
Metanálise para Modelos de Regressão

Laryssa Vieira dos Santos

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Laryssa Vieira dos Santos

Metanálise para Modelos de Regressão

Dissertação apresentada ao Departamento de Estatística, DEs-UFSCar e ao Instituto de Ciências Matemáticas e de Computação, ICMC-USP, como parte dos requisitos para obtenção do título de Mestra em Estatística junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

Área de Concentração: Estatística

Orientador: Prof. Dr. Adriano Polpo de Campos

UFSCar/USP - São Carlos

Outubro de 2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar
Processamento Técnico
com os dados fornecidos pelo(a) autor(a)

S237m Santos, Laryssa Vieira dos
Metanálise para Modelos de Regressão / Laryssa
Vieira dos Santos. -- São Carlos : UFSCar, 2016.
48 p.

Dissertação (Mestrado) -- Universidade Federal de
São Carlos, 2016.

1. Metanálise. 2. Inferência Bayesiana. I. Título.



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a defesa de dissertação de mestrado da candidata Laryssa Vieira dos Santos, realizada em 28/10/2016:

Prof. Dr. Adriano Polpo de Campos
UFSCar

Prof. Dr. Eduardo Yoshio Nakano
UnB

Prof. Dr. Marcio Luis Lanfredi Viola
UFSCar

Certifico que a sessão de defesa foi realizada com a participação à distância do membro Prof. Dr. Eduardo Yoshio Nakano e, depois das arguições e deliberações realizadas, o participante à distância está de acordo com o conteúdo do parecer da comissão examinadora redigido no relatório de defesa da aluna Laryssa Vieira dos Santos.

Prof. Dr. Adriano Polpo de Campos
Presidente da Comissão Examinadora
UFSCar

Resumo

A metanálise tem sido amplamente utilizada em estudos médicos especialmente em revisões sistemáticas de ensaios clínicos aleatorizados. Para modelos de regressão a técnica ainda é muito escassa e limitada. Geralmente, trata-se apenas de uma medida baseada nas médias de estimativas pontuais dos diferentes estudos, perdendo-se muita informação dos dados originais. Atualmente, torna-se cada vez mais fundamental o uso da metanálise para sumarizar estudos de mesmo objetivo, em razão do avanço da ciência e o desejo de usar o menor número de seres humanos em ensaios clínicos. Utilizando uma medida metanalítica Bayesiana, o objetivo é propor um método genérico e eficiente para realizar metanálise em modelos de regressão.

Palavras-chave: Metanálise, Inferência Bayesiana

Sumário

1	Introdução	2
1.1	Objetivo	3
1.2	Organização do trabalho	3
2	Conceitos de Metanálise	3
3	Metodologia	6
3.1	Metanálise via Modelo Bayesiano Hierárquico	6
3.2	Medida Metanalítica	11
3.2.1	Abordagem Frequentista da Metanálise	16
3.2.2	Comparação entre os métodos	17
3.3	Aplicação Geral	19
4	Metanálise em Modelos de Regressão	26
4.1	Introdução	26
4.2	Metanálise em Modelos de regressão	27
4.3	Aplicação Dados Reais	33
4.3.1	Análise de regressão	34
4.3.2	Metanálise em modelos de regressão	40
5	Considerações Finais	45
	Referências Bibliográficas	47

Lista de Tabelas

3.1	Proporção de sucessos no uso da droga antidepressiva SAMe.	20
3.2	Exemplo Berry: Medidas descritivas à posteriori.	24
4.1	Estimativas dos parâmetros do modelo Binomial Negativo - PCE.	36
4.2	Análise de desvio - PCE.	36
4.3	Estimativas dos parâmetros do modelo - MNPCE.	38
4.4	Análise de desvio - MNPCE.	39

Lista de Figuras

3.1	<i>Esquema Bayesiano Hierárquico.</i>	7
3.2	<i>Curvas de nível a posteriori de (μ, σ) da metanálise.</i>	10
3.3	<i>Distribuições a posteriori marginal de μ e σ.</i>	10
3.4	<i>(a)Curvas de nível da distribuição conjunta a posteriori de (μ, σ^2) em cada estudo; (b)Curvas de nível da distribuição a posteriori metanalítica de (μ, σ^2); (c)Gráfico tridimensional da distribuição a posteriori metanalítica de (μ, σ^2); (d)Região HPD de 95% de credibilidade para (μ, σ^2).</i>	15
3.5	<i>Forest Plot para os dados do Exemplo 3.2</i>	17
3.6	<i>Distribuições a posteriori marginal de μ.</i>	18
3.7	<i>Forest Plot para proporção de sucessos no uso da droga SAME.</i>	20
3.8	<i>Esquema Bayesiano Hierárquico Aplicação Geral.</i>	22
3.9	<i>Média a posteriori de F, geradora da proporção dos sucessos. É a densidade média com respeito a distribuição a posteriori de (a,b).</i> . .	23
3.10	<i>(a) Distribuições a posteriori de p para os nove estudos. (b) Distribuição a posteriori metanalítica de p.</i>	24
4.1	<i>Densidade a posteriori marginal de β_0, β_1 e σ^2.</i>	30
4.2	<i>Reta de Regressão para cada estudo e média da medida metanalítica.</i> .	31
4.3	<i>Curva de nível da densidade preditiva de Y baseada na medida metanalítica.</i>	31
4.4	<i>Densidade preditiva de Y baseada na medida metanalítica.</i>	32
4.5	<i>Gráfico normal de probabilidades referente ao modelo BN ajustado aos dados.</i> . .	36
4.6	<i>Gráfico para verificação de pontos atípicos (PCE).</i>	37

4.7	<i>Gráfico para verificação de suposições dos resíduos (PCE).</i>	37
4.8	<i>Gráfico normal de probabilidades referente ao modelo BN ajustado aos dados.</i>	38
4.9	<i>Gráfico para verificação de pontos atípicos (MNPCE).</i>	39
4.10	<i>Gráfico para verificação de suposições dos resíduos (MNPCE).</i>	39
4.11	<i>Distribuição a posteriori de μ condicional as covariáveis para variável resposta PCE/NCE.</i>	42
4.12	<i>Medida metanalítica analisada para variável PCE/NCE.</i>	42
4.13	<i>Distribuição a posteriori de μ condicional as covariáveis para variável resposta MNPCE.</i>	43
4.14	<i>Medida metanalítica analisada para variável MNPCE.</i>	44

Capítulo 1

Introdução

A metanálise é uma técnica estatística que combina resultados de diferentes estudos independentes com o objetivo de obter uma estimativa mais precisa do efeito comum a todos os estudos. A medida metanalítica é o principal parâmetro de interesse e sintetiza os resultados dos estudos envolvidos.

Esta metodologia tem se mostrado uma ferramenta muito útil na área médica, principalmente, do ponto de vista de redução do tamanho amostral de experimentos, pois através da metanálise é possível combinar resultados de experimentos prévios e obter conclusões importantes antes de realizar-se um novo experimento. A redução pode dar-se no caso em que a metanálise aponta, por exemplo, que um tratamento A é melhor que B. No caso de realizar um estudo comprobatório, possivelmente o tamanho amostral não precisará ser tão grande quanto o de costume. Outra possibilidade é em situações claras de que um tratamento não é bom, não há nem a necessidade de se realizar um experimento comprobatório, evitando gastos desnecessários. Quando as unidades amostrais do experimento são seres vivos, a redução do número de amostras traz um ganho significativo.

A medida metanalítica de Martins [12], por ser uma distribuição de probabilidade no espaço paramétrico, preserva as informações a posteriori sobre os parâmetros de interesse. No caso da metanálise frequentista, o foco está na média dos estudos, levando muitas vezes a perda de informação sobre a variabilidade e outros momentos importantes. Então, neste caso, utiliza-se de técnicas para estimar a variância. Entretanto, tais métodos tratam geralmente apenas da variância do estimador pontual

da média da distribuição amostral. Desta forma, a metanálise tradicional falha na obtenção de um modelo de regressão.

1.1 Objetivo

O objetivo principal do trabalho é desenvolver um método metanalítico geral para modelos de regressão (por exemplo, análise de regressão linear múltipla, análise de sobrevivência, modelos lineares generalizados, ...) baseado na medida metanalítica proposta por Martins [12]. Pretende-se avaliar a qualidade do método desenvolvido fazendo comparações com outras soluções para o problema de regressão com mais de um estudo, sendo eles o modelo Bayesiano hierárquico e um modelo de regressão que considera os diferentes estudos como uma covariável. Cabe ressaltar que o intuito é trabalhar somente com dados completos de todos os estudos e não apenas medidas resumo. Entretanto, também será investigado o caso onde somente as medidas resumo estão disponíveis, verificando se existe uma solução viável para tal. Como objetivo final será reanalisado o problema de genotoxicidade ambiental em roedores selvagens (Bueno, 2000), utilizando a medida metanalítica a ser desenvolvida e comparando com os resultados obtidos em Bueno [4].

1.2 Organização do trabalho

Além desta introdução que traz o conceito geral de metanálise, a dissertação conta com mais quatro capítulos que retratam as diferentes abordagens e a discussão dos métodos Frequentista e Bayesiano como forma de resolução da técnica.

O Capítulo 2 faz referência aos conceitos da metanálise e as formas como a mesma é realizada. No Capítulo 3, estuda-se principalmente o texto de Martins [12], acrescentando o conhecimento sobre o tratamento da metanálise via modelo Bayesiano hierárquico e a proposta de uma medida metanalítica Bayesiana baseada na mistura de distribuições a posteriori do parâmetro de interesse de cada estudo pertencente à metanálise. Neste momento, o trabalho traz uma alternativa mais informativa daquela vista usualmente, uma vez que a medida metanalítica proposta é uma distribuição de probabilidades e não apenas uma simples medida-resumo da

característica de interesse. Ademais, esta medida também será útil para encontrar similaridades entre os estudos envolvidos na metanálise. Neste capítulo são dados exemplos simulados e um exemplo com dados reais disponível em Berry [3].

No Capítulo 4, dedicamo-nos a aplicar a técnica apresentada no capítulo anterior em modelos de regressão. A fim de exemplificar a teoria, foi desenvolvido um exemplo de aplicação simples com dados simulados e um posteriormente um exemplo com dados reais.

O Capítulo 5 destina-se às considerações finais da dissertação.

Vale ressaltar que a metodologia proposta no capítulo 3 pode ser utilizada com qualquer distribuição a priori e função de verossimilhança e, para os cálculos Bayesianos, utilizou-se o método de simulação de Metropolis-Hastings.

Capítulo 2

Conceitos de Metanálise

Uma revisão sistemática é um tipo de pesquisa planejada para resumir diferentes estudos com a finalidade de responder a uma pergunta específica. A grande disponibilidade de informações científicas e a necessidade de aproveitar o volume dessas informações fazem com que novas tecnologias de sistematização sejam criadas. Dessa forma, segundo Martins [12], diversos autores vem utilizando-se de métodos apropriados para identificar, selecionar e avaliar de forma crítica dados envolvidos nos estudos. Em uma revisão sistemática, quando os dados dos diferentes estudos são combinados de maneira quantitativa, denomina-se de técnica de metanálise. A metodologia é uma forma de combinar estudos realizados sob diferentes condições, com distintos níveis de precisão e por grupos de pesquisadores de diferentes regiões, países e formações. Assim, espera-se conclusões mais amplas do que as obtidas pelos estudos que compõem a sistematização.

O papel da metanálise de sumarizar estudos publicados torna-se cada dia mais fundamental e podemos definí-la como uma revisão sistemática quantitativa que combina, de forma estatística, pelo menos dois estudos para produzir uma estimativa única. O método da combinação quantitativa de diferentes estudos surgiu na década de 30 com os trabalhos de Fisher [8], Cochran [5] e Pearson [15], porém o termo metanálise foi utilizado pela primeira vez apenas em 1976 por Glass [9] ao definí-la como uma análise estatística de uma coleção de resultados provenientes de estudos individuais, com o propósito de complementar o que já havia sido descrito.

Segundo Martinez [13], a metanálise tem sido amplamente utilizada em estu-

dos médicos, especialmente em revisões sistemáticas de ensaios clínicos aleatorizados. Para Glass [9], metanálise é uma análise de análises, ou seja, uma análise estatística que visa combinar resultados já encontrados em análises anteriores de diferentes estudos de mesmo interesse. Atallah [2] acredita que uma revisão sistemática bem orientada por profissionais experientes não leva menos de 3 meses, podendo ultrapassar o período de 1 ano, sendo praticamente inviável de ser realizada apenas por um pesquisador. Assim, para que se tenha uma metanálise de qualidade, é importante que os estudos a serem combinados tenham sido avaliados com cautela para que haja confiança na associação.

Existem dois tipos de metanálise em toda a literatura: Metanálise baseada na Literatura e Metanálise Caso a Caso.

DerSimonian e Laird [6] definem metanálise como uma análise estatística de uma coleção de resultados analíticos com o propósito de integrar seus achados. Ao combinar resultados de muitos estudos deste tipo, tem-se como objetivo aumentar o poder das conclusões e contornar restrições inerentes aos estudos. De um modo geral, a metanálise definida anteriormente é o que chamamos de metanálise baseada na literatura. Nela, usualmente, considera-se apenas os resultados de cada estudo a fim de combiná-los por meio de algum método. Ou seja, extrai-se de cada estudo uma medida de efeito, combinando-se e encontrando uma única medida metanalítica que, geralmente, nada mais é do que uma média ponderada das medidas de efeito de cada estudo.

Entretanto, há diversas limitações no uso desse tipo de metanálise, tais quais: estudos publicados podem utilizar métodos diferentes, criando dificuldades na combinação dos resultados; diferenças no planejamento do estudo podem dificultar a justificativa do uso da técnica; entre outras. Assim, de forma alternativa à metanálise baseada na literatura, tem-se a metanálise caso a caso, que trabalha com os dados individuais de cada estudo, isto é, ao invés de trabalhar com os resultados dos estudos publicados, os dados originais de cada estudo são solicitados diretamente aos pesquisadores responsáveis pelo estudo em questão, são reanalisados e, se apropriados, combinados na metanálise.

De acordo com Dutton [7], apesar dessa abordagem levar mais tempo que a metanálise convencional, ela oferece benefícios quanto a qualidade dos dados e

com os tipos de análises que podem ser realizadas. Dutton acredita ainda que a metanálise caso a caso deveria ser utilizada em todas as situações, pois possibilita todos os tipos de análise dos dados fornecendo resultados mais precisos e, por isso, esta metanálise é considerada o *padrão ouro de metanálise*. Contudo, ainda assim, existem circunstâncias em que não se pode ter acesso aos dados dos estudos em questão por motivos simples, como a falta de colaboração ou perda de dados.

Graficamente, os resultados de uma metanálise usual são apresentados em um gráfico chamado Forest Plot. Essa representação é elaborada de modo a comparar os efeitos de tratamentos em estudos quantitativos, especialmente do tipo ensaios clínicos controlados e randomizados. O termo "forest" vem da ideia de que o gráfico se assemelha a uma floresta de linhas. Nos exemplos dos próximos capítulos há exemplos de gráficos do tipo Forest Plot.

Por fim é importante ressaltar que para que se tenha uma metanálise de qualidade é importante que os estudos a serem combinados tenham sido avaliados com cautela para que haja confiança na união. Dessa forma, é preciso que critérios de seleção de estudos sejam estabelecidos [10].

Capítulo 3

Metodologia

Neste capítulo, apresenta-se todas as metodologias necessárias para o desenvolvimento de uma metanálise e desenvolve-se a construção de um modelo Bayesiano hierárquico. O objetivo principal é apresentar uma medida metanalítica Bayesiana baseada na mistura das distribuições a posteriori do parâmetro de interesse de cada estudo pertencente à metanálise, isto é, esta medida metanalítica proposta é uma distribuição de probabilidade da quantidade de interesse e não apenas uma simples medida-resumo da mesma, o que acarreta melhor conclusão do problema.

Alguns exemplos são desenvolvidos e discutidos para maior entendimento dos métodos propostos.

3.1 Metanálise via Modelo Bayesiano Hierárquico

Nesta seção aborda-se a metanálise via modelo Bayesiano hierárquico. Para isso, define-se primeiramente o modelo Bayesiano hierárquico e, na sequência, apresenta-se a metanálise via este procedimento.

A inferência Bayesiana baseia-se na especificação de um modelo probabilístico para os dados observados X , dado um vetor de parâmetros desconhecidos θ , tendo então a função de verossimilhança $L(\theta|X)$. Assume-se que θ é aleatório, com função a priori denotada por $\pi(\theta)$. Pelo teorema de Bayes, a inferência sobre θ é baseada na distribuição a posteriori. A distribuição a posteriori de θ é dada por:

$$\pi(\theta|X) = \frac{L(\theta|X)\pi(\theta)}{\int L(\theta|X)\pi(\theta)d\theta}, \quad (3.1)$$

em que Θ denota o espaço paramétrico de θ .

De (3.1), $\pi(\theta|X)$ é proporcional à multiplicação da função de verossimilhança e da priori, envolvendo a contribuição dos dados observados por meio de $L(\theta|X)$ e a contribuição da informação a priori quantificada por $\pi(\theta)$. Assim, a posteriori é:

$$\pi(\theta|X) \propto L(\theta|X)\pi(\theta). \quad (3.2)$$

Quando utilizamos a abordagem Bayesiana de forma hierárquica, estamos especificando a distribuição a priori do parâmetro em etapas.

Considerando uma metanálise composta por J estudos, temos um parâmetro geral (θ), e os parâmetros para cada um dos estudos. A Figura 3.1 ilustra como é o esquema hierárquico:

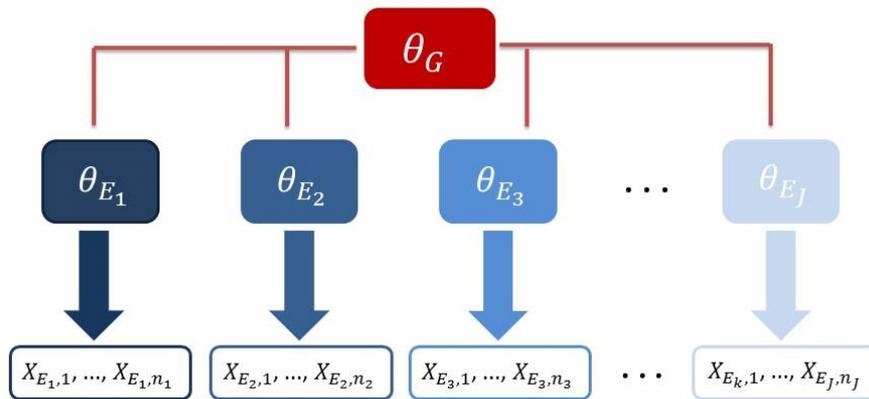


Figura 3.1: *Esquema Bayesiano Hierárquico.*

Dessa forma a priori para θ é dada por:

$$\begin{aligned}\pi(\theta) &= \pi(\theta_G)\pi(\theta_{E_1}, \dots, \theta_{E_J}|\theta_G) \\ &= \pi(\theta_G) \prod_{j=1}^J \pi(\theta_{E_j}|\theta_G),\end{aligned}\tag{3.3}$$

em que:

θ_G é o parâmetro geral;

$\theta_{E_1}, \dots, \theta_{E_J}$ são os parâmetros de cada um dos $j = 1, \dots, J$ estudos;

Da figura (3.1):

X_{E_j, n_j} são as observações dos $j = 1, \dots, J$ estudos;

n_j é o tamanho do j -ésimo estudo.

A seguir, exemplifica-se o uso de modelos Bayesianos hierárquicos por meio de dados simulados.

Exemplo 3.1 *Suponha dois estudos sobre o mesmo assunto em que a variável de interesse, em cada um deles, seja normalmente distribuída, com média μ e variância σ^2 , onde os parâmetros estão também relacionados à mesma característica. Os dados dos estudos foram gerados de uma distribuição normal, em que as observações do Estudo 1 segue de uma distribuição normal com média $\mu_1 = -3$ e variância $\sigma_1^2 = 15$, $X \sim N(-3, 15)$, e as observações do Estudo 2 de uma distribuição normal com média $\mu_2 = 3$ e variância $\sigma_2^2 = 12$, $Y \sim N(3, 12)$. O tamanho amostral de ambos os estudos é igual a 30 ($n_x = n_y = 30$).*

Geradas as amostras X e Y , temos que a média amostral e a variância do estudo 1 são respectivamente, $\bar{x} = \sum_{i=1}^{n_x} \frac{x_i}{n_x} = -3,18$ e $s_x^2 = \sum_{i=1}^{n_x} \frac{(x_i - \bar{x})^2}{n_x - 1} = 14,43$, e para o estudo 2, $\bar{y} = \sum_{i=1}^{n_y} \frac{y_i}{n_y} = 3,61$ e $s_y^2 = \sum_{i=1}^{n_y} \frac{(y_i - \bar{y})^2}{n_y - 1} = 8,36$.

Como priori, deve-se estabelecer os níveis de hierarquia dos parâmetros. Das informações dadas acima, temos:

\implies **1º nível hierárquico é dado por:**

$\mu_i \sim N(\mu, V_\mu)$ e $\sigma_i \sim \text{Gama}(a, b)$, em que $i = 1, 2$.

Porém, é necessário agora assumir uma distribuição de probabilidades para μ , V_μ , σ e V_σ , uma vez que, na abordagem Bayesiana, os parâmetros a serem estimados são variáveis aleatórias.

⇒ **2º nível hierárquico** é dado por:

$$\mu \sim N(0, 3), V_\mu \sim \text{Gama}(3, 1), \sigma \sim \text{Gama}((14^2)/3, 14/3), V_\sigma \sim \text{Gama}(3, 1).$$

Portanto, as priors para os parâmetros podem ser expressas da seguinte maneira:

$$\begin{aligned} \pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \mu, \sigma, V_\mu, V_\sigma) &= \pi(\mu_1|\mu, V_\mu)\pi(\mu_2|\mu, V_\mu)\pi(\sigma_1|\sigma, V_\sigma)\pi(\sigma_2|\sigma, V_\sigma) \\ &\pi(\mu)\pi(\sigma)\pi(V_\mu)\pi(V_\sigma). \end{aligned} \quad (3.4)$$

Logo, a posteriori de $(\mu_1, \mu_2, \sigma_1, \sigma_2, \mu, \sigma, V_\mu, V_\sigma)$ dado X, Y é

$$\begin{aligned} \pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \mu, \sigma, V_\mu, V_\sigma|X, Y) &\propto \pi(\mu_1, \mu_2, \sigma_1, \sigma_2, \mu, \sigma, V_\mu, V_\sigma) \\ &L(\mu_1, \mu_2, \sigma_1, \sigma_2, \mu, \sigma, V_\mu, V_\sigma|X, Y). \end{aligned} \quad (3.5)$$

Na Figura 3.2 apresenta-se a curva de nível da distribuição a posteriori conjunta de μ e σ e na Figura 3.3 suas marginais. Originalmente sabemos que os dados dos estudos foram simulados com médias totalmente opostas, -3 e 3 , e variâncias grandes. O método Bayesiano hierárquico considera que os parâmetros dos estudos pertençam a uma mesma população. Com isso, é possível notar na figura que o modelo hierárquico não consegue distinguir essa grande diferença entre as médias dos estudos, mostrando uma média em torno de zero, o que traz uma evidência de inverdade como conclusão geral dos estudos.

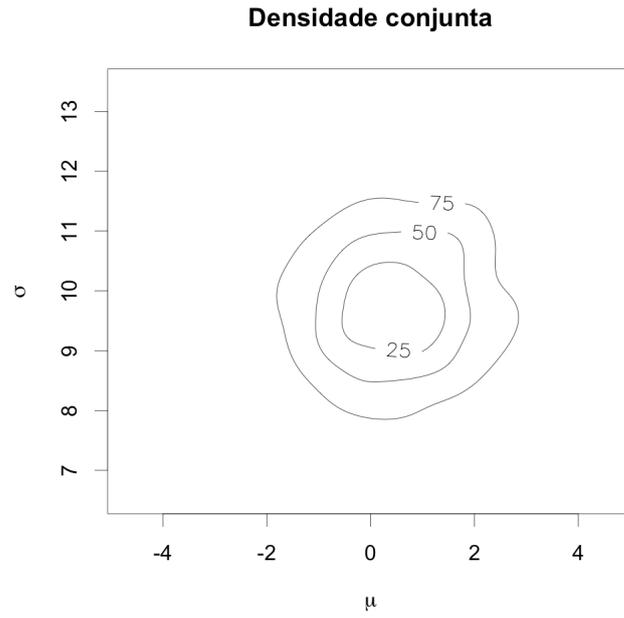


Figura 3.2: *Curvas de nível a posteriori de (μ, σ) da metanálise.*

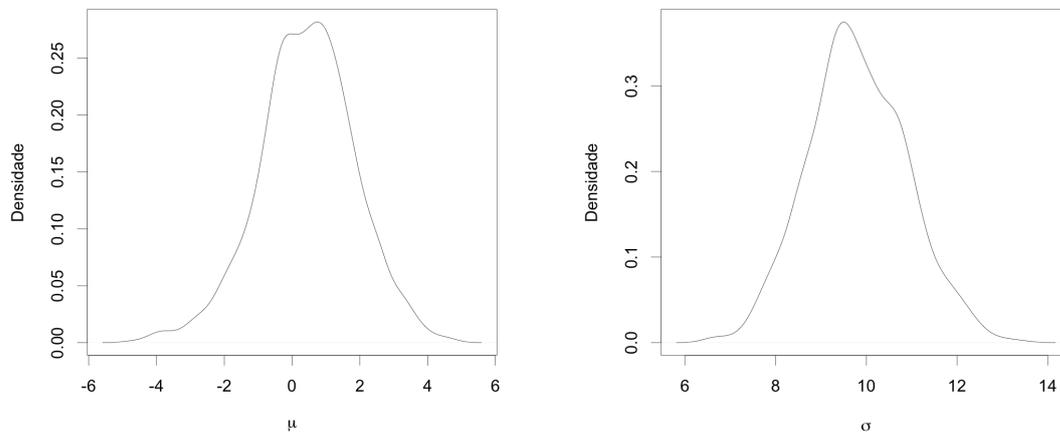


Figura 3.3: *Distribuições a posteriori marginal de μ e σ .*

3.2 Medida Metanalítica

Considere uma metanálise de J estudos independentes que investigam certa característica de interesse θ . Para o j -ésimo estudo ($j = 1, \dots, J$), suponha que $\mathbf{X}_j = (X_{j1}, X_{j2}, \dots, X_{jn_j})^t$ denota um vetor de variáveis aleatórias de tamanho n_j , com função densidade de probabilidade $f(x_j|\theta)$ no espaço de probabilidade $(\mathcal{X}, \mathcal{F}, \mathcal{P})$. Considerando $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{n_jj})$ o vetor observado de X_j , tem-se que a função de verossimilhança para o j -ésimo estudo é dada por: $L_j(\theta|\mathbf{x}_j) = \prod_{i=1}^{n_j} f(x_{ij}; \theta)$.

Admitindo a metodologia Bayesiana, tem-se a função a priori de θ dada por $\pi(\theta)$ e a distribuição a posteriori de θ sendo $\pi_j(\theta|\mathbf{x}_j) \propto L_j(\theta|\mathbf{x}_j)\pi(\theta)$. Em geral a função a priori é a mesma para todos os estudos pois independe de onde o experimento foi conduzido. Então, a medida metanalítica proposta é a mistura das J distribuições a posteriori de θ , produzindo uma única distribuição para θ , que é chamada de distribuição a posteriori metanalítica de θ . A mistura de densidades é uma técnica utilizada em alguns casos para obter conclusões a respeito de θ [16].

Portanto, a distribuição a posteriori metanalítica de θ é dada por:

$$\pi_M(\theta|\mathbf{x}) = \sum_{j=1}^J \omega_j \pi_j(\theta|\mathbf{x}_j), \quad (3.6)$$

em que ω_j é o peso do j -ésimo estudo. Estes pesos são fixos e podem representar a importância dos estudos, o tamanho amostral, etc. Inicialmente, considere $\omega_j = \frac{n_j}{\sum_{i=1}^J n_j}$.

A média e a variância a posteriori metanalítica para θ são, respectivamente,

$$E_M(\theta|\mathbf{x}) = \sum_{i=1}^J \omega_j E_j[\theta|\mathbf{x}_j] \quad e \quad (3.7)$$

$$Var_M(\theta|\mathbf{x}) = \sum_{i=1}^J \omega_j^2 Var_j[\theta|\mathbf{x}_j]. \quad (3.8)$$

A medida metanalítica proposta é uma medida geral, isto é, qualquer função a priori e qualquer função de verossimilhança podem ser utilizadas em (3.6). Para

que se tenha uma metanálise de qualidade é importante esclarecer que, além de uma técnica estatística apropriada, é fundamental que os estudos a serem combinados tenham sido avaliados com cuidado para que haja confiança em sua associação.

Vale ressaltar que o método estudado neste trabalho não discute a questão de heterogeneidade ou homogeneidade entre os estudos, uma vez que sempre se leva em consideração a variabilidade existente entre os estudos incluídos na metanálise.

A seguir aborda-se um exemplo do método proposto.

Exemplo 3.2 *Considere o mesmo cenário do Exemplo 3.1, onde a variável de interesse é normalmente distribuída com média μ e variância σ^2 . Os dados do estudo 1 foram gerados de uma distribuição Normal com média -3 e variância 15 , $X \sim \text{Normal}(-3, 15)$, e do estudo 2, de uma distribuição Normal com média 3 e variância 12 , $Y \sim \text{Normal}(3, 12)$. O tamanho amostral de ambos os estudos é igual a 30 ($n_x = n_y = 30$). A média e a variância amostral do estudo 1 são $\bar{x} = \sum_{i=1}^{n_x} \frac{x_i}{n_x} = -3,18$ e $s_x^2 = \sum_{i=1}^{n_x} \frac{(x_i - \bar{x})^2}{n_x - 1} = 14,44$ e do estudo 2 são, respectivamente, $\bar{y} = \sum_{i=1}^{n_y} \frac{y_i}{n_y} = 3,62$ e $s_y^2 = \sum_{i=1}^{n_y} \frac{(y_i - \bar{y})^2}{n_y - 1} = 8,37$*

Dispondo das informações acima, a função de verossimilhança para os estudos 1 e 2 são, respectivamente,

$$L_x(\mu, \sigma^2 | \mathbf{x}) \propto (\sigma^2)^{-n_x/2} \exp \left\{ -\frac{1}{2\sigma^2} [(n_x - 1)s_x^2 + n_x(\mu - \bar{x})^2] \right\} \quad e \quad (3.9)$$

$$L_y(\mu, \sigma^2 | \mathbf{y}) \propto (\sigma^2)^{-n_y/2} \exp \left\{ -\frac{1}{2\sigma^2} [(n_y - 1)s_y^2 + n_y(\mu - \bar{y})^2] \right\}. \quad (3.10)$$

Admitindo a priori de Jeffreys para $(\mu; \sigma^2)$, ou seja, $\pi(\mu; \sigma^2) \propto (\sigma^2)^{-1}$, e considerando a função de verossimilhança de (3.9) e (3.10), as distribuições conjuntas a posteriori de μ e σ^2 em ambos os estudos seguem distribuição Normal-Gama-Inversa, dadas por:

$$\pi(\mu, \sigma^2 | \mathbf{X}) = \frac{\sqrt{n_x}}{\sigma\sqrt{2\pi}} \frac{[(n_x - 1)s_x^2/2]^{n_x/2}}{\Gamma(n_x/2)} \left(\frac{1}{\sigma^2}\right)^{\frac{n_x+1}{2}+1} \exp\left\{-\frac{(n_x - 1)s_x^2 + n_x(\mu - \bar{x})^2}{2\sigma^2}\right\} \quad e \quad (3.11)$$

$$\pi(\mu, \sigma^2 | \mathbf{Y}) = \frac{\sqrt{n_y}}{\sigma\sqrt{2\pi}} \frac{[(n_y - 1)s_y^2/2]^{n_y/2}}{\Gamma(n_y/2)} \left(\frac{1}{\sigma^2}\right)^{\frac{n_y+1}{2}+1} \exp\left\{-\frac{(n_y - 1)s_y^2 + n_y(\mu - \bar{y})^2}{2\sigma^2}\right\}, \quad (3.12)$$

ou seja, $(\mu, \sigma^2 | \mathbf{X}) \sim NIG(\bar{x}, n_x, \frac{n_x-1}{2}, \frac{(n_x-1)s_x^2}{2})$ e $(\mu, \sigma^2 | \mathbf{Y}) \sim NIG(\bar{y}, n_y, \frac{n_y-1}{2}, \frac{(n_y-1)s_y^2}{2})$.

Os valores esperados a posteriori e a matriz de variância-covariância a posteriori de μ e σ^2 para (3.11) são, respectivamente:

$$E[\mu, \sigma^2 | \mathbf{X}] = \begin{pmatrix} \bar{x} \\ \frac{(n_x-1)s_x^2}{(n_x-3)} \end{pmatrix}, n_x > 3 \quad e \quad (3.13)$$

$$Var[\mu, \sigma^2 | \mathbf{X}] = \begin{pmatrix} \frac{(n_x-1)s_x^2}{(n_x-3)n_x} & 0 \\ 0 & \frac{(n_x-1)2s_x^2}{(n_x-3)^2(n_x-5)} \end{pmatrix}, n_x > 5. \quad (3.14)$$

Para (3.12) segue análogo.

Utilizando (3.11) e (3.12), para obter a distribuição a posteriori metanalítica de μ e σ^2 , sugere-se misturar as distribuições a posteriori de $(\mu; \sigma^2)$ associadas a cada estudo, conforme 3.6. Dessa forma,

$$\pi_M(\mu, \sigma^2 | \mathbf{X}, \mathbf{Y}) = \frac{n_x \pi(\mu, \sigma^2 | \mathbf{X}) + n_y \pi(\mu, \sigma^2 | \mathbf{Y})}{n_x + n_y}, \quad (3.15)$$

sendo aqui considerados como pesos os tamanhos amostrais n_x referente ao estudo 1 e n_y ao estudo 2.

Na Figura 3.4(a), estão as curvas de nível das distribuições a posteriori conjunta de (μ, σ^2) para cada estudo, permitindo observar os gráficos tridimensionais agora em apenas duas dimensões. Observa-se que a diferença entre as médias a posteriori parece grande. Na Figura 3.4(b), estão as curvas de nível da distribuição a posteriori metanalítica conjunta de (μ, σ^2) que é dada pela mistura das distribuições a posteriori de (μ, σ^2) associadas aos dois estudos em questão e, na Figura 3.4(c), está o gráfico tridimensional da distribuição a posteriori metanalítica dada em (3.15).

Portanto, podemos concluir, pela observação dos gráficos acima, que a medida metanalítica proposta é uma distribuição de (μ, σ^2) e não apenas uma simples medida-resumo. Além disso, neste caso a medida metanalítica de (μ, σ^2) é uma distribuição bimodal, o que sugere a existência de dois grupos de estudos distintos, sendo necessário uma melhor investigação do problema.

Na Figura (3.4)d encontra-se a estimação por região HPD (Highest Posterior Density) com 95% de credibilidade, ou seja, a menor região que contém o verdadeiro valor de (μ, σ^2) , com probabilidade maior ou igual a 95%. Em outras palavras, as regiões de maior densidade nos gráficos são aquelas mais prováveis aos "verdadeiros" parâmetros.

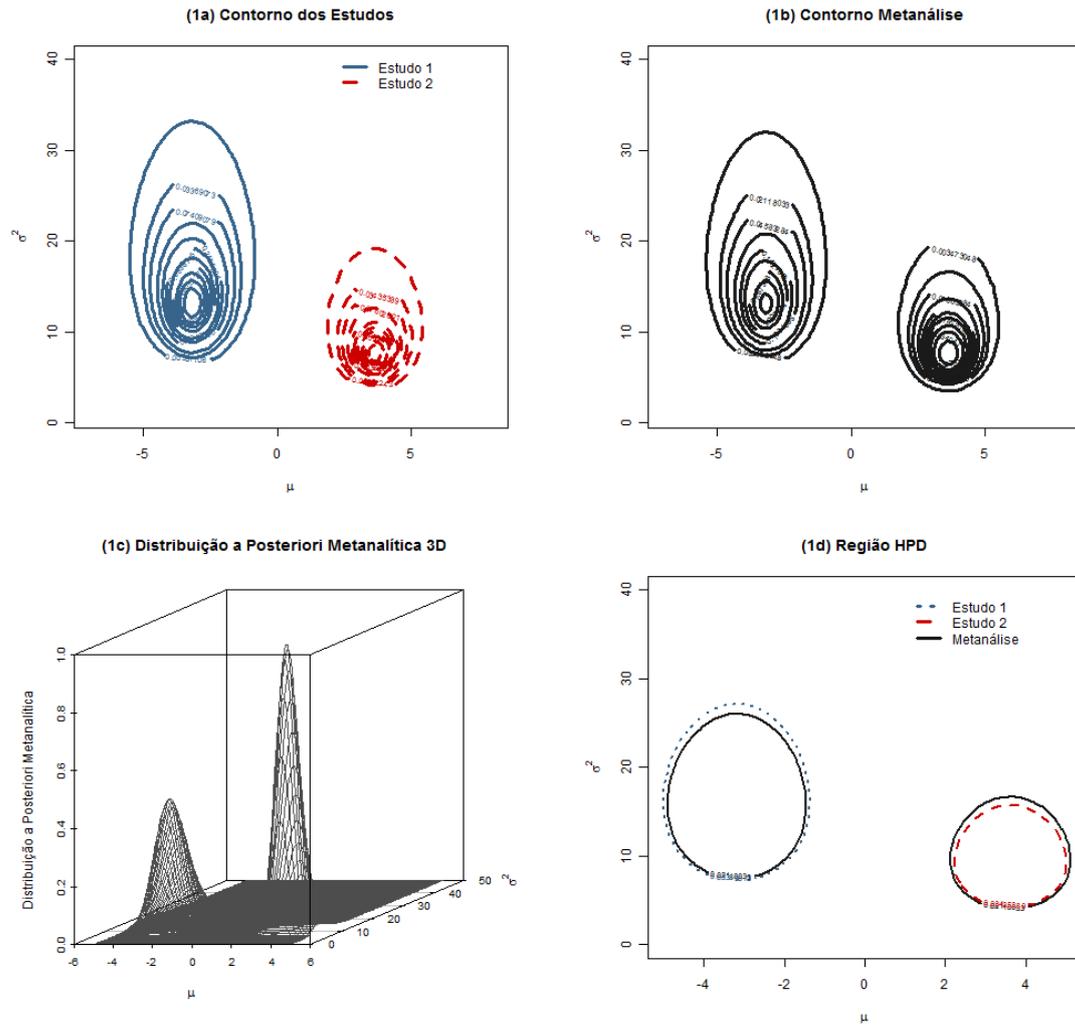


Figura 3.4: (a) Curvas de nível da distribuição conjunta a posteriori de (μ, σ^2) em cada estudo; (b) Curvas de nível da distribuição a posteriori metanalítica de (μ, σ^2) ; (c) Gráfico tridimensional da distribuição a posteriori metanalítica de (μ, σ^2) ; (d) Região HPD de 95% de credibilidade para (μ, σ^2) .

Vale ressaltar, neste exemplo, que a distribuição a posteriori conjunta de μ e σ^2 depende dos dados apenas por meio das estatísticas suficientes $(n_x, \bar{x}, \sigma_x^2)$ e $(n_y, \bar{y}, \sigma_y^2)$. Neste casos, quando utilizamos a distribuição metanalítica proposta, temos que a evidência fornecida pela estatística suficiente é idêntica à fornecida pelos dados. Dessa forma, realizar uma metanálise baseada na literatura é equivalente a realizar uma metanálise caso a caso, sem perda de qualidade.

3.2.1 Abordagem Frequentista da Metanálise

O método mais usual de metanálise, método de DerSimonian e Laird, também conhecido como Metanálise baseada na Literatura, baseia-se numa média ponderada das medidas de efeito de cada um dos estudos como descrito em (3.16).

Trata-se de um método em que há a possibilidade de estimar medidas metanalíticas sem a necessidade de pressupor que os estudos que compõem a metanálise são homogêneos.

Considerando que $\hat{\theta}_j$ a medida de efeito para j -ésimo estudo, a medida metanalítica θ_{MDL} dada por DerSimonian e Laird é estimada por:

$$\hat{\theta}_{MDL} = \frac{\sum_{j=1}^J \omega_j^* \hat{\theta}_j}{\sum_{j=1}^J \omega_j^*}, \quad (3.16)$$

em que:

O peso para o j -ésimo estudo é dado por: $\omega_j^* = \frac{1}{\tau^2 + \omega_j^{-1}}$,

com $\omega_j = \frac{1}{\sigma_j^2}$; $\hat{\tau}^2 = \frac{Q - (J-1)}{\sum_{j=1}^J \omega_j - \frac{\sum_{j=1}^J \omega_j^2}{\sum_{j=1}^J \omega_j}}$; $Q = \sum_{j=1}^J \omega_j (\hat{\theta}_j - \hat{\theta}_M)^2$;

$\hat{\theta}_M = \frac{\sum \omega_j \hat{\theta}_j}{\sum \omega_j}$; $\hat{\theta}_j$ é a medida de efeito do j -ésimo estudo; $j = 1, \dots, k$; τ é a variabilidade estimada entre os estudos.

Exemplo 3.3 Continuação Exemplo 3.2 com Abordagem Frequentista

Se a medida metanalítica fosse calculada pelo método de DerSimonian e Laird, baseado em (3.16), o método seria uma média ponderada das medidas de efeito dos estudos e este valor seria igual a 0,71.

A medida metanalítica pelo método tem intervalo de 95% de confiança igual a $[-5,91 ; 7,31]$. Os resultados dessa metanálise baseada na literatura estão resumidos no gráfico Forest Plot, Figura 3.5.

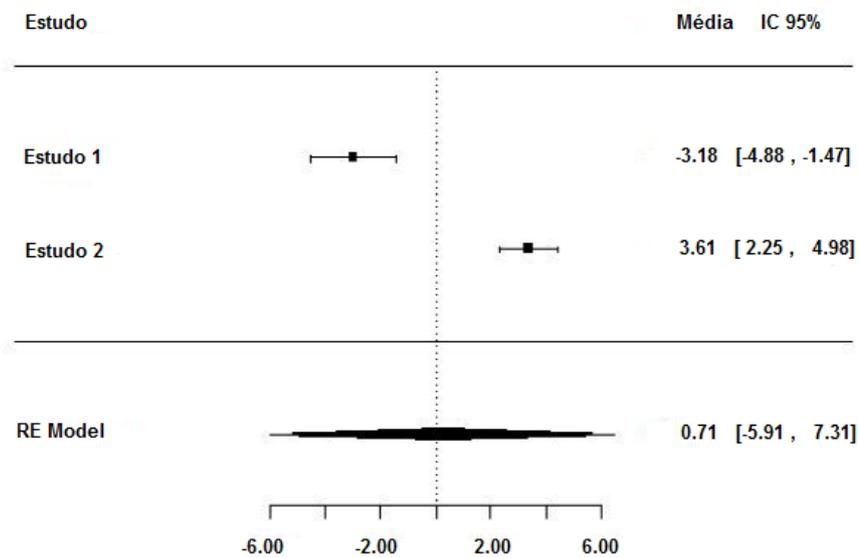


Figura 3.5: Forest Plot para os dados do Exemplo 3.2

3.2.2 Comparação entre os métodos

A fim de aproximar o método frequentista e o método proposto, para eventuais comparações, é interessante encontrar a distribuição a posteriori metanalítica apenas em função de μ .

As distribuições a posteriori marginal de μ , nos estudos 1 e 2, são $\mu|\mathbf{x} \sim$

$t_{n_x-1}(\bar{x}, \hat{\sigma}^2/n_x)$ e $\mu|\mathbf{y} \sim t_{n_y-1}(\bar{y}, \hat{\sigma}^2/n_y)$. Para encontrar a medida metanalítica de μ utiliza-se a equação (3.15) e esta é apresentada na Figura 3.6(a).

É importante destacar que a distribuição a posteriori metanalítica marginal de μ também é uma distribuição bimodal com intervalo HPD de 95% de credibilidade dado pelos intervalos disjuntos $[-4, 62; -1, 73]$ e $[2, 43; 4, 79]$. Também pode-se observar que o valor encontrado pela abordagem frequentista pelo método de DerSimonian e Laird, 0,71, está fora dos intervalos HPD.

Podemos dessa forma concluir que, pelo fato da distribuição metanalítica ser bimodal, encontrar a média a posteriori de μ não traz muita informação. Por outro lado, a medida metanalítica com enfoque Bayesiano que foi proposta sugere a existência de dois grupos distintos, deixando claro que este resultado obtido permite conclusões mais amplas e mais informativas do que os resultados do método usual de metanálise.

Pela Figura 3.6(b) fica claro que o resultado obtido pelo método proposto permite conclusões mais amplas e traz mais informações quanto aos estudos envolvidos na metanálise.

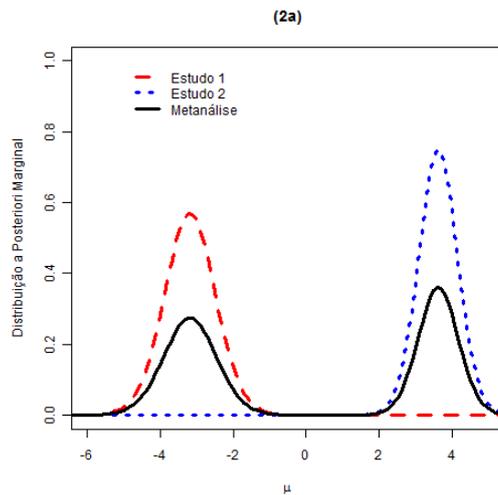


Figura 3.6: Distribuições a posteriori marginal de μ .

É importante deixar claro que o objetivo até este momento não é fazer comparação entre os métodos, até porque os métodos não são comparáveis, mas sim apontar possíveis falhas da metodologia usual de metanálise e alertar para uma possível forma de resolver estes problemas.

3.3 Aplicação Geral

Nesta seção, discute-se um exemplo apresentado por Berry [1989] que utilizou uma metanálise composta de nove estudos envolvendo a droga antidepressiva SAMe.

A variável aleatória de interesse X representa o sucesso no uso da droga antidepressiva, ou seja,

$$\begin{cases} X = 1 : \text{Sucesso no uso da droga com probabilidade } p, \\ X = 0 : \text{Não Sucesso no uso da droga com probabilidade } 1 - p. \end{cases}$$

Então, $X|p$ segue uma distribuição Bernoulli(p) e sua distribuição de probabilidade é dada por:

$$P(X = x|p) = p^x(1 - p)^{1-x}, x = 0, 1 \quad (3.17)$$

Assim, para uma amostra aleatória X_1, \dots, X_n de X , o número de tentativas que resulta em sucesso, $Y = \sum_{i=1}^n X_i$ segue uma distribuição Binomial(n, p).

Os dados da Tabela 3.1 contém os resultados para os nove estudos, cujo número de pacientes no j -ésimo estudo é n_j e o número observado de sucessos dado por y_j .

Tabela 3.1: Proporção de sucessos no uso da droga antidepressiva SAMe.

Estudo j	y_j	n_j	y_j/n_j
1	20	20	1,00
2	4	10	0,40
3	11	16	0,69
4	10	19	0,53
5	5	14	0,36
6	36	46	0,78
7	9	10	0,90
8	7	9	0,78
9	4	6	0,67
Total	106	150	0,71

Abordagem Frequentista

No Forest Plot (Figura 3.7) observa-se os resultados obtidos pela metanálise proposta por DerSimonian e Laird [6], cuja medida metanalítica encontrada é igual a 0,72 com intervalo de 95% de confiança, igual a [0,65 ; 0,79]. Portanto, a proporção combinada de sucessos no uso da droga é de 72%.

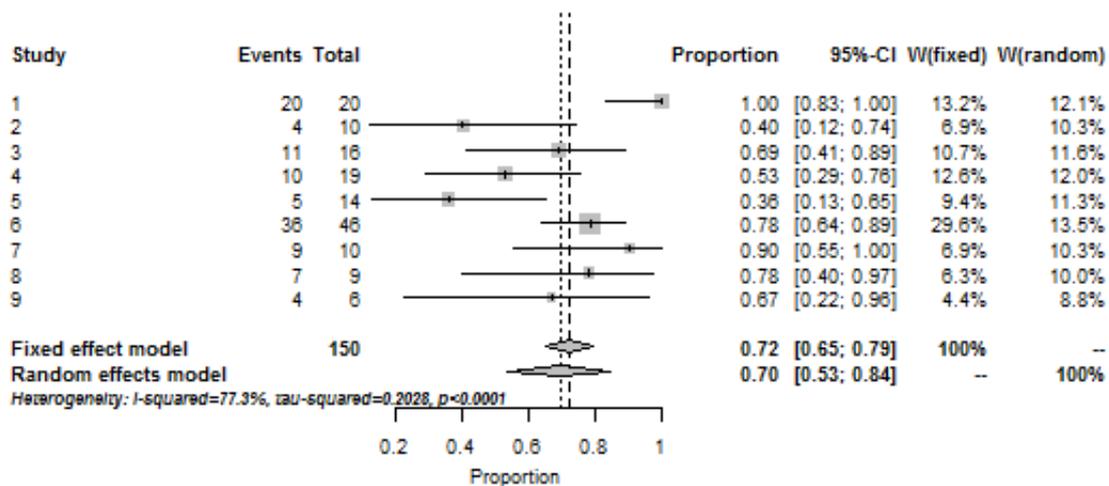


Figura 3.7: Forest Plot para proporção de sucessos no uso da droga SAMe.

Abordagem Bayesiana Hierárquica

Sob o enfoque Bayesiano hierárquico, considera-se que cada proporção de sucesso (parâmetro) em cada estudo seja selecionado de uma mesma população. Assim, para quantificarmos a informação sobre a população, precisamos de uma distribuição de probabilidade da distribuição populacional. (Berry [3])

Dessa forma, a variável aleatória Y_j é o número de sucessos entre os n_j pacientes do estudo j , tal que, Y_j segue uma distribuição Binomial(n_j, p_j) para $j = 1, \dots, 9$, com n_j constante e conhecido e p_j o parâmetro a ser estimado. Então, a função de verossimilhança para $\mathbf{p} = (p_1, \dots, p_9)$ é dada por:

$$L(\mathbf{p}|\mathbf{y}) \propto \prod_{j=1}^9 p_j^{y_j} (1 - p_j)^{n_j - y_j}. \quad (3.18)$$

Suponha que p_1, \dots, p_9 seja uma amostra aleatória da distribuição F , que é aleatória e assumida como uma Beta(a, b) a priori, onde a e b são desconhecidos, temos a distribuição a priori de p_j (de 1º nível) é:

$$\pi(p_j|a, b) = B(a, b)^{-1} p_j^{a-1} (1 - p_j)^{b-1}, \quad (3.19)$$

em que $a > 0$ e $b > 0$ e $B(a; b) = [\int_0^1 p_j^{a-1} (1 - p_j)^{b-1}]^{-1}$.

Em outras palavras, F é a distribuição geradora dos p 's.

Considera-se (3.19) como a priori de 1º nível. Assume-se no 2º nível uma distribuição conjunta de probabilidade para a e b , dada por $\pi(a, b)$. Neste exemplo, será considerada uma priori Imprópria (vaga):

$$\pi(a, b) \propto 1. \quad (3.20)$$

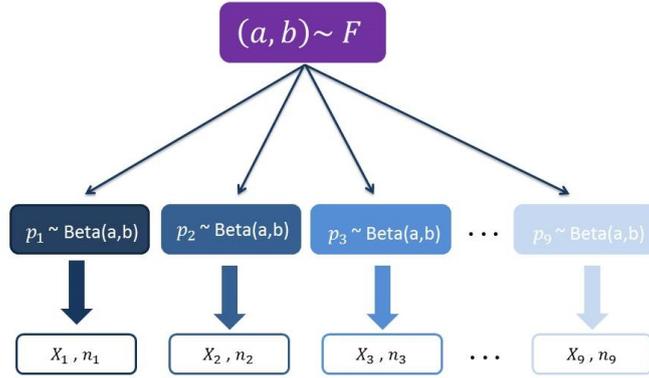


Figura 3.8: *Esquema Bayesiano Hierárquico Aplicação Geral.*

Logo, a distribuição a posteriori (própria) de (a, b, p) dado y é da forma:

$$\begin{aligned} \pi(a, b, \mathbf{p} | \mathbf{y}) &\propto \pi(a, b) \prod_{j=1}^9 \pi(p_j | a, b) L(p_j | \mathbf{y}) \\ &\propto \prod_{j=1}^9 p_j^{y_j} (1 - p_j)^{n_j - y_j} B(a, b)^{-1} p_j^{a-1} (1 - p_j)^{b-1}. \end{aligned} \quad (3.21)$$

A Figura 3.9 mostra a densidade estimada (F) a posteriori mostra média igual a 0,67, moda igual a 0,75 e região HPD de 95% de credibilidade igual a [0,35; 0,99].

Observa-se uma grande massa em torno de 75% que decai nas extremidades. Ou seja, a importância do Estudo 1 (100% de sucesso) é minimizada no modelo hierárquico Bayesiano.

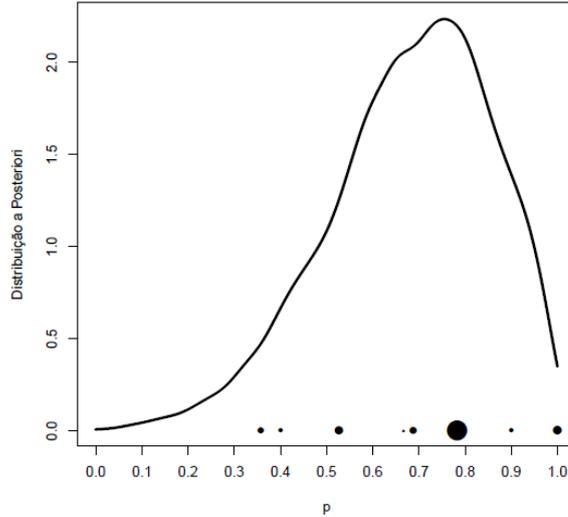


Figura 3.9: Média a posteriori de F , geradora da proporção dos sucessos. É a densidade média com respeito a distribuição a posteriori de (a, b) .

Abordagem Bayesiana Medida Metanalítica

A medida metanalítica é calculada pela metodologia proposta em (3.6). Então, para o estudo j , com $(j = 1, 2, \dots, 9)$, suponha que Y_j são variáveis aleatórias, condicionalmente independentes dado p , com distribuição Binomial(n_j, p). A função de verossimilhança para o j -ésimo estudo é

$$L_j(\mathbf{y}_j|p) \propto p^{y_j}(1-p)^{n_j-y_j}, \quad (3.22)$$

Considerando que a distribuição a priori de p é Uniforme $(0, 1)$ e a função de verossimilhança é dada em (3.22), a distribuição a posteriori de p , para o j -ésimo estudo, é dada por:

$$\pi_j(p|\mathbf{y}_j) = B(n_j + 2, y_j + 1)p^{y_j}(1-p)^{n_j-y_j}, \quad (3.23)$$

ou seja, $p|\mathbf{y}_j \sim \text{Beta}(y_j + 1, n_j - y_j + 1)$, $j = 1, \dots, 9$. Encontrada a distribuição a posteriori de p para cada estudo, o próximo passo será a obtenção da

distribuição metanalítica de p .

Com base na Equação (3.6), considerando $\omega_j = \frac{n_j}{n}$ com $n = \sum_{j=1}^9 n_j$, a distribuição a posteriori metanalítica de p , $\pi_M(p|\mathbf{y})$, é

$$\pi_M(p|\mathbf{y}) = \sum_{j=1}^9 \frac{\Gamma(n_j + 2)}{\Gamma(y_j + 1)\Gamma(n_j - y_j + 1)} p^{y_j} (1 - p)^{n_j - y_j} \omega_j. \quad (3.24)$$

Na Tabela 3.2 encontram-se as principais medidas descritivas da distribuição metanalítica. O intervalo HPD de 95% de probabilidade é de $[0,31; 1,00]$.

Tabela 3.2: Exemplo Berry: Medidas descritivas à posteriori.

Medida	Mediana	Moda	Média
Valor	0,72	0,78	0,69

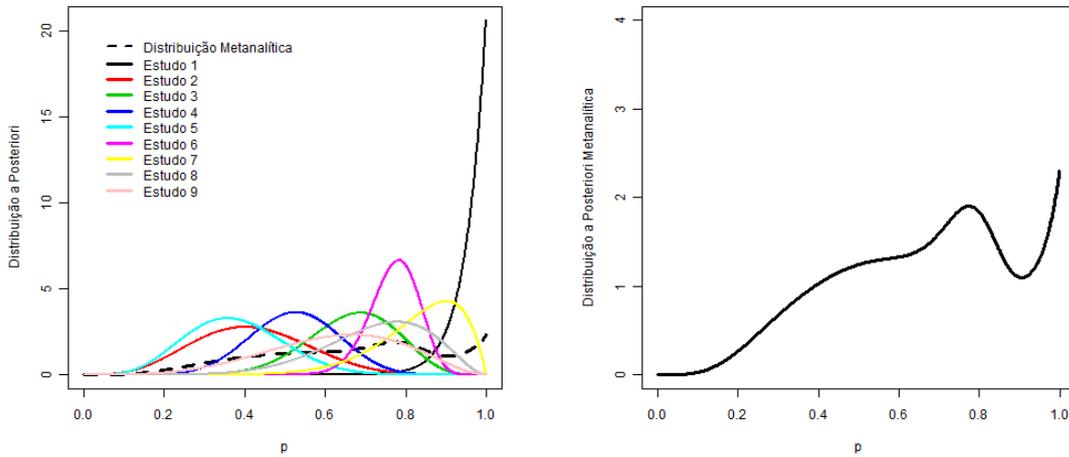


Figura 3.10: (a) Distribuições a posteriori de p para os nove estudos. (b) Distribuição a posteriori metanalítica de p .

Na Figura 3.10(a) estão as distribuições a posteriori de p para cada estudo pertencente à metanálise e na Figura 3.10(b) está a distribuição metanalítica de p , dada pela mistura das distribuições a posteriori de p dos nove estudos em questão. Nota-se que o comportamento da distribuição a posteriori metanalítica de p não

é simétrico, sugerindo a existência de uma distribuição trimodal, sendo os valores 0.78, 0.50, 1, os máximos locais.

Discussão

As metodologias propostas não podem ser comparadas por diferirem conceitualmente. Portanto, nenhum tipo de métrica será usado para comparar os métodos.

Ao se observar apenas os dados da Tabela 3.1, nota-se que, provavelmente, existe um efeito no uso da droga SAME e que este efeito, aparentemente, difere em alguns estudos.

Há estudos com percentual de sucesso no uso da droga muito alto, como no Estudo 7 com 90% e no Estudo 1 com 100% de sucesso e alguns estudos com baixo percentual de sucesso no uso da mesma droga, como o caso do Estudo 5 com apenas 36%. Portanto, é nítido que a proporção de sucesso difere entre os estudos e a análise de heterogeneidade entre os estudos já não faz sentido. É importante ressaltar que o método proposto vai além de observar a média ponderada dos estudos pertencentes à metanálise, como é o caso do método de DerSimonian e Laird.

Neste exemplo, a metodologia proposta permite observar o comportamento do parâmetro de interesse através de sua distribuição a posteriori metanalítica. Também não há preocupação com a existência ou não da heterogeneidade entre os estudos, uma vez que esta já está incorporada neste modelo, diferentemente dos métodos usuais.

Capítulo 4

Metanálise em Modelos de Regressão

Neste capítulo, apresenta-se uma Metanálise em Modelos de Regressão. O objetivo principal é apresentar a medida metanalítica Bayesiana baseada na mistura das distribuições a posteriori do parâmetro de interesse de cada modelo de regressão correspondente aos estudo pertencente à metanálise.

4.1 Introdução

Sabe-se que um modelo linear clássico tem a seguinte forma:

$$Y = \mu + \epsilon,$$

em que:

$Y_{n \times 1}$ é o vetor de variáveis resposta;

$\mu = E(Y) = X\beta$ componente sistemático;

\mathbf{X} é uma matriz $n \times p$ de variáveis explicativas;

$\beta = (\beta_0, \beta_1, \dots, \beta_p)^t$ vetor de parâmetros;

$\epsilon = (\epsilon_1, \dots, \epsilon_n)^t$: erro aleatório, com $\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$; $i = 1, \dots, n$

Contudo, em muitos casos, a estrutura aditiva do componente sistemático e do componente aleatório não é satisfeita. Além disso, não há razão para restringir a estrutura simples dada por $\mu = E(Y) = X\beta$ no caso do componente sistemático e nem restringir a distribuição normal para o componente aleatório.

Segundo Paula [14], uma proposta mais interessante e inovadora sobre modelos lineares foi apresentada por Nelder e Wedderburn (1972). Eles propuseram os modelos lineares generalizados (MLGs). A ideia básica consiste em ter mais opções para a distribuição da variável resposta, bem como dar maior flexibilidade para a relação funcional entre a média da variável resposta e o preditor linear η . Ou seja, define-se em MLG uma distribuição para a variável resposta que representa os dados e não uma distribuição para o erro aleatório.

Para uma amostra formada por n observações, o modelo linear generalizado envolve os três componentes:

- i) Componente aleatório: representado por um conjunto de variáveis aleatórias independentes provenientes de uma mesma distribuição que faz parte da família exponencial na forma canônica;
- ii) Componente sistemático: as variáveis explicativas entram na forma de uma soma linear de seus efeitos: $\eta = \mathbf{X}\beta$, em que:
 $\mathbf{X} = (x_1, \dots, x_n)^t$ a matriz do modelo, $\beta = (\beta_1, \dots, \beta_p)^t$ o vetor de parâmetros;
- iii) Função de ligação: uma função que liga o componente aleatório ao componente sistemático, ou seja, relaciona a média ao preditor linear: $\eta_i = g(\mu_i)$.

4.2 Metanálise em Modelos de regressão

Baseado no que se apresentou no Capítulo 3, pretende-se aqui aplicar a medida metanálise em um modelo de regressão.

Seja J o número de estudos da metanálise e Y_j o efeito observado no estudo j , $j = 1, 2, \dots, J$. O modelo de regressão que será explorado neste trabalho, é dado por:

$$Y_j = X_j\beta + \epsilon_j, \tag{4.1}$$

em que: $\epsilon_j \stackrel{iid}{\sim} N(0, \sigma^2 I)$;

$\epsilon_j = (\epsilon_{j1}, \dots, \epsilon_{jn})$: erro aleatório associado ao estudo j ;

$Y_j = (y_{j1}, y_{j2}, \dots, y_{jn})^t$ vetor de variáveis respostas correspondentes ao j – simo estudo com n indivíduos;

$$X_j \text{ matriz de covariáveis, } X_j = \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{pmatrix}; \quad e$$

$\beta_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jp})$: vetor de parâmetros da regressão.

A fim de exemplificar o uso da Metanálise em modelos de regressão, considere um simples exemplo apresentado a seguir.

Exemplo 4.1 *Suponha o seguintes modelo de regressão referentes a dois estudos, de tamanhos $n_1 = 100$ e $n_2 = 150$.*

Modelo Estudo 1: $Y_1 = \beta_0 + \beta_1 X_1 + \epsilon_1$; e

Modelo Estudo 2: $Y_2 = \beta_0 + \beta_1 X_2 + \epsilon_2$.

Para a geração dos dados neste exemplo, considerou-se no estudo 1 $\beta_0 = 1$, $\beta_1 = 0,5$, $\sigma = 0,2$ e $X_1 \sim Normal(6, 1)$. No estudo 2, $\beta_0 = 8$, $\beta_1 = -0,8$, $\sigma = 0,3$ e $X_2 \sim Normal(6, 1)$.

Considere que os modelos de regressão são como descrito acima 4.1. Logo, a função de verossimilhança para o vetor de parâmetros $(\beta_0, \beta_1, \sigma^2)$ é dada por:

$$\begin{aligned} L(\beta|Y, X) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-1}{2\sigma^2} [(y_i - \beta_0 - \beta_1 X_j) - 0]^2 \right\} \quad (4.2) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp \frac{-1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 - \beta_1 X_j)]^2. \end{aligned}$$

Assumindo que β_0 , β_1 e σ^2 são independentes, tais que $\beta_0 \sim N(0, 10)$, $\beta_1 \sim N(0, 10)$, $\sigma \sim InvGamma(2, 2)$, a priori para $(\beta_0, \beta_1, \sigma^2)$ pode ser assumida como:

$$\begin{aligned}\pi(\beta_0, \beta_1, \sigma^2) &= \pi(\beta_0)\pi(\beta_1)\pi(\sigma^2), \\ \implies \pi(\beta_0, \beta_1, \sigma^2) &\sim N(0, 10)N(0, 10)InvGamma(2, 2).\end{aligned}\tag{4.3}$$

Como foi sugerido no Capítulo 3, podemos abordar a metodologia da mistura metanalítica para obter uma mistura das distribuições a posteriori de $(\beta_0, \beta_1, \sigma^2)$ de cada estudo, ou seja,

$$\begin{aligned}\pi(\beta_0, \beta_1, \sigma^2 | Y_1, Y_2, X_1, X_2) &= \\ &= w_1\pi_1(\beta_0, \beta_1, \sigma^2 | Y_1, X_1) + w_2\pi_2(\beta_0, \beta_1, \sigma^2 | Y_2, X_2),\end{aligned}\tag{4.4}$$

em que: $w_1 = \frac{n_1}{(n_1+n_2)}$ e $w_2 = \frac{n_2}{(n_1+n_2)}$.

Na Figura 4.1 encontram-se as funções densidades a posteriori marginal dos parâmetros β_0, β_1 e σ^2 . Pode-se observar que o comportamento das mesma refletem uma densidade bimodal, o que indica que os estudos apresentam comportamentos diferentes. Entretanto, para uma análise mais completa tais densidades a posteriori são insuficientes para tomar conclusões sobre os estudos. Veja que não se pode fazer afirmações se todas as maiores modas são do mesmo estudo, necessitando assim de uma análise que considere todos os parâmetros aos mesmo tempo.

Na Figura 4.2, os pontos em vermelho representam os dados do estudo 1 e os pontos em verde, os dados do estudo 2. As retas vermelha e verde são referentes ao ajuste da regressão utilizando somente os dados do estudo 1 e 2, respectivamente. A reta preta corresponde a média da distribuição metanalítica a posteriori. Observa-se que o comportamento dessa reta em preto não representa o comportamento de nenhum dos estudos. Portanto, construiu-se a distribuição preditiva baseada na medida metanalítica (veja Figura 4.3). Note que a densidade preditiva abrange os pontos dos dois estudos, tendo maior concentração no centro e seguindo com maior

massa as retas de cada estudo individual. Na Figura 4.4 apresenta-se o gráfico tridimensional da densidade preditiva.

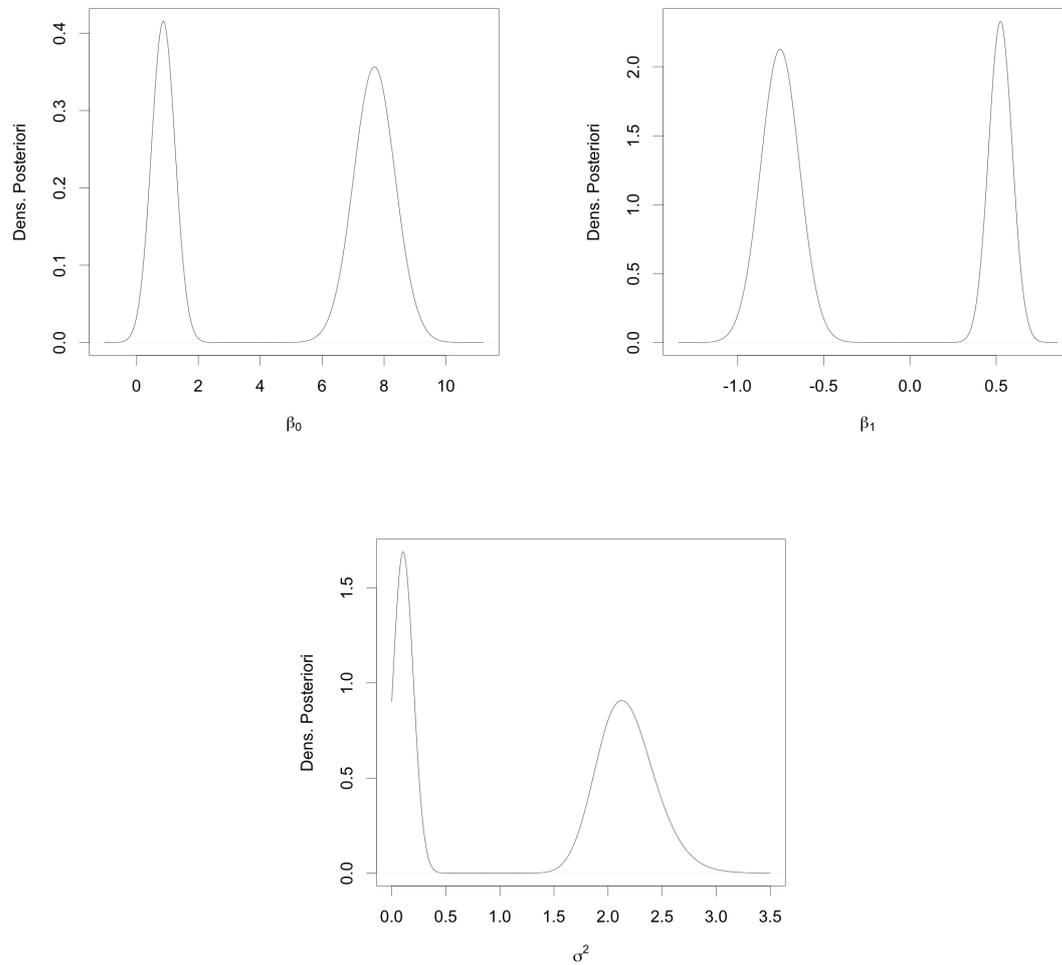


Figura 4.1: Densidade a posteriori marginal de β_0 , β_1 e σ^2 .

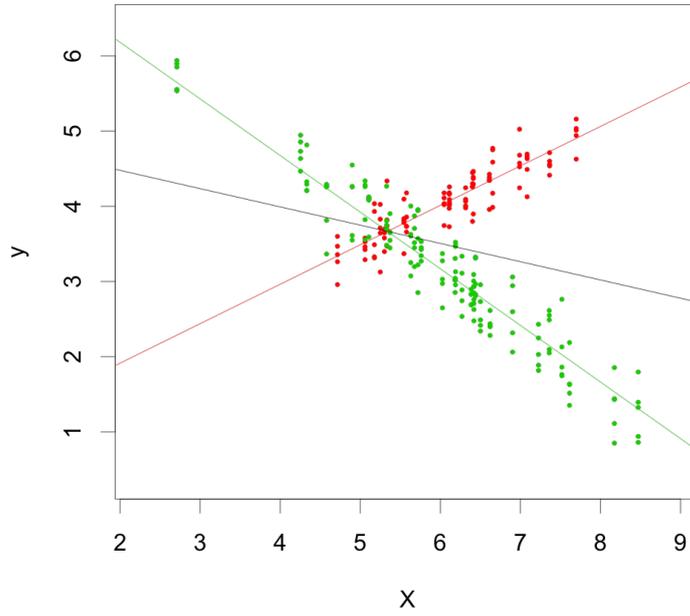


Figura 4.2: *Reta de Regressão para cada estudo e média da medida metanalítica.*

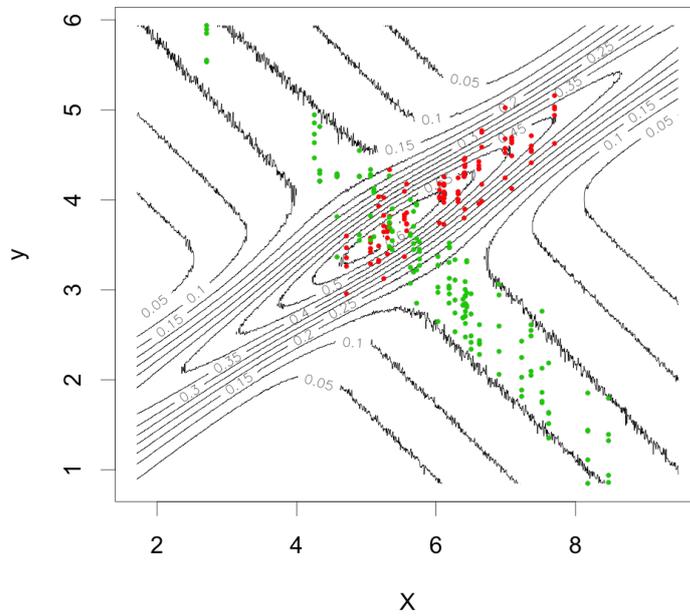


Figura 4.3: *Curva de nível da densidade preditiva de Y baseada na medida metanalítica.*

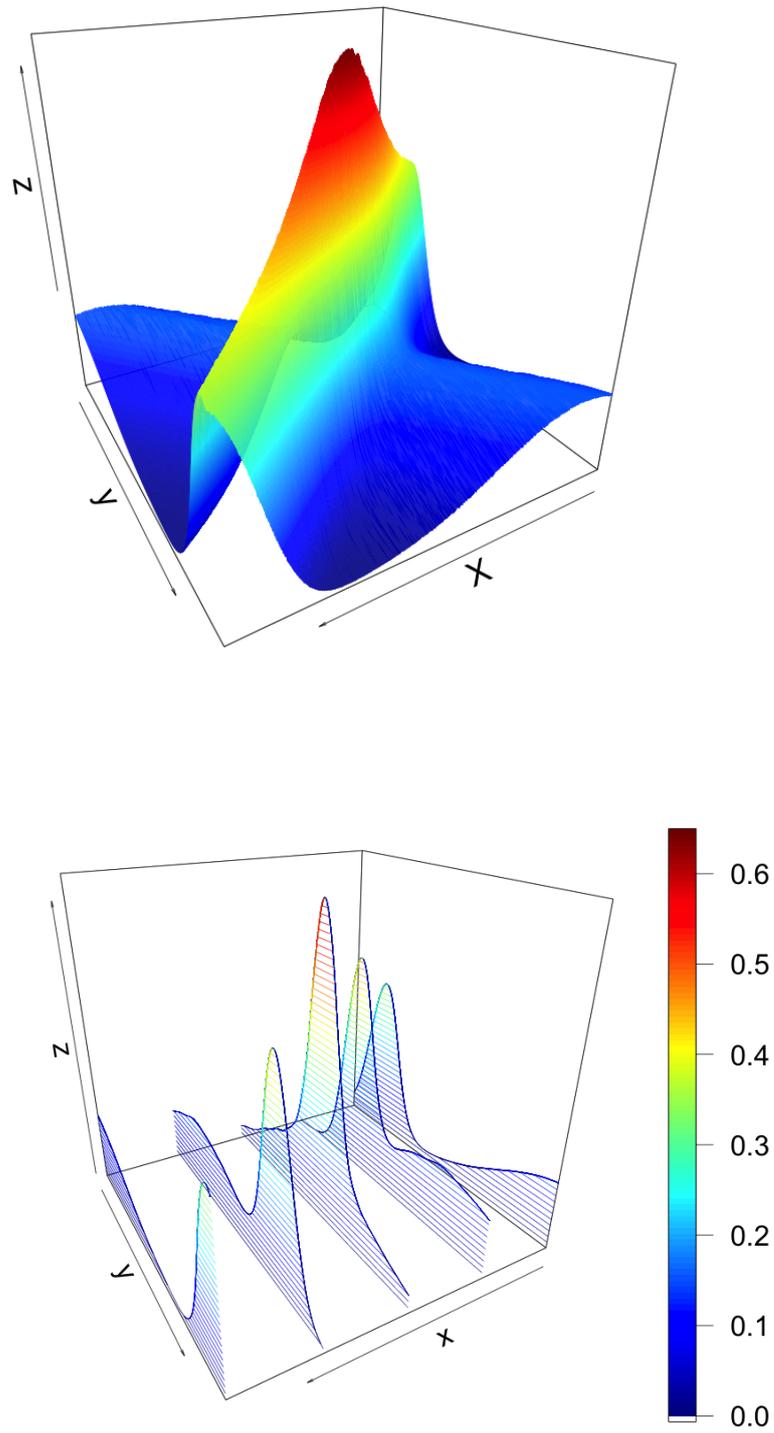


Figura 4.4: *Densidade preditiva de Y baseada na medida metanalítica.*

4.3 Aplicação Dados Reais

Nesta seção discute-se um exemplo apresentado por Bueno [4] no qual foram analisados os endpoints citogenéticos de duas espécies de roedores silvestres: 'Akodon montensis' and 'Oryzomys nigripes – living', observadas em três áreas: uma área industrial, na agricultura e em uma área de preservação. Os estudos foram realizados no estado de Santa Catarina, no vale do Itajaí.

Características dos estudos

Foram feitos 3 estudos em paralelo com as mesmas duas espécies de roedores: 'AM - Akodon montensis' and 'ON - Oryzomys nigripes – living'. Na área de preservação (AP - Estudo 1) foram estudados animais não expostos a poluição ambiental; na agricultura (CA - Estudo 2) foram analisados animais expostos aos efeitos genotóxicos de pesticidas usados na plantação de arroz e na área industrial (AI - Estudo 3), animais expostos ao resíduo industrial.

Após a captura dos animais, os mesmos foram pesados, marcados quanto ao sexo e injetados com uma solução. Logo em seguida foram mortos por inalação de um gás tóxico. Cada roedor foi usado para estimar todos os endpoints citogenéticos.

A análise dos endpoints citogenéticos em organismos expostos a produtos químicos em seu ambiente natural pode ajudar na detecção precoce de dano genotóxico. Esta metodologia foi utilizada para avaliar o impacto de compostos ambientais que ocorrem no Vale do Itajaí, estado de Santa Catarina, sul do Brasil.

O objetivo dos estudos foi avaliar o desempenho dos endpoints a fim de estabelecer um perfil genotóxico de cada área, avaliando: a proporção de células (PCE/NCE) policromática/normocromáticas; o índice mitótico (IM); a frequência de células micronucleadas tanto na medula óssea quanto no sangue periférico e a frequência de células com aberrações cromossômicas na medula óssea.

Como propósito final dos estudos, pretendeu-se entender se a observação do endpoint citogenético é suficiente para entender o efeito dos agentes ambientais no material genético dos animais.

Os métodos apresentados nos capítulos anteriores serão abordados de forma a exemplificar a medida metanalítica Bayesiana como forma de desenvolver uma

metanálise para modelos de regressão e, assim, entender se existe diferença entre os estudos.

O conjunto de dados [4] tem disponível as seguintes variáveis:

- a) frequência de PCE entre os 100 primeiros NCE observados em cada lâmina (PCE/NCE);
- b) frequência de células micronucleadas em eritrócitos policromáticos (MNPCE) dentre os 2.000 que foram contados ;
- c) frequência de eritrócitos normocromáticos (MNNCE) dentre os 2.000 que foram analisados;
- d) frequência de células mortas (DC) dentre 2.000;
- e) frequência de eritrócitos de sangue periférico micronucleados (MNPBE) dentre os 4.000 eritrócitos;
- f) frequência de Metáfases (fase mitótica) entre as 2.000 células observadas (META).

Nesta seção, apresentamos as análises das variáveis PCE/NCE e MNPCE.

4.3.1 Análise de regressão

Considera-se as duas espécies (AM e ON), os três estudos (AP: área de preservação, CA: campo de arroz, AI: área industrial) e o sexo como covariáveis de uma análise de regressão linear simples e, como variável resposta, testaremos as frequências observadas de PCE/NCE e MNPCE apenas.

Observe que estas variáveis resposta são quantitativas discreta, portanto temos um problema de dados de contagem, no qual utilizamos o modelo com distribuição Binomial Negativa para a variável de interesse.

Genericamente, a interpretação de Modelos Lineares Generalizados com função de ligação logarítmica ou genericamente chamados de modelos log-lineares, que usamos neste trabalho, apresentam a mesma forma de interpretação dos coeficientes estimados, independente da distribuição adotada para a variável resposta. E, em geral temos um modelo com ligação logarítmica.

Ajuste

A fim de realizar uma análise de regressão, seja Y_{ijkl} a frequência de endpoints no i -ésimo rato referente ao j -ésimo estudo ($j = 1, 2, 3$), a k -ésima espécie ($k = 1, 2$), ao l -ésimo sexo ($l = 1, 2$). Além disso, Y_{ijkl} são variáveis iid, que segue uma distribuição Binomial Negativa, $Y_{ijkl} \sim BN(\mu_{ijkl}, \phi)$.

Utilizou-se como casela de referência: Espécie AM, Sexo Masculino e Estudo 3 (AI).

Por meio do gráfico de probabilidade normal com envelopes simulados para um ajuste da distribuição Binomial Negativa, Figuras 4.5 e 4.8, observa-se que o modelo produziu um bom ajuste.

- **Variável PCE/NCE**

O modelo de regressão ajustado, assumindo uma distribuição binomial negativa para a variável resposta PCE/NCE, é dado pela expressão:

$$\hat{y} = \exp [4.03 - 0,03especieON - 0,03sexo + 0,16estudoAP + 0,28estudoCA]$$

Na Tabela 4.1 encontram-se os valores estimado, bem como a significância dos coeficientes associados as covariáveis do modelo. Observa-se que não há covariáveis significativas no modelo, ou seja, não podemos afirmar nada quanto a diferença de estudos. Analisando o p-valor apresentado na Tabela 4.2, temos que ao nível de 5% de significância há evidências para não rejeitar que os parâmetros não são significativo ao modelo.

A fim de verificar se o ajuste de um modelo é adequado, podemos comparar o valor da Deviance com os percentis da distribuição χ^2 com $(n - p) = 62$ graus de liberdade, em que n é número total de indivíduos e p é o posto da matriz do modelo e nível de significância α . Na Tabela 4.2, como o valor da Deviance, igual a 69,587, é inferior ao valor da $\chi^2_{(n-p),\alpha} = 81,381$, podemos dizer que há evidências a um nível aproximado de 100α % de confiança e que o modelo proposto está bem ajustado aos dados.

A análise de diagnóstico do modelo não apresenta nenhum problema quanto aos resíduos do modelo e também não há pontos atípicos.

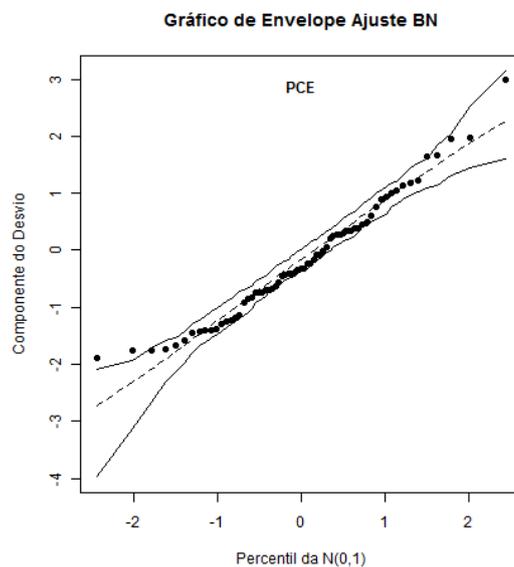


Figura 4.5: Gráfico normal de probabilidades referente ao modelo BN ajustado aos dados.

Tabela 4.1: Estimativas dos parâmetros do modelo Binomial Negativo - PCE.

Coefficientes	Estimativa	Erro Padrão	P-valor
Intercepto	4,030	0,148	< 2e-16
Espécie ON	-0,033	0,155	0,8066
Sexo M	-0,031	0,144	0,8270
Estudo AP	0,166	0,164	0,3091
Estudo CA	0,281	0,169	0,0964

Tabela 4.2: Análise de desvio - PCE.

	DF	Deviance	P(> Chi)
Intercepto		72,443	
Espécie	1	72,400	0,8364
Sexo	1	72,354	0,8292
Estudo	2	69,587	0,2504

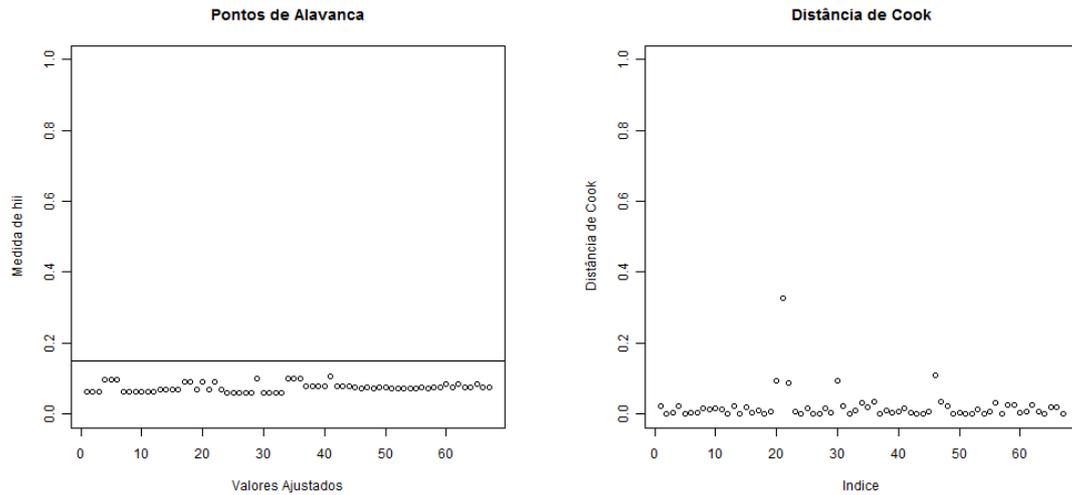


Figura 4.6: Gráfico para verificação de pontos atípicos (PCE).

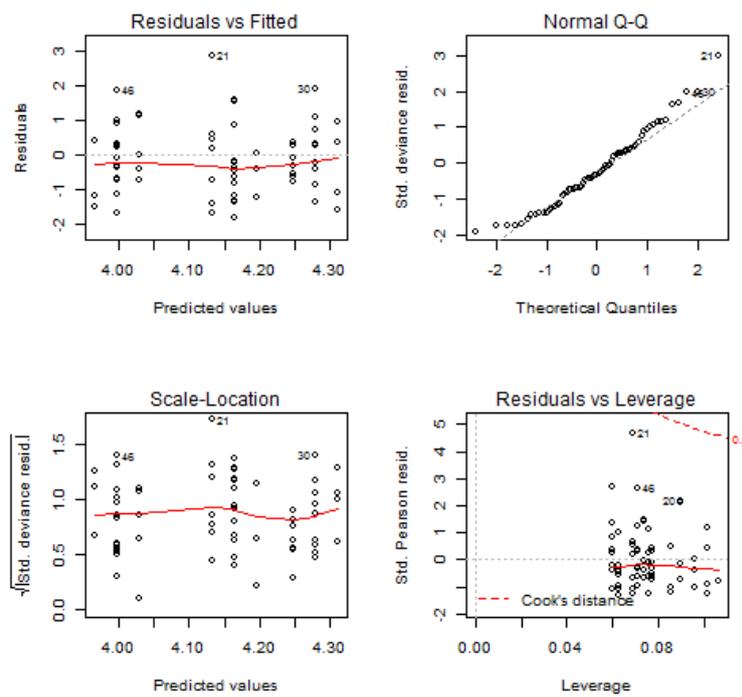


Figura 4.7: Gráfico para verificação de suposições dos resíduos (PCE).

- Variável MNPCE

Para o modelo referente a variável resposta MNPCE, o modelo ajustado é

dado por:

$$\hat{y} = \exp 2,29 + +0,60\text{especieON} + 0,18\text{sexo} - 0,30\text{estudoAP} - 0,04\text{estudoCA}$$

Pela Tabela 4.3, adotando um nível de significância de 10%, nota-se que apenas os fatores Espécie ON e Estudo AP são significativos.

Observa-se, pelo p-valor mostrado na Tabela 4.4, que ao nível de significância 5% de significância há evidências para não rejeitar a hipótese nula de que $\beta_1 = 0$, o que indica que a covariável Espécie é significativa ao modelo, enquanto Sexo e Estudo não são.

Quanto ao ajuste do modelo, há evidências de que modelo proposto é adequado, uma vez que o valor da Deviance igual a 70,134 é menor do que o valor $\chi^2_{(n-p),\alpha} = 81,381$.

Pelos Figuras 4.10 e 4.9, tem-se que o modelo não apresenta pontos atípicos, bem como também não há violação de suposições resíduos.

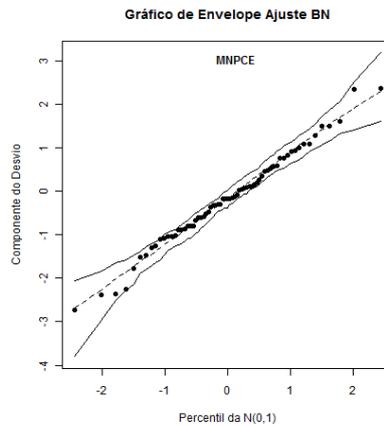


Figura 4.8: Gráfico normal de probabilidades referente ao modelo BN ajustado aos dados.

Tabela 4.3: Estimativas dos parâmetros do modelo - MNPCE.

Coefficientes	Estimativa	Erro Padrão	P-valor
Intercepto	2,294	0,153	< 2e-16
Espécie ON	0,604	0,138	1,2e-05
Sexo M	0,180	0,149	0,225
Estudo AP	-0,304	0,169	0,071
Estudo CA	-0,045	0,172	0,791

Tabela 4.4: Análise de desvio - MNPCE.

	DF	Deviance	P(> Chi)
Intercepto		91,556	
Espécie	1	74,491	3,613e-05
Sexo	1	73,743	0,3869
Estudo	2	70,134	0,1646

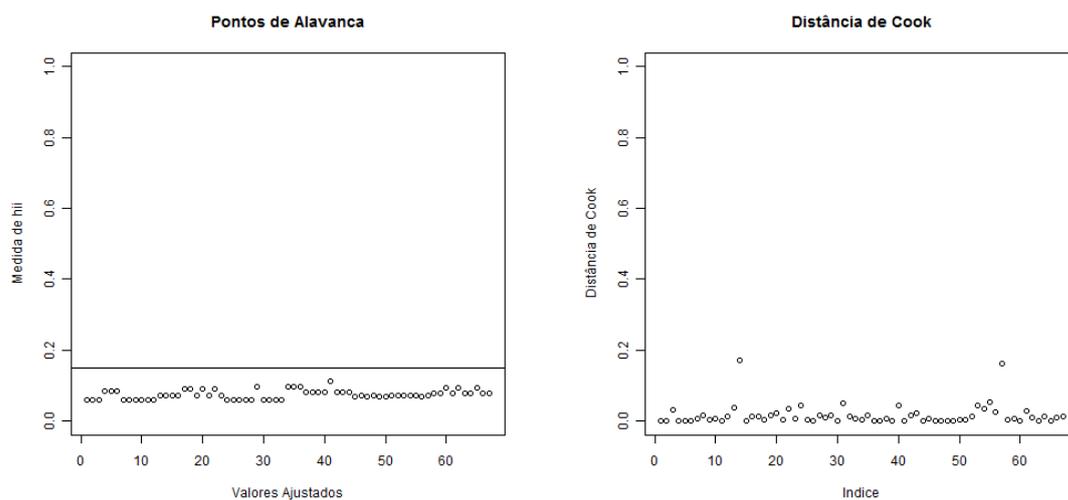


Figura 4.9: Gráfico para verificação de pontos atípicos (MNPCE).

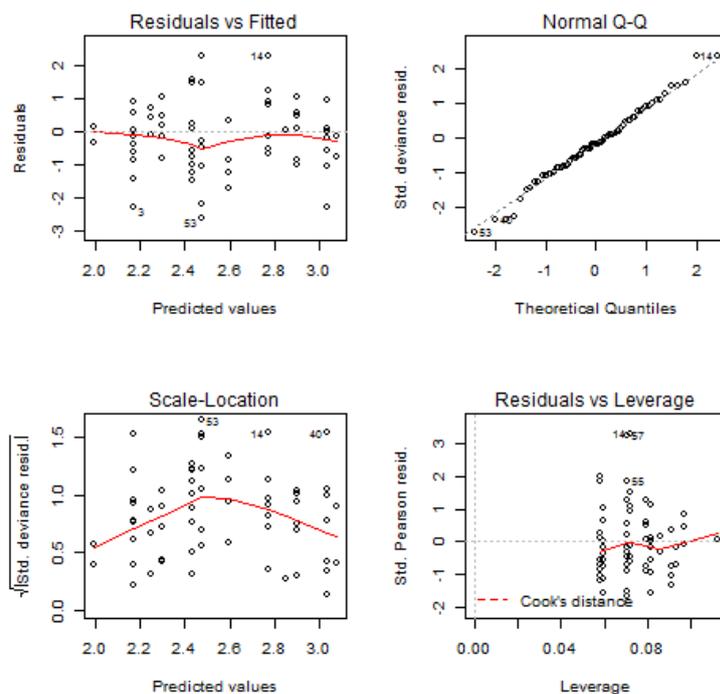


Figura 4.10: Gráfico para verificação de suposições dos resíduos (MNPCE).

4.3.2 Metanálise em modelos de regressão

Considere agora a existência de um modelo para cada estudo. Ou seja, J o número de estudos da metanálise e Y_j o efeito observado no estudo j (com $j = 1, 2, \dots, J$). Logo, o modelo é dado por:

$$Y_j \sim BN(\beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j}, \phi)$$

Para os dados, tem-se os seguintes modelos de regressão, de tamanhos $n_1 = 23$, $n_2 = 21$, $n_3 = 23$.

A função de verossimilhança para o vetor de parâmetros $\theta = (\beta_0, \beta_1, \beta_2, \phi)$ é dada por:

$$L_j(\theta | Y_j, X_{1j}, X_{2j}) = \sum_{i=1}^{n_j} \left[\log \left\{ \frac{\Gamma(\phi + y_{ij})}{\Gamma(y_{ij} + 1)\Gamma(\phi)} \right\} + \phi \log(\phi) + y_{ij} \log(\mu_{ij}) - (\phi + y_{ij}) \log(\mu_{ij} + \phi) \right], \quad (4.5)$$

em que: $\mu_{ij} = \exp(\beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij})$

Assumindo que $\beta_0, \beta_1, \beta_2$ e ϕ são independentes, e que $\beta_0, \beta_1, \beta_2 \sim N(0, 3^2)$, $\phi \sim \text{Gamma}(1/100, 1/100)$, uma priori para $(\beta_0, \beta_1, \beta_2, \phi)$ pode ser assumida como:

$$\begin{aligned} \pi(\beta_0, \beta_1, \beta_2, \phi) &= \pi(\beta_0)\Pi(\beta_1)\Pi(\beta_2)\pi(\phi) \quad (4.6) \\ \implies \Pi(\beta_0, \beta_1, \beta_2, \phi) &\sim N(0, 3^2)N(0, 3^2)N(0, 3^2)\text{Gamma}(1/100, 1/100). \end{aligned}$$

Dessa forma, pode-se obter a posteriori para cada um dos estudos da seguinte forma:

$$\pi_j(\theta | Y_j, X_{1j}, X_{2j}) \propto L_j(\theta | Y_j, X_{1j}, X_{2j})\Pi(\theta) \quad (4.7)$$

Podemos abordar a metodologia da mistura metanalítica para obter uma mistura das distribuições a posteriori de $(\beta_0, \beta_1, \beta_2, \phi)$ de cada estudo. Ou seja,

$$\pi(\boldsymbol{\beta}, \phi | \mathbf{Y}, X) = \tag{4.8}$$

$$w_1 \pi_1(\beta_0, \beta_1, \beta_2, \phi | \mathbf{Y}, X) + \pi_2(\beta_0, \beta_1, \beta_2, \phi | \mathbf{Y}, X)$$

- **Variável resposta PCE/NCE**

Na Figura 4.11 encontra-se a média do modelo de cada estudo (curvas vermelha = Estudo 1 (AP); verde = Estudo 2 (CA); azul = Estudo 3 (AI)) e também o resultado da medida metanalítica (curva preta). Observa-se, por exemplo, que os estudos apresentam comportamentos bem diferentes quanto a média. Analisando ratos da Espécie ON e sexo masculino tem-se que os Estudos 2 e 3 apresentam médias menores do que no Estudo 1. Ainda do sexo masculino, mas da Espécie AM a maior média se encontra em ratos do Estudo 2. Nota-se que para os ratos da Espécie ON e do sexo feminino os estudos 2 e 3 tem comportamentos semelhantes, sendo o estudo 1 o de maior média. Quanto a medida metanalítica, chama a atenção ratos da Espécie ON e do sexo feminino, pois observa-se que a curva da medida tem comportamento quase que bimodal, refletindo esses dois estudo semelhantes e o estudo 1 claramente diferente. Considerando todos os estudos combinados, pela Figura 4.12, não podemos afirmar que as espécies e o sexo apresentam diferenças. Apenas há indícios de que a maior diferença possa ocorrer entre as espécies AM e ON no sexo feminino.

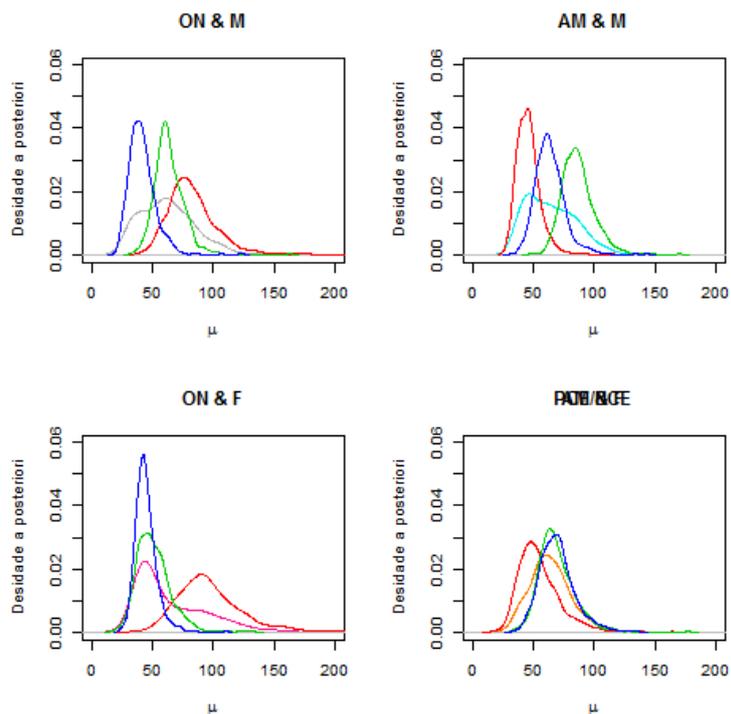


Figura 4.11: Distribuição a posteriori de μ condicional as covariáveis para variável resposta PCE/NCE.

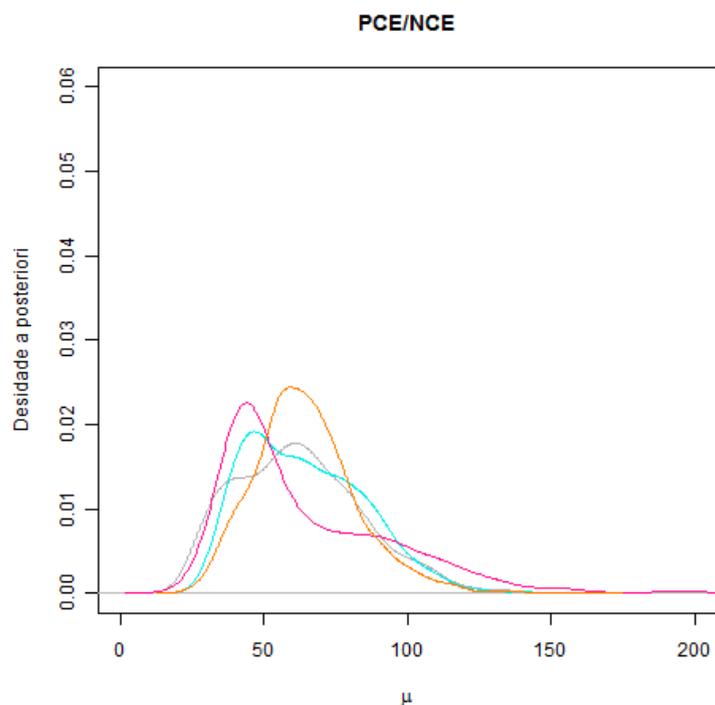


Figura 4.12: Medida metanalítica analisada para variável PCE/NCE.

- Variável resposta MNPCE

Para a variável MNPCE tem-se resultados diferentes ao compararmos com a variável PCE/NCE.

Na Figura 4.13 encontra-se a média do modelo de cada estudo e da medida metanalítica. Nota-se que o estudo 1 (curva vermelha) se diferencia dos demais quando comparado com as combinações dos níveis de espécie e sexo, sendo a contagem de células MNPCE menor na espécie AM.

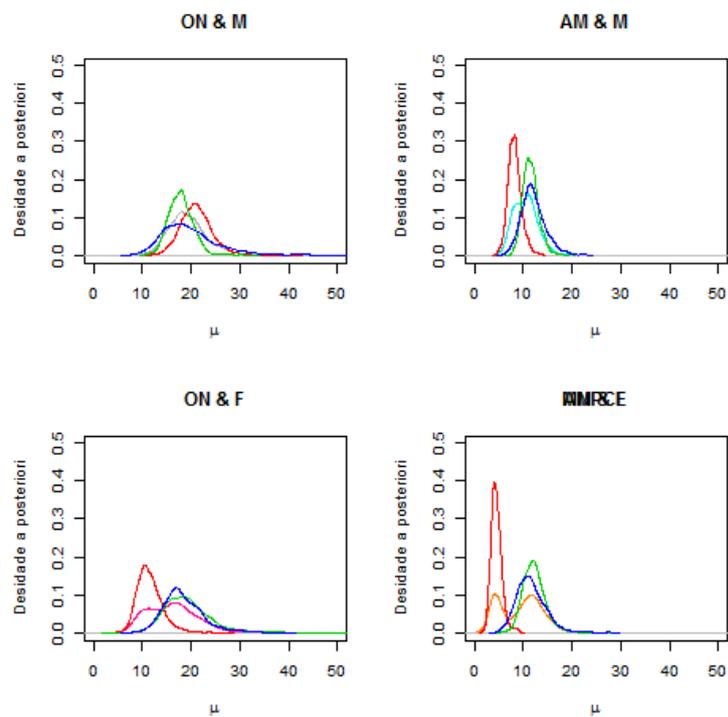


Figura 4.13: Distribuição a posteriori de μ condicional as covariáveis para variável resposta MNPCE.

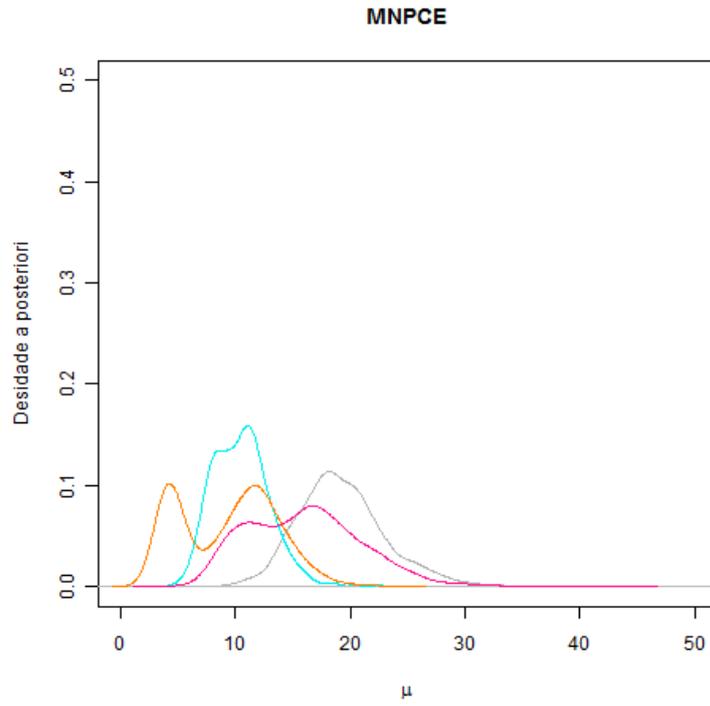


Figura 4.14: *Medida metanalítica analisada para variável MNPCE.*

Capítulo 5

Considerações Finais

Neste trabalho concentram-se problemas de multi centros, ou multi estudos, no qual o interesse é realizar metanálise. Na área médica, estudos de metanálise são muito utilizados por pesquisadores (ver Atallah [2], Berry [3]) e, por isso, entende-se que a técnica seja um problema atual e de interesse da sociedade.

O modelo de DerSimonian e Laird, metanálise baseada na literatura, é o método mais utilizado para a realização de uma metanálise. Ele incorpora a variabilidade existente entre os estudos na ponderação das medidas de efeito de cada estudo para, então, produzir uma medida metanalítica. Dessa forma, tal método tende a dar maior peso aos estudos de menor variabilidade, onde observa-se que a medida metanalítica resultante é pobre, pode ser viesada e não fornece muita informação sobre o parâmetro em questão. No Exemplo 3.2, referente aos dados normais simulados, o modelo de mistura finita utilizado em metanálise não auxiliou na interpretação dos resultados e, além disso, não considerou a variabilidade existente entre os estudos.

A metodologia Bayesiana Hierárquica é também uma das formas de tratar esses problemas de multi centros. Como visto no Exemplo 4.1, desenvolvido com dados de Berry [3], essa abordagem apresentou uma variabilidade a posteriori para a proporção de sucessos coerente com os estudos individuais. Entretanto, acreditamos que essa metodologia não consegue assegurar as diferenças entre os estudos, pois, por exemplo, o estudo com 100% de cura aparenta não ter sido bem representado

na distribuição a posteriori. Portanto, é necessário a procura de outras formas de realizar a metanálise que considerem o problema de diferentes maneiras e capture melhor esse tipo de informação.

Sendo assim, uma medida metanalítica Bayesiana foi proposta [[12]] como alternativa aos métodos utilizados. Esta medida é uma mistura de distribuições a posteriori do parâmetro de interesse, θ , de cada estudo. É importante ressaltar que esse método proposto produz, como medida metanalítica, uma distribuição de probabilidade a posteriori de θ e, por meio desta distribuição, podem ser calculadas medidas, tais como a esperança e a variância a posteriori θ , além de outras medidas. Portanto, a metodologia proposta é muito mais informativa, intuitiva e precisa que as existentes.

Observou-se que a metodologia proposta é uma medida geral, o que faz com que ela possa ser utilizada desde o problema mais trivial de metanálise até problemas mais complexos. Além disso, quando há estatística suficiente disponível, realizar uma metanálise baseada na literatura é equivalente a realizar uma metanálise caso a caso, sem perda de informação. Isso resulta em um ganho de qualidade para a metanálise baseada na literatura, uma vez que a metanálise caso a caso é considerada o "padrão ouro".

É importante deixar claro que o objetivo não é a comparação dos métodos usuais e Bayesiano, uma vez que os mesmos não são comparáveis. Mas o intuito é evidenciar possíveis falhas de uma metodologia amplamente utilizada na metanálise e chamar a atenção para a forma como o mesmo problema pode ser resolvido.

No Capítulo 4 a técnica metanalítica Bayesiana foi tratada em modelos de regressão com dados simulados. O intuito dessa aplicação foi entender melhor como se comporta a medida metanalítica Bayesiana em casos como este e a partir disso aplicá-la em dados reais. Descreveu-se em detalhes a metanálise para modelos de regressão e a técnica da medida metanalítica foi aplicada no problema de genotoxicidade ambiental em roedores selvagens (Bueno [4]). É possível notar que, neste caso, os estudos apresentam comportamento igual quando utilizada a variável PCE/NCE e comportamento distinto quando analisada a variável MNPCE.

Referências Bibliográficas

- [1] ALDERSON, P; GREEN, S. *Cochrane Collaboration open learning material for reviewers*. The Cochrane Collaboration. 2002.
- [2] ATALLAH, AN. *Revisão sistemática da literatura e metanálise. Medicina baseada em evidências: fundamentos da pesquisa clínica*. Lemos-Editorial, 1998.
- [3] BERRY, Donald A. *A Bayesian Approach to Multicenter Trials and Metaanalysis*. National Science Foundation, University of Minnesota, Washington, 1989
- [4] BUENO, A. M. S.; PEREIRA, C. A. B. e RABELLO-GAY, N. “Environmental Genotoxicity Evaluation Using Cytogenetic End Points in Wild Rodents”, 2000.
- [5] COCHRAN, WG. *Problems arising in the analysis of a series of similar experiments*. Journal of Royal Statistical Society Supplement, 1937.
- [6] DERSIMONIAN, R and Laird, *Meta-Analysis in Clinical Trials. Controlled Clinical Trials*, 1986.
- [7] DUTTON, M. *Individual patient-level data meta-analysis: A comparison of methods for the diverse populations collaboration data set*. PHD thesis, 2011.
- [8] FISHER, RA. *Statistical Methods for Research Workers*. London, Oliver and Boyd, 4th ed. 1, 1932.
- [9] GLASS, GV. *Primary, secondary, and meta-analysis of research*. Educ Res, 1976.
- [10] JUSTO, LP; Soares, BGO e Calil, HM. *Revisão sistemática, metanálise e medicina baseada em evidências: considerações conceituais*. Jornal Brasileiro de Psiquiatria, 2005.

- [11] LEANDRO, G. *Meta-analysis in Medical Research*. The handbook for the understanding and practice of meta-analysis BMJ Books, Blackwell Publishing, 2005.
- [12] MARTINS, Camila B. *Metanálise caso a caso sob a perspectiva bayesiana*. Tese apresentada para obtenção de título de doutor em Estatística. São Paulo, Instituto de Matemática e Estatística, 2013.
- [13] MARTINEZ, EZ. *Metanálise de ensaios clínicos controlados aleatorizados: aspectos quantitativos*. Revista Medicina (Ribeirão Preto), 2007.
- [14] PAULA, A., Gilberto. *Modelos de Regressão com apoio computacional*. Disponível em: http://www.ime.unicamp.br/cnaber/Livro_MLG.pdf. Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2010.
- [15] PEARSON, ES. *The probability integral transformation for testing goodness of fit and combining independent tests of significance*. Biometrika, 1938.
- [16] O'HAGAN, A. *Baysian Inference. Kendall's advanced theory of statistics*. 1991.
- [17] R. Disponível em: <http://www.r-project.org>, versão 2.8.1.