

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

**A NECESSIDADE DE CLASSIFICAÇÕES REPETIDAS NO  
MODELO DE REGRESSÃO LOGÍSTICA COM ERROS NA  
VARIÁVEL RESPOSTA**

**DANILO GILBERTO DE OLIVEIRA VALADARES**

Belo Horizonte, 04 de fevereiro de 2019

Danilo Gilberto de Oliveira Valadares

**A NECESSIDADE DE CLASSIFICAÇÕES REPETIDAS NO  
MODELO DE REGRESSÃO LOGÍSTICA COM ERROS NA  
VARIÁVEL RESPOSTA**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Estatística.

Áreas de concentração: Modelos de regressão e aplicações.

Orientador: Prof. Dr. Roberto da Costa Quinino

Co-orientadora: Prof.<sup>a</sup>. Dr.<sup>a</sup>. Magda Carvalho Pires

Belo Horizonte

Instituto de Ciências Exatas da UFMG,

2019

# **A NECESSIDADE DE CLASSIFICAÇÕES REPETIDAS NO MODELO DE REGRESSÃO LOGÍSTICA COM ERROS NA VARIÁVEL RESPOSTA**

**Danilo Gilberto de Oliveira Valadares**

Dissertação defendida em 04, de fevereiro de 2019, pela Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais, como parte dos requisitos necessários à obtenção do título de “**Mestre em Estatística**”.

---

Prof. Dr. Roberto da Costa Quinino – UFMG – Orientador

---

Prof.<sup>a</sup>. Dr.<sup>a</sup>. Magda Carvalho Pires – UFMG – Co-Orientadora

---

Prof. Dr. Frederico Rodrigues Borges da Cruz – UFMG – Examinador

---

Prof.<sup>a</sup>. Dr.<sup>a</sup>. Linda Lee Ho – USP – Examinador

“É extraordinário que esta ciência, que surgiu da análise dos jogos de azar, tenha se tornado o mais importante objeto do conhecimento humano. (...) as mais importantes questões da vida são na verdade, em sua grande maioria, apenas problemas de probabilidade.”

(Laplace, Pierre-Simon, 1812)

## **AGRADECIMENTOS**

O presente trabalho não poderia chegar a um porto sem a preciosa ajuda e paciência de várias pessoas.

Em primeiro lugar, não posso deixar de agradecer ao meu orientador, Professor Dr. Roberto Quinino, por toda a paciência, empenho e sentido prático com que sempre me orientou durante todo o mestrado e antes mesmo do início dele. Afirmo, sem medo, que sem a sua ajuda, o caminho teria sido muito mais tortuoso.

Desejo igualmente agradecer à minha co-orientadora Prof.<sup>a</sup> Dr.<sup>a</sup> Magda Pires, por todas as sugestões e correções, sempre pertinentes.

Por último, quero agradecer à minha família, aos amigos e aos colegas pelo apoio e amizade incondicionais que sempre recebi.

Meus sinceros agradecimentos.

## RESUMO

Estimadores de máxima verossimilhança de um modelo de regressão logística com erros de classificação na variável resposta são, reconhecidamente, viciados quando os erros são ignorados. A introdução de parâmetros de erros de classificação na função de verossimilhança pode resolver, satisfatoriamente, o problema do vício. Porém, conforme discutido neste trabalho, o aumento da variabilidade dos estimadores pode comprometer o processo de decisão. Assim, o problema pode ser minimizado com a introdução de informação adicional. Mostra-se que a realização de classificações repetidas na variável resposta da amostra ou em parte dela pode ser uma solução para as diminuições simultâneas do vício e da variabilidade dos estimadores.

**Palavras Chaves:** Resposta binária; regressão logística; erros de classificação; classificações repetidas.

## ABSTRACT

Maximum likelihood estimators for the logistic regression model with misclassification in the response variable are extremely biased when error probabilities are ignored. If misclassification parameters are incorporated in the likelihood function, the bias of the estimators will be satisfactorily reduced, however, there would be a considerable increase in variability, which would reduce the quality of the decision-making process. To minimize the problem, there is a need to introduce additional information. It will be demonstrated that the realization of repeated measures in the response variable, or in part of it, can reduce bias and variability of the estimators, simultaneously.

**Keywords:** Binary outcome; logistic regression; misclassification; repeated classifications.

## LISTA DE ILUSTRAÇÕES

FIGURA 1: Gráficos das distribuições dos parâmetros estimados para o modelo de regressão logística binário nos casos sem erros de classificação (SEC) e com erros ignorados (IEC) na variável resposta. ....	11
FIGURA 2: Diagramas de caixa dos parâmetros do modelo CECCR( $m = 1, n_r = n$ ). ....	24
FIGURA 3: Diagramas de caixa dos parâmetros do modelo CECCR( $m = [3; 5], n_r = n$ ). ..	27



## LISTA DE TABELAS

TABELA 1: Distribuição de probabilidade para um caso não-identificável.....	18
TABELA 2: Cenários para geração de dados e análise de desempenho de cada modelo.....	211
TABELA 3: Comparação entre os modelos SEC( $n = 500$ ) e IEC( $n = 500$ ). .....	244
TABELA 4: Comparação entre os modelos IEC( $n = 1000$ ) e CECCR( $m = 1, n_r = n, n = 1000$ ).....	25
TABELA 5: Comparação entre os modelos CECCR( $m = 1, n_r = n, n = 500$ ) e CECCR( $m = 3, n_r = n, n = 500$ ).....	27
TABELA 6: Comparação entre os modelos SEC( $n = 500$ ) e CECCR( $m = 5, n_r = n, n = 500$ ).....	28
TABELA 7: Comparação entre os modelos CECCR( $m = 1, n_r = n, n = 1000$ ) e CECCR( $m = 3, n_r = n, n = 50$ ). .....	28
TABELA 8: Comparação entre os modelos CECCR( $m = 1, n_r = n, n = 1000$ ) e CECCR( $m = 3, n_r = n, n = 100$ ).....	29
TABELA 9: Comparação entre os modelos CECCR( $m = [1; 3], n_r = [0,3n; 0,7n; n], n = 500$ ).....	30
TABELA 10: Frequência de classificações como "perfeita" após 5 classificações. ....	33
TABELA 11: Estimativas dos parâmetros para os dados das lentes. ....	33
TABELA 12: Estimativas das probabilidades de se obter lentes “perfeitas”.....	34
TABELA 13: Simulações para o modelo regressão logística sem erros de classificação na variável resposta (SEC). .....	39
TABELA 14: Simulações para o modelo de regressão logística com erros de classificação na variável resposta e ignorados (IEC).....	40
TABELA 15: Simulações para o modelo de regressão logística com verossimilhança modificada para acomodar erros de medidas CECCR( $m = [1; 3; 5], n_r = n$ ). .....	41
TABELA 16: Simulações para o modelo de regressão logística com erros de medida e reclassificação parcial CECCR( $m = [3; 5], n_r = 0,3n$ ).....	43
TABELA 17: Simulações para o modelo de regressão logística com erros de medida e reclassificação parcial CECCR( $m = [3; 5], n_r = 0,7n$ ).....	45

## SUMÁRIO

1. INTRODUÇÃO .....	10
2. MODELOS PROBABILÍSTICOS.....	13
2.1 Modelo de regressão logística sem erro de classificação .....	13
2.2 Modelo de regressão logística com erros de classificação na função de verossimilhança .....	14
2.3 Modelo de regressão logística com erros de classificação na função de verossimilhança e reclassificação dos itens na amostra .....	15
2.4 Modelo de regressão logística com erros de classificação na função verossimilhança e classificação parcial da amostra.....	16
3. ESTUDO DE SIMULAÇÃO .....	19
3.1 Tamanho amostral.....	20
3.2 Erros de classificação.....	20
3.3 Análise única e reclassificação de um dado item .....	20
3.4 Análise com parte da amostra reclassificada .....	21
3.5 Geração dos dados .....	21
3.6 Critérios de avaliação.....	22
3.7 Resultados.....	23
3.7.1 Resultados utilizando informação adicional.....	26
4. ANÁLISE DE UM BANCO DE DADOS REAL .....	32
5. COMENTÁRIOS FINAIS .....	35
REFERÊNCIAS .....	36
APÊNDICE A – Resultados das simulações.....	38
ANEXO A – Procedimento Bootstrap utilizado .....	47

## 1. INTRODUÇÃO

Modelos de regressão clássicos para dados binários, como a regressão logística, ignoram erros de medidas nas variáveis envolvidas. Assumir tal condição como verdadeira é uma restrição muito forte na análise em situações práticas e pode comprometer o processo de decisão e inferência.

Nesse sentido, erros de medida podem ocorrer em sistemas de controle de qualidade, em que há a possibilidade de erro na classificação de peças, como “conforme” ou “não-conforme”. Ou, em estudos de comportamento social, no qual um participante pode responder incorretamente à questão. Outros exemplos são na análise de dados históricos e na detecção de doenças através de exames.

Sendo assim, constata-se que esses erros podem levar à considerável vício no processo de estimação dos parâmetros caso sejam ignorados; veja, por exemplo, Gustafson (2004).

Cabe ressaltar que a análise dos “erros de medida” em dados binários será feita sobre o erro ao reportar o verdadeiro valor da variável resposta em uma categoria diferente da real. Desse modo, não serão considerados os erros de medida nas covariáveis do modelo de regressão.

A FIGURA 1 mostra uma simulação de uma regressão logística para dois conjuntos de dados: (1) a variável resposta não possui erros de classificação (SEC); (2) os erros de classificação existem, mas são ignorados (IEC). Perceba que as estimativas para o caso IEC são viciadas e distantes do verdadeiro valor, representado pela linha vertical em cada gráfico.

Procurando resolver esse problema referente ao vício dos estimadores, vários pesquisadores incorporaram os erros de classificação desconhecidos como parâmetros adicionais na função de verossimilhança, em uma perspectiva frequentista, quais sejam: Manski (1988); Magder e Hughes (1997); Bollinger e David (1997); Hausman, Abrevaya e Scott-Morton (1998); Cheng e Hsueh (2003); Liu e Zhang (2017); Meyer e Mittag (2017). Métodos bayesianos também foram desenvolvidos, como por exemplo: Paulino, Soares e Neuhaus (2003);

McInturff *et al.* (2004); Gerlach e Stamey (2007); Naranjo *et al.* (2014); Pires e Quinino (2018).

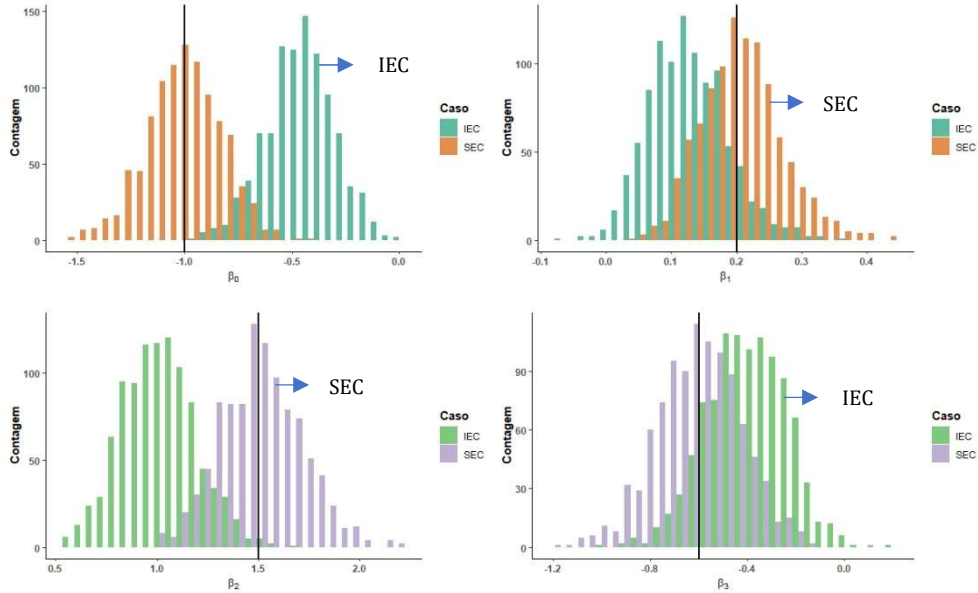


FIGURA 1: Gráficos da distribuição dos parâmetros estimados para o modelo de regressão logística binário nos casos sem erros de classificação (SEC) e com erros ignorados (IEC) na variável resposta.

Assim, será utilizado o modelo proposto por Hausman, Abrevaya e Scott-Morton (1998), modificado do modelo de regressão clássico com a inclusão de mais dois parâmetros: o falso-negativo ou erro do tipo I ( $e_1$ ) e o falso-positivo ou erro do tipo II ( $e_2$ ). O primeiro ocorre quando se observa o valor 0 tendo o valor real 1; já o último, ao se observar o valor 1, quando, na verdade, tem-se o valor 0 como real.

Nesse contexto, é interessante pensar que o trabalho de Hausman, Abrevaya e Scott-Morton (1998), e, posteriormente, ampliado por Liu e Zhang (2017), demonstraram a capacidade do novo modelo em estimar tanto os parâmetros originais da regressão, quanto os erros de classificação. Nesses dois artigos, os parâmetros foram estimados por máxima verossimilhança e os estimadores diminuem, consideravelmente, o vício das estimativas. No entanto, a variabilidade dos estimadores aumenta substancialmente.

Ainda acerca dos estudos de Liu e Zhang (2017), importa ressaltar que criaram, também, um pacote no R chamado de “logistic4p” para facilitar o uso de modelos com erros de classificação. Esse pacote é capaz de estimar as taxas dos erros de classificação, os

coeficientes de regressão e os respectivos erros-padrão. Em verdade, possui certas limitações, já que não contempla o caso de classificações repetidas em um mesmo item e, em algumas vezes, ocorrem problemas de convergência, em geral, quando o tamanho amostral é pequeno.

Resumindo, temos um duplo problema: ignorar os erros de classificação no processo de inferência, implica considerável vício na estimação dos parâmetros. Por outro lado, introduzir os erros de classificação na função de verossimilhança minimiza o vício, mas aumenta de forma considerável a variabilidade.

Uma outra abordagem para a solução desse problema é o uso de classificações repetidas na variável resposta. Por meio de simulação Monte Carlo, chega-se à conclusão de que o vício e a variabilidade do processo diminuem, simultaneamente. Observa-se, também, que a realização de classificações repetidas apenas em parte da amostra pode ser suficiente para gerar um processo de estimação confiável e mais econômico.

Ademais, destaca-se que o uso de classificações repetidas não é uma abordagem nova. Sem a presença de covariáveis, Blischke (1962, 1964) discutiram sobre estimadores para a proporção de interesse sujeita a erros de classificação. A propósito, em recente artigo, Pires e Quinino (2018), utilizaram a abordagem bayesiana com distribuição *a priori* informativa e incorporaram classificações repetidas em modelos de regressão binomial em que a variável resposta apresenta erros de classificação. Nesse trabalho, discute-se: (1) a abordagem frequentista; (2) o problema de não usar classificações repetidas; (3) construção da função de verossimilhança sem a necessidade de que as classificações repetidas estejam presentes em todos os elementos da amostra.

No Capítulo 2, serão apresentados os modelos de regressão logística com erro de classificação. Já no Capítulo 3, será dada atenção ao estudo de simulação e seus resultados. Finalmente, uma aplicação de dados reais estará descrita no Capítulo 4, seguida de discussão e comentários no Capítulo 5.

Deve ser destacado que, para todas as análises, o programa utilizado foi o R, que se trata de uma linguagem e ambiente para análise computacional estatística e pode ser instalado gratuitamente pelo site [www.r-project.org](http://www.r-project.org).

## 2. MODELOS PROBABILÍSTICOS

Os erros que serão abordados nesse trabalho referem-se aos erros reportados para a variável resposta, apenas. Ressalta-se que são erros constantes, dependem do estado real do sujeito e independentes dos valores das covariáveis. Sobre os erros de medida nas covariáveis, não serão considerados.

Nesse cenário, considere  $\tilde{Y}_i$  como o valor real da variável resposta binária do item  $i$  ( $i = 1, \dots, n$ ).  $\mathbf{X}$  é a matriz ( $n \times p$ ) de covariáveis, na qual a  $i$ -ésima linha é representada por  $\mathbf{X}_i = (X_{1i}, X_{2i}, \dots, X_{pi})$ . E, também,  $\{(y_i, x_i), i = 1, \dots, n\}$  o conjunto de dados com  $n$  observações. Os parâmetros desconhecidos  $\beta_0, \dots, \beta_p$  representam a associação das covariáveis com a variável resposta  $\tilde{Y}$ . Assim, deseja-se estimar  $P(\tilde{Y}_i = 1 | X_i) = F_i$ .

Em relação aos modelos de regressão que serão estudados, têm “logit” como função de ligação.

Finalmente, o método da máxima verossimilhança será o método usado para a estimação dos parâmetros de todos os modelos. Em geral, os estimadores de máxima verossimilhança apresentam bom desempenho para amostras finitas e são assintoticamente eficientes e consistentes (Casella e Berger, 2002).

### 2.1 Modelo de regressão logística sem erro de classificação

O modelo clássico de regressão logística está definido em (1):

$$\left\{ \begin{array}{l} \tilde{Y}_i \sim \text{Bernoulli}(F_i) \\ F_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}} \\ \eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} \end{array} \right. \quad (1)$$

Considerando uma amostra de tamanho  $n$ , a função de verossimilhança de (1) está representada em (2) e a função log-verossimilhança em (3):

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n F_i^{y_i} (1 - F_i)^{1 - y_i} = \prod_{i=1}^n \frac{e^{(X_i' \boldsymbol{\beta})^{y_i}}}{1 + e^{X_i' \boldsymbol{\beta}}} \quad (2)$$

$$l(\boldsymbol{\beta}) = \ln[L(\boldsymbol{\beta})] = \sum_{i=1}^n y_i X_i' \boldsymbol{\beta} - \ln(1 + e^{X_i' \boldsymbol{\beta}}) \quad (3)$$

## 2.2 Modelo de regressão logística com erros de classificação na função de verossimilhança

Na hipótese de erros de classificação na variável resposta, o conjunto observado  $y_1, \dots, y_n$  pode diferir dos valores reais  $\tilde{y}_1, \dots, \tilde{y}_n$ . Como o erro de classificação é função apenas do estado real, tem-se:

$$\begin{aligned} P(Y_i = 0 | \tilde{Y}_i = 1) &= e_1 \\ P(Y_i = 1 | \tilde{Y}_i = 0) &= e_2 \\ P(Y_i = 0 | \tilde{Y}_i = 0) &= (1 - e_2) \\ P(Y_i = 1 | \tilde{Y}_i = 1) &= (1 - e_1) \end{aligned}$$

Para considerar os erros de classificação, é necessário encontrar a distribuição de  $Y$  e pela regra da probabilidade total tem-se:

$$\begin{aligned} P(Y_i = 1) &= q_i = P(Y_i = 1 | \tilde{Y}_i = 1)P(\tilde{Y}_i = 1) + P(Y_i = 1 | \tilde{Y}_i = 0)P(\tilde{Y}_i = 0) \\ &= (1 - e_1)F_i + e_2(1 - F_i) \\ P(Y_i = 1) &= q_i = e_2 + (1 - e_2 - e_1)F_i \end{aligned}$$

A especificação do modelo binário para acomodar erros de classificação é:

$$\begin{cases} Y_i \sim \text{Bernoulli}(q_i) \\ q_i = e_2 + (1 - e_2 - e_1)F_i \\ F_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}} \\ \eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} \end{cases} \quad (4)$$

A função de verossimilhança para uma amostra de  $n$  valores de  $Y_i$  de (4) é dada por:

$$L(e_1, e_2, \boldsymbol{\beta}) = \prod_{i=1}^n q_i^{y_i} (1 - q_i)^{1 - y_i} \quad (5)$$

E a função log-verossimilhança:

$$l(\boldsymbol{\beta}) = \ln[L(e_1, e_2, \boldsymbol{\beta})] = \sum_{i=1}^n \{y_i \ln(q_i) + (1 - y_i) \ln(1 - q_i)\} \quad (6)$$

### 2.3 Modelo de regressão logística com erros de classificação na função de verossimilhança e reclassificação dos itens na amostra

Agora, considere a possibilidade de cada observação  $i$  ser classificada independentemente  $m$  vezes. Seja  $S_{ij}$  ( $i = 1, \dots, n; j = 1, \dots, m$ ) uma variável Bernoulli correspondendo à  $j$  – ésima classificação da  $i$  – ésima observação. O número de vezes que a observação  $i$  recebeu o valor 1 é  $s_i$  ( $S_i = \sum_{j=1}^m S_{ij}$ ,  $i = 1, 2, \dots, n$ ) e pode ser visto como uma mistura de duas binomiais.

$$P(S_i = s_i) = P(S_i = s_i | \tilde{Y}_i = 1)P(\tilde{Y}_i = 1) + P(S_i = s_i | \tilde{Y}_i = 0)P(\tilde{Y}_i = 0)$$

$$P(S_i = s_i) = F_i \binom{m}{s_i} (1 - e_1)^{s_i} e_1^{m - s_i} + (1 - F_i) \binom{m}{s_i} e_2^{s_i} (1 - e_2)^{m - s_i}$$

A especificação completa do modelo é dada em (7):



$$\left\{ \begin{array}{l} Y_i = \sum_{k=1}^2 w_k \pi_k(y) \\ w_1 = F_i; w_2 = (1 - F_i) \\ \pi_1(y) \sim \text{Binomial} [m; (1 - e_1)] \\ \pi_2(y) \sim \text{Binomial} (m; e_2) \\ F_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}} \\ \eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi} \end{array} \right. \quad (7)$$

A função verossimilhança de (7) para uma amostra de tamanho  $n$  é encontrada em Pires e Quinino (2018):

$$\begin{aligned} L(e_1, e_2, \boldsymbol{\beta}) &= \prod_{i=1}^n F_i \binom{m}{s_i} (1 - e_1)^{s_i} e_1^{m-s_i} + (1 - F_i) \binom{m}{s_i} e_2^{s_i} (1 - e_2)^{m-s_i} \\ &= \prod_{i=1}^n \left[ \frac{e^{X_i' \boldsymbol{\beta}}}{1 + e^{X_i' \boldsymbol{\beta}}} \binom{m}{s_i} (1 - e_1)^{s_i} e_1^{m-s_i} + \left( 1 - \frac{e^{X_i' \boldsymbol{\beta}}}{1 + e^{X_i' \boldsymbol{\beta}}} \right) \binom{m}{s_i} e_2^{s_i} (1 - e_2)^{m-s_i} \right] \end{aligned} \quad (8)$$

E a função log-verossimilhança por:

$$\begin{aligned} l(\boldsymbol{\beta}) &= \ln[L(e_1, e_2, \boldsymbol{\beta})] \\ &= \sum_{i=1}^n \ln \left\{ \frac{e^{X_i' \boldsymbol{\beta}}}{1 + e^{X_i' \boldsymbol{\beta}}} \binom{m}{s_i} (1 - e_1)^{s_i} e_1^{m-s_i} + \left( 1 - \frac{e^{X_i' \boldsymbol{\beta}}}{1 + e^{X_i' \boldsymbol{\beta}}} \right) \binom{m}{s_i} e_2^{s_i} (1 - e_2)^{m-s_i} \right\} \end{aligned} \quad (9)$$

## 2.4 Modelo de regressão logística com erros de classificação na função verossimilhança e classificação parcial da amostra

Suponha, agora, que  $n_r$  sujeitos são classificados independentemente  $m$  vezes e os restantes  $(n - n_r)$  são classificados apenas uma única vez. Ou seja, supõe-se que uma parte da amostra ( $n_r$ ) seja classificada  $m$  vezes e a outra parte uma única vez. Assim, a função de verossimilhança pode ser escrita como:

$$L(e_1, e_2, \boldsymbol{\beta}) = \prod_{i=1}^{n_r} \left\{ F_i \binom{m}{s_i} (1 - e_1)^{s_i} e_1^{m-s_i} + (1 - F_i) \binom{m}{s_i} e_2^{s_i} (1 - e_2)^{m-s_i} \right\} \prod_{i=n_r+1}^n \left\{ q_i^{y_i} (1 - q_i)^{1-y_i} \right\} \quad (10)$$

Note que a expressão (10) generaliza trabalhos anteriores. Se  $n_r = 0$ , então a função de verossimilhança se reduz àquela utilizada nos trabalhos, por exemplo, de Magder e Hughes (1997); Liu e Zhang (2017); Paulino, Soares e Neuhaus (2003); e McInturff *et al.* (2004). Se  $n_r = n$ , logo a função de verossimilhança é equivalente à utilizada em Pires e Quinino (2018). Além disso, quando  $e_1 = e_2 = 0$ ,  $m = 1$  e  $n_r = 0$ , a função de verossimilhança (10) é igual àquela do modelo de regressão logística tradicional em que a variável resposta não está sujeita a erros de classificação.

Os estimadores de máxima verossimilhança são os valores de  $\boldsymbol{\beta}$ ,  $e_1$ ,  $e_2$  que maximizam (2, 5, 8, 10), ou equivalentemente, o logaritmo natural dos respectivos.

Uma característica importante da função de verossimilhança é que a identificabilidade garante que o valor máximo é único (Newey e McFadden, 1994). Nestes termos, é importante avaliar a identificabilidade do modelo. Demonstra-se a propriedade de identificabilidade do modelo proposto, reescrevendo a função de verossimilhança como:

$$L = \prod_{i=1}^n P(S_i = s_i)^{\delta_i} P(Y_i = y_i)^{1-\delta_i}$$

em que  $\delta_i = 1$  indica a presença de classificações repetidas para a observação  $i$  e  $\delta_i = 0$ , caso contrário. Pelos resultados de Hausman, Abrevaya e Scott-Morton (1998) e Pires e Quinino (2018), ambos os termos são identificáveis se  $e_1 + e_2 < 1$  e, portanto, o modelo apresentado em (10) também é (localmente) identificável.

Para melhor entendimento da necessidade de tal restrição, considere  $\theta = (e_1 = 0,1; e_2 = 0,2; \beta_0 = -0,6; \beta_1 = 3,0)$  e  $\tilde{\theta} = (e_1 = 0,8; e_2 = 0,9; \beta_0 = 0,6; \beta_1 = -3,0)$ . A TABELA 1 apresenta a contribuição de diferentes observações para a função de verossimilhança. Dessa maneira, observe que, para um mesmo resultado amostral  $s_i$  e com

$\mathbf{x}_i = [1; 1,3]$ , tem-se  $P_\theta(S_i = s_i) = P_{\tilde{\theta}}(S_i = s_i)$ , o que caracterizaria o problema da falta de identificabilidade. Esse comportamento é observado sempre que  $\tilde{e}_1 = (1 - e_2)$ ,  $\tilde{e}_2 = (1 - e_1)$ ,  $\tilde{\boldsymbol{\beta}} = -\tilde{\boldsymbol{\beta}}$ . Portanto, a restrição  $e_1 + e_2 < 1$  é necessária para garantir a identificabilidade local.

TABELA 1

Distribuição de probabilidade para um caso não-identificável.

<b>Observação</b>	<b><math>m</math></b>	<b><math>S_i</math></b>	<b><math>P_\theta(S_i = s_i)</math></b>	<b><math>P_{\tilde{\theta}}(S_i = s_i)</math></b>
1	1	0	0,1249	0,1249
2	1	1	0,8751	0,8751
3	3	0	0,0192	0,0192
4	3	1	0,0397	0,0397
5	3	2	0,2378	0,2378
6	3	3	0,7034	0,7034

### 3. ESTUDO DE SIMULAÇÃO

No capítulo anterior, os modelos que serão usados para obter a estimação do conjunto de parâmetros desconhecidos  $\varphi = (\boldsymbol{\beta}, e_1, e_2)$  foram descritos. Neste capítulo, simulações Monte Carlo serão utilizadas para avaliar o impacto dos erros de classificação no processo de estimação da regressão logística. O objetivo desse estudo de simulação é duplo. (1) Demonstrar a influência dos erros de classificação na estimação dos parâmetros. (2) Avaliar o desempenho de classificações repetidas na amostra ou em parte dela. Para uma comparação mais eficaz, inclui-se a situação em que não existe a presença dos erros de classificação, que será utilizada como a referência a ser atingida.

Sendo assim, os dados foram gerados de 3 covariáveis:  $X_1$  tem distribuição lognormal (0, 1);  $X_2$  é uma variável dummy com probabilidade de 50% de ter o valor 1;  $X_3$  tem distribuição uniforme no intervalo (-1,1). Os coeficientes da regressão são:  $\boldsymbol{\beta} = [\beta_0; \beta_1; \beta_2; \beta_3] = [-1; 0,2; 1,5; -0,6]$ . Como  $F_i = F(X_i' \boldsymbol{\beta})$ , a escolha dessas covariáveis e desses coeficientes foi baseado em Hausman, Abrevaya e Scott-Morton (1998), e também para garantir valores ao longo de todo o intervalo (0, 1) para  $F_i$ .

Além disso, serão considerados como potenciais influenciadores na estimação dos parâmetros os seguintes fatores: tamanho da amostra, taxas de erros de classificação ( $e_1, e_2$ ) e presença ou não de classificação repetida na variável resposta. Resumidamente, consideram-se fatores de tamanhos pequenos, médios e grandes.

Os principais cenários avaliados foram: (1) não existem erros de classificação; (2) os erros de classificação existem, mas são ignorados; (3) os erros de classificação existem e são inseridos como parâmetros adicionais na função de verossimilhança; (4) os erros de classificação existem e são inseridos como parâmetros adicionais na função de verossimilhança com uso de classificações repetidas; (5) os erros de classificação existem e são inseridos como parâmetros adicionais na função de verossimilhança, utilizando classificações repetidas apenas em parte da amostra (30% e 70%).

Com isso, o objetivo foi comparar os resultados entre todos eles e permitir ao leitor uma escolha mais econômica e que, simultaneamente, apresentasse vício e variabilidade mais próximos à situação em que não houvesse erros de classificação na variável dependente.

A saber, o número de replicação Monte Carlo foi fixo e definido como  $R = 1000$ .

A função utilizada do programa R para a estimação dos parâmetros por máxima verossimilhança foi a “NLMINB”. NLMINB é um método de otimização quasi-Newton para limites definidos ou não. O código foi baseado na biblioteca Fortran e maiores detalhes encontrados em Dennis, Gay e Welsch (1977).

### **3.1 Tamanho amostral**

O trabalho de Hausman, Abrevaya e Scott-Morton (1998) considerou simulações com tamanhos amostrais  $n = 5000$ . Já, Liu e Zhang (2017) consideraram tamanhos de  $n = [1000; 2000; 5000]$ . Segundo os autores, valores menores do que esses apresentaram convergência muito baixa e não foram reportados.

Entretanto, do ponto de vista prático, essas amostras são muito grandes, nem sempre disponíveis e de alto custo econômico e operacional. Assim, os tamanhos amostrais do trabalho foram definidos em  $n = [50; 100; 500; 1000]$ .

### **3.2 Erros de classificação**

As taxas dos erros de classificação assumiram os seguintes valores:  $e_i = [0,05; 0,10; 0,20]$ ,  $i (i = 1, 2)$ . Mantendo todas as outras variáveis fixas, 9 combinações diferentes são possíveis.

### **3.3 Análise única e reclassificação de um dado item**

Para um tamanho amostral fixo, a condição de repetir a classificação de cada item traz mais informação experimental do que a condição de uma única análise do item. Assim, espera-se que a condição de reanálise leve a estimadores melhores, isto é, menor vício e variabilidade. Além disso, quanto maior o número de repetições, mais informação estará disponível.

Para analisar o efeito em reclassificar ou não, os valores considerados foram:  $\mathbf{m} = [1; 3; 5]$ . Observe que quando  $m = 1$ , tem-se uma única observação para cada item da amostra.

### 3.4 Análise com parte da amostra reclassificada

As frações consideradas para as reclassificações dos itens de uma amostra foram  $\mathbf{n}_r = [0,3n; 0,7n; n]$ . Como exemplo, para  $n_r = 0,3n$ , 30% da amostra é reclassificada  $m$  vezes, e o restante da amostra é classificado uma única vez.

A TABELA 2 apresenta o resumo de todos os cenários a serem avaliados:

TABELA 2

Cenários para geração de dados e análise de desempenho de cada modelo.

Tamanho amostral ( $n$ )	Erros de classificação ( $e_1; e_2$ )	Nº de classificações de cada item ( $m$ )	Fração reclassificada( $n_r$ )
50	0,05	1	0,3n
100	0,10	3	0,7n
500	0,20	5	$n$
1000			

### 3.5 Geração dos dados

Antes de gerar qualquer dado, uma semente aleatória foi definida e fixada para que o leitor interessado possa reproduzir exatamente os mesmos resultados.

O procedimento utilizado para gerar o conjunto de observações  $(y_i, x_i)$  para análise dos modelos foi:

1. Defina o tamanho amostral:  $\mathbf{n} = [50; 100; 500; 1000]$ ;
2. Defina a quantidade de análise (repetição) de cada item:  $\mathbf{m} = [1; 3; 5]$ ;
3. Defina os valores de  $\mathbf{e}_j = [0,05; 0,10; 0,20]$ ,  $j(j = 1,2)$ ;

4. Defina a fração da amostra que será reclassificada:  $\mathbf{n}_r = [0,3n; 0,7n; n]$ ;
5. Gere, aleatoriamente, os  $n$  valores de cada covariável de acordo com cada distribuição definida e obtenha a matriz de planejamento  $n \times p$ ;
6. Considere a  $i$  – ésima linha da matriz gerada em 5,  $i(i = 1, \dots, n)$ ;
7. Calcule  $\eta_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$ ;
8. Calcule  $F_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$ ;
9. Gere uma variável aleatória Bernoulli com probabilidade de sucesso  $F_i$  (assim foi obtido o vetor de valores reais  $\tilde{Y}$ );
10. Para gerar o vetor observável ( $\mathbf{Y}$ ) com erros de classificação:
  - a. Para o conjunto de observação até o valor ( $n_r$ ):
    - i. Se  $\tilde{Y}_i = 1 \rightarrow$  gerar  $m$  variáveis aleatórias Bernoulli com probabilidade de sucesso  $(1 - e_1)$ ;
    - ii. Se  $\tilde{Y}_i = 0 \rightarrow$  gerar  $m$  variáveis aleatórias Bernoulli com probabilidade de sucesso  $(e_2)$ ;
  - b. Para o conjunto de observação de valor  $(n_r + 1)$  até  $n$ :
    - i. Se  $\tilde{Y}_i = 1 \rightarrow$  gerar uma variável aleatória Bernoulli com probabilidade de sucesso  $(1 - e_1)$ ;
    - ii. Se  $\tilde{Y}_i = 0 \rightarrow$  gerar uma variável aleatória Bernoulli com probabilidade de sucesso  $(e_2)$ ;
11. Voltar ao item 5 e repetir esse procedimento  $R$  vezes.

### 3.6 Critérios de avaliação

O desempenho de cada modelo foi avaliado de acordo com a média, o vício relativo, o erro-padrão e o erro quadrático médio.

Considere  $\varphi$  como o valor real do parâmetro,  $\hat{\varphi}_r$  o valor do parâmetro estimado na replicação  $r$ ,  $R$  o número total de replicações Monte Carlo e  $\hat{\varphi}$  como a média das estimativas obtidas:

$$\hat{\varphi} = \frac{\sum_{r=1}^R \hat{\varphi}_r}{R}$$

O vício relativo será dado por:

$$vicio = \left( \frac{\hat{\varphi} - \varphi}{|\varphi|} \right) * 100$$

O desvio-padrão:

$$\hat{\sigma} = \sqrt{\frac{\sum_{r=1}^R (\hat{\varphi}_r - \hat{\varphi})^2}{(R - 1)}}$$

Já o erro quadrático médio:

$$E. Q. M. = \frac{\sum_{r=1}^R (\hat{\varphi}_r - \varphi)^2}{R}$$

### 3.7 Resultados

Os resultados de todas as simulações estão no APÊNDICE A. Devido à variedade de cenários simulados, não haverá discussão de todos nessa seção. E como o padrão observado foi similar, um exemplo será escolhido como representativo. Para demonstrar o desempenho dos modelos, foram escolhidos e fixados os valores das taxas de erro  $e_1 = 0,10$  e  $e_2 = 0,20$ .

Quando ignoramos os erros de classificação equivocadamente (IEC), o vício das estimativas pode ser severo e comprometer o processo de decisão. A TABELA 3 compara a situação em que não existem erros de classificação na variável resposta (SEC) com o caso IEC para  $n = 500$ ,  $m = 1$ ,  $e_1 = 0,10$ ,  $e_2 = 0,20$ . Observe que  $e_1 = e_2 = 0$  para o caso SEC. Pode-se observar que a magnitude do vício relativo para o caso IEC é superior a 30%, o que leva a um substancial aumento do erro médio quadrático quando comparado ao caso SEC, uma vez que os desvios-padrão nos dois casos são similares. Essa condição mantém-se para tamanhos amostrais maiores e a constatação desse vício motivou vários trabalhos na literatura, como descrito na introdução.

Com objetivo de diminuir tal vício, os parâmetros relativos aos erros de classificação foram inseridos na função de verossimilhança, gerando-se um caso particular da expressão (10) quando  $m = 1$ ,  $n_r = n$ , que será denominado como CECCR( $m = 1, n_r = n$ ).



TABELA 3

Comparação entre os modelos SEC( $n = 500$ ) e IEC( $n = 500$ ).

	Real	SEC				IEC			
		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
$\beta_0$	-1,0	-1,0122	0,1752	-1,22	0,0308	-0,4507	0,1705	54,93	0,3308
$\beta_1$	0,2	0,2064	0,0616	3,21	0,0038	0,1293	0,0600	-35,37	0,0086
$\beta_2$	1,5	1,5126	0,2052	0,84	0,0422	1,0226	0,2006	-31,83	0,2681
$\beta_3$	-0,6	-0,6115	0,1718	-1,92	0,0296	-0,4000	0,1626	33,33	0,0664

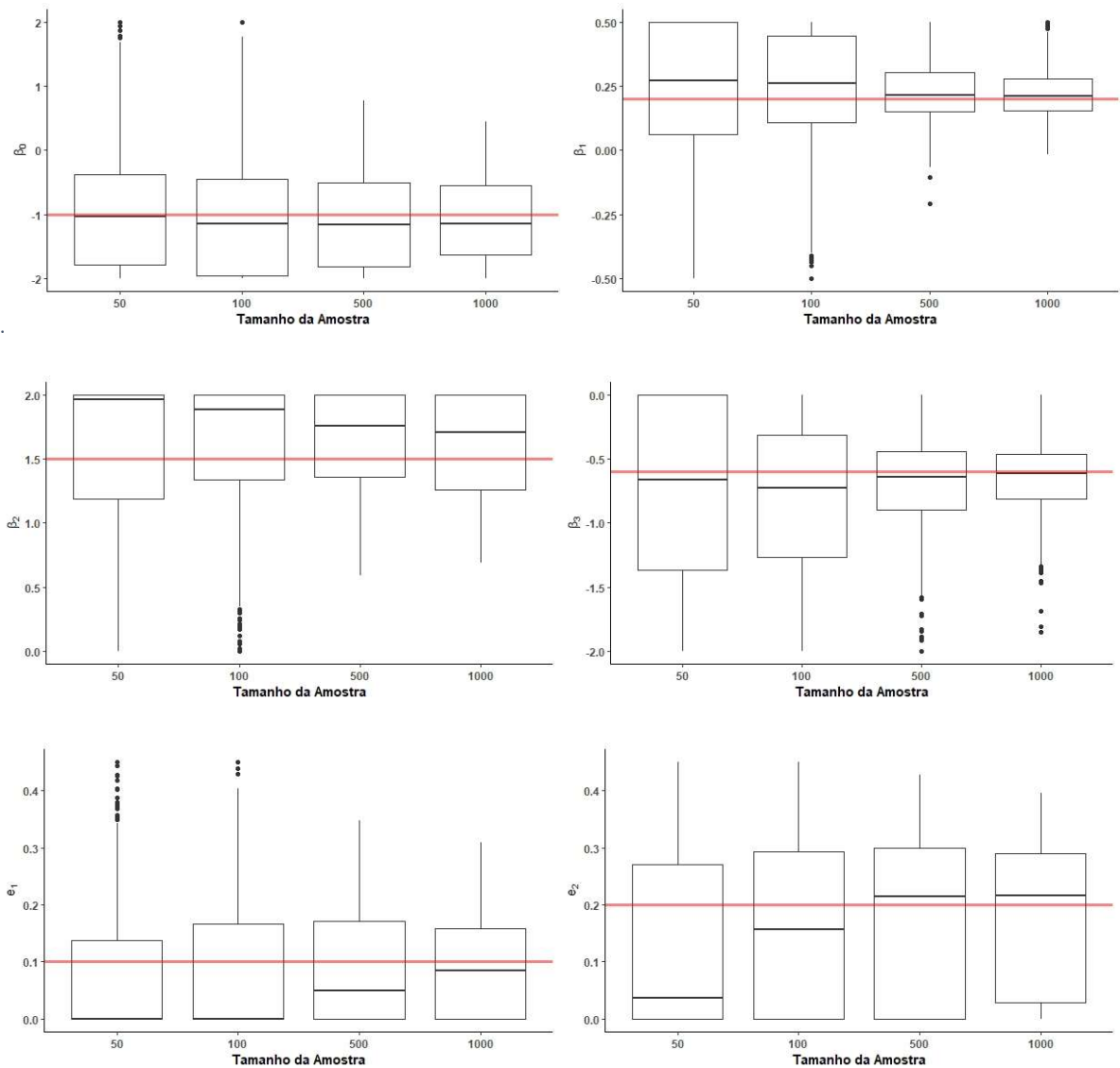


FIGURA 2: Diagramas de caixa dos parâmetros do modelo CECCR( $m = 1, n_r = n$ ).

A respeito do caso  $CECCR(m = 1, n_r = n)$ , a FIGURA 2 mostra que o vício dos parâmetros diminui à medida que o tamanho amostral aumenta. Outro ponto a destacar é que se observou um grande e comprometedor desvio-padrão das estimativas, principalmente para amostras menores.

A TABELA 4 compara a situação  $IEC(n = 1000)$  com a  $CECCR(m = 1, n_r = n, n = 1000)$ . Observe que a nova função de verossimilhança, incorporando os parâmetros de erros de classificação, gerou estimadores de máxima verossimilhança decididamente com vícios inferiores quando comparados à situação em que os erros de classificação são ignorados. Ressalta-se que todos os vícios relativos são inferiores a 15% e tendem a aproximar do caso SEC. Este resultado está de acordo com trabalhos anteriores, como, por exemplo, Magder e Hughes (1997); Hausman, Abrevaya e Scott-Morton (1998); e Liu e Zhang (2017).

TABELA 4

Comparação entre os modelos  $IEC(n = 1000)$  e  $CECCR(m = 1, n_r = n, n = 1000)$ .

	Real	IEC				CECCR( $m = 1, n_r = n$ )			
		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
$\beta_0$	-1,0	-0,4506	0,1132	54,94	0,3147	-1,1114	0,6146	-11,14	0,3898
$\beta_1$	0,2	0,1271	0,0412	-36,46	0,0070	0,2272	0,0980	13,59	0,0103
$\beta_2$	1,5	1,0174	0,1392	-32,18	0,2523	1,6105	0,3849	7,37	0,1602
$\beta_3$	-0,6	-0,3881	0,1157	35,31	0,0583	-0,6519	0,2602	-8,65	0,0703
$e_1$	0,1					0,0875	0,0838	-12,49	0,0072
$e_2$	0,2					0,1820	0,1260	-8,98	0,0162

Com negativa surpresa, observou-se, também, um grande aumento do desvio-padrão das estimativas, levando a, pelo menos, dobrar a magnitude quando comparado aos casos SEC e IEC. Isso conduz a um problema efetivo de decisão, pois, se por um lado o vício diminui, por outro lado a variabilidade dos estimadores aumenta substancialmente. Nesse sentido, observe que, se o critério do erro quadrático médio for usado, fica-se inclinado a optar pelo modelo que ignora os erros de classificação (caso IEC), pois, com tal hipótese, as estimativas de  $\beta_0$ ,  $\beta_1$  e  $\beta_3$  apresentam valores de EQMs inferiores. Apenas a estimativa do EQM de  $\beta_2$  apresentou valor inferior para caso  $CECCR(m = 1, n_r = n)$  quando comparado ao IEC, mas, ainda assim, em patamares similares. Tais situações permitiram concluir que a utilização do modelo  $CECCR(m = 1, n_r = n)$  ainda pode comprometer o processo de inferência e decisão.

Tendo isso em vista, discute-se, na seção 3.7.1, alternativas para minimizar a influência dos erros de classificação. Assim, a melhoria será obtida por meio do uso de informações adicionais que podem ser oriundas de classificações repetidas da variável repostada.

### 3.7.1 Resultados utilizando informação adicional

Amplie o exemplo discutido anteriormente para considerar  $m$  classificações independentes em cada item da amostra  $CECCR(m = [3; 5], n_r = n)$ . Nessa perspectiva, a FIGURA 3 mostra a dispersão dos dados para uma dada variação no tamanho amostral e no número de reclassificações. Os tamanhos escolhidos foram  $n = [50; 500]$ . Os erros de classificação foram mantidos fixos em  $e_1 = 0,10$  e  $e_2 = 0,20$ .

As simulações mostram que, ao aumentar o número de reclassificações, a dispersão dos dados, o vício relativo e o erro quadrático médio diminuem. As mesmas afirmações podem ser feitas para o caso em que o número de repetições é fixo e aumenta-se o tamanho amostral.

Agora, considere o cenário em que  $CECCR(m = 3, n_r = n, n = 500)$ ,  $e_1 = 0,10$ ,  $e_2 = 0,20$ . A TABELA 5 mostra que a realização de três classificações diminuiu os resultados do vício, desvio-padrão e EQM quando comparados ao caso  $CECCR(m = 1, n_r = n)$ . Resultados ainda melhores são obtidos com mais classificações repetidas. A TABELA 6 apresenta o caso com  $CECCR(m = 5, n_r = n, n = 500)$ , e percebe-se que a influência dos erros de classificação foi praticamente eliminada quando comparamos com o caso  $SEC(n = 500)$ . Podem-se estender esses resultados para os outros cenários com diferentes erros de classificação e tamanhos amostrais.

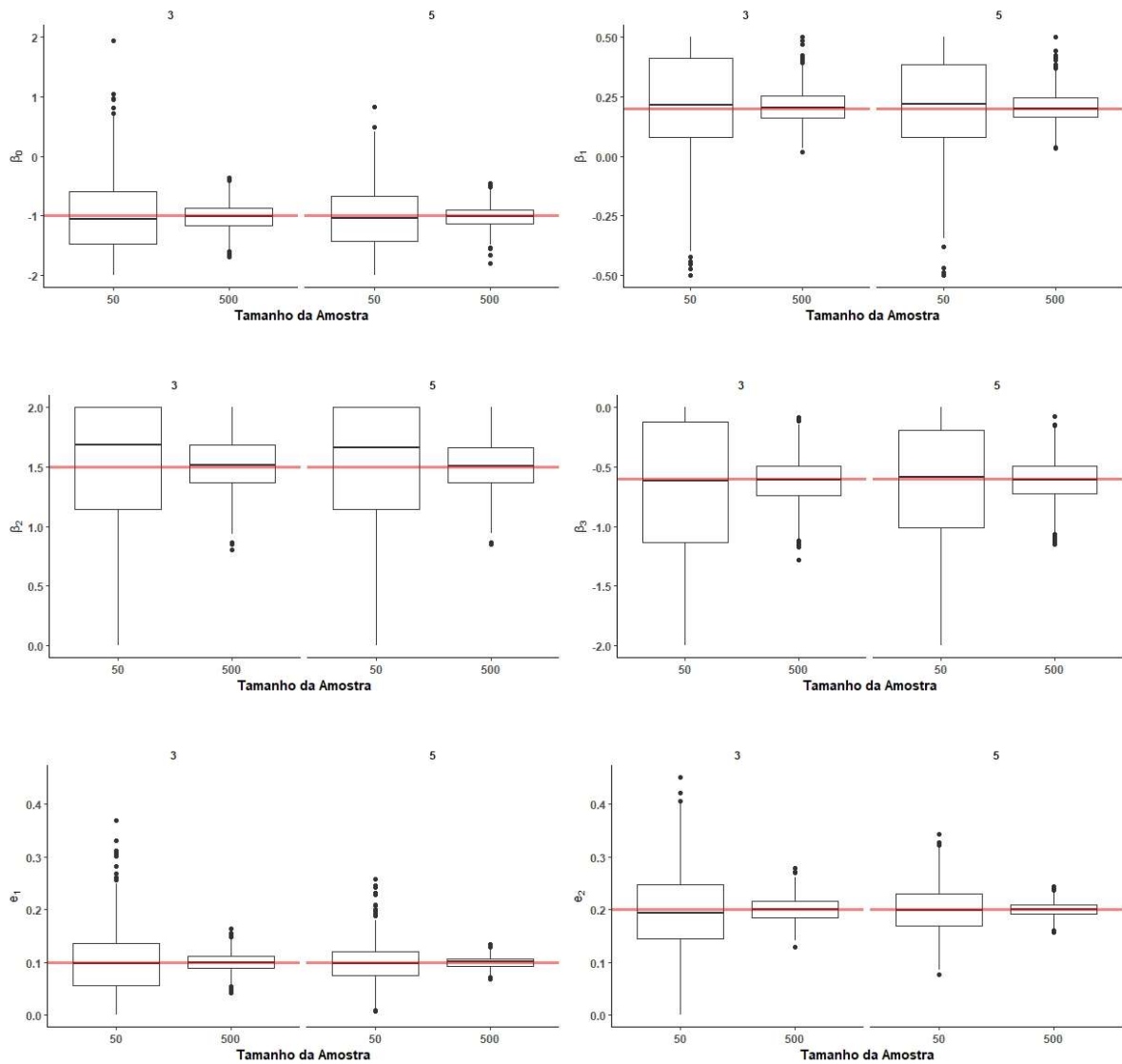


FIGURA 3: Diagramas de caixa dos parâmetros do modelo  $CECCR(m = [3; 5], n_r = n)$ .

TABELA 5

Comparação entre os modelos  $CECCR(m = 1, n_r = n, n = 500)$  e  $CECCR(m = 3, n_r = n, n = 500)$ .

	Real	$CECCR(m = 1, n_r = n)$				$CECCR(m = 3, n_r = n)$			
		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
$\beta_0$	-1,0	-1,1146	0,7018	-11,46	0,5052	-1,0194	0,2194	-1,94	0,0485
$\beta_1$	0,2	0,2321	0,1176	16,07	0,0148	0,2081	0,0707	4,05	0,0051
$\beta_2$	1,5	1,6445	0,3814	9,63	0,1662	1,5197	0,2302	1,31	0,0533
$\beta_3$	-0,6	-0,6971	0,3478	-16,19	0,1303	-0,6158	0,1911	-2,64	0,0367
$e_1$	0,1	0,0868	0,0961	-13,17	0,0094	0,1004	0,0170	0,40	0,0003
$e_2$	0,2	0,1804	0,1359	-9,80	0,0188	0,1997	0,0226	-0,15	0,0005

TABELA 6

Comparação entre os modelos SEC( $n = 500$ ) e CECCR( $m = 5, n_r = n, n = 500$ ).

	Real	SEC				CECCR( $m = 5, n_r = n$ )			
		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
$\beta_0$	-1,0	-1,0122	0,1752	-1,22	0,0308	-1,0155	0,1838	-1,55	0,0340
$\beta_1$	0,2	0,2064	0,0616	3,21	0,0038	0,2071	0,0652	3,57	0,0043
$\beta_2$	1,5	1,5126	0,2052	0,84	0,0422	1,5146	0,2143	0,97	0,0461
$\beta_3$	-0,6	-0,6115	0,1718	-1,92	0,0296	-0,6143	0,1767	-2,38	0,0314
$e_1$	0,1					0,1002	0,0102	0,25	0,0001
$e_2$	0,2					0,2001	0,0139	0,03	0,0002

Como visto, a situação CECCR( $m = 1, n_r = n$ ) apresenta EQMs superiores ao caso CECCR( $m = 3, n_r = n$ ). Contudo, o caso avaliado CECCR( $m = 3, n_r = n$ ) utilizou 1500 classificações totais, considerando que foi trabalhado com uma amostra  $n = 500$  e  $m = 3$ . Apesar disso, vale ressaltar que o caso CECCR( $m = 3, n_r = n, n = 50$ ) teve resultados inferenciais próximos ao CECCR( $m = 1, n_r = n, n = 1000$ ), ou seja, com maior tamanho amostral analisado, independente das taxas de erros de medida. Isso pode ser visto na TABELA 7.

A TABELA 8 demonstra o cenário CECCR( $m = 3, n_r = n, n = 100$ ) comparado com o CECCR( $m = 1, n_r = n, n = 1000$ ).

TABELA 7

Comparação entre os modelos CECCR( $m = 1, n_r = n, n = 1000$ ) e CECCR( $m = 3, n_r = n, n = 50$ ).

	Real	CECCR( $m = 1, n_r = n$ )				CECCR( $m = 3, n_r = n$ )			
		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
$\beta_0$	-1,0	-1,1114	0,6146	-11,14	0,3898	-1,0171	0,6230	-1,71	0,3880
$\beta_1$	0,2	0,2272	0,0980	13,59	0,0103	0,2125	0,2309	6,25	0,0534
$\beta_2$	1,5	1,6105	0,3849	7,37	0,1602	1,5122	0,5446	0,81	0,2964
$\beta_3$	-0,6	-0,6519	0,2602	-8,65	0,0703	-0,6953	0,5992	-15,88	0,3678
$e_1$	0,1	0,0875	0,0838	-12,49	0,0072	0,1000	0,0621	-0,04	0,0039
$e_2$	0,2	0,1820	0,1260	-8,98	0,0162	0,1984	0,0771	-0,81	0,0059

TABELA 8

Comparação entre os modelos CECCR( $m = 1, n_r = n, n = 1000$ ) e CECCR( $m = 3, n_r = n, n = 100$ ).

	Real	CECCR ( $m = 1, n_r = n$ )				CECCR( $m = 3, n_r = n$ )			
		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
$\beta_0$	-1,0	-1,1114	0,6146	-11,14	0,3898	-1,0481	0,4635	-4,81	0,2170
$\beta_1$	0,2	0,2272	0,0980	13,59	0,0103	0,2246	0,1728	12,31	0,0304
$\beta_2$	1,5	1,6105	0,3849	7,37	0,1602	1,5241	0,4397	1,61	0,1937
$\beta_3$	-0,6	-0,6519	0,2602	-8,65	0,0703	-0,6688	0,4333	-11,47	0,1923
$e_1$	0,1	0,0875	0,0838	-12,49	0,0072	0,0987	0,0403	-1,27	0,0016
$e_2$	0,2	0,1820	0,1260	-8,98	0,0162	0,1996	0,0523	-0,22	0,0027

Assim, quando possível, deve-se escolher a opção com reclassificação da amostra. A análise de 1000 itens no modelo CECCR( $m = 1, n_r = n$ ) teve resultados semelhantes às 300 análises ( $n * m = 100$  amostras \* 3 reclassificações de cada item) do modelo CECCR( $m = 3, n_r = n$ ).

Em face desse contexto, entende-se que o uso de classificações repetidas diminui a influência dos erros de classificação quanto ao aumento do vício e à variabilidade dos estimadores.

Como já mencionado, o uso do modelo CECCR( $m = [3; 5], n_r = n$ ) requer reclassificações para todos os  $n$  sujeitos, o que pode demandar tempo e recursos econômicos que não estão disponíveis. Neste sentido, avaliou-se o impacto da realização de classificações repetidas apenas em parte da amostra. Classificações repetidas nas frações de 30% e 70% da amostra foram consideradas CECCR( $m = [3; 5], n_r = [0,3n; 0,7n]$ ).

A TABELA 9 compara os desempenhos dos modelos nos cenários  $m = [1; 3], n = 500, e_1 = 0,10, e_2 = 0,20, n_r = [0,3n; 0,7n; n]$ . Como pode ser observado, os critérios de avaliação para caso CECCR( $m = 3, n_r = 0,7n$ ) são melhores que o caso CECCR( $m = 3, n_r = 0,3n$ ). Entretanto, as realizações de classificações repetidas parciais CECCR( $m = 3, n_r = [0,3n; 0,7n]$ ) tem resultados inferiores ao caso em que realizamos classificações repetidas em toda a amostra CECCR( $m = 3, n_r = n$ ). Contudo, constituem-se uma opção razoável, pois são melhores do que o caso CECCR( $m = 1, n_r = n$ ), ou seja, apresentam EQM, vício e dispersão inferiores.

TABELA 9

Comparação entre os modelos CECCR( $m = [1; 3]$ ,  $n_r = [0,3n; 0,7n; n]$ ,  $n = 500$ ).

	<b>Caso</b>	<b>Média</b>	<b>Desvio</b>	<b>Vício(%)</b>	<b>EQM</b>
<b><math>\beta_0 = -1,0</math></b>	CECCR( $m = 1, n_r = n$ )	-1,1146	0,7018	-11,46	0,5052
	CECCR( $m = 3, n_r = 0,3n$ )	-1,0390	0,3113	-3,90	0,0983
	CECCR( $m = 3, n_r = 0,7n$ )	-1,0256	0,2450	-2,56	0,0606
	CECCR( $m = 3, n_r = n$ )	-1,0194	0,2194	-1,94	0,0485
<b><math>\beta_1 = 0,2</math></b>	CECCR( $m = 1, n_r = n$ )	0,2321	0,1176	16,07	0,0148
	CECCR( $m = 3, n_r = 0,3n$ )	0,2101	0,0892	5,06	0,0081
	CECCR( $m = 3, n_r = 0,7n$ )	0,2091	0,0772	4,56	0,0060
	CECCR( $m = 3, n_r = n$ )	0,2081	0,0707	4,05	0,0051
<b><math>\beta_2 = 1,5</math></b>	CECCR( $m = 1, n_r = n$ )	1,6445	0,3814	9,63	0,1662
	CECCR( $m = 3, n_r = 0,3n$ )	1,5294	0,2802	1,96	0,0793
	CECCR( $m = 3, n_r = 0,7n$ )	1,5223	0,2512	1,49	0,0635
	CECCR( $m = 3, n_r = n$ )	1,5197	0,2302	1,31	0,0533
<b><math>\beta_3 = -0,6</math></b>	CECCR( $m = 1, n_r = n$ )	-0,6971	0,3478	-16,19	0,1303
	CECCR( $m = 3, n_r = 0,3n$ )	-0,6224	0,2353	-3,73	0,0558
	CECCR( $m = 3, n_r = 0,7n$ )	-0,6173	0,2045	-2,89	0,0421
	CECCR( $m = 3, n_r = n$ )	-0,6158	0,1911	-2,64	0,0367
<b><math>e_1 = 0,10</math></b>	CECCR( $m = 1, n_r = n$ )	0,0868	0,0961	-13,17	0,0094
	CECCR( $m = 3, n_r = 0,3n$ )	0,0976	0,0313	-2,39	0,0010
	CECCR( $m = 3, n_r = 0,7n$ )	0,0995	0,0201	-0,50	0,0004
	CECCR( $m = 3, n_r = n$ )	0,1004	0,0170	0,40	0,0003
<b><math>e_2 = 0,20</math></b>	CECCR( $m = 1, n_r = n$ )	0,1804	0,1359	-9,80	0,0188
	CECCR( $m = 3, n_r = 0,3n$ )	0,2018	0,0407	0,88	0,0017
	CECCR( $m = 3, n_r = 0,7n$ )	0,2010	0,0266	0,50	0,0007
	CECCR( $m = 3, n_r = n$ )	0,1997	0,0226	-0,15	0,0005

Com base em todas as simulações realizadas, a conclusão geral é que se deve utilizar, no mínimo, três classificações repetidas e em, pelo menos, 70% da amostra para minimizar a

influência dos erros de classificação nas estimações dos parâmetros da regressão logística. A simples incorporação dos erros de classificação na função de verossimilhança com uma única classificação torna o vício das estimativas aceitável, mas aumenta muito a variabilidade das estimativas, o que compromete o processo de inferência dos parâmetros do modelo. Para uma melhor compreensão, no Capítulo 4, um exemplo numérico é descrito, procurando enfatizar a necessidade da realização das classificações repetidas.



## 4. ANÁLISE DE UM BANCO DE DADOS REAL

O exemplo discutido nesta seção é o apresentado em Pires e Quinino (2018), quando o modelo logística de regressão com reclassificação na variável resposta foi avaliado sob a perspectiva Bayesiana.

Em termos mais específicos, relata-se que uma indústria que produz lentes para óculos contactou o Departamento de Estatística da UFMG para avaliar o aumento do número de lentes defeituosas após a diversificação da fonte de matéria-prima ("Região": 0 ou 1) e a utilização de um laboratório alternativo construído para a produção de lentes ("Laboratório": 0 ou 1). As lentes foram submetidas a uma inspeção visual para detectar defeitos como arranhões, rachaduras, espessura fora do padrão, entre outros. Se uma lente tiver um ou mais defeitos, ela foi classificada como "defeituoso"; caso contrário, foi classificada como "perfeito". Quatro inspetores participaram do estudo de avaliação e realizaram cinco classificações independentes repetidas de 400 lentes selecionadas aleatoriamente da produção. A seleção foi feita em quantidades iguais para cada combinação de fonte de matéria-prima (R) e laboratório (L) ( $R_0L_0 = R_0L_1 = R_1L_0 = R_1L_1 = 100$ ). A alocação das lentes para os inspetores foi randomizada. É bastante razoável considerar que o mecanismo de classificação não está relacionado com os covariáveis e as probabilidades de erro são as mesmas para todos os inspetores, uma vez que são periodicamente e igualmente treinados.

É descrito na TABELA 10 um resumo dos dados obtidos após cinco classificações independentes repetidas. Observou-se que, por exemplo, entre cinco classificações repetidas, 13 lentes  $L_0R_0$  não obtiveram classificação "perfeita", enquanto 34 lentes  $L_1R_1$  alcançaram 4 classificações "perfeitas".

Neste sentido, o objetivo é ajustar um modelo de regressão logística com as variáveis independentes dummies Região ( $R = 0, 1$ ) e laboratório ( $L = 0, 1$ ). Em outras palavras, desejamos estimar a probabilidade  $P(\tilde{Y}_i | \mathbf{X})$  dada por:

$$P(\widehat{Y}_i = 1) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 R + \hat{\beta}_2 L)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 R + \hat{\beta}_2 L)} \quad (11)$$

TABELA 10

Frequência de classificações como "perfeita" após 5 classificações.

Número de classificações como "perfeita"	Frequência			
	$L_0R_0$	$L_0R_1$	$L_1R_0$	$L_1R_1$
0	13	8	5	1
1	11	9	2	1
2	5	4	2	2
3	3	9	5	11
4	20	27	31	34
5	48	43	55	51

A estimação de todos os parâmetros está disponível na TABELA 11. Para melhor avaliação do efeito das classificações repetidas, considerou-se o caso em que todas as lentes foram classificadas independentemente cinco vezes  $CECCR(m = 5, n_r = n)$ , cuja análise utilizou apenas a primeira classificação  $CECCR(m = 1, n_r = n)$ , e o caso em que 70% da amostra foi reclassificada cinco vezes e o restante uma única vez  $CECCR(m = 5, n_r = 0,7n)$ , escolhidos aleatoriamente. As estimativas pontuais foram obtidas pelo estimador de máxima verossimilhança. Já o desvio-padrão e o intervalo de confiança foram obtidos utilizando a metodologia Bootstrap. A propósito, a descrição do procedimento Bootstrap utilizado neste trabalho é apresentado no ANEXO A.

TABELA 11

Estimativas dos parâmetros para os dados das lentes.

	$CECCR(m = 5, n_r = n)$				$CECCR(m = 5, n_r = 0,7n)$				$CECCR(m = 1, n_r = n)$			
	Média	Desvio	IC	Perc. (95%)	Média	Desvio	IC	Perc. (95%)	Média	Desvio	IC	Perc. (95%)
<i>Interc.</i>	0,8610	0,2160	0,4530	1,3100	0,8370	0,2366	0,4260	1,3500	0,9405	0,5625	-0,0443	2,0100
<i>Lab.</i>	1,7937	0,4113	1,0800	2,6900	1,9144	0,4865	1,0900	3,0600	2,7496	1,3687	0,6820	4,0000
<i>Reg.</i>	0,5722	0,3207	-0,0339	1,2300	0,3913	0,3382	-0,2810	1,0400	0,3866	0,6952	-0,6380	1,8900
$e_1$	0,1041	0,0081	0,0885	0,1200	0,1060	0,0100	0,0870	0,1260	0,0826	0,0637	0,0000	0,1760
$e_2$	0,1519	0,0262	0,1050	0,2060	0,1660	0,0303	0,1110	0,2300	0,1314	0,0163	0,0000	0,3500

É possível observar que a não realização de reclassificações dos itens aumenta muito o desvio padrão, podendo comprometer o processo de inferência. Logo, sugere-se um esforço para realizar classificações repetidas em, pelo menos, parte da amostra. Aliás, pode-se encontrar o intervalo de confiança percentílico Bootstrap para probabilidade de lentes perfeitas para cada um dos quatro cruzamentos entre Região e Laboratório. Para tanto, usam-se os intervalos dos parâmetros indicados na TABELA 11 e a expressão (11). Baseado no estudo de simulação, podemos considerar que o caso com  $CECCR(m = 5, n_r = n)$  é a melhor alternativa. Em relação ao uso  $CECCR(m = 1, n_r = n)$ , deve ser feito com bastante prudência, pois apresenta grande variabilidade de resultados, o que pode levar a decisões equivocadas sobre o percentual de lentes perfeitas, considerando o laboratório e a região.

TABELA 12

Estimativas das probabilidades de se obter lentes “perfeitas”.

	CECCR( $m = 5, n_r = n$ )			CECCR( $m = 5, n_r = 0,7n$ )			CECCR( $m = 1, n_r = n$ )		
	Probabili- dade	IC Perc. (95%)		Probabili- dade	IC Perc. (95%)		Probabili- dade	IC Perc. (95%)	
$L_0R_0$	0,7028	0,6113	0,7875	0,6978	0,6049	0,7912	0,7192	0,4889	0,8818
$L_0R_1$	0,8073	0,6032	0,9268	0,7735	0,5361	0,9160	0,7903	0,3357	0,9801
$L_1R_0$	0,9343	0,8224	0,9820	0,9399	0,8199	0,9879	0,9756	0,6542	0,9975
$L_1R_1$	0,9618	0,8174	0,9946	0,9586	0,7746	0,9957	0,9833	0,4999	0,9996

## 5. COMENTÁRIOS FINAIS

Constatou-se que, ao incorporar os parâmetros de erros de classificação na função de verossimilhança de um modelo de regressão logística com erros na variável resposta, os estimadores podem apresentar grande variabilidade nas estimativas, apesar de serem pouco viciados quando comparados a situação que se ignora os erros.

O uso de reclassificações das unidades amostrais ou, pelo menos, em parte delas aproxima os valores das estimativas para o caso sem erro na variável resposta. Entende-se, então, que é recomendável insistir na obtenção de informação adicional sobre os erros de classificação para minimizar sua influência.

Demonstrou-se que o modelo  $CECCR(m = 3, n_r = n)$  traz resultados semelhantes ao modelo  $CECCR(m = 1, n_r = n)$ , mesmo quando o primeiro tem tamanho amostral bem menor. Se o número de classificações é equivalente ( $n * m$ ) em ambos os casos, o  $CECCR(m = 3, n_r = n)$  é superior em todos os critérios de avaliação. Assim, do ponto de vista econômico, o modelo com reclassificação será, também, a escolha.

A metodologia utilizada é bastante genérica e pode ser usada para abordar uma variedade de problemas práticos.

Para um trabalho futuro, duas generalizações são necessárias para atender situações práticas: (1) considerar os erros de classificação como diferenciais (dependendo ou não de variáveis preditoras); (2) a variável resposta pode apresentar mais de duas categorias e os erros de classificação podem diferir para cada categoria, bem como serem influenciados ou não pelas variáveis preditoras.

## REFERÊNCIAS

- BLISCHKE, W. R. Estimating the Parameters of Mixtures of Binomial Distributions. *Journal of the American Statistical Association*, v. 59, n. 306, p. 510–528, jun. 1964.
- BLISCHKE, W. R. Moment Estimators for the Parameters of a Mixture of Two Binomial Distributions. *The Annals of Mathematical Statistics*, v. 33, n. 2, p. 444–454, jun. 1962.
- BOLLINGER, C. R.; DAVID, M. H. Modeling Discrete Choice with Response Error: Food Stamp Participation. *Journal of the American Statistical Association*, v. 92, n. 439, p. 827–835, set. 1997.
- CASELLA, G.; BERGER, R. L. *Statistical inference*. 2nd ed ed. Australia; Pacific Grove, CA: Thomson Learning, 2002.
- CHENG, K. F.; HSUEH, H. M. Estimation of a Logistic Regression Model with Mismeasured Observations. *Statistica Sinica*, v. 13, n. 1, p. 111–127, 2003.
- DAVISON, A. C.; HINKLEY, D. V. *Bootstrap methods and their application*. Cambridge: Cambridge University Press, 1997.
- DENNIS, J.; GAY, D.; WELSCH, R. An Adaptive Nonlinear Least Square Algorithm., no w0196. Cambridge, MA: National Bureau of Economic Research, ago. 1977. Disponível em: <<http://www.nber.org/papers/w0196.pdf>>. Acesso em: 12 dez. 2018.
- EFRON, B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, v. 7, n. 1, p. 1–26, 1979.
- EFRON, B.; HASTIE, T. *Computer age statistical inference: algorithms, evidence, and data science*. New York, NY: Cambridge University Press, 2016. (Institute of Mathematical Statistics monographs).
- GERLACH, R.; STAMEY, J. Bayesian model selection for logistic regression with misclassified outcomes. *Statistical Modelling: An International Journal*, v. 7, n. 3, p. 255–273, out. 2007.
- GUSTAFSON, P. *Measurement error and misclassification in statistics and epidemiology: impacts and Bayesian adjustments*. Boca Raton: Chapman & Hall/CRC, 2004.

- HAUSMAN, J.; ABREVAYA, J.; SCOTT-MORTON, F. M. Misclassification of the dependent variable in a discrete-response setting. *Journal of Econometrics*, v. 87, n. 2, p. 239–269, 1998.
- LIU, H.; ZHANG, Z. Logistic regression with misclassification in binary outcome variables: a method and software. *Behaviormetrika*, v. 44, n. 2, p. 447–476, 1 jul. 2017.
- MAGDER, L. S.; HUGHES, J. P. Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology*, v. 146, n. 2, p. 195–203, 15 jul. 1997.
- MANSKI, C. F. Identification of Binary Response Models. *Journal of the American Statistical Association*, v. 83, n. 403, p. 729–738, 1988.
- MCINTURFF, P. et al. Modelling risk when binary outcomes are subject to error. *Statistics in Medicine*, v. 23, n. 7, p. 1095–1109, 15 abr. 2004.
- MEYER, B. D.; MITTAG, N. Misclassification in binary choice models. *Journal of Econometrics*, v. 200, n. 2, p. 295–311, out. 2017.
- NARANJO, L. et al. Addressing misclassification for binary data: probit and t-link regressions. *Journal of Statistical Computation and Simulation*, v. 84, n. 10, p. 2187–2213, 3 out. 2014.
- NEWKEY, W. K.; MCFADDEN, D. Chapter 36 Large sample estimation and hypothesis testing. *Handbook of Econometrics*. [S.l.]: Elsevier, 1994. v. 4. p. 2111–2245. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1573441205800054>>. Acesso em: 8 nov. 2018.
- PAULINO, C. D.; SOARES, P.; NEUHAUS, J. Binomial regression with misclassification. *Biometrics*, v. 59, n. 3, p. 670–675, set. 2003.
- PIRES, M. C.; QUININO, R. C. Repeated responses in misclassification binary regression: A Bayesian approach. *Statistical Modelling*, p. 1471082X1877339, 11 jun. 2018.

## **APÊNDICE A – Resultados das simulações**

Nas tabelas abaixo, estão os resultados de todas as simulações efetuadas e estudadas.

TABELA 13

Simulações para o modelo regressão logística sem erros de classificação na variável resposta (SEC).

Tamanho amostra	Erros de Medida		$\beta_0 = 1,0$				$\beta_1 = 0,2$				$\beta_2 = 1,5$				$\beta_3 = -0,6$			
	$e_1$	$e_2$	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
50	0,05	0,05	-1,1636	0,6591	-16,36	0,4608	0,2717	0,2772	35,83	0,0819	1,6406	0,7332	9,37	0,5568	-0,6719	0,6485	-11,98	0,4253
50		0,10	-1,1071	0,6647	-10,71	0,4528	0,2472	0,2852	23,59	0,0835	1,6674	0,7442	11,16	0,5813	-0,6698	0,6333	-11,64	0,4055
50		0,20	-1,1875	0,6841	-18,75	0,5026	0,2649	0,2959	32,44	0,0917	1,6858	0,7390	12,39	0,5802	-0,6460	0,6537	-7,66	0,4291
50	0,10	0,05	-1,1636	0,6591	-16,36	0,4608	0,2717	0,2772	35,83	0,0819	1,6406	0,7332	9,37	0,5568	-0,6719	0,6485	-11,98	0,4253
50		0,10	-1,1071	0,6647	-10,71	0,4528	0,2472	0,2852	23,59	0,0835	1,6674	0,7442	11,16	0,5813	-0,6698	0,6333	-11,64	0,4055
50		0,20	-1,1875	0,6841	-18,75	0,5026	0,2649	0,2959	32,44	0,0917	1,6858	0,7390	12,39	0,5802	-0,6460	0,6537	-7,66	0,4291
50	0,20	0,05	-1,1636	0,6591	-16,36	0,4608	0,2717	0,2772	35,83	0,0819	1,6406	0,7332	9,37	0,5568	-0,6719	0,6485	-11,98	0,4253
50		0,10	-1,1071	0,6647	-10,71	0,4528	0,2472	0,2852	23,59	0,0835	1,6674	0,7442	11,16	0,5813	-0,6698	0,6333	-11,64	0,4055
50		0,20	-1,1875	0,6841	-18,75	0,5026	0,2649	0,2959	32,44	0,0917	1,6858	0,7390	12,39	0,5802	-0,6460	0,6537	-7,66	0,4291
100	0,05	0,05	-1,0724	0,4088	-7,24	0,1722	0,2299	0,1706	14,94	0,0300	1,5701	0,4668	4,67	0,2226	-0,6321	0,4064	-5,34	0,1661
100		0,10	-1,0646	0,4267	-6,46	0,1861	0,2267	0,1581	13,37	0,0257	1,5616	0,4956	4,11	0,2492	-0,6354	0,4361	-5,90	0,1913
100		0,20	-1,0817	0,4346	-8,17	0,1953	0,2329	0,1693	16,46	0,0297	1,5833	0,4796	5,55	0,2368	-0,6461	0,4206	-7,68	0,1789
100	0,10	0,05	-1,0724	0,4088	-7,24	0,1722	0,2299	0,1706	14,94	0,0300	1,5701	0,4668	4,67	0,2226	-0,6321	0,4064	-5,34	0,1661
100		0,10	-1,0646	0,4267	-6,46	0,1861	0,2267	0,1581	13,37	0,0257	1,5616	0,4956	4,11	0,2492	-0,6354	0,4361	-5,90	0,1913
100		0,20	-1,0817	0,4346	-8,17	0,1953	0,2329	0,1693	16,46	0,0297	1,5833	0,4796	5,55	0,2368	-0,6461	0,4206	-7,68	0,1789
100	0,20	0,05	-1,0724	0,4088	-7,24	0,1722	0,2299	0,1706	14,94	0,0300	1,5701	0,4668	4,67	0,2226	-0,6321	0,4064	-5,34	0,1661
100		0,10	-1,0646	0,4267	-6,46	0,1861	0,2267	0,1581	13,37	0,0257	1,5616	0,4956	4,11	0,2492	-0,6354	0,4361	-5,90	0,1913
100		0,20	-1,0817	0,4346	-8,17	0,1953	0,2329	0,1693	16,46	0,0297	1,5833	0,4796	5,55	0,2368	-0,6461	0,4206	-7,68	0,1789
500	0,05	0,05	-1,0100	0,1712	-1,00	0,0294	0,2055	0,0644	2,76	0,0042	1,5084	0,1976	0,56	0,0391	-0,6000	0,1674	0,01	0,0280
500		0,10	-1,0216	0,1745	-2,16	0,0309	0,2095	0,0642	4,74	0,0042	1,5115	0,1967	0,77	0,0388	-0,6001	0,1743	-0,01	0,0304
500		0,20	-1,0122	0,1752	-1,22	0,0308	0,2064	0,0616	3,21	0,0038	1,5126	0,2052	0,84	0,0422	-0,6115	0,1718	-1,92	0,0296
500	0,10	0,05	-1,0100	0,1712	-1,00	0,0294	0,2055	0,0644	2,76	0,0042	1,5084	0,1976	0,56	0,0391	-0,6000	0,1674	0,01	0,0280
500		0,10	-1,0216	0,1745	-2,16	0,0309	0,2095	0,0642	4,74	0,0042	1,5115	0,1967	0,77	0,0388	-0,6001	0,1743	-0,01	0,0304
500		0,20	-1,0122	0,1752	-1,22	0,0308	0,2064	0,0616	3,21	0,0038	1,5126	0,2052	0,84	0,0422	-0,6115	0,1718	-1,92	0,0296
500	0,20	0,05	-1,0100	0,1712	-1,00	0,0294	0,2055	0,0644	2,76	0,0042	1,5084	0,1976	0,56	0,0391	-0,6000	0,1674	0,01	0,0280
500		0,10	-1,0216	0,1745	-2,16	0,0309	0,2095	0,0642	4,74	0,0042	1,5115	0,1967	0,77	0,0388	-0,6001	0,1743	-0,01	0,0304
500		0,20	-1,0122	0,1752	-1,22	0,0308	0,2064	0,0616	3,21	0,0038	1,5126	0,2052	0,84	0,0422	-0,6115	0,1718	-1,92	0,0296
1000	0,05	0,05	-1,0046	0,1163	-0,46	0,0135	0,2018	0,0422	0,91	0,0018	1,4995	0,1320	-0,04	0,0174	-0,6075	0,1249	-1,25	0,0156
1000		0,10	-1,0095	0,1211	-0,95	0,0148	0,2039	0,0426	1,93	0,0018	1,5052	0,1415	0,35	0,0200	-0,6009	0,1222	-0,14	0,0149
1000		0,20	-1,0102	0,1198	-1,02	0,0144	0,2037	0,0435	1,87	0,0019	1,5103	0,1417	0,68	0,0202	-0,6001	0,1184	-0,01	0,0140
1000	0,10	0,05	-1,0046	0,1163	-0,46	0,0135	0,2018	0,0422	0,91	0,0018	1,4995	0,1320	-0,04	0,0174	-0,6075	0,1249	-1,25	0,0156
1000		0,10	-1,0095	0,1211	-0,95	0,0148	0,2039	0,0426	1,93	0,0018	1,5052	0,1415	0,35	0,0200	-0,6009	0,1222	-0,14	0,0149
1000		0,20	-1,0102	0,1198	-1,02	0,0144	0,2037	0,0435	1,87	0,0019	1,5103	0,1417	0,68	0,0202	-0,6001	0,1184	-0,01	0,0140
1000	0,20	0,05	-1,0046	0,1163	-0,46	0,0135	0,2018	0,0422	0,91	0,0018	1,4995	0,1320	-0,04	0,0174	-0,6075	0,1249	-1,25	0,0156
1000		0,10	-1,0095	0,1211	-0,95	0,0148	0,2039	0,0426	1,93	0,0018	1,5052	0,1415	0,35	0,0200	-0,6009	0,1222	-0,14	0,0149
1000		0,20	-1,0102	0,1198	-1,02	0,0144	0,2037	0,0435	1,87	0,0019	1,5103	0,1417	0,68	0,0202	-0,6001	0,1184	-0,01	0,0140



TABELA 14

Simulações para o modelo de regressão logística com erros de classificação na variável resposta e ignorados (IEC).

Tamanho amostra	Erros de Medida		$\beta_0 = 1,0$				$\beta_1 = 0,2$				$\beta_2 = 1,5$				$\beta_3 = -0,6$			
	$e_1$	$e_2$	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
50	0,05	0,05	-1,0279	0,6613	-2,79	0,4376	0,2394	0,2760	19,68	0,0776	1,4441	0,7217	-3,73	0,5234	-0,5815	0,6179	3,09	0,3818
50		0,10	-0,8009	0,6120	19,91	0,4139	0,2038	0,2728	1,90	0,0743	1,3855	0,7158	-7,64	0,5249	-0,5497	0,6602	8,38	0,4380
50		0,20	-0,5476	0,6239	45,24	0,5936	0,2191	0,2990	9,53	0,0897	1,2694	0,6988	-15,37	0,5411	-0,4425	0,6375	26,26	0,4308
50	0,10	0,05	-1,0696	0,6576	-6,96	0,4369	0,2092	0,2653	4,62	0,0704	1,3653	0,7059	-8,98	0,5159	-0,5521	0,6061	7,98	0,3693
50		0,10	-0,8469	0,6078	15,31	0,3924	0,1766	0,2663	-11,71	0,0714	1,2746	0,6944	-15,03	0,5326	-0,5059	0,6447	15,68	0,4241
50		0,20	-0,5792	0,6162	42,08	0,5564	0,1899	0,2897	-5,03	0,0839	1,1554	0,6872	-22,97	0,5905	-0,3925	0,6238	34,58	0,4318
50	0,20	0,05	-1,1896	0,6644	-18,96	0,4770	0,1730	0,2632	-13,49	0,0699	1,2047	0,7003	-19,69	0,5771	-0,4734	0,6062	21,10	0,3831
50		0,10	-0,9658	0,6081	3,42	0,3706	0,1420	0,2554	-29,02	0,0685	1,0967	0,6875	-26,88	0,6349	-0,4347	0,6248	27,55	0,4173
50		0,20	-0,6665	0,5986	33,35	0,4692	0,1458	0,2716	-27,08	0,0766	0,9536	0,6614	-36,42	0,7355	-0,3165	0,6081	47,25	0,4498
100	0,05	0,05	-0,9415	0,3842	5,85	0,1509	0,1964	0,1645	-1,81	0,0270	1,3926	0,4517	-7,16	0,2154	-0,5506	0,4023	8,23	0,1641
100		0,10	-0,7664	0,3972	23,36	0,2122	0,1879	0,1558	-6,06	0,0244	1,2976	0,4624	-13,49	0,2546	-0,5252	0,4017	12,47	0,1668
100		0,20	-0,4639	0,4019	53,61	0,4488	0,1813	0,1661	-9,33	0,0279	1,1781	0,4646	-21,46	0,3193	-0,4637	0,4041	22,72	0,1817
100	0,10	0,05	-0,9926	0,3822	0,74	0,1460	0,1787	0,1579	-10,65	0,0254	1,3032	0,4438	-13,12	0,2355	-0,5141	0,4049	14,31	0,1711
100		0,10	-0,8184	0,3981	18,16	0,1913	0,1714	0,1497	-14,31	0,0232	1,2061	0,4484	-19,60	0,2872	-0,4875	0,3888	18,76	0,1637
100		0,20	-0,5079	0,3964	49,21	0,3992	0,1587	0,1577	-20,66	0,0265	1,0741	0,4517	-28,39	0,3852	-0,4225	0,3937	29,58	0,1863
100	0,20	0,05	-1,1055	0,3888	-10,55	0,1622	0,1472	0,1519	-26,39	0,0258	1,1471	0,4328	-23,53	0,3116	-0,4461	0,3976	25,66	0,1816
100		0,10	-0,9219	0,3938	7,81	0,1610	0,1343	0,1420	-32,87	0,0245	1,0555	0,4358	-29,64	0,3874	-0,4218	0,3865	29,70	0,1810
100		0,20	-0,5956	0,3862	40,44	0,3126	0,1245	0,1508	-37,75	0,0284	0,8962	0,4357	-40,25	0,5542	-0,3473	0,3917	42,12	0,2172
500	0,05	0,05	-0,8836	0,1719	11,64	0,0431	0,1742	0,0620	-12,91	0,0045	1,3307	0,1952	-11,28	0,0667	-0,5208	0,1661	13,20	0,0338
500		0,10	-0,7286	0,1674	27,14	0,1017	0,1704	0,0649	-14,78	0,0051	1,2503	0,1916	-16,65	0,0991	-0,4870	0,1673	18,84	0,0407
500		0,20	-0,4106	0,1710	58,94	0,3766	0,1505	0,0624	-24,77	0,0063	1,1195	0,2042	-25,37	0,1865	-0,4430	0,1637	26,16	0,0514
500	0,10	0,05	-0,9290	0,1714	7,10	0,0344	0,1545	0,0626	-22,77	0,0060	1,2462	0,1934	-16,92	0,1017	-0,4849	0,1646	19,18	0,0403
500		0,10	-0,7725	0,1668	22,75	0,0795	0,1514	0,0629	-24,32	0,0063	1,1593	0,1924	-22,71	0,1530	-0,4510	0,1633	24,83	0,0488
500		0,20	-0,4507	0,1705	54,93	0,3308	0,1293	0,0600	-35,37	0,0086	1,0226	0,2006	-31,83	0,2681	-0,4000	0,1626	33,33	0,0664
500	0,20	0,05	-1,0449	0,1701	-4,49	0,0309	0,1275	0,0600	-36,23	0,0088	1,1045	0,1931	-26,36	0,1936	-0,4250	0,1656	29,16	0,0580
500		0,10	-0,8750	0,1669	12,50	0,0434	0,1216	0,0580	-39,20	0,0095	1,0045	0,1905	-33,03	0,2818	-0,3852	0,1617	35,80	0,0723
500		0,20	-0,5471	0,1662	45,29	0,2327	0,0991	0,0544	-50,43	0,0131	0,8531	0,1940	-43,13	0,4561	-0,3315	0,1641	44,75	0,0990
1000	0,05	0,05	-0,8780	0,1141	12,20	0,0279	0,1713	0,0409	-14,33	0,0025	1,3208	0,1293	-11,95	0,0488	-0,5304	0,1210	11,60	0,0195
1000		0,10	-0,7208	0,1146	27,92	0,0911	0,1636	0,0418	-18,19	0,0031	1,2496	0,1366	-16,69	0,0813	-0,4856	0,1190	19,06	0,0272
1000		0,20	-0,4094	0,1143	59,06	0,3619	0,1458	0,0431	-27,12	0,0048	1,1193	0,1440	-25,38	0,1656	-0,4284	0,1179	28,60	0,0433
1000	0,10	0,05	-0,9243	0,1151	7,57	0,0190	0,1527	0,0415	-23,64	0,0040	1,2349	0,1285	-17,68	0,0868	-0,4915	0,1204	18,09	0,0263
1000		0,10	-0,7634	0,1140	23,66	0,0690	0,1438	0,0407	-28,08	0,0048	1,1591	0,1343	-22,72	0,1342	-0,4483	0,1169	25,28	0,0367
1000		0,20	-0,4506	0,1132	54,94	0,3147	0,1271	0,0412	-36,46	0,0070	1,0174	0,1392	-32,18	0,2523	-0,3881	0,1157	35,31	0,0583
1000	0,20	0,05	-1,0376	0,1156	-3,76	0,0148	0,1240	0,0392	-38,02	0,0073	1,0932	0,1310	-27,12	0,1826	-0,4276	0,1192	28,73	0,0439
1000		0,10	-0,8671	0,1173	13,29	0,0314	0,1160	0,0382	-42,02	0,0085	1,0066	0,1367	-32,89	0,2621	-0,3861	0,1171	35,65	0,0594
1000		0,20	-0,5469	0,1129	45,31	0,2181	0,0977	0,0390	-51,15	0,0120	0,8480	0,1343	-43,47	0,4431	-0,3175	0,1131	47,08	0,0926

TABELA 15

Simulações para o modelo de regressão logística com verossimilhança modificada para acomodar erros de medidas CECCR( $m = [1; 3; 5]$ ,  $n_r = n$ ).

Tamanho amostra	Erros de Medida		Reclassificações (m)	$\beta_0 = 1,0$				$\beta_1 = 0,2$				$\beta_2 = 1,5$				$\beta_3 = -0,6$				$e_1$				$e_2$					
	$e_1$	$e_2$		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio
50	0,05	0,05	1	-1,1742	0,7553	-17,42	0,6002	0,2526	0,2614	26,32	0,0710	1,6691	0,4870	11,27	0,2655	-0,8654	0,6743	-44,24	0,5247	0,0699	0,1145	39,76	0,0135	0,0824	0,1187	64,90	0,0151		
50	0,05	0,05	3	-1,0363	0,5140	-3,63	0,2652	0,2310	0,2034	15,49	0,0423	1,4911	0,5094	-0,59	0,2593	-0,6765	0,5301	-12,76	0,2865	0,0506	0,0302	1,13	0,0009	0,0504	0,0320	0,84	0,0010		
50	0,05	0,05	5	-1,0298	0,5006	-2,98	0,2512	0,2310	0,1997	15,49	0,0408	1,4839	0,5047	-1,07	0,2547	-0,6715	0,5219	-11,92	0,2772	0,0511	0,0206	2,18	0,0004	0,0494	0,0205	-1,12	0,0004		
50	0,05	0,10	1	-1,0804	0,8126	-8,04	0,6661	0,2235	0,2785	11,73	0,0780	1,6454	0,5179	9,70	0,2891	-0,8779	0,6989	-46,32	0,5652	0,0578	0,0991	15,61	0,0099	0,1072	0,1371	7,16	0,0188		
50	0,10	0,10	3	-0,9795	0,5306	2,05	0,2816	0,2060	0,2163	2,98	0,0468	1,5039	0,5032	0,26	0,2530	-0,6951	0,5396	-15,85	0,2999	0,0499	0,0336	-0,26	0,0011	0,0981	0,0460	-1,88	0,0021		
50	0,10	0,10	5	-0,9727	0,5038	2,73	0,2543	0,2092	0,2041	4,59	0,0417	1,4923	0,5014	-0,51	0,2512	-0,6764	0,5194	-12,73	0,2754	0,0504	0,0208	0,86	0,0004	0,0979	0,0300	-2,15	0,0009		
50	0,05	0,20	1	-0,9150	0,8864	8,50	0,7921	0,2237	0,2845	11,85	0,0814	1,6106	0,5611	7,37	0,3268	-0,8185	0,7076	-36,41	0,5479	0,0526	0,0942	5,24	0,0089	0,1378	0,1625	-31,08	0,0302		
50	0,20	0,20	3	-1,0336	0,5633	-3,36	0,3181	0,2170	0,2206	8,49	0,0489	1,5156	0,5300	1,04	0,2809	-0,6716	0,5687	-11,93	0,3282	0,0507	0,0423	1,36	0,0018	0,1997	0,0691	-0,14	0,0048		
50	0,20	0,20	5	-1,0425	0,5122	-4,25	0,2639	0,2032	0,2193	10,65	0,0417	1,5156	0,5048	1,04	0,2549	-0,6567	0,5342	-9,44	0,2883	0,0503	0,0230	0,57	0,0005	0,1999	0,0417	-0,05	0,0017		
50	0,10	0,05	1	-1,2028	0,7914	-20,28	0,6669	0,2266	0,2769	13,31	0,0773	1,6419	0,5154	9,46	0,2855	-0,8713	0,6793	-45,22	0,5346	0,0855	0,1273	-14,51	0,0164	0,0793	0,1143	58,65	0,0139		
50	0,05	0,05	3	-1,0331	0,5222	-3,31	0,2735	0,2316	0,2049	15,80	0,0429	1,4887	0,5176	-0,75	0,2678	-0,6865	0,5358	-14,42	0,2943	0,0990	0,0434	-1,00	0,0019	0,0500	0,0340	-0,09	0,0012		
50	0,05	0,05	5	-1,0273	0,5028	-2,73	0,2533	0,2305	0,1999	15,23	0,0408	1,4824	0,5078	-1,17	0,2579	-0,6759	0,5237	-12,65	0,2798	0,1004	0,0291	0,42	0,0008	0,0494	0,0212	-1,21	0,0005		
50	0,10	0,10	1	-1,1119	0,8422	-11,19	0,7211	0,2110	0,2882	5,52	0,0831	1,5936	0,5555	6,24	0,3170	-0,8494	0,7034	-41,57	0,5565	0,0762	0,1170	-23,81	0,0142	0,1020	0,1310	1,98	0,0172		
50	0,10	0,10	3	-0,9696	0,5517	3,04	0,3050	0,2055	0,2209	2,77	0,0488	1,4977	0,5157	-0,16	0,2567	-0,7071	0,5591	-17,85	0,3237	0,1018	0,0477	1,83	0,0023	0,0975	0,0498	-2,50	0,0025		
50	0,10	0,10	5	-0,9709	0,5085	2,91	0,2591	0,2086	0,2059	4,30	0,0424	1,4941	0,5047	-0,39	0,2545	-0,6831	0,5253	-13,84	0,2826	0,1018	0,0288	1,77	0,0008	0,0977	0,0317	-2,32	0,0010		
50	0,10	0,20	1	-0,9372	0,9329	6,28	0,8733	0,2086	0,2987	4,29	0,0892	1,5571	0,5997	3,81	0,3625	-0,7964	0,7183	-32,74	0,5540	0,0718	0,1126	-28,18	0,0135	0,1361	0,1585	-31,93	0,0292		
50	0,20	0,20	3	-1,0171	0,6230	-1,71	0,3880	0,2125	0,2309	6,25	0,0534	1,5122	0,5446	0,81	0,2964	-0,6953	0,5992	-15,88	0,3678	0,1000	0,0621	-0,04	0,0039	0,1984	0,0771	-0,81	0,0059		
50	0,20	0,20	5	-1,0387	0,5288	-3,87	0,2809	0,2174	0,2080	8,71	0,0435	1,5114	0,5147	0,76	0,2648	-0,6627	0,5459	-10,46	0,3017	0,0995	0,0347	-0,45	0,0012	0,2003	0,0452	0,17	0,0020		
50	0,20	0,05	1	-1,2664	0,8082	-26,64	0,7234	0,2045	0,2977	2,26	0,0886	1,5130	0,5977	0,86	0,3571	-0,8343	0,7033	-39,05	0,5490	0,1206	0,1553	-39,68	0,0304	0,0704	0,1032	40,80	0,0111		
50	0,05	0,3	1	-1,0318	0,5674	-3,18	0,3226	0,2294	0,2163	14,71	0,0476	1,4873	0,5277	-0,84	0,2784	-0,6990	0,5596	-16,50	0,3227	0,1981	0,0654	-0,94	0,0043	0,0496	0,0421	-0,81	0,0018		
50	0,05	0,5	1	-1,0255	0,5123	-2,55	0,2628	0,2290	0,2039	14,50	0,0424	1,4805	0,5081	-1,30	0,2583	-0,6800	0,5343	-13,34	0,2916	0,2011	0,0417	0,53	0,0017	0,0499	0,0239	-0,19	0,0006		
50	0,20	0,10	1	-1,1358	0,9117	-13,58	0,8488	0,1830	0,3114	-8,49	0,0972	1,4923	0,6235	-0,52	0,3885	-0,8425	0,7200	-40,42	0,5767	0,1230	0,1524	-38,51	0,0291	0,0913	0,1180	-8,67	0,0140		
50	0,10	0,3	1	-0,9586	0,6380	4,14	0,4084	0,2041	0,2383	2,07	0,0568	1,4775	0,5578	-1,50	0,3113	-0,7443	0,6018	-24,05	0,3826	0,1995	0,0717	-0,23	0,0051	0,0962	0,0605	-3,76	0,0037		
50	0,10	0,3	3	-0,9608	0,5311	3,92	0,2833	0,2025	0,2152	1,26	0,0463	1,4844	0,5262	-1,04	0,2768	-0,6898	0,5449	-14,97	0,3047	0,2022	0,0427	1,08	0,0018	0,0979	0,0357	-2,06	0,0013		
50	0,20	0,20	1	-0,9618	0,9983	3,82	0,9971	0,1848	0,3193	-7,61	0,1021	1,4270	0,6593	-4,87	0,4396	-0,7830	0,7344	-30,50	0,5724	0,1109	0,1445	-44,54	0,0288	0,1241	0,1457	-37,93	0,0270		
50	0,20	0,3	1	-0,9510	0,7653	4,90	0,5874	0,2000	0,2635	0,01	0,0693	1,4924	0,5934	-0,50	0,3519	-0,7321	0,6528	-22,02	0,4432	0,1987	0,0871	-0,67	0,0076	0,1896	0,0917	-5,20	0,0085		
50	0,20	0,5	1	-1,0279	0,5793	-2,79	0,3361	0,2180	0,2209	9,02	0,0491	1,5095	0,5367	0,63	0,2879	-0,6942	0,5871	-15,70	0,3532	0,2006	0,0510	0,28	0,0026	0,1987	0,0525	-0,65	0,0027		
100	0,05	0,1	1	-1,2609	0,6627	-26,09	0,5068	0,2509	0,1995	25,43	0,0424	1,7561	0,3529	17,07	0,1900	-0,8235	0,5511	-37,25	0,3534	0,0667	0,1024	33,43	0,0107	0,0942	0,1077	88,43	0,0136		
100	0,05	0,3	1	-1,0320	0,3699	-3,20	0,1377	0,1520	0,1922	9,62	0,0235	1,5198	0,3803	1,32	0,1449	-0,6345	0,3883	-5,76	0,1518	0,0494	0,0203	-1,21	0,0004	0,0489	0,0207	-2,20	0,0004		
100	0,05	0,5	1	-1,0331	0,3639	-3,31	0,1334	0,2193	0,1500	9,67	0,0228	1,5170	0,3756	1,13	0,1412	-0,6331	0,3818	-5,51	0,1467	0,0494	0,0141	-1,14	0,0002	0,0492	0,0142	-1,57	0,0002		
100	0,05	0,10	1	-1,1635	0,7426	-16,35	0,5777	0,2576	0,1989	28,81	0,0428	1,7314	0,3966	15,42	0,2107	-0,8456	0,5753	-40,93	0,3910	0,0711	0,1000	42,15	0,0104	0,1152	0,1265	15,15	0,0162		
100	0,10	0,3	1	-1,0314	0,4025	-3,14	0,1628	0,2198	0,1480	9,90	0,0223	1,5125	0,4233	0,84	0,1792	-0,6398	0,4152	-6,64	0,1738	0,0502	0,0217	0,32	0,0005	0,1001	0,0325	0,13	0,0011		
100	0,10	0,5	1	-1,0296	0,3911	-2,96	0,1537	0,2200	0,1455	9,98	0,0216	1,5086	0,4154	0,58	0,1725	-0,6356	0,4040	-5,93	0,1643	0,0501	0,0142	0,25	0,0002	0,0993	0,0205	-0,68	0,0004		
100	0,05	0,20	1	-1,0285	0,8251	-2,85	0,6810	0,2553	0,2152	27,65	0,0493	1,6614	0,4573	10,76	0,2350	-0,8119	0,6046	-35,32	0,4100	0,0575	0,0888	15,07	0,0079	0,1612	0,1563	-19,38	0,0259		
100	0,20	0,3	1	-1,0495	0,4276	-4,95	0,1851	0,2236	0,1658	11,81	0,0280	1,5293	0,4165	1,95	0,1742	-0,6578	0,4137	-9,64	0,1743	0,0485	0,0273	-3,01	0,0007	0,2002	0,0473	0,08	0,0022		
100	0,20	0,5	1	-1,0336	0,3886	-3,36	0,1520	0,2200	0,1520	10,02	0,0235	1,5259	0,3931	1,73	0,1551	-0,6494	0,3937	-8,24	0,1573	0,0499	0,0155	-0,30	0,0002	0,1998	0,0289	-0,11	0,0008		
100	0,10	0,05	1	-1,2969	0,6750	-29,69	0,5433	0,2411	0,2155	20,55	0,0481	1,7150	0,3876	14,33	0,1963	-0,8087	0,5729	-34,78	0,3714	0,0870	0,1200	-13,00	0,0145	0,0912	0,1044	82,30	0,0126		
100	0,05	0,3	1	-1,0334	0,3829	-3,34	0,1476	0,2177	0,1551	8,84	0,0243	1,5216	0,3849	1,44	0,1484	-0,6357	0,3969	-5,96	0,1586	0,0983	0,0284	-1,68	0,0008	0,0491	0,0231	-1,74	0,0005		
100	0,05	0,5	1	-1,0334	0,3679	-3,34	0,1364	0,2191	0,1508	9,56	0,0231	1,5169	0,3755	1,13	0,1412	-0,6331	0,3859	-5,51	0,1499	0,0993	0,0196	-0,67	0,0004	0,0493	0,0149	-1,45	0,0002		
100	0,10	0,10	1	-1,1938	0,7917	-19,38	0,6637	0,2514	0,2050	25,69	0,0446	1,6938	0,4273	12,92	0,2200	-0,8359	0,5827	-39,32	0,3949	0,0930	0,1152	-6,99	0,0133	0,1111	0,1224	11,06	0,0151		
100	0,10	0,3	1	-1,0327	0,4151	-3,27	0,1732	0,2221	0,1515	11,06	0,0234	1,5159	0,4280	1,06	0,1832	-0,6496	0,4244	-8,27											



TABELA 16

Simulações para o modelo de regressão logística com erros de medida e reclassificação parcial  $CECCR(m = [3; 5], n_r = 0,3n)$ .

Tamanho amostra	Erros de Medida		Reclassificações (m)	$\beta_0 = 1,0$				$\beta_1 = 0,2$				$\beta_2 = 1,5$				$\beta_3 = -0,6$				$e_1$				$e_2$			
	$e_1$	$e_2$		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
50	0,05	0,05	3	-1,0219	0,5524	-2,19	0,3053	0,2319	0,2153	15,97	0,0473	1,4816	0,5244	-1,23	0,2750	-0,6940	0,5710	-15,6742	0,33	0,0508	0,0516	1,55	0,0027	0,0469	0,0542	-6,18	0,0029
50		0,05	5	-1,0118	0,5307	-1,18	0,2815	0,2284	0,2125	14,19	0,0459	1,4728	0,5261	-1,81	0,2772	-0,6824	0,5560	-13,7331	0,32	0,0521	0,0391	4,26	0,0015	0,0473	0,0374	-5,41	0,0014
50		0,10	3	-0,9502	0,6113	4,98	0,3758	0,2050	0,2345	2,49	0,0550	1,4970	0,5435	-0,20	0,2951	-0,7256	0,5975	-20,9294	0,37	0,0488	0,0560	-2,49	0,0031	0,0889	0,0721	-11,10	0,0053
50	0,05	0,10	5	-0,9580	0,5697	4,20	0,3260	0,2065	0,2283	3,25	0,0521	1,4810	0,5377	-1,27	0,2892	-0,7052	0,5700	-17,5288	0,34	0,0483	0,0381	-3,43	0,0015	0,0938	0,0531	-6,19	0,0029
50		0,20	3	-0,9684	0,7324	3,16	0,5368	0,2189	0,2474	9,44	0,0615	1,5179	0,5615	1,19	0,3153	-0,7037	0,6270	-17,2760	0,40	0,0542	0,0656	8,41	0,0043	0,1805	0,1062	-9,74	0,0116
50		0,20	5	-1,0131	0,6192	-1,31	0,3832	0,2171	0,2310	8,56	0,0536	1,5032	0,5485	0,21	0,3006	-0,6821	0,6058	-13,6755	0,37	0,0504	0,0419	0,81	0,0018	0,1954	0,0734	-2,29	0,0054
50	0,10	0,05	3	-1,0329	0,5721	-3,29	0,3281	0,2255	0,2280	12,74	0,0526	1,4898	0,5404	-0,68	0,2919	-0,7210	0,5911	-20,1659	0,36	0,0939	0,0708	-6,10	0,0051	0,0498	0,0609	-0,45	0,0037
50		0,05	5	-1,0022	0,5476	-0,22	0,2995	0,2196	0,2249	9,81	0,0509	1,4739	0,5450	-1,74	0,2974	-0,7000	0,5796	-16,6664	0,35	0,0992	0,0523	-0,79	0,0027	0,0475	0,0402	-5,02	0,0016
50		0,10	3	-0,9590	0,6447	4,10	0,4168	0,2024	0,2436	1,19	0,0593	1,4900	0,5659	-0,66	0,3200	-0,7394	0,6219	-23,2376	0,41	0,0937	0,0756	-6,32	0,0058	0,0890	0,0779	-10,99	0,0062
50	0,10	0,10	5	-0,9562	0,5986	4,38	0,3598	0,2001	0,2396	0,04	0,0574	1,4747	0,5577	-1,68	0,3114	-0,7176	0,5898	-19,5956	0,36	0,0960	0,0567	-4,00	0,0032	0,0936	0,0568	-6,35	0,0033
50		0,20	3	-0,9549	0,7915	4,51	0,6279	0,2650	0,2650	3,52	0,0702	1,5055	0,5928	0,37	0,3511	-0,7199	0,6671	-19,9911	0,46	0,0971	0,0898	-2,94	0,0081	0,1810	0,1153	-9,52	0,0137
50		0,20	5	-0,9858	0,6664	1,42	0,4438	0,2070	0,2469	3,48	0,0610	1,4904	0,5744	-0,64	0,3297	-0,6935	0,6354	-15,5910	0,41	0,0977	0,0631	-2,33	0,0040	0,1938	0,0783	-3,08	0,0062
50	0,20	0,05	3	-1,0597	0,6521	-5,97	0,4284	0,2143	0,2580	7,16	0,0667	1,4660	0,5811	-2,27	0,3385	-0,7402	0,6371	-23,3637	0,43	0,1771	0,1007	-11,47	0,0107	0,0529	0,0676	5,88	0,0046
50		0,05	5	-1,0012	0,5880	-0,12	0,3454	0,2163	0,2453	8,15	0,0604	1,4578	0,5700	-2,81	0,3264	-0,7158	0,6205	-19,3063	0,40	0,1943	0,0731	-2,87	0,0054	0,0475	0,0425	-4,95	0,0018
50		0,10	3	-0,9590	0,7535	4,10	0,5688	0,1866	0,2777	-6,71	0,0772	1,4593	0,6151	-2,71	0,3797	-0,7840	0,6729	-30,6652	0,49	0,1783	0,1094	-10,83	0,0124	0,0892	0,0853	-10,79	0,0074
50	0,20	0,10	5	-0,9510	0,6568	4,90	0,4334	0,1923	0,2635	-3,87	0,0694	1,4538	0,6000	-3,08	0,3617	-0,7423	0,6437	-23,7175	0,43	0,1911	0,0790	-4,43	0,0063	0,0932	0,0623	-6,80	0,0039
50		0,20	3	-0,9558	0,8977	4,42	0,8070	0,1936	0,3009	-3,18	0,0905	1,4918	0,6453	-0,55	0,4161	-0,7738	0,7221	-28,9593	0,55	0,1803	0,1193	-9,83	0,0146	0,1793	0,1231	-10,36	0,0156
50		0,20	5	-0,9892	0,7781	1,08	0,6050	0,1927	0,2904	-3,65	0,0843	1,4792	0,6201	-1,38	0,3846	-0,7356	0,6995	-22,6001	0,51	0,1880	0,0873	-6,01	0,0078	0,1944	0,0906	-2,78	0,0082
100	0,05	0,05	3	-1,0225	0,3934	-2,25	0,1551	0,2147	0,1620	7,34	0,0264	1,5181	0,4019	1,21	0,1617	-0,6412	0,4183	-6,8607	0,18	0,0480	0,0355	-3,90	0,0013	0,0472	0,0383	-5,51	0,0015
100		0,05	5	-1,0183	0,3912	-1,83	0,1532	0,2132	0,1586	6,62	0,0253	1,5127	0,3997	0,85	0,1598	-0,6365	0,4139	-6,0871	0,17	0,0485	0,0254	-2,93	0,0006	0,0475	0,0260	-5,06	0,0007
100		0,10	3	-1,0317	0,4633	-3,17	0,2154	0,2193	0,1652	9,64	0,0276	1,5085	0,4465	0,57	0,1992	-0,6531	0,4564	-8,8481	0,21	0,0479	0,0380	-4,11	0,0014	0,0981	0,0555	-1,85	0,0031
100	0,05	0,10	5	-1,0237	0,4334	-2,37	0,1882	0,2190	0,1608	9,48	0,0262	1,5012	0,4396	0,08	0,1931	-0,6457	0,4403	-7,6165	0,20	0,0487	0,0256	-2,53	0,0007	0,0978	0,0362	-2,16	0,0013
100		0,20	3	-1,0441	0,5412	-4,41	0,2945	0,2317	0,1845	15,83	0,0350	1,5287	0,4710	1,91	0,2224	-0,6736	0,4893	-12,2666	0,24	0,0475	0,0459	-4,96	0,0021	0,1925	0,0782	-3,73	0,0062
100		0,20	5	-1,0437	0,4822	-4,37	0,2342	0,2279	0,1785	13,93	0,0326	1,5235	0,4595	1,57	0,2115	-0,6672	0,4612	-11,1986	0,22	0,0487	0,0283	-2,68	0,0008	0,1987	0,0521	-0,64	0,0027
100	0,10	0,05	3	-1,0276	0,4306	-2,76	0,1860	0,2138	0,1693	6,91	0,0288	1,5207	0,4211	1,38	0,1776	-0,6522	0,4440	-8,6973	0,20	0,0953	0,0491	-4,69	0,0024	0,0475	0,0411	-5,01	0,0017
100		0,05	5	-1,0214	0,4113	-2,14	0,1695	0,2141	0,1676	7,04	0,0283	1,5150	0,4148	1,00	0,1721	-0,6473	0,4334	-7,8840	0,19	0,0980	0,0343	-2,00	0,0012	0,0477	0,0273	-4,68	0,0007
100		0,10	3	-1,0454	0,5080	-4,54	0,2599	0,2231	0,1755	11,53	0,0313	1,5226	0,4633	1,51	0,2150	-0,6712	0,4778	-11,8705	0,23	0,0963	0,0567	-3,70	0,0032	0,0993	0,0621	-0,69	0,0039
100	0,10	0,10	5	-1,0255	0,4546	-2,55	0,2071	0,2207	0,1688	10,34	0,0289	1,5054	0,4587	0,36	0,2102	-0,6549	0,4548	-9,1473	0,21	0,0981	0,0355	-1,85	0,0013	0,0976	0,0381	-2,40	0,0015
100		0,20	3	-1,0574	0,6170	-5,74	0,3837	0,2323	0,1963	16,16	0,0395	1,5297	0,4925	1,98	0,2432	-0,6926	0,5291	-15,4348	0,29	0,0940	0,0674	-6,03	0,0046	0,1940	0,0870	-3,01	0,0076
100		0,20	5	-1,0474	0,5167	-4,74	0,2690	0,2265	0,1871	13,23	0,0357	1,5242	0,4735	1,61	0,2246	-0,6787	0,4924	-13,1192	0,25	0,0970	0,0415	-2,96	0,0017	0,1995	0,0556	-0,27	0,0031
100	0,20	0,05	3	-1,0302	0,5011	-3,02	0,2518	0,2092	0,1895	4,58	0,0360	1,5146	0,4533	0,97	0,2055	-0,6646	0,4836	-10,7592	0,24	0,1892	0,0735	-5,38	0,0055	0,0486	0,0474	-2,83	0,0022
100		0,05	5	-1,0173	0,4454	-1,73	0,1984	0,2097	0,1831	4,83	0,0336	1,5051	0,4413	0,34	0,1945	-0,6520	0,4655	-8,6739	0,22	0,1952	0,0486	-2,38	0,0024	0,0486	0,0304	-2,80	0,0009
100		0,10	3	-1,0328	0,6029	-3,28	0,3643	0,2167	0,2011	8,33	0,0407	1,5181	0,4906	1,21	0,2407	-0,6951	0,5297	-15,8445	0,29	0,1895	0,0848	-5,23	0,0073	0,0963	0,0711	-3,71	0,0051
100	0,20	0,10	5	-1,0191	0,4964	-1,91	0,2465	0,2138	0,1891	6,92	0,0359	1,5021	0,4887	0,14	0,2386	-0,6659	0,4933	-10,9830	0,25	0,1954	0,0536	-2,28	0,0029	0,0979	0,0441	-2,13	0,0019
100		0,20	3	-1,0430	0,7398	-4,30	0,5486	0,2208	0,2264	10,41	0,0517	1,5206	0,5313	1,38	0,2824	-0,7183	0,5960	-19,7126	0,37	0,1826	0,0924	-8,68	0,0088	0,1885	0,1000	-5,76	0,0101
100		0,20	5	-1,0362	0,5975	-3,62	0,3579	0,2211	0,2110	10,54	0,0449	1,5185	0,5044	1,23	0,2545	-0,7003	0,5498	-16,7085	0,31	0,1923	0,0607	-3,84	0,0037	0,1974	0,0642	-1,30	0,0041

TABELA 16

Simulações para o modelo de regressão logística com erros de medida e reclassificação parcial  $CECCR(m = [3; 5], n_r = 0,3n)$ .

Tamanho amostra	Erros de Medida		Reclassificações (m)	$\beta_0 = 1,0$				$\beta_1 = 0,2$				$\beta_2 = 1,5$				$\beta_3 = -0,6$				$e_1$				$e_2$			
	$e_1$	$e_2$		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
500	0,05	0,05	3	-1,0162	0,1963	-1,62	0,0387	0,2067	0,0715	3,33	0,0051	1,5129	0,2214	0,86	0,0491	-0,6029	0,1852	-0,4852	0,03	0,0498	0,0158	-0,47	0,0002	0,0503	0,0164	0,60	0,0003
500		0,05	5	-1,0148	0,1933	-1,48	0,0375	0,2063	0,0712	3,15	0,0051	1,5112	0,2188	0,75	0,0479	-0,6019	0,1839	-0,3191	0,03	0,0499	0,0114	-0,29	0,0001	0,0502	0,0116	0,36	0,0001
500		0,10	3	-1,0341	0,2148	-3,41	0,0473	0,2123	0,0766	6,13	0,0060	1,5141	0,2249	0,94	0,0507	-0,5982	0,1979	0,2996	0,04	0,0487	0,0175	-2,65	0,0003	0,1007	0,0243	0,72	0,0006
500	0,05	0,10	5	-1,0277	0,2037	-2,77	0,0422	0,2126	0,0758	6,29	0,0059	1,5122	0,2231	0,81	0,0499	-0,5972	0,1959	0,4656	0,04	0,0494	0,0120	-1,10	0,0001	0,0994	0,0163	-0,58	0,0003
500		0,20	3	-1,0315	0,2724	-3,15	0,0751	0,2102	0,0814	5,11	0,0067	1,5204	0,2657	1,36	0,0709	-0,6210	0,2180	-3,4983	0,05	0,0487	0,0225	-2,67	0,0005	0,2017	0,0366	0,85	0,0013
500		0,20	5	-1,0128	0,2296	-1,28	0,0528	0,2085	0,0783	4,25	0,0062	1,5134	0,2522	0,89	0,0637	-0,6157	0,2098	-2,6134	0,04	0,0502	0,0132	0,38	0,0002	0,2001	0,0229	0,03	0,0005
500	0,10	0,05	3	-1,0148	0,2080	-1,48	0,0435	0,2068	0,0776	3,41	0,0061	1,5196	0,2338	1,30	0,0550	-0,6064	0,1963	-1,0718	0,04	0,1005	0,0245	0,47	0,0006	0,0497	0,0178	-0,51	0,0003
500		0,05	5	-1,0143	0,2001	-1,43	0,0402	0,2059	0,0759	2,96	0,0058	1,5157	0,2271	1,05	0,0517	-0,6039	0,1932	-0,6420	0,04	0,1003	0,0164	0,30	0,0003	0,0501	0,0120	0,18	0,0001
500		0,10	3	-1,0376	0,2326	-3,76	0,0554	0,2128	0,0824	6,41	0,0070	1,5149	0,2412	0,99	0,0583	-0,5995	0,2096	0,0776	0,04	0,0980	0,0264	-1,98	0,0007	0,1010	0,0260	0,96	0,0007
500	0,10	0,10	5	-1,0296	0,2114	-2,96	0,0455	0,2136	0,0797	6,79	0,0065	1,5098	0,2364	0,65	0,0559	-0,5976	0,2049	0,3945	0,04	0,0990	0,0168	-1,00	0,0003	0,0995	0,0168	-0,51	0,0003
500		0,20	3	-1,0390	0,3113	-3,90	0,0983	0,2101	0,0892	5,06	0,0081	1,5294	0,2802	1,96	0,0793	-0,6224	0,2353	-3,7284	0,06	0,0976	0,0313	-2,39	0,0010	0,2018	0,0407	0,88	0,0017
500		0,20	5	-1,0151	0,2478	-1,51	0,0616	0,2077	0,0841	3,85	0,0071	1,5168	0,2657	1,12	0,0708	-0,6171	0,2249	-2,8497	0,05	0,0997	0,0183	-0,32	0,0003	0,2003	0,0249	0,14	0,0006
500	0,20	0,05	3	-1,0209	0,2452	-2,09	0,0605	0,2105	0,0880	5,26	0,0078	1,5284	0,2532	1,89	0,0649	-0,6125	0,2204	-2,0912	0,05	0,2004	0,0355	0,22	0,0013	0,0496	0,0224	-0,86	0,0005
500		0,05	5	-1,0206	0,2153	-2,06	0,0467	0,2083	0,0841	4,15	0,0071	1,5214	0,2430	1,42	0,0594	-0,6073	0,2120	-1,2086	0,04	0,2002	0,0223	0,08	0,0005	0,0503	0,0135	0,67	0,0002
500		0,10	3	-1,0419	0,2898	-4,19	0,0856	0,2161	0,0948	8,03	0,0092	1,5269	0,2743	1,79	0,0759	-0,6023	0,2378	-0,3767	0,06	0,1983	0,0403	-0,87	0,0016	0,1000	0,0320	-0,02	0,0010
500	0,20	0,10	5	-1,0343	0,2393	-3,43	0,0584	0,2157	0,0887	7,87	0,0081	1,5143	0,2629	0,95	0,0693	-0,5985	0,2247	0,2452	0,05	0,1992	0,0237	-0,39	0,0006	0,0995	0,0187	-0,50	0,0003
500		0,20	3	-1,0466	0,4124	-4,66	0,1721	0,2139	0,1056	6,94	0,0113	1,5433	0,3149	2,89	0,1009	-0,6306	0,2842	-5,0929	0,08	0,1964	0,0471	-1,79	0,0022	0,1998	0,0491	-0,12	0,0024
500		0,20	5	-1,0183	0,2882	-1,83	0,0833	0,2098	0,0965	4,89	0,0094	1,5211	0,2950	1,40	0,0874	-0,6219	0,2644	-3,6487	0,07	0,1991	0,0268	-0,47	0,0007	0,1999	0,0282	-0,07	0,0008
1000	0,05	0,05	3	-1,0129	0,1345	-1,29	0,0182	0,2049	0,0476	2,43	0,0023	1,5036	0,1484	0,24	0,0220	-0,6100	0,1354	-1,6640	0,02	0,0502	0,0116	0,40	0,0001	0,0506	0,0123	1,29	0,0002
1000		0,05	5	-1,0107	0,1295	-1,07	0,0169	0,2046	0,0470	2,31	0,0022	1,5015	0,1458	0,10	0,0212	-0,6095	0,1339	-1,5821	0,02	0,0501	0,0081	0,22	0,0001	0,0501	0,0079	0,20	0,0001
1000		0,10	3	-1,0163	0,1529	-1,63	0,0236	0,2059	0,0504	2,94	0,0026	1,5084	0,1635	0,56	0,0268	-0,5982	0,1411	0,3038	0,02	0,0496	0,0126	-0,73	0,0002	0,0993	0,0173	-0,67	0,0003
1000	0,05	0,10	5	-1,0154	0,1418	-1,54	0,0203	0,2052	0,0497	2,60	0,0025	1,5077	0,1602	0,52	0,0257	-0,5983	0,1390	0,2825	0,02	0,0498	0,0083	-0,49	0,0001	0,0997	0,0110	-0,32	0,0001
1000		0,20	3	-1,0264	0,1886	-2,64	0,0362	0,2071	0,0575	3,56	0,0034	1,5216	0,1882	1,44	0,0358	-0,6040	0,1576	-0,6747	0,02	0,0493	0,0149	-1,37	0,0002	0,2010	0,0255	0,49	0,0007
1000		0,20	5	-1,0131	0,1597	-1,31	0,0256	0,2058	0,0550	2,89	0,0031	1,5147	0,1818	0,98	0,0332	-0,6027	0,1502	-0,4483	0,02	0,0502	0,0086	0,38	0,0001	0,1994	0,0158	-0,31	0,0002
1000	0,10	0,05	3	-1,0131	0,1455	-1,31	0,0213	0,2052	0,0522	2,61	0,0028	1,5031	0,1551	0,21	0,0240	-0,6092	0,1409	-1,5285	0,02	0,1000	0,0168	0,03	0,0003	0,0505	0,0133	0,93	0,0002
1000		0,05	5	-1,0111	0,1372	-1,11	0,0189	0,2049	0,0513	2,44	0,0026	1,5016	0,1524	0,11	0,0232	-0,6086	0,1388	-1,4351	0,02	0,1000	0,0112	-0,04	0,0001	0,0500	0,0082	-0,10	0,0001
1000		0,10	3	-1,0163	0,1643	-1,63	0,0272	0,2060	0,0543	3,00	0,0030	1,5106	0,1721	0,71	0,0297	-0,5992	0,1496	0,1273	0,02	0,0999	0,0181	-0,10	0,0003	0,0991	0,0187	-0,90	0,0004
1000	0,10	0,10	5	-1,0163	0,1486	-1,63	0,0223	0,2054	0,0529	2,68	0,0028	1,5089	0,1685	0,60	0,0284	-0,5987	0,1466	0,2115	0,02	0,0998	0,0116	-0,20	0,0001	0,0996	0,0113	-0,39	0,0001
1000		0,20	3	-1,0327	0,2141	-3,27	0,0469	0,2083	0,0629	4,16	0,0040	1,5267	0,2028	1,78	0,0418	-0,6061	0,1689	-1,0150	0,03	0,0989	0,0221	-1,07	0,0005	0,2011	0,0275	0,57	0,0008
1000		0,20	5	-1,0158	0,1708	-1,58	0,0294	0,2066	0,0595	3,30	0,0036	1,5171	0,1910	1,14	0,0367	-0,6032	0,1598	-0,5396	0,03	0,1004	0,0125	0,37	0,0002	0,1996	0,0169	-0,21	0,0003
1000	0,20	0,05	3	-1,0146	0,1749	-1,46	0,0308	0,2072	0,0600	3,58	0,0036	1,5066	0,1752	0,44	0,0307	-0,6081	0,1544	-1,3418	0,02	0,2008	0,0250	0,41	0,0006	0,0500	0,0169	-0,02	0,0003
1000		0,05	5	-1,0132	0,1502	-1,32	0,0227	0,2056	0,0572	2,80	0,0033	1,5036	0,1703	0,24	0,0290	-0,6070	0,1509	-1,1653	0,02	0,2004	0,0158	0,19	0,0003	0,0499	0,0089	-0,18	0,0001
1000		0,10	3	-1,0172	0,2094	-1,72	0,0441	0,2086	0,0628	4,29	0,0040	1,5139	0,1985	0,92	0,0396	-0,6003	0,1690	-0,0531	0,03	0,1995	0,0268	-0,26	0,0007	0,0984	0,0228	-1,58	0,0005
1000	0,20	0,10	5	-1,0177	0,1696	-1,77	0,0290	0,2072	0,0597	3,59	0,0036	1,5115	0,1888	0,77	0,0357	-0,5992	0,1618	0,1288	0,03	0,1998	0,0162	-0,09	0,0003	0,0994	0,0129	-0,56	0,0002
1000		0,20	3	-1,0491	0,2907	-4,91	0,0868	0,2127	0,0768	6,37	0,0061	1,5395	0,2349	2,63	0,0567	-0,6058	0,1989	-0,9628	0,04	0,1979	0,0334	-1,06	0,0011	0,2009	0,0339	0,43	0,0011
1000		0,20	5	-1,0316	0,2123	-3,16	0,0460	0,2099	0,0694	4,95	0,0049	1,5267	0,2173	1,78	0,0479	-0,6024	0,1847	-0,3946	0,03	0,1995	0,0191	-0,25	0,0004	0,2001	0,0203	0,06	0,0004

TABELA 17

Simulações para o modelo de regressão logística com erros de medida e reclassificação parcial  $CECCR(m = [3; 5], n_r = 0,7n)$ .

Tamanho amostra	Erros de Medida		Reclassi- ficações (m)	$\beta_0 = 1,0$				$\beta_1 = 0,2$				$\beta_2 = 1,5$				$\beta_3 = -0,6$				$e_1$				$e_2$			
	$e_1$	$e_2$		Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
50	0,05	0,05	3	-1,0302	0,5294	-3,02	0,2809	0,2304	0,2080	15,20	0,0442	1,4893	0,5190	-0,71	0,2693	-0,6786	0,5385	-13,09	0,2958	0,0517	0,0366	3,49	0,0013	0,0493	0,0371	-1,50	0,0014
50	0,05	0,05	5	-1,0233	0,5107	-2,33	0,2611	0,2285	0,2053	14,23	0,0429	1,4762	0,5171	-1,59	0,2677	-0,6695	0,5310	-11,58	0,2865	0,0512	0,0247	2,45	0,0006	0,0495	0,0243	-0,91	0,0006
50	0,05	0,10	3	-0,9832	0,5568	1,68	0,3099	0,2092	0,2226	4,58	0,0496	1,5121	0,5166	0,81	0,2668	-0,7036	0,5687	-17,26	0,3339	0,0492	0,0398	-1,65	0,0016	0,0964	0,0543	-3,57	0,0030
50	0,05	0,10	5	-0,9734	0,5330	2,66	0,2845	0,2122	0,2142	6,09	0,0460	1,4910	0,5159	-0,60	0,2660	-0,6821	0,5422	-13,68	0,3005	0,0499	0,0244	-0,21	0,0006	0,0971	0,0354	-2,93	0,0013
50	0,20	0,20	3	-1,0285	0,6054	-2,85	0,3669	0,2264	0,2233	13,21	0,0505	1,5085	0,5479	0,56	0,2999	-0,6819	0,5904	-13,65	0,3549	0,0515	0,0490	3,04	0,0024	0,1950	0,0809	-2,50	0,0066
50	0,20	0,20	5	-1,0298	0,5399	-2,98	0,2921	0,2230	0,2115	11,48	0,0452	1,4954	0,5289	-0,30	0,2795	-0,6638	0,5587	-10,63	0,3159	0,0503	0,0283	0,51	0,0008	0,1991	0,0487	-0,45	0,0024
50	0,10	0,05	3	-1,0318	0,5376	-3,18	0,2897	0,2316	0,2163	15,80	0,0477	1,4946	0,5319	-0,36	0,2827	-0,6989	0,5554	-16,48	0,3179	0,1003	0,0499	0,26	0,0025	0,0493	0,0397	-1,34	0,0016
50	0,05	0,5	5	-1,0192	0,5150	-1,92	0,2653	0,2275	0,2095	13,74	0,0446	1,4810	0,5286	-1,27	0,2795	-0,6809	0,5413	-13,48	0,2992	0,1009	0,0348	0,86	0,0012	0,0492	0,0254	-1,53	0,0006
50	0,10	0,10	3	-0,9789	0,5838	2,11	0,3409	0,2078	0,2318	3,88	0,0537	1,5021	0,5347	0,14	0,2856	-0,7191	0,5905	-19,85	0,3625	0,0990	0,0569	-0,99	0,0032	0,0965	0,0588	-3,53	0,0035
50	0,10	0,10	5	-0,9735	0,5483	2,65	0,3010	0,2096	0,2184	4,82	0,0477	1,4943	0,5275	-0,38	0,2780	-0,6950	0,5515	-15,83	0,3128	0,1008	0,0356	0,83	0,0013	0,0970	0,0379	-2,98	0,0014
50	0,20	0,3	3	-1,0062	0,6940	-0,62	0,4812	0,2194	0,2434	9,70	0,0596	1,5025	0,5763	0,17	0,3318	-0,7065	0,6221	-17,74	0,3979	0,0986	0,0718	-1,36	0,0051	0,1928	0,0912	-3,59	0,0084
50	0,20	0,5	5	-1,0367	0,5703	-3,67	0,3263	0,2184	0,2207	9,19	0,0490	1,4965	0,5478	-0,24	0,2998	-0,6662	0,5765	-11,03	0,3364	0,0975	0,0418	-2,46	0,0018	0,2003	0,0535	0,13	0,0029
50	0,20	0,05	3	-1,0375	0,6022	-3,75	0,3636	0,2262	0,2346	13,12	0,0557	1,4940	0,5474	-0,40	0,2994	-0,7113	0,5908	-18,55	0,3611	0,1953	0,0765	-2,33	0,0059	0,0496	0,0493	-0,84	0,0024
50	0,05	0,5	5	-1,0191	0,5416	-1,91	0,2934	0,2275	0,2206	13,75	0,0494	1,4771	0,5388	-1,52	0,2906	-0,6893	0,5685	-14,89	0,3308	0,2009	0,0509	0,47	0,0026	0,0496	0,0283	-0,85	0,0008
50	0,10	0,3	3	-0,9718	0,6687	2,82	0,4475	0,2009	0,2527	0,47	0,0638	1,4785	0,5724	-1,44	0,3277	-0,7542	0,6292	-25,70	0,4192	0,1944	0,0832	-2,82	0,0069	0,0952	0,0688	-4,77	0,0048
50	0,20	0,10	5	-0,9658	0,5787	3,42	0,3358	0,2036	0,2309	1,82	0,0533	1,4832	0,5554	-1,12	0,3085	-0,7085	0,5762	-18,08	0,3435	0,2007	0,0529	0,37	0,0028	0,0971	0,0430	-2,88	0,0019
50	0,20	0,3	3	-0,9469	0,8322	5,31	0,6947	0,2043	0,2730	2,13	0,0745	1,4765	0,6268	-1,57	0,3930	-0,7252	0,6760	-20,87	0,4722	0,1934	0,0992	-3,32	0,0099	0,1841	0,1063	-7,96	0,0115
50	0,20	0,5	5	-1,0185	0,6561	-1,85	0,4304	0,2198	0,2390	9,89	0,0574	1,5030	0,5651	0,20	0,3190	-0,6983	0,6253	-16,38	0,4002	0,1984	0,0631	-0,78	0,0040	0,1958	0,0648	-2,12	0,0042
100	0,05	0,05	3	-1,0238	0,3829	-2,38	0,1470	0,2161	0,1562	8,07	0,0246	1,5193	0,3890	1,28	0,1515	-0,6368	0,3984	-6,13	0,1599	0,0493	0,0249	-1,45	0,0006	0,0481	0,0251	-3,75	0,0006
100	0,05	0,05	5	-1,0239	0,3772	-2,39	0,1427	0,2163	0,1543	8,13	0,0241	1,5140	0,3849	0,93	0,1482	-0,6359	0,3952	-5,98	0,1573	0,0492	0,0167	-1,68	0,0003	0,0486	0,0170	-2,85	0,0003
100	0,10	0,3	3	-1,0257	0,4216	-2,57	0,1783	0,2193	0,1541	9,67	0,0241	1,5056	0,4312	0,37	0,1858	-0,6397	0,4349	-6,62	0,1905	0,0498	0,0255	-0,39	0,0007	0,0994	0,0378	-0,62	0,0014
100	0,05	0,10	5	-1,0282	0,4070	-2,82	0,1663	0,2205	0,1518	10,24	0,0234	1,5059	0,4215	0,39	0,1775	-0,6361	0,4208	-6,02	0,1782	0,0500	0,0166	0,08	0,0003	0,0994	0,0244	-0,64	0,0006
100	0,10	0,20	3	-1,0518	0,4611	-5,18	0,2151	0,2264	0,1716	13,22	0,0301	1,5254	0,4394	1,69	0,1935	-0,6732	0,4475	-12,19	0,2054	0,0484	0,0324	-3,11	0,0011	0,1991	0,0548	-0,43	0,0030
100	0,20	0,5	5	-1,0371	0,4169	-3,71	0,1750	0,2236	0,1612	11,82	0,0265	1,5177	0,4232	1,18	0,1792	-0,6593	0,4208	-9,88	0,1804	0,0496	0,0187	-0,77	0,0004	0,1996	0,0341	-0,21	0,0012
100	0,10	0,05	3	-1,0239	0,4007	-2,39	0,1609	0,2155	0,1620	7,77	0,0265	1,5178	0,3965	1,19	0,1574	-0,6432	0,4129	-7,20	0,1722	0,0981	0,0340	-1,86	0,0012	0,0483	0,0282	-3,46	0,0008
100	0,10	0,05	5	-1,0243	0,3861	-2,43	0,1495	0,2169	0,1575	8,44	0,0251	1,5135	0,3899	0,90	0,1520	-0,6404	0,4030	-6,73	0,1638	0,0992	0,0231	-0,77	0,0005	0,0486	0,0174	-2,88	0,0003
100	0,10	0,3	3	-1,0291	0,4469	-2,91	0,2004	0,2226	0,1596	11,30	0,0260	1,5124	0,4418	0,83	0,1952	-0,6513	0,4447	-8,54	0,2002	0,0997	0,0379	-0,27	0,0014	0,0987	0,0411	-1,27	0,0017
100	0,10	0,10	5	-1,0280	0,4177	-2,80	0,1751	0,2201	0,1550	10,06	0,0244	1,5051	0,4341	0,34	0,1883	-0,6416	0,4298	-6,93	0,1863	0,0993	0,0236	-0,69	0,0006	0,0995	0,0257	-0,45	0,0007
100	0,20	0,3	3	-1,0542	0,5073	-5,42	0,2600	0,2288	0,1783	14,41	0,0326	1,5237	0,4644	1,58	0,2160	-0,6861	0,4727	-14,35	0,2306	0,0985	0,0488	-1,46	0,0024	0,1985	0,0613	-0,74	0,0038
100	0,20	0,5	5	-1,0348	0,4379	-3,48	0,1927	0,2201	0,1678	10,06	0,0285	1,5143	0,4387	0,95	0,1924	-0,6621	0,4299	-10,35	0,1885	0,0992	0,0270	-0,82	0,0007	0,1998	0,0369	-0,08	0,0014
100	0,20	0,05	3	-1,0302	0,4359	-3,02	0,1907	0,2160	0,1743	8,00	0,0306	1,5218	0,4226	1,45	0,1789	-0,6566	0,4482	-9,44	0,2039	0,1961	0,0507	-1,93	0,0026	0,0486	0,0341	-2,82	0,0012
100	0,05	0,5	5	-1,0253	0,3991	-2,53	0,1598	0,2160	0,1668	7,98	0,0280	1,5138	0,4020	0,92	0,1616	-0,6408	0,4259	-6,80	0,1829	0,1990	0,0332	-0,52	0,0011	0,0491	0,0195	-1,84	0,0004
100	0,10	0,3	3	-1,0232	0,5111	-2,32	0,2615	0,2217	0,1789	10,85	0,0325	1,5154	0,4641	1,03	0,2154	-0,6645	0,4765	-10,75	0,2310	0,1980	0,0597	-1,01	0,0036	0,0969	0,0499	-3,15	0,0025
100	0,20	0,10	5	-1,0245	0,4324	-2,45	0,1874	0,2175	0,1636	8,73	0,0270	1,5027	0,4497	0,18	0,2020	-0,6498	0,4481	-8,29	0,2031	0,1986	0,0349	-0,69	0,0012	0,0996	0,0286	-0,37	0,0008
100	0,20	0,3	3	-1,0570	0,6175	-5,70	0,3842	0,2223	0,1985	11,17	0,0399	1,5248	0,4983	1,65	0,2487	-0,7064	0,5276	-17,73	0,2894	0,1914	0,0728	-4,32	0,0054	0,1962	0,0756	-1,90	0,0057
100	0,20	0,5	5	-1,0280	0,4968	-2,80	0,2473	0,2177	0,1792	8,86	0,0324	1,5154	0,4684	1,03	0,2194	-0,6794	0,4694	-13,23	0,2264	0,1976	0,0410	-1,22	0,0017	0,1983	0,0433	-0,87	0,0019

TABELA 17

Simulações para o modelo de regressão logística com erros de medida e reclassificação parcial  $CECCR(m = [3; 5], n_r = 0,7n)$ .

Tamanho amostra	Erros de Medida			Reclassificações (m)	$\beta_0 = 1,0$				$\beta_1 = 0,2$				$\beta_2 = 1,5$				$\beta_3 = -0,6$				$e_1$				$e_2$			
	$e_1$	$e_2$			Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM	Média	Desvio	Vício (%)	EQM
500	0,05	0,05	3	-1,0117	0,1822	-1,17	0,0333	0,2067	0,0672	3,34	0,0046	1,5084	0,2107	0,56	0,0444	-0,6009	0,1756	-0,15	0,0308	0,0497	0,0104	-0,55	0,0001	0,0501	0,0111	0,13	0,0001	
500	0,05	0,05	5	-1,0103	0,1785	-1,03	0,0319	0,2064	0,0667	3,20	0,0045	1,5064	0,2077	0,43	0,0431	-0,5997	0,1744	0,05	0,0304	0,0498	0,0074	-0,31	0,0001	0,0501	0,0077	0,11	0,0001	
500		0,10	3	-1,0269	0,1925	-2,69	0,0377	0,2103	0,0701	5,13	0,0050	1,5130	0,2110	0,87	0,0447	-0,6005	0,1889	-0,08	0,0356	0,0496	0,0118	-0,80	0,0001	0,1005	0,0161	0,49	0,0003	
500	0,05	0,10	5	-1,0243	0,1861	-2,43	0,0352	0,2105	0,0686	5,26	0,0048	1,5128	0,2078	0,85	0,0433	-0,6017	0,1854	-0,29	0,0344	0,0499	0,0078	-0,11	0,0001	0,0998	0,0106	-0,17	0,0001	
500		0,20	3	-1,0224	0,2232	-2,24	0,0502	0,2092	0,0718	4,58	0,0052	1,5158	0,2409	1,06	0,0582	-0,6171	0,1929	-2,85	0,0374	0,0496	0,0149	-0,73	0,0002	0,2010	0,0243	0,50	0,0006	
500		0,20	5	-1,0134	0,1945	-1,34	0,0380	0,2077	0,0682	3,87	0,0047	1,5142	0,2271	0,95	0,0517	-0,6144	0,1855	-2,40	0,0346	0,0503	0,0085	0,56	0,0001	0,2002	0,0153	0,10	0,0002	
500	0,10	0,05	3	-1,0127	0,1875	-1,27	0,0353	0,2070	0,0708	3,51	0,0051	1,5127	0,2193	0,85	0,0482	-0,6028	0,1834	-0,46	0,0336	0,0997	0,0152	-0,30	0,0002	0,0499	0,0122	-0,24	0,0001	
500		0,05	5	-1,0100	0,1815	-1,00	0,0330	0,2064	0,0687	3,20	0,0048	1,5069	0,2126	0,46	0,0452	-0,6004	0,1789	-0,06	0,0320	0,0997	0,0102	-0,26	0,0001	0,0500	0,0080	0,06	0,0001	
500		0,10	3	-1,0278	0,2028	-2,78	0,0419	0,2106	0,0735	5,30	0,0055	1,5128	0,2193	0,85	0,0482	-0,6022	0,1948	-0,37	0,0379	0,0995	0,0170	-0,47	0,0003	0,1006	0,0173	0,60	0,0003	
500	0,10	0,10	5	-1,0242	0,1908	-2,42	0,0370	0,2105	0,0702	5,26	0,0050	1,5123	0,2129	0,82	0,0454	-0,6043	0,1896	-0,72	0,0359	0,0998	0,0108	-0,20	0,0001	0,0998	0,0110	-0,16	0,0001	
500		0,20	3	-1,0256	0,2450	-2,56	0,0606	0,2091	0,0772	4,56	0,0060	1,5223	0,2512	1,49	0,0635	-0,6173	0,2045	-2,89	0,0421	0,0995	0,0201	-0,50	0,0004	0,2010	0,0266	0,50	0,0007	
500		0,20	5	-1,0154	0,2050	-1,54	0,0422	0,2076	0,0712	3,82	0,0051	1,5170	0,2365	1,14	0,0562	-0,6156	0,1920	-2,61	0,0371	0,1002	0,0117	0,23	0,0001	0,2004	0,0164	0,21	0,0003	
500	0,20	0,05	3	-1,0140	0,2097	-1,40	0,0441	0,2088	0,0768	4,40	0,0060	1,5184	0,2317	1,23	0,0540	-0,6036	0,1987	-0,61	0,0395	0,2005	0,0230	0,25	0,0005	0,0495	0,0155	-1,04	0,0002	
500		0,05	5	-1,0136	0,1896	-1,36	0,0361	0,2078	0,0731	3,88	0,0054	1,5114	0,2190	0,76	0,0481	-0,6008	0,1873	-0,14	0,0351	0,2000	0,0144	0,01	0,0002	0,0502	0,0089	0,32	0,0001	
500		0,10	3	-1,0298	0,2292	-2,98	0,0534	0,2126	0,0805	6,28	0,0066	1,5192	0,2404	1,28	0,0581	-0,6041	0,2149	-0,68	0,0462	0,2003	0,0257	0,14	0,0007	0,1002	0,0210	0,17	0,0004	
500	0,20	0,10	5	-1,0230	0,2005	-2,30	0,0407	0,2107	0,0735	5,33	0,0055	1,5116	0,2243	0,77	0,0504	-0,6042	0,1984	-0,69	0,0394	0,2001	0,0156	0,06	0,0002	0,0998	0,0122	-0,24	0,0001	
500		0,20	3	-1,0351	0,3172	-3,51	0,1017	0,2123	0,0885	6,15	0,0080	1,5307	0,2822	2,05	0,0805	-0,6253	0,2391	-4,21	0,0578	0,1993	0,0310	-0,37	0,0010	0,2005	0,0324	0,25	0,0010	
500		0,20	5	-1,0224	0,2299	-2,24	0,0533	0,2093	0,0781	4,67	0,0062	1,5202	0,2532	1,35	0,0644	-0,6186	0,2101	-3,10	0,0445	0,1996	0,0173	-0,19	0,0003	0,2006	0,0188	0,31	0,0004	
1000	0,05	0,05	3	-1,0079	0,1241	-0,79	0,0154	0,2031	0,0448	1,55	0,0020	1,5020	0,1400	0,13	0,0196	-0,6089	0,1305	-1,49	0,0171	0,0502	0,0074	0,31	0,0001	0,0501	0,0081	0,24	0,0001	
1000		0,05	5	-1,0061	0,1206	-0,61	0,0146	0,2024	0,0444	1,20	0,0020	1,5008	0,1379	0,05	0,0190	-0,6091	0,1289	-1,51	0,0167	0,0500	0,0051	0,09	0,0000	0,0499	0,0053	-0,14	0,0000	
1000		0,10	3	-1,0113	0,1335	-1,13	0,0179	0,2048	0,0462	2,39	0,0022	1,5070	0,1527	0,47	0,0233	-0,5992	0,1326	0,14	0,0176	0,0501	0,0083	0,14	0,0001	0,0994	0,0111	-0,58	0,0001	
1000	0,05	0,10	5	-1,0108	0,1271	-1,08	0,0163	0,2044	0,0450	2,21	0,0020	1,5059	0,1487	0,39	0,0221	-0,5985	0,1292	0,25	0,0167	0,0501	0,0055	0,29	0,0000	0,0997	0,0072	-0,29	0,0001	
1000		0,20	3	-1,0152	0,1549	-1,52	0,0242	0,2056	0,0515	2,78	0,0027	1,5142	0,1674	0,95	0,0282	-0,6000	0,1407	0,00	0,0198	0,0500	0,0097	0,06	0,0001	0,1999	0,0176	-0,07	0,0003	
1000		0,20	5	-1,0114	0,1359	-1,14	0,0186	0,2044	0,0482	2,22	0,0023	1,5107	0,1574	0,71	0,0249	-0,5989	0,1327	0,18	0,0176	0,0502	0,0058	0,35	0,0000	0,2000	0,0108	0,02	0,0001	
1000	0,10	0,05	3	-1,0097	0,1306	-0,97	0,0171	0,2041	0,0477	2,06	0,0023	1,5017	0,1440	0,12	0,0207	-0,6085	0,1333	-1,41	0,0178	0,1001	0,0110	0,08	0,0001	0,0501	0,0088	0,17	0,0001	
1000		0,05	5	-1,0061	0,1249	-0,61	0,0156	0,2028	0,0463	1,38	0,0022	1,4996	0,1406	-0,02	0,0197	-0,6080	0,1308	-1,34	0,0172	0,1000	0,0074	0,01	0,0001	0,0498	0,0055	-0,34	0,0000	
1000		0,10	3	-1,0109	0,1393	-1,09	0,0195	0,2042	0,0487	2,10	0,0024	1,5078	0,1589	0,52	0,0253	-0,5999	0,1368	0,01	0,0187	0,1000	0,0117	-0,03	0,0001	0,0993	0,0121	-0,69	0,0001	
1000	0,10	0,10	5	-1,0113	0,1305	-1,13	0,0172	0,2039	0,0463	1,97	0,0022	1,5072	0,1532	0,48	0,0235	-0,5994	0,1321	0,10	0,0174	0,1001	0,0077	0,13	0,0001	0,0998	0,0075	-0,25	0,0001	
1000		0,20	3	-1,0183	0,1723	-1,83	0,0300	0,2062	0,0555	3,08	0,0031	1,5168	0,1778	1,12	0,0319	-0,6025	0,1486	-0,41	0,0221	0,1001	0,0145	0,15	0,0002	0,2000	0,0190	0,00	0,0004	
1000		0,20	5	-1,0128	0,1432	-1,28	0,0207	0,2046	0,0510	2,32	0,0026	1,5125	0,1624	0,84	0,0265	-0,5998	0,1364	0,04	0,0186	0,1004	0,0083	0,40	0,0001	0,2001	0,0116	0,06	0,0001	
1000	0,20	0,05	3	-1,0109	0,1441	-1,09	0,0209	0,2052	0,0517	2,62	0,0027	1,5040	0,1543	0,27	0,0238	-0,6098	0,1425	-1,64	0,0204	0,2007	0,0171	0,36	0,0003	0,0500	0,0110	-0,06	0,0001	
1000		0,05	5	-1,0064	0,1296	-0,64	0,0168	0,2029	0,0483	1,46	0,0023	1,5005	0,1459	0,03	0,0213	-0,6088	0,1357	-1,47	0,0185	0,2005	0,0105	0,27	0,0001	0,0498	0,0060	-0,31	0,0000	
1000		0,10	3	-1,0100	0,1617	-1,00	0,0262	0,2065	0,0541	3,26	0,0030	1,5081	0,1757	0,54	0,0309	-0,5980	0,1490	0,33	0,0222	0,2001	0,0176	0,07	0,0003	0,0987	0,0144	-1,32	0,0002	
1000	0,20	0,10	5	-1,0119	0,1407	-1,19	0,0199	0,2050	0,0495	2,51	0,0025	1,5072	0,1644	0,48	0,0271	-0,5994	0,1389	0,10	0,0193	0,2002	0,0108	0,11	0,0001	0,0996	0,0085	-0,37	0,0001	
1000		0,20	3	-1,0232	0,2211	-2,32	0,0494	0,2082	0,0638	4,09	0,0041	1,5249	0,1992	1,66	0,0403	-0,6010	0,1674	-0,17	0,0280	0,2000	0,0223	0,02	0,0005	0,1994	0,0231	-0,29	0,0005	
1000		0,20	5	-1,0227	0,1674	-2,27	0,0285	0,2067	0,0568	3,34	0,0033	1,5202	0,1797	1,34	0,0326	-0,5994	0,1503	0,10	0,0226	0,2000	0,0123	-0,02	0,0002	0,2003	0,0135	0,17	0,0002	

## ANEXO A – Procedimento Bootstrap utilizado

Em Estatística, Bootstrap é qualquer método de reamostragem e foi proposto por Bradley Efron em 1979. É utilizado para aproximar a distribuição de uma estatística, com base em reamostragens de uma amostra aleatória, permitindo estimar o erro padrão, bem como construir intervalos de confiança e realizar testes de hipóteses sobre parâmetros de interesse.

O significado do trabalho de Efron (1979), foi muito bem expresso por Davison e Hinkley (1997): “a publicação em 1979 do primeiro artigo de Bradley Efron sobre os métodos de Bootstrap foi um grande evento na Estatística, sintetizando algumas das ideias anteriores de reamostragem e a criação de um novo quadro para análise estatística baseada em simulação. A ideia de substituir as aproximações complicadas e muitas vezes imprecisas para as tendências, variâncias e outras medidas de incerteza por simulação de computador pegou a imaginação de pesquisadores teóricos e usuários de métodos estatísticos”. Maiores detalhes sobre a metodologia Bootstrap podem ser encontrados em Efron, Bradley e Hastie (2016).

O procedimento Bootstrap utilizado neste trabalho foi o não-paramétrico:

**Passo 01:** Considere o banco de dados disponível  $\{(y_i, x_i), i = 1, \dots, n\}$  com  $n$  observações;

**Passo 02:** Faça uma amostra Bootstrap com amostragem dos pares com reposição  $(x, y)_b^* = \{(x_1, y_1)^*, \dots, (x_n, y_n)^*\}$  para  $(b = 1, \dots, B)$ . O número de reamostragem Bootstrap foi  $B = 5000$ .

**Passo 03:** Para cada  $b = 1, \dots, B$ , estime as estatísticas Bootstrap  $\hat{\varphi}_1^*, \dots, \hat{\varphi}_B^*$ , na qual  $\hat{\varphi}_b^* = (\hat{\beta}^*, \hat{e}_1^*, \hat{e}_2^*)$  ajustando o modelo em (16);

**Passo 04:** Estime os valores da média e erro-padrão Bootstrap de  $\hat{\varphi} = (\hat{\beta}, \hat{e}_1, \hat{e}_2)$ , tais que:

$$\bar{\varphi}^* = \frac{\sum_{b=1}^B \hat{\varphi}_b^*}{B}$$

$$se(\hat{\varphi}) = \sqrt{\frac{\sum_{b=1}^B (\hat{\varphi}_b^* - \bar{\varphi}^*)^2}{B - 1}}$$



**Passo 05:** Estime o intervalo de confiança  $(1 - \alpha)100\%$  Bootstrap percentílico por meio dos percentis:  $(\hat{\varphi}_I, \hat{\varphi}_S) = \left( \hat{\varphi}_{(b)}^{(\frac{\alpha}{2})}, \hat{\varphi}_{(b)}^{(1-\frac{\alpha}{2})} \right)$ .