MAE0524 - Análise Bayesiana de dados - primeiro semestre de 2019
Professora: Márcia D'Elia Branco

LISTA 5

1) Um caso especial do modelo linear geral é o modelo de análise de variância usado para comparação de médias de K populações independentes normais com variâncias iguais.

a) Considere n_j o tamanho da j-ésima amostra, com $j = 1, \ldots, K$ e $\mu_1, \ldots, \mu_K$ as médias populacionais. Escreva o modelo de análise de variância na estrutura do modelo linear geral, especificando a matriz de planejamento X e relacionando as médias populacionais com o vetor de parâmetros β.

Usando os resultados apresentados em aula para o ML e considerando suposições adequadas, responda a questão a seguir:

b)Um conjunto de 19 porcos foi subdividido em 4 grupos, cada um dos quais foi alimentado durante certo período com diferentes tipos de ração. Ao fim desse período os porcos foram pesados e os valores são apresentados na tabela abaixo.

Tipo de Ração	Peso (kg)				
A	133.8	125.2	143.1	128.9	135.7
B	151.2	149.0	162.7	145.8	153.5
C	225.8	224.6	220.4	212.3	
D	193.4	185.3	182.8	188.5	198.6

b.1)Compare as médias dos pesos dos grupos C e D.

b.2) Faça uma análise de variância bayesiana do conjunto total de dados.

2) Considere o modelo de regressão linear simples com erros t-Student independentes. Obtenha a distribuição *a posteriori* do vetor (α, β) condicional a σ^2 e $W = (w_1, \ldots, w_n)$, em que W é o vetor das variáveis latentes usadas na construção hierárquica do modelo.

3) Deseja-se ajustar o modelo não linear $\beta_0 x^{\beta_1}$, relacionando o peso (y) com o comprimento (x) de determinada espécie de peixe. Para isso será utilizada uma linearização do modelo dada por:

$$log(y) = \beta_0^* + \beta_1 log(x) + \epsilon,$$

em que $\beta_0^* = log(\beta_0)$ e $\epsilon \sim N(0, \sigma^2)$.

Na tabela abaixo apresentamos o comprimento (em cm) e o peso (em kg) de 25 exemplares de um peixe capturados na costa sul do Brasil.

x	10.2	19.0	22.3	23.8	24.9	26.1	26.9	27.9	28.5	29.2	30.1	31.0	31.7
y	0.01	0.09	0.12	0.18	0.19	0.23	0.24	0.31	0.33	0.32	0.33	0.40	0.45
x	32.4	32.9	33.3	34.1	34.6	35.5	36.3	37.0	38.0	39.0	39.6	43.0	
y	0.49	0.67	0.48	0.54	0.63	0.59	0.65	0.50	0.81	0.87	0.71	0.91	

a) Construa os diagramas de dispersão de y versus x e de $log(y)$ versus $log(x)$. Comente.

b) Considerando *a priori* de Jeffreys, especifique a distribuição *a posteriori* conjunta de (β_0^*, β_1). Simule uma amostra de tamanho $M = 1000$ desta distribuição.

c) Com base na amostra simulada em (b), obtenha uma amostra de (β_0, β_1). Aproxime as distribuições *a posteriori* marginais via histograma dos dados simulados. Obtenha os intervalos de credibilidade 0.90 para β_0 e β_1.

d) Considere que a razão $\frac{\beta_1}{(10^5\beta_0)}$ seja de interesse para estudos sobre biologia da espécie. Aproxime a distribuição *a posteriori* dessa quantidade via histograma, obtenha estimativas pontuais para ela, bem como, o intervalo de credibilidade 0.90.

4) Considere um experimento onde os indivíduos indicam o número de eventos estressantes (y_i) aos quais foram submetidos em determinado mês (i). Os dados são apresentados na tabela a seguir

Mês	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
y	15	11	14	17	5	11	10	4	8	10	7	9	11	3	6	1	1	4

Vamos considerar o modelo de regressão loglinear Poisson:

$$log\lambda_i = \beta_0 + \beta_1 i$$

em que $Y_i \mid \lambda_i \sim Poisson(\lambda_i)$.

(a) Sob distribuição *a priori* uniforme (imprópria) para (β_0, β_1). Escreva uma função no **R** para calcular o logaritmo da posteriori.

(b) Use o algoritmo de metropolis com proposta passeio aleatório para simular uma amostra de tamanho 1000 da distribuição *a posteriori* de β_1. Apresente os valores simulados usando o gráfico *box-plot*.

(c) Obtenha um intervalo de credibilidade 0.9 para β_1.

(d) Considerando o intervalo construído em (c), você diria que há eviências para rejeitar a hipótese $H_0 : \beta_1 = 0$?

5) Num estudo para implantação de turbinas eólicas numa dada zona mediu-se a velocidade X do vento (em m/s) a uma dada altura ao longo de várias ocasiões, obtendo-se os dados $x = (x_1, \ldots, x_n)$. O modelo utilizado para descrever X é o modelo Weibull, cuja f.d.p. é proporcional a

$$\delta \alpha x^{\alpha-1} \exp\left[-\delta x^{\alpha}\right] I_{(0,\infty)}(x) \ , \ \delta > 0, \alpha > 0.$$

Considere as seguintes distribuições *a priori*: $\delta \sim Gama(a,b)$ e $\alpha \sim Log - normal(c,d)$, onde $a, b, c, b > 0$.

Obtenha as distribuições condicionais completas e indique como deveria ser implementado o algoritmo de Gibbs para obter uma amostra da distribuição *a posteriori*.

6) Seja $Y \mid p \sim Bin(n,p)$ e $\theta = log\left(\frac{p}{1-p}\right)$. Considere θ com distribuição *a priori* Normal com média zero e desvio padrão 0.25. Para $n = 5$ e $y = 5$, determine a probabilidade *a posteriori* de $\theta > 0$ usando as duas aproximações indicadas a seguir. (Note que $\theta > 0$ equivale a $p > 0.5$)

(a) Use a aproximação normal para a densidade *a posteriori*.

(b) Use o algoritmo de Metropolis-Hasting com proposta passeio aleatório. No algoritmo considere o desvio padrão da proposta igual a duas vezes o desvio padrão da aproximação normal obtida em (a).

Entrega: 18/06/2019