Paper 792-2017

Pricing a Self-Funded Health Plan by Applying Generalized Linear Models Using SAS® Enterprise Guide®

Luiz Carlos da Silva Leão, Universidade Federal Fluminense (UFF)

ABSTRACT

This paper explores the utilization of medical services, which has a characteristic exponential distribution. Because of this characteristic, a variable generalized linear model can be applied to it to obtain self-managed health plan rates. This approach is different from what is generally used to set the rates of health plans. This new methodology is characterized by capturing qualitative elements of exposed participants that old ratemaking methods are not able to capture. Moreover, this paper also uses generalized linear models to estimate the number of days that individuals remain hospitalized. The method is expanded in a project in SAS® Enterprise Guide® in which the utilization of medical services by the base during the years 2012, 2013, 2014, and 2015 (the last year of the base) is compared with the Hospital Cost Index of Variation. The results show that, among the variables chosen for the model, the income variable has an inverse relationship with the risk of health care expenses. Individuals with higher earnings tend to use fewer services offered by the health plan. Male individuals have a higher expenditure than female individuals, and this is reflected in the rate statistically determined. Finally, the model is able to generate tables with rates that can be charged to plan participants for health plans that cover all average risks.

Keywords: Actuary - GLM - Regression - Risk - Health - Ratemaking

INTRODUCTION

In Brazilian Health Plans market, self-managed operators are those in which the company itself or other organization establishes and manages with non-profit purpose the health care program for its beneficiaries (RN 279 ANS).

These operators have a tradition of applying simple methods from the statistical-actuarial point of view to ratemake their health plans. Two of them are a fixed percentage of the participant salary or a co-participation single contribution table.

Given that the traditional medical utilization of a health plan follow an exponential distribution family (Jones, 2010), which was confirmed with the database used for this article, it was possible to apply the Generalized Linear Models (GLM) to obtain the values of rate that should be charged to participants of a health plan's self-management mode in order to obtain a new type of ratemaking.

This methodology considers the individual profile of each participant (individual risk), similar to what is done today in the open health insurance market.

It was used the results of a medical plan that has 300.069 lives through the period of 2012 to 2014, both including, with outpatient (low) and inpatient (high) costs.

Additionally, it was also modeled the time of staying in hospital of the same plan.

Finally, the model was able to generate tables with rates that can be charged to plan participants covering all average risk.

GENERALIZED LINEAR MODELS

Are functions of the type (Jong and Heller, 2008):

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

That involves a response variable (y), or explanatory variables or covariates (x) and a random sample of n independent observations.

$$2)\, g(\mu) = \eta = X\beta, \quad E(y) = \mu$$

$$3)\, \eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p,\, n > p$$

Having:

1. The random component of the model distributed, according to an exponential distribution of the family.
2. Enters the explanatory variables in the form of a linear structure, constituting the systematic component of the model.
3. A connection is made between and through a monotonic and differentiable link function. This link function (link) describes the relationship between the mean of the response variable and the covariates.

Steps to model in GLM:

1. Chooses the distribution f(y).
2. Picks up the link function $g(\mu)$, which can be simplified by choosing the parameter corresponds to canonical f(y).
3. Picks up the explanatory variables in terms of $g(\mu)$ in which can be modeled.
4. Collect up $y_1, y_2, y_3, y_4, ...y_n$ observations and corresponding $x_1, x_2, x_3, x_4, ...x_n$.
5. Adjusts the model by estimating β, which is usually done by a statistical software.
6. Since β is estimated, it is verified whether the model fit and explanatory variables are important to determine μ.

Note: If $g(\mu) = \theta$, then g is called the canonical corresponding link for $a(\theta)$. The commonly used links are:

Table 1

Link Function	$g(\mu)$	**Canonical Link For**
identity	μ	Normal
log	$\ln \mu$	Poisson
power	μ^{-1}	Gamma
logit	$\ln\left\{\frac{\pi}{1-\pi}\right\}$	Binomial

Table 1 Common Used Links

METHODOLOGY

The database used refers to the utilization of medical information from January 2012 to December 2014 with data from outpatient medical expenses (including this Dental Expenses) and inpatient medical expenses.

First, it was elaborated a project in SAS® Enterprise Guide statistical software, in which the medical use of base was consolidated into a new one; the years 2012, 2013 and 2014 were taken monetarily for the year 2015 (the last year of the base) with the Hospital Costs' Index of Variation:

Chart 1

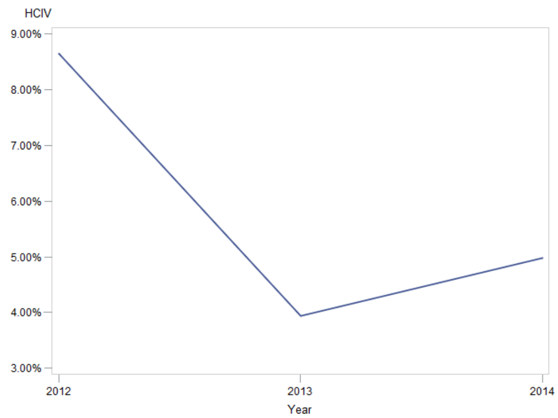


Chart 1 Hospital Costs' Index of Variation observed through 2012 to 2014

Those steps were taken toobtain a largest and most consistent statistical basis, considering the Large Numbers Law, whereby the larger the sample the greater the consistency of statistical analysis (DeGroot & Schervish, 1931).

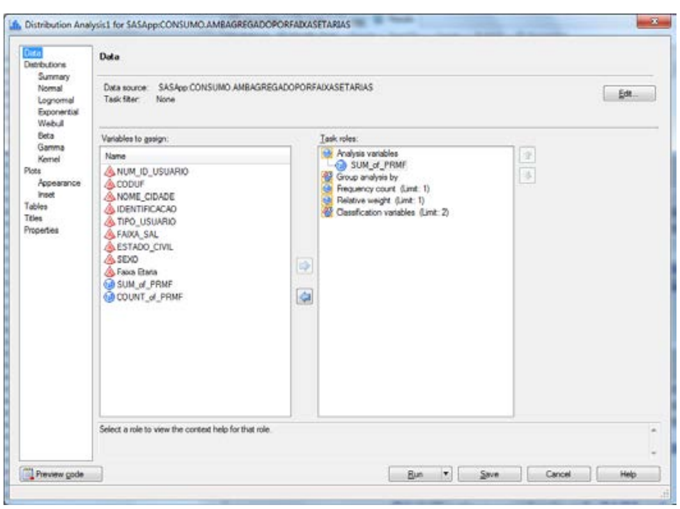
Key contributions to Hospital Costs' Index of Variation are an aging population, the introduction of new technology and services, devices and drugs, and more intense utilization.

The purpose of this article is to find an explanatory model for outpatient and inpatient expenditures in order to make a rate for the plan in question. Therefore, it was necessary to analyze the statistical behavior of these expenses.

It was made an analysis of the behavior of outpatient and inpatient expenses per individual and the sum of total medical spending for each registration because it was analyzed several small expenditures of same interventions.

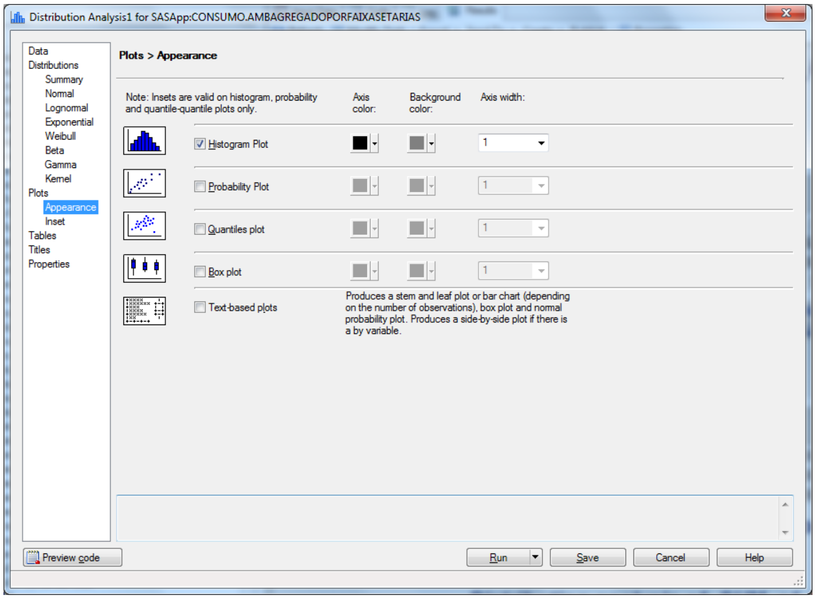
First, some important statistical information:

Display 1



Display 1

Display 2



Display 2

Chart 2

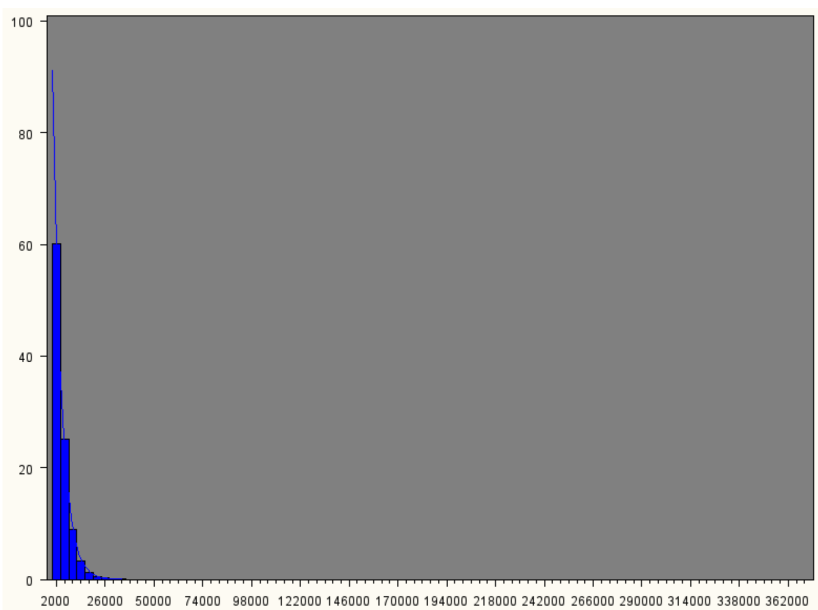
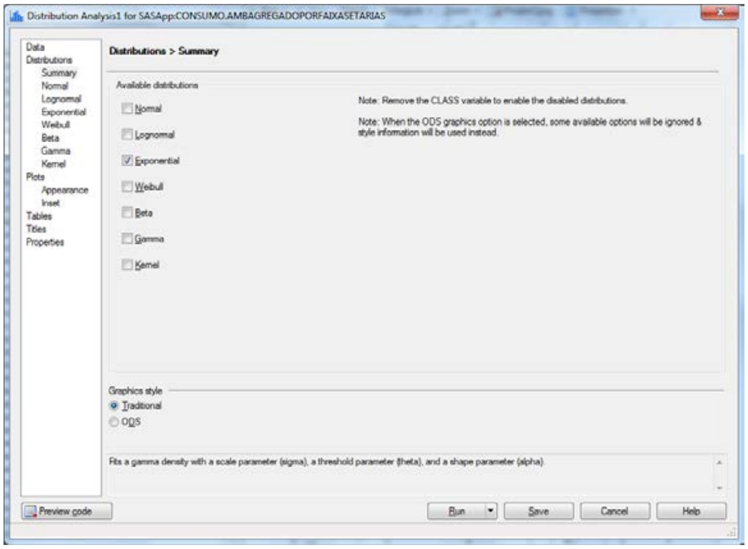


Chart 2 Histogram of Outpatient Expenses

The probability distribution of outpatient expenses follows with a high level of significance a exponential family distribution, as will be proved below.

It can be seen in the following table the results of analysis of the probability distribution obtained in SAS®, adjustment test for Exponential Distribution all p-values of tests Kolgomorov-Smirnov (DeGroot & Schervish, 1931), Cramer-von-Mises (Anderson, 1962) and Anderson-Darling (Anderson, 1954). All of them are under 0.001 percent, in fact, in the non-rejection region of the statistical hypothesis tests.

Display 3



Display 3

There is no statistical evidence to reject the null hypothesis that the distribution of expenses Outpatient follows an exponential distribution. Hypothesis of the Test:

- H0: The distribution of Outpatient expenses follows an exponential distribution.
- H1: The distribution of Outpatient expenses follows another distribution than the exponential distribution.

Table 2

Goodness-of-Fit Tests for Exponential Distribution				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	D	0.020240	Pr > D	<0.001
Cramer-von Mises	W-Sq	45.303557	Pr > W-Sq	<0.001
Anderson-Darling	A-Sq	311.752270	Pr > A-Sq	<0.001

Table 2 Adjustment Test for Outpatient Expenses

Below are the percentiles of the exponential distribution set. Notice the small difference between the observed and estimated distribution with the statistical software:

Table 3

Percentiles for Exponential Distribution		
Percent	**Quantile**	
	Observed	**Estimated**
1.0	58.0280	44.0884
5.0	260.6607	225.0750
10.0	538.8349	462.3383
25.0	1376.9633	1262.4204
50.0	3097.3913	3041.7238
75.0	5861.2291	6083.4631
90.0	9629.5776	10104.4238
95.0	12611.9733	13146.1631

Percentiles for Exponential Distribution		
Percent	Quantile	
	Observed	Estimated
99.0	20735.9411	20208.8630

Table 3 Exponential Distribution Percentiles of Outpatient Expenses

This checking, that the Outpatient expenses follow an exponential distribution, is critical for applying GLM.

Chart 3

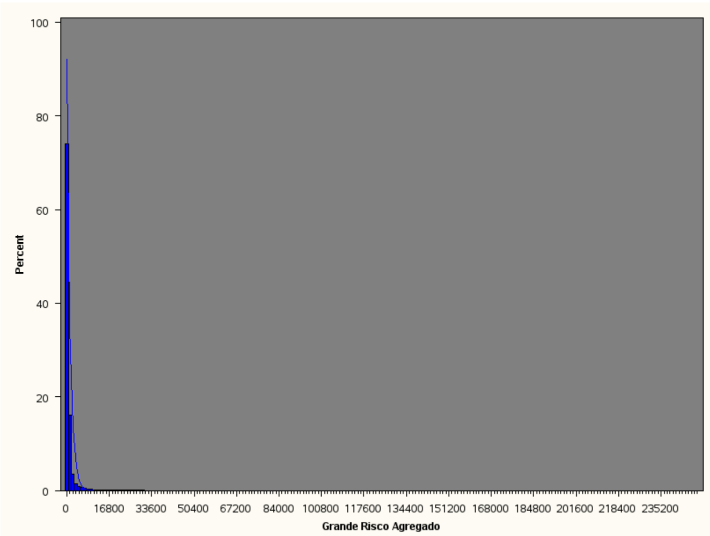
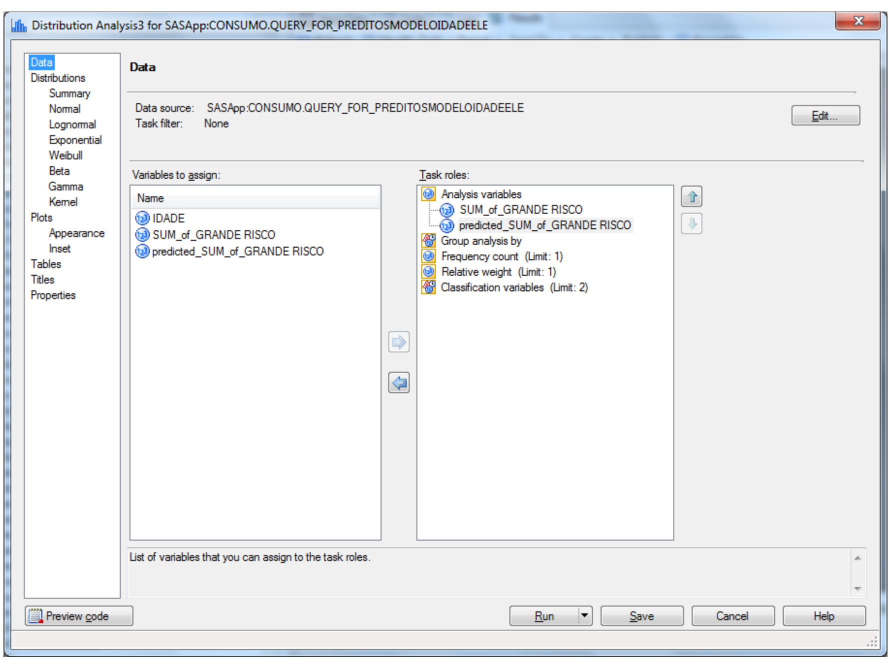


Chart 3 Histogram of Inpatient Expenses

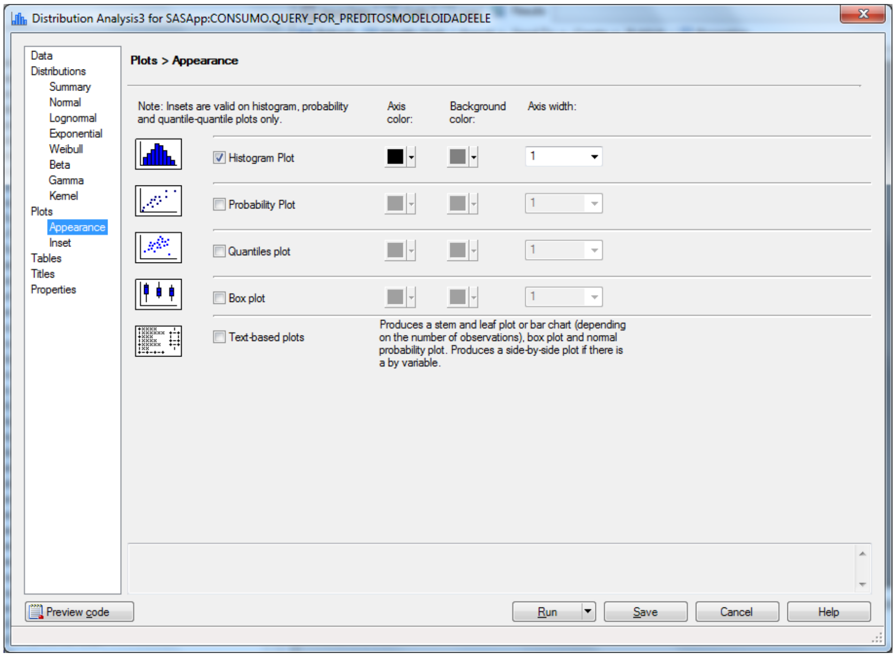
The probability distribution of the costs of inpatient also follows with a high level of significance an exponential family distribution, as proved below.

Display 4



Display 4

Display 5

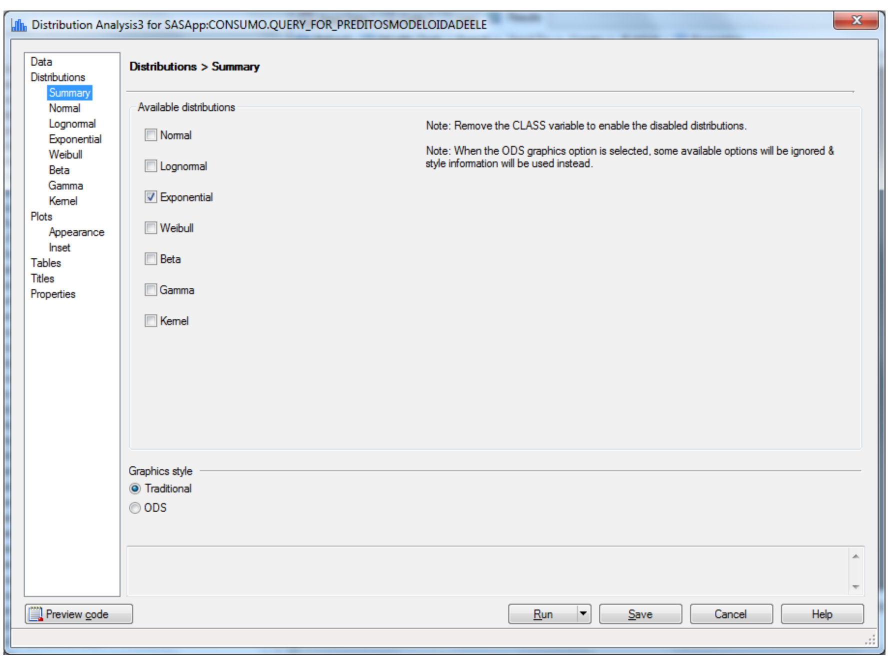


Display 5

It can be seen in the table below the result of analysis of the probability distribution obtained in SAS® Adjustment test for Exponential Distribution, all p-values of tests Kolgomorov-Smirnov (DeGroot & Schervish, 1931), Cramer-von-Mises (Anderson, 1962) and Anderson-Darling (Anderson, 1954). All of them are under 0.001 percent, in fact, in the non-rejection region of the statistical hypothesis tests.

There is no statistical evidence to reject the null hypothesis that the distribution of the inpatient costs follows an exponential distribution.

Display 6



Display 6

Hypothesis of the Test:

- H0: The distribution of inpatient expenses follows an exponential distribution.
- H1: The distribution of inpatient expenses internment follows another distribution than the exponential distribution.

Table 4

Goodness-of-Fit Tests for Exponential Distribution				
Test	**Statistic**		**p Value**	
Kolmogorov-Smirnov	D	0.3243	Pr > D	<0.001
Cramer-von Mises	W-Sq	3838.0489	Pr > W-Sq	<0.001
Anderson-Darling	A-Sq	20401.1260	Pr > A-Sq	<0.001

Table 4 Adjustment Test for Inpatient Expenses

Below are the percentiles of the exponential distribution set:

Table 5

Percentiles for Exponential Distribution		
Percent	**Quantile**	
	Observed	**Estimated**
1.0	151.890	251.376
5.0	449.750	1284.064
10.0	964.210	2637.861
25.0	2599.274	7203.035
50.0	5782.339	17355.531
75.0	16270.568	34711.337
90.0	54477.995	57654.465
95.0	108670.962	75010.271
99.0	340942.897	115309.205

Table 5 Exponential Distribution Percentiles of Outpatient Expenses

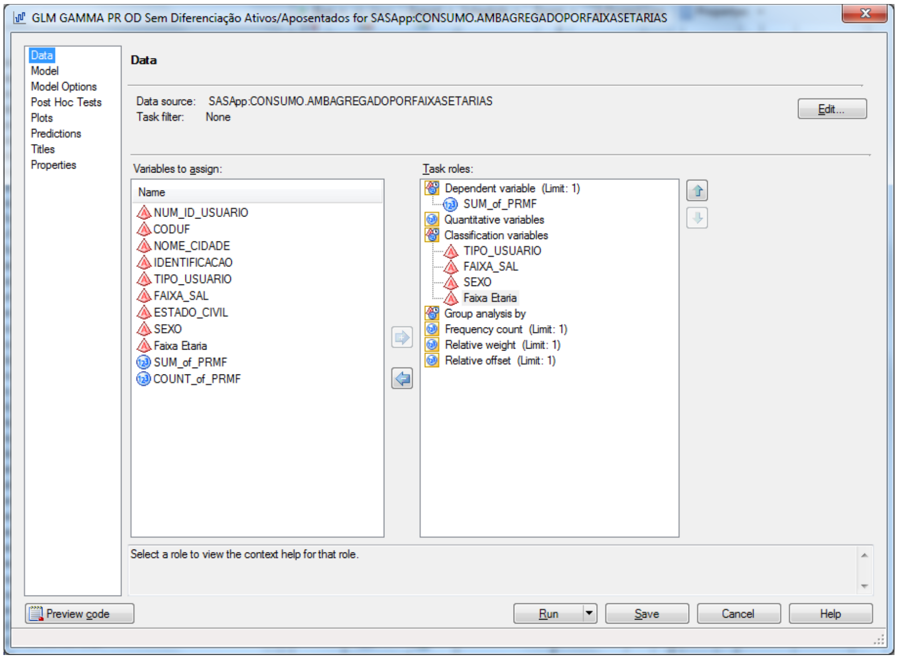
This checking, that the inpatient expenses follow an exponential distribution, will be key for modeling GLM.

Interesting is that the phenomenon described above has been already observed by the world scientific community, and may be cited by (Griswold *et al*, 2004) and by (Jones, 2010).

MODELING

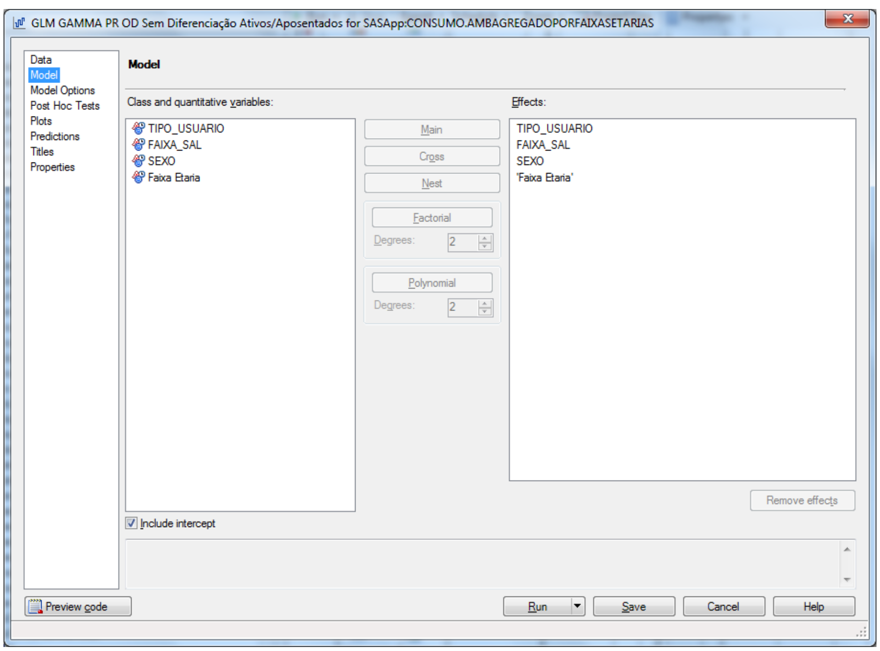
In this article, it is used the GLM for modeling the outpatient e inpatient medical costs. The GLM should be used when the response variable is in the exponential family. This modeling is used to evaluate and quantify the relationship between the response variable (y) and the explanatory variables (x's).

Display 7



Display 7

Display 8



Display 8

Furthermore, it differs from the linear model in two aspects:

- The distribution of the response variable is a chosen exponential family (Poisson, Binomial, Gamma, Normal, Negative Binomial and Normal Inverse).
- The transformation of the mean of the response variable is linearly related to the explanatory variables.

The purpose of the model is to explain the medical expenses for qualitative variables of each participant in the plan to reach the corresponding rate.

The parameters chosen to the model which had better statistical performance were Age Groups, Gender, User Type and Salary Range:

Table 6

Age Groups	0 to 18 years
	19 to 23 years
	24 to 28 years
	29 to 33 years
	34 to 38 years
	39 to 43 years
	44 to 48 years
	49 to 53 years
	54 to 58 years
	Greater than 58 years
Gender	Female
	Male
User Type	01 - Holder
	02 - Spouse or Partner (a)
	03 - Son (a) less than 24 years
	04 - Other (Parents, Grandparents, Aggregates)
Salary Range	1 – Up to R$ 882,14
	2 - Up to R$ 1.628,57
	3 - Up to R$ 3.257,14
	4 - Up to R$ 6.075,07
	5 - Up to R$ 12.150,14
	6 - Up to R$ 12.150,14

Table 6 Parameters chosen for modeling

OUTPATIENT EXPENSES

The Outpatient Expenses modeling results are as follows:

```
PROC GENMOD DATA=WORK.SORTTempTableSorted
       PLOTS(ONLY)=ALL
;
  CLASS TIPO_USUARIO FAIXA_SAL SEXO "Faixa Etaria"n
  ;
  MODEL SUM_of_PRMF=    TIPO_USUARIO FAIXA_SAL SEXO "Faixa Etaria"n
       /
       LINK=LOG
       DIST=GAMMA
```

```
    ;

    OUTPUT
OUT=CONSUMO.PREDITOSMODELOFAIXASETARIAISPR(LABEL="Generalized
Linear Models predictions and statistics for
CONSUMO.AMBAGREGADOPORFAIXASETARIAS")
        PREDICTED=predicted_SUM_of_PRMF ;
RUN; QUIT;
```

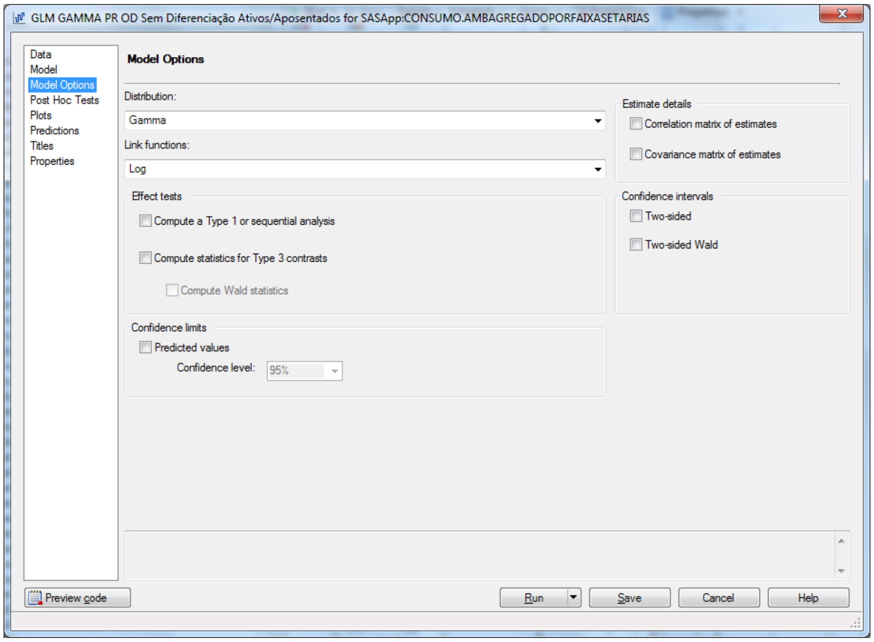
Table 7

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	8.4856	0.0072	8.4715	8.4997	1392133	<.0001
USER_TYPE	01	1	0.2440	0.0065	0.2313	0.2566	1429.42	<.0001
USER_TYPE	02	1	0.2258	0.0068	0.2125	0.2392	1093.71	<.0001
USER_TYPE	03	1	0.3137	0.0216	0.2713	0.3560	210.59	<.0001
USER_TYPE	05	0	0.0000	0.0000	0.0000	0.0000	.	.
SALARY	1	1	-0.4461	0.0276	-0.5002	-0.3919	260.83	<.0001
SALARY	2	1	-0.2770	0.0160	-0.3083	-0.2457	300.75	<.0001
SALARY	3	1	-0.2088	0.0072	-0.2229	-0.1946	838.52	<.0001
SALARY	4	1	-0.1210	0.0050	-0.1308	-0.1111	577.88	<.0001
SALARY	5	1	-0.0463	0.0046	-0.0552	-0.0373	102.95	<.0001
SALARY	6	0	0.0000	0.0000	0.0000	0.0000	.	.
GENDER	F	1	0.2900	0.0041	0.2820	0.2980	5038.06	<.0001
GENDER	M	0	0.0000	0.0000	0.0000	0.0000	.	.
AGE	0-18	1	-0.8932	0.0217	-0.9358	-0.8506	1689.57	<.0001
AGE	19-23	1	-0.9813	0.0213	-1.0230	-0.9395	2121.82	<.0001
AGE	24-28	1	-0.9512	0.0071	-0.9650	-0.9374	18203.3	<.0001
AGE	29-33	1	-0.8433	0.0074	-0.8577	-0.8289	13101.8	<.0001
AGE	34-38	1	-0.6160	0.0091	-0.6339	-0.5981	4537.62	<.0001
AGE	39-43	1	-0.4807	0.0091	-0.4986	-0.4628	2777.53	<.0001
AGE	44-48	1	-0.3305	0.0073	-0.3449	-0.3161	2026.76	<.0001
AGE	49-53	1	-0.2122	0.0069	-0.2258	-0.1987	938.55	<.0001
AGE	54-58	1	-0.1435	0.0067	-0.1567	-0.1304	456.51	<.0001
AGE	> 58	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	1.2442	0.0030	1.2385	1.2501		

Table 7 Results of Modeling Outpatient Expenses

In fact, the best GLM Model for Outpatient expenditure was Gamma Distribution, Link Function: Log by the Deviance Criteria. It can be seen that the p-values of the Wald test are all in non-rejection region, confirming that the variables were statistically well selected. According to (Bruin, 2006), if the model fits the data well, the ratio of the Deviance to DF (Degrees of Freedom), Value/DF, should be about one.

Display 9



Display 9

The table of Goodness of Fit is as follows:

Table 8

Criterion	**DF**	**Value**	**Value/DF**
Deviance	28E4	257084.2791	0.9058
Scaled Deviance	28E4	319877.0484	1.1271
Pearson Chi-Square	28E4	442588.8582	1.5595
Scaled Pearson X2	28E4	550691.0734	1.9404
Log Likelihood		-2636846.852	
Full Log Likelihood		-2636846.852	
AIC (smaller is better)		5273733.7045	
AICC (smaller is better)		5273733.7075	
BIC (smaller is better)		5273944.8268	

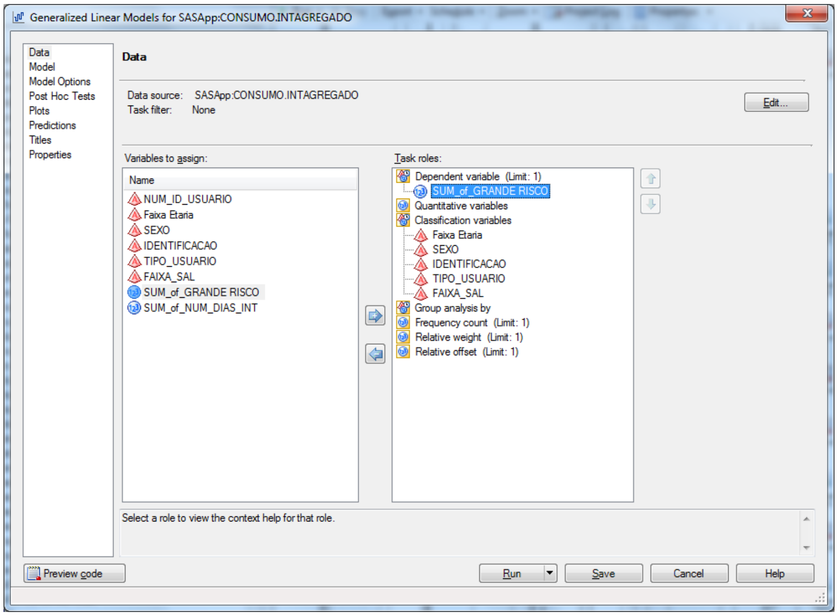
Table 8 Criteria for assessing Goodness of Fit for outpatient expenses

As it can be seen above, the ratio Value/DF was 0.9058, showing, then, that the model for outpatient Eexpenses, by the deviance criteria, was well adjusted.

INPATIENT EXPENSES

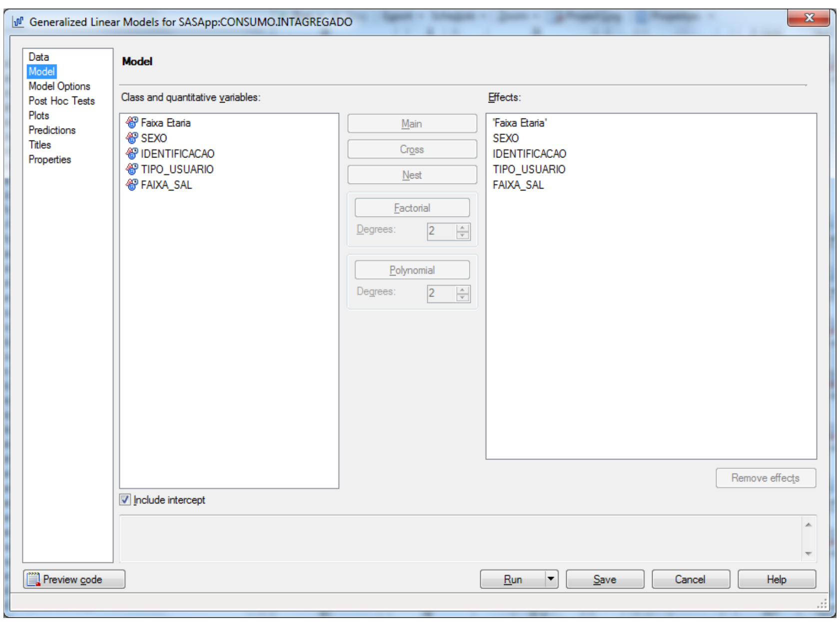
The Inpatient Expenses modeling results are as follows:

Display 10



Display 10

Display 11



Display 11

Table 9

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	10.8728	0.0174	10.8387	10.9069	389793	<.0001

GENDE	F	1	-0.1987	0.0114	-0.2210	-0.1764	306.05	<.0001
GENDER	M	0	0.0000	0.0000	0.0000	0.0000	.	.
AGE	0-18	1	-1.4441	0.0672	-1.5758	-1.3124	461.82	<.0001
AGE	19-23	1	-1.3592	0.0662	-1.4890	-1.2294	421.02	<.0001
AGE	24-28	1	-1.6330	0.0207	-1.6737	-1.5924	6199.58	<.0001
AGE	29-33	1	-1.3552	0.0209	-1.3962	-1.3142	4193.70	<.0001
AGE	34-38	1	-1.2962	0.0249	-1.3449	-1.2474	2713.91	<.0001
AGE	39-43	1	-1.1098	0.0255	-1.1597	-1.0598	1896.71	<.0001
AGE	44-48	1	-0.9316	0.0200	-0.9708	-0.8924	2164.88	<.0001
AGE	49-53	1	-0.7588	0.0182	-0.7945	-0.7230	1731.41	<.0001
AGE	54-58	1	-0.5501	0.0171	-0.5836	-0.5165	1032.93	<.0001
AGE	> 58	0	0.0000	0.0000	0.0000	0.0000	.	.
USER_TYPE	01	1	-0.3167	0.0154	-0.3469	-0.2865	422.75	<.0001
USER_TYPE	02	1	-0.3962	0.0161	-0.4277	-0.3648	609.16	<.0001
USER_TYPE	03	1	-0.3734	0.0664	-0.5035	-0.2433	31.65	<.0001
USER_TYPE	04	0	0.0000	0.0000	0.0000	0.0000	.	.
SALARY	1	1	0.6750	0.0700	0.5378	0.8122	93.04	<.0001
SALARY	2	1	0.3927	0.0366	0.3210	0.4645	115.10	<.0001
SALARY	3	1	0.2364	0.0187	0.1996	0.2731	159.17	<.0001
SALARY	4	1	0.1047	0.0142	0.0769	0.1325	54.60	<.0001
SALARY	5	1	0.0046	0.0133	-0.0213	0.0306	0.12	<.0001
SALARY	6	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		1	0.5529	0.0022	0.5486	0.5573		

Table 9 Results of Modeling Inpatient Expenses

The best GLM Model for Inpatient expenses was Gamma Distribution, Link Function: Log by the Deviance Criteria. It can be seen that the p-values of the Wald test are also in the Non-Rejection Region, confirming that the model is well adapted for the chosen explanatory variables.

Table 10

Criterion	**DF**	**Value**	**Value/DF**
Deviance	87E3	197627.2154	2.2629
Scaled Deviance	87E3	109271.5437	1.2512
Pearson Chi-Square	87E3	749177.1876	8.5784
Scaled Pearson X2	87E3	414233.1694	4.7431
Log Likelihood		-939838.7406	
Full Log Likelihood		-939838.7406	
AIC (smaller is better)		1879717.4811	

AICC (smaller is better)		1879717.4908	
BIC (smaller is better)		1879905.0352	

Table 10 Criteria for assessing Goodness of Fit inpatient expenses

As it can be seen above, the ratio Value/DF was 2.2629, showing that, then, the model for Inpatient Expenses was, by the deviance criteria, well adjusted (value/DF most approximate to one, when compared to other models).

In fact, it is clear the matching between the histograms of observed inpatient expenses and the estimated inpatient expenses:

Chart 4

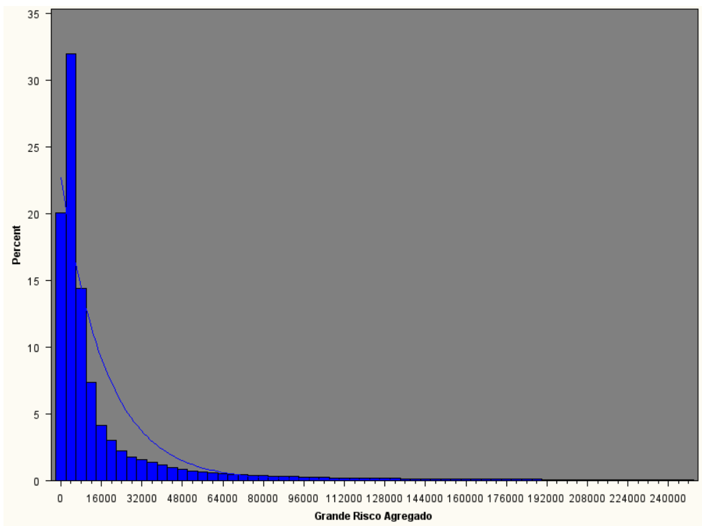


Chart 4 Histogram of inpatient expenditures with observed data

Chart 5

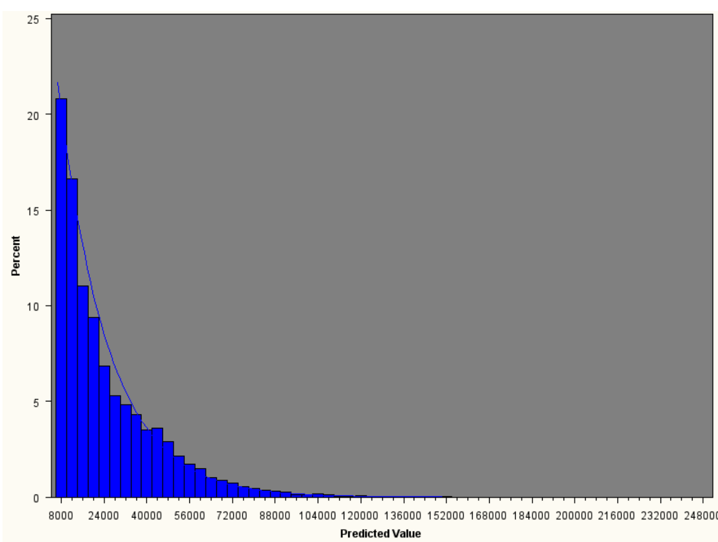
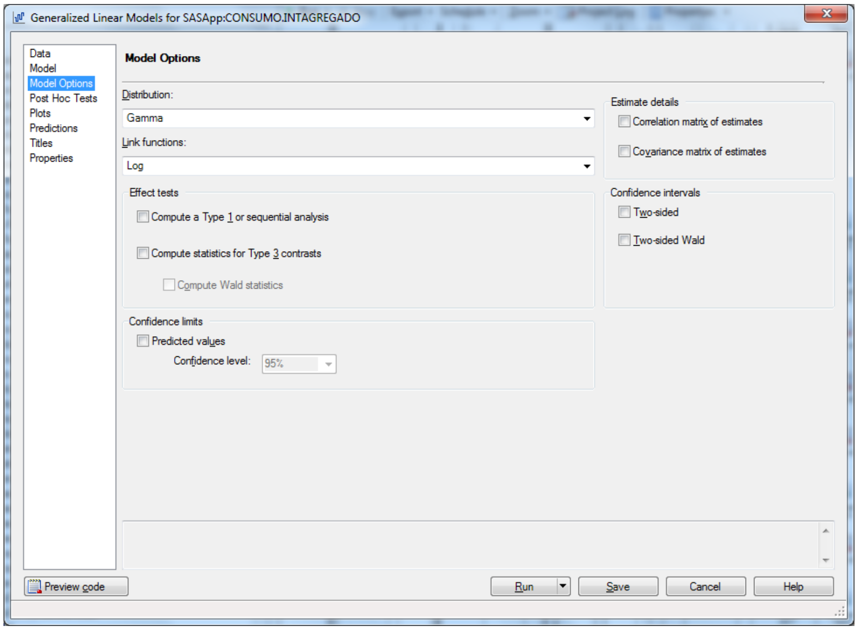


Chart 5 Histogram of inpatient expenditures with estimated data by GLM

Display 12



Display 12

Then, with the values of the parameters determined in the modeling above, whose coverage includes the outpatient and the inpatient expenses, finally, we can obtain the rating tables from the fundamental formula of GLM, adding the corresponding amounts determined in both modeling by the corresponding GLM formulas:

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

2. $\log(\mathbb{E}(H_i)) = \eta_i = \alpha + \beta D_i$

3. $\hat{\mu}_i = \exp(\hat{\eta}_i)$

4. Where α , β are the parameters obtained and $\hat{\mu}_i$ is the estimated rate by GLM.

Table 11

	Up to R$ 882,14	**Up to R$ 1.628,57**	**Up to R$ 3.257,14**	**Up to R$ 6.075,07**	**Up to R$ 12.150,14**	**Greater than R$ 12.150,14**
0-18	R$ 237,65	R$ 188,52	R$ 166,77	R$ 152,12	R$ 142,80	R$ 143,83
19-23	R$ 254,95	R$ 200,80	R$ 176,78	R$ 160,40	R$ 149,87	R$ 150,72
24-28	R$ 199,23	R$ 159,05	R$ 141,23	R$ 129,40	R$ 121,97	R$ 123,00
29-33	R$ 258,83	R$ 204,97	R$ 181,13	R$ 165,00	R$ 154,72	R$ 155,78
34-38	R$ 278,98	R$ 222,68	R$ 197,75	R$ 181,15	R$ 170,73	R$ 172,17
39-43	R$ 334,45	R$ 266,27	R$ 236,08	R$ 215,90	R$ 203,17	R$ 204,78
44-48	R$ 398,60	R$ 316,93	R$ 280,78	R$ 256,52	R$ 241,18	R$ 243,02

49-53	R$ 471,38	R$ 373,88	R$ 330,70	R$ 301,58	R$ 283,10	R$ 285,12
54-58	R$ 573,88	R$ 452,48	R$ 398,70	R$ 362,13	R$ 338,52	R$ 340,52
> = 59	R$ 968,23	R$ 752,87	R$ 657,43	R$ 590,77	R$ 547,17	R$ 548,75

Table 11 Table Rate Single Gender Male Modeling GLM User Type: Holder

Indeed, it is observed that individuals with lower salary must contribute with higher rates to the plan. This occurs because the model captured that individuals with lower income tend to utilize more plan than individuals with higher income, so their rates are more expensive (similar to cars' insurance, young people generally have an insurance premium more expensive than older individuals).

The salary range of "Up to R$ 882,14", according to the model, represents a so high medical risk that its rate was higher than the salary of its individual band, being necessary to rearrange the rates according to the orientation of the standard RN 63/2003 Brazilian National Health Plans Agency (ANS).

COMPLEMENTING STUDY: TIME FOR STAYING IN HOSPITAL

The Study of Time for Staying in Hospital (number of days hospitalized) is another important indicator for health plans, especially in regard to the provision of the plan. Thus, given that the available database has this variable (number of days hospitalized) by each participant, GLM was applied and the results are as follows.

First, the results obtained were satisfactory.

Indeed, it was observed that the number of days hospitalized also follows an exponential distribution.

```
PROC GENMOD DATA=WORK.SORTTempTableSorted
       PLOTS(ONLY)=ALL
;
  CLASS "Faixa Etaria"n SEXO IDENTIFICACAO TIPO_USUARIO FAIXA_SAL
  ;
  MODEL SUM_of_NUM_DIAS_INT= "Faixa Etaria"n SEXO IDENTIFICACAO
TIPO_USUARIO FAIXA_SAL
       /
       LINK=LOG
       DIST=GAMMA
  ;
RUN; QUIT;
```

The results of the GLM with Poisson distribution and Log Link Function, both chosen by the Deviance criteria, are:

Table 12

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	3.7661	0.0062	3.7540	3.7782	374019	<.0001
GENDER	F	1	-0.4352	0.0027	-0.4405	-0.4300	26271.0	<.0001
GENDER	M	0	0.0000	0.0000	0.0000	0.0000	.	.
AGE	0-18	1	-0.6739	0.0242	-0.7213	-0.6264	775.39	<.0001
AGE	19-23	1	-1.0952	0.0235	-1.1413	-1.0491	2172.37	<.0001

AGE	24-28	1	-1.5191	0.0063	-1.5315	-1.5067	57475.4	<.0001
AGE	29-33	1	-1.0329	0.0064	-1.0455	-1.0204	26039.3	<.0001
AGE	34-38	1	-0.8734	0.0089	-0.8909	-0.8559	9604.56	<.0001
AGE	39-43	1	-0.6530	0.0081	-0.6690	-0.6371	6460.37	<.0001
AGE	44-48	1	-0.4048	0.0055	-0.4155	-0.3940	5455.01	<.0001
AGE	49-53	1	-0.4682	0.0049	-0.4778	-0.4586	9163.67	<.0001
AGE	54-58	1	-0.5442	0.0042	-0.5524	-0.5360	16876.2	<.0001
AGE	> 58	0	0.0000	0.0000	0.0000	0.0000	.	.
USER_TYPE	01	1	-0.8966	0.0029	-0.9023	-0.8909	95338.5	<.0001
USER_TYPE	02	1	-0.9846	0.0034	-0.9912	-0.9779	84353.2	<.0001
USER_TYPE	03	1	-0.8025	0.0239	-0.8493	-0.7558	1132.00	<.0001
USER_TYPE	04	0	0.0000	0.0000	0.0000	0.0000	.	.
SALARY	1	1	1.2270	0.0094	1.2086	1.2454	17032.6	<.0001
SALARY	2	1	0.9414	0.0061	0.9293	0.9534	23532.6	<.0001
SALARY	3	1	0.6104	0.0042	0.6023	0.6186	21421.1	<.0001
SALARY	4	1	0.2801	0.0037	0.2729	0.2873	5815.68	<.0001
SALARY	5	1	0.0885	0.0034	0.0817	0.0952	663.66	<.0001
SALARY	6	0	0.0000	0.0000	0.0000	0.0000	.	.
IDENTIFICATION	00	1	-0.4839	0.0047	-0.4931	-0.4748	10743.9	<.0001
IDENTIFICATION	01	1	-0.1139	0.0039	-0.1215	-0.1062	857.79	<.0001
IDENTIFICATION	02	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

Table 12 Results of Modeling of Time in Hospital

Table 13

Criterion	**DF**	**Value**	**Value/DF**
Deviance	28E4	256861.9993	0.9051
Scaled Deviance	28E4	319851.9733	1.1270
Pearson Chi-Square	28E4	436836.9988	1.5392
Scaled Pearson X2	28E4	543962.0359	1.9167
Log Likelihood		-2636708.512	
Full Log Likelihood		-2636708.512	
AIC (smaller is better)		5273461.0239	
AICC (smaller is better)		5273461.0275	
BIC (smaller is better)		5273693.2584	

Table 13 Criteria For Assessing Goodness of Fit Time To Stay in Hospital

As can be seen above, the ratio Value/DF was 0.9051, showing that in fact the model for Time for staying in Hospital was well adjusted. The model obtained good results. All p-values of the Wald test resulted in values lower than 0.1%, all of them in non-rejection region.

By applying the model, we can infer the following:

- Gender male individuals tend to stay longer in hospitals.
- Individuals aged "Greater than 58 years" tend to stay longer in hospital.
- Individuals with lower income tend to stay longer in hospital.
- Retirees and survivor's benefits owners tend to stay longer in hospital than active employees, kept constant all other variables.
- There is no such difference in the length of stay in hospital if the person is the holder, spouse or child, kept constant all other variables.

Below, an example of how many days it is expected that an individual, male gender, age group "Greater than 58 years", "Holder", with an income above R$ 12,150.14 and active, remain hospitalized:

Table 14

Modeling Residence Time in Hospital - Testing 1 (one) Individual		
Parameters		**0 (if false) or 1 (if true) for the variables of the participant**
Gender	Female	0
	Male	1
Ages	0 to 18	0
	19 to 23	0
	24 to 28	0
	29 to 33	0
	34 to 38	0
	39 to 43	0
	44 to 48	0
	49 to 53	0
	54 to 58	0
	> 58	1
User Type	01 - holder	1
	02 - spouse or partner	0
	03 - child under 24 years	0
	04 - others	0
Salary Range	1 - Up to R $ 882.14 (1.3 MSB)	0
	2 - Up to R $ 1,628.57 (2.4 MSB)	0
	3 - Up to R $ 3,257.14 (4.8 MSB)	0
	4 - Up to R $ 6,075.07 (9.6 MSB)	0
	5 - Up to R $ 12,150.14 (19.2 MSB)	0

	6 - More than R $ 12,150.14	1
Identification	00 - Active	1
	01 - Retired	0
	02 - Survivor's Benefit	0

Table 14 Test with 1 (one) individual model of Time Remain in Hospital

Table 15

Number of Days Admitted	10.87 days	
	-95%	+95%
Confidence Interval by Wald	10 days	12 days

Table 15 Results of Testing 1 (one) individual model of Time Remain in Hospital

Thus, it is expected that an individual with the characteristics described above remain 10.87 days hospitalized.

With this model it is possible to estimate how many days, from his economic and social characteristics, a person remains hospitalized.

This model can be used to support decision-making methods in management of health plans, hospitals or companies sponsoring health benefits to their employees.

CONCLUSION

Statistical software SAS Enterprise Guide proved to be an excellent tool for modelling GLM, given its simplicity and computational efficiency in the treatment of large databases.

GLM was very well adapted for the health plan database analysed. With a statistical software, it could be possible to estimate a ratemaking model that considers the different characteristics between the plan participants according to the current practices in the insurance market.

The GLM could also estimate a number of days that is expected a person remains hospitalized. This modeling can be used for numerous experiments with various implications.

Unlike the common practice of self-management market in health insurances, which is applying a percentage of participant's salary or collecting values in a fixed table, by this model, each participant with their individual characteristics will have her or his specific rate, considering the degree of risk of each plan (levels of utilization), similar to what is done in the insurance market in general.

From the point of view of actuarial technical provisions, this modeling differs from current practices in self-management health plans, because it tends to collect a different amount of rates.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Luiz Carlos da Silva Leão
Universidade Federal Fluminense (UFF)
luizcarlosleao@id.uff.br