

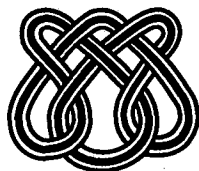
UNIVERSIDADE DE SÃO PAULO

**RELAÇÃO ENTRE O USO DE HORMÔNIO E
CÂNCER EM MULHERES: UMA APLICAÇÃO
DE META-ANÁLISE SOB UM ENFOQUE
BAYESIANO**

**JORGE ALBERTO ACHCAR
VIVIANE CARLA FORTULAN**

Nº 54

NOTAS



Instituto de Ciências Matemáticas de São Carlos

Instituto de Ciências Matemáticas e de Computação

ISSN - 0103-2577

**RELAÇÃO ENTRE O USO DE HORMÔNIO E
CÂNCER EM MULHERES: UMA APLICAÇÃO
DE META-ANÁLISE SOB UM ENFOQUE
BAYESIANO**

**JORGE ALBERTO ACHCAR
VIVIANE CARLA FORTULAN**

Nº 54

**NOTAS DO ICM C
Série Estatística**

**São Carlos
Dez./1998**

ABSTRACT

In this paper, considering a Bayesian approach to meta-analysis, we analyse medical data from different laboratories to study the possible link between the use of estrogen and cancer of the breast and endometrium. A special model is considered to analyse the data, and posterior summaries of interest are obtained using Gibbs Sampling algorithms. The proposed methodology is illustrated with two data sets introduced by Larose & Dey (1995).

RESUMO

Neste artigo, considerando um enfoque Bayesiano para meta-análise, analisamos dados médicos de laboratórios diferentes para estudar a possível relação entre a reposição de estrógeno e câncer nos seios, ovários e útero. Um modelo especial é considerado para analisar os dados, e sumários a posteriori de interesse são obtidos usando o algoritmo amostrador de Gibbs. A metodologia proposta é ilustrada com dois conjuntos de dados introduzidos por Larose & Dey (1995).

Palavras-chave: meta-análise, uso de estrógeno, câncer, análise Bayesiana, amostrador de Gibbs.

1. INTRODUÇÃO

Nos últimos anos tem havido muita pesquisa médica para avaliar a possível associação entre a incidência de câncer nos seios, ovários e útero e o uso de hormônios (ver por exemplo, Adami, 1992; Kaufman, 1991; ou Brinton & Schairer, 1993).

Como existe muitos pesquisadores independentes em diferentes centros de pesquisa envolvidos neste projeto de pesquisa médica, uma combinação dos resultados obtidos por estes pesquisadores deve ser considerada para uma análise estatística. A técnica estatística para combinar informação de várias fontes é chamada meta-análise (ver por exemplo, Hedges & Olkin, 1985).

Além da comparação de vários grupos de pacientes recebendo ou não hormônio (estrógeno), também devem ser consideradas alguns possíveis fatores de risco como a duração do tratamento com hormônios, o tipo de hormônio usado, a obesidade e outros fatores.

Em cada laboratório que é feito um estudo deste tipo são determinados os riscos relativos (ver por exemplo, Fleiss, 1973) de pacientes contraírem câncer recebendo ou não estrógeno.

O risco relativo que mede a associação entre o uso de estrógeno e a incidência de câncer é estimado por,

$$RR = \frac{ad}{bc} \quad (1)$$

onde a é o número de pacientes com câncer entre as usuárias de estrógeno; b é o número de pacientes sem câncer entre as usuárias de estrógeno; c é o número de pacientes com câncer entre as não usuárias de estrógeno e d é o número de pacientes sem câncer entre as não usuárias de estrógeno.

Observar que se $RR > 1$, temos um indicativo de que a chance de uma mulher ter câncer (seios, ovários e útero) é maior para as usuárias de hormônio.

Na tabela 1, temos um exemplo de dados desse tipo (ver Larose & Dey, 1995) coletados em alguns laboratórios de pesquisa médica dos Estados Unidos, envolvendo um estudo sobre câncer em ovários e útero.

Laborat6rios	Resposta	Dura76o			Tipo de Estr6geno		Obesidade	
		log(RR)	<12	[12,60)	≥ 60	Conjugado	Outros	N6o
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	-	-

Tabela 1: Dados para laborat6rios diferentes (incid6ncia de c6ncer de 6tero e ov6rios)

Na tabela 1, observamos que para v6rios laborat6rios h6 muita informa76o perdida para as covari6veis.

Para analisar os dados da tabela 1, podemos considerar o uso de meta-an6lise sob o enfoque Bayesiano (ver por exemplo, DuMouchel & Harris, 1983; ou Larose & Dey, 1995).

Assumir o seguinte modelo para analisar os dados da tabela 1:

$$\tilde{y} = X \tilde{\beta} + W \tilde{\delta} + \tilde{\varepsilon}, \quad (2)$$

onde \tilde{y} 6 um vetor ($n \times 1$) das respostas dadas por $y = \log(RR)$; X 6 uma matriz ($n \times 3$) de covari6veis dada por,

$$X = \begin{pmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_{1n} & x_{2n} & x_{3n} \end{pmatrix} = \begin{pmatrix} \tilde{x}_1 & \tilde{x}_2 & \tilde{x}_3 \end{pmatrix} \quad (3)$$

onde \tilde{x}_1 6 o vetor da covari6vel denotando a propor76o do tempo de tratamento maior ou igual 6 60 meses; \tilde{x}_2 6 o vetor da covari6vel denotando a propor76o de usu6rias de estrog6neo conjugado e \tilde{x}_3 6 o vetor da covari6vel denotando a propor76o de obesas em cada estudo; $\tilde{\beta}$ 6 o

vetor (3 x 1) de parâmetros de regressão $\beta_m, m=1,2,3$; δ é o vetor (2 x 1) de parâmetros $\delta_i, i=1,2$ que mede o efeito dos dois grupos diferentes (com informação completa nas covariáveis, ou não); W é uma matriz (n x 2) dada por,

$$W = \begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ w_{1n} & w_{2n} \end{pmatrix} = \begin{pmatrix} w_{\sim 1} & w_{\sim 2} \end{pmatrix} \quad (4)$$

onde $w_{ji} = 1$ (0); $j=1,2; i=1,2,\dots,n$ indica informação completa (ou não) para as covariáveis, e ε é o vetor de erros com elementos $\varepsilon_i, i=1,2,\dots,n$.

Assumir que as variáveis aleatórias $\varepsilon_i, i=1,2,\dots,n$ sejam independentes identicamente distribuídas com distribuição normal com média zero e variância σ^2 .

Associado ao modelo (2) com os dados da tabela 1, temos:

$$X = \begin{pmatrix} 0.42 & 0.77 & 0.41 \\ 0.12 & 0.33 & 0.18 \\ 0.50 & 0.73 & 0.26 \\ 0.47 & 0.88 & - \\ 0.64 & - & 0.16 \\ 0.21 & - & 0.16 \\ 0.53 & - & 0.24 \\ 0.52 & 0.54 & - \end{pmatrix}; \quad W = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad (5)$$

Para as observações perdidas em X , podemos considerar um estimador simplificado dado pela média das observações completas em cada vetor $x_j, j=1,2,3$; ou usar um método de simulação para gerar essas observações perdidas.

Larose & Dey (1995) consideram uma extensão da modelo de componentes de variância hierárquico de DuMouchel & Harris (1983) para analisar os dados da tabela 1, usando métodos de Monte Carlo em cadeias de Markov (MCMC) para uma análise Bayesiana.

Neste artigo assumimos o modelo (2) e uma análise Bayesiana não-hierárquica usando o algoritmo Gibbs Sampling (ver por exemplo, Gelfand & Smith, 1990) para determinar as quantidades a posteriori de interesse .

Além do uso de meta-análise sob o enfoque Bayesiano para relacionar o uso de hormônio e a incidência de câncer no útero e ovários, também consideramos o modelo (2) para analisar um conjunto de dados de vários institutos de pesquisa introduzido por Larose & Dey (1995) relacionando a incidência de câncer nos seios e o uso de hormônios associados com vários fatores de risco.

2. UMA ANÁLISE BAYESIANA PARA OS DADOS DA TABELA 1

A função de verossimilhança para os parâmetros $\underline{\beta}, \underline{\delta}$ e σ^2 do modelo (2) é dada por,

$$L(\underline{\beta}, \underline{\delta}, \sigma^2) = (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \left(\underline{y} - X\underline{\beta} - W\underline{\delta}\right)' \left(\underline{y} - X\underline{\beta} - W\underline{\delta}\right)\right\} \quad (6)$$

onde \underline{y}' denota a transposta do vetor \underline{y} .

Para uma análise Bayesiana, assumir as seguintes distribuições a priori para os parâmetros:

- (i) $\delta_l \sim N(0, \alpha_l^2); l=1,2; \alpha_l$ conhecidos;
- (ii) $\beta_m \sim N(0, b_m^2); m=1,2,3; b_m$ conhecidos; (7)
- (iii) $\sigma^2 \sim IG(c, d); c, d$ conhecidos,

onde $N(a, b^2)$ denota uma distribuição normal com média a e variância b^2 e $IG(c, d)$ denota uma distribuição gama inversa com média $d/(c-1)$ e variância $d^2/(c-1)^2(c-2)$. Também assumir independência a priori entre os parâmetros

Assumindo $\alpha_1 = \alpha_2 = \alpha$ e $b_1 = b_2 = b_3 = b$, a distribuição a posteriori conjunta para $\underline{\beta}, \underline{\delta}$ e σ^2 é dada por:

$$\pi(\underline{\beta}, \underline{\delta}, \sigma^2 | \underline{y}) \propto (\sigma^2)^{-\left(\frac{n}{2}+c+1\right)} \exp\left\{-\frac{1}{2a^2} \underline{\delta}' \underline{\delta} - \frac{1}{2b^2} \underline{\beta}' \underline{\beta}\right\} \times \exp\left\{-\frac{1}{\sigma^2} \left[d + \frac{1}{2} (\underline{y}' \underline{y} - 2 \underline{y}' \underline{X} \underline{\beta} - 2 \underline{y}' \underline{W} \underline{\delta} + \underline{\beta}' \underline{X}' \underline{X} \underline{\beta} + 2 \underline{\beta}' \underline{X}' \underline{W} \underline{\delta} + \underline{\delta}' \underline{W}' \underline{W} \underline{\delta}) \right] \right\} \quad (8)$$

As distribuições condicionais necessárias para o algoritmo Gibbs Sampling são dadas por:

$$\begin{aligned} \text{(i)} \quad \sigma^2 | \underline{\beta}, \underline{\delta}, \underline{y} &\sim \text{IG}\left[\frac{n}{2} + c; d + \frac{1}{2} \sum_{i=1}^n (y_i - x_{1i} \beta_1 - x_{2i} \beta_2 - x_{3i} \beta_3 - w_{1i} \delta_1 - w_{2i} \delta_2)^2\right] \\ \text{(ii)} \quad \beta_1 | \beta_{(1)}, \underline{\delta}, \sigma^2, \underline{y} &\sim \text{N}\left\{\frac{k_{12}}{k_{11}}; \frac{1}{k_{11}}\right\}; \\ \text{(iii)} \quad \beta_2 | \beta_{(2)}, \underline{\delta}, \sigma^2, \underline{y} &\sim \text{N}\left\{\frac{k_{22}}{k_{21}}; \frac{1}{k_{21}}\right\}; \\ \text{(iv)} \quad \beta_3 | \beta_{(3)}, \underline{\delta}, \sigma^2, \underline{y} &\sim \text{N}\left\{\frac{k_{32}}{k_{31}}; \frac{1}{k_{31}}\right\}; \\ \text{(v)} \quad \delta_1 | \delta_2, \underline{\beta}, \sigma^2, \underline{y} &\sim \text{N}\left\{\frac{r_{12}}{r_{11}}; \frac{1}{r_{11}}\right\}; \\ \text{(vi)} \quad \delta_2 | \delta_1, \underline{\beta}, \sigma^2, \underline{y} &\sim \text{N}\left\{\frac{r_{22}}{r_{21}}; \frac{1}{r_{21}}\right\}. \end{aligned} \quad (9)$$

onde $\beta_{(i)} = (\beta_j, \beta_k); j \neq i, k \neq i; i, j, k = 1, 2, 3;$

$$k_{j1} = \frac{1}{b^2} + \frac{b_{jj}}{\sigma^2}; \quad j=1, 2, 3;$$

$$k_{12} = \frac{a_1}{\sigma^2} - \frac{b_{12} \beta_2}{\sigma^2} - \frac{b_{31} \beta_3}{\sigma^2} - \frac{c_{11} \delta_1}{\sigma^2} - \frac{c_{12} \delta_2}{\sigma_2};$$

$$k_{22} = \frac{a_2}{\sigma^2} - \frac{b_{12} \beta_1}{\sigma^2} - \frac{b_{23} \beta_3}{\sigma^2} - \frac{c_{21} \delta_1}{\sigma^2} - \frac{c_{22} \delta_2}{\sigma_2};$$

$$k_{32} = \frac{a_3}{\sigma^2} - \frac{b_{31} \beta_1}{\sigma^2} - \frac{b_{23} \beta_2}{\sigma^2} - \frac{c_{31} \delta_1}{\sigma^2} - \frac{c_{32} \delta_2}{\sigma_2};$$

$$r_{j1} = \frac{1}{a^2} + \frac{e_{jj}}{\sigma^2}; \quad j=1,2;$$

$$r_{12} = \frac{d_1}{\sigma^2} - \frac{c_{11}\beta_1}{\sigma^2} - \frac{c_{21}\beta_2}{\sigma^2} - \frac{c_{31}\beta_3}{\sigma^2} - \frac{\delta_2 e_{12}}{\sigma_2};$$

$$r_{22} = \frac{d_2}{\sigma^2} - \frac{c_{12}\beta_1}{\sigma^2} - \frac{c_{22}\beta_2}{\sigma^2} - \frac{c_{32}\beta_3}{\sigma^2} - \frac{\delta_1 e_{12}}{\sigma_2};$$

$$a_j = \sum_{i=1}^n y_i x_{ji}; \quad j=1,2,3;$$

$$b_{jk} = \sum_{i=1}^n x_{ji} x_{ki}; \quad j,k=1,2,3;$$

$$c_{jk} = \sum_{i=1}^n x_{ji} w_{ki}; \quad j=1,2,3; k=1,2;$$

$$d_j = \sum_{i=1}^n y_i w_{ji}; \quad j=1,2;$$

$$e_{jk} = \sum_{i=1}^n w_{ji} w_{ki}; \quad j,k=1,2;$$

Para as observações perdidas nos vetores de covariáveis x_i , $i=1,2,3$, podemos gerar valores em cada iteração do algoritmo Gibbs Sampling a partir da distribuição $N(\bar{x}_i, s_i^2)$ onde \bar{x}_i é a média amostral das observações completas em x_i e s_i^2 é a variância amostral.

Considerando o modelo (2) para os dados da tabela 1, e as distribuições a priori (7) para os parâmetros do modelo, com $a=2$, $b=2$, $c=1$, $d=1$, geramos cinco cadeias de Gibbs cada uma com 2000 iterações, sendo as 400 primeiras observações eliminadas (“burn-in samples”). A convergência das amostras de Gibbs foram monitoradas usando o método de Gelman & Rubin (1992) que usa a técnica de análise de variância para determinar se iterações adicionais são necessárias. Para cada parâmetro consideramos uma iteração em cada 10. Na tabela 2 temos os sumários a posteriori e os valores \hat{R} (critério de Gelman & Rubin). Para todos os parâmetros observamos a convergência do algoritmo Gibbs Sampling ($\sqrt{\hat{R}} < 1.1$ para todos os parâmetros).

Observamos que as covariáveis x_1 (duração do tratamento); x_2 (tipo de hormônio) e x_3 (obesidade) indicam efeitos significativos no risco relativo (observar que os intervalos de credibilidade não incluem zero).

Parâmetros	Média a posteriori	Intervalo de Credibilidade (95%)	\hat{R}
β_1	0.8272	(0.3096 ; 1.3507)	1.0038
β_2	0.7816	(0.2992 ; 1.2406)	1.0321
β_3	3.4257	(2.0255 ; 4.7415)	1.0084
δ_1	-0.7992	(-1.1637 ; -0.3771)	1.0028
δ_2	-0.3377	(-0.6465 ; 0.0190)	1.0024
σ^2	0.1452	(0.0534 ; 0.3497)	1.0030

Tabela 2: Sumários a posteriori (x_1 é o vetor duração maior ou igual à 60 meses)

Nas figuras 1 e 2 temos as representações gráficas dos sumários a posteriori os gráficos das distribuições a posteriori marginais aproximadas pelas amostras geradas pelo algoritmo de Gibbs e gráficos de convergência das amostras geradas.

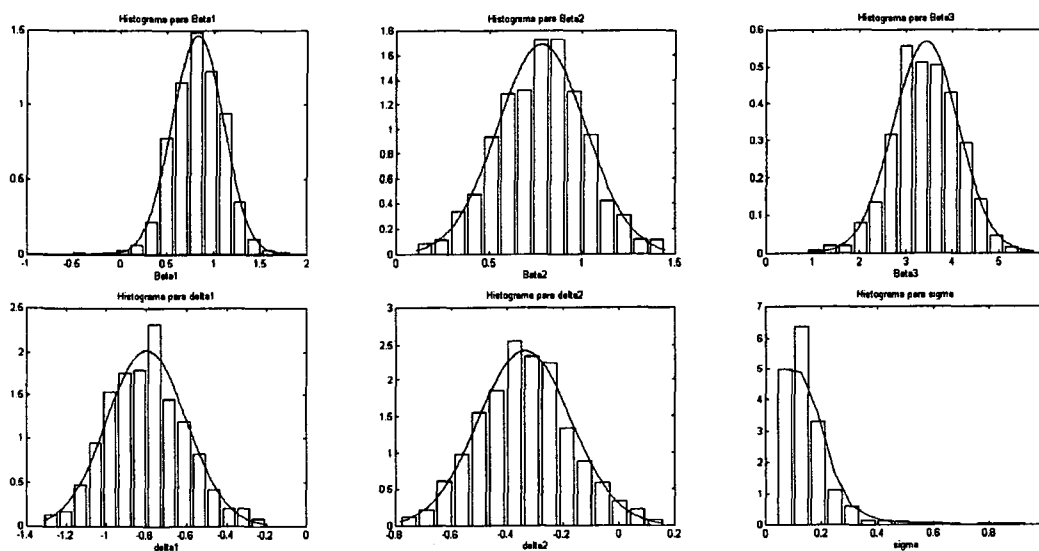


Figura 1: Densidades a posteriori marginais (duração maior do que 60 meses)

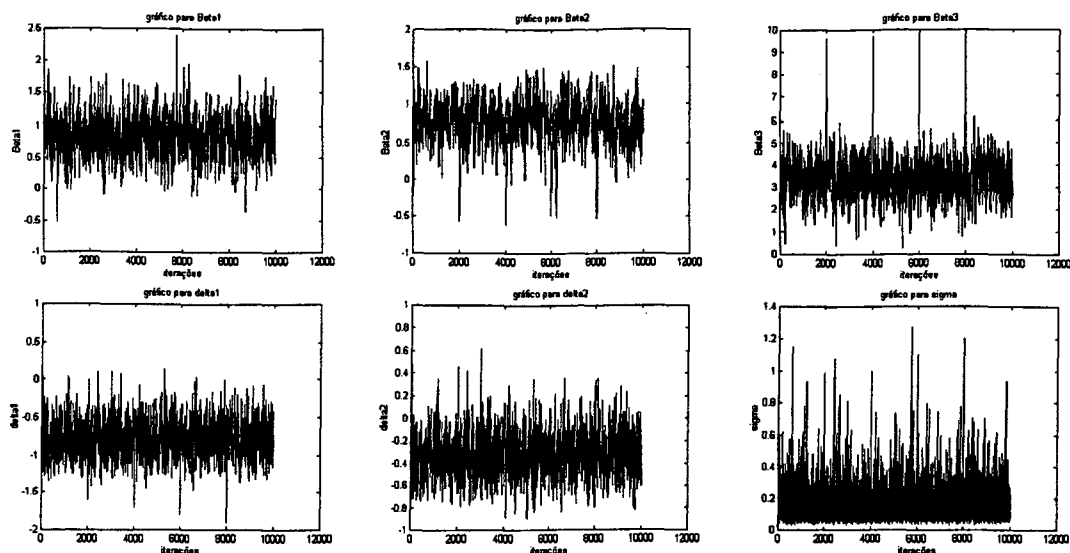


Figura 2: Convergência das amostras geradas (duração maior do que 60 meses)

Também poderíamos considerar o vetor de covariáveis x_1 como a proporção de usuárias de estrógeno por pelo menos 36 meses em lugar dos 60 meses considerados previamente. Essa proporção pode ser medida pelo número médio das usuárias de estrógeno no intervalo de 12 a 60 meses adicionado das usuárias de estrógeno por pelo menos 60 meses. Dessa forma, a primeira coluna da matriz X dada em (5) é substituída por, $x_1 = (0.63, 0.19, 0.65, 0.63, 0.75, 0.45, 0.69, 0.68)$. Na tabela 3 temos os sumários a posteriori e os valores de \hat{R} considerando x_1 a proporção de usuárias de estrógeno por pelo menos 36 meses.

Parâmetros	Média a posteriori	Intervalo de Credibilidade (95%)	\hat{R}
β_1	1.2017	(0.6790 ; 1.6967)	0.9999
β_2	0.5064	(-0.0114 ; 0.9635)	1.0102
β_3	3.1148	(1.5899 ; 4.6636)	1.0121
δ_1	-0.8419	(-1.1955 ; -0.4106)	0.9991
δ_2	-0.4653	(-0.8008 ; -0.1214)	0.9983
σ^2	0.1519	(0.0556 ; 0.3872)	1.0035

Tabela 3: Sumários a posteriori (x_1 é o vetor de duração maior ou igual à 36 meses)

Nas figuras 3 e 4, temos as representações gráficas dos sumários a posteriori.

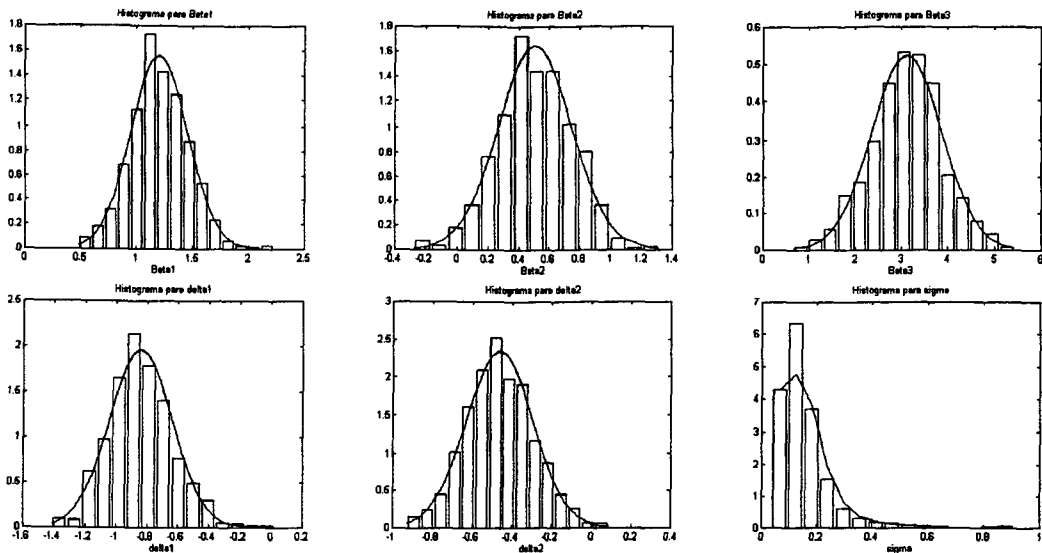


Figura 3: Densidades a posteriori marginais (duração maior do que 36 meses)

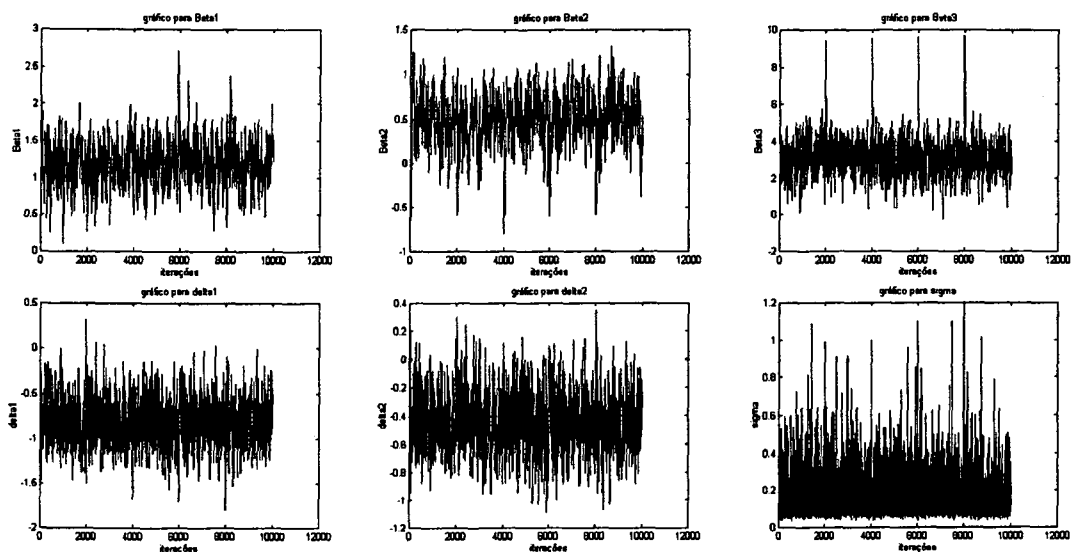


Figura 4: Convergência das amostras geradas (duração maior do que 36 meses)

A partir dos resultados obtidos na tabela 3, observamos uma concordância com os resultados obtidos na tabela 2, isto é, ainda indicando um efeito positivo da duração do tempo usando estrógeno (36 meses), tipo de estrógeno e obesidade.

Esses resultados também estão de acordo com os resultados obtidos por Larose & Dey (1995) assumindo um modelo Bayesiano hierárquico de efeitos aleatórios agrupados, que mostra significância para o fator de risco duração do tratamento e incidência de câncer no útero e ovário, significância para o tipo de estrógeno usado (conjugado versus não conjugado). Larose & Dey (1995) não encontram significância para o fator obesidade, o que contradiz o esperado pela literatura médica. Em nossa análise com o modelo (2) encontramos significância para o fator de risco obesidade.

3. RELAÇÃO ENTRE O USO DE ESTRÓGENO E A INCIDÊNCIA DE CASOS DE CÂNCER NOS SEIOS

Na tabela 4, temos um conjunto de dados introduzidos por Larose & Dey (1995) envolvendo vários estudos sobre a relação da incidência de câncer nos seios e o uso de hormônios.

Para analisar os dados da tabela 4, assumir o modelo (2) onde \underline{y} é o vetor das respostas $y = \log(RR)$, RR denota o risco relativo de uma mulher Ter câncer nos seios entre as usuárias de estrógeno; X é a matriz ($n \times 7$) de covariáveis dada por $X = \left(\begin{matrix} \underline{x}_1 & \underline{x}_2 & \dots & \underline{x}_7 \end{matrix} \right)$ onde \underline{x}_1 é o vetor de covariáveis representando a proporção de mulheres que tiveram pelo menos três filhos; \underline{x}_2 denota a proporção de mulheres que tiveram o primeiro filho com pelo menos 30 anos; \underline{x}_3 denota a proporção de mulheres que tiveram a primeira menstruação com pelo menos 15 anos; \underline{x}_4 denota a proporção de mulheres com tipo de menopausa cirúrgica; \underline{x}_5 denota a proporção de mulheres com história de casos de câncer na família; \underline{x}_6 denota a proporção de mulheres que usaram estrógeno por pelo menos 10 anos e \underline{x}_7 denota a proporção de mulheres que entraram na menopausa com pelo menos 50 anos. Também consideramos $W = \left(\begin{matrix} w_1 & w_2 \end{matrix} \right)$ onde $w_{ji} = 1$ (0); $j=1,2; i=1,2,\dots,n$ indica informação completa (ou não) para as covariáveis.

Labora- tórios	Resposta	Nº de Filhos			Idade no Primeiro Parto			Idade na Primeira Menstruação		
		0	1-2	≥3	<20	20-29	≥30	<12	12-14	≥15
1	0.0245	345	683	539	155	848	219	287	1031	242
2	0.0746	576	1266	1230	387	2556	376	590	2098	408
3	0.0827	204	731	770	113	1116	271	294	1194	201
4	-0.0419	331	981	1678	721	1740	136	666	1987	352
5	0.0185	359	733	844	146	1120	311	317	1361	250
6	0.0172	210	440	422	81	613	168	194	691	181
7	0.1793	-	-	-	-	-	-	202	717	403
8	0.1487	112	298	340	53	433	121	172	467	109
9	0.3338	197	411	473	-	-	-	-	-	-

Labora- tórios	Tipo de Menopausa		Câncer nos Seios na Família		Duração do Tratamento com Estrógeno			Idade no Início da Menopausa		
	Natural	Cirúrg.	Positivo	Negativo	≤1	1-9	≥10	<40	40-49	≥50
1	600	244	130	1438	49	108	45	42	243	370
2	2134	961	335	2761	187	259	105	571	1156	1239
3	1274	432	244	1416	177	142	29	164	695	808
4	1913	1101	279	1639	186	632	137	-	-	-
5	2435	1002	498	1440	589	1025	536	304	879	761
6	-	-	85	982	115	308	89	168	378	378
7	942	380	-	-	100	211	177	126	629	557
8	-	-	92	656	-	-	-	96	276	376
9	1081	201	70	1011	63	103	41	-	-	-

Tabela 4: Dados de laboratórios diferentes (incidência de câncer nos seios)

Assim a matriz X é dada por:

$$X = \begin{pmatrix} 0.34 & 0.18 & 0.15 & 0.29 & 0.08 & 0.22 & 0.56 \\ 0.40 & 0.11 & 0.13 & 0.31 & 0.11 & 0.19 & 0.42 \\ 0.45 & 0.18 & 0.12 & 0.25 & 0.15 & 0.08 & 0.48 \\ 0.56 & 0.05 & 0.12 & 0.36 & 0.14 & 0.14 & - \\ 0.44 & 0.20 & 0.13 & 0.29 & 0.26 & 0.25 & 0.39 \\ 0.39 & 0.19 & 0.17 & - & 0.08 & 0.17 & 0.41 \\ - & - & 0.30 & 0.29 & - & 0.36 & 0.42 \\ 0.45 & 0.20 & 0.15 & - & 0.12 & - & 0.50 \\ 0.44 & - & - & 0.16 & 0.06 & 0.20 & - \end{pmatrix} \quad (10)$$

Com as mesmas suposições consideradas anteriormente e com as mesmas distribuições a priori para os parâmetros dadas em (7), onde β_m , $m=1,2,\dots,7$ são os parâmetros de regressão associados com as 7 covariáveis, as distribuições condicionais necessárias para o algoritmo Gibbs Sampling são dadas por:

$$\begin{aligned}
 \text{(i)} \quad \sigma^2 | \underset{\sim}{\beta}, \underset{\sim}{\delta}, \underset{\sim}{y} &\sim \text{IG} \left[\frac{n}{2} + c; d + \frac{1}{2} \sum_{i=1}^n (y_i - x_{1i}\beta_1 - x_{2i}\beta_2 - x_{3i}\beta_3 - x_{4i}\beta_4 - x_{5i}\beta_5 - \right. \\
 &\quad \left. - x_{6i}\beta_6 - x_{7i}\beta_7 - w_{1i}\delta_1 - w_{2i}\delta_2)^2 \right] \\
 \text{(ii)} \quad \beta_1 | \underset{\sim}{\beta}_{(1)}, \underset{\sim}{\delta}, \underset{\sim}{\sigma}^2, \underset{\sim}{y} &\sim \text{N} \left\{ \frac{k_{12}}{k_{11}}; \frac{1}{k_{11}} \right\}; \\
 \text{(iii)} \quad \beta_2 | \underset{\sim}{\beta}_{(2)}, \underset{\sim}{\delta}, \underset{\sim}{\sigma}^2, \underset{\sim}{y} &\sim \text{N} \left\{ \frac{k_{22}}{k_{21}}; \frac{1}{k_{21}} \right\}; \\
 \text{(iv)} \quad \beta_3 | \underset{\sim}{\beta}_{(3)}, \underset{\sim}{\delta}, \underset{\sim}{\sigma}^2, \underset{\sim}{y} &\sim \text{N} \left\{ \frac{k_{32}}{k_{31}}; \frac{1}{k_{31}} \right\}; \\
 \text{(v)} \quad \beta_4 | \underset{\sim}{\beta}_{(4)}, \underset{\sim}{\delta}, \underset{\sim}{\sigma}^2, \underset{\sim}{y} &\sim \text{N} \left\{ \frac{k_{42}}{k_{41}}; \frac{1}{k_{41}} \right\}; \\
 \text{(vi)} \quad \beta_5 | \underset{\sim}{\beta}_{(5)}, \underset{\sim}{\delta}, \underset{\sim}{\sigma}^2, \underset{\sim}{y} &\sim \text{N} \left\{ \frac{k_{52}}{k_{51}}; \frac{1}{k_{51}} \right\}; \\
 \text{(vii)} \quad \beta_6 | \underset{\sim}{\beta}_{(6)}, \underset{\sim}{\delta}, \underset{\sim}{\sigma}^2, \underset{\sim}{y} &\sim \text{N} \left\{ \frac{k_{62}}{k_{61}}; \frac{1}{k_{61}} \right\}; \\
 \text{(viii)} \quad \beta_7 | \underset{\sim}{\beta}_{(7)}, \underset{\sim}{\delta}, \underset{\sim}{\sigma}^2, \underset{\sim}{y} &\sim \text{N} \left\{ \frac{k_{72}}{k_{71}}; \frac{1}{k_{71}} \right\}; \\
 \text{(ix)} \quad \delta_1 | \underset{\sim}{\delta}_2, \underset{\sim}{\beta}, \underset{\sim}{\sigma}^2, \underset{\sim}{y} &\sim \text{N} \left\{ \frac{r_{12}}{r_{11}}; \frac{1}{r_{11}} \right\}; \\
 \text{(x)} \quad \delta_2 | \underset{\sim}{\delta}_1, \underset{\sim}{\beta}, \underset{\sim}{\sigma}^2, \underset{\sim}{y} &\sim \text{N} \left\{ \frac{r_{22}}{r_{21}}; \frac{1}{r_{21}} \right\}.
 \end{aligned} \tag{11}$$

onde $\underset{\sim}{\beta}_{(i)} = (\beta_j, \beta_k)$; $j \neq i$, $k \neq i$; $i, j, k = 1, 2, 3, 4, 5, 6, 7$;

$$k_{j1} = \frac{1}{b^2} + \frac{b_{jj}}{\sigma^2}; \quad j=1, 2, \dots, 7;$$

$$k_{12} = \frac{a_1}{\sigma^2} - \frac{b_{12}\beta_2}{\sigma^2} - \frac{b_{13}\beta_3}{\sigma^2} - \frac{b_{14}\beta_4}{\sigma^2} - \frac{b_{15}\beta_5}{\sigma^2} - \frac{b_{16}\beta_6}{\sigma^2} - \frac{b_{17}\beta_7}{\sigma^2} - \frac{c_{11}\delta_1}{\sigma^2} - \frac{c_{12}\delta_2}{\sigma_2},$$

$$k_{22} = \frac{a_2}{\sigma^2} - \frac{b_{12}\beta_1}{\sigma^2} - \frac{b_{23}\beta_3}{\sigma^2} - \frac{b_{24}\beta_4}{\sigma^2} - \frac{b_{25}\beta_5}{\sigma^2} - \frac{b_{26}\beta_6}{\sigma^2} - \frac{b_{27}\beta_7}{\sigma^2} - \frac{c_{21}\delta_1}{\sigma^2} - \frac{c_{22}\delta_2}{\sigma_2},$$

$$k_{32} = \frac{a_3}{\sigma^2} - \frac{b_{13}\beta_1}{\sigma^2} - \frac{b_{23}\beta_2}{\sigma^2} - \frac{b_{34}\beta_4}{\sigma^2} - \frac{b_{35}\beta_5}{\sigma^2} - \frac{b_{36}\beta_6}{\sigma^2} - \frac{b_{37}\beta_7}{\sigma^2} - \frac{c_{31}\delta_1}{\sigma^2} - \frac{c_{32}\delta_2}{\sigma_2},$$

$$k_{42} = \frac{a_4}{\sigma^2} - \frac{b_{14}\beta_1}{\sigma^2} - \frac{b_{23}\beta_2}{\sigma^2} - \frac{b_{34}\beta_3}{\sigma^2} - \frac{b_{45}\beta_5}{\sigma^2} - \frac{b_{46}\beta_6}{\sigma^2} - \frac{b_{47}\beta_7}{\sigma^2} - \frac{c_{41}\delta_1}{\sigma^2} - \frac{c_{42}\delta_2}{\sigma_2},$$

$$k_{52} = \frac{a_5}{\sigma^2} - \frac{b_{15}\beta_1}{\sigma^2} - \frac{b_{25}\beta_2}{\sigma^2} - \frac{b_{35}\beta_3}{\sigma^2} - \frac{b_{45}\beta_4}{\sigma^2} - \frac{b_{56}\beta_6}{\sigma^2} - \frac{b_{57}\beta_7}{\sigma^2} - \frac{c_{51}\delta_1}{\sigma^2} - \frac{c_{52}\delta_2}{\sigma_2},$$

$$k_{62} = \frac{a_6}{\sigma^2} - \frac{b_{16}\beta_1}{\sigma^2} - \frac{b_{26}\beta_2}{\sigma^2} - \frac{b_{36}\beta_3}{\sigma^2} - \frac{b_{46}\beta_4}{\sigma^2} - \frac{b_{56}\beta_5}{\sigma^2} - \frac{b_{67}\beta_7}{\sigma^2} - \frac{c_{61}\delta_1}{\sigma^2} - \frac{c_{62}\delta_2}{\sigma_2},$$

$$k_{72} = \frac{a_7}{\sigma^2} - \frac{b_{17}\beta_1}{\sigma^2} - \frac{b_{27}\beta_2}{\sigma^2} - \frac{b_{37}\beta_3}{\sigma^2} - \frac{b_{47}\beta_4}{\sigma^2} - \frac{b_{57}\beta_5}{\sigma^2} - \frac{b_{67}\beta_6}{\sigma^2} - \frac{c_{71}\delta_1}{\sigma^2} - \frac{c_{72}\delta_2}{\sigma_2},$$

$$r_{j1} = \frac{1}{a^2} + \frac{e_{jj}}{\sigma^2}; \quad j=1,2;$$

$$r_{12} = \frac{d_1}{\sigma^2} - \frac{c_{11}\beta_1}{\sigma^2} - \frac{c_{21}\beta_2}{\sigma^2} - \frac{c_{31}\beta_3}{\sigma^2} - \frac{c_{41}\beta_4}{\sigma^2} - \frac{c_{51}\beta_5}{\sigma^2} - \frac{c_{61}\beta_6}{\sigma^2} - \frac{c_{71}\beta_7}{\sigma^2} - \frac{\delta_2 e_{12}}{\sigma_2},$$

$$r_{22} = \frac{d_2}{\sigma^2} - \frac{c_{12}\beta_1}{\sigma^2} - \frac{c_{22}\beta_2}{\sigma^2} - \frac{c_{32}\beta_3}{\sigma^2} - \frac{c_{42}\beta_4}{\sigma^2} - \frac{c_{52}\beta_5}{\sigma^2} - \frac{c_{62}\beta_6}{\sigma^2} - \frac{c_{72}\beta_7}{\sigma^2} - \frac{\delta_2 e_{12}}{\sigma_2},$$

$$a_j = \sum_{i=1}^n y_i x_{ji}; \quad j=1,2,\dots,7;$$

$$b_{jk} = \sum_{i=1}^n x_{ji} x_{ki}; \quad j,k=1,2,\dots,7;$$

$$c_{jk} = \sum_{i=1}^n x_{ji} w_{ki}; \quad j=1,2,\dots,7; k=1,2;$$

$$d_j = \sum_{i=1}^n y_i w_{ji}; \quad j=1,2;$$

$$e_{jk} = \sum_{i=1}^n w_{ji} w_{ki}; \quad j,k=1,2;$$

Para as observações perdidas na matriz X dada em (10) usamos o mesmo procedimento de geração usado para os dados da tabela 1.

Com as distribuições a priori (7) com $a=0.2$, $b=0.1$, $c=1$, $d=1$, geramos cinco cadeias de Gibbs cada uma com 2000 iterações, e após descartar as 400 primeiras observações, tomamos cada décima observação na nossa amostra final. Assim em cinco cadeias, temos uma amostra de 800 observações.

Na tabela 5, temos os sumários a posteriori e os valores de \hat{R} (critério de Gelman & Rubin). Para todos os parâmetros, observamos a convergência do algoritmo Gibbs Sampling ($\sqrt{\hat{R}} < 1.1$ para todos os parâmetros).

Parâmetros	Média a posteriori	Intervalo de Credibilidade (95%)	\hat{R}
β_1	0.0081	(-0.0124 ; 0.0304)	1.0025
β_2	0.0036	(-0.0157 ; 0.0226)	1.0007
β_3	0.0044	(-0.0145 ; 0.0253)	1.0012
β_4	0.0025	(-0.0174 ; 0.0224)	1.0018
β_5	0.0013	(-0.0183 ; 0.0220)	1.0034
β_6	0.0051	(-0.0149 ; 0.0261)	1.0041
β_7	0.0090	(-0.0113 ; 0.0319)	1.0002
δ_1	0.0196	(-0.0329 ; 0.0690)	1.0055
δ_2	0.0551	(0.0054 ; 0.0927)	1.0022
σ^2	0.2030	(0.0835 ; 0.4550)	1.0030

Tabela 5: Sumários a posteriori

Na figura 5 temos os gráficos das distribuições a posteriori marginais aproximadas pelas amostras geradas pelo algoritmo de Gibbs.

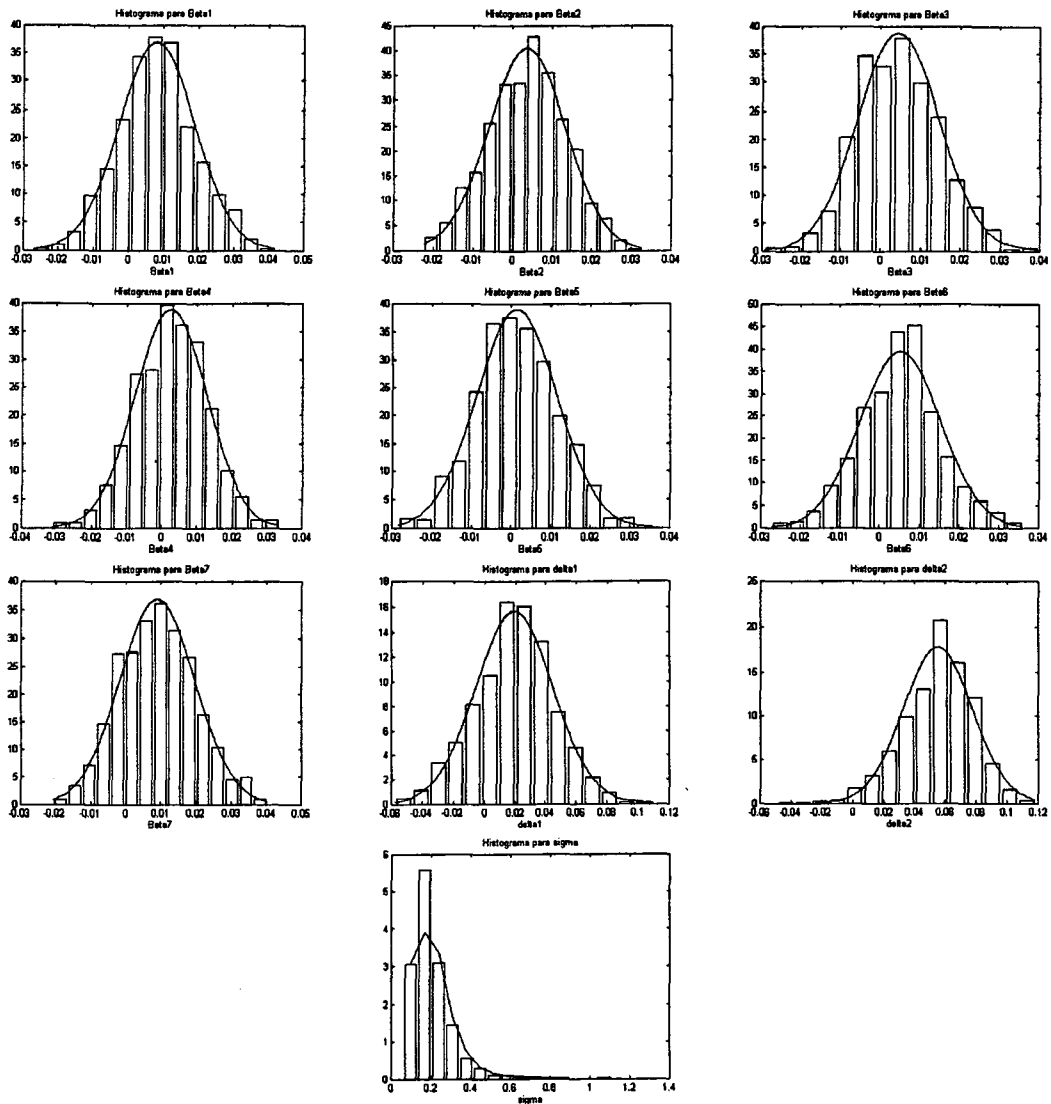


Figura 5: Densidades a posteriori marginais

A partir dos resultados dados na tabela 5, observamos que as covariáveis x_1, x_2, \dots, x_7 não apresentam efeitos significativos na resposta $y = \log(RR)$ onde RR é o risco relativo de câncer entre as usuárias de estrógeno.

4. ALGUMAS CONCLUSÕES

A possível relação entre incidência de câncer em mulheres (seios, útero e ovários) e terapia de hormônios é um problema de grande interesse médico. O uso de meta-análise permite a combinação dos resultados provenientes de vários estudos independentes relacionando terapia de substituição de hormônios e câncer.

Os resultados obtidos a partir do modelo (2) proposto para analisar os dados podem ser estendidos para outros conjuntos de dados. Larose & Dey (1995) obtiveram resultados similares, considerando uma versão do modelo hierárquico de componentes de variância proposto por DuMouchel & Harris (1983).

O uso de métodos de simulação MCMC torna viável o desenvolvimento Bayesiano.

REFERÊNCIAS

- ADAMI, H. O. (1992). "Long-Term consequences of estrogen and estrogen-progestin replacement", *Cancer Causes and Control*, 3, 83-90.
- BRINTON, L. A.; SCHAIRER, C. (1993). "Estrogen replacement therapy and breast cancer risk", *Epidemiologic Reviews*, 15, 66-79.
- DuMOUCHEL, W. H.; HARRIS, J. E. (1983). "Bayes methods for combining the results of cancer studies in humans and other species (with discussion)", *Journal of the American Statistical Association*, 78, 293-315
- FLEISS, J. L.(1973). "Statistical methods for rates & proportions", John Wiley & Sons.
- GELFAND, A. E.; SMITH, A. F. M. (1990). "Sampling-based approaches to calculating marginal densities", *Journal of the American Statistical Association*, 85, 398-409.
- GELMAN, A. E.; RUBIN, D. (1992). "Inference from iterative simulation using multiple sequences", *Statistical Science*, 7, 457-472.
- HEDGES, L. V.; OLKIN, I. (1985). "Statistical methods for meta-analysis", Academic Press.
- KAUFMAN, D. W. et al (1991). "Estrgen replacement therapy and the risk of breast cancer: results from the case control surveillance study", *American Journal of Epidemiology*, 134, 1375-1385.
- LAROSE, D. T.; DEY, D. K. (1995). "Is estrogen linked to cancers of the breast and endometrium? New meta-analyses using Bayesian grouped random effects model", relatório técnico 95-35, departamento de estatística, Universidade de Connecticut, Storrs.

NOTAS DO ICMC

SÉRIE ESTATÍSTICA

- 053/98 CID, J. E.R.; ACHCAR, J.A. – Bayesian inference for nonhomogeneous poisson processes in software reliability models assuming nonmonotonic intensity functions.
- 052/98 ANDRADE FILHO, M.G; MIZOI, M.F. – Aplicação de MCMC na estimação de máxima verossimilhança para processos AR(p) e MA(q).
- 051/98 RODRIGUES, J. – Bayesian analysis for the accelerated life tests with informative prior distributions obtained from fixed stresses.
- 050/98 ACHCAR, J.A.; ANDRADE, .M.G.; LOIBEL, S. - A bayesian analysis for homogeneous poisson processes with change-points.
- 049/98 ACHCAR, J.A; PEREIRA, G.A. - Use of mixture of exponential power distributions for interval-censored survival data in presence of covariates.
- 048/98 CID, J.E.R.; ACHCAR, J.A. - Software reliability considering the superposition of non-homogeneous Poisson processes in the presence of a covariate.
- 047/98 ACHCAR, J.A.; PEREIRA, G.A. - Use of exponential power distributions for mixture models in the presence of covariates.
- 046/98 ANDRADE, M.G.; HUTTER, C.F.F. - Teste de sazonalidade para função de autocorrelação de processos auto-regressivos periódicos - PAR (pm)
- 045/98 ACHCAR, J.A.; ANDRADE, M.G.; LOIBEL, S. - Weibull hazard function with a change-point: a bayesian approach using Markov chain Monte carlo methods.
- 044/97 ACHCAR, J.A.; PEREIRA, G.A. - Bayesian analysis of mixture models for survival data: some computational aspects.