

Modelo de Programação Linear Aplicado ao Planeamento de Recursos em Sistemas de Filas de Espera

Linear Programming Model Applied to the Resource Scheduling at Queue Systems

Manuel do Carmo

Universidade Europeia, Laureate International Universities
Estrada da Correia, 53, 1500-210 Lisboa, Portugal
manuel.carmo@universidadeeuropeia.pt
e CIMA-UE, Universidade de Évora
Évora, Portugal

Paulo Jorge da Silva

Timepod, Lisboa, Portugal
paulo.silva@timepod.pt

Paulo Infante

ECT/DMAT, Universidade de Évora,
e CIMA-UE, Universidade de Évora
Évora, Portugal
pinfante@uevora.pt

Resumo — A pressão para manter (ou aumentar) o nível de competitividade das empresas conduz à exigência, cada vez maior, em conseguir uma gestão eficaz, sustentada pela variedade de recursos existentes, que tendem a dificultar positivamente o trabalho do gestor. No caso particular das filas de espera, o aspeto mais relevante prende-se com a combinação da quantidade de servidores multitarefa designados para atender, e com a procura que varia ao longo de um dia de trabalho. O objetivo primordial é ir ao encontro das expectativas criadas pelo cliente, nomeadamente no que diz respeito à opinião do serviço que foi prestado. O contributo que este trabalho se propõe dar consiste em planear o número de servidores adequado a uma determinada procura, sem esquecer, ao nível da gestão, a valorização que está associada à libertação dos recursos para desempenhar outras tarefas relacionadas com o negócio. Recorremos à programação linear para determinar a quantidade ótima de servidores destacados para servir clientes, dependendo, simultaneamente da sua disponibilidade horária para cumprir essa função e da procura estimada. O modelo é submetido a quatro cenários reais distintos, cada um refletindo uma determinada realidade de atendimento para, dessa forma, podermos avaliar o seu comportamento e eventuais ganhos face a situações reais ocorridas nas mesmas condições. Os resultados obtidos mostram a aplicabilidade do modelo, pois revela, em condições muito semelhantes às de uma situação real (traduzidas num dos cenários), diferenças muito pouco relevantes.

Palavras Chave – programação linear; planeamento; filas de espera.

Abstract — The pressure to maintain (or increase) the level of competitiveness in companies leads to an ever-increasing requirement for effective management supported by the variety of existing resources, which tend to interfere into the work of the manager. Particularly, in the case of waiting queues, the most relevant aspect is the combination of the number of multitasking servers assigned to attend and the demand that varies over the course of a day's work. The main objective is to meet the expectations created by the client regarding the opinion of the service that was provided. The contribution that this work proposes to give consists in planning the number of servers adapted to a given demand, not forgetting the valuation associated with the release of resources to perform other tasks related to the business. We use linear programming to determine the optimal number of the servers to serve customers, depending on their hourly availability to fulfill this function and the estimated demand. The model is submitted to four distinct real scenarios, each one reflecting a certain service reality, so that we can evaluate its behavior and possible gains against real situations occurring under the same conditions. The results show the credibility of the model since it reveals in conditions very similar to those of a real situation (translated in one of the scenarios), very little relevant differences.

Keywords – linear programming; scheduling; queue.

I. INTRODUÇÃO

Uma das primeiras coisas que pode causar perturbação a um cliente, quando este se depara com uma fila de espera presencial, é o tempo para ser servido. “*Espero ou não espero?*” é a dúvida que habitualmente surge no momento da chegada. A decisão costuma envolver a contagem do número de pessoas que se encontram à sua frente e a consequente estimativa do tempo para as servir. Na prática, este processo de decisão revela-se da maior importância, ao envolver custos de ordem psicológica, pois enquanto os clientes aguardam pela sua vez, estes estão a aceitar prescindir de um recurso escasso: tempo. Por outro lado, o fornecedor do serviço deve garantir que nenhum dos seus recursos se vá manter desocupado, pois ao ter clientes em espera na fila, são estes que vão “pagar” com o seu tempo o privilégio de serem servidos transferindo-se, assim, os custos do fornecedor para o cliente. Esperar por um serviço pode ser uma experiência desagradável, que cria clientes insatisfeitos e influencia negativamente a sua opinião quanto à qualidade de serviço prestado [1]. A questão de fundo é a relação que existe entre a experiência da espera e a avaliação que o cliente faz do serviço prestado [1], pois este pode tomar a decisão de não mais voltar a recorrer aos serviços da empresa, arrançando uma outra alternativa. Mesmo que não o façam, de facto, algumas pessoas detestam tanto esperar que estão dispostas a contratar outras pessoas para esperar por elas [2].

Consequentemente, as empresas são pressionadas para providenciar um melhor serviço, mais rápido, e a desenvolver ações que reduzam os efeitos negativos que a demora pode causar nos clientes. Ao consegui-lo, estão não só a corresponder positivamente à expectativa criada pelo cliente em relação à atenção e tratamento que a empresa vai dar ao seu assunto, como também a ganhar consciência que a opinião formada quanto à qualidade geral do atendimento, deve contemplar a experiência da espera, para além da imagem corporativa, instalações, cortesia dos funcionários e organização dos serviços, a experiência da espera.

A interpretação do momento em que o cliente entra no estado de espera é variada. Uma das abordagens mais populares é feita pela percepção que o cliente tem do tempo que está à espera [3]. Se não é possível controlar a aleatoriedade da chegada de clientes, nem o verdadeiro tempo necessário para os servir, então tenta-se controlar a percepção que o cliente tem do tempo que julga ter estado à espera,

por exemplo disponibilizando informação acerca da duração do tempo de espera e informação do tamanho da fila [4]. Assim, pensamos haver espaço para explorar a possibilidade de aumentar a eficácia do serviço prestado na vertente de quem tem, em primeiro lugar, a missão de servir o cliente, isto é, o operador, definindo um modelo de gestão que permita simplificar e planear antecipadamente os recursos humanos que melhor correspondem à procura *versus* o custo subjacente que, necessariamente, lhe está associado. Esta visão é validada pela perspectiva defendida em [5], que demonstra que antecipando o início das horas de trabalho dos servidores para “*poucos minutos*” antes da “*hora certa*” (segundo [5], uma hora com incrementos de 5, 10 ou 15 minutos), “*(...) pode ter um impacto dramático nos tempos de espera do cliente por períodos extensos.*”. Isto reforça a ideia de que o grau de subjetividade, que está associado à capacidade instalada para assistir à tarefa “*servir cliente*”, pode não interferir nem comprometer a coordenação e planificação eficiente de todo o esforço envolvido em servir clientes. O gestor desempenha, por isso, um papel crucial na decisão do momento em que deve ocorrer a entrada em serviço de um ou mais servidores, importância essa acrescida pelo facto de se criar um efeito colateral relacionado com a aversão que os servidores demonstraram pelo facto do turno não ter início à “*hora certa*”. Mas isso é uma questão marginal a este trabalho, que não vamos abordar.

Assim, começamos por descrever a metodologia, incluindo o enquadramento do problema num conjunto de cenários reais que dispõem de um grupo de rácios que medem o comportamento e eficácia do modelo para cada uma das situações enunciadas. Apresentamos os resultados mais significativos, onde é feita uma análise crítica e resumida, terminando com as principais conclusões que, em certa medida, justificam o contributo deste trabalho para a comunidade.

II. METODOLOGIA

Os dados obtidos que constituem a nossa população de estudo são provenientes de um *SGFE* (*Sistema de Gestão de Filas de Espera*) em produção, cuja principal função está associada à gestão e registo de eventos ocorridos no âmbito de um sistema de atendimento, desde a chegada do cliente à fila, com a emissão de uma senha com um número, até ao momento em que esse número é apresentado no ecrã de chamada a indicar que se

deve deslocar ao servidor disponível para o servir. Pretendemos, pois, destacar que os dados que constituem a base do estudo refletem situações reais de atendimento e não são o resultado de nenhuma simulação realizada em laboratório, contrariamente a alguns estudos analisados, como, por exemplo, é feito em [4], [6], [7] e [8].

Foram detetadas situações anómalas, que não refletiam a realidade do atendimento. Por exemplo, o servidor nem sempre informava o sistema do momento seguinte ao terminar o serviço, acontecendo erradamente alguns serviços ficarem com uma duração de várias horas. O oposto, ou seja, serviços com uma duração reduzida também foram identificados. Situações como as referidas, bem como outras, foram identificadas e, posteriormente, removidas do domínio ou corrigidas com vista à definição final da amostra-alvo. Simultaneamente, dada a extensão do conjunto de resultados, realizámos uma abordagem em vários passos. Começámos por realizar uma análise exploratória, obtendo medidas de tendência central, de dispersão, de forma (assimetria e curtose), às variáveis utilizadas no modelo. Pudemos gerar novas variáveis (p.e., a variável “*tempo entre chegadas*” foi obtida a partir do registo da “*data e hora de chegada*” do cliente) relevantes para o estudo e foi possível construir e executar *queries*, na base de dados relacional, com o objetivo de apurar ou esclarecer incoerências nos dados originais (p.e., só foram considerados serviços com duração superior a 0.25 e inferior a 12 minutos). De seguida aplicámos o teste do *Qui-Quadrado de independência* para avaliar a existência de associação entre os principais grupos: “*tipo de fila*”, “*dia da semana*” e “*servidor*” no que diz respeito à característica “*duração do serviço*”. Para a característica “*tempo entre chegadas*”, considerámos as variáveis do “*tipo de fila*” e “*dia da semana*”, permitindo, desta forma perceber a existência de influências em duas das principais variáveis do modelo. Analisámos gráficos *pp-plot* e aplicámos o teste de *Kolmogorov-Smirnov (K-S)* de ajustamento às variáveis “*tempo entre chegadas*” e “*duração do serviço*” para averiguar se podíamos assumir a distribuição exponencial para estes tempos. Este procedimento permitiu-nos perceber se estávamos na presença do tradicional modelo *M/M/n* (sistema de fila de espera com chegadas de *Poisson*, tempos de serviço exponenciais e *n* postos de serviço em paralelo). Desta forma, ao reduzir a diversidade dos grupos das variáveis mais significativas da amostra, estamos não só a simplificar os parâmetros que vamos

utilizar na análise estatística, como também pensamos que ela contribua para demonstrar, de uma forma mais clara e precisa, as possíveis soluções para o problema de programação linear a propor.

O conjunto de pressupostos que permitiram a aplicação da nossa solução assentaram na estipulação de um padrão de qualidade de serviço, definido pela gestão do atendimento, com base na simplificação da mensuração do custo pecuniário associado ao salário, através da sua transformação em horas. Assim, foi formulado um problema de programação linear que minimiza o custo de atendimento numa base de hora/serviço. Para este estudo assumimos três hipóteses:

1. os servidores menos onerosos são aqueles que têm como tarefa primária servir clientes, independentemente do serviço solicitado;
2. ser possível estimar o número mínimo de clientes a servir em cada dia útil da semana;
3. ser possível estimar o número de horas que os servidores têm atribuídas à tarefa “*servir clientes*”.

A função objetivo do modelo de programação linear, para o problema exposto, identifica as variáveis para a tomada da decisão e respeita as seguintes condições:

- função objetivo: $\text{Min } f(X)$;
- restrições técnicas e lógicas ao modelo.

Assim, os resultados obtidos permitirão concluir se o modelo é viável (*feasible*), aferir o número ótimo de atendimentos de cada servidor em cada dia da semana e planificar uma tabela de alocação de recursos humanos otimizada. Com as hipóteses apresentadas, abordaremos esta questão através da criação de quatro cenários que ilustram a procura dos clientes e disponibilidade de cinco servidores ao longo da semana, com o intuito de elaborar o escalonamento de recursos humanos mais eficiente em função desses dois parâmetros. Pretendemos refletir nesses cenários as seguintes situações:

(I) previsão da realização de x serviços no início da semana (segunda-feira) com redução gradual da procura ao longo da semana (ex. semana típica com base na evidência estatística da amostra);

(II) mesma tendência de procura que em (I), duplicando a quantidade prevista de serviços a realizar (p.e. período de férias);

(III) mesma tendência de procura que em (I), com indisponibilidade total por parte dos dois servidores

que têm designada como tarefa primária servir clientes (p.e. ausência do servidor por motivo de doença);

(IV) procura estabilizada ao longo da semana, à exceção de um dia evidenciando um aumento de 100% face aos restantes dias (p.e. data limite para pagamento de impostos no final do mês).

Outros cenários poderiam ser construídos para refletir outras realidades associadas ao atendimento. Contudo, identificámos apenas quatro por julgarmos serem aqueles que podem ilustrar as vantagens e desvantagens do modelo adotado. Para isso serão utilizados rácios de comparação entre os resultados obtidos pela aplicação do modelo de programação linear construído e situações similares encontradas na base de dados real, por forma a medir os eventuais ganhos na utilização da nossa proposta. Para avaliar, no contexto considerado, a variação entre o número médio de servidores proposto, em cada cenário, e o número médio de servidores encontrados a servir, vamos considerar o rácio (1), denominado a partir de agora como Q_1 , e que representa uma medida de variação relativa do número médio de servidores, em %, tendo como referência o número médio de servidores proposto em cada cenário, dado por:

$$Q_1 = \frac{n^\circ \text{ médio serv. sit. real} - n^\circ \text{ médio serv. cenário}}{n^\circ \text{ médio serv. cenário}} \times 100 \quad (1)$$

Para avaliar a variação entre a parametrização do número médio de horas alocadas ao atendimento, em cada cenário, e o número médio de horas detetadas na amostra para desempenhar a mesma tarefa, vamos considerar o rácio (2), denominado a partir de agora como Q_2 , e que representa uma medida de variação relativa do número médio de horas alocadas ao atendimento, em %, tendo como referência o número médio de horas parametrizadas em cada cenário, dado por:

$$Q_2 = \frac{n^\circ \text{ médio horas sit. real} - n^\circ \text{ médio horas cenário}}{n^\circ \text{ médio horas cenário}} \times 100 \quad (2)$$

Com estas medidas de desempenho, tentamos aferir as diferenças mais significativas do que seria a situação ideal com a realidade, apoiando a gestão em dois aspetos, considerados decisivos para o sucesso das funções que desempenha: antecipar quantos servidores são necessários e quantas horas do seu horário de trabalho devem ser alocadas ao desempenho de dois tipos de tarefas: “*servir clientes*” e “*outras*”. O modelo de programação linear proposto corresponde, assim, à minimização da função:

$$\text{Min } f(X) = \sum_{k=1}^n \sum_{i=2}^6 \sum_{j=1}^m X_{kij}, \quad (3)$$

onde $X_{kij} \geq 0$ é a quantidade de atendimentos do servidor j no dia útil i para o tipo de fila k , m a ordem do último servidor considerado apto para realizar serviços e n a ordem do último tipo de fila considerada (a ordem numérica incremental subjacente ao índice deste somatório é, na realidade, uma representação de um grupo de letras do alfabeto português (de A a Z), também ela incremental).

A primeira restrição diz respeito ao número mínimo estimado de clientes que os servidores no seu todo devem servir em cada dia útil. O mesmo é dizer que o somatório do número de serviços prestados pelos servidores deve ser maior ou igual à estimativa da procura diária dos clientes. Deste modo temos:

$$\sum_{k=1}^n \sum_{j=1}^m X_{kij} \geq P_i, \quad (4)$$

onde $X_{kij} \geq 0$, m e n são definidos como anteriormente e $P_i \geq 0$ a quantidade mínima diária de clientes a servir no dia útil i ($2 \leq i \leq 6$) com $n \in \{1,2,3,4\}$ e $m \in \{1,2,3,4,5\}$.

A segunda restrição diz respeito à quantidade de tempo alocado a cada servidor para dedicar à tarefa de atendimento em função do tipo de fila e dia útil da semana. Assim, temos:

$$\frac{1}{m} \times \sum_{k=1}^n \sum_{j=1}^m X_{kij} \leq D_{kij}, \quad (5)$$

onde $X_{kij} \geq 0$, m e n são definidos como anteriormente, $D_{kij} \geq 0$ representa a disponibilidade diária (horas) do servidor j no dia útil i ($2 \leq i \leq 6$) para o tipo de fila k e $1/m$ o coeficiente, constante, aplicado a todas as variáveis de decisão, cujo valor é determinado pela quantidade de servidores utilizados, com $n \in \{1,2,3,4\}$ e $m \in \{1,2,3,4,5\}$. O cálculo da quantidade de atendimentos atribuída a cada servidor em cada dia da semana é feito depois de executado o modelo efetuando a soma dos valores calculados, no âmbito do grupo de variáveis de decisão que está associado a cada servidor/dia de semana.

Para a execução deste trabalho recorreremos ao *Microsoft SQL Server Management Studio*® para a geração das novas variáveis e na construção e execução de *queries* na base de dados relacional. Na execução do modelo de programação linear, em que está assente a resposta ao problema formulado, utilizamos o *LPSolve 5.5*.

III. APRESENTAÇÃO E DISCUSSÃO DE RESULTADOS

Num modelo de filas de espera ($M/M/n$) o número de chegadas, por unidade de tempo, segue uma distribuição de *Poisson* e a duração do atendimento uma distribuição exponencial negativa. Com recurso ao software *SPSS 23.0*[®], essas premissas não se verificaram. Considerando o teste de *K-S*, para análise às duas variáveis obtivemos um valor- $p=0.001(<0.05)$, o que nos levou a não poder admitir a distribuição exponencial para o “tempo entre chegadas” e para a “duração de serviço”.

Quando efetuamos o teste do *Qui-Quadrado*, concluímos que há evidência estatística para concluir que o “tempo entre chegadas” esteja associado ao “dia da semana” e ao “tipo de fila” (valor- $p=0.001<0.05$) e que a variável “duração de serviço” esteja associada ao “tipo de fila” e ao “servidor” (valor- $p=0.001<0.05$), mas não ao “dia da semana” (valor- $p=0.203<0.05$).

Nas condições impostas na metodologia, e considerando os rácios (1) e (2) para o cenário (I), os resultados obtidos apresentam-se na *Tabela I*.

TABELA I. VALORES DE Q_1 E DE Q_2 PARA O CENÁRIO (I).

Volume Serviços	Nº médio de horas de serviço por servidor	Nº médio de servidores alocados	Nº médio de horas de serviço por servidor	Nº médio de servidores alocados	Q_1	Q_2
	- real -	- real -	cenário (I)	cenário (I)		
26	3.7 h	3	8 h	2	50	-53.8
28	3.4 h	3.5	8 h	2	75	-57.5
29	4.3 h	3	8 h	2	50	-46.3
30	3.7 h	3.3	8 h	2	65	-53.8

Ainda nas mesmas condições, e rácios, obtivemos os resultados para o cenário (II), que apresentamos na *Tabela II*. Relativamente ao cenário (III), os resultados obtidos para Q_1 e Q_2 são apresentados na *Tabela III*.

Dentro da lógica que sustenta os quatro cenários construídos, o rácio Q_1 indica ganhos relevantes para quase todas as situações, à exceção do cenário (III), onde o ganho é nulo. Analisando com mais detalhe o cenário (III), aquele que de entre os quatro apresenta Q_1 com os valores mais baixos quando comparamos o número médio de servidores utilizados entre a situação real e a proposta, verificamos que esse número é praticamente igual, acompanhando Q_2 também essa tendência. Tal resulta do facto de neste cenário os servidores que habitualmente estariam dedicados à tarefa de “servir clientes” estarem indisponíveis e, em função disso, os servidores que

os substituíram terem menos tempo alocado para essa tarefa.

Nos restantes três cenários, a quantidade de servidores sugerida pelo modelo é sempre menor à da realidade representada pela amostra, situando-se os ganhos entre os 41.7% (cenário (I)) e os 91.7% (cenário (IV)).

TABELA II. VALORES DE Q_1 E DE Q_2 PARA O CENÁRIO (II).

Volume Serviços	Nº médio de horas de serviço por servidor	Nº médio de servidores alocados	Nº médio de horas de serviço por servidor	Nº médio de servidores alocados	Q_1	Q_2
	- real -	- real -	cenário (II)	cenário (II)		
52	4.3 h	4.3	6.7 h	3	41.7	-35.8
56	4.8 h	4.8	6.7 h	3	58.3	-28.4
58	4.5 h	5.1	6.7 h	3	71.4	-32.8
60	4.7 h	4.7	6.7 h	3	57.6	-29.9

Para o cenário (IV), os resultados obtidos para os rácios considerados, são apresentados na *Tabela IV*.

TABELA III. VALORES DE Q_1 E DE Q_2 PARA O CENÁRIO (III).

Volume Serviços	Nº médio de horas de serviço por servidor	Nº médio de servidores alocados	Nº médio de horas de serviço por servidor	Nº médio de servidores alocados	Q_1	Q_2
	- real -	- real -	cenário (II)	cenário (II)		
26	3.7 h	3	3.3 h	3	0.0	12.1
28	3.4 h	3.5	3.5 h	3	16.7	-2.9
29	4.3 h	3	4.2 h	3	0.0	2.4
30	3.7 h	3.3	4.2 h	3	10.0	-11.9

No cenário (IV) é onde se prevê uma maior situação de *stress*, e onde pensamos que o modelo sai reforçado pelo facto de propor uma solução que exige significativamente menos recursos em relação ao número detetado na amostra, perante condições idênticas.

TABELA IV. VALORES DE Q_1 E DE Q_2 PARA O CENÁRIO (IV).

Volume Serviços	Nº médio de horas de serviço por servidor	Nº médio de servidores alocados	Nº médio de horas de serviço por servidor	Nº médio de servidores alocados	Q_1	Q_2
	- real -	- real -	cenário (II)	cenário (II)		
60	4.7 h	4.7	5.0 h	3	57.6	-6.0
64	4.4 h	5.0	5.3 h	3	66.7	-17.0
67	4.5 h	5.8	5.5 h	3	91.7	-18.1
68	4.4 h	5.5	5.5 h	3	82.1	-20.0
120	5.2 h	6.8	8.0 h	4	70.5	-35.0

Em virtude de na amostra a quantidade de servidores ser tipicamente superior à média de servidores propostos pelo modelo, consideramos natural que na situação real o tempo médio alocado à tarefa “servir

clientes” seja menor. Como resultado desta situação, tipicamente, Q_2 apresenta percentagens que denotam, em qualquer dos cenários, uma faixa média de tempo para prestar serviço sempre superior à encontrada em situação real.

É para nós uma informação merecedora de destaque o facto do cenário (III), aquele que mais se assemelha às características de atendimento real, ou seja, o que considera a maior diversidade de servidores face ao número de serviços estimados, ser o que apresenta o menor desvio Q_2 com um valor de, apenas, -11.9%. Na prática, estamos a falar numa diferença de trinta minutos. No outro extremo temos o cenário (I), justificado, segundo o nosso ponto de vista, pelo planeamento prever a existência de dois servidores alocados à tarefa primária “*servir clientes*”, durante um período de oito horas cada. Uma explicação para tal desequilíbrio pode ser dada pela maneira *ad hoc* como, em ambiente real, o atendimento é gerido, com os recursos a terem todos como tarefa primária “*servir clientes*”, agindo de acordo com a conveniência ou a maior proximidade do local de atendimento, com a chegada de novos clientes à fila vazia e subsequente incremento da mesma. Este modo de funcionamento vai registar uma maior diversidade de servidores a atender e com menos investimento de tempo por parte de cada um.

Pelos resultados obtidos podemos concluir que no cenário (III) os valores dos rácios Q_1 e Q_2 são, praticamente, idênticos aos de uma situação real, ou seja, as diferenças são pouco significativas. Desse modo, os resultados gerados para outras conjecturas podem ser suficientemente consistentes de modo a darem um contributo relevante para a planificação dos recursos afetos à tarefa “*servir clientes*”. Assim, julgamos que uma abordagem menos reativa aos acontecimentos, substituída por uma atitude mais refletida e programada da atividade de atendimento, terá um impacto assinalável na eficiência do trabalho realizado diariamente na instituição, permitindo deslocar os recursos mais valiosos (ou valiosos), por exemplo, para tarefas mais técnicas ou complexas.

IV. CONCLUSÕES

Os resultados apresentados, e analisados, são parte de um conjunto de dados mais alargado, que não apresentamos por limitações de espaço.

Assim, tendo em conta os mesmos, e pelas comparações feitas com situações idênticas às ocorridas em contexto real, estamos convictos que o nosso trabalho aplicado ao contexto de gestão de

atendimento dará um contributo fiável na área de escalonamento de recursos multitarefa, fornecendo indicadores que suportam o gestor no processo de tomada de decisão, em particular na antecipação de quantos servidores são necessários e quantas horas do horário de trabalho devem ser alocadas ao desempenho de dois tipos de tarefas – “*servir clientes*” e “*outras*”. Existindo dependência entre a duração de serviço e o servidor, como demonstrado pelo teste do *Qui-Quadrado* (valor- $p=0.001<0.05$), a eficácia desta decisão pode ser aumentada se for ponderada a capacidade que cada servidor, ou seja, cada colaborador, dispõe para melhorar o tempo de resposta a um ou vários tipos de serviço. Esta conclusão é reforçada pelo facto dos dados da amostra que sustentam este trabalho serem reais, em vez de serem produzidos artificialmente através de simulação. Acreditamos que a “*realidade*” obtida por simulação não contribui para um grau de fiabilidade semelhante ao de trabalhos como o nosso, suportados em factos reais. Estamos, por isso, convencidos que os resultados alcançados para os cenários idealizados são credíveis e minimizam os custos, ao alocar a quantidade ótima de recursos face aos objetivos traçados. No futuro pretendemos estudar esta situação recorrendo a um dos modelos económicos que se podem encontrar na literatura.

AGRADECIMENTOS

Manuel do Carmo e Paulo Infante são membros do Centro de Investigação em Matemática e Aplicações (UID/MAT/04674/2013), financiado pela Fundação para a Ciência e Tecnologia (FCT).

REFERÊNCIAS BIBLIOGRÁFICA

- [1] Taylor, S. “Waiting for Service: The Relationship Between Delays and Evaluation”, Vol.2, Journal of Marketing, 58, 1994, pp. 56-67.
- [2] Geist, W. “They’re Hiring Others to Stand in Line, 2011, retrieved 04/06/2011, 11:29.
- [3] Jacob, H. “Subjective vs Objective Time Measures: A Note on the Perception of Time in Consumer Behavior”, Vol. 1, Journal of Consumer Research, 11, 1984, pp. 615-618.
- [4] Hui, M. K.; Tse, D. K. “What to tell consumers in waits of different lengths: An integrative model of service evaluation”, Vol. 2, Journal of Marketing, 60, 1996, pp. 81-90.
- [5] Martinich, J. S. “The Critical Few Minutes in Scheduling Time-Varying Queuing Systems”, Vol. 3, Decision Sciences, 33, 2002, pp. 415-431.
- [6] Susan, H. X.; Long, G.; Ou, J. “Service Performance Analysis and Improvement for a Ticket Queue with Balking Customers”, Vol. 6, Management Science, 53, 2007, pp. 971-990.
- [7] Pazgal, A.; Radas, S. “Comparison of Customer Balking and Reneging Behavior to Queuing Theory Predictions: An Experimental Study”, Vol. 8, Computers Operations Research, 35, 2008, pp. 2537-2548.
- [8] Perel, N.; Yechiali, U. “Queues with Slow Servers and Impatient Customers, Vol. 1, European Journal of Operational Research, 201, 2010, pp.247-258.