

MODELOS GENERALIZADOS ADITIVOS PARA LOCAÇÃO, ESCALA E FORMA NUMA ANÁLISE DE CUSTOS DE PROCEDIMENTOS HOSPITALARES DE UMA OPERADORA DE PLANOS DE SAÚDE

Rafaela de Sousa KIENEN¹
Cesar Augusto TACONELI²

- **RESUMO:** A modelagem estatística dos custos associados a procedimentos médicos e hospitalares constitui uma tarefa bastante complexa, sendo comum verificar distribuições fortemente assimétricas, presença de múltiplos fatores associados, efeitos não lineares e variâncias heterogêneas para diferentes níveis (valores) dos fatores considerados. A análise desse tipo de dados requer a utilização de métodos estatísticos que permitam tratar adequadamente dados com tais características. Neste contexto, o presente trabalho apresenta uma aplicação de Modelos Generalizados Aditivos para Locação, Escala e Forma na análise dos custos dos partos realizados por uma operadora de planos de saúde no município de Curitiba-PR em 2013. Por meio da utilização dessa classe de modelos, foi possível considerar uma maior variedade de distribuições para a resposta, incorporar aos modelos efeitos aleatórios e modelar conjuntamente parâmetros de locação e dispersão usando covariáveis. Baseado em diferentes critérios de qualidade de ajuste e análise de resíduos, verificou-se que as distribuições que proporcionaram melhores ajustes foram a *Box-Cox power exponential* (BCPE) e a *Skew-t type 3* (ST3). Para modelos ajustados com essas duas distribuições, fatores como o tipo de parto, local de internação e tipo de acomodação mostraram-se associados aos custos dos procedimentos. O médico responsável pelo parto foi incluído aos modelos por meio de efeitos aleatórios, permitindo identificar a contribuição dos médicos para o custo final dos partos e avaliar a correlação entre os efeitos produzidos pelas duas distribuições. Os efeitos de médico e local de internação foram considerados também na modelagem da dispersão dos custos, verificando-se que tal especificação contribuiu para um melhor ajuste do modelo.
- **PALAVRAS-CHAVE:** Custos de partos; efeitos aleatórios; modelos aditivos generalizados para locação, escala e forma; planos de saúde.

1 Introdução

Plano privado de Assistência à saúde é definido pela Lei 9.656, de 3 de junho de 1998, como: “prestação continuada de serviços ou cobertura de custos assistenciais a preço pré ou pós-estabelecido, por prazo indeterminado, com a finalidade de garantir, sem

¹ Universidade Federal do Paraná - UFPR, Setor de Ciências Exatas, Curso de Estatística, Caixa Postal 19.081, CEP: 81.531-990, Curitiba, Paraná, Brasil. Email: rafaela_kienen@hotmail.com

² Universidade Federal do Paraná, Setor de Ciências Exatas, Departamento de Estatística, Caixa Postal 19.081, CEP: 81.531-990, Curitiba, Paraná, Brasil. Email: cetacneli@gmail.com

limite financeiro, a assistência à saúde, pela faculdade de acesso e atendimento por profissionais ou serviços de saúde.” (Lei nº 9.656, 1998, art. 1).

Os cidadãos brasileiros, em busca de acesso aos serviços de saúde e diante da ausência de condições do Estado em disponibilizá-los, cada vez mais têm recorrido aos planos de assistência médica particular. Dados da Agência Nacional de Saúde (ANS) indicam isso, sendo que nos últimos dez anos, no Brasil, o número de beneficiários aumentou cerca de 49%. Entretanto, no mesmo período a quantidade de operadoras ativas com beneficiários decresceu 28%, restando, em 2013, 1268 operadoras ativas (ANS, 2013). Paralelamente a esse movimento do mercado, as despesas assistenciais têm aumentado cada vez mais, o que impacta diretamente no valor de venda repassado aos beneficiários.

A incorporação de novas tecnologias e o envelhecimento da população ajudam a explicar porque os custos com saúde aumentam acima da inflação geral (KAISER FOUNDATION, 2012). A principal fonte geradora de custo das operadoras de saúde são as internações hospitalares, que contemplam custos inerentes a materiais, medicamentos, honorários médicos, acomodação e exames, dentre outros. As internações correspondem a aproximadamente 62% dos gastos das operadoras, segundo dados do Instituto de Estudos de Saúde Suplementar (IESS). No período de julho/10 a junho/11 o Índice de Variação do Custo Médico Hospitalar, medido pelo IESS (VCMH/IESS), atingiu o ápice da série histórica (16,6%).

Um dos procedimentos que demandam internação, produzindo elevados custos, é o parto. Os custos inerentes aos partos apresentam alta variabilidade associada, uma vez que em diversos casos os recém-nascidos precisam de internações de longa duração gerando, consequentemente, altos custos para as operadoras. Nesse contexto, o levantamento e análise dos dados referentes a tais procedimentos médicos se fazem necessários para as operadoras de planos de saúde, pois com base nessas informações torna-se possível identificar fatores associados a elevados custos nos partos, servindo como base para o planejamento financeiro e redução de custos.

A análise dos custos de procedimentos médicos e hospitalares (em particular, custos de partos) requer a utilização de modelos flexíveis, capazes de lidar com distribuições assimétricas, elevado grau de dispersão e relações não lineares, dentre outros. Neste contexto, os Modelos Aditivos Generalizados para Localização, Escala e Forma (*Generalized Additive Models for Location, Scale and Shape – GAMLSS*) (RIGBY e STASINOPOULOS, 2005) constituem uma alternativa de modelagem pertinente, por contemplar uma grande variedade de distribuições (discretas e contínuas), proporcionar a modelagem conjunta de parâmetros de localização, dispersão e forma, além da possibilidade de incorporar efeitos aleatórios e termos não paramétricos.

Com base nos custos de partos, realizados por médicos ginecologistas e obstetras, durante o ano de 2013 e cobertos por uma operadora de planos de saúde situada no município de Curitiba-PR, considera-se aqui a modelagem usando GAMLSS com o objetivo de identificar fatores geradores de maiores custos e assim desenvolver um mecanismo de direcionamento que contribua com a gestão dos custos de procedimentos médicos e hospitalares.

2 Materiais e métodos

2.1 Descrição da amostra e recursos computacionais

O banco de dados analisado foi disponibilizado por uma operadora de planos de saúde do município de Curitiba-PR, elaborado a partir de arquivos eletrônicos enviados pelos prestadores de serviço, contendo todas as informações registradas nos prontuários médicos. É composto por todas as informações disponíveis referentes aos partos cobertos pela operadora durante o ano de 2013, totalizando 3026 procedimentos. Para cada parto realizado, dispõe-se das seguintes informações: Tipo de parto (normal ou cesariana), tipo de acomodação (enfermaria ou apartamento), local de internação (hospital) em que o parto foi realizado, médico responsável pelo parto, se a idade da mãe configurava um fator de risco (sim, caso a mãe tivesse menos de 16 ou mais de 35 anos e não, caso contrário), a operadora que dava cobertura à gestante (a operadora principal possui três operadoras credenciadas) e o custo total do procedimento de parto, em reais.

Algumas restrições foram impostas à base de dados, conforme os objetivos da análise. Primeiramente, tendo-se em vista que um dos objetivos é avaliar a contribuição dos médicos para o custo dos partos, foram excluídos da base partos realizados por médicos com menos de 10 procedimentos no ano, a fim de se obter estimativas razoavelmente precisas para o efeito dos médicos. Vale salientar que, em alguns casos, a informação do médico que atendeu a paciente não está disponível, sendo considerado como responsável o médico que coordena o corpo clínico do local de internação.

Outra restrição imposta à base foi quanto ao número de dias de internação da gestante e à ocorrência de internação da criança na Unidade de Terapia Intensiva (UTI). Foram analisados apenas partos com tempo de permanência igual ou inferior a três dias e sem a ocorrência de internação em UTI, pois se entende que os partos que divergem destas características configuram eventos excepcionais, apresentando algum tipo de intercorrência. Aplicadas as restrições, restaram 2512 partos para análise.

Todas as análises foram realizadas utilizando o software R, versão 3.0.1 (R CORE TEAM, 2014). O pacote GAMLSS (RIGBY e STASINOPOULOS, 2005) foi usado no ajuste dos modelos generalizados aditivos para locação, escala e forma.

2.2 Modelos generalizados aditivos para locação, escala e forma

Para a definição de um GAMLSS, considere y_1, y_2, \dots, y_n variáveis aleatórias independentes com função (densidade) de probabilidade $f(y_i | \theta_i)$ sendo $\theta_i = (\mu_i, \sigma_i, \nu_i, \tau_i)$ um vetor de parâmetros relacionados às variáveis explicativas. É possível modelar distribuições com até quatro parâmetros sendo que, de acordo com a notação apresentada, os dois primeiros parâmetros (μ e σ) usualmente se referem à locação e à dispersão da distribuição, sendo os outros dois (ν e τ) parâmetros de forma.

Dando sequência à definição, sejam $g_k(\cdot)$, $k = 1, 2, 3, 4$, funções de ligação monótonas que relacionam os parâmetros às partes sistemáticas do modelo. Um GAMLSS fica definido por:

$$\begin{aligned}
g_1(\boldsymbol{\mu}) &= \mathbf{X}_1 \boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1} \gamma_{j1} \\
g_2(\boldsymbol{\sigma}) &= \mathbf{X}_2 \boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2} \gamma_{j2} \\
g_3(\mathbf{v}) &= \mathbf{X}_3 \boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3} \gamma_{j3} \\
g_4(\boldsymbol{\tau}) &= \mathbf{X}_4 \boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4} \gamma_{j4}
\end{aligned}$$

sendo:

$\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, \mathbf{v} e $\boldsymbol{\tau}$ vetores de dimensão n ;

$\boldsymbol{\beta}'_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{J_k})$ vetor de dimensão J_k ;

\mathbf{X}_k uma matriz de delineamento de dimensão $n \times J_k$;

\mathbf{Z}_{jk} uma matriz de delineamento de dimensão $n \times q_{jk}$;

γ_{jk} vetor aleatório de tamanho q_{jk} para o qual é usual assumir $\gamma_{jk} \sim \text{Normal}_{q_{jk}}(0, \mathbf{G}_{jk}^{-1})$;

\mathbf{G}_{jk}^{-1} é a inversa de $\mathbf{G}_{jk}(\lambda_{jk})$, que pode depender de um vetor de hiperparâmetros λ_{jk} .

Efeitos aleatórios e componentes não paramétricos podem ser incorporados especificando-se adequadamente a parte sistemática do modelo (STASINOPOULOS e RIGBY, 2007). Os parâmetros são estimados via máxima verossimilhança, com base na seguinte função de verossimilhança penalizada:

$$l_p = l - \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^{J_k} \lambda_{jk} \gamma'_{jk} \mathbf{G}_{jk} \gamma_{jk},$$

sendo $l = \sum_{i=1}^n \ln f(y_i | \theta_i)$ a função de log-verossimilhança.

No presente estudo, efeitos aleatórios foram inseridos nos modelos de regressão com o objetivo de incorporar os efeitos dos médicos ao custo do procedimento. A inclusão de tais efeitos possibilita avaliar a variabilidade entre os custos dos partos de diferentes médicos e acomodar possível correlação de custos intra-médico. A estimação dos efeitos aleatórios em GAMLSS utiliza um argumento bayesiano empírico, assumindo distribuição normal (possivelmente imprópria) para os β'_s e para os efeitos aleatórios, com as estimativas obtidas pela moda da distribuição a posteriori. Com base na independência dos y'_{ij} s (condicional a $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$) e independência (e normalidade) para os efeitos aleatórios, Rigby e Stasinopoulos (2009) demonstram que o método de estimação resultante é equivalente à maximização de uma verossimilhança penalizada. Informações mais detalhadas quanto a procedimentos bayesianos empíricos podem ser encontradas, por exemplo, em Casella (1985) e Carlin e Louis (2000).

2.3 Procedimento de análise e ajuste dos modelos generalizados aditivos para locação, escala e forma

Para a modelagem dos custos dos partos foram consideradas, num primeiro momento, as distribuições Normal e Gama. Posteriormente, com o objetivo de identificar

distribuições que proporcionassem melhores ajustes, foram consideradas as demais distribuições contínuas com suporte no conjunto dos reais (ou reais positivos) disponíveis no pacote GAMLSS. Nesta primeira etapa, foram ajustados modelos contemplando os efeitos de todas as covariáveis disponíveis além do efeito aleatório correspondente ao médico responsável pelo parto. Para aquelas distribuições que apresentaram melhores ajustes, foi verificada ainda a possibilidade de inclusão de covariáveis na modelagem do parâmetro de dispersão.

Assim, considere y_{ij} o custo do parto da gestante i atendida pelo médico j , \mathbf{x}_i seu vetor de covariáveis e $g_1(\cdot)$ e $g_2(\cdot)$ funções de ligação. Assume-se aqui que $y_{ij} \sim f(y_{ij} | \boldsymbol{\theta}_{ij})$, $\boldsymbol{\theta}'_{ij} = (\mu_{ij}, \sigma_{ij}, \nu, \tau)$ e:

$$\begin{aligned} g_1(\mu_{ij}) &= \mathbf{x}_{i1}\boldsymbol{\beta}_1 + \nu_j \\ g_2(\sigma_{ij}) &= \mathbf{x}_{i2}\boldsymbol{\beta}_2 + \omega_j \end{aligned}$$

sendo $\nu_j \sim Normal(0, \sigma_1^2)$ e $\omega_j \sim Normal(0, \sigma_2^2)$ referentes aos efeitos dos médicos na locação e na dispersão dos custos, respectivamente, com ν_j e ω_j independentes. As funções de ligação utilizadas foram aquelas sugeridas e implementadas como padrão por Stasinopoulos e Rigby (2007).

Os ajustes foram avaliados e comparados com base na análise dos resíduos e segundo diferentes critérios de qualidade de ajuste, como o Critério de Informação de Akaike - AIC (Akaike, 1974) e o Critério de Informação Bayesiano - BIC (Schwarz, 1978). Para um modelo em que \hat{l} é a log-verossimilhança maximizada e df o número de graus de liberdade, tais critérios ficam definidos, respectivamente, por:

$$\begin{aligned} AIC &= -2\hat{l} + 2df \\ BIC &= -2\hat{l} + \ln(n)df \end{aligned}$$

Repare que os critérios diferem quanto à penalização aplicada aos graus de liberdade usados no ajuste do modelo. Em ambos os casos, valores menores são indicadores de modelos mais parcimoniosos.

Adicionalmente, para o diagnóstico do ajuste foram utilizados os resíduos quantílicos aleatorizados, propostos por Dunn e Smyth (1996), que são definidos por $r_i = \phi^{-1}(u_i)$, onde ϕ^{-1} é a inversa da função de distribuição acumulada da normal padrão e $u_i = F(y_i | \hat{\boldsymbol{\theta}}_i)$ é calculado com base na função distribuição acumulada do modelo proposto. Se as suposições do modelo são satisfeitas, estes resíduos são normalmente distribuídos com variância constante. Assim, a análise de resíduos pode ser realizada, por exemplo, com base em gráficos probabilísticos normais e gráficos de resíduos versus valores ajustados (ou resíduos versus índice).

Stasinopoulos e Rigby (2007) apresentam uma extensa relação de distribuições de probabilidades, discretas e contínuas, com diferentes suportes, que podem ser consideradas na definição de um GAMLSS. A título de exemplo, estão implementadas no pacote GAMLSS distribuições pertencentes à família Box Cox simétrica, extensões

assimétricas da distribuição t – *Student*, distribuições com suporte no intervalo 0-1 como a beta, beta inflacionada e generalizações. Duas distribuições, em particular, mostraram-se adequadas e foram utilizadas no presente trabalho: a distribuição Box Cox Exponencial Potência (*Box-Cox Power Exponential* – BCPE) e a *skew-t type 3* – ST3. A opção por dar continuidade à análise com duas distribuições, ao invés de considerar apenas aquela que proporciona melhor ajuste, tem como objetivo a comparação dos resultados produzidos pelos dois modelos.

A distribuição BCPE pertence à família de distribuições Box-Cox simétricas, que podem ser deduzidas a partir de uma transformação baseada em Box e Cox (1964), envolvendo uma classe de distribuições simétricas truncadas. É definida por quatro parâmetros, μ , σ , ν e τ , sendo o primeiro relacionado à mediana da distribuição, o segundo um parâmetro de dispersão e, os outros dois, parâmetros de assimetria e curtose. Informações detalhadas a respeito da distribuição BCPE e das demais distribuições da família Box-Cox simétrica podem ser encontradas em Fumes (2014) e Voudouris *et al* (2012).

A distribuição ST3, por sua vez, é uma das cinco versões assimétricas para a distribuição t disponíveis no pacote GAMLSS. Também é definida por quatro parâmetros, sendo que o parâmetro de localização (μ) corresponde à moda da distribuição e está relacionado de forma aditiva à média. A distribuição ST3 tem suporte no conjunto dos reais. Stasinopoulos e Rigby (2007) e Fernandes e Steel (1998) detalham as expressões e propriedades das distribuições t assimétricas.

3 Resultados e discussão

Para o conjunto de 2512 partos efetivamente analisados, foi verificado um custo médio de R\$2.522,15 e desvio padrão R\$763,54. A Figura 1 apresenta um histograma para a distribuição geral dos custos e boxplots para os custos segundo o médico responsável pelo procedimento. No primeiro, verifica-se que a distribuição apresenta assimetria à direita. Já no segundo é possível identificar grande heterogeneidade nos custos associados a diferentes médicos, tanto em relação à localização quanto em relação à dispersão. Complementando essa etapa inicial da análise, a Tabela 1 apresenta uma análise descritiva dos custos segundo as categorias das covariáveis consideradas. Algumas diferenças podem ser destacadas, por exemplo, com relação ao local de internação, sendo que o custo médio varia de R\$1.831,14 a R\$3.172,99 dentre os seis locais. Adicionalmente, uma diferença de apenas R\$16,28 foi verificada nos custos médios associados a gestantes com idade de risco para aquelas que não tinham idade de risco.

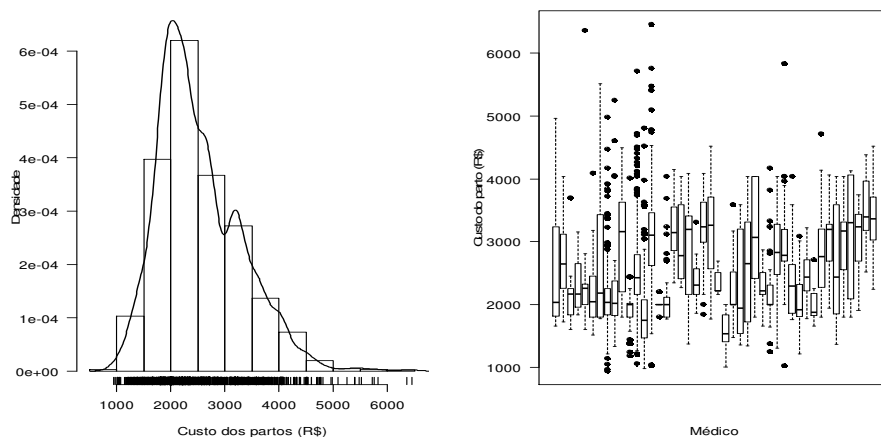


Figura 1 - Distribuição dos custos dos partos.

Tabela 1 - Descritiva dos custos dos partos segundo as categorias das covariáveis

Variável	Categoria	Frequência	Média	Mediana	Desvio
Tipo de parto	Cesariana	2.272	2.535,56	2.349,00	737,38
	Normal	240	2.395,19	2.103,60	969,87
Local de internação	1	426	2.083,22	2.004,10	481,91
	2	261	1.961,77	2.006,00	235,18
	3	589	2.473,82	2.323,07	567,44
	4	302	1.831,14	1.735,65	519,39
	5	49	2.408,04	1.980,00	1.035,15
	6	885	3.172,99	3.178,00	623,58
Risco associado à idade	Não	2.025	2.519,00	2.327,90	767,65
	Sim	487	2.535,28	2.356,66	746,85
Tipo de acomodação	Apartamento	1.040	2.785,27	2.685,20	798,85
	Enfermaria	1.472	2.336,25	2.216,89	678,90
Operadora	A	272	2.340,65	2.166,00	690,74
	D	1.668	2.544,83	2.349,00	776,92
	P	572	2.542,33	2.349,00	746,69

Dando sequência à análise, foram ajustados modelos de regressão com diferentes distribuições especificadas para a variável resposta. Ao considerar as distribuições contínuas disponíveis no pacote GAMLSS, verificou-se, com base em medidas de qualidade de ajuste e gráficos de resíduos que as duas que proporcionaram melhores

ajustes foram a BCPE e a ST3. Para efeito de comparação com os ajustes produzidos pelas distribuições Normal e Gama, a Tabela 2 apresenta as deviances, AIC's e BIC's gerados pelos modelos correspondentes às quatro distribuições. Pode-se notar, com base nesses resultados, menores valores produzidos pelas distribuições BCPE e ST3 em relação aos produzidos pelas distribuições Normal e Gama para ambos os critérios.

Tabela 2 - Medidas de qualidade de ajuste para os seis modelos de regressão considerados.

Modelo ajustado	Deviance	AIC	BIC	Graus de liberdade
Normal	38.549,74	38.609,52	38.783,75	29,89
Gama	38.041,00	38.098,28	38.265,23	28,64
BCPE (covariáveis apenas para μ)	37.406,82	37.486,45	37.718,53	39,81
ST3 (covariáveis apenas para μ)	37.614,48	37.701,06	37.953,40	43,29
BCPE (covariáveis para μ e σ)	37.053,98	37.222,51	37.713,69	84,27
ST3 (covariáveis para μ e σ)	37.087,53	37.213,81	37.581,82	63,14

Adicionalmente foram ajustados modelos em que também o parâmetro de dispersão depende de covariáveis (local de internação e médico responsável pelo parto). Deve-se ressaltar que essas duas variáveis foram consideradas nesta etapa da modelagem com base nos resultados da análise descritiva e na própria percepção dos analistas quanto à existência de uma acentuada variação dos custos dos partos associada a tais fatores. Ainda com base na Tabela 2 é possível verificar diminuição nos valores das estatísticas de qualidade de ajuste resultantes da inclusão dessas variáveis para modelar a dispersão do custo.

A Figura 2 apresenta gráficos para os resíduos quantílicos aleatorizados produzidos pelos modelos ajustados com as quatro distribuições consideradas. Ao analisar a distribuição dos resíduos, fica evidente a melhor adequação dos ajustes proporcionados pelas distribuições BCPE e ST3, verificando-se resíduos distribuídos, quase na totalidade, entre -3 e 3, com variância aproximadamente constante e boa aderência à distribuição Normal.

Com base nos resultados apresentados, foram escolhidos, para sequência da análise, modelos com distribuição BCPE e ST3, incluindo as variáveis tipo de parto, local de internação, tipo de acomodação, operadora e médico responsável pelo parto (efeito aleatório) na modelagem do parâmetro de locação e local de internação e efeito aleatório de médico na modelagem da dispersão. Cumpre informar que tentativas adicionais foram realizadas considerando as distribuições Normal e Gama, utilizando transformações e inclusão de covariáveis na modelagem da dispersão, sem alcançar resultados satisfatórios. Deve-se ressaltar ainda que a variável referente ao risco associado à idade da gestante não apresentou efeito significativo no custo do parto, segundo o teste de Wald, para ambos os modelos. Isso justificou sua exclusão da análise.

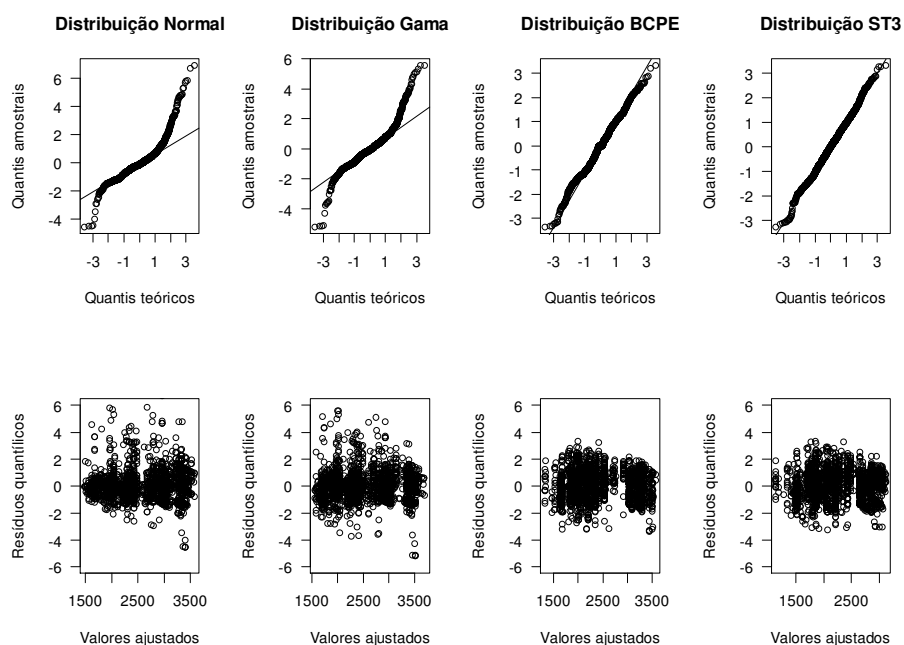


Figura 2 - Gráficos de resíduos para os modelos de regressão ajustados para as distribuições Normal, Gama, BCPE e ST3.

A Tabela 3 apresenta o resumo dos modelos de regressão ajustados com as distribuições BCPE e ST3. Pode-se observar que embora as estimativas dos parâmetros variem de um modelo para o outro, as conclusões, em geral, são as mesmas, verificando-se menor custo associado a partos normais, realizados no hospital 4 e com acomodação em enfermaria. O hospital 6 está associado a maiores custos, bem como partos vinculados à operadora 3 (particularmente quanto ao modelo com a distribuição ST3). Com relação à dispersão, para ambas as distribuições têm-se menor efeito associado ao hospital 2. É importante destacar que não se esperava que as estimativas associadas às duas distribuições fossem idênticas, uma vez que as parametrizações dos dois modelos são distintas.

Os efeitos estimados associados aos médicos, tanto na locação quanto na dispersão da distribuição, são apresentados na Figura 3. É importante verificar nessa figura a variação dos efeitos e a correlação existente entre os efeitos gerados pelos dois modelos. Há uma boa correlação entre os efeitos estimados sob os dois modelos, sobretudo quanto ao efeito do médico na dispersão. Algumas exceções podem ser verificadas, particularmente, com relação a médicos com efeito quase nulo na locação, segundo o modelo BCPE, mas com efeito bastante positivo segundo o modelo ST3.

Complementando a análise, foram identificadas as características dos partos contidos na base que produziram, segundo o modelo ST3, maior estimativa para o custo

mediano. O mesmo foi feito em relação aos partos com menor estimativa de custo mediano. Esses resultados são apresentados na Tabela 4.

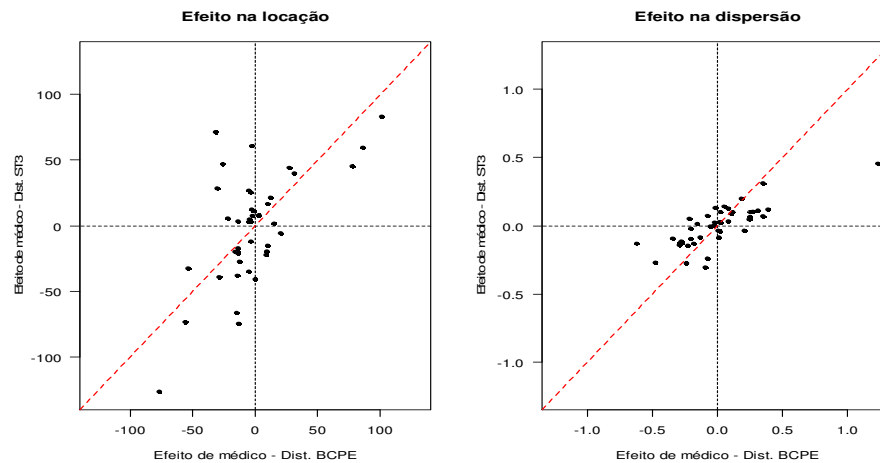


Figura 3 – Efeitos estimados para os médicos (na locação e na dispersão) segundo os modelos BCPE e ST3.

Tabela 3 - Resumo dos modelos de regressão ajustados.

Parâmetro	Distribuição BCPE				Distribuição ST3		
	Termo	Estimat.	Erro padrão	<i>t</i>	Estimat.	Erro padrão	<i>t</i>
μ	Intercepto	1993,92	21,86	91,2	2041,02	5,81	351,1
	Parto Normal	-388,24	16,04	-24,2	-256,20	13,15	-19,48
	Hosp. 2	12,38	13,83	0,89	-18,90	4,50	-4,20
	Hosp. 3	333,66	16,70	19,98	398,34	5,85	68,09
	Hosp. 4	-306,50	20,40	-15,0	-154,88	10,22	-15,16
	Hosp. 5	81,47	65,48	1,24	21,53	28,59	0,75
	Hosp. 6	962,48	19,38	49,66	1300,46	8,24	157,7
	Acomod. Enferm.	-206,11	9,04	-22,8	-200,78	1,17	-171,4
	Operad. 2	2,01	20,15	0,10	-12,68	5,33	-2,38
	Operad. 3	53,34	21,29	2,51	188,89	5,44	34,72
σ	Intercepto	5,07	0,05	112,16	-1,82	0,06	-31,79
	Hosp. 2	-0,84	0,07	-11,80	-1,60	0,09	-17,23
	Hosp. 3	0,32	0,06	5,31	-0,01	0,08	-0,20
	Hosp. 4	0,37	0,07	5,13	0,58	0,09	6,57
	Hosp. 5	0,62	0,11	5,72	0,62	0,18	3,51
	Hosp. 6	0,93	0,06	15,57	0,29	0,07	4,16
ν		0,47	0,01	31,47	-0,33	0,10	-3,33
τ		1,23	0,05	23,76	-0,34	0,03	-11,93

Tabela 4 - Estimativas e características dos 10 partos com maiores (e menores) custos medianos estimados segundo o modelo de regressão com distribuição ST3.

Tipo parto	Local de internação	Médico	Acomod.	Operadora	Custo mediano	P(Custo <2500)	P(Custo >4000)
Cesárea	6	133	Apartam.	3	3.315,78	0,0129	0,1060
Cesárea	6	58	Apartam.	3	3.317,12	0,0469	0,1780
Cesárea	6	169	Apartam.	2	3.319,66	0,0374	0,1650
Cesárea	6	64	Apartam.	2	3.327,83	0,0393	0,1722
Cesárea	6	77	Apartam.	3	3.328,16	0,0226	0,1402
Cesárea	6	93	Apartam.	3	3.349,21	0,0424	0,1881
Cesárea	6	174	Apartam.	3	3.351,80	0,0308	0,1699
Cesárea	6	132	Apartam.	3	3.356,85	0,0381	0,1854
Cesárea	6	169	Apartam.	3	3.377,98	0,0301	0,1821
Cesárea	6	64	Apartam.	3	3.386,16	0,0318	0,1897
Normal	4	31	Enferm.	1	1.301,56	0,9741	0,0027
Normal	4	31	Enferm.	2	1.303,43	0,9739	0,0027
Normal	4	31	Enferm.	3	1.361,76	0,9706	0,0029
Normal	4	69	Apartam.	2	1.395,86	0,9875	0,0009
Normal	2	105	Enferm.	2	1.427,36	0,9934	0,0004
Normal	1	28	Enferm.	2	1.478,55	0,9959	0,0002
Normal	4	77	Apartam.	2	1.499,88	0,9766	0,0001
Normal	4	31	Apartam.	1	1.512,13	0,9595	0,0035
Normal	4	31	Apartam.	2	1.513,99	0,9593	0,0035
Normal	1	28	Enferm.	3	1.536,87	0,9952	0,0002

Também foram estimadas as probabilidades de custos inferiores a R\$2.500,00 e superiores a R\$4.000,00 para esses 20 partos. É importante notar que, dentre os dez partos mais “caros”, todos são do tipo cesárea, realizados no hospital 6 e com padrão apartamento. Os mais “baratos”, por sua vez, correspondem a partos normais, em sua maioria com padrão enfermagem e realizados no hospital 4. Pode-se observar que a probabilidade estimada de um parto custar menos de R\$2.500,00 varia de 0,0129 a 0,9952, enquanto a probabilidade estimada de um parto com custo superior a R\$4.000,00 varia de 0,0002 a 0,1897. Embora estimativas referentes à distribuição BCPE não sejam apresentadas, por brevidade, destaca-se aqui que os resultados foram bastante semelhantes com relação às características dos partos.

Conclusões

A análise dos custos de procedimentos médicos e hospitalares requer a utilização de modelos estatísticos adequados, capazes de lidar com distribuições fortemente assimétricas, elevada dispersão, acomodar adequadamente efeitos aleatórios e não lineares. Nesse contexto, o presente trabalho apresenta a análise dos custos de partos cobertos por uma operadora do município de Curitiba-PR no ano de 2013. O interesse em modelar os custos dos partos justifica pelo fato de tais procedimentos envolverem internação, gerando elevados custos para a operadora. Como covariáveis foram

consideradas características do parto, da gestante, além do local de internação e do médico responsável pelo procedimento.

Verificou-se, inicialmente, que o uso das distribuições normal e gama não proporcionaram ajustes adequados. Dentre as distribuições contínuas implementadas no pacote GAMLSS do software R, destacaram-se, produzindo melhores ajustes, a *Box-Cox Power Exponential* (BCPE) e a *Skew-t Type 3*(ST3). Foram modelados conjuntamente os parâmetros de locação (com base em todas as variáveis disponíveis) e dispersão (com base nos efeitos associados ao local de internação e médico). O médico responsável pelo parto foi incluído nos modelos por meio de um par de efeitos aleatórios (um para cada parâmetro).

Para ambas as distribuições, foi verificado maior custo para partos do tipo cesárea e acomodação em apartamento. Adicionalmente, verificou-se grande variação associada aos hospitais em que os procedimentos foram realizados. A idade da mãe, incorporada aos modelos por meio de uma variável dicotômica que expressava o risco do parto, não apresentou efeito significativo e foi excluída dos modelos. A extração das estimativas dos efeitos dos médicos permitiu identificar médicos associados a maiores custos, sendo que os efeitos gerados pelos dois modelos apresentaram boa correlação.

Dessa forma, o presente trabalho mostra a conveniência em se considerar a utilização de modelos generalizados aditivos para locação, escala e forma na análise dos custos de procedimentos médicos e hospitalares.

KIENEN, R.S.; TACONELI, C.A. Generalized additive models for location, scale and shape in an analysis of hospital procedures costs financed by a health insurance company. *Rev. Bras. Biom.*, v.33, n.3, p.330-342, 2015.

- *ABSTRACT: Statistical modeling of costs associated with medical and hospital procedures is a very complex task, being common to find heavily skewed distributions, multiple associated factors, nonlinear effects and heterogeneous variances. Analysis of such data requires the use of statistical methods that properly handle data with such features. In this context, this work presents an application of Generalized Additive Model application for Location, Scale and Shape (GAMLSS) in the analysis of child-birth costs, financed by a health insurance company, in Curitiba-PR, 2013. It was possible to consider a wider variety of distributions for the random component, to incorporate random effects and jointly modeling location and dispersion parameters using covariates. It was found that Box-Cox power exponential (BCPE) and Skew-T type 3 (ST3) distributions provided better fits. For models fitted with these two distributions, factors such as child-birth type, place of hospitalization and accommodation type shown to be related to the proceedings costs. The doctor responsible for child-birth was included to the models by random effects, allowing identifying their contribution to the final cost and assessing the correlation between the effects produced by the two distributions. The effects of medical and place of hospitalization were also considered in modeling the costs dispersion, verifying that such specification contributed to a better fit.*
- *KEYWORDS: Child-birth costs; generalized additive models for location; scale and shape; random effects; health insurance.*

Referências

- AGÊNCIA NACIONAL DE SAÚDE (ANS). Dados Gerais. Rio de Janeiro, RJ, 2013. Disponível em: <<http://www.ans.gov.br/materiais-para-pesquisas/perfil-do-setor/dados-gerais>> Acesso em: 14/08/2014.
- AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. Notre Dame, v.19, n.6, p.716-723, 1974.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*. Malden, v.26, n.2, p.211-252, 1964.
- CARLIN, B P.; L, Thomas A. *Bayes and empirical Bayes methods for data analysis*. New York: Chapman & Hall, 2000. 419p.
- CASELLA, G. An introduction to empirical Bayes data analysis. *The American Statistician*, v.39, n. 2, p.83-87, 1985.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, v.5, n.3, p.236-244, 1996.
- FERNÁNDEZ, C.; STEEL, M. F. J. On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, v.93, n.441, p.359-371, 1998.
- FUMES, G. *Modelos Box-Cox simétricos e aplicações a dados nutricionais*. 2014. 99f. Tese (Doutorado em Ciências) – Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2014.
- KAISER FAMILY FOUNDATION. Health Care Costs: A Primer Key Information on Health Care Costs and their Impact. Mai. 2012. Disponível em: <http://kaiserfamilyfoundation.files.wordpress.com/2013/01/7670-03.pdf> Acesso em: 20/07/2014.
- RIGBY, R. A.; STASINOPOULOS, D. M. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, v.54, n. 3, p. 507-554, 2005.
- RIGBY, R. A.; STASINOPOULOS, D. M. *A flexible regression approach using GAMLSS in R*. London Metropolitan University, London, 2009. 282p.
- SCHWARZ, G. Estimating the dimension of a model. *The annals of statistics*, v.6, n.2, p.461-464, 1978.
- STASINOPOULOS, D. M; RIGBY, R. A. Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, v.23, n.7, p.1-46, 2007.
- TEAM, R. Core. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2014.
- VOUDOURIS, V.; GILCHRIST, R.; RIGBY, R. A.; SEDGWICK, J.; STASINOPOULOS, D. M. Modelling skewness and kurtosis with BCPE density in GAMLSS. *Journal of Applied Statistics*, v.39, n.6, p.1279-1293, 2012.

Recebido em 10.03.2015

Aprovado após revisão em 02.06.2015