

Departamento de Estatística - Universidade de Brasília

# **Modelo Tobit em uma abordagem bayesiana**

Matheus Yair Medeiros Porto

Brasília

2017



Matheus Yair Medeiros Porto

## **Modelo Tobit em uma abordagem bayesiana**

Projeto apresentado para obtenção do título de Bacharel em Estatística ao Departamento de Estatística da Universidade de Brasília

Orientador: Prof. Dr. Leandro Tavares Correia

Brasília  
2017



# Resumo

Este trabalho apresenta aplicações para o modelo de regressão tobit, considerando uma abordagem bayesiana. As estimativas pontuais e seus respectivos intervalos HPD foram encontrados realizando as inferências bayesianas com o uso das técnicas de MCMC (Markov Chain Monte Carlo). Posteriormente, essa metodologia será aplicada em conjuntos de dados reais, atestando a validade da mesma por meio de técnicas de diagnóstico, e também pela comparação dos resultados obtidos com estudos feitos com abordagem frequentista para o mesmo banco de dados. As estimativas geradas, gráficos e simulações foram elaborados pelo software R.

**Palavras-chave:** Modelo Tobit, MCMC, Intervalo HPD, Inferência Bayesiana.



# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>7</b>
<b>1.1</b>	<b>Objetivo</b>	<b>8</b>
1.1.1	Objetivo Geral	8
1.1.2	Objetivos Específicos	8
<b>1.2</b>	<b>Metodologia</b>	<b>9</b>
<b>2</b>	<b>MODELO DE REGRESSÃO LINEAR MÚLTIPLO</b>	<b>11</b>
<b>2.1</b>	<b>Estimação dos parâmetros</b>	<b>11</b>
<b>2.2</b>	<b>Estimadores de Mínimos Quadrados</b>	<b>12</b>
<b>2.3</b>	<b>Propriedades dos Estimadores</b>	<b>12</b>
<b>2.4</b>	<b>Técnicas de Diagnóstico</b>	<b>13</b>
2.4.1	Alavancagem	13
2.4.2	Resíduos	13
2.4.3	Distância de Cook	14
<b>2.5</b>	<b>Exemplo</b>	<b>14</b>
<b>3</b>	<b>MODELO DE REGRESSÃO LINEAR BAYESIANA</b>	<b>17</b>
<b>3.1</b>	<b>O Teorema de Bayes</b>	<b>17</b>
<b>3.2</b>	<b>Regressão Bayesiana</b>	<b>17</b>
<b>3.3</b>	<b>Distribuição a Priori</b>	<b>18</b>
<b>3.4</b>	<b>Função de Verossimilhança</b>	<b>19</b>
<b>3.5</b>	<b>Distribuição a Posteriori</b>	<b>19</b>
<b>3.6</b>	<b>Estimação Intervalar</b>	<b>21</b>
<b>3.7</b>	<b>Técnicas de Diagnóstico</b>	<b>21</b>
3.7.1	Resíduos bayesianos Studentizados	21
3.7.2	Comparação de Modelos	22
3.7.2.1	DIC	22
3.7.2.2	CPO	23
3.7.3	Pontos de Influência	23
3.7.3.1	Distância de Kullback-Leibler	23
<b>3.8</b>	<b>Exemplo</b>	<b>24</b>
<b>4</b>	<b>DADOS LIMITADOS</b>	<b>27</b>
<b>4.1</b>	<b>Truncamento</b>	<b>27</b>
4.1.1	Distribuição Truncada	27
<b>4.2</b>	<b>Censura</b>	<b>27</b>

4.2.1	Distribuição Censurada . . . . .	28
<b>4.3</b>	<b>Modelo Tobit . . . . .</b>	<b>28</b>
4.3.1	Estimação Bayesiana . . . . .	28
4.3.2	Aumento de Dados (Data Augmentation) e Amostrador de Gibbs . . . . .	30
<b>5</b>	<b>ESTUDO DE SIMULAÇÕES . . . . .</b>	<b>33</b>
<b>5.1</b>	<b>Simulação 1 . . . . .</b>	<b>33</b>
<b>5.2</b>	<b>Simulação 2 . . . . .</b>	<b>34</b>
<b>6</b>	<b>APLICAÇÕES EM DADOS REAIS . . . . .</b>	<b>37</b>
<b>6.1</b>	<b>Relações Extraconjugais . . . . .</b>	<b>37</b>
<b>6.2</b>	<b>Horas trabalhadas por mulheres casadas . . . . .</b>	<b>40</b>
<b>7</b>	<b>CONCLUSÃO . . . . .</b>	<b>45</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>47</b>
	<b>APÊNDICE A – RESÍDUOS DO MODELO TOBIT . . . . .</b>	<b>49</b>
	<b>ANEXO A – R SCRIPT . . . . .</b>	<b>51</b>



# 1 Introdução

No contexto da modelagem estatística é comum a ocorrência de casos em que a variável resposta está limitada na extremidade (sendo ela inferior ou superior). Nesses casos os modelos de regressão lineares usuais não são os mais adequados. Podemos contornar esse problema ao interpretarmos o valor extremo da variável resposta como uma censura; assim, a ideia popularizada por Tobin (1958), denominada Modelo Tobit, que é o modelo de principal foco desse projeto.

Como exemplo de dados reais, vemos o caso da Vacinação no Haiti, estudado por Moulton e Haulsey (1995). Os dados consistem de uma amostra de 330 crianças com até um ano de idade, que foram vacinadas contra o sarampo, onde a concentração de anticorpos foi medida. Nos laboratórios onde as medições foram realizadas, existia um limite de detecção mínimo (LMD) de 0.1 mm/l. Isso causa uma censura para concentrações menores que 0.1. Para ilustrar o problema, temos na figura (1) o gráfico de uma densidade normal padrão contendo um limite inferior igual a -1.5 sobreposta a uma normal padrão.

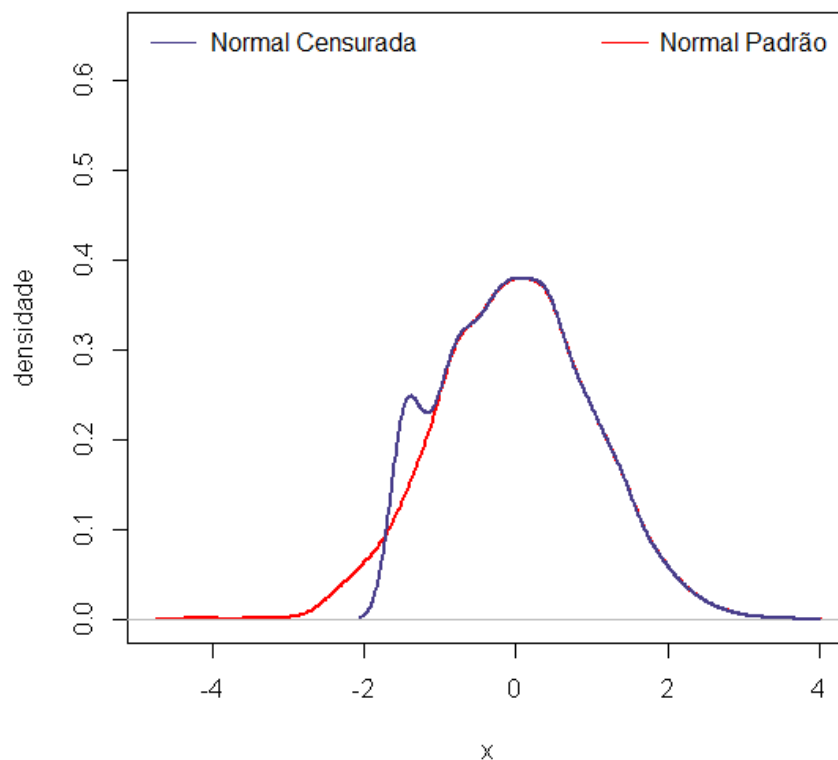


Figura 1 – Comparação de uma distribuição Normal padrão e uma distribuição Normal com censura inferior no ponto -1.5.

Podemos notar, na figura (1), a diferença entre o gráfico, de uma distribuição normal censurada e uma distribuição normal sem censura.

Além do exemplo citado, podemos aplicar o Modelo Tobit em diversas áreas, como econometria, biometria, ensaios clínicos etc.

A estrutura da variável resposta no Modelo Tobit normal é apresentada a seguir. Considere um conjunto de dados com limite inferior no valor  $c$ . Sem perda de generalidade, a variável observada  $y$  pode ser expressa com o auxílio de uma variável latente  $y^*$  (não observada) da seguinte forma:

$$y_i = \begin{cases} c, & \text{se } y^* \leq c \\ y^*, & \text{se } y^* > c \end{cases} \quad (1.1)$$

sendo  $y^* \sim \text{Normal}(\mu; \sigma^2)$ .

Para modelagem e análise do problema prático, usaremos a abordagem Bayesiana do Modelo Tobit. Como justificativa, podemos argumentar a praticidade na estimação de funções dos parâmetros usando técnicas de Monte Carlo, assim como o acréscimo de informações valiosas ao modelo por meio de distribuições *a priori*. Além disso, evitamos o uso de aproximações assintóticas. Usaremos também o método de Aumento de dados (Data Augmentation), que pode ser visto como uma forma de completar os dados presentes na censura. Com isso teremos a distribuição completa do modelo, deixando de ser uma distribuição deformada como vemos no exemplo da figura (1). Posteriormente, no Capítulo 4, veremos com mais clareza esse método.

## 1.1 Objetivo

### 1.1.1 Objetivo Geral

- Análise e modelagem de dados com a variável resposta limitada.

### 1.1.2 Objetivos Específicos

- Uso da Inferência Bayesiana para a modelagem de dados.
- Estudo de distribuições censuradas e estrutura do Modelo Tobit.
- Uso de métodos computacionais, como o algoritmo Monte Carlo Cadeia de Markov (MCMC).
- Aplicação de dados reais.

## 1.2 Metodologia

O projeto será constituído inicialmente por um capítulo de motivação ao tema abordado no trabalho, seguido por um capítulo onde revisaremos o modelo de regressão linear clássico e o modelo de regressão Bayesiano, sendo o segundo utilizado nesse projeto. Será feito o estudo das distribuições censuradas, seguido pela apresentação do Modelo Tobit, que é o principal foco deste trabalho. A modelagem Bayesiana do Modelo Tobit será aplicada com o auxílio de métodos computacionais (método MCMC). Por fim, aplicaremos o modelo em dados reais. Todas as análises serão realizadas com o auxílio do software livre R, por meio, principalmente, do pacote *BayesCR*.



## 2 Modelo de Regressão Linear Múltiplo

Considere o modelo geral de regressão linear múltiplo,

$$y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i(p-1)} + \epsilon_i \quad (2.1)$$

Tendo  $Y_i$  como a variável resposta,  $X_{i(p-1)}$  variáveis preditoras,  $\beta_{(p-1)}$  os parâmetros regressores e  $\epsilon_i$  representando o erro aleatório. As variáveis presentes no modelo podem ser definidas em forma matricial

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1(p-1)} \\ 1 & x_{21} & \cdots & x_{2(p-1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & x_{n(p-1)} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

Assim fica definido  $Y$  como um vetor  $n \times 1$ ,  $X$  uma matrix  $n \times p$  e o vetor  $\beta$   $p \times 1$ . Então o modelo é escrito na forma matricial:

$$Y = X\beta + \epsilon,$$

com os seguintes pressupostos: os erros  $\epsilon$  são IID com  $\epsilon \sim N(0; \sigma^2 I)$ .

### 2.1 Estimação dos parâmetros

Supondo que o modelo de regressão linear múltiplo segue uma distribuição normal multivariada, de ordem  $n$ , com vetor de média e matriz de covariância dados por, respectivamente:

$$\begin{aligned} E[y | x] &= E[X\beta + \epsilon] \\ &= E[X\beta] + E[\epsilon] \\ &= X\beta + 0 \\ &= X\beta \end{aligned} \quad (2.2)$$

$$\begin{aligned} Var[y | x] &= Var[X\beta + \epsilon] \\ &= Var[X\beta] + Var[\epsilon] \\ &= \sigma^2 I \end{aligned} \quad (2.3)$$

Ambos os parâmetros  $\beta$  e  $\sigma^2$  são desconhecidos, e por isso precisam ser estimados.

## 2.2 Estimadores de Mínimos Quadrados

Esse método consiste na escolha do valor que minimiza a soma de quadrados das diferenças. Para o parâmetro  $\beta$  a função de mínimos quadrados a ser analisada é:

$$\begin{aligned}
 SQ(y - X\beta) &= \sum (y_i - x_i'\beta)^2 \\
 &= (Y - X\beta)'(Y - X\beta) \\
 &= (Y' - \beta'X')(Y - X\beta) \\
 &= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta
 \end{aligned} \tag{2.4}$$

Note que  $Y'X\beta$  e  $\beta'X'Y$  são iguais, pois são matrizes escalares e uma é transposta da outra. Com isso obtemos,

$$SQ(\beta) = Y'Y - 2(Y'X\beta) + \beta'X'X\beta \tag{2.5}$$

Derivando a função em termos de  $\beta$ , temos

$$\begin{aligned}
 \frac{\partial Y'Y}{\partial \beta} - 2\frac{\partial Y'X\beta}{\partial \beta} + \frac{\partial \beta'X'X\beta}{\partial \beta} &= 0 - 2(Y'X)' + 2X'X\beta \\
 &= -2(X'Y) + 2X'X\beta
 \end{aligned} \tag{2.6}$$

Agora é calculado o vetor  $\beta$  que minimiza a função: igualamos a equação a 0 e denominamos o vetor resultante de  $\hat{\beta}$ :

$$\begin{aligned}
 -2(X'Y) + 2X'X\hat{\beta} &= 0 \\
 X'X\hat{\beta} &= X'Y \\
 (X'X)^{-1}X'X\hat{\beta} &= (X'X)^{-1}X'Y \\
 \hat{\beta} &= (X'X)^{-1}X'Y
 \end{aligned} \tag{2.7}$$

Logo a solução da função para o método de mínimos quadrado desde que  $(X'X)^{-1}$  exista é

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{2.8}$$

## 2.3 Propriedades dos Estimadores

Assumindo que  $E(\epsilon) = 0$  e  $Var(\epsilon) = \sigma^2 I_n$ , temos que o estimador de  $\beta$ ,  $\hat{\beta}$  é um estimador sem viés

$$E(\hat{\beta}) = \beta \tag{2.9}$$

$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1} \tag{2.10}$$

Um estimador de  $\sigma^2$  é dado por

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n - p} \tag{2.11}$$

que é o resultado da divisão entre a soma de quadrados dos resíduos dividido por  $(n - p)$ , sendo  $p$  a dimensão de  $\beta$ . Logo substituindo o  $\sigma^2$  pelo seu estimador  $\widehat{\sigma}^2$ , encontramos o estimador da variância de  $\widehat{\beta}$ ,

$$\widehat{Var}(\widehat{\beta}) = \widehat{\sigma}^2(X'X)^{-1}$$

## 2.4 Técnicas de Diagnóstico

Após estimarmos os parâmetros de um modelo específico, é necessária a verificação do modelo para sabermos se o mesmo é, de fato, o mais adequado para a situação e, para termos certeza de que os resultados são confiáveis. A Análise dos Resíduos consiste em um conjunto de técnicas utilizadas para investigar a adequabilidade de um modelo de regressão com base nos resíduos. Caso os resíduos mostrem algo errado no modelo, conclusões equivocadas podem ser obtidas.

Por isso a Análise de Diagnóstico se torna tão importante para a validação de um modelo. Serão utilizados as seguintes metodologias para essa análise:

### 2.4.1 Alavancagem

Seja  $\mathbf{H}$  a matriz simétrica de projeção ortogonal de vetores do  $\mathfrak{R}^n$  no subespaço gerado pelas colunas da matriz, onde  $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$  é simétrica e idempotente. Seus elementos são denotados por  $h_{ii}$  porém usaremos apenas sua diagonal principal, sendo assim denotado apenas por  $h_i$ . Esses elementos desempenham um papel importante para as técnicas de diagnóstico, pois são utilizados para verificar a discrepância existente entre os valores de  $X$  da  $i$ -ésima observação e da média geral de  $X$  e como base para outras inúmeras técnicas. Pela existência da idempotência da matriz  $\mathbf{H}$  temos que  $\text{posto}(\mathbf{H}) = \text{tr}(\mathbf{H}) = \sum_{i=1}^n h_i = p$ .

Em um bom modelo ajustado, esperamos que todos os pontos exerçam a mesma influência sobre os valores ajustados. Dado isso, é esperado que  $h_i$  esteja próximo de  $\frac{\text{tr}(\mathbf{H})}{n} = \frac{p}{n}$ . Com isso examinamos os pontos onde  $h_i \geq \frac{2p}{n}$ . Esses pontos são conhecidos como pontos de alavancagem. Para um conjunto de dados relativamente grande um ponto de alavancagem é considerado alto a partir de 0.5.

### 2.4.2 Resíduos

O resíduo para a  $i$ -ésima observação pode ser definido por meio da equação  $\varepsilon_i = y_i - \widehat{y}_i$ , onde  $\widehat{y}_i$  é o valor esperado para a  $i$ -ésima observação. Os resíduos são usados para a detecção de valores discrepantes para a variável  $Y$ , e o sinal de  $\varepsilon$  indica a direção dessa discrepância. Contudo esses resíduos não são muito informativos, pois não possuem variância constante. Como podemos ver,  $Var(\varepsilon_i) = \sigma^2(1 - h_i)$ , a variância dos resíduos depende de  $h_i$ . Com o intuito de promover uma comparação mais apropriada, os resíduos

são padronizados da seguinte forma:

$$\varepsilon_i^* = \frac{\varepsilon_i}{\sqrt{\sigma^2(1 - h_i)}}$$

Porém como  $\varepsilon_i$  não é independente da variância  $\sigma^2$ ,  $\varepsilon_i^*$  ainda não segue uma distribuição t-student como esperado. Então resolvemos o problema da dependência substituindo  $\sigma^2$  por  $\sigma_{(i)}^2$  que é o erro quadrático médio sem a  $i$ -ésima observação do modelo. O índice ( $i$ ) indica a  $i$ -ésima observação que será excluída do modelo. Assim temos que o novo resíduo padronizado é dado por

$$t_i = \frac{\varepsilon_i}{\sigma_{(i)}\sqrt{1 - h_i}}$$

e segue uma distribuição  $t_{n-p}$ .

### 2.4.3 Distância de Cook

É uma medida de diagnóstico usada para encontrar pontos que estão influenciando negativamente a estimação dos parâmetros da regressão do modelo. É uma combinação entre as observações dos valores de alavancagem e resíduos, e é dada por:

$$D_i = \frac{\varepsilon_i^2}{p\sigma^2} \left( \frac{h_i}{(1 - h_i)^2} \right)$$

Existem diferentes formas de interpretar essa medida: observações maiores que  $4/n$  ou valores acima de 0.5 são considerados altos, por exemplo.

## 2.5 Exemplo

Foram extraídos os dados do censo do IBGE de 2000, onde estão representados os dados referentes ao número médio de anos de estudo e a renda média mensal para cada Unidade da Federação. Foi considerado o seguinte modelo

$$RENDA_i = \beta_0 + \beta_1 ANOS + \varepsilon_i,$$

onde RENDA - Renda média mensal do chefe ou chefes do domicílio. ANOS - Número médio de anos de estudo do chefe ou chefes do domicílio.

Os dados estão representados na seguinte tabela:



Tabela 1 – Renda e escolaridade média dos chefes de domicílio no Brasil em 2000.

UF	Anos	Renda
AC	4.50	526
AL	3.70	454
AM	5.50	627
AP	6.00	683
BA	4.10	460
CE	4.00	448
DF	8.20	1499
ES	5.70	722
GO	5.50	689
MA	3.60	343
MG	5.40	681
MS	5.70	731
MT	5.40	775
PA	4.70	536
PB	3.90	423
PE	4.60	517
PI	3.50	383
PR	6.00	782
RJ	7.10	970
RN	4.50	513
RO	4.90	662
RR	5.70	685
RS	6.40	800
SC	6.30	814
SE	4.30	462
SP	6.80	1076
TO	4.50	520

Foram obtidas as estimativas dos parâmetros e seus respectivos erros-padrão, são eles  $\beta_0 = -381,28$ ,  $\beta_1 = 199,82$  e  $69,40$  e  $13,03$  respectivamente. Isso nos indica que o coeficiente angular da reta é bastante significativa. Podemos interpretar que esse é o aumento esperado na renda média do chefe do domicílio se o tempo de escolaridade aumentar em uma unidade (um ano). Já a estimativa para o parâmetro  $\sigma$  foi igual a  $77,22$ .

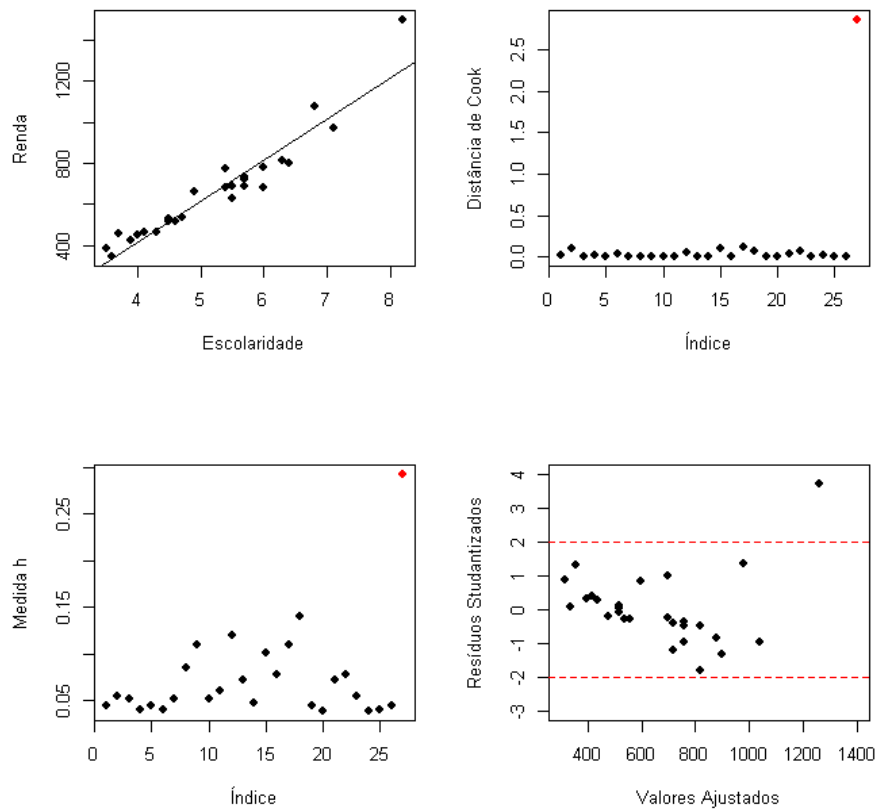


Figura 2 – Gráficos das medidas de diagnóstico para ps dados da tabela 1.

Analisando os gráficos e os resultados obtidos pelas técnicas de diagnóstico, podemos reparar que com o aumento de renda os dados têm maior variabilidade, o que indica que o modelo possui heterocedasticidade com o aumento da variável "Renda". Também podemos observar que a observação 27, que representa a UF do Distrito Federal (DF), possui uma grande discrepância no modelo, pois é um ponto de alavanca, outlier e influente no exemplo. Por esse motivo, poderia ser feito, por exemplo, o estudo de um novo modelo sem essa observação para checarmos se com tal mudança o modelo se encaixaria melhor aos dados, ou uma regressão polinomial, ou transformação log para estabilizar a variância.

## 3 Modelo de Regressão Linear Bayesiana

### 3.1 O Teorema de Bayes

No contexto da estatística Bayesiana, o teorema de Bayes é a ferramenta inicial do processo. Se considerarmos dois eventos  $A$  e  $B$ , a probabilidade de observarmos os dois simultaneamente é dada por:

$$P(A \cap B) = P(B|A) P(A)$$

Também pode ser escrito como

$$P(B \cap A) = P(A|B) P(B)$$

Ou seja, temos que, se  $P(A \cap B) = P(B \cap A)$ :

$$P(A \cap B) = P(B|A) P(A) = P(A|B) P(B)$$

Se rearranjarmos, temos que:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)} \quad (3.1)$$

O Teorema de Bayes propõe, por meio de estudos já realizados sobre o assunto, uma caracterização da aprendizagem, ou seja, o estudo do evento  $A$  não é iniciado do zero existe uma probabilidade inicial, chamada de probabilidade *a priori*. Com o estudo de informações adicionais, que são denotadas pelo evento  $B$ , temos uma mudança sobre os conhecimentos do evento  $A$ . Isso afeta a probabilidade *a priori* e através do teorema de Bayes encontramos a probabilidade *a posteriori*. Podemos estender essa mesma ideia para tratarmos a estimação dos parâmetros de interesse. Supondo  $\theta$  como o vetor de parâmetros e  $x$  sendo o vetor dos eventos de interesse, o Teorema de Bayes é definido como

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} \propto p(x|\theta)p(\theta), \quad (3.2)$$

onde a proporcionalidade  $\propto$  na equação significa que para o modelo  $\frac{1}{p(x)}$  é apenas uma constante e será ignorada quando observarmos  $p(\theta|x)$  como função de  $\theta$ . Depois de entendido o Teorema de Bayes, podemos avançar para os tópicos de estatística bayesiana.

### 3.2 Regressão Bayesiana

Considere uma amostra com  $n$  observações independentes  $(y_i, x_{1i}, \dots, x_{pi}), i = 1, \dots, n$ , num conjunto de  $p$  variáveis aleatórias  $Y, X_1, \dots, X_p$  com distribuição conjunta desconhecida. Comumente esse tipo de caso é analisado pelo modelo de regressão onde assumimos

que a relação entre a variável  $Y$  e as variáveis  $X_1, \dots, X_p$  é a seguinte

$$Y = X\beta + \varepsilon \quad (3.3)$$

, onde  $Y$  é o vetor  $(Y_1, \dots, Y_n)'$  e  $X$  a matriz  $n \times p$  onde o  $(i,j)$ -ésimo termo é o  $x_{ji}$ ,  $\varepsilon$  é um vetor aleatório de média  $0_n$ , matriz de variância-covariância  $\sigma^2 \mathbf{I}$ , com  $\sigma^2$  constante, e o vetor  $\beta$  são os parâmetros de interesse a ser estimado.

Um dos estimadores que podem ser usados para  $\beta$  é o estimador de mínimos quadrados, equivalente ao estimador encontrado pelo Método de Máxima Verossimilhança, supondo  $\varepsilon$  seja normal. Assim, temos que  $\hat{\beta} = (X^T X)^{-1} X^T y$ .

Agora, estimaremos o  $\beta$  por uma ótica bayesiana. Podemos especificar uma distribuição *a priori* para o parâmetro e a distribuição *a posteriori*, sendo avaliada com base nos dados da amostra, e então estimamos  $\beta$ .

### 3.3 Distribuição a Priori

A distribuição *a priori* representa a informação que temos sobre o parâmetro desconhecido  $\theta$  do nosso modelo, que ao ser combinado com a distribuição de probabilidade dos nossos dados resulta na distribuição *a posteriori* que é usada para futuras decisões ou inferências sobre  $\theta$ . Existem diferentes métodos de especificação da distribuição *a priori*. Serão tratadas nesta seção as prioris conjugadas e não-informativas.

- Distribuição a priori conjugada

Dizemos que uma distribuição é conjugada se as distribuições *a priori* e as distribuições *a posteriori* pertencem a mesma classe de distribuição. Sendo assim para o conhecimento de  $\theta$  ser atualizado é preciso apenas uma mudança nos parâmetros indexadores da família de distribuições *a priori*, chamados de hiperparâmetros, para serem diferenciados dos parâmetros já existentes de  $\theta$ .

- Distribuição a priori não-informativa

Distribuições a priori não-informativas começaram a ser usadas quando o objetivo era representar formalmente a "ignorância". Porém agora existe uma tendência de observá-las como opção para casos onde a informação *a priori* é insuficiente que torne difícil eliminar uma distribuição subjetiva considerada adequada. Esse tipo de distribuição é usada como forma de comparar os resultados da inferência clássica onde são usados somente informações da amostra.

### 3.4 Função de Verossimilhança

A função de verossimilhança tem papel fundamental, seja na inferência clássica ou na inferência bayesiana, como responsável por conter a informação dada pela amostra. A função é calculada fixando  $X$  e variando os valores do parâmetro  $\theta$ . Ela registra o quão verossímil os dados estão desse parâmetro. A função de verossimilhança é definida por

$$L(\theta) = L(\theta|X) = \prod_{i=1}^n f(x_i|\theta). \quad (3.4)$$

Para o caso de uma distribuição normal multivariada com média  $X\beta$  e variância  $\sigma^2$ , temos:

$$L(\beta, \sigma^2|X) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[ -\frac{1}{2}(y - X\beta)'(\sigma^2 I)^{-1}(y - X\beta) \right] \quad (3.5)$$

Será mostrado mais a frente o método para o cálculo da função de verossimilhança para o modelo a ser abordado nesse trabalho.

### 3.5 Distribuição a Posteriori

A distribuição *a posteriori* é o resultado obtido pela combinação entre a função de verossimilhança e a distribuição *a priori* do modelo. Nesse sub-capítulo iremos especificar dois métodos de obtenção da distribuição *a posteriori*, a primeira utilizando o método de Jeffreys para distribuições *a priori* não-informativas e a segunda para o caso de uma distribuição *a priori* conjugada.

- Distribuição *a posteriori* quando *a priori* é não-informativa

Precisamos especificar a distribuição *a priori* dos parâmetros desconhecidos da regressão. Para esse caso será utilizada a distribuição *a priori* não informativa usual. Considerando independência entre os parâmetro, temos

$$p(\beta) \propto 1, \quad p(\sigma^2) = \frac{1}{\sigma^2}, \quad \text{equivalente a } p(\beta, \sigma^2) = \frac{1}{\sigma^2}.$$

A distribuição *a posteriori* de  $\beta$  dado  $\sigma^2$  é dada então por

$$\beta|\sigma^2, y \sim N_p(\hat{\beta}, \sigma^2(X'X)^{-1}),$$

onde  $\hat{\beta} = (X'X)^{-1}X'y$  é o estimador de  $\beta$  obtido pelo MMQ (Método de Mínimos Quadrado). A distribuição marginal *a posteriori* de  $\sigma^2$  segue uma Gama Invertida

$$\sigma^2|y \sim GI(k/2, ks^2/2)$$

onde  $k = n - p$  e  $ks^2 = (y - X\hat{\beta})'(y - X\hat{\beta})$  é chamada soma dos quadrados residual. Segue então que a distribuição marginal *a posteriori* de  $\beta$  é uma t-Student não central p-variada com três parâmetros, dada por

$$\beta|y \sim t_p(k; \hat{\beta}, s^2(X'X)^{-1})$$

onde  $k$  são os graus de liberdade, vetor de média  $\hat{\beta}$  e o fator escalar  $s^2(X'X)^{-1}$ .

- Distribuição *a posteriori* quando *a priori* é conjugada

Consideramos agora um distribuição *a priori* conjugada, com  $(\beta, \sigma^2)$  distribuídos pela *a priori* dada por

$$p(\beta|\sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2\sigma^2} (\beta - \beta_0)' V_0^{-1} (\beta - \beta_0) \right],$$

$$p(\sigma^2) \propto (\sigma^2)^{c_0/2+1} \exp \left[ -\frac{d_0}{2\sigma^2} \right].$$

Ou seja,  $\beta|\sigma^2 \sim N_p(\beta^*, \sigma V_0)$  e  $\sigma^2 \sim GI(\frac{c_0}{2}, \frac{d_0}{2})$ .

Para obtermos a distribuição *a posterior* basta combinar essas informações com a função de verossimilhança, que é dada por:

$$L(\beta, \sigma^2) \propto (\sigma^2)^{-n/2} \exp \left[ -\frac{1}{2\sigma^2} [ks^2 + (\beta_0 - \hat{\beta})' X' X (\beta_0 - \hat{\beta})] \right].$$

Ao combinarmos essas informações obtemos a expressão para a distribuição *a posteriori*

$$\begin{aligned} p(\beta, \sigma^2|y) &\propto (\sigma^2)^{-(\frac{n+k}{2} + \frac{c_0}{2} + 1)} \times \\ &\exp \left[ -\frac{1}{2\sigma^2} \left( (\beta - \beta^*)' V_0^{-1} (\beta - \beta^*) + (\beta^* - \hat{\beta})' X' X (\beta^* - \hat{\beta}) \right) \right] \times \\ &\exp \left( -\frac{ks^2}{2\sigma^2} \right) \exp \left( -\frac{d_0}{2\sigma^2} \right) \end{aligned}$$

Agora usamos uma técnica para completar fórmulas quadráticas com um certo trabalho algébrico (cálculos detalhados em Paulino, 1983) e definimos

$$\begin{aligned} V_1 &= V_0^{-1} + X'X, \quad \beta_1 = V_1^{-1}(V_0^{-1}\beta^* + X'X\hat{\beta}) \\ c_1 &= c_0 + n, \quad d_1 = d_0 + \beta^{*'} V_0^{-1} \beta^* + y'y - \beta_1 V_1^{-1} \beta_1 \end{aligned}$$

Sendo assim, é provada que a distribuição *a posteriori* dos parâmetros  $(\beta, \sigma^2)$  é Normal Multivariada e Gama Invertida, respectivamente, tal que

$$p(\beta|\sigma^2, y) \propto \exp \left[ -\frac{1}{2\sigma^2} (\beta - \beta_1)' V_1 (\beta - \beta_1) \right],$$

$$p(\sigma^2) \propto (\sigma^2)^{c_1/2+1} \exp \left[ -\frac{d_1}{2\sigma^2} \right],$$

ou equivalentemente,  $\beta|\sigma^2, y \sim N_p(\beta_1, \sigma^2 V_1^{-1})$  e  $\sigma^2|y \sim GI(\frac{c_1}{2}, \frac{d_1}{2})$ .

## 3.6 Estimação Intervalar

Existe na inferência bayesiana um equivalente ao Intervalo de Confiança do caso frequentista, chamado de Intervalos de Credibilidade. Eles fornecem a probabilidade *a posteriori* de um parâmetro a sua escolha pertencer ao intervalo  $C$ . Um subconjunto de  $\Theta$ , é intervalo de credibilidade de  $100(1 - \alpha)\%$  para  $\theta$  se  $P(\theta \in C) \geq (1 - \alpha)$ . Como existem inúmeros intervalos de credibilidade que podem ser construídos, iremos adotar somente um para ser usado durante o estudo, o Intervalo HPD (Highest Posterior Density), que se justifica, segundo Ehlers (2007), por ser o intervalo de comprimento mínimo possível, pois são obtidos tomando-se os valores  $\theta$  com maior densidade a posteriori.

- **Intervalo HPD:** Um intervalo de credibilidade de  $(1 - \alpha)\%$  para  $\theta$  é HPD se  $c = \theta \in \Theta : \pi(\theta|(\mathbf{y})) \geq k(\alpha)$  onde  $k(\alpha)$  é a maior constante tal que  $P(\theta \in C) \geq 1 - \alpha$ .

São usados, em geral, métodos computacionais para construir o intervalo HPD. Após a obtenção do intervalo temos, por definição, que todos os pontos fora do intervalo terão densidade menor que os pontos internos. Ao obtermos os intervalos para cada coeficiente do modelo, precisamos verificar se o 0 está no intervalo, pois se estiver não teremos evidências para dizer que o parâmetro é significativo no modelo.

## 3.7 Técnicas de Diagnóstico

A importante função que as técnicas de diagnóstico tem no contexto de modelagem estatística é inegável. Podemos usar muitos métodos para tal diagnóstico, tal como análise de resíduos e seus gráficos para encontrarmos possíveis observações que influenciam o modelo. Contudo, na abordagem bayesiana essa análise não é tão reconhecida. Porém existem técnicas de diagnóstico que ajudam na modelagem bayesiana. Algumas delas serão vistas nesse subcapítulo.

### 3.7.1 Resíduos bayesianos Studentizados

Os resíduos bayesianos Studentizados são definidos por Paulino et. al (2003) como

$$r_i(\theta) = \frac{y_i - E(y_i|\mathbf{y})}{\sqrt{Var(y_i|\mathbf{y})}}, \quad (3.6)$$

que é função do parâmetro  $\theta$  e, portanto, tem distribuição *a posteriori* definida. Temos que  $\mathbf{y}$  representa os dados, e  $E(y_i|\mathbf{y})$  é obtida por

$$E(y_i|\mathbf{y}) = \int y_i p(y_i|\mathbf{y}) dy_i = \int E(y_i|\theta) \pi(\theta|\mathbf{y}) d\theta$$

e

$$E(y_i^2|\mathbf{y}) = \int y_i^2 p(y_i|\mathbf{y}) dy_i = \int [Var(y_i|\theta) + [E(y_i|\theta)]^2 \pi(\theta|\mathbf{y})] d\theta.$$

O cálculo para a obtenção de tais valores são computados usando a amostra  $\{\theta_{(j)}^*, j = 1, \dots, m\}$  da distribuição a *posteriori* de  $\pi(\theta|\mathbf{y})$ , então é feita a média do valor esperado de cada observação para cada  $\theta_{(j)}^*$ , ou seja

$$E(y_i|\mathbf{y}) = \frac{1}{m} \sum_{j=1}^m E(y_i|\theta)$$

e

$$E(y_i^2|\mathbf{y}) = \frac{1}{m} \sum_{j=1}^m (Var(y_i|\theta) + [E(y_i|\theta)]^2)$$

Notamos que os resíduos não são independentes, pois suas distribuições a posteriori dependerem todas de  $\theta$ . Por isso o uso dessa ferramenta se torna mais informal. Usamos também gráficos para avaliarmos o comportamento dos resíduos, onde plotamos os valores dos resíduos e definimos um limite para termos noção das observações que são outliers no modelo. É razoável esperar que em modelos bem ajustados os resíduos estejam entre  $-2$  e  $2$ .

### 3.7.2 Comparação de Modelos

Existe no contexto da estatística bayesiana uma série de distribuições a priori que poderão ser candidatas ao modelo. Devido as diferentes interpretações do problema, uma série de distribuições *a priori* poderão ser candidatas ao modelo, sendo os modelos representados por

$$B = \{p_i(x|\theta_i), \pi_i(\theta_i), i \in I\}$$

Para isso existem métodos que especificam critérios para decidir qual o melhor candidato. Os métodos que serão citados a seguir são os critérios DIC (Deviance Information Criterion e CPO (Conditional Predictive Ordinate).

#### 3.7.2.1 DIC

Talvez seja o método mais comum usado para a escolha de modelos na inferência bayesiana. É uma generalização dos critérios frequentistas BIC e AIC. Spiegelhalter et al. (2002) propôs o DIC baseado no princípio de que DIC = "qualidade de ajuste" + "complexidade". DIC então é definido por

$$DIC = \bar{D} + p_D \tag{3.7}$$

Onde  $p_D$  é a medida de 'número efetivo de parâmetros' sendo denotada por:

$$p_D = E_{\theta|\mathbf{y}}[-2 \log f(y|\theta)] + 2 \log[f(y|\hat{\theta})] \tag{3.8}$$

Tomando  $f(y)$  como um termo totalmente especificado e padronizado que é função apenas de  $Y$ , ou seja, depende apenas dos dados,  $p_D$  pode ser escrito como

$$p_D = \overline{D(\theta)} - D(\bar{\theta}), \tag{3.9}$$



onde  $\overline{D(\theta)} = \frac{1}{m} \sum_{j=1}^m D(\theta_j)$  e  $D(\theta) = -2 \sum_{i=1}^n \log[f(y_i|\theta)]$  é a estatística Deviance.

Assim como nos métodos frequentistas AIC e BIC, o melhor modelo a ser escolhido, dentro do conjunto de modelos possíveis será aquele com o menor valor de DIC.

### 3.7.2.2 CPO

Esse é mais um dos inúmeros métodos de comparação de modelos para um determinado conjunto de dados. O CPO (Conditional Predictive Ordinate) é uma estatística baseada no critério de validação cruzada para comparação de modelos. Em Gelfand et. al (1992) podem ser vistas as propriedades detalhadas da estatística CPO e aplicações em alguns modelos. A estatística CPO para a  $i$ -ésima observação é descrita por

$$CPO_i = p(y_i|y_{(-i)}) = \int_{\theta \in \Theta} f(y_i|\theta) \pi(\theta|y_{(-i)}) d\theta = \left( \int_{\theta \in \Theta} \frac{\pi(\theta|y)}{f(y_i|\theta)} d\theta \right)^{-1}, \quad (3.10)$$

onde  $p(y_i|y_{(-i)})$  representa a distribuição de  $y_i$  dada o resto das observações da amostra e  $y_{(-i)}$  os dados sem a  $i$ -ésima observação. Um pequeno valor dessa estatística representa uma observação que não está ajustada ao modelo. Isso ocorre porque os valores obtidos são um indicador do valor da função de verossimilhança de cada observação dado o restante das observações. Podemos ainda estimar a  $CPO_i$  por Monte Carlo, onde só é preciso usar uma única amostra de MCMC para a distribuição *a posterior*  $\pi(\theta|y)$ . Então a aproximação de Monte Carlo para a estatística  $CPO_i$ , de uma amostra com tamanho  $M$ , é dada por:

$$\widehat{CPO}_i = \left( \frac{1}{M} \sum_{m=1}^M \frac{1}{f(y_i|\theta_m)} \right)^{-1} \quad (3.11)$$

Para a análise dos valores do  $CPO_i$  é usado a LPML (Log-Marginal Pseudo Likelihood), onde  $LPLM = \sum_{i=1}^n \log(\widehat{CPO}_i)$ . Quanto maior for o valor de  $CPO_i$  melhor é o ajuste feito pelo modelo em questão.

## 3.7.3 Pontos de Influência

Assim como na regressão clássica, existem formas para encontrarmos pontos que são influentes nos resultados obtidos pela análise estatística. Iremos fazer, para o caso bayesiano, o método de exclusão de observações, onde são excluídas observações uma a uma e comparamos os resultados para descobrirmos qual observação é a mais influente no modelo. O método escolhido é a Distância de Kullback-Leibler.

### 3.7.3.1 Distância de Kullback-Leibler

A distância de Kullback-Leibler é uma medida não simétrica usada como ferramenta para medir a diferença entre duas distribuições de probabilidade,  $p(x)$  e  $q(x)$ , para uma mesma variável  $x$ . Chamada geralmente de distância KL, a diferença entre  $q(x)$  e  $p(x)$

mede a informação perdida quando é usada a distribuição  $q(x)$  como uma aproximação de  $p(x)$ . Definimos  $D_{KL}$  pela seguinte equação:

$$D_{KL}(p(x)||q(x)) = \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}$$

Temos que  $D_{KL}(p(x)||q(x))$  é uma medida sempre positiva, ou seja,  $D_{KL}(p(x)||q(x)) \geq 0$   $D_{KL}(p(x)||q(x)) = 0$ , se e somente se  $p(x) = q(x)$ , e por ser uma medida não simétrica  $D_{KL}(p(x)||q(x)) \neq D_{KL}(q(x)||p(x))$ .

Para ser aplicado no caso bayesiano de diagnóstico de pontos influentes é necessário definirmos algumas mudanças. Considere  $\pi_1 = \pi[\theta|y_{(-i)}]$  a distribuição *a posteriori* de  $\theta$  retirada a  $i$ -ésima observação de  $y$  e  $\pi_2 = \pi[\theta|y]$  a distribuição *a posteriori* de  $\theta$  com  $y$  completo. A distância entre  $\pi_1$  e  $\pi_2$  é dada por:

$$D_{KL}(\pi_2||\pi_1) = - \int \pi[\theta|y] \log \left( \frac{\pi[\theta|y]}{\pi[\theta|y_{(-i)}]} \right) d\theta \quad (3.12)$$

Ou seja,  $D_{KL}(\pi_2||\pi_1)$  mede o efeito da retirada da  $i$ -ésima observação na distribuição *a posteriori* conjunta de  $\theta$ . Podemos simplificar essa equação deixando-a em função da equação 3.11 (já vista anteriormente) dada por

$$\begin{aligned} D_{KL}(\pi_2||\pi_1) &= \log E_{\theta|y}(f(y_i|\theta)^{-1}) + E_{\theta|y}(\log f(y_i|\theta)) \\ &= -\log(CPO_i) + E_{\theta|y}(\log f(y_i|\theta)). \end{aligned} \quad (3.13)$$

A prova para tal igualdade se encontra em Garay et al. (2011).

Além da distância de Kullback-Leibler existem na literatura outras distâncias para detectar pontos de influência na literatura (distância L, distância J, distância qui-quadrado). Podemos ler sobre em Bolfarine et. al. Ressaltando que por ser um método para medidas de pontos influentes, é necessário definir um valor limite para determinar se um ponto é ou não influente. Para isso nesse projeto iremos usar a definição proposta por Peng e Dey (1995) e Vidal e e Castro (2010), onde definiram para as diferentes medidas de pontos de influência a seguinte função:

$$d_g(p) = \frac{g(2p) + g(2(1-p))}{2} \quad (3.14)$$

Para o nosso caso, onde será usado a distância de Kullback-Leibler a função  $g(p) = \log(p)$ , e será usado o valor  $p = 0.8$ , com isso podemos afirmar que quando  $d_{KL} \geq 0.22$  a  $i$ -ésima observação será considerada como influente para o modelo.

### 3.8 Exemplo

Para fins de comparação iremos usar o mesmo conjunto de dados do exemplo de regressão clássica, que pode ser visto na tabela (1). Foi usado o pacote MCMCpack, do

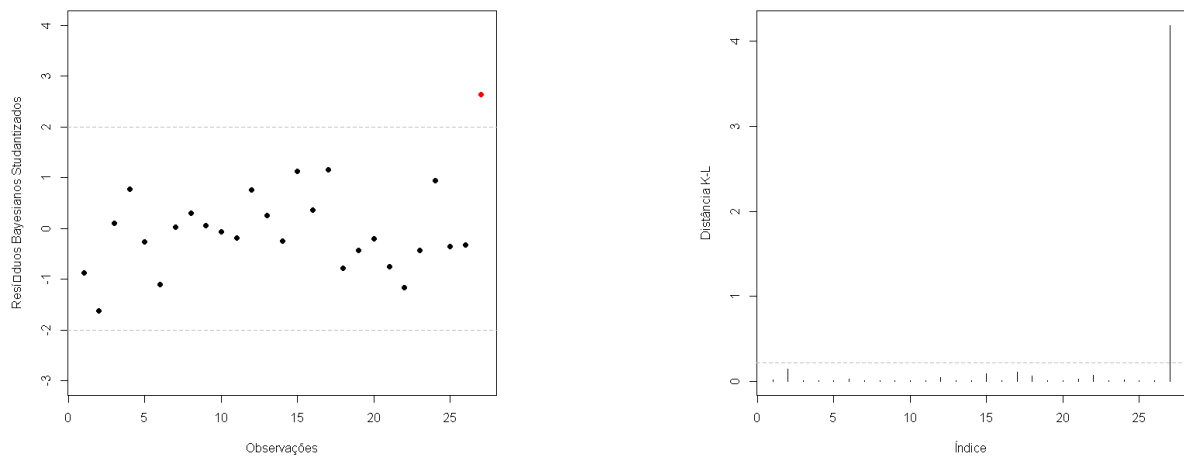
R Studio, para calcularmos as estimativas do intercepto e dos parâmetros  $(\beta, \sigma^2)$ . Foi escolhida a distribuição Uniforme Imprópria para *a priori* de  $\beta$  e para  $\sigma^2$  uma Gama Invertida com  $c = 0.0001$  e  $d = 0.0001$ . Os resultados foram obtidos através da média a posteriori das estimativas de cada cadeia de Markov, sendo muito semelhantes ao caso clássico.

**Comparação**

	Bayesiano	Clássico
Intercepto	-380.3	-381.28
$\beta$	199.6	199.83
$\sigma^2$	6507.7	5962.68

Na tabela (3.8) reparamos que para as estimativas do intercepto e de  $\beta$  os valores são próximos. Porém a estimativa de  $\sigma^2$  destoa bastante, sendo que para o caso bayesiano a estimativa é bem maior.

Foi feito também a análise de diagnóstico para o caso bayesiano, onde testamos os resíduos e se o modelo usado é adequado pela distância de Kullback-Leibler. A seguir estão os gráficos para os resíduos e a adequabilidade do modelo usado via distância de Kullback-Leibler.



Podemos notar que a observação 27, que representa o Distrito Federal, destoa muito do modelo em questão, tanto para a análise dos resíduos studentizados bayesianos e para a distância de Kullback-Leibler. Sendo assim obtemos a mesma conclusão que o diagnóstico feito para o caso frequentista.



## 4 Dados Limitados

O Modelo Tobit é um modelo de regressão linear usado quando algumas observações da variável resposta encontram-se incompletas devido a algum tipo de censura. Porém existem duas causas para ocorrência de tal evento, quando ocorre *truncamento* e quando ocorre *censura*. Primeiramente iremos esclarecer a diferença entre esses dois casos.

### 4.1 Truncamento

O truncamento ocorre (na maioria das vezes) quando é feito um estudo com base em uma base de dados já usada anteriormente, de onde são retiradas apenas as informações desejadas. Com isso algumas observações ficam indisponíveis na variável resposta e nas variáveis regressoras.

#### 4.1.1 Distribuição Truncada

São ditas distribuições truncadas as partes de uma distribuição não truncada que está acima ou abaixo de um determinado valor. Temos a seguir a distribuição normal truncada. Dado que  $x$  segue uma distribuição normal com média  $\mu$  e variância  $\sigma^2$ , então

$$P(x > a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right) = 1 - \Phi(\alpha),$$

onde  $\alpha = (a - \mu)/\sigma$  e  $F(\cdot)$  é a f.d.a. (Função de Distribuição Acumulada) da normal padrão. A densidade de uma distribuição normal truncada é dada por

$$f(x|x > 0) = \frac{f(x)}{1 - F(\alpha)} = \frac{(2\pi\sigma^2)^{-1/2} \exp^{-(x-\mu)^2/(2\sigma^2)}}{1 - F(\alpha)} = \frac{\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)}{1 - F(\alpha)}$$

onde  $f(\cdot)$  é a f.d.p.(Função de Distribuição de Probabilidade) da normal padrão.

### 4.2 Censura

A censura ocorre quando para certas observações da amostra os dados sobre a variável resposta não se encontram disponíveis em sua totalidade, por estarem limitadas, mas diferentemente do truncamento, temos os dados sobre as variáveis regressoras disponíveis. A censura é essencialmente um defeito na coleta da amostra, seja ela pela falta de informação ou por não ser possível medir as observações por algum tipo de limitação. Então os valores contidos nessa área de limitação são reportados todos com o mesmo valor.

### 4.2.1 Distribuição Censurada

Assim como para a distribuição truncada, as distribuições censuradas são derivadas de distribuições conhecidas e sem censura. Quando temos uma distribuição censurada à esquerda em um certo valor 'c', ou seja, as observações menores ou iguais ao valor limite c são substituídas por c, e assim a variável resposta "completa" é apresentada como:

$$y_i = \begin{cases} c, & \text{se } y_i^* \leq c \\ y_i^*, & \text{se } y_i^* > c \end{cases}$$

Seja  $Y^*$  uma variável contínua com f.d.p.  $f(\cdot)$  e f.d.a.  $F(\cdot)$ . Teremos para as variáveis censuradas,

$$f(y_i) = [f(y_i^*)]^{I_i} [F(c)]^{(1-I_i)} \quad i = 1, \dots, n.,$$

onde  $c$  é uma constante e

$$I_i = \begin{cases} 1, & \text{se } y_i^* > c \\ 0, & \text{se } y_i^* \leq c \end{cases}$$

$i = 1, \dots, n$ . Para o caso da distribuição normal censurada basta tomar  $f = \phi$  e  $F = \Phi$ .

## 4.3 Modelo Tobit

Nesse trabalho abordaremos o Modelo Tobit. O nome faz referência a Tobin(1958), artigo pelo qual esse modelo foi popularizado, Trata-se de um modelo de regressão onde são estimadas as relações de uma variável dependente limitada. Por se tratarem de dados censurados temos algumas observações que são parcialmente observadas, ou seja, não estão representadas com seu valor real e sim com o valor da censura. O modelo de regressão Tobit é comparável ao modelo clássico,

$$y_i^* = x_i' \beta + \varepsilon_i. \quad (4.1)$$

Suponha um limite inferior no valor 0, ou seja censura à esquerda. A variável latente  $y_i^*$  será observada da seguinte maneira

$$y_i = \begin{cases} 0, & \text{se } y_i^* \leq 0 \\ y_i^*, & \text{se } y_i^* > 0, \end{cases}$$

onde  $y_i^* \sim F(y_i^* | x_i, \theta)$  com função de densidade  $f(y | y_i^* | x_i, \theta)$ .

### 4.3.1 Estimação Bayesiana

O procedimento é análogo ao modelo linear, mas como existem particularidades existentes no Modelo Tobit, alguns cálculos precisam ser adaptados ou feitos com ajuda computacional.

A seguir vemos o caso da obtenção da função de verossimilhança, para censura no valor  $c$ :

$$L(\theta) = \prod_{i=1}^n \left( f(y_i|x_i, \theta)^{I_i} F(c|x_i, \theta)^{1-I_i} \right), \quad (4.2)$$

onde  $I_i = 1$  quando a observação está com o valor exato e  $I_i = 0$  quando a observação está censurada à esquerda  $\theta$  representa o vetor de parâmetros  $\beta$  e  $\sigma^2$ .

A função de log-verossimilhança para a regressão censurada à esquerda é dada por:

$$\ln L = \sum_{i:y_i > c} -\frac{1}{2} \left[ \log(2\pi) + \ln \sigma^2 + \frac{(y_i - x'_i \beta)^2}{\sigma^2} \right] + \sum_{y_i < c} \ln \left[ F\left(\frac{x'_i \beta}{\sigma}\right) \right]. \quad (4.3)$$

Podemos ver que a primeira parte da função é semelhante à verossimilhança para as observações que não estão presentes na censura e a segunda se refere às probabilidades de censura das observações censuradas. Esse resultado se aplica apenas a censura à esquerda. Para a censura à direita, com censura no valor  $c$  temos:

$$L(\theta) = \prod_{i=1}^n \left( f(y_i|x_i, \theta)^{I_i} \{1 - F(c|x_i, \theta)\}^{1-I_i} \right), \quad (4.4)$$

onde  $I_i = 1$  quando a observação está com o valor exato e  $I_i = 0$  quando a observação está censurada à direita, A função de log-verossimilhança para a regressão censurada à esquerda é dada por:

$$\ln L = \sum_{i:y_i < c} \ln \frac{1}{\sigma} \phi \left[ \frac{y_i - x'_i \beta}{\sigma} \right] + \sum_{y_i < c} \ln \left[ 1 - F\left(\frac{c - x'_i \beta}{\sigma}\right) \right]. \quad (4.5)$$

Pela existência da censura não é possível obter a função de verossimilhança na forma completa. Para contornarmos esse problema usamos o método de aumento de dados para fazermos a imputação da parte negativa da função. Combinando tal método com o Amostrador de Gibbs é possível então termos a distribuição *a posteriori* de  $\theta$ .

- Distribuição *a priori*

O processo de especificação da distribuição *a priori* para o Modelo Tobit não difere dos métodos já explicados no Capítulo 3, página 16, onde podemos encontrar a distribuição *a priori* por meio de prioris conjugadas ou prioris não-informativas.

- Distribuição *a posteriori*

Diferentemente da distribuição *a priori*, para a distribuição *a posteriori* o processo de obtenção é ligeiramente diferente, pois como foi discutido anteriormente no presente capítulo, a função de verossimilhança é alterada por não apresentar uma forma completa. Logo os métodos para a especificação da distribuição *a posteriori*, dada por  $\pi(\theta|x) \propto L(\theta; x)\pi(\theta)$ , não podem ser, em um primeiro momento, empregados, pois a função de

verossimilhança é dividida em duas partes. Sendo assim não é possível encontrar a forma fechada para uma distribuição. Porém será possível aplicarmos os métodos já discutidos ao encontrarmos a função de verossimilhança na forma completa.

### 4.3.2 Aumento de Dados (Data Augmentation) e Amostrador de Gibbs

Conhecendo a função de verossimilhança e dado a distribuição *a priori* de  $\theta$  seria possível calcular a distribuição *a posteriori*. Porém nossa função de verossimilhança por estar incompleta não pode ser calculada de uma forma direta, pois não resultaria numa distribuição de forma fechada. Com isso introduzimos o método aumento de dados em conjunto com o Amostrador de Gibbs.

O Amostrador de Gibbs é um processo de simulações estocásticas via cadeias de Markov, onde a função geradora é formada por densidades condicionais completas, tornando possível gerar amostras de uma distribuição marginal para casos onde a geração direta da posteriori conjunta é complicada. Para o nosso caso queremos gerar  $\pi(\theta|y^*)$  ( $\pi(\beta, \sigma^2|y^*)$ ), o que é impossibilitado pelo fato da função de verossimilhança não se encontrar na forma completa. Porém a distribuição *a priori* dos parâmetros são dadas,  $\beta \sim N(\beta_0, \sigma V_0)$  e  $\sigma^2 \sim GI(c_0/2, d_0/2)$ . Com isso é possível gerarmos as condicionais  $\pi(\beta|\sigma^2, y^*)$  e  $\pi(\sigma^2|\beta, y^*)$ . O algoritmo de Gibbs então se dá por meio de seguidas gerações dessas distribuições condicionais. A seguir temos o algoritmo de como o é feito o processo iterativo (Ganerman, 1996):

1. o processo é iniciado em  $j = 1$ . Primeiramente partimos dos valores *a priori* de  $\theta^{(0)} = (\beta^{(0)}, \sigma^{2(0)})$ ;
2. obtém-se a partir de  $\theta^{(j-1)}$  um novo valor para  $\theta^{(j)} = (\beta^{(j)}, \sigma^{2(j)})$  por meio das gerações de valores

$$\begin{aligned}\beta^{(j)} &\sim \pi(\beta|\sigma^{2(j-1)}, y^*) \\ \sigma^{2(j)} &\sim \pi(\sigma^2|\beta^{(j-1)}, y^*);\end{aligned}$$

3. o processo então vai de  $j$  para  $j + 1$  volta para o passo (2) e assim por diante até atingir a convergência.

Assumida uma convergência para uma distribuição que esteja aleatoriamente próxima à distribuição desejada  $\pi(\theta|y)$ , e, após a convergência todas as gerações de uma mesma cadeia serão distribuições convergentes a amostra de  $\pi(\theta|y)$ . Dado a convergência das cadeias, à amostra desejada é retirada tomando-se saltos entre os valores gerados, evitando a dependência linear entre os valores gerados.

No processo de aumento de dados são introduzidas  $n$  variáveis latentes independentes  $Y_1^*, \dots, Y_n^*$ , onde em geral a distribuição *a posteriori* de  $\theta$  dado  $Y^* = (Y_1^*, \dots, Y_n^*)$  segue:

$$\pi(\theta|y_i^*, x) \propto \pi(\theta) \prod_{i=1}^n f(y_i^*|x_i, \theta). \quad (4.6)$$



O aumento de dados nos leva a  $Y_i^*s \sim N(x_i'\beta, \sigma)$ :

$$(y_i^*|y, x, \theta) \sim y_i I(y_i > 0) + \left[ \frac{f(y_i^*|x_i, \theta)}{F(0|x_i, \theta)} I(y_i^* \leq 0) \right] I(y_i = 0) \quad (4.7)$$

ou seja,  $y_i^* = y_i$  se  $y_i > 0$  e  $(y_i^*|x_i, y_i, \theta) \sim \frac{f(y_i^*|x_i, \theta)}{F(0|x_i, \theta)} I(y_i^* \leq 0)$  se  $y_i = 0$ . Observe que se os  $Y_i^*s$  são conhecidos e uma priori normal multivariada é escolhida para  $\theta$ , então a distribuição *a posterior* de  $\theta$  pode ser derivada utilizando os resultados da regressão linear normal bayesiana (Albert and Chib, 1993). Os  $Y_i^*s$  são desconhecidos mas, dado  $Y_i$ , a distribuição de  $Y_i^*$  segue uma normal truncada. Com isso já podemos fazer a ligação com o amostrador de Gibbs, para então obter a distribuição que irá convergir para a *a posterior* real de  $\theta$ .



## 5 Estudo de simulações

Tendo a finalidade de avaliar o desempenho do modelo proposto, ou seja, analisar se para o caso de dados censurados as definições e suposições estão sendo cumpridas e se suas estimativas *a posteriori* estão seguindo o modelo original. Os procedimentos computacionais foram desenvolvidos utilizando o software R, com a ajuda do pacote *BayesCR*. Para as simulações foram geradas amostras de tamanho  $n$ , de um modelo linear censurado, seguindo (4.1), onde  $\epsilon_i \sim N(0, \sigma^2 = 1)$ ,  $\beta^T = (\beta_0, \beta_1) = (1, 2)$   $x'_i = (1, x_i)$ , sendo cada  $x_i^T$  uma observação obtida por uma amostra de uma distribuição  $U(-1, 1)$  e um nível de censura fixado  $p$ .

### 5.1 Simulação 1

Nesse estudo o objetivo é analisar a taxa de acerto médio das estimativas obtidas para o modelo. Para isso foram considerados diferentes níveis de censura para  $p$  (10%, 20%, 30%, 40%, 50%, 60% e 70%), onde para cada nível de censura foram geradas 100 amostras de Monte Carlo e foram calculadas suas estimativas *a posteriori* dos parâmetros.

censura	Parâmetros	Estimativa	Desvio Padrão	% valor real
10%	$\hat{\beta}_0$	0.9996429	0.10902382	92%
	$\hat{\beta}_1$	1.963516	0.1984743	91%
	$\hat{\sigma}^2$	1.019397	0.1663651	92%
20%	$\hat{\beta}_0$	0.9835375	0.10544863	95%
	$\hat{\beta}_1$	2.009927	0.1790368	97%
	$\hat{\sigma}^2$	1.008252	0.1877581	92%
30%	$\hat{\beta}_0$	0.9861090	0.09902344	99%
	$\hat{\beta}_1$	2.003582	0.1861714	96%
	$\hat{\sigma}^2$	1.086538	0.2002262	92%
40%	$\hat{\beta}_0$	0.9698883	0.11933453	92%
	$\hat{\beta}_1$	2.057882	0.2539729	92%
	$\hat{\sigma}^2$	1.080632	0.2159631	92%
50%	$\hat{\beta}_0$	0.9683830	0.17419893	97%
	$\hat{\beta}_1$	2.034584	0.2824929	98%
	$\hat{\sigma}^2$	1.048598	0.2499068	95%
60%	$\hat{\beta}_0$	0.9569816	0.17613150	93%
	$\hat{\beta}_1$	2.060665	0.3007020	93%
	$\hat{\sigma}^2$	1.087558	0.2481077	97%
70%	$\hat{\beta}_0$	0.9683830	0.17419893	90%
	$\hat{\beta}_1$	2.034584	0.2824929	97%
	$\hat{\sigma}^2$	1.048598	0.2499068	94%

Foram calculados o valor médio das estimativas de  $\beta$  e  $\sigma^2$ , assim como a média dos desvios padrão das estimativas. Também foi calculada a percentagem de vezes que o intervalo de credibilidade HPD *a posteriori* de 95% contém o verdadeiro valor do parâmetro. Os resultados estão apresentados na tabela acima.

Podemos observar na tabela 5.1 que as estimativas estão muito estáveis, ou seja, perto dos valores reais dos parâmetros. Também é possível observar que para os maiores níveis de censura a percentagem de vezes que o intervalo de credibilidade *a posteriori* de 95% contém o verdadeiro valor do parâmetro  $\sigma$  aumenta, apesar de apresentar, em média, estimativas mais distantes do valor real do parâmetro.

## 5.2 Simulação 2

Nesta simulação procuramos avaliar visualmente a qualidade das estimativas para diferentes tamanhos de amostras ( $n = 50, 100$  e  $300$ ), por meio das densidades *a posteriori* dos parâmetros. As diferentes amostras possuem os mesmos valores reais para os parâmetros,  $\beta_0 = 1, \beta_1 = 2$  e  $\sigma = 1$ . Utilizamos apenas dois níveis de censura, 30% e 50%, por serem os níveis de censura que apresentaram os melhores resultados para a simulação anterior.

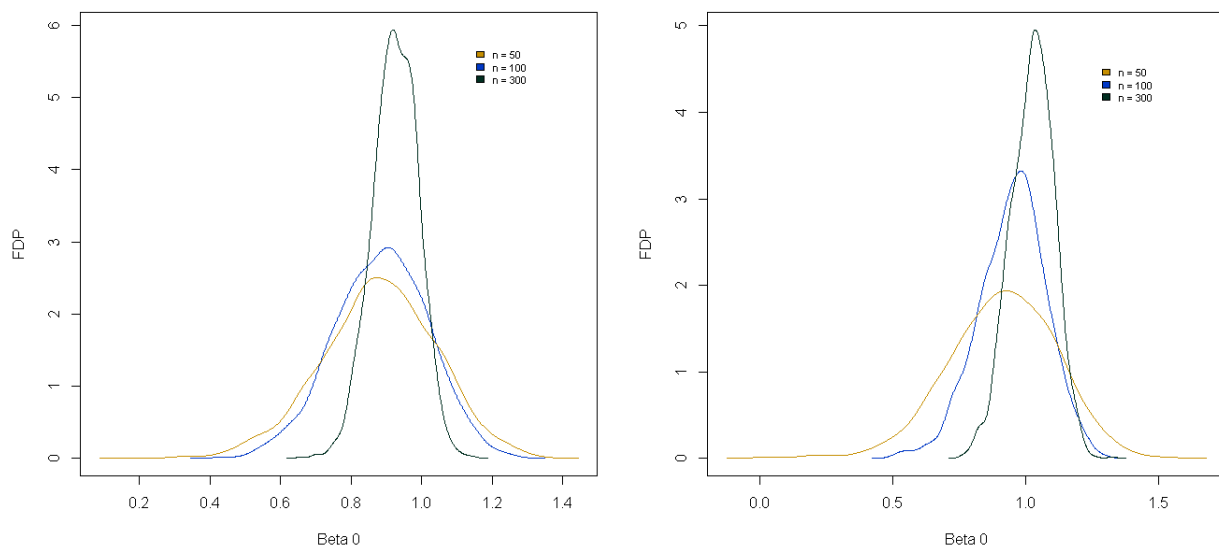


Figura 3 – Comparação das estimativas para  $\beta_0$  entre os níveis de censura 30% e 50%.

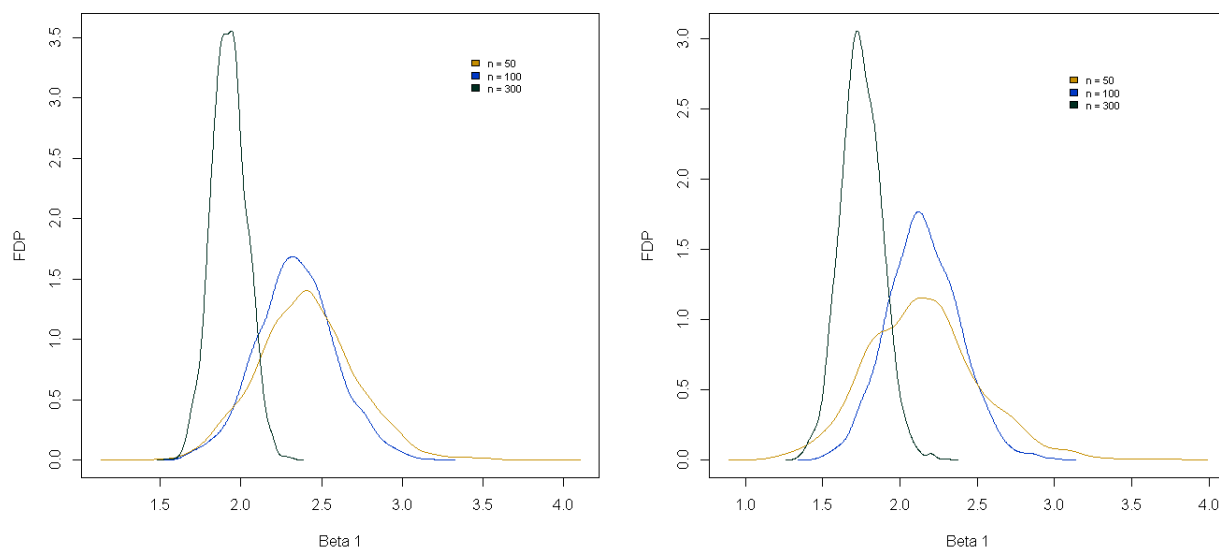


Figura 4 – Comparação das estimativas para  $\beta_1$  entre os níveis de censura 30% e 50%.

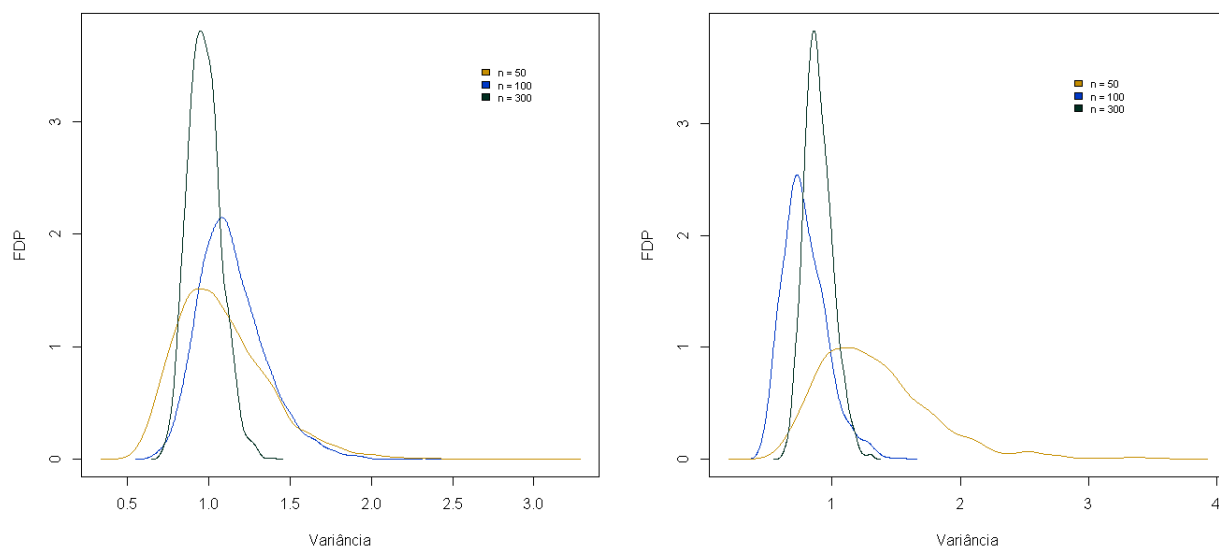


Figura 5 – Comparação das estimativas para  $\sigma^2$  entre os níveis de censura 30% e 50%.

Percebe-se que as estimativas foram, em geral, melhores para o nível de 50% de censura. Isso pode ser uma evidência de que o método se ajusta melhor a níveis altos de censura, ou pode ter ocorrido apenas por causa das amostras geradas. O parâmetro que obteve as estimativas com melhores resultados foi  $\sigma^2$  onde podemos observar que o valor da estimativa, para 30% e 50%, se encontra mais próximo do valor real.



## 6 Aplicações em dados reais

Esse capítulo abordará a aplicação da metodologia discutida nos Capítulos 3 e 4. Os dados escolhidos para a aplicação foram retirados de estudos já realizados, e serão revisitados com uma metodologia diferente.

Foi especificada a mesma distribuição *a priori* para as aplicações subsequentes, onde foi utilizada a priori obtida na função *Bayes.CR*, onde é assumido que  $\beta \sim N_p(b_0, S_\beta)$ , em que  $b_0$  é o vetor  $p \times 1$  de hiperparâmetros fixados e  $S_\beta$  é uma matriz definida positiva conhecida, de ordem  $p \times p$ . Também é suposto que o o parâmetro de escala  $\sigma^2$  segue uma distribuição gama inversa, isto é,  $\sigma^{-2} \sim \text{Gamma}(a/2, b/2)$ . Para o nosso caso em específico os valores utilizados foram,  $b_0^T = 0_p$ ,  $S_\beta = 10^3 I_p$ ,  $a = 2$  e  $b = 0.2$ .

### 6.1 Relações Extraconjugais

O primeiro conjunto de dados foi retirado do estudo feito por Fair (1978) onde foram utilizados dados de uma pesquisa feita em 1969 pela *Psychology Today* (PT), que publicou um questionário sobre sexo em Julho de 1969, onde os leitores, tanto homens quanto mulheres, foram convidados a enviar suas resposta. Cerca de 20.000 respostas foram recebidas. Somente 601 delas eram utilizáveis, e com uma grande quantidade de respostas 0 (450 observações, que representa quase 75% das respostas). Nesse estudo as respostas iguais a 0 foram consideradas como respostas censuradas. Apesar de ser impossível ter valores negativos de relações extraconjugais a censura foi utilizada porque devido as combinações das estimativas dos parâmetros seria explicável termos valores negativos, o que indicaria ser altamente improvável esses indivíduos com valores negativos terem relações extraconjugais. A variável resposta ( $Y$ ) representa o número de casos extraconjugais no último ano. No conjunto de dados original existem 8 variáveis, mas para esse estudo serão utilizadas somente 4 variáveis no modelo além da variável resposta. São elas:

- $Y$ : número de casos extraconjugais no ano anterior.
- $X_1$ : anos de casamento.
- $X_2$ : idade.
- $X_3$ : religiosidade (escala de 1 (ateu) a 5 (frequenta regularmente)).
- $X_4$ : avaliação do casamento (escala de 1(muito infeliz) a 5 (muito feliz))

Tabela 2 – Inferências bayesianas para o modelo.

Parâmetros	Estimativa	EP	Intervalo HPD 95%	
$\hat{\beta}_0$	9.11380	2.69002	3.86954	14.39072
$\hat{\beta}_1$	-0.16355	0.08035	-0.32462	-0.01749
$\hat{\beta}_2$	0.55159	0.14319	0.26250	0.81876
$\hat{\beta}_3$	-1.76029	0.42126	-2.57106	-0.92805
$\hat{\beta}_4$	-2.31521	0.42414	-3.08964	-1.50341
$\hat{\sigma}^2$	72.67587	10.29531	54.77223	93.09366

A tabela 2 mostra as inferências bayesianas para o modelo proposto. É observado que nenhum coeficiente aparenta ser não significativo para o modelo. Podemos analisar isso pelo intervalo HPD e ver que o 0 não está incluso em nenhum dos intervalos.

Note que apenas a estimativa do coeficiente  $\hat{\beta}_2$  obteve valor positivo. Com isso temos um indicativo que cônjugues de idade mais avançadas tendem a ter mais casos extraconjugais. Analisando as estimativas para os outros coeficiente observamos que o que mais influencia negativamente (ter menos casos extraconjugais) na quantidade de relações extraconjugais é a variável avaliação do casamento; ou seja, quanto mais alta for a avaliação menor é a variável resposta, seguida pela variável religiosidade; cônjugues que são praticantes da sua respectiva religião tendem a ter menos casos do que os não praticantes ou ateus. A variável anos de casamento apesar de possuir uma estimativa negativa influencia muito pouco o modelo, sendo possível observar pelo intervalo HPD que ela está muito próxima do 0, ou seja, beirando ser não significativa para o mesmo.

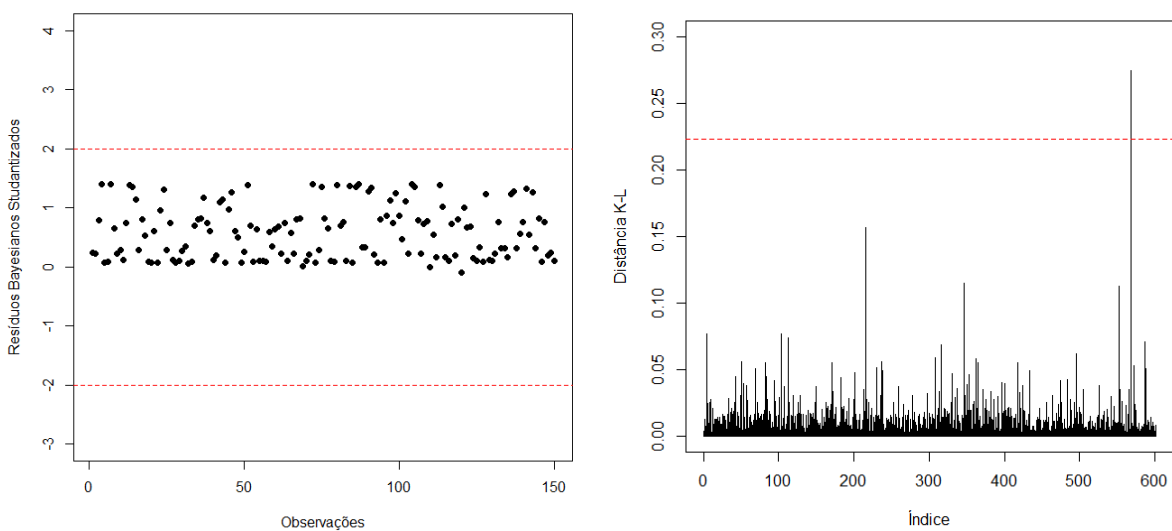


Figura 6 – Gráfico dos resíduos studentizados e da distância de Kullback-Leibler

Acima podemos ver os gráficos das medidas de diagnóstico para o modelo. O primeiro



gráfico mostra os resíduos das observações contínuas dos dados. À primeira vista parece estar errado, pois os resíduos estão distribuídos apenas na parte positiva, mas é algo esperado; uma vez que mais de 75% das observações são censuradas a média é arrastada para perto do valor da censura; com isso os valores esperados acabam sendo subestimados. Para o gráfico da distância de Kullback-Leibler existe uma única observação que ultrapassa o limite estabelecido para o modelo. A observação em questão é a 568 do banco de dados. O indivíduo tem 52 anos, 7 anos de casamento, 4 na escala de religiosidade, 5 na escala de avaliação do casamento e teve 12 casos extraconjugais no ano anterior, com isso podemos explicar a distorção dessa observação no conjunto de dados, pois ele teve um alto número de casos extraconjugais mesmo tendo avaliado o casamento em 5 "muito feliz" e tendo religiosidade 4, que representam variáveis com coeficientes negativos. Ou seja, quanto maior avaliado esperamos um menor valor para a variável resposta. Mesmo com essa observação discrepante é possível dizer que o modelo está adequado.

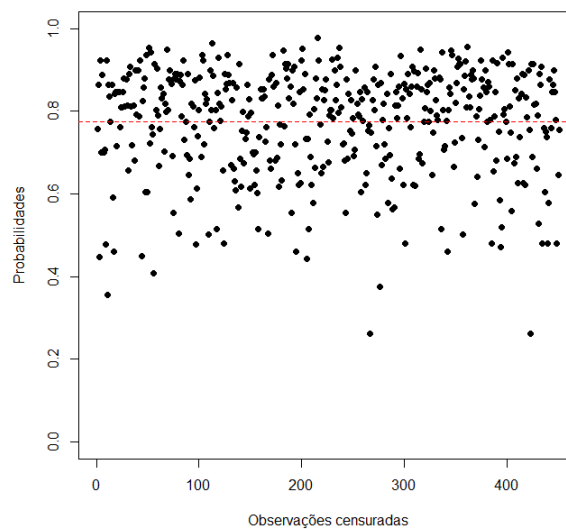


Figura 7 – Gráfico para  $p(y < 0|y)$

Como esperado, vemos na figura 7 que as probabilidades a posteriori das observações censuradas (do modelo após o aumento de dados) sejam menores que o valor da censura, são em geral altas, sendo que a maioria está no intervalo entre 0.8 e 1.

Tabela 3 – Comparação das estimativas dos parâmetros Inferência Clássica e Bayesiana

Parâmetros	Clássico		Bayesiano	
	Estimativa	Erro Padrão	Estimativa	Erro Padrão
$\hat{\beta}_0$	9.08	2.66	9.11	2.69
$\hat{\beta}_1$	-0.16	0.08	-0.16	0.08
$\hat{\beta}_2$	0.54	0.13	0.55	0.14
$\hat{\beta}_3$	-1.72	0.41	-1.76	0.42
$\hat{\beta}_4$	-2.26	0.41	-2.32	0.42
$\hat{\sigma}$	8.27	0.55	8.50	0.34

Na tabela 3 vemos as estimativas de diferentes metodologias para o mesmo banco de dados. O caso clássico foi analisado por Bolfarine et. al (2013) e na comparação vemos que os resultados obtidos nesse capítulo estão bem próximos, quanto para a estimativa pontual e para o seu respectivo erro padrão. Por isso podemos dizer que os resultados obtidos foram iguais, ou muito semelhantes. Então não perdemos nada no nosso modelo ao usarmos a metodologia bayesiana para o Modelo Tobit nesse caso.

## 6.2 Horas trabalhadas por mulheres casadas

O segundo conjunto de dados é uma replicação parcial da análise feita por Mroz (1987), onde foi estudado um modelo para a oferta de trabalho feminino para mulheres casadas. Os dados apesar de serem facilmente encontrados, estão na maioria das vezes incompletos. A sub-amostra utilizada por Mroz para a análise foi retirada do Panel Study of Income Dynamics de 1975, da Universidade de Michigan. Este conjunto de dados contém 753 observações (sendo 44% delas observações consideradas censuradas) e 22 variáveis. Assim como a primeira aplicação apesar de ser impossível termos valores negativos de horas trabalhadas a censura foi utilizada porque devido as combinações das estimativas dos parâmetros seria explicável termos valores negativos, o que indicaria ser altamente improvável essas mulheres terem trabalhado no ano da pesquisa. Para essa replicação iram ser usadas apenas 6 variáveis fora a variável resposta. São elas:

- $Y$ : horas trabalhadas / 100.
- $X_1$ : quantidade de filhos com menos de 6 anos.
- $X_2$ : quantidade de filhos com idade entre 7 e 18 anos.
- $X_3$ : idade.
- $X_4$ : anos de educação.
- $X_5$ : anos de experiência no mercado de trabalho.

Tabela 5 – Inferências bayesianas para o modelo reduzido.

Parâmetros	Estimativa	EP	Intervalo HPD 95%	
$\hat{\beta}_0$	13.32222	3.82884	5.77715	20.82381
$\hat{\beta}_1$	-9.23227	1.12609	-11.69604	-7.20392
$\hat{\beta}_3$	-0.60861	0.06863	-0.73390	-0.46690
$\hat{\beta}_4$	0.74470	0.20271	0.35428	1.13687
$\hat{\beta}_5$	0.80813	0.06445	0.69139	0.94322
$\hat{\sigma}^2$	131.12270	9.84459	112.57073	150.95659

- $X_6$ : (renda familiar - salário da mulher  $\times$  horas trabalhadas) / 1000.

No banco de dados original a variável resposta (horas trabalhadas) se encontrava inteira, mas para o nosso estudo a dividimos por 100, pois assim é justificável o uso de censura para quando a variável resposta for igual a 0. Temos em Bolfarine et. al (2013), no Capítulo 2, outra análise para esse mesmo conjunto de dados onde foi utilizado o Modelo Tobit com excesso de zeros porque não é viável, para o caso desse banco de dados, considerar censura no valor 0, pois a média está muito distante desse valor.

Tabela 4 – Inferências bayesianas para o modelo.

Parâmetros	Estimativa	EP	Intervalo HPD 95%	
$\hat{\beta}_0$	13.61992	4.49438	4.91237	22.86926
$\hat{\beta}_1$	-9.33866	1.11457	-11.55169	-7.23985
$\hat{\beta}_2$	-0.17912	0.39950	-0.94832	0.59585
$\hat{\beta}_3$	-0.59604	0.07464	-0.74274	-0.44702
$\hat{\beta}_4$	0.86099	0.22232	0.39822	1.27455
$\hat{\beta}_5$	0.77419	0.06927	0.62907	0.89985
$\hat{\beta}_6$	-0.08376	0.04500	-0.16596	0.00694
$\hat{\sigma}^2$	131.30415	9.81127	113.05615	151.17816

Na tabela 4 estão apresentados as inferências obtidas para o modelo. Ao observarmos o intervalo HPD é possível notar que dois coeficientes não são significantes, são eles  $\beta_2$  e  $\beta_6$ , que representam as respectivas variáveis; quantidade de filhos com idade entre 7 e 18 anos e a renda familiar menos o produto do salário da mulher com as horas trabalhadas, dividido por 1000. Com isso iremos remodelar os dados retirando essas duas variáveis do modelo, e então será feita a comparação entre os modelos para sabermos se, de fato, ocorre uma melhora ao retirarmos tais variáveis.

Na tabela 5 podemos ver que as estimativas pontuais para as variáveis em comum entre o modelo completo e modelo reduzido estão bem próximas, com todas permanecendo significantes para o modelo.

Tabela 6 – Medidas para comparação dos modelos

	LPML	DIC
Modelo Completo	-1862.256	3724.532
Modelo Reduzido	-1862.204	3724.154

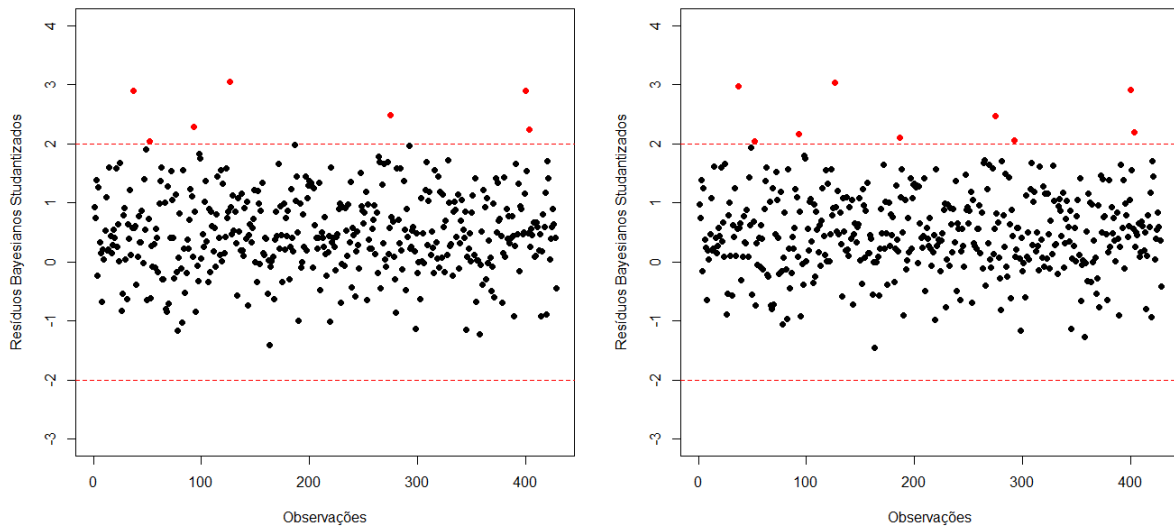
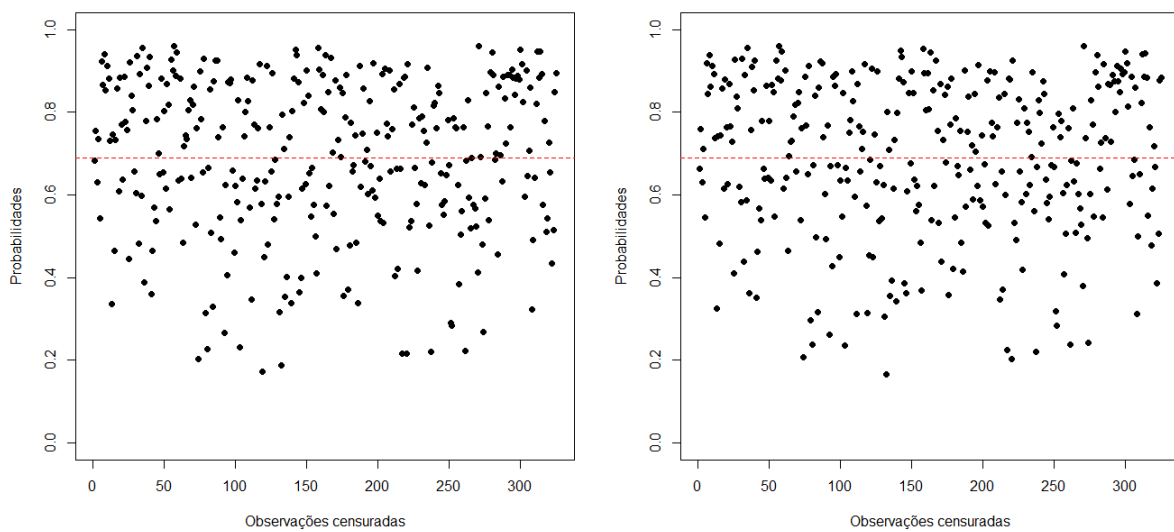


Figura 8 – Gráficos dos resíduos studentizados para a parte contínua do modelo completo e do modelo reduzido, respectivamente

Figura 9 – Gráfico para  $p(y < 0 | \mathbf{y})$  do modelo completo e do modelo reduzido, respectivamente

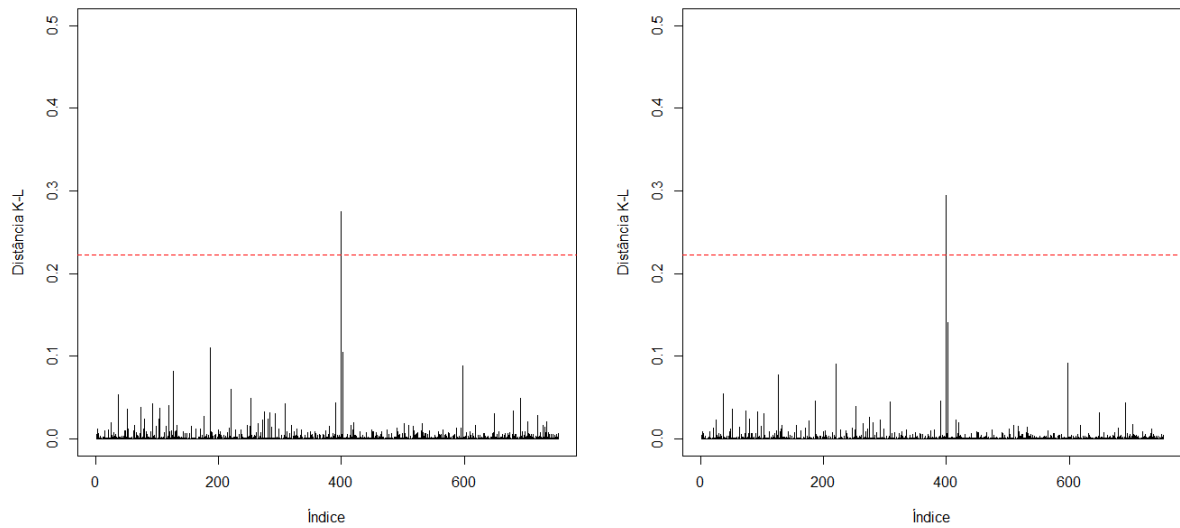


Figura 10 – Gráficos da distância de Kullback-Leibler do modelo completo e do modelo reduzido, respectivamente

Nas figuras 8, 10, 9 tabela 6 os modelos são comparados. Podemos analisar que no gráfico dos resíduos studentizados para as observações contínuas não acontece uma grande mudança em como as observações estão organizadas. O único acontecimento perceptível é a aparição de 2 novas observações na área fora dos limites. Porém no modelo completo elas já se encontravam bem próximas do limite. Na figura 9 podemos ver que, assim como na aplicação anterior, acontece o esperado; as probabilidades de censura das observações censuradas estão agrupadas no intervalo 0.8 a 1. Na figura 10 é possível observar que a única observação considerada influente para o modelo completo continuou sendo influente e passou a ter um valor maior. Isso pode ser explicado ao checarmos essa observação; essa mulher teve na variável resposta um valor de 38.33, 4 filhos, sendo 2 menores de 6 anos e outros dois maiores, 38 anos, 15 anos de educação, 17 anos de experiência e 18.43 da variável  $x_6$ . Como podemos ver ela pode ser uma observação discrepante por ter um alto valor da variável resposta, bem acima da média, mesmo possuindo 2 filhos menores de 6 anos, o que diminuiria bastante o valor esperado de uma observação, apesar dessa observação.

Para concluirmos qual modelo é o melhor para o caso em questão existem os métodos de comparação de modelos discutidos no Capítulo 3. Na tabela 6 podemos ver que o modelo reduzido tem o menor valor absoluto para ambos os métodos. Por fim, podemos dizer que para o caso em questão, o modelo que melhor explica a variável resposta (quantidade de horas trabalhadas)/100 é o modelo reduzido, com as variáveis  $X_1$  quantidade de filhos com menos de 6 anos,  $X_3$  idade,  $X_4$  anos de educação e  $X_5$  anos de experiência de trabalho. Porém por ser uma diferença mínima podemos dizer que os dois modelos estão adequados

para explicar as observações.

## 7 Conclusão

O trabalho abordado retrata o uso da estatística Bayesiana para o Modelo Tobit, mas para tal aplicação foram definidos previamente o necessário para conseguir colocar em prática o que foi pretendido. Em seguida foram feitas simulações para avaliar se, de fato, a aplicação do método como estava sendo feito era viável. Após realizada as simulações o resultado obtido foi satisfatório e prosseguimos para a aplicação em dados reais.

Ao escolhermos os conjuntos de dados que são usados no trabalho, optamos por escolher dados que já foram usados para estudos anteriores na área de regressão de dados censurados, mas com a abordagem frequentista, sendo possível assim uma comparação dos resultados, para, por fim, analisarmos o desempenho do método abordado no trabalho frente a métodos que já são utilizados na literatura.

Para o primeiro conjunto de dados estudado fizemos a comparação com os resultados obtidos na análise feita em Bolfarine et. al (2013) e o resultado obtido foi bastante animador, pois pela comparação podemos notar que tanto as estimativas quanto seus respectivos erros padrão estão próximos, dando mais indícios de que a abordagem bayesiana do Modelo Tobit é factível.

Na análise do segundo conjunto de dados, foi dado um peso maior para a parte de comparação de modelos e análise de pontos de influência, pois foram rodados dois modelos para o mesmo problema e prosseguimos com a comparação de modelos, por meio das medidas de comparação LPLM e DIC.





# Referências

- [1] ALBERT, J. H & CHIB S. (1993). Bayes inference via gibbs sampling of auto-regressive time series subject to markov mean and variance shifts. *Journal of Business & Economic Statistics*, 11(1):1-15.
- [2] BOLFARINE, H., SANTOS, B., CORREIA, L., MARTINEZ, G., GOMEZ, H., & BAZAN, J. (2013). *Modelos de Regressão com Respostas Limitadas e Censuradas*. 13ª Escola de Modelos de Regressão.
- [3] BOX, G. E. & TIAO, G. C. (2011). *Bayesian inference in statistical analysis, volume 40* John Wiley & Sons
- [4] CHIB, S. (1992). Bayes inference in the tobit censored regression model. **Journal of Econometrics**, 51(1-2):79-99.
- [5] CLARKE, B. (1996). Implications of reference priors for prior information and for sample size. *Journal of the American Statistical Association*, 91(433):173–184.
- [6] FAIR, R. C. (1978). A theory of extramarital affairs. *Political Economy*, 86:45-61.
- [7] GANERMAN, D. (1996). Simulação estocástica via cadeias de markov. *12º Sinape, Caxambu/MG*, pages 101–139.
- [8] GARAY, A. M. (2014). *Modelos de regressão para dados censurados sob distribuições assimétricas*, Tese (Doutorado) - Estatística, Universidade de São Paulo.
- [9] GARAY, A. M., LACHOS, V. H., & BOLFARINE, H. (2011). On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics & Data Analysis*, 55:1304–1318.
- [10] GOLDSTEIN, M. (1976). Bayesian analysis of regression problems. *Biometrika*, 63(1):51–58.
- [11] GREENE, W. H. (2003). *Econometric Analysis*. Pearson.
- [12] MOULTON, L. H. & HAULSEY, N. A. (1995). A mixture model with detection limits for regression analyses of antibody response to vaccine. *Biometrics*, 51:1570–1578.
- [13] MROZ, L. H. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica: Journal of Econometric Society*, pág. 765-799.

- 
- [14] PAULINO, C. D., TURKMAN, M. A. A., & MURTEIRA, B. (2003). *Estatística Bayesiana*. Fundação Calouste Gulbekian.
- [15] SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P., & VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.
- [16] TOBIN, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26:24–36.
- [17] WEISBERG, S. (2005). **Applied Linear Regression**. John Wiley & Sons.
- [18] R Development Core Team (2009) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [19] "Tobit Analysis | Stata Data Analysis Examples." IDRE Stats. Acessado em 10 de Junho, 2017. <https://stats.idre.ucla.edu/stata/dae/tobit-analysis> .

## APÊNDICE A – Resíduos do modelo tobit

No Capítulo 6 foram apresentados os gráficos dos resíduos studentizados apenas para a parte contínua (observações sem a presença de censura) dos bancos de dados. Isso foi utilizado porque não há muito valor em analisarmos os resíduos da parte censurada do modelo, devido a presença de censura essas observações não admitem valores menores que o valor da censura, fazendo com que a  $E(y_i|\mathbf{y})$  ao calcularmos o resíduo seja igual a 0 quando o valor esperado é negativo (veja na tabela 11 e 12). Por isso foi feito, como alternativa de analisarmos a parte censurada do modelo, a probabilidade da  $i$ -ésima observação censurada ser menor que 0, após realizado o processo de aumento de dados.

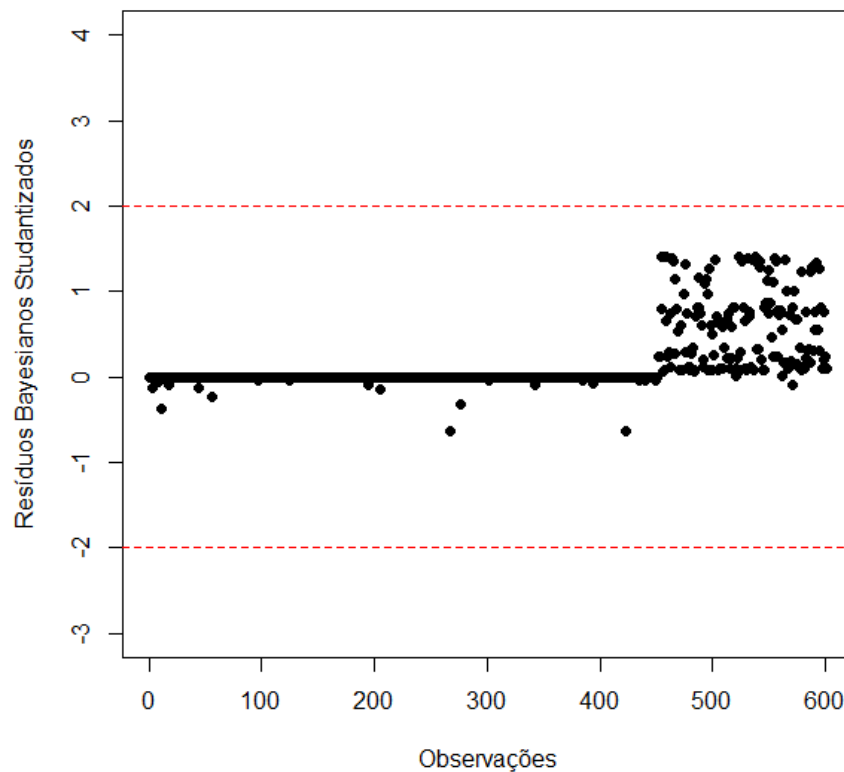


Figura 11 – Gráficos dos resíduos studentizados para o modelo dos dados extraconjugais

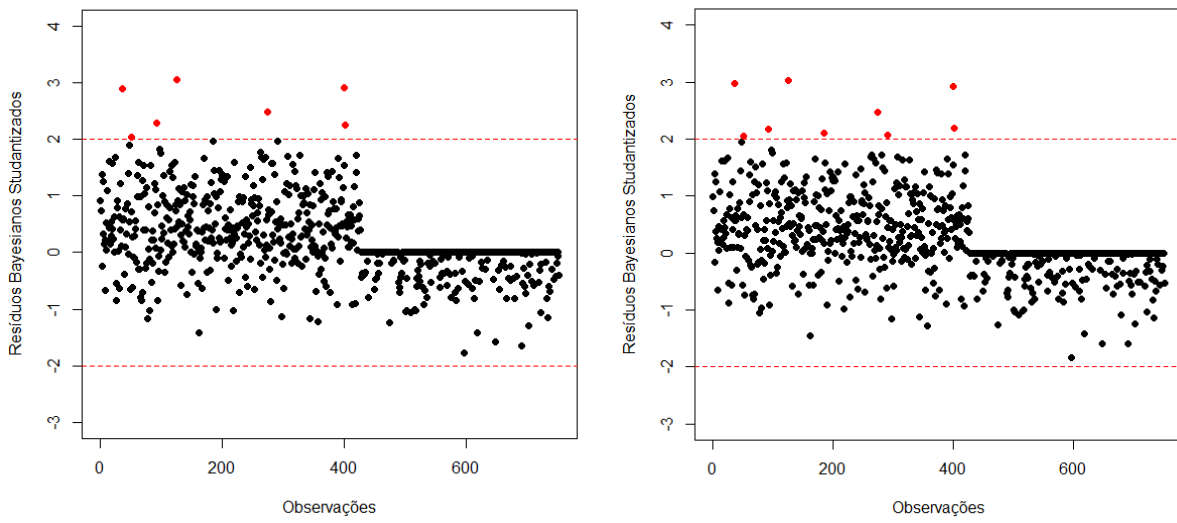


Figura 12 – Gráficos dos resíduos studentizados do modelo completo e do modelo reduzido, respectivamente, dos dados de horas trabalhadas por mulheres casada

# ANEXO A – R Script

```
##### PACOTES #####
library(MCMCpack)
library(BayesCR)
require(HDInterval)
library(AER)
library(truncnorm)
library(foreign)

##### EXEMPLO #####

cens <- read.table("https://www.ime.usp.br/~giapaula/censo.dat", quote="\")
colnames(cens) <- c("UF","Anos","Renda")

mod = lm(Renda ~ Anos, data = cens)
summary(mod)

dim(mod$beta)
dim(posterior)

posterior <- MCMCregress(Renda~Anos,data=cens,burnin = 1000,
                        mcmc = 10000,thin = 1,c0 =0.0001,d0 = .0001)

# RESIDUOS #

anos = cens$Anos
estim = posterior[1:10000,]
vetor<-matrix(0,nrow=27,ncol=10000)
vetor2<-matrix(0,nrow=27,ncol=10000)
for(i in 1:27){
  vetor[i,]<-posterior[,1]+cens$Anos[i]*posterior[,2]
  vetor2[i,]<-posterior[,3] + (posterior[,1]+cens$Anos[i]*posterior[,2])^2
}
var<-apply(vetor2,1,mean)-(apply(vetor,1,mean))^2
```

```
resid<-(cens$Renda-apply(vetor,1,mean))/sqrt(var)

# CPO #
vetor<-matrix(0,nrow=27,ncol=10000)
for(i in 1:27){
  vetor[i,]<-1/(dnorm(cens$Renda[i],posterior[,1]+cens$Anos[i]
  *posterior[,2],sqrt(posterior[,3])))
}

lpml <- sum(log((apply(vetor,1,mean))^-1)))

cpoi = apply(vetor,1,mean)^(-1)

# KULLBACK-LEIBLER #

vetor<-matrix(0,nrow=27,ncol=10000)
for(i in 1:27){
  vetor[i,]<-log(dnorm(cens$Renda[i],posterior[,1]+cens$Anos[i]
  *posterior[,2],sqrt(posterior[,3])))
}
kli = apply(vetor,1,mean)
KL = -log(cpoi) + kli

##### SIMULACAO #####

ST1 <- list()
for (i in 1:100) {

x <- runif(100,min = -1, max = 1)
y = 1 + 2*x + rnorm(100,0,1)
xi <- cbind(1,x)

sim1 = y
sim1[sim1 < quantile(sim1, .1)] = quantile(sim1, .1)
```

```
cc1 <- y
cc1[cc1 > quantile(y, .1)] = 0
cc1[cc1 != 0] = 1

ST1[[i]] = Bayes.CR(cc1, xi, sim1, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 1000, n.chains = 2,
chain = "TRUE",spacing = 1)
}

ST2 <- list()
for(i in 1:100){
x <- runif(100,min = -1, max = 1)
y = 1 + 2*x + rnorm(100,0,1)
xi <- cbind(1,x)

sim2 = y
sim2[sim2 < quantile(sim2, .2)] = quantile(sim2, .2)

cc2 <- y
cc2[cc2 > quantile(y, .2)] = 0
cc2[cc2 != 0] = 1

ST2[[i]] = Bayes.CR(cc2, xi, sim2, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 1000, n.chains = 2,
chain = "TRUE",spacing = 1)
}

ST3 <- list()
for(i in 1:100){
x <- runif(100,min = -1, max = 1)
y = 1 + 2*x + rnorm(100,0,1)
xi <- cbind(1,x)
sim3 = y
sim3[sim3 < quantile(sim3, .3)] = quantile(sim3, .3)

cc3 <- y
cc3[cc3 > quantile(y, .3)] = 0
cc3[cc3 != 0] = 1
```

```
ST3[[i]] = Bayes.CR(cc3, xi, sim3, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL, n.thin = 10, burnin = 100, n.iter = 1000, n.chains = 2,
chain = "TRUE", spacing = 1)
}
```

```
ST4 <-list()
for(i in 1:100){
x <- runif(100,min = -1, max = 1)
y = 1 + 2*x + rnorm(100,0,1)
xi <- cbind(1,x)

sim4 = y
sim4[sim4 < quantile(sim4, .4)] = quantile(sim4, .4)
```

```
cc4 <- y
cc4[cc4 > quantile(y, .4)] = 0
cc4[cc4 != 0] = 1
```

```
ST4[[i]] = Bayes.CR(cc4, x, sim4, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL, n.thin = 10, burnin = 100, n.iter = 1000, n.chains = 2,
chain = "TRUE", spacing = 1)
}
```

```
ST5 <- list()
for(i in 1:100) {
x <- runif(100,min = -1, max = 1)
y = 1 + 2*x + rnorm(100,0,1)
xi <- cbind(1,x)

sim5 = y
sim5[sim5 < quantile(sim5, .5)] = quantile(sim5, .5)
```

```
cc5 <- y
cc5[cc5 > quantile(y, .5)] = 0
cc5[cc5 != 0] = 1
```

```
ST5[[i]] = Bayes.CR(cc5, x, sim5, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL, n.thin = 10, burnin = 100, n.iter = 1000, n.chains = 2,
```



```
chain = "TRUE",spacing = 1)
}

ST6 <- list()
for(i in 1:100) {
x <- runif(100,min = -1, max = 1)
y = 1 + 2*x + rnorm(100,0,1)
xi <- cbind(1,x)

sim6 = y
sim6[sim6 < quantile(sim6, .6)] = quantile(sim6, .6)

cc6 <- y
cc6[cc6 > quantile(y, .6)] = 0
cc6[cc6 != 0] = 1

ST6[[i]] = Bayes.CR(cc6, x, sim6, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 1000, n.chains = 2,
chain = "TRUE",spacing = 1)
}

ST7 <- list()
for(i in 1:100) {
x <- runif(100,min = -1, max = 1)
y = 1 + 2*x + rnorm(100,0,1)
xi <- cbind(1,x)

sim7 = y
sim7[sim7 < quantile(sim7, .5)] = quantile(sim7, .5)

cc7 <- y
cc7[cc7 > quantile(cc7, .5)] = 0
cc7[cc7 != 0] = 1

ST7[[i]] = Bayes.CR(cc7, x, sim7, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 1000, n.chains = 2,
chain = "TRUE",spacing = 1)
}
```

```
saveRDS(ST1,"ST1.rds")
saveRDS(ST2,"ST2.rds")
saveRDS(ST3,"ST3.rds")
saveRDS(ST4,"ST4.rds")
saveRDS(ST5,"ST5.rds")
saveRDS(ST6,"ST6.rds")
saveRDS(ST7,"ST7.rds")

ST1 <- readRDS("ST1.rds")
ST2 <- readRDS("ST2.rds")
ST3 <- readRDS("ST3.rds")
ST4 <- readRDS("ST4.rds")
ST5 <- readRDS("ST5.rds")
ST6 <- readRDS("ST6.rds")
ST7 <- readRDS("ST7.rds")

#      MEDIAS ESTIMADAS      #

beta01 = matrix()
beta11 = matrix()
sigma1 = matrix()
covb01 = matrix()
covb11 = matrix()
covs1 = matrix()

for(i in 1:100){
  beta01[i]<-mean(ST1[[i]]$beta[,1])
  beta11[i]<-mean(ST1[[i]]$beta[,2])
  sigma1[i]<-mean(ST1[[i]]$sigma2)
  if
  ((hdi(ST1[[i]]$beta[,1], credMass = .95)[1] <= 1) &&
  (1 <= hdi(ST1[[i]]$beta[,1], credMass = .95)[2]))
    {covb01[i] = 1} else {
    covb01[i] = 0
  }
  if
```

```

((hdi(ST1[[i]]$beta[,2], credMass = .95)[1] <= 2) &&
(2 <= hdi(ST1[[i]]$beta[,2], credMass = .95)[2]))
{covb11[i] = 1} else {
  covb11[i] = 0
}
if
((hdi(ST1[[i]]$sigma2, credMass = .95)[1] <= 1) &&
(1 <= hdi(ST1[[i]]$sigma2, credMass = .95)[2]))
{covs1[i] = 1} else {
  covs1[i] = 0
}
}

mbeta11 = mean(beta11)
sdbeta11 = sd(beta11)
mbeta01 = mean(beta01)
sdbeta01 = sd(beta01)
msigma1 = mean(sigma1)
sdsig1 = sd(sigma1)

beta02 = matrix()
beta12 = matrix()
sigma2 = matrix()
covb02 = matrix()
covb12 = matrix()
covs2 = matrix()

for(i in 1:100){
  beta02[i]<-mean(ST2[[i]]$beta[,1])
  beta12[i]<-mean(ST2[[i]]$beta[,2])
  sigma2[i]<-mean(ST2[[i]]$sigma2)
  if
  ((hdi(ST2[[i]]$beta[,1], credMass = .95)[1] <= 1) &&
  (1 <= hdi(ST2[[i]]$beta[,1], credMass = .95)[2]))
  {covb02[i] = 1} else {
    covb02[i] = 0
  }
  if
  ((hdi(ST2[[i]]$beta[,2], credMass = .95)[1] <= 2) &&

```

```

(2 <= hdi(ST2[[i]]$beta[,2], credMass = .95)[2]))
{covb12[i] = 1} else {
  covb12[i] = 0
}
}
if
((hdi(ST2[[i]]$sigma2, credMass = .95)[1] <= 1) &&
(1 <= hdi(ST2[[i]]$sigma2, credMass = .95)[2]))
{covs2[i] = 1} else {
  covs2[i] = 0
}
}
mbeta12 = mean(beta12)
sdbeta12 = sd(beta12)
mbeta02 = mean(beta02)
sdbeta02 = sd(beta02)
msigma2 = mean(sigma2)
sdsig2 = sd(sigma2)

beta03 = matrix()
beta13 = matrix()
sigma3 = matrix()
covb03 = matrix()
covb13 = matrix()
covs3 = matrix()

for(i in 1:100){
  beta03[i]<-mean(ST3[[i]]$beta[,1])
  beta13[i]<-mean(ST3[[i]]$beta[,2])
  sigma3[i]<-mean(ST3[[i]]$sigma2)
  if
  ((hdi(ST3[[i]]$beta[,1], credMass = .95)[1] <= 1) &&
  (1 <= hdi(ST3[[i]]$beta[,1], credMass = .95)[2]))
  {covb03[i] = 1} else {
    covb03[i] = 0
  }
  if
  ((hdi(ST3[[i]]$beta[,2], credMass = .95)[1] <= 2) &&
  (2 <= hdi(ST3[[i]]$beta[,2], credMass = .95)[2]))
  {covb13[i] = 1} else {

```

```
covb13[i] = 0
}
if
((hdi(ST3[[i]]$sigma2, credMass = .95)[1] <= 1) &&
(1 <= hdi(ST3[[i]]$sigma2, credMass = .95)[2]))
{covs3[i] = 1} else {
  covs3[i] = 0
}
}
mbeta13 = mean(beta13)
sdbeta13 = sd(beta13)
mbeta03 = mean(beta03)
sdbeta03 = sd(beta03)
msigma3 = mean(sigma3)
sdsig3 = sd(sigma3)

beta04 = matrix()
beta14 = matrix()
sigma4 = matrix()
covb04 = matrix()
covb14 = matrix()
covs4 = matrix()

for(i in 1:100){
  beta04[i]<-mean(ST4[[i]]$beta[,1])
  beta14[i]<-mean(ST4[[i]]$beta[,2])
  sigma4[i]<-mean(ST4[[i]]$sigma2)
  if
  ((hdi(ST4[[i]]$beta[,1], credMass = .95)[1] <= 1) &&
  (1 <= hdi(ST4[[i]]$beta[,1], credMass = .95)[2]))
  {covb04[i] = 1} else {
    covb04[i] = 0
  }
  if
  ((hdi(ST4[[i]]$beta[,2], credMass = .95)[1] <= 2) &&
  (2 <= hdi(ST4[[i]]$beta[,2], credMass = .95)[2]))
  {covb14[i] = 1} else {
    covb14[i] = 0
  }
}
```

```

}
if
((hdi(ST4[[i]]$sigma2, credMass = .95)[1] <= 1) &&
(1 <= hdi(ST4[[i]]$sigma2, credMass = .95)[2]))
{covs4[i] = 1} else {
  covs4[i] = 0
}
}
mbeta14 = mean(beta14)
sdbeta14 = sd(beta14)
mbeta04 = mean(beta04)
sdbeta04 = sd(beta04)
msigma4 = mean(sigma4)
sdsig4 = sd(sigma4)

beta05= matrix()
beta15 = matrix()
sigma5 = matrix()
covb05 = matrix()
covb15 = matrix()
covs5 = matrix()

for(i in 1:100){
  beta05[i]<-mean(ST7[[i]]$beta[,1])
  beta15[i]<-mean(ST7[[i]]$beta[,2])
  sigma5[i]<-mean(ST7[[i]]$sigma2)
  if
  ((hdi(ST5[[i]]$beta[,1], credMass = .95)[1] <= 1) &&
  (1 <= hdi(ST5[[i]]$beta[,1], credMass = .95)[2]))
  {covb05[i] = 1} else {
    covb05[i] = 0
  }
  if
  ((hdi(ST5[[i]]$beta[,2], credMass = .95)[1] <= 2) &&
  (2 <= hdi(ST5[[i]]$beta[,2], credMass = .95)[2]))
  {covb15[i] = 1} else {
    covb15[i] = 0
  }
  if

```

```
((hdi(ST5[[i]]$sigma2, credMass = .95)[1] <= 1) &&
(1 <= hdi(ST5[[i]]$sigma2, credMass = .95)[2]))
{covs5[i] = 1} else {
  covs5[i] = 0
}
}
mbeta15 = mean(beta15)
sdbeta15 = sd(beta15)
mbeta05 = mean(beta05)
sdbeta05 = sd(beta05)
msigma5 = mean(sigma5)
sdsig5 = sd(sigma5)

beta06 = matrix()
beta16 = matrix()
sigma6 = matrix()
covb06 = matrix()
covb16 = matrix()
covs6 = matrix()

for(i in 1:100){
  beta06[i] <- mean(ST6[[i]]$beta[,1])
  beta16[i] <- mean(ST6[[i]]$beta[,2])
  sigma6[i] <- mean(ST6[[i]]$sigma2)
  if
  ((hdi(ST6[[i]]$beta[,1], credMass = .95)[1] <= 1) &&
  (1 <= hdi(ST6[[i]]$beta[,1], credMass = .95)[2]))
  {covb06[i] = 1} else {
    covb06[i] = 0
  }
  if
  ((hdi(ST6[[i]]$beta[,2], credMass = .95)[1] <= 2) &&
  (2 <= hdi(ST6[[i]]$beta[,2], credMass = .95)[2]))
  {covb16[i] = 1} else {
    covb16[i] = 0
  }
  if
  ((hdi(ST6[[i]]$sigma2, credMass = .95)[1] <= 1) &&
```

```

(1 <= hdi(ST6[[i]]$sigma2, credMass = .95)[2]))
{covs6[i] = 1} else {
  covs6[i] = 0
}
}
mbeta16 = mean(beta16)
sdbeta16 = sd(beta16)
mbeta06 = mean(beta06)
sdbeta06 = sd(beta06)
msigma6 = mean(sigma6)
sdsig6 = sd(sigma6)

beta07 = matrix()
beta17 = matrix()
sigma7 = matrix()
covb07 = matrix()
covb17 = matrix()
covs7 = matrix()

for(i in 1:100){
  beta07[i]<-mean(ST7[[i]]$beta[,1])
  beta17[i]<-mean(ST7[[i]]$beta[,2])
  sigma7[i]<-mean(ST7[[i]]$sigma2)
  if
  ((hdi(ST7[[i]]$beta[,1], credMass = .95)[1] <= 1) &&
  (1 <= hdi(ST7[[i]]$beta[,1], credMass = .95)[2]))
  {covb07[i] = 1} else {
    covb07[i] = 0
  }
  if
  ((hdi(ST7[[i]]$beta[,2], credMass = .95)[1] <= 2) &&
  (2 <= hdi(ST7[[i]]$beta[,2], credMass = .95)[2]))
  {covb17[i] = 1} else {
    covb17[i] = 0
  }
  if
  ((hdi(ST7[[i]]$sigma2, credMass = .95)[1] <= 1) &&
  (1 <= hdi(ST7[[i]]$sigma2, credMass = .95)[2]))
  {covs7[i] = 1} else {

```



```
        covs7[i] = 0
    }

}

mbeta17 = mean(beta17)
sdbeta17 = sd(beta17)
mbeta07 = mean(beta07)
sdbeta07 = sd(beta07)
msigma7 = mean(sigma7)
sdsig7 = sd(sigma7)

# TABELA - VALORES PARAMETROS #

betas0 = c(mbeta01,mbeta02, mbeta03, mbeta04, mbeta05, mbeta06, mbeta07)
varbeta0 = c(sdbeta01,sdbeta02,sdbeta03,sdbeta04,sdbeta05,sdbeta06,sdbeta07)
betas1 = c(mbeta11,mbeta12, mbeta13, mbeta14, mbeta15, mbeta16, mbeta17)
varbeta1 = c(sdbeta11,sdbeta12,sdbeta13,sdbeta14,sdbeta15,sdbeta16,sdbeta17)
sigmas = c(msigma1,msigma2, msigma3, msigma4, msigma5, msigma6, msigma7)
varsigma = c(sdsig1,sdsig2,sdsig3,sdsig4,sdsig5,sdsig6,sdsig7)

sum(covb01)
sum(covb11)
sum(covs1)

sum(covb02)
sum(covb12)
sum(covs2)

sum(covb03)
sum(covb13)
sum(covs3)

sum(covb04)
sum(covb14)
sum(covs4)

sum(covb05)
sum(covb15)
sum(covs5)
```

```
sum(covb06)
sum(covb16)
sum(covs6)

sum(covb07)
sum(covb17)
sum(covs7)

# GRAFICOS #

plot(c(10,20,30,40,50,60,70),betas0, type = 'l',ylim=c(0.5, 1.5), main = "Beta 0",
ylab = "MC Média", xlab = "Porcentagem de Censura" )
abline(h=1,lty=2, col="blue")

plot(c(10,20,30,40,50,60,70),betas1, type = 'l',ylim=c(1.5, 2.5), main = "Beta 1"
,ylab = "MC Média", xlab = "Porcentagem de Censura" )
abline(h=2,lty=2, col="blue")

plot(c(10,20,30,40,50,60,70),sigmas, type = 'l',ylim=c(0.5, 1.5), main = "Beta 1"
,ylab = "MC Média", xlab = "Porcentagem de Censura" )
abline(h=1,lty=2, col="blue")

##### GRAFICOS SIMULACAO 2 #####

x <- runif(100,min = -1, max = 1)
y = 1 + 2*x + rnorm(100,0,1)
xi <- cbind(1,x)

sim = y
sim[sim < quantile(sim, .3)] = quantile(sim, .3)

cc <- y
cc[cc > quantile(y, .3)] = 0
cc[cc != 0] = 1

CT1 = Bayes.CR(cc, xi, sim, cens=1, dist="Normal", prior = "Hierar",
```

---

```
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 8000, n.chains = 2,  
chain = "TRUE",spacing = 1)  
saveRDS(CT1,"CT1.rds")
```

```
x <- runif(100,min = -1, max = 1)  
y = 1 + 2*x + rnorm(100,0,1)  
xi <- cbind(1,x)
```

```
sim = y  
sim[sim < quantile(sim, .6)] = quantile(sim, .6)
```

```
cc <- y  
cc[cc > quantile(y, .6)] = 0  
cc[cc != 0] = 1
```

```
CT2 = Bayes.CR(cc, xi, sim, cens=1, dist="Normal", prior = "Hierar",  
hyper = NULL, n.thin = 10, burnin = 100, n.iter = 8000, n.chains = 2,  
chain = "TRUE",spacing = 1)  
saveRDS(CT2,"CT2.rds")
```

```
x <- runif(50,min = -1, max = 1)  
y = 1 + 2*x + rnorm(50,0,1)  
xi <- cbind(1,x)
```

```
sim = y  
sim[sim < quantile(sim, .3)] = quantile(sim, .3)
```

```
cc <- y  
cc[cc > quantile(y, .3)] = 0  
cc[cc != 0] = 1
```

```
cT1 = Bayes.CR(cc, xi, sim, cens=1, dist="Normal", prior = "Hierar",  
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 8000, n.chains = 2,  
chain = "TRUE",spacing = 1)  
saveRDS(cT1,"ccT1.rds")
```

```
x <- runif(50,min = -1, max = 1)
```

```
y = 1 + 2*x + rnorm(50,0,1)
xi <- cbind(1,x)
```

```
sim = y
sim[sim < quantile(sim, .6)] = quantile(sim, .6)
```

```
cc <- y
cc[cc > quantile(y, .6)] = 0
cc[cc != 0] = 1
```

```
cT2 = Bayes.CR(cc, xi, sim, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL, n.thin = 10, burnin = 100, n.iter = 8000, n.chains = 2,
chain = "TRUE", spacing = 1)
saveRDS(cT2, "ccT2.rds")
```

```
x <- runif(300, min = -1, max = 1)
y = 1 + 2*x + rnorm(300,0,1)
xi <- cbind(1,x)
```

```
sim = y
sim[sim < quantile(sim, .3)] = quantile(sim, .3)
```

```
cc <- y
cc[cc > quantile(y, .3)] = 0
cc[cc != 0] = 1
```

```
TT1 = Bayes.CR(cc, xi, sim, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL, n.thin = 10, burnin = 100, n.iter = 8000, n.chains = 2,
chain = "TRUE", spacing = 1)
saveRDS(TT1, "TT1.rds")
```

```
x <- runif(300, min = -1, max = 1)
y = 1 + 2*x + rnorm(300,0,1)
xi <- cbind(1,x)
```

```
sim = y
sim[sim < quantile(sim, .6)] = quantile(sim, .6)
```

```
cc <- y
```

```
cc[cc > quantile(y, .6)] = 0
cc[cc != 0] = 1
```

```
TT2 = Bayes.CR(cc, xi, sim, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 8000, n.chains = 2,
chain = "TRUE",spacing = 1)
saveRDS(TT2,"TT2.rds")
```

```
d1 = density(CT1$beta[,1])
d2 = density(cT1$beta[,1])
d3 = density(TT1$beta[,1])
```

```
plot(range(d1$x, d2$x, d3$x), range(d1$y, d2$y, d3$y), type = "n",
xlab = "Beta 0",ylab = "FDP")
lines(d1, col = "#0038c6",type = "l")
lines(d2, col = "#c68e00",type = "l")
lines(d3, col = "#002e20",type = "l")
# create value labels
cyl.f <- factor(levels= c(50,100,300),
                labels = c("n = 50", "n = 100", "n = 300"))
# add legend via mouse click
legend(locator(1), levels(cyl.f), fill=c("#c68e00","#0038c6","#002e20"),
       ,bty="n",cex=0.7,pt.cex=0.7)
```

```
d1 = density(CT1$beta[,2])
d2 = density(cT1$beta[,2])
d3 = density(TT1$beta[,2])
```

```
plot(range(d1$x, d2$x, d3$x), range(d1$y, d2$y, d3$y), type = "n",
xlab = "Beta 1",ylab = "FDP")
lines(d1, col = "#0038c6",type = "l")
lines(d2, col = "#c68e00",type = "l")
lines(d3, col = "#002e20",type = "l")
# adicionando legenda pelo clique #
legend(locator(1), levels(cyl.f), fill=c("#c68e00","#0038c6","#002e20"),
       ,bty="n",cex=0.7,pt.cex=0.7)
```

```
d1 = density(CT1$sigma2)
d2 = density(cT1$sigma2)
d3 = density(TT1$sigma2)

plot(range(d1$x, d2$x, d3$x), range(d1$y, d2$y, d3$y), type = "n",
      xlab = "Variância", ylab = "FDP")
lines(d1, col = "#0038c6", type = "l")
lines(d2, col = "#c68e00", type = "l")
lines(d3, col = "#002e20", type = "l")
legend(locator(1), levels(cyl.f), fill=c("#c68e00", "#0038c6", "#002e20"),
       , bty="n", cex=0.7, pt.cex=0.7)

d1 = density(CT2$beta[,1])
d2 = density(cT2$beta[,1])
d3 = density(TT2$beta[,1])

plot(range(d1$x, d2$x, d3$x), range(d1$y, d2$y, d3$y), type = "n",
      xlab = "Beta 0", ylab = "FDP")
lines(d1, col = "#0038c6", type = "l")
lines(d2, col = "#c68e00", type = "l")
lines(d3, col = "#002e20", type = "l")
legend(locator(1), levels(cyl.f), fill=c("#c68e00", "#0038c6", "#002e20"),
       , bty="n", cex=0.7, pt.cex=0.7)

d1 = density(CT2$beta[,2])
d2 = density(cT2$beta[,2])
d3 = density(TT2$beta[,2])

plot(range(d1$x, d2$x, d3$x), range(d1$y, d2$y, d3$y), type = "n",
      xlab = "Beta 1", ylab = "FDP")
lines(d1, col = "#0038c6", type = "l")
lines(d2, col = "#c68e00", type = "l")
lines(d3, col = "#002e20", type = "l")
legend(locator(1), levels(cyl.f), fill=c("#c68e00", "#0038c6", "#002e20"),
       , bty="n", cex=0.7, pt.cex=0.7)

d1 = density(CT2$sigma2)
d2 = density(cT2$sigma2)
```

```
d3 = density(TT2$sigma2)

plot(range(d1$x, d2$x, d3$x), range(d1$y, d2$y, d3$y), type = "n",
     xlab = "Variância",ylab = "FDP")
lines(d1, col = "#0038c6",type = "l")
lines(d2, col = "#c68e00",type = "l")
lines(d3, col = "#002e20",type = "l")
legend(locator(1), levels(cyl.f), fill=c("#c68e00","#0038c6","#002e20"),
      ,bty="n",cex=0.7,pt.cex=0.7)

##### DADOS REAIS #####

# DADOS AFFAIR #
data("Affairs")

aff <- Affairs[,c(3,4,6,9)]

y <- Affairs[,1]
x <- cbind(1,aff)

cc <- y
cc[c(1:451)] <- 1
cc[c(452:601)] <- 0

mod = Bayes.CR(cc, x, y, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 8000, n.chains = 2,
chain = "TRUE",criteria = "TRUE",spacing = 1)
summary(mod)

# DADOS MROZ #
mroz = read.dta("C:\\Users\\aluno\\Documents\\My SAS Files 1.0\\mroz.dta")
mroz = read.dta("C:\\Users\\matheus.porto\\Documents\\TCC\\mroz.dta")

y <- mroz[,2]/100
x <-cbind(1,mroz[,c(3:6,19,20)])

cc <- y
cc[c(429:753)] <- 1
```

```

cc[c(1:428)] <- 0
mod1 = Bayes.CR(cc, x, y, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 8000, n.chains = 2,
chain = "TRUE",criteria = "TRUE",spacing = 1)

x0 <-x[,c(1,2,4,5,6)]
mod0 = Bayes.CR(cc, x0, y, cens=1, dist="Normal", prior = "Hierar",
hyper = NULL,n.thin = 10, burnin = 100, n.iter = 8000, n.chains = 2,
chain = "TRUE",criteria = "TRUE",spacing = 1)

##### ESTADISTICAS AFFAIR #####
# RESIDUOS #

teste<-matrix(0,nrow=601,ncol=1580)
teste2<-matrix(0,nrow=601,ncol=1580)
for(i in 1:601){
  teste[i,]<-mod$beta[,1]+x[i,2]*mod$beta[,2]+x[i,3]*mod$beta[,3]+x[i,4]
  *mod$beta[,4]+x[i,5]*mod$beta[,5]
  teste2[i,]<-mod$sigma2 +
  (mod$beta[,1]+x[i,2]*mod$beta[,2]+x[i,3]*mod$beta[,3]+x[i,4]*mod$beta[
  ,4]+x[i,5]*mod$beta[,5])^2
}

esp = apply(teste,1,mean)
esp[esp<0] = 0
esp2 = apply(teste2,1,mean)

var<-apply(teste2,1,mean)-(esp)^2

resid<-(y-esp)/sqrt(var)

# CPO #
teste<-matrix(0,nrow=601,ncol=1580)

for(i in 1:451){
  teste[i,]<-1/(pnorm(0,mod$beta[,1]*x[i,1]+x[i,2]*mod$beta[,2]+x[i,3]
  *mod$beta[,3]+x[i,4]*mod$beta[,4]+x[i,5]*mod$beta[,5],sqrt(mod$sigma2)

```



```

    ))
  }

for(i in 1:150){
  teste[451+i,]<-1/(dnorm(y[451+i],mod$beta[,1]+x[451+i,2]*mod$beta[,2]
+x[451+i,3]*mod$beta[,3]+x[451+i,4]*mod$beta[,4]+x[451+i,5]
*mod$beta[,5],sqrt(mod$sigma2)))
}
lpml = sum(log((apply(teste,1,mean))^(-1)))

cpoi = apply(teste,1,mean)^(-1)

# KULLBACK-LEIBLER #

teste<-matrix(0,nrow=601,ncol=1580)

for(i in 1:451){
  teste[i,]<-log(pnorm(0,mod$beta[,1]*x[i,1]+x[i,2]*mod$beta[,2]+x[i,3]
*mod$beta[,3]+x[i,4]*mod$beta[,4]+x[i,5]*mod$beta[,5],
sqrt(mod$sigma2)))
}

for(i in 1:150){
  teste[451+i,]<-log(dnorm(y[451+i],mod$beta[,1]+x[451+i,2]*mod$beta[,2]
+x[451+i,3]*mod$beta[,3]+x[451+i,4]*mod$beta[,4]+x[451+i,5]
*mod$beta[,5],sqrt(mod$sigma2)))
}

kli = apply(teste,1,mean)
KL = -log(cpoi) + kli

# DIC #
vetor<-matrix(0,nrow=601,ncol=1580)
vetor2 <- matrix(0,nrow=601,ncol=1)
for(i in 1:451){
  vetor[i,]<- -2*log(pnorm(0,mod$beta[,1]*x[i,1]+x[i,2]*mod$beta[,2]
+x[i,3]*mod$beta[,3]+x[i,4]*mod$beta[,4]+x[i,5]*mod$beta[,5],
sqrt(mod$sigma2)))
}

```

```

for(i in 1:150){
  vetor[451+i,]<- -2*log(dnorm(y[451+i],mod$beta[,1]+x[451+i,2]
  *mod$beta[,2]+x[451+i,3]*mod$beta[,3]+x[451+i,4]*mod$beta[,4]
  +x[451+i,5]*mod$beta[,5],sqrt(mod$sigma2)))
}

for(i in 1:451){
  vetor2[i,]<- -2*log(pnorm(0,9.10380+x[i,2]*(-.16411)+x[i,3]*.55324
  +x[i,4]*(-1.75351)+x[i,5]*(-2.31507),sqrt(72.84652)))
}

for(i in 1:150){
  vetor2[451+i,]<- -2*log(dnorm(y[451+i],9.10380+x[451+i,2]*(-.16411)
  +x[451+i,3]*.55324+x[451+i,4]*(-1.75351)+x[451+i,5]*(-2.31507),
  sqrt(72.84652)))
}

pd <- mean(apply(vetor,2,sum)) - sum(vetor2)
dic <- sum(vetor2) + 2*pd

# GRAFICOS #

vetor <-matrix(0,nrow=451,ncol=1580)
for(i in 1:451){
  vetor[i,]<-pnorm(0,mod$beta[,1]*x[i,1]+x[i,2]*mod$beta[,2]+x[i,3]
  *mod$beta[,3]+x[i,4]*mod$beta[,4]+x[i,5]*mod$beta[,5],sqrt(mod$sigma2))
}

graf = apply(vetor,1,mean)
plot(graf, xlab = "Observações censuradas", ylab = "Probabilidades",
ylim = c(0,1) )
points(graf,pch = 19)

resid = resid[c(452:601)]
plot(resid, xlab="Observações",
ylab="Resíduos Bayesianos Studantizados",ylim=c(-3,4))
abline(h=c(-2,2),lwd=1, lty=2, col = "red")
points(resid,pch = 19)

```

```

pp = c(which(resid>2))
points(pp,resid[resid>2],pch = 19,col="red")

plot(KL,type = 'h', xlab = "Índice",
ylab = "Distância K-L",ylim = c(0,0.3) )
abline(h=(-log(0.8*2)-log(2*.2))/2,lwd=1, lty=2, col = "red")

#### MROZ ####
# RESIDUOS #

teste<-matrix(0,nrow=753,ncol=1580)
teste2<-matrix(0,nrow=753,ncol=1580)
for(i in 1:753){
  teste[i,]<-mod1$beta[,1]+x[i,2]*mod1$beta[,2]+x[i,3]*mod1$beta[,3]
  +x[i,4]*mod1$beta[,4]+x[i,5]*mod1$beta[,5]+x[i,6]*mod1$beta[,6]
  +x[i,7]*mod1$beta[,7]
}
for(i in 1:753){
  teste2[i,]<-mod1$sigma2
  +(mod1$beta[,1]+x[i,2]*mod1$beta[,2]+x[i,3]*mod1$beta[,3]+x[i,4]
  *mod1$beta[,4]+x[i,5]*mod1$beta[,5]+x[i,6]*mod1$beta[,6]+x[i,7]
  *mod1$beta[,7])^2
}

esp = apply(teste,1,mean)
esp[esp<0] = 0
esp2 = apply(teste2,1,mean)

var<-apply(teste2,1,mean)-(esp)^2

resid<-(y-esp)/sqrt(var)

# CPO #
teste<-matrix(0,nrow=753,ncol=1580)

for(i in 429:753){
  teste[i,]<-1/(pnorm(0,mod1$beta[,1]+x[i,2]*mod1$beta[,2]+x[i,3]
  *mod1$beta[,3]+x[i,4]*mod1$beta[,4]+x[i,5]*mod1$beta[,5]+x[i,6]
  *mod1$beta[,6]+x[i,7]*mod1$beta[,7],sqrt(mod1$sigma2)))
}

```

```

}

for(i in 1:428){
  teste[i,]<-1/(dnorm(y[i],mod1$beta[,1]+x[i,2]*mod1$beta[,2]+x[i,3]
  *mod1$beta[,3]+x[i,4]*mod1$beta[,4]+x[i,5]*mod1$beta[,5]+x[i,6]
  *mod1$beta[,6]+x[i,7]*mod1$beta[,7],sqrt(mod1$sigma2)))
}

lpml = sum(log((apply(teste,1,mean))^(-1)))

cpoi = apply(teste,1,mean)^(-1)

# KULLBACK-LEIBLER #

teste<-matrix(0,nrow=753,ncol=1580)
for(i in 429:753){
  teste[i,]<-log(pnorm(0,mod1$beta[,1]+x[i,2]*mod1$beta[,2]+x[i,3]
  *mod1$beta[,3]+x[i,4]*mod1$beta[,4]+x[i,5]*mod1$beta[,5]+x[i,6]
  *mod1$beta[,6]+x[i,7]*mod1$beta[,7],sqrt(mod1$sigma2)))
}

for(i in 1:428){
  teste[i,]<-log(dnorm(y[i],mod1$beta[,1]+x[i,2]*mod1$beta[,2]+x[i,3]
  *mod1$beta[,3]+x[i,4]*mod1$beta[,4]+x[i,5]*mod1$beta[,5]+x[i,6]
  *mod1$beta[,6]+x[i,7]*mod1$beta[,7],sqrt(mod1$sigma2)))
}

kli = apply(teste,1,mean)
KL = -log(cpoi) + kli

# DIC #
vetor<-matrix(0,nrow=753,ncol=1580)
vetor2 <- matrix(0,nrow=753,ncol=1)
for(i in 429:753){
  vetor[i,]<- -2*log(pnorm(0,mod1$beta[,1]+x[i,2]*mod1$beta[,2]+x[i,3]
  *mod1$beta[,3]+x[i,4]*mod1$beta[,4]+x[i,5]*mod1$beta[,5]+x[i,6]
  *mod1$beta[,6]+x[i,7]*mod1$beta[,7],sqrt(mod1$sigma2)))
}

for(i in 1:428){

```

```

vetor[i,]<- -2*log(dnorm(y[i],mod1$beta[,1]+x[i,2]*mod1$beta[,2]
+x[i,3]*mod1$beta[,3]+x[i,4]*mod1$beta[,4]+x[i,5]*mod1$beta[,5]
+x[i,6]*mod1$beta[,6]+x[i,7]*mod1$beta[,7],sqrt(mod1$sigma2)))
}

for(i in 429:753){
  vetor2[i,]<- -2*log(pnorm(0,9.10380+x[i,2]*(-.16411)+x[i,3]*.55324
+x[i,4]*(-1.75351)+x[i,5]*(-2.31507),sqrt(72.84652)))
}

for(i in 1:428){
  vetor2[451+i,]<- -2*log(dnorm(y[451+i],9.10380+x[451+i,2]*(-.16411)
+x[451+i,3]*.55324+x[451+i,4]*(-1.75351)+x[451+i,5]*(-2.31507),
sqrt(72.84652)))
}

pd <- mean(apply(vetor,2,sum)) - sum(vetor2)
dic <- sum(vetor2) + 2*pd

# GRAFICOS #

vetor <-matrix(0,nrow=325,ncol=1580)
for(i in 1:325){
  vetor[i,]<-pnorm(0,mod1$beta[,1]+x[i,2]*mod1$beta[,2]+x[i,3]
*mod1$beta[,3]+x[i,4]*mod1$beta[,4]+x[i,5]*mod1$beta[,5]+x[i,6]
*mod1$beta[,6]+x[i,7]*mod1$beta[,7],sqrt(mod1$sigma2))
}

graf = apply(vetor,1,mean)
plot(graf, xlab = "Observações censuradas", ylab = "Probabilidades",
ylim = c(0,1) )
points(graf,pch = 19)

resid = resid[c(1:428)]
plot(resid, xlab="Observações",
ylab="Resíduos Bayesianos Studentizados",ylim=c(-3,4))
abline(h=c(-2,2),lwd=1, lty=2, col = "red")
points(resid,pch = 19)
pp = c(which(resid>2))

```

```

points(pp,resid[resid>2],pch = 19,col="red")

plot(KL,type = 'h', xlab = "Índice", ylab = "Distância K-L",
ylim = c(0,0.5) )
abline(h=(-log(0.8*2)-log(2*.2))/2,lwd=1, lty=2, col = "red")

#### MROZ REDUZIDO ####
# RESIDUOS #

teste<-matrix(0,nrow=753,ncol=1580)
teste2<-matrix(0,nrow=753,ncol=1580)
for(i in 1:753){
  teste[i,]<-mod0$beta[,1]+x0[i,2]*mod0$beta[,2]+x0[i,3]*mod0$beta[,3]
  +x0[i,4]*mod0$beta[,4]+x0[i,5]*mod0$beta[,5]
}
for(i in 1:753){
  teste2[i,]<-mod1$sigma2 + (mod0$beta[,1]+x0[i,2]*mod0$beta[,2]+x0[i,3]*mod0$beta[,3]+x0[i,4]*mod0$beta[,4]+x0[i,5]*mod0$beta[,5])^2
}

esp = apply(teste,1,mean)
esp[esp<0] = 0
esp2 = apply(teste2,1,mean)

var<-apply(teste2,1,mean)-(esp)^2

resid<-(y-esp)/sqrt(var)

# CPO #
teste<-matrix(0,nrow=753,ncol=1580)

for(i in 429:753){
  teste[i,]<-1/(pnorm(0,mod0$beta[,1]+x0[i,2]*mod0$beta[,2]+x0[i,3]*mod0$beta[,3]+x0[i,4]*mod0$beta[,4]+x0[i,5]*mod0$beta[,5],
  sqrt(mod1$sigma2)))
}

for(i in 1:428){

```

---

```

teste[i,]<-1/(dnorm(y[i],mod0$beta[,1]+x0[i,2]*mod0$beta[,2]+x0[i,3]
*mod0$beta[,3]+x0[i,4]*mod0$beta[,4]+x0[i,5]*mod0$beta[,5],
sqrt(mod1$sigma2)))
}
lpml = sum(log((apply(teste,1,mean))^-1)))

cpoi = apply(teste,1,mean)^(-1)

# KULLBACK-LEIBLER #

teste<-matrix(0,nrow=753,ncol=1580)
for(i in 429:753){
  teste[i,]<-log(pnorm(0,mod0$beta[,1]+x0[i,2]*mod0$beta[,2]+x0[i,3]
*mod0$beta[,3]+x0[i,4]*mod0$beta[,4]+x0[i,5]*mod0$beta[,5],
sqrt(mod1$sigma2)))
}

for(i in 1:428){
  teste[i,]<-log(dnorm(y[i],mod0$beta[,1]+x0[i,2]*mod0$beta[,2]
+x0[i,3]*mod0$beta[,3]+x0[i,4]*mod0$beta[,4]+x0[i,5]*mod0$beta[,5],
sqrt(mod1$sigma2)))
}

kli = apply(teste,1,mean)
KL = -log(cpoi) + kli

# DIC #
vetor<-matrix(0,nrow=753,ncol=1580)
vetor2 <- matrix(0,nrow=753,ncol=1)
for(i in 429:753){
  vetor[i,]<- -2*log(pnorm(0,mod0$beta[,1]+x0[i,2]*mod0$beta[,2]
+x0[i,3]*mod0$beta[,3]+x0[i,4]*mod0$beta[,4]+x0[i,5]*mod0$beta[,5],
sqrt(mod1$sigma2)))
}

for(i in 1:428){
  vetor[i,]<- -2*log(dnorm(y[i],mod0$beta[,1]+x0[i,2]*mod0$beta[,2]
+x0[i,3]*mod0$beta[,3]+x0[i,4]*mod0$beta[,4]+x0[i,5]*mod0$beta[,5],
sqrt(mod1$sigma2)))
}

```

```

}

for(i in 429:753){
  vetor2[i,]<- -2*log(pnorm(0,9.10380+x[i,2]*(-.16411)+x[i,3]*.55324
  +x[i,4]*(-1.75351)+x[i,5]*(-2.31507),sqrt(72.84652)))
}

for(i in 1:428){
  vetor2[451+i,]<- -2*log(dnorm(y[451+i],9.10380+x[451+i,2]*(-.16411)
  +x[451+i,3]*.55324+x[451+i,4]*(-1.75351)+x[451+i,5]*(-2.31507),
  sqrt(72.84652)))
}

pd <- mean(apply(vetor,2,sum)) - sum(vetor2)
dic <- sum(vetor2) + 2*pd

# GRAFICOS #

vetor <-matrix(0,nrow=325,ncol=1580)
for(i in 1:325){
  vetor[i,]<-pnorm(0,mod0$beta[,1]+x0[i,2]*mod0$beta[,2]+x0[i,3]
  *mod0$beta[,3]+x0[i,4]*mod0$beta[,4]+x0[i,5]*mod0$beta[,5],
  sqrt(mod1$sigma2))
}

graf = apply(vetor,1,mean)
plot(graf, xlab = "Observações censuradas", ylab = "Probabilidades",
ylim = c(0,1) )
points(graf,pch = 19)

plot(resid, xlab="Observações",
ylab="Resíduos Bayesianos Studentizados",ylim=c(-3,4))
abline(h=c(-2,2),lwd=1, lty=2, col = "red")
points(resid,pch = 19)
pp = c(which(resid>2))
points(pp,resid[resid>2],pch = 19,col="red")

plot(KL,type = 'h', xlab = "Índice", ylab = "Distância K-L",
ylim = c(0,0.5) )

```



```
abline(h=(-log(0.8*2)-log(2*.2))/2,lwd=1, lty=2, col = "red")
```