

ANALYSIS OF HEALTHCARE COST ASSISTANCE OF A BRAZILIAN INSURANCE COMPANY

Marcelo Coelho de Sá – (Universidade Federal do Rio Grande do Norte, RN, Brasil) – marcelocs@camed.com.br

José Alfredo Ferreira Costa – (Universidade Federal do Rio Grande do Norte, RN, Brasil) – jafcosta@gmail.com

Mariana Rodrigues Almeida – (Universidade Federal do Rio Grande do Norte, RN, Brasil) – almeidamariana@yahoo.com

Elias Antonio Borges de Abreu – (Fundação Getúlio Vargas, RJ, Brasil) – elias@riskmed.com.br

The relevance of rising healthcare costs is a main topic in complementary health companies in Brazil. In 2011, these expenses consumed more than 80% of the monthly health insurance in Brazil. Considering the administrative costs, it is observed that the companies operating in this market work, on average, at the threshold between profit and loss. This paper presents results after an investigation of the welfare costs of a health plan company in Brazil. It was based on the KDD process and explorative Data Mining. A diversity of results is presented, such as data summarization, providing compact descriptions of the data, revealing common features and intrinsic observations. Among the key findings was observed that a small portion of the population is responsible for the most demanding of resources devoted to health care.

Keywords: Healthcare costs, health insurance, data mining.

ANÁLISE DOS CUSTOS ASSISTENCIAIS DE UMA OPERADORA DE PLANO DE SAÚDE NO BRASIL

A relevância do aumento dos custos assistenciais é um fenômeno que consiste em uma das principais discussões na área da saúde suplementar. Em 2011, estas despesas representaram mais de 80% das mensalidades dos planos de saúde no Brasil. Considerando as despesas administrativas, observa-se que as empresas atuantes neste mercado trabalham, em média, no limiar entre lucro e prejuízo. Este artigo, com base no processo de descoberta de conhecimento (KDD) e mineração de dados, realizou uma investigação dos custos assistenciais de uma operadora de plano de saúde. Os resultados propiciam descrições compactas dos dados, revelando características comuns e intrínsecas das observações. Dentre as principais conclusões observa-se que uma reduzida parcela da população é responsável por demandar a maior parte dos recursos destinados aos cuidados com saúde.

Palavras-chave: Custos assistenciais, plano de saúde, mineração de dados.

1. Introdução

Os planos de saúde no Brasil representam uma importante via de prestação dos serviços de assistência à saúde, sendo observado em março de 2012, aproximadamente, 48 milhões de brasileiros vinculados a algum plano de assistência médica com ou sem odontologia (ANS, BRASIL, 2012).

Na economia brasileira, o mercado de plano de saúde consiste em um importante segmento faturando em 2011, aproximadamente, R\$ 82 bilhões, cerca de 2% do Produto Interno Bruto (PIB) do Brasil (IBGE, BRASIL, 2011; ANS, BRASIL, 2012).

A regulação estatal desse mercado foi introduzida por meio de Leis específicas que dispõem sobre os planos e seguros privados de assistência à saúde e criaram a Agência Nacional de saúde suplementar (ANS) (SILVA, 2003; MALTA *et al.*, 2004; ALBUQUERQUE *et al.*, 2008).

A concepção da regulação está relacionada com a melhoria da eficiência do setor de saúde suplementar, o que deste modo, posiciona o Estado como agente complementar à ação do mercado, assegurando sua viabilidade. Em suma, o novo marco regulatório estabeleceu novas regras, concebendo um arranjo organizacional que ampliou os direitos e garantias dos beneficiários e tornou mais rígidas as normas para atuação das operadoras de planos de saúde (MALTA *et al.*, 2004; COSTA, 2008; NISHIJIMA; POSTALI; FAVA, 2011).

Na visão de Carvalho e Cecílio (2007), a regulamentação do setor foi fruto da articulação dos movimentos dos consumidores de planos de saúde, dos portadores de patologia e dos médicos que se sentiam ameaçados. Os primeiros, em relação aos seus direitos de assistência, os últimos, em relação à sua autonomia profissional, tendo em vista a racionalização crescente das práticas das operadoras, atuando na lógica do mercado.

Dentre as principais discussões na área da saúde suplementar, destaca-se a importância da avaliação dos custos com a assistência à saúde. A relevância do aumento dos custos assistenciais é um fenômeno mundial, num contexto onde a sua evolução possui tendência crescente de participação no PIB das nações (MCLEOD; GROBLER, 2010; AUERBACH; KELLERMANN, 2011; BRADLEY *et al.*, 2011; MACKENBACH; MEERDING; KUNST, 2011; TANGCHAROENSATHIEN *et al.*, 2011).

Os custos assistenciais correspondem aos valores gastos com a assistência a saúde e são classificados, basicamente, em consultas, exames, procedimentos terapêuticos, materiais, medicamentos, gastos com internações e procedimentos odontológicos (ZUCCHI; NERO; MALIK, 2000; ANS, BRASIL, 2012).

Em 2011, conforme a Tabela 1, estas despesas consumiram 81,6% da receita total auferida com as mensalidades dos beneficiários dos planos de saúde no Brasil. Entre os anos de 2004 e 2011, verifica-se que esta representatividade variou entre 79,1% a 82,3%, sendo a média observada no período de 81,3%.

TABELA 1 – Relação entre os custos assistências e receita das operadoras

ANO	2004	2005	2006	2007	2008	2009	2010	2011	MÉDIA
REPRESENTATIVIDADE	81,3%	81,4%	79,7%	80,6%	80,4%	83,0%	81,2%	82,4%	81,3%

Fonte: Adaptado de ANS (BRASIL, 2012)

Considerando que as despesas administrativas em 2011 consumiram 15,4% das receitas auferidas, calcula-se que, no geral, as operadoras de plano de saúde no Brasil obtiveram o baixo índice de lucro, aproximadamente, de 3%, sinalizando que as empresas atuantes neste mercado trabalham, em média, no limiar entre lucro e prejuízo (SÁ *et al.*, 2010; ANS, BRASIL, 2012).

Não obstante a realidade mundial, os custos assistenciais no setor de saúde suplementar estão em escala ascendente. Entre o período de 2007 a 2011, a Tabela 2 demonstra o aumento do gasto médio com consultas e internações, respectivamente 25% e 55%.

Para Leal e Matos (2009), observa-se nos planos de saúde uma trajetória crescente dos custos assistências, em consonância ao comportamento internacional dos gastos em saúde.

TABELA 2 – Evolução do custo médio assistencial por tipo de gasto

ANO	CONSULTAS		INTERNAÇÃO	
	CUSTO	% VAR.	CUSTO	% VAR.
2007	R\$ 36,91	0,0%	R\$ 3.220	0,0%
2008	R\$ 40,30	9,2%	R\$ 3.480	8,1%
2009	R\$ 40,26	-0,1%	R\$ 3.844	10,5%
2010	R\$ 42,54	5,7%	R\$ 4.621	20,2%
2011	R\$ 46,12	8,4%	R\$ 4.979	7,7%

Fonte: Adaptado de ANS (BRASIL, 2012)

A Tabela 3 exibe o custo assistencial per capita, em que se verifica no período entre 2006 e 2011 que o custo médio por beneficiário nas operadoras de assistência médica com ou sem odontologia aumentou, aproximadamente, 60%.

Na perspectiva de Abreu (2004, p. 13), “os custos da assistência médica elevam-se mais e mais a cada dia tornando o seu financiamento cada vez mais insustentável, seja o financiamento do setor público seja do sistema de saúde suplementar.”

TABELA 3 – Evolução do custo per capita

ANO	CUSTO TOTAL	BENEFICIÁRIOS	CUSTO PER CAPITA	% VAR.
2006	R\$ 33.265.334.420	36.990.226	R\$ 899	0,0%
2007	R\$ 41.202.893.366	38.906.880	R\$ 1.059	17,8%
2008	R\$ 47.839.015.388	40.928.208	R\$ 1.169	10,4%
2009	R\$ 53.514.799.485	42.291.509	R\$ 1.265	8,3%
2010	R\$ 59.229.725.227	45.839.262	R\$ 1.292	2,1%
2011	R\$ 67.903.060.139	47.492.330	R\$ 1.430	10,7%

Fonte: Adaptado de ANS (BRASIL, 2012)

Em razão da relevância da discussão dos custos crescentes em saúde, é de suma importância aprofundar as investigações sobre os custos assistenciais em saúde suplementar no Brasil. Diante do contexto apresentado, a presente pesquisa tem como objetivo investigar os custos assistenciais de uma operadora de grande porte situada na Região Nordeste do Brasil, analisando os beneficiários em dois grupos de custeio assistencial - baixos e altos custos assistenciais.

De tal modo, pretende-se aplicar o processo de descoberta de conhecimento na base de dados da operadora, do modo que o conhecimento gerado contribua na identificação de tendências e fatores relevantes, aprimorando a gestão dos custos assistenciais e, por conseguinte, norteados as operadoras atuantes para uma posição mercadológica sustentável e segura.

Este artigo está estruturado em cinco seções, contando com a introdução. A seção dois apresenta uma revisão da literatura sobre análise de grandes bases de dados. Em seguida, na seção três são descritos os aspectos metodológicos da pesquisa. A seção quatro apresenta os resultados obtidos e a discussão da pesquisa. Conclusões são apresentadas na seção cinco.

2. Análise de grandes bases de dados

A constante evolução da computação, em especial às tecnologias de armazenamento e transmissão de dados, tem proporcionado às organizações a capacidade de armazenar grandes volumes de dados operacionais, o que torna lento, subjetivo e custoso a transformação dos dados em informações úteis, por meio dos métodos convencionais, tais como planilhas eletrônicas e gerenciadores de banco de dados (SFERRA; CORRÊA, 2003; MATTOZO, 2007; MACEDO; MATOS, 2010; LIN *et al.*, 2008; BARROS *et al.*, 2011).

Nesse sentido, Gorgônio e Costa (2007, p. 67) acrescentam que:

O processo de análise e visualização de grandes volumes de dados na forma de registros, descritos por vários atributos e armazenados em um banco de dados é uma tarefa não trivial, tanto em função do grande número de registros normalmente existentes nesses bancos de dados, como pela grande quantidade de informações presentes em cada registro.

Diante da complexidade da análise de grandes volumes de dados e da noção de que o conhecimento é uma potencial vantagem para as organizações, torna-se relevante a evolução das tecnologias, o que direcionou pesquisas neste sentido e fez emergir na ciência um campo denominado descoberta de conhecimento em bases de dados, geralmente referenciado na literatura por *Knowledge Discovery in Data Bases* (KDD) (OLIVEIRA, 2000; MELANDA, 2004).

2.1 Descoberta de conhecimento em bases de dados

A definição de KDD aceita por diversos pesquisadores é a fornecida por Fayyad, Piatetsky-Shapiro e Smyth (1996) que designa a descoberta de conhecimento em bases de dados como um processo interativo e iterativo, não trivial de identificação de padrões válidos, novos, potencialmente úteis, compreensíveis e embutidos nos dados, envolvendo numerosos passos, com muitas decisões sendo feitas pelo usuário (CUNHA, 2010).

Na visão de Romão (2002, p. 42):

O termo não trivial significa que envolve algum mecanismo de busca ou inferência, e não qualquer processamento de dados direto de uma quantidade pré-definida. Nessa definição, um conjunto de dados representa fatos enquanto que os padrões podem ser interpretados como uma expressão em alguma linguagem capaz de descrever um subconjunto de dados ou um modelo aplicável a este subconjunto. Os padrões descobertos devem ser válidos diante de novos dados com algum grau de certeza. Estes padrões podem ser considerados conhecimento dependendo de sua natureza.

O KDD consiste em um processo capaz de extrair conhecimento valioso, previamente desconhecido, contido em grandes volumes de dados armazenados em *data warehouses*, revelando ao usuário relacionamentos inesperados e resumindo os dados em novas formas que são compreensíveis e úteis (CIOS; MOORE, 2002; CHEN *et al.*, 2011).

A Figura 1 ilustra o processo KDD, cujos objetos representam uma visão geral das nove etapas constituintes. O fluxo não é linear, isto é, envolve a iteração entre as etapas constituintes, conforme o resultado observado ao longo do processo. O tempo despendido em todo o processo não é distribuído de forma equitativa para todas as etapas (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; CABENA *et al.*, 1998; MATTOZO, 2007).

Em outras palavras, a sequência entre as etapas não é rigorosa. Dependendo do resultado observado em cada etapa ou em uma etapa particular, há possibilidade de retorno ou avanço, sendo as relações existentes indicadas pelas flechas. Ressalta-se que o processo de mineração

pode continuar após uma solução ter sido encontrada, logo as minerações subsequentes são beneficiadas pelas experiências anteriores.

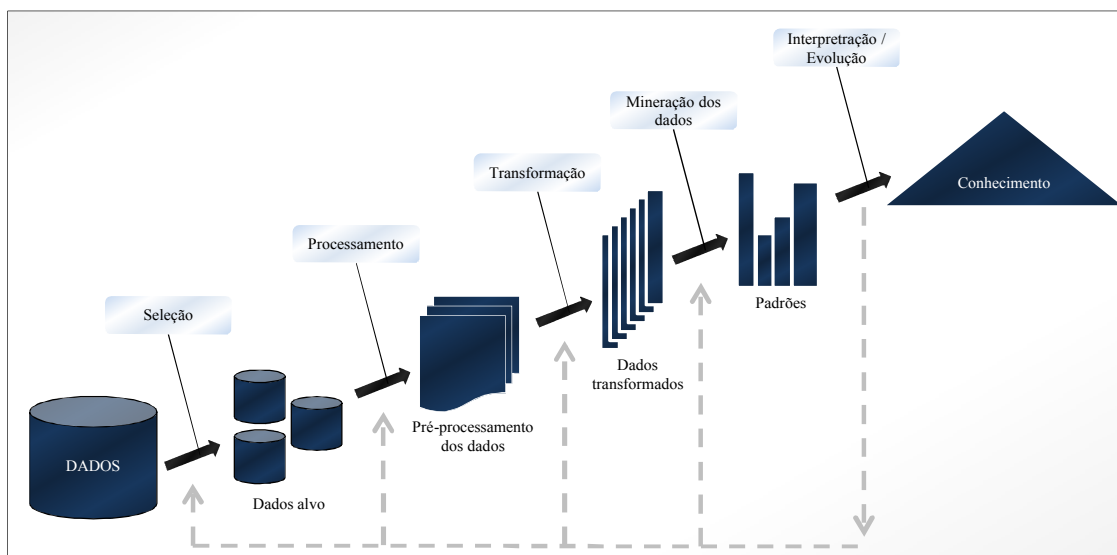


FIGURA 1 – Visão geral do processo KDD

Fonte: Adaptada de Fayyad, Piatetsky-Shapiro e Smyth (1996)

Dependendo da complexidade do trabalho ao qual o pesquisador se propôs, uma determinada etapa pode exigir um maior tempo que outra. A literatura referente ao processo KDD, no entanto, aponta a mineração dos dados, ou *data mining* (DM), como a etapa em que é normalmente despendido o menor tempo, sendo a preparação dos dados a etapa em que se tem o maior custo, cerca de 60% (CABENA *et al.*, 1998; GOEBEL; GRUENWALD, 1999).

É importante destacar que a literatura científica não converge plenamente nas definições de KDD e DM. Há autores designando os termos como sinônimos, fazendo menção que se referem ao mesmo processo (MELANDA, 2004; PADHY; MISHRA; PANIGRAHI, 2012), enquanto outros adotam em suas pesquisas que DM é parte integrante do KDD, consistindo a principal etapa deste abrangente processo (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; FERNANDES NETO; MATTOZO; COSTA, 2008; COLARES, 2011; ELEZI *et al.*, 2011; BRIGNOLI *et al.*, 2012; SVORAY *et al.*, 2012).

Para Galvão e Marin (2009), inicialmente, a expressão Mineração de Dados surge como um sinônimo de KDD, mas é apenas uma das etapas da descoberta de conhecimento em bases de dados no processo global do KDD.

Com crivo, observou-se na literatura uma predominância de pesquisas baseadas na definição de Fayyad, Piatetsky-Shapiro e Smyth (1996), indicando uma vertente científica nesse sentido. No presente trabalho, os termos são usados com significados distintos, em que DM refere-se a uma etapa do processo KDD.

2.2 Etapas do processo de descoberta de conhecimento em bases de dados

O processo KDD desenvolve-se em nove etapas, sendo estas: aprendizagem do domínio, seleção dos dados, pré-processamento dos dados, transformação dos dados, definição das técnicas de DM, seleção da técnica de DM e parâmetros a serem aplicados, mineração dos dados, interpretação dos padrões descobertos e, por fim, consolidação do conhecimento (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

Dessa maneira, o processo KDD tem início com a aprendizagem do domínio da aplicação que consiste na compreensão dos aspectos condicionantes do negócio, avaliação da situação atual,

percepção do conhecimento relevante e, por fim, delineamento dos objetivos pretendidos (SANTOS; AMARAL; PIMENTA, 1999; SFERRA; CORRÊA, 2003).

Para o entendimento satisfatório do domínio, Oliveira (2000, p. 14) cita que:

Alguns aspectos devem ser levados em consideração na tentativa de se definir a esfera de ação de uma aplicação, como: definição dos objetivos e dos requisitos da extração de conhecimento, identificação das possíveis fontes de dados, pesquisa de conhecimento prévio existente, modo de distribuição do conhecimento extraído, incorporação do conhecimento extraído num sistema de tomada de decisão e, por fim, estudo da viabilidade e custos da aplicação do processo.

O processo KDD é interativo e iterativo, envolvendo as etapas, conforme observado na Figura 1. Com um domínio especificado, tem-se início a etapa da localização e definição das fontes de dados e os conjuntos de dados que são apropriados e relacionados ao domínio (OLIVEIRA, 2000; COMBES *et al.*, 2008; MACEDO; MATOS, 2010).

Com objetivo de assegurar a qualidade dos dados selecionados, após a etapa da seleção dos dados, inicia-se a o pré-processamento e limpeza destes. Nesta etapa é verificada a existência de dados incorretos, isto é, inconsistentes, para os quais é necessária a definição de uma estratégia de tratamento, tais como a padronização dos valores dos atributos, remoção de registros duplicados, tratamento e/ou eliminação de ruídos e valores ausentes (OLIVEIRA, 2000; KALAVATHY; SURESH; AKHILA, 2007; COMBES *et al.*, 2008).

A quarta etapa consiste na transformação dos dados que objetiva a procura de configurações apropriadas que melhor represente os dados, proporcionando uma redução na quantidade de registros e dimensões e, ainda, resolvendo eventuais limitações que os métodos de DM possuam (SANTOS; AMARAL; PIMENTA, 1999; KALAVATHY; SURESH; AKHILA, 2007).

Durante a transformação dos dados - quarta etapa - tem-se a busca por configurações apropriadas que melhor represente os dados, proporcionando uma redução na quantidade de registros e dimensões e, ainda, resolvendo eventuais limitações dos métodos de DM (SANTOS; AMARAL; PIMENTA, 1999; KALAVATHY; SURESH; AKHILA, 2007).

Dentre as diversas transformações existentes, encontram-se a normalização mínima-máxima, padronização z-escore, raiz quadrada, logarítmica e arco-seno (BARTLETT, 1947; BRAGA, 2005; LAROSE, 2005; BIANCONI *et al.*, 2008; AL-SHALABI, 2011).

A quinta etapa baseia-se na definição de quais técnicas de DM satisfazem os objetivos pretendidos. Por conseguinte, na sexta etapa, é selecionada uma técnica de DM, entre as elencadas na etapa cinco e são definidos os parâmetros mais apropriados a serem utilizados (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996; SANTOS; AMARAL; PIMENTA, 1999).

Por sua vez, com os dados selecionados, limpos, transformações e definida a técnica de DM e seus parâmetros, inicia-se a sétima e principal etapa - mineração de dados - que será descrita com maiores detalhes no tópico 2.3 (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996; COMBES *et al.*, 2008).

Ressalta-se que não há técnica de mineração de dados universal ou ótima. Na realidade o pesquisador tem a sua disposição uma variedade de técnicas de DM, capazes de produzir diferentes resultados, cujas interpretações podem gerar conhecimentos em diversas áreas. A técnica selecionada deve ser aquela mais apropriada para solução do problema em questão, produzindo um desempenho satisfatório na concepção do conhecimento. Acrescenta-se a possibilidade de se utilizar duas ou mais técnicas de DM para o atingimento do objetivo proposto (ROMÃO, 2002).

Na oitava etapa, após a aplicação da técnica de DM, os padrões descobertos são interpretados e avaliados em relação ao objetivo proposto, sendo identificadas e apresentadas as melhores informações. O propósito do resultado não consiste somente em visualizar, gráfica ou logicamente o rendimento do DM, mas, também, em filtrar a informação a ser apresentada, eliminando padrões redundantes ou irrelevantes que podem surgir no processo (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; SFERRA; CORRÊA, 2003).

Na perspectiva de Oliveira (2000, p. 25):

Caso o procedimento não seja validado, então, provavelmente, deve-se retornar às etapas anteriores e tentar refazer-las ou melhorá-las. Esta iteração pode ocorrer até que se obtenham resultados aceitáveis ou concluir-se que não seja possível extrair conhecimento relevante dos dados.

A nona e última etapa do processo KDD é a consolidação do conhecimento. Nesta etapa, o conhecimento obtido deve agregar valor ao sistema de apoio a tomada de decisão ou, simplesmente, documentado e relatado às partes interessadas, sendo comparado com conhecimentos anteriores (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; OLIVEIRA, 2000; MELANDA, 2004).

A exposição da metodologia remete a conclusão de que o KDD condiciona o processo da descoberta de conhecimento de um modo estruturado, organizado e documentado, garantindo, por conseguinte, maior celeridade, segurança e confiabilidade nos resultados obtidos.

2.3 Mineração de dados

A mineração de dados é o núcleo do processo KDD. Nesta etapa é realizada a exploração e análise da base de dados, a fim de que relacionamentos observados revelem padrões e regras úteis para as instituições (DIAS, 2001; KOH; TAN, 2011; YOO *et al.*, 2012).

As técnicas de DM podem ser aplicadas a tarefas de classificação, estimativa, associação, segmentação e sumarização. A tarefa de classificação tem como objetivo elaborar um modelo que possa ser aplicado a dados não classificados, categorizando-os em classes previamente definidas. A tarefa de estimativa consiste na definição de um valor para uma variável contínua, diferenciando-se da classificação, por atribuir um valor real ao invés de um atributo nominal ou categórico (SOARES JUNIOR; QUINTELA, 2005; WASAN; BHATNAGAR; KAUR, 2006; ARORA; BENIWAL, 2012).

A tarefa de associação consiste em determinar o quanto a presença de um conjunto de atributos influencia a existência de outro conjunto distinto, ou seja, identifica as relações existentes entre os eventos em uma determinada ocasião (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; MELANDA, 2004; COLARES, 2011; YOO *et al.*, 2012).

A segmentação visa a partição de um grupo de observações heterogêneas em vários subgrupos - *clusters* - mais homogêneos. Os registros são agrupados ou segregados de acordo com o grau de semelhança, diferenciando-se da classificação por não haver classes predefinidas (DIAS, 2001; WASAN; BHATNAGAR; KAUR, 2006; GORGÔNIO; COSTA, 2007; KALAVATHY; SURESH; AKHILA, 2007; YOO *et al.*, 2012).

Por fim, a tarefa de sumarização, por meio de análises exploratórias e descrições compactas dos subconjuntos dos dados, tem como objetivo a organização e resumo das observações, de modo que seja possível a visualização das principais características estruturais (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; JACKSON, 2002; SOARES JUNIOR; QUINTELA, 2005; KALAVATHY; SURESH; AKHILA, 2007). A Tabela 4 sintetiza as tarefas realizadas pelo processo de DM.

TABELA 4 – Tarefas realizadas por técnicas de mineração de dados

TAREFA	DESCRIÇÃO	EXEMPLOS
Classificação	Constrói um modelo de algum tipo que possa ser aplicado a dados não classificados a fim de categorizá-los em classes	• Classificar pedidos de crédito • Esclarecer pedidos de seguros fraudulentos • Identificar a melhor forma de tratamento de um paciente
Estimativa	Usada para definir um valor para alguma variável contínua desconhecida	• Estimar o número de filhos ou a renda total de uma • Estimar o valor em tempo de vida de um cliente • Estimar a probabilidade de que um paciente morrerá • Prever a demanda de um consumidor para um novo
Associação	Usada para determinar quais itens tendem a co-	• Determinar quais os produtos costumam ser
Segmentação	Processo de partição de uma população heterogênea em vários subgrupos ou grupos mais homogêneos	• Agrupar clientes por região do país • Agrupar clientes com comportamento de compra • Agrupar seções de usuários Web para prever comportamento futuro de usuário
Sumarização	Envolve métodos para encontrar uma descrição compacta para um subconjunto de dados	• Tabular o significado e desvios padrão para todos os itens de dados • Derivar regras de síntese

Fonte: Dias (2001, p. 11)

3. Metodologia

A metodologia KDD é empregada para o objetivo proposto, conduzindo o processo de descoberta de conhecimento, sendo a etapa da mineração dos dados baseada na técnica da análise exploratória dos dados – sumarização. A sumarização dos dados propicia, através das análises quantitativas, descrições compactas dos dados, revelando características comuns e intrínsecas das observações.

A natureza deste trabalho é quantitativa exploratória com características de um estudo de caso. De tal modo, a análise exploratória dos dados compreendeu nesta pesquisa a tabulação dos dados, a construção de tabelas de frequências, o cálculo dos totais, médias, proporções, teste de Mann-Witney para diferenças entre as médias, variâncias, desvios-padrões, o coeficiente de assimetria, teste de normalidade Kolmogorov-Smirnov (KS) e, por fim, o coeficiente de correlação de Sperman.

A elaboração da tabela de frequência baseou-se no método de Sturges para definição da quantidade de classes, sendo consideradas amplitudes iguais e diferentes para os intervalos (TOLEDO; OVALLE, 1985; PALMUTI, 2012). O teste de Mann-Witney foi utilizado para a comparação das médias dos grupos, em virtude dos dados em estudo não possuírem distribuição aproximadamente normal.

A medida de assimetria empregada para aferição da distorção foi baseada no terceiro momento da distribuição (OLIVEIRA *et al.*, 1997; SILVA *et al.*, 2004). Por fim, o coeficiente de correlação de Spearman foi utilizado por não requer a suposição da relação linear entre as variáveis, ao contrário do coeficiente de Pearson (LIRA, 2004; FIGUEIREDO FILHO; SILVA JÚNIOR, 2009).

Foram utilizadas observações contidas em um *data warehouse* Oracle de uma operadora de plano de saúde de grande porte situada na região Nordeste do Brasil. Os dados compreendem um período entre os anos de 2008 e 2011, referindo-se às características pessoais, características do plano de saúde e utilização médica de, aproximadamente, 100.000 beneficiários.

Com a fonte de dados localizada, procedeu-se com a extração dos dados alvo para a pesquisa, por meio da linguagem *Structed Query Language* (SQL), eliminando-se nesta etapa as observações (beneficiários) que contenham atributos com valores omissos, duplicados,

distorcidos e redundantes que não representam a realidade. Eliminaram-se, também, os beneficiários com planos de saúde com cobertura assistencial reduzida, em razão das distorções causadas nos resultados.

A amostra selecionada final foi de 29.796 beneficiários. Com base no método descrito em Marotti *et al.* (2008, p. 192), a quantidade de observações a ser utilizada nesta pesquisa, capaz de fornecer um nível de significância de 0,01 seria de, aproximadamente, 1.800 beneficiários. Entretanto, dada a enorme capacidade de processamento dos recursos computacionais disponíveis e da amostra resultante ser tratada, esta significativa quantidade de observações, isto é, 29.796 beneficiários, foi considerada em sua totalidade no trabalho, tornando a amostra alvo de estudo extremamente representativa da população, aproximando a pesquisa a um censo e produzindo, por conseguinte, resultados precisos e confiáveis, com um grau de erro mínimo (FREITAS *et al.*, 2000; MAROTTI *et al.*, 2008).

Os atributos disponíveis permitiram a elaboração de uma base de dados contendo sete variáveis, conforme a Tabela 5. As variáveis utilizadas são quantidade de consultas eletiva, quantidade de consultas de urgência, custo com exames, custo com terapias, custo com materiais e medicamentos, custo assistencial hospitalar e custo assistencial total.

De tal modo, as variáveis um a cinco referem-se, exclusivamente, a eventos ocorridos em consultórios médicos, consultórios de urgência em hospitais, clínicas médicas e laboratórios, compreendendo estes os gastos assistenciais ambulatoriais. A variável seis corresponde aos custos assistenciais ocorridos em regime de internação hospitalar e a variável sete afere todo o custo despendido com a assistência à saúde.

TABELA 5 – Variáveis utilizadas na pesquisa

VARIÁVEL	DESCRIÇÃO	OBSERVAÇÃO
1	Consulta Eletiva	Quantidade de consultas realizadas em consultórios médicos e clínicas
2	Consulta Urgência	Quantidade de consultas realizadas em pronto-socorros de hospitais
3	Exames	Custo com exames realizados em consultórios, clínicas e laboratórios
4	Terapias	Custo com terapias realizadas em consultórios, clínicas e laboratórios
5	Materiais e Medicamentos	Custo com materiais e medicamentos utilizados em consultórios, clínicas e laboratórios
6	Custo Hospitalar	Custo total realizado em hospitais, exceto em pronto-socorros
7	Custo Assistencial Total	Custo total do beneficiário

Fonte: Elaborado pelo autor

A análise exploratória dos dados foi realizada nas observações originais, isto é, sem transformações para que os resultados observados fossem elucidativos para o objetivo proposto. Na análise da distribuição do custeio assistencial total, no entanto, utilizaram-se as configurações nos dados normalização mínima-máxima, padronização z-escore, raiz quadrada, logarítmica e arco-seno, para transformação das observações originais e posterior comparação.

Nesta pesquisa, a fim de investigar os beneficiários com altos custos assistenciais, categorizaram-se os indivíduos em dois grupos: baixo e alto custo assistencial. Compreende o grupo dos beneficiários com baixos custos aqueles que utilizaram serviços de assistência à saúde, no ano de 2011, em até R\$ 10.000, sendo, por conseguinte, os beneficiários com custos acima deste valor, denominados de alto custo assistencial.

As Tabelas, Gráficos, cálculos estatísticos e modelagens foram realizados com auxílio de dois renomados programas: Microsoft Excel 2010 e *Statistical Package for Social Sciences* 17 (SPSS).

4. Resultados e discussões

Em razão da elevada quantidade de análises, os resultados e discussões são apresentados em subtópicos, objetivando uma melhor exposição e compreensão dos mesmos. O subtópico 4.1 apresenta o resultado da análise da distribuição do custeio assistencial total e a aplicação das técnicas de transformação. O subtópico 4.2 traz o teste de normalidade das variáveis. No subtópico 4.3 é apresentada a análise do custeio assistencial por grupo de custo. Em seguida, o subtópico 4.4 apresenta a evolução do custeio por tipo de despesa e, por fim, o subtópico 4.5 demonstra a correlação das variáveis.

4.1 Distribuição do custo assistencial total

A base de dados elaborada para esta pesquisa possui, entre outras características, o custo assistencial total dos 29.796 beneficiários no ano de 2011. A frequência em cada classe desta variável, descrita na Tabela 6, exibe a elevada assimetria à direita do custeio assistencial (DEMERS, 2004; BRITO, 2005; BERTSIMAS *et al.*, 2008; ANDRADE; MAIA; RODRIGUES, 2010), em que a primeira classe detém 98,84% das observações. Em consonância, verifica-se um coeficiente de assimetria de 18,15 e a rejeição da hipótese da normalidade dos dados, por meio do teste KS a um nível de significância de 0,01.

TABELA 6 – Frequência dos custos assistenciais com amplitudes iguais nas classes

CLASSE	LIMITE		BENEFICIÁRIOS	FREQUÊNCIA BENEFICIÁRIOS	
	INFERIOR	SUPERIOR		SIMPLES	ACUMULADA
1	0	R\$ 24.782	29.451	98,84%	98,84%
2	R\$ 24.783	R\$ 49.564	193	0,65%	99,49%
3	R\$ 49.565	R\$ 74.347	74	0,25%	99,74%
4	R\$ 74.348	R\$ 99.129	30	0,10%	99,84%
5	R\$ 99.130	R\$ 123.912	20	0,07%	99,91%
6	R\$ 123.913	R\$ 148.694	10	0,03%	99,94%
7	R\$ 148.695	R\$ 173.477	4	0,01%	99,95%
8	R\$ 173.478	R\$ 198.259	3	0,01%	99,96%
9	R\$ 198.260	R\$ 223.042	2	0,01%	99,97%
10	R\$ 223.043	R\$ 247.825	3	0,01%	99,98%
11	R\$ 247.826	R\$ 272.607	2	0,01%	99,99%
12	R\$ 272.608	R\$ 297.390	1	0,00%	99,99%
13	R\$ 297.391	R\$ 322.172	0	0,00%	99,99%
14	R\$ 322.173	R\$ 346.955	2	0,01%	100,00%
15	R\$ 346.956	R\$ 371.737	0	0,00%	100,00%
16	R\$ 371.738	R\$ 396.520	1	0,00%	100,00%
TOTAL			29.796	100,0%	

Fonte: Elaborado pelo autor

Diante desta elevada assimetria, a segunda Tabela de frequências elaborada considerou diferentes amplitudes entre as classes. De tal forma, a Tabela 7 revela na primeira classe que 5,72% dos beneficiários não utilizaram nenhum serviço de assistência à saúde, em contraste as maiores classe, em que pequenas quantidades de beneficiários produziram elevados custos assistenciais.

Neste sentido, com base nos custos assistenciais totais, calcularam-se as medidas estatísticas média, mediana e coeficiente de variação, respectivamente, com valores de R\$ 2.117, R\$ 671 e 417%, possibilitando inferir que a maioria absoluta dos beneficiários possui custos

assistenciais inferiores a média, no entanto, há indivíduos com custos muito elevados (KANAMURA; VIANA, 2007).

TABELA 7 – Frequência dos custos assistenciais com amplitudes diferentes nas classes

CLASSE	LIMITE		BENEFICIÁRIOS	FREQUÊNCIA		CUSTO ASSISTENCIAL		
	INFERIOR	SUPERIOR		SIMPLES	ACUMULADA	CLASSE	REPRESENTATIVIDADE CLASSE	ACUMULADA
1	R\$ 0	R\$ 0	1.705	5,72%	5,72%	R\$ 0	0,00%	0,00%
2	R\$ 1	R\$ 50	714	2,40%	8,12%	R\$ 27.744	0,04%	0,04%
3	R\$ 51	R\$ 100	1.449	4,86%	12,98%	R\$ 102.899	0,16%	0,21%
4	R\$ 101	R\$ 250	3.436	11,53%	24,51%	R\$ 592.109	0,94%	1,15%
5	R\$ 251	R\$ 500	4.921	16,52%	41,03%	R\$ 1.814.887	2,88%	4,02%
6	R\$ 501	R\$ 1.000	6.422	21,55%	62,58%	R\$ 4.637.233	7,35%	11,38%
7	R\$ 1.001	R\$ 2.500	6.700	22,49%	85,07%	R\$ 10.475.057	16,61%	27,99%
8	R\$ 2.501	R\$ 5.000	2.516	8,44%	93,51%	R\$ 8.737.921	13,86%	41,85%
9	R\$ 5.001	R\$ 10.000	1.089	3,65%	97,17%	R\$ 7.384.993	11,71%	53,56%
10	R\$ 10.001	R\$ 20.000	418	1,40%	98,57%	R\$ 5.893.293	9,35%	62,91%
11	R\$ 20.001	R\$ 30.000	145	0,49%	99,06%	R\$ 3.549.039	5,63%	68,53%
12	R\$ 30.001	R\$ 50.000	130	0,44%	99,49%	R\$ 5.177.511	8,21%	76,75%
13	R\$ 50.001	R\$ 100.000	103	0,35%	99,84%	R\$ 6.951.083	11,02%	87,77%
14	R\$ 100.001	R\$ 150.000	30	0,10%	99,94%	R\$ 3.542.075	5,62%	93,39%
15	R\$ 150.001	R\$ 250.000	12	0,04%	99,98%	R\$ 2.298.050	3,64%	97,03%
16	R\$ 250.001	R\$ 400.000	6	0,02%	100,00%	R\$ 1.872.039	2,97%	100,00%
TOTAL			29.796		100,0%	R\$ 63.055.933	100,0%	

Fonte: Elaborado pelo autor

A análise da relação da concentração do custeio assistencial nos beneficiários com maiores custos assistenciais, descrita na Tabela 8, permite observar que apenas 1% dos beneficiários é responsável por 32,2% de todo o custeio assistencial da população e 3% gastaram, aproximadamente, 50%.

Dessa maneira, é possível concluir que uma reduzida parcela da população é responsável por demandar a maior parte dos recursos destinados aos cuidados com saúde (ENGEL; KORFF; KATON, 1996; BRITO, 2005). A pesquisa de Dove *et al.* (2003) corrobora, ao apontar que na amostra em estudo, 1% dos indivíduos são responsáveis por 21% de todo o custeio assistencial e que 13% consomem 77% de todos os recursos assistenciais.

TABELA 8 – Concentração do custeio nos beneficiários com maiores custos assistenciais

% BENEFICIÁRIOS	% CUSTO ASSISTENCIAL
1%	32,2%
2%	41,7%
3%	47,2%
5%	54,2%
10%	65,1%
20%	77,2%
30%	84,6%
50%	93,5%
75%	98,8%
100%	100,0%

Fonte: Elaborado pelo autor

Com objetivo de transformar a variável custo assistencial total e assim modificar a estrutura dos dados, elaborou-se a Tabela 9. Verificaram-se as configurações normalização mínima-máxima e padronização z-escore como eficientes na redução da dimensionalidade, no entanto, não modificaram a distribuição da densidade dos dados, não interferindo, por conseguinte, na elevada assimetria.

As demais configurações reduziram a dimensão dos dados e assimetria da distribuição. A transformação logarítmica, no entanto, verificou-se a mais eficiente, pois, em relação à

distribuição original dos dados, a configuração com logaritmo reduziu a escala dos dados ([menor;maior]) de [0;396.505] para [0;5,60]. No tocante a assimetria, esta foi significativamente reduzida de 18,15 para -1,38, contudo, o teste KS na distribuição transformada rejeitou a um nível de significância de 0,01, a hipótese de normalidade dos dados.

TABELA 9 – Transformação da variável custo assistencial total

CLASSE	DADOS ORIGINAIS	TRANSFORMAÇÃO			
		NORMALIZAÇÃO MÍNIMA-MÁXIMA	PADRONIZAÇÃO Z-SCORE	LOGARÍTMICA	ARCO-SENO
1	98,8%	98,8%	98,8%	5,7%	5,7%
2	0,6%	0,6%	0,6%	0,0%	0,0%
3	0,2%	0,2%	0,2%	0,0%	0,0%
4	0,1%	0,1%	0,1%	0,1%	0,0%
5	0,1%	0,1%	0,1%	3,4%	0,9%
6	0,0%	0,0%	0,0%	5,9%	4,5%
7	0,0%	0,0%	0,0%	11,7%	8,9%
8	0,0%	0,0%	0,0%	21,3%	17,6%
9	0,0%	0,0%	0,0%	24,5%	26,0%
10	0,0%	0,0%	0,0%	16,0%	20,9%
11	0,0%	0,0%	0,0%	7,4%	9,8%
12	0,0%	0,0%	0,0%	2,3%	3,4%
13	0,0%	0,0%	0,0%	1,0%	1,2%
14	0,0%	0,0%	0,0%	0,6%	0,7%
15	0,0%	0,0%	0,0%	0,2%	0,2%
16	0,0%	0,0%	0,0%	0,0%	0,0%
MENOR	R\$ 0	R\$ 0,00	-R\$ 0,24	R\$ 0,00	R\$ 0,00
MEDIANA	R\$ 671	R\$ 0,00	-R\$ 0,16	R\$ 2,83	R\$ 7,20
MÉDIA	R\$ 2.117	R\$ 0,01	R\$ 0,00	R\$ 2,70	R\$ 6,87
MAIOR	R\$ 396.505	R\$ 1,00	R\$ 44,72	R\$ 5,60	R\$ 13,58
ASSIMETRIA	18,15	19,19	18,15	-1,38	-1,62

Fonte: Elaborado pelo autor

4.2 Normalidade das variáveis

A base de dados elaborada para esta pesquisa, conforme a Tabela 5, possui sete variáveis com valores aferidos entre os anos de 2008 e 2011. Com o objetivo de verificar a normalidade das variáveis, elaborou-se a Tabela 10 contendo o coeficiente de assimetria e o teste de normalidade KS a um nível de significância de 0,01. De tal modo, ambas estatísticas apontam que todas as variáveis estudadas possuem distribuição assimétrica, ou seja, os dados não seguem uma distribuição normal.

TABELA 10 – Teste de normalidade das variáveis

VARIÁVEL	ANO	COEFICIENTE DE ASSIMETRIA	KOLGOMOROV-SMIRNOV	
			ESTATÍSTICA	SIGNIFICÂNCIA
Consultas eletivas	2008	1,63	0,15	0,00
	2009	1,65	0,15	0,00
	2010	1,72	0,15	0,00
	2011	1,68	0,15	0,00
Consultas urgência	2008	8,82	0,26	0,00
	2009	6,33	0,25	0,00
	2010	14,29	0,27	0,00
	2011	9,23	0,26	0,00
Exames	2008	2,59	0,22	0,00
	2009	2,79	0,22	0,00
	2010	3,00	0,22	0,00
	2011	2,61	0,21	0,00
Terapias	2008	76,93	0,46	0,00
	2009	153,27	0,47	0,00
	2010	55,08	0,44	0,00
	2011	36,11	0,41	0,00
Materiais e medicamentos	2008	58,49	0,48	0,00
	2009	65,81	0,48	0,00
	2010	92,40	0,48	0,00
	2011	59,86	0,48	0,00
Custo assistencial hospitalar	2008	21,75	0,45	0,00
	2009	16,82	0,44	0,00
	2010	28,45	0,45	0,00
	2011	23,84	0,45	0,00
Custo assistencial total	2008	16,85	0,39	0,00
	2009	15,50	0,39	0,00
	2010	21,30	0,41	0,00
	2011	18,15	0,41	0,00

Fonte: Elaborado pelo autor

4.3 Beneficiários por grupo de custeio assistencial

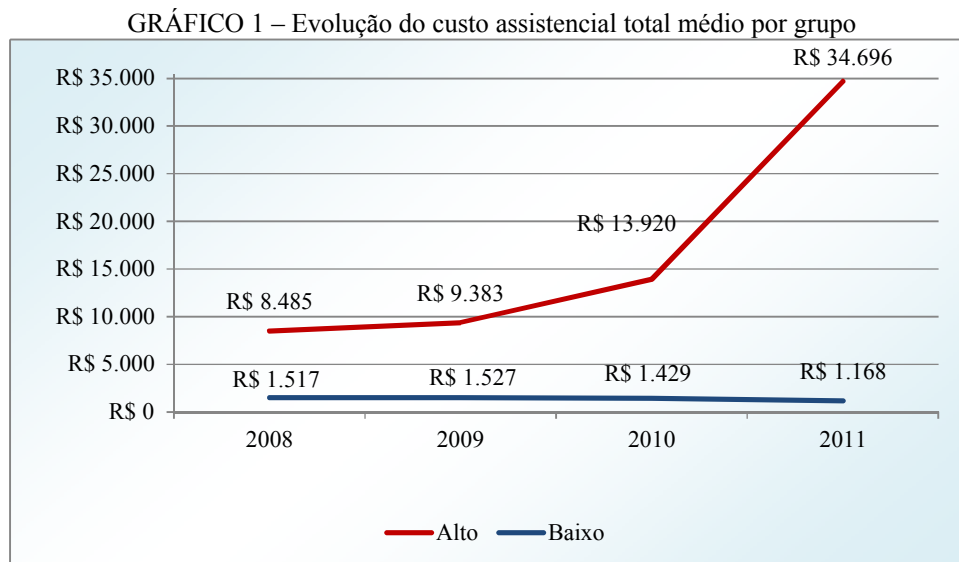
A Tabela 11 exibe o resultado da classificação dos beneficiários de acordo com o custo assistencial ocorrido em 2011, sendo R\$ 10.000 o limite entre os grupos, conforme os procedimentos metodológicos. O primeiro grupo do custeio assistencial - baixo custo - contém 97,17% dos beneficiários em estudo, sendo estes responsáveis por 53,58% de todo o custeio assistencial. O segundo grupo contém apenas 2,83% dos beneficiários, entretanto, esta pequena parcela é responsável por 46,42% do custeio assistencial.

TABELA 11 – Grupos de beneficiários

GRUPO	LIMITE		PONTO MÉDIO	BENEFICIÁRIOS	FREQUÊNCIA BENEFICIÁRIOS		CUSTO ASSISTENCIAL		
	INFERIOR	SUPERIOR			SIMPLES	ACUMULADA	CLASSE	REPRESENTATIVIDADE	
Baixo	R\$ 0	R\$ 10.000	R\$ 1.168	28.952	97,17%	97,17%	R\$ 33.802.453	53,58%	53,58%
Alto	R\$ 10.001	R\$ 400.000	R\$ 34.696	844	2,83%	100,00%	R\$ 29.283.090	46,42%	100,00%
TOTAL			R\$ 2.117	29.796		100,0%	R\$ 63.085.543	100,0%	200,0%

Fonte: Elaborado pelo autor

O Gráfico 1 exibe a evolução do custeio assistencial médio dos grupos, entre os anos de 2008 a 2011. Constatase que os indivíduos considerados como alto custo, desde 2008, possuem tendência crescente do custeio assistencial, distanciando-se dos beneficiários com baixos custos. Nesse sentido, o teste não paramétrico de Mann-Whitney rejeita, a um nível de significância de 0,01, a hipótese da igualdade das médias entre os grupos, em todos os anos analisados.



Fonte: Elaborado pelo autor

A análise do Gráfico 1 remete a inferência de que os beneficiários com altos custos assistenciais em 2011 agravaram o seu estado de saúde ao longo dos anos anteriores, por conseguinte, demandando maiores cuidados assistenciais, principalmente entre os anos de 2010 e 2011, em que o custo assistencial médio aumentou 150%.

Destarte, conclui-se que há evidências da possibilidade da identificação dos beneficiários que ensejam altos custos assistenciais em ano, com base no comportamento assistencial destes indivíduos nos anos anteriores, sendo que o gerenciamento deste pequeno grupo de beneficiários pode resultar em reduções significativas do custeio assistencial (NAESSENS, 2005; MOTORU; JOHNSON; LIU, 2007; BJARNADÓTTIR, 2008).

4.4 Análise do custeio assistencial por tipo de despesa

A análise dos custos por tipo de despesa fornece uma visualização mais detalhada da evolução do custeio assistencial, permitindo evidências da influencia de cada componente sobre o custeio assistencial total.

A Tabela 12 exibe a quantidade média de consultas eletivas, consultas de urgência, o custo médio com exames, terapias, materiais, medicamentos e gastos hospitalares, segregados por grupo. Observa-se que os beneficiários com altos custos em 2011 demandaram uma quantidade crescente de consultas nos anos anteriores, principalmente entre os anos de 2010 e 2011.

Na mesma perspectiva, os exames, materiais, medicamentos e gastos hospitalares possuem custos crescentes no grupo dos beneficiários com altos custos assistenciais, até se tornarem indivíduos com altos custos assistenciais em 2011.

TABELA 12 – Custeio assistencial por tipo de despesa

VARIÁVEL	GRUPO	ANO			
		2008	2009	2010	2011
Consulta eletiva	Alto	8,60	8,74	8,87	10,42
	Baixo	5,40	5,36	5,08	5,09
	Total	5,49	5,46	5,18	5,24
Consulta urgência	Alto	2,04	2,28	2,50	3,47
	Baixo	1,33	1,37	1,23	1,29
	Total	1,35	1,40	1,27	1,35
Exames	Alto	R\$ 981	R\$ 1.087	R\$ 1.276	R\$ 1.701
	Baixo	R\$ 446	R\$ 463	R\$ 461	R\$ 491
	Total	R\$ 461	R\$ 481	R\$ 484	R\$ 525
Terapias	Alto	R\$ 31	R\$ 24	R\$ 25	R\$ 74
	Baixo	R\$ 13	R\$ 15	R\$ 13	R\$ 16
	Total	R\$ 14	R\$ 16	R\$ 14	R\$ 18
Materiais e medicamentos	Alto	R\$ 497	R\$ 477	R\$ 686	R\$ 1.572
	Baixo	R\$ 54	R\$ 47	R\$ 50	R\$ 28
	Total	R\$ 67	R\$ 59	R\$ 68	R\$ 72
Custo assistencial hospitalar	Alto	R\$ 4.618	R\$ 4.376	R\$ 7.316	R\$ 24.526
	Baixo	R\$ 608	R\$ 607	R\$ 558	R\$ 263
	Total	R\$ 721	R\$ 714	R\$ 750	R\$ 951

Fonte: Elaborado pelo autor

As terapias, no entanto, divergem dos demais tipos de despesa, pois não revelam essa tendência de custeio crescente entre os anos de 2008 e 2010, entretanto, converge em 2011 quando apresenta um significativo aumento de, aproximadamente, 200%.

Diante do até então exposto, com exceção das terapias, há indícios de que as variáveis estudadas possuem força preditiva para o grupo de beneficiários com altos custos assistenciais.

Em uma abordagem estatística, o teste Mann-Whitney rejeita, a um nível de significância de 0,01, a hipótese da igualdade das médias entre os grupos, em todos os anos analisados, para os tipos de despesa estudados, evidenciando que os beneficiários com baixos e altos custos assistenciais possuem distribuições diferentes em cada tipo de despesa assistencial.

Em outras palavras, os beneficiários com baixos e altos custos assistenciais possuem perfis distintos de utilização médica nos anos analisados.

A Tabela 13 traz a concentração dos custos assistenciais por tipo de despesa. Na base em estudo, em 2011, as consultas representaram 19,6% dos custos assistenciais totais. Em consonância, a ANS, Brasil (2012) denota que este tipo de despesa representou para o setor de saúde suplementar no mesmo ano, 17,6% das despesas totais, enquanto Severo (2010), com base em dados de uma operadora do Rio Grande do Sul, aponta que as consultas corresponderam a 16% de todo o custeio assistencial.

De tal forma, constata-se nos trabalhos investigados que os custos com consultas representaram, em média, 17,3% dos custos totais, constituindo um dos principais componentes da despesa assistencial total.

Cumprir pontuar, com efeito do estudo dos trabalhos relacionados na Tabela 13, verificam-se diferenças na forma de agrupar as despesas assistenciais, o que, por conseguinte, produziu limitações analíticas. Dessa forma, salienta-se que as despesas com materiais e medicamentos

em ANS, Brasil (2012) estão alocadas em outros agrupamentos, que a pesquisa de Severo *et al.* (2010) agrupou exames e terapias em um único tipo de despesa (SADT) e o trabalho de Ramos (2001) agrupou todo o custeio assistencial em dois tipos, ambulatorial e hospitalar.

TABELA 13 – Custeio assistencial por tipo de despesa

TIPO DE DESPESA	% CUSTO ASSISTENCIAL			
	Base em estudo - Ano 2011	ANS, Brasil (2012)	Severo (2010)	Ramos (2001)
Consultas	19,6%	17,6%	16,0%	49,5%
Exames	31,2%	21,3%	15,0%	
Terapias	0,8%	4,7%		
Materiais e medicamentos	3,4%	-	3,0%	
Custo assistencial hospitalar	44,9%	41,9%	64,0%	50,5%
TOTAL	100,0%	85,5%	98,0%	100,0%

Fonte: Elaborado pelo autor

Nos trabalhos constantes na Tabela 13 não houve consenso no tocante às despesas com exames e terapias. Quanto aos custos com materiais e medicamentos, estes representaram nesta pesquisa 3,4% do custeio assistencial total, estando em consonância ao observado por Severo *et al.* (2010, p. 10). As despesas hospitalares, por sua vez, apresentaram percentuais próximos na base de dados em estudo, na ANS, Brasil (2012) e em Ramos (2001), revelando que as despesas realizadas em hospitais constitui, aproximadamente, metade de todos os custos assistenciais.

Por fim, conclui-se, no entanto, uma comparação das despesas assistenciais dos trabalhos relacionados na Tabela 13 é incipiente, em decorrência das diferentes origens dos dados e formas de agrupamento das despesas assistenciais.

4.5 Correlação

A Tabela 14 aponta a correlação entre as variáveis estudadas no ano de 2011, em que os índices apurados são todos estatisticamente significantes a um nível de 0,01. As correlações mais fortes são o custeio assistencial total e as consultas eletivas ($r = 0,79$; $p < 0,001$) e o custeio assistencial total com os exames ($r = 0,89$; $p < 0,001$).

TABELA 14 – Correlação entre as variáveis no ano de 2011

VARIÁVEL	CONSULTA ELETIVA	CONSULTA URGÊNCIA	EXAMES	TERAPIAS	MATERIAL E MEDICAMENTO	CUSTO ASSISTENCIAL HOSPITALAR	CUSTO ASSISTENCIAL TOTAL
Consulta eletiva	1,00						
Consulta urgência	0,27	1,00					
Exames	0,71	0,26	1,00				
Terapias	0,38	0,14	0,40	1,00			
Materiais e medicamentos	0,32	0,61	0,40	0,16	1,00		
Custo assistencial hospitalar	0,28	0,20	0,32	0,11	0,20	1,00	
Custo assistencial total	0,79	0,39	0,89	0,40	0,46	0,54	1,00

Fonte: Elaborado pelo autor

A Tabela 15, por sua vez, designa a correlação do custeio assistencial total em 2011 com as demais variáveis, considerando para estas os custos assistenciais ocorridos nos anos anteriores, isto é, 2008 a 2010. Observa-se a moderada correlação entre custo assistencial total em 2011 e as consultas eletivas e os exames ocorridos entre os anos de 2008 e 2010, sendo observada para as demais variáveis uma fraca correlação. Em específico, a correlação do custo assistencial total em 2011 com as consultas eletivas em 2010 é de 0,53. Todos os valores são estatisticamente significantes a um nível de 0,01.

TABELA 15 – Correlação entre as variáveis no ano de 2011

VARIÁVEL	CONSULTA ELETIVA	CONSULTA URGÊNCIA	EXAMES	TERAPIAS	MATERIAL E MEDICAMENTO	CUSTO ASSISTENCIAL HOSPITALAR
2010	0,53	0,23	0,54	0,25	0,30	0,20
2009	0,49	0,19	0,51	0,26	0,25	0,20
2008	0,46	0,16	0,49	0,24	0,23	0,19

Fonte: Elaborado pelo autor

Cabe ressaltar que a correlação, apesar de medir o grau de associação entre duas variáveis, não exprime uma relação de causa-efeito, isto é, não afirma quem varia em função de quem, o que não permite, por conseguinte, inferências mais conclusivas (FIGUEIREDO FILHO; SILVA JÚNIOR, 2009).

Desse modo, apesar dos coeficientes de correlação nortear para os apontamentos realizados é necessário um maior aprofundamento sobre as causas e consequências da relação entre as variáveis. As análises obtidas, no entanto, são satisfatórias para os devido propósito ao exprimir que há indícios da associação entre as variáveis estudadas.

Por conseguinte, em face do indício da relação entre as variáveis e da falta de normalidade das distribuições infere-se, a priori, que não é possível a aplicação de diversas técnicas de análise multivariada de dados, tal o método análise discriminante descrito em Sá *et al.* (2010), em decorrência da quebra de dois importantes pressupostos, a ausência de multicolineariedade e a normalidade multivariada.

Neste trabalho destaca-se a complexidade das pesquisas científicas, cujas análises exploratórias são realizadas com base em dados de custo assistencial. As principais dificuldades enfrentadas consistem na elevada dimensionalidade dos dados, nas diversas fontes de dados que compõem o *data warehouse*, na presença de registros incompletos ou inconsistentes, a extrema dificuldade da caracterização matemática neste tipo de pesquisa e a manutenção do sigilo das informações, de modo que a ética não seja violada (CIOS; MOORE, 2002; WASAN; BHATNAGAR; KAUR, 2006; HARRISON, 2008; KOH; TAN, 2011).

Acrescenta-se a dificuldade na comparação dos resultados obtidos com outros trabalhos semelhantes, em decorrência das diferentes terminologias adotadas pelas operadoras atuantes no mercado e a ANS (CIOS; MOORE, 2002; WASAN; BHATNAGAR; KAUR, 2006; HARRISON, 2008; KOH; TAN, 2011).

5. Conclusões

O presente trabalho delineou, inicialmente, o setor de saúde suplementar expondo o desafio das operadoras de plano de saúde frente ao aumento dos custos assistenciais, destacando a importância da sua investigação. Nesse sentido, apresentou-se conceitualmente o processo de análise de grandes bases de dados, em especial a metodologia *Knowledge Discovery in Data Bases*, cuja sustentação teórica circunscreve o objeto de pesquisa, isto é, a grande base de dados utilizada na pesquisa.

Em meio aos elementos resultantes do referencial teórico, a tarefa de *Data Mining* sumarização dos dados é a definida como ideal para o objetivo desta pesquisa. Esse interesse é motivado principalmente em decorrência desta tarefa propiciar descrições compactas dos dados, revelando características comuns e intrínsecas das observações.

Os resultados deste estudo demonstraram que o custo assistencial total possui uma distribuição de densidade dos dados muito assimétrica à direita, designando a elevada presença de beneficiários com baixos custos assistenciais e uma minoria ao longo da cauda (maiores classes), responsável por grande parte de todo o custo assistencial.

A busca por modificar a estrutura da distribuição do custo assistencial total e tornar os dados simétricos demonstrou que a transformação logarítmica foi a mais eficiente nesse sentido, reduzindo significativamente a assimetria da distribuição de 18,15 para -1,38. Mesmo diante de tal redução, o teste KS rejeitou a normalidade dos dados transformados a um nível de significância de 0,01. A análise da normalidade das demais variáveis, por meio do teste KS a um nível de significância de 0,01, revelou que todas as variáveis estudadas possuem distribuição assimétrica, ou seja, os dados não seguem uma distribuição normal.

O experimento realizado verificou que 5,72% dos beneficiários não utilizaram nenhum serviço de assistência à saúde, em contraste as maiores classe, em que pequenas quantidades de beneficiários produziram elevados custos assistenciais. Constatou-se que apenas 1% dos beneficiários utilizaram 32,2% de todo o custeio assistencial da população e que 3% gastaram, aproximadamente, 50%. Conclui-se, por conseguinte, que uma reduzida parcela da população é responsável por demandar a maior parte dos recursos destinados aos cuidados com saúde.

O estudo comparativo dos beneficiários com baixos e altos custos assistenciais no ano de 2011 remete a conclusão de que os dois grupos possuem perfis de utilização assistencial distintos, não somente no ano de 2011, mas também nos anos anteriores. Os beneficiários com altos custos vinham, desde 2008, em uma tendência crescente do custeio assistencial, culminando em 2011 com custos por indivíduo acima de R\$ 10.000,00.

De tal modo, infere-se que os beneficiários com altos custos assistenciais em 2011, são indivíduos que agravaram o seu estado de saúde ao longo dos anos anteriores, por conseguinte, demandando maiores cuidados assistenciais, principalmente entre os anos de 2010 e 2011.

Recomenda-se as operadoras de planos de saúde a adoção de programas de medicina preventiva para os beneficiários com altos custos assistenciais, em virtude de que o gerenciamento deste pequeno grupo de beneficiários pode resultar em reduções significativas do custeio assistencial.

O detalhamento dos custos assistenciais permitiu verificar no grupo dos beneficiários com altos custos que houve entre os anos de 2008 a 2011 uma utilização crescente por consultas eletivas, consultas de urgência, exames, materiais, medicamentos e gastos hospitalares. As terapias, no entanto, foi o único tipo de despesa assistencial em que não se constatou essa tendência. Acrescenta-se que os custos hospitalares foram as maiores despesas, representando 45% de todo o custo com a assistência a saúde.

Para analisar a relação entre as variáveis utilizou-se o coeficiente de correlação de Spearman. Observou-se que as correlações mais fortes no ano de 2011 são o custeio assistencial total e as consultas eletivas ($r = 0,79$; $p < 0,001$) e o custeio assistencial total com os exames ($r = 0,89$; $p < 0,001$).

A análise da correlação do custo assistencial total em 2011 com as demais variáveis, considerando para estas os custos assistenciais ocorridos nos anos anteriores, apresentou índices de correlação menores, sendo observada uma correlação moderada para as consultas eletivas e os exames ocorridos entre os anos de 2008 e 2010. Em outras palavras, o custo assistencial total dos beneficiários em 2011 está moderadamente correlacionado com os custos com consultas eletivas e exames entre os anos de 2008 e 2010.

Desse modo, conclui-se que há evidências da possibilidade da identificação dos beneficiários que ensejam altos custos assistenciais em ano, com base no comportamento assistencial destes indivíduos nos anos anteriores.

Acrescenta-se que para este objetivo, em face do indício da correlação entre as variáveis e da não normalidade das distribuições dos dados infere-se, a priori, que não é possível a aplicação de diversas técnicas de análise multivariada de dados, tal como a análise discriminante descrita em Sá *et al.* (2010), em decorrência da quebra de dois importantes pressupostos, a ausência de multicolineariedade e a normalidade multivariada.

Análises exploratórias foram realizadas com base em dados de custeio assistencial à saúde de beneficiários. Dentre as principais dificuldades enfrentadas, destaca-se a elevada dimensionalidade dos dados e dificuldade da caracterização matemática neste tipo de pesquisa. A aplicação do processo KDD na base de dados da operadora foi satisfatória para os objetivos propostos, gerando informações úteis para a gestão dos custos assistenciais, identificando padrões, tendências e fatores relevantes. A visualização integrada dos resultados apresentados e discutidos no presente trabalho constitui valioso conhecimento, capaz de auxiliar os gestores do mercado de saúde suplementar durante o processo de tomada de decisão.

Esta pesquisa consistiu em um marco para futuras minerações, em potencial, aplicações de *data mining* que identifiquem os beneficiários que irão incorrer em altos custos assistenciais (predição), possibilitando o tratamento preventivo dos mesmos e, por conseguinte, a redução dos custos assistenciais.

Referências

- ABREU, Elias Antônio Borges de. *Gestão do risco saúde populacional através da avaliação do perfil de risco individual*. 2004. 88 f. Dissertação (Mestrado em Sistema de Gestão) - Universidade Federal Fluminense, Niterói, 2004.
- ALBUQUERQUE, Ceres; PIOVESAN, Márcia Franke; SANTOS, Isabela Soares; MARTINS, Ana Cristina Marques; FONSECA, Artur Lourenço; SASSON, Daniel; SIMÕES, Kelly de Almeida. A situação atual do mercado da saúde suplementar no Brasil e apontamentos para o futuro. *Revista Ciência & Saúde Coletiva*, v. 13, n. 5, p.1421-1430, 2008.
- AL-SHALABI, Luai. Knowledge Discovery Process: Guide Lines for New Researchers. *Journal of Artificial Intelligence*, v. 4, n. 1, p. 21-28, 2011.
- ANDRADE, Mônica Viegas; MAIA, Ana Carolina; RODRIGUES, Cristina Guimarães. Indicadores de gastos com serviços médicos no setor de saúde suplementar no Brasil: o caso Sabesp. In: CONGRESSO DA ASSOCIAÇÃO LATINO-AMERICANA DE POPULAÇÃO, 4, 2010, Havana, Cuba. *Anais...* 2010.
- BRASIL. Agência Nacional de Saúde Suplementar - ANS. *Caderno de Informação de Saúde Suplementar*, Rio de Janeiro, Rio de Janeiro, Junho/2012.
- ARORA, Jitender; BENIWAL, Sunita. Classification and Feature Selection Techniques in Data Mining. *International Journal of Engineering Research & Technology*, v. 1, n. 6, p. 1-7, 2012.
- AUERBACH, David I.; KELLERMANN, Arthur L. A Decade Of Health Care Cost Growth Has Wiped Out Real Income Gains For An Average US Family. *Health Affairs*, v. 30, n. 9, p. 1630-1636, 2010.
- BARROS, Everton Fernando; ROMÃO, Wesley; CONSTANTINO, Ademir Aparecido; SOUZA, Celso Lara de. Pré-processamento para mineração de dados sobre beneficiários de planos de saúde suplementar. *Journal of Health Informatics*, v. 3, n. 1, p. 19-26, 2011.

- BARTLETT, M. S. The Use of Transformations. *Biometrics Journal*. v. 3, n. 1, p. 39-52, 1947.
- BERTSIMAS, Dimitris; BJARNADÓTTIR, Margrét V.; KANE, Michael A.; KRYDER, J. Christian; PANDEY, Rudra; VEMPALA, Santosh; WANG, Grant. Algorithmic Prediction of Health-Care Costs. *Operations Research Journal*, v. 56, n. 6, p. 1382–1392, 2008.
- BIANCONI, André; GOVONE, José Silvio; ZUBEN, Cláudio Jose Von; PIÃO, Antonio Carlos Simões; PIZANO, Marcos Aparecido; ALBERTI, Luís Fernando. Transformação de Dados e Implicações da Utilização do Teste de Kruskal-Wallis em Pesquisas Agroecológicas. Pesticidas: *Revista de Ecotoxicologia e Meio Ambiente*, v. 18, n. 0, p. 27-34, 2008.
- BJARNADÓTTIR, Margrét Vilborg. *Data-driven Approach to Health Care: Applications Using Claims Data*. 2008. 130 f. Tese (Doutorado em Pesquisa Operacional) – Massachusetts Institute of Technology, Massachusetts, 2008.
- BRADLEY, Elizabeth H.; ELKINS, Benjamin R.; HERRIN, Jeph; ELBEL, Brian. Health and social services expenditures: associations with health outcomes. *BMJ Quality & Safety Journal*, v. 20, n. 10, p. 826-831, 2011.
- BRAGA, Luis Paulo Vieira. *Introdução à Mineração de Dados*. 2º ed. Revista e ampliada. Rio de Janeiro: E-Papers Serviços, 2005.
- BRIGNOLII, Juliano Tonizetti; JUNIORII, Egon Sewald; MIGUEZIII, Viviane Brandão; SANTOSIV, Neri dos; SPANHOLV, Fernando. A Intervenção Humana na Qualificação de Processos de Data Mining: Estudo de Caso Em Uma Base de Dados Hipotética. *Revista Eletrônica de Biblioteconomia e Ciência da Informação*, v. 17, n. 1, p. 110-124, 2012.
- BRITO, Rubens José Amaral de. *Demanda Por Serviços de Saúde: O Caso Sabesp*. 2005. 147 f. Dissertação (Mestrado em Economia) - Universidade Federal de Minas Gerais, Belo Horizonte, 2005.
- CABENA, Peter; HADJNIAN, Pablo; STADLER, Rolf; ZANASI, Alessandro. *Discovering Data Mining: From Concept to Implementation*. 1º ed. New Jersey: Prentice Hall, 1998.
- CARVALHO, Eurípedes Balsanuf; CECÍLIO, Luiz Carlos de Oliveira. A regulamentação do setor de saúde suplementar no Brasil: a reconstrução de uma história de disputas. *Revista Cadernos de Saúde Pública*, v. 23, n. 9, p. 2167-2177, 2007.
- CHEN, Yongdang; WANG, Yang; XIAO, Xiao; SHI, Meihong. Knowledge Discovery Technology Based on Access Information Mining on Knowledge Warehouse. In: INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS AND KNOWLEDGE DISCOVERY, 8, v. 4, p. 1285-1288, 2011.
- CIOŚ, Krzysztof J.; MOORE, G. William. Uniqueness of Medical Data Mining. *Artificial Intelligence in Medicine Journal*, v. 26, n. 1, p. 1-24, 2002.
- COLARES, Peterson Fernandes. *Processo de Indução e Ranqueamento de Árvores de Decisão Sobre Modelos Olap*. 2011. 109 f. Dissertação (Mestrado em Ciência da Computação) - Pontifícia Universidade Católica do Rio Grande do Sul, 2005.
- COMBES, C.; MESKENS, N.; RIVAT, C.; VANDAMME, J.-P. Using a KDD process to forecast the duration of surgery. *International Journal of Production Economics*, v. 112, n. 1, p. 279-293, 2008.
- COSTA, Nilson do Rosário. O regime regulatório e o mercado de planos de saúde no Brasil. *Revista Ciência & Saúde Coletiva*, v. 13, n. 5, p. 1453-1462, 2008.

- CUNHA, João Roberto Del Ducca. *Implementação de uma Abordagem de Extração e Otimização de Regras Fuzzy Utilizando Sistemas Imunológicos Artificiais*. 2010. 99 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Itajubá, Itajubá, 2010.
- DEMERS, Marie. The predictability of individual primary care costs and its impact on managed care plans. *Health Policy Journal*, v. 68, n. 3, p. 345-352, 2004.
- DIAS, Maria Madalena. *Um Modelo de Formalização do Processo de Desenvolvimento de Sistemas de Descoberta de Conhecimento em Banco de Dados*. 2001. 197 f. Tese (Doutorado em Engenharia de Produção) - Universidade Federal de Santa Catarina, Florianópolis, 2001.
- DOVE, Henry G.; DUNCAN, Ian; ROBB, Arthur. A prediction model for targeting low-cost, high-risk members of managed care organizations. *The American Journal of Managed Care*, v. 9, n. 5, p. 381-389, 2003.
- ELEZI, Fatos; SHARAFI, Armin; MIRSON, Alexander; WOLF, Petra; KRCMAR, Helmut; LINDEMANN, Udo. A Knowledge Discovery In Databases (KDD) Approach For Extracting Causes Of Iterations In Engineering Change Orders. In: ASME 2011 INTERNATIONAL DESIGN ENGINEERING TECHNICAL CONFERENCES & COMPUTERS AND INFORMATION IN ENGINEERING CONFERENCE, Washington, EUA, 2011.
- ENGEL, Charles C.; KORFF, Michael Von; KATON, Wayne. Back pain in primary care: predictors of high health-care costs. *Journal Of The International Association For The Study Of Pain*, v. 65, n. 2-3, p. 197-204, 1996.
- FAYYAD, Usama; PIATETSKY-SHAPIO, Gregory; SMYTH, Padhraic. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, v. 17, n. 3, p. 37-54, 1996.
- FERNANDES NETO, André Pedro; MATTOZO, Teófilo Camara; COSTA, José Alfredo Ferreira. Uma Aplicação de Mineração de Dados no Gerenciamento do Churn em Serviços de Banda Larga de Telecomunicações. In: CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 4, 2010, Niterói, RJ, Brasil. *Anais...* 2008.
- FIGUEIREDO FILHO, Dalson Britto; SILVA JÚNIOR, José Alexandre da. Desvendando os Mistérios do Coeficiente de Correlação de Pearson. *Revista Política Hoje*, v. 18, n. 1, p. 115-146, 2009.
- FREITAS, Henrique; OLIVEIRA, Mírian; SACCOL, Amarolinda Zanela; MOSCAROLA, Jean. O Método de Pesquisa Survey. *Revista de Administração*, v. 35, n. 3, p. 105-112, 2000.
- GALVÃO, Noemi Dreyer; MARIN, Heimar de Fátima. Técnica de mineração de dados: uma revisão da literatura. *Revista Acta Paulista de Enfermagem*, v. 22, n. 5, p. 686-90, 2009.
- GOEBEL, Michael; GRUENWALD, Le. A Survey Of Data Mining And Knowledge Discovery Software Tools. *SIGKDD Explorations*, v. 1, n. 1, p. 20-33, 1999.
- GORGÔNIO, Flavius L.; COSTA, José Alfredo Ferreira. Análise de Agrupamentos Distribuída Através de Múltiplos Mapas Auto-Organizáveis. In: WORKSHOP EM ALGORITMOS E APLICAÇÕES DE MINERAÇÃO DE DADOS, 3, 2007, João Pessoa, PB. *Anais...* 2007.
- HARRISON, James H. Introduction to the Mining of Clinical Data. *Clinics in Laboratory*, v. 28, n. 1, p. 1-7, 2008.
- JACKSON, Joyce. Data Mining: A Conceptual Overview. *Communications of the Association for Information Systems*, v. 8, n. 19, p. 267-296, 2002.

- KALAVATHY, R.; SURESH, R.M.; AKHILA, R. KDD And Data Mining. In: INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATION TECHNOLOGY IN ELECTRICAL SCIENCES, 1, v. 2, p.1105-1110, 2007.
- KANAMURA, Alberto Hideki; VIANA, Ana Luiza D'Ávila. Gastos elevados em plano privado de saúde: com quem e em quê. *Revista de Saúde Pública*, São Paulo, n. 41, v. 5, p. 814-820, 2007.
- KOH, Hian Chye; TAN, Gerald. Data Mining Applications in Healthcare. *Journal of Healthcare Information Management*, v. 19, n. 2, p. 64-72, 2005.
- LAROSE, Daniel T. Discovering Knowledge In Data: An Introduction to Data Mining. 1º ed. New Jersey: Wiley, 2005.
- LEAL, Rodrigo Mendes; MATOS, João Boaventura Branco de. Planos de Saúde: uma Análise dos Custos Assistenciais e seus Componentes. *Revista de Administração de Empresas*, v. 49, n. 4, p. 447-458, 2009.
- LIN, Chinho; LIN, Chun-Mei; LI, Sheng-Tun; KUO, Shu-Ching. Intelligent physician segmentation and management based on KDD approach. *Expert Systems with Applications Journal*, v. 34, n. 3, p. 1963–1973, 2008.
- LIRA, Sachiko Araki. *Análise de Correlação: Abordagem Teórica e de Construção dos Coeficientes com Aplicações*. 2004. 196 f. Dissertação (Mestrado em Ciências) – Universidade Federal do Paraná, Curitiba, 2004.
- MACEDO, Dayana Carla de; MATOS, Simone Nasser. Extração de Conhecimento Através da Mineração de Dados. *Revista de Engenharia e Tecnologia*, v. 2, n. 2, p. 22-30, 2010.
- MACKENBACH, Johan P.; MEERDING, Willem J.; KUNST, Anton E. Economic costs of health inequalities in the European Union. *Journal of Epidemiology & Community Health*, v. 65, n. 5, p. 412e419, 2011.
- MALTA, Deborah Carvalho; CECÍLIO, Luiz Carlos de Oliveira; MERHY, Emerson Elias; FRANCO, Túlio Batista; JORGE, Alzira de Oliveira; COSTA, Mônica Aparecida. Perspectivas da regulação na saúde suplementar diante dos modelos assistenciais. *Revista Ciência & Saúde Coletiva*, v. 9, n. 2, p. 433-444, 2004.
- MAROTTI, Juliana; GALHARDO, Alessandra Pucci Mantelli; FURUYAMA, Ricardo Jun; PIGOZZO, Mônica Nogueira; CAMPOS, Tomie Nakakuki de; LAGANÁ, Dalva Cruz. Amostragem em Pesquisa Clínica : Tamanho da Amostra. *Revista de Odontologia da Universidade Cidade de São Paulo*, v. 20, n. 2, p. 186-194, 2008.
- MATTOZO, Teófilo Camara. *Análise de Desempenho de Vendas em Telecomunicações Utilizando Técnicas de Mineração de Dados*. 2007. 115 f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal do Rio Grande do Norte, Natal, 2007.
- MCLEOD, Heather; GROBLER, Pieter. Risk equalisation and voluntary health insurance: The South Africa experience. *Health Policy Journal*, v. 98, n. 1, p. 27–38, 2010.
- MELANDA, Edson Augusto. *Pós-processamento de regras de associação*. 2004. 130 f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Universidade de São Paulo, São Carlos, 2004.
- MOTURU, Sai T.; JOHNSON, William G.; LIU, Huan. Predicting Future High-Cost Patients: A Real-World Risk Modeling Application. In: IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE – BIBM, 2007, Silicon Valley, EUA. *Proceedings...*2007.

NAESSENS, James M; BAIRD, Macaran A.; HOUTEN, Holly K. Van; VANNES, David J.; CAMPBELL, Claudia R. Predicting Persistently High Primary Care Use. *Annals of Family Medicine*, v. 3, n. 4, 2005.

NISHIJIMA, Marislei; POSTALI, Fernando Antonio Slaibe; FAVA, Vera Lúcia. Consumo de Serviços Médicos e Marco Regulatório no Mercado de Seguro de Saúde Brasileiro. *Revista Pesquisa e Planejamento Econômico*, v. 41, n. 3, p. 509-531, 2011.

OLIVEIRA, André Luís Meira de; SOUZA, Clarice Castro; BOEHME, Patrick; SAMOHYL, Robert Wayne. A Normalização de Distribuições Não-Normais Através da Transformação de Box-Cox e Alguns Comentários Sobre a Avaliação de Qualidade. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO – ENEGEP, 17, 1997, Gramado, RS, Brasil. *Anais...* 1997.

OLIVEIRA, Robson Butaca Tadorelli de. *O processo de extração de conhecimento de base de dados apoiado por agentes de software*. 2000. 104 f. Tese (Mestrado em Ciências de Computação e Matemática Computacional) – Universidade de São Paulo, São Carlos, 2000.

PADHY, Neelamadhab; MISHRA, Pragnyan; PANIGRAHI, Rasmita. The Survey of Data Mining Applications and Feature Scope. *International Journal of Computer Science, Engineering and Information Technology*, v. 2, n. 3, p. 43-58, 2012.

PALMUTI, Claudio Silva. *Mensuração do Risco de Crédito Através de Análise Estatística Multivariada*. 2012. 89 f. Dissertação (Mestrado em Administração) - Faculdade Campo Limpo Paulista, Campo Limpo Paulista, 2012.

RAMOS, Carlos Alberto. Aspectos Distributivos do Gasto Público na Saúde. *Revista Planejamento e Políticas Públicas*, n. 24, p. 189-268, 2001.

ROMÃO, Wesley. *Descoberta de Conhecimento Relevante em Banco de Dados sobre Ciência e Tecnologia*. 2002. 238 f. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Santa Catarina, Florianópolis, 2002.

SA, Marcelo Coelho de; SILVA, Ilaydiany Cristina Oliveira da; COSTA, José Alfredo Ferreira; DIOGO NETO, Jose Andre; SIQUEIRA, Antonia Maria dos Santos. Método de Predição de Custos Assistenciais: Considerações no Gerenciamento de Risco na Área da Saúde Suplementar no Brasil. In: SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO – SIMPEP, 17, 2010, Bauru, SP, Brasil. *Anais...* 2010.

SANTOS, Maribel Yasmina; AMARAL, Luís; PIMENTA, Pedro. A descoberta de conhecimento em bases de dados geográficas através da explicitação semântica. In: CONGRESS AND EXIBITION OF LATIN AMERICA GEO-PROCESSING USERS, 5, 1999, Salvador, BA, Brasil. *Proceedings...* 1999.

SEVERO, Eliana Andrea; RANNA, Darla Silvana Risson; KAMARGO, Maria Emilia; OLEA, Pelayo Munhoz. Gestão de Custos Assistenciais: Um Estudo de Caso no Setor De Saúde. In: CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 6, 2010, Niterói, RJ, Brasil. *Anais...* 2010.

SFERRA, Heloisa Helena; CORRÊA, Ângela M. C. Jorge. Conceitos e Aplicações de Data Mining. *Revista de Ciência & Tecnologia*, v. 11, n. 22, p. 19-34, 2003.

SILVA, Hudson Pacífico da. Estado, regulação e saúde: considerações sobre a regulação econômica do mercado de saúde suplementar. *Revista Leituras de Economia Política*, v. 7, n. 1, p. 193-226, 2003.

SILVA, Alberto Luis da; ROMÃO, Verônica Maria Costa; SILVA, Tarcísio Ferreira da; SILVA, Antônio Simões; SAMOHYL, Robert Wayne. Normalidade e Correlação das

Observações GPS. In: SIMPÓSIO DE CIÊNCIAS GEODÉSICAS E TECNOLOGIAS DA GEOSINFORMAÇÃO, 1, 2004, Recife, PB, Brasil. *Anais...* 2004.

SOARES JUNIOR, Jair Sampaio; QUINTELLA, Rogério Hermida. Descoberta de conhecimento em bases de dados públicas: uma proposta de estruturação metodológica. *Revista de Administração Pública*, v. 39, n. 5, p. 1077-1107, 2005.

SVORAY, Tal; MICHAILOV, Evgenia; COHEN, Avraham; ROKAH, Lior; STURM, Arnon. Predicting gully initiation: comparing data mining techniques, analytical hierarchy processes and the topographic threshold. *Earth Surface Processes and Landforms Journal*, v. 37, n. 6, p. 607-619, 2012.

TANGCHAROENSATHIEN, Viroj; PATCHARANARUMOL, Walaiporn; IR, Por; ALJUNID, Syed Mohamed; MUKTI, Ali Ghufon; AKKHAVONG, Kongsap; BANZON, Eduardo; HUONG, Dang Boi; THABRANY, Hasbullah; MILLS, Anne. Health-financing reforms in southeast Asia: challenges in achieving universal coverage. *The Lancet*, v. 377, n. 9768, p. 863-873, 2011.

TOLEDO, Geraldo Luciano; OVALLE, Ivo Izidoro. *Estatística Básica*. 2º ed. São Paulo: Atlas, 1985.

WASAN, Siri Krishan; BHATNAGAR, Vasudha; KAUR, Harleen. The Impact Of Data Mining Techniques On Medical Diagnostics. *Data Science Journal*, v. 5, p. 119-126, 2006.

YOO, Illhoi; ALAFAIREET, Patricia; MARINOV, Miroslav; PENA-HERNANDEZ, Keila; GOPIDI, Rajitha; CHANG, Jia-Fu; HUA, Lei. Data Mining in Healthcare and Biomedicine: A Survey of the Literature. *Journal of Medical Systems*, v. 36, n. 4, p. 2431-2448, 2012.

ZUCCHI, Paola; NERO, Carlos Del; MALIK, Ana María. Gastos em Saúde: Os Fatores que Agem na Demanda e na Oferta dos Serviços de Saúde. *Revista Saúde e Sociedade*, v. 9, n. 1-2, p.127-150, 2000.