

Prevendo a demanda de ligações em um call center por meio de um modelo de Regressão Múltipla

Marco Aurélio Carino Bouzada
Eduardo Saliby



Resumo

Este trabalho descreve – por meio do estudo de um caso – o problema da previsão de demanda de chamadas para um determinado produto no call center de uma grande empresa brasileira do setor – a Contax – e como ele foi abordado com o uso de Regressão Múltipla com variáveis dummy. Depois de destacar e justificar a importância do tema, o estudo apresenta uma breve revisão de literatura acerca de métodos de previsão de demanda e de sua aplicação em call centers. O caso é descrito, contextualizando, inicialmente, a empresa estudada e descrevendo, a seguir, a forma como ela lida com o problema de previsão de demanda de chamadas para o produto 103 – serviços relacionados à telefonia fixa. Um modelo de Regressão Múltipla com variáveis dummy é, então, desenvolvido para servir como base do processo de previsão de demanda proposto. Este modelo utiliza informações disponíveis capazes de influenciar a demanda, tais como o dia da semana, a ocorrência ou não de feriado e a proximidade da data com eventos críticos, como a chegada da conta à residência do cliente e seu vencimento; e apresentou ganhos de acurácia da ordem de 3 pontos percentuais para o período estudado, quando comparado com a ferramenta anteriormente em uso.

Palavras-chave: Call center. Previsão de demanda. Regressão Múltipla.

1 Introdução

Call centers consistem em centros operacionais, instalados para utilizar telecomunicação e tecnologias computacionais, para automatizar uma quantidade volumosa de diferentes atividades e serviços telefônicos, tanto recebidos quanto originados pelo centro. Os centros do tipo *inbound*, onde as chamadas são originadas pelos clientes, são caracterizados como um sistema constituído de várias pessoas (atendentes), que recebem ligações telefônicas de outras pessoas, normalmente clientes – ainda que potenciais, que desejam obter alguma informação, contratar algum serviço, comprar algum produto, responder a alguma pesquisa, atualizar dados, registrar ocorrências ou fazer reclamações, entre outras demandas (GROSSMAN et al., 2001; HAWKINS et al., 2001).

Sakamoto (2001) define *call center* como sendo o lugar onde os chamados são feitos ou recebidos em grandes quantidades com o objetivo de apoiar processos de vendas, *marketing*, serviços ao cliente, suporte técnico ou outra atividade específica.

Para Minghelli (2002), um *call center* pode ser entendido como um centro integrado de contato entre empresas e

consumidores estabelecido de forma remota e/ou virtual, por meio do uso da tecnologia.

Do ponto de vista operacional, um *call center* caracteriza-se, basicamente, por um conjunto de teleoperadores, reunidos em um espaço e interligados ao público de uma empresa por sistemas de telefonia, recebendo e realizando ligações para atender, tirar dúvidas, ouvir reclamações, resolver problemas ou vender produtos e serviços. Do ponto de vista do negócio, um *call center* é um meio de relacionamento entre empresas e seus clientes, que tem como objetivos: estreitar esse relacionamento; melhorar, acelerar e desburocratizar processos; promover a fidelização do cliente com a empresa por meio das melhorias no atendimento; e aumentar receitas (WEISSHUHN, 2004).

Bapat e Pruitte Jr. (1998) argumentam que o *call center* é a arma estratégica mais visível de uma empresa. É uma frente de batalha de negócios na qual milhões de dólares em produtos e serviços são comprados, vendidos e negociados. Também é um local onde milhares de clientes são conquistados e perdidos instantaneamente. Como empresas líderes se tornam mais criativas em disseminar

informação e agregar valor aos seus consumidores através de linhas telefônicas, é natural que elas enxerguem o *call center* como seu ponto de acesso ao mercado.

Embora, por definição, o *call center* seja uma forma de relacionamento sem contato presencial, seu uso vem se tornando cada vez mais intenso por empresas de diversos setores da economia. A indústria, o comércio e uma infinidade de empresas de serviços utilizam-se deste tipo de recurso para atender, vender, contatar e entender melhor seus clientes (WEISSHUHN, 2004).

Segundo Azevedo e Caldas (2002), um *call center* consiste, por essência, em uma atividade do setor de serviços, que funciona como um elo entre a empresa e a sua base de clientes. Os autores acrescentam que as inovações tecnológicas nesse setor têm sido muito grandes, provocando uma reestruturação ampla no funcionamento dessas organizações.

Dezenas de bilhões de dólares foram gastos em *call centers* na última metade da década de 1990. O crescimento dessa indústria naquela década foi na faixa de 20% ao ano e é esperado que esse crescimento se mantenha nesse patamar durante o início deste século. Na economia de hoje, os *call centers* não apenas se tornaram os pontos primários de contato entre clientes e empresas, mas também um grande investimento para muitas organizações (GROSSMAN et al., 2001; WEINBERG; BROWN; STROUD, 2006).

A indústria de *call centers* tem crescido bastante nos últimos anos, em particular no Brasil. O mercado doméstico já movimenta acentuadas cifras financeiras. Boa parte dos custos envolvidos na administração das centrais de atendimento é oriunda dos recursos humanos, por isso, é vital – do ponto de vista da competitividade do negócio – gerenciá-los de forma eficiente.

Mas os gerentes e planejadores dos *call centers* têm um emprego muito mais difícil hoje em dia do que no passado. Com mais produtos e serviços sendo criados, disponibilizados no mercado, vendidos e assistidos, em relação a momentos anteriores, eles precisam entender o que está havendo nos *call centers* para saber como as suas características influenciam os indicadores de desempenho. Certamente a utilização de simples palpites, experiências de tentativa e erro, intuição ou *softwares* de funcionamento invisível para o usuário não representam a melhor alternativa para os gerentes das centrais de empresas que querem ser bem sucedidas (MEHROTRA; PROFOZICH; BAPAT, 1997).

Considerando-se que o custo de pessoal representa aproximadamente 70% do total da indústria, a matéria-prima para o processo de gerenciamento eficiente (níveis satisfatórios de serviço a custos razoáveis) de um *call center* consiste na previsão da demanda de ligações e da consequente carga de trabalho, que precisam, então, ser estimadas da forma mais acurada possível. Essa necessidade confere à previsão de demanda de ligações o *status* de atividade crítica durante a busca pela eficiência e pela efetividade do negócio. Uma previsão acurada dos parâmetros do

sistema é um pré-requisito para um nível de serviço consistente, representando um dos passos mais importantes do processo como um todo (HALL; ANTON, 1998; ALAM, 2002; GANS; KOOLE; MANDELBAUM, 2003; REYNOLDS, 2005; BOUZADA, 2006; WEINBERG; BROWN; STROUD, 2006).

Em vista disso, esse trabalho procura descrever e atacar o problema de uma grande empresa brasileira de *call centers* que, apesar de dispor de uma série de informações potencialmente úteis para a tarefa de prever a demanda de ligações, não consegue cumprir essa tarefa de forma satisfatoriamente acurada, em parte por não incluir sistematicamente essas úteis informações no seu processo de previsão de demanda.

2 Revisão de literatura

Uma grande ênfase tem sido dada à melhoria do processo de tomada de decisão em organizações governamentais e empresariais. Em vista do ambiente altamente competitivo do atual mundo de negócios, tornou-se objetivo básico alocar de maneira otimizada os sempre escassos recursos disponíveis nas diversas atividades concorrentes. No caso dos *call centers*, este objetivo é viabilizado, em parte, por meio de um processo de previsão mais preciso. Nos anos 70, devido a limitações técnicas e principalmente de recursos computacionais, os administradores utilizavam frequentemente seus próprios sentimentos e intuições para auxiliar no processo de tomada de decisão. Hoje eles estão complementando esse *feeling* em relação à indústria ou à economia com a utilização de técnicas de previsão, tendência reforçada pela difusão do uso de microcomputadores no ambiente empresarial (HANKE, 1992).

Os supervisores dos *call centers*, responsáveis pelo seu planejamento operacional, precisam dimensionar eficientemente as suas equipes de modo a proporcionar níveis de serviço satisfatórios a custos razoáveis. A administração apropriada das centrais requer a estimação de algumas características operacionais que levam ao dimensionamento recomendado da equipe de atendimento; dentre estas características, destaca-se a demanda de ligações de clientes, questão a ser permanentemente abordada neste tipo de negócio (WEINBERG; BROWN; STROUD, 2006; KLUNGLE; MALUCHNIK, 1997).

Segundo Matan e Nourbakhsh (1998), para gerar previsões acuradas dos volumes futuros de chamadas, os gerentes dos *call centers* precisam juntar informações de várias fontes, desde previsões do mercado até, eventualmente, a duração dos intervalos comerciais veiculados na mídia. No entanto, parte importante das informações provém das estatísticas do histórico de chamadas que, usualmente, podem ser ajustadas para as tendências atuais ou esperadas.

De acordo com Mehrotra (1997), os modelos de séries temporais têm sido a abordagem mais utilizada para o

tratamento da previsão de chamadas telefônicas. No entanto, outros modelos de previsão foram aplicados com sucesso em alguns ambientes de *call center*, como, por exemplo, centrais de vendas baseando sua previsão de chamadas nas remessas de produtos recebidas e nas propagandas enviadas aos potenciais clientes.

A atividade de previsão é inerentemente difícil no caso dos *call centers* por causa do tamanho relativamente pequeno dos intervalos de tempo utilizados na prática. Técnicas de previsão estabelecidas, como o alisamento exponencial de Winters e a regressão linear múltipla, são úteis para este tipo de negócio. Geralmente, usar uma abordagem de previsão apropriada irá reduzir pela metade os erros de previsão. No entanto, muitos *call centers* encontram dificuldades nesse processo por conta do conhecimento técnico necessário para adequar estas técnicas aos complexos padrões de chamadas e ao aspecto pouco organizado dos dados (GROSSMAN et al., 2001).

A previsão do volume de chamadas é um importante aspecto das operações em *call centers*, sendo necessária para a previsão de carga de trabalho, alocação de mão de obra e planejamento de capacidade. Em geral, existem algumas regularidades nos dados de chegada das chamadas, como: a dependência entre os dias, ou seja, a correlação entre os volumes de dias consecutivos e possíveis ciclos semanais, mensais e até anuais nos volumes de ligações; e a dependência intradiária, ou seja, a correlação entre chegadas em diferentes períodos de tempo – manhã, tarde ou noite – dentro do mesmo dia (SHEN; HUANG, 2005).

O processo de previsão consiste tanto em arte quanto em ciência. Em arte, porque o futuro, afinal de contas, está sendo previsto e a acurácia da previsão vai depender em parte do julgamento e da experiência de quem estiver conduzindo o processo. Mas também em ciência: um processo matemático, passo a passo, que utiliza a história passada para prever eventos futuros. Assim, um entendimento de técnicas estatísticas especializadas é necessário para o processo ser bem feito. Os administradores que possuem *softwares* de gerenciamento de força de trabalho, que automatizam o processo de previsão, não podem pensar que isso apenas é suficiente; entender estes cálculos é tão crítico quanto possuir o *software*: não só para verificar a acurácia dos resultados como também, e talvez de forma mais importante, para explicar os números para a gerência. Ou seja, mesmo tendo as ferramentas, vale a pena entender os fundamentos do processo de previsão (REYNOLDS, 2005; BOUZADA, 2006).

Weinberg, Brown e Stroud (2006) propõem, em seu artigo, um modelo multiplicativo para prever taxas intradiárias de chegadas de chamadas ao *call center* de um banco comercial norte-americano. Adicionalmente, qualquer conhecimento anterior baseado em experiência passada do gerente pode ser incorporado ao modelo, assim como

componentes observados somente *a posteriori*. Comparando as previsões para o dia seguinte fornecidas pelo modelo dos autores e por modelos estatísticos clássicos, eles puderam perceber melhorias de até 25% por parte do seu modelo em relação aos padrões.

Em geral, se a variável a ser predita está fortemente correlacionada com outras variáveis que estão sujeitas a grandes variações, um modelo multivariado é necessário (SILVER, 2000).

3 Descrição do caso

3.1 A empresa

A Contax surgiu no final de 2000, como extensão natural dos negócios da Telemar, em um setor da economia que pouco investia em tecnologia e qualificação de atendimento, para ajudar seus clientes na gestão operacional do serviço de atendimento, agregando valor no relacionamento com os consumidores finais (CONTAX, 2006).

No Brasil, atualmente, é a empresa de maior crescimento nessa indústria, tendo crescido quase 60% em 2005, com um faturamento de R\$ 1.129 milhões. Caracteriza-se como a maior empresa do ramo em número de posições de atendimento e como a segunda maior em termos de faturamento e número de funcionários, dentro do território nacional (OUTSOURCING, 2005).

A firma faz mais de 100 milhões de contatos por mês, por telefone, correio, *e-mail*, torpedos, Internet e *chat*. São mais de 22 mil posições de atendimento e quase 50 mil funcionários distribuídos nas 16 centrais situadas em dez cidades brasileiras: Brasília, Salvador, Recife, Fortaleza, Porto Alegre, Belo Horizonte, Nova Lima (MG), Rio de Janeiro, Niterói (RJ) e São Paulo. Esta última cidade abriga o mais moderno complexo de *contact centers* da América Latina (CONTAX, 2006).

A Contax é uma empresa de capital 100% nacional e hoje opera com mais de 40 clientes, sendo o principal a própria Telemar, que responde por aproximadamente 60% do faturamento. Os principais produtos relacionados a este cliente são:

- 102 – que recebe ligações dos clientes que buscam auxílio à lista telefônica;
- Velox – que engloba o suporte técnico e o *help desk* para os clientes do serviço de Internet de banda larga da Telemar;
- 103 – que abrange clientes que buscam serviços relacionados à telefonia fixa, como mudança de endereço de cobrança e dúvidas ou problemas com a conta telefônica;
- Suporte técnico e reparo de defeitos da Telemar;
- OI – que abrange todo o atendimento para os clientes dos serviços de telefonia móvel da Telemar.

3.2 O processo atual para a previsão de demanda

Na opinião da equipe de operação da empresa, a busca pela satisfação do cliente ainda é realizada de forma ineficiente. De fato, há um consenso de que os custos oriundos dessa busca de satisfação ainda estão altos, apresentando grande potencial de redução. O setor responsável por melhorar a eficiência operacional da empresa é o de Planejamento de Tráfego.

A principal ferramenta que este setor utiliza para auxiliar as decisões gerenciais de planejamento operacional é o *software* TotalView, totalmente integrado à infraestrutura de telefonia. O produto consiste em uma ferramenta de força de trabalho utilizada para dimensionar e controlar o atendimento, que auxilia os gerentes – entre outras atividades – na previsão de demanda (por horário e dia da semana).

De acordo com o gerente de Planejamento de Tráfego da empresa, a principal preocupação consiste em conseguir melhorar as previsões de demanda de ligações. A acurácia das previsões não é boa para alguns produtos, e ele próprio avalia que – principalmente para o produto 103 – a empresa dispõe de todas as informações que poderiam fornecer uma melhor previsão, estando apenas carente de um método mais adequado. Essas informações são basicamente as quantidades de contas enviadas para os clientes em cada data, normalmente 5-6 dias antes do vencimento, e as suas próprias datas de vencimento. A impressão da gerência e da coordenação do produto 103 é que a demanda de ligações para o produto é influenciada pelos eventos de chegada da conta e pelo seu próprio vencimento, datas em que haveria maior acesso ao serviço por parte dos clientes. O problema é que o sistema utilizado para previsão de demanda não leva em conta essas informações.

A magnitude do erro absoluto percentual médio, segundo a ótica diária, incorrido durante o processo de previsão de demanda para o produto 103 era da ordem de 5%, em meados de 2006. Na opinião dos responsáveis pela previsão de demanda, esse indicador de erro encontra-se em um patamar baixo, sendo considerado mais do que aceitável. No entanto, estes responsáveis entendem que existe um potencial de melhoria no processo, de forma que esses erros podem se tornar ainda menores.

O primeiro passo do processo de previsão de demanda para o produto em questão consiste na geração de uma previsão básica, realizada pelo *software* de gerenciamento de força de trabalho, o TotalView. Um dos seus módulos é o de *forecast*, que trabalha com o histórico de demanda e faz a previsão de ligações baseado nas 13 últimas semanas, calculando uma média ponderada, com pesos diferenciados, sendo maiores para as semanas mais recentes e menores para as mais remotas. A previsão é gerada para cada intervalo ou bloco de meia hora, levando em consideração também a sazonalidade do dia da semana. Assim, por exemplo, para calcular a demanda de ligações da próxima terça-feira, no período de 10h30-11h, o TotalView considera

o histórico de demanda desse mesmo intervalo de meia hora, calculando uma média ponderada da demanda das últimas 13 terças-feiras.

O segundo passo consiste em uma crítica, por parte dos analistas da coordenação, a esses números gerados pelo TotalView. O *software* busca o histórico do dia em questão nas últimas 13 semanas, mas se nesse histórico houver alguma informação contaminada por algum evento específico, como, por exemplo, um jogo da Copa do Mundo que acarreta redução na quantidade de ligações, uma limpeza deve ser feita pelo analista. Este deve fazer uso do seu bom senso e poder analítico para, se perceber algum desvio grande no histórico, realizar o ajuste manualmente, modificando ou expurgando do histórico o dado discrepante, e gerar novamente a previsão de forma pontual. Este acompanhamento acontece diariamente.

Depois da geração da prévia da previsão por meio do número inicial do TotalView e da crítica inicial da equipe, o terceiro passo consiste em incorporar aos valores previstos o possível impacto causado por alguns eventos específicos, como a segunda-feira de carnaval ou uma ocorrência diferente em algum sábado, por exemplo. A maneira de incorporar tais impactos depende do evento ser novo ou desconhecido e do prévio conhecimento das suas consequências.

Alguns dos eventos especiais podem ter seu impacto previsto de uma forma melhor, quando a equipe de analistas é subsidiada por algumas premissas recebidas de outros setores capazes de impactar a demanda de chamadas, como: previsão de crescimento da base de clientes; implantação de novos planos e serviços na telefonia fixa, como o plano de minutos, por exemplo.

Essas premissas também envolvem o comportamento dos ciclos das contas telefônicas dos clientes, uma informação oriunda do setor de faturamento da Telemar e que ajuda a previsão de demanda do produto 103. Essa informação consiste, basicamente, nas quantidades de contas enviadas para os clientes em cada data, normalmente 5-6 dias antes do vencimento. Ela é útil porque a demanda de ligações para o produto é influenciada, segundo a equipe de previsão, pelos eventos de chegada da conta no destino e seu vencimento, datas em que haveria maior acesso ao serviço por parte dos clientes, para fins de esclarecimento e/ou reclamação.

Na verdade, durante todo o período de 4-5 dias em que a conta está com o cliente (desde a chegada até o vencimento), ocorre um acréscimo no volume de ligações, mas o maior impacto acontece mesmo no momento da chegada da conta na residência do cliente, na opinião da coordenação responsável pelo produto. Segundo os analistas, o impacto pós-vencimento existe quando o cliente que esqueceu de pagar a conta liga para saber como proceder, mas é pequeno, não sendo significativo em termos de aumento de demanda.

O problema é que o TotalView não leva em conta isso; a sazonalidade do dia da semana é considerada, mas não a distância do dia em relação à chegada da conta ou ao seu

vencimento. Por exemplo, a conta de um determinado cliente vence no dia 10 de cada mês, e não na segunda quarta-feira; então, no mês em que o dia 10 caísse em uma terça-feira, esse dia da semana deveria ser mais impactado do que os outros, enquanto que no mês em que o dia 10 caísse em uma quarta-feira, o impacto maior deveria ser incluído neste dia. Mas o TotalView considera, em termos de sazonalidade, todas as terças-feiras da mesma forma, seguindo o mesmo raciocínio para os outros dias da semana. Além disso, há meses em que as contas não são enviadas exatamente 5 dias antes do vencimento, e a chegada da conta ao destino também deve ser considerada como um evento especial.

Essas informações são levadas em conta no processo de previsão, mas não de forma automática. Toda a análise é feita pela equipe de previsão, e não pelo *software*, que:

- recebe e avalia o cronograma da área de faturamento, com a quantidade de contas sendo postadas e vencendo em cada data;
- verifica no histórico o impacto dos eventos de chegada e vencimento das contas, para o ciclo, ou seja, o grupamento de contas em cada filial, em questão;
- acrescenta, por fora do TotalView, o impacto histórico desses eventos, como, por exemplo, 5% e 3%, respectivamente, ao volume inicialmente previsto pela ferramenta;
- critica os resultados;
- retroalimenta o *software*, ajustando a previsão que havia sido previamente gerada.

Durante esse processo, os analistas não levam em conta um possível impacto da véspera e da antevéspera do vencimento da conta e nem do período pós-vencimento; apenas o efeito dos eventos chegada e vencimento.

A previsão é elaborada mensalmente, sempre com 45 dias de antecedência em relação à necessidade de contratação de operadores. Este prazo se justifica, para este produto, por causa do processo de validação com o cliente, do tempo de seleção do novo funcionário e dos 30 dias necessários para o treinamento do operador nas suas futuras atividades.

Fisicamente, existem centrais de atendimento do 103 no Rio de Janeiro, em Minas Gerais e no Ceará. Mas, em termos organizacionais, tudo funciona como se houvesse uma central única de atendimento, abrangendo potencialmente os clientes de todos os estados. Já que a demanda se superpõe para todas as centrais, torna-se necessário apenas elaborar uma previsão consolidada para todo o Brasil, como se houvesse uma central única de atendimento.

3.3 O processo sugerido para a previsão de demanda

3.3.1 Proposição do processo

Atualmente, a previsão básica é extraída do TotalView e a coordenação realiza toda essa análise no Excel. A opinião

geral é que as informações do faturamento poderiam ser usadas de uma forma mais sistemática para poder ser construído um método de previsão mais acurado, levando em conta todos esses aspectos e não apenas a sazonalidade do dia da semana, como faz o TotalView.

A ideia é fugir mais dessa pós-análise não metódica no Excel – que depende muito da intuição do analista – e poder contar com uma ferramenta que apoie a equipe neste momento, usando as informações do faturamento no processo de elaboração das previsões, para que se possa evitar o trabalho manual, que aumenta a possibilidade de erro. A ferramenta não tem a pretensão de substituir a intuição e experiência da equipe de previsão, mas servir de apoio à sua tomada de decisão. Ou seja, a opinião dos especialistas não deve ser descartada; ela deve, sim, ser utilizada para enriquecer o modelo que busca, basicamente, objetivar as informações subjetivas que são relevantes para o processo de previsão.

Operacionalmente, a equipe de previsão coletaria, com a antecedência usual, os *inputs* necessários oriundos do setor de faturamento e alimentaria a ferramenta de previsão, ambientada em uma planilha de Excel, com eles. O *output* da ferramenta consistiria na quantidade de ligações demandada para cada data futura.

O método escolhido foi o da Regressão Múltipla, pois é capaz de capturar o impacto específico de cada evento e informação na demanda de cada dia. Por exemplo, qual o impacto, na demanda de ligações, do fato do dia em questão cair em uma quinta-feira? Qual o impacto do dia consistir em um feriado? E de chegarem 200 mil contas para os clientes no dia em questão? E do dia ser a véspera do vencimento de 500 mil contas?

A ideia é verificar a distância do dia específico em relação aos eventos especiais, a chegada e o vencimento das contas, e calcular para quantas contas esse dia é o C (dia de chegada), o V (dia do vencimento), o $V - 1$ (véspera do vencimento), o $V - 2$ (antevéspera do vencimento), ... , $V + 1$ (dia seguinte ao vencimento), $V + 2$,

3.3.2 Aplicação do modelo

De posse do cronograma de faturamento fornecido pela equipe de previsão, para o período de julho de 2005 a julho de 2006, foi possível consolidar as informações em termos da data de chegada e data de vencimento das contas, conforme apresentado parcialmente nas Tabelas 1 e 2, respectivamente.

Para poder ser desenvolvido o modelo de Regressão Múltipla para prever a quantidade de ligações em cada data, foi preciso ter acesso ao histórico desta grandeza. Por motivo de sigilo estratégico, esses valores não puderam ser revelados, em sua plena magnitude; no entanto, a eles foi aplicada uma constante de proporcionalidade, e os resultados, de julho de 2005 até março de 2006, fornecidos pela empresa, estão apresentados graficamente na Figura 1.

Como se observa, a série não apresenta tendência de crescimento ou decréscimo, mas apenas um comportamento errático, dispensando, assim, a necessidade de inclusão de uma variável associada à passagem do tempo.

De posse de todas essas informações foi possível montar uma planilha com os dados necessários para a construção do modelo de Regressão Múltipla dentro do qual o comportamento da variável dependente quantidade de ligações em um dia específico poderia estar sendo explicado – inicialmente – por 10 variáveis independentes:

Tabela 1. Contas agrupadas por data de chegada prevista, setembro a outubro de 2005.

Data de chegada prevista	Total
06/9/05	3.360.569
13/9/05	801.143
14/9/05	432.670
15/9/05	709.437
26/9/05	1.315.342
27/9/05	820.855
28/9/05	1.555.097
29/9/05	2.506.038
04/10/05	1.808.735
05/10/05	1.167.447
07/10/05	800.105
11/10/05	1.890.364
18/10/05	709.437
26/10/05	1.731.518
28/10/05	587.118
Total	20.195.875

Tabela 2. Contas agrupadas por data de vencimento, setembro a outubro de 2005.

Vencimento	Total
04/9/05	2.142.215
07/9/05	1.918.920
11/9/05	2.177.552
13/9/05	798.630
15/9/05	1.456.656
17/9/05	801.143
20/9/05	432.670
23/9/05	79.437
01/10/05	2.136.197
04/10/05	2.142.215
07/10/05	1.918.920
11/10/05	2.177.552
13/10/05	798.630
15/10/05	1.456.656
17/10/05	801.143
20/10/05	432.670
23/10/05	79.437
Total	23.010.643

- dia da semana;
- feriado – se o dia específico consiste em um feriado ou não;
- chegada (C) – quantas contas estão previstas para chegarem ao cliente no dia específico;
- vencimento (V) – quantas contas vencem no dia específico;
- V – 3 – para quantas contas aquele dia específico corresponde a 3 dias antes do vencimento, ou, em outras palavras, quantas contas irão vencer 3 dias depois da data específica;
- V – 2 – para quantas contas aquele dia específico corresponde à antevéspera do vencimento;
- V – 1 – para quantas contas aquele dia específico corresponde à véspera do vencimento;
- V + 1 – para quantas contas aquele dia específico corresponde a 1 dia depois do vencimento, ou, em outras palavras, quantas contas terão vencido na véspera daquela data específica;
- V + 2 – para quantas contas aquele dia específico corresponde a 2 dias depois do vencimento;
- V + 3 – para quantas contas aquele dia específico corresponde a 3 dias depois do vencimento.

A Tabela 3 a seguir apresenta, em parte, a planilha de dados em questão.

Cabe salientar que as variáveis independentes dia da semana e feriado foram trabalhadas como sendo do tipo *dummy* (0 = Não; 1 = Sim). Foi criada uma variável *dummy* para a ocorrência de feriado e seis variáveis *dummy* para o dia da semana: domingo, segunda-feira, terça-feira, quarta-feira, quinta-feira e sexta-feira.

Dessa forma, para o dia 28/2/2006, por exemplo, a variável *dummy* feriado assumiu o valor 1, por se tratar da terça-feira de Carnaval; as variáveis *dummy* domingo, segunda-feira, quarta-feira, quinta-feira e sexta-feira assumiram o valor 0 enquanto que a variável *dummy* terça-feira assumiu o valor 1. Seguindo o mesmo raciocínio, para o dia 4/2/2006, por exemplo, a variável *dummy* feriado assumiu o valor 0; e as variáveis *dummy* domingo, segunda-feira, terça-feira, quarta-feira, quinta-feira e sexta-feira assumiram o valor 0, caracterizando, portanto e por exclusão, o dia em questão como sendo um sábado.

Após rodar a regressão, a primeira análise aponta para um modelo com um potencial de explicação muito bom, já que o R-quadrado ajustado indica que 84% da variação dos dados pode ser explicada. Neste caso, pode-se inferir a real influência das variáveis explicativas na quantidade de ligações e aceitar o modelo com elevada confiança, já que o nível de significância da estatística F deu igual a 0,0000.

A contribuição de cada variável independente é expressa pelos coeficientes estimados, juntamente com os respectivos *p-values* (ou valores p), apresentados na Tabela 4.

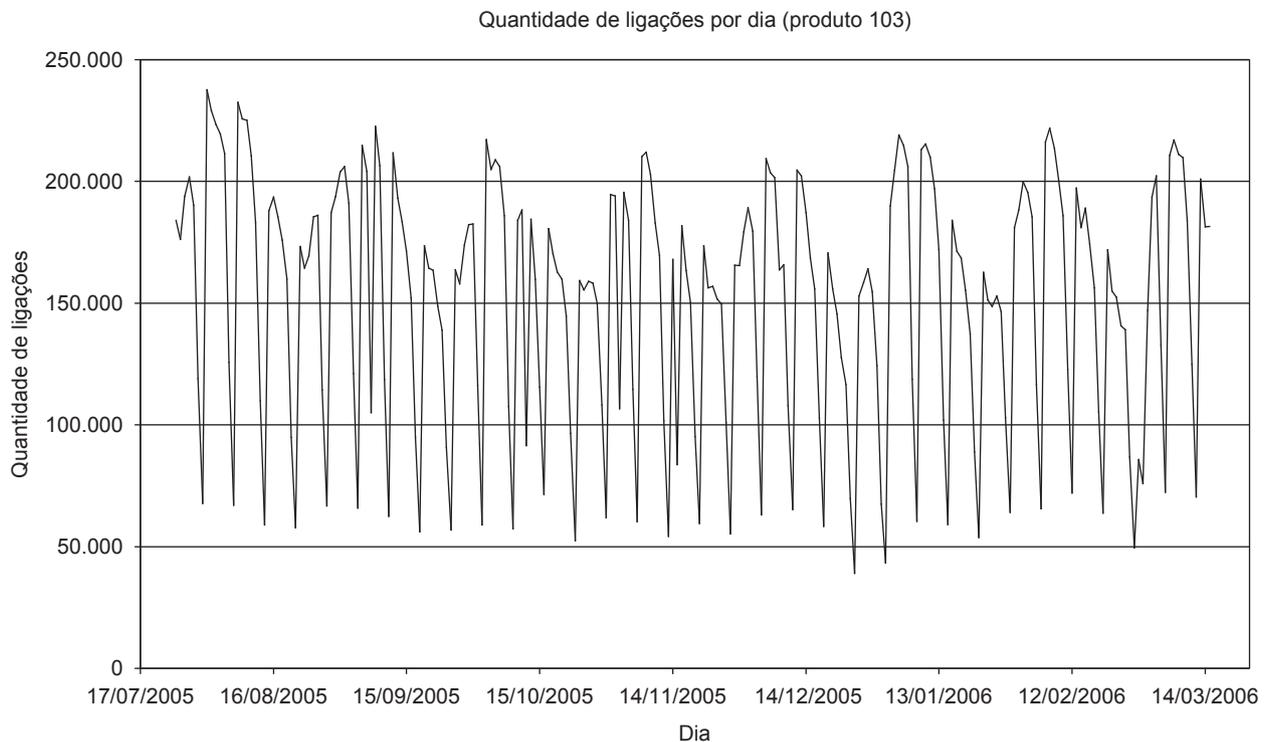


Figura 1. Quantidade de ligações por dia, julho de 2005 a março de 2006.

Na Tabela 4, observa-se um elevado valor p para a variável $V + 3$, indicando que a quantidade de contas para as quais aquele dia específico corresponde a 3 dias depois do vencimento não pode ser vista como um bom predictor para a quantidade de ligações na data em questão.

Assim sendo, o efeito pós-vencimento parece, então, acabar já no segundo dia após o vencimento da conta, já que os valores p de $V + 1$ e $V + 2$ são baixos; em outras palavras, os clientes ainda ligam para o 103 um ou dois dias após o vencimento da conta, mas esse impacto desaparece a partir do terceiro dia, deixando de ser significativo.

A variável $V + 3$ pode estar contaminando o modelo e convém retirá-la da análise. Sem ela, um novo modelo de regressão foi gerado, continuando com um potencial muito bom e até ligeiramente melhor, já que 85% da variação dos dados pode ser explicada por ele.

Os coeficientes de regressão estimados, juntamente com os respectivos valores p , estão apresentados na Tabela 5.

Como pode ser observado, os valores p das outras variáveis mudaram muito pouco e, para este modelo modificado, apenas duas variáveis apresentam coeficientes angulares com valores p superiores a 5%: Chegada (C) e $V - 2$.

Em relação a esta última variável, ela será mantida no modelo, porque o valor p do seu coeficiente está abaixo de um valor também bastante utilizado como nível de significância nos testes de hipóteses (10%) e pelo fato das variáveis $V - 3$ e $V - 1$ apresentarem-se também como bons predictores. Ora, não faria muito sentido manter no modelo $V - 3$, retirar $V - 2$ e manter $V - 1$, pois a continuidade estaria sendo quebrada; seria o mesmo que afirmar que os

clientes ligam para o 103 três dias antes do vencimento e na sua véspera, mas não na antevéspera!

Já a variável Chegada (C) também será mantida no modelo, apesar de apresentar um valor p maior do que 10%, pelo fato de ter sido identificada, na opinião subjetiva e calcada na experiência da equipe de planejamento, como crítica e muito impactante na demanda de ligações, contrariando o resultado objetivo que indica um coeficiente não significativo para essa variável; também é válido ressaltar que seu valor p (14,30%) não ficou tão acima de um nível de significância considerado bom (5%).

Essa intervenção subjetiva em um modelo tão objetivo traz à tona a já mencionada importância da utilização da opinião de especialistas na construção de um processo de previsão, mesmo nos casos em que esse é essencialmente sistemático.

Desta forma, a retirada da variável $V + 3$ foi benéfica ao modelo, tornando-o mais parcimonioso e permitindo uma interpretação dos coeficientes obtidos.

O primeiro coeficiente corresponde ao valor da variável dependente quantidade de ligações quando todas as variáveis independentes assumirem o valor zero; ou seja, quando domingo, segunda-feira, terça-feira, quarta-feira, quinta-feira, sexta-feira, feriado, C, $V - 3$, $V - 2$, $V - 1$, V, $V + 1$ e $V + 2$ forem zero – ou ainda, quando o dia for sábado, não feriado e não consistir da data de chegada, nem de nenhuma data dentro do intervalo de 6 dias em torno do vencimento de alguma conta, a quantidade de ligações demandadas prevista é de 92.693.

Tabela 3. Planilha de dados do modelo de Regressão Múltipla para prever a quantidade de ligações, fevereiro de 2006.

Dia	Dia da semana	Feriado?	Chegada (C)	V - 3	V - 2	V - 1	Vencimento (V)	V + 1	V + 2	V + 3	Ligações
1/2/2006	4a. feira	Não	2.506.038	2.142.215	-	-	2.136.197	-	-	-	199.908
2/2/2006	5a. feira	Não	-	-	2.142.215	-	-	2.136.197	-	-	195.497
3/2/2006	6a. feira	Não	1.122.892	-	-	2.142.215	-	-	2.136.197	-	185.542
4/2/2006	Sábado	Não	-	1.918.920	-	-	2.142.215	-	-	2.136.197	116.657
5/2/2006	Domingo	Não	-	-	1.918.920	-	-	2.142.215	-	-	65.558
6/2/2006	2a. feira	Não	587.354	-	-	1.918.920	-	-	2.142.215	-	216.104
7/2/2006	3a. feira	Não	1.484.473	-	-	-	1.918.920	-	-	2.142.215	221.831
8/2/2006	4a. feira	Não	1.238.119	2.177.552	-	-	-	1.918.920	-	-	213.584
9/2/2006	5a. feira	Não	801.143	-	2.177.552	-	-	-	1.918.920	-	200.359
10/2/2006	6a. feira	Não	-	798.630	-	2.177.552	-	-	-	1.918.920	185.882
11/2/2006	Sábado	Não	-	-	798.630	-	2.177.552	-	-	-	122.764
12/2/2006	Domingo	Não	-	1.456.656	-	798.630	-	2.177.552	-	-	72.033
13/2/2006	2a. feira	Não	-	-	1.456.656	-	798.630	-	2.177.552	-	197.231
14/2/2006	3a. feira	Não	432.670	801.143	-	1.456.656	-	798.630	-	2.177.552	181.092
15/2/2006	4a. feira	Não	709.437	-	801.143	-	1.456.656	-	798.630	-	188.990
16/2/2006	5a. feira	Não	-	-	-	801.143	-	1.456.656	-	798.630	174.253
17/2/2006	6a. feira	Não	-	432.670	-	-	801.143	-	1.456.656	-	156.317
18/2/2006	Sábado	Não	-	-	432.670	-	-	801.143	-	1.456.656	105.306
19/2/2006	Domingo	Não	-	709.437	-	432.670	-	-	801.143	-	63.827
20/2/2006	2a. feira	Não	-	-	-	-	432.670	-	-	801.143	171.921
21/2/2006	3a. feira	Não	-	-	709.437	-	-	432.670	-	-	155.035
22/2/2006	4a. feira	Não	2.136.197	-	-	709.437	-	-	432.670	-	152.537
23/2/2006	5a. feira	Não	1.555.097	-	-	-	709.437	-	-	432.670	140.621
24/2/2006	6a. feira	Não	2.506.038	-	-	-	-	709.437	-	-	139.056
25/2/2006	Sábado	Não	-	-	-	-	-	-	709.437	-	86.746
26/2/2006	Domingo	Não	-	2.136.197	-	-	-	-	-	709.437	49.520
27/2/2006	2a. feira	Sim	-	-	-	-	-	-	-	-	85.790
28/2/2006	3a. feira	Sim	-	-	-	2.136.197	-	-	-	-	75.801

Esse número pode ser encarado como uma quantidade básica, à qual devem ser adicionados os efeitos da ocorrência das variáveis independentes, quantificados pelos seus respectivos coeficientes angulares.

Assim, o fato do dia ser domingo contribui com uma diminuição de 41.103 ligações previstas em relação ao número básico (92.693); a segunda-feira espera, segundo o modelo, 79.874 ligações a mais do que o dia básico (sábado); a terça-feira espera 75.900 ligações a mais do que o sábado; a quarta-feira espera 73.908 ligações a mais; a quinta-feira, 73.082 a mais; e a sexta-feira, 61.898 ligações a mais do que o dia básico.

Com essas informações, é possível elaborar um gráfico – apresentado na Figura 2 – que permite uma melhor visualização da sazonalidade do dia da semana.

A evolução da demanda de ligações ao longo da semana, considerando dias sem a ocorrência de feriados e não próximos da chegada ou vencimento de contas, começa com um valor bem baixo no domingo, apresentando um grande aumento na segunda-feira; a demanda vai caindo suavemente ao longo da semana, tendo uma queda um pouco mais brusca na sexta-feira e bem mais intensa no sábado, antes de voltar ao patamar de domingo.

De forma análoga aos dias da semana, o fato de um dia ser feriado reduz em 71.345 a quantidade prevista de ligações; e cada conta telefônica prevista para chegar à casa do cliente naquele dia aumenta em 0,0034 a previsão da quantidade de ligações demandadas para o dia em questão. Assim se, por exemplo, 100.000 contas estiverem chegando em um dia específico, é de se esperar que isso implique em um aumento de $100.000 \times 0,0034 = 340$ ligações no contingente diário. Se esse número for examinado de forma percentual, é possível concluir que, em média, 0,34% (= 0,0034) dos clientes ligam para o Telemar no dia em que recebem a sua conta telefônica.

Em relação aos dias que circundam a data do vencimento, é análogo o raciocínio que explica as consequências: cada conta vencendo 3 dias depois, ou seja, para as quais a data em questão representa $V - 3$, aumenta em 0,0052 a quantidade de ligações; cada conta vencendo 2 dias depois aumenta em 0,0042 a quantidade de ligações; cada conta vencendo no dia seguinte aumenta em 0,0072 essa quantidade; cada conta vencendo no dia em questão aumenta em 0,0077 a mesma quantidade; cada conta vencendo na véspera aumenta em 0,0071 a quantidade; e cada conta vencendo na antevéspera aumenta em 0,0060 a quantidade prevista de ligações.

Naturalmente, os coeficientes não têm apenas significado isolado, mas também podem e devem ser analisados em conjunto: se as consequências dos efeitos dia da semana, ocorrência de feriado e data crítica em relação à chegada ou vencimento da conta forem incorporadas ao valor previsto para um dia básico, será possível estabelecer uma equação para prever a quantidade demandada de ligações em um determinado dia em função das características da data em

Tabela 4. Coeficientes estimados das variáveis independentes do modelo de Regressão Múltipla para prever a quantidade de ligações na data em questão.

	<i>Coeficientes</i>	<i>valor-P</i>
Interseção	92.545	0,00%
Domingo	(41.228)	0,00%
2a. Feira	79.882	0,00%
3a. Feira	75.694	0,00%
4a. Feira	73.812	0,00%
5a. Feira	73.068	0,00%
6a. Feira	61.763	0,00%
Feriado	(71.088)	0,00%
Chegada (C)	0,0035	13,57%
V - 3	0,0052	2,21%
V - 2	0,0042	7,81%
V - 1	0,0068	0,47%
Vencimento (V)	0,0074	0,19%
V + 1	0,0071	0,30%
V + 2	0,0064	0,75%
V + 3	0,0009	68,36%

Tabela 5. Coeficientes estimados das variáveis independentes do modelo (sem V+3) de Regressão Múltipla para prever a quantidade de ligações na data em questão.

	<i>Coeficientes</i>	<i>valor-P</i>
Interseção	92.693	0,00%
Domingo	(41.103)	0,00%
2a. Feira	79.874	0,00%
3a. Feira	75.900	0,00%
4a. Feira	73.908	0,00%
5a. Feira	73.082	0,00%
6a. Feira	61.898	0,00%
Feriado	(71.345)	0,00%
Chegada (C)	0,0034	14,30%
V - 3	0,0052	2,11%
V - 2	0,0042	7,86%
V - 1	0,0072	0,13%
Vencimento (V)	0,0077	0,06%
V + 1	0,0071	0,30%
V + 2	0,0060	0,71%

questão. Dessa maneira, a quantidade de ligações pode ser prevista conforme a Equação 1.

Equação do modelo para prever a quantidade de ligações:

$$\begin{aligned} \text{Quantidade prevista de ligações} = & 92.693 - 41.103 \\ & \times \text{domingo} + 79.874 \times \text{segunda-feira} + 75.900 \times \text{terça-} \\ & \text{feira} + 73.908 \times \text{quarta-feira} + 73.082 \times \text{quinta-feira} \\ & + 61.898 \times \text{sexta-feira} - 71.345 \times \text{feriado} + 0,0034 \times \\ & C + 0,0052 \times V-3 + 0,0042 \times V-2 + 0,0072 \times V-1 + \\ & 0,0077 \times V + 0,0071 \times V + 1 + 0,0060 \times V + 2 \end{aligned} \quad (1)$$

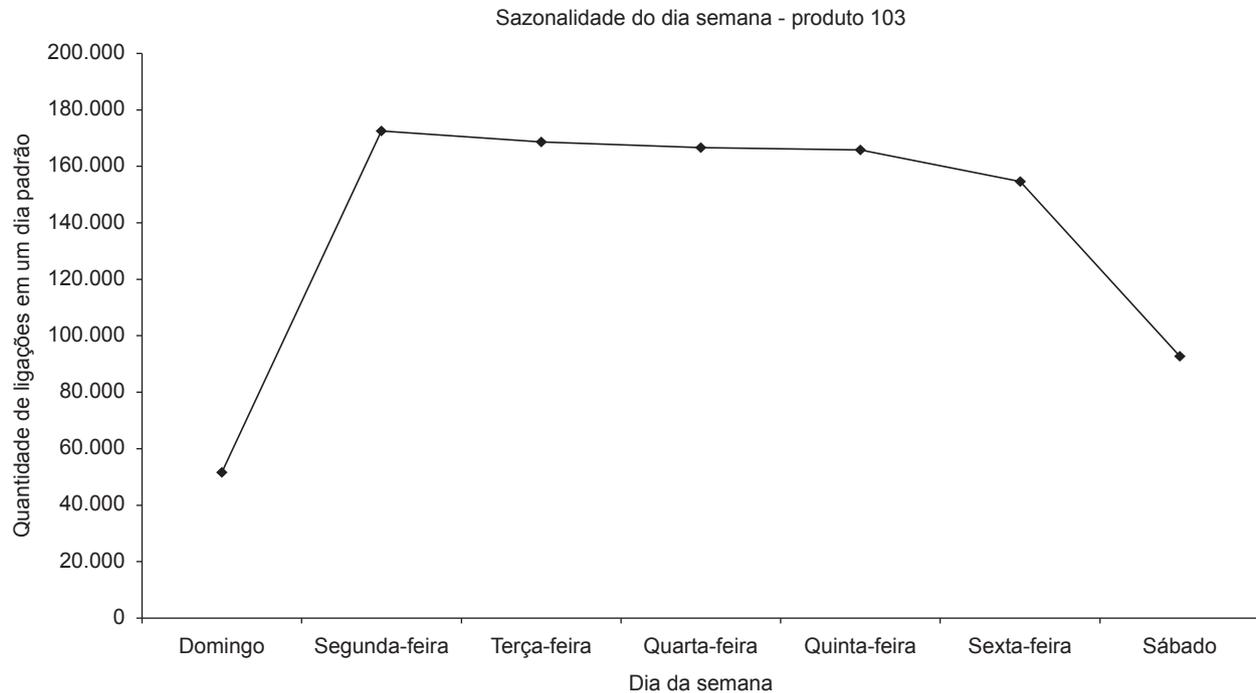


Figura 2. Comportamento sazonal do dia da semana – modelo sem $V + 3$.

- as sete primeiras variáveis são do tipo *dummy*, assumindo o valor 0 (não) ou 1 (sim);
- as sete últimas representam, respectivamente: a quantidade de contas chegando no dia; vencendo 3 dias depois; 2 dias depois; 1 dia depois; no próprio dia; 1 dia antes; e 2 dias antes.

3.3.3 Considerações finais sobre o modelo obtido

De posse dessa equação, então, é possível calcular quanto o modelo de regressão teria previsto para a quantidade de ligações em cada um dos dias acerca dos quais o histórico disponibiliza o valor real desta grandeza. Desta forma, pode-se comparar os valores que teriam sido previstos com os que efetivamente ocorreram, de maneira que um erro de previsão possa ser mensurado. Essa comparação pode ser visualizada na Figura 3 a seguir.

A diferença entre os valores reais e os gerados pelo modelo também é conhecida como resíduo. Para ser usado um modelo de Regressão Múltipla, é necessário supor que a correlação entre os resíduos, ao longo do espectro das variáveis independentes, é zero; ou seja, que os resíduos são independentes entre si, não apresentando, conseqüentemente, autocorrelação serial (CORRAR; PAULO; DIAS FILHO, 2007).

Segundo estes autores, uma forma de verificar a validade dessa suposição é por meio do teste de Durbin-Watson. No modelo que está sendo analisado, foi encontrado o valor 1,838 para a Estatística DW.

De acordo com Corrar, Paulo e Dias Filho (2007), essa estatística obtida deve ser comparada com valores críticos que variam em função do nível de significância do teste, do tamanho da amostra e do número de variáveis independentes.

Com 5% de significância, 234 dados na amostra e 14 variáveis independentes, o valor crítico inferior (d_L) obtido foi 1,661 e o valor crítico superior (d_U) encontrado foi 1,914 (Valores encontrados em tabela disponível no endereço <<http://www.stanford.edu/~clint/bench/dwcrit.htm>>. Acesso em: 24/1/2009).

Como o valor obtido para a Estatística DW (1,838) se situa entre os dois valores críticos encontrados, o teste para verificar a existência de autocorrelação dos resíduos é não conclusivo: não é possível garantir que os resíduos são independentes, mas também é incorreto afirmar que existe uma relação de dependência entre eles (CORRAR; PAULO; DIAS FILHO, 2007).

Sendo assim, parece válido examinar o gráfico do resíduo contra cada variável independente a fim de procurar uma suposta autocorrelação e a identificação de uma eventual lei de formação nos valores dos resíduos à medida que o valor da variável independente avança. A Figura 4 a seguir ilustra o caso da variável *dummy* quarta-feira.

Como pode ser observado, não parece existir uma lei de formação; ou seja, a magnitude dos resíduos não parece estar dependendo do valor da variável independente (quarta-feira).

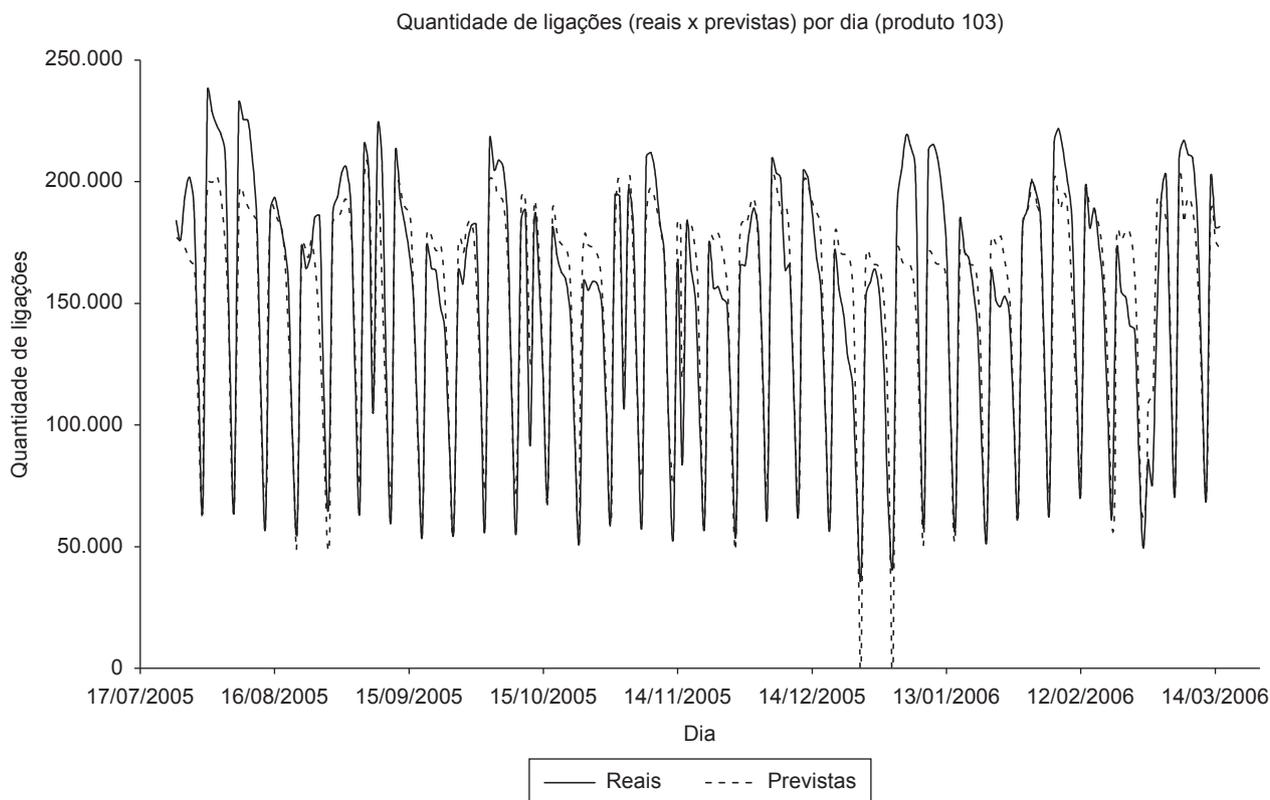


Figura 3. Valores reais multiplicados por uma constante e previstos pelo modelo para a quantidade de ligações por dia, julho de 2005 a março de 2006.

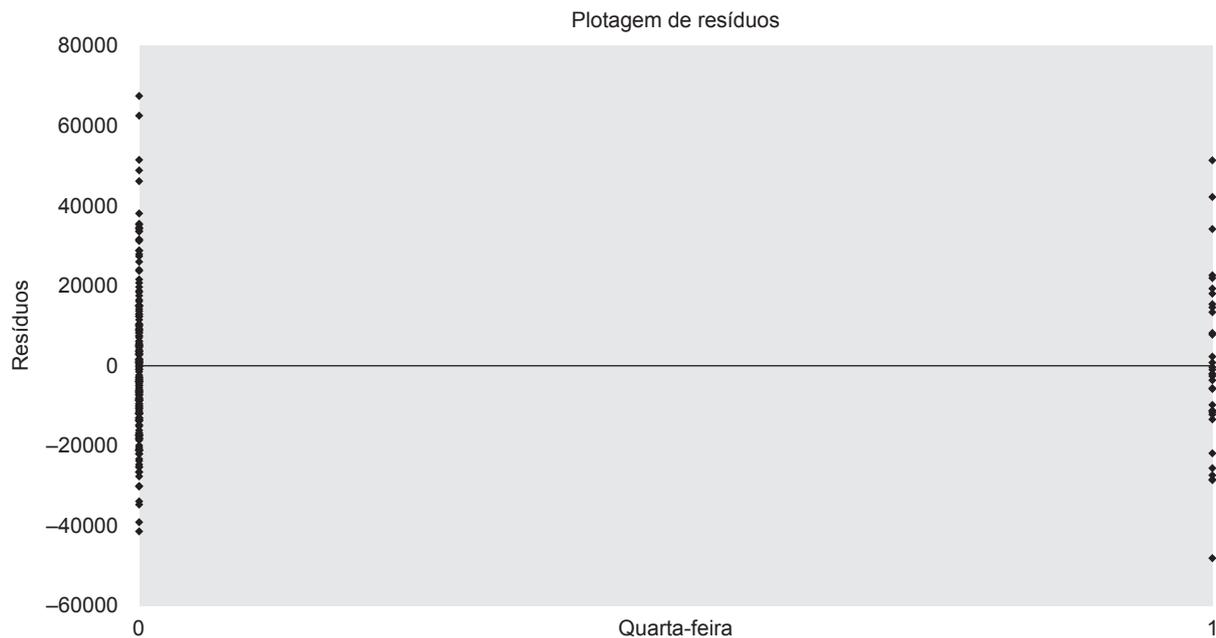


Figura 4. Resíduos do modelo \times variável dummy quarta-feira.

Os resíduos apresentaram comportamento bem parecido em relação às outras variáveis *dummy* (feriado e as referentes aos outros dias da semana). E já que o modelo trabalha com 14 variáveis independentes, seria um tanto inviável apresentar aqui tal quantidade de gráficos. Dessa forma, serão apresentadas graficamente apenas as relações entre os resíduos e as outras variáveis independentes mais marcantes: Chegada e Vencimento. Tais relações podem ser visualizadas nas Figuras 5 e 6, respectivamente, a seguir.

O comportamento dos resíduos em relação às outras variáveis (V-3, V-2, V-1, V+1 e V+2) é bastante similar ao verificado nas figuras 5 e 6: parecem estar distribuídos aleatoriamente em torno do eixo das abscissas, não apresentando nenhuma tendência.

Tudo isso gera uma segurança maior para os resultados do modelo, já que o teste para a verificação de autocorrelação serial dos resíduos foi não conclusivo. É claro que essa

última análise realizada é meramente visual, carecendo de um cientificismo maior. Mas é importante proceder dessa forma principalmente em situações em que os critérios objetivos apontam para uma indefinição, como foi o caso.

Já em relação à qualidade preditiva do modelo, foi possível perceber, na Figura 3 anterior, a menos de alguns pontos específicos, uma boa aderência visual dos valores previstos em relação aos dados reais. Esta aderência foi aqui formalizada pelo MAPE, o *Mean Absolute Percentual Error* ou erro médio percentual absoluto.

A título de ilustração, a Tabela 6 apresenta dados relativos ao mês de dezembro de 2005, onde constam a quantidade real de ligações, os valores que o modelo de regressão teria previsto para esta grandeza, assim como os erros percentuais absolutos oriundos desta previsão. Pode-se calcular que o modelo teria errado em 15,5%, em média, na previsão dos volumes diários de ligações durante o mês em questão.

Tabela 6. Valores reais, previstos e erros do modelo de Regressão Múltipla para prever a quantidade de ligações, dezembro de 2005.

Dia	Dia da semana	Feriado?	Ligações	Ligações (previstas pelo modelo)	Erro de previsão do modelo
1/12/2005	5a. feira	Não	189.179	193.288	2,2%
2/12/2005	6a. feira	Não	179.607	178.716	0,5%
3/12/2005	Sábado	Não	119.738	121.003	1,1%
4/12/2005	Domingo	Não	63.118	77.992	23,6%
5/12/2005	2a. feira	Não	209.398	203.208	3,0%
6/12/2005	3a. feira	Não	203.525	197.677	2,9%
7/12/2005	4a. feira	Não	201.497	186.681	7,4%
8/12/2005	5a. feira	Não	163.731	190.634	16,4%
9/12/2005	6a. feira	Não	165.708	175.358	5,8%
10/12/2005	Sábado	Não	108.033	112.476	4,1%
11/12/2005	Domingo	Não	65.321	71.680	9,7%
12/12/2005	2a. feira	Não	204.689	201.258	1,7%
13/12/2005	3a. feira	Não	202.289	196.754	2,7%
14/12/2005	4a. feira	Não	187.186	188.341	0,6%
15/12/2005	5a. feira	Não	168.579	185.651	10,1%
16/12/2005	6a. feira	Não	155.912	170.652	9,5%
17/12/2005	Sábado	Não	102.895	109.900	6,8%
18/12/2005	Domingo	Não	58.292	59.078	1,3%
19/12/2005	2a. feira	Não	170.636	180.520	5,8%
20/12/2005	3a. feira	Não	155.824	172.655	10,8%
21/12/2005	4a. feira	Não	145.547	170.260	17,0%
22/12/2005	5a. feira	Não	127.727	169.413	32,6%
23/12/2005	6a. feira	Não	116.546	155.684	33,6%
24/12/2005	Sábado	Não	69.839	93.699	34,2%
25/12/2005	Domingo	Sim	39.038	(18.895)	148,4%
26/12/2005	2a. feira	Não	152.909	172.567	12,9%
27/12/2005	3a. feira	Não	158.507	168.593	6,4%
28/12/2005	4a. feira	Não	164.109	166.601	1,5%
29/12/2005	5a. feira	Não	154.690	165.775	7,2%
30/12/2005	6a. feira	Não	124.406	154.591	24,3%
31/12/2005	Sábado	Não	67.541	92.693	37,2%

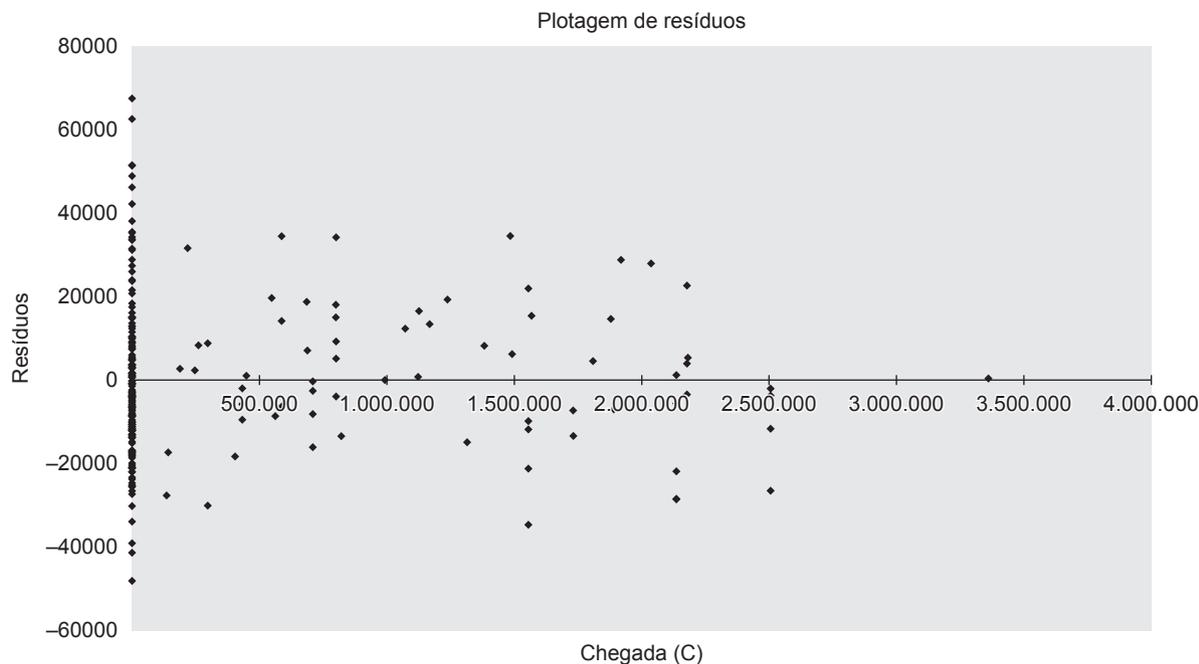


Figura 5. Resíduos do modelo × variável Chegada.

Mas dezembro é um mês atípico, com um comportamento geralmente mais difícil de prever. Tal afirmação pode ser verificada neste exemplo particular a partir da Figura 7 a seguir, no qual é possível perceber que todos os outros meses relacionados, com exceção de janeiro, outro período atípico, apresentam erros médios de previsão pelo menos 4 pontos percentuais menores do que o erro médio de dezembro de 2005.

Em média, o modelo está incorrendo num erro de previsão da ordem de 10,98%. Embora não possa ser considerado um valor elevado, ele é substancialmente maior que os 5% em que a Coordenação de Planejamento costuma incorrer na previsão de demanda do produto 103. No entanto, cabem aqui algumas observações.

É válido lembrar, por exemplo, que, durante o processo padrão de previsão, os analistas recebem os valores previstos pelo TotalView e acrescentam a eles os impactos esperados por eventos e premissas, como o comportamento do ciclo de contas e a ocorrência de feriados, assim como outros eventos especiais relacionados ao produto, como o plano de minutos, campanhas na mídia, crescimento de base e outros potenciais aspectos impactantes. Os 5% de erro de previsão só são conseguidos após a consideração de todas essas subjetividades e dentro de um processo diário de revisão da previsão.

Ora, o modelo de regressão proposto visa substituir apenas os valores sugeridos pelo TotalView, acrescidos da influência dos feriados e do comportamento dos ciclos de contas; nunca foi a ideia substituir o papel do analista encarregado da previsão – dotado de toda a sua valiosa

experiência e subjetividade – mas sim de fornecer-lhe um subsídio complementar para a tomada de decisão. Assim, e já que o analista estará apto a acrescentar aos valores sugeridos pelo modelo de regressão os impactos que julgar relevantes, é mais justo comparar os erros de previsão do modelo aqui proposto com os erros oriundos do módulo de previsão do TotalView, com o intuito de verificar a ocorrência ou não de melhorias no processo de previsão.

Assim sendo, o MAPE do TotalView para o mesmo período analisado foi de 13,54%, superior, portanto, aos 10,98% do modelo de regressão.

Além disso, uma análise mais detalhada dos erros de previsão do modelo proposto revela que alguns dos dias mais difíceis de prever e, portanto, com erros de previsão potencialmente maiores, consistem em feriados ou datas que se comportam quase como um feriado. Por exemplo, os dias 24/12/2005 e 31/12/2005 não foram classificados como feriados, mas, certamente também não são dias comuns e devem gerar uma demanda de ligações inferior ao de um dia padrão. O problema é que o modelo de regressão não levou em conta isso, já que era possível classificar um dia apenas como sendo feriado ou não; e já que esses dias não eram efetivamente feriados, foram tratados pelo modelo como dias comuns. Muito provavelmente por causa disso, suas demandas foram superestimadas e seus erros de previsão foram bem acima da média: 34,2% e 37,2%, respectivamente, conforme a Tabela 6 anterior.

Problemas semelhantes, mas talvez em menor escala, certamente ocorreram em alguns dias do ano que consistiram em feriados locais, mas não nacionais. Como o modelo

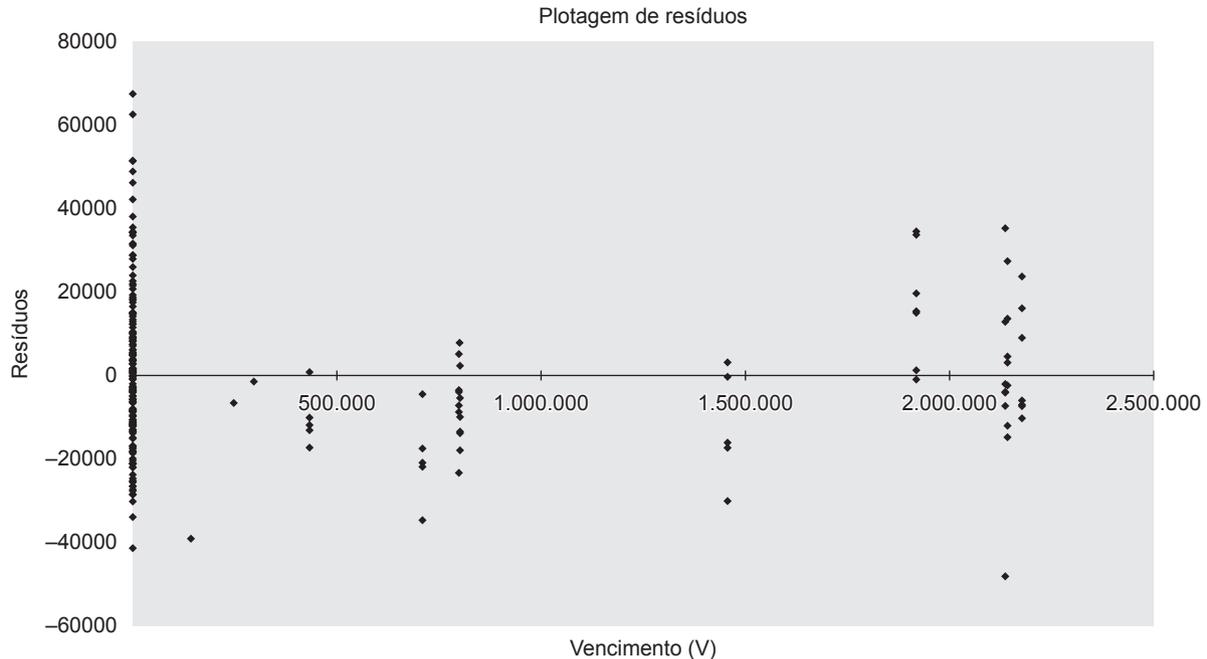


Figura 6. Resíduos do modelo \times variável Vencimento.

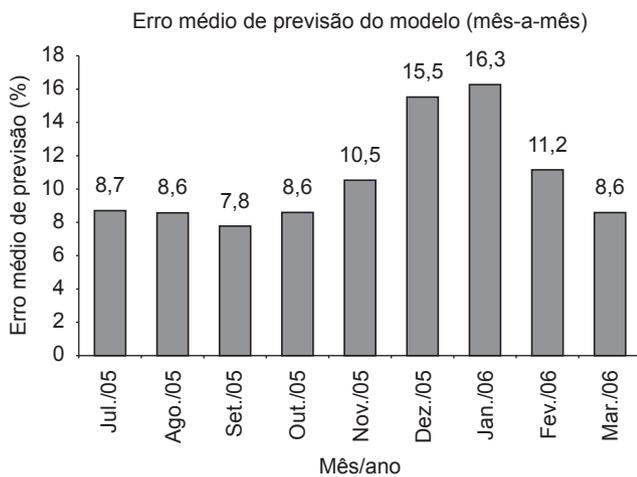


Figura 7. Erro médio mensal do modelo de Regressão Múltipla para prever a quantidade de ligações, julho de 2005 a março de 2006.

propõe uma previsão consolidada da demanda, só foi possível considerar os feriados nacionais; por exemplo, o dia 20/1/2006, um feriado local na cidade do Rio de Janeiro, teve uma redução na demanda nacional provocada pela redução na demanda da cidade mas, ao ser tratado como um dia normal pelo modelo, teve sua demanda superestimada. Esse problema também se observa nos feriados prolongados e nos dias que se situam entre um feriado e o fim de semana, em que a demanda foge do padrão.

Um problema ainda mais grave ocorreu com o dia 25/12/2005. Seu erro de previsão foi de quase 150%, já que a sua demanda prevista assumiu um valor negativo, - 18.895

ligações, conforme a Tabela. Isso aconteceu porque o dia em questão apresentava dois efeitos negativos e redutores de demanda, que foram somados pelo modelo de natureza aditiva: ser domingo e ser feriado. Mas, na prática, o efeito de um dia ser feriado certamente é reduzido se ele cair em um domingo ou sábado; e o modelo não levou isso em conta, até porque só é possível considerar um dia como sendo feriado ou não, e não como uma espécie de feriado moderado.

Na verdade, esses problemas fizeram com que os erros de previsão aumentassem, mas isso não quer dizer que o processo futuro de previsão terá que incorrer nesses erros de grande magnitude quando ocorrerem dias com comportamentos diferentes. Na prática, quando os analistas de previsão estiverem utilizando o modelo aqui proposto, eles poderão manipular os valores da variável *dummy* feriado para os dias diferentes do padrão.

Assim, por exemplo, o dia 24/12 pode receber manualmente o valor 0,7 para a sua variável feriado, ao invés de ser tratado como um dia normal; o dia 20/01 pode receber o valor 0,4 da mesma forma; um feriado que caia num domingo pode ter o seu valor da variável feriado reduzido de 1 para 0,6.

Depois desta manipulação, basta considerar as informações do setor de faturamento da Telemar sobre os ciclos de contas e usar a equação 1 apresentada anteriormente para prever a quantidade de ligações para o produto 103, a cada dia. Uma vez gerados os *outputs* da equação, os analistas podem incorporar os impactos esperados pelos eventos especiais, como campanhas de mídias, novos serviços etc. e, dessa forma, reduzir substancialmente os erros de previsão.

4 Conclusões

Paralelo ao crescimento que a indústria de *call centers* tem apresentado no Brasil e no mundo, esse tipo de negócio vem experimentando um grande aumento na complexidade de suas operações, que dificultam sua gestão sem o uso de ferramentas quantitativas adequadas.

Foi constatado que a ferramenta de previsão (TotalView) utilizada pela empresa estudada – a Contax – não leva em conta algumas informações que poderiam refinar o processo de previsão. Assim, no caso específico do produto 103, que diz respeito a serviços de telefonia fixa e problemas com a conta telefônica, não incluir a ocorrência de feriados, datas de vencimento e de chegada das contas na residência dos clientes, resulta no desperdício de uma informação útil e de fácil acesso capaz de reduzir os erros de previsão.

Levando estes efeitos em conta, foi desenvolvido um modelo de Regressão Múltipla com variáveis *dummy* para prever o volume diário de ligações. Vale lembrar que esse modelo foi construído com o objetivo de complementar – e nunca substituir! – a experiência e a intuição dos especialistas da empresa.

As previsões fornecidas pelo modelo de regressão foram comparadas aos dados reais de demanda e revelaram uma acurácia superior àquela obtida apenas pelo TotalView. Além disso, a maneira como o modelo foi desenvolvido permite sua fácil incorporação ao processo de previsão da empresa.

Finalmente, alguns tópicos não abordados no presente trabalho representam potenciais oportunidades de estudo futuros, tais como:

- A consideração de regionalidades seria um aspecto interessante a ser tratado em estudos futuros. Por exemplo, o impacto na demanda decorrente de uma

data representar o dia da chegada da conta telefônica para 200.000 clientes do Rio de Janeiro pode ser diferente do impacto proporcionado pela chegada da conta para 200.000 clientes da Bahia. Seria interessante a elaboração de um modelo capaz de levar em conta uma eventual diferença entre o comportamento cultural dos cariocas e baianos em relação à chegada da conta; e que não tratasse a demanda de forma consolidada nacionalmente e nem considerasse os impactos de forma homogênea.

- Outra sugestão consistiria na elaboração de um modelo de previsão de demanda capaz de capturar, entender e prever os efeitos de uma possível sazonalidade ao longo do ano, além da já capturada sazonalidade ao longo da semana. A consequência de tal efeito poderia ser capturada se a disponibilidade de informações abrangesse um período maior de tempo e se fossem criadas variáveis *dummy* para os meses do ano, assim como foi feito para os dias da semana.
- Poder-se-ia também tentar desenvolver um modelo para a previsão intradiária da demanda de ligações, já que a maioria dos *call centers* utiliza os blocos de 30 minutos como unidade básica de programação.
- Uma última sugestão aponta para a necessidade de extrapolar o uso dos modelos de regressão – como o aqui apresentado – para prever a quantidade de chamadas de outros produtos da própria Contax, como o Velox, OI, 102 ou suporte técnico.
- Assim sendo, quer pela sua crescente importância no ambiente atual de negócios, quer pela diversidade de problemas que apresenta, a gestão eficiente de *call centers* representa uma área fértil para a aplicação de ferramentas quantitativas para a sua melhor operação e planejamento.

Forecasting a call center demand using a Multiple Regression model

Abstract

This work describes – with the aid of a case study – a demand forecast problem for a specific product reported to the call center of a large Brazilian company in an industry called Contax, and the way it was approached with the use of Multiple Regression using dummy variables. After highlighting and justifying the studied matter relevance, the article presents a small literature review regarding demand forecast methods and their use in the call center industry. The case is described presenting the studied company and the way it deals with the Forecasting Demand for a telephone all center regarding telephone services products. Therefore, a Multiple Regression with dummy variables model was developed to work as the basis of the proposed demand forecast process. This model uses available data capable of influencing the demand such as the week day, occurrence of holidays, and the date of critical events such as the date on which the bill is sent and the date of payment collect. The model presented an improvement of Demand Forecasting Accuracy of 0.3% in the studied period when compared to the previously tool in use

Keywords: Call center. Demand Forecast. Multiple Regression.

Referências bibliográficas

- ALAM, M. **Using call centers to deliver public services**: house of commons paper. London: The Stationery Office Books, 2002.
- AZEVEDO, M.; CALDAS, M. Seriam os call centers os sweatshops do século XXI? In: ENCONTRO DE ESTUDOS ORGANIZACIONAIS, 2002, Recife. **Anais...** Recife: ANPAD, 2002. p. 1-15.
- BAPAT, V.; PRUITTE Jr., E. Using simulation in call centers. In: WINTER SIMULATION CONFERENCE, 1998, Washington. **Proceedings...** Los Alamitos: IEEE Computer Society Press, 1998. p. 1395-1399.
- BOUZADA, M. **O uso de ferramentas quantitativas em call centers: o caso Contax**. 2006. 256p. Tese (Doutorado em Administração de Empresas) – Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- CONTAX. **Contax Contact Center**. Disponível em: <www.contax.net.br>. Acesso em: 19 de Julho de 2006.
- CORRAR, L.; PAULO, E.; DIAS FILHO, J. **Análise Multivariada para os Cursos de Administração, Ciências Contábeis e Economia**. São Paulo: Atlas, 2007.
- GANS, N.; KOOLE, G.; MANDELBAUM, A. Telephone call centers: tutorial, review and research prospects. **Manufacturing and Service Operations Management**, v. 5, n. 2, p. 79-141, 2003.
- GROSSMAN, T. et al. Call centers. In: GASS, S.; HARRIS, C. (Eds.). **Encyclopedia of Operations Research and Management Science**. Boston: Kluwer Academic Publishers, 2001. p. 73-76.
- HALL, B.; ANTON, J. Optimizing your call center through simulation. White Paper, Oct. 1998. Disponível em: <http://www.erlang.com.br/artigos/Optimizing%20Your%20Call%20Center%20Through%20Simulation.pdf>
- HANKE, J. **Business forecasting**. New Jersey: Prentice Hall, 1992.
- HAWKINS, L. et al. **Planning guidance document for US call centers**. Maryland: Information Technology Support Center, 2001.
- KLUNGLE, R.; MALUCHNIK, J. The role of simulation in call center management. In: MSUG CONFERENCE, 1997, Michigan. **Proceedings...** Michigan: MSUG, 1997. p. 1-10.
- MATAN, O.; NOURBAKHSI, I. **Playing the numbers**: using ACD statistics for workforce management. [S.L.]: Telemarketing & Call Center Solutions, Mar 1998. Disponível em: <http://www.encyclopedia.com/doc/1P3-33178823.html >.
- MEHROTRA, V. Ringing Up Big Business. **OR/MS Today**, v. 24, n. 4, p. 18-24, 1997.
- MEHROTRA, V.; PROFOZICH, D.; BAPAT, V. **Simulation**: the best way to design your call center. [S.L.]: Telemarketing & Call Center Solutions, Nov 1997. Disponível em: <http://findarticles.com/p/articles/mi_qa3700/is_199711/ai_n8778658/>.
- MINGHELLI, G. **Call center**: estudo de casos múltiplos em empresas de telefonia do estado do Rio Grande do Sul. 2002. 64p. Monografia (Bacharelado em Informática) - Universidade do Vale do Rio dos Sinos, São Leopoldo.
- OUTSOURCING. **Ranking**. Disponível em: <www.callcenter.inf.br>. Acesso em: 20 de Setembro de 2005.
- REYNOLDS, P. **Forecasting fundamentals**: the art and science of predicting call center workload. [S.L.]: Technology Marketing Corporation. White Paper, Feb 17, 2005. Disponível em: <http://www.tmcnet.com/channels/workforce-optimization/workforce-optimization-articles/forecasting-call-center-workload.htm>.
- SAKAMOTO, K. **Como implementar um call center**. São Paulo: Happy Few, 2001.
- SHEN, H.; HUANG, J. Analysis of call centre arrival data using singular value decomposition. **Applied Stochastic Models in Business and Industry**, v. 21, n. 3, p. 251-263, 2005.
- SILVER, M. **Estatística para Administração**. São Paulo: Atlas, 2000.
- WEINBERG, J.; BROWN, L.; STROUD, J. **Bayesian Forecasting of an Inhomogeneous Poisson Process with Applications to Call Center Data**. Pennsylvania: University of Pennsylvania, 2006.
- WEISSHUHN, W. **Análise de custos na formação de preços em uma empresa de Call Centers**: um estudo de caso. 2004. 125p. Dissertação (Mestrado em Administração de Empresas) – Universidade Federal do Rio de Janeiro, Rio de Janeiro.

Sobre os autores

Marco Aurélio Carino Bouzada

Mestrado em Administração e Desenvolvimento Empresarial (MADE), Universidade Estácio de Sá (UNESA) – Av. Presidente Vargas, 642, sala 2207, Centro, CEP 20071-001 Rio de Janeiro – RJ
e-mail: marco.bouzada@estacio.br

Eduardo Saliby

COPPEAD/UFRJ, Universidade Federal do Rio de Janeiro
R. Pascoal Lemme, 355, CEP 21941-918, Rio de Janeiro – RJ
e-mail: saliby@coppead.ufrj.br

Recebido: 28/1/2009

Aceito: 5/6/2009