avaliar a variabilidade entre estudos a partir de componentes de variância.

O modelo de efeito fixo pode ser usado quando não há heterogeneidade significativa entre os estudos, ou seja, há variabilidade apenas dentro dos estudos e não entre eles. Caso haja heterogeneidade significativa entre os estudos, deve-se optar pelo uso do modelo de efeito aleatório, no qual estão presentes as variâncias dentro e entre os estudos.

Este problema pode ser evitado assumindo um efeito aleatório δ_i , o qual explica a variabilidade entre estudos.

Assim, temos o modelo,

$$y_i = \mu + \delta_i + \varepsilon_i,$$

na qual μ é a média geral, $\delta_i \sim N(0, \tau^2)$ e $\varepsilon_i \sim N(0, S_i^2)$.

O modelo de efeito fixo é um caso especial do aleatório em que a variância entre os estudos é nula, ou seja, $\tau^2 = 0$.

Assumindo θ uma variável aleatória contínua com distribuição a *priori* $\pi(\theta)$, a distribuição a *posteriori* para θ

Capítulo 3

Meta-análise Bayesiana em um estudo médico.

3.1 Uma análise Bayesiana do efeito de uma nova droga antiepilética.

Muitas vezes, o objetivo principal do pesquisador é comparar uma nova droga com uma já existente, procurando estabelecer se essa nova droga traz sucesso maior, menor ou igual que a existente na cura de uma determinada doença. Outras vezes, o objetivo é ajudar na melhoria do paciente com a doença.

Como ilustração vamos considerar um conjunto de dados, introduzido por *Van Belle* (1992), de 15 relatos independentes de experiências clínicas de uma nova droga antiepilética. A variável resposta (y_i : número de pacientes melhores em cada estudo, considerando n_i pacientes) foi definida como a redução acima de 50% na frequência da doença do paciente (ver *Achcar et al.*, 2001).

Assim temos uma resposta dicotômica, ou seja, se o novo tratamento melhora a situação do paciente em mais de 50%, recebe o valor 1, caso contrário a resposta é 0. Desta forma, o número de pacientes com resposta positiva (y_i) para cada estudo tem distribuição binomial com parâmetros n_i e p_i , ou seja, $y_i \sim b(n_i, p_i)$, $0 < p_i < 1$, $i = 1, 2, \dots, N$.

Temos interesse em comparar um tratamento novo com um tratamento usual, levando em consideração N estudos independentes. Os dados da Tabela 3.1 são de dois grupos

distintos que estão sendo observados. Os seis primeiros estudos são abertos ou sobre o tratamento novo, que permitem investigar o efeito do tratamento sem outros efeitos psicológicos. Os nove últimos são de estudos fechados ou sobre o tratamento usual, que são usados para controlar experiências clínicas, onde o paciente sabe que está testando um novo medicamento. Estes estudos oferecem uma estrutura adequada para incorporar informação, com uma *priori* apropriada para melhorar a precisão de inferência. Na coluna 1 temos o número de cada estudo, na coluna 2 a ordem cronológica do estudo, na coluna 3 o número total de pacientes (n_i) e na coluna 4 o número de pacientes com resposta positiva (y_i).

Tabela 3.1: Dados de 15 estudos independentes sobre a eficácia de uma nova droga antiepilética.

	Estudo	Original	n_i	y_i
Estudos abertos	1	1	30	17
	2	7	16	8
	3	9	69	41
	4	10	23	13
	5	11	42	32
	6	15	151	90
Estudos fechados	7	2	20	5
	8	3	20	9
	9	4	17	3
	10	5	15	7
	11	6	18	8
	12	8	17	9
	13	12	19	1
	14	13	51	12
	15	14	59	17

Uma análise clássica dos dados é comparada com uma análise Bayesiana, considerando intervalos de confiança e intervalos de credibilidade, respectivamente.

3.2 Análise Clássica

Inicialmente, vamos definir algumas variáveis.

Seja,

n_i : número total de pacientes que completaram o estudo i ;

y_i : número de pacientes com resposta positiva em cada estudo i ; (3.0)

\hat{p}_i : proporção observada de pacientes melhores durante o estudo i ;

$\hat{\theta}_i = \log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)$ é o logaritmo da razão de risco (logito) de pacientes com resposta positiva, $i = 1, \dots, N$ ($N = 15$).

Considerando que as respostas são dicotômicas e os estudos são independentes, $y_i \sim b(n_i, p_i)$ e a função de verossimilhança para p_1, \dots, p_N é dada por,

$$L(p_1, \dots, p_N) = \prod_{i=1}^N \binom{n_i}{y_i} p_i^{y_i} (1-p_i)^{n_i-y_i} \propto \prod_{i=1}^N \left(\frac{p_i}{1-p_i}\right)^{y_i} (1-p_i)^{n_i}.$$

O logaritmo da função de verossimilhança para p_1, \dots, p_N é dado por,

$$\begin{aligned} l(p_1, \dots, p_N) &= \log L(p_1, \dots, p_N) \\ &\propto \sum_{i=1}^N \left[y_i \log\left(\frac{p_i}{1-p_i}\right) + n_i \log(1-p_i) \right] \\ &\propto \sum_{i=1}^N [y_i [\log(p_i) - \log(1-p_i)] + n_i \log(1-p_i)]. \end{aligned}$$

Da equação $\frac{\partial l}{\partial p_i} = 0$ tem-se,

$$\frac{y_i}{p_i} + \frac{y_i}{(1-p_i)} - \frac{n_i}{(1-p_i)} = 0,$$

Daí, temos que o estimador de máxima verossimilhança de p_i é dado por,

$$\hat{p}_i = \frac{y_i}{n_i}. \quad (3.1)$$

e, conseqüentemente, obtemos os estimadores $\hat{\theta}_i$, onde $\hat{\theta}_i = \log\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right)$.

Sabemos que $y_i \sim b(n_i, p_i)$ e $\hat{p}_i = \frac{y_i}{n_i}$; $i = 1, \dots, N$ e pelas propriedades da distribuição binomial de y_i , temos que o valor esperado e a variância de y_i são dados por,

$$\begin{aligned} E(y_i) &= n_i p_i \\ \text{Var}(y_i) &= n_i p_i (1 - p_i). \end{aligned}$$

Logo,

$$\begin{aligned} E(\hat{p}_i) &= p_i \\ \text{Var}(\hat{p}_i) &= \frac{p_i(1 - p_i)}{n_i}. \end{aligned}$$

Um intervalo com 95% de confiança para p_i é definido a partir do fato que

$$Z_i = \frac{y_i - n_i \hat{p}_i}{\sqrt{n_i \hat{p}_i (1 - \hat{p}_i)}} \stackrel{a}{\sim} N(0, 1), \quad (3.2)$$

ou,

$$Z_i = \frac{\hat{p}_i - p_i}{\sqrt{\frac{\hat{p}_i(1 - \hat{p}_i)}{n_i}}} \stackrel{a}{\sim} N(0, 1).$$

A probabilidade de Z_i estar contida no intervalo $(-1.96, 1.96)$, é:

$$P\{-1.96 \leq Z_i \leq 1.96\} = 0.95. \quad (3.3)$$

Assim, um intervalo com 95% de confiança para cada proporção p_i , $i = 1, \dots, N$, é dada por,

$$\left(\hat{p}_i - 1.96 \sqrt{\frac{\hat{p}_i(1 - \hat{p}_i)}{n_i}}; \hat{p}_i + 1.96 \sqrt{\frac{\hat{p}_i(1 - \hat{p}_i)}{n_i}} \right).$$

Usando os dados da Tabela 3.1 e as equações (3.0) e (3.1) obtemos os valores estimados para p_i e θ_i , $i = 1, \dots, N$. A Tabela 3.2 mostra os valores estimados de p_i e θ_i e os intervalos de confiança.

Tabela 3.2: Valores estimados para p_i e θ_i e intervalosde 95% de confiança para p_i ($IC(p_i)$).

	Estudo	\hat{p}_i	$\hat{\theta}_i$	$IC(p_i)$
Estudos abertos	1	0.5666	0.2680	(0.3893; 0.7439)
	2	0.5000	0.0000	(0.2550; 0.7450)
	3	0.5942	0.3814	(0.4783; 0.7100)
	4	0.5652	0.2623	(0.3626; 0.7678)
	5	0.7619	1.1631	(0.6330; 0.8907)
	6	0.5960	0.3888	(0.5177; 0.6742)
Estudos fechados	7	0.2500	-1.0986	(0.0602; 0.4397)
	8	0.4500	-0.2007	(0.2319; 0.6680)
	9	0.1765	-1.5402	(-0.0047; 0.3576)
	10	0.4667	-0.1334	(0.2141; 0.7191)
	11	0.4444	-0.2233	(0.2148; 0.6740)
	12	0.5294	0.1177	(0.2921; 0.7666)
	13	0.0526	-2.8910	(-0.0477; 0.1530)
	14	0.2353	-1.1786	(0.1188; 0.3517)
	15	0.2881	-0.9046	(0.1725; 0.4037)

$IC(p_i)$: intervalo de confiança para p_i , $i = 1, \dots, 15$.

Observamos na Tabela 3.2 que os valores de \hat{p}_i para os estudos abertos apresentam, em geral melhoria em mais de 50% dos pacientes, ao contrário dos estudos fechados. Lembre-se que a resposta é significativa se o novo medicamento proporciona uma redução superior a 50% na frequência da doença nos pacientes.

3.3 Análise Bayesiana

Na Meta-análise Bayesiana sugerida como uma possível solução para obtenção de estimativas combinadas mais confiáveis, tem como objetivo combinar estimativas através

da Meta-análise, utilizando um modelo Bayesiano hierárquico. O modelo Bayesiano hierárquico proposto por *Larose e Dey* (1997), será usado para descrever e analisar os dados da Tabela 3.1. Na análise Bayesiana hierárquica, os parâmetros das distribuições dos parâmetros, chamados hiperparâmetros, são desconhecidos, mas a distribuição é conhecida (ver *Gelman e Rubin*, 1992). Na maioria das vezes, o problema é estruturado de forma que existem relações probabilísticas a respeito dos parâmetros e hiperparâmetros, que podem ser incorporados ao modelo através de distribuições a *priori* hierárquicas.

Para os dados da Tabela 3.1, consideramos, o modelo Bayesiano hierárquico e as distribuições a *priori* dadas por,

$$\begin{aligned} y_i &\sim b(n_i, p_i), \quad 0 < p_i < 1, \quad i = 1, 2, \dots, N; \\ \theta_i &= \log\left(\frac{p_i}{1-p_i}\right), \quad i = 1, 2, \dots, N; \\ \theta_i &\sim N(\mu + \delta_k, \sigma_k^2), \quad -\infty < \theta_i < +\infty, \quad k = 1, 2; \\ \sigma_k^2 &\sim IG(a_k, b_k), \quad \sigma_k^2 > 0, \quad a_k \text{ e } b_k \text{ conhecidos}, \quad k = 1, 2; \\ \mu &\sim N(0, c), \quad -\infty < \mu < +\infty \text{ e } c \text{ conhecido}; \\ \delta_k &\sim N(0, d), \quad -\infty < \delta_k < +\infty; \quad k = 1, 2 \text{ e } d \text{ conhecido}. \end{aligned}$$

Observe que θ_i é o logito estimado de melhora no i -ésimo estudo, logo é conveniente assumir uma distribuição normal, já que os logits tem domínio definido em toda reta. y_i tem uma distribuição binomial, θ_i tem uma distribuição normal, σ_k^2 uma distribuição gama inversa, μ e δ_k têm distribuições normais.

Como citado anteriormente, temos dois grupos distintos no conjunto de dados e, desta forma, a estrutura da média de θ_i contém uma média geral μ mais um efeito aleatório específico para cada grupo δ_k , $k = 1, 2$, sendo que $k = 1$ refere-se aos estudos abertos e $k = 2$ aos estudos fechados.

De (3.0) podemos definir p_i em termos de θ_i , isto é,

$$\theta_i = \log\left(\frac{p_i}{1-p_i}\right), \text{ isto é, } e^{\theta_i} = \frac{p_i}{1-p_i}, \text{ então, } p_i(1+e^{\theta_i}) = e^{\theta_i}.$$

Assim,

$$p_i = \frac{e^{\theta_i}}{1+e^{\theta_i}}. \quad (3.4)$$

Substituindo o valor de $p_i = \frac{e^{\theta_i}}{1+e^{\theta_i}}$ em (3.1), temos que a função de verossimilhança para $\theta_1, \dots, \theta_N$ é dada por,

$$\begin{aligned}
L(\theta_1, \dots, \theta_N) &\propto \prod_{i=1}^N \left(\frac{e^{\theta_i}}{1 + e^{\theta_i}} \right)^{y_i} \left(1 - \frac{e^{\theta_i}}{1 + e^{\theta_i}} \right)^{n_i - y_i} \\
&\propto \prod_{i=1}^N \frac{e^{\theta_i y_i}}{(1 + e^{\theta_i})^{n_i}},
\end{aligned} \tag{3.5}$$

para $N = N_1 + N_2$, N_1 é o número de estudos abertos e N_2 é o número de estudos fechados.

Distribuição a priori conjunta

Assumindo independência a *priori*, a distribuição a *priori* conjunta para $\theta_1, \dots, \theta_N$, σ_1^2 , σ_2^2 , μ , δ_1 e δ_2 é dada por,

$$\begin{aligned}
\pi(\theta_1, \dots, \theta_N, \sigma_1^2, \sigma_2^2, \mu, \delta_1, \delta_2) &\propto (\sigma_1^2)^{-\left(\frac{N_1}{2}a_1+1\right)} \times (\sigma_2^2)^{-\left(\frac{N_2}{2}a_2+1\right)} \\
&\times \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{N_1} (\theta_i - \mu - \delta_1)^2 \right\} \\
&\times \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=N_1+1}^N (\theta_i - \mu - \delta_2)^2 \right\} \\
&\times \exp \left\{ -\frac{1}{b_1\sigma_1^2} - \frac{1}{b_2\sigma_2^2} - \frac{\mu^2}{2c} - \frac{\delta_1^2}{2d} - \frac{\delta_2^2}{2d} \right\}.
\end{aligned} \tag{3.6}$$

Distribuição a posteriori conjunta

Combinando a distribuição a *priori* conjunta (3.6) com a função de verossimilhança dada em (3.5), obtemos a distribuição a *posteriori* conjunta para $\theta_1, \dots, \theta_N$, σ_1^2 , σ_2^2 , μ , δ_1 e δ_2 , dada por,

$$\begin{aligned}
\pi(\theta_1, \dots, \theta_N, \sigma_1^2, \sigma_2^2, \mu, \delta_1, \delta_2 | \mathbf{D}) &\propto (\sigma_1^2)^{-\left(\frac{N_1}{2}+a_1+1\right)} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{N_1} (\theta_i - \mu - \delta_1)^2 \right\} \\
&\times (\sigma_2^2)^{-\left(\frac{N_2}{2}+a_2+1\right)} \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=N_1+1}^N (\theta_i - \mu - \delta_2)^2 \right\} \\
&\times \exp \left\{ -\frac{1}{b_1\sigma_1^2} - \frac{1}{b_2\sigma_2^2} - \frac{\mu^2}{2c} - \frac{\delta_1^2}{2d} - \frac{\delta_2^2}{2d} \right\} \\
&\times \prod_{i=1}^N \frac{e^{\theta_i y_i}}{(1 + e^{\theta_i})^{n_i}}.
\end{aligned}$$

Utilizando o método de simulação MCMC (Monte Carlo em Cadeias de Markov) e considerando os algoritmos Gibbs Sampling e Metropolis-Hastings encontramos as distribuições condicionais necessárias para gerar valores da distribuição a *posteriori* conjunta.

Distribuições condicionais

As distribuições condicionais a serem utilizadas com os algoritmos Gibbs Sampling e Metropolis-Hastings, são dadas a seguir.

A distribuição condicional para μ , dados os outros parâmetros é,

$$\begin{aligned} \pi(\mu|\boldsymbol{\delta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{N_1} (\theta_i - \mu - \delta_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=N_1+1}^N (\theta_i - \mu - \delta_2)^2 - \frac{\mu^2}{2c} \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{N_1} (\mu - \mu_i^{(1)})^2 - \frac{1}{2\sigma_2^2} \sum_{i=N_1+1}^N (\mu - \mu_i^{(2)})^2 - \frac{\mu^2}{2c} \right\}, \end{aligned}$$

na qual $\mu_i^{(1)} = \theta_i - \delta_1$ e $\mu_i^{(2)} = \theta_i - \delta_2$. Isto é,

$$\begin{aligned} \pi(\mu|\boldsymbol{\delta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} \left(N_1\mu^2 - 2\mu \sum_{i=1}^{N_1} \mu_i^{(1)} \right) - \frac{1}{2\sigma_2^2} \left(N_2\mu^2 - 2\mu \sum_{i=N_1+1}^N \mu_i^{(2)} \right) - \frac{\mu^2}{2c} \right\} \\ &\propto \exp \left\{ -\frac{1}{2 \left(\frac{N_1}{\sigma_1^2} + \frac{N_2}{\sigma_2^2} + \frac{1}{c} \right)^{(-1)}} \left[\mu - \frac{\left(\frac{N_1}{\sigma_1^2} (\bar{\theta}_1 - \delta_1) + \frac{N_2}{\sigma_2^2} (\bar{\theta}_2 - \delta_2) \right)}{\left(\frac{N_1}{\sigma_1^2} + \frac{N_2}{\sigma_2^2} + \frac{1}{c} \right)} \right]^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2 \left(\frac{\sigma_1^2 \sigma_2^2}{N_1 \sigma_2^2 + N_1 \sigma_1^2 + \frac{\sigma_1^2 \sigma_2^2}{c}} \right)} \left[\mu - \frac{N_1 \sigma_2^2 (\bar{\theta}_1 - \delta_1) + N_2 \sigma_1^2 (\bar{\theta}_2 - \delta_2)}{N_1 \sigma_2^2 + N_1 \sigma_1^2 + \frac{\sigma_1^2 \sigma_2^2}{c}} \right]^2 \right\}. \end{aligned}$$

Portanto,

$$\mu|\boldsymbol{\delta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D} \sim N \left\{ \frac{N_1 \sigma_2^2 (\bar{\theta}_1 - \delta_1) + N_2 \sigma_1^2 (\bar{\theta}_2 - \delta_2)}{N_1 \sigma_2^2 + N_2 \sigma_1^2 + \frac{\sigma_1^2 \sigma_2^2}{c}}, \frac{\sigma_1^2 \sigma_2^2}{N_1 \sigma_2^2 + N_2 \sigma_1^2 + \frac{\sigma_1^2 \sigma_2^2}{c}} \right\}, \quad (3.7)$$

na qual $\boldsymbol{\delta} = (\delta_1, \delta_2)$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$, $\mathbf{D} = (y_1, \dots, y_N)$, $\bar{\theta}_1 = \frac{\sum_{i=1}^{N_1} \theta_i}{N_1}$ e $\bar{\theta}_2 = \frac{\sum_{i=N_1+1}^N \theta_i}{N_2}$.

A distribuição condicional para δ_1 , dados os outros parâmetros é,

$$\begin{aligned}\pi(\delta_1|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) &\propto \exp\left\{-\frac{1}{2\sigma_1^2} \sum_{i=1}^{N_1} (\theta_i - \mu - \delta_1)^2 - \frac{\delta_1^2}{2d}\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma_1^2} \sum_{i=1}^{N_1} (\delta_1 - \mu_i^{(0)})^2 - \frac{\delta_1^2}{2d}\right\},\end{aligned}$$

na qual $\mu_i^{(0)} = \theta_i - \mu$. Isto é,

$$\begin{aligned}\pi(\delta_1|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) &\propto \exp\left\{-\frac{1}{2\sigma_1^2} \left(N_1\delta_1^2 - 2\delta_1 \sum_{i=1}^{N_1} \mu_i^{(0)}\right) - \frac{\delta_1^2}{2d}\right\} \\ &\propto \exp\left\{-\frac{1}{2\left(\frac{N_1}{\sigma_1^2} + \frac{1}{d}\right)^{(-1)}} \left[\delta_1^2 - \frac{\left(\frac{1}{\sigma_1^2} (N_1(\bar{\theta}_1 - \mu))\right)}{\frac{N_1}{\sigma_1^2} + \frac{1}{d}}\right]^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\left(\frac{\sigma_1^2}{N_1 + \frac{\sigma_1^2}{d}}\right)} \left[\delta_1^2 - \frac{N_1(\bar{\theta}_1 - \mu)}{N_1 + \frac{\sigma_1^2}{d}}\right]^2\right\}.\end{aligned}$$

Portanto,

$$\delta_1|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D} \sim N\left\{\frac{N_1(\bar{\theta}_1 - \mu)}{N_1 + \frac{\sigma_1^2}{d}}, \frac{\sigma_1^2}{N_1 + \frac{\sigma_1^2}{d}}\right\},$$

na qual $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$, $\mathbf{D} = (y_1, \dots, y_N)$ e $\bar{\theta}_1 = \frac{\sum_{i=1}^{N_1} \theta_i}{N_1}$.

A distribuição condicional para δ_2 , dados os outros parâmetros é,

$$\begin{aligned}\pi(\delta_2|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) &\propto \exp\left\{-\frac{1}{2\sigma_2^2} \sum_{i=N_1+1}^N (\theta_i - \mu - \delta_2)^2 - \frac{\delta_2^2}{2d}\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma_2^2} \sum_{i=N_1+1}^N (\delta_2 - \mu_i^{(0)})^2 - \frac{\delta_2^2}{2d}\right\},\end{aligned}$$

na qual $\mu_i^{(0)} = \theta_i - \mu$.

Isto é,

$$\begin{aligned}
\pi(\delta_2|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) &\propto \exp\left\{-\frac{1}{2\sigma_2^2}\left(N_2\delta_2^2 - 2\delta_2\sum_{i=N_1+1}^N\mu_i^{(0)}\right) - \frac{\delta_2^2}{2d}\right\} \\
&\propto \exp\left\{-\frac{1}{2\left(\frac{N_2}{\sigma_2^2} + \frac{1}{d}\right)^{(-1)}}\left[\delta_2^2 - \frac{\left(\frac{1}{\sigma_2^2}(N_2(\bar{\theta}_2 - \mu))\right)}{\frac{N_2}{\sigma_2^2} + \frac{1}{d}}\right]^2\right\} \\
&\propto \exp\left\{-\frac{1}{2\left(\frac{\sigma_2^2}{N_2 + \frac{\sigma_2^2}{d}}\right)}\left[\delta_2^2 - \frac{N_2(\bar{\theta}_2 - \mu)}{N_2 + \frac{\sigma_2^2}{d}}\right]^2\right\}.
\end{aligned}$$

Portanto,

$$\delta_2|\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D} \sim N\left\{\frac{N_2(\bar{\theta}_2 - \mu)}{N_2 + \frac{\sigma_2^2}{d}}, \frac{\sigma_2^2}{N_2 + \frac{\sigma_2^2}{d}}\right\},$$

na qual $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$, $\mathbf{D} = (y_1, \dots, y_N)$ e $\bar{\theta}_2 = \frac{\sum_{i=N_1+1}^N \theta_i}{N_2}$.

A distribuição condicional para σ_1^2 , dados os outros parâmetros é,

$$\begin{aligned}
\pi(\sigma_1^2|\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\theta}, \mathbf{D}) &\propto (\sigma_1^2)^{-\left(\frac{N_1}{2} + a_1 + 1\right)} \times \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{N_1}(\theta_i - \mu - \delta_1)^2\right\} \times \exp\left\{-\frac{1}{b_1\sigma_1^2}\right\} \\
&\propto (\sigma_1^2)^{-\left(\frac{N_1}{2} + a_1 + 1\right)} \times \exp\left\{-\frac{1}{\sigma_1^2}\left[\frac{1}{b_1} + \frac{1}{2}\sum_{i=1}^{N_1}(\theta_i - (\mu + \delta_1))^2\right]\right\}.
\end{aligned}$$

Portanto,

$$\sigma_1^2|\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\theta}, \mathbf{D} \sim IG\left[\frac{N_1}{2} + a_1, \left(\frac{1}{b_1} + \frac{1}{2}\sum_{i=1}^{N_1}(\theta_i - (\mu + \delta_1))^2\right)^{-1}\right],$$

na qual $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ e $\mathbf{D} = (y_1, \dots, y_N)$.

A distribuição condicional para σ_2^2 , dados os outros parâmetros é,

$$\begin{aligned}
\pi(\sigma_2^2|\boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\theta}, \mathbf{D}) &\propto (\sigma_2^2)^{-\left(\frac{N_2}{2} + a_2 + 1\right)} \times \exp\left\{-\frac{1}{2\sigma_2^2}\sum_{i=N_1+1}^N(\theta_i - \mu - \delta_2)^2\right\} \times \exp\left\{-\frac{1}{b_2\sigma_2^2}\right\} \\
&\propto (\sigma_2^2)^{-\left(\frac{N_2}{2} + a_2 + 1\right)} \times \exp\left\{-\frac{1}{\sigma_2^2}\left[\frac{1}{b_2} + \frac{1}{2}\sum_{i=N_1+1}^N(\theta_i - (\mu + \delta_2))^2\right]\right\}.
\end{aligned}$$

Portanto,

$$\sigma_2^2 | \boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\theta}, \mathbf{D} \sim IG \left[\frac{N_2}{2} + a_2, \left(\frac{1}{b_2} + \frac{1}{2} \sum_{i=N_1+1}^N (\theta_i - (\mu + \delta_2))^2 \right)^{-1} \right],$$

na qual $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$ e $\mathbf{D} = (y_1, \dots, y_N)$.

A distribuição condicional para θ_i , dados os outros parâmetros é,

$$\begin{aligned} \pi(\theta_i | \boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}_{(i)}, \mathbf{D}) &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{N_1} (\theta_i - (\mu + \delta_1))^2 \right\} \times \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=N_1+1}^N (\theta_i - (\mu + \delta_2))^2 \right\} \\ &\times \prod_{i=1}^N \frac{e^{\theta_i y_i}}{(1 + e^{\theta_i})^{n_i}} \\ &\propto e^{\theta_i y_i - \frac{(\theta_i - (\mu + \delta_k))^2}{2\sigma_k^2}} \times \frac{1}{(1 + e^{\theta_i})^{n_i}}. \end{aligned}$$

Assim,

$$\pi(\theta_i | \boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}_{(i)}, \mathbf{D}) \propto e^{\theta_i y_i - \frac{(\theta_i - (\mu + \delta_k))^2}{2\sigma_k^2}} \times \frac{1}{(1 + e^{\theta_i})^{n_i}}. \quad (3.8)$$

Notamos que não obtemos uma distribuição condicional conhecida para θ_i . Por essa razão, o algoritmo Metropolis-Hastings para gerar os θ_i 's será empregado. O núcleo considerado para θ_i segue uma distribuição normal, assim temos:

$$\pi(\theta_i | \boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}_{(i)}, \mathbf{D}) \propto N(\mu + \delta_k, \sigma_k^2) \times \Psi(\theta_i), \quad (3.9)$$

na qual $\Psi(\theta_i) = \exp \{ \theta_i y_i - n_i \ln(1 + e^{\theta_i}) \}$, $\boldsymbol{\delta} = (\delta_1, \delta_2)$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$, $\boldsymbol{\theta}_{(i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_N)$ e $\mathbf{D} = (y_1, \dots, y_N)$.

Com a parametrização original, temos dificuldades para conseguir a convergência para o algoritmo Gibbs Sampling, especialmente para os parâmetros μ e δ_k , $k = 1, 2$ que são parâmetros desconhecidos. Desta maneira, consideramos uma reparametrização para uma reanálise dos dados dispostos na Tabela 3.1 definida da seguinte maneira $\xi_k = \mu + \delta_k$, $k = 1, 2$, na qual $\mu + \delta_k$ é a média do parâmetro θ_i , assim ξ_k segue uma distribuição normal com média 0 e variância $(c + d)$. A distribuição a *posteriori* conjunta para $\theta_1, \dots, \theta_N$, σ_1^2 , σ_2^2 , ξ_1 e ξ_2 , é dada por,

$$\begin{aligned}
\pi(\theta_1, \dots, \theta_N, \sigma_1^2, \sigma_2^2, \xi_1, \xi_2 | \mathbf{D}) &\propto (\sigma_1^2)^{-\left(\frac{N_1}{2} + a_1 + 1\right)} \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{N_1} (\theta_i - \xi_1)^2 \right\} \times \\
&(\sigma_2^2)^{-\left(\frac{N_2}{2} + a_2 + 1\right)} \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=N_1+1}^N (\theta_i - \xi_2)^2 \right\} \times \\
&\exp \left\{ -\frac{1}{b_1\sigma_1^2} - \frac{1}{b_2\sigma_2^2} - \frac{\xi_1^2}{2c} - \frac{\xi_2^2}{2d} \right\} \times \\
&\prod_{i=1}^N \frac{e^{\theta_i y_i}}{(1 + e^{\theta_i})^{n_i}}.
\end{aligned}$$

Com essa reparametrização temos as distribuições condicionais para σ_k^2 , θ_i , ξ , na qual $\xi = (\xi_1, \xi_2)$, são determinadas.

A distribuição condicional para σ_1^2 e σ_2^2 , dados os outros parâmetros é,

$$\sigma_1^2 | \xi, \theta, \mathbf{D} \sim IG \left[\frac{N_1}{2} + a_1, \left(\frac{1}{b_1} + \frac{1}{2} \sum_{i=1}^{N_1} (\theta_i - \xi_1)^2 \right)^{-1} \right] \quad (3.10)$$

$$\sigma_2^2 | \xi, \theta, \mathbf{D} \sim IG \left[\frac{N_2}{2} + a_2, \left(\frac{1}{b_2} + \frac{1}{2} \sum_{i=N_1+1}^N (\theta_i - \xi_2)^2 \right)^{-1} \right] \quad (3.11)$$

Mesmo com a reparametrização, não conseguimos identificar a distribuição condicional de θ_i com uma distribuição conhecida. Por essa razão, o algoritmo Metropolis-Hastings para gerar os θ_i 's foi novamente empregado. Dessa maneira, a distribuição condicional para θ_i dados os outros parâmetros é dada por,

$$\pi(\theta_i | \xi, \sigma^2, \theta_{(i)}, \mathbf{D}) \propto e^{\theta_i y_i - \frac{(\theta_i - \xi_k)^2}{2\sigma_k^2}} \times \frac{1}{(1 + e^{\theta_i})^{n_i}}. \quad (3.12)$$

O núcleo considerado para θ_i continua sendo de uma distribuição normal, assim temos:

$$\pi(\theta_i | \xi, \sigma^2, \theta_{(i)}, \mathbf{D}) \propto \text{Normal}(\xi_k, \sigma_k^2) \times \Psi(\theta_i), \quad (3.13)$$

na qual $\Psi(\theta_i) = \exp \{ \theta_i y_i - n_i \ln(1 + e^{\theta_i}) \}$.

A distribuição condicional para ξ_1 , dados os outros parâmetros é,

$$\pi(\xi_1|\boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) \propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{N_1} (\theta_i - \xi_1)^2 \right\} \times \exp \left\{ -\frac{\xi_1^2}{2c} \right\}.$$

Isto é,

$$\begin{aligned} \pi(\xi_1|\boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} [N_1\xi_1^2 - 2\xi_1 N_1\bar{\theta}_1] - \frac{\xi_1^2}{2c} \right\} \\ &\propto \exp \left\{ -\frac{\xi_1^2}{2} \left[\frac{N_1}{\sigma_1^2} + \frac{1}{c} \right] + \frac{\xi_1 N_1\bar{\theta}_1}{\sigma_1^2} \right\}. \end{aligned}$$

Portanto,

$$\pi(\xi_1|\boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) \propto \exp \left\{ -\frac{1}{2 \left(\frac{\sigma_1^2}{c} + N_1 \right)} \left[\xi_1 - \frac{N_1\bar{\theta}_1}{\frac{\sigma_1^2}{c} + N_1} \right]^2 \right\}.$$

Assim,

$$\xi_1|\boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D} \sim N \left\{ \frac{N_1\bar{\theta}_1}{\frac{\sigma_1^2}{c} + N_1}; \frac{\sigma_1^2}{\frac{\sigma_1^2}{c} + N_1} \right\}, \quad (3.14)$$

na qual $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$, $\mathbf{D} = (y_1, \dots, y_N)$ e $\bar{\theta}_1 = \frac{\sum_{i=1}^{N_1} \theta_i}{N_1}$.

A distribuição condicional para ξ_2 dados os outros parâmetros é dada por,

$$\pi(\xi_2|\boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) \propto \exp \left\{ -\frac{1}{2\sigma_2^2} \sum_{i=N_1+1}^N (\theta_i - \xi_2)^2 \right\} \times \exp \left\{ -\frac{\xi_2^2}{2d} \right\}.$$

Isto é,

$$\begin{aligned} \pi(\xi_2|\boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) &\propto \exp \left\{ -\frac{1}{2\sigma_2^2} [N_2\xi_2^2 - 2\xi_2 N_2\bar{\theta}_2] - \frac{\xi_2^2}{2d} \right\} \\ &\propto \exp \left\{ -\frac{\xi_2^2}{2} \left[\frac{N_2}{\sigma_2^2} + \frac{1}{d} \right] + \frac{\xi_2 N_2\bar{\theta}_2}{\sigma_2^2} \right\}. \end{aligned}$$

Portanto,

$$\pi(\xi_2|\boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D}) \propto \exp \left\{ -\frac{1}{2 \left(\frac{\sigma_2^2}{d} + N_2 \right)} \left[\xi_2 - \frac{N_2\bar{\theta}_2}{\frac{\sigma_2^2}{d} + N_2} \right]^2 \right\}.$$

Assim,

$$\xi_2 | \boldsymbol{\sigma}^2, \boldsymbol{\theta}, \mathbf{D} \sim N \left\{ \frac{N_2 \bar{\theta}_2}{\frac{\sigma_2^2}{d} + N_2}; \frac{\sigma_2^2}{\frac{\sigma_2^2}{d} + N_2} \right\}, \quad (3.15)$$

na qual $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$, $\mathbf{D} = (y_1, \dots, y_N)$ e $\bar{\theta}_2 = \frac{\sum_{i=N_1+1}^N \theta_i}{N_2}$.

Considerando os dados da Tabela 3.1 e as distribuições *a priori* para os parâmetros do modelo, com $a_1 = a_2 = 1$, $b_1 = b_2 = 0.1$ e $(c + d) = 0.001$ assumindo independência e *prioris* não informativas para os parâmetros, geramos duas cadeias de Gibbs Sampling cada uma com 110000 iterações. Para cada série consideramos saltos de tamanho dez, e finalmente obtemos uma amostra de tamanho 16000. As 30000 primeiras observações foram eliminadas ("burn in"). A convergência das amostras de Gibbs Sampling foi monitorada usando o Método de *Gelman e Rubin* (1992) ($\sqrt{\hat{R}} < 1.1$ para todos os parâmetros). A Tabela 3.3 mostra os sumários *a posteriori* e os valores $\sqrt{\hat{R}}$ (critério de *Gelman e Rubin*, 1992).

Tabela 3.3: Sumários a *posteriori*

Parâmetros	Média a posteriori	95% de intervalos de credibilidade	$\sqrt{\widehat{R}}$
ξ_1	0.441	(0.062, 0.814)	1.00
ξ_2	-0.832	(-1.299, -0.384)	0.99
τ_1	15.11	(1.016, 31.05)	1.00
τ_2	8.412	(1.033, 31.05)	0.99
p_1	0.591	(0.462, 0.708)	1.00
p_2	0.578	(0.419, 0.71)	0.99
p_3	0.600	(0.504, 0.692)	0.99
p_4	0.593	(0.456, 0.716)	1.00
p_5	0.677	(0.567, 0.793)	0.99
p_6	0.599	(0.528, 0.669)	0.99
p_7	0.2836	(0.154, 0.424)	1.00
p_8	0.3686	(0.228, 0.546)	1.00
p_9	0.2586	(0.120, 0.404)	1.00
p_{10}	0.3696	(0.221, 0.567)	1.00
p_{11}	0.3659	(0.228, 0.549)	0.99
p_{12}	0.3983	(0.247, 0.602)	1.00
p_{13}	0.2082	(0.0712, 0.346)	1.00
p_{14}	0.2627	(0.166, 0.366)	1.00
p_{15}	0.2944	(0.201, 0.396)	0.99

Da Tabela 3.3 note que $\tau_1 = \frac{1}{\sigma_1^2}$ e $\tau_2 = \frac{1}{\sigma_2^2}$, também os índices de *Gelman e Rubin* ($\sqrt{\widehat{R}}$), dos quais assumimos a convergência do algoritmo Gibbs Sampling.

Na Tabela 3.3 notamos que $\sigma_1^2 < \sigma_2^2$, indicando que estudos abertos formam um grupo mais homogêneo do que os estudos fechados. Observamos também que para estudos abertos, a média a *posteriori* para p_1, \dots, p_6 está no intervalo (0.57; 0.67), e para os estudos fechados a média a *posteriori* para p_7, \dots, p_{15} está no intervalo (0.20; 0.39), que representa

as médias dos pacientes com resposta positiva. Como em estudos abertos a média dos pacientes com resposta positiva é maior que 0.5 e em estudos fechados é menor que 0.5, concluímos que os estudos abertos têm uma maior eficácia do que aos estudos fechados, e estes resultados estão de acordo com os resultados obtidos por *Larose e Dey* (1997). Concluímos também que os estudos abertos formam um grupo mais homogêneo, do que os estudos fechados, pois apresentam um intervalo menor para a média a *posteriori*.

A diferença entre estudos abertos e fechados também pode ser verificada através de uma estimativa da média para θ_i , dada por $\xi_k = \mu + \delta_k$, $k = 1, 2$. Para estudos abertos, p_i se reduz a $\frac{e^{0.441}}{1+e^{0.441}} = 0.6085$ e para os estudos fechados, p_i se reduz a $\frac{e^{-0.832}}{1+e^{-0.832}} = 0.3032$, onde 0.441 é a média a *posteriori* de $\mu + \delta_1$ e (-0.832) é a média a *posteriori* de $\mu + \delta_2$. As médias a *posteriori* ponderadas para estudos abertos e para estudos fechados são dadas respectivamente por (ver *Fortulan*, 1999),

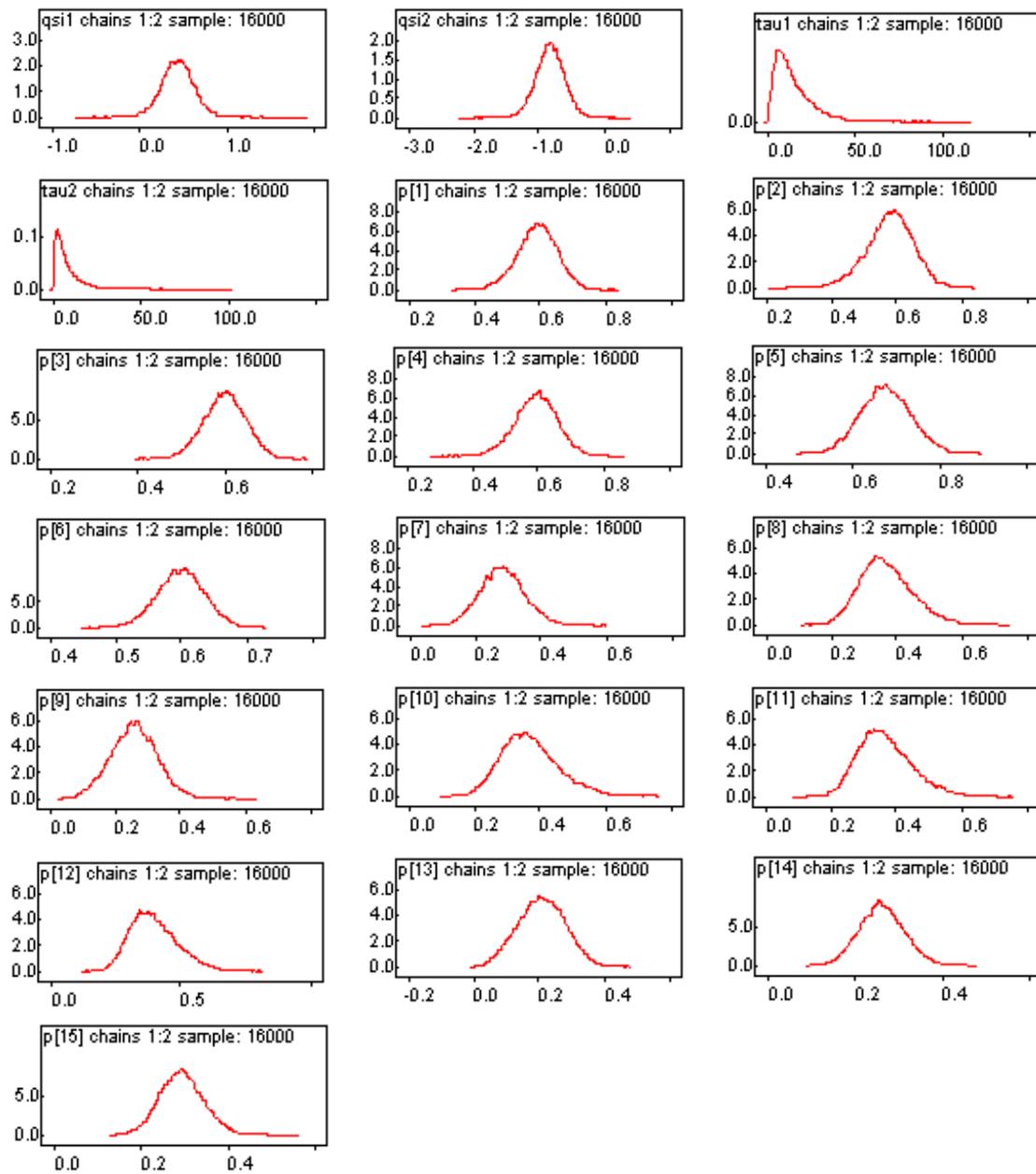
$$MP_{\text{abertos}} = \frac{\sum_{i=1}^6 n_i p_i}{\sum_{i=1}^6 n_i} = 0.6069$$

$$MP_{\text{fechados}} = \frac{\sum_{i=7}^{15} n_i p_i}{\sum_{i=7}^{15} n_i} = 0.3011.$$

Com esses resultados notamos que as médias a *posteriori* $\xi_k = \mu + \delta_k$, $k = 1, 2$, em relação a média ponderada, estão muito próximas. Portanto, concluímos, tanto pela análise clássica quanto pela Bayesiana, que os resultados para os estudos abertos sustentam a eficácia do medicamento.

3.3.1 Análise gráfica.

A Figura 3.1 mostra as formas das distribuições a *posteriori*, que podem ser interpretadas considerando que estas representam uma aproximação com relação à verdadeira distribuição dos parâmetros. Da Figura 3.2 observamos que os parâmetros apresentam-se não correlacionados.

Figura 3.1: Densidades *a posteriori* marginais.

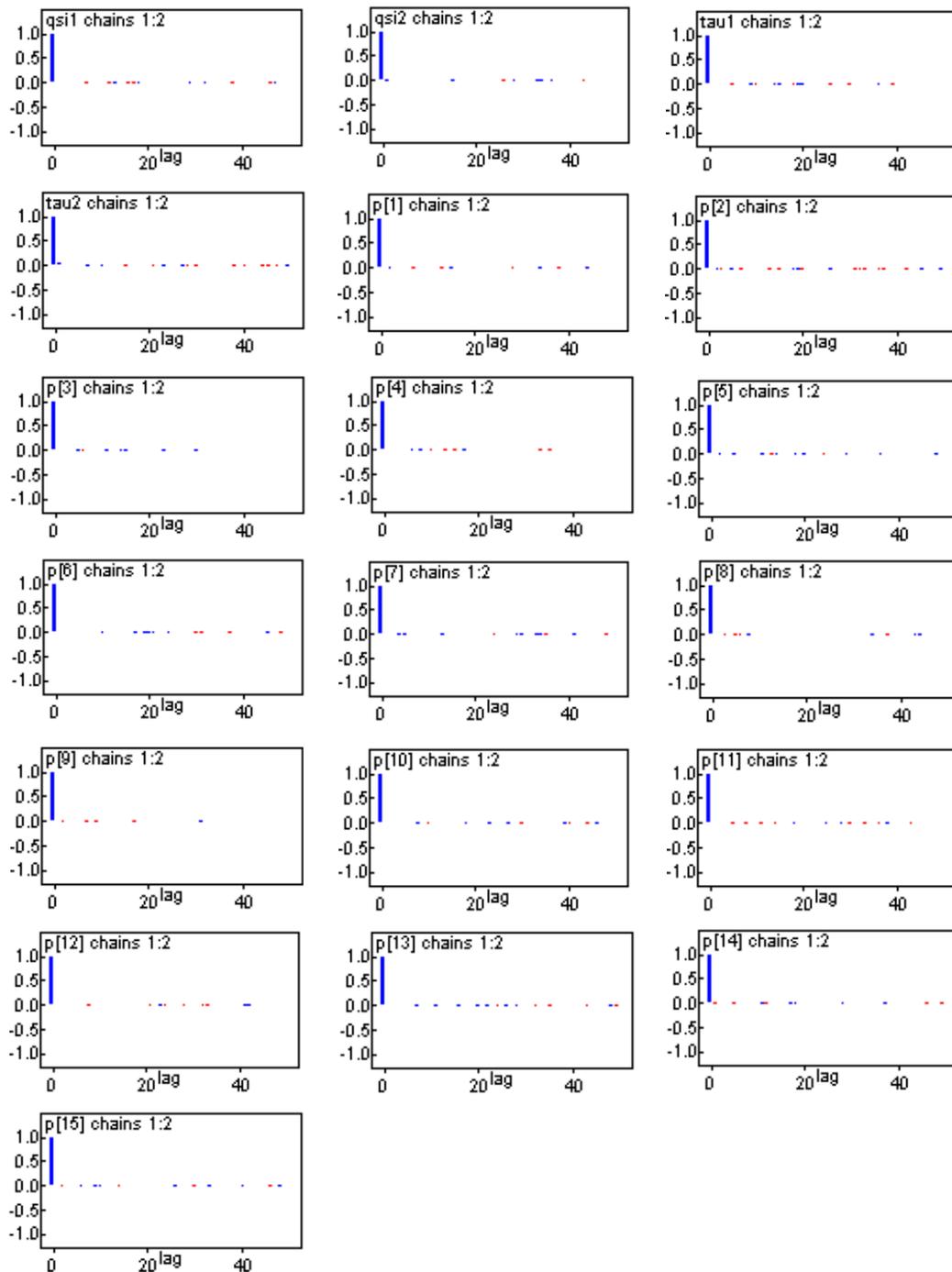
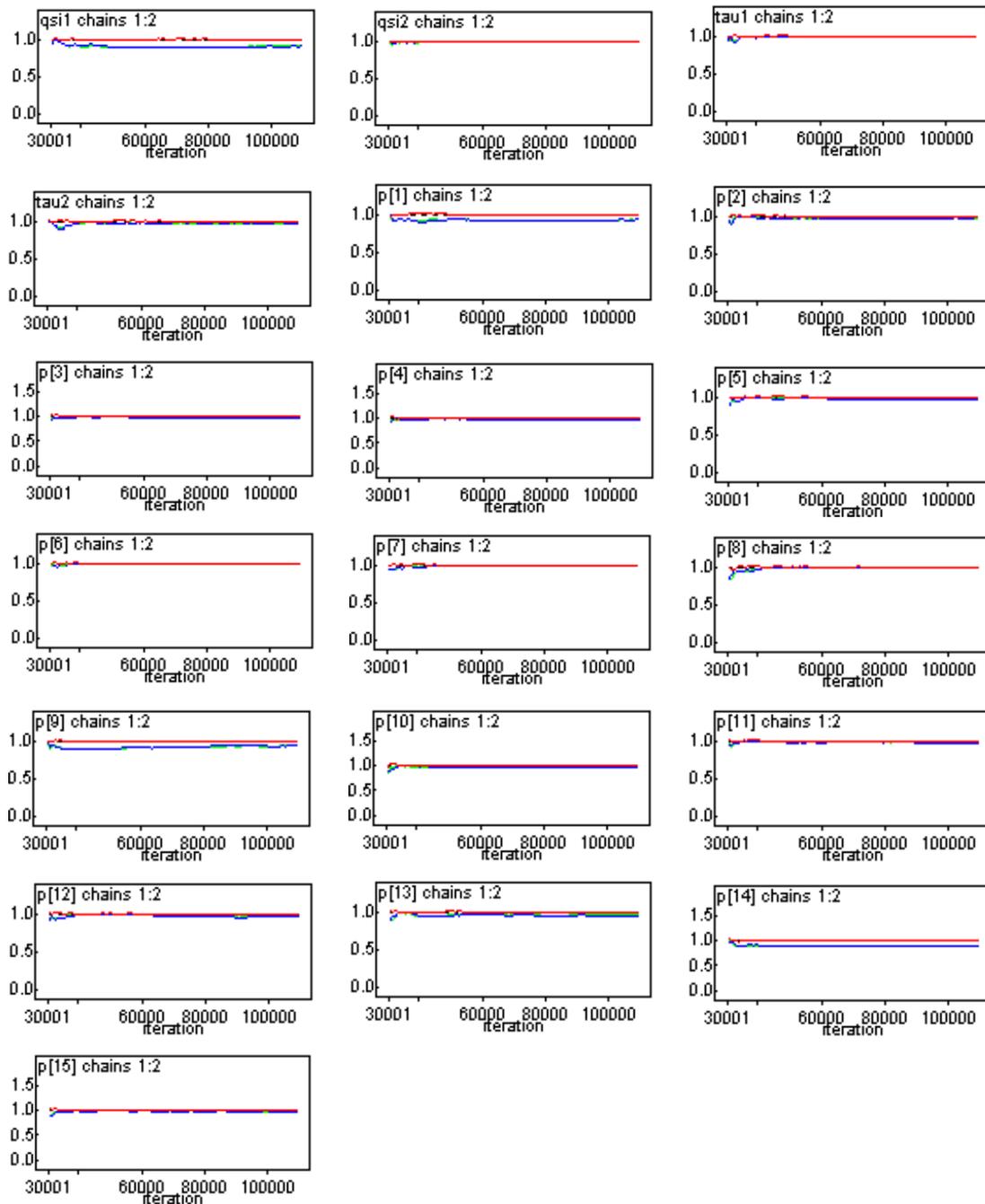


Figura 3.2: Gráficos de autocorrelações.

A Tabela 3.3 mostra a convergência dos parâmetros a partir dos índices de *Gelman e Rubin* que estão de acordo com o valor recomendado que é aproximadamente 1. Isto pode ser verificado através da Figura 3.3 que à medida que aumenta o número de iterações, temos um indicativo de convergência para os parâmetros do modelo. Na Figura 3.4, os gráficos sugerem convergência para todos os parâmetros.

Figura 3.3: Gráficos dos índices R de *Gelman e Rubin*.

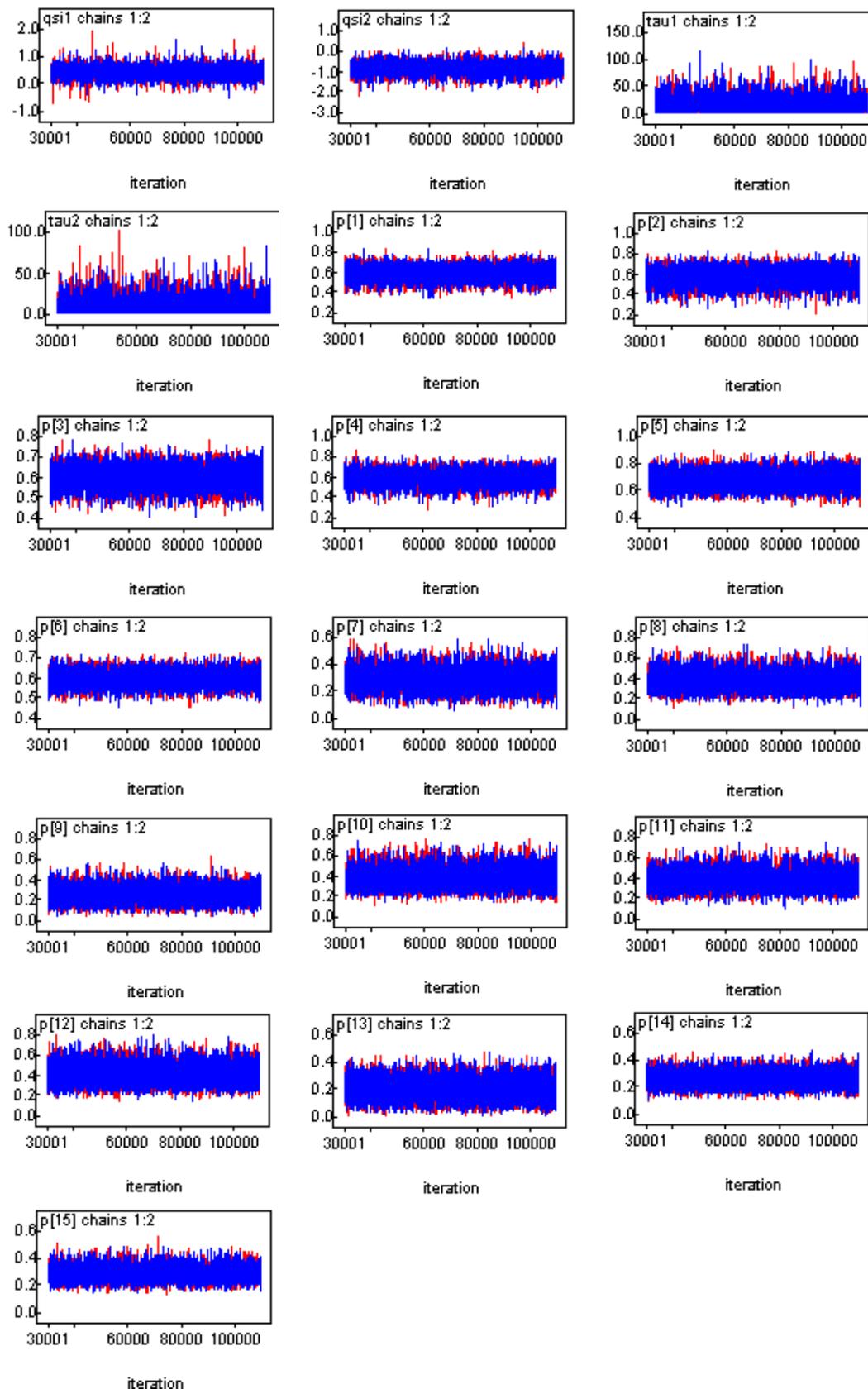


Figura 3.4: Gráficos das séries temporais.

Capítulo 4

Relação entre o uso de estrógeno e a incidência de câncer no endométrio.

Na literatura médica, há indícios que o uso da terapia de reposição de estrógeno pode aumentar o risco do câncer no endométrio. Uma possibilidade de redução do risco é a remoção de ambos os ovários, antes da menopausa.

Nos últimos anos, várias pesquisas médicas estão direcionadas para o estudo do efeito do estrógeno, pois existe a possibilidade desse hormônio contribuir para o desenvolvimento do câncer no endométrio (ver *Adami (1992)*, *Brinton e Schaiier (1993)* ou *Kaufman et al. (1991)*). Como existem pesquisadores independentes, em diferentes centros, envolvidos neste projeto de pesquisa médica, uma combinação dos resultados obtidos por pesquisadores pode ser conveniente para uma análise estatística.

Câncer no endométrio apresenta um risco significativo para a saúde e a Meta-análise pode ajudar a esclarecer muitos resultados contraditórios, entre os estudos que relatam uma possível associação entre estrógeno e câncer no endométrio.

Para investigar se o uso de estrógeno está produzindo câncer no endométrio, alguns investigadores têm recorrido à Meta-análise. Além do uso do estrógeno, podemos analisar outros potenciais fatores de risco como duração da terapia (tratamento com hormônios), tipo de hormônio usado, idade aproximada da menopausa, idade da primeira gravidez, obesidade, entre muitos outros.

Em cada laboratório onde é feito um estudo deste tipo, são determinados os riscos relativos (ver *Fleiss, 1973*), ou medidas do grau de associação entre o uso de estrógeno e

a incidência de câncer, sendo muitas vezes aproximado por uma probabilidade relativa.

Um modelo Bayesiano hierárquico, que é um sistema muito flexível para combinar informação de vários estudos usados em Meta-análise, é baseado no logaritmo do risco relativo de pacientes contraírem câncer, recebendo ou não estrógeno. O risco relativo de usuárias de estrógeno, com respeito a câncer no endométrio, pode ser calculado de forma a seguir. Seja,

A : número de pacientes com câncer;

\bar{A} : número de pacientes sem câncer;

B : número de pacientes usuárias de estrógeno;

\bar{B} : número de pacientes não usuárias de estrógeno.

Além disso, seja,

- a : número de casos expostos ou de pacientes com câncer entre as usuárias de estrógeno;

- b : número de controles expostos ou de pacientes sem câncer entre as usuárias de estrógeno;

- c : número de casos não expostos ou de pacientes com câncer entre as não usuárias de estrógeno;

- d : número de controles não expostos ou de pacientes sem câncer entre as não usuárias de estrógeno.

A partir das definições acima, temos a seguinte tabela.

Tabela 4.1: Distribuição de pacientes segundo presença de câncer e uso do estrógeno.

	A	\bar{A}	$Total$
B	a	b	$a + b$
\bar{B}	c	d	$c + d$
$Total$	$a + c$	$b + d$	n

Assim, definimos os riscos para dois grupos. O primeiro é definido a partir de pacientes que são usuárias de estrógeno e o segundo é definido a partir de pacientes que não são usuárias de estrógeno.

Para pacientes usuárias de estrógeno:

$$O_B = \frac{P(A|B)}{P(\bar{A}|B)}.$$

Para pacientes não usuárias de estrógeno:

$$O_{\bar{B}} = \frac{P(A|\bar{B})}{P(\bar{A}|\bar{B})}.$$

A medida de associação entre O_B e $O_{\bar{B}}$ é

$$O = \frac{O_B}{O_{\bar{B}}} = \frac{P(A|B)}{P(\bar{A}|B)} \times \frac{P(\bar{A}|\bar{B})}{P(A|\bar{B})}. \quad (4.1)$$

Temos, $n = a + b + c + d$, assim,

$$P(A|B) = \frac{a}{a+b}$$

$$P(\bar{A}|B) = \frac{b}{a+b}$$

$$P(\bar{A}|\bar{B}) = \frac{d}{c+d}$$

$$P(A|\bar{B}) = \frac{c}{c+d}$$

Desta forma, o **risco relativo** (RR) é dado por,

$$RR = O = \frac{a \times d}{b \times c}. \quad (4.2)$$

Observe-se que se $RR > 1$, temos um indicativo de que a chance de uma mulher ter câncer (endométrio) é maior para usuárias de estrógeno.

O objetivo principal desse estudo é obter conclusões quantitativas relacionadas aos efeitos de estrógeno sobre câncer no endométrio, e principalmente analisar o efeito de outras covariáveis sobre o risco de câncer.

4.1 O uso de estrógeno e a incidência de câncer no endométrio.

Para pesquisarmos se existe alguma relação entre estrógeno e câncer no endométrio, vamos analisar os dados da Tabela 4.2, introduzidos por *Larose e Dey* (1997) sobre 8 diferentes centros médicos que avaliam o índice de câncer no endométrio. Para análise dos dados, consideramos uma Meta-análise sob o enfoque Bayesiano (ver *Achcar et al.*, 2001). Uma análise Bayesiana destes dados foi introduzida por *Larose e Dey* (1997), assumindo uma extensão de componentes da variância hierárquica do modelo de *DuMouchel e Harris* (1983).

Tabela 4.2: Dados sobre câncer no endométrio.

Estudos	Resposta	Duração do Trat. com Estrógeno (em meses)			Tipo de Estrógeno		Obesidade	
		<12	[12,60)	≥60	Conjugado	Outro	Não	Sim
1	2.08	22	55	56	137	40	186	129
2	-0.06	68	12	11	68	134	584	126
3	0.76	49	70	121	334	122	405	146
4	1.58	18	25	39	74	10	-	-
5	0.87	11	17	49	-	-	722	140
6	0.71	57	85	39	-	-	991	188
7	1.24	26	61	98	-	-	596	190
8	1.99	8	17	27	54	45	-	-

Na Tabela 4.2, observamos que para alguns laboratórios há informações perdidas para as covariáveis e uma solução encontrada foi considerar a média das observações completas em cada vetor x_q , $q = 1, 2, 3$. De forma alternativa, poderíamos gerar as observações perdidas em cada ciclo do algoritmo Gibbs Sampling (ver por exemplo *Larose e Dey*, 1997 ou *Larose e Dey*, 1995).

Para analisar os dados da Tabela 4.2, vamos considerar o modelo de regressão dado por,

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\omega}\delta + \boldsymbol{\varepsilon}, \quad (4.3)$$

na qual $\varepsilon \stackrel{iid}{\sim} N(0, \sigma^2)$.

Seja \mathbf{y} o vetor ($n \times 1$) das respostas dadas por $y = \log(RR)$, onde (RR) denota o risco relativo de uma mulher ter câncer no endométrio entre as usuárias de estrogênio;

Seja X a matriz de covariáveis ($n \times 3$) dada por, $X = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, ou,

$$X = \begin{pmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots \\ x_{1n} & x_{2n} & x_{3n} \end{pmatrix}, \quad (4.4)$$

na qual

\mathbf{x}_1 : vetor de proporções de duração de tratamento maior ou igual a 60 meses;

\mathbf{x}_2 : vetor de proporções de usuárias de estrogênio conjugado;

\mathbf{x}_3 : vetor de proporções de indivíduos obesos em cada estudo;

$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$: vetor de parâmetros de regressão;

$\boldsymbol{\delta}$: mede a variabilidade entre os estudos, ou seja, δ é uma variável "dummy", que mede o efeito dos dois grupos diferentes (com informação completa nas covariáveis, ou não);

$\boldsymbol{\omega}$: vetor ($n \times 1$) dado por $\boldsymbol{\omega}' = (\omega_1, \dots, \omega_n)$, onde $\omega_i = 1$ ou $\omega_i = 0$ para $i = 1, 2, \dots, n$, indicando informação completa (ou não) para as covariáveis, tal que,

$$\boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{pmatrix}. \quad (4.5)$$

$\boldsymbol{\varepsilon}$: vetor de erros com elementos ε_i , $i = 1, 2, \dots, n$ ($n=8$).

Associando o modelo (4.3) com os dados da Tabela 4.2, temos,

$$X = \begin{pmatrix} 0.42 & 0.77 & 0.41 \\ 0.12 & 0.33 & 0.18 \\ 0.50 & 0.73 & 0.26 \\ 0.47 & 0.88 & - \\ 0.64 & - & 0.16 \\ 0.21 & - & 0.16 \\ 0.53 & - & 0.24 \\ 0.52 & 0.54 & - \end{pmatrix} = \begin{pmatrix} 0.42 & 0.77 & 0.41 \\ 0.12 & 0.33 & 0.18 \\ 0.50 & 0.73 & 0.26 \\ 0.47 & 0.88 & 0.24 \\ 0.64 & 0.65 & 0.16 \\ 0.21 & 0.65 & 0.16 \\ 0.53 & 0.65 & 0.24 \\ 0.52 & 0.54 & 0.24 \end{pmatrix}; \boldsymbol{\omega} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad (4.6)$$

na qual os valores das variáveis x_{11} e x_{32} são obtidas por, $x_{11} = \frac{56}{22+55+56} = 0.42$ e $x_{32} = \frac{126}{584+126} = 0.18$, seguindo o mesmo para as outras variáveis. Observe-se que para as observações perdidas, usamos as médias aritméticas dos valores observados, ou seja, usando os valores da matriz X , temos $x_{25} = \frac{0.77+0.33+0.73+0.88+0.54}{5} = 0.65$ considerando o mesmo valor para x_{26} , x_{27} e para $x_{34} = \frac{0.41+0.18+0.26+0.16+0.16+0.24}{6} = 0.24$ considerando o mesmo valor para x_{38} .

A partir da (4.3) e (4.6), usamos o algoritmo Gibbs Sampling (*Gelfand e Smith, 1990*) para determinar as quantidades *a posteriori* de interesse.

4.2 Análise Bayesiana

A partir do modelo (4.3), temos que $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \delta \omega_i + \epsilon_i$; $i = 1, 2, \dots, n$; e que $y_i \sim N\{\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \delta \omega_i; \sigma^2\}$ são independentes.

A função de verossimilhança para $\boldsymbol{\theta} = (\delta, \beta_1, \beta_2, \beta_3, \sigma^2)$ é dada por,

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2} (\epsilon_i - 0)^2\right\} \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \epsilon_i^2\right\}, \end{aligned} \quad (4.7)$$

na qual $\epsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta \omega_i$, para $i = 1, 2, \dots, n$ ($n = 8$).

Para uma análise Bayesiana, consideramos independência *a priori* entre os parâmetros e assumimos as seguintes distribuições *a priori* para δ , β_1 , β_2 , β_3 e σ^2 :

$\delta \sim N(0; a^2)$ com a conhecido;

$\beta_m \sim N(0; b_m^2)$ com $m = 1, 2, 3$; b_m conhecido;

$\sigma^2 \sim IG[e, f]$ com e e f conhecidos.

Assumindo independência *a priori* entre os parâmetros e considerando $b_1 = b_2 = b_3 = b$, podemos expressar a distribuição *a priori* conjunta para os parâmetros β , δ e σ^2 por,

$$\pi(\boldsymbol{\theta}) \propto \exp\left\{-\frac{\delta^2}{2a^2}\right\} \exp\left\{-\frac{(\beta_m^2)}{2b^2}\right\} (\sigma^2)^{-(e+1)} e^{-\frac{f}{\sigma^2}}, \quad (4.8)$$

na qual $\boldsymbol{\theta}$ é o vetor de todos os parâmetros do modelo e $m = 1, 2, 3$.

A distribuição *a posteriori* conjunta para $\delta, \beta_1, \beta_2, \beta_3$ e σ^2 é dada por,

$$\pi(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) \propto \exp\left\{-\frac{\delta^2}{2a^2}\right\} \exp\left\{-\frac{(\beta_m^2)}{2b^2}\right\} (\sigma^2)^{-(e+1)} e^{-\frac{f}{\sigma^2}} (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \epsilon_i^2\right\}, \quad (4.9)$$

na qual $m = 1, 2, 3$ e ϵ_i é dada em (4.7).

As distribuições condicionais a serem utilizadas com os algoritmos Gibbs Sampling e Metropolis-Hastings, são dadas a seguir.

A distribuição condicional para σ^2 , dados os outros parâmetros é,

$$\pi(\sigma^2|\boldsymbol{\theta}_{(\sigma^2)}, \mathbf{x}, \mathbf{y}) \propto (\sigma^2)^{-(e+\frac{n}{2}+1)} \exp\left\{-\frac{1}{\sigma^2} \left[f + \frac{1}{2} \sum_{i=1}^n \epsilon_i^2\right]\right\}.$$

Isto é,

$$\sigma^2|\boldsymbol{\theta}_{(\sigma^2)}, \mathbf{x}, \mathbf{y} \sim IG\left[e + \frac{n}{2}; f + \frac{1}{2} \sum_{i=1}^n \epsilon_i^2\right], \quad (4.10)$$

na qual $\epsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta \omega_i$, para $i = 1, 2, \dots, n$.

A distribuição condicional para δ , dados os outros parâmetros é,

$$\pi(\delta|\boldsymbol{\theta}_{(\delta)}, \mathbf{x}, \mathbf{y}) \propto \exp\left\{-\frac{\delta^2}{2a^2}\right\} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (\delta \omega_i - \mu_i^{(0)})^2\right\},$$

na qual $\mu_i^{(0)} = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i}$, $i = 1, 2, \dots, n$.

Isto é,

$$\begin{aligned}
 \pi(\delta | \boldsymbol{\theta}_{(\delta)}, \mathbf{x}, \mathbf{y}) &\propto \exp\left\{-\frac{\delta^2}{2a^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\left[\delta^2 \sum_{i=1}^n \omega_i^2 - 2\delta \sum_{i=1}^n \mu_i^{(0)} \omega_i\right]\right\} \\
 &\propto \exp\left\{-\frac{1}{2\left(\frac{1}{a^2} + \frac{1}{\sigma^2} \sum_{i=1}^n \omega_i^2\right)^{(-1)}} \left[\delta - \frac{\sum_{i=1}^n \mu_i^{(0)} \omega_i}{\sigma^2 \left[\frac{1}{a^2} + \frac{1}{\sigma^2} \sum_{i=1}^n \omega_i^2\right]}\right]^2\right\} \\
 &\propto \exp\left\{-\frac{1}{2\left(\frac{a^2 \sigma^2}{\sigma^2 + a^2 \sum_{i=1}^n \omega_i^2}\right)} \left[\delta - \frac{a^2 \sum_{i=1}^n \mu_i^{(0)} \omega_i}{\sigma^2 + a^2 \sum_{i=1}^n \omega_i^2}\right]^2\right\}.
 \end{aligned}$$

Isto é,

$$\delta | \boldsymbol{\theta}_{(\delta)}, \mathbf{x}, \mathbf{y} \sim N \left\{ \frac{a^2 \sum_{i=1}^n \mu_i^{(0)} \omega_i}{\sigma^2 + a^2 \sum_{i=1}^n \omega_i^2}, \frac{a^2 \sigma^2}{\sigma^2 + a^2 \sum_{i=1}^n \omega_i^2} \right\}, \quad (4.11)$$

na qual $\mu_i^{(0)} = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i}$, $i = 1, 2, \dots, n$.

A distribuição condicional para β_m , dados os outros parâmetros é,

$$\pi(\beta_m | \boldsymbol{\theta}_{(\beta_m)}, \mathbf{x}, \mathbf{y}) \propto \exp\left\{-\frac{\beta_m^2}{2b_m^2}\right\} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n \left(\beta_m X_{mi} - \mu_i^{(m)}\right)^2\right\},$$

na qual $\mu_i^{(m)} = y_i - \delta \omega_i - \sum_{\substack{j=1 \\ j \neq m}}^3 \beta_j X_{ji}$, para $i = 1, 2, \dots, n$; $m = 1, 2, 3$.

Isto é,

$$\begin{aligned}
 \pi(\beta_m | \boldsymbol{\theta}_{(\beta_m)}, \mathbf{x}, \mathbf{y}) &\propto \exp\left\{-\frac{\beta_m^2}{2b^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\left[\beta_m^2 \sum_{i=1}^n X_{mi}^2 - 2\beta_m \sum_{i=1}^n \mu_i^{(m)} X_{mi}\right]\right\} \\
 &\propto \exp\left\{-\frac{1}{2\left(\frac{1}{b^2} + \frac{1}{\sigma^2} \sum_{i=1}^n X_{mi}^2\right)^{(-1)}} \left[\beta_m - \frac{\sum_{i=1}^n \mu_i^{(m)} X_{mi}}{\sigma^2 \left(\frac{1}{b^2} + \frac{1}{\sigma^2} \sum_{i=1}^n X_{mi}^2\right)}\right]^2\right\} \\
 &\propto \exp\left\{-\frac{1}{2\left(\frac{b^2\sigma^2}{\sigma^2 + b^2} \sum_{i=1}^n X_{mi}^2\right)} \left[\beta_m - \frac{b^2 \sum_{i=1}^n \mu_i^{(m)} X_{mi}}{\sigma^2 + b^2 \sum_{i=1}^n X_{mi}^2}\right]^2\right\}.
 \end{aligned}$$

Isto é,

$$\beta_m | \boldsymbol{\theta}_{(\beta_m)}, \mathbf{x}, \mathbf{y} \sim N \left\{ \frac{b^2 \sum_{i=1}^n \mu_i^{(m)} X_{mi}}{\sigma^2 + b^2 \sum_{i=1}^n X_{mi}^2}; \frac{b^2\sigma^2}{\sigma^2 + b^2 \sum_{i=1}^n X_{mi}^2} \right\},$$

na qual $\mu_i^{(m)} = y_i - \delta\omega_i - \sum_{\substack{j=1 \\ j \neq m}}^3 \beta_j X_{ji}$, para $i = 1, 2, \dots, n$; $m = 1, 2, 3$.

Desta forma, as distribuições condicionais para β_1 , β_2 e β_3 individualmente são dadas por:

$$\beta_1 | \boldsymbol{\theta}_{(\beta_1)}, \mathbf{x}, \mathbf{y} \sim N \left\{ \frac{b^2 \sum_{i=1}^n \mu_i^{(1)} X_{1i}}{\sigma^2 + b^2 \sum_{i=1}^n X_{1i}^2}; \frac{b^2\sigma^2}{\sigma^2 + b^2 \sum_{i=1}^n X_{1i}^2} \right\}, \quad (4.12)$$

na qual $\mu_i^{(1)} = y_i - \beta_2 X_{2i} - \beta_3 X_{3i} - \delta\omega_i$, para $i = 1, 2, \dots, n$;

$$\beta_2 | \boldsymbol{\theta}_{(\beta_2)}, \mathbf{x}, \mathbf{y} \sim N \left\{ \frac{b^2 \sum_{i=1}^n \mu_i^{(2)} X_{2i}}{\sigma^2 + b^2 \sum_{i=1}^n X_{2i}^2}; \frac{b^2 \sigma^2}{\sigma^2 + b^2 \sum_{i=1}^n X_{2i}^2} \right\}, \quad (4.13)$$

na qual $\mu_i^{(2)} = y_i - \beta_1 X_{1i} - \beta_3 X_{3i} - \delta\omega_i$, para $i = 1, 2, \dots, n$;

$$\beta_3 | \boldsymbol{\theta}_{(\beta_3)}, \mathbf{x}, \mathbf{y} \sim N \left\{ \frac{b^2 \sum_{i=1}^n \mu_i^{(3)} X_{3i}}{\sigma^2 + b^2 \sum_{i=1}^n X_{3i}^2}; \frac{b^2 \sigma^2}{\sigma^2 + b^2 \sum_{i=1}^n X_{3i}^2} \right\}, \quad (4.14)$$

na qual $\mu_i^{(3)} = y_i - \beta_1 X_{1i} - \beta_2 X_{2i} - \delta\omega_i$, para $i = 1, 2, \dots, n$.

Interpretação dos resultados.

Considerando o modelo (4.3) para os dados da Tabela 4.2, e as distribuições *a priori* para os parâmetros do modelo, com $a = b = e = f = 1$ assumindo independência *a priori* não informativa entre os parâmetros, geramos duas cadeias de Gibbs Sampling. De cada uma das cadeias com 110000 iterações e considerando saltos de tamanho dez, e finalmente obtemos uma amostra de tamanho 16000, após as 30000 primeiras observações serem eliminadas ("burn in"). A convergência das amostras de Gibbs Sampling foi monitorada usando o Critério de *Gelman e Rubin* (1992) ($\sqrt{\hat{R}} < 1.1$ para todos os parâmetros). A Tabela 4.3 mostra os sumários *a posteriori* e os valores $\sqrt{\hat{R}}$ (Critério de *Gelman e Rubin*, 1992).

Tabela 4.3: Sumários *a posteriori* para os dados da Tabela 4.2 e o modelo (4.3).

Parâmetros	Média a posteriori	Intervalos de 95% de credibilidade	$\sqrt{\widehat{R}}$
β_1	0.214	(-1.262; 1.669)	1.00
β_2	-1.192	(-2.372; -0.0103)	0.99
β_3	9.71	(7.874; 11.53)	0.99
δ	-1.168	(-1.941; -0.372)	1.00
τ	3.333	(0.9664; 7.162)	0.99

Após obter os resultados da Tabela 4.3, vamos analisar se há alguma relação entre a doença e seus fatores de risco.

Dos intervalos de credibilidade na Tabela 4.3, podemos observar os fatores que indicam efeitos significativos no risco relativo, ou seja, que levam ao aumento da incidência de câncer entre as usuárias de estrógeno. Esses são o tipo de estrógeno usado pelas mulheres, indicado pela variável x_2 , a obesidade, indicada pela variável x_3 . A variabilidade entre os estudos indicando se há informações perdidas ou não, representada por δ , e a variabilidade representada por τ , o inverso de σ^2 . O fator duração do tratamento com estrógeno, indicado pela variável x_1 , não apresenta efeito significativo sobre o risco relativo (intervalos de 95% de credibilidade inclui o zero).

Também temos na Tabela 4.3 os índices de *Gelman e Rubin*. Por esses índices observamos convergência do algoritmo Gibbs Sampling.

A Figura 4.1 mostra os gráficos das distribuições *a posteriori* e a Figura 4.2 mostra os gráficos das autocorrelações, onde as densidades se comportam de acordo com as distribuições *a priori* especificadas e os parâmetros apresentam-se não correlacionados.

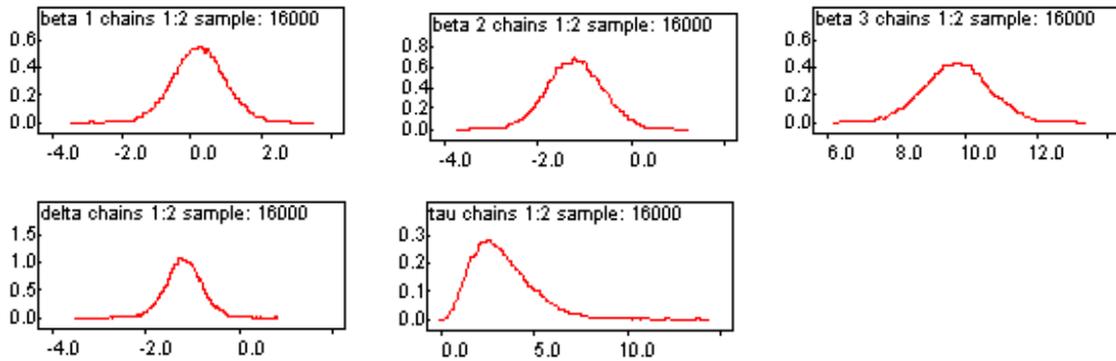


Figura 4.1: Densidades *a posteriori* marginais.

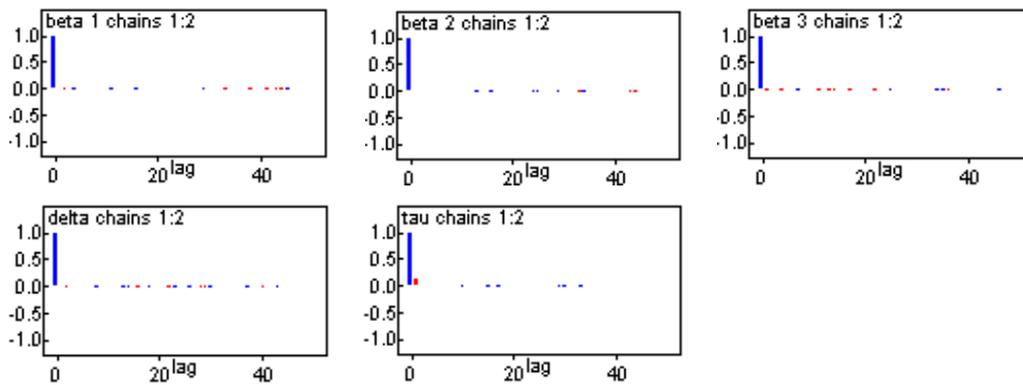


Figura 4.2: Gráficos de autocorrelações.

A Tabela 4.3 mostra que os índices de convergência *Gelman e Rubin* que estão abaixo do limite 1 recomendado, e as Figuras 4.3 e 4.4 mostram que à medida que aumenta o número de iterações, há um indicativo de convergência para todos os parâmetros do modelo.

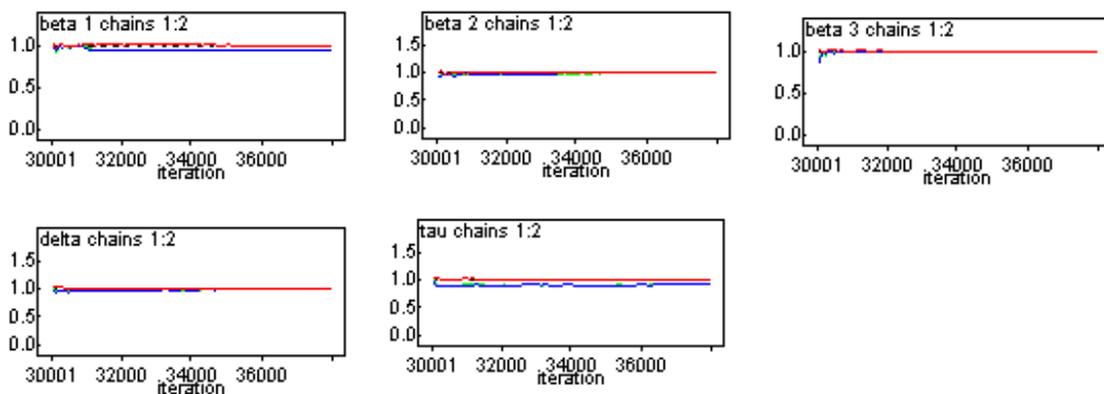


Figura 4.3: Gráficos dos índices R de *Gelman e Rubin*.

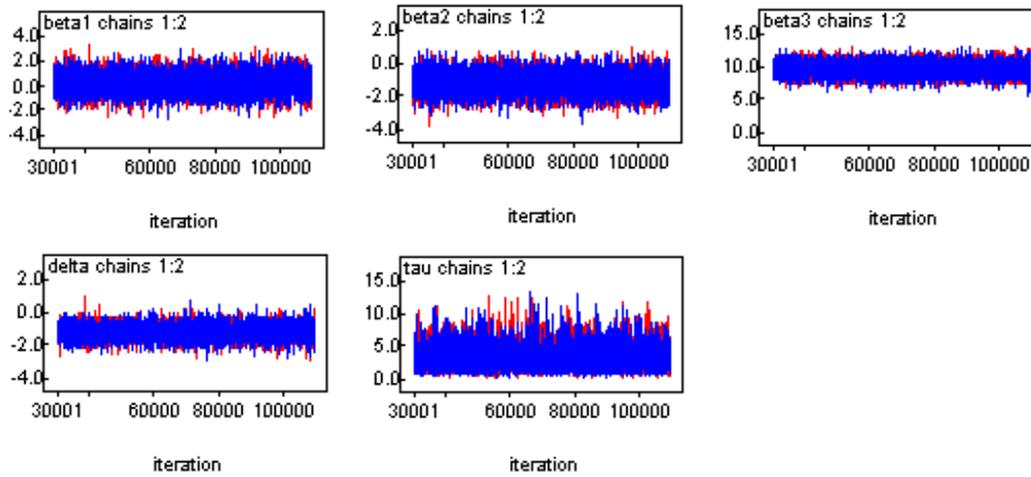


Figura 4.4: Gráficos das séries temporais.

Como motivação para o próximo capítulo, vamos construir um histograma dos resíduos ou mesmo dos dados originais, para verificar como estes se comportam. Considerando o modelo $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \delta \omega_i + \varepsilon_i$ para os dados das Tabelas 4.1 e 4.2. Os valores dos resíduos são: 0.095, -0.275, 0.165, 0.199, -0.043, -0.113, -0.425, 0.196.

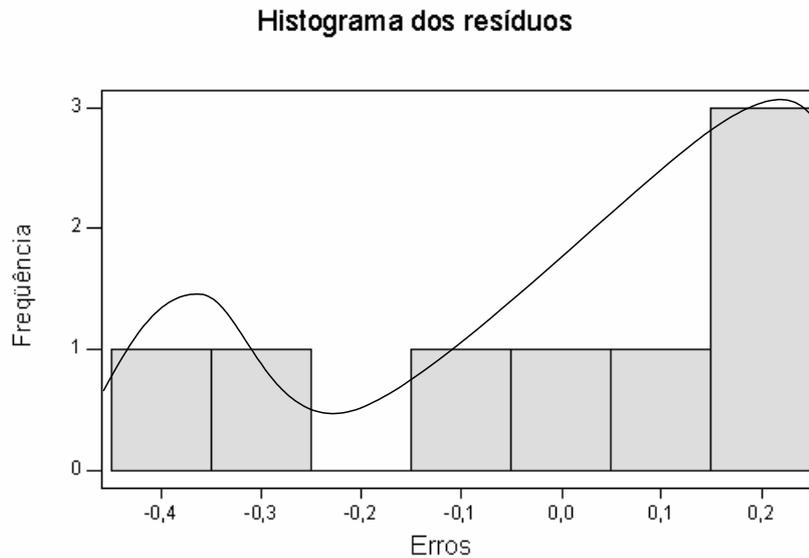


Figura 4.5: Histograma dos resíduos.

O histograma dos resíduos (Figura 4.5), mostra que os resíduos possuem aproximadamente uma mistura de duas distribuições normais. Portanto, uma reanálise dos dados usando uma mistura de duas distribuições normais para o erro do modelo é sugerida.

Capítulo 5

Modelo com mistura finita de distribuições normais para o erro.

Neste capítulo introduzimos um modelo assumindo uma mistura finita de distribuições normais para o erro, na presença de covariáveis.

As misturas finitas de distribuições são aplicadas a populações compostas de mais de uma componente, ou seja, usual em problemas práticos, quando os valores da variável aleatória correspondem a duas ou mais condições diferentes.

Uma primeira aplicação de mistura de distribuições ocorreu em 1890, quando *W.R. Weldon* consultou *Karl Pearson* sobre um experimento que estava fazendo (ver *Everitt e Hand*, 1981). Os dados deste experimento eram da razão entre o tamanho da cabeça e o comprimento do corpo inteiro, para 1000 caranguejos. Fazendo uma análise gráfica, observou-se assimetria, a qual *Weldon* sugeriu que poderia decorrer do fato da amostra conter dois tipos de caranguejos e não terem sido etiquetados quando coletados.

Em termos matemáticos, a distribuição sugerida por *Pearson* (1894) para as medidas dos caranguejos é a seguinte: $f(x) = pN(\mu_1, \sigma_1) + (1 - p)N(\mu_2, \sigma_2)$, sendo que p é a proporção de caranguejos de espécie que seguem a distribuição $N(\mu_1, \sigma_1^2)$ e $(1 - p)$ é a proporção de caranguejos da espécie que seguem a distribuição $N(\mu_2, \sigma_2^2)$. O objetivo é estimar a proporção p de observações geradas da distribuição normal ($N(\mu_1, \sigma_1)$). Temos também o exemplo das distribuições de altura em uma população de adultos, que reflete a mistura de indivíduos masculino e feminino na população. A mistura de modelos pode ser usada em situações deste tipo, na qual a população é dividida em subpopulações.

5.1 Modelo com mistura de distribuições para o erro.

Considere-se o modelo de regressão (4.3), na qual os erros têm uma função densidade de probabilidade dada por,

$$f(\varepsilon_i) = \sum_{j=1}^J p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2). \quad (5.1)$$

A expressão (5.1) representa uma mistura de J modelos, onde, $p_j > 0$ com $\sum_{j=1}^J p_j = 1$. Os parâmetros p_1, p_2, \dots, p_J são chamados de proporções da mistura, $f_j(\varepsilon_i | \mu_j, \sigma_j^2)$ são as componentes da mistura e vamos assumir $f_j(\varepsilon_i | \mu_j, \sigma_j^2) \sim N(\mu_j, \sigma_j^2)$, $j = 1, 2$ já que é usual assumirmos distribuição normal para os erros, para melhor ajuste dos dados. Assume-se que os erros são independentes. Também vamos supor que $p_1 \mu_1 + \dots + p_J \mu_J = 0$. Com apenas dois componentes, temos $p_1 \mu_1 + p_2 \mu_2 = 0$, isto é, $\mu_2 = -\frac{p_1 \mu_1}{p_2}$. Observar que $f_j(\varepsilon_i | \mu_j, \sigma_j^2) \propto \sigma_j^{-1} \exp \left\{ -\frac{1}{2\sigma_j^2} (\varepsilon_i - \mu_j)^2 \right\}$.

A distribuição a *posteriori* conjunta para $\boldsymbol{\theta}$ (vetor dos parâmetros) é dada por,

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \pi(\boldsymbol{\theta}) \times L(\boldsymbol{\theta}), \quad (5.2)$$

na qual $\pi(\boldsymbol{\theta})$ é a distribuição a *priori* conjunta dos parâmetros e a função de verossimilhança para $\boldsymbol{\theta}$ é dada por,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \left\{ \sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2) \right\}. \quad (5.3)$$

Buscando simplificar a distribuição a *posteriori* e as distribuições condicionais completas, necessárias para o algoritmo Gibbs Sampling, serão introduzidas variáveis latentes. As variáveis latentes são usadas para classificar as observações em relação aos componentes, além de simplificar as condicionais através da eliminação do somatório da função de verossimilhança em (5.3).

Neste caso, vamos considerar uma mistura com dois modelos, ou seja, $J = 2$. Dessa forma, temos um vetor $Z_i = (Z_{i1}, Z_{i2})$, $i = 1, 2, \dots, n$, para cada observação e $Z_{i1} \sim \text{Bernoulli}(h_{i1})$, na qual h_{i1} é a probabilidade de sucesso de cada componente da

distribuição de *Bernoulli*, definida por (ver *Achcar, Mazucheli e Kass, 2001* e *Diebolt e Robert, 1994*),

$$h_{i1} = \frac{p_1 f_1(\varepsilon_i | \mu_1, \sigma_1^2)}{\sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)}, \quad (5.4)$$

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

Desse modo, assumimos que a distribuição condicional para Z_i tem uma distribuição de *Bernoulli* com probabilidade de sucesso h_{i1} , ou seja, $\pi(Z_{i1}) \propto \prod_{i=1}^n h_{i1}^{Z_{i1}} (1 - h_{i1})^{Z_{i2}}$, $i = 1, \dots, n$, na qual

$$Z_{i1} = \begin{cases} 0, & \text{com probabilidade } (1 - h_{i1}); \\ 1, & \text{com probabilidade } h_{i1}. \end{cases},$$

$Z_{i1} + Z_{i2} = 1$ e Z_{i1} faz com que a variável Z em cada observação i seja atribuída a somente uma das componentes da mistura.

A partir da distribuição de *Bernoulli* e da equação (5.4), temos a distribuição conjunta para (Z_1, \dots, Z_n) dada por,

$$\begin{aligned} \pi(Z_1, \dots, Z_n) &\propto \prod_{i=1}^n h_{i1}^{Z_{i1}} \times (1 - h_{i1})^{Z_{i2}} \\ &\propto \prod_{i=1}^n \left(\frac{p_1 f_1(\varepsilon_i | \mu_1, \sigma_1^2)}{\sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)} \right)^{Z_{i1}} \times \left(1 - \frac{p_1 f_1(\varepsilon_i | \mu_1, \sigma_1^2)}{\sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)} \right)^{Z_{i2}} \\ &\propto \prod_{i=1}^n \left(\frac{p_1 f_1(\varepsilon_i | \mu_1, \sigma_1^2)}{\sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)} \right)^{Z_{i1}} \times \left(\frac{p_2 f_2(\varepsilon_i | \mu_2, \sigma_2^2)}{\sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)} \right)^{Z_{i2}} \\ &\propto \prod_{i=1}^n \frac{[p_1 f_1(\varepsilon_i | \mu_1, \sigma_1^2)]^{Z_{i1}} \times [p_2 f_2(\varepsilon_i | \mu_2, \sigma_2^2)]^{Z_{i2}}}{\left[\sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2) \right]^{Z_{i1} + Z_{i2}}} \\ &\propto \frac{\prod_{i=1}^n \prod_{j=1}^2 \{p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)\}^{Z_{ij}}}{\prod_{i=1}^n \left\{ \sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2) \right\}}. \end{aligned}$$

Isto é,

$$\pi(Z_1, \dots, Z_n) \propto \frac{\prod_{i=1}^n \prod_{j=1}^2 \{p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)\}^{Z_{ij}}}{\prod_{i=1}^n \left\{ \sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2) \right\}} \quad (5.5)$$

Combinando as equações (5.3) e (5.5), temos a distribuição a *posteriori* conjunta para $(\boldsymbol{\theta}, \mathbf{Z})$ dada por,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{Z} | \mathbf{y}) &\propto \pi(\boldsymbol{\theta}) \times \frac{\prod_{i=1}^n \prod_{j=1}^2 \{p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)\}^{Z_{ij}}}{\prod_{i=1}^n \left\{ \sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2) \right\}} \times \prod_{i=1}^n \left\{ \sum_{j=1}^2 p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2) \right\} \\ &\propto \pi(\boldsymbol{\theta}) \times \prod_{i=1}^n \prod_{j=1}^2 [p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)]^{Z_{ij}}. \end{aligned}$$

Isto é,

$$\pi(\boldsymbol{\theta}, \mathbf{Z} | \mathbf{y}) \propto \pi(\boldsymbol{\theta}) \times \prod_{i=1}^n \prod_{j=1}^2 [p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)]^{Z_{ij}}. \quad (5.6)$$

Comparando a equação (5.6) com a (5.2), observa-se que a função de verossimilhança é definida por,

$$L(\boldsymbol{\theta}, \mathbf{Z}) = \prod_{i=1}^n \prod_{j=1}^2 [p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)]^{Z_{ij}} \quad (5.7)$$

Como ilustração, vamos analisar os dados da Tabela 4.1 com o modelo (4.3) e com uma mistura de duas distribuições normais para o erro.

Neste caso, a distribuição a *posteriori* conjunta para $\boldsymbol{\theta} = (p_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \beta_1, \beta_2, \beta_3, \delta)$ é dada por (5.6), na qual $\varepsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta \omega_i$.

Distribuição a *priori* conjunta

Assumindo independência entre as seguintes distribuições a *priori* para os parâmetros:

$$\pi(\delta, \beta_1, \beta_2, \beta_3) \propto \text{constante};$$

$$\sigma_j^2 \sim IG[a_j, b_j]; \quad a_j \text{ e } b_j \text{ conhecidos } j = 1, 2;$$

$\mu_j \sim N[c_j; d_j^2]$; c_j e d_j conhecidos $j = 1, 2$;

$p_1 \sim \text{Beta}[e, f]$; e e f conhecidos.

Isto é, a distribuição a *priori* conjunta é dada por,

$$\pi(\boldsymbol{\theta}) \propto \left\{ \prod_{j=1}^2 (\sigma_j^2)^{-(a_j+1)} e^{-\frac{b_j}{\sigma_j^2}} \right\} \times \left\{ \prod_{j=1}^2 \exp\left(-\frac{1}{2d_j^2} (\mu_j - c_j)^2\right) \right\} \times p_1^{e-1} (1-p_1)^{f-1}, \quad (5.8)$$

na qual $\boldsymbol{\theta} = (p_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \beta_1, \beta_2, \beta_3, \delta)$.

Observa-se que devemos assumir distribuições a *priori* informativas para μ_j ; $j = 1, 2$; ou seja, para assegurar a identificabilidade do modelo.

Distribuição a *posteriori* conjunta

A distribuição a *posteriori* conjunta para $\boldsymbol{\theta}$ e \mathbf{Z} é dada por,

$$\begin{aligned} \pi(\boldsymbol{\theta}, \mathbf{Z} | \mathbf{y}) &\propto \pi(\boldsymbol{\theta}) \times \prod_{i=1}^n \prod_{j=1}^2 [p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)]^{Z_{ij}} \\ &\propto \left\{ \prod_{j=1}^2 (\sigma_j^2)^{-(a_j+1)} e^{-\frac{b_j}{\sigma_j^2}} \right\} \times \left\{ \prod_{j=1}^2 \exp\left(-\frac{1}{2d_j^2} (\mu_j - c_j)^2\right) \right\} \\ &\quad \times p_1^{e-1} (1-p_1)^{f-1} \times \prod_{i=1}^n \prod_{j=1}^2 [p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)]^{Z_{ij}}, \end{aligned} \quad (5.9)$$

na qual $\boldsymbol{\theta} = (p_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \beta_1, \beta_2, \beta_3, \delta)$.

Distribuições condicionais

A partir da distribuição a *posteriori* conjunta para $\boldsymbol{\theta}$, determinamos as distribuições condicionais utilizando o algoritmo Gibbs Sampling.

A distribuição condicional para β_1 dados os outros parâmetros é dada por,

$$\pi(\beta_1 | \boldsymbol{\theta}_{(\beta_1)}, \mathbf{Z}, \mathbf{y}) \propto \prod_{i=1}^n \prod_{j=1}^2 [p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)]^{Z_{ij}},$$

na qual $\boldsymbol{\theta}_{(\beta_1)}$ é o vetor de todos os parâmetros, exceto β_1 , $\varepsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta \omega_i$, $i = 1, \dots, n$ e

$$f_j(\varepsilon_i|\mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2\sigma_j^2}(\varepsilon_i - \mu_j)^2\right\} \propto \exp\left\{-\frac{1}{2\sigma_j^2}(\varepsilon_i - \mu_j)^2\right\}.$$

Isto é,

$$\begin{aligned} \pi(\beta_1|\boldsymbol{\theta}_{(\beta_1)}, \mathbf{Z}, \mathbf{y}) &\propto \prod_{i=1}^n \prod_{j=1}^2 \left[\exp\left\{-\frac{1}{2\sigma_j^2}(\varepsilon_i - \mu_j)^2\right\} \right]^{Z_{ij}} \\ &\propto \exp\left\{-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^2 \frac{Z_{ij}}{\sigma_j^2} (\varepsilon_i - \mu_j)^2\right\}. \end{aligned}$$

Assim,

$$\pi(\beta_1|\boldsymbol{\theta}_{(\beta_1)}, \mathbf{Z}, \mathbf{y}) \propto \exp\left\{-\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2\right\}.$$

Vamos utilizar uma transformação para simplificar a distribuição condicional. Seja, $(\varepsilon_i - \mu_1)^2 = (\beta_1 x_{1i} - \lambda_{1i}^{(1)})^2$, na qual $\lambda_{1i}^{(1)} = y_i - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta\omega_i - \mu_1$. Assim, $\varepsilon_i - \mu_1^2 = (\beta_1 x_{1i} - \lambda_{1i}^{(1)})^2 = \beta_1^2 x_{1i}^2 - 2\beta_1 x_{1i} \lambda_{1i}^{(1)} + \lambda_{1i}^{(1)2}$. Seja, $(\varepsilon_i - \mu_2)^2 = (\beta_1 x_{1i} - \lambda_{1i}^{(2)})^2$, na qual $\lambda_{1i}^{(2)} = y_i - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta\omega_i - \mu_2$. Assim, $\varepsilon_i - \mu_2^2 = (\beta_1 x_{1i} - \lambda_{1i}^{(2)})^2 = \beta_1^2 x_{1i}^2 - 2\beta_1 x_{1i} \lambda_{1i}^{(2)} + \lambda_{1i}^{(2)2}$. Então,

$$\begin{aligned} \pi(\beta_1|\boldsymbol{\theta}_{(\beta_1)}, \mathbf{Z}, \mathbf{y}) &\propto \exp\left\{-\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2\right\} \\ &\propto \exp\left\{-\frac{\beta_1^2}{2} \left[\frac{\sum_{i=1}^n Z_{i1} x_{1i}^2}{\sigma_1^2} + \frac{\sum_{i=1}^n Z_{i2} x_{1i}^2}{\sigma_2^2} \right] + \beta_1 \left[\frac{\sum_{i=1}^n Z_{i1} x_{1i} \lambda_{1i}^{(1)}}{\sigma_1^2} + \frac{\sum_{i=1}^n Z_{i2} x_{1i} \lambda_{1i}^{(2)}}{\sigma_2^2} \right]\right\}. \end{aligned}$$

Definindo:

$$\begin{aligned} A_1 &= \frac{1}{\sigma_1^2} \sum_{i=1}^n Z_{i1} x_{1i}^2 + \frac{1}{\sigma_2^2} \sum_{i=1}^n Z_{i2} x_{1i}^2 \\ B_1 &= \frac{1}{\sigma_1^2} \sum_{i=1}^n Z_{i1} x_{1i} \lambda_{1i}^{(1)} + \frac{1}{\sigma_2^2} \sum_{i=1}^n Z_{i2} x_{1i} \lambda_{1i}^{(2)} \end{aligned}$$

Assim,

$$\begin{aligned}\pi(\beta_1 | \boldsymbol{\theta}_{(\beta_1)}, \mathbf{Z}, \mathbf{y}) &\propto \exp\left\{-\frac{\beta_1^2}{2}A_1 + \beta_1 B_1\right\} \\ &\propto \exp\left\{-\frac{1}{2A_1^{-1}}\left[\beta_1 - \frac{B_1}{A_1}\right]^2\right\}.\end{aligned}$$

Então,

$$\beta_1 | \boldsymbol{\theta}_{(\beta_1)}, \mathbf{Z}, \mathbf{y} \sim N\left\{\frac{B_1}{A_1}; A_1^{-1}\right\}.$$

Portanto,

$$\beta_1 | \boldsymbol{\theta}_{(\beta_1)}, \mathbf{Z}, \mathbf{y} \sim N\left\{\frac{\sigma_2^2 \sum_{i=1}^n Z_{i1} x_{1i} \lambda_{1i}^{(1)} + \sigma_1^2 \sum_{i=1}^n Z_{i2} x_{1i} \lambda_{1i}^{(2)}}{\sigma_2^2 \sum_{i=1}^n Z_{i1} x_{1i}^2 + \sigma_1^2 \sum_{i=1}^n Z_{i2} x_{1i}^2}; \frac{\sigma_1^2 \sigma_2^2}{\sigma_2^2 \sum_{i=1}^n Z_{i1} x_{1i}^2 + \sigma_1^2 \sum_{i=1}^n Z_{i2} x_{1i}^2}\right\}, \quad (5.10)$$

na qual $\boldsymbol{\theta}_{(\beta_1)}$ é o vetor de todos os parâmetros, exceto β_1 .

A distribuição condicional para β_2 dados os outros parâmetros é dada por,

$$\pi(\beta_2 | \boldsymbol{\theta}_{(\beta_2)}, \mathbf{Z}, \mathbf{y}) \propto \prod_{i=1}^n \prod_{j=1}^2 [p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)]^{Z_{ij}},$$

na qual $\boldsymbol{\theta}_{(\beta_2)}$ é o vetor de todos os parâmetros, exceto β_2 , $\varepsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta\omega_i$, $i = 1, \dots, n$ e

$$f_j(\varepsilon_i | \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2\sigma_j^2}(\varepsilon_i - \mu_j)^2\right\} \propto \exp\left\{-\frac{1}{2\sigma_j^2}(\varepsilon_i - \mu_j)^2\right\}.$$

Isto é,

$$\begin{aligned}\pi(\beta_2 | \boldsymbol{\theta}_{(\beta_2)}, \mathbf{Z}, \mathbf{y}) &\propto \prod_{i=1}^n \prod_{j=1}^2 \left[\exp\left\{-\frac{1}{2\sigma_j^2}(\varepsilon_i - \mu_j)^2\right\} \right]^{Z_{ij}} \\ &\propto \exp\left\{-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^2 \frac{Z_{ij}}{\sigma_j^2} (\varepsilon_i - \mu_j)^2\right\}.\end{aligned}$$

Assim,

$$\pi(\beta_2 | \boldsymbol{\theta}_{(\beta_2)}, \mathbf{Z}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2 \right\}.$$

Seja, $(\varepsilon_i - \mu_1)^2 = (\beta_2 x_{2i} - \lambda_{2i}^{(1)})^2$, na qual $\lambda_{2i}^{(1)} = y_i - \beta_1 x_{1i} - \beta_3 x_{3i} - \delta \omega_i - \mu_1$. Assim, $(\varepsilon_i - \mu_1)^2 = (\beta_2 x_{2i} - \lambda_{2i}^{(1)})^2 = \beta_2^2 x_{2i}^2 - 2\beta_2 x_{2i} \lambda_{2i}^{(1)} + \lambda_{2i}^{(1)2}$. Seja, $(\varepsilon_i - \mu_2)^2 = (\beta_2 x_{2i} - \lambda_{2i}^{(2)})^2$, na qual $\lambda_{2i}^{(2)} = y_i - \beta_1 x_{1i} - \beta_3 x_{3i} - \delta \omega_i - \mu_2$. Assim, $(\varepsilon_i - \mu_2)^2 = (\beta_2 x_{2i} - \lambda_{2i}^{(2)})^2 = \beta_2^2 x_{2i}^2 - 2\beta_2 x_{2i} \lambda_{2i}^{(2)} + \lambda_{2i}^{(2)2}$. Então,

$$\begin{aligned} \pi(\beta_2 | \boldsymbol{\theta}_{(\beta_2)}, \mathbf{Z}, \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2 \right\} \\ &\propto \exp \left\{ -\frac{\beta_2^2}{2} \left[\frac{\sum_{i=1}^n Z_{i1} x_{2i}^2}{\sigma_1^2} + \frac{\sum_{i=1}^n Z_{i2} x_{2i}^2}{\sigma_2^2} \right] + \beta_2 \left[\frac{\sum_{i=1}^n Z_{i1} x_{2i} \lambda_{2i}^{(1)}}{\sigma_1^2} + \frac{\sum_{i=1}^n Z_{i2} x_{2i} \lambda_{2i}^{(2)}}{\sigma_2^2} \right] \right\}. \end{aligned}$$

Definindo:

$$A_2 = \frac{1}{\sigma_1^2} \sum_{i=1}^n Z_{i1} x_{2i}^2 + \frac{1}{\sigma_2^2} \sum_{i=1}^n Z_{i2} x_{2i}^2$$

$$B_2 = \frac{1}{\sigma_1^2} \sum_{i=1}^n Z_{i1} x_{2i} \lambda_{2i}^{(1)} + \frac{1}{\sigma_2^2} \sum_{i=1}^n Z_{i2} x_{2i} \lambda_{2i}^{(2)}$$

Assim,

$$\begin{aligned} \pi(\beta_2 | \boldsymbol{\theta}_{(\beta_2)}, \mathbf{Z}, \mathbf{y}) &\propto \exp \left\{ -\frac{\beta_2^2}{2} A_2 + \beta_2 B_2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2A_2^{-1}} \left[\beta_2 - \frac{B_2}{A_2} \right]^2 \right\}. \end{aligned}$$

Então,

$$\beta_2 | \boldsymbol{\theta}_{(\beta_2)}, \mathbf{Z}, \mathbf{y} \sim N \left\{ \frac{B_2}{A_2}; A_2^{-1} \right\}.$$

Portanto,

$$\beta_2 | \boldsymbol{\theta}_{(\beta_2)}, \mathbf{Z}, \mathbf{y} \sim N \left\{ \frac{\sigma_2^2 \sum_{i=1}^n Z_{i1} x_{2i} \lambda_{2i}^{(1)} + \sigma_1^2 \sum_{i=1}^n Z_{i2} x_{2i} \lambda_{2i}^{(2)}}{\sigma_2^2 \sum_{i=1}^n Z_{i1} x_{2i}^2 + \sigma_1^2 \sum_{i=1}^n Z_{i2} x_{2i}^2}; \frac{\sigma_1^2 \sigma_2^2}{\sigma_2^2 \sum_{i=1}^n Z_{i1} x_{2i}^2 + \sigma_1^2 \sum_{i=1}^n Z_{i2} x_{2i}^2} \right\}, \quad (5.11)$$

na qual $\boldsymbol{\theta}_{(\beta_2)}$ é o vetor de todos os parâmetros, exceto β_2 .

A distribuição condicional para β_3 dados os outros parâmetros é dada por,

$$\pi(\beta_3 | \boldsymbol{\theta}_{(\beta_3)}, \mathbf{Z}, \mathbf{y}) \propto \prod_{i=1}^n \prod_{j=1}^2 [p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)]^{Z_{ij}},$$

na qual $\boldsymbol{\theta}_{(\beta_3)}$ é o vetor de todos os parâmetros, exceto β_3 , $\varepsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta\omega_i$, $i = 1, \dots, n$ e

$$f_j(\varepsilon_i | \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{1}{2\sigma_j^2} (\varepsilon_i - \mu_j)^2\right\} \propto \exp\left\{-\frac{1}{2\sigma_j^2} (\varepsilon_i - \mu_j)^2\right\}.$$

Isto é,

$$\begin{aligned} \pi(\beta_3 | \boldsymbol{\theta}_{(\beta_3)}, \mathbf{Z}, \mathbf{y}) &\propto \prod_{i=1}^n \prod_{j=1}^2 \left[\exp\left\{-\frac{1}{2\sigma_j^2} (\varepsilon_i - \mu_j)^2\right\} \right]^{Z_{ij}} \\ &\propto \exp\left\{-\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^2 \frac{Z_{ij}}{\sigma_j^2} (\varepsilon_i - \mu_j)^2\right\}. \end{aligned}$$

Assim,

$$\pi(\beta_3 | \boldsymbol{\theta}_{(\beta_3)}, \mathbf{Z}, \mathbf{y}) \propto \exp\left\{-\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2\right\}.$$

Seja, $(\varepsilon_i - \mu_1)^2 = (\beta_3 x_{3i} - \lambda_{3i}^{(1)})^2$, na qual $\lambda_{3i}^{(1)} = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \delta\omega_i - \mu_1$. Assim, $(\varepsilon_i - \mu_1)^2 = (\beta_3 x_{3i} - \lambda_{3i}^{(1)})^2 = \beta_3^2 x_{3i}^2 - 2\beta_3 x_{3i} \lambda_{3i}^{(1)} + \lambda_{3i}^{(1)2}$. Seja, $(\varepsilon_i - \mu_2)^2 = (\beta_3 x_{3i} - \lambda_{3i}^{(2)})^2$, na qual $\lambda_{3i}^{(2)} = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \delta\omega_i - \mu_2$. Assim, $(\varepsilon_i - \mu_2)^2 = (\beta_3 x_{3i} - \lambda_{3i}^{(2)})^2 = \beta_3^2 x_{3i}^2 - 2\beta_3 x_{3i} \lambda_{3i}^{(2)} + \lambda_{3i}^{(2)2}$. Então,

$$\begin{aligned} \pi(\beta_3 | \boldsymbol{\theta}_{(\beta_3)}, \mathbf{Z}, \mathbf{y}) &\propto \exp\left\{-\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2\right\} \\ &\propto \exp\left\{-\frac{\beta_3^2}{2} \left[\frac{\sum_{i=1}^n Z_{i1} x_{3i}^2}{\sigma_1^2} + \frac{\sum_{i=1}^n Z_{i2} x_{3i}^2}{\sigma_2^2} \right] + \beta_3 \left[\frac{\sum_{i=1}^n Z_{i1} x_{3i} \lambda_{3i}^{(1)}}{\sigma_1^2} + \frac{\sum_{i=1}^n Z_{i2} x_{3i} \lambda_{3i}^{(2)}}{\sigma_2^2} \right]\right\}. \end{aligned}$$

Definindo:

$$A_3 = \frac{1}{\sigma_1^2} \sum_{i=1}^n Z_{i1} x_{3i}^2 + \frac{1}{\sigma_2^2} \sum_{i=1}^n Z_{i2} x_{3i}^2$$

$$B_3 = \frac{1}{\sigma_1^2} \sum_{i=1}^n Z_{i1} x_{3i} \lambda_{3i}^{(1)} + \frac{1}{\sigma_2^2} \sum_{i=1}^n Z_{i2} x_{3i} \lambda_{3i}^{(2)}$$

Assim,

$$\begin{aligned} \pi(\beta_3 | \boldsymbol{\theta}_{(\beta_3)}, \mathbf{Z}, \mathbf{y}) &\propto \exp \left\{ -\frac{\beta_3^2}{2} A_3 + \beta_3 B_3 \right\} \\ &\propto \exp \left\{ -\frac{1}{2A_3^{-1}} \left[\beta_3 - \frac{B_3}{A_3} \right]^2 \right\}. \end{aligned}$$

Então,

$$\beta_3 | \boldsymbol{\theta}_{(\beta_3)}, \mathbf{Z}, \mathbf{y} \sim N \left\{ \frac{B_3}{A_3}; A_3^{-1} \right\}.$$

Portanto,

$$\beta_3 | \boldsymbol{\theta}_{(\beta_3)}, \mathbf{Z}, \mathbf{y} \sim N \left\{ \frac{\sigma_2^2 \sum_{i=1}^n Z_{i1} x_{3i} \lambda_{3i}^{(1)} + \sigma_1^2 \sum_{i=1}^n Z_{i2} x_{3i} \lambda_{3i}^{(2)}}{\sigma_2^2 \sum_{i=1}^n Z_{i1} x_{3i}^2 + \sigma_1^2 \sum_{i=1}^n Z_{i2} x_{3i}^2}; \frac{\sigma_1^2 \sigma_2^2}{\sigma_2^2 \sum_{i=1}^n Z_{i1} x_{3i}^2 + \sigma_1^2 \sum_{i=1}^n Z_{i2} x_{3i}^2} \right\}, \quad (5.12)$$

na qual $\boldsymbol{\theta}_{(\beta_3)}$ é o vetor de todos os parâmetros, exceto β_3

A distribuição condicional para δ dados os outros parâmetros é dada por,

$$\pi(\delta | \boldsymbol{\theta}_{(\delta)}, \mathbf{Z}, \mathbf{y}) \propto \prod_{i=1}^n \prod_{j=1}^2 [p_j f_j(\varepsilon_i | \mu_j, \sigma_j^2)]^{Z_{ij}},$$

na qual $\boldsymbol{\theta}_{(\delta)}$ é o vetor de todos os parâmetros, exceto δ , $\varepsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta \omega_i$, $i = 1, \dots, n$ e

$$f_j(\varepsilon_i | \mu_j, \sigma_j^2) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left\{ -\frac{1}{2\sigma_j^2} (\varepsilon_i - \mu_j)^2 \right\} \propto \exp \left\{ -\frac{1}{2\sigma_j^2} (\varepsilon_i - \mu_j)^2 \right\}.$$

Isto é,

$$\begin{aligned}\pi(\delta|\boldsymbol{\theta}_{(\delta)}, \mathbf{Z}, \mathbf{y}) &\propto \prod_{i=1}^n \prod_{j=1}^2 \left[\exp \left\{ -\frac{1}{2\sigma_j^2} (\varepsilon_i - \mu_j)^2 \right\} \right]^{Z_{ij}} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^2 \frac{Z_{ij}}{\sigma_j^2} (\varepsilon_i - \mu_j)^2 \right\}.\end{aligned}$$

Assim,

$$\pi(\delta|\boldsymbol{\theta}_{(\delta)}, \mathbf{Z}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2 \right\}.$$

Seja, $(\varepsilon_i - \mu_1)^2 = (\delta\omega_i - \lambda_{4i}^{(1)})^2$, na qual $\lambda_{4i}^{(1)} = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \mu_1$. Assim, $(\varepsilon_i - \mu_1)^2 = (\delta\omega_i - \lambda_{4i}^{(1)})^2 = \delta^2\omega_i^2 - 2\delta\omega_i\lambda_{4i}^{(1)} + \lambda_{4i}^{(1)2}$. Seja, $(\varepsilon_i - \mu_2)^2 = (\delta\omega_i - \lambda_{4i}^{(2)})^2$, na qual $\lambda_{4i}^{(2)} = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \mu_2$. Assim, $(\varepsilon_i - \mu_2)^2 = (\delta\omega_i - \lambda_{4i}^{(2)})^2 = \delta^2\omega_i^2 - 2\delta\omega_i\lambda_{4i}^{(2)} + \lambda_{4i}^{(2)2}$. Então,

$$\begin{aligned}\pi(\delta|\boldsymbol{\theta}_{(\delta)}, \mathbf{Z}, \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2 \right\} \\ &\propto \exp \left\{ -\frac{\delta^2}{2} \left[\frac{\sum_{i=1}^n Z_{i1}\omega_i^2}{\sigma_1^2} + \frac{\sum_{i=1}^n Z_{i2}\omega_i^2}{\sigma_2^2} \right] + \delta \left[\frac{\sum_{i=1}^n Z_{i1}\omega_i\lambda_{4i}^{(1)}}{\sigma_1^2} + \frac{\sum_{i=1}^n Z_{i2}\omega_i\lambda_{4i}^{(2)}}{\sigma_2^2} \right] \right\}.\end{aligned}$$

Definindo:

$$A_4 = \frac{1}{\sigma_1^2} \sum_{i=1}^n Z_{i1}\omega_i^2 + \frac{1}{\sigma_2^2} \sum_{i=1}^n Z_{i2}\omega_i^2$$

$$B_4 = \frac{1}{\sigma_1^2} \sum_{i=1}^n Z_{i1}\omega_i\lambda_{4i}^{(1)} + \frac{1}{\sigma_2^2} \sum_{i=1}^n Z_{i2}\omega_i\lambda_{4i}^{(2)}$$

Assim,

$$\begin{aligned}\pi(\delta|\boldsymbol{\theta}_{(\delta)}, \mathbf{Z}, \mathbf{y}) &\propto \exp \left\{ -\frac{\delta^2}{2} A_4 + \delta B_4 \right\} \\ &\propto \exp \left\{ -\frac{1}{2A_4^{-1}} \left[\delta - \frac{B_4}{A_4} \right]^2 \right\}.\end{aligned}$$

Então,

$$\delta|\boldsymbol{\theta}_{(\delta)}, \mathbf{Z}, \mathbf{y} \sim N \left\{ \frac{B_4}{A_4}, A_4^{-1} \right\}.$$

Portanto,

$$\delta|\boldsymbol{\theta}_{(\delta)}, \mathbf{Z}, \mathbf{y} \sim N \left\{ \frac{\sigma_2^2 \sum_{i=1}^n Z_{i1} \omega_i \lambda_{4i}^{(1)} + \sigma_1^2 \sum_{i=1}^n Z_{i2} \omega_i \lambda_{4i}^{(2)}}{\sigma_2^2 \sum_{i=1}^n Z_{i1} \omega_i^2 + \sigma_1^2 \sum_{i=1}^n Z_{i2} \omega_i^2}; \frac{\sigma_1^2 \sigma_2^2}{\sigma_2^2 \sum_{i=1}^n Z_{i1} \omega_i^2 + \sigma_1^2 \sum_{i=1}^n Z_{i2} \omega_i^2} \right\}, \quad (5.13)$$

na qual $\boldsymbol{\theta}_{(\delta)}$ é o vetor de todos os parâmetros, exceto δ .

A distribuição condicional para μ_1 dados os outros parâmetros é dada por,

$$\pi(\mu_1 | \boldsymbol{\theta}_{(\mu_1)}, \mathbf{Z}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2d_1^2} (\mu_1 - c_1)^2 \right\} \times \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 \right\},$$

na qual $\boldsymbol{\theta}_{(\mu_1)}$ é o vetor de todos os parâmetros, exceto μ_1 , $\varepsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta \omega_i$, $i = 1, \dots, n$.

Portanto,

$$\begin{aligned} \pi(\mu_1 | \boldsymbol{\theta}_{(\mu_1)}, \mathbf{Z}, \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2d_1^2} (\mu_1 - c_1)^2 \right\} \times \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^n Z_{i1} (\mu_1^2 - 2\mu_1 \varepsilon_i) \right\} \\ &\propto \exp \left\{ -\frac{\mu_1^2}{2} \left[\frac{1}{d_1^2} + \frac{\sum_{i=1}^n Z_{i1}}{\sigma_1^2} \right] + \mu_1 \left[\frac{c_1}{d_1^2} + \frac{\sum_{i=1}^n Z_{i1} \varepsilon_i}{\sigma_1^2} \right] \right\}. \end{aligned}$$

Definindo:

$$A_5 = \frac{1}{d_1^2} + \frac{\sum_{i=1}^n Z_{i1}}{\sigma_1^2}$$

$$B_5 = \frac{c_1}{d_1^2} + \frac{\sum_{i=1}^n Z_{i1} \varepsilon_i}{\sigma_1^2} = \frac{c_1 \sigma_1^2 + d_1^2 \sum_{i=1}^n Z_{i1} \varepsilon_i}{d_1^2 \sigma_1^2}$$

Assim,

$$\begin{aligned} \pi(\mu_1 | \boldsymbol{\theta}_{(\mu_1)}, \mathbf{Z}, \mathbf{y}) &\propto \exp \left\{ -\frac{\mu_1^2}{2} A_5 + \mu_1 B_5 \right\} \\ &\propto \exp \left\{ -\frac{1}{2A_5^{-1}} \left[\mu_1 - \frac{B_5}{A_5} \right]^2 \right\}. \end{aligned}$$

Então,

$$\mu_1 | \boldsymbol{\theta}_{(\mu_1)}, \mathbf{Z}, \mathbf{y} \sim N \left\{ \frac{B_5}{A_5}; A_5^{-1} \right\}.$$

Portanto,

$$\mu_1 | \boldsymbol{\theta}_{(\mu_1)}, \mathbf{Z}, \mathbf{y} \sim N \left\{ \frac{c_1 \sigma_1^2 + d_1^2 \sum_{i=1}^n Z_{i1} \varepsilon_i}{\sigma_1^2 + d_1^2 \sum_{i=1}^n Z_{i1}}; \frac{d_1^2 \sigma_1^2}{\sigma_1^2 + d_1^2 \sum_{i=1}^n Z_{i1}} \right\}, \quad (5.14)$$

na qual $\boldsymbol{\theta}_{(\mu_1)}$ é o vetor de todos os parâmetros, exceto μ_1 e $\varepsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta \omega_i$, $i = 1, \dots, n$.

Nota: observar da relação $\mu_1 p_1 + \mu_2 p_2 = 0$, que $\mu_2 p_2 = -\mu_1 p_1$ ou $\mu_2 = -\frac{\mu_1 p_1}{p_2}$ e $p_2 = 1 - p_1$.

Daí, gerar μ_1 de $\pi(\mu_1 | \boldsymbol{\theta}_{(\mu_1)}, \mathbf{Z}, \mathbf{y})$ e calcular μ_2 .

A distribuição condicional para σ_1^2 dados os outros parâmetros é dada por,

$$\pi(\sigma_1^2 | \boldsymbol{\theta}_{(\sigma_1^2)}, \mathbf{Z}, \mathbf{y}) \propto (\sigma_1^2)^{-(a_1 + \frac{n}{2} + 1)} \exp \left\{ -\frac{1}{\sigma_1^2} \left[b_1 + \frac{1}{2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 \right] \right\}.$$

Isto é,

$$\sigma_1^2 | \boldsymbol{\theta}_{(\sigma_1^2)}, \mathbf{Z}, \mathbf{y} \sim IG \left[a_1 + \frac{n}{2}; b_1 + \frac{1}{2} \sum_{i=1}^n Z_{i1} (\varepsilon_i - \mu_1)^2 \right], \quad (5.15)$$

na qual $\boldsymbol{\theta}_{(\sigma_1^2)}$ é o vetor de todos os parâmetros, exceto σ_1^2 e $\varepsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta \omega_i$, $i = 1, \dots, n$.

A distribuição condicional para σ_2^2 dados os outros parâmetros é dada por,

$$\pi(\sigma_2^2 | \boldsymbol{\theta}_{(\sigma_2^2)}, \mathbf{Z}, \mathbf{y}) \propto (\sigma_2^2)^{-(a_2 + \frac{n}{2} + 1)} \exp \left\{ -\frac{1}{\sigma_2^2} \left[b_2 + \frac{1}{2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2 \right] \right\}.$$

Isto é,

$$\sigma_2^2 | \boldsymbol{\theta}_{(\sigma_2^2)}, \mathbf{Z}, \mathbf{y} \sim IG \left[a_2 + \frac{n}{2}; b_2 + \frac{1}{2} \sum_{i=1}^n Z_{i2} (\varepsilon_i - \mu_2)^2 \right], \quad (5.16)$$

na qual $\boldsymbol{\theta}_{(\sigma_2^2)}$ é o vetor de todos os parâmetros, exceto σ_2^2 e $\varepsilon_i = y_i - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} - \delta \omega_i$, $i = 1, \dots, n$.

A distribuição condicional para p_1 dados os outros parâmetros é dada por,

$$\begin{aligned} \pi(p_1 | \boldsymbol{\theta}_{(p_1)}, \mathbf{Z}, \mathbf{y}) &\propto p_1^{e-1} (1-p_1)^{f-1} \times \left[\prod_{i=1}^n p_1^{Z_{i1}} \right] \left[\prod_{i=1}^n (1-p_1)^{Z_{i2}} \right] \\ &\propto p_1^{e + \sum_{i=1}^n Z_{i1} - 1} (1-p_1)^{f + \sum_{i=1}^n Z_{i2} - 1}. \end{aligned}$$

Isto é,

$$p_1 | \boldsymbol{\theta}_{(p_1)}, \mathbf{Z}, \mathbf{y} \sim \text{Beta}\left(e + \sum_{i=1}^n Z_{i1}; f + \sum_{i=1}^n Z_{i2}\right), \quad (5.17)$$

na qual $\boldsymbol{\theta}_{(p_1)}$ é o vetor de todos os parâmetros, exceto p_1 .

Nota: $p_2 = 1 - p_1$; basta atualizar p_2 a partir do valor gerado para p_1 .

Consideramos os dados da Tabela 4.1 e o modelo $y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + \omega_i\delta + \epsilon_i$, $i = 1, 2, \dots, n$, dado em (4.3).

Com a distribuição *a priori* para os parâmetros do modelo, assumindo os valores para os hiperparâmetros $a_1 = a_2 = b_1 = b_2 = 1$, $c_1 = -0.35$, $d_1 = 20$ e $e = f = 3$, geramos duas cadeias de Gibbs, cada uma com 60000 iterações. Para cada série, consideramos saltos de tamanho dez, e daí obtemos uma amostra de tamanho 10000. Eliminamos 10000 primeiras observações geradas ("burn in"). A convergência das amostras de Gibbs foi monitorada usando o Critério de *Gelman e Rubin* (1992), ($\sqrt{\hat{R}} < 1.1$ para todos os parâmetros) e métodos gráficos. O valor aproximado de $c_1 = -0.35$ foi obtido através do histograma dos resíduos (Figura 5.1).

Dessa forma, encontramos os resultados (ver Tabela 5.1) de interesse.

Tabela 5.1: Sumários a *posteriori*

para os dados na Tabela 4.2 e o modelo 4.3.

Parâmetros	Média a posteriori	Intervalos de 95% de credibilidade	$\sqrt{\widehat{R}}$
β_1	0.259	(-1.357;1.875)	0.99
β_2	-1.07	(-2.681;0.544)	0.99
β_3	9.734	(7.845;11.6)	0.99
δ	-1.085	(-2.028;-0.116)	0.99
μ_1	-0.353	(-0.788;0.082)	0.99
μ_2	0.338	(-0.062;1.353)	1.00
p_1	0.431	(0.136;0.755)	0.99
p_2	0.568	(0.245;0.865)	0.99
τ_1	2.181	(0.242;5.9)	1.00
τ_2	2.251	(0.265;5.813)	0.99

Na Tabela 5.1, podemos observar que os fatores que indicam efeitos significativos no risco relativo, ou seja, que levam ao aumento da incidência de câncer entre as usuárias de estrógeno estão relacionados com a obesidade indicada pela variável x_3 , com a variabilidade entre os estudos indicando informações perdidas ou não, indicado por δ , com a proporção de resíduos de tipo 1, indicada por p_1 , com a proporção de resíduos de tipo 2, indicada por p_2 e com as variâncias indicadas por τ_1 e τ_2 , na qual τ_1 é definida pelo inverso de σ_1^2 assim como para τ_2 , estas conclusões são obtidas através dos intervalos de credibilidade, na qual não incluem o zero. Analisando o fator duração do tratamento com estrógeno indicado pela variável x_1 , o tipo de estrógeno usado pelas mulheres indicado pela variável x_2 e as médias indicadas por μ_1 e μ_2 , observamos que estas não apresentam indicações de efeitos significativos do risco relativo (intervalos de 95% de credibilidade inclui o zero).

Também temos na Tabela 5.1 os índices de *Gelman e Rubin*. Por esses índices observamos convergência do algoritmo Gibbs Sampling.

5.1.1 Análise gráfica.

As distribuições *a posteriori* para os parâmetros de interesse estão especificadas na Figura 5.1, onde estas se comportam de acordo com as distribuições *a priori* especificadas e na Figura 5.2 os parâmetros apresentam-se não correlacionados.

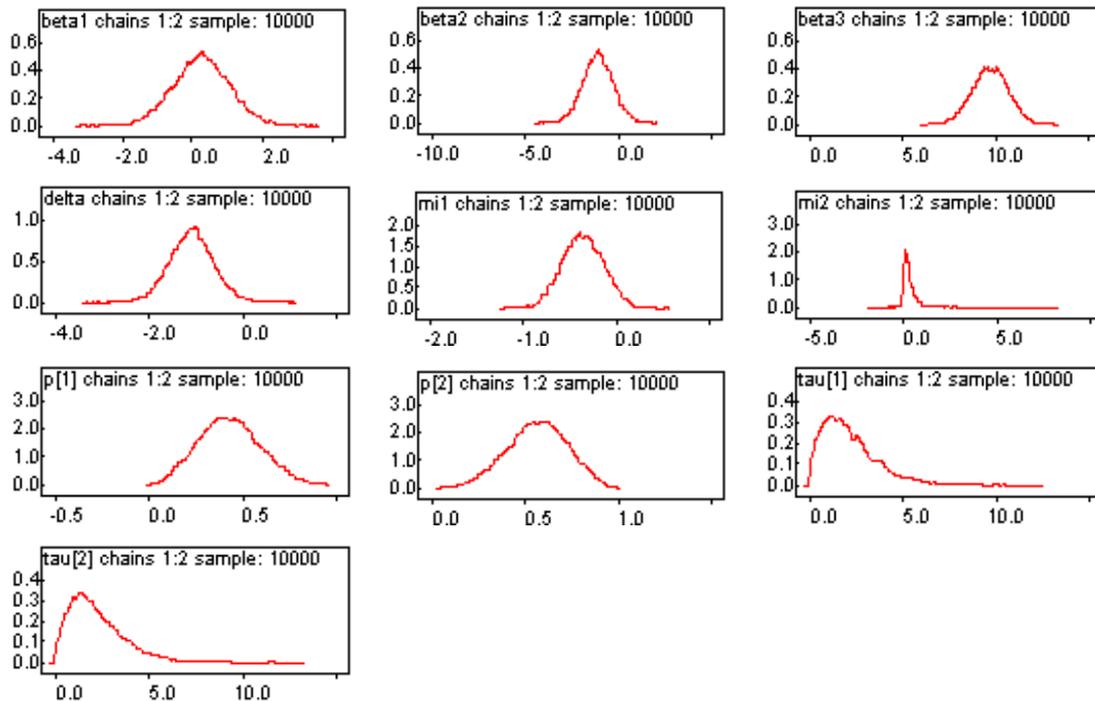
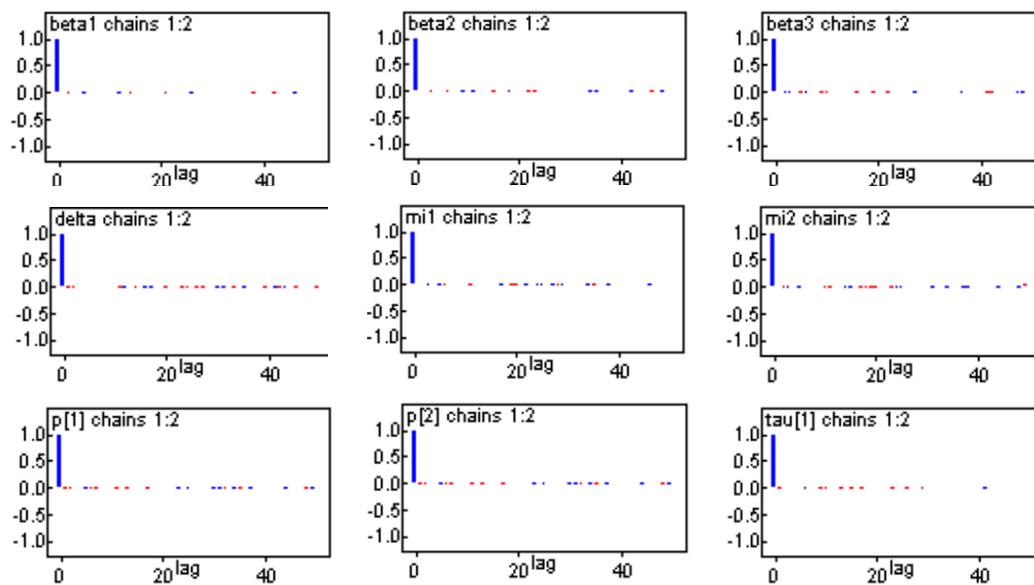


Figura 5.1: Densidades *a posteriori* marginais.



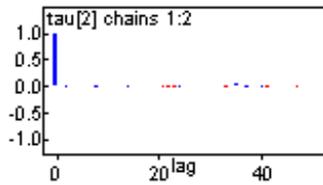


Figura 5.2: Gráficos de autocorrelações.

A Tabela 5.1 mostra os índices de *Gelman e Rubin* que estão de acordo com o valor recomendado que é aproximadamente 1, e na Figura 5.3 observa-se que à medida que aumenta o número de iterações, temos um indicativo de convergência para todos os parâmetros do modelo. A Figura 5.4, mostra uma indicação de convergência para todos os parâmetros, através do gráfico de séries temporais.

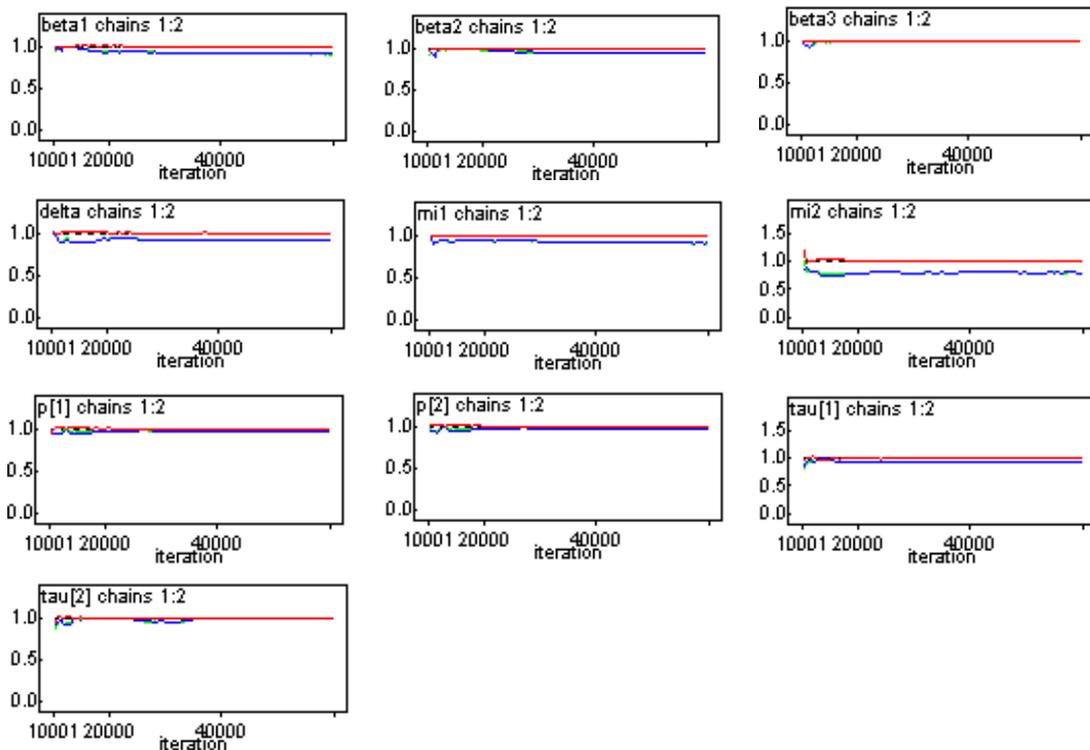


Figura 5.3: Gráficos dos índices R de *Gelman e Rubin*.

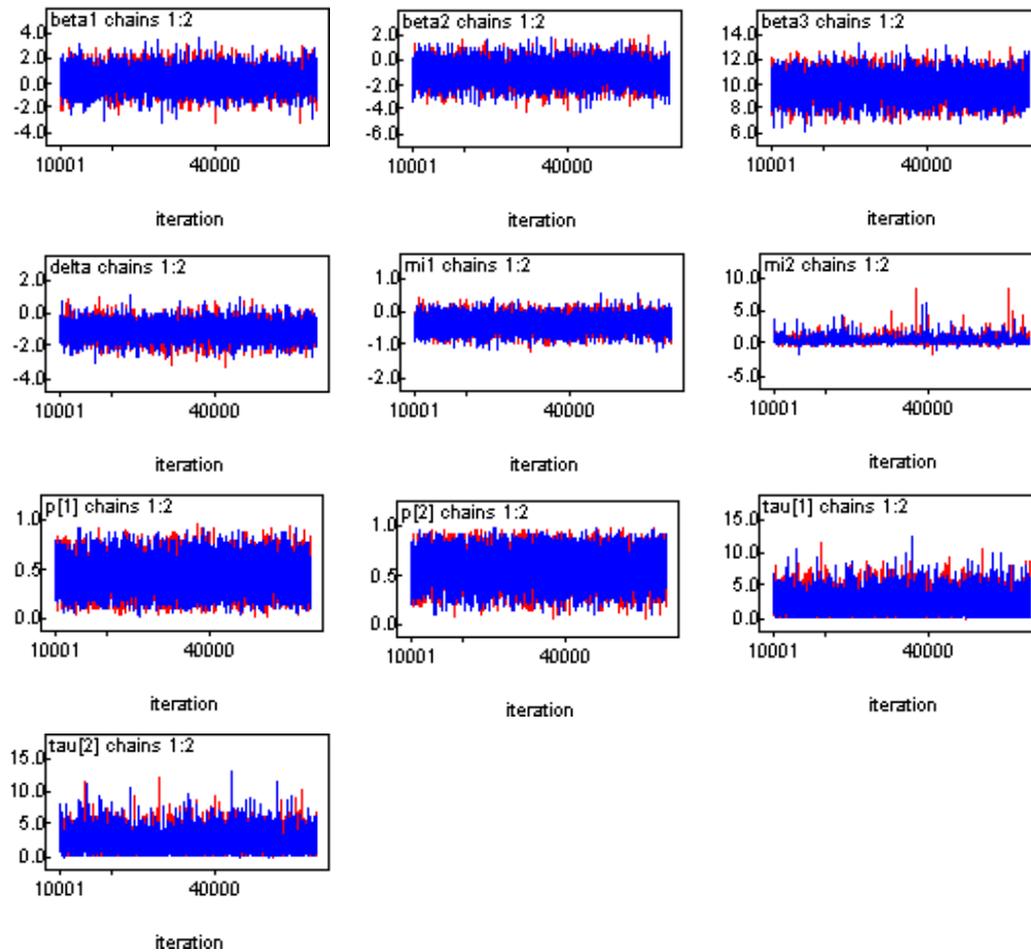


Figura 5.4: Gráficos das séries temporais.

5.2 Critério de seleção de Modelos.

Para uma análise detalhada dos dados, devemos sempre considerar o problema da validade do modelo. Isto é, o problema da avaliação e escolha do modelo que melhor represente a situação em estudo. Existem vários critérios baseados em aproximações ao fator de Bayes (BF) para discriminar dois modelos M_1 e M_2 , em especial o critério BIC (*Bayesian Information Criterion*) e o critério DIC (*Deviance Information Criterion*). Vamos considerar como critério de seleção o critério BIC (ver *Paulino et al.*, 2003).

O critério BIC é definido por,

$$\widehat{BIC}_i = -2 \times E[\ln(L(\theta_i|x, M_i))] + p_i \times \ln(n),$$

na qual n é o tamanho da amostra e p_i é o número de parâmetros sob o modelo M_i ,

$i = 1, 2$. Escolhe-se o modelo que apresenta maior valor de \widehat{BIC}_i , $i = 1, 2$.

Seja, M_1 o modelo com uma distribuição normal para o erro e M_2 é o modelo com uma mistura de duas distribuições normais para o erro. Os valores para os dois modelos são $\widehat{BIC}_1 = 15.741$ para M_1 e $\widehat{BIC}_2 = 31.772$ para M_2 .

De acordo com esses valores de \widehat{BIC} , o modelo M_2 com uma mistura de duas distribuições normais para o erro é o melhor modelo para análise dos dados da Tabela 4.2.

Capítulo 6

Meta-Análise de efeitos de aspirina em ataques cardíacos.

Na década de 70, seis grandes experimentos clínicos aleatorizados multicêntricos sobre a eficiência do uso da aspirina no tratamento de pacientes, após um ataque cardíaco (infarto do miocárdio, IM) por todas as causas, foram conduzidos na Europa e nos Estados Unidos. Neste capítulo vamos analisar o conjunto de dados dispostos na Tabela 6.1, na qual temos o número de pacientes e taxa de mortalidade para todas as causas desses seis experimentos controlados e aleatorizados. Em particular, vamos apresentar uma análise clássica baseada em testes de hipóteses e intervalos de confiança e compará-la com uma análise Bayesiana baseada em intervalos de credibilidade. Temos como objetivo principal comparar o uso de aspirina e placebo em pacientes após um ataque cardíaco.

As seis pesquisas em consideração incluem duas realizadas na Inglaterra dirigidas por ELWOOD ET. AL. (1974) denominada de UK-1 e por ELWOOD & SWETNAN (1979), UK-2. Uma realizada na Alemanha e Áustria, o "German-Autrian Multicenter Study", GAMS, BREDDIN ET. AL.(1979). Outra pesquisa é o "Coronary Drug Project Aspirin Study", CDPA (1976). A Quinta é a "Persantine-Aspirin Reinfarction Study Research Group", PARIS (1980). E a última "Aspirin Myocardial Infarction Study Research Group", AMIS (1980) (ver *Costa*, 1999).

A Tabela 6.1 mostra o número de pacientes e a taxa de mortalidade para os dois tratamentos em cada estudo.

Tabela 6.1: Dados de uso de aspirina e placebo em pacientes após um ataque cardíaco.

Estudo	Aspirina		Placebo	
	Número de pacientes	Taxa de mortalidade (%)	Número de pacientes	Taxa de mortalidade (%)
UK-1	615	7,97	624	10,74
CDPA	758	5,80	771	8,30
GAMS	317	8,52	309	10,36
UK-2	832	12,26	850	14,82
PARIS	810	10,49	406	12,81
AMIS	2267	10,85	2257	9,70
Geral	5599	9,88	5217	10,73

Fonte: DRAPER, D.; GAVER, D.; GOEL, P.; GREENHOUSE, J.; HEDGES, L.; MORRIS, C.; TUCKER, J.; WATERNAUX (1993). *Combining Information: Statistical Issues and opportunities in research*, Contemporary Statistics series, n.1, American Statistical Association, Alexandria, Va.

A inspeção dos dados mostra índices de mortalidade semelhantes para os dois grupos, havendo uma diferença em sentido contrário no último estudo (AMIS), na qual o índice de mortalidade para aspirina é mais ou menos próximo de outros resultados obtidos, em particular com os resultados obtidos em PARIS. No entanto, embora o índice de mortalidade para placebo em AMIS não apresente também uma grande diferença em relação aos outros obtidos, a menos de UK-2, ele é menor do que o seu índice obtido para aspirina.

6.1 Análise Clássica

Vamos considerar dados de dois grupos, na qual y_{1i} e y_{2i} são os números de mortes de pacientes que usam aspirina ou placebo, respectivamente.

Seja,

n_{1i} : número de pacientes que tomaram aspirina no hospital i ;

n_{2i} : número de pacientes que tomaram placebo no hospital i .

Assumir que p_{1i} é a proporção de mortes sob o tratamento com aspirina e p_{2i} é a proporção de mortes sob o tratamento com placebo. Para cada elemento da amostra, consideramos a ocorrência de um sucesso, caso a morte do paciente se verifique, e de um fracasso, caso contrário. Desta forma y_{1i} e y_{2i} seguem distribuições Binomiais com parâmetros n_{1i} , p_{1i} e n_{2i} , p_{2i} , respectivamente.

Logo, $E(y_{1i}) = n_{1i}p_{1i}$, $E(y_{2i}) = n_{2i}p_{2i}$, $Var(y_{1i}) = n_{1i}p_{1i}(1 - p_{1i})$ e $Var(y_{2i}) = n_{2i}p_{2i}(1 - p_{2i})$.

Sejam \hat{p}_{1i} e \hat{p}_{2i} os estimadores das probabilidades de morte para cada grupo e estudo (hospital). As frequências relativas \hat{p}_{1i} e \hat{p}_{2i} são por sua vez,

$$\hat{p}_{1i} = \frac{y_{1i}}{n_{1i}} \text{ e } \hat{p}_{2i} = \frac{y_{2i}}{n_{2i}}, \quad (6.1)$$

na qual $i = 1, \dots, 6$ (número de hospital).

Assim os valores esperados e as variâncias das variáveis aleatórias \hat{p}_{1i} e \hat{p}_{2i} são dadas por,

$$\begin{aligned} E(\hat{p}_{1i}) &= E\left(\frac{y_{1i}}{n_{1i}}\right) = \frac{1}{n_{1i}}n_{1i}p_{1i} = p_{1i} \\ E(\hat{p}_{2i}) &= E\left(\frac{y_{2i}}{n_{2i}}\right) = \frac{1}{n_{2i}}n_{2i}p_{2i} = p_{2i} \\ Var(\hat{p}_{1i}) &= Var\left(\frac{y_{1i}}{n_{1i}}\right) = \frac{1}{n_{1i}^2}n_{1i}p_{1i}(1 - p_{1i}) = \frac{p_{1i}(1 - p_{1i})}{n_{1i}} \\ Var(\hat{p}_{2i}) &= Var\left(\frac{y_{2i}}{n_{2i}}\right) = \frac{1}{n_{2i}^2}n_{2i}p_{2i}(1 - p_{2i}) = \frac{p_{2i}(1 - p_{2i})}{n_{2i}}. \end{aligned} \quad (6.2)$$

Como, n_{1i} e n_{2i} são valores suficientemente grandes a distribuição Binomial pode ser aproximada pela distribuição normal. Observe que $y_n = X_1 + X_2 + \dots + X_n$, onde cada X_i tem uma distribuição de Bernoulli, com média $\mu = p$ e variância $\sigma^2 = p(1 - p)$.

Observar que $y_n = n\bar{X}$, onde pelo TLC, \bar{X} terá distribuição aproximadamente normal, com média p e variância $\sigma^2 = \frac{p(1-p)}{n}$. Logo, a transformação $y_n = n\bar{X}$ terá a distribuição $y_n \sim N(np, np(1-p))$ (ver *Bussab e Moretin, 2003*). Observe que \bar{X} é a própria variável \hat{p} . Assim, a distribuição aproximadamente normal para \hat{p} é dada por,

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right). \quad (6.3)$$

Dessa forma, as distribuições para os estimadores de máxima verossimilhança \hat{p}_{1i} e \hat{p}_{2i} são dadas por,

$$\begin{aligned} \hat{p}_{1i} &\stackrel{a}{\sim} N\left\{p_{1i}, \frac{p_{1i}(1-p_{1i})}{n_{1i}}\right\} \\ \hat{p}_{2i} &\stackrel{a}{\sim} N\left\{p_{2i}, \frac{p_{2i}(1-p_{2i})}{n_{2i}}\right\}, \end{aligned} \quad (6.4)$$

na qual $i = 1, \dots, 6$.

Assumimos que \hat{p}_{1i} e \hat{p}_{2i} são independentes e $\hat{p}_i = \hat{p}_{2i} - \hat{p}_{1i}$ (diferenças nas taxas de mortalidade entre os dois tratamentos), $i = 1, \dots, 6$.

Com os dados da Tabela 6.1 e de acordo com a expressão (6.1), obtemos os resultados para y_{1i} , \hat{p}_{1i} , y_{2i} , \hat{p}_{2i} e \hat{p}_i , apresentados na Tabela 6.2.

Tabela 6.2: Dados de y_{1i} , \hat{p}_{1i} , y_{2i} , \hat{p}_{2i} e \hat{p}_i .

i	y_{1i}	\hat{p}_{1i}	y_{2i}	\hat{p}_{2i}	\hat{p}_i
1	49,02	0,0797	67,02	0,1074	0,0277
2	43,96	0,0580	63,99	0,0830	0,0250
3	27,01	0,0852	32,01	0,1036	0,0184
4	102,00	0,1226	125,97	0,1482	0,0256
5	84,97	0,1049	52,01	0,1281	0,0232
6	245,97	0,1085	218,93	0,0970	-0,0115
Geral	553,18	0,0988	559,78	0,1073	0,0085

As variáveis aleatórias $\hat{p}_{2i} - \hat{p}_{1i}$ são independentes e têm distribuição normal, com média $p_{2i} - p_{1i}$ e variância dada por $\frac{p_{1i}(1-p_{1i})}{n_{1i}} + \frac{p_{2i}(1-p_{2i})}{n_{2i}}$, isto é,

$$\hat{p}_i = \hat{p}_{2i} - \hat{p}_{1i} \stackrel{a}{\sim} N \left\{ p_{2i} - p_{1i}; \frac{p_{1i}(1-p_{1i})}{n_{1i}} + \frac{p_{2i}(1-p_{2i})}{n_{2i}} \right\}. \quad (6.5)$$

Logo, a variável padronizada é dada por,

$$Z_i = \frac{(\hat{p}_{2i} - \hat{p}_{1i}) - (p_{2i} - p_{1i})}{\sqrt{\frac{\hat{p}_{1i}(1-\hat{p}_{1i})}{n_{1i}} + \frac{\hat{p}_{2i}(1-\hat{p}_{2i})}{n_{2i}}}} \stackrel{a}{\sim} N(0, 1), \quad (6.6)$$

na qual Z_i é a diferença das taxas de mortalidade dividido pelo erro padrão da diferença entre duas proporções.

Seja V_i a variância de \hat{p}_i e SE_i o erro padrão da diferença de mortalidade para o estudo i , $i = 1, \dots, 6$.

Assim,

$$SE_i = \sqrt{V_i} = \sqrt{\frac{\hat{p}_{1i}(1-\hat{p}_{1i})}{n_{1i}} + \frac{\hat{p}_{2i}(1-\hat{p}_{2i})}{n_{2i}}}. \quad (6.7)$$

Como ilustração, vamos calcular o valor do erro padrão para $i = 2$ (CDPA), usando os dados das Tabelas 6.1 e 6.2,

$$\begin{aligned} SE_2 &= \sqrt{\frac{0,058(1-0,058)}{758} + \frac{0,083(1-0,083)}{771}} \\ &= 0,013065 \end{aligned}$$

Portanto, $SE_2 = 0,013065$, que equivale à 1,31%.

Teste de hipótese

Denotamos por H_{0i} a hipótese a ser testada, e por H_{1i} a hipótese alternativa. O teste irá levar à aceitação ou rejeição da hipótese H_{0i} , o que corresponde, portanto, respectivamente, à negação ou afirmação de H_{1i} .

Para a tomada de decisão, as hipóteses a serem testadas para cada estudo são definidas por,

$$\begin{cases} H_{0i} : p_i = 0 \\ H_{1i} : p_i > 0 \end{cases},$$

sob a hipótese nula de que não há diferença na mortalidade entre aspirina e placebo.

A estatística de teste sob H_{0i} é dada por,

$$Z_i = \frac{\hat{p}_{2i} - \hat{p}_{1i}}{\sqrt{\frac{\hat{p}_{1i}(1-\hat{p}_{1i})}{n_{1i}} + \frac{\hat{p}_{2i}(1-\hat{p}_{2i})}{n_{2i}}} \sim N(0, 1). \quad (6.8)$$

Assumindo um nível de significância igual à 5%, a hipótese H_{0i} definida acima será rejeitada quando $|Z_i| \geq 1.96$ para cada proporção p_i , $i = 1, \dots, 6$.

Interpretação

A Tabela 6.3 mostra o resultado dos testes de hipóteses dado pela expressão (6.8), considerando um nível $\alpha = 5\%$ de significância.

Tabela 6.3: Resultado da análise para os 6 hospitais.

i	$p_i\%$	$SE_i(\%)$	Z_i	$p - value$
1	2,77	1,65	1,68	0,047*
2	2,50	1,31	1,91	0,028*
3	1,84	2,34	0,79	0,215
4	2,56	1,67	1,53	0,063
5	2,32	1,98	1,17	0,121
6	-1,15	0,90	-1,27	0,898
Geral	0,85	0,59	1,45	0,074

Em termos de diferenças significativas, somente o CDPA (estudo 2) apresenta um nível fortemente significativo ($p - value = 0.028$), e os estudos (UK-1 (estudo 1) e UK-2 (estudo 4)), revelam alguma significância (0.047 e 0.063 respectivamente). O resultado geral apresenta uma significância de 0.074, sugerindo que a aspirina propicia uma certa diminuição no índice de mortalidade.

6.2 Análise Bayesiana

Vamos fazer o uso da Meta-análise para combinar as informações dos 6 resultados usando métodos Bayesianos, pois este captura características individuais de cada estudo.

A natureza diversa dos diferentes estudos, tanto em termos de delineamento como em relação aos métodos empregados na realização de cada um deles é a principal dificuldade para combinar os resultados. Então, para solucionar esta dificuldade incorporamos na análise a variação entre os estudos.

Existem duas formas de combinar as informações em Meta-análise, de acordo com a modelagem da variação entre as diferentes estimativas, a qual pode ter variáveis fixas ou aleatórias e a diferença entre as duas é a forma com que a variabilidade dos resultados é tratada (ver *Egger et al.*, 1997). A modelagem da variação entre os diferentes estudos é uma etapa importante na condução de uma Meta-análise. Nos modelos de efeito fixo assume-se que o tamanho populacional é um valor fixo, enquanto que nos modelos de efeito aleatório assume-se uma variável aleatoriamente distribuída com média e variâncias próprias.

6.2.1 Modelo com efeitos aleatórios

Na presença de um efeito aleatório, assumimos o modelo hierárquico,

$$y_i = \theta_i + \varepsilon_i, \quad (6.9)$$

na qual $\theta_i \stackrel{iid}{\sim} N\{\theta, \sigma^2\}$, $y_i = \hat{p}_i$ é a diferença entre as taxas (proporções de mortes sob o tratamento com aspirina e com placebo) de mortalidade e ε_i é o erro estocástico com uma certa distribuição, freqüentemente assumida como normal com média 0 e variância V_i , de tal forma que os ε_i sejam independentes, assumindo V_i conhecido, $i = 1, \dots, 6$.

Dada as diferenças nos estudos e protocolos de tratamentos, cada estudo tem um efeito aleatório θ_i , e as diferenças y_i são observações aleatórias normais com $y_i \sim N\{\theta_i, \sigma^2 + V_i\}$ e a normalidade para y_i é justificada pelo grande tamanho amostral dentro de cada estudo (ver *Draper*, 2000).

Distribuição a priori

Assumindo uma análise Bayesiana hierárquica e as seguintes distribuições para os parâmetros θ e σ^2 :

$$\theta \sim N\{a, b^2\}; \quad a \text{ e } b \text{ conhecidos;}$$

$\sigma^2 \sim IG [c, d]$; c e d conhecidos.

Assumindo independência *a priori* entre os parâmetros, $\theta_1, \dots, \theta_n, \theta, \sigma^2$, a distribuição *a priori* conjunta é dada por,

$$\pi(\theta_1, \dots, \theta_n, \theta, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\theta_i - \theta)^2 \right\} \times \exp \left\{ -\frac{d}{\sigma^2} - \frac{1}{2b^2} (\theta - a)^2 \right\} \times (\sigma^2)^{-(c+1)}, \quad (6.10)$$

na qual $i = 1, \dots, n$ ($n = 6$).

Função de verossimilhança

A função de verossimilhança para os θ_i , $i = 1, \dots, 6$; é dada por,

$$\begin{aligned} L(\theta_1, \dots, \theta_n) &= \prod_{i=1}^n f(\epsilon_i | \theta_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi V_i}} \exp \left\{ -\frac{\epsilon_i^2}{2V_i} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta_i)^2}{V_i} \right\}, \end{aligned} \quad (6.11)$$

na qual $\epsilon_i = y_i - \theta_i$, ($n = 6$).

Distribuição a *posteriori* conjunta

A distribuição a *posteriori* conjunta é dada por,

$$\begin{aligned} \pi(\theta_1, \dots, \theta_n, \theta, \sigma^2 | \mathbf{y}) &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\theta_i - \theta)^2 \right\} \times \exp \left\{ -\frac{d}{\sigma^2} - \frac{1}{2b^2} (\theta - a)^2 \right\} \times (\sigma^2)^{-(c+1)} \\ &\times \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta_i)^2}{V_i} \right\}. \end{aligned} \quad (6.12)$$

Distribuições condicionais

A partir da distribuição a *posteriori* conjunta para $\theta_1, \dots, \theta_n, \theta, \sigma^2$, determinamos as distribuições condicionais.

A distribuição condicional para θ dados os outros parâmetros é dada por,

$$\begin{aligned}\pi(\theta|\mathbf{v}(\theta), \mathbf{y}) &\propto \exp\left\{-\frac{1}{2b^2}(\theta - a)^2\right\} \times \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(\theta_i - \theta)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\left(\frac{1}{b^2} + \frac{n}{\sigma^2}\right)^{(-1)}}\left[\theta - \frac{\left(\frac{a\sigma^2}{b^2} + \sum_{i=1}^n\theta_i\right)}{\sigma^2\left(\frac{1}{b^2} + \frac{n}{\sigma^2}\right)}\right]^2\right\}.\end{aligned}$$

Assim,

$$\pi(\mathbf{v}(\theta), \mathbf{y}) \propto \exp\left\{-\frac{1}{2\left(\frac{b^2\sigma^2}{\sigma^2+nb^2}\right)}\left[\theta - \frac{a\sigma^2 + b^2\sum_{i=1}^n\theta_i}{\sigma^2 + nb^2}\right]^2\right\}.$$

Isto é,

$$\theta|\mathbf{v}(\theta), \mathbf{y} \sim N\left\{\frac{a\sigma^2 + b^2\sum_{i=1}^n\theta_i}{\sigma^2 + nb^2}; \frac{b^2\sigma^2}{\sigma^2 + nb^2}\right\}, \quad (6.13)$$

na qual $\mathbf{v}(\theta)$ é o vetor de todos os parâmetros, exceto θ .

A distribuição condicional para σ^2 dados os outros parâmetros é dada por,

$$\begin{aligned}\pi(\sigma^2|\mathbf{v}(\sigma^2), \mathbf{y}) &\propto (\sigma^2)^{-(c+\frac{n}{2}+1)} e^{-\frac{d}{\sigma^2}} \times \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(\theta_i - \theta)^2\right\} \\ &\propto (\sigma^2)^{-(c+\frac{n}{2}+1)} \exp\left\{-\frac{1}{\sigma^2}\left[d + \frac{1}{2}\sum_{i=1}^n(\theta_i - \theta)^2\right]\right\}.\end{aligned}$$

Isto é,

$$\sigma^2|\mathbf{v}(\sigma^2), \mathbf{y} \sim IG\left[c + \frac{n}{2}; d + \frac{1}{2}\sum_{i=1}^n(\theta_i - \theta)^2\right], \quad (6.14)$$

na qual $\mathbf{v}(\sigma^2)$ é o vetor de todos os parâmetros, exceto σ^2 .

A distribuição condicional para θ_i dados os outros parâmetros é dada por,

$$\begin{aligned}\pi(\theta_i|\mathbf{v}(\theta_i), \mathbf{y}) &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n(\theta_i - \theta)^2\right\} \exp\left\{-\frac{1}{2}\sum_{i=1}^n\frac{(y_i - \theta_i)^2}{V_i}\right\} \\ &\propto \exp\left\{-\frac{1}{2\left(\frac{1}{\sigma^2} + \frac{1}{V_i}\right)^{(-1)}}\left[\theta_i - \frac{\left(\theta + \frac{\sigma^2 y_i}{V_i}\right)}{\sigma^2\left(\frac{1}{\sigma^2} + \frac{1}{V_i}\right)}\right]^2\right\}.\end{aligned}$$

Assim,

$$\pi(\theta_i | \mathbf{v}_{(\theta_i)}, \mathbf{y}) \propto \exp \left\{ -\frac{1}{2 \left(\frac{\sigma^2 V_i}{V_i + \sigma^2} \right)} \left[\theta_i - \frac{\theta V_i + \sigma^2 y_i}{V_i + \sigma^2} \right]^2 \right\}.$$

Isto é,

$$\theta_i | \mathbf{v}_{(\theta_i)}, \mathbf{y} \sim N \left\{ \frac{\theta V_i + \sigma^2 y_i}{V_i + \sigma^2}; \frac{\sigma^2 V_i}{V_i + \sigma^2} \right\}, \quad (6.15)$$

na qual $\mathbf{v}_{(\theta_i)}$ é o vetor de todos os parâmetros, exceto θ_i ,. $i = 1, \dots, n$ ($n = 6$).

Interpretação

Considerando os dados das Tabelas 6.1 e 6.2 e o modelo (6.9), obtemos os sumários a *posteriori* para os parâmetros de interesse. Assumindo os valores para os hiperparâmetros do modelo dados por, $a = c = d = 1$ e $b = 0.01$, geramos duas cadeias de Gibbs cada uma com 80000 iterações; para cada série, consideramos saltos de tamanho dez, e finalmente obtemos uma amostra de tamanho 11000. As 30000 primeiras observações foram eliminadas ("burn in"). A convergência das amostras de Gibbs foram monitoradas usando o critério de *Gelman e Rubin* (1992), ($\sqrt{\hat{R}} < 1.1$ para todos os parâmetros).

A Tabela 6.4 mostra os sumários a *posteriori* para os parâmetros.

Tabela 6.4: Sumários a *posteriori*.

Estimadores	Valores estimados	Intervalos de 95% de credibilidade	$\sqrt{\hat{R}}$
θ_1	2.543	(1.521, 3.602)	0.99
θ_2	2.314	(1.11, 3.552)	1.00
θ_3	1.882	(1.095, 2.662)	0.99
θ_4	2.393	(1.392, 3.438)	0.99
θ_5	2.245	(1.358, 3.144)	1.00
θ_6	0.679	(-1.389, 2.324)	0.99
θ	2.003	(0.877, 2.996)	1.00
τ	1.452	(0.216, 4.327)	1.00

Na Tabela 6.4, observamos que os parâmetros $\theta_1, \dots, \theta_5$ e τ , na qual τ é definido pelo inverso de σ^2 indicam efeitos significativos, ou seja, o uso de aspirina diminuiu o índice de

mortalidade. Ao contrário dos outros parâmetros, θ_6 não apresenta indicações de efeito significativo (o intervalo de 95% de credibilidade inclui o zero). Isto também pode ser observado na Tabela 6.1, onde somente no estudo 6 (AMIS) não há diminuição da taxa de mortalidade para pacientes recebendo aspirina.

Também temos na Tabela 6.4 os índices de *Gelman e Rubin*; por esses índices observamos a convergência do algoritmo Gibbs Sampling.

6.2.2 Análise gráfica.

Gráficos das distribuições *a posteriori* para os parâmetros de interesse estão expostos na Figura 6.1, os quais mostram que as mesmas se comportam de acordo com as distribuições *a priori* especificadas e na Figura 6.2 temos os gráficos onde os parâmetros apresentam-se não correlacionados.

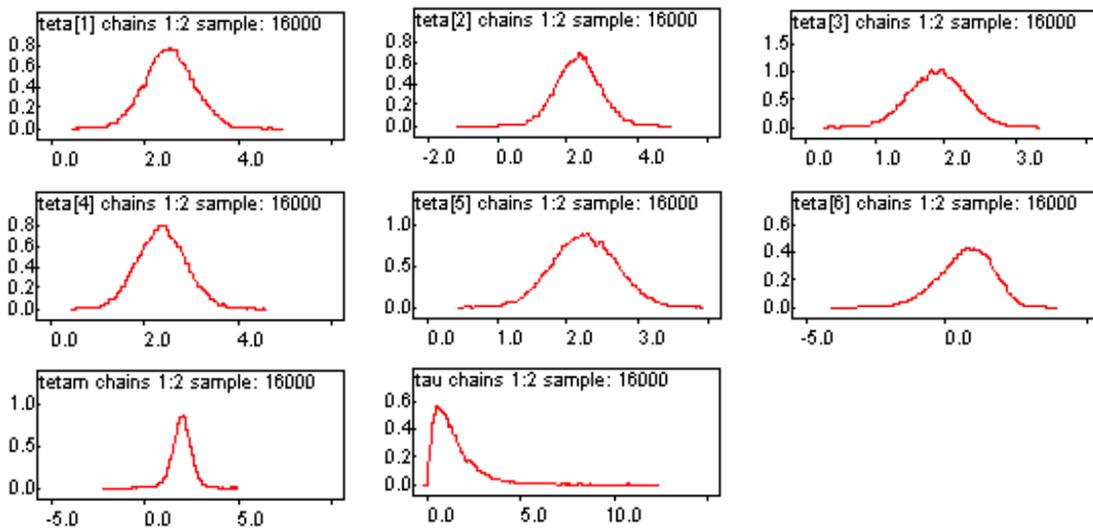


Figura 6.1: Densidades *a posteriori* marginais.

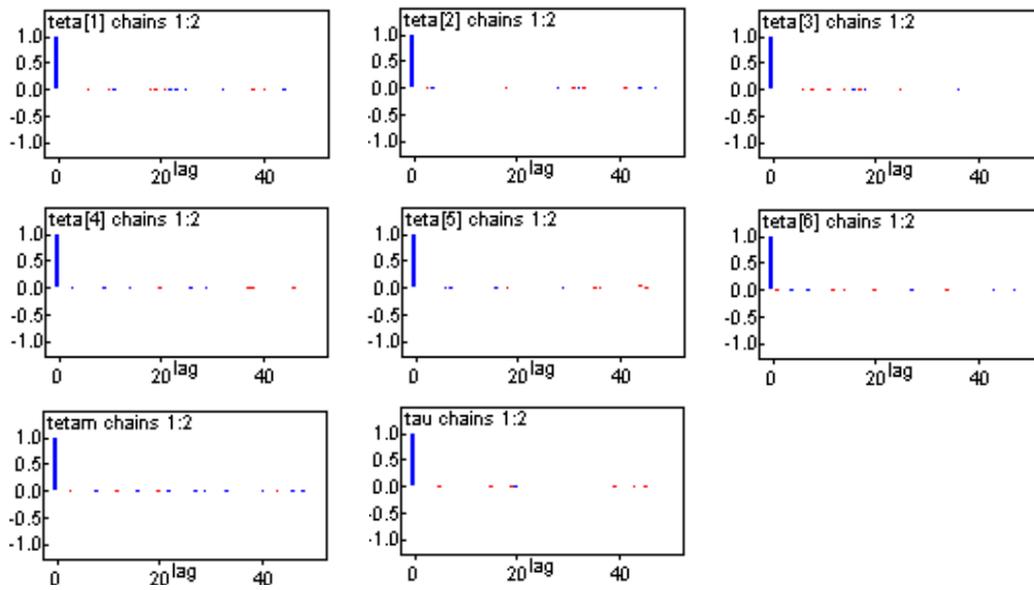


Figura 6.2: Gráficos de autocorrelações.

A Tabela 6.4 mostra os índices de *Gelman e Rubin* que estão de acordo com o valor recomendado que é aproximadamente 1, e através da Figura 6.3 observa-se que à medida que aumenta o número de iterações, temos um indicativo de convergência para todos os parâmetros do modelo. A Figura 6.4, mostra uma indicação de convergência para todos os parâmetros através do gráfico de séries temporais.

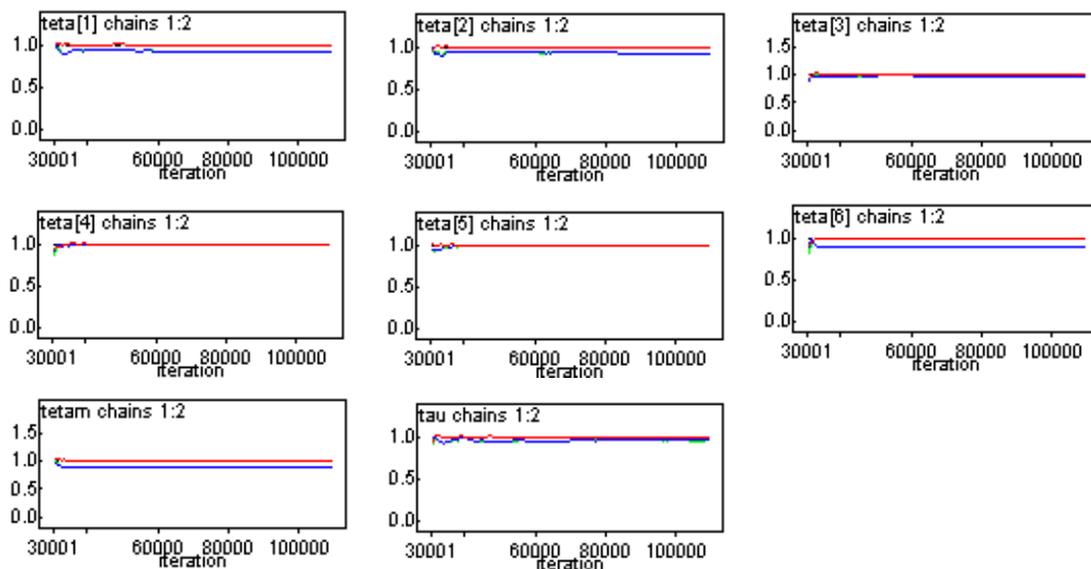


Figura 6.3: Gráficos dos índices R de *Gelman e Rubin*.

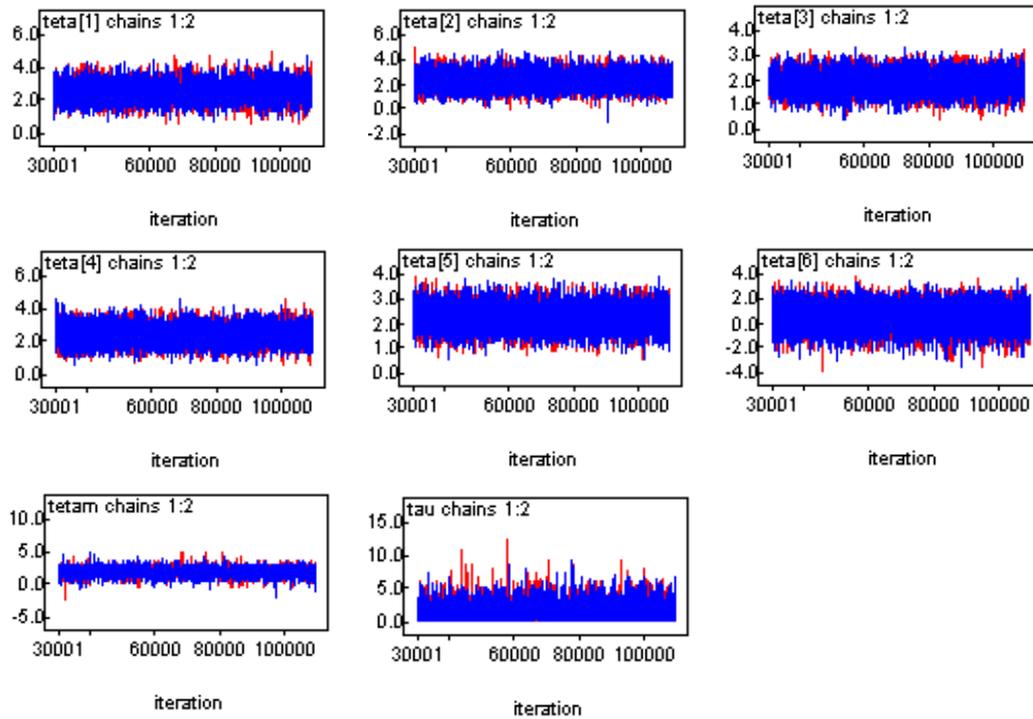


Figura 6.4: Gráficos das séries temporais.

6.3 Modelo clássico de efeitos aleatórios.

Considerando o modelo (6.9), na qual θ_i , ε_i são variáveis aleatórias e θ_i são efeitos aleatórios e tais que $\theta_i \sim N\{\theta; \sigma^2\}$ e $\varepsilon_i \sim N\{0; V_i\}$, com V_i conhecidos, $i = 1, \dots, 6$ e θ_i e ε_i são independentes.

A partir do modelo (6.9), a média e a variância de y_i são dadas por,

$$\begin{aligned} E(y_i) &= E(\theta_i + \varepsilon_i) = E(\theta_i) + E(\varepsilon_i) = \theta + 0 = \theta \\ \text{var}(y_i) &= \text{var}(\theta_i + \varepsilon_i) = \text{var}(\theta_i) + \text{var}(\varepsilon_i) = \sigma^2 + V_i. \end{aligned}$$

Dessa forma, obtemos a distribuição de y_i dada por,

$$y_i \sim N\{\theta; \sigma^2 + V_i\}, \quad (6.16)$$

na qual $i = 1, \dots, n$ ($n = 6$).

Muitas vezes queremos considerar conjuntamente o estimador e a precisão com que

se estima o parâmetro. A forma usual de se fazer isto é através dos chamados *intervalos de confiança*, obtidos através dos estimadores de máxima verossimilhança. Para isso calculamos a função de verossimilhança e conseqüentemente seus estimadores.

Função de verossimilhança

A função de verossimilhança para θ e σ^2 é dada por,

$$\begin{aligned} L(\theta, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi(\sigma^2 + V_i)}} \exp \left\{ -\frac{1}{2(\sigma^2 + V_i)} (y_i - \theta)^2 \right\} \\ &\propto \prod_{i=1}^n (\sigma^2 + V_i)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta)^2}{(\sigma^2 + V_i)} \right\}. \end{aligned} \quad (6.17)$$

O log-verossimilhança é dado por,

$$l(\theta, \sigma^2) \propto -\frac{1}{2} \sum_{i=1}^n \log(\sigma^2 + V_i) - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta)^2}{(\sigma^2 + V_i)}. \quad (6.18)$$

As derivadas primeiras são definidas por,

$$\begin{aligned} \frac{\partial l}{\partial \theta} &= \sum_{i=1}^n \frac{(y_i - \theta)}{(\sigma^2 + V_i)} = 0 \\ \frac{\partial l}{\partial \sigma^2} &= -\frac{1}{2} \sum_{i=1}^n \frac{1}{\sigma^2 + V_i} + \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta)^2}{(\sigma^2 + V_i)^2} = 0. \end{aligned} \quad (6.19)$$

Os valores de θ e σ^2 foram estimados pelo método da máxima verossimilhança dados por *Draper (2000)*,

$$\sum_{i=1}^n \frac{(y_i - \theta)}{(\sigma^2 + V_i)} = 0, \text{ isto é } \sum_{i=1}^n \hat{W}_i (y_i - \hat{\theta}) = 0.$$

Então,

$$\hat{\theta} = \frac{\sum_{i=1}^n y_i \hat{W}_i}{\sum_{i=1}^n \hat{W}_i}, \quad (6.20)$$

na qual $\hat{W}_i = \frac{1}{\hat{\sigma}^2 + V_i}$.

Para σ^2 , temos:

$$\sum_{i=1}^n \frac{1}{\hat{\sigma}^2 + V_i} = \sum_{i=1}^n \frac{(y_i - \hat{\theta})^2}{(\hat{\sigma}^2 + V_i)^2},$$

isto é,

$$\sum_{i=1}^n \left[\frac{1}{\hat{\sigma}^2 + V_i} - \frac{1}{(\hat{\sigma}^2 + V_i)^2} + \frac{1}{(\hat{\sigma}^2 + V_i)^2} \right] = \sum_{i=1}^n \frac{(y_i - \hat{\theta})^2}{(\hat{\sigma}^2 + V_i)^2}$$

ou,

$$\sum_{i=1}^n \left[\hat{W}_i^2 (\hat{\sigma}^2 + V_i - 1) + \hat{W}_i^2 \right] = \sum_{i=1}^n \hat{W}_i^2 (y_i - \hat{\theta})^2,$$

o que leva a,

$$\hat{\sigma}^2 \sum_{i=1}^n \hat{W}_i^2 + \sum_{i=1}^n \hat{W}_i^2 V_i = \sum_{i=1}^n \hat{W}_i^2 (y_i - \hat{\theta})^2.$$

Portanto,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{W}_i^2 \left[(y_i - \hat{\theta})^2 - V_i \right]}{\sum_{i=1}^n \hat{W}_i^2}. \quad (6.21)$$

Intervalos de confiança aproximados

Consideramos intervalos de confiança aproximados para os parâmetros θ e σ^2 baseados na distribuição assintótica dos estimadores de máxima verossimilhança $\hat{\theta}$ de θ e $\hat{\sigma}^2$ de σ^2 .

Assim, as distribuições assintóticas para $\hat{\theta}$ e $\hat{\sigma}^2$ são definidas por,

$$\begin{aligned} \hat{\theta} &\stackrel{a}{\sim} N \left\{ \theta; I_o(\hat{\theta})^{-1} \right\} \\ \hat{\sigma}^2 &\stackrel{a}{\sim} N \left\{ \sigma^2; I_o(\hat{\sigma}^2)^{-1} \right\}. \end{aligned} \quad (6.22)$$

Desta forma, temos a informação de *Fisher* (ver *Fisher*, 1932) observada para θ e σ^2 dada por,

$$\begin{aligned} I_o(\hat{\theta}) &= -\frac{\partial^2 l}{\partial \theta^2} = \sum_{i=1}^n \frac{1}{(\sigma^2 + V_i)} = \sum_{i=1}^n \omega_i \\ I_o(\hat{\sigma}^2) &= -\frac{\partial^2 l}{\partial (\sigma^2)^2} = -\frac{1}{2} \sum_{i=1}^n \frac{1}{(\sigma^2 + V_i)^2} + \sum_{i=1}^n \frac{(y_i - \theta)^2}{(\sigma^2 + V_i)^3} = -\frac{1}{2} \sum_{i=1}^n \omega_i^2 + \sum_{i=1}^n \omega_i^3 (y_i - \theta)^2. \end{aligned} \quad (6.23)$$

Logo, substituindo os resultados de (6.23) em (6.22), temos:

$$\hat{\theta} \stackrel{a}{\sim} N \left\{ \theta; \frac{1}{\sum_{i=1}^n \hat{W}_i} \right\} \quad (6.24)$$

$$\hat{\sigma}^2 \stackrel{a}{\sim} N \left\{ \sigma^2; \frac{1}{\left(-\frac{1}{2} \sum_{i=1}^n \hat{W}_i^2 + \sum_{i=1}^n \hat{W}_i^3 (y_i - \theta)^2 \right)} \right\}.$$

Assim, a distribuição normal padronizada para Z_1 é dada por,

$$Z_1 = \frac{\hat{\theta} - \theta}{\sqrt{I_o(\hat{\theta})^{-1}}} = \frac{\hat{\theta} - \theta}{\sqrt{\left(\sum_{i=1}^n \hat{W}_i \right)^{-1}}} \stackrel{a}{\sim} N(0, 1). \quad (6.25)$$

Fixando $\alpha = 5\%$ observamos pela tabela da distribuição normal padronizada Z_1 que,

$$P(-1.96 < Z_1 < 1.96) = 0.95.$$

Substituindo o valor de Z_1 (6.25), obtemos o intervalo de 95% de confiança para $\hat{\theta}$, que é a estimativa do efeito da aspirina no índice de mortalidade. Portanto, o intervalo de confiança para $\hat{\theta}$ é definido por,

$$\hat{\theta} \pm 1.96 \sqrt{\left(\sum_{i=1}^n \hat{W}_i \right)^{-1}}. \quad (6.26)$$

Da mesma forma, temos a distribuição normal padronizada para Z_2 dada por,

$$Z_2 = \frac{\hat{\sigma}^2 - \sigma^2}{\sqrt{I_o(\hat{\sigma}^2)^{-1}}} = \frac{\hat{\sigma}^2 - \sigma^2}{\sqrt{\left(-\frac{1}{2} \sum_{i=1}^n \hat{W}_i^2 + \sum_{i=1}^n \hat{W}_i^3 (y_i - \theta)^2 \right)^{-1}}} \stackrel{a}{\sim} N(0, 1). \quad (6.27)$$

Fixando $\alpha = 5\%$ observamos pela tabela da distribuição normal padronizada Z_2 que,

$$P(-1.96 < Z_2 < 1.96) = 0.95.$$

Substituindo o valor de Z_2 (6.27), obtemos o intervalo de confiança para $\hat{\sigma}^2$ dado por,

$$\hat{\sigma}^2 \pm 1.96 \sqrt{\left(-\frac{1}{2} \sum_{i=1}^n \hat{W}_i^2 + \sum_{i=1}^n \hat{W}_i^3 (y_i - \theta)^2 \right)^{-1}}. \quad (6.28)$$

Interpretação

A Tabela 6.5 mostra os resultados dos estimadores de máxima verossimilhança para θ e σ^2 .

Tabela 6.5: Resultados dos estimadores de máxima verossimilhança.

Estimadores	Valores estimados	Erros padrões	Intervalo de 95% de confiança
θ_1	1.9543	1.0133	(-0.0235; 3.8934)
θ_2	1.9697	0.9169	(0.1767; 3.7302)
θ_3	1.5482	1.1256	(-0.6304; 3.7068)
θ_4	1.8689	1.0179	(-0.1153; 3.8188)
θ_5	1.711	1.0781	(-0.3831; 3.7766)
θ_6	-0.29	0.7369	(-1.7009; 1.1665)
θ	1.4513	0.8144	(-0.1450; 3.0477)
σ^2	1.5013	1.6752	(-1.7821; 4.7847)

Analisando os estimadores $\theta_1, \dots, \theta_6$, observamos que apenas os estimadores θ_1 e θ_2 são significativos para o modelo. Sob o ponto de vista Bayesiano há evidência fortemente a favor da utilidade da aspirina para reduzir a mortalidade, após um ataque cardíaco. Vários artigos mostram que a aspirina ajuda a diminuir o risco de ataques cardíacos e que é, inclusive, mais benéfica do que se pensava.

Capítulo 7

Meta-análise em Educação.

Raudenbush, 1984 estudou dezenove experimentos publicados entre 1966 e 1974, estimando o efeito da expectativa do professor associada ao QI dos estudantes. Estes dezenove experimentos independentes estavam divididos em dois grupos de estudantes.

Temos, grupo 1 (grupo experimental) formado por alunos escolhidos aleatoriamente entre os alunos com grande potencial (QI's elevados) e o grupo 2 (grupo controle) formado por alunos escolhidos aleatoriamente entre os alunos com baixo potencial (QI's baixos).

O tamanho estimado do efeito do QI no desempenho dos estudantes é definido por,

$$y_i = \frac{\bar{T}_i - \bar{C}_i}{SD_{i:p}}, \quad (7.1)$$

na qual \bar{T}_i é a média amostral das notas referente ao grupo 1, \bar{C}_i é a média amostral das notas referente ao grupo 2 e $SD_{i:p}$ é o erro padrão ponderado; para o grupo 1 e para o grupo 2 os erro padrões são definidos respectivamente por,

$$SD_{1i} = \sqrt{\frac{\sum_{j=1}^{n_i} (T_{ij} - \bar{T}_i)^2}{n_i - 1}} \text{ e } SD_{2i} = \sqrt{\frac{\sum_{j=1}^{m_i} (C_{ij} - \bar{C}_i)^2}{m_i - 1}}, \quad (7.2)$$

na qual n_i é o número de estudantes no grupo 1 e m_i é o número de estudantes no grupo 2.

Desta forma, o erro padrão ponderado $(SD_{i:p})^2$ é definido por,

$$(SD_{i:p})^2 = \frac{(n_i - 1)(SD_{1i})^2 + (m_i - 1)(SD_{2i})^2}{n_i + m_i - 2},$$

para $i = 1, \dots, 19$.

O contato prévio do professor com os alunos pode diluir o efeito do QI no desempenho dos alunos, quanto maior for o contato prévio, menor será a diferença entre os dois grupos de estudantes, isto é, menor o efeito do QI. Neste caso temos uma covariável x_i definida pelo tempo ou duração do contato prévio entre o professor e o aluno (dado em semanas).

Podemos observar na Tabela 7.1 os dados dos dezenove experimentos indicados na coluna 1, na coluna 2 temos a covariável x_i indicando o tempo do contato prévio entre o professor e o aluno, variando de 0 à 3 semanas, na coluna 3 o tamanho do efeito (y_i) variando de -0.32 até 1.18 e na coluna 4, temos o erro padrão de y_i .

Tabela 7.1: Resultados de dezenove experimentos estimando o efeito da expectativa do professor associada ao QI dos estudantes.

$Estudos(i)$	x_i	y_i	$\sqrt{V_i}$
1	2	0.03	0.125
2	3	0.12	0.147
3	3	-0.14	0.167
4	0	1.18	0.373
5	0	0.26	0.369
6	3	-0.06	0.103
7	3	-0.02	0.103
8	3	-0.32	0.220
9	0	0.27	0.164
10	1	0.80	0.251
11	0	0.54	0.302
12	0	0.18	0.223
13	1	-0.02	0.289
14	2	0.23	0.290
15	3	-0.18	0.159
16	3	-0.06	0.167
17	1	0.30	0.139
18	2	0.07	0.094
19	3	-0.07	0.174

7.1 Análise Bayesiana

Vamos fazer o uso da Meta-análise para combinar as informações dos dezenove experimentos usando métodos Bayesianos.

Cada estudo tem um efeito aleatório θ_i , e os tamanhos y_i são observações aleatórias normais com distribuição $y_i \sim N\{\theta_i, V_i\}$. Na presença de um efeito aleatório, assumimos

o modelo estatístico,

$$y_i = \theta_i + \varepsilon_i,$$

na qual ε_i é uma variável erro com uma dada distribuição, frequentemente assumida como normal com média 0 e variância V_i , de tal forma que os ε_i sejam independentes, assumindo V_i conhecido (dados na Tabela 7.1), $i = 1, \dots, 19$.

Uma análise gráfica do tamanho do efeito (y_i) e o tempo de contato (x_i) é fundamental para verificar como estes se relacionam.

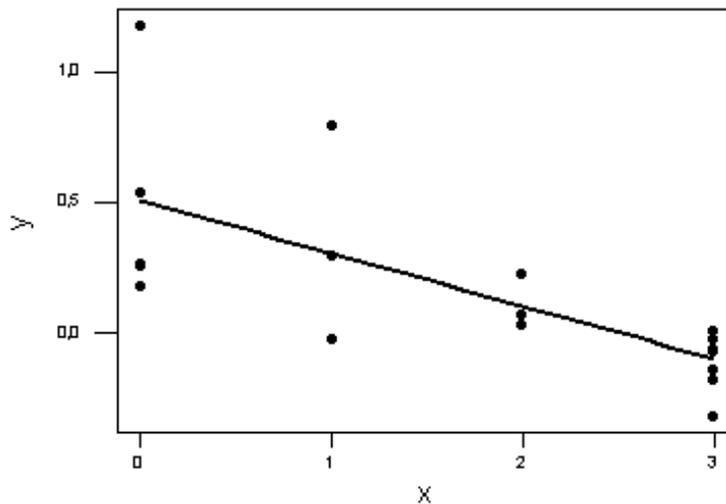


Figura 7.1. Gráfico do tamanho do efeito estimado versus tempo de contato.

Podemos observar na Figura 7.1 que y_i está relacionada com x_i numa forma linear; desta forma, *Bryk e Raudenbush* (ver *Draper*, 2000) propuseram o modelo hierárquico dado por: $\theta_i | \alpha, \beta, \sigma^2 \stackrel{indep}{\sim} N \{ \alpha + \beta x_i; \sigma^2 \}$, na qual θ_i são os efeitos aleatórios.

Para melhorar o desempenho do algoritmo Gibbs Sampling vamos considerar covariáveis centralizadas. Desta forma, o modelo de regressão passa a ser definido por,

$$y_i = \alpha + \beta (x_i - \bar{x}) + \varepsilon_i,$$

na qual $y_i \sim N \{ \alpha + \beta (x_i - \bar{x}), V_i \}$.

Distribuição a *priori*

Assumir uma análise Bayesiana hierárquica e as seguintes distribuições a *priori* para os parâmetros α , β e σ^2 .

$$\alpha \sim N\{a; b^2\}; a, b \text{ conhecidos};$$

$$\beta \sim N\{c; d^2\}; c, d \text{ conhecidos};$$

$$\sigma^2 \sim IG[e; f]; e, f \text{ conhecidos}.$$

Assumindo independência a *priori* entre os parâmetros, a distribuição a *priori* conjunta é dada por,

$$\begin{aligned} \pi(\alpha, \beta, \sigma^2 | \mathbf{y}, \mathbf{x}) &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\theta_i - \alpha - \beta(x_i - \bar{x}))^2 \right\} (\sigma^2)^{-(e+\frac{n}{2}+1)} \\ &\exp \left\{ -\frac{f}{\sigma^2} - \frac{1}{2d^2}(\beta - c)^2 - \frac{1}{2b^2}(\alpha - a)^2 \right\}. \end{aligned}$$

Função de verossimilhança

A função de verossimilhança para os parâmetros θ_i , $i = 1, \dots, 19$ é dada por,

$$\begin{aligned} L(\theta_1, \dots, \theta_n) &= \prod_{i=1}^n f(\epsilon_i | \theta_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi V_i}} \exp \left\{ -\frac{\epsilon_i^2}{2V_i} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta_i)^2}{V_i} \right\}, \end{aligned} \quad (7.3)$$

na qual $\epsilon_i = y_i - \theta_i$, $i = 1, \dots, n$ ($n = 19$).

Distribuição a *posteriori* conjunta

A distribuição a *posteriori* conjunta para $\theta_1, \dots, \theta_n$, α , β , σ^2 é dada por,

$$\begin{aligned} \pi(\theta_1, \dots, \theta_n, \alpha, \beta, \sigma^2 | \mathbf{y}, \mathbf{x}) &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\theta_i - \alpha - \beta(x_i - \bar{x}))^2 \right\} \times \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta_i)^2}{V_i} \right\} \\ &(\sigma^2)^{-(e+\frac{n}{2}+1)} \times \exp \left\{ -\frac{f}{\sigma^2} - \frac{1}{2d^2}(\beta - c)^2 - \frac{1}{2b^2}(\alpha - a)^2 \right\}. \end{aligned} \quad (7.4)$$

Distribuições condicionais

A partir da distribuição a *posteriori* conjunta para $\theta_1, \dots, \theta_n, \alpha, \beta, \sigma^2$, determinamos as distribuições condicionais necessárias para o algoritmo Gibbs Sampling.

A distribuição condicional para α dados os outros parâmetros é dada por,

$$\begin{aligned} \pi(\alpha | \mathbf{v}_{(\alpha)}, \mathbf{y}, \mathbf{x}) &\propto \exp \left\{ -\frac{1}{2b^2} (\alpha - a)^2 \right\} \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\theta_i - \alpha - \beta((x_i - \bar{x})))^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2 \left(\frac{1}{b^2} + \frac{n}{\sigma^2} \right)^{(-1)}} \left[\alpha - \frac{\left(\frac{a}{b^2} + \frac{\sum_{i=1}^n \mu_i^{(1)}}{\sigma^2} \right)}{\left(\frac{1}{b^2} + \frac{n}{\sigma^2} \right)} \right]^2 \right\}. \end{aligned}$$

Assim,

$$\pi(\alpha | \mathbf{v}_{(\alpha)}, \mathbf{y}, \mathbf{x}) \propto \exp \left\{ -\frac{1}{2 \left(\frac{b^2 \sigma^2}{\sigma^2 + nb^2} \right)} \left[\alpha - \frac{a\sigma^2 + b^2 \sum_{i=1}^n \mu_i^{(1)}}{\sigma^2 + nb^2} \right]^2 \right\}$$

Isto é,

$$\alpha | \mathbf{v}_{(\alpha)}, \mathbf{y}, \mathbf{x} \sim N \left\{ \frac{a\sigma^2 + b^2 \sum_{i=1}^n \mu_i^{(1)}}{\sigma^2 + nb^2}; \frac{b^2 \sigma^2}{\sigma^2 + nb^2} \right\}, \quad (7.5)$$

na qual $\mu_i^{(1)} = \theta_i - \beta(x_i - \bar{x})$ e $\mathbf{v}_{(\alpha)}$ é o vetor de todos os parâmetros, exceto α .

A distribuição condicional para β dados os outros parâmetros é dada por,

$$\begin{aligned} \pi(\beta | \mathbf{v}_{(\beta)}, \mathbf{y}, \mathbf{x}) &\propto \exp \left\{ -\frac{1}{2d^2} (\beta - c)^2 \right\} \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\theta_i - \alpha - \beta((x_i - \bar{x})))^2 \right\} \\ &\propto \exp \left\{ -\frac{1}{2 \left(\frac{1}{d^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sigma^2} \right)^{(-1)}} \left[\beta - \frac{\left(\frac{c}{d^2} + \frac{\sum_{i=1}^n (x_i - \bar{x}) \mu_i^{(2)}}{\sigma^2} \right)}{\left(\frac{1}{d^2} + \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sigma^2} \right)} \right]^2 \right\}. \end{aligned}$$

Assim,

$$\pi(\beta|\mathbf{v}_{(\beta)}, \mathbf{y}, \mathbf{x}) \propto \exp \left\{ -\frac{1}{2 \left(\frac{d^2 \sigma^2}{\sigma^2 + d^2 \sum_{i=1}^n (x_i - \bar{x})} \right)} \left[\beta - \frac{c\sigma^2 + d^2 \sum_{i=1}^n (x_i - \bar{x}) \mu_i^{(2)}}{\sigma^2 + d^2 \sum_{i=1}^n (x_i - \bar{x})} \right]^2 \right\}.$$

Isto é,

$$\beta|\mathbf{v}_{(\beta)}, \mathbf{y}, \mathbf{x} \sim N \left\{ \frac{c\sigma^2 + d^2 \sum_{i=1}^n (x_i - \bar{x}) \mu_i^{(2)}}{\sigma^2 + d^2 \sum_{i=1}^n (x_i - \bar{x})}; \frac{d^2 \sigma^2}{\sigma^2 + d^2 \sum_{i=1}^n (x_i - \bar{x})} \right\}, \quad (7.6)$$

na qual $\mu_i^{(2)} = \theta_i - \alpha$ e $\mathbf{v}_{(\beta)}$ é o vetor de todos os parâmetros, exceto β .

A distribuição condicional para σ^2 dados os outros parâmetros é dada por,

$$\begin{aligned} \pi(\sigma^2|\mathbf{v}_{(\sigma^2)}, \mathbf{y}, \mathbf{x}) &\propto (\sigma^2)^{-(e+\frac{n}{2}+1)} e^{-\frac{f}{\sigma^2}} \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\theta_i - \alpha - \beta((x_i - \bar{x})))^2 \right\} \\ &\propto (\sigma^2)^{-(e+\frac{n}{2}+1)} \exp \left\{ -\frac{1}{\sigma^2} \left[f + \frac{1}{2} \sum_{i=1}^n (\theta_i - \alpha - \beta((x_i - \bar{x})))^2 \right] \right\}. \end{aligned}$$

Isto é,

$$\sigma^2|\mathbf{v}_{(\sigma^2)}, \mathbf{y}, \mathbf{x} \sim IG \left[e + \frac{n}{2}; f + \frac{1}{2} \sum_{i=1}^n (\theta_i - \alpha - \beta((x_i - \bar{x})))^2 \right], \quad (7.7)$$

na qual $\mathbf{v}_{(\sigma^2)}$ é o vetor de todos os parâmetros, exceto σ^2 .

A distribuição condicional para θ_i dados os outros parâmetros é dada por,

$$\begin{aligned} \pi(\theta_i|\mathbf{v}_{(\theta_i)}, \mathbf{y}, \mathbf{x}) &\propto \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (\theta_i - \alpha - \beta((x_i - \bar{x})))^2 \right\} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \frac{(y_i - \theta_i)^2}{V_i} \right\} \\ &\propto \exp \left\{ -\frac{1}{2 \left(\frac{1}{\sigma^2} + \frac{1}{V_i} \right)^{(-1)}} \left[\theta_i - \frac{\sum_{i=1}^n \left(\frac{\mu_i^{(3)}}{\sigma^2} + \frac{y_i}{V_i} \right)}{\frac{1}{\sigma^2} + \frac{1}{V_i}} \right]^2 \right\}. \end{aligned}$$

Assim,

$$\pi(\theta_i | \mathbf{v}(\theta_i), \mathbf{y}, \mathbf{x}) \propto \exp \left\{ -\frac{1}{2 \left(\frac{\sigma^2 V_i}{V_i + \sigma^2} \right)} \left[\theta_i - \sum_{i=1}^n \left(\frac{\mu_i^{(3)} V_i + \sigma^2 y_i}{V_i + \sigma^2} \right) \right]^2 \right\}.$$

Isto é,

$$\theta_i | \mathbf{v}(\theta_i), \mathbf{y}, \mathbf{x} \sim N \left\{ \sum_{i=1}^n \frac{\mu_i^{(3)} V_i + \sigma^2 y_i}{V_i + \sigma^2}; \frac{\sigma^2 V_i}{V_i + \sigma^2} \right\}, \quad (7.8)$$

na qual $\mu_i^{(3)} = \alpha + \beta(x_i - \bar{x})$ e $\mathbf{v}(\theta_i)$ é o vetor de todos os parâmetros, exceto θ_i , $i = 1, \dots, n$ ($n = 19$).

Interpretação

Considerando os dados da Tabela 7.1 e as distribuições *a priori* para os parâmetros do modelo dados em (7.1.1), com $a = 0$, $c = 0$, $b = 10$, $d = 10$, $e = 0.1$ e $f = 0.001$ e assumindo independência *a priori* não informativa para os parâmetros, geramos duas cadeias de Gibbs Sampling cada uma com 10000 iterações. Para cada série consideramos saltos de tamanho dez, e finalmente obtemos uma amostra de tamanho 12000. As 40000 primeiras observações foram eliminadas ("burn in"). A convergência das amostras de Gibbs Sampling foram monitoradas usando o Método de *Gelman e Rubin* (1992) ($\sqrt{\hat{R}} < 1.1$ para todos os parâmetros). A Tabela 7.2 mostra os sumários *a posteriori* e os valores $\sqrt{\hat{R}}$ (critério de *Gelman e Rubin*, 1992).

Tabela 7.2: Sumários a *posteriori*.

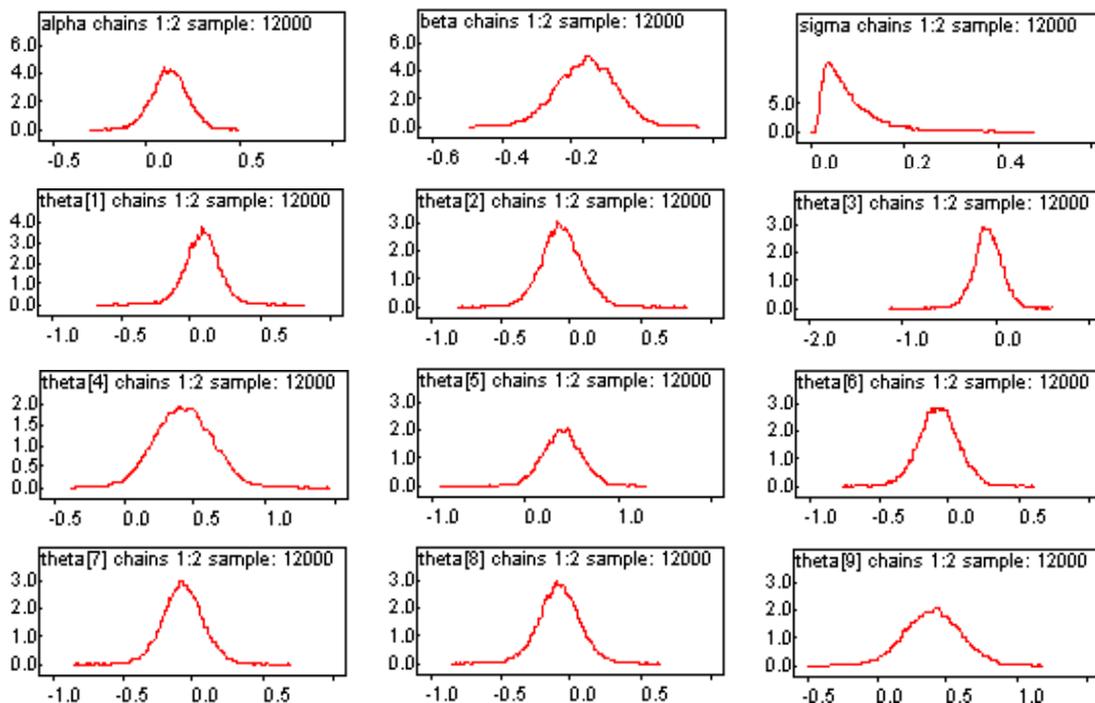
Parâmetros	Média a posteriori	Intervalos de 95% de credibilidade	$\sqrt{\hat{R}}$
α	0.1299	(-0.0558; 0.3141)	1.00
β	-0.16	(-0.3249; -0.00029)	1.00
σ^2	0.0816	(0.022; 0.2278)	0.99
θ_1	0.0834	(-0.1663; 0.3265)	1.00
θ_2	-0.0630	(-0.3465; 0.2355)	0.99
θ_3	-0.0771	(-0.3718; 0.2169)	0.99
θ_4	0.4256	(0.029; 0.836)	0.99
θ_5	0.405	(0.00095; 0.8138)	1.00
θ_6	-0.0726	(-0.3542; 0.2123)	1.00
θ_7	-0.0692	(-0.3533; 0.2202)	0.99
θ_8	-0.0823	(-0.3874; 0.2145)	1.00
θ_9	0.401	(0.0056; 0.7957)	1.00
θ_{10}	0.2652	(-0.0264; 0.5738)	1.00
θ_{11}	0.4104	(0.0089; 0.8182)	0.99
θ_{12}	0.3999	(0.0008; 0.7984)	1.00
θ_{13}	0.2411	(-0.0583; 0.5366)	0.99
θ_{14}	0.0928	(-0.161; 0.3581)	1.00
θ_{15}	-0.0785	(-0.3752; 0.2109)	1.00
θ_{16}	-0.0710	(-0.3661; 0.2239)	0.99
θ_{17}	0.2507	(-0.0365; 0.5395)	1.00
θ_{18}	0.0874	(-0.1537; 0.323)	0.99
θ_{19}	-0.0726	(-0.3734; 0.2213)	1.01

Comparando os valores encontrados na Tabela 7.2 com aqueles da Tabela 7.1, verifica-se que os valores obtidos foram coerentes. Observa-se que quando o aluno não tem nenhum contado com o professor ($x_i = 0$), isso para θ_4 , θ_5 , θ_9 , θ_{11} e θ_{12} , o efeito é significativo,

ou seja, as expectativas dos professores são altas, variando de 0.39 à 0.42 o efeito do QI. Considerando $x_i = 1$ e $x_i = 2$, observa-se que as estimativas são positivas, variando de 0.08 à 0.26, apesar de ser positivo, o intervalo inclui o zero, ou seja, não são significativos para o modelo. Para $x_i = 3$, ou seja, contato de três semanas entre aluno e professor, temos um menor efeito do QI; neste caso os parâmetros são não significativos para o modelo, pois todas as estimativas são negativas com intervalos incluindo o zero. Isto também pode ser visto na Figura 7.1, quanto maior o contato prévio do aluno com o professor, menor o efeito estimado. Analisando os resultados da Tabela 7.2, a convergência dos parâmetros pode ser confirmada através do critério proposto por *Gelman e Rubin*, visto que os valores obtidos estão próximos de 1.

7.1.1 Análise gráfica.

Gráficos das distribuições *a posteriori* para os parâmetros de interesse estão expostos na Figura 7.2, os quais mostram que as mesmas se comportam de acordo com as distribuições *a priori* especificadas e na Figura 7.3 temos os gráficos de autocorrelações, onde apresentam-se não correlacionados à medida que aumenta o número de iterações.



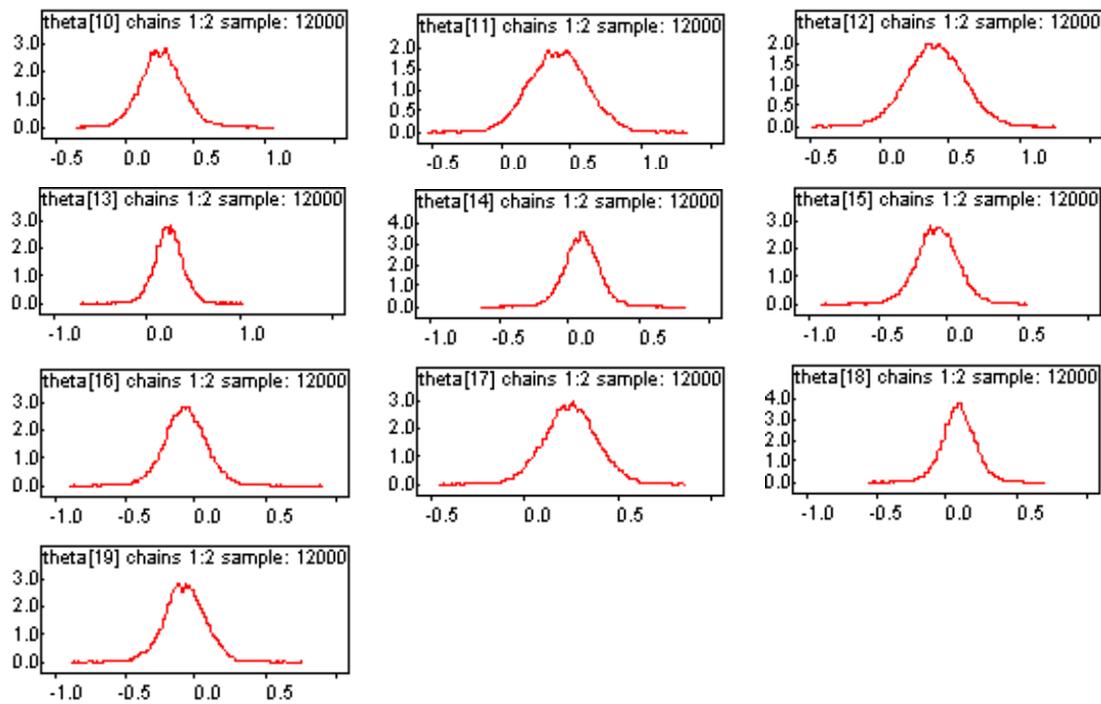
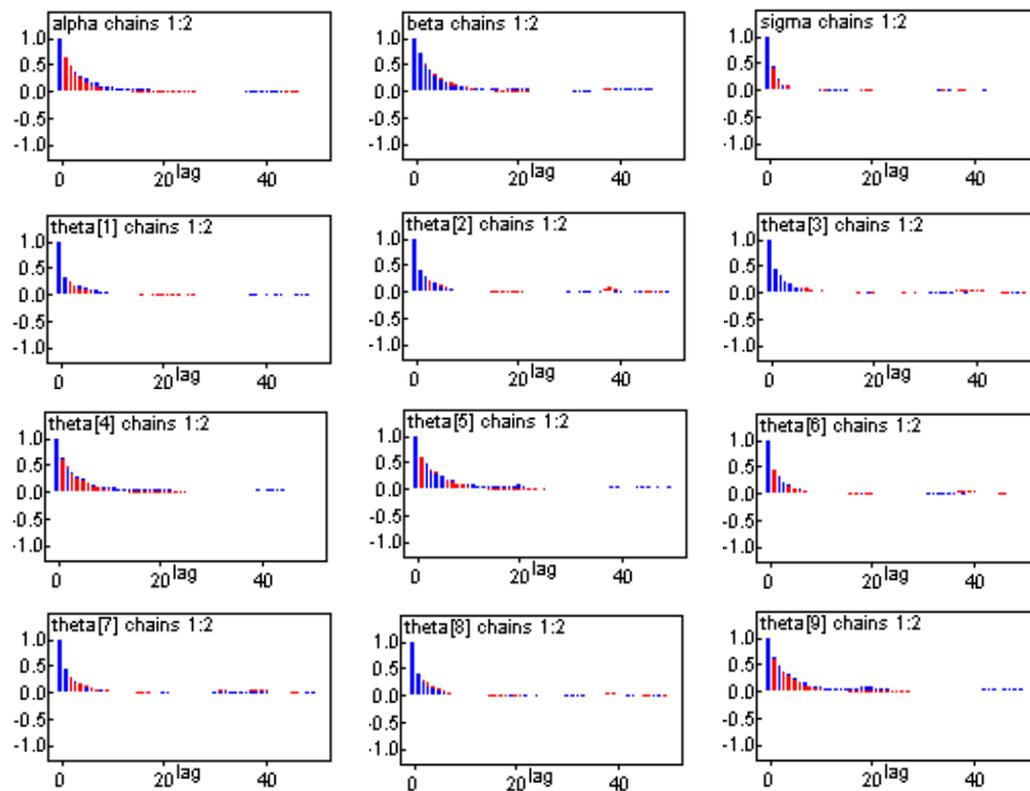


Figura 7.2: Densidades *a posteriori* marginais.



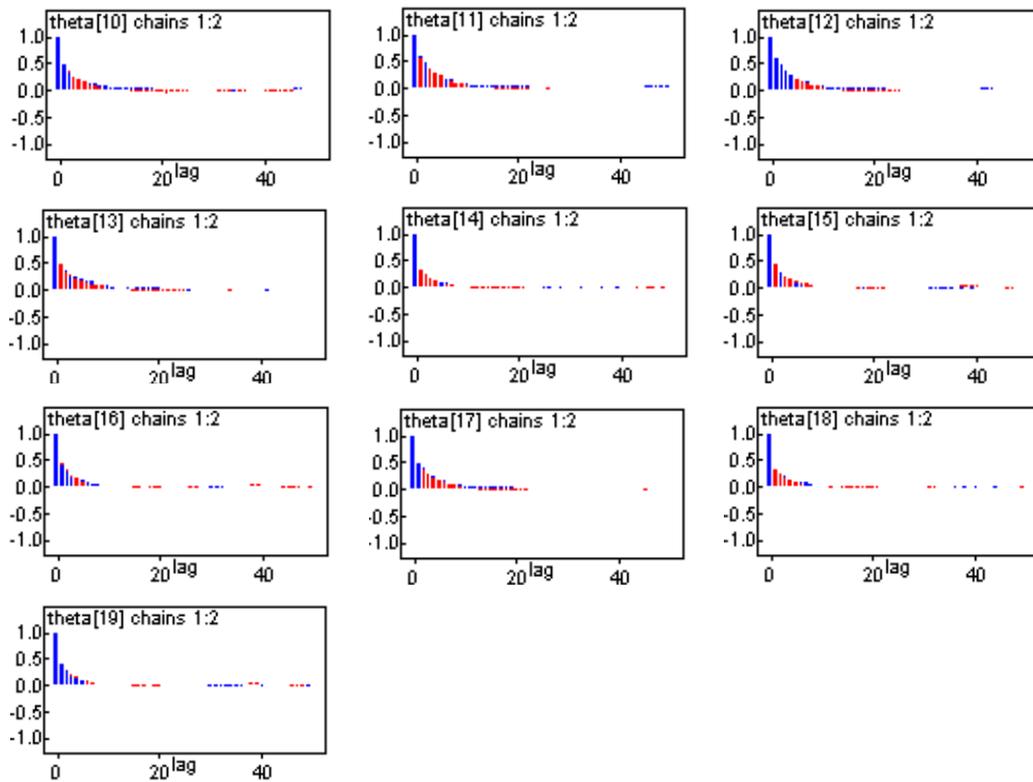
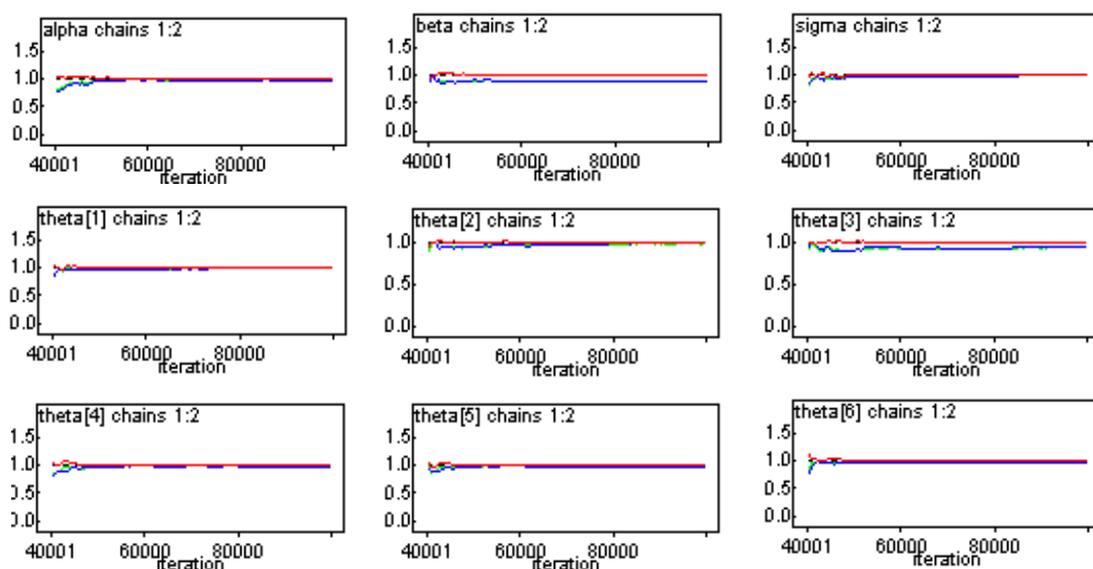


Figura 7.3: Gráficos de autocorrelações.

A Tabela 7.2 mostra os índices de *Gelman e Rubin* que estão de acordo com o valor recomendado que é aproximadamente 1, e através da Figura 7.4 observa-se que à medida que aumenta o número de iterações, temos um indicativo de convergência para todos os parâmetros do modelo. A Figura 7.5, mostra uma indicação de convergência para todos os parâmetros através do gráfico de séries temporais.



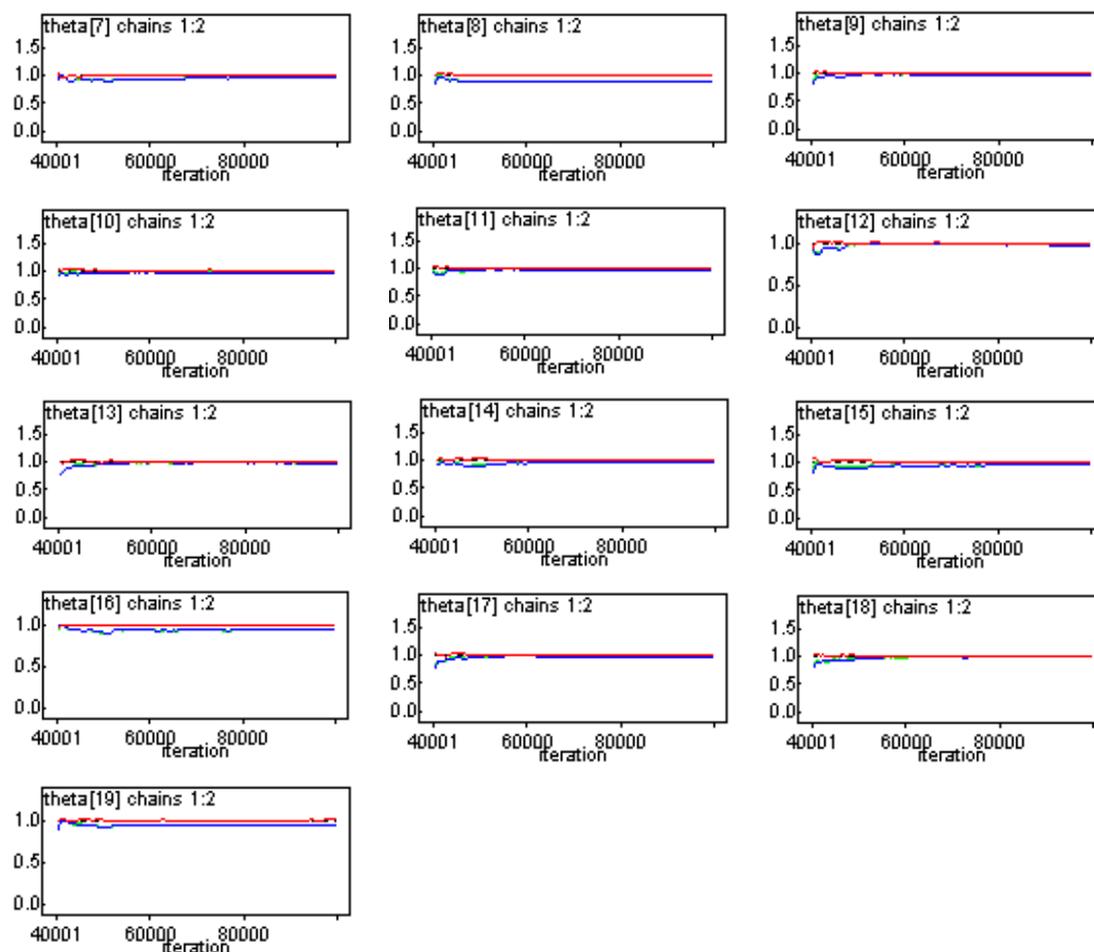
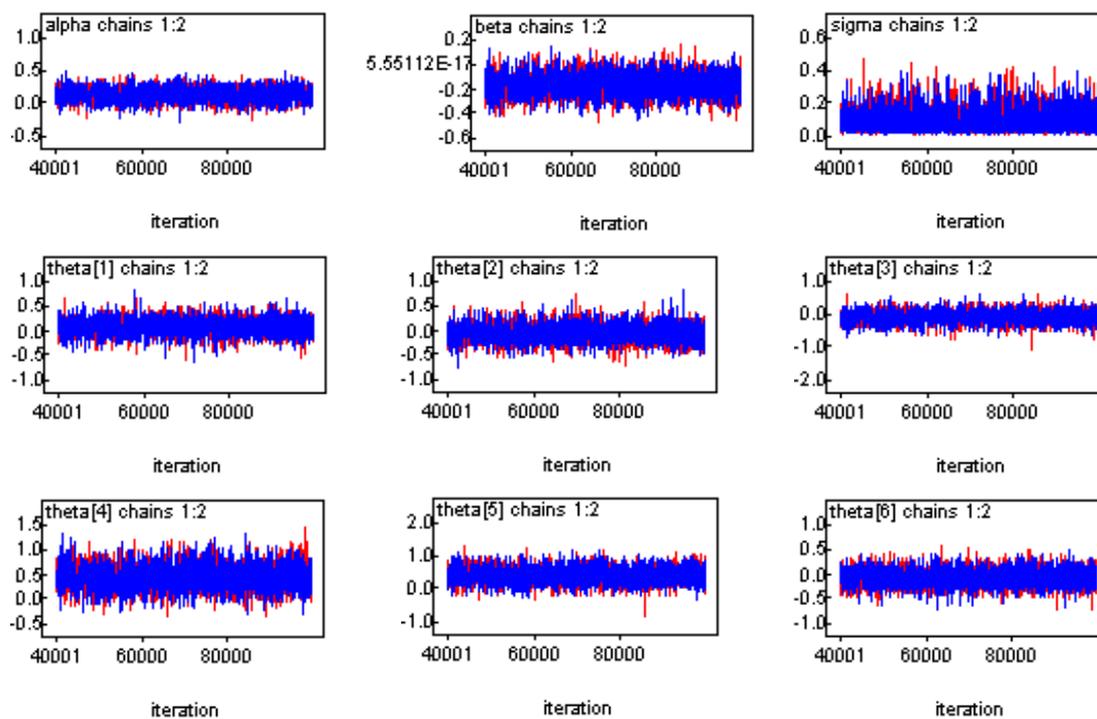


Figura 7.4: Gráficos dos índices R de *Gelman e Rubin*.



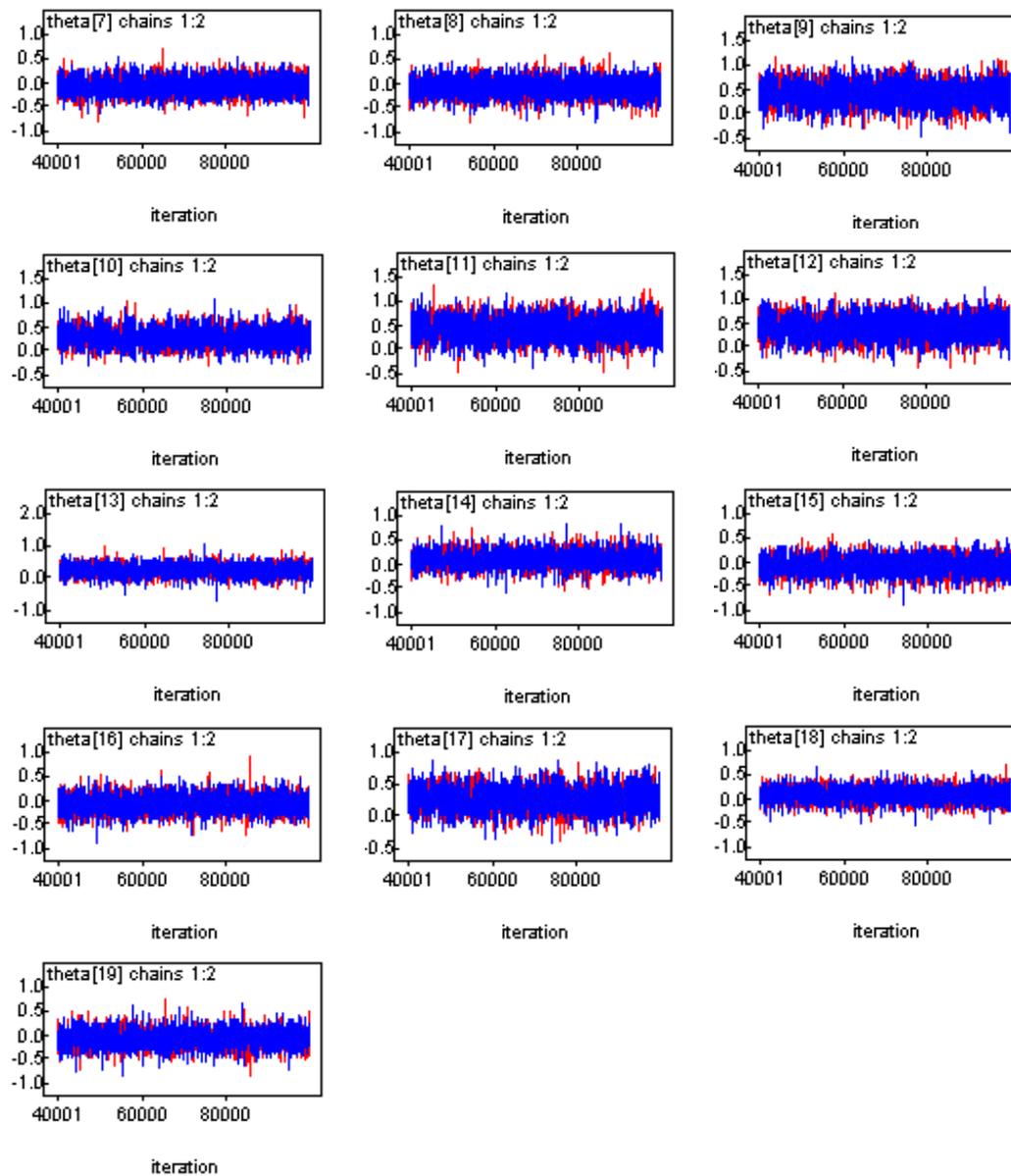


Figura 7.5: Gráficos das séries temporais.

Capítulo 8

Conclusão final.

A metodologia da Meta-análise é recomendável para a obtenção de valores combinados de resultados provenientes de vários estudos independentes, em virtude de permitir a comparação e combinação de resultados distintos, possibilitando uma conclusão geral sobre o corpo da pesquisa. A Meta-análise tem sido frequentemente empregada na área médica, principalmente na comparação de grupos de estudo, utilizando, na maioria das vezes a modelagem hierárquica, e nas aplicações, podemos observar que a utilização do modelo Bayesiano hierárquico é bastante conveniente e tem sido muito útil para avaliar se existe alguma associação entre uma determinada doença e seus fatores de risco e também para analisar o efeito de aspirina em ataques cardíacos, assim como exemplificamos neste trabalho.

O uso de métodos Bayesianos está tornando-se cada vez mais popular em Meta-análise. Em especial, a simulação de métodos baseada em MCMC tem sido uma alternativa prática de grande interesse em Meta-análise, onde o modelo proposto pode ter também muitos parâmetros, e a inferência clássica usual pode ser difícil ou, às vezes, impossível para ser obtida.

Além do uso de modelos propostos considerado neste trabalho, estes podem ser aplicados para muitas outras situações de combinações de estudos, em diferentes aplicações.

Apêndice A

Algoritmo Gibbs Sampling

O Amostrador de Gibbs é um esquema iterativo de amostragem de uma Cadeia de Markov cujo núcleo de transição é formado pelas distribuições condicionais completas.

Para descrever esse algoritmo, supor que $\pi(\theta)$ seja a distribuição de interesse, para, $\theta = (\theta_1, \dots, \theta_p)$, e que seja praticamente impossível gerar amostras diretamente de $\pi(\theta)$; se, as densidades condicionais a posteriori $\pi_i(\theta_i) = \pi(\theta|\theta_{-i}), i = 1, \dots, p$

estiverem disponíveis, então podemos gerar amostras de $\pi(\theta)$ através de sucessivas gerações dessas distribuições condicionais $\pi_i(\theta_i)$ (ver *Casella e George, 1992* e *Gelfand, (1995)*).

Este método pode ser descrito da seguinte maneira:

Passo1: Inicialize o contador da cadeia $k = 1$ e escolha valores iniciais para $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})$;

Passo2: Obtenha um novo valor para $\theta^{(k)} = (\theta_1^{(k)}, \dots, \theta_p^{(k)})$ a partir de $\theta^{(k-1)}$ através de sucessivas gerações de valores,

$$\begin{aligned}\theta_1^{(k)} &\sim \pi(\theta_1|\theta_2^{(k-1)}, \dots, \theta_p^{(k-1)}) \\ \theta_2^{(k)} &\sim \pi(\theta_2|\theta_1^{(k)}, \dots, \theta_p^{(k-1)}) \\ \theta_3^{(k)} &\sim \pi(\theta_3|\theta_1^{(k)}, \theta_2^{(k)}, \dots, \theta_p^{(k-1)}) \\ &\dots \\ \theta_p^{(k)} &\sim \pi(\theta_p|\theta_1^{(k)}, \theta_2^{(k)}, \theta_3^{(k)}, \dots, \theta_{p-1}^{(k-1)})\end{aligned}$$

Passo3: Atualize o contador de k para $k + 1$ e retorne ao passo 2 até atingir a convergência.

Apêndice B

Algoritmo Metropolis-Hastings

Este algoritmo (ver *Chib e Greenberg, 1994*) é usado para gerar amostras de uma densidade de forma indireta, quando não se sabe gerar diretamente. Assim, seja a distribuição de interesse denotada por π e supor que seja muito difícil ou impossível gerar amostras diretamente de π . Se as densidades condicionais a *posteriori* não tiverem formas conhecidas, neste caso, usamos o algoritmo Metropolis-Hastings.

Neste algoritmo, se no estágio k a cadeia está no estado x , o próximo estado da cadeia é encontrado gerando um valor candidato y de uma densidade arbitrária $q(x, y)$. Se $q(x, y)$ satisfaz a condição de reversibilidade da cadeia, então:

$$q(x, y)\pi(x) = q(y, x)\pi(y)$$

e assim, temos um núcleo de transição satisfatório e todos os valores gerados serão aceitos. Mas, no entanto, para algum (x, y) podemos ter que

$$q(x, y)\pi(x) > q(y, x)\pi(y)$$

nesse caso, a cadeia se move de x para y com maior frequência do que de y para x . Para corrigir isso, o número de movimentos de x para y é reduzido introduzindo a probabilidade de ocorrer este movimento $\alpha(x, y) < 1$. Se o movimento ocorre, então $x^{(k+1)} = y$, caso contrário $x^{(k+1)} = x^{(k)}$.

Assim, temos a relação:

$$q(x, y)\pi(x)\alpha(x, y) = q(y, x)\pi(y).$$

A partir da relação acima, temos que a condição de reversibilidade da cadeia é satisfeita por construção e a probabilidade de aceitação do valor gerado é:

$$\alpha(x, y) = \begin{cases} \min \left\{ \frac{q(y, x)\pi(y)}{q(x, y)\pi(x)}, 1 \right\}, & \text{se } q(x, y)\pi(x) > 0 \\ 1 & \text{caso contrário} \end{cases}$$

Com as definições acima o algoritmo segue da seguinte forma:

Passo1: Inicie o contador da cadeia $k = 1$ e escolha o valor inicial para x , $x^{(0)}$;

Passo2: Mova a cadeia para um novo valor y gerado da densidade $q(x^{(k-1)}, y)$;

Passo3: Calcule a probabilidade de aceitação do movimento, $\alpha(x^{(k-1)}, y)$ e compara com um valor u gerado de uma distribuição uniforme $[0,1]$. Se o movimento for aceito, ou seja, $u \leq \alpha(x^{(k-1)}, y)$, então $x^{(k)} = y$, se não for aceito, $x^{(k)} = x^{(k-1)}$ e a cadeia não se move;

Passo4: Mude o contador de k para $k+1$ e retorne ao passo 2 até obter uma distribuição estacionária.

Apêndice C

Verificação de convergência

Uma forma de verificar a convergência dos algoritmos Gibbs Sampling e Metropolis-Hastings é gerar várias cadeias com diferentes condições iniciais para se certificar que se trata de uma cadeia de Markov irredutível, ou seja, que existe uma distribuição de equilíbrio. Para verificar a convergência do algoritmo, podemos considerar várias técnicas, tais como, (ver *Gelfand e Smith, 1990*), sugerem o uso de técnicas gráficas para a verificação da convergência do seguinte modo:

Gerava-se um grande número de iterações N em m cadeias paralelas. Com a amostra gerada, construa um histograma. O mesmo procedimento é repetido com $N + k$ iterações. Se não houver diferença perceptível entre os gráficos, então pode-se concluir pela convergência das cadeias.

Outra possibilidade para verificar a convergência é proposta por *Gelman e Rubin (1992)*. Este método é baseado em técnicas de análise de variância e sugere a convergência da cadeia apenas quando a variância entre as cadeias ($m \geq 2$) for bem menor que a variância dentro de cada cadeia ou, quando os histogramas das cadeias misturadas são similares ao de cada uma delas isoladamente.

Esse método pode ser descrito da seguinte maneira:

Considerando m cadeias paralelas em uma função real $t(\theta)$, tem-se m trajetórias $\{t_i^{(1)}, t_i^{(2)}, \dots, t_i^{(n)}\}$, $i = 1, \dots, m$ para cada t . Portanto, podem ser obtidas a variância

entre as cadeias B e a variância dentro das cadeias ω , dadas por:

$$B = \frac{n}{m-1} \sum_{i=1}^m (\bar{t}_i - \bar{t})^2 \text{ e } \omega = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (t_i^{(j)} - \bar{t}_i)^2$$

na qual, \bar{t}_i é a média das observações da cadeia i e \bar{t} é a média dessas médias, $i = 1, \dots, m$.

Sob convergência, todos $m \times n$ valores são gerados da *posteriori* e a variância t pode ser estimada de forma não viciada por

$$\hat{\sigma}^2 = \frac{n-1}{n}(\omega) + \frac{1}{n}(B)$$

e a média desejada pode ser estimada por $\hat{\mu} = \bar{t}$.

Um indicador da convergência é dado pela redução potencial estimada de escala $R = \sqrt{\hat{\sigma}^2/\omega}$. Assim, à medida que n cresce, R convergirá para 1. Logo, R pode ser usado como indicador da convergência pela avaliação de sua proximidade a 1, e então, admitindo que a convergência ocorreu, admitimos também que as amostras das iterações selecionadas são independentes e identicamente distribuídas sobre a distribuição desejada.

Apêndice D

Alguns programas desenvolvidos no WinBugs

1) Estes programas são referentes ao exemplo introduzido no capítulo 3 de aplicação de uma nova droga antiepilética. Os programas foram feitos para cada grupo, ou seja, o primeiro programa se refere aos estudos abertos e o segundo programa aos estudos fechados.

- Programa 1

```
model
{
for (i in 1:N1)
{
y[i] ~ dbin (p[i],n[i])
logit(p[i]) <- theta[i]
theta[i] ~ dnorm(qsi1, tau1)
}
qsi1 ~ dnorm(0,0.001)
tau1 ~ dgamma(1,0.1)
sigma1 <- 1 / sqrt(tau1)
}
list(y=c(17,8,41,13,32,90), n=c(30,16,69,23,42,151), N1=6)
list(qsi1=0, tau1=1)
```

- Programa 2

Modelo

```
{
for (i in 1: N2)
{
y[i] ~ dbin (p[i],n[i])
logit(p[i]) <- theta[i]
theta[i] ~ dnorm(qsi2, tau2)
}
qsi2 ~ dnorm(1,0.1)
tau2 ~ dgamma(1,1)
}
list(y=c(5,9,3,7,8,9,1,12,17), n=c(20,20,17,15,18,17,19,51,59), N2=9)
list(qsi2=0, tau2=1)
```

2) Este programa é referente ao exemplo introduzido no capítulo 4 da relação entre o uso de estrógeno e a incidência de câncer no endométrio. Quando o erro do modelo de regressão possui uma distribuição normal.

Modelo

```
{
for (i in 1: N) {
y[i] ~ dnorm (mu[i], tau)
mu[i] <- beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+delta*w[i]
}
beta1 ~ dnorm (0.2,1)
beta2 ~ dnorm (-1.2,1)
beta3 ~ dnorm (9.7,1)
delta ~ dnorm (-1.12,1)
tau ~ dgamma (1,1)
}
list(y=c(2.08,-0.06,0.76,1.58,0.87,0.71,1.24,1.99),
x1=c(0.42,0.12,0.50,0.47,0.64,0.21,0.53,0.52),
```

```
x2=c(0.77,0.33,0.73,0.88,0.65,0.65,0.65,0.54),
```

```
x3=c(0.41,0.18,0.26,0.24,0.16,0.16,0.24,0.24),
```

```
w=c(1,1,1,0,0,0,0), N=8)
```

```
list(beta1=0.2, beta2=-1.2, beta3=9.7, delta=-1.12,tau=1)
```

3) Este programa é referente ao exemplo introduzido no capítulo 5 usando mistura de duas distribuições para o erro do modelo.

Modelo

```
{
```

```
for(i in 1: N)
```

```
{
```

```
y[i] ~ dnorm(mu[i], tau[T[i]])
```

```
mu[i] <- lambda[T[i], i]
```

```
T[i] ~ dcat(P[])
```

```
}
```

```
P[1:2] ~ ddirch(gamma[])
```

```
beta1 ~ dnorm(0.2,1)
```

```
beta2 ~ dnorm(-1.2,1)
```

```
beta3 ~ dnorm(9.7,1)
```

```
delta ~ dnorm(-1.12,1)
```

```
mi1 ~ dnorm(-0.35,30)
```

```
mi2 ~ dnorm(0.15,30)
```

```
tau[1] ~ dgamma(1,1)
```

```
tau[2] ~ dgamma(1,1)
```

```
p1 ~ dbeta(5,5)
```

```
p2 ~ dbeta(5,5)
```

```
for( i in 1 : N )
```

```
{
```

```
lambda[2,i]<-lambda[1,i]
```

```
lambda[1,i]<-beta1*x1[i]+beta2*x2[i]+beta3*x3[i]+delta*w[i]
```

```
} }
```

```
list(y=c(2.08,0.06,0.76,1.58,0.87,0.71,1.24,1.99), N=8, gamma=c(1,1),
```

```

x1=c(0.42,0.12,0.50,0.47,0.64,0.21,0.53,0.52),
x2=c(0.77,0.33,0.73,0.88,0.65,0.65,0.65,0.54),
x3=c(0.41,0.18,0.26,0.24,0.16,0.16,0.24,0.24),
T=c(1,NA,NA,NA,NA,NA,NA,2), w=c(1,1,1,0,0,0,0,0))
list(beta1=0.2, beta2=1.2, beta3=9.7, delta=-1.12)

```

4) Este programa é referente ao exemplo do efeito de aspirina em ataques cardíacos, introduzido no capítulo 6.

```

Modelo
{
for (i in 1: N)
{
teta[i] ~dnorm (tetam, tau)
y[i] ~dnorm (teta[i], V[i])
}
tetam ~dnorm(1,0.01)
tau ~dgamma (1,1)
}
list(y=c(2.77,2.50,1.84,2.56,2.31,-1.15),
      V=c(2.7225,1.7161,5.4756,2.7889,3.9402,0.810), N=6)
list(teta=c(0,0,0,0,0,0), tau=1)

```

5) Este programa é referente ao exemplo de Meta-análise em Educação, introduzido no capítulo 7.

```

model
{
for (i in 1:N)
{
thetam[i] <- alpha + beta * (x[i] - a)
theta[i] ~dnorm(thetam[i], tau)
SE[i] <-1/ V[i]
y[i] ~dnorm (theta[i], SE[i])
}
}

```

```
alpha ~ dnorm(0, 0.001)
beta ~ dnorm(0, 0.001)
tau ~ dgamma(0.1,0.001)
sigma <- 1 / sqrt(tau)
}
list(N=19, a=1.7368, y=c(0.03,0.12,-0.14,1.18,0.26,-0.06,-0.02,
-0.32,0.27,0.80,0.54,0.18,-0.02,0.23,-0.18,-0.06,0.30,0.07,-0.07),
x=c( 2,3,3,0,0,3,3,3,0,1,0,0,1,2,3,3,1,2,3),
V=c(0.125, 0.147, 0.167, 0.373, 0.369, 0.103, 0.103, 0.220,
0.164, 0.251, 0.302, 0.223, 0.289, 0.290, 0.159, 0.167,
0.139, 0.094, 0.174))
list(theta=c(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0), alpha=1, tau=1, beta=1)
```

Referências Bibliográficas

- [1] ACHCAR, J.A.; MAZUCHELI, J.; FORTULAN, V. C. (2001). *Meta-analysis: A Bayesian Approach*. Brazilian Journal of Probability and Statistics, 15, pp. 33-47.
- [2] ACHCAR, J.A.; MAZUCHELI, J.; KASS, R. (2001). *Use of mixture of distributions for the error in regression models for life time data*. Brazilian Journal of Probability and Statistics, 15, pp. 155-168.
- [3] ADAMI, H. O. (1992). *Long-term consequences of estrogen and estrogen-progestin replacement*. Cancer Causes and Control, 2, 83-90.
- [4] BIGGERSTAFF, B. J.; TWEEDIE, R. L. (1997). *Incorporating Variability in Estimates of Heterogeneity in the Random Effects Model in Meta-Analysis*. *Statistics in Medicine*, Vol. 16, 753-768.
- [5] BOLFARINE, H.; SANDOVAL, M. C. (2001). *Introdução à Inferência Estatística* Rio de Janeiro-RJ: Coleção Matemática Aplicada, Sociedade Brasileira de Matemática.
- [6] BOX, G. E. P.; TIÃO, G. C. (1992). *Bayesian Inference in Statistical Analysis*, Wiley Classics Library Edition Published.
- [7] BRINTON, L. A.; SCHAIER, C. (1993). *Estrogen replacement therapy and breast cancer risk*. *Epidemiologic Reviews*, 15, 66-79.
- [8] BUSSAB, W. O.; MORETTIN, P. A. (2003). *Estatística básica*. 5.ed. São Paulo: Saraiva, 526p.
- [9] CASELLA, G.; GEORGE, E. I. (1992). *Explaining the Gibbs Sampler*. *The American Statistician*, 46, 3, 167-174.

- [10] CHIB, S.; GREENBERG, E. (1994). *Understanding the Metropolis-Hastings Algorithm*, American Statistician, 49, 4, 327-335.
- [11] COSTA, P. A. B. (1999). *Um enfoque segundo a teoria de conjuntos difusos para Meta-análise*. Florianópolis, 140p Tese (Doutorado) - Universidade Federal de Santa Catarina.
- [12] DIEBOLT, J.; ROBERT, C. P. (1994). *Estimation of Finite Mixture Distributions through Bayesian Sampling*. Journal Royal Statistical Society, Vol. 56, No.2, pp. 363-375.
- [13] DRAPER D. (2000). *Bayesian Hierarchical Modeling*. Department of Mathematical Sciences University of Bath, England, Technical Report.
- [14] DRAPER, D.; GAVER, D.; GOEL, P.; GREENHOUSE, J.; HEDGES, L.; MORRIS, C.; TUCKER, J.; WATERNAUX (1993). *Combining Information: Statistical Issues and opportunities in research*, Contemporary Statistics series, n.1, American Statistical Association, Alexandria, Va.
- [15] DUMOUCHEL W. H.; HARRIS J. E. (1983). *Bayes Methods for Combining the Results of Cancer Studies in Humans and Other Species*. Journal of the American Statistical Association, 78, 382, 293-315.
- [16] EGGER, M.; SMITH, G.D.; PHILIPS, A.N. (1997). *Meta-analysis: Principles and procedures*. British Journal of Medicine, vol. 315, p.1533-1537.
- [17] EVERITT, B.; HAND, D. J. (1981). *Finite Mixture Distributions*. London; New York: Chapman and Hall.
- [18] FAGARD, R. H.; STAESSEN, J. A.; THIJS, L. (1996). *Advantages and disadvantages of the meta-analysis approach*. Journal of Hypertension, v.14, suppl.2,p.9-13.
- [19] FISHER, R. A. (1932). *Statistical methods for research works* (4th ed.). London: Oliver & Boyd.
- [20] FLEISS, J. L. (1973). *Statistical methods for rates e proportions*. John Wiley & Sons.

- [21] FORTULAN, V. C. (1999). *Meta-análise: Um enfoque bayesiano*. São Carlos, 133p Dissertação (Mestrado) - Instituto de Ciências Matemáticas de São Carlos, Universidade de São Paulo.
- [22] GELFAND, A. E.; SMITH, A. F. M. (1990). *Sampling-based approaches to calculating marginal densities*. Journal of the American Statistical Association, 85, 398-409.
- [23] GELFAND, A. E. (1995). *Gibbs Sampling*. Encyclopedia of Statistical Science.
- Gelfand, A.E. (1995). Gibbs sampling (a contribution to the Encyclopedia of Statistical Science). Preprint, University of Connecticut. <http://citeseer.ist.psu.edu/gelfand95gibbs.html>
- [24] GELMAN, A.B.; CARLIM, J.S.; STERN, H.S.; RUBIN; D.B. (1995). *Bayesian Data Analysis*. CHAPMAN & HALL/CRC, London New York Washington; D.C.
- [25] GELMAN, A. B.; RUBIN, D. B. (1992). *Inference from iterative simulation using multiple sequences*, Statistical Science, 4, 457-511.
- [26] GIANNOTTI, J. D. G.; PACKER, I. U.; MERCADANTE, M. E. Z. (2002). *Meta-analysis for estimates of genetic correlation between birth and weaning weights of cattle*. Sci. agric. (Piracicaba, Brazil.), Vol.59, No.3, p.435-440.
- [27] GLASS, G. V. (1976). *Primary, secondary, and meta-analysis of research*. Educational Researcher, 5, 3-8.
- [28] HEDGES, L. V.; OLKIN, I. (1985). *Statistical methods for meta-analysis*. Orlando, F. L. Academic Press, 369p.
- [29] KAUFMAN, D. W.; PALMER, J. R.; MOUZON, J.; ROSENBERG, L.; STOLLEY, P. D.; WARSHAUER, M. E.; ZAUBER, A. G.; SHAPIRO, S. (1991). *Estrogen replacement therapy and the risk of breast cancer: results from the case control surveillance study*. American Journal of Epidemiology, Vol. 134, No. 12, 1375-1385.
- [30] LAROSE, D. T., DEY, D. K. (1997). *Grouped Random Effects Models for Bayesian Meta-Analysis*. Statistics in Medicine, Vol. 16, 1817-1829.

- [31] LAROSE, D. T., DEY, D. K. (1995). *Modelling Dependent Covariate Subclass Effects in Bayesian Meta-Analysis*. Relatório Técnico, University of Connecticut, Storrs.
- [32] LI, Z.; BEGG, C.B. (1994). *Random effects for combining results from controlled and uncontrolled studies in a meta-analysis*. Journal of the American Statistical Association, v.89, p.1523-1527.
- [33] PAULINO, C. D.; TURKMAN M. A. A.; MURTEIRA B. (2003). *Estatística Bayesiana*. Edição da Fundação Calouste Gulbenkian.
- [34] SILLIMAN, N. P. (1997). *Hierarchical Selection Models With Applications in Meta-Analysis*. Journal of the American Statistical Association, Vol. 92, No. 439, 926-936.
- [35] VAN BELLE, G. (1992). *A Meta-analysis of the effectiveness of progabide a an antiepileptic drug*. Unpublished manuscript.