

## **CADERNOS DO IME – Série Estatística**

Universidade do Estado do Rio de Janeiro - UERJ  
Rio de Janeiro - RJ - Brasil  
ISSN impresso 1413-9022 / ISSN on-line 2317-4536 - v. 33, p. 01 - 17, 2012

# **TARIFICAÇÃO DE UM PLANO DE SAÚDE AUTOGESTÃO APLICANDO OS MODELOS LINEARES GENERALIZADOS**

Helano Silva Eugênio de Souza  
Instituto Brasileiro de Mercado de Capitais (IBMEC)  
helanosouza@petrobras.com.br

Luiz Carlos da Silva Leão  
Instituto de Matemática e Estatística (UERJ)  
luiz.leao@petrobras.com.br

### **Resumo**

*O presente trabalho aplicou os Modelos Lineares Generalizados para fazer o processo de tarifação de um plano de saúde autogestão. O trabalho investigou que os custos médicos seguem uma distribuição de probabilidade da família exponencial. Assim sendo, foi possível aplicar os Modelos Lineares Generalizados a fim de se obter uma nova tarifa, que é diferente do que geralmente tem sido praticado no mercado de planos de saúde na modalidade autogestão. Essa nova metodologia é caracterizada por capturar por elementos qualitativos dos participantes expostos que processos de tarifação antigos não eram capazes de capturar. Os resultados mostraram que, dentre as variáveis escolhidas para o modelo, a variável renda tem uma relação inversa com o risco de gastos médicos. Indivíduos com maiores salários tendem a utilizar o plano de saúde menos que a média. Indivíduos do Gênero Masculino têm gastos médicos superiores aos dos indivíduos do Gênero Feminino e isso se reflete na tarifa atuarialmente determinada. Finalmente, o modelo foi capaz de gerar tabelas de tarifas que poderão ser cobradas dos participantes cobrindo todo o risco médio (valor esperado de utilização médica).*

**Palavras-chave:** Atuária, Saúde, Tarifação.

## 1. Introdução

No mercado brasileiro de Planos de Saúde, operadoras de autogestão são aquelas nas quais a própria empresa estabelece e gerencia, sem propósito de gerar lucro, o plano de saúde de seus empregados segundo a RN 279/2011 da ANS, Agência Nacional de Saúde Suplementar (2011). Pode-se mencionar o Plano CASSI do Banco do Brasil, o Plano AMS, da Petrobras S/A.

Essas operadoras têm a tradição de aplicar métodos simples, do ponto de vista atuarial e estatístico, para tarifar seus planos de saúde. Dois destes métodos são: a aplicação de um percentual fixo do salário do participante ou uma tabela única de coparticipação contributiva.

Dado que a distribuição de probabilidade dos gastos com utilização médica de planos de saúde tradicionalmente segue uma distribuição da família exponencial (Jones, 2010), o que foi constatado com a experiência de utilização médica utilizada para este artigo, foi possível aplicar os Modelos Lineares Generalizados para se obter valores de tarifa a serem cobradas aos participantes de um plano de saúde autogestão, obtendo assim uma nova forma de precificá-lo.

Esta metodologia considera o perfil de cada participante (risco individual), similar ao que é praticado hoje no mercado aberto de planos de saúde, levando em consideração variáveis como idade, gênero, renda, dentre outras.

Foi considerada para este artigo uma base de dados de um plano de saúde autogestão de 300.069 vidas, no período de exposição de 2007 a 2010 inclusive, segregados em gastos ambulatoriais (consultas, exames, etc.) e de internação (cirurgias, etc.).

Finalmente, pela manipulação dos resultados dos MLG (Modelos Lineares Generalizados) estatisticamente obtidos, a modelagem proposta foi capaz de gerar tabelas de contribuição contendo tarifas que poderão ser cobradas dos participantes do plano, cobrindo todo o risco médio (Valor esperado de Sinistros do Plano).

Podemos citar Roosevelt (1994) e Fleischacker *et al.* (1994) como trabalhos que já discorreram sobre precificação de risco em saúde pela metodologia dos Modelos Lineares Generalizados e Griswold *et al.* (2004) que compara algumas metodologias, tais como mínimos quadrados ordinários, modelos lognormais e COX PHM.

## 2. Modelos Lineares Generalizados

De acordo com Dobson (2001), Modelos Lineares Generalizados são utilizados para avaliar e quantificar a relação entre uma variável resposta ( $y$ ) e variáveis explicativas ( $x_1, x_2, \dots, x_n$ ). Os MLG diferem do modelo linear clássico em dois aspectos:

- a) A distribuição de probabilidade da variável resposta é escolhida da família exponencial (Poisson, Binomial, Gamma, Normal, Binomial Negativa e Gaussiana Inversa);
- b) Uma transformação do valor esperado da variável resposta é linearmente relacionada com as variáveis explicativas;

Funções da Família Exponencial são do tipo (1):

$$f(y) = c(y, \phi) \exp \left\{ \frac{y\theta - a(\theta)}{\phi} \right\} \quad (1)$$

Isso envolve uma variável resposta ( $y$ ), variáveis explicativas ( $x$ ) e uma amostra aleatória de ( $n$ ) observações independentes:

$$\begin{aligned} g(\mu) &= \eta = X^T \beta, \quad E(y) = \mu \\ \eta &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p, \quad n > p \end{aligned}$$

Tendo:

- Componente aleatório do modelo, distribuído de acordo com a distribuição da família exponencial;
- Variáveis explicativas da estrutura linear, constituindo o componente sistemático do modelo;
- A conexão feita através de uma função link monotônica e diferenciável. Esta função link descreve a relação entre o valor esperado da variável resposta e as variáveis explicativas.

Passos para modelagem com MLG:

- a) Escolha da distribuição de  $f(y)$ ;
- b) Determinar a função de ligação  $g(\mu)$ , que pode ser simplificado escolhendo o parâmetro correspondente da canônica  $f(y)$ ;
- c) Determinar as variáveis explanatórias em termos de  $g(\mu)$  sobre as quais o modelo será estimado;

- d) Coletar as  $y_1, y_2, y_3, y_4, \dots, y_n$  observações e os correspondentes  $x_1, x_2, x_3, x_4, \dots, x_n$  ;
- e) Ajustar o modelo estimando  $\beta$ , o que é geralmente feito por um software estatístico;
- f) Com  $\beta$  estimado, é verificado se o modelo ficou ajustado e se as variáveis explicativas são importantes para determinar  $\mu$ .

Nota: Se  $g(\mu) = \theta$ , então  $g$  é chamado de ligação canônica correspondente para  $a(\theta)$ .

As funções de ligação mais comumente utilizadas estão na tabela 1.

Tabela 1 – Funções de ligação comumente utilizadas

<b>Função de Ligação</b>	$g(\mu)$	<b>Ligação Canônica para</b>
Identidade	$\mu$	Normal
Log	$\ln \mu$	Poisson
Power	$\mu^{-1}$	Gamma
Logit	$\ln \left\{ \frac{\pi}{1-\pi} \right\}$	Binomial

Fonte: Dobson (2001)

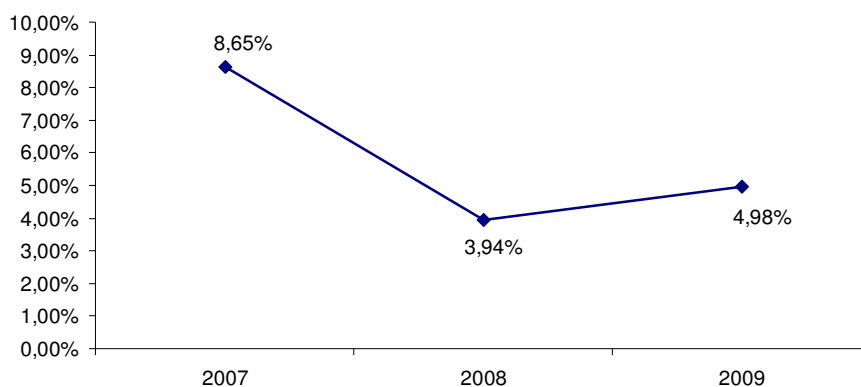
### 3. Metodologia

A base de dados utilizada refere-se a um plano de 300.069 vidas expostas no período de Janeiro de 2007 a Dezembro de 2010, correspondendo a despesas ambulatoriais (inclusive de odontologia) e despesas de internação. Primeiramente foi criado um projeto num software estatístico de licença paga SAS (2012) e depois criado um projeto no software estatístico de licença livre R (2005). Todos os dados foram consolidados para uma massa única, ou seja, de 4 (quatro) anos de exposição ao risco.

Para assumir a posição correta das despesas médicas no tempo, as despesas referentes aos anos de 2007, 2008 e 2009 foram atualizadas monetariamente para o ano de 2010 (último ano da análise considerado). Esta atualização foi feita através do Índice

de Variação dos Custos Médico-Hospitalares – VCMH<sup>1</sup> Cechin *et al.* (2004) observada na mesma experiência estatística do plano (Figura 1).

Figura 1 – Variação dos Custos Médico-Hospitalares observado de 2007 a 2009



Fonte: Experiência própria do plano

Os passos descritos foram adotados para se obter uma massa de dados maior e mais consistente, considerando a Lei dos Grandes Números, segundo a qual quanto maior a amostra, mais consistente a análise estatística DeGroot (2011).

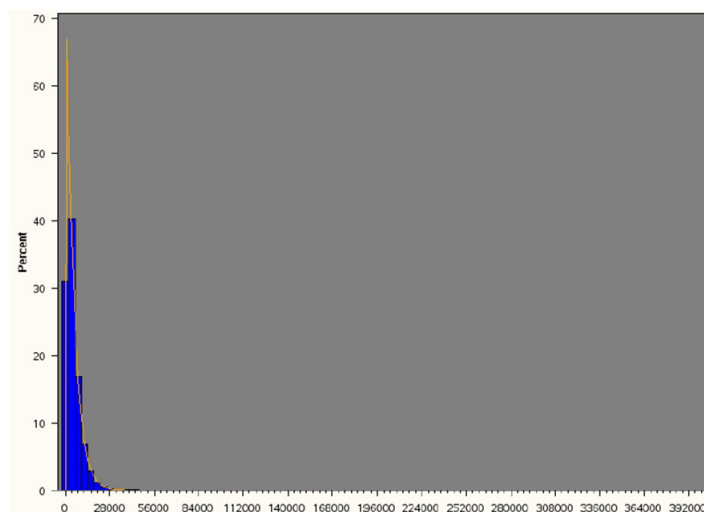
O propósito principal deste artigo é encontrar um modelo explanatório para despesas ambulatoriais e de internação para tarifar um plano de saúde autogestão. Para tanto, foi necessário analisar o comportamento estatístico destas variáveis aleatórias (despesa ambulatorial e de internação total no período *per capita*).

Como pode ser observado na Figura 2 e estatisticamente observado a seguir, a distribuição de probabilidade das despesas ambulatoriais segue, com alto nível de significância, uma distribuição de probabilidade da família exponencial.

A Tabela 2 mostra os resultados da análise da distribuição de probabilidade obtida com teste de aderência do software estatístico para distribuição Gamma. Neste teste, os p-valores dos testes de aderência de Kolgomorov-Smirnov DeGroot (2011) e Cramer-von-Mises Anderson (1962), resultaram abaixo de 0,001 por cento, na região de não rejeição das hipóteses estatísticas do teste.

<sup>1</sup> O Índice de Variação dos Custos Médico-Hospitalares é obtido calculando-se a variação do total de despesas *per capita* entre dois anos consecutivos. A expertise interna do provedor dos dados sugere que, considerando uma cesta definida de itens médicos (como é feito na medição de índices de inflação comuns), três elementos chaves compõem a VCMH: variação dos preços da cesta corrente, variação da demanda pela cesta corrente e mudanças nos itens da cesta. Assim sendo, pode-se enunciar a inflação médica, novas tecnologias, mudanças no regulamento do plano, mudança no perfil etário da população, regulação governamental dentre outros, como componentes chaves.

Figura 2 – Histograma de Despesas Ambulatoriais



Fonte: Base de Dados

As hipóteses do teste são:

- $H_0$ : A Distribuição de Probabilidade das Despesas Ambulatoriais segue uma distribuição Gamma;
- $H_1$ : A Distribuição de Probabilidade das Despesas Ambulatoriais segue outra distribuição que não a Gamma;

Considerando os resultados, não há evidência estatística para rejeitar a hipótese nula.

Tabela 2 – Resultado do Teste de Aderência Despesas Ambulatoriais

Teste de Aderência para Distribuição Gamma				
Testes	Estatística		p-Valor	
Kolmogorov-Smirnov	D	0,014003	Pr > D	<0,001
Cramer-von Mises	W-Sq	15,129909	Pr > W-Sq	<0,001

Fonte: Elaboração própria

A Tabela 3 abaixo mostra uma diferença pequena entre os valores observados e estimados pelo software estatístico, denotando o bom ajustamento da distribuição Gamma.

Esta inferência, de que as despesas ambulatoriais seguem, com alto nível de confiança, uma distribuição de probabilidade Gamma, é crítica para aplicação dos modelos lineares generalizados.

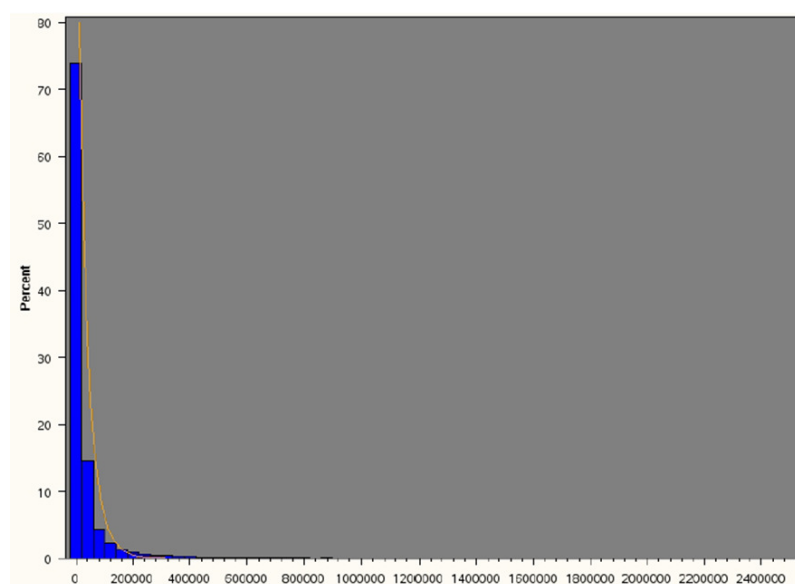
Tabela 3 – Percentis da distribuição Gamma da variável aleatória Despesa Ambulatorial

Percentis para Distribuição Gamma		
Percentil	Percentil	
	Observado	Estimado
1,0	682,370	731,709
5,0	3.095,427	3.209,225
10,0	6.375,972	6.207,281
25,0	15.988,565	15.706,024
50,0	35.311,429	35.744,351
75,0	66.345,495	68.882,428
90,0	108.869,620	111.876,220
95,0	142.037,432	144.103,302
99,0	232.919,522	218.420,473

Fonte: Elaboração própria

Como pode ser observado na Figura 3 e estatisticamente observado abaixo, a distribuição de probabilidade das despesas de internação segue, com alto nível de significância, uma distribuição de probabilidade da família exponencial.

Figura 3 – Histograma de Despesas de Internação



Fonte: Base de Dados

A Tabela 4 mostra os resultados da análise da distribuição de probabilidade obtida com teste de aderência do software estatístico para distribuição Gamma. Neste teste, os p-valores dos testes de aderência de Kolgomorov-Smirnov, DeGroot (2011), e

Cramer-von-Mises, Anderson (1962), resultaram abaixo de 0,001 por cento, na região de não rejeição das hipóteses estatísticas do teste.

As hipóteses do teste são:

- $H_0$ : A Distribuição de Probabilidade das Despesas de Internação segue uma distribuição Gamma;
- $H_1$ : A Distribuição de Probabilidade das Despesas de Internação segue outra distribuição que não a Gamma;

Tabela 4 – Resultado do Teste de Aderência Despesas de Internação

Teste de Aderência para Distribuição Gamma				
Teste	Estatística		p-Valor	
Kolmogorov-Smirnov	D	0,346080	Pr > D	<0,001
Cramer-von Mises	W-Sq	4.711,501060	Pr > W-Sq	<0,001

Fonte: Elaboração própria

A Tabela 5 abaixo mostra uma diferença pequena entre os valores observados e estimados pelo software estatístico, denotando o bom ajustamento da distribuição Gamma.

Tabela 5 – Percentis da distribuição Gamma da variável aleatória Despesa de Internação

Percentis para Distribuição Gamma		
Percentil	Percentil	
	Observado	Estimado
1,0	195,135	119,452
5,0	595,753	900,596
10,0	1.281,669	2.178,596
25,0	3.399,607	7.329,916
50,0	7.560,317	20.612,864
75,0	21.526,483	45.467,496
90,0	72.083,501	79.977,694
95,0	142.323,181	106.697,087
99,0	447.781,281	169.796,233

Fonte: Elaboração própria

Esta inferência, de que as despesas com internação seguem, com alto nível de confiança, uma distribuição de probabilidade Gamma, também é crítica para aplicação dos modelos lineares generalizados.

Estes fenômenos descritos acima já foram observados por Griswold *et. al* (2004) e Jones (2010).



#### 4. Modelagem

Neste artigo, serão usados os modelos lineares generalizados para modelagem das despesas médicas ambulatoriais e de internação. O MLG deve ser utilizado quando a variável resposta ( $y$ ) pertence à família exponencial.

O propósito do modelo é explicar as despesas médicas através de variáveis qualitativas de cada participante a fim de obter a tarifa estatisticamente correspondente.

Os parâmetros testados para o modelo que tiveram a melhor performance estatística pelo critério do teste de Wald<sup>2</sup> foram: Idade, Gênero, Tipo de Usuário e Salário (tabela 6).

Tabela 6 – Parâmetros escolhidos para modelagem pelo critério do teste de Wald

Faixa Etárias	0 a 18 anos
	19 a 23 anos
	24 a 28 anos
	29 a 33 anos
	34 a 38 anos
	39 a 43 anos
	44 a 48 anos
	49 a 53 anos
	54 a 58 anos
	Maior que 58 anos
Gênero	Feminino
	Masculino
Tipo de Usuário	01 – Titular
	02 – Esposo(a) ou Parceiro(a)
	03 – Filho(a) menor de 24 anos
	05 - Outros (Pais, Avós, Agregados)
Faixa Salarial <sup>3</sup>	1 – Menos que R\$ 882,14
	2 – Acima de R\$ 882,14 e menos que R\$ 1.628,57
	3 – Acima de R\$ 1.628,57 e menos que R\$ 3.257,14
	4 – Acima de R\$ 3.257,14 e menos que R\$ 6.075,07
	5 – Acima de R\$ 6.075,07 e menos que R\$ 12.150,14
	6 – Acima de R\$ 12.150,14

Fonte: Elaboração própria

<sup>2</sup> Para testar em simultâneo hipóteses sobre várias combinações lineares dos parâmetros, usa-se a estatística de Wald, Engle (1983)

<sup>3</sup> Real (R\$) – Moeda brasileira. Salários posicionados em 2010.

#### 4.1. Despesas Ambulatoriais

Os resultados da modelagem estatística das despesas ambulatoriais estão descritos na tabela 7.

Tabela 7 – Resultados da Modelagem da Despesa Ambulatorial

Análise de Máxima Verossimilhança dos Parâmetros Estimados								
Parâmetro		GL	Estimativa	Desvio Padrão	Intervalos de Confiança Wald 95%		Estatística Qui-Quadrado Wald	Prob > Qui-Quadrado
Intercepto		1	8,4856	0,0072	8,4715	8,4997	1392133	<,0001
Tipo de Usuário	1	1	0,2440	0,0065	0,2313	0,2566	1429,42	<,0001
	2	1	0,2258	0,0068	0,2125	0,2392	1093,71	<,0001
	3	1	0,3137	0,0216	0,2713	0,3560	210,59	<,0001
	5	0	0,0000	0,0000	0,0000	0,0000	.	.
Faixa Salarial	1	1	-0,4461	0,0276	-0,5002	-0,3919	260,83	<,0001
	2	1	-0,2770	0,0160	-0,3083	-0,2457	300,75	<,0001
	3	1	-0,2088	0,0072	-0,2229	-0,1946	838,52	<,0001
	4	1	-0,1210	0,0050	-0,1308	-0,1111	577,88	<,0001
	5	1	-0,0463	0,0046	-0,0552	-0,0373	102,95	<,0001
	6	0	0,0000	0,0000	0,0000	0,0000	.	.
Gênero	Feminino	1	0,2900	0,0041	0,2820	0,2980	5038,06	<,0001
	Masculino	0	0,0000	0,0000	0,0000	0,0000	.	.
Faixa Etária	0 a 18	1	-0,8932	0,0217	-0,9358	-0,8506	1689,57	<,0001
	19 a 23	1	-0,9813	0,0213	-1,0230	-0,9395	2121,82	<,0001
	24 a 28	1	-0,9512	0,0071	-0,9650	-0,9374	18203,3	<,0001
	29 a 33	1	-0,8433	0,0074	-0,8577	-0,8289	13101,8	<,0001
	34 a 38	1	-0,6160	0,0091	-0,6339	-0,5981	4537,62	<,0001
	39 a 43	1	-0,4807	0,0091	-0,4986	-0,4628	2777,53	<,0001
	44 a 48	1	-0,3305	0,0073	-0,3449	-0,3161	2026,76	<,0001
	49 a 53	1	-0,2122	0,0069	-0,2258	-0,1987	938,55	<,0001
	54 a 58	1	-0,1435	0,0067	-0,1567	-0,1304	456,51	<,0001
	> 58	0	0,0000	0,0000	0,0000	0,0000	.	.

Fonte: Elaboração própria

O melhor ajustamento MLG para as despesas ambulatoriais foi Distribuição de Probabilidade Variável Resposta: Distribuição Gamma, função de ligação: Log pelo critério do *Deviance* (Dobson, 2001). Pode ser observado que todos os p-valores do Teste de Wald<sup>4</sup> se encontram na região de não rejeição, denotando a significância das

<sup>4</sup> Sobre a hipótese nula de independência, a estatística qui-quadrado de Wald aproximadamente segue uma distribuição qui-quadrado, equivalente ao teste estatístico qui-quadrado: um coeficiente particular da regressão é zero, dado todos os preditores do modelo.

variáveis explicativas escolhidas: Tipo de Usuário, Faixa Salarial, Gênero e Faixa Etária, Jong e Heller (2008).

De acordo com Bruin (2006), se o modelo se ajusta bem aos dados, a razão entre o *Deviance* e o número de Graus de Liberdade deve ser em torno de 1 (um).

Tabela 8 – Tabela de Critérios para o acessar o Teste de Aderência das Despesas Ambulatoriais

Critérios para acessar o Teste de Aderência			
Critério	Graus de Liberdade	Valor	Valor / Graus de Liberdade
Deviance	28E4	257084,2791	0,9058
Deviance Ajustado	28E4	319877,0484	1,1271
Qui-quadrado de Pearson	28E4	442588,8582	1,5595
Pearson Ajustado	28E4	550691,0734	1,9404
Log de Verossimilhança		2636846,8520	
Log de Verossimilhança Completo		2636846,8520	
AIC		5273733,7045	
AICC		5273733,7045	
BIC		5273944,8268	

Fonte: Elaboração própria

Como pode ser observada na Tabela 8, a razão Deviance sobre número de Graus de Liberdade foi de 0,9058 denotando que a modelagem das Despesas Ambulatoriais está bem ajustada para a distribuição de probabilidade escolhida (Gamma) e função de ligação (Log) pelo critério proposto por Bruin (2006).

#### 4.2. Despesas de Internação

Os resultados da modelagem estatística das despesas de internação estão na tabela 9.

Tabela 9 – Resultados da Modelagem da Despesa de Internação

Análise de Máxima Verossimilhança dos Parâmetros Estimados								
Parâmetro		GL	Estimativa	Desvio Padrão	Intervalos de Confiança Wald 95 %		Estatística Qui-Quadrado Wald	Prob > Qui-Quadrado
Intercepto		1	10,8728	0,0174	10,8387	10,9069	389793	<,0001
Gênero	Feminino	1	-0,1987	0,0114	-0,2210	-0,1764	306,05	<,0001
	Masculino	0	0,0000	0,0000	0,0000	0,0000	.	.
Faixa Etária	0 a 18	1	-1,4441	0,0672	-1,5758	-1,3124	461,82	<,0001
	19 a 23	1	-1,3592	0,0662	-1,4890	-1,2294	421,02	<,0001
	24 a 28	1	-1,6330	0,0207	-1,6737	-1,5924	6199,58	<,0001
	29 a 33	1	-1,3552	0,0209	-1,3962	-1,3142	4193,7	<,0001

	34 a 38	1	-1,2962	0,0249	-1,3449	-1,2474	2713,91	<,0001
	39 a 43	1	-1,1098	0,0255	-1,1597	-1,0598	1896,71	<,0001
	44 a 48	1	-0,9316	0,0200	-0,9708	-0,8924	2164,88	<,0001
	49 a 53	1	-0,7588	0,0182	-0,7945	-0,7230	1731,41	<,0001
	54 a 58	1	-0,5501	0,0171	-0,5836	-0,5165	1032,93	<,0001
	> 58	0	0,0000	0,0000	0,0000	0,0000	.	.
Tipo de Usuário	1	1	-0,3167	0,0154	-0,3469	-0,2865	422,75	<,0001
	2	1	-0,3962	0,0161	-0,4277	-0,3648	609,16	<,0001
	3	1	-0,3734	0,0664	-0,5035	-0,2433	31,65	<,0001
	5	0	0,0000	0,0000	0,0000	0,0000	.	.
Faixa Salarial	1	1	0,6750	0,0700	0,5378	0,8122	93,04	<,0001
	2	1	0,3927	0,0366	0,3210	0,4645	115,1	<,0001
	3	1	0,2364	0,0187	0,1996	0,2731	159,17	<,0001
	4	1	0,1047	0,0142	0,0769	0,1325	54,6	<,0001
	5	1	0,0046	0,0133	-0,0213	0,0306	0,12	<,0001
	6	0	0,0000	0,0000	0,0000	0,0000	.	.

Fonte: Elaboração própria

Pode ser observado que todos os p-valores do Teste de Wald se encontram na região de não rejeição, denotando significância das variáveis explicativas escolhidas.

Tabela 10 – Tabela de Critérios para o acessar o Teste de Aderência das Despesas de Internação

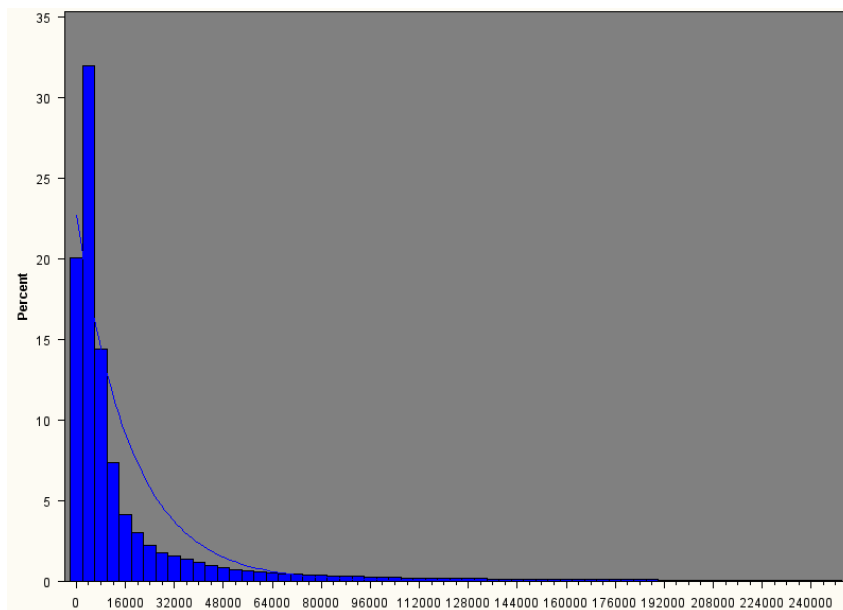
Critérios para acessar o Teste de Aderência			
Critério	Graus de Liberdade	Valor	Valor / Graus de Liberdade
Deviance	87E3	197.627,2154	2,2629
Deviance Ajustado	87E3	109.271,5437	1,2512
Qui-quadrado de Pearson	87E3	749.177,1876	8,5784
Pearson Ajustado	87E3	414.233,1694	4,7431
Log de Verossimilhança		-939838,7406	
Log de Verossimilhança Completo		-939838,7406	
AIC		1879717,4811	
AICC		1879717,4908	
BIC		1879905,0352	

Fonte: Elaboração própria

Como pode ser observado na Tabela 10, a razão Deviance sobre número de Graus de Liberdade foi de 2,2629, denotando que a modelagem das Despesas de Internação está bem ajustada.

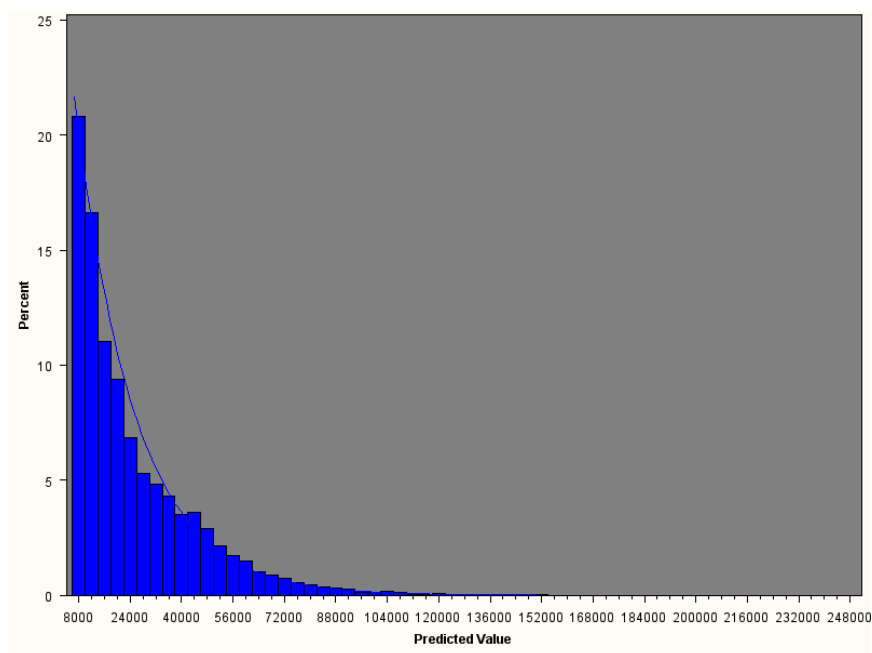
Analisando as figuras 4 e 5 abaixo, é perceptível o alinhamento entre os histogramas da despesa observada de internação e as estimadas pelos dados da modelagem MLG:

Figura 4 – Histograma de Despesas de Internação com dados observados



Fonte: Base de Dados

Figura 5 – Histograma de Despesas de Internação com dados estimados



Fonte: Base de Dados

Com os parâmetros estimados nas modelagens acima, cuja cobertura inclui despesas ambulatoriais e de internação, é possível obter as tabelas de tarifas pela fórmula fundamental do MLG, adicionando os respectivos valores (estimado ambulatorial mais estimado de internação):

$$H_i \sim \text{Gamma}$$

Onde  $H_i$  foi a distribuição Gamma escolhida para ambas as despesas.

Aplicando a transformação MLG com a função de ligação log (2):

$$\log(E(H_i)) = \eta_i = \alpha + \sum_{n=1}^p \beta_n D_n \quad (2)$$

Onde  $\alpha$  são os interceptos,  $\beta_n$  são os parâmetros obtidos nas Tabelas 7 e 9 através do MLG,  $D_n$  são variáveis dummies que calibram o modelo.

Seguindo, o valor esperado de  $H_i$  é igual à exponencial de  $\eta_i$ :

$$E(H_i) = \exp(\hat{\eta}_i) = \hat{\mu}_i$$

$$T = \frac{\hat{\mu}_i}{12 \times 4} \quad (3)$$

A tarifa  $T$  é obtida (3) dividindo-se  $\hat{\mu}_i$  por:

- 12 (doze): porque o período de exposição é de 12 meses e a tarifa deve ser calculada na base mensal, segundo Kass *et al.* (2008);
- 4 (quatro): porque o período de coleta da database, como mencionado no capítulo 2, é de 4 anos de exposição, segundo Kass *et al.* (2008).

Segue na Tabela 11 uma das 8 tabelas distintas com as tarifas estimadas via MLG para um plano de saúde autogestão (a primeira linha se refere às faixas salariais explicadas na Tabela 6). Neste exemplo, considerou-se o gênero feminino e o tipo de usuário titular:

Tabela 11 – Tabela de Tarifa MLG/ Gênero: Feminino / Tipo de Usuário: Titular

	1	2	3	4	5	6
<b>0 - 18</b>	R\$ 349,13	R\$ 282,67	R\$ 253,27	R\$ 234,32	R\$ 222,79	R\$ 225,27
<b>19 - 23</b>	R\$ 372,27	R\$ 298,48	R\$ 265,82	R\$ 244,28	R\$ 230,90	R\$ 233,04
<b>24 - 28</b>	R\$ 294,26	R\$ 240,20	R\$ 216,29	R\$ 201,22	R\$ 192,23	R\$ 194,65
<b>29 - 33</b>	R\$ 379,70	R\$ 306,71	R\$ 274,42	R\$ 253,49	R\$ 240,69	R\$ 243,26
<b>34 - 38</b>	R\$ 412,00	R\$ 336,27	R\$ 302,79	R\$ 281,66	R\$ 269,07	R\$ 272,45
<b>39 - 43</b>	R\$ 492,85	R\$ 400,95	R\$ 360,30	R\$ 334,43	R\$ 318,88	R\$ 322,69

<b>44 - 48</b>	R\$ 586,74	R\$ 476,49	R\$ 427,73	R\$ 396,55	R\$ 377,72	R\$ 382,13
<b>49 - 53</b>	R\$ 692,42	R\$ 560,46	R\$ 502,09	R\$ 464,45	R\$ 441,53	R\$ 446,41
<b>54 - 58</b>	R\$ 838,77	R\$ 673,54	R\$ 600,42	R\$ 552,37	R\$ 522,59	R\$ 527,60
<b>&gt; = 59</b>	R\$ 1.398,66	R\$ 1.102,06	R\$ 970,69	R\$ 880,97	R\$ 823,42	R\$ 828,18

Fonte: Elaboração própria

É observado que indivíduos com menores salários devem contribuir com tarifas mais altas para o plano. Isto ocorre porque o modelo capturou que indivíduos com menor renda tendem a utilizar mais o plano que indivíduos com maiores ganhos.

Devido a isto, suas tarifas são mais elevadas (similar ao que acontece com seguros de carros, onde indivíduos mais jovens têm um prêmio de seguro mais elevado que indivíduos mais idosos).

## 5. Conclusão

Os Modelos Lineares Generalizados se ajustaram muito bem à base de dados de custos médicos analisada.

Usando ferramentas estatísticas foi possível estimar um modelo de tarifação que considera as diferentes características entre os participantes do plano, de acordo com as práticas vigentes no mercado de seguros.

Ao contrário das práticas comuns, no mercado de seguros de saúde autogestão, vem sendo aplicado um percentual do salário do participante ou valores de uma tabela fixa de contribuição. Por este modelo, cada participante, com suas características individuais (idade, salário, gênero, etc.) tem sua tarifa específica, considerando o seu grau de risco no plano (níveis de utilização médica), semelhante ao que é feito no mercado aberto de seguro saúde.

## Referências

- ANDERSON, T. W. On the Distribution of the Two-Sample Cramer-von Mises Criterion (PDF). **The Annals of Mathematical Statistics** (Institute of Mathematical Statistics) 33 (3): 1148–1159. doi:10.1214/aoms/1177704477. ISSN 0003-4851. 1962. (Acesso em Junho de 2009).
- BRUIN, J. **Newtest: command to compute new test**. UCLA: Academic Technology Services, Statistical Consulting Group. 2006. Disponível em: <http://www.ats.ucla.edu/stat/stata/ado/analysis/> (Acesso em 14 de Junho de 2012).
- CECHIN, J.; MARTINS, C. B.; LEITE, F. **Variação dos Custos Médicos-Hospitalares**. IEES – Instituto de Estudos de Saúde Suplementar, Agosto/2009. Disponível em:

<http://www.iess.org.br/html/TD00252009VCMH.pdf> (Acesso em 25 de Novembro de 2012).

DEGROOT, M. H. **Probability and Statistics**. 4<sup>th</sup> ed. (233-234), 2011.

DOBSON A. J. **An Introduction to Generalized Linear Models**. 2<sup>nd</sup> ed.. Taylor e Francis, 2001.

ENGLE, Robert F.. Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. Intriligator, M. D.; and Griliches, Z. *Handbook of Econometrics II*. Elsevier. pp. 796–801. 1983

FLEISCHACKER, P.R. DISCENZA J.A, HUEY M.S. **Actuarial Issues Related to Pricing Health Plans Under Health Care Reform**, American Academy of Actuaries, Julho, 1994 <http://www.actuary.org/pdf/health/mono10.pdf> (Acesso em 2 de Julho de 2013)

GRISWOLD M.; PARMIGIANI G.; POTOSKY A.; LIMPSOMB J. Analysing Health Care Costs: A Comparison of Statistical Methods Motivated by Medicare Colorectal Cancer Charges **Biostatistics** (2004) 1,1, pp. 1-23. Disponível em: <http://www.biostat.jhsph.edu/project/seermed/paper/pap033104.pdf> (Acesso em 14 de Junho de 2012).

JONES, A. M. **Models for Health Care**. Econometrics and Data Group The University of York Heath. 2010. Disponível em: [http://www.york.ac.uk/res/herc/documents/wp/10\\_01.pdf](http://www.york.ac.uk/res/herc/documents/wp/10_01.pdf) (Acesso em 14 de Junho de 2012).

JONG P. D.; HELLER G. Z. **Generalized Linear Models for Insurance Data**. Cambridge University Press, 2008.

KAAS R.; GOOVAERTS M.; DHAENE J.; DENUIT M., **Modern Actuarial Risk Theory**. 1<sup>st</sup> ed. Springer Verlag Berlin Heidelberg, 2008.

**R: A Language and Environment for Statistical Computing** R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2005. Disponível em: <http://www.R-project.org> (Acesso em 26 de Novembro de 2012).

**RESOLUÇÃO NORMATIVA Nº63/2003** ANS Agência Nacional de Saúde Suplementar Disponível em: [http://www.ans.gov.br/index2.php?option=com\\_legislacao&view=legislacao&task=TextoLei&format=raw&id=748](http://www.ans.gov.br/index2.php?option=com_legislacao&view=legislacao&task=TextoLei&format=raw&id=748). (Acesso em 26 de Novembro de 2012 em 2012).

**RESOLUÇÃO NORMATIVA Nº279/2011** ANS Agência Nacional de Saúde Suplementar Disponível em: [http://www.ans.gov.br/index2.php?option=com\\_legislacao&view=legislacao&task=TextoLei&format=raw&id=1898](http://www.ans.gov.br/index2.php?option=com_legislacao&view=legislacao&task=TextoLei&format=raw&id=1898). (Acesso em 26 de Novembro de 2012).

ROOSEVELT C.M.J. **Estimating Claim Settlement Values Using GLM** Casualty Actuarial Society, 1994 <http://casualtyactuarialsociety.org/pubs/dpp/dpp04/04dpp291.pdf> (Acesso em 2 de Julho de 2013).

**SAS Annotated Output Poisson Regression**. UCLA: Statistical Consulting Group. Disponível em: [http://www.ats.ucla.edu/stat/sas/output/sas\\_poisson\\_output.htm](http://www.ats.ucla.edu/stat/sas/output/sas_poisson_output.htm) (Acesso em 26 de Novembro de 2012).



## PRICING A SELF-MANAGED HEALTH PLAN APPLYING GENERALIZED LINEAR MODELS

### Abstract

*This paper applied Generalized Linear Models to make the pricing process of a self-managed Health Plan. It investigated that medical costs have a probability distribution of the exponential family. Accepting this assumption, it could be applied to this variable a Generalized Linear Model to obtain new self-managed health plan's tariffs, different from what has been generally practiced in the market of these plans. This new methodology is characterized by capturing qualitative elements of exposed participants that old pricing methods were not able to capture. The results showed that, among the variables chosen for the model, the income variable has an inverse relationship with the risk of healthcare expenses. Individuals with higher earnings tend to utilize the health plan less than usual. Male gender individuals have higher expenditures than female ones and it is reflected in the tariff statistically determined. Finally, the model was able to generate tariff tables that can be charged to plan's participants covering all average risk (expected medical utilization).*

**Key-words:** Actuary, Health, Pricing.