

**Regressão logística com
censura nas variáveis explicativas**

Elivane da Silva Victor

DISSERTAÇÃO APRESENTADA AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO TÍTULO DE
MESTRE EM CIÊNCIAS

Programa: Estatística

Orientador: Prof. Dr. Julio da Motta Singer

Durante parte do desenvolvimento deste trabalho a autora recebeu auxílio financeiro
da CAPES

São Paulo, maio de 2015

Regressão logística com censura nas variáveis explicativas

Esta versão da dissertação contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 14/05/2015. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Julio da Motta Singer (orientador) - IME-USP
- Prof^a. Dr^a. Lucia Pereira Barroso - IME-USP
- Prof^a. Dr^a. Hildete Prisco Pinheiro - UNICAMP

Resumo

Victor, E. S. **Regressão logística com censura nas variáveis explicativas**. 2015. 46 f. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.

Consideramos o ajuste de modelos de regressão logística com variáveis explicativas contínuas e censuradas à esquerda por um limite de detecção, aplicados a conjuntos de dados pequenos ou moderados, motivados por um estudo que teve como objetivo investigar a associação entre marcadores biológicos e o diagnóstico de endometriose. Por meio de um estudo de simulação, foram comparados sete métodos de substituição das observações censuradas, a estimação por máxima verossimilhança considerando as censuras e um modelo ajustado de forma usual, com a variável explicativa sem censura. As diferenças entre os métodos foram mais evidentes em amostras com maior proporção de censuras e com maiores valores para o coeficiente associado à variável explicativa. A substituição pelo limite de detecção apresentou o pior desempenho na maioria dos casos, de acordo com as estimativas de viés. O método de máxima verossimilhança considerando a censura apresentou estimativas com menores vieses, principalmente nos casos de maior proporção de censura.

Palavras-chave: Regressão logística, limite de detecção, censura, marcadores biológicos.

Abstract

Victor, E. S. **Logistic regression with censoring in the explanatory variables.** 2015. 46 p. Dissertação (Mestrado) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2015.

We considered the adjustment of logistic regression models with continuous left-censored explanatory variables due to a limit of detection, applied to small or moderate datasets, motivated by a study which aimed to investigate the association between biomarkers and the diagnosis of endometriosis. Through a simulation study, we compared seven methods of substitution of the censored observations, maximum likelihood estimation considering censoring, and a model adjusted as usual with uncensored explanatory variable. The differences between the methods were more evident in samples with higher proportion of censoring and with higher values for the coefficient associated to the explanatory variable. Substitution by limit of detection presented the worst performance in most cases, according to bias estimates. Maximum likelihood estimation provided estimates with lower bias, especially in cases of higher proportion of censoring.

Keywords: Logistic regression, limit of detection, censoring, biomarkers.

Sumário

1	Introdução	1
2	Métodos para tratamento de censuras	13
2.1	Técnicas de substituição dos valores censurados	13
2.2	Utilização de estimadores de máxima verossimilhança	15
3	Simulação	19
4	Aplicação	29
5	Conclusões	35
A	Dados utilizados como exemplo	39
B	Código para o ajuste por máxima verossimilhança	41
C	Resultados obtidos com as simulações	45
	Referências Bibliográficas	95

Capítulo 1

Introdução

O endométrio é uma mucosa que reveste a cavidade uterina, sensível às alterações do ciclo menstrual e onde o óvulo se implanta após a fertilização. A endometriose é uma moléstia em que o tecido endometrial adere a outras regiões do abdômen. Com diferentes níveis de gravidade e causas ainda não claramente estabelecidas, a doença acomete cerca de 10% a 15% das mulheres em idade reprodutiva e ocorre mais comumente nos ovários, peritônio pélvico e ligamentos útero-sacros ou, de forma menos comum, em locais como o intestino, reto, bexiga, diafragma, vagina e parede abdominal (Podgaec & Abrão, 2004).

Algumas pacientes com endometriose são assintomáticas; no entanto, a maioria apresenta sintomas, sendo os principais dismenorreia, dor pélvica crônica, infertilidade, dispareunia de profundidade, além de sintomas intestinais e urinários cíclicos, como dor ou sangramento ao evacuar ou urinar durante o período menstrual (Bellelis et al., 2010).

O padrão áureo no diagnóstico da endometriose é a vídeolaparoscopia, que possibilita a visualização das lesões sugestivas da doença e a obtenção

de uma amostra tecidual que confirme, pela análise histológica, esta suspeita (Carvalho, 2012). A necessidade de métodos diagnósticos invasivos como laparoscopia e biópsia é uma limitação para o diagnóstico correto e oportuno da endometriose. Marcadores bioquímicos precisos e confiáveis trariam um grande avanço nessa área, mas na prática o marcador ideal ainda não existe (FEBRASGO, 2010).

Em diferentes centros envolvidos com o estudo da endometriose, várias substâncias foram testadas para identificar algum método de avaliação laboratorial que permitisse o diagnóstico e o seguimento do tratamento das pacientes sem ter de recorrer à videolaparoscopia. É possível citar como exemplos: anticorpos antiendométrio, propeptídeo protocollágeno tipo III, proteína C reativa, anticorpos anticardiolipina, proteína sérica amilóide A, CA19-9, CA15-3, antígeno carcinoembrionário, alfa-feto-proteína e beta-2-microglobulina, entre outras (Abrão et al., 1997).

Num estudo realizado na Faculdade de Medicina da USP, parte dos objetivos foi saber se as concentrações do marcador tumoral CA125 e da proteína C reativa (PCR) podem ser utilizadas como fator prognóstico de endometriose em mulheres em idade reprodutiva (Abrão et al., 1997). Quinze mulheres de um grupo controle e 35 com endometriose foram avaliadas quanto às concentrações de CA125 e PCR durante a menstruação e 10 dias após seu término, além de outras variáveis.

As concentrações observadas de CA125 variaram de valores entre 0 e 0,5 UI/mL e 893,0 UI/mL durante a menstruação e 524,9 UI/mL 10 dias após a menstruação. Para o PCR, foram observados valores entre 0 e 0,5

$\mu\text{g/mL}$ e $32,0 \mu\text{g/mL}$ durante a menstruação e $9,6 \mu\text{g/mL}$ em 10 dias após a menstruação. Parte dos dados do estudo estão apresentados no Apêndice A. Tanto os valores de CA125 como os de PCR foram mensurados até o limite mínimo de 0,5 unidades, pois o método utilizado para mensuração não possui precisão suficiente para quantificar as concentrações abaixo desse valor. Valores abaixo deste limite são ditos censurados.

É importante diferenciar a censura do truncamento, que limita os dados mais severamente, excluindo observações com base nas características da variável em estudo. Em uma amostra truncada, todos os casos em que a variável apresenta valor menor que 100, por exemplo, serão excluídos. A amostra truncada tem seu tamanho diminuído e a amostra censurada, não (Long, 1997).

As proporções de censura são apresentadas na Tabela 1.1.

Tabela 1.1: Distribuição de censuras para CA125 e PCR

	Grupo			
	Controle		Doente	
	censuras	%	censuras	%
CA 125 durante	2	13,3	0	0,0
CA 125 após 10 dias	0	0,0	1	2,9
PCR durante	7	46,7	5	14,3
PCR após 10 dias	7	46,7	8	22,9
Total de observações	15		35	

Na ausência de valores censurados pelo limite de detecção, modelos de regressão logística poderiam ser utilizados para investigar a relação entre os marcadores e a presença de endometriose. Em particular, esses modelos

podem ser escritos como

$$\ln\left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right] = \boldsymbol{\beta}^T \mathbf{x} \quad (1.1)$$

em que $\mathbf{x} = (\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2)^T$ com \mathbf{x}_1 o vetor representando a concentração de CA125 e \mathbf{x}_2 o vetor representando a concentração de PCR, mensurados durante ou após 10 dias da menstruação, e $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \beta_2)$ é o vetor de parâmetros. O termo $\pi(\mathbf{x})$ representa a probabilidade de uma mulher com concentração x_1 de CA125 e x_2 de PCR apresentar endometriose. Além disso, $\exp(\beta_1)$ representa a alteração esperada na chance de ser classificada como endometriose associada ao aumento de uma UI/mL na concentração de CA125, mantida constante a concentração de PCR, e $\exp(\beta_2)$ representa a alteração esperada na chance de endometriose associada ao aumento de um $\mu\text{g/mL}$ na concentração de PCR, mantida constante a concentração de CA125. Alternativamente, poderiam ser incluídos os valores de CA125 e PCR mensurados antes e após 10 dias em um mesmo modelo, considerando então a possível entre as medidas realizadas nos dois momentos e incluindo o momento das mensurações como covariável.

A presença de valores censurados pelo limite de detecção nas variáveis explicativas não permite o ajuste do modelo (1.1) sem alguma modificação para acomodá-los. É necessária a substituição dos valores censurados ou a utilização de técnicas de ajuste que contemplem a presença das censuras.

Num contexto mais geral, Gilliom & Helsel (1986) utilizaram simulação para comparar oito metodologias para a estimação de medidas resumo em amostras de tamanho 10, 25 e 50 censuradas por limite de detecção à

esquerda. Os métodos considerados incluíram:

- i) substituição dos valores censurados por 0;
- ii) substituição dos valores censurados pelo limite de detecção (LD);
- iii) substituição por valores obtidos sob uma distribuição uniforme entre 0 e LD ;
- iv) substituição por valores extrapolados a partir de modelos de regressão linear aplicados aos dados originais e transformados logaritmicamente (ver detalhes na seção 2.1);
- v) estimação por máxima verossimilhança a partir dos dados não censurados, assumindo distribuições normal, lognormal e delta para a variável censurada (Aitchison, 1955).

Gilliom & Helsel (1986) concluem que o método mais robusto para a estimação da média, do desvio padrão e dos quartis foi a substituição por valores extrapolados a partir de modelos de regressão linear aplicados aos dados observados transformados logaritmicamente.

Beal (2001) considera sete métodos para análise de dados com censuras no contexto de farmacocinética, em que as concentrações plasmáticas mensuradas em vários momentos após a utilização da droga são utilizadas para a estimação de dois parâmetros: o volume de distribuição aparente, que é o volume necessário para a droga estar homogeneamente distribuída entre o sangue e os tecidos, e a depuração, que é a taxa de eliminação da droga do organismo e parâmetro essencial para definição da posologia. Os

métodos considerados e comparados por meio de um estudo de simulação são:

- M1) descarte das observações em que há censura e uso dos valores restantes como uma amostra completa e sem restrições;
- M2) descarte dos valores abaixo do LD e uso dos valores restantes como uma amostra truncada, sendo a verossimilhança dos valores restantes calculada condicionalmente aos valores serem maiores que o LD ;
- M3) estimação utilizando toda a amostra e considerando que os valores abaixo do LD são censurados, sendo a verossimilhança da parte censurada da amostra calculada condicionalmente aos valores serem menores que o LD ;
- M4) estimação como em M3, porém assumindo que os valores censurados são todos positivos;
- M5) imputação dos valores abaixo do LD por $LD/2$ e estimação considerando a amostra como real;
- M6) substituição da primeira observação por $LD/2$ e descarte das demais observações do indivíduo em casos de amostras repetidas em um mesmo indivíduo;
- M7) Imputação dos valores censurados por 0 e estimação considerando a amostra como completa.

A partir dos resultados das simulações e medidas de erro médio percentual e raiz quadrada do erro quadrático médio, [Beal \(2001\)](#) conclui que

o método M3 apresenta vantagens em relação aos demais. No entanto, o ganho na utilização deste método em relação ao método M2 diminui à medida que a proporção de censuras diminui.

Com o objetivo principal de comparar o desempenho dos métodos M2 e M3, considerados por [Beal \(2001\)](#), [Senn et al. \(2012\)](#) aplicam os dois métodos para a estimação da média e do erro padrão da concentração de um determinado fármaco, considerando uma medida de concentração por indivíduo e a presença de censura devida a limite de detecção.

A partir da aplicação de métodos a um conjunto de dados reais de concentração plasmática de Voriconazol, [Senn et al. \(2012\)](#) verificaram que o método M2 produz erros padrão muito maiores para as estimativas e que as diferenças nas estimativas são grandes entre os dois métodos. Os autores conseguem quantificar de forma teórica a perda ao utilizar o M2 em comparação com o método M3, assim como a perda ao utilizar o método M3 em comparação com a utilização da informação total, sem censura, e observam que a perda de informação ao utilizar o método M3 em comparação com a informação total é menor que a perda esperada ao utilizar o método M2 em comparação com o método M3. Os autores concluem que o método M3 é superior ao método M2 na estimação.

[Schisterman et al. \(2006\)](#) avaliaram as limitações de um modelo de regressão logística com censura à esquerda numa única variável explicativa contínua (X), por meio de um estudo de simulação. Foram consideradas duas distribuições para a variável explicativa: uma distribuição normal bimodal ([Prasad, 1955](#)), com parametrização dada por $N(5, 2^2)\theta_i +$

$N(12, 2^2)(1 - \theta_i)$, com θ_i variável aleatória com distribuição Bernoulli em que $P(\theta_i = 1) = \frac{1}{2}$ e uma distribuição gama, comumente assumida para marcadores biológicos devido à assimetria, com parâmetros de escala e de formato iguais a 4. Para a distribuição gama da variável explicativa foram considerados os valores 0,3 e 0,5 para o parâmetro β do modelo (1.1) com apenas uma variável explicativa (correspondentes a razões de chances 1,35 e 4,48) e para a distribuição bimodal foram considerados valores 0,3, 0,5 e 0,7 para o parâmetro β do mesmo modelo (razões de chances 1,35, 1,65 e 2,01).

Para cada distribuição e valor do parâmetro foram simulados 10000 conjuntos de dados de tamanho 300 e cada conjunto teve 25%, 50% e 75% dos menores valores censurados, de modo que o efeito da censura pudesse ser avaliado em comparação com situações em que não há censura. Os valores censurados foram então substituídos por $E(X|X < LD)$, 0, LD e $LD/2$, com LD representando o limite de detecção.

Para a distribuição gama, a substituição por $E(X|X < LD)$ produziu menores vieses e variâncias em comparação com aqueles obtidos com a substituição por 0 para todas as proporções de censura e para os dois valores do parâmetro β considerados. Para a distribuição normal bimodal, os melhores resultados se alternam entre substituição por 0 e por $E(X|X < LD)$, quaisquer que fossem as proporções de censura e valores dos parâmetros considerados. Os autores observaram também que tanto o viés quanto a variância aumentam à medida que se aumenta a proporção de observações censuradas, tanto para a distribuição gama quanto para a

distribuição normal bimodal. Além disso, mantendo a mesma proporção de censura e o mesmo valor do parâmetro, na maioria dos casos, o viés e a variância foram maiores sob a distribuição gama. As alternativas de substituição por LD ou por $LD/2$ produziram viés e variância maiores que as demais substituições, independentemente das proporções de censura ou da distribuição considerada.

Em um estudo de simulação, [Cole et al. \(2009\)](#) compararam o desempenho de cinco métodos para a estimação de razões de chances a partir de modelos de regressão logística em que a única variável explicativa é contínua e censurada à esquerda. Esses autores consideraram os seguintes métodos:

- i) exclusão dos valores censurados;
- ii) substituição dos valores censurados por LD ;
- iii) substituição dos valores censurados por $LD/\sqrt{2}$;
- iv) substituição dos valores censurados por $LD/2$;
- v) estimação por máxima verossimilhança.

Nesse trabalho, foram considerados dois tamanhos de amostra (500 e 2000) e três proporções de censura (25%, 50% e 75%), com a variável explicativa de distribuição lognormal com mediana 1 e primeiro e terceiro quartis iguais a 0,5 e 2. Os resultados mostraram que a estimação por máxima verossimilhança forneceu estimativas não enviesadas e intervalos de confiança com coberturas observadas de 95% e 96%, próximas à cobertura

nominal de 95%. As substituições por LD , $LD/\sqrt{2}$ ou $LD/2$, consideradas pelo autor por serem comuns nas aplicações, forneceram estimativas de razões de chances em que o viés aumenta com a proporção de censura e com coberturas observadas entre 54% e 96% para uma cobertura nominal de 95%.

Rao (1996) cita que a utilização de amostras com dados imputados como se fossem amostras completas é prática comum. Este procedimento pode levar a variâncias subestimadas e, conseqüentemente, inferências errôneas, já que a variabilidade adicional devida à imputação dos valores censurados não é levada em consideração. Rao (1996) aborda a utilização de imputação múltipla e do método *jackknife* para estimação da variância e indica que o método *jackknife* deve ser preferido, por apresentar menor custo computacional que o método de imputação múltipla.

Considerando o ajuste de modelos de regressão logística com variáveis explicativas censuradas à esquerda, os estudos mais informativos são o de Schisterman et al. (2006) e o de Cole et al. (2009). No entanto, os métodos comparados nos dois estudos não coincidem, sendo que o método de melhor desempenho em um estudo não foi abordado no outro, assim como o contrário. Além disso, os estudos consideram o desempenho das técnicas aplicadas em amostras de tamanhos grandes, 300 em Schisterman et al. (2006) e 500 e 2000 em Cole et al. (2009), pouco comuns em estudos biomédicos. Para a melhor avaliação do desempenho das diferentes técnicas de substituição ou de ajuste contemplando censuras, investigamos, por meio de um estudo de simulação, o comportamento de tais técnicas apli-

casas a amostras pequenas ou moderadas, da ordem de 50, 100 ou 200 pacientes. Nesse contexto, consideramos um modelo de regressão logística com uma única variável explicativa contínua, com distribuição lognormal e censurada à esquerda. A distribuição lognormal foi escolhida devido à assimetria, comum a marcadores biológicos, e aderência aos dados do exemplo considerado neste trabalho.

No capítulo 2 descrevemos os métodos considerados neste trabalho. No capítulo 3 apresentamos os resultados de um estudo de simulação realizado com o intuito de comparar esses métodos. No capítulo 4 apresentamos a aplicação das técnicas consideradas a parte de um conjunto de dados reais do estudo de [Abrão et al. \(1997\)](#), dispostos no Apêndice A, e no capítulo 5 apresentamos as conclusões obtidas, além de sugestões para futuras pesquisas.

Capítulo 2

Métodos para tratamento de censuras

Consideramos o modelo de regressão logística (1.1) com uma única variável explicativa contínua X censurada à esquerda por um limite de detecção (LD), ou seja, em que $Y_i|X_i$ segue uma distribuição Bernouli com probabilidade de sucesso $\pi_i = [1 + \exp(-\beta_0 - \beta_1 x_i)]^{-1}$. Supondo que a amostra observada seja composta por n observações, consideramos que $m < n$ são censuradas (menores que LD).

2.1 Técnicas de substituição dos valores censurados

Nas técnicas de substituição, os valores censurados são imputados na amostra, de modo que o modelo de regressão logística é ajustado considerando os valores observados e imputados como uma amostra completa, com o erro padrão das estimativas calculado pelo método *jackknife*, contemplando a variabilidade adicional devida aos valores imputados (Rao, 1996).

Dentre as técnicas de substituição dos valores censurados, consideramos:

- i) imputação por 0;

- ii) imputação pelo limite de detecção (LD);
- iii) imputação por metade do limite de detecção (LD/2);
- iv) imputação por $E(X|X < LD)$;
- v) imputação por valores aleatoriamente escolhidos sob a distribuição que melhor se aproxima da distribuição da variável explicativa, condicionada ao intervalo $[0, LD]$;
- vi) imputação por valores aleatoriamente escolhidos sob uma distribuição uniforme ($U[0, LD]$);
- vii) imputação por valores estimados a partir de modelos de regressão;

Para a opção iv, estudada por [Schisterman et al. \(2006\)](#), selecionamos a distribuição e os parâmetros que melhor se adequam aos valores não censurados de X por meio de medidas descritivas, histogramas, gráficos de comparação de quantis e também por intermédio de testes de aderência a algumas distribuições ([Ricci, 2005](#)). Definida a distribuição que melhor se adequa aos dados observados, calculamos o valor esperado de X condicionado ao limite de detecção considerado, $E(X|X < LD)$, e utilizamos esse valor na substituição de todos os valores censurados.

Na imputação por valores estimados a partir de modelos de regressão (vii), proposta por [Gilliom & Helsel \(1986\)](#), os dados observados de X (originais ou transformados logaritmicamente) são utilizados para o ajuste de modelos de regressão linear pelo método de mínimos quadrados, que servem para extrapolar os valores da amostra abaixo do LD . Neste método,

consideramos uma distribuição normal para as n observações, censuradas ou não. Posteriormente, ordenamos as $n - m$ observações não censuradas da amostra e calculamos os escores normais associados aos postos destas observações, a saber:

$$z_i = \phi^{-1}(r_i/n + 1) \quad (2.1)$$

em que r_i representa o posto da i -ésima observação, ϕ^{-1} , a inversa da função acumulada da distribuição normal padrão e $i = m + 1, m + 2, \dots, n$.

Um modelo de regressão linear é ajustado pelo método de mínimos quadrados ordinários tendo como variável resposta os dados observados e como variável explicativa os escores z_i calculados. Após o ajuste do modelo, tendo atribuído aleatoriamente os postos z_1, z_2, \dots, z_m aos dados censurados, utilizamos o modelo para estimar os m valores censurados, sendo qualquer valor negativo substituído por zero. Os m valores estimados são utilizados para a substituição das observações censuradas.

2.2 Utilização de estimadores de máxima verossimilhança

Para avaliar a associação entre a variável explicativa X (censurada) e a variável resposta dicotômica Y adotamos o método proposto por [Cole et al. \(2009\)](#) e supomos que a variável X segue uma distribuição lognormal com média μ e variância σ^2 desconhecidas.

Em uma amostra de n unidades, em que X foi censurada pelo limite de

detecção em $m < n$ unidades, a função de verossimilhança para a amostra ordenada com m unidades censuradas colocadas nas primeiras posições é

$$\begin{aligned} L(\beta_0, \beta_1, \mu, \sigma; \mathbf{x}, \mathbf{y}) &= \left[\prod_{i=m+1}^n [f(y|x_i)g(x_i)]^{(1-\delta_i)} \right] \times \left[\prod_{i=1}^m \left[\int_0^{LD} f(y|x)g(x)dx \right]^{(\delta_i)} \right] \\ &= \left[\prod_{i=m+1}^n [\pi_i^{y_i}(1-\pi_i)^{(1-y_i)}g(x_i)]^{(1-\delta_i)} \right] \times \left[\prod_{i=1}^m \left[\int_0^{LD} \pi_i^{y_i}(1-\pi_i)^{(1-y_i)}g(x)dx \right]^{(\delta_i)} \right] \end{aligned} \quad (2.2)$$

com $\pi_i = \pi_i(x) = [1 + \exp(-\beta_0 - \beta_1 x_i)]^{-1}$. A função log-verossimilhança correspondente é

$$\begin{aligned} \mathbf{l}(\beta_0, \beta_1, \mu, \sigma; \mathbf{x}, \mathbf{y}) &= \\ \sum_{i=1}^n \log &\left[[\pi_i^{y_i}(1-\pi_i)^{(1-y_i)}g(x_i)](1-\delta_i) + \left[\int_0^{LD} \pi_i^{y_i}(1-\pi_i)^{(1-y_i)}g(x)dx \right] \delta_i \right] \end{aligned} \quad (2.3)$$

com

$$\delta_i = \begin{cases} 1 & \text{se } x_i \leq LD \\ 0 & \text{se } x_i > LD \end{cases}$$

e

$$g(x) = \frac{1}{x\sigma\sqrt{2\pi}} \times \exp \left\{ -\frac{1}{2} \left[\frac{\ln(X) - \mu}{\sigma} \right]^2 \right\}, x > 0.$$

(2.4)

A integral indicada em (2.3) não possui solução explícita e, com isso, os estimadores de máxima verossimilhança são obtidos numericamente, com o uso de um método de otimização como o considerado em [Nelder & Mead \(1965\)](#). Neste trabalho, esse procedimento foi concretizado por meio da função `optim()` da programoteca `stats` do pacote R.

O código disponibilizado por [Cole et al. \(2009\)](#) para o ajuste por máxima verossimilhança foi modificado para este trabalho, de modo a contemplar a distribuição lognormal para os dados. Detalhes são apresentados no Apêndice B.

Capítulo 3

Simulação

Realizamos um estudo de simulação (Burton et al., 2006) para avaliar e comparar o desempenho dos seguintes métodos para lidar com valores censurados no ajuste de modelos de regressão logística:

M0) substituição por 0;

MLD) substituição por LD ;

MLD2) substituição por $LD/2$;

ME) substituição por $E(X|X < LD)$;

MU) substituição por valores escolhidos aleatoriamente de acordo com distribuição ($U[0, LD]$);

MN) substituição por valores escolhidos aleatoriamente de acordo com distribuição lognormal truncada no intervalo $[0; LD]$;

MR) substituição a partir de modelo de regressão aplicado aos postos da variável explicativa (Gilliom & Helsel, 1986);

MMV) Utilização de máxima verossimilhança.

Para a escolha dos valores de coeficientes utilizados na simulação, ajustamos um modelo de regressão logística simples considerando os dados de proteína C reativa após 10 dias da menstruação e de classificação quanto à endometriose apresentados no Apêndice A, substituindo os valores censurados de concentração de proteína C reativa pelo limite de detecção, 0,5. Esta variável foi escolhida por apresentar a maior proporção de observações abaixo do limite de detecção. A estimativa do coeficiente associado à proteína C reativa após 10 dias da menstruação é 0,75. Para o intercepto a estimativa é -1,1.

Os valores da variável explicativa na simulação foram obtidos a partir de uma distribuição lognormal, com média e desvio padrão em escala logarítmica iguais a 0,60 e 1,00, respectivamente próximos aos valores observados para a proteína C reativa após 10 dias da menstruação em escala logarítmica (0,60 e 0,96). Em resumo, os cenários considerados na simulação foram:

- 1) tamanhos de amostra (n) iguais a 50, 100 e 200;
- 2) proporções de censura iguais a 5%, 10%, 25% e 50%;
- 3) intercepto do modelo (β_0) igual a -1,00;
- 4) coeficientes associados à variável explicativa (β_1) iguais a 0,25, 0,75 e 1,25;

Um dos valores de β_1 na simulação foi próximo ao observado com a proteína C reativa (0,75) e os demais foram obtidos com a adição ou subtração de 0,50 a este valor. Para o intercepto o valor foi fixado em -1,00.

Para cada combinação de n e β_1 , foram gerados 1000 conjuntos de dados, de acordo com os seguintes passos:

- geramos valores de x_i , supondo que $X_i \sim LN(\mu_x; \sigma_x^2), i = 1, \dots, n$;
- calculamos os valores de $\pi_i = [1 + \exp(-\beta_0 - \beta_1 x_i)]^{-1}, i = 1, \dots, n$;
- geramos valores de y_i , supondo que $Y_i \sim Bernouli(\pi_i), i = 1, \dots, n$;
- a partir de cada conjunto de dados obtidos para X , geramos um novo conjunto de dados, eliminando 5%, 10%, 25% ou 50% dos menores valores do conjunto;
- estimamos os valores de β_0 , de β_1 e obtivemos os erros padrão das estimativas a partir do conjunto de dados completo, de forma usual;
- estimamos os valores de β_0 , de β_1 e obtivemos os erros padrão das estimativas a partir do conjunto de dados censurado considerando as oito metodologias citadas no início do capítulo, utilizando o método *jackknife* na estimação dos erros padrão nos casos de substituição;

Para avaliar o desempenho dos métodos descritos na seção 2.1, calculamos o viés relativo das estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$ obtidas a partir dos conjuntos de dados censurados e também do conjunto de dados completo. Todas medidas

foram acompanhadas de seus erros padrão. O viés relativo foi calculado como $\sum_{i=1}^{1000} [1000^{-1} \times (\hat{\beta}_i - \beta/\beta)]$ e apresentado em porcentagem, em que $\hat{\beta}_i$ é o valor estimado para o parâmetro β a partir de uma amostra i de tamanho n . Para avaliar se algum dos métodos apresenta estimativas mais próximas dos valores reais dos parâmetros e se os erros padrão estimados são próximos aos obtidos com o ajuste considerando o conjunto de dados completo, construímos gráficos *box plot* com as estimativas obtidas a partir das 1000 iterações em todas os 36 cenários da simulação.

Considerando $\beta_1 = 0,25$ e a estimação de $\hat{\beta}_0$, observamos valores próximos de viés relativo para todos os métodos quando a proporção de censura é de 5% ou 10%, para qualquer dos tamanhos de amostra considerados. As diferenças são evidentes para 25% ou 50% de censura, em que observamos variações de viés relativo de até -2,6% a 38,9% para um mesmo cenário. O método de substituição pelo limite de detecção (MLD) apresenta os maiores vieses relativos e os métodos M0, ME e MR se destacam por apresentarem os menores vieses (Tabela C.1). Esse comportamento foi similar nos casos de estimação de $\hat{\beta}_1$, também considerando $\beta_1 = 0,25$, sendo que os métodos M0 e MR, em geral, apresentam os menores vieses (Tabela C.2).

As estimativas dos parâmetros para 5% ou 10% de censura, apresentadas nas figuras C.1, C.2, C.5, C.6, C.9 e C.10, são similares entre os métodos e, para as estimativas de erros padrão, o método MMV se destaca pela

semelhança às estimativas obtidas com o conjunto de dados completo.

Considerando 25% de censura e amostra de tamanho 50, ainda para $\beta_1 = 0,25$ (Figura C.3), observamos que à exceção do método MLD, as estimativas obtidas pelos diferentes métodos são próximas ao verdadeiro valor do parâmetro, sendo que o método MLD se destaca dos demais com relação às estimativas dos parâmetros e também de seus erros padrão. Nos casos de 25% de censura e amostras de tamanho 100 ou 200 (figuras C.7 e C.11), os métodos M0, MLD e MR são os que mais se distanciam do resultado obtido com o conjunto de dados completo para estimativas dos parâmetros e de seus erros padrão.

As maiores diferenças entre os métodos são observadas com 50% de censura (figuras C.4, C.8 e C.12), sendo os métodos MLD, M0 e MR os métodos que mais diferem do valor do parâmetro e das estimativas obtidas com o conjunto de dados completo, para os três tamanhos de amostra considerados. Além destes, o método ME tem desempenho inferior aos demais métodos para amostras de tamanho 100 e 200.

Considerando $\beta_1 = 0,75$, o método MLD apresenta os maiores vieses para todos os tamanhos de amostra, proporções de censura e parâmetros considerados. Para a estimação com 25% ou 50% de censura os métodos M0, MU, ME e MR apresentam maiores vieses na estimação de β_0 , assim como os métodos M0 e MR na estimação de β_1 . Para amostras de tamanho 200 e 25% ou 50% de censura o método MLD2 apresenta vieses

próximos de zero tanto para os dois parâmetros (tabelas C.3 e C.4).

As estimativas dos parâmetros considerando $\beta_1 = 0,75$ e 5% ou 10% de censura, dispostas nas figuras C.13, C.14, C.17, C.18, C.21 e C.22, são similares entre os métodos, como observamos com $\beta_1 = 0,25$. O método MMV é o mais próximo ao ajuste considerando o conjunto de dados completo quanto às estimativas para os erros padrão dos parâmetros.

Para 25% de censura (figuras C.15, C.19 e C.23), os métodos com estimativas mais próximas às obtidas com ajuste com o conjunto de dados completo são MLD2, MU, MN e MMV, e o método que mais difere dos demais quanto às estimativas dos parâmetros e de seus erros padrão é o MLD.

As maiores diferenças entre os métodos com $\beta_1 = 0,75$ ocorrem com 50% de censura (figuras C.16, C.20 e C.24), sendo os métodos M0, MLD e MR os mais distintos para todos os tamanhos de amostra. Nos casos com amostra de tamanho 50 ou 100 observamos valores discrepantes nas estimativas do método MLD e, para possibilitar a comparação dos métodos por meio de gráficos, limitamos os dados ao intervalo $[-20,20]$.

Considerando $\beta_1 = 1,25$, o método MLD apresenta maiores vieses relativos quando a proporção de censuras é 50%, para os três tamanhos de amostra e os dois parâmetros considerados. O viés calculado com as estimativas do método MLD chega a 1328,7%, com amostra de tamanho 50. Chama a atenção que nos casos de amostras de tamanho 100 e 200 com

5% de censura o método a apresentar menor viés é o MLD.

Na estimação de $\hat{\beta}_0$ e para censuras de até 25%, os vieses relativos são maiores para os métodos M0, ME e MR, à exceção do caso de 5% de censura e amostra de tamanho 50, em que o método MLD apresentou maior viés (Tabela C.5).

Para $\hat{\beta}_1$, observamos que os métodos M0, MLD, ME e MR apresentam os maiores vieses relativos, exceto nos casos com amostra de tamanho 200 e 5% ou 10% de censura, em que o menor viés relativo é o do método MLD. Para amostras de tamanho 50, os métodos com menores vieses se alternam entre M0, MU, MN e MMV. Além disso, nas amostras de tamanho 100 ou 200 o método MMV apresenta vieses menores, ou muito próximos dos menores, em 9 dos 10 cenários considerados (Tabela C.6).

As estimativas dos parâmetros com $\beta_1 = 1, 25$ e 5% de censura, apresentadas nas figuras C.25, C.29 e C.33, são similares entre os métodos e o método MMV é o mais similar ao ajuste com conjunto de dados completo quanto às estimativas para os erros padrão dos parâmetros, como observamos com os outros valores de β_1 considerados.

Para 10% de censura (figuras C.26, C.30 e C.34) e 25% de censura com amostras de tamanho 50 (Figura C.27), os métodos com estimativas mais próximas às obtidas com conjunto de dados completo são MLD, MU, MN e MMV e os métodos MLD e MN são os que mais se destacam dos demais quanto às estimativas dos erros padrão.

Para 25% de censura e amostras de tamanho 100 e 200 (figuras C.31 e C.35), o método com estimativas mais próximas aos valores dos parâmetros é o MMV, seguido de MN e MLD2, e o método que mais se destaca dos demais quanto às estimativas dos parâmetros e de seus erros padrão é o MLD.

Com 50% de censura e $\beta_1 = 1, 25$ (figuras C.28, C.32 e C.36) o método com estimativas mais próximas às obtidas com o conjunto de dados completo é o MMV. Nos casos com amostra de tamanho 50 ou 100 observamos valores discrepantes nas estimativas do método MLD e, para possibilitar a comparação dos métodos por meio de gráficos, limitamos os dados ao intervalo $[-20, 20]$.

O método MMV apresenta vieses relativos maiores para amostras de tamanho 50 em relação às amostras de tamanho 100 ou 200, para os três valores de β_1 considerados. Para os demais métodos o viés relativo aumenta à medida que a proporção de censuras aumenta nas amostras, sendo os métodos M0, MLD, ME e MR os mais afetados pelo aumento da proporção de censuras. Na maioria dos cenários os vieses dos métodos M0, ME e MR aumentam com o aumento do tamanho de amostra, enquanto o comportamento inverso é observado para os demais métodos.

Em geral, o método de máxima verossimilhança apresenta os menores vieses ou valores próximos aos menores vieses observados. Este método apresenta erros padrão estimados mais próximos aos obtidos com o con-

junto de dados completo.

Buscando entender se a adição de um termo aleatório ao valor substituído no método de substituição pelo valor esperado condicionado ao limite de detecção (ME) e também a utilização de distribuição lognormal no método de substituição a partir de modelo de regressão aplicado aos postos da variável explicativa (MR) poderiam melhorar o desempenho destes métodos em comparação aos resultados obtidos com o ajuste com banco de dados completo e com o método de estimação por máxima verossimilhança (MMV), realizamos um novo estudo de simulação considerando três métodos de ajuste do modelo 1.1, com variável explicativa única e censurada à esquerda, como anteriormente. Os métodos considerados foram:

ME) substituição por $E(X|X < LD)$ adicionado a um termo aleatório de distribuição $Normal(0, (0, 1)^2)$;

MR) substituição a partir de modelo de regressão aplicado aos postos da variável explicativa supondo distribuição lognormal

MMV) Utilização de máxima verossimilhança.

Os cenários para esse estudo foram selecionados dentre os de maior discrepância entre os métodos no estudo anterior, a saber:

- 1) tamanhos de amostra (n) iguais a 50, 100 e 200;
- 2) proporções de censura iguais a 25% e 50%;

- 3) intercepto do modelo (β_0) igual a -1,00;
- 4) coeficiente associado à variável explicativa (β_1) igual a 1,25;

O número de iterações considerados foi 500, e observamos que, para os casos de 25% de censura os métodos MMV e MR apresentam vieses e coeficientes estimados similares aos do ajuste com conjunto de dados completo, enquanto o método ME se destaca desses (Tabela C.7 e figuras C.37, C.39 e C.41). As estimativas dos erros padrão dos parâmetros obtidas com o método MMV são mais próximas às obtidas com o conjunto de dados completo, em comparação com o método MR.

Para 50% de censura (Tabela C.7 e figuras C.38, C.40 e C.42) as estimativas dos parâmetros e vieses obtidos com o método MR se destacam dos obtidos com MV, sendo esse o método com resultados mais próximos aos obtidos com o conjunto de dados completo, no cenário com 50% de censura.

Capítulo 4

Aplicação

Neste capítulo aplicaremos os métodos estudados no capítulo anterior aos dados do estudo de [Abrão et al. \(1997\)](#), que motivou este trabalho. Neste estudo foram observadas 50 mulheres entre 1993 e 1995, sendo 15 do grupo controle e 35 do grupo com endometriose. Os dados do estudo são apresentados no Apêndice A e nesta aplicação consideraremos um modelo de regressão logística (1.1) com única variável explicativa, a concentração sérica de proteína C reativa mensurada 10 dias após a menstruação que apresenta 15% de suas observações censuradas pelo limite de detecção de 0,5.

Observamos maiores valores de concentração 10 dias após a menstruação no grupo com endometriose (Figura C.40), que indica associação entre o marcador e a condição estudada.

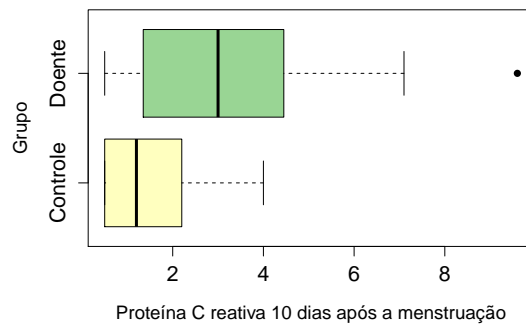


Figura 4.1: Concentração sérica de PCR 10 dias após a menstruação

O estudo de simulação apresentado no capítulo anterior considera uma distribuição lognormal para a variável explicativa. Se uma variável possui distribuição lognormal, a transformação logarítmica desta apresentará distribuição normal. Avaliamos então a distribuição da concentração de PCR após a transformação logarítmica nos dois grupos de estudo e concluimos pela adequação à distribuição normal (Figuras C.41 e C.42). Para esta avaliação os valores censurados não foram considerados.

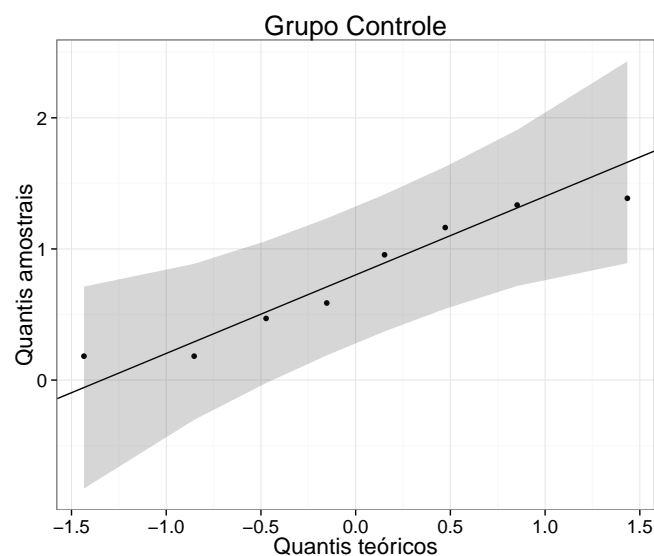


Figura 4.2: Gráfico QQ para o logaritmo da concentração sérica de PCR 10 dias após a menstruação no grupo Controle

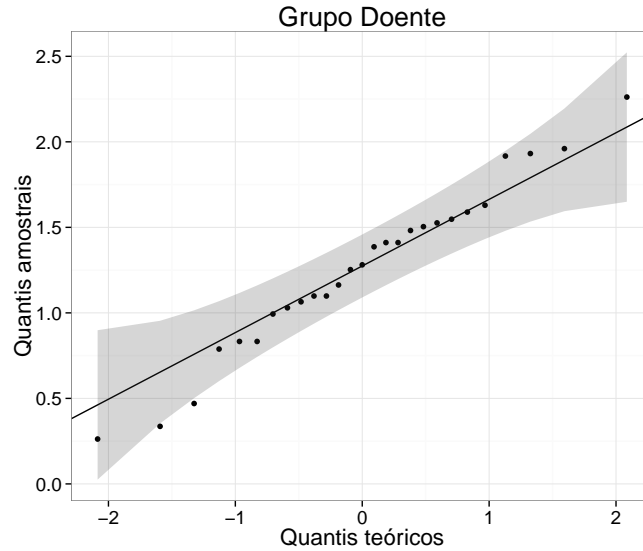


Figura 4.3: Gráfico QQ para o logaritmo da concentração sérica de PCR 10 dias após a menstruação no grupo Doente

As estimativas de máxima verossimilhança foram obtidas por meio do código apresentado no Apêndice B e apresentam valor $-0,22$ para o intercepto (erro padrão: $0,50$) e $0,50$ para o coeficiente associado à concentração de PCR (erro padrão: $0,21$). Aplicamos também os métodos de substituição estudados no capítulo anterior, com estimativas para os parâmetros obtidas por ajuste usual de um modelo de regressão logística e estimativas para os erros padrão obtidas pelo método *jackknife*. Além desses métodos, consideramos nesta aplicação o ajuste com a exclusão das observações com valor de PCR 10 dias após a menstruação censurado (M^*), sendo a amostra restante de tamanho 35.

As estimativas são apresentadas na Tabela 4.1, em que observamos que a razão de chances de classificação com endometriose associada ao acréscimo de uma unidade no valor de PCR em 10 dias após a menstruação é

estimada pelo método de máxima verossimilhança em 1,64, com intervalo de confiança de 95% de 1,09 a 2,47. Apesar de o método de máxima verossimilhança apresentar melhor desempenho no estudo de simulação, o ajuste pelos métodos de substituição considerados produz estimativas similares para β_0 e β_1 e para seus erros padrão, de modo que não há alteração na inferência com a escolha de um ou outro método. Já o ajuste excluindo as observações com valor de PCR censurado apresenta estimativas distantes das obtidas com os demais métodos, além de erros padrão maiores.

Tabela 4.1: Estimativas obtidas para a relação entre PCR (10 dias após a menstruação) e classificação de endometriose.

Parâmetro	Método	Estimativa	Erro padrão	Razão de chances	IC 95%
β_0	M0	-0,09	0,48		
	MLD	-0,33	0,54		
	MLD2	-0,20	0,50		
	ME	-0,11	0,48		
	MU	-0,17	0,50		
	MN	-0,29	0,52		
	MR	-0,11	0,49		
	MMV	-0,22	0,50		
	M*	-1,06	1,11		
β_1	M0	0,46	0,18	1,58	(1,10 a 2,25)
	MLD	0,53	0,22	1,70	(1,10 a 2,63)
	MLD2	0,49	0,20	1,63	(1,10 a 2,43)
	ME	0,46	0,19	1,58	(1,10 a 2,28)
	MU	0,48	0,20	1,61	(1,09 a 2,37)
	MN	0,52	0,22	1,68	(1,10 a 2,56)
	MR	0,46	0,19	1,58	(1,09 a 2,28)
	MMV	0,50	0,21	1,64	(1,09 a 2,47)
	M*	0,74	0,38	2,11	(1,00 a 4,45)

Nota: IC: Intervalo de confiança; M*: ajuste excluindo as observações censuradas.

Capítulo 5

Conclusões

Os resultados do estudo de simulação realizado mostraram que a alternativa de substituição dos dados censurados utilizando o valor do limite de detecção não apresenta resultados satisfatórios em modelos de regressão logística aplicados a amostras de tamanho pequeno ou moderado, entre 50 e 200.

Para os casos com 5% ou 10% de censura os oito métodos considerados apresentaram vieses relativos similares, sendo as diferenças entre os métodos evidentes para maiores proporções de censura, sendo os métodos de substituição por 0 (M0), pelo limite de detecção (MLD), pelo valor esperado (ME) e por valores obtidos a partir do ajuste de um modelo de regressão (MR) os mais afetados pelo aumento da proporção de censuras.

O método de máxima verossimilhança (MMV) apresenta vieses relativos maiores para amostras de tamanho 50 em relação às amostras de tamanho 100 ou 200, para os três valores de β_1 considerados. Na maioria dos cenários os vieses dos métodos M0, ME e MR aumentam com o

aumento do tamanho de amostra, enquanto o comportamento inverso é observado para os demais métodos.

Em geral, o método de máxima verossimilhança apresenta os menores vieses ou valores próximos aos menores vieses observados. Este método apresenta erros padrão estimados mais próximos dos obtidos com o ajuste considerando a amostra completa, sem censuras.

Considerando a avaliação do desempenho dos métodos de substituição pelo valor esperado adicionado a um termo aleatório, observamos que esta alteração ao método não contribui para o seu desempenho. Já a substituição dos valores censurados por valores estimados a partir de um modelo de regressão considerando a distribuição lognormal apresenta resultados satisfatórios nos casos de 25% de censura considerados, sendo a estimação por máxima verossimilhança o método com resultados mais próximos aos obtidos com o conjunto de dados completo considerando 50% de censura.

Observamos vieses relativos menores e estimativas mais próximas às obtidas considerando o conjunto de dados completo na estimação pelo método de máxima verossimilhança para maiores proporções de censura e maiores tamanhos de amostra, em comparação com os demais métodos.

Concluimos que, apesar de seu uso comum em aplicações com dados censurados, o método de substituição pelo valor do limite de detecção deve ser preterido, já que possui os maiores valores de viés na maioria dos cenários considerados neste trabalho.

Apesar de o método de máxima verossimilhança apresentar indícios de melhor desempenho, a aplicação a dados reais considerada mostra que o ajuste pelos métodos de substituição produz estimativas similares aos do método de máxima verossimilhança, sem alteração na inferência com a escolha de um ou outro método. Observamos discrepâncias somente ao considerar a exclusão das observações com concentração de PCR censurada, em que as estimativas para os parâmetros não são similares às demais e os erros padrão são maiores.

Neste trabalho estudamos o desempenho de métodos para ajuste de um modelo de regressão logística com variável explicativa única, censurada à esquerda e com distribuição lognormal, que é um cenário específico e uma limitação do estudo.

Para um melhor entendimento do ajuste de modelos de regressão logística com censura nas variáveis explicativas são necessários outros estudos que avaliem o efeito da censura em uma variável explicativa na presença de outras variáveis explicativas, censuradas e não censuradas. Também é necessário estudar o ajuste considerando outras distribuições para a variável explicativa censurada e também a utilização de métodos não paraméricos, além de outros valores para os parâmetros.

Apêndice A

Dados utilizados como exemplo

Tabela A.1: Concentrações de CA125 e PCR observadas em mulheres sadias e com endometriose (Abrão et al., 1997).

Grupo	Paciente	CA125-Durante	CA125-10 dias	PCR-Durante	PCR-10 dias
Controle	1	15,9	6,1	2,0	3,2
Controle	2	2,9	2,9	<0,5	<0,5
Controle	3	48,6	24,0	3,2	1,6
Controle	4	7,2	7,9	2,2	1,2
Controle	5	24,1	19,8	3,4	1,2
Controle	6	<0,5	0,6	1,2	<0,5
Controle	7	14,0	13,2	3,8	3,8
Controle	8	16,2	11,1	<0,5	<0,5
Controle	9	20,9	11,9	<0,5	2,6
Controle	10	13,8	6,3	2,0	1,8
Controle	11	7,5	6,1	2,4	<0,5
Controle	12	21,9	13,2	<0,5	<0,5
Controle	13	13,5	11,0	<0,5	<0,5
Controle	14	1,6	4,0	<0,5	4,0
Controle	15	<0,5	7,2	<0,5	<0,5
Doente	1	47,5	17,0	4,7	2,3
Doente	2	47,0	45,1	6,6	4,4
Doente	3	85,9	51,3	12,0	6,8
Doente	4	69,4	27,6	4,7	<0,5
Doente	5	30	22,3	<0,5	2,3
Doente	6	35,5	31,0	<0,5	<0,5
Doente	7	50,0	28,9	3,6	<0,5
Doente	8	35,5	25,4	<0,5	<0,5
Doente	9	62,6	27,5	3,5	3,2
Doente	10	80,9	36,9	5,8	1,4
Doente	11	177,2	61,6	4,6	3,6
Doente	12	76,6	24,0	2,9	<0,5
Doente	13	60,2	35,8	27,0	7,1
Doente	14	141,4	94,5	3,5	2,9
Doente	15	27,4	<0,5	1,6	<0,5
Doente	16	73,8	57,1	<0,5	9,6
Doente	17	29,1	7,8	4,6	4,5
Doente	18	61,3	45,2	7,9	4,6
Doente	19	218,7	52,2	4,5	<0,5
Doente	20	57,5	28,0	1,8	1,6
Doente	21	219,0	205,0	1,5	4,9
Doente	22	74,0	63,0	20,0	3,5
Doente	23	286,8	80,2	3,8	3,0
Doente	24	113,6	111,7	3,7	1,3
Doente	25	94,2	95,9	16,0	5,1
Doente	26	115,3	115,1	16,5	4,0
Doente	27	893,0	211	13,0	4,7
Doente	28	335,4	87,7	8,8	2,7
Doente	29	292,0	25,0	<0,5	<0,5
Doente	30	186,5	92,9	26,0	6,9
Doente	31	550,0	524,9	32,0	4,1
Doente	32	68,8	23,1	7,0	2,2
Doente	33	54,8	33,4	4,5	2,8
Doente	34	78,4	42,4	25,0	3,0
Doente	35	363,5	132,2	19,0	4,1

Apêndice B

Código para o ajuste por máxima verossimilhança

Apresentamos o código utilizado para o ajuste por máxima verossimilhança de um modelo de regressão logística com uma única variável explicativa censurada à esquerda por um limite de detecção. Para o ajuste, os dados devem estar organizados em 3 colunas representando as seguintes variáveis:

- Variável resposta binária que identifica as observações com ou sem a condição de interesse
- Variável numérica censurada, sendo que os valores censurados são substituídos pelo limite de detecção observado (LD)
- Variável binária que identifica as observações com censura, com o valor 1, ou sem censura, com o valor 0

Precisamos especificar o valor do limite de detecção LD observado. Com estas informações, o código apresentado a seguir pode ser utilizado para o

ajuste.

```

1
2 ajustemv <- function(ve=ve, vr=vr, id=id, LD=LD) {
3
4   #--- Descrição:
5   #-----
6   #--- Estima os parâmetros e erros padrão para um modelo de
7   #--- regressão logística com variável explicativa numérica
8   #--- contínua única e com valores censurados à esquerda por
9   #--- um limite de detecção.
10  #---
11  #--- Os dados devem estar disponíveis pelo comando attach( ).
12  #---
13  #--- Argumentos obrigatórios:
14  #-----
15  #--- ve: variável explicativa numérica, contínua e positiva,
16  #--- com valores censurados substituídos pelo valor do
17  #--- limite de detecção observado.
18  #---
19  #--- vr: variável resposta binária de interesse.
20  #---
21  #--- id: Variável dicotômica que assume valor 1 caso a
22  #--- observação tenha valor censurado e 0 caso contrário
23  #---
24  #--- LD: Valor positivo correspondente ao
25  #--- limite de detecção observado
26  #---
27  #--- Saída esperada
28  #-----
29  #--- Retorna os valores estimados para os dois parâmetros do
30  #--- modelo e seus erros padrão, obtidos pelo ajuste por
31  #--- máxima verossimilhança, como exemplo abaixo.
32  #---
33  #---           [,1]           [,2]           [,3]
34  #---           [1,] "Parâmetro"       "Estimativa" "Erro padrão"
35  #---           [2,] "intercepto"      "x.xxx"      "x.xxx"
36  #---           [3,] "variável explicativa" "x.xxx"      "x.xxx"
37  #---
38  #--- Referência:
39  #-----
40  #--- Cole, Stephen R and Chu, Haitao and Nie, Lei and Schisterman,
41  #--- Enrique F. (2009). Estimating the odds ratio when exposure
42  #--- has a limit of detection. International journal of
43  #--- epidemiology. 38,1674-1680.
44  #---
45  #--- Exemplo de utilização:
46  #-----
47  #--- ajustemv(x, y, delta, LD)
48  #---
49
50  int.p1<-function(ve) {1/(1+exp(-b0-b1*ve))*dlnorm(ve, c0, c1)}
51  int.p0<-function(ve) {1/(1+exp(b0+b1*ve))*dlnorm(ve, c0, c1)}
52  logitln <- function(para) {
53    b0 <-< para[1]
54    b1 <-< para[2]
55    c0 <-< para[3]
56    c1 <-< para[4]
57    pve <- 1/(1+exp(-b0-b1*ve))

```

```

58   logL.1 <- sum((1-id)*(vr*log(pve)+(1-vr)*log(1-pve)+log(dlnorm(ve
      ,c0,c1))))
59   logL.2 <- sum(id*vr*log(integrate(int.p1,lower=0,upper=LD)$value)
      +
60           id*(1-vr)*log(integrate(int.p0,lower=0,upper=LD)$
      value))
61   return(-(logL.1+logL.2))
62 }
63 fit = try(optim(par=c(-1, 1, 0, 1), fn=logitln,hessian = T))
64 coef0 = fit$par[1]
65 coef1 = fit$par[2]
66 se0 = sqrt(diag(solve(fit$hessian)))[1];se1 = sqrt(diag(solve(fit$
      hessian)))[2]
67 res <- rbind(cbind("Parâmetro", "Estimativa","Erro padrão"), cbind(
      "intercepto",round(coef0,3),round(se0,3)), cbind("variável
      explicativa",round(coef1,3), round(se1,3)))
68 return(res)
69 }

```


Apêndice C

Resultados obtidos com as simulações

Tabela C.1: Vieses relativos percentuais (erros padrão) de $\hat{\beta}_0$ para $\beta_1 = 0,25$

n	% Censura	Método								
		Completo	M0	MLD	MLD2	ME	MU	MN	MR	MMV
50	5%	7,2% (1,6%)	6,2% (1,6%)	7,6% (1,6%)	6,9% (1,6%)	6,2% (1,6%)	6,9% (1,6%)	7,2% (1,6%)	6,2% (1,6%)	7,2% (1,6%)
	10%	11,5% (1,7%)	9,3% (1,6%)	12,6% (1,7%)	11,0% (1,7%)	9,5% (1,6%)	11,0% (1,7%)	11,5% (1,7%)	9,3% (1,6%)	11,6% (1,7%)
	25%	8,6% (1,6%)	-0,5% (1,4%)	14,2% (1,6%)	6,7% (1,5%)	1,7% (1,5%)	6,4% (1,5%)	8,2% (1,6%)	-0,5% (1,4%)	8,5% (1,6%)
	50%	10,2% (1,6%)	-13,4% (1,3%)	38,9% (2,1%)	8,8% (1,5%)	-2,6% (1,4%)	7,2% (1,5%)	8,7% (1,6%)	-13,1% (1,3%)	11,6% (1,6%)
100	5%	5,4% (1,1%)	4,7% (1,1%)	5,8% (1,1%)	5,2% (1,1%)	4,8% (1,1%)	5,2% (1,1%)	5,4% (1,1%)	4,7% (1,1%)	5,5% (1,1%)
	10%	5,4% (1,1%)	3,3% (1,1%)	6,3% (1,1%)	4,8% (1,1%)	3,5% (1,1%)	4,8% (1,1%)	5,3% (1,1%)	3,3% (1,1%)	5,4% (1,1%)
	25%	4,5% (1,1%)	-3,2% (1,0%)	9,4% (1,1%)	3,1% (1,0%)	-1,3% (1,0%)	2,9% (1,0%)	4,3% (1,1%)	-3,2% (1,0%)	4,6% (1,1%)
	50%	4,2% (1,1%)	-17,1% (0,9%)	28,0% (1,3%)	2,9% (1,0%)	-7,3% (1,0%)	1,6% (1,1%)	2,8% (1,1%)	-16,9% (0,9%)	4,8% (1,1%)
200	5%	2,5% (0,8%)	1,9% (0,8%)	2,8% (0,8%)	2,3% (0,8%)	1,9% (0,8%)	2,3% (0,8%)	2,5% (0,8%)	1,9% (0,8%)	2,5% (0,8%)
	10%	1,6% (0,7%)	-0,3% (0,7%)	2,4% (0,8%)	1,1% (0,7%)	-0,1% (0,7%)	1,1% (0,7%)	1,6% (0,7%)	-0,3% (0,7%)	1,6% (0,7%)
	25%	3,7% (0,7%)	-3,5% (0,7%)	8,2% (0,8%)	2,3% (0,7%)	-1,8% (0,7%)	2,1% (0,7%)	3,5% (0,7%)	-3,5% (0,7%)	3,7% (0,7%)
	50%	2,1% (0,7%)	-17,8% (0,6%)	23,6% (0,9%)	0,9% (0,7%)	-8,6% (0,7%)	-0,5% (0,7%)	1,1% (0,7%)	-17,8% (0,6%)	2,5% (0,7%)

Nota: Resultados obtidos com 1000 iterações. n: tamanho da amostra.

Tabela C.2: Vieses relativos percentuais (erros padrão) de $\hat{\beta}_1$ para $\beta_1 = 0,25$

n	% Censura	Método								
		Completo	M0	MLD	MLD2	ME	MU	MN	MR	MMV
50	5%	17,2% (2,0%)	16,1% (2,0%)	17,6% (2,0%)	16,9% (2,0%)	16,2% (2,0%)	16,9% (2,0%)	17,2% (2,0%)	16,1% (2,0%)	17,2% (2,0%)
	10%	17,4% (2,2%)	15,0% (2,2%)	18,4% (2,2%)	16,8% (2,2%)	15,3% (2,2%)	16,9% (2,2%)	17,4% (2,2%)	15,0% (2,2%)	17,4% (2,2%)
	25%	14,8% (1,9%)	6,1% (1,8%)	19,7% (2,1%)	13,0% (1,9%)	8,3% (1,8%)	12,5% (1,9%)	14,4% (1,9%)	6,1% (1,8%)	14,7% (1,9%)
	50%	16,5% (2,0%)	-3,3% (1,6%)	39,4% (2,8%)	15,5% (2,0%)	6,0% (1,8%)	12,5% (2,0%)	14,2% (2,0%)	-3,1% (1,6%)	18,0% (2,1%)
100	5%	8,4% (1,4%)	7,8% (1,4%)	8,7% (1,4%)	8,3% (1,4%)	7,8% (1,4%)	8,3% (1,4%)	8,4% (1,4%)	7,8% (1,4%)	8,5% (1,4%)
	10%	11,1% (1,4%)	9,0% (1,3%)	12,0% (1,4%)	10,6% (1,4%)	9,3% (1,3%)	10,6% (1,4%)	11,0% (1,4%)	9,0% (1,3%)	11,1% (1,4%)
	25%	8,1% (1,3%)	1,2% (1,2%)	12,0% (1,3%)	6,8% (1,2%)	3,0% (1,2%)	6,5% (1,2%)	7,8% (1,3%)	1,2% (1,2%)	8,2% (1,3%)
	50%	7,6% (1,3%)	-9,2% (1,0%)	24,7% (1,7%)	6,7% (1,3%)	-1,3% (1,2%)	4,2% (1,3%)	5,3% (1,3%)	-9,1% (1,0%)	8,2% (1,3%)
200	5%	3,8% (0,9%)	3,2% (0,9%)	4,0% (0,9%)	3,6% (0,9%)	3,2% (0,9%)	3,6% (0,9%)	3,8% (0,9%)	3,2% (0,9%)	3,8% (0,9%)
	10%	3,0% (0,9%)	1,4% (0,9%)	3,7% (0,9%)	2,6% (0,9%)	1,6% (0,9%)	2,6% (0,9%)	3,0% (0,9%)	1,4% (0,9%)	3,0% (0,9%)
	25%	5,3% (0,9%)	-0,9% (0,8%)	8,8% (0,9%)	4,2% (0,9%)	0,7% (0,8%)	3,9% (0,9%)	5,1% (0,9%)	-0,9% (0,8%)	5,4% (0,9%)
	50%	3,0% (0,9%)	-12,0% (0,7%)	17,5% (1,1%)	2,2% (0,9%)	-4,8% (0,8%)	-0,3% (0,9%)	1,4% (0,9%)	-12,0% (0,7%)	3,4% (0,9%)

Nota: Resultados obtidos com 1000 iterações. n: tamanho da amostra.

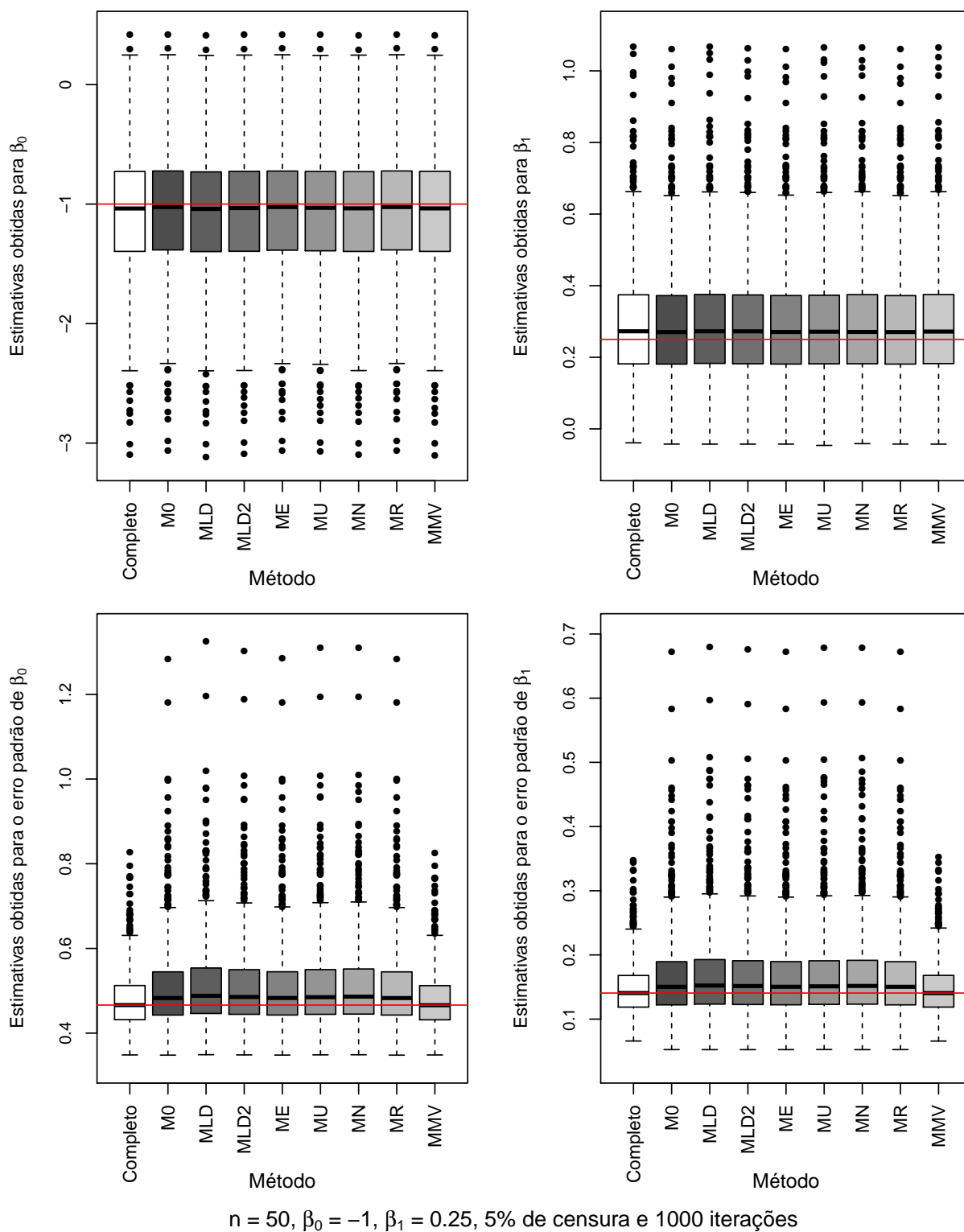


Figura C.1: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 5% de censura e $\beta_1 = 0,25$.

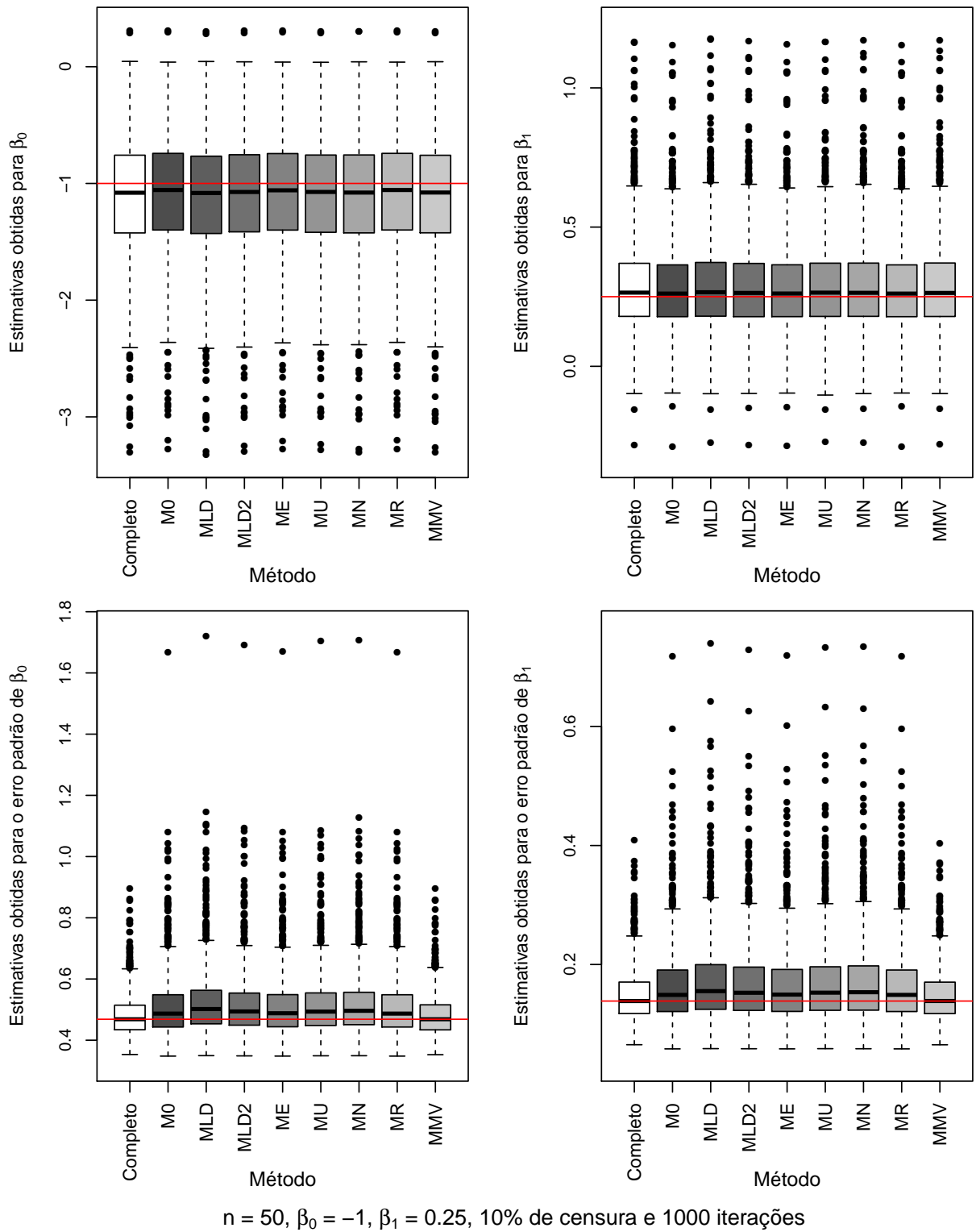


Figura C.2: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 10% de censura e $\beta_1 = 0,25$.

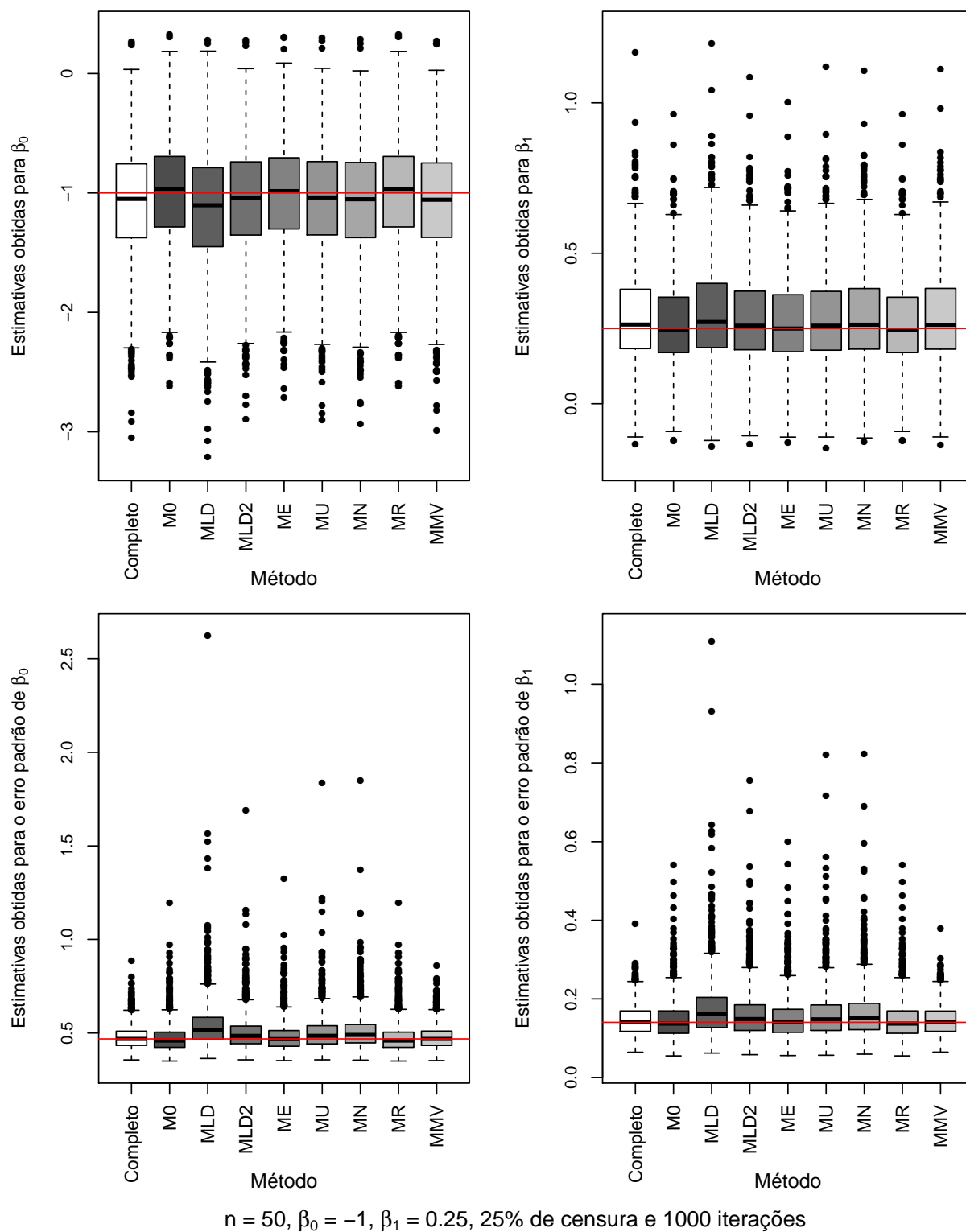


Figura C.3: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 25% de censura e $\beta_1 = 0,25$.

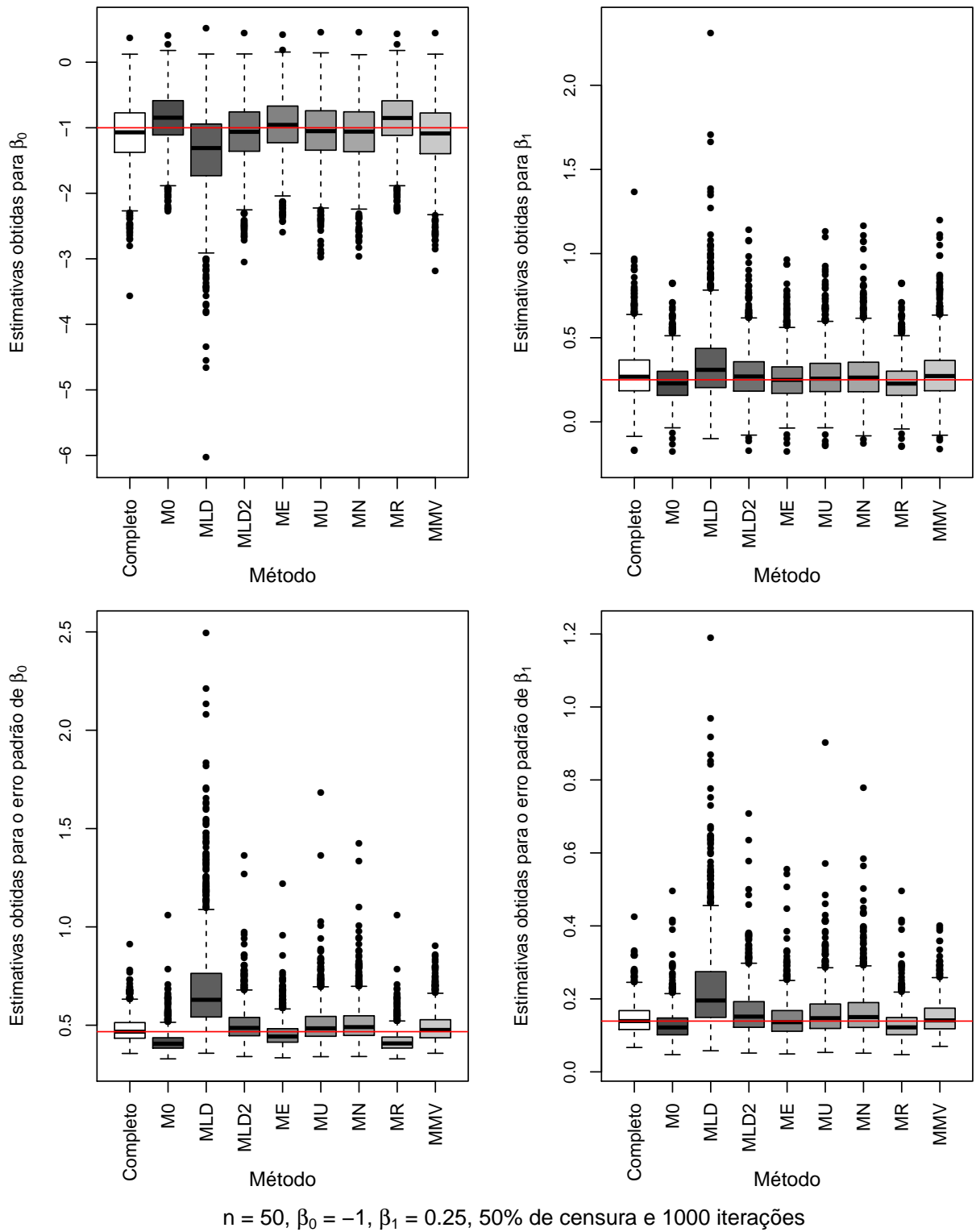


Figura C.4: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 50% de censura e $\beta_1 = 0,25$.

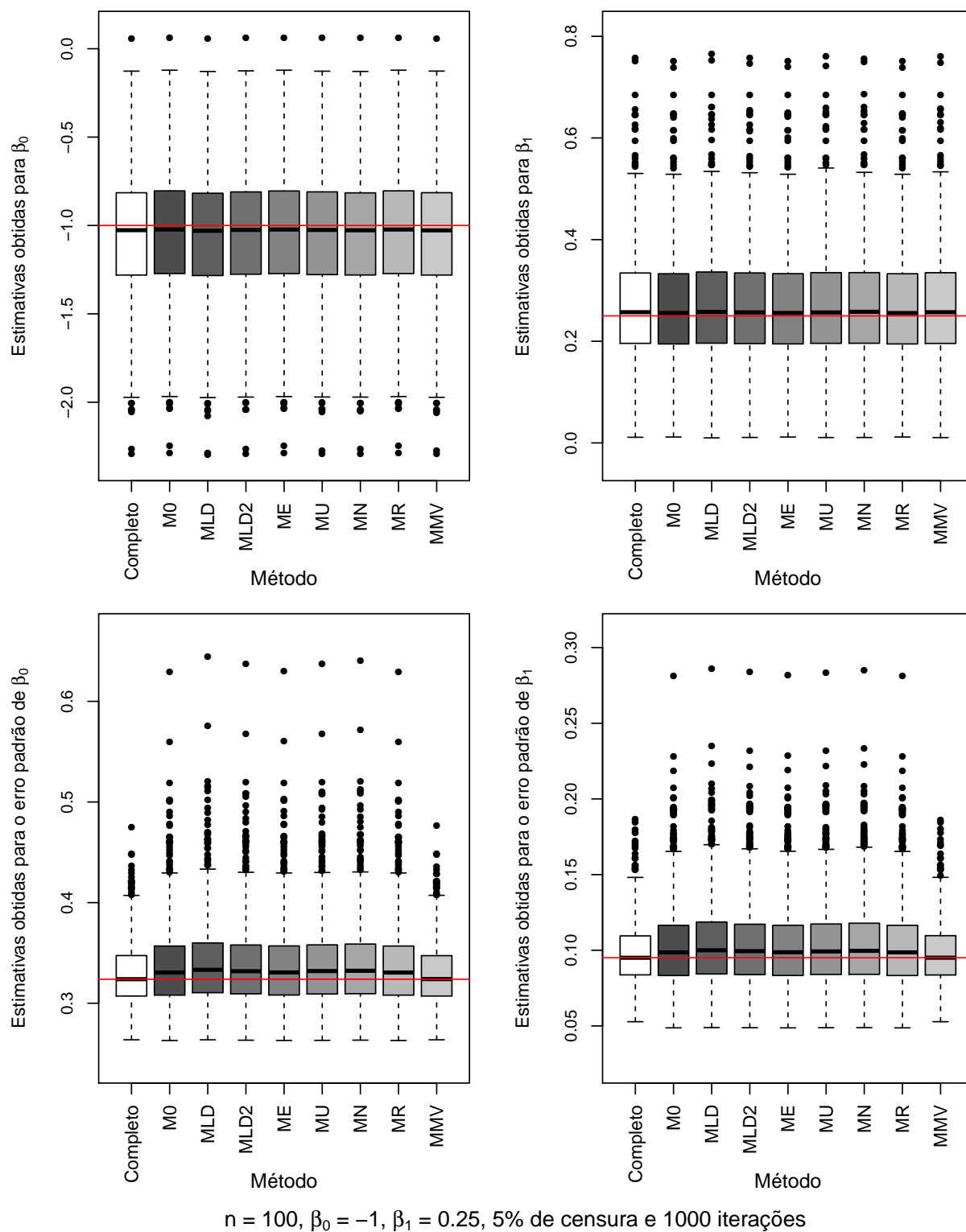


Figura C.5: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 5% de censura e $\beta_1 = 0,25$.

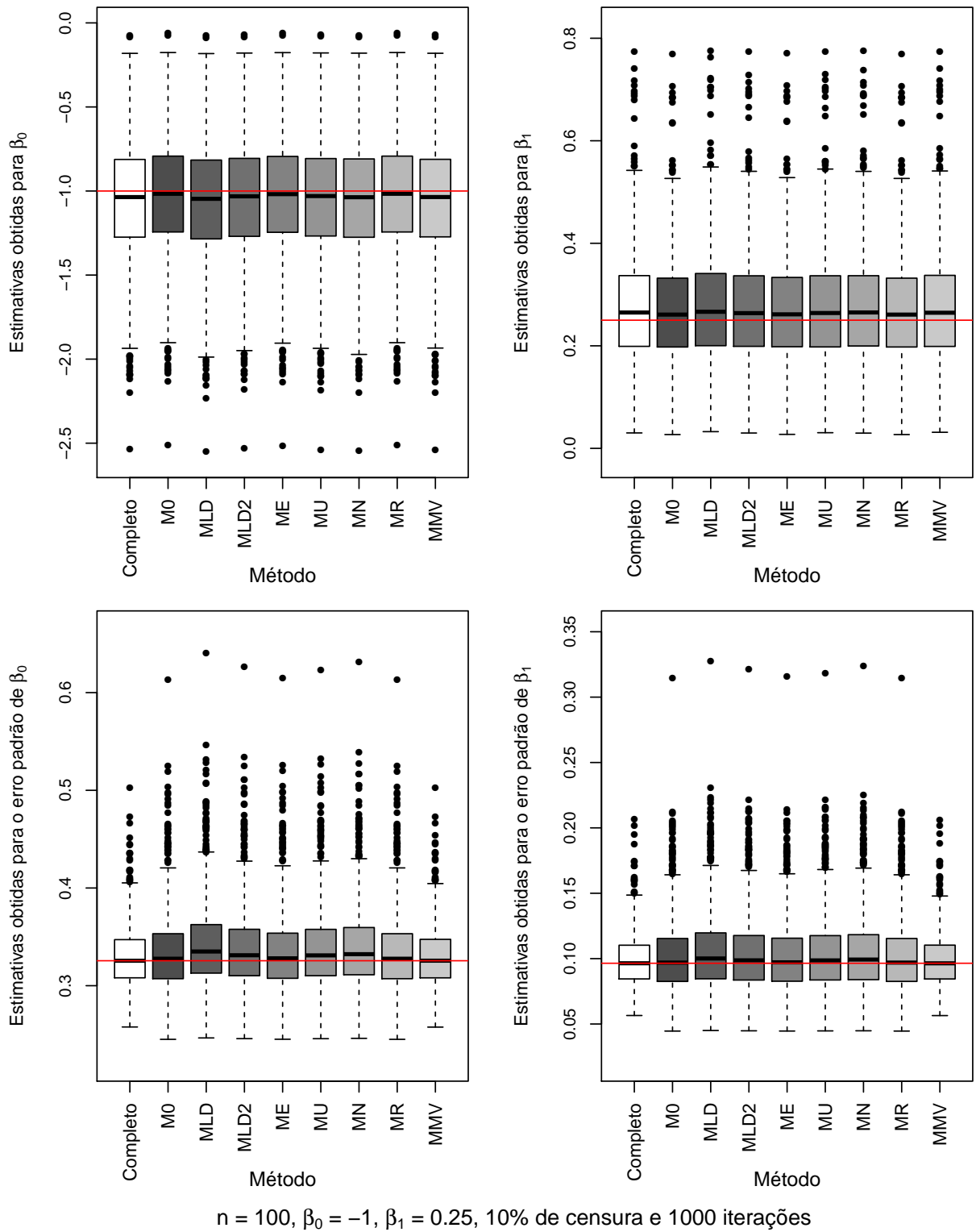


Figura C.6: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 10% de censura e $\beta_1 = 0,25$.

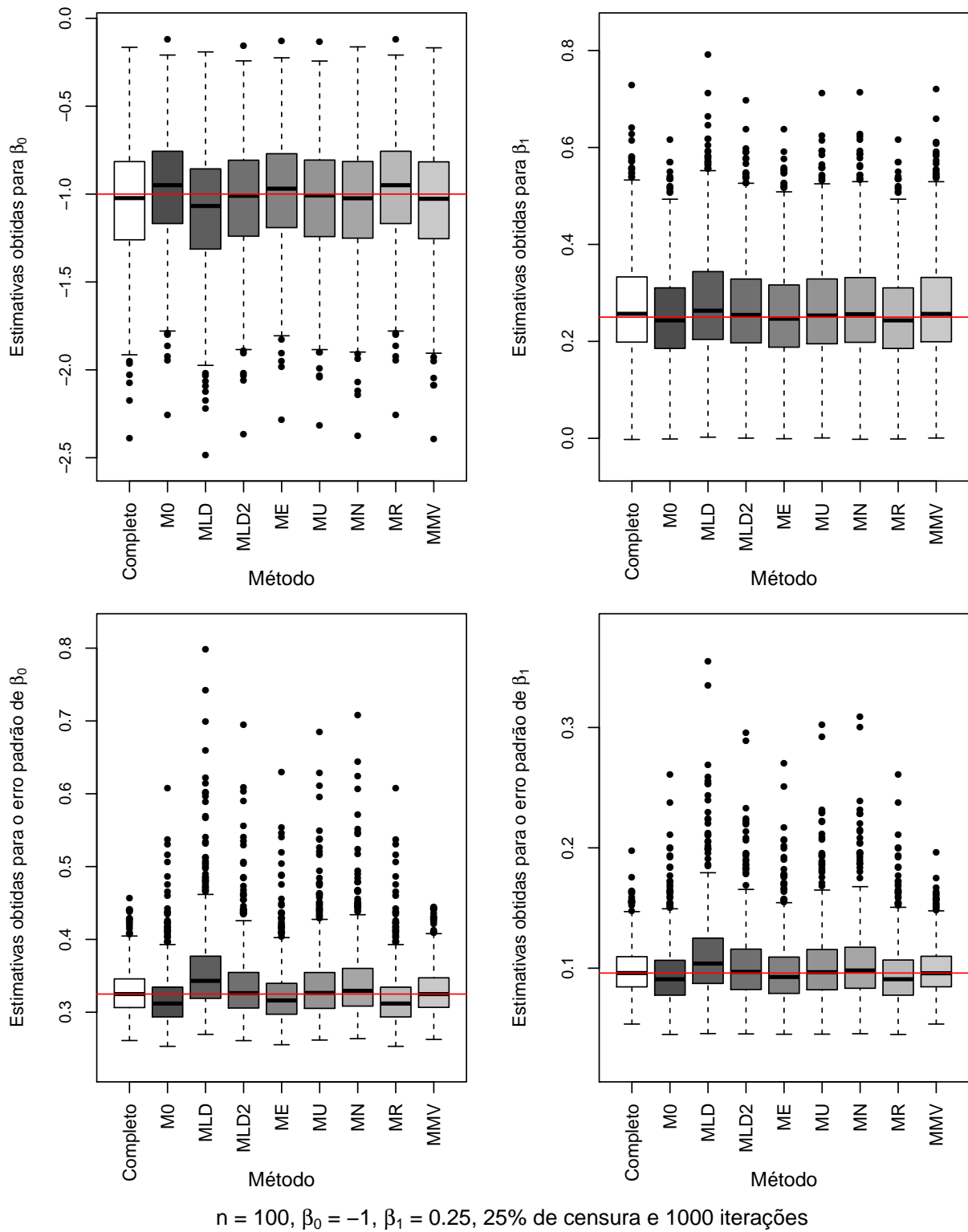
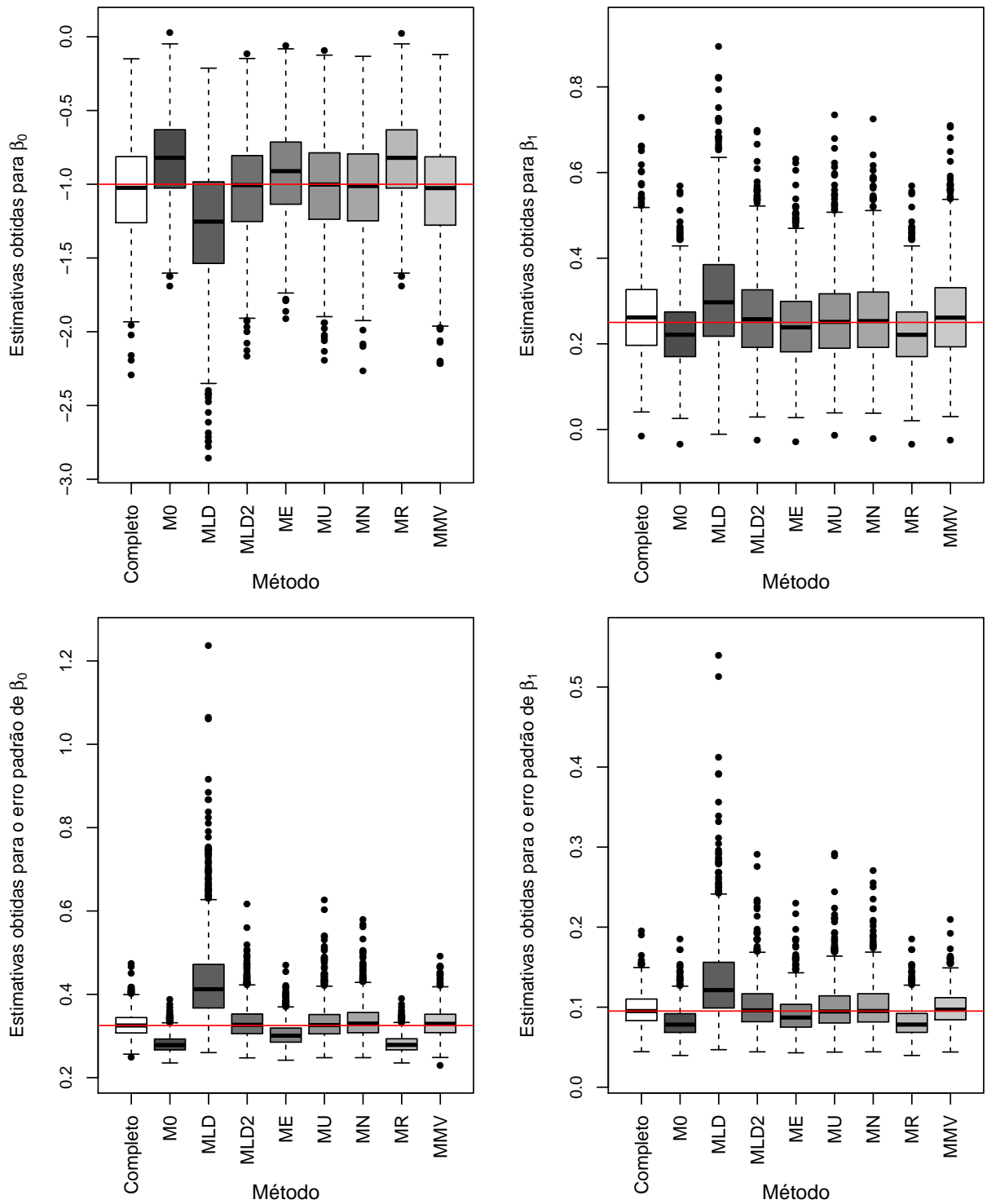


Figura C.7: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 25% de censura e $\beta_1 = 0,25$.



$n = 100, \beta_0 = -1, \beta_1 = 0.25, 50\%$ de censura e 1000 iterações

Figura C.8: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 50% de censura e $\beta_1 = 0,25$.

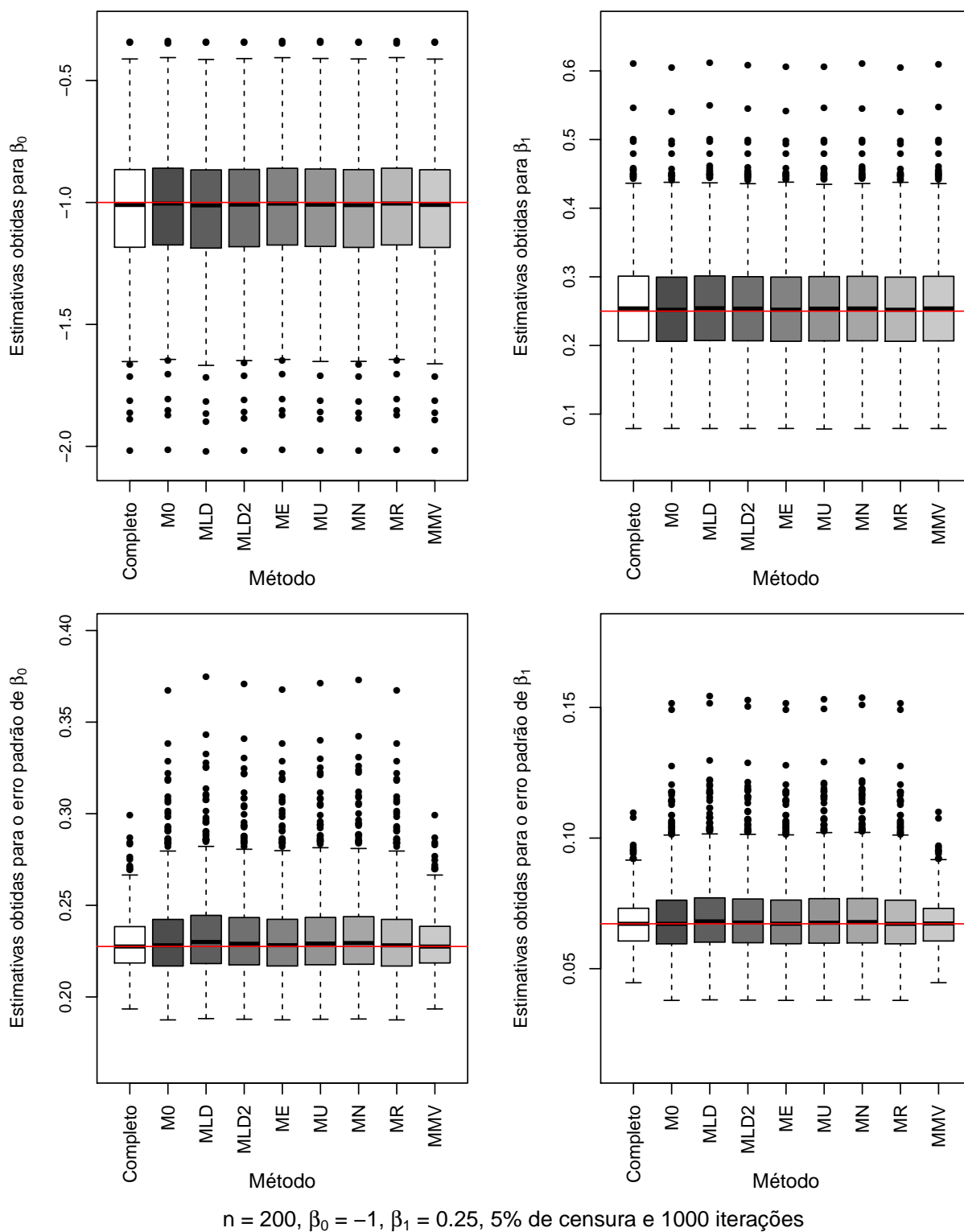


Figura C.9: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 5% de censura e $\beta_1 = 0,25$.

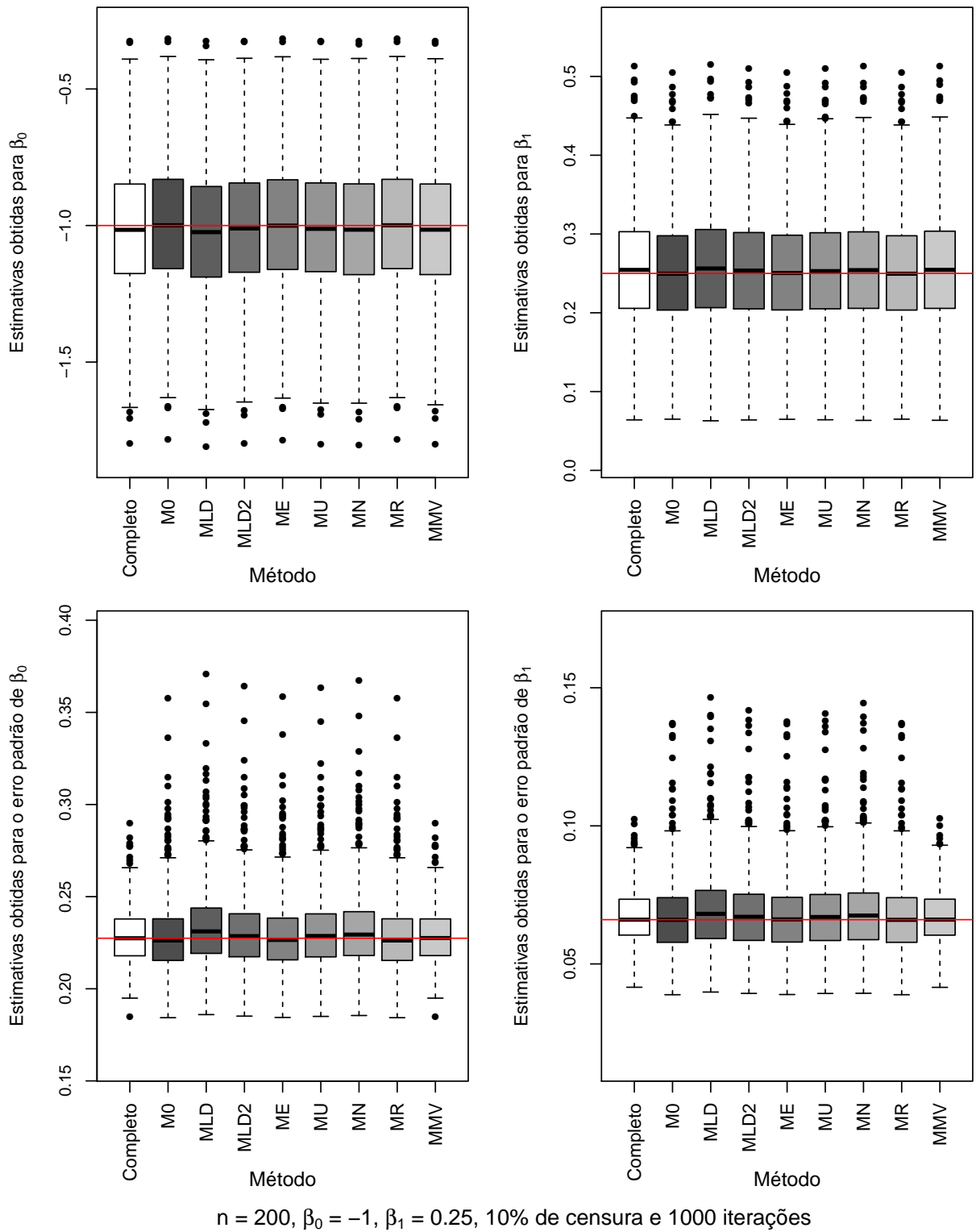


Figura C.10: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 10% de censura e $\beta_1 = 0,25$.

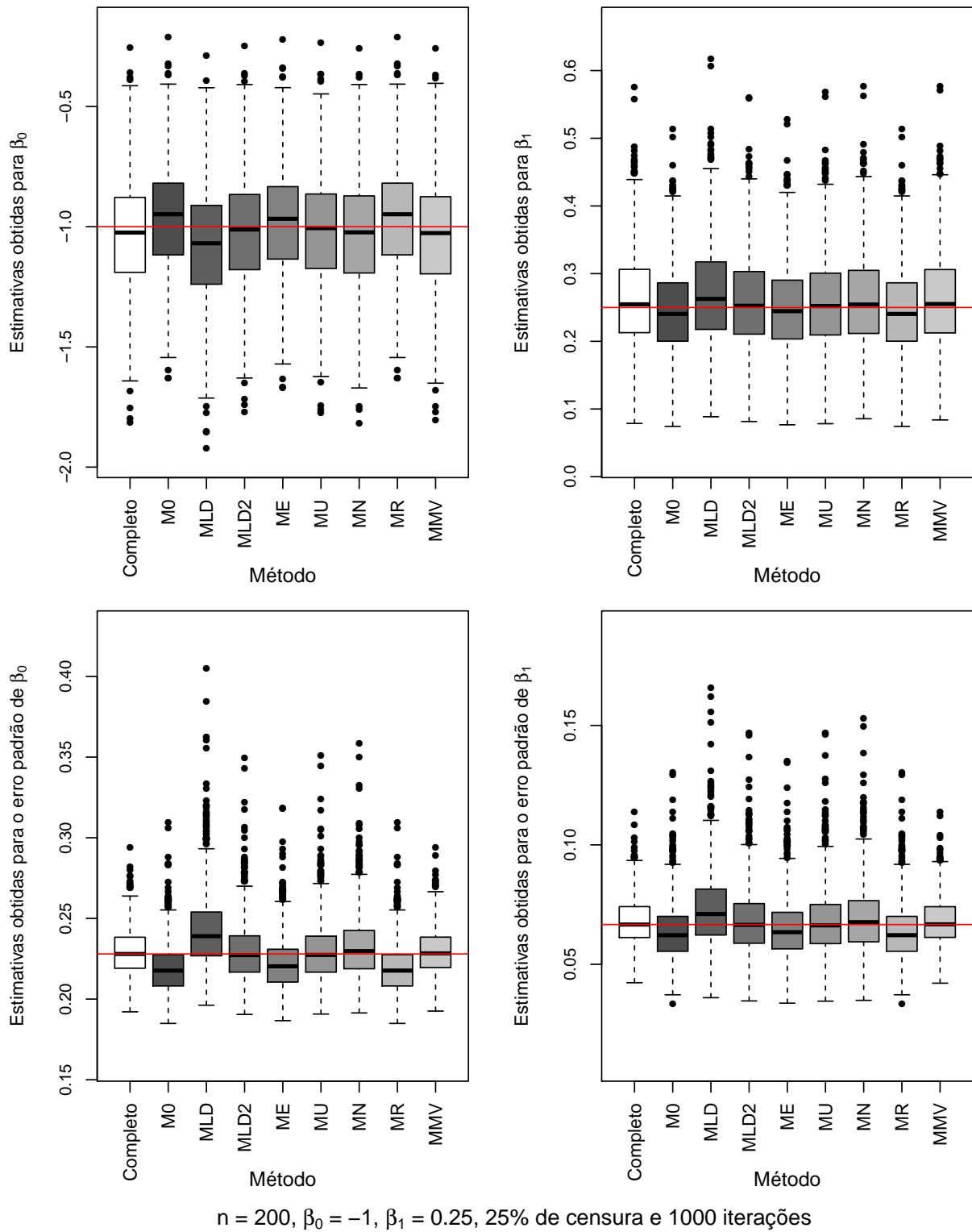
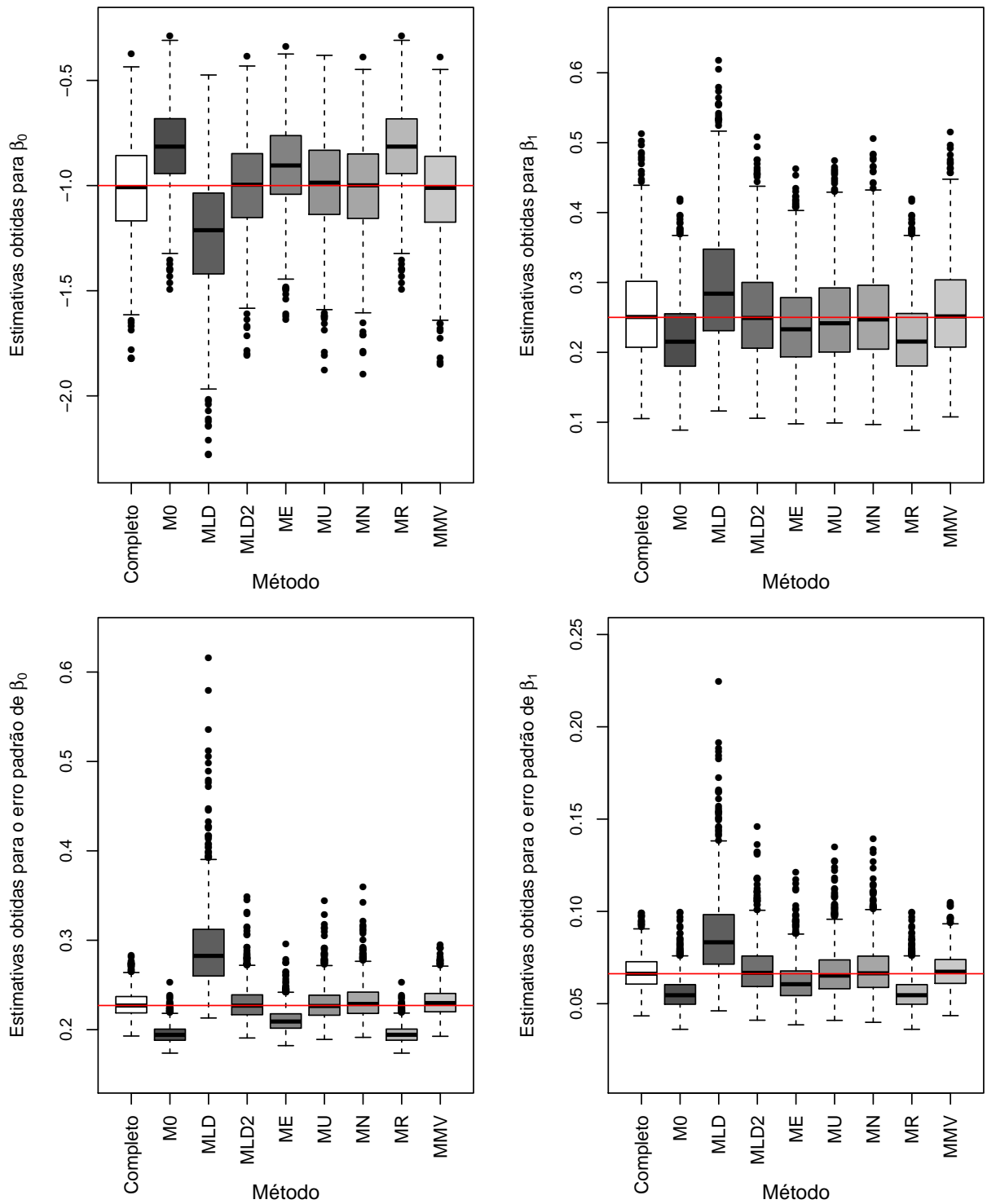


Figura C.11: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 25% de censura e $\beta_1 = 0,25$.



$n = 200, \beta_0 = -1, \beta_1 = 0.25, 50\%$ de censura e 1000 iterações

Figura C.12: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostrass de tamanho 200, 50% de censura e $\beta_1 = 0,25$.

Tabela C.3: Vieses relativos percentuais (erros padrão) de $\hat{\beta}_0$ para $\beta_1 = 0,75$

n	% Censura	Método								
		Completo	M0	MLD	MLD2	ME	MU	MN	MR	MMV
50	5%	13,4% (2,0%)	8,3% (2,0%)	15,2% (2,0%)	11,8% (2,0%)	8,6% (2,0%)	11,9% (2,0%)	13,1% (2,0%)	8,3% (2,0%)	13,1% (2,0%)
	10%	10,5% (2,1%)	-0,2% (2,0%)	15,6% (2,2%)	7,7% (2,1%)	1,0% (2,0%)	7,6% (2,1%)	10,4% (2,1%)	-0,2% (2,0%)	10,6% (2,1%)
	25%	10,6% (2,0%)	-26,9% (1,6%)	38,4% (2,4%)	1,7% (1,8%)	-18,7% (1,7%)	-0,3% (1,8%)	7,8% (1,9%)	-26,8% (1,6%)	10,4% (2,0%)
	50%	9,5% (2,0%)	-68,4% (1,4%)	597,7% (399,0%)	5,8% (2,5%)	-35,8% (1,6%)	-13,6% (1,9%)	-5,0% (1,9%)	-67,6% (1,3%)	16,2% (2,3%)
100	5%	4,8% (1,4%)	1,1% (1,4%)	6,3% (1,4%)	3,8% (1,4%)	1,3% (1,4%)	3,7% (1,4%)	4,8% (1,4%)	1,1% (1,4%)	4,8% (1,4%)
	10%	3,7% (1,4%)	-5,5% (1,3%)	8,1% (1,4%)	1,4% (1,4%)	-4,5% (1,3%)	1,2% (1,4%)	3,6% (1,4%)	-5,5% (1,3%)	3,8% (1,4%)
	25%	5,7% (1,4%)	-26,5% (1,2%)	29,3% (1,6%)	-0,9% (1,3%)	-19,2% (1,2%)	-2,4% (1,3%)	4,0% (1,3%)	-26,5% (1,2%)	6,4% (1,4%)
	50%	4,4% (1,3%)	-68,5% (1,0%)	140,9% (4,4%)	-0,2% (1,3%)	-38,0% (1,1%)	-15,9% (1,3%)	-9,2% (1,3%)	-68,3% (1,0%)	7,8% (1,5%)
200	5%	3,2% (0,9%)	-0,2% (0,9%)	4,5% (0,9%)	2,3% (0,9%)	0,0% (0,9%)	2,2% (0,9%)	3,1% (0,9%)	-0,2% (0,9%)	3,2% (0,9%)
	10%	3,2% (1,0%)	-5,7% (1,0%)	7,3% (1,0%)	0,9% (1,0%)	-4,8% (1,0%)	0,7% (1,0%)	3,1% (1,0%)	-5,7% (1,0%)	3,2% (1,0%)
	25%	1,8% (1,0%)	-29,0% (0,8%)	23,4% (1,2%)	-4,5% (1,0%)	-22,0% (0,9%)	-6,2% (1,0%)	0,5% (1,0%)	-29,0% (0,8%)	2,1% (1,0%)
	50%	3,3% (0,9%)	-67,6% (0,7%)	122,4% (2,3%)	-1,5% (0,9%)	-37,8% (0,7%)	-15,9% (0,9%)	-10,0% (0,9%)	-67,5% (0,7%)	5,3% (1,0%)

Nota: Resultados obtidos com 1000 iterações. n: tamanho da amostra.

Tabela C.4: Vieses relativos percentuais (erros padrão) de $\hat{\beta}_1$ para $\beta_1 = 0,75$

n	% Censura	Método								
		Completo	M0	MLD	MLD2	ME	MU	MN	MR	MMV
50	5%	11,9% (1,4%)	8,8% (1,3%)	13,0% (1,4%)	11,0% (1,4%)	9,0% (1,4%)	11,1% (1,4%)	11,8% (1,4%)	8,8% (1,3%)	11,8% (1,4%)
	10%	11,3% (1,5%)	4,9% (1,4%)	14,1% (1,6%)	9,7% (1,5%)	5,6% (1,4%)	9,5% (1,5%)	11,1% (1,5%)	4,9% (1,4%)	11,4% (1,5%)
	25%	11,2% (1,4%)	-9,1% (1,0%)	25,4% (1,8%)	6,4% (1,3%)	-4,7% (1,1%)	4,3% (1,3%)	8,9% (1,3%)	-9,1% (1,0%)	11,1% (1,4%)
	50%	10,6% (1,4%)	-21,0% (1,3%)	332,4% (234,0%)	11,4% (2,3%)	-6,9% (1,7%)	-6,6% (1,2%)	-1,7% (1,2%)	-21,7% (0,9%)	15,1% (1,6%)
100	5%	5,7% (0,9%)	3,5% (0,9%)	6,5% (0,9%)	5,1% (0,9%)	3,6% (0,9%)	5,0% (0,9%)	5,7% (0,9%)	3,5% (0,9%)	5,6% (0,9%)
	10%	4,4% (0,9%)	-0,8% (0,9%)	6,7% (1,0%)	3,1% (0,9%)	-0,2% (0,9%)	3,0% (0,9%)	4,3% (0,9%)	-0,8% (0,9%)	4,5% (0,9%)
	25%	5,3% (0,9%)	-11,4% (0,7%)	16,6% (1,1%)	2,0% (0,9%)	-7,5% (0,8%)	0,6% (0,9%)	3,9% (0,9%)	-11,4% (0,7%)	5,7% (1,0%)
	50%	5,5% (0,9%)	-24,3% (0,6%)	61,9% (2,7%)	4,2% (1,0%)	-11,6% (0,8%)	-10,0% (0,8%)	-5,7% (0,8%)	-24,2% (0,6%)	7,6% (1,0%)
200	5%	3,3% (0,6%)	1,4% (0,6%)	4,0% (0,6%)	2,8% (0,6%)	1,5% (0,6%)	2,7% (0,6%)	3,3% (0,6%)	1,4% (0,6%)	3,3% (0,6%)
	10%	3,0% (0,7%)	-2,0% (0,6%)	5,1% (0,7%)	1,8% (0,7%)	-1,4% (0,6%)	1,6% (0,7%)	2,9% (0,7%)	-2,0% (0,6%)	3,0% (0,7%)
	25%	2,4% (0,7%)	-13,2% (0,5%)	12,4% (0,8%)	-0,7% (0,6%)	-9,6% (0,6%)	-2,3% (0,6%)	1,3% (0,7%)	-13,2% (0,5%)	2,5% (0,7%)
	50%	3,4% (0,7%)	-25,4% (0,4%)	49,6% (1,4%)	1,6% (0,7%)	-13,2% (0,5%)	-11,1% (0,5%)	-7,7% (0,6%)	-25,4% (0,4%)	4,4% (0,7%)

Nota: Resultados obtidos com 1000 iterações. n: tamanho da amostra.

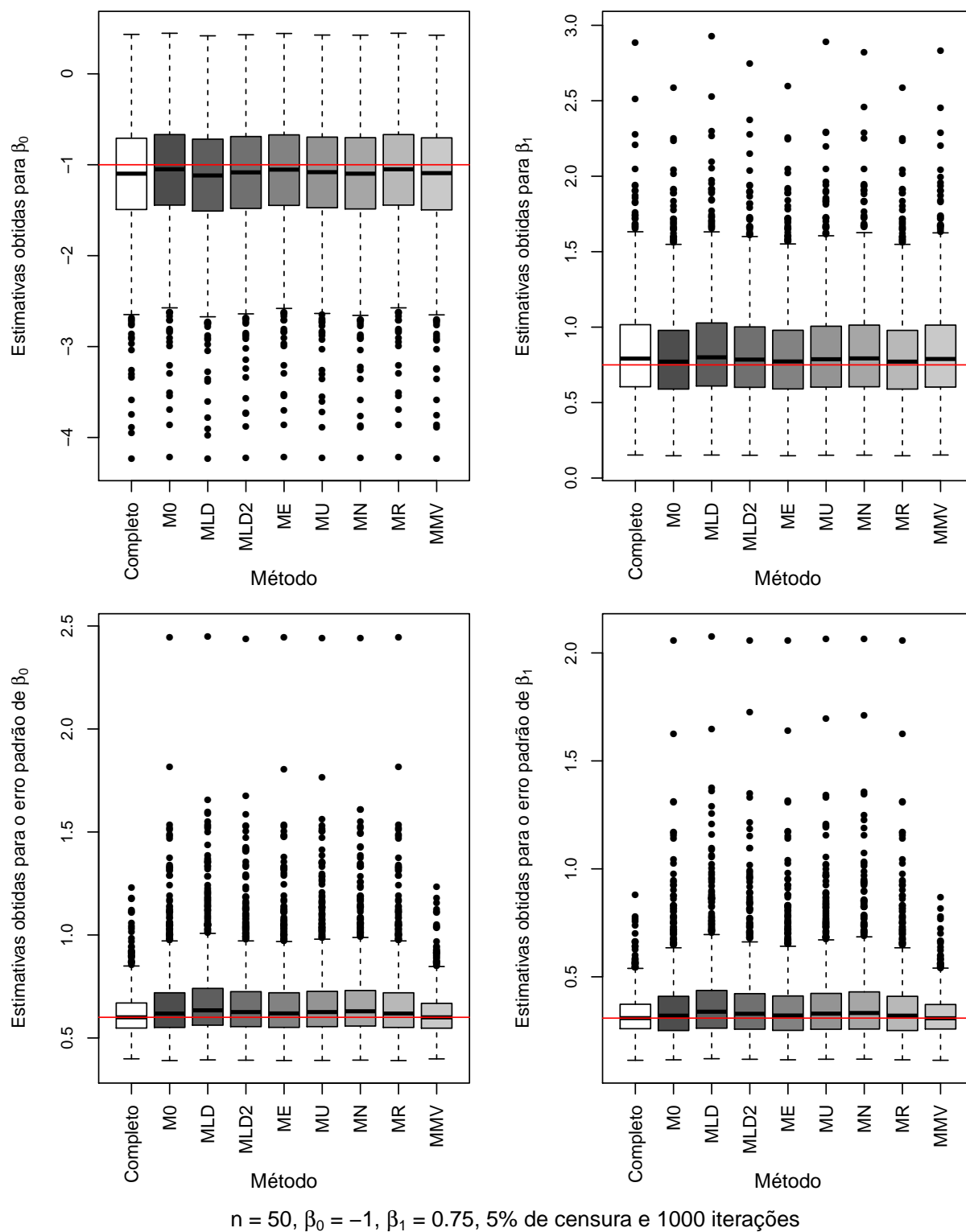


Figura C.13: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 5% de censura e $\beta_1 = 0,75$.

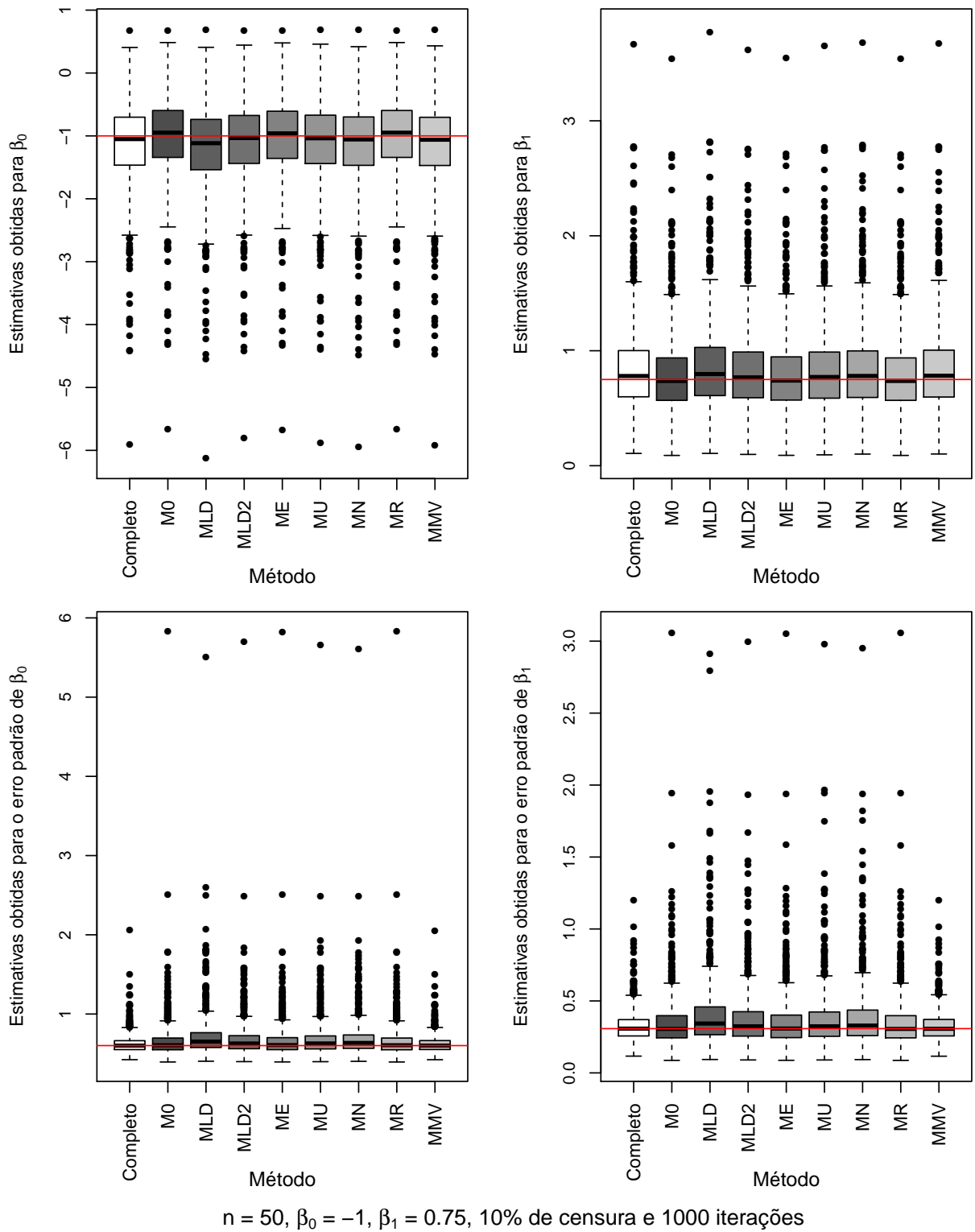


Figura C.14: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 10% de censura e $\beta_1 = 0,75$.

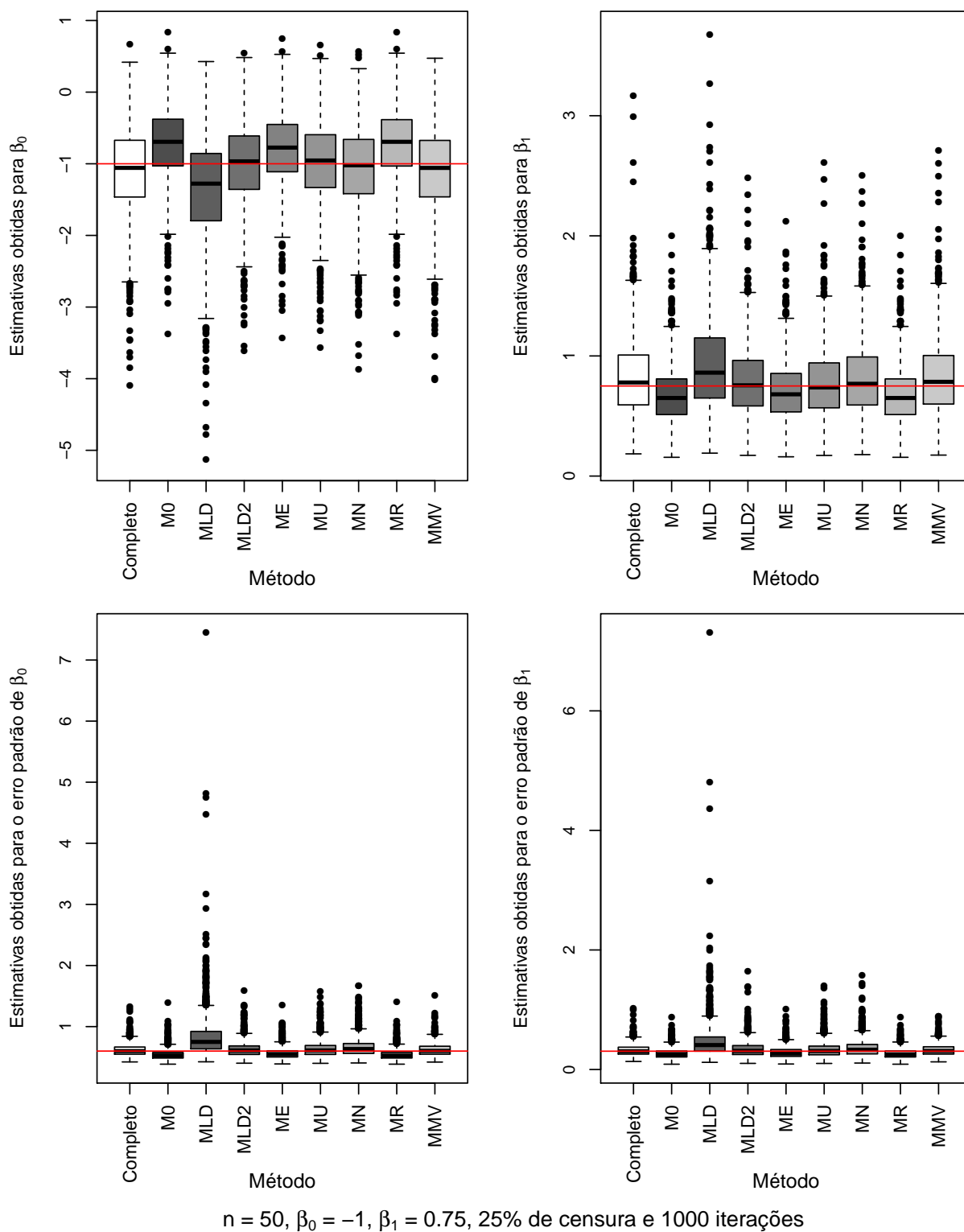
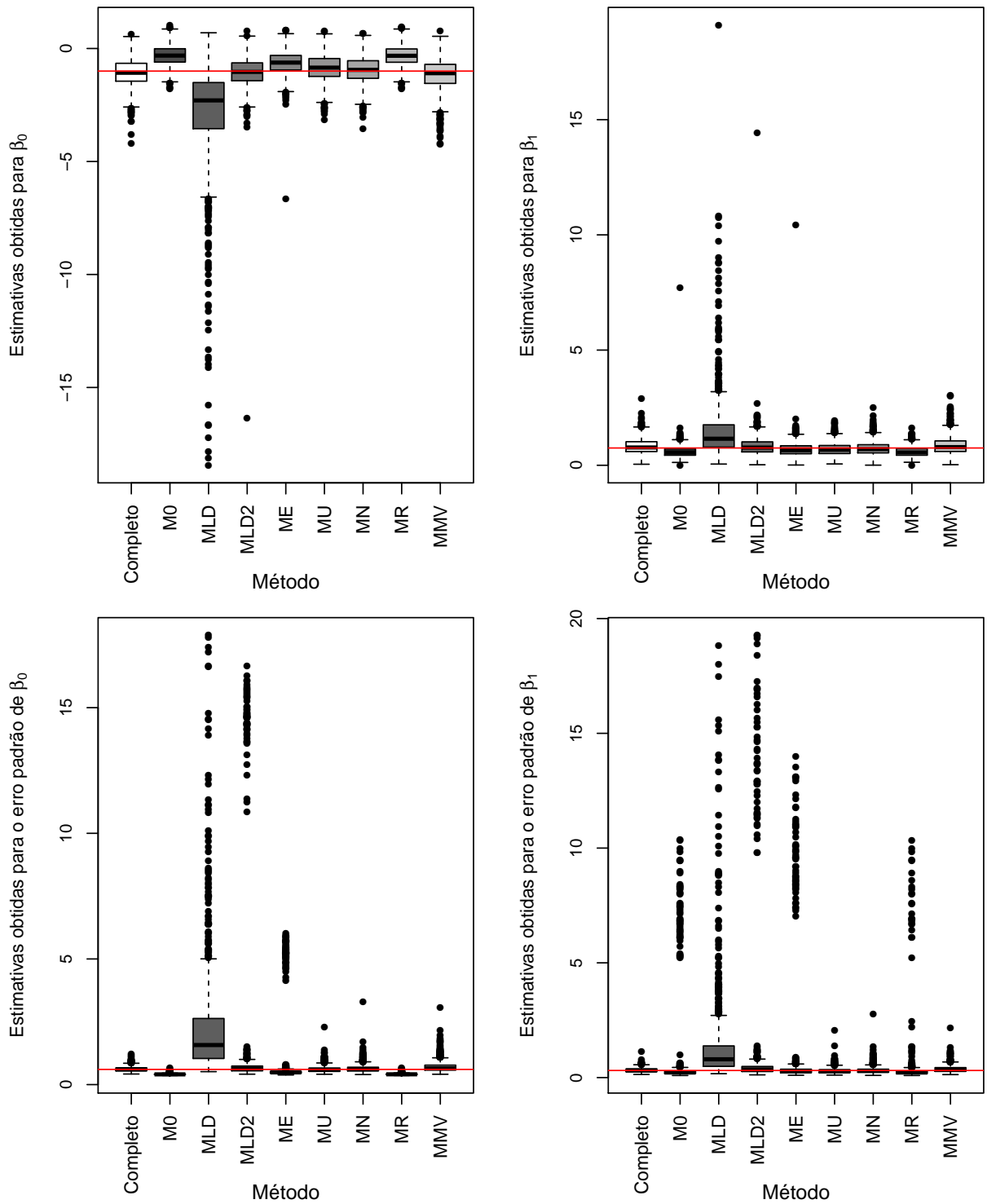


Figura C.15: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 25% de censura e $\beta_1 = 0,75$.



$n = 50, \beta_0 = -1, \beta_1 = 0.75, 50\%$ de censura e 1000 iterações. Dados truncados em $[-20,20]$

Figura C.16: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 50% de censura e $\beta_1 = 0,75$.

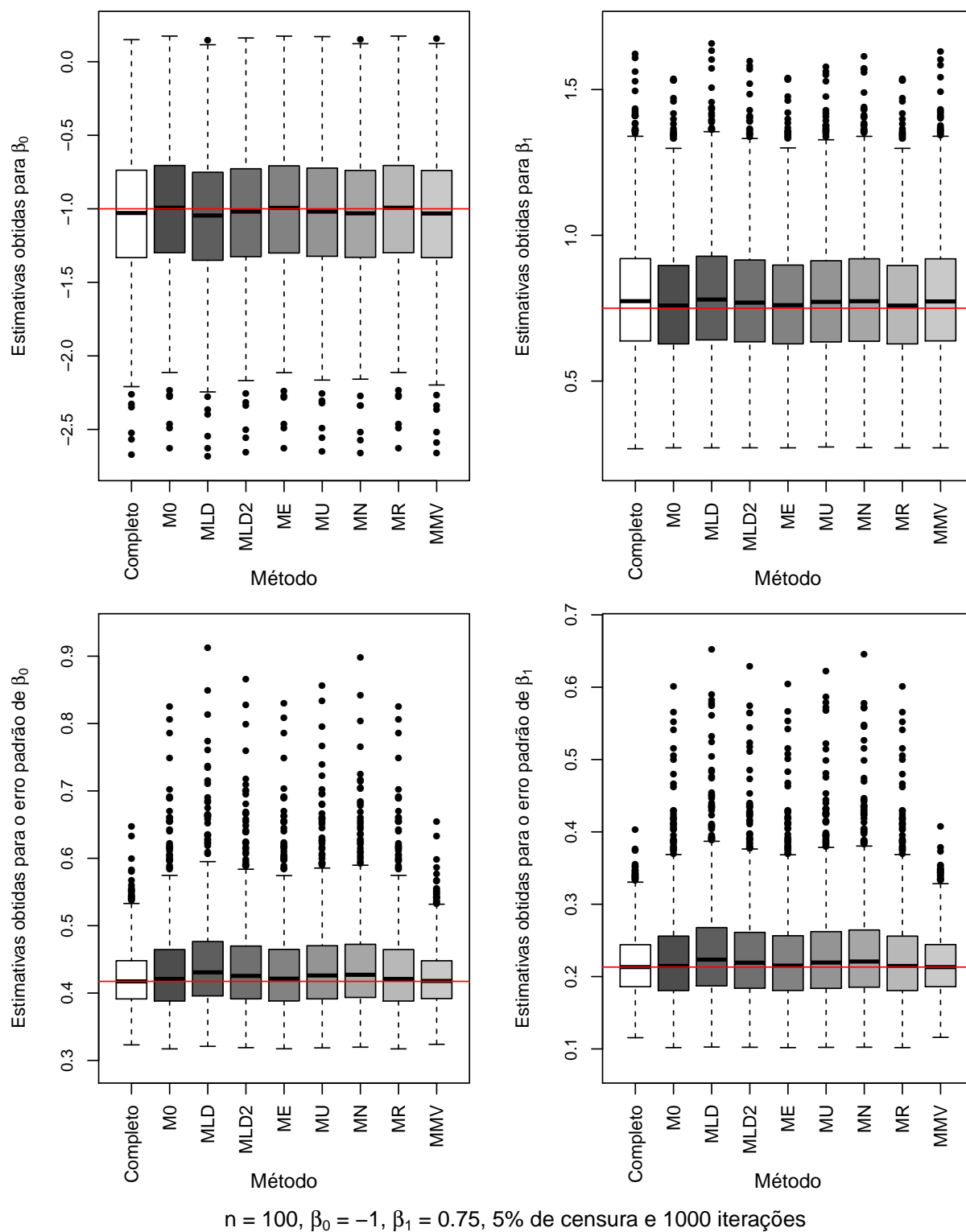


Figura C.17: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 5% de censura e $\beta_1 = 0,75$.

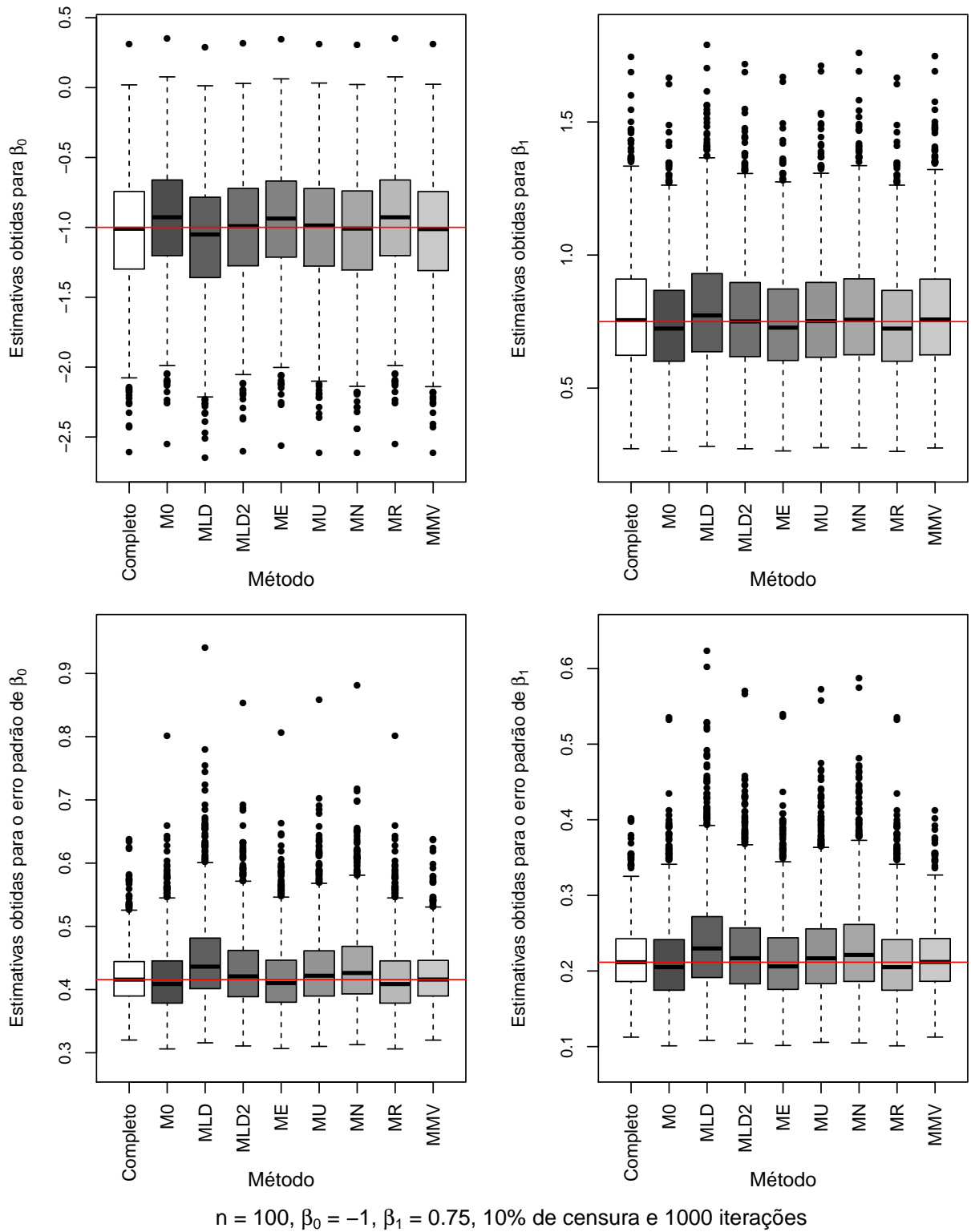


Figura C.18: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 10% de censura e $\beta_1 = 0,75$.

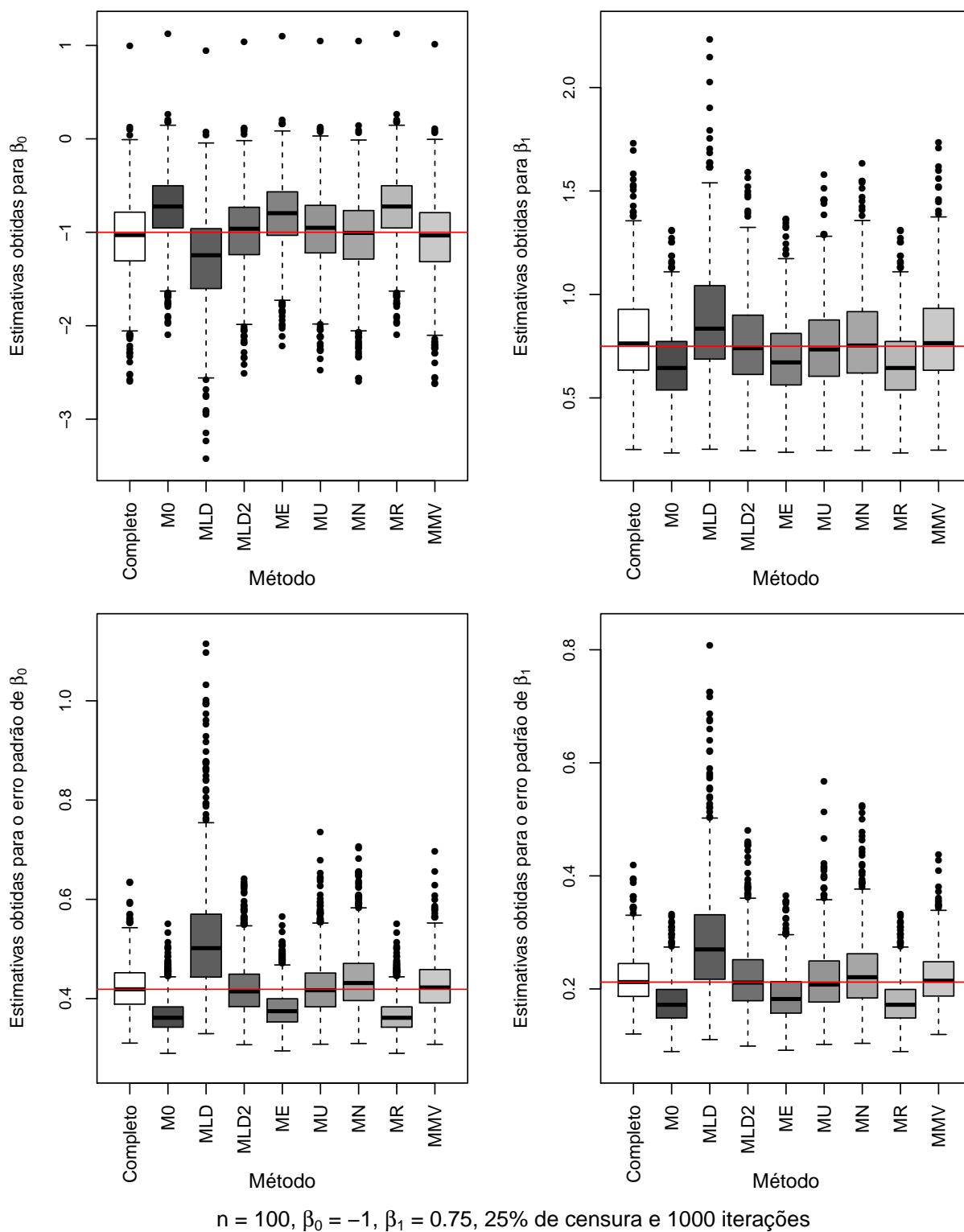
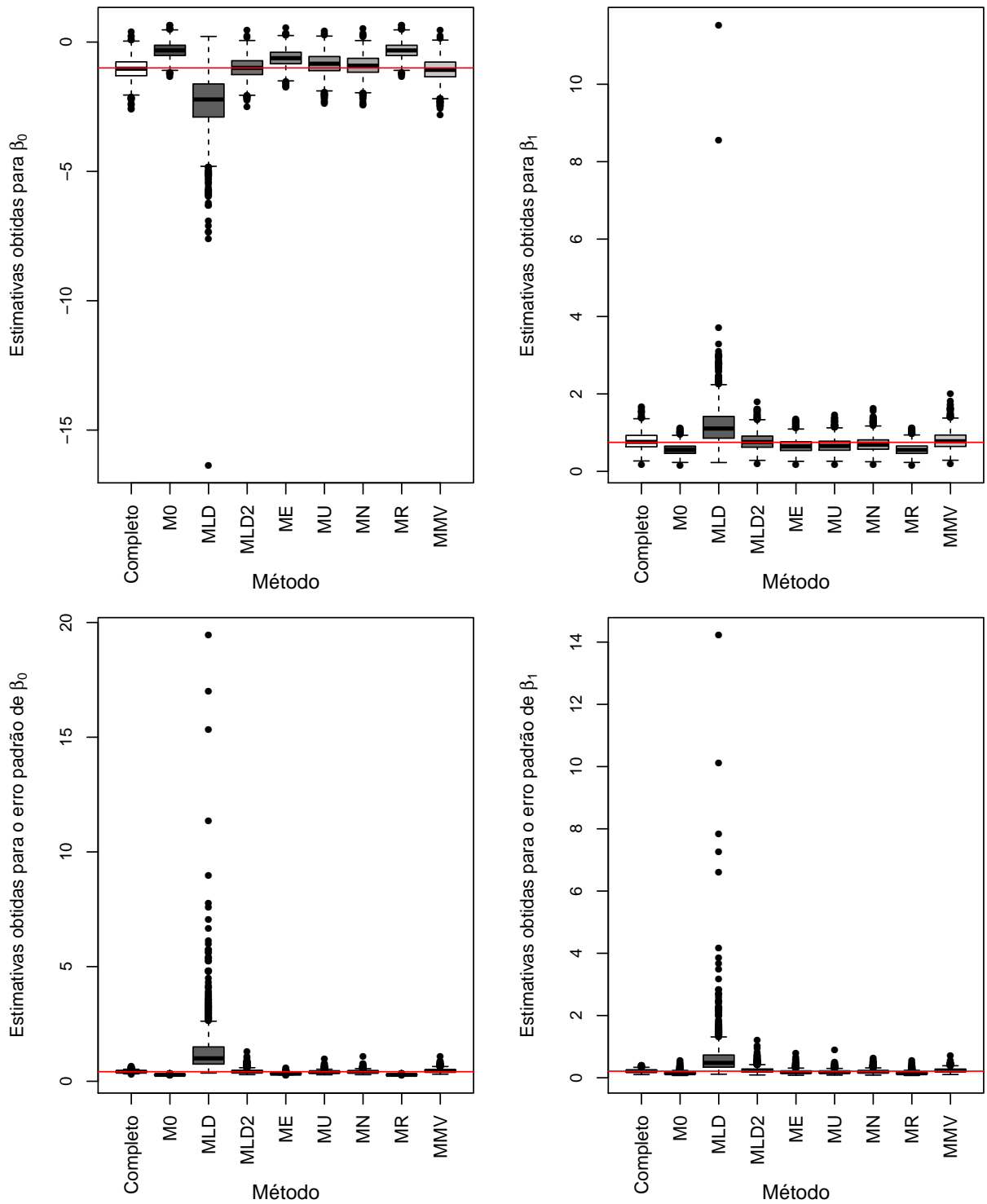


Figura C.19: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 25% de censura e $\beta_1 = 0,75$.



$n = 100$, $\beta_0 = -1$, $\beta_1 = 0.75$, 50% de censura e 1000 iterações. Dados truncados em $[-20,20]$

Figura C.20: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 50% de censura e $\beta_1 = 0,75$.

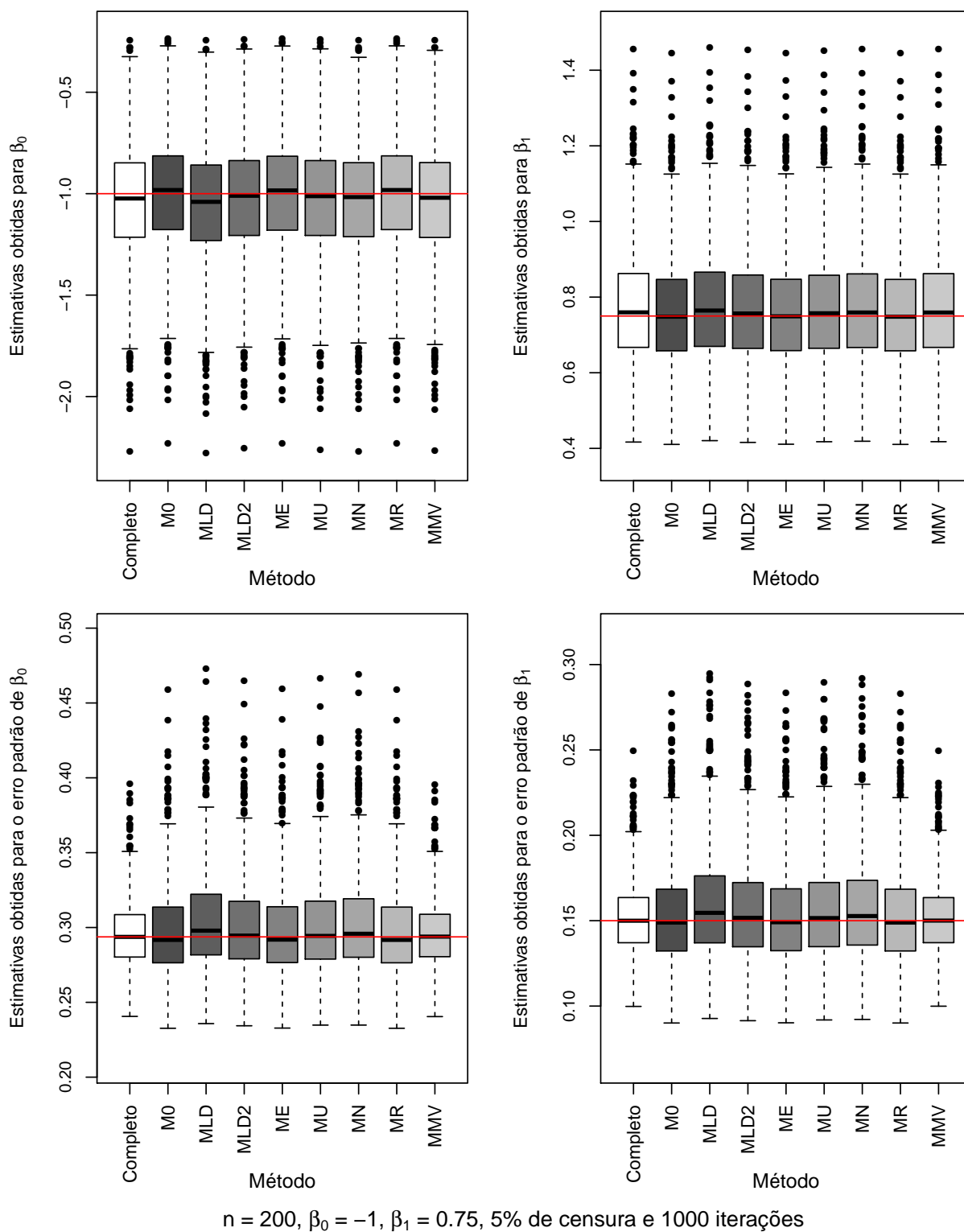
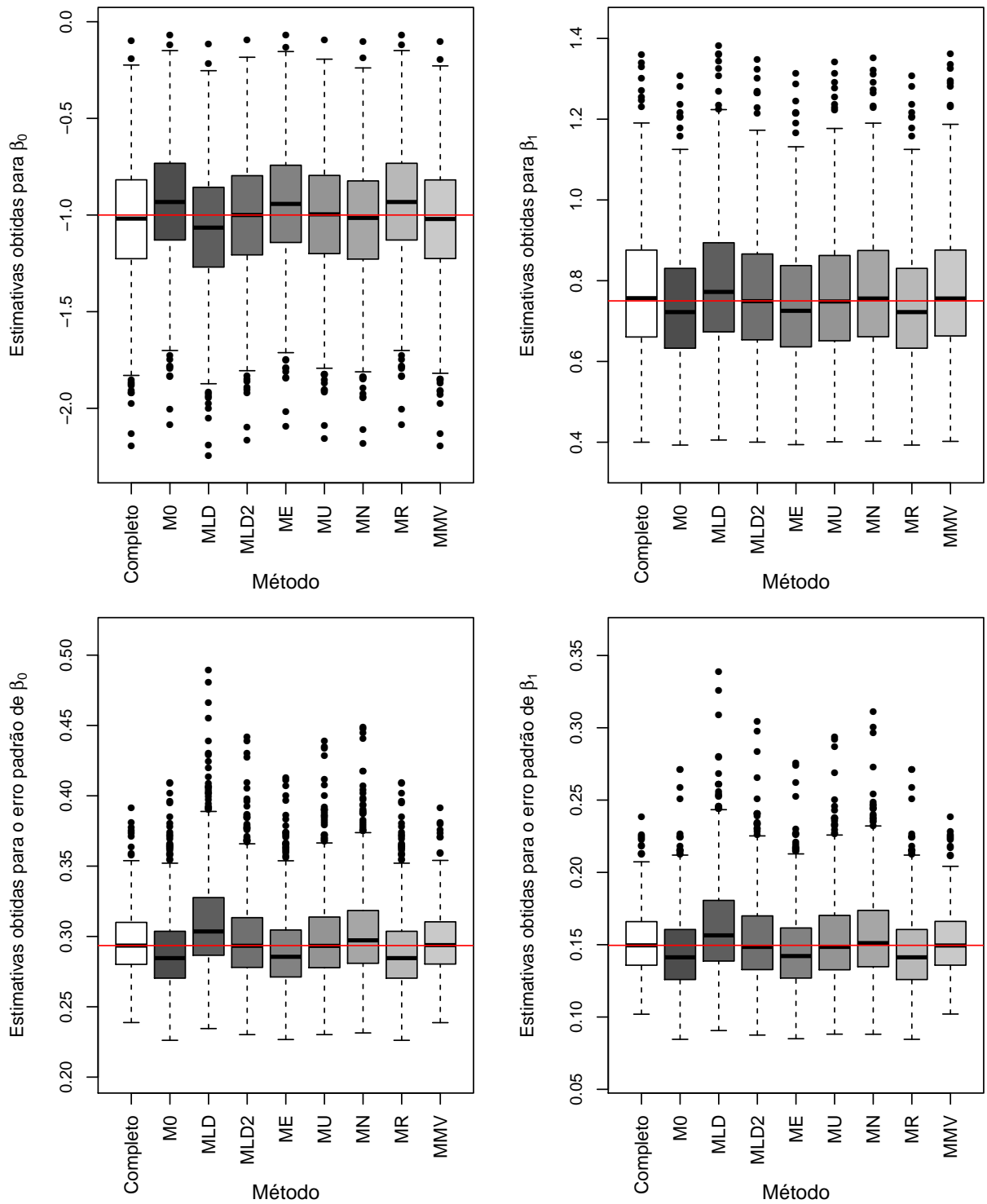


Figura C.21: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 5% de censura e $\beta_1 = 0,75$.



$n = 200, \beta_0 = -1, \beta_1 = 0.75, 10\%$ de censura e 1000 iterações

Figura C.22: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 10% de censura e $\beta_1 = 0,75$.

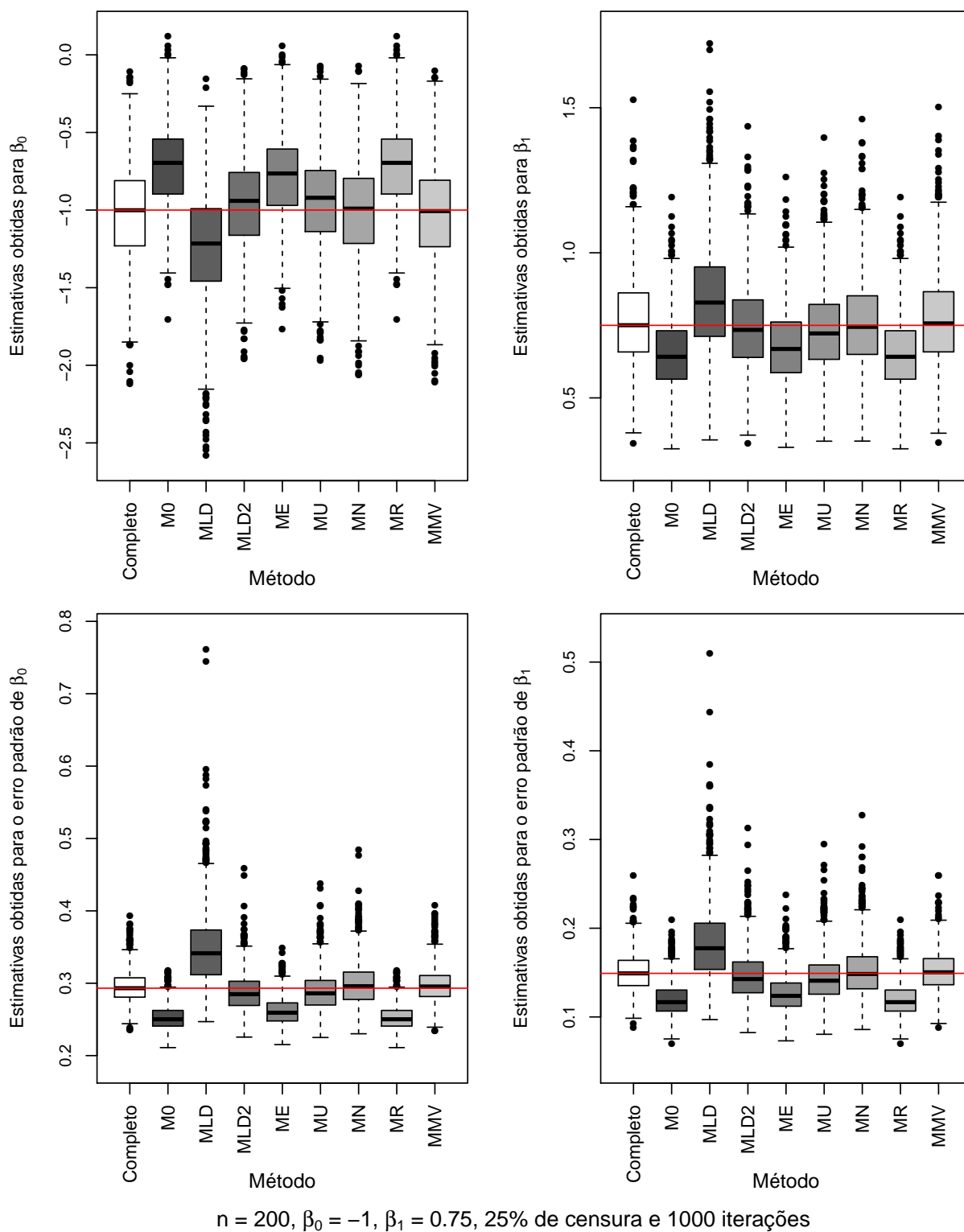
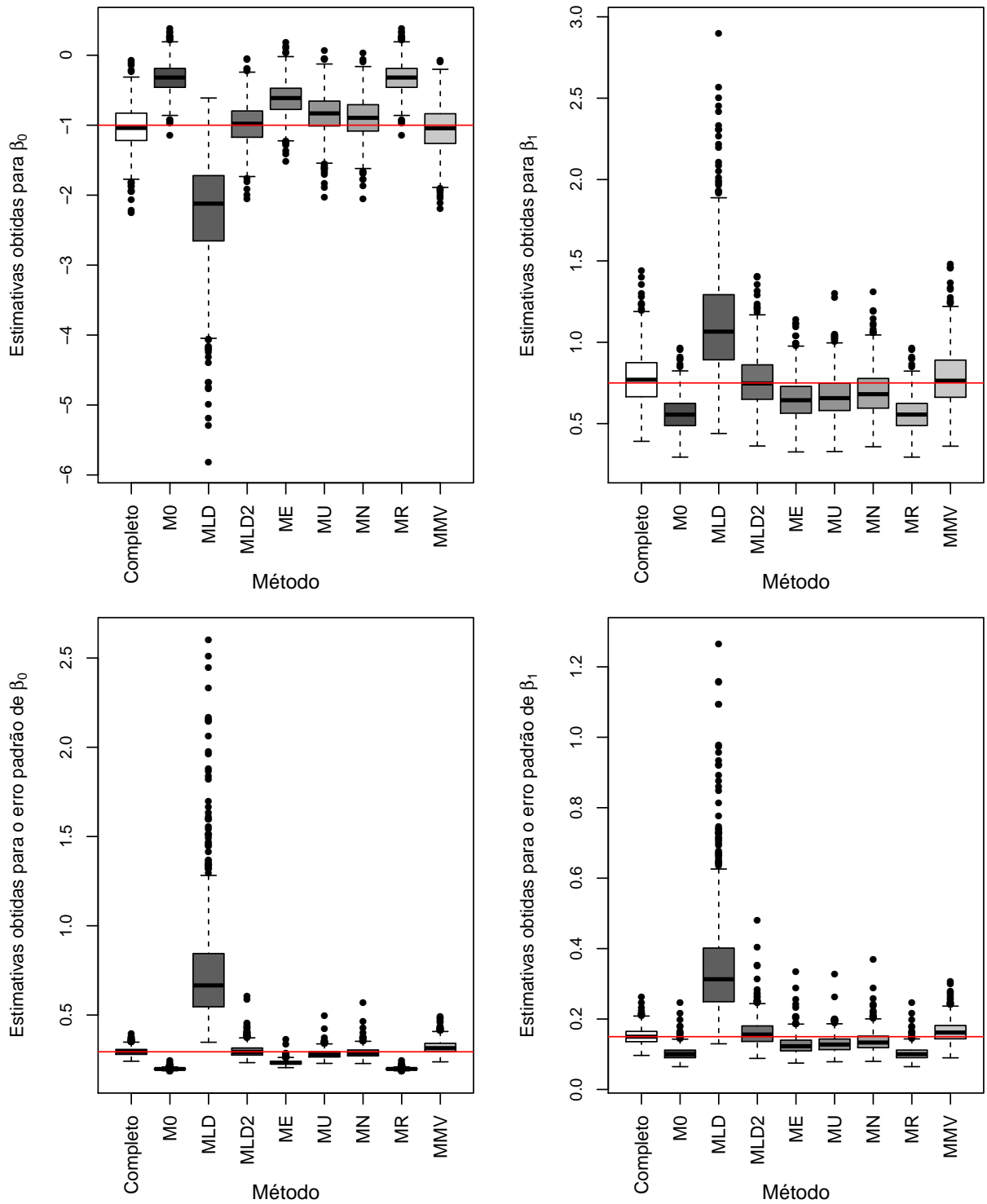


Figura C.23: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 25% de censura e $\beta_1 = 0,75$.



$n = 200, \beta_0 = -1, \beta_1 = 0.75, 50\%$ de censura e 1000 iterações

Figura C.24: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 50% de censura e $\beta_1 = 0,75$.

Tabela C.5: Vieses relativos percentuais (erros padrão) de $\hat{\beta}_0$ para $\beta_1 = 1, 25$

n	% Censura	Método								
		Completo	M0	MLD	MLD2	ME	MU	MN	MR	MMV
50	5%	8,4% (2,3%)	-4,4% (2,2%)	12,5% (2,4%)	4,0% (2,3%)	-3,7% (2,3%)	3,8% (2,3%)	7,5% (2,3%)	-4,4% (2,2%)	7,3% (2,3%)
	10%	6,1% (2,3%)	-18,2% (2,1%)	18,3% (2,5%)	-0,6% (2,2%)	-15,7% (2,1%)	-1,4% (2,2%)	5,5% (2,3%)	-18,2% (2,1%)	6,1% (2,3%)
	25%	5,4% (2,4%)	-64,2% (1,7%)	71,7% (3,6%)	-13,1% (2,1%)	-50,7% (1,8%)	-18,3% (2,2%)	-2,3% (2,3%)	-64,1% (1,7%)	4,4% (2,4%)
	50%	-3,0% (2,3%)	-119,4% (1,3%)	1328,7% (310,0%)	-0,3% (5,2%)	-69,9% (2,4%)	-54,5% (1,9%)	-39,3% (2,0%)	-118,4% (1,4%)	1,5% (2,7%)
100	5%	-1,8% (1,6%)	-10,3% (1,6%)	1,6% (1,6%)	-4,2% (1,6%)	-9,9% (1,6%)	-4,4% (1,6%)	-1,9% (1,6%)	-10,3% (1,6%)	-2,1% (1,6%)
	10%	-0,9% (1,6%)	-21,7% (1,5%)	8,9% (1,7%)	-6,5% (1,5%)	-19,6% (1,5%)	-7,0% (1,5%)	-1,5% (1,6%)	-21,7% (1,5%)	-1,0% (1,6%)
	25%	-4,1% (1,5%)	-63,6% (1,2%)	47,9% (2,2%)	-17,7% (1,4%)	-51,4% (1,2%)	-22,7% (1,4%)	-9,0% (1,5%)	-63,6% (1,2%)	-3,3% (1,5%)
	50%	-2,6% (1,5%)	-118,2% (0,9%)	903,4% (384,9%)	-5,5% (2,6%)	-70,7% (1,3%)	-52,0% (1,3%)	-41,3% (1,4%)	-117,8% (0,9%)	2,8% (1,8%)
200	5%	-3,6% (1,1%)	-11,4% (1,1%)	-0,5% (1,1%)	-5,8% (1,1%)	-11,0% (1,1%)	-5,8% (1,1%)	-3,6% (1,1%)	-11,4% (1,1%)	-3,8% (1,1%)
	10%	-6,0% (1,0%)	-25,6% (1,0%)	3,0% (1,1%)	-11,2% (1,0%)	-23,6% (1,0%)	-11,8% (1,0%)	-6,6% (1,0%)	-25,6% (1,0%)	-6,6% (1,0%)
	25%	-4,8% (1,1%)	-63,1% (0,8%)	43,9% (1,5%)	-17,9% (1,0%)	-51,0% (0,9%)	-23,4% (1,0%)	-10,0% (1,1%)	-63,1% (0,8%)	-4,7% (1,1%)
	50%	-4,8% (1,1%)	-118,3% (0,7%)	292,7% (6,7%)	-11,7% (1,1%)	-72,8% (0,8%)	-54,4% (0,9%)	-41,8% (1,0%)	-118,2% (0,7%)	-2,8% (1,3%)

Nota: Resultados obtidos com 1000 iterações. n: tamanho da amostra.

Tabela C.6: Vieses relativos percentuais (erros padrão) de $\hat{\beta}_1$ para $\beta_1 = 1, 25$

n	% Censura	Método								
		Completo	M0	MLD	MLD2	ME	MU	MN	MR	MMV
50	5%	8,0% (1,4%)	1,6% (1,2%)	9,9% (1,4%)	5,9% (1,3%)	2,0% (1,2%)	5,6% (1,3%)	7,5% (1,3%)	1,6% (1,2%)	7,4% (1,3%)
	10%	6,7% (1,3%)	-5,0% (1,1%)	12,2% (1,4%)	3,5% (1,2%)	-3,8% (1,1%)	2,9% (1,2%)	6,2% (1,3%)	-5,0% (1,1%)	6,5% (1,3%)
	25%	6,1% (1,3%)	-22,8% (0,8%)	33,7% (2,1%)	-1,6% (1,2%)	-17,2% (0,9%)	-5,6% (1,1%)	1,5% (1,3%)	-22,7% (0,8%)	5,6% (1,3%)
	50%	-0,6% (1,2%)	-31,6% (2,1%)	510,2% (130,6%)	6,1% (4,0%)	-16,4% (2,8%)	-30,7% (0,8%)	-23,5% (0,9%)	-25,6% (10,0%)	-0,1% (1,3%)
100	5%	1,1% (0,9%)	-3,0% (0,8%)	2,6% (0,9%)	0,0% (0,9%)	-2,8% (0,8%)	-0,2% (0,9%)	1,0% (0,9%)	-3,0% (0,8%)	0,8% (0,9%)
	10%	0,6% (0,9%)	-9,0% (0,8%)	4,7% (1,0%)	-1,9% (0,8%)	-8,0% (0,8%)	-2,3% (0,8%)	0,2% (0,9%)	-9,0% (0,8%)	0,5% (0,9%)
	25%	-0,9% (0,8%)	-24,7% (0,6%)	18,8% (1,2%)	-6,3% (0,8%)	-19,8% (0,6%)	-10,1% (0,7%)	-4,3% (0,8%)	-24,7% (0,6%)	-0,6% (0,9%)
	50%	-0,5% (0,8%)	-33,5% (1,0%)	322,6% (143,4%)	0,4% (1,9%)	-19,4% (1,3%)	-29,5% (0,6%)	-25,2% (0,6%)	-34,0% (0,8%)	1,4% (1,0%)
200	5%	-1,9% (0,6%)	-5,6% (0,6%)	-0,6% (0,6%)	-2,9% (0,6%)	-5,4% (0,6%)	-3,0% (0,6%)	-2,0% (0,6%)	-5,6% (0,6%)	-2,1% (0,6%)
	10%	-2,6% (0,6%)	-11,4% (0,5%)	1,1% (0,6%)	-4,8% (0,5%)	-10,5% (0,5%)	-5,3% (0,5%)	-2,9% (0,6%)	-11,4% (0,5%)	-3,0% (0,6%)
	25%	-2,3% (0,6%)	-25,3% (0,4%)	15,4% (0,8%)	-7,4% (0,5%)	-20,5% (0,4%)	-11,3% (0,5%)	-5,8% (0,5%)	-25,3% (0,4%)	-2,4% (0,6%)
	50%	-2,6% (0,6%)	-36,1% (0,4%)	84,1% (2,7%)	-5,2% (0,7%)	-23,1% (0,5%)	-31,7% (0,4%)	-26,2% (0,4%)	-36,1% (0,4%)	-2,6% (0,7%)

Nota: Resultados obtidos com 1000 iterações. n: tamanho da amostra.

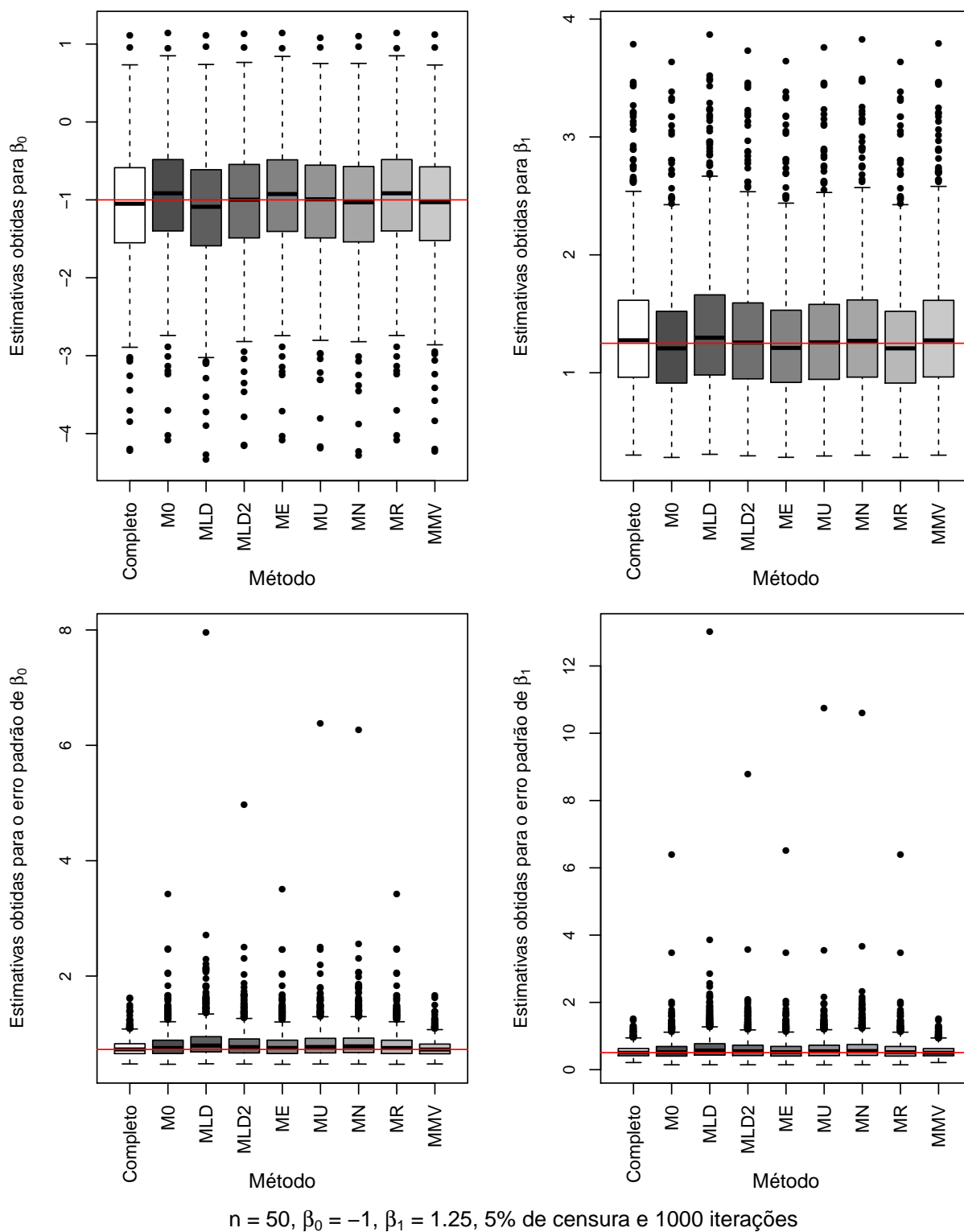
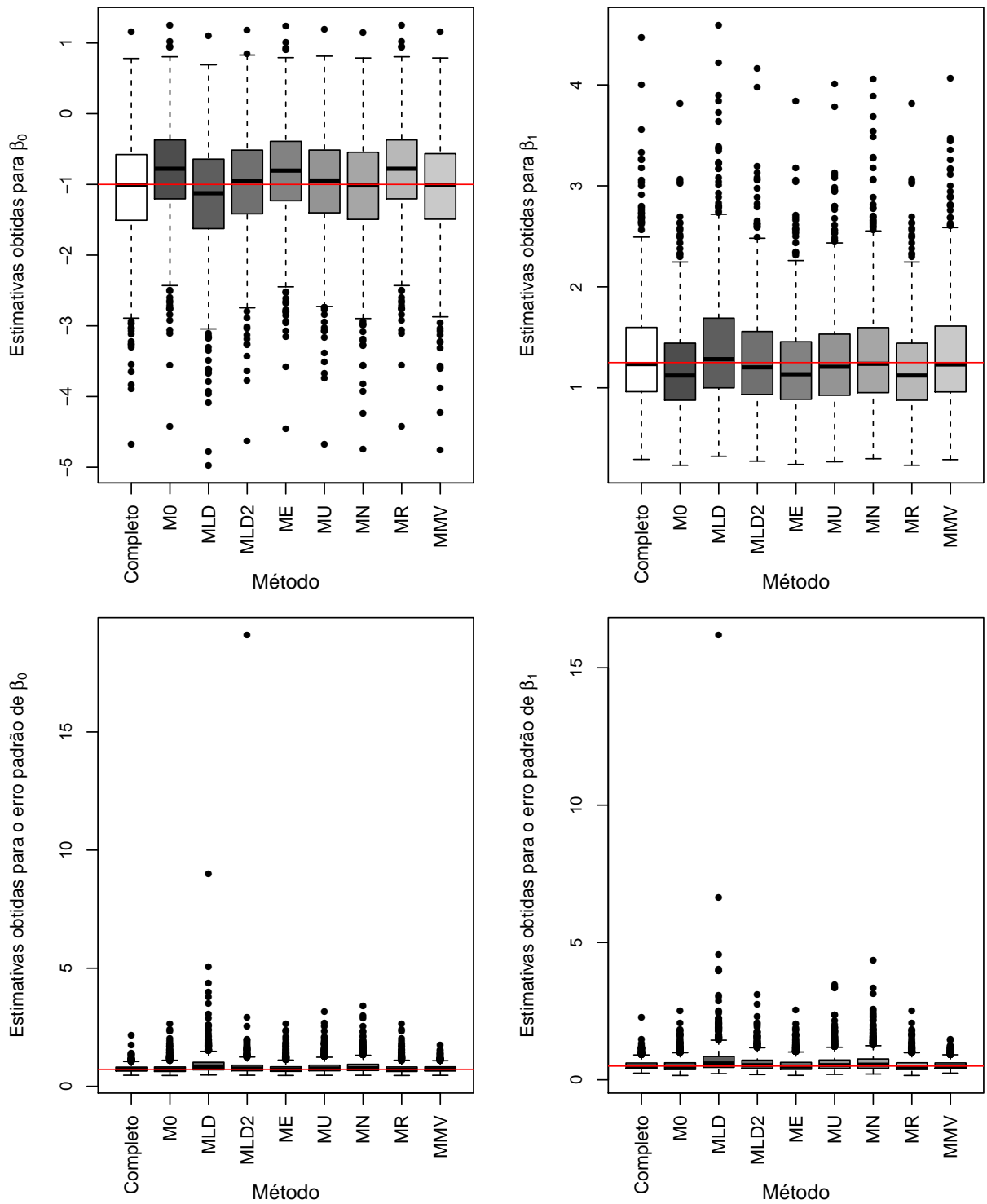


Figura C.25: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 5% de censura e $\beta_1 = 1,25$.



$n = 50, \beta_0 = -1, \beta_1 = 1.25, 10\%$ de censura e 1000 iterações. Dados truncados em $[-20,20]$

Figura C.26: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 10% de censura e $\beta_1 = 1,25$.

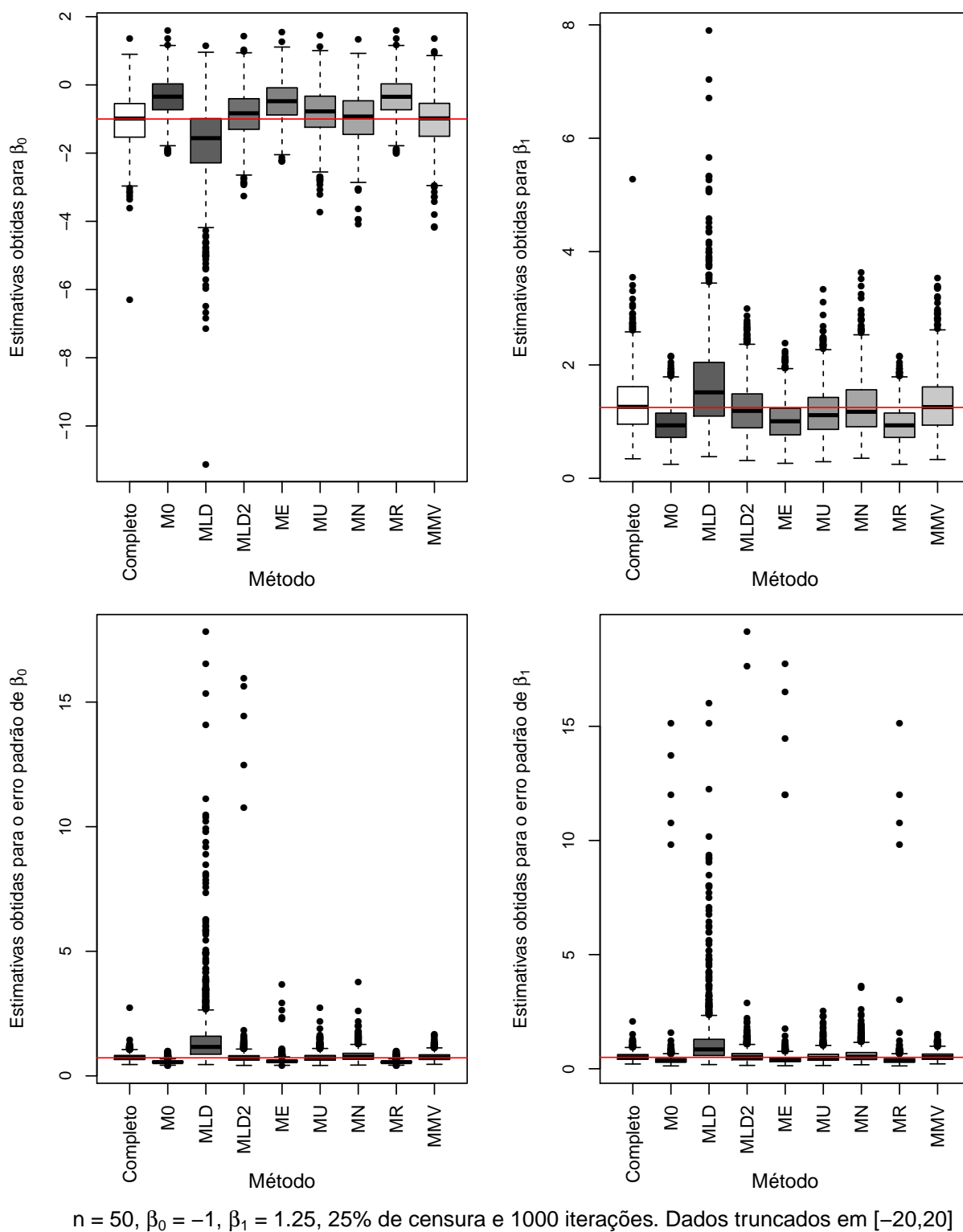
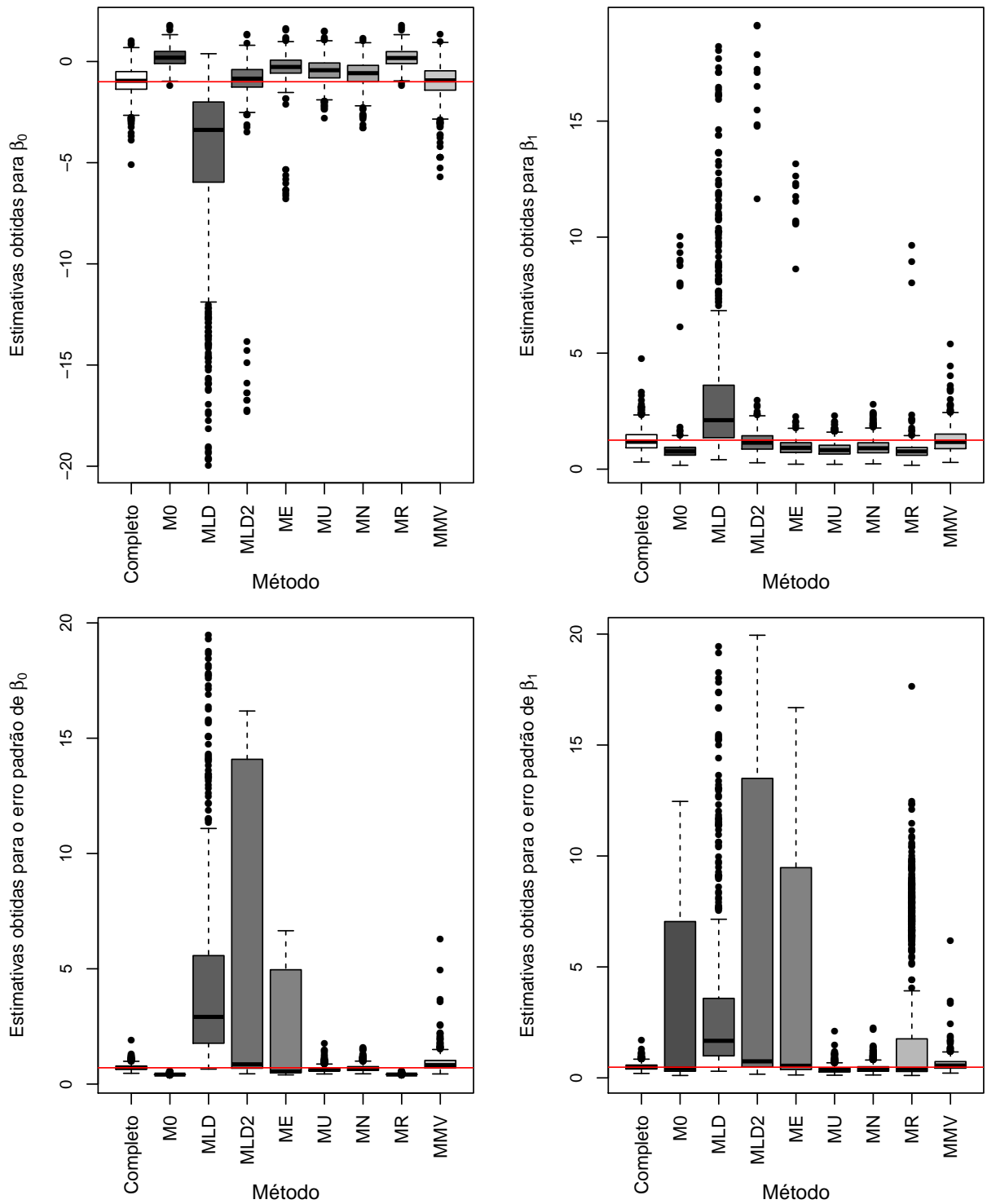
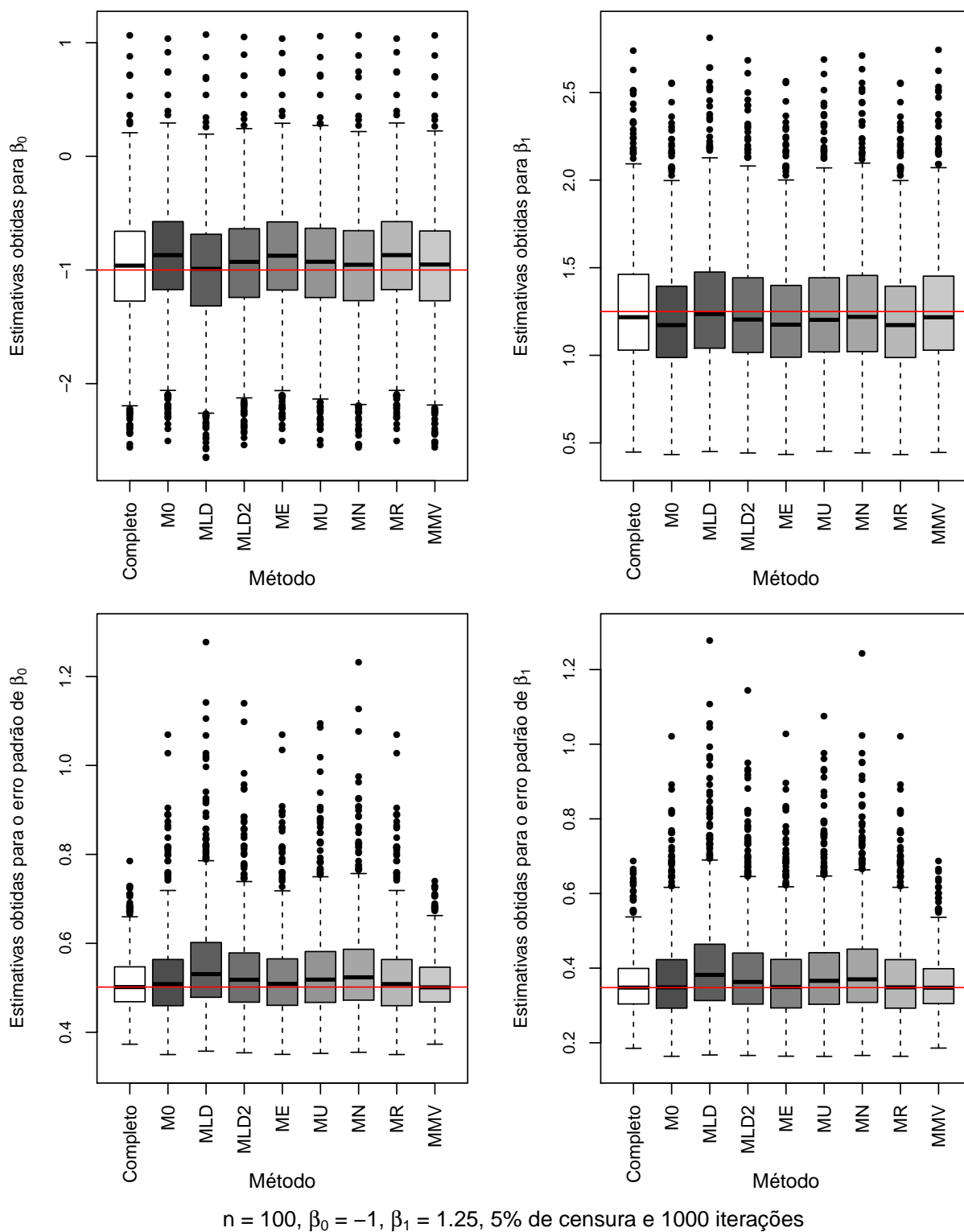


Figura C.27: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 25% de censura e $\beta_1 = 1,25$.



$n = 50, \beta_0 = -1, \beta_1 = 1.25, 50\%$ de censura e 1000 iterações. Dados truncados em $[-20,20]$

Figura C.28: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 50% de censura e $\beta_1 = 1,25$.



$n = 100, \beta_0 = -1, \beta_1 = 1.25, 5\%$ de censura e 1000 iterações

Figura C.29: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 5% de censura e $\beta_1 = 1,25$.

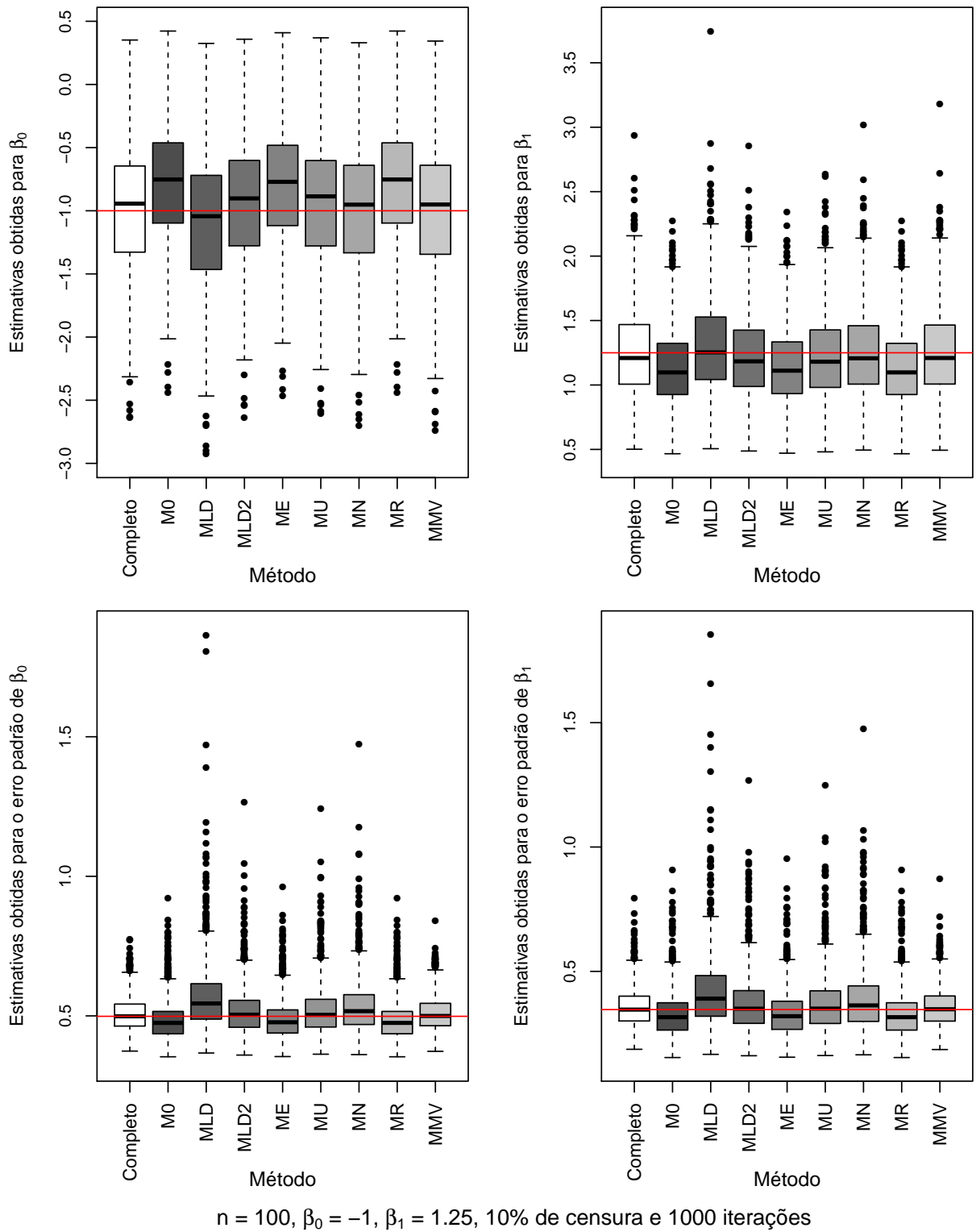


Figura C.30: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 10% de censura e $\beta_1 = 1,25$.

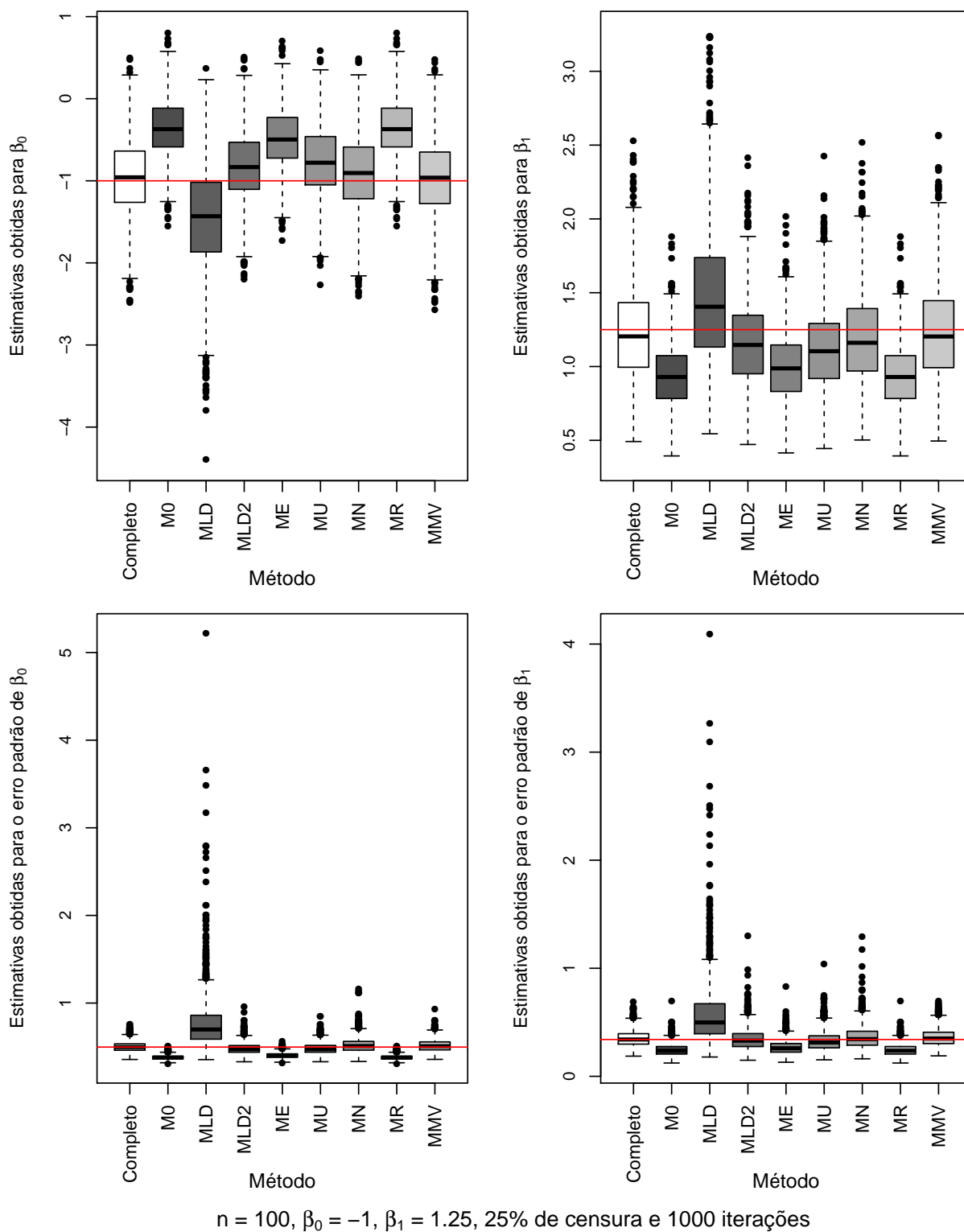
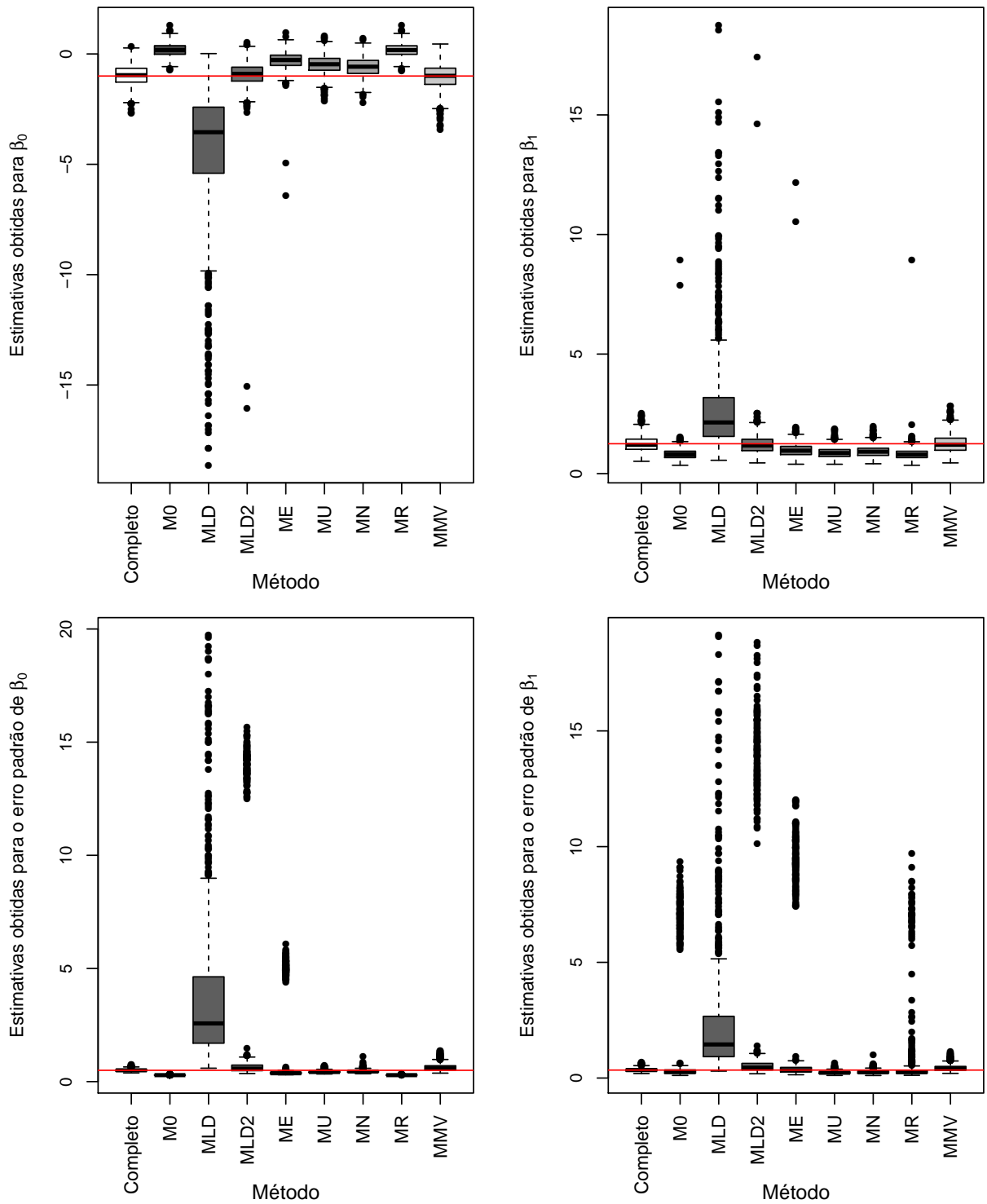


Figura C.31: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 25% de censura e $\beta_1 = 1,25$.



$n = 100, \beta_0 = -1, \beta_1 = 1.25, 50\%$ de censura e 1000 iterações. Dados truncados em $[-20,20]$

Figura C.32: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 50% de censura e $\beta_1 = 1,25$.

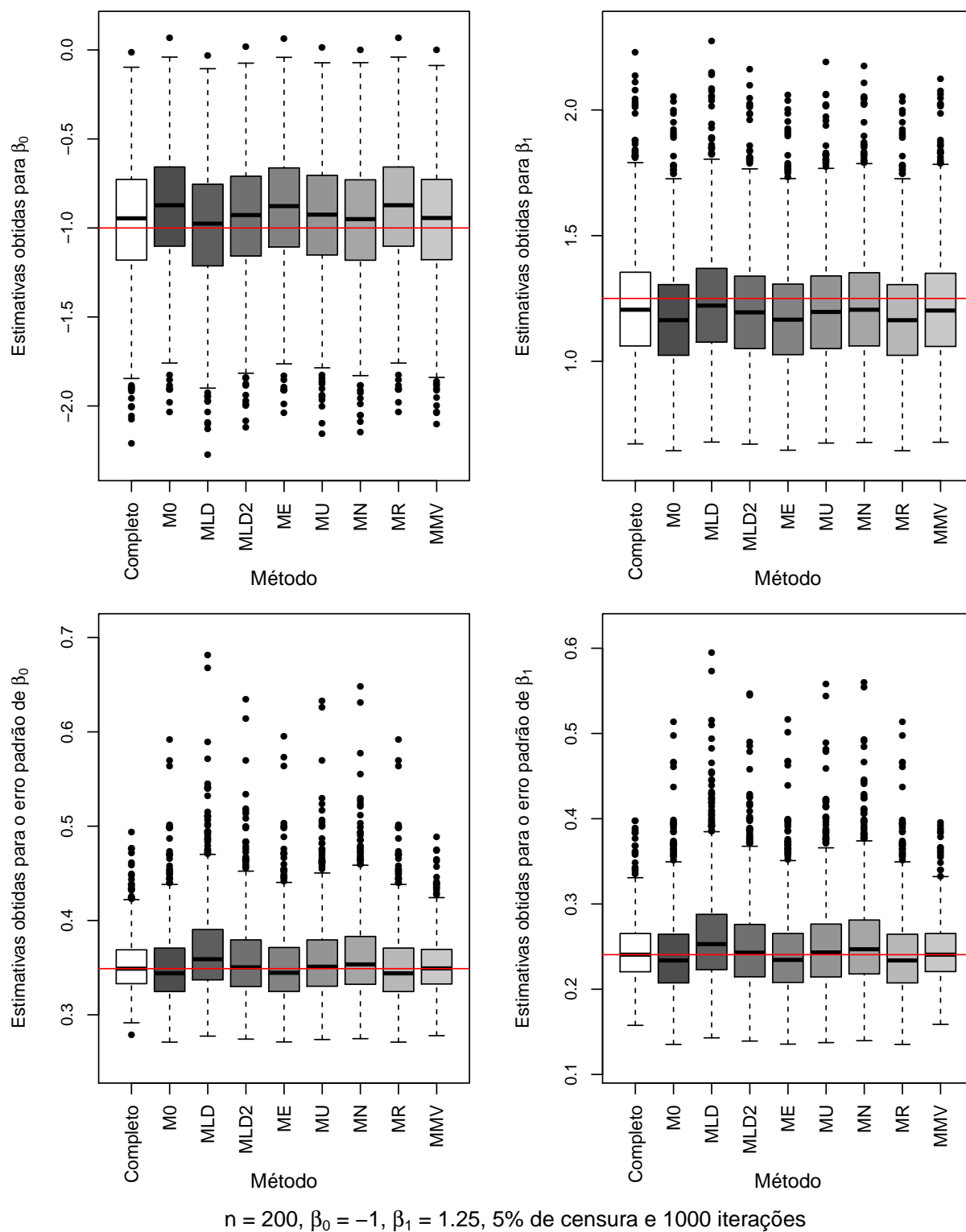


Figura C.33: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 5% de censura e $\beta_1 = 1,25$.

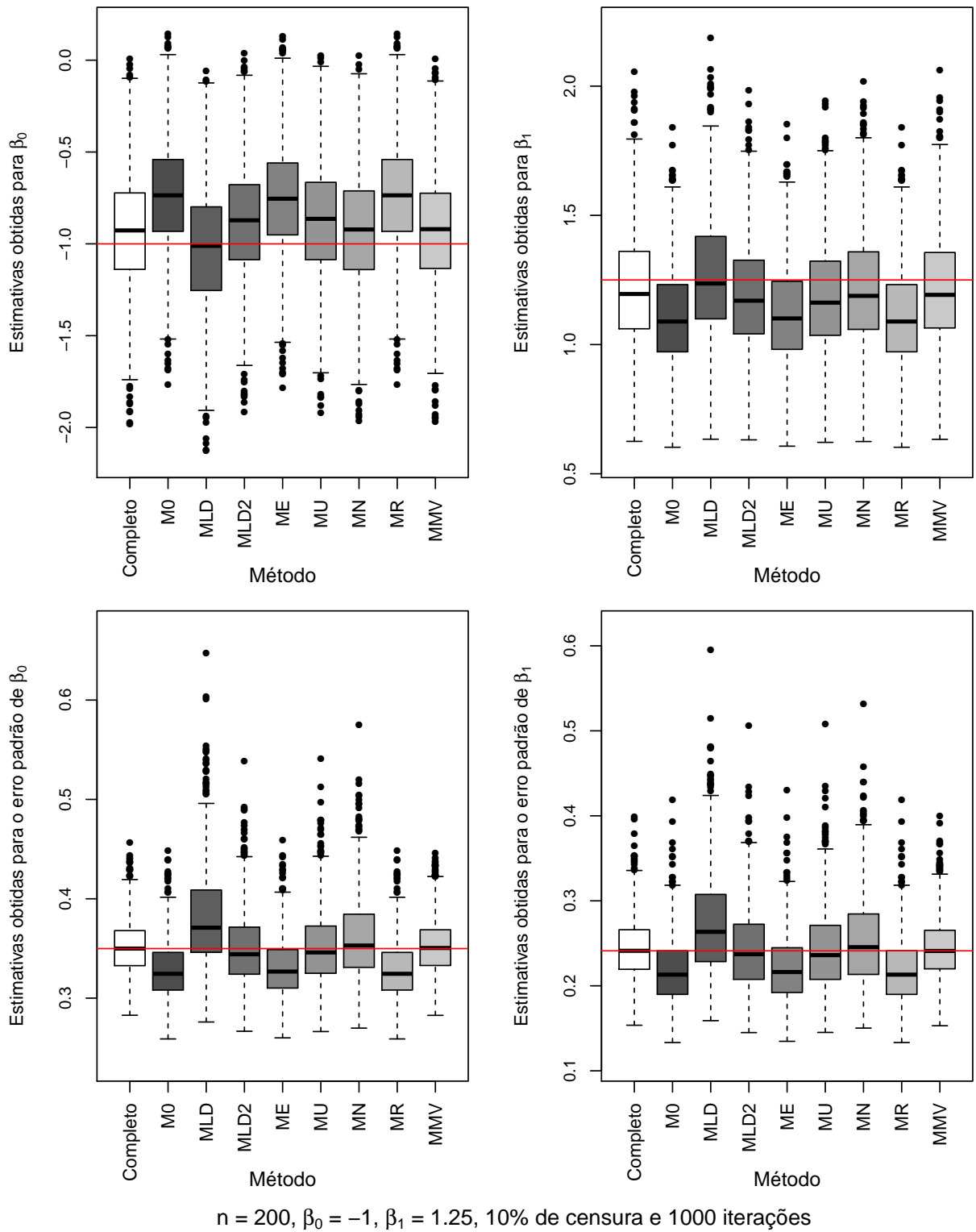


Figura C.34: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 10% de censura e $\beta_1 = 1,25$.

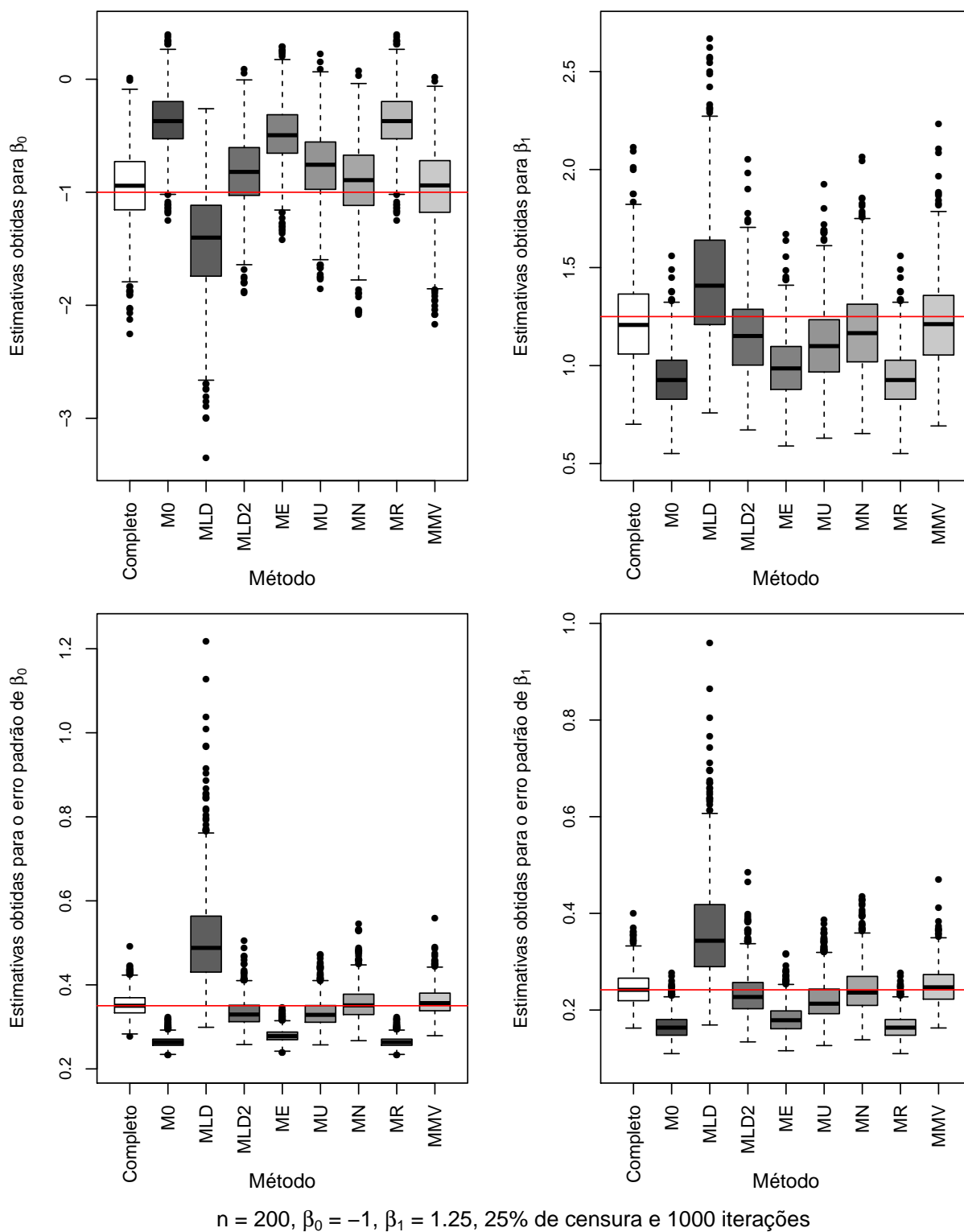
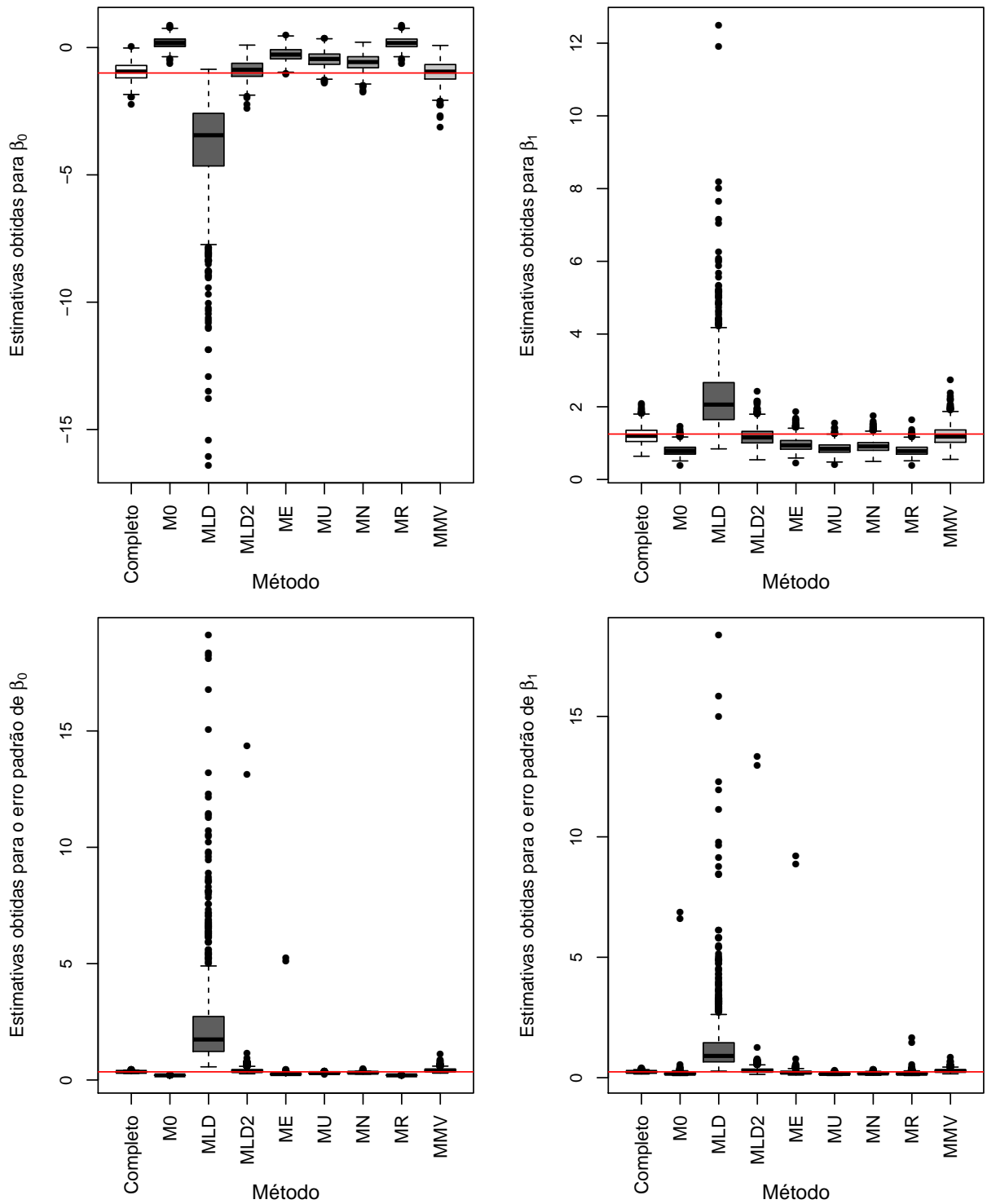


Figura C.35: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 25% de censura e $\beta_1 = 1,25$.



$n = 200, \beta_0 = -1, \beta_1 = 1.25, 50\%$ de censura e 1000 iterações. Dados truncados em $[-20,20]$

Figura C.36: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 50% de censura e $\beta_1 = 1,25$.

Tabela C.7: Vieses relativos percentuais (erros padrão) de $\hat{\beta}_0$ e $\hat{\beta}_1$ para $\beta_1 = 1, 25$

n	% censura	Parâmetro	Método			
			Completo	ME	MR	MMV
50	25%	β_0	0.3% (2.4%)	-52.4% (1.9%)	-1.9% (2.4%)	0.2% (2.4%)
		β_1	2.9% (1.2%)	-18.6% (0.9%)	1.1% (1.2%)	2.9% (1.3%)
	50%	β_0	1.3% (2.2%)	-64.6% (2.3%)	-55.9% (1.8%)	4.2% (2.4%)
		β_1	-0.1% (1.2%)	-13.7% (3.1%)	-25.9% (0.8%)	0.0% (1.2%)
100	25%	β_0	-1.1% (1.6%)	-48.4% (1.3%)	-2.6% (1.6%)	-0.3% (1.6%)
		β_1	0.1% (0.8%)	-18.8% (0.6%)	-1.2% (0.8%)	0.4% (0.9%)
	50%	β_0	-1.0% (1.5%)	-69.9% (1.3%)	-61.2% (1.2%)	2.8% (1.7%)
		β_1	-0.2% (0.8%)	-19.8% (1.2%)	-27.0% (0.5%)	0.7% (0.9%)
200	25%	β_0	-1.2% (1.1%)	-48.6% (0.9%)	-2.8% (1.2%)	-0.5% (1.2%)
		β_1	-0.5% (0.6%)	-19.4% (0.4%)	-1.8% (0.6%)	-0.3% (0.6%)
	50%	β_0	-3.3% (1.1%)	-72.7% (0.9%)	-62.3% (0.9%)	-1.6% (1.3%)
		β_1	-1.4% (0.6%)	-22.2% (0.5%)	-26.8% (0.4%)	-1.2% (0.7%)

Nota: Resultados obtidos com 500 iterações. n: tamanho da amostra.

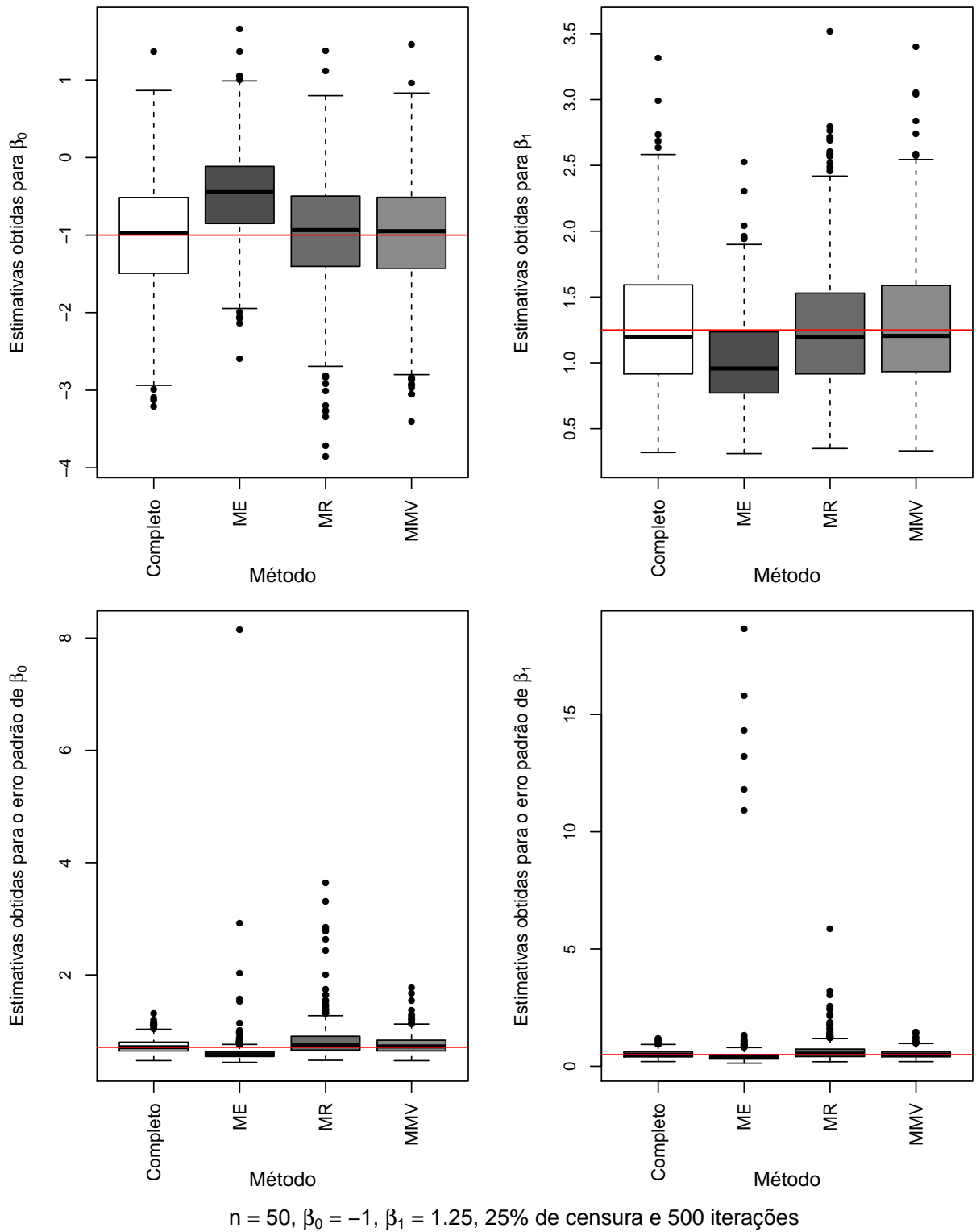


Figura C.37: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 25% de censura e $\beta_1 = 1,25$.

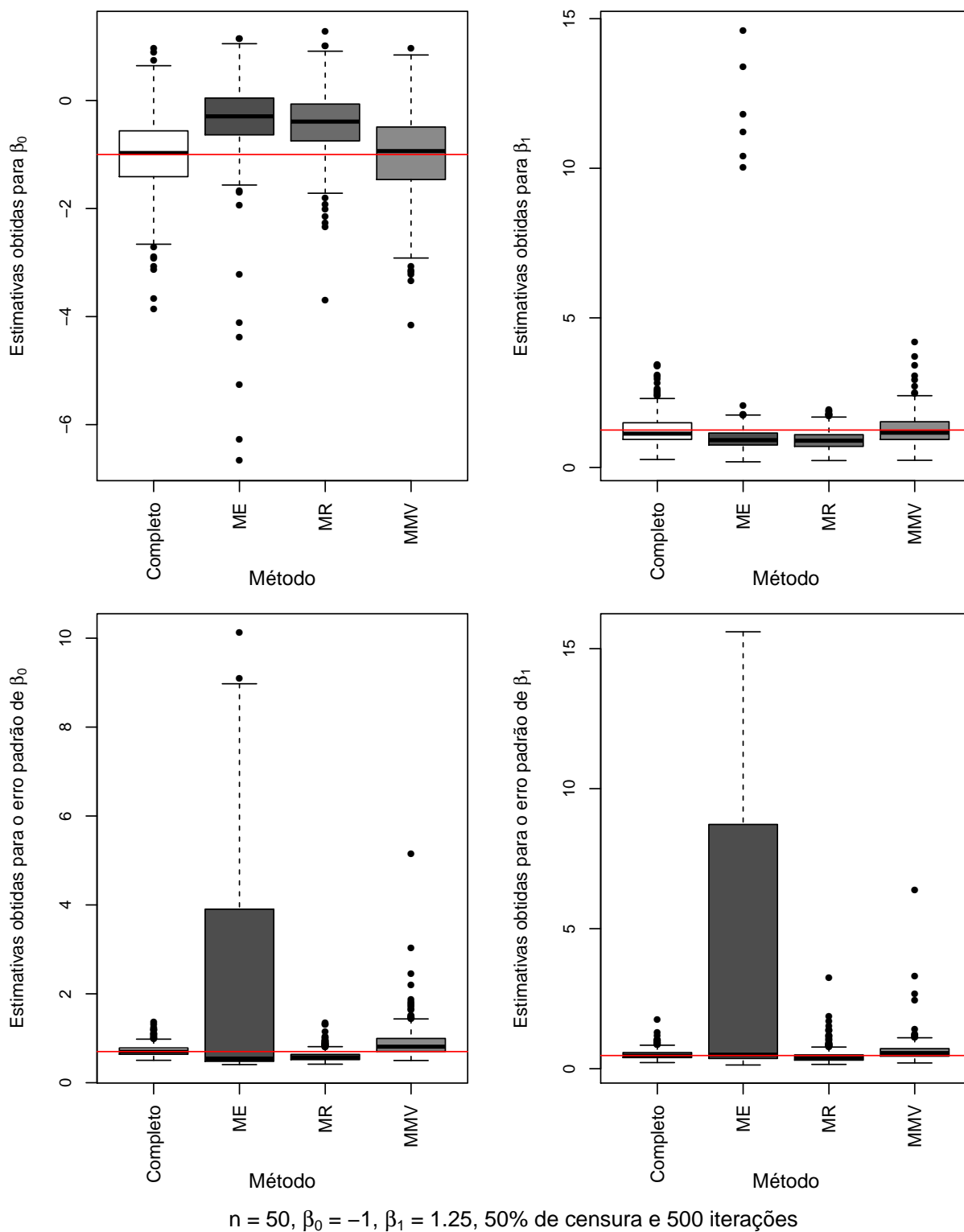


Figura C.38: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 50, 50% de censura e $\beta_1 = 1,25$.

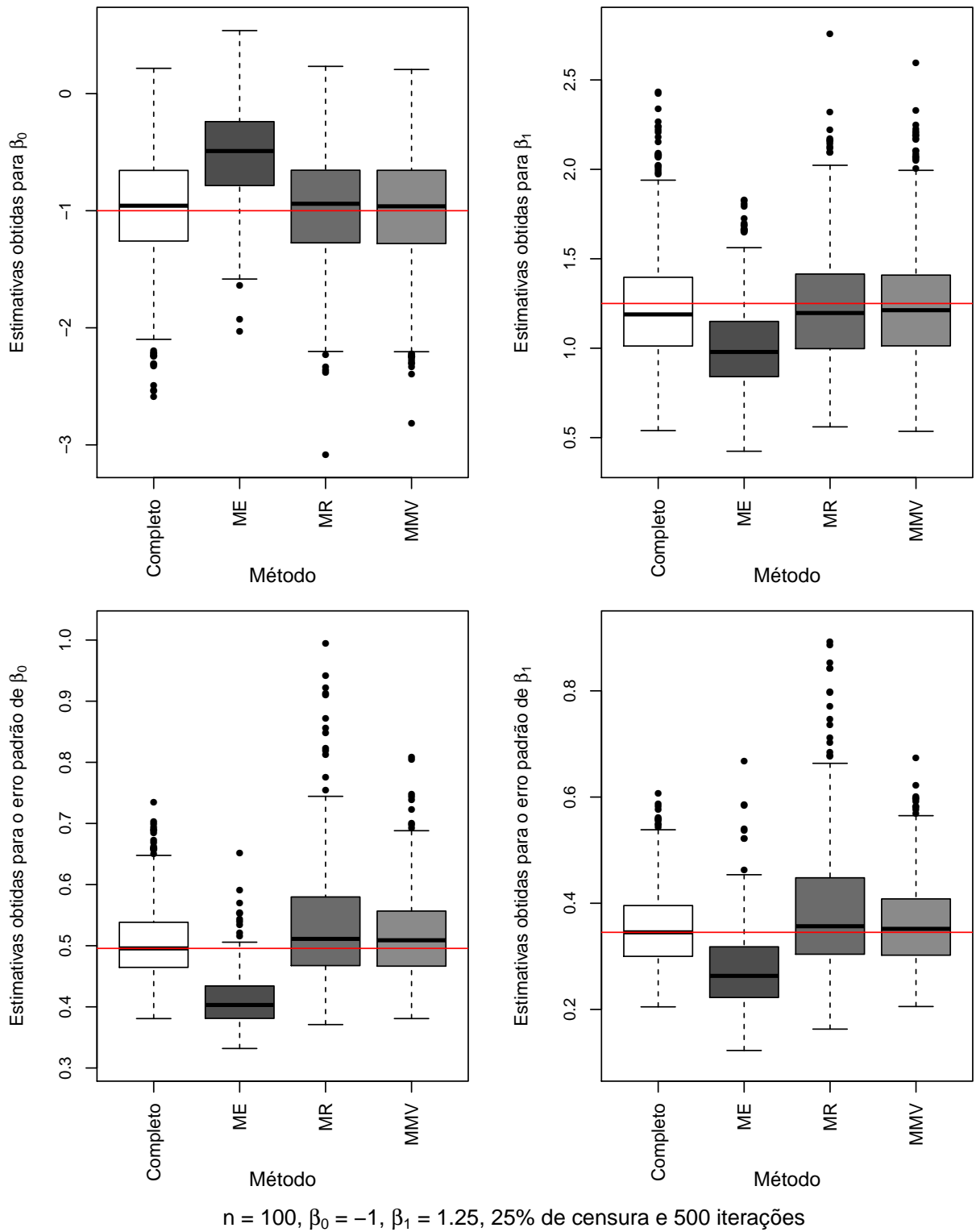


Figura C.39: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 25% de censura e $\beta_1 = 1,25$.

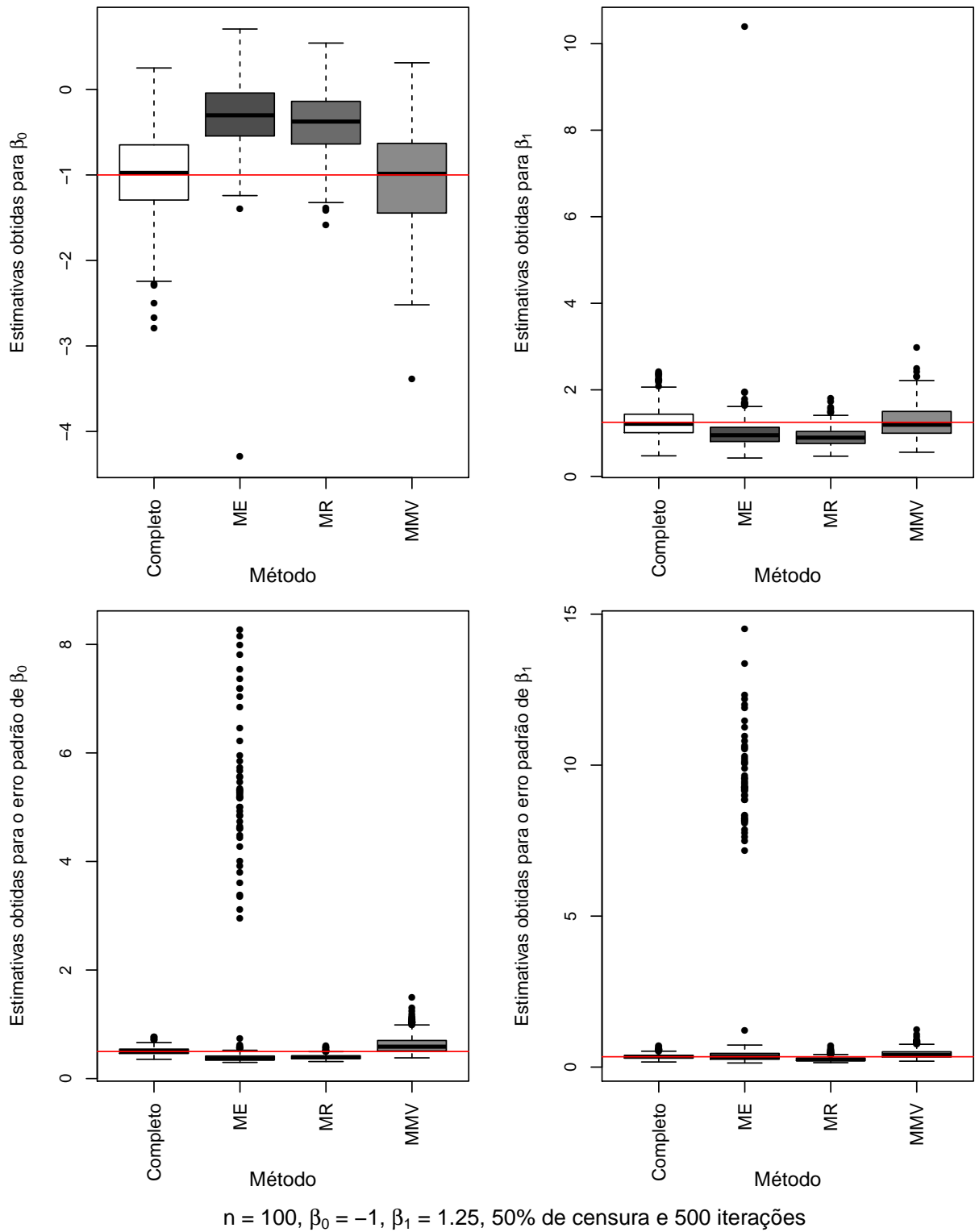


Figura C.40: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 100, 50% de censura e $\beta_1 = 1,25$.

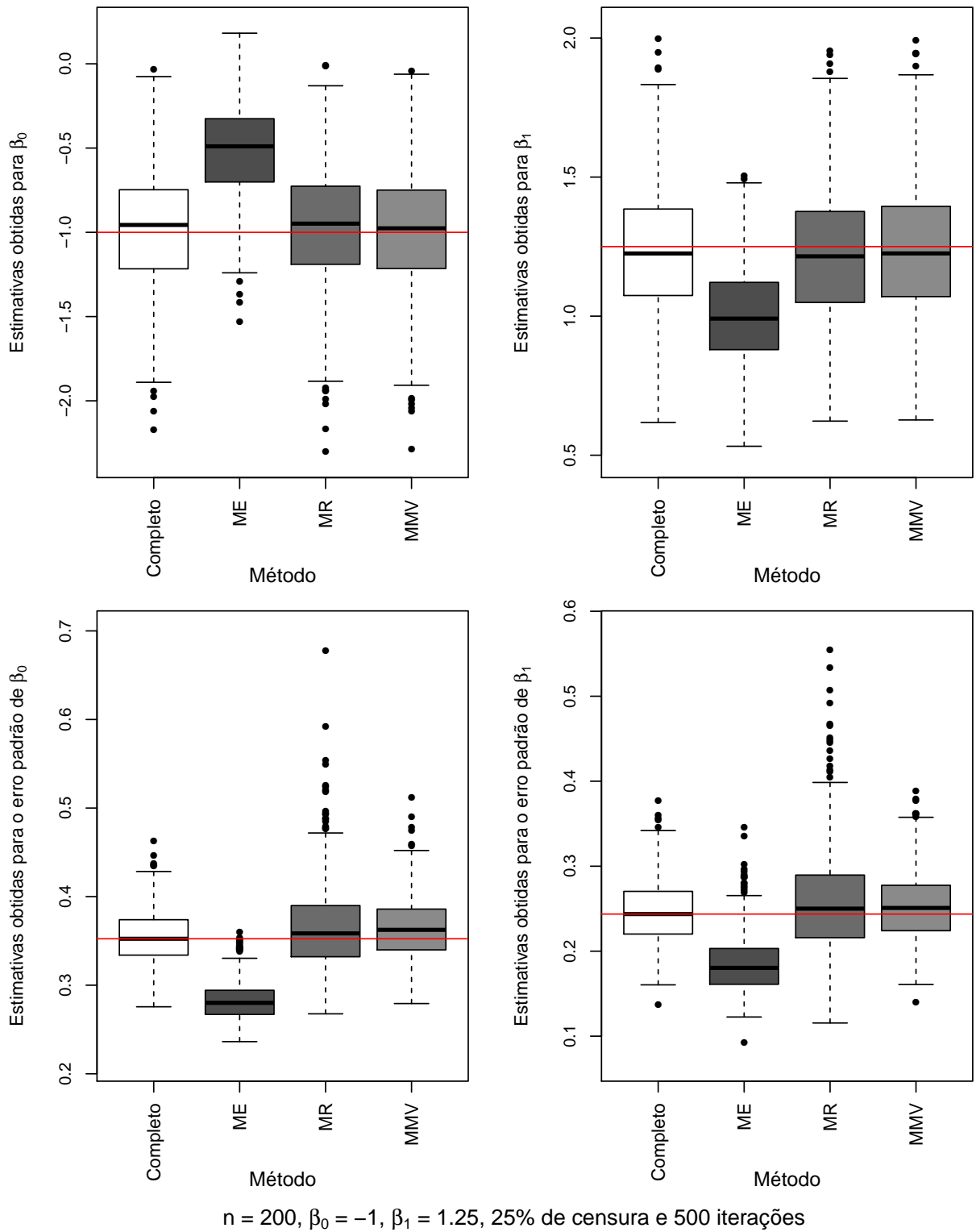


Figura C.41: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 25% de censura e $\beta_1 = 1,25$.

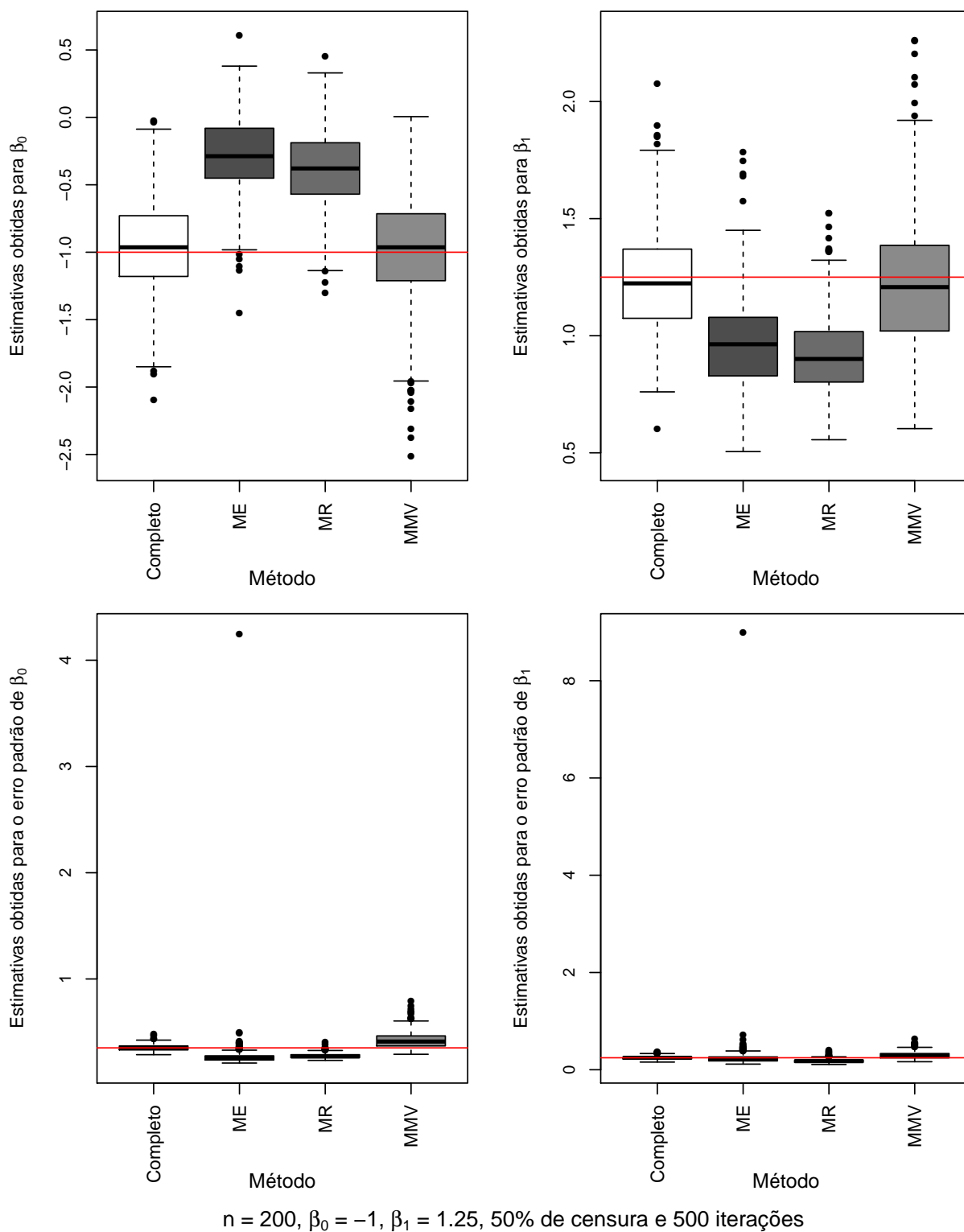


Figura C.42: Valores obtidos para $\hat{\beta}_0$ e $\hat{\beta}_1$ considerando amostras de tamanho 200, 50% de censura e $\beta_1 = 1,25$.

Referências Bibliográficas

- Abrão, M. S., Podgaec, S., Martorelli Filho, B., Ramos, L. O., Pinotti, J. A. and Oliveira, R. M. (1997). The use of biochemical markers in the diagnosis of pelvic endometriosis. *Human Reproduction* **12**, 2523–2527. [2](#), [11](#), [29](#), [40](#)
- Aitchison, J. (1955). On the distribution of a positive random variable having a discrete Probability mass at the origin. *Journal of the american statistical association* **50**, 901–908. [5](#)
- Beal, S. L. (2001). Ways to fit a PK model with some data below the quantification limit. *Journal of Pharmacokinetics and Pharmacodynamics* **28**, 481–504. [5](#), [6](#), [7](#)
- Bellelis, P., Dias Júnior, J., Podgaec, S., Gonzales, M., Baracat, E. and Abrão, M. (2010). Aspectos epidemiológicos e clínicos da endometriose pélvica: uma série de casos. *Revista da Associação Médica Brasileira* **56**, 467–71. [1](#)
- Burton, A., Altman, D. G., Royston, P. and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine* **25**, 4279–4292. [19](#)
- Carvalho, L. F. P. (2012). Avaliação dos marcadores de estresse oxidativo em pacientes com endometriose pélvica. Tese de doutorado – programa de ginecologia e obstetrícia da faculdade de medicina. Universidade de São Paulo. [2](#)
- Cole, S. R., Chu, H., Nie, L. and Schisterman, E. F. (2009). Estimating the odds ratio when exposure has a limit of detection. *International Journal of Epidemiology* **38**, 1674–1680. [9](#), [10](#), [15](#), [17](#)
- FEBRASGO (2010). *Diagnóstico laboratorial da endometriose. Manual de orientação*

- em Endometriose*. Federação Brasileira das Associações de Ginecologia e Obstetrícia. <http://www.febrasgo.org.br> [acessado em 04/ 02/2014]. chapter 7, pp. 43–47. 2
- Gilliom, R. J. and Helsel, D. R. (1986). Estimation of distributional parameters for censored trace level water quality data: 1. Estimation techniques. *Water Resources Research* **22**, 135–146. 4, 5, 14, 19
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables (Advanced Quantitative Techniques in the Social Sciences)*. Vol. 7. Sage Publications, USA.. Thousand Oaks. 3
- Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal* **7**, 308–313. 17
- Podgaec, S. and Abrão, M. S. (2004). Endometriose: aspectos atuais do diagnóstico e tratamento. *Rev Bras Med* **61**, 41–6. 1
- Prasad, A. (1955). Bi-modal distributions derived from the normal distribution. *Sankhyā: The Indian Journal of Statistics* **14**, 369–374. 7
- Rao, J. (1996). On variance estimation with imputed survey data. *Journal of the American Statistical Association* **91**, 499–506. 10, 13
- Ricci, V. (2005). *Fitting distributions with R*. Contributed Documentation available on CRAN. <http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf> [acessado em 20/ 08/2014]. 14
- Schisterman, E. F., Vexler, A., Whitcomb, B. W. and Liu, A. (2006). The limitations due to exposure detection limits for regression models. *American Journal of Epidemiology* **163**, 374–383. 7, 10, 14
- Senn, S., Holford, N. and Hockey, H. (2012). The ghosts of departed quantities: approaches to dealing with observations below the limit of quantitation. *Statistics in Medicine* **31**, 4280–4295. 7