



ARTIGO ORIGINAL

INDICADORES ANTECEDENTES NA ÁREA DE ACIDENTES DE TRABALHO NA PREVIDÊNCIA SOCIAL BRASILEIRA

Antônio Carlos Magalhães da Silva¹

Universidade Estácio de Sá, Rio de Janeiro/RJ – Brasil

Paulo Roberto da Costa Vieira²

Universidade Estácio de Sá, Rio de Janeiro/RJ – Brasil

Jonias dos Santos Bueno³

Universidade Estácio de Sá, Rio de Janeiro/RJ – Brasil

RESUMO – Indicadores antecedentes na área de acidentes de trabalho na previdência social brasileira. O propósito deste estudo foi gerar um modelo multivariado sobre os benefícios acidentários, na categoria de acidentes de trabalho na Previdência Social. Foi realizada uma pesquisa de campo a partir de um *survey* de experiência através de um questionário, destinado aos especialistas da DATAPREV (Empresa de Tecnologia e Informações da Previdência Social) e do INSS (Instituto Nacional do Seguro Social). Foi utilizado um modelo de regressão logística na metodologia. Neste estudo foram avaliados todos os benefícios por incapacidade concedidos pela Previdência Social após junho/1994, sendo o modelo adequado somente a quatro modalidades de benefícios por incapacidade. O modelo visa a geração de escalas (*scores*) de risco para os benefícios, possibilitando ações mais pontuais dos órgãos de controle e Auditoria do Ministério da Previdência e Assistência Social, através da ampliação dos critérios inerentes à seleção por amostragem nos processos, e verificação de conformidade dos processos de concessão, além de manutenção dos benefícios por incapacidade regidos pelo Regime Geral de Previdência Social (RGPS) gerenciado pelo Ministério de Previdência e Assistência Social. Os valores gerados para os beneficiários e passíveis de recuperação, correspondem a R\$ 1,34 bilhões para os 65.784 beneficiários e R\$ 108 milhões para 2.597 beneficiários com *scores* entre 99% e 100% de chance de constarem em situação irregular, segundo os critérios vigentes de conformidade da Previdência Social Brasileira. Os resultados encontrados foram robustos de uma forma geral, indicando que tal técnica pode ser aplicada como indicador de benefícios concedidos de forma indevida.

Palavras-chave: Benefícios por incapacidade; Acidentes de trabalho; Métodos Quantitativos.

ABSTRACT – Chronology indicators of accidents work at brazilian social security. The purpose of this study was to generate a multivariate model on accident benefits in the category of industrial accidents in Social Security. A field survey from a survey of experience through a questionnaire was conducted, aimed at specialists DATAPREV (Enterprise Information Technology and Social Security) and the INSS (National Social Security Institute). One logistic regression model was used in the methodology. This study evaluated all disability benefits granted by the Social Security after June / 1994, with the appropriate model only four types of disability benefits. The model aims to generate scales (*scores*) of risk to benefits, allowing more specific actions of the organs of control and audit of the Ministry of Social Security, through the expansion of the criteria inherent in the sample selection processes, and verification compliance procedures for granting, and maintenance of disability benefits governed by the General Social Security System (RGPS) managed by the Ministry of Welfare and Assistance Social. The values generated for beneficiaries and potentially recoverable, corresponding to R\$ 1,34 billion for 65,784 beneficiaries and R\$ 108 millions for 2,597 beneficiaries with *scores* between 99% and 100% chance to appear in an irregular situation, according to the standard criteria for compliance Brazilian Social Security. The results were generally robust, indicating that this technique can be applied as an indicator of providing benefits improperly.

Keywords: Incapacity benefits; Accidents at work; Quantitative Methods.

¹ Professor do Mestrado em Administração da Universidade Estácio de Sá. Bolsista com Auxílio APQ-1 da FAPERJ

² Professor do Mestrado em Administração da Universidade Estácio de Sá

³ Auditor da DataPrev.

1 INTRODUÇÃO

Os processos de verificação promovidos pelas auditorias, bem como as regras de negócio gerenciadas pelos sistemas de Tecnologia da Informação da Previdência Social Brasileira, estão fortemente embasados pelos aspectos legais e pelas definições gerenciais promovidas pelo Instituto Nacional do Seguro Social (INSS), considerando as exceções verificadas no decorrer dos anos e as previsões dos técnicos, baseados em informações governamentais, sobre possíveis tendências de comportamento dos beneficiários, da sociedade, do mercado de trabalho e das transformações sociais e suas relações com as atividades remuneradas e vínculos previdenciários e assistenciais. Tais preocupações são fundamentadas pela missão da Previdência Social Brasileira, que deve garantir os recursos aos seus associados, seja pelo cumprimento do prazo de contribuições previsto no momento da sua filiação ao sistema, decorrente de moléstia que impeça o desempenho de atividade remunerada ou na forma de pensões aos seus dependentes em caso de falecimento do titular vinculado ao RGPS (Regime Geral de Previdência Social).

Este artigo apresentará o processo de geração e testes de um modelo estatístico, capaz de gerar indicadores antecedentes, que permitam a verificação da conformidade dos benefícios acidentários, considerando o score de risco e a similaridade das características em relação aos benefícios cessados pelos órgãos de controle, em função das irregularidades constatadas em análises pretéritas ou através de análises pontuais, realizadas de forma complementar às verificações realizadas pelos Sistemas de Tecnologia da Informação e Comunicação (TIC) utilizados pelo INSS na gestão dos processos de concessão e manutenção dos benefícios relativos à área acidentária.

O modelo gerado pretende apresentar os impactos da aplicação de um modelo estatístico, segundo a análise projetiva, na detecção de características comuns entre os benefícios por incapacidade cessados e os mantidos pela Previdência Social Brasileira, além dos montantes financeiros envolvidos e passíveis de recuperação ou economizados pela Previdência Social, relacionados aos pagamentos realizados aos segurados decorrentes de acidentes ou doenças profissionais.

A relevância do estudo para a sociedade consistirá na ampliação da eficiência na análise dos processos de concessão e manutenção de benefícios, com possibilidades de resultados em economia de recursos públicos, gerados através de análise criteriosa do acervo de benefícios, que possibilitará melhorias na avaliação das concessões, manutenções ou dos cancelamentos dos benefícios, através da verificação mais abrangente da conformidade dos benefícios existentes na categoria de Benefícios por Incapacidade.

A ação contribuirá para a melhoria dos sistemas da Previdência Social através da incorporação da metodologia proposta aos sistemas existentes, bem como aos sistemas que integrarão o Novo Modelo de Gestão (NMG), além disso contribuirá também para redução do déficit previdenciário e para a melhoria dos processos de Auditoria da Previdência Social e da DATAPREV, a partir de análises globais e definições de padrões e tendências sobre as informações previdenciárias; embasará as análises de conformidade e avaliações globais que serão mantidas permanentemente (monitoramento contínuo e auditoria contínua), em comparação com os procedimentos utilizados atualmente sobre amostragens e critérios legais.

O modelo gerado buscará também a redução dos valores de custeio relativos à Previdência Social, já que os segurados que requerem aposentadorias antes da conclusão do prazo de contribuição e idade mínima (Benefícios não-programados), contribuem para o déficit da Previdência Social Brasileira, uma vez que o Sistema Previdenciário baseia-se em um fundo de repartição simples, ou seja, os contribuintes atuais financiam os beneficiários existentes no momento.

2 METODOLOGIA

Segundo Boaventura (2007), o ponto de partida para a pesquisa deve ser desdobrado em questões ou hipóteses, fundamentadas na literatura e operacionalizadas pela metodologia e claramente definidos os objetivos a serem alcançados, através do desdobramento em questões que possibilitem a sua

compreensão e funcionalidade, permitindo a utilização de instrumentos, técnicas e de processos de investigação.

De acordo com Rodrigues (2007, p.42), a pesquisa de campo busca fontes primárias “no mundo dos acontecimentos não provocados e nem controlados pelo pesquisador, que se caracteriza por desenrolar-se em ambiente natural”. Utiliza o procedimento de observação direta do objeto estudado no meio em que é próprio, sem a interferência do pesquisador.

Segundo Cooper et Schindler (2003), o Survey de experiência é uma entrevista onde devem ser obtidas as idéias dos especialistas em relação às questões ou aspectos importantes do assunto tratado e descobrir o que é importante dentro do campo de conhecimento da pessoa. O formato investigativo que será utilizado deve ser flexível para explorar as várias possibilidades que surgirem durante a entrevista.

Após o survey de experiência, foram enviados 200 questionários aos especialistas selecionados por júris no âmbito da Previdência Social e DATAPREV, questionários contendo 20 questões sobre o nível de precedência das variáveis identificadas juntos aos especialistas, que deveriam assinalar, segundo a sua percepção, um dos 7 (sete) níveis da escala Likert que melhor representasse a sua opinião quanto à utilização de cada variável na geração do modelo. Foram respondidos 127 questionários.

3 REFERENCIAL TEÓRICO

As previsões têm o propósito de tornar um pensamento público, sendo fundamentadas por três razões: satisfazer a curiosidade, embasar os tomadores de decisão e gerar consenso entre as pessoas. A previsão como uma forma de consenso está ligada às formas de gerenciamento das organizações e suas movimentações estratégicas, através de cenários futuros e previsões de longo prazo (Makridakis, 1990, p. 55).

Segundo Schwartz (2000), os cenários concebidos através dos enfoques probabilísticos podem ser classificados em Cenário Projetivo e Cenário Prospectivo.

O Cenário Projetivo tem uma única visão de futuro, destina-se predominantemente aos ambientes de baixa ou pouca turbulência e são aconselháveis para projeções de curto prazo, realizando projeções baseadas no comportamento do passado e do presente, considerando que não ocorrerão fatos que alterarão as tendências identificadas. Em relação aos Cenários Prospectivos, possibilitam diversas visões do futuro, projetam tendências, procuram captar rupturas e descontinuidades, adaptando-se melhor aos ambientes turbulentos e aos de longo prazo.

As informações armazenadas, em sua grande maioria, possuem valiosas relações sobre tendências e padrões que podem ser utilizados para melhorar as decisões de negócios (Rezende, 2003). Os cenários preditivos são obtidos pela extrapolação dos fatos passados e consideram que as forças que moldaram o passado e construíram o presente continuarão a atuar e modelar o futuro.

Para Börjeson et al. (2005), o cenário projetivo mostra o que irá ocorrer, o cenário prospectivo expõe o que pode ocorrer, enquanto o cenário normativo indica como atingir um objetivo específico.

As áreas de estatística e mineração de dados estão fortemente ligadas, já que ambas têm como objetivo encontrar padrões e regularidades nos dados, além do fato de que a maioria dos algoritmos utilizados está voltada para a produção de conjuntos de declarações sobre dependências locais entre variáveis de interesse do que propriamente para a construção de modelos globais que incluam todas as variáveis de interesse do problema (Glymour, Madigan, Pregibon e Smyth, 1997 apud Rezende, 2003).

Segundo Hair et al. (2005, p.231), quando a variável dependente tem apenas dois grupos, a regressão logística pode ser a escolhida por diversas razões. Este estudo busca a identificação de benefícios que possam ou não apresentar irregularidades, portanto, devendo ou não integrar os processos de Auditoria. Ainda segundo Hair e al. (2005, p. 232), a regressão logística se aproxima da tarefa de predição de probabilidade de maneira mais semelhante à encontrada em regressão múltipla,

diferindo no sentido de que a Regressão Logística prevê diretamente a probabilidade de um evento ocorrer.

Uma das principais vantagens associadas ao uso da Regressão Logística está na relativa facilidade com que ela pode explicar e prever a ocorrência de determinados fenômenos nas diversas áreas do conhecimento, tais como economia, administração, contabilidade, sociologia e medicina. Pode-se afirmar que o modelo logístico se presta a elucidar dois grandes objetivos: identificar a que grupo certos objetos, pessoas ou fenômenos pertencem e estimar a probabilidade de que eles possam se enquadrar nesta ou naquela categoria (Corrar, 2007).

Foi adotada a técnica de levantamento de dados como método de coleta, tendo sido realizada através da análise dos trabalhos já executados pela Auditoria do MPAS e DATAPREV e da comparação com os resultados obtidos sobre as informações extraídas do acervo da Previdência Social, após a aplicação do método estatístico proposto por este estudo.

Após a verificação do nível de precisão atingido pelo modelo, este foi extrapolado para todo o acervo de benefícios das categorias de benefícios por incapacidade da Previdência Social Brasileira, buscando identificar os benefícios que se encontram em situação divergente, considerando os fatores comuns identificados pelo modelo e as informações relativas às outras categorias de benefícios onde o modelo foi aplicado, uma vez que o modelo em estudo, por ser um estudo de caso, destina-se apenas ao estabelecimento dos scores sobre as quatro espécies que compõem o universo do estudo.

Foram selecionados todos os benefícios nas situações de “CESSADO” pela Auditoria do INSS e pela Inspeção do INSS, com o propósito de confirmar a situação dos mesmos a partir da análise das características (dimensões) que integraram os benefícios (variáveis explicativas), tendo como resultado uma variável dicotômica (INCORRETO ou CORRETO), segundo afirmado por Hair e al. (2005, p. 235), “o pesquisador que se defronta com uma variável dicotômica não precisa apelar para métodos elaborados para acomodar as limitações da Regressão Múltipla e nem precisa empregar Análise Discriminante”.

Para realizar este processo, o conteúdo da amostra foi dividido em 10 classes iguais com um mínimo de 20 observações em cada classe, segundo Hosmer e Lemeshow apud Hair et al. (2005). Em seguida, foram selecionadas as dimensões constituintes dos benefícios por Incapacidade (60% dos casos, ainda segundo Hosmer e Lemeshow) e sobre estas informações foi realizado o processo de Regressão Logística (logit).

Uma vez obtido e validado, o modelo foi aplicado sobre todos os benefícios por incapacidade, requisitados à Previdência Social após junho/1994.

Após o processamento dos questionários específicos, tornou-se possível à realização da seleção das características dos benefícios, identificadas através de um survey de experiência com os especialistas do MPAS e DATAPREV, a partir de informações sobre segurados titulares de benefícios na categoria de Benefícios Acidentários.

Após reuniões, com os especialistas da Área de Gestão e Manutenção de Benefícios da Dataprev, foi obtido o conjunto das variáveis que na opinião dos especialistas, deveria ser analisado com o propósito de identificar o perfil dos beneficiários que mantiveram em algum momento um vínculo com a Previdência Social através de um Benefício por Incapacidade, mesmo estando na condição de cessado no mês de janeiro do ano de 2008, data limite que este estudo considera para a posição dos benefícios. Foram identificadas as seguintes variáveis:

ID_SEG_DIB – Idade do segurado no momento da requisição do benefício por incapacidade e do respectivo início do benefício.

ID_SEG_2008 – Idade do segurado em janeiro de 2008 (data que limitará a avaliação promovida pelo presente estudo).

CS_SEXO – Código atribuído pela Previdência Social Brasileira para indicar o sexo do segurado.

TP_CONTRIB – Tempo de contribuição do segurado junto à Previdência Social desde a sua filiação ao mesmo, no momento da solicitação do benefício por incapacidade.

CS_ESPECIE – Classificador numérico de utilização interna pela Previdência Social para definir os tipos de benefícios abrangidos pela Previdência Social aos seus segurados.

CS_FORMA_FILIACAO – Classificador numérico utilizado pela Previdência Social para categorizar os tipos de vínculos da filiação do segurado ao RGPS (Regime Geral de Previdência Social).

CS_CBO – Classificador utilizado pela Previdência Social para identificar a atividade profissional desempenhada pelo segurado (Classificação Brasileira de Ocupações), segundo as definições do Ministério do Trabalho e Emprego, disponíveis em <http://www.mtecbo.gov.br/busca.asp> (acessado em 04/07/2008).

QT_DEP_IR – Quantidade de dependentes para Imposto de Renda informada pelo Segurado à Previdência Social e vigente em seu vínculo com o RGPS.

QT_SM_RMI – Quantidade de Salários Mínimos relativa ao valor do benefícios recebido pela Previdência Social em decorrência de uma incapacidade para exercer atividade profissional (Renda Mensal Inicial).

CS_ACIDENTE – Classificador utilizado pela Previdência Social para retratar o tipo de acidente que motivou a concessão do benefício por Incapacidade ora prestado.

CS_RAMO_ATIVIDADE – Classificador utilizado pela Previdência Social para retratar o setor empresarial no qual o segurado exerce a atividade profissional.

QT_REVISAO_RMI – Quantidade de revisões solicitadas pelo segurado ou promovidas automaticamente pela Previdência Social no valor definido como a renda que o segurado recebe mensalmente (RMI – Renda Mensal Inicial).

IN_PERICIA_MEDICA – Indicador da realização de uma Perícia realizada através de um médico ou junta médica da Previdência Social ou conveniada a esta, para verificar ocorrência do acidente ao segurado, a sua capacidade laboral ou ainda o tempo necessário para a sua recuperação, quando uma nova perícia deverá ser realizada, avaliando a sua recuperação ou um novo período de recuperação ou a concessão de sua aposentadoria em caráter definitivo devido à impossibilidade do segurado continuar desenvolvendo atividade profissional.

CS_ULTIMO_LT – Classificador utilizado pela Previdência Social para retratar o local (UF - Unidade da Federação) em que o segurado desenvolveu a atividade profissional remunerada.

CS_SISTEMA_CONCES - Classificador utilizado pela Previdência Social para retratar o sistema de concessão utilizado para conceder o benefício por incapacidade ao segurado, este classificar corresponde às diversas formas de solicitação do benefício (internet, Agência da Previdência Social, dentre outros).

CS_BATIMENTO_CNIS – Classificador que retrata o resultado da verificação da situação do segurado em relação aos vínculos e remunerações mantidas pelo segurado, decorrentes de sua atividade profissional remunerada. As informações sobre os vínculos e remunerações dos segurados são mantidas no Cadastro Nacional de Informações Sociais (CNIS), esta base de informações é mantida e

atualizada a partir de informações de vínculos e remunerações enviadas pela Caixa Econômica Federal (CEF) e por informações demográficas repassadas pelo Ministério do Trabalho e Emprego. As informações são gerenciadas pela DATAPREV (Empresa de Tecnologia e Informações da Previdência Social), que através de definição legal, somente poderá conceder benefícios após examinar a situação dos segurados junto ao CNIS, evitando sempre que possível o “ônus da prova” (o segurado só apresentará documentação quando necessário ao processo de concessão de benefícios).

CS_MOT_DESP – Classificador utilizado pela Previdência Social para identificar os depachos de natureza administrativa (sob a análise do INSS) ou judicial (resultante de ações judiciais impetradas por segurados em circunstâncias que necessitam de análise mediadora entre o plano de benefícios e o entendimento dos segurados solicitantes dos benefícios ou as alterações nos mesmos).

QT_ANOS_RECEBE – Quantidade de anos que o beneficiário recebeu ou recebe o benefício por incapacidade garantido pelo Regime Geral de Previdência Social.

As informações para a realização do estudo foram obtidas através de um processo de extração sobre o conjunto de 14.286.545 benefícios contidos no cadastro de benefícios em Janeiro/2008, tendo sido selecionadas as informações sobre os benefícios cessados por irregularidades constatadas pelos Órgãos de Inspeção e Auditoria do INSS (Instituto Nacional do Seguro Social).

Para este estudo, foram selecionados os 1.211 benefícios cessados e outros 1.211 benefícios na situação de Ativos a partir do conjunto de 1.420.673 benefícios por incapacidade ativos concedidos e mantidos pelo INSS e concedidos após junho de 1994.

Para executar um estudo de Regressão Logística, deve ser verificada a correlação entre as variáveis métricas explicativas (Variáveis Independentes), característica denominada como multicolinearidade, evitando que o modelo seja influenciado pelas correlações e os processos stepwise sejam incorretamente aplicados na retirada ou na inclusão de variáveis em modelos com alto poder preditivo (Hair et al (2005); Tabachnick e Fidell (2001); Corrar (2007)).

Tomou-se a decisão de promover agregações nas categorias de natureza não-métrica, utilizou-se o critério de agregação para categorias que não ultrapassassem 100 ocorrências, na amostra de 2.622 benefícios, sendo, portanto agrupadas em

categoria definida neste estudo como “OUTROS” e não há qualquer associação às categorias utilizadas no âmbito do INSS e DATAPREV em relação aos seus significados e descrições.

Antes da geração do modelo logit através do software estatístico, foi executado o processo de separação da amostra em dois conjuntos com respectivamente 60% e 40% das ocorrências, este processo é identificado por Hosmer e Lemeshow (1989) como amostra de treinamento (60%) e amostra de teste (40%), enquanto que para Hair et al. (2005) estas amostras são denominadas como Amostra de Análise (60%) e Amostra de Validação (40%), para Corrar et al. (2007) as amostras são tratadas conforme a denominação de Hosmer e Lemeshow (1989).

Na tabela 1, consta o sumário das informações processadas na geração do modelo, onde foram consideradas 1588 observações na amostra de treinamento, sobre as quais foi gerado o modelo e 1030 observações relativas à amostra de teste, sobre a qual o modelo foi validado.

Tabela 1 - Sumário do processamento do modelo

Unweighted Cases(a)		N	Percent
Amostra de Treinamento	Incluído na análise	1.588	60,60%
	Missing Cases	4	0,20%
	Total	1.592	60,70%
Amostra de Testes		1.030	39,30%
Total		2.622	100,00%

A classificação da variável dicotômica correspondente à situação dos benefícios foi analisada como 0 (zero) para os benefícios ativos e 1 (um) para os benefícios considerados como irregulares pelos Órgãos de controle e cessados por irregularidade, a partir da determinação destes órgãos.

Na amostra de treinamento o modelo consegue classificar corretamente os 819 benefícios na situação de cessados corretamente, porém classifica incorretamente os 769 benefícios na situação de ativos como cessados, atingindo um percentual de classificação de 51,57% de correção. Em relação à amostra de testes, todos os

benefícios foram classificados como ativos, embora 492 destes tenham sido cessados pelos Órgãos de controle, correspondendo ao percentual de classificação correto de 48%.

Através da execução do teste utilizando a técnica stepwise com o método Likelihood value, onde as variáveis independentes serão incluídas gradualmente e retiradas em função do nível de influência na probabilidade (likelihood value) sobre a relação entre os valores examinados e os previstos através da estatística qui-quadrado e da verossimilhança, considerando os níveis definidos para aceitação dos valores (5%) e de retirada dos mesmos (10%), considerando uma distribuição normal com 2 desvios-padrão (-1,96 a +1,96 sobre a curva normal).

No tabela 2 encontram-se demonstrados os coeficientes pseudo-r² de Cox e Snell (0,688) e o pseudo-r² Nagelkerke (0,918). O nível de significância do modelo está retratado pelos coeficientes pseudo-r² tanto o de Cox e Snell que possui o seu melhor valor próximo a 1 (um), mesmo sem poder atingi-lo. Já para o coeficiente pseudo-r² de Nagelkerke, pode atingir o valor máximo igual a 1 (um) (HAIR et al, 2005;CORRAR et al., 2007).

Considerando os níveis atingidos, o modelo possui alto poder preditivo em relação à amostra utilizada, uma vez que o de Cox e Snell consegue explicar 68,8% do log da razão de chance, enquanto o Nagelkerke consegue explicar 91,8% do log da razão de chance, ambos após a inclusão das variáveis explicativas independentes no modelo, através do processo stepwise, considerando como referência o modelo contendo apenas a constante (CORRAR et al., 2007, p. 308).

Tabela 2 - Resumo do modelo

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	1407,309(a)	0,393	0,524
2	1128,145(b)	0,491	0,655
3	637,626(b)	0,626	0,835
4	535,059(b)	0,649	0,866
5	495,774(b)	0,658	0,878

6	469,565(b)	0,664	0,885
7	443,211(b)	0,669	0,893
8	430,095(b)	0,672	0,896
9	359,875(b)	0,686	0,915
10	350,494(b)	0,688	0,918

Tabela 3 - Classificação final do modelo

		Casos selecionados (Treinamento)			Casos não-selecionados (Teste)			
		STATUS		Percentual Correto	STATUS		Percentual Correto	
		ATIVO	CESSADO		ATIVO	CESSADO		
Passo 10	STATUS	ATIVO	747	22	97,14	524	9	98,31
		CESSADO	34	785	95,85	25	467	94,92
		Percentual geral			96,47			96,68

Em relação ao observado na tabela 3 o modelo gerado é capaz de classificar os benefícios da amostra de validação com 96,47% de precisão em relação à sua situação (ativo ou cessado), representada pela variável dicotômica STATUS, enquanto na amostra de teste o nível de precisão obtido foi de 96,68%, segundo Hair et al. (2005), o nível de predição do modelo deverá ser o obtido na classificação da amostra de teste, definidas segundo Hair et al. (2005) como amostra de análise (60%) e amostra de validação (40%).

Na amostra de análise (60% do total) o erro do tipo I (benefícios cessados que foram classificados como ativos) foi de 4,15% (34 casos) e o erro do tipo II (benefícios ativos que foram classificados como cessados) foi de 2,86% (22 casos).

Na amostra de validação (40% do total) o erro do tipo I foi de 5,08% (25 casos) e o erro do tipo II foi de 1,69% (9 casos).

Diante dos custos de classificação descritos acima, observa-se que o modelo não penaliza os segurados que possuem benefícios ativos, já que o percentual de classificações de ativos como cessados (erros tipo II) é proporcionalmente muito inferior ao erro tipo I, tanto na amostra de análise quanto na amostra de validação.

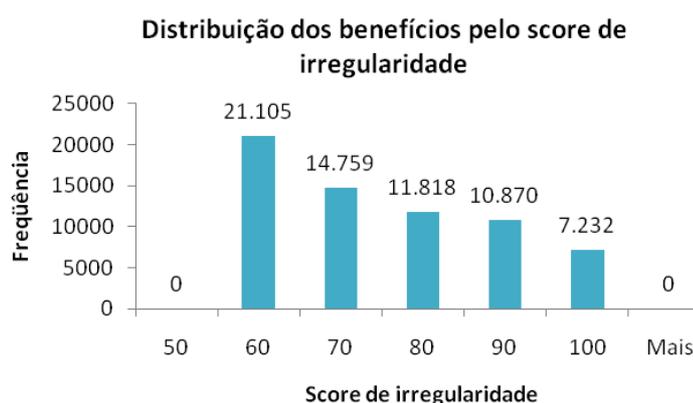
O modelo final apresenta 4 variáveis métricas e 7 categóricas. Em sua fase inicial o estudo possuía 20 variáveis independentes, o método stepwise (likelihood ratio) utilizado na geração do modelo através do software estatístico, desconsiderou 9 (nove) variáveis a partir das significâncias definidas para entrada (5%) e para eliminação das variáveis (10%).

4 CONCLUSÕES E RECOMENDAÇÕES

Após a aplicação do modelo gerado no universo dos benefícios por incapacidade na situação de ativo em dezembro de 2007 (1.399.296 de recebedores), foram identificados 65.784 benefícios na situação de ativos e segundo o modelo, recomenda-se uma análise criteriosa dos benefícios identificados, em função do alto grau similaridade dos benefícios identificados em relação aos benefícios cessados por irregularidade.

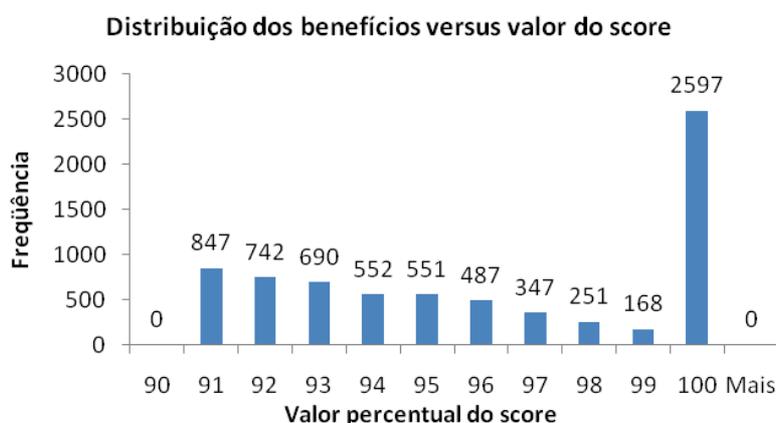
Conforme observado na figura 1, há 7.232 benefícios com probabilidade (score) entre 90% e 100% de similaridade em relação aos benefícios por incapacidade cessados por irregularidade.

Figura 1 - Histograma dos benefícios com score entre 50% e 100%



A figura 2 apresenta a distribuição dos benefícios com score, resultante da aplicação do modelo, entre 90% e 100%.

Figura 2 - Distribuição dos benefícios na faixa de risco entre 90% e 100%



A seguir são apresentados os valores gerados para pagamento dos benefícios identificados pelo modelo com score superior a 50%. Os valores foram calculados a partir de uma solução de tecnologia desenvolvida para a execução deste estudo. Na apuração dos valores, foi considerada a quantidade de salários-mínimos recebida pelo segurado em cada competência (mensalmente), convertida para o valor do salário-mínimo em Reais (R\$) nas respectivas competências. A soma dos valores correspondentes ao recebimento de cada benefício, valores gerados pela Previdência Social para pagamento, é apresentada através do agrupamento das variáveis explicativas identificadas pelo modelo de regressão logística gerado.

São apresentados os valores para os benefícios que possuem scores entre 50% e 100%.

Especie	Descrição	Quantidade	Valor gerado (R\$)	% do Valor gerado
31	AUXILIO DOENÇA PREVIDENCIARIO	53.070	958.028.602,87	71,75%
32	APOSENTADORIA INVALIDEZ PREVIDENCIARIA	10.726	299.110.452,61	22,40%
91	AUXILIO DOENÇA POR ACIDENTE DO TRABALHO	804	25.031.575,66	1,87%
92	APOSENT. INVALIDEZ ACIDENTE TRABALHO	1.184	52.982.047,44	3,97%

Total		65.784	1.335.152.678,58	
-------	--	--------	------------------	--

Em relação aos valores de recebimento como renda inicial, após a concessão dos benefícios, foram identificados 18.345 benefícios com renda até um salário-mínimo, correspondendo a 27,89% dos 65.784 benefícios selecionados, também foram identificados 44.771 benefícios com valores entre 1 e 5 salários-mínimos, correspondendo a 68,06% do conjunto selecionado pelo modelo estatístico.

A pesquisa sobre a conformidade dos benefícios por incapacidade baseou-se nas opiniões dos funcionários especialistas da DATAPREV e INSS, elegendo-se um conjunto de características que foi confirmado através de um survey de experiência, identificando-se 20 características que foram submetidas ao modelo de regressão logística tendo como referência um conjunto de 1.211 benefícios por incapacidade cessados por irregularidade, a partir de um conjunto de 14.286.545 de benefícios, segundo os órgãos de controle da Previdência Social. A amostra probabilística sistemática contendo 1.211 benefícios acidentários na situação de ativos foi obtida a partir do conjunto 2.373.704 de benefícios, sendo considerados 1.420.673 como integrantes do universo de benefícios por incapacidade ativos.

O modelo logístico resultante do estudo considerou 10 variáveis explicativas independentes e classificou com 96,68% de precisão a situação dos benefícios como ativos ou cessados (decorrentes de irregularidade), selecionando 65.784 benefícios com níveis de score entre 50% e 100%, que totalizam o valor de R\$ 1.335.152.678,58 disponibilizados para pagamento ao longo da vigência destes benefícios, a partir de um universo de 1.420.673 benefícios ativos na competência dezembro/2007, correspondendo o nível de acerto a 4,63% do universo na referida competência.

O nível de precisão na classificação dos quantitativos identificados e os valores monetários apresentados são evidências de que há critérios comuns entre os benefícios cessados por irregularidade e os benefícios ativos, capazes de propiciar a verificação de conformidade dos benefícios ativos a partir das características resultantes da análise dos órgãos de controle da Previdência Social e que este

modelo multivariado, se utilizado nos processos de concessão e manutenção de benefícios por incapacidade, poderá evitar o pagamento indevido de 4,63% dos benefícios, além de auxiliar nas ações de auditoria contínua, possibilitando a análise de benefícios em momentos diferentes, a partir da análise por faixas de score atribuídas pelo modelo, podendo representar ações de recuperação de valores já pagos aos segurados ou evitar o pagamento indevido de valores financeiros da ordem de R\$ 1,335 bilhão.

O estudo apresentou o mapeamento de 10 características explicativas, a partir de um universo de 20 variáveis independentes, tendo sido identificadas 4 variáveis métricas e 6 taxonômicas, resultando em uma variável dependente (score) com valores entre 0,50 e 1,00 com referencial de corte (cut-off) de 0,50, classificando como benefícios incorretos àqueles que apresentaram score superior a 0,50.

Dentre as variáveis explicativas taxonômicas foram identificadas características que poderão ser incorporadas aos sistemas de tecnologia da informação da Previdência Social, contribuindo para a redução dos riscos de concessão ou pagamentos indevidos aos segurados na área de benefícios por incapacidade. Os maiores riscos identificados são relativos aos benefícios da categoria de Auxílio Doença Previdenciário (espécie 31), detentores de vínculo como desempregado, que desenvolvam atividades no setor industrial, que exerçam atividade na Região Nordeste, que tenha sido concedidos através dos sistemas de concessão descentralizada ou do sistema de concessão Prisma/SUB e cujas informações demográficas ou contributivas não tenham sido confirmadas pelo Cadastro Nacional de Informações Sociais (CNIS) no momento da concessão do benefício.

Dentre as variáveis explicativas métricas identificadas pelo modelo, para cada alteração unitária ocorre uma redução ou aumento na chance de um benefício ter a sua situação alterada entre correto ou incorreto. Para a variável VL_MR_ATU (Valor recebido pelo segurado) a cada unidade acrescentada, ocorre a redução de 0,4% na chance de um benefício tornar-se cessado. A variável QT_REVISAO_RMI (Quantidade de revisões solicitadas sobre a quantidade de salários-mínimos no início do benefício), a cada acréscimo unitário, ocorre a redução de 76,15% das chances de um benefício assumir o status de cessado. Para a variável Anos_recebidos

(Tempo de recebimento do benefício em anos), ocorre a redução de 61,20% nas chances de um benefício por incapacidade assumir o status de cessado, a cada ano de recebimento do benefício.

As variáveis demográficas relativas aos segurados, não apresentaram nenhum poder preditivo no modelo resultante, sendo relevantes na geração do modelo apenas as características empregatícias e concessórias em poder da Previdência Social.

O modelo gerado permite que neste estudo seja desconsiderada a hipótese (H0) de que não há critérios comuns nos benefícios por incapacidade cessados que possibilitem a verificação da conformidade legal dos benefícios por incapacidade concedidos e mantidos. O resultado do modelo assegura e confirma a aceitação da hipótese (H1) que segundo a análise das características dos benefícios cessados, possibilita a geração de Indicadores Antecedentes que serão aplicados na verificação de conformidade dos benefícios mantidos pela Previdência Social Brasileira.

O modelo gerado apresenta característica conservadora em suas previsões, uma vez que os erros de classificação mais significativos não penalizam os segurados, uma vez que foram identificados erros (tipo I) de 2,86% e 1,69% nos benefícios ativos que foram classificados como incorretos (cessados), enquanto foram identificados erros (tipo II) de 4,15% e 5,08% para os benefícios cessados (incorretos) que foram classificados como ativos, neste caso necessitando que outras ações sejam tomadas, uma vez que o nível de precisão do modelo não contempla estas situações.

Através da aplicação do modelo resultante do estudo, foi possível identificar que há 7.232 benefícios com scores entre 90% e 100%, resultando em R\$ 174.225.911,24, fato que evidencia a execução de avaliações detalhadas sobre estes benefícios, porém, também foram evidenciados 2.597 benefícios com scores entre 99% e 100%, resultando em R\$ 109.986.938,27 gerados para pagamento dos segurados, indicando que as ações de acompanhamento intensivo e avaliação detalhada

deveriam ser aplicadas, prioritariamente, sobre este último conjunto de benefícios, em função dos valores envolvidos e do quantitativo evidenciado, seguindo o princípio da materialidade.

REFERÊNCIAS

ALMEIDA, F.C.; SIQUEIRA, J.O. Comparação entre Regressão Logística e Redes Neurais na previsão de falência de bancos Brasileiros. Terceiro Congresso Brasileiro de Redes Neurais. V. 4. Florianópolis, 1997. Disponível em: <http://www.fia.com.br/proinfo/artigos/Florianopolis.PDF> acessado em 11/07/2008.

ALTMAN, Edward I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, v. 23, setembro de 1968.

BOAVENTURA, Edivaldo M. Metodologia da Pesquisa. São Paulo: Atlas, 2007.

BORJESON, L., HOJER, M., DREBORG, K. H., EKVALL, T., FINNVEDEN, G. "Towards a User's Guide to Scenarios: a Report on Scenario Type and Scenario Techniques", Environmental Strategies Research. Stockholm: Royal Institute of Technology, 2005. Disponível em: <<http://www.infra.kth.se/fms>>

Acesso em: 05/02/2008.

BRASIL. Livro Branco da Previdência Social. Ministério da Previdência e Assistência Social. Brasília, 2002. 152. p.

BRITO, Giovani. A.S.; ASSAF NETO, A. Modelo de Classificação de Risco de Crédito de Empresas. *Revista de Contabilidade e Finanças (RCF – USP)*, São Paulo, v.19, n.46, p.18-29, jan./abr. 2008.

BUARQUE, S. C., Metodologia e técnicas de construção de cenários globais e regionais. IPEA, Textos para discussão, fev.2003.

CARVALHO, Alexandre Pinto de; NERI, Marcelo Cortes. Mobilidade Ocupacional e Raça: Origens, Destinos e Riscos dos Afro-Brasileiros. Rio de Janeiro: Fundação Getúlio Vargas, 2000. Disponível em: <HTTP://virtualbib.fgv.br/dspace/handle/123456789/408>. Acessado em: 05/07/2008.

COOPER, Donald R.;SCHINDLER, Pamela S. Métodos de Pesquisa em Administração. Tradução de Luciana de Oliveira da Rocha. 7ed, Porto Alegre: Bookman, 2003.

CORRAR, Luiz J.; PAULO, Edilson; DIAS FILHO, José Maria. Análise Multivariada. São Paulo: Atlas, 2007.

COX, D.R. Notes on Some Aspects of Regression Analysis. *Journal of the Royal Statistical Society*, vol.131, n.3, p.265-279, 1968.

COX, D.R.; HINKLEY, D. V.; Theoretical statistics., London: Chapman & Hall, 1986. 174 p.

- COX, D. R.; SNELL, E.J. Analysis of binary data. 2. Ed. London: Champman and Hall, 1989.
- DATAPREV. Empresa de Tecnologia e Informações da Previdência Social. Anuário Estatístico Da Previdência Social: Suplemento Histórico – V. 1(1980.1996) – Brasília 1997.
- DATAPREV. O Modelo Previdenciário Brasileiro. Rio de Janeiro: 1979.
- DATAPREV.SUIBE – Sistema Unificado de Informações sobre Benefícios [site intrane WWW-SUIBE]. Rio de Janeiro: março/2007.
- FRÁGUAS JUNIOR, Renério; ALVES, Tânia. C. T. F. Depressão no Hospital Geral:Estudo de 136 casos. Revista da Associação Médica Brasileira. V.48. n. 3. 2002. p. 225-230.
- GIL, Antonio Carlos. Como elaborar projetos e Pesquisa. 3ª. Ed. São Paulo: Atlas;1995:58.
- GIL, Antônio Carlos. Métodos e Técnicas de Pesquisa Social. 5ª.ed. São Paulo: Atlas, 1999.
- GIMENO, Suely Godoy A.; SOUZA, J.M.P. Utilização de estratificação e modelo de regressão logística na análise de dados de estudos caso-controle. Revista de Saúde Pública. v.29. n.4.1995.p.283-289.
- GUIMARÃES, I.A; NETO, A.C. Reconhecimento de Padrões: Metodologias estatísticas em crédito ao consumidor. Rae-eletrônica. V.1. Número 2, jul-dez/2002. Disponível em: <http://www.rae.com.br/electronica/index.cfm?FuseAction=Artigo&ID=1215&Secao=FINANÇAS2&Volume=1&Numero=2&Ano=2002>.
- HAIR, Joseph .F.Jr; ANDERSON, Rolph.E; TATHAM, Ronald. L; BLACK, William. C. Análise Multivariada de dados. Tradução Adonai Schlup Sant'Anna e Anselmo Chaves Neto. - 5ª. Ed. - Porto Alegre: Bookman, 2005.
- HOSMER, D.W.; LEMESHOW, S.; Applied logistic regression. New York: John Wiley, 1989, 307 p.
- LAKATOS, Eva Maria; MARCONI, Marina de Andrade. Fundamentos da Metodologia Científica. 6ª. Ed. São Paulo: Atlas, 2007a.
- LAKATOS, Eva Maria; MARCONI, Marina de Andrade. Técnicas de Pesquisa. 6ª. Ed. São Paulo: Atlas, 2007b.
- LAKATOS, Eva Maria; MARCONI, Marina de Andrade. Metodologia Científica. 5ª. Ed. São Paulo: Atlas, 2008.
- MAKRIDAKIS, SPYROS G., WHEELWRIGHT, S. C., McGEE, V. E. Forecasting, Methods and Application. 2 ed. New York: John Wiley & Sons, 1983.
- MAKRIDAKIS, SPYROS,G. Forecasting, Planning and Strategy for 21st Century. New York: THE FREE PRESS, 1990. 279p.
- MARQUES, H. R.; MANFROI, J.; CASTILHO, M. A.; NOAL, M. L. Metodologia da pesquisa e do trabalho científico. 2. ed. CAMPO GRANDE: EDITORA UCDB, 2006. v. 1. 130 p.

- MATIAS-PEREIRA, José. Manual de Metodologia da Pesquisa Científica. São Paulo: Atlas, 2007.
- MENARD, SCOTT. Applied logistic regression analysis. 1995.
- MICHEL, Maria Helena. Metodologia e Pesquisa Científica em Ciências Sociais. São Paulo: Atlas, 2005.
- MINUSSI, João Alberto; DAMASCENA, Cláudio; NESS-JUNIOR, Walter Lee. Um modelo preditivo de Solvência Utilizando Regressão Logística. In: ENANPAD, 25. 2001, Campinas, Anais. Rio de Janeiro: ANPAD, 2001.
- MORITZ, G. Planejamento por cenários prospectivos: a construção de um referencial metodológico baseado em casos. Tese (Doutorado) – UFSC, Florianópolis, 2004.
- MPAS. Dinâmica Socio-econômica e Previdência Sistêmica. Brasília: 1999.
- MPAS. Empresa de Tecnologia e Informações da Previdência Social. ANUÁRIO ESTATÍSTICO DA PREVIDÊNCIA SOCIAL – Ano 1 (1988.1992) – Brasília 1993.
- MPAS. Empresa de Tecnologia e Informações da Previdência Social. ANUÁRIO ESTATÍSTICO DA PREVIDÊNCIA SOCIAL – Ano 03. ANO 2006
- MPAS. Empresa de Tecnologia e Informações da Previdência Social. Boletim Estatístico da Previdência Social - Vol. 12 Nº 07. Brasília. MPAS, 2006.
- MPAS. Previdência, Assistência Social e Combate à Pobreza. Brasília, 2000 196 p. (Previdência Social, série Debates) ISBN-85-88219-05-0.
- MPAS. Regulamento da Previdência Social. Decreto 3048 de 06/05/1999. Brasília 1999, 334. p.
- MPAS. Seguridade e Desenvolvimento: Um projeto para o Brasil. Brasília:2001.
- NAGELKERKE, N. J. D. A note on a general definition of the coefficient of determination. Biometrika, p. 691-692, 1991.
- ONUSIC, L.M.; KASSAI, Silvia; VIANA, A. B. N. Comparação dos resultados de utilização de análise por envoltória de dados e regressão logística em modelos de previsão de insolvência: Um estudo aplicado a empresas brasileiras. Facep Pesquisa – V. 7 – N. 1, 2004. Disponível em: http://www.facef.br/facefpesquisa/2004/nr1/2_ONUSIC_KASSAI_VIANA.pdf. Acessado em 11/07/2008.
- PEREIRA, Orlando Mansur; NESS JUNIOR, Walter Lee. O modelo e-score de previsão de Falências para Empresas de Internet. Rio de Janeiro: Fundação Getúlio Vargas, 2008. Disponível em: <HTTP://virtualbib.fgv.br/dspace/handle/123456789/1231>. Acessado em: 10/09/2008.
- REZENDE, Solange Oliveira. Sistemas Inteligentes. São Paulo: Manole, 2003.
- RODRIGUES, Rui Martinho. Pesquisa Acadêmica. São Paulo: Atlas, 2007.
- ROESH, Sylvia Maria Azevedo. Projeto de estágio do curso de Administração: guia para pesquisas, projetos, estágios e trabalhos de conclusão de curso. São Paulo: Atlas, 1996.

- SAMPAIO, Rosana. F.; NAVARRO, Albert; MARTIN, M. Incapacidades laborales: problemas em la reinserción al trabajo. Cadernos de Saúde Pública. Rio de Janeiro, out-dez, 1999.
- SANTOS, Narcisa M. G. dos. Um estudo prospectivo sobre a Previdência Social Brasileira. Rio de Janeiro, 2004. Tese – Universidade Federal do Rio de Janeiro, COPPE – 229 p.
- SANTOS¹, Alcione M.; SEIXAS, J. M.; PEREIRA, B. B.; MEDRONHO, R. A. Utilizando Redes Neurais artificiais e Regressão Logística na Predição da Hepatite A. Revista Brasileira de Epidemiologia. V.8. n.2, p. 117-126, 2008.
- SCHWARTZ, P. A arte da visão de longo prazo: planejando o futuro em um mundo de incertezas. São Paulo: Best Seller, 2000.
- SIMON, Herbert Alexander. Comportamento Administrativo. Rio de Janeiro: Fundação Getúlio Vargas, 1957.
- SIMON, Herbert Alexander. A capacidade de decisão e de liderança. Rio de Janeiro: Editora Fundo de Cultura, 1960.
- SINCICH, Terry. Business Statistics by Example. 5ed. New Jersey: Prentice-Hall, 1996.
- SINPAS. Reorganização e racionalização da Previdência Social. Lei no. 6439 de 01/09/1977. Brasília. MPAS. 1977.
- STEVENSON, Willian J. Estatística Aplicada à Administração. São Paulo: Harbra, 2001.
- TABACHNICK, Barbara G.; FIDELL, Linda. S. Using multivariate statistics. New York: HarperCollins, 2001.
- THOMAS, W.; COOK, R. D.; Assessing influence on predictions from generalized linear models technometrics, Alexandria, v.32, p.59-65, 1990.
- VENTURA, Magda Maria. O estudo de Caso como Modalidade de Pesquisa. Rio de Janeiro: Socerj, 2007.
- Yin, R. K. Case Study Research: (1984). Design and Methods. Vol. 5. Applied Social Research Methods, ed. Leonard Bickman. Beverly Hills, CA: Sage, 1984.
- YIN, Robert K. Estudo de caso: planejamento e métodos. 2^a. ed. Porto Alegre: Bookman, 2001.