MODELAGEM ESTATÍSTICA DE FATORES DE RISCO DO INFARTO NO MIOCÁRDIO

Julia Helena Maia do Nascimento
hnascimentojulia@gmail.com

Margarete Afonso de Sousa
margarete.afonso.sousa@gmail.com

Narley Worllos do Carmo Netto
eng.narley@gmail.com

Fernando Luiz Cyrino Oliveira
cyrino@puc-rio.br

Julia Lima Fleck
jfleck@puc-rio.br
Programa de Pós-Graduação em Engenharia Industrial - PUC Rio
Rua Marquês de São Vicente, 225 – Gávea, Rio de Janeiro – RJ

RESUMO

O infarto agudo no miocárdio (IAM) é a principal causa de mortes no Brasil e há diversos fatores que contribuem para o aumento das chances de sua incidência. O objetivo desse trabalho foi desenvolver um modelo de regressão logística multivariada para quantificar a influência de fatores no desfecho de pacientes internados com infarto no miocárdio no estado do Rio de Janeiro e através do *bootstrap* gerar novas amostras para simular novos cenários para validar o modelo proposto. Para isso gerou-se 1000 amostras, estimando novos parâmetros para comparar os intervalos de confiança. Valores menores que 0,055 de p-valor foram considerados como estatisticamente significativos em todas as etapas do estudo. O estudo considerou pacientes adultos, que estiveram internados por até 30 dias nos anos de 2014 a 2015. O intervalo de confiança gerado pelo *bootstrap* confirma que as variáveis são estatisticamente significantes no desfecho de pacientes internados com IAM no Rio de Janeiro.

PALAVRAS CHAVE. Infarto agudo no miocárdio, Regressão Logística, *Bootstrap*.

ABSTRACT

Acute myocardial infarction (AMI) is the leading cause of death in Brazil and there are several factors that contribute to the increased chances of its incidence. The aim of this study was to develop a multivariate logistic regression model to quantify the influence of factors in the outcome of hospitalized patients with myocardial infarction in the state of Rio de Janeiro and through out bootstrap generate new samples to simulate new scenarios to validate the proposed model. To do so 1000 samples were generated, estimating new parameters to compare the confidence intervals. Values less than 0.055 p-value were considered as statistically significant at all stages of the study. The study considered adult patients who were hospitalized up to 30 days in the years 2014 to 2015. The confidence interval generated by the bootstrap confirms variables are statistically significant in the outcome of hospitalized patients with AMI in Rio de Janeiro.

KEYWORDS. Acute myocardial infarction, Logistic regression, *Bootstrap*.

https://proceedings.science/p/85487?lang=pt-br

1. Introdução

De acordo com a Organização Pan-Americana da Saúde (OPAS), as Doenças Cardiovasculares (DCV) estão entre as principais causas de morte no mundo. O Infarto Agudo do Miocárdio (IAM) é primeira causa de morte no Brasil, de acordo com a base de dados do Departamento de Informática do SUS, que registra cerca de 100 mil óbitos anuais devidos à doença [DATASUS, 2014].

Em 2016, 17 milhões de pessoas foram vítimas de problemas coronarianos, como ataques cardíacos. No Brasil, 300 mil pessoas sofrem infartos todos os anos, em 30% dos casos o ataque cardíaco é fatal [AGÊNCIA BRASIL, 2017].

Segundo o DATASUS (2014), os fatores de risco tradicionais que causam o infarto no miocárdio são fatores modificáveis, como glicemia, trigliceridemia, colesterolemia e obesidade, fatores que estão ligados a uma alimentação desequilibrada rica em gorduras, carboidratos, sal, alimentos processados, uso de álcool, de cigarro e de outras drogas, e outras situações recorrentes de estresse e sedentarismo. Enquanto que fatores não modificáveis são etnia, sexo e idade e histórico familiar de doenças cardiovasculares, estes podem favorecer o aparecimento da IAM até em indivíduos com uma boa alimentação [DOMÍNGUEZ *et al*. 1998].

Compreender o quanto esses fatores modificáveis impactam o risco de morte de quem pode apresentar infarto no miocárdio é de extrema importância para tentar mitigar a sua incidência, trabalhando no nível de atenção primário junto a população de modo a prevenir a doença reduzindo esses principais fatores de risco. Também é importante investigar a eficiência de internações em hospitais públicos e privados a fim de realocar recursos.

A distribuição de recursos principalmente na saúde pública como leitos hospitalares, remédios, recursos financeiros e humanos é caracterizada por uma ineficiência tão acentuada que deveria merecer um pouco mais de atenção por parte dos tomadores de decisões da área da saúde. Sendo assim, esse estudo representa um passo inicial cujos resultados podem auxiliar no direcionamento de esforços de investigação relativos à determinação de políticas públicas de distribuição de recursos na saúde.

Percebe-se uma complexidade crescente dos problemas tratados pela ciência, muitas são as dificuldades em transformar dados em informações úteis. Neste contexto, várias técnicas estatísticas têm sido utilizadas para análise de dados e auxílio à tomada de decisão na área de saúde, sendo assim o presente estudo tem como objetivo a aplicação da regressão logística a fim de definir a significância das variáveis que contribuem para a mortalidade das pessoas por infarto no miocárdio e através do *bootstrap* gerar novas amostras simulando novos cenários para validar modelo proposto.

Além desta introdução, o trabalho é composto por uma seção de embasamento teórico acerca da metodologia utilizada seguida da aplicação dessa metodologia no contexto estudado. Em seguida, são feitas as discussões dos resultados alcançados. Por fim, seguem-se as conclusões acerca do estudo.

2. Revisão da Literatura

2.1 Regressão Logística Múltipla

A Regressão Logística é um modelo estatístico usado quando se deseja explicar uma variável resposta categórica em função de variáveis explicativas quantitativas ou qualitativas [HOSMER e LEMESHOW, 1989].

Na Regressão Logística múltipla consideremos um conjunto de n variáveis independentes expresso pelo vetor $\mathbf{X}=(x_1, x_2, x_3 ... x_n)$, sendo assim temos que $E(Y|\mathbf{X})$, sendo $E(Y|\mathbf{X})$ lida como "valor esperado de Y dado $\mathbf{X}=x_n$".

Neste caso, $E(Y|\mathbf{X})=$ é dado pela equação (1).

$$\Pi(\mathbf{X}) = \frac{e^{\beta 0}+e^{\beta 1 x 1}+e^{\beta 2 x 2}+\cdots+e^{\beta n x n}}{1+e^{\beta 0}+e^{\beta 1 x 1}+e^{\beta 2 x 2}+\cdots+e^{\beta n x n}} \quad (1)$$

https://proceedings.science/p/85487?lang=pt-br

De forma a linearizar o modelo de regressão logística a transformação *logit* é aplicada. Tendo assim o *logit* da Regressão logística múltipla é dado pela equação (2).

$$g(X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \quad (2)$$

Um conceito muito importante na regressão logística é o de chance, usualmente chamado pelo nome em inglês *odds*. A razão de duas chances é chamada *odds ratios*, que é a chance de Y ser 1 em X pela razão da chance de Y ser 0 em X, simplificadamente é dada pela equação (3).

$$odds\ ratios = e^{\beta} \quad (3)$$

2.2 *Bootstrap*

O *bootstrap* é um método de simulação para inferência estatística utilizado para obter um novo conjunto de dados através da reamostragem dos dados originais com reposição. Os dados gerados através de *bootstrap* $\left(x_{1,}^{*}\ x_{2,}^{*} \ldots, x_{n,}^{*}\right)$ é a reamostragem dos dados de origem $\left(x_{1,}\ x_{2,}\ \ldots, x_{n,}\right)$ com reposição onde um dado pode aparecer uma única vez, ou até mais de uma vez (EFRON e TIBISHIRANI, 1993).

Nesse artigo utilizamos um número de replicações por *bootstrap* B = 1000 para estimar os intervalos de confiança dos β. Efron e Tibishirani, (1993) indicam um valor ≥ 500 ou 1000 para garantir uma variação pequena dos valores estimados no intervalo de confiança. Os artigos mostrados na Tabela 1 também embasaram a escolha do B.

https://proceedings.science/p/85487?lang=pt-br

Autores	Ano	Título	Objetivo	Qtd variáveis explicativas na RLM	Amostra	Nº. Repet. Boots.
Farnaz Zandvakili et *al.*,	2017	Maternal Outcomes Associated with Caesarean versus Vaginal Delivery	Determinar o modo de partos e resultados maternos em um hospital no Irã, durante um ano.	10	5984	1000
R. Patell *et al.*	2017	Identifying predictors for bleeding in hospitalized cancer patients: A cohort study	Melhorar a previsão do risco de sangramento com o objetivo de ajudar os médicos na definição do tratamento em paciente com câncer	8	3358	1000
Kim *et al*	2017	Prediction model for non-curative resection of endoscopic submucosal dissection in patients with early gastric cancer	Criar um modelo para prever a ressecção não-curativa da ESD em pacientes com EGC	10	1639	2000
Lee *et al*	2017	Toward an MRI-based nomogram for the prediction of transperineal prostate biopsy outcome: A physician and patient decisio ntool	Desenvolver e validar internamente um nomograma usando variáveis derivadas de imagem de ressonância magnética biparamétrica (B-MRI) para a predição de câncer de próstata na biopsia de próstata guiada por setor transperineal (TPSB).	4	615	200
Haddad *et al*	2015	Preoperative multivariable prognostic models for prediction of survival and major complications following surgical resection of renal cell carcinoma with suprahepatic caval tumor thrombus	Desenvolver um modelo multivariado preoperativo para previsão de sobrevivência e risco de complicações para paciente com trombose supra-hepatica	21	132	100

Tabela 1: Artigos que citam a aplicação de *boostrap* na área de saúde

2.3 Aplicação da Regressão Logística e Bootstrap na área da Saúde

Os modelos de regressão logística são aplicados intensamente em várias áreas de conhecimento e, em especial, na área da saúde. A condição de se estudar variáveis respostas binárias em função de um conjunto de fatores explicativos tem se tornado cada vez mais comum em estudos epidemiológicos, que tem como o objetivo investigar a relação entre algumas variáveis preditivas e uma variável-desfecho do tipo dicotômica [SOUZA, 2013].

Haddad *et al.* (2015) desenvolveu um modelo de regressão logística multivariado pré-operativo para previsão de sobrevivência e risco de complicações para paciente com trombose supra-hepática e utilizou o *bootstrap* com reamostragem para calibrar nomogramas. Já Patell *et al.* (2017) utilizou o modelo de regressão logística para identificar os fatores de risco de sangramento

e validou-os com *bootstrap*, a fim de melhorar a previsão do risco de sangramento de forma a ajudar os médicos na definição do tratamento em pacientes com câncer.

Kim *et al.* (2017) criou um modelo para prever a ressecção não-curativa da Dissecção Endoscópica Submucosa em pacientes com câncer gástrico em estado inicial, no qual foi criado um sistema de pontuação baseado em um modelo de regressão logística e validação de através de *bootstrap.*

Lee *et al*. (2017) desenvolveu e validou internamente um nomograma usando variáveis derivadas de imagem de ressonância magnética biparamétrica (B-MRI) para a predição de câncer de próstata na biopsia de próstata guiada por setor transperineal (TPSB) e utilizou a regressão logística univariável e multivariável para determinar preditores de resultados de câncer de próstata e utilizou o *bootstrap* para validar internamente seu modelo de regressão logística.

Zandvakili *et al.* (2017) para determinar resultados pós-parto como morte materna, sangramento, ruptura uterina, entre outros empregou vários modelos de regressão logística utilizando como variáveis explicativas tipo de partos como cesária, parto normal, parto normal com episiotomia e parto vaginal operativo. E para estimar os parâmetros e aumentar o nível de significância, utilizou o *bootstrap* para gerar 1000 amostras.

3. Pacientes e métodos

3.1 Origem dos dados

Neste trabalho foram utilizados dados do Sistema de Informações Hospitalares do Sistema Único de Saúde (SIHSUS). A base de dados utilizada foi a Autorizações de Internações Hospitalares (AIH) que, na íntegra, possui mais de 100 milhões de registros que tratam de internações financiadas pelo SUS em todo território brasileiro. Os dados são anonimizados e disponibilizados online através do site *www.datasus.gov.br*.

3.2 População

Este estudo restringiu a análise à: adultos (> 17 anos), no período de 2014 a 2015, internados no estado do Rio de Janeiro, que permaneceram até 30 dias no hospital e que foram internados ou sofreram em algum momento da internação infarto agudo no miocárdio. Os códigos considerados da doença foram, de acordo com Código Internacional de Doenças (CID 10), as famílias I21 (I21.0, I21.1, I21.2, I21.3, I21.4, I21.9), I22 (I22.0, I22.1, I22.8, I22.9) e I23 (I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6, I23.8). Após a aplicação dos filtros foram consideradas para o modelo o total de 9530 internações.

3.3 Variáveis candidatas ao modelo

Domínguez *et al*. (1998) cita como fatores de risco que favorecem a ocorrência do IAM o avanço da idade, a hipertensão arterial, o tabagismo, o sedentarismo, o sexo (relacionando com o estado civil), a ocorrência de infartos anteriores e outros. O trabalho de Domínguez *et al*. (1998) norteou a seleção das variáveis candidatas ao modelo, que foram escolhidas de acordo com os campos disponíveis na AIH respeitadas as limitações da base.

Além disso, as organizações de saúde também podem contribuir indiretamente para o desfecho da doença pois hospitais públicos e privados se distinguem por diferenças quanto aos objetivos ligados à competição e à lucratividade, à administração e aos mecanismos de controle e regulação, à incorporação tecnológica, à eficiência e ao processo do cuidado prestado aos pacientes. Essas disparidades entre esses dois tipos de hospitais podem ser devido a diferenças no que tange aos recursos da organização [MARTINS *et al.,* 2004].

Dessa forma, as variáveis selecionadas para o modelo de regressão logística foram:

Desfecho (DES): Variável resposta que representa se o paciente morreu ou sobreviveu.

Faixa de idade (ID): Idade dos pacientes internados com incidência de infarto.

Sexo (SEX): Sexo dos pacientes internados com incidência de infarto do Miocárdio.

Tipo de hospital (TH): Variável que define o tipo de hospital.

Faixas de internação (FI): Faixas de dias em que o paciente ficou internado. As faixas foram categorizadas de acordo com Kristoffersen *et al.*, (2013).

As variáveis são utilizadas na regressão logística são todas categóricas. A quantidade de pacientes associados a elas, assim como as categorias correspondentes a cada variável estão descritas na Tabela 2.

Variável	Categoria	Morreu = 1 (%)	Sobreviveu = 0 (%)	Total	%
Faixa de idade em anos (ID)	18 - 30	9 (0,09)	69 (0,72)	78	0,82
	31 - 64	802 (8,42)	4326 (45,39)	5128	53,81
	65 - 84	1243 (13,04)	2640 (27,70)	3883	40,75
	85 - 99	239 (2,51)	202 (2,12)	441	4,63
				9530	**100**
Sexo (SEX)	Homem	1253 (13,15)	4550 (47,74)	5803	60,89
	Mulher	1040 (10,91)	2687 (28,20)	3727	39,11
				9530	**100**
Tipo de hospital (TH)	Privado	485 (5,09)	3316 (34,80)	3801	39,88
	Público	1808 (18,97)	3921 (41,14)	5729	60,12
				9530	**100**
Faixas de internação em dias (FI)	0 a 4	1459 (15,31)	3500 (36,73)	4959	52,04
	5 a 8	354 (3,71)	1507 (15,81)	1861	19,53
	9 a 12	196 (2,06)	916 (9,61)	1112	11,67
	13 a 30	284 (2,98)	1314 (13,79)	1598	16,77
				9530	**100**

Tabela 2: Resumo descritivo da amostra

Para prever a chance de mortalidade de um paciente foi proposto o modelo representado pela equação (4):

$$\text{DES} = \beta_0 + \beta_1 \text{SEX} + \beta_2 \text{ID} + \beta_3 \text{TH} + \beta_4 \text{FI} \qquad (4)$$

Base (intercepto - β_0): homem, atendido em hospital privado, adulto que ficou internado entre 0-4 dias.

3.4 Análises

Os filtros na base de dados foram aplicados com o software gerenciador de bando de dados pgAdmin III e o "fit" do modelo, procedimentos de *bootstrap* e demais análises foram realizadas no software R versão 3.3.2.

4. Resultados

A Tabela 3 mostra o resumo dos coeficientes do modelo de regressão logística para os dados reais. As categorias com betas positivos possuem mais chances de morte (y = 1) em relação à categoria base da mesma variável. Por exemplo, mulheres possuem mais chances de morrer que homens assim como indivíduos com incidência de infarto no miocárdio que são internados em hospitais públicos possuem mais chances de morrer em relação aos internados em hospitais privados. O beta representa o logaritmo das chances (*odds ratio*).

https://proceedings.science/p/85487?lang=pt-br

Quase todas as faixas apresentaram p-valores menores que 0,001, portanto foram consideradas significativas para este modelo. A faixa “novos adultos” (pacientes entre 18 e 30 anos) da variável “faixa de idade” apresentou p-valor mais alto que as demais variáveis, no entanto não se rejeitou a hipótese de que essa faixa não é significativa ao modelo, pois a quantidade amostral é pequena, acredita-se que em uma amostra maior a faixa se tornaria mais significante.

	Beta	Erro padrão	Z-valor	P-valor
Intercepto	-2,32444	0,06302	-36,886	< 0,001
Mulheres	0,18965	0,05338	3,553	< 0,001
Hospital Público	1,5585	0,06249	24,938	< 0,001
Idosos (65 a 84)	1,01758	0,05496	18,514	< 0,001
"Muito" idosos (85 a 99)	1,83858	0,11142	16,501	< 0,001
"Novos" adultos (18 a 30)	-0,70113	0,36528	-1,919	0,05493
Faixa 2 (5 a 8 diárias)	-1,06622	0,07355	-14,497	< 0,001
Faixa 3 (9 a 12)	-1,21504	0,09136	-13,299	< 0,001
Faixa 4 (13 a 30)	-1,25861	0,07941	-15,85	< 0,001

Tabela 3: Resumo dos coeficientes da regressão logística

Após a geração dos intervalos de confiança para os betas do modelo resumido anteriormente na Tabela 3, realizou-se a reamostragem de dados utilizando o *bootstrap* (B= 1000) com reposição e aplicação dos novos dados no modelo proposto à cada reamostragem. A partir da média dos betas foram gerados os intervalos de confiança (95%) por quartis.

	Amostra real			Amostra *bootstrap*		
	2,5%	97,5%	OR	2,5%	97,5%	OR
Intercepto	-2,449	-2,202	0,098	-2,461	-2,197	0,098
Mulheres	0,085	0,294	1,209	0,087	0,291	1,209
Hospital Público	1,437	1,682	4,752	1,427	1,686	4,743
Idosos (65 a 84)	0,910	1,126	2,766	0,912	1,130	2,770
Muito idosos (85 a 99)	1,621	2,058	6,288	1,580	2,078	6,269
"Novos" adultos (18 a 30)	-1,484	-0,035	0,496	-1,762	-0,050	0,480
Faixa 2 (5 a 8)	-1,211	-0,923	0,344	-1,209	-0,919	0,344
Faixa 3 (9 a 12)	-1,396	-1,038	0,297	-1,401	-1,037	0,296
Faixa 4 (13 a 30)	-1,416	-1,104	0,284	-1,401	-1,103	0,284

Tabela 4: Intervalos de confiança e *odds ratio*

Na Tabela 4 comparam-se os intervalos de confiança dos betas da regressão e *odds ratio* ajustados para melhor interpretação. Os intervalos de confiança não passam por 0, ou seja, não é possível descartar a hipótese nula (não significância), logo se não passa por zero descarta-se a hipótese de que a variável não é significante para o modelo.

4.1 Análise de sensibilidade

A curva ROC, apresentada no Gráfico 1 a seguir, mostra a acurácia para todos os limiares do modelo através da taxa de verdadeiros positivos e falsos positivos, o indicador usado para aferir a acurácia é a AUC – “Area under curve”. A acurácia foi de 0,74923. Quanto maior a área sob a curva, mais próximo do topo superior esquerdo a curva se encontrará, portanto mais próximo de

1,0. Os valores esperados para esse indicador são maiores que 0,5 pois nesse caso o modelo prevê apenas o que realmente aconteceu.

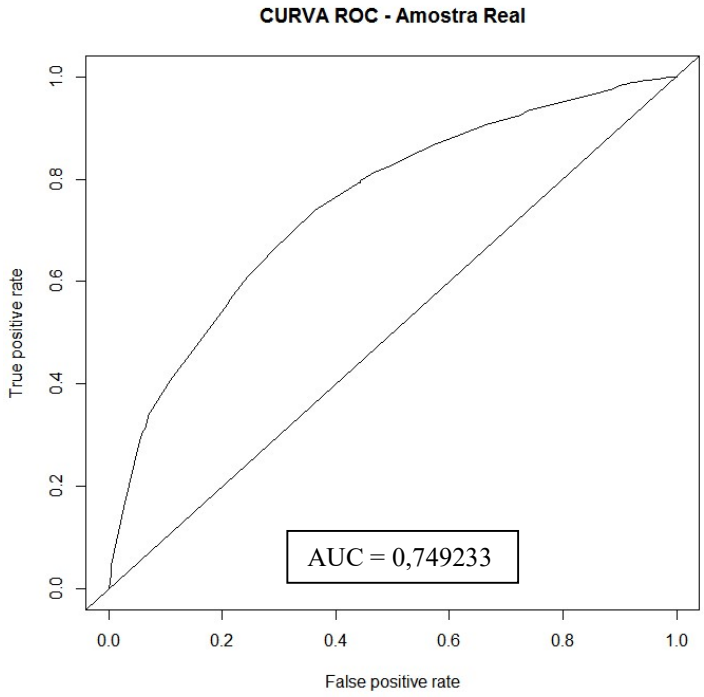


Gráfico 1: Curva ROC

Para aferir a acurácia do modelo utilizando os dados oriundos da reamostragem, também se calculou as AUCs para cada umas das 1000 novas amostras. A tabela 5 mostra que a variação da acurácia para os dados simulados variou entre 0,7308 e 0,7678. A média foi de 0,7494, valor de acurácia acima do esperado (0,5). Para valores acima de 0,5 interpreta-se que o modelo prevê melhor que a realidade.

Min.	0,7308
1º quad.	0,7454
Mediana	0,7494
Média	0,7494
3º quad.	0,7534
Max.	0,7678

Tabela 5: Variação da acurácia para os dados simulados

4.2 Discussão

Este estudo analisou a relevância de variáveis que influenciam no desfecho de hospitalizações financiadas pelo Governo Federal por infarto agudo no miocárdio ocorridas no estado do Rio de Janeiro no decorrer dos anos de 2014 e 2015 e utilizou as variáveis de sexo, faixas de idade, faixas de dias de internação e tipo de hospital onde a internação aconteceu para propor um modelo de regressão logística multivariada.

Conforme mostra a Tabela 3, os intervalos de confiança gerados após o *bootstrap* foram bastante aproximados dos gerados pelos dados reais. O mesmo acontece com a *odds ratio*, taxa que representa a chance de um paciente que sobreviveu (y = 0) mudar para o grupo que morreu (y = 1).

Dessa forma, após validar os parâmetros do modelo através do intervalo de confiança do *bootstrap*, é possível fazer inferências com mais segurança pois o IC garante a confiabilidade do modelo além de confirmar que o tamanho da amostra é suficiente para que o modelo faça previsões com acurácia.

De acordo com os resultados, mulheres possuem 1,2 mais chances de morrer do que homens, apesar da incidência de infarto agudo ter sido maior em homens (60,89%) conforme mostrado na Tabela 1. O modelo também evidencia que indivíduos internados em hospitais de gestão pública possuem 4,7 vezes mais chances de morrer em relação aos internados em hospitais de gestão privada. Os gastos referentes às internações em hospitais privados são repassados aos hospitais pelo Sistema Único de Saúde (SUS), dessa forma, de acordo com o modelo, mostra-se que para casos de infarto, os atendimentos em hospitais de gestão privada são mais eficazes que os atendimentos em hospitais públicos.

Para a variável "faixas de idade" pode-se inferir, a partir do modelo, que quanto mais idoso, maior as chances de morrer após a incidência da doença, ainda que a maior incidência amostral acorreu em adultos, na faixa entre 31 e 64 anos de idade. Indivíduos "muito idosos" (85 – 99) possuem 6,28 mais chances de morrer do que os "adultos" (31 – 64), que representam a categoria de base comparativa do modelo. Os "novos adultos" (18 – 30) apresentam chances menores de morrer após a internação por infarto no miocárdio em relação aos "adultos".

A variável "faixas de internação", referente ao período que o paciente ficou internado, apresentou que quando maior a quantidade de diárias de internação, menores são as chances de morte. Esse resultado é esperado dada a característica fulminante da doença assim como a necessidade de se realizar um atendimento de urgência veloz e eficaz.

5. Considerações finais

O intervalo de confiança gerado pela simulação (*bootstrap* com reposição) confirma que o modelo proposto considera variáveis estatisticamente significantes no desfecho (morte ou sobrevivência) de pacientes internados com infarto no miocárdio no estado do Rio de Janeiro.

O resultado prático do trabalho pode ser traduzido em investigações da causa de por que em hospitais públicos as chances de morrer são tão maiores em relação a hospitais privados a fim de equiparar o atendimento. Existe também a possibilidade de criar sistemas de prioridade em atendimento de pessoas por sexo e idade (mulheres e idosos são mais urgentes do que homens adultos).

Como proposta de trabalhos futuros sugere-se incluir mais variáveis ao modelo, como comorbidades (doenças mútuas), utilizar outras técnicas de simulação para validar o modelo e também reduzir a amostra gradualmente e validar o modelo com o *bootstrap* através do intervalo de confiança e assim estimar qual o tamanho de amostra é possível gerar resultados confiáveis neste modelo.

Referências

AGÊNCIA BRASIL (2017). *Doenças cardiovasculares são principal causa de morte no mundo.* URL:http://www.brasil.gov.br/saude/2017/09/doencas-cardiovasculares-sao-principal-causa-de-morte-no-mundo. Acessado: 2017-11-01.

DATASUS (2016). SIHSUS - Sistema de Informações Hospitalares do SUS. URL:http://www.datasus.gov.br.

DATASUS (2014). Infarto agudo do miocárdio é primeira causa de mortes no País, revela dados do DATASUS. Disponível em: URL:http://datasus.saude.gov.br/noticias/atualizacoes/559-infarto-agudo-do-miocardio-e-primeira-causa-de-mortes-no-pais-revela-dados-do-datasus.%20Acesso%20em%20Ago%202017. Acesso em: 17 out. 2017.

DOMÍNGUEZ *et al.* (1998). *Algunos factores de riesgo que favorecen la aparición del infarto agudo del miocárdio*. Revista Cubana de Medicina General Integral.

EFRON, B. & TIBSHIRANI, R. J. (1993), *An Introduction to the Bootstrap*, Vol. 57 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York, United States.

HADDAD, A. Q. *et al.* (2015). *Preoperative multivariable prognostic models for prediction of survival and major complications following surgical resection of renal cell carcinoma with suprahepatic caval tumor thrombus. Urologic Oncology: Seminars and Original Investigations*. Volume 33, Issue 9, September 2015, Pages 388.e1-388.e9

HOSMER, David; LEMESHOW, Stanley. (1989) *Applied Logistic Regression*. United States: John Wiley & Sons.

KIM, E. H. *et al.* (2017). *Prediction model for non-curative resection of endoscopic submucosal dissection in patients with early gastric cancer.* Gastrointestinal Endoscopy. Volume 85, Issue 5, May 2017, Pages 976-983

LEE, S. *et al.* (2017). *Toward an MRI-based nomogram for the prediction of transperineal prostate biopsy outcome: A physician and patient decision tool.* Urologic Oncology: Seminars and Original Investigations. Volume 35, Issue 11, November 2017, Pages 664.e11-664.e18

MARTINS, M. et al. (2004). *Mortalidade hospitalar e tempo de permanência: comparação entre hospitais públicos e privados na região de Ribeirão Preto, São Paulo, Brasil*. *Cad. Saúde Pública* [online]. 2004, vol.20, suppl.2, pp. S268-S282.

PATELL, R. et al. (2017). *Identifying predictors for bleeding in hospitalized cancer patients: A cohort study.* Thrombosis Research. Volume 158, October 2017, Pages 38-43.

SANTOS, A. C. (2009). Sistema de Sistema de Informações Hospitalares do Sistema Único de Saúde: documentação do sistema para auxiliar o uso das suas informações.

SOUZA, L. D. A. (2013). *Uso de regressão logística para identificar os fatores de risco Associados à ocorrência de anomalias congênitas em recém-nascidos.* João Pessoa, 2013. 37p. Monografia (Bacharelado em Estatística) – Universidade Federal da Paraíba.

ZANDVAKILI F. *et al*. (2017). *Maternal Outcomes Associated with Caesarean versus Vaginal Delivery.* Obstetrics and Gynaecology Section. 2017, jul, vol.11, Issue 7, p. QC01.