

Detecção de dados atípicos no estudo de tentativa de fraude no abastecimento de frotas utilizando o programa Seis Sigma

Bruno de Almeida Scheibler (UFRGS)

Vera Lúcia Milani Martins (PPGEP/UFRGS)

Márcia Elisa Soares Echeveste (IM-Estatística/PPGEP/UFRGS)

Resumo

Buscando reduzir os custos relativos ao abastecimento dos veículos, as empresas estão analisando o comportamento dos gastos dependentes do fator humano. Dessa forma, torna-se necessária uma ferramenta que auxilie no estudo de tentativa de fraude no abastecimento de frotas devido a vieses humanos, como comportamento de consumo dependente do motorista e/ou do estabelecimento. Este artigo propõe uma ferramenta, baseado na metodologia Seis Sigma, para detectar os dados atípicos no processo de abastecimento dos veículos e a partir dessa detecção, auxiliar na identificação das tentativas de fraudes. A metodologia foi aplicada em uma amostra coletada na base de dados de uma empresa de gestão de frotas. O resultado identificou quais fatores influenciam comportamentos diferenciados de um padrão estimado probabilisticamente, o que pode representar indícios de fraudes. A identificação destes fatores orienta estudos posteriores, priorizando a análise entre as diversas combinações de fatores existentes.

Palavras-chave: fraude, abastecimento, gestão de frotas, seis sigma.

1 Introdução

O transporte rodoviário e o uso de veículos durante a prestação de serviços correspondem a uma parcela considerável do custo total de um produto final na cadeia produtiva. O abastecimento dessas frotas compõe até 40% do custo total da frota, que também inclui os custos de manutenção, assistência 24h, documentação e multas (Revista Brasil Econômico, 2012). As empresas cujos custos diretos consideram a operação dos veículos necessitam realizar uma gestão do abastecimento da frota, de forma a controlar o seu gasto.

Dentre os diversos sistemas existentes de gestão do abastecimento de frotas, um dos mais utilizados é o de cartões magnéticos, que fornecem os dados através das máquinas do tipo POS (Point of Sale), encontrados nos postos de combustíveis. Este sistema realiza uma verificação dos dados fornecidos, sendo possível autorizar ou não o pagamento do abastecimento realizado. A definição dos parâmetros que restringem a transação é feita pelo gestor da frota, no caso de haver um cargo específico, ou pelo responsável pelo abastecimento da frota. Os parâmetros podem ser referentes a diversos indicadores, tais como desempenho, preço abastecido, tipo de combustível, data e local.

Visto que as frotas das companhias são compostas de uma grande quantidade de veículos, a gestão individual do abastecimento torna-se inviável. Portanto, os parâmetros aceitáveis de abastecimento são estabelecidos para diferentes grupos de veículos. Os veículos podem variar de modelo, ano de fabricação, local de atuação, motoristas que os utilizam e de peso da carga (BIELLI; BIELLI; ROSSI, 2011).

Portanto, a elaboração dos limites de controle dos grupos de veículos depende destas características que provocam variabilidade nas medidas. O desconhecimento desta variabilidade, por margem de segurança, faz com que os gestores permitam limites mais amplos do que realmente seria necessário, com objetivo de não correr o risco de altas taxas de não autorização nos abastecimentos.

Com as constantes melhorias tecnológicas dos sistemas de gestão de abastecimento, as indústrias estão expandindo o controle de qualidade para a área humana da empresa (ERICSSON, 2001). A dificuldade de monitorar o impacto causado pelos funcionários no abastecimento dos veículos permite que ocorram desvios inesperados, incluindo a ocorrência de fraudes (TRNKA, 2010).

A presença de fraude em qualquer processo produtivo impacta dentro e fora da empresa, trazendo prejuízos imensuráveis. Características de produto como confiabilidade e qualidade percebida, quando não atingem os

resultados esperados, fazem com que o cliente mude a sua posição perante a empresa. Portanto, é imprescindível que se tenha um método eficiente de detecção de fraudes que esteja de acordo com o cenário onde será aplicado (LI et al., 2008). Neste estudo de gestão de frotas considera-se que o motorista possa causar diferença no comportamento do consumo. Contudo, não é viável quantificar diretamente a relação do motorista com o veículo.

O estudo desta interação ocorre por meio de uma análise das variáveis do próprio abastecimento, cujos desvios tenham grande probabilidade de ter a intervenção humana (MERCEDES; JAIME, 2003). Todas as transações (autorizadas ou não) realizadas através do POS geram uma gama de dados que possibilitam o estudo posterior para monitoramento. Com base nos dados dos registros é possível identificar as causas de variação de comportamento de consumo. A proposta deste trabalho é desenvolver uma ferramenta, baseado na metodologia Seis Sigma, para detectar os dados atípicos no processo de abastecimento dos veículos e a partir dessa detecção, auxiliar na identificação das tentativas de fraudes.

A utilização do método proposto como uma ferramenta de auxílio na gestão de prevenção de fraudes pode trazer benefícios como redução dos gastos, transparência no abastecimento, diferencial competitivo e indicadores precisos (LI et al., 2008). Esses benefícios justificam o investimento do desenvolvimento do método.

Este artigo está composto pelas seguintes etapas: primeiramente é feita uma introdução do tema e do conteúdo do estudo. No referencial teórico, os conceitos de transporte, gestão de frotas, fraudes no abastecimento de frotas e Seis Sigma são contextualizados. Em seguida, serão descritos os procedimentos metodológicos realizados para elaborar o estudo e, por fim, os resultados obtidos serão discutidos e as conclusões finais serão feitas.

2 Referencial Teórico

Para se compreender melhor o impacto que uma fraude no abastecimento pode causar, será feita primeiramente uma contextualização da participação do consumo de combustíveis na economia do Brasil e das empresas, para então concluir a importância da realização de uma gestão de frota eficiente e apresentar os métodos disponíveis para a realização do controle de fraudes.

2.1 Panorama do setor de transporte no Brasil

O meio de transporte predominantemente utilizado no Brasil é o transporte rodoviário. Conforme a Tabela 1, este setor representou 61,1% do TKU (Tonelada Quilômetro Útil) total no Brasil em 2011. Mais de 60% do transporte de cargas é realizado através da malha rodoviária (ATLAS DO TRANSPORTE 2006).

Tabela 1: Matriz do transporte de cargas

Modal	TKU (milhões)	Participação (%)
Rodoviário	485.625	61,1
Ferrovário	164.809	20,7
Aquaviário	108.000	13,6
Dutoviário	33.300	4,2
Aéreo	3.169	0,4
Total	794.903	100

Fonte: (adaptado da Revista CNT Transporte Atual)

Apesar da geografia brasileira favorecer o uso de ferrovias (TEDESCO et al., 2011), o setor de transporte no país está sofrendo uma redução nos recursos investidos para desenvolvimento (CNT, 2012). No que tange ao consumo de combustível, o transporte ferroviário utiliza apenas 20% do volume consumido pelo transporte rodoviário (BALDEZ, 2011). Percebe-se, assim, que há uma exploração ineficiente dos recursos disponíveis.

Em 2008, os custos de transporte no Brasil constituíram 6,9% do PIB. Nos Estados Unidos, neste mesmo ano, a relação foi de 5,4% (CNT, 2012). Diante deste cenário, a gestão de frota possui um papel importante na economia do Brasil, buscando reduzir os custos operacionais dos veículos, incluindo o custo devido ao abastecimento (VALENTE; PASSAGLIA; NOVAES, 2003). O setor de transporte, portanto, apresenta um custo significativo para a economia do Brasil.

2.2 Gestão de frotas

De acordo com Valente et al. (2003), gestão de frota significa “reger, administrar ou gerenciar um conjunto de veículos pertencentes a uma mesma empresa”. Dentre os diversos serviços considerados nesta gestão, está o abastecimento dos veículos.

Visto que a gestão da frota de uma empresa é essencial para que esta permaneça no mercado, deve-se ter de forma clara a composição do Custo Total da Frota (TCO – Total Cost Ownership) para se tomar ações de redução. Quando se trata de Transporte Rodoviário de Cargas (TRC), o abastecimento compõe entre 40% e 60% do valor do frete (CNT, 2010).

A presença de fraude no abastecimento da frota, além de apresentar uma perda inerente, traz prejuízos imensuráveis como a imprecisão na coleta de dados e a distorção nos indicadores de abastecimento. Frente a esta situação, será discutido nos itens a seguir o benefício de um sistema *online* de gestão do abastecimento e as formas de se monitorar este sistema.

2.3 Gestão de fraude no abastecimento de frotas

Devido à distância entre os postos de combustíveis e as filiais das empresas, a fraude ocorre no momento em que o abastecimento é realizado. O gestor da frota, por sua vez, está localizado dentro da empresa e tem a responsabilidade de gerenciar e pagar o valor abastecido (LADEIRA; BERTE; ROMANZINI, 2011).

Originalmente, o controle dos abastecimentos era realizado de forma manual, reembolsando os motoristas a partir da quilometragem informada no hodômetro dos veículos ou através da coleta de Cupons Fiscais. A redução dos custos de telecomunicações e de TI possibilita o uso de sistemas *online* para controlar em tempo real o abastecimento da frota. Estas ferramentas auxiliam a tomada de decisões e aumentam a eficiência dos veículos (REGAN; MAHMASSANI; JAILLET, 1996).

Os sistemas existentes de gestão de frotas possibilitam consultar posteriormente os dados obtidos no momento da transação. Estes dados servem de apoio para análises da operação e possíveis estudos estatísticos, como a utilização das ferramentas de Seis Sigma.

2.4 Seis Sigma

De acordo com Reis (2011), Seis Sigma pode ser considerado como “uma metodologia que utiliza métodos estatísticos e não-estatísticos, integrados em uma sequência lógica, que são aplicados usualmente através de uma abordagem de gestão de projetos com o objetivo de atingir elevados níveis de desempenho, que tipicamente podem ser medidos e expressos através da variação do processo...”. O processo, neste caso, é o abastecimento da frota de uma empresa.

O gestor de frota, tendo como suporte a metodologia Seis Sigma nos parâmetros que restringem o abastecimento, está estabelecendo um nível de erro tolerável de aproximadamente 0,002 ppm (partes por milhão) (PEREZ-WILSON, 1999). A Figura 1 exemplifica os parâmetros de desempenho (km/l) mínimo e máximo, configurados com seis desvios-padrão em relação à média.

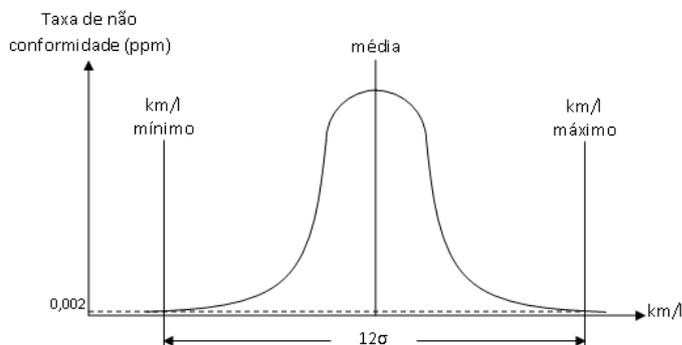


Figura 1: Seis Sigma na variável desempenho (km/l) (fonte: adaptado de REIS, 2011)

2.5 Métodos e Técnicas Estatísticas

Os métodos de suporte a análise citados na definição de Seis Sigma podem ser exemplificados por ferramentas como o Controle Estatístico de Processos (CEP), Análise de Variância (ANOVA) e Projeto de Experimentos. Pode-se aplicar tal metodologia com a integração dos métodos através do ciclo DMAIC (*Define, Measure, Analysis, Improve e Control*). Trnka (2010) implementou a detecção de fraudes em subsídios financeiros para fazendas na etapa *Control* do DMAIC. Segundo Reis (2011), as etapas do DMAIC são: (i) Definição: definição de oportunidades; (ii) Medição: medição dos processos; (iii) Análise: determinação das causas; (iv) Melhoria: aperfeiçoamento dos processos e obtenção de resultados; e (v) Controle: manutenção dos ganhos obtidos. Para desenvolver este estudo, serão utilizadas apenas algumas ferramentas deste ciclo: SIPOC (*Supplier, Input, Process, Output e Client*), Diagrama de Pareto e Função Perda de Qualidade de Taguchi.

A sigla SIPOC significa *Supplier* (Fornecedor), *Input* (Entrada), *Process* (Processo), *Output* (Saída) e *Client* (Cliente). Preenchendo esses cinco itens ao longo das etapas do processo em análise, é possível identificar as variáveis passíveis de mensuração.

Outra ferramenta importante para a análise é o Diagrama de Pareto, que faz uma priorização no diagnóstico, sendo utilizado na priorização de projetos e na detecção das causas vitais de defeitos de fabricação (JURAN; GODFREY, 2000). Muitos estudos de Seis Sigma buscam conhecer as fontes de variação da variável de resposta observada.

2.5.1 Função Perda de Qualidade

A Função Perda de Qualidade, proposta por Taguchi (2004), calcula em termos financeiros a perda obtida por fornecer ao consumidor um produto fora das especificações. Segundo Taguchi (2004), a perda de qualidade é causada pela insatisfação dos seus consumidores, ou seja, é a perda obtida após a venda do produto, a longo prazo. O autor também afirma que a perda de qualidade pode estar associada às características do produto.

Um produto pode possuir a característica de menor-é-melhor, maior-é-melhor ou nominal-é-melhor. Neste último caso, significa que há uma meta finita a se atingir e limites de especificação inferiores e superiores a essa meta. Porém, o controle de qualidade baseado apenas em estar dentro ou fora dos limites de especificação é uma medida inadequada devido à pobreza de informações fornecidas, visto que a variação entre os limites de especificação é ignorada (TAGUCHI et al., 2004).

A perda aumenta continuamente conforme se distancia da média, inclusive dentro dos parâmetros mínimo e máximo. A Figura 2 exemplifica a curva da Perda de Qualidade para a variável desempenho (km/l) onde todos os desvios em relação à média são considerados.

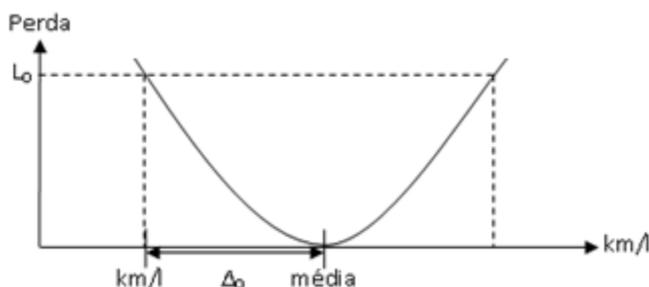


Figura 2: Perda de Qualidade para o desempenho (fonte: adaptado de Taguchi, 2004)

No caso de variáveis do tipo nominal-é-melhor, quando se está monitorando duas ou mais repetições para cada combinação de fatores, deve-se utilizar a Equação 1 para calcular a perda de qualidade:

$$L(y_i) = k[\sigma_i^2 + (\bar{y}_i - m)^2] \quad (1)$$

onde k é uma constante de proporcionalidade, σ_i^2 é a variância das “ j ” repetições em torno da média e $(\bar{y}_i - m)^2$ é o desvio da média em relação à meta. O índice “ i ” equivale à i -ésima combinação de fatores e “ j ” equivale à j -ésima repetição. Portanto, para reduzir a perda, é possível reduzir a variabilidade em relação à média

ou ajustar a média à meta. No caso de abastecimento de frotas, pode-se optar por controlar médias de desempenho (km/l).

Quando a variável é do tipo menor-é-melhor, deve-se calcular a perda de cada y_i conforme a Equação 2:

$$L(y_i) = k \left(\sum_{j=1}^m \frac{y_{ij}^2}{m} \right) \quad (2)$$

onde k é uma constante de proporcionalidade e y_{ij} equivale à j -ésima repetição da i -ésima combinação de fatores.

Os procedimentos aplicados neste estudo contemplam a utilização das ferramentas descritas neste capítulo. A integração de tais ferramentas será descrita a seguir no procedimento metodológico.

3 Procedimento Metodológico

A seguir, será descrito em três seções o procedimento metodológico do estudo. Primeiramente, será feita uma descrição do cenário onde será realizado o estudo. Em seguida, serão caracterizados os métodos de pesquisa e de trabalho.

3.1 Descrição do cenário

O estudo será realizado com base nos dados capturados por uma empresa de gestão de frotas. Esta empresa atua em nível nacional e oferece uma solução completa em gestão sustentável de frotas. Atualmente é líder no mercado, com mais de 9.000 clientes no país, administrando mais de 520.000 veículos. Possui uma rede credenciada de mais de 12.000 postos e 7.000 oficinas.

O principal produto do portfólio, Sistema de Abastecimento, será utilizado para fornecer os dados necessários para a detecção de dados atípicos no abastecimento da frota do cliente escolhido como exemplo. Este sistema, através de um meio de pagamento eletrônico, aceito na rede credenciada de postos no Brasil, controla o abastecimento através da sistemática de gestão por exceção. Todas as transações que estiverem fora dos parâmetros operacionais definidos pelo gestor da frota são auditadas ou protegidas evitando que haja algum desvio no abastecimento. O fluxo do sistema é o seguinte:

1. Configuração dos parâmetros - são definidos os parâmetros restritivos da frota como, por exemplo: km/l mínimo e máximo, capacidade do tanque, R\$/l mínimo e máximo e intervalo de horas entre cada transação.
2. Abastecimento - abastecimento efetivo do veículo com o combustível desejado.
3. Transação - após o abastecimento, deve-se passar o cartão magnético (fornecido pela empresa de gestão de frotas) no POS (*Point of Sale*) e identificar o motorista através de matrícula e senha ou o veículo através da placa (a identificação varia conforme o tipo de cartão adquirido). Por fim, são registrados manualmente no POS os seguintes dados: hodômetro do veículo, litros abastecidos, valor total e tipo de combustível. O sistema verifica o cadastro de placa/motorista, limite de crédito, parâmetros de inconsistência e de restrições. Se os dados estiverem de acordo, o sistema aprova o pagamento da transação e emite o comprovante, que deve ser assinado pelo motorista. Caso contrário, o sistema reprovará o pagamento.
4. Relatórios - imediatamente após a aprovação ou reprovação da transação, são disponibilizados no sistema *online* relatórios detalhando as últimas transações.

3.2 Caracterização do método de pesquisa

O método de pesquisa utilizado neste estudo é de natureza aplicada, pois a metodologia proposta será aplicada nos dados coletados de um dos clientes da empresa de gestão de frotas. A sua abordagem, portanto, é quantitativa: tanto os *inputs* (obtidos na base de dados da empresa) quanto os *outputs* do estudo são variáveis mensuráveis. O objetivo do método de pesquisa é explicativo, pois, de acordo com Gil (2002), será analisado o comportamento de cada variável e fator para detectar a probabilidade de fraude no abastecimento.

3.3 Caracterização do método de trabalho

O estudo será aplicado em sete etapas, descritas no Quadro 1. No período de observação selecionado, outubro de 2012, foram registradas 64.223 transações, correspondentes a 5.915 veículos, 5.846 motoristas e 328 estabelecimentos. É importante observar que nem todos os veículos foram conduzidos por todos os motoristas e que nem todos os motoristas abasteceram em todos os estabelecimentos.

Quadro 1: Etapas do Procedimento Metodológico

Eta- pa	Objetivo	Ferramenta	Responsável	Entrada	Saída
I	Identificação das variáveis	SIPOC e <i>brainstorming</i>	Colaboradores da empresa	Processo de abastecimento	<i>Inputs</i> e <i>outputs</i> do processo
II	Seleção das variáveis	Pareto e <i>brainstorming</i>	Colaboradores da empresa	<i>Inputs</i> e <i>outputs</i> do processo	Variáveis priorizadas
III	Identificação dos fatores	SIPOC	Colaboradores da empresa	Variáveis priorizadas	Principais fatores
IV	Coleta de dados	<i>Business Intelligence</i>	Autor deste artigo	Variáveis, fatores, período desejado e cliente desejado	Dados coletados
V	Análise dos fatores	Análise da população	Autor deste artigo	Fatores e dados coletados	Fatores que devem ser descartados do estudo
VI	Identificação dos dados atípicos	Função Perda de Qualidade	Autor deste artigo	Variáveis, fatores e dados coletados	Dados atípicos e respectivas probabilidades de ocorrências
VII	Ordenamento das observações com as maiores probabilidades de ser um dado atípico	Seis sigmas	Autor deste artigo	<i>Outliers</i> e respectivas probabilidades de ocorrências	Ordenamento dos dados atípicos

Fonte: elaborado pelo autor

Estas sete etapas, no contexto do ciclo DMAIC, compõem apenas os três primeiros passos (Definir, Medir e Analisar). As etapas seguintes (Melhoria e Controle) devem ser realizadas conforme a gestão de cada empresa. É importante ressaltar que é necessário monitorar e revisar o resultado para manter as melhorias realizadas em vigor.

A aplicação da metodologia proposta acima possibilita identificar os dados atípicos do processo de gestão de frotas e avaliar a probabilidade de fraude associada a cada dado identificado. A seguir, será descrita a aplicação da metodologia para a detecção de dados atípicos no estudo de tentativa de fraude no abastecimento de frotas. Os resultados obtidos também serão analisados a seguir.

4 Resultados e Discussões

Nesta seção, serão apresentadas as ações realizadas em cada etapa proposta no Procedimento Metodológico e em seguida os resultados serão discutidos e avaliados.

4.1 Identificação das variáveis

A partir do fluxo do sistema de abastecimento, descrito no Procedimento Metodológico, foi realizado um *brainstorming* com os colaboradores da empresa de gestão de frotas para definir os *inputs* das etapas “Configuração dos parâmetros” e “Transação”, identificados pelos *suppliers* “Gestor da frota” e “Motorista”, respectivamente. Esses *inputs* são as potenciais variáveis, que serão priorizadas na Etapa II. A Tabela 2 apresenta o resultado do SIPOC realizado para o processo de abastecimento.

Tabela 2: SIPOC do processo de abastecimento

Supplier	Input	Process	Output	Client
Gestor da frota	km/l máx e mín	Sistema cadastra as parametrizações para a frota do cliente	Parametrizações cadastradas e em funcionamento instantaneamente	Gestor da frota
	Capacidade tanque			
	R\$/l máx e mín			
	Hora/transação			
	Dia, hora, dia semana			
	Qtd transações/ veículo			
	Serviços			
Motorista	km	Sistema verifica cadastro de placa/motorista, limite de crédito, parâmetros de inconsistência e de restrições	Transação aprovada	Gestor da frota
	litros		km inválida	
	R\$/l		nº matrícula inexistente	
	R\$ e litro de óleo		Σ transações > limite de crédito	
	Demais serviços		consumo < permitido	
	Matrícula		consumo > permitido	
	Tipo de combustível		outros	

Fonte: elaborado pelo autor

4.2 Seleção das variáveis

Para definir quais serão as variáveis utilizadas no estudo, foi elaborado um Diagrama de Pareto ordenando quantas vezes cada *input* não obteve o *output* “Transação aprovada”, ou seja, quais são as variáveis responsáveis pela maior parte dos abastecimentos fora dos parâmetros definidos pelo gestor da frota. A Tabela 3 apresenta os motivos das transações reprovadas registradas entre janeiro de 2012 e outubro de 2012. Para simplificar a visualização, a Tabela 3 apresenta apenas os motivos que compõem a fração de 70% do percentual acumulado, ou seja, a curva A definida pelos colaboradores da empresa para orientar a priorização das variáveis.

Tabela 3: Motivos das transações reprovadas entre janeiro de 2012 e outubro de 2012

Motivo da Ocorrência	Qtd. Transações Protegidas	% Total	% Total acum.
km inválida	888.977	21,24%	21,24%
nº matrícula inexistente	477.728	11,41%	32,65%
Σ transações > limite de crédito	406.998	9,72%	42,37%
consumo < permitido	381.282	9,11%	51,48%
consumo > permitido	370.345	8,85%	60,32%
senha inexistente	363.190	8,68%	69,00%

A Figura 3 apresenta o Diagrama de Pareto com os 15 principais motivos das transações reprovadas e a porcentagem acumulada de ocorrências. Os demais motivos foram agrupados como “OUTROS”.

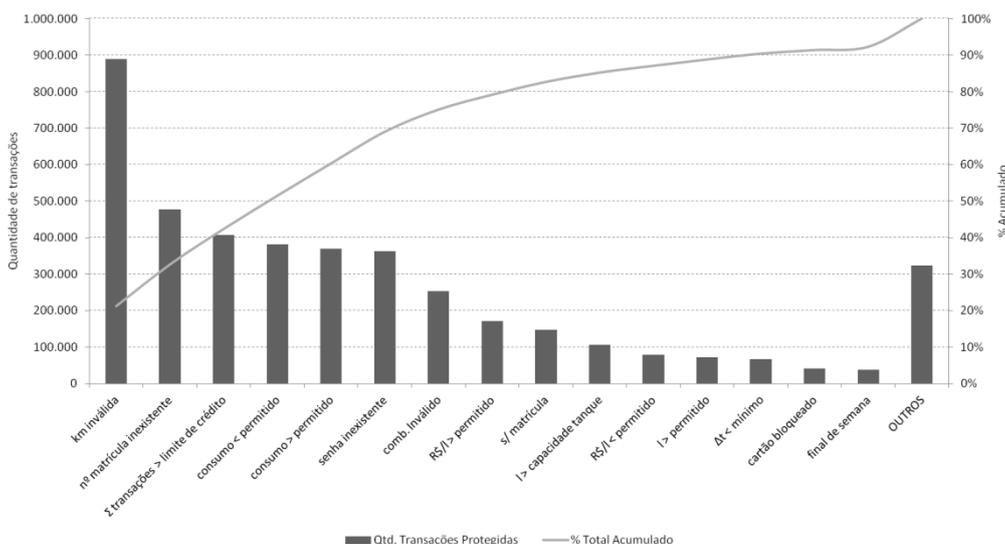


Figura 3: Diagrama de Pareto (fonte: elaborado pelo autor)

Foi utilizado o princípio de Pareto para definir quais são os principais motivos de reprovação das transações. Com base nisso, realizou-se um *brainstorming* com os colaboradores da empresa de gestão de frotas para definir as relações de *inputs* da Tabela 2 que correspondem a cada motivo de reprovação e priorizaram-se as seguintes variáveis para serem mensuradas no estudo:

- Variável 1: R\$/l (preço pago pelo combustível);
- Variável 2: Litros abastecidos/semana; e
- Variável 3: Quilometragem rodada/semana.

4.3 Identificação dos fatores

Foi utilizado o resultado do SIPOC realizado na Etapa I para definir, em conjunto com os colaboradores da empresa de gestão de frotas, quais são os fatores que podem influenciar no resultado das variáveis priorizadas na Etapa II. Os fatores selecionados foram os seguintes: (i) placa; (ii) motorista; e (iii) estabelecimento.

4.4 Coleta de dados

Para realizar a coleta de dados, foram utilizadas como amostra as transações de um dos principais clientes da empresa de gestão de frotas, pertencente ao ramo de telecomunicações. O Quadro 2 exemplifica a estrutura da coleta de dados realizada através do *Business Intelligence* (BI), que resultou em 64.223 transações.

Quadro 2: Estrutura para a coleta de dados no BI (fonte: elaborado pelo autor)

Cód. Transação	Data	Placa	Motorista	Estabelecimento	R\$	km Rodada	Litros
1	dd/mm/aaaa	AAA1111	Motorista 1	Estabelecimento 1	R\$ 0,00	0	0,00

Fonte: elaborado pelo autor

Analisando os dados coletados, verificou-se que, de todas as transações, mais de 84% é oriundo dos veículos leves (Gol, Celta, Palio, etc) e, desta fração, aproximadamente 80% corresponde ao combustível "gasolina comum". Esta proporção equivale a 81,82% do total pago. Deste modo, para a aplicação da metodologia e para se ter uma amostra com um mesmo perfil de operação (desempenho médio e quilometragem rodada média), as análises a seguir foram realizadas para os abastecimentos feitos por veículos leves com gasolina comum, resultando em um total de 42.852 observações das 64.223 transações ocorridas no período contemplado no estudo. Desconsiderando as repetições de combinações de fatores, o total obtido é de 10.068 observações.

4.5 Análise dos fatores

Foi identificado que os fatores "placa" e "motorista" possuem uma relação muito próxima: dos 5.027 motoristas considerados na amostra restringida na etapa anterior, apenas 7,26% utilizou mais de um veículo; dos 5.059 veículos considerados, por sua vez, somente 18,56% foram utilizados por mais de um motorista.

Desta forma, a maior parte dos abastecimentos foi realizada por uma única combinação de "placa" *versus* "motorista". Este fato justifica a eliminação de um dos fatores em questão. Dado que as tentativas de fraude dependem da intervenção humana, é preferível que o resultado final deste estudo apresente os motoristas às placas. Assim sendo, nesta etapa e nas etapas seguintes serão considerados apenas dois fatores: "motorista" e "estabelecimento".

4.6 Identificação dos dados atípicos

Antes de iniciar os cálculos, foi necessário tratar os dados coletados, ou seja, remover os dados inconsistentes. Em relação à variável "km/semana", por exemplo, foram desconsiderados os valores negativos e iguais a zero, visto que são resultados de erros de digitação no POS.

As variáveis "l/semana" e "km/semana" foram consideradas como variáveis do tipo nominal-é-melhor. Portanto, foi utilizada a Equação 1 para obter as suas perdas. A variável "R\$/l", por outro lado, é do tipo menor-é-melhor, dado que é preferível pelos gestores de frota pagar pelo menor preço. Neste caso, foi utilizada a Equação 2.

Primeiramente, a Função Perda de Taguchi foi calculada para cada variável, desconsiderando a constante de proporcionalidade “ k ”, visto que não se deseja quantificar as perdas financeiramente e sim estabelecer uma hierarquia entre os valores. As perdas calculadas para cada variável são referentes a cada nível de “ i ”, ou seja, a cada combinação dos fatores “motorista” e “estabelecimento”.

No caso das variáveis do tipo nominal-é-melhor, a meta “ m ”, equivalente à média de todas as observações da variável em questão, é comparada com a média das repetições do nível “ i ” que está sendo calculado. Desta forma, considera-se um cenário onde a média geral não está associada à ocorrência de fraudes. Não é considerado o cenário pessimista, onde a fraude ocorre de forma generalizada tal que a média geral estaria distorcida.

Para se obter uma hierarquia única entre os níveis de “ i ”, foi calculada a proporção de cada perda em relação ao total da variável correspondente. Feito isto, obteve-se o percentual de perda acumulada de cada “ i ”, somando as perdas relativas de cada variável. Assim, pôde-se somar perdas de diferentes escalas, atribuindo o mesmo peso para cada variável, de acordo com a Equação 3:

$$\% \text{ perda acumulada } (i) = \frac{L(y_{i1})}{\sum_{i=1}^p L(y_{i1})} + \frac{L(y_{i2})}{\sum_{i=1}^p L(y_{i2})} + \frac{L(y_{i3})}{\sum_{i=1}^p L(y_{i3})} \quad (3)$$

onde os índices 1, 2 e 3 se referem às variáveis “R\$/l”, “l/semana” e “km/semana”, respectivamente.

Ao calcular o percentual de perda acumulada, os níveis de atipicidade em cada variável são acumulados para cada y_i e, portanto, tem-se o grau de atipicidade total para cada “motorista/estabelecimento”.

4.7 Ordenamento das observações com as maiores probabilidades de ser um dado atípico

A obtenção do percentual de perda acumulada na etapa anterior para cada “ i ” foi analisada, identificado que os valores seguem a distribuição Gama com três parâmetros. Segue abaixo os valores dos parâmetros, aproximados pelo Minitab®:

- Forma (*shape*): 1,67822;
- Escala (*scale*): 0,00013; e
- Limiar (*threshold*): 0,00009.

A Figura 4 apresenta o histograma do percentual de perda acumulada, com uma linha aproximando a curva de Gama com os parâmetros descritos acima. Pode-se observar que os dados acompanham o comportamento da curva de Gama.

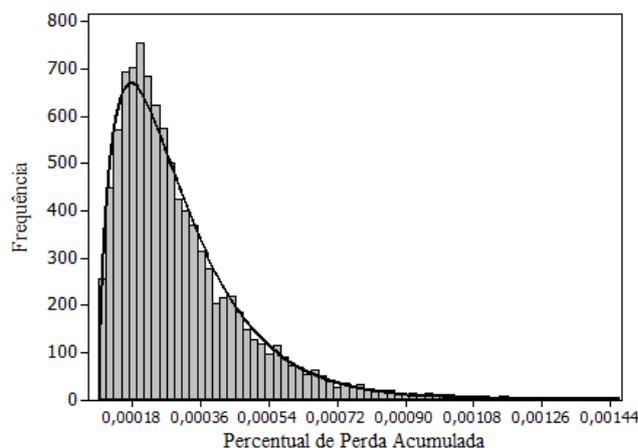


Figura 4: Histograma da perda total com aproximação da curva de Gama com três parâmetros (fonte: elaborado pelo autor)

Ajustando à metodologia Seis Sigma, o percentual de perda acumulada cuja probabilidade acumulada for maior do que aproximadamente 99,87% deve ser considerado como dado atípico. Desta forma, foram identificadas 15 combinações de “motorista/estabelecimento” cujos percentuais de perda acumulada são atípicas. A Tabela 8 lista essas combinações com suas respectivas perdas acumuladas e probabilidades acumuladas. Para manter o sigilo dos motoristas e dos estabelecimentos, os seus nomes foram substituídos por pseudônimos.

Tabela 4: Combinações de motorista/estabelecimento fora da margem de seis sigmas

Motorista / Estabelecimento	% perda acumulada	Probabilidade acumulada
Motorista 3572 / Posto 38	0,0014311	99,98%
Motorista 4906 / Posto 50	0,0013938	99,98%
Motorista 791 / Posto 104	0,0013909	99,97%
Motorista 3899 / Posto 134	0,0012756	99,94%
Motorista 3497 / Posto 90	0,0012743	99,94%
Motorista 4667 / Posto 97	0,001274	99,94%
Motorista 249 / Posto 78	0,0012699	99,94%
Motorista 4413 / Posto 58	0,0012411	99,93%
Motorista 2492 / Posto 130	0,001224	99,92%
Motorista 3464 / Posto 104	0,0012187	99,91%
Motorista 4139 / Posto 68	0,0012096	99,91%
Motorista 2644 / Posto 92	0,00119	99,89%
Motorista 3693 / Posto 143	0,001176	99,88%
Motorista 100 / Posto 132	0,0011687	99,88%
Motorista 3937 / Posto 72	0,0011631	99,87%

Fonte: elaborado pelo autor

Observando o resultado da Tabela 4, pode-se identificar a recorrência do estabelecimento “Posto 104”. Este fato sugere uma análise separada dos fatores. O resultado final orienta o estudo de tentativas de fraudes no abastecimento, indicando as quinze principais combinações de motoristas e estabelecimentos a serem analisadas, entre 5.027 motoristas e 165 estabelecimentos, no caso da amostra selecionada.

Analisando detalhadamente o histórico de abastecimentos do “Motorista 3572” no “Posto 38”, relação que representa o maior percentual de perda acumulada, pôde-se identificar indícios de fraude. O histórico apresenta uma variação de “km/semana” maior do que 300% entre a primeira e a segunda semana de abastecimento, sendo que o indicador “litros/semana” não variou significativamente neste mesmo período. Assim, pode-se concluir que o “Motorista 3572” percorreu uma quilometragem menor na primeira semana, porém consumiu o mesmo volume abastecido na segunda semana. Além disso, os preços praticados estão entre os mais altos da amostra considerada. As demais combinações da Tabela 8 também apresentam um comportamento atípico nos seus históricos de abastecimento, apresentando indícios de fraudes.

5 Conclusões

O processo de detecção de fraudes no abastecimento de frotas se tornou parte integrante da gestão de frotas, auxiliando na busca pela redução do Custo Total da Frota. Dados os diversos fatores que influenciam direta e indiretamente nos indicadores de abastecimento, há a necessidade de uma ferramenta que oriente o estudo de tentativas de fraudes. Este artigo teve como objetivo desenvolver uma ferramenta, baseado na metodologia Seis Sigma, para detectar os dados atípicos no processo de abastecimento dos veículos e a partir dessa detecção, auxiliar na identificação das tentativas de fraudes.

No estudo realizado, a ferramenta identificou os cenários que mais se enquadram como atípicos e que apresentam indícios de fraude. Esta classificação facilitará o estudo aprofundado do histórico de abastecimento da frota em questão, que fará parte das etapas de Melhoria e de Controle do ciclo DMAIC. Repetindo este ciclo diversas vezes, o nível de gestão da frota e, conseqüentemente, o controle do custo de abastecimento, estarão cada vez mais precisos.

O procedimento pode ser aplicado a outros indicadores, como km/l e l/transação, que não foram abordados no estudo apresentado e pode inclusive ser automatizado através da sua programação. Sugere-se que, em trabalhos futuros, as perdas sejam quantificadas financeiramente, definindo os valores para a constante de proporcionalidade “*k*” proposta por Taguchi. Para atribuir um peso diferente a cada variável, aconselha-se o uso da Função Perda Quadrática Multivariada, abordada por Ribeiro e Elsayed (1995).

Referências Bibliográficas

- ATLAS DO TRANSPORTE 2006. Confederação Nacional do Transporte. Disponível em: <<http://www.cnt.org.br/Paginas/Atlas-do-Transporte.aspx>>. Acessado em: 20 mai. 2012.
- BALDEZ, L. H. Entrevista. Revista ANTT, v. 3, n. 2, nov. 2011. ISSN: 2177-6571.
- BIELLI, M.; BIELLI, A.; ROSSI, R. Trends in Models and Algorithms for Fleet Management. *Procedia Social and Behavioral Sciences*. Elsevier, n. 20, p. 4-18, 2011.
- CNT. Confederação Nacional do Transporte. Economia em Foco. Brasília, 10 jan. 2012.
- CNT. Confederação Nacional do Transporte. Informe Econômico do Setor de Transporte. Brasília, 2. ed., nov. 2010.
- CNT. Confederação Nacional do Transporte. Revista CNT Transporte Atual. Brasília, ano XVIII, n. 200, mai. 2012.
- Empresa Jornalística Econômico S.A. Revista Brasil Econômico. São Paulo, 14, 15 e 16 set. 2012.
- ERICSSON, E. Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transportation Research Part D*. Elsevier, p. 325-345, 2001.
- GIL, A. C. Como elaborar projetos de pesquisa. São Paulo: Atlas, 4. ed., 2002.
- JURAN, J. M.; GODFREY A. B. Juran's Quality Handbook. McGraw-Hill, 5. ed., 2000.
- LADEIRA, W. J.; BERTE, R. S.; ROMANZINI, A. P. Business-to-business e Controle Total de Frota: um estudo na CTF Technologies. *Gestão Contemporânea*, Porto Alegre, ano 8, n. 9, p. 103-128, jan./jun. 2011. Disponível em: <<http://seer2.fapa.com.br/index.php/arquivo>>. Acessado em: 20 mai. 2012.
- LI, J.; HUANG, K.; JIN, J.; SHI, J. A survey on statistical methods for health care fraud detection. *Health Care Management Science*. Springer US, v. 11, n. 3, p. 275-287, set. 2008.
- MERCEDES, R. M.; JAIME, C. G. Reduced consumption and environment pollution in Mexico by optimal technical driving of heavy motor vehicles. *Fuel and Energy Abstracts*, v. 44, n. 3, mai. 2003.
- PEREZ-WILSON, M. Six sigma: understanding the concept, implications and challenges. *Advanced Systems Consultants*, 1999.
- REGAN, A. C.; MAHMASSANI, H. S.; JAILLET, P. Dynamic Decision Making for Commercial Fleet Operations Using Real-Time Information. *Transportation Research Record*, n. 1537, p. 91-97, 1996. ISSN: 0361-1981.
- REIS, D. A. F. Curso Seis Sigma Green Belts. Flemming Consultoria Ltda, v.1, 2011.
- RIBEIRO, J. L.; ELSAYED, E. A. A case study on process optimization using the gradient loss function. *International Journal of Production Research*, v. 33, n. 12, p. 3233-3248, 1995.
- TAGUCHI, G.; CHOWDHURY, S.; WU, Y. Taguchi's Quality Engineering Handbook. John Wiley & Sons, 1. ed., 2004.
- TEDESIO, G. M. I.; VILLELA, T. M. A.; GRANEMANN, S. R.; FORTES, J. A. A. S. Mercado de Transporte Rodoviário de Cargas no Brasil. Revista ANTT, v. 3, n. 2, nov. 2011. ISSN: 2177-6571.
- TRNKA, A. Six Sigma Methodology with Fraud Detection. *Advances in Data Networks, Communications, Computers*. *Advances in Data Networks Communications Computers-Proceedings*, p. 162-165, 2010.
- VALENTE, A. M.; PASSAGLIA, E.; NOVAES, A. G. Gerenciamento de Transporte e Frotas. São Paulo: Cengage Learning, 1. ed., 2003.