LEVANTAMENTO ESTATÍSTICO DO PERFIL DE UM *CALL CENTER* VISANDO O AJUSTE DE UM MODELO DE FILAS

Helinton A. L. Barbosa
Departamento de Estatística – ICEx – UFMG
Av. Antônio Carlos, 6627, 31270-901 – Belo Horizonte – MG
e-mail

Frederico R. B. Cruz
Departamento de Estatística – ICEx – UFMG
Av. Antônio Carlos, 6627, 31270-901 – Belo Horizonte – MG
e-mail: fcruz@ufmg.br

Resumo

Neste artigo é apresentada uma metodologia estatística para levantamento do perfil de um *call center* visando o ajuste de um modelo de filas finitas. Para encontrar este perfil, são realizadas análises descritivas em estratificações mensais, diárias e horárias dos dados, além de testes de hipóteses não-paramétricos. A idéia é obter valores baseados nos quais teremos modelos aderentes à situação real, de forma que simulações possam ser realizadas posteriormente, com o objetivo de encontrar pontos de operação economicamente vantajosos para *call centers* e sistemas estocásticos similares.

Palavras-Chaves: Estatísticas descritivas; Testes de hipóteses não-paramétricos; Teoria de filas.

Abstract

This paper summarizes a statistical analysis of a record of call center operations. The goal is to adjust a finite queueing model. Several statistical techniques are adapted for the analysis of the data. The characteristics deduced from the statistical analysis will form in the future the building blocks for theoretically interesting and practically useful mathematical models for call center operations and similar stochastic systems.

Keywords: Descriptive statistics; Non-parametric statistics; Queueing theory.

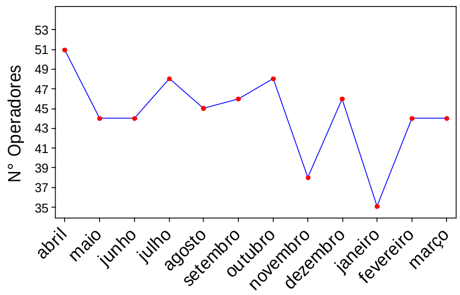
1. INTRODUÇÃO

Um *call center* é uma central que recebe chamadas telefônicas. O dimensionamento correto deste sistema promove entre outros fatores, a otimização de custos de operação, a minimização do tempo que o cliente espera até o atendimento *etc*. Para alcançar esta dimensão ótima, acreditamos que modelos de filas possam ser de grande ajuda, promovendo grandes benefícios às empresas e aos clientes. Neste estudo, analisaremos descritivamente um *call center* de tamanho médio, com vários servidores e grande capacidade de chamadas que aguardam atendimento. Esta análise descritiva tem o objetivo de traçar o perfil do sistema. Testaremos as suposições levantadas usando alguns testes de hipóteses não-paramétricos. Os valores serão usados na estimação de parâmetros de um modelo de filas conhecido como filas $M/G/c/k$, em que se adotando a notação de Kendall [1], o M indica que a chegada é um processo markoviano, G representa um tempo de serviço com distribuição geral, c se refere ao número de servidores e, por fim, a capacidade total do sistema é restrita a k usuários.

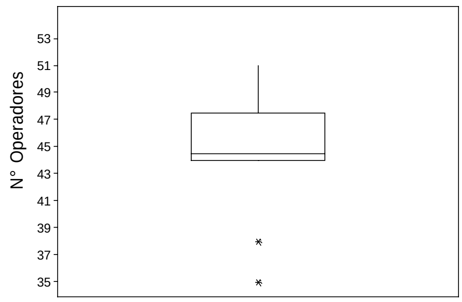
Este relatório encontra-se organizado da seguinte forma: Na Seção 2, temos a análise descritiva dos dados, organizada em periodicidade mensal, diária e horária, além de testes de hipóteses não-paramétricos. Na Seção 3 concluímos este relatório, apresentando resultados importantes obtidos e algumas observações finais.

2. ANÁLISE DESCRITIVA

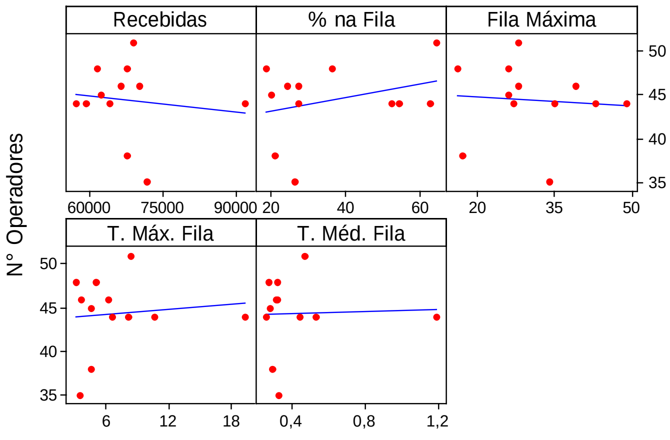
Inicialmente, informamos que este *call center* funciona de segunda a sexta-feira, das 8:00 às 22:00 horas, em dois turnos de trabalho. O primeiro turno vai das 8:00 às 15:00 horas e o segundo turno vai das 15:00 às 22:00 horas. No sábado, o funcionamento vai das 8:00 às 14:00 horas em um único turno. A análise mensal será feita com uma amostra de doze meses, entre abril de 2005 e março de 2006. Posteriormente, será utilizada uma estratificação em dias e, em seguida, por horas. O objetivo desta estratificação é entender o comportamento tanto no nível estratégico (dados mensais), quanto tático (dados diários e horários). Com isso, um modelo poderá ser ajustado, com grande confiabilidade, pois está baseado em valores extremos. As variáveis analisadas são o número de chamadas recebidas (recebidas); o percentual de chamadas do sistema que passaram pela fila (% na fila); o número máximo de chamadas na fila simultaneamente (fila máxima); o tempo máximo, em minutos, que a chamada aguarda na fila até atendimento (tempo máx. de fila.); o tempo médio, em minutos, que a chamada aguarda na fila até atendimento (tempo méd. de fila.) e a eficiência do *call center*, que é uma medida do número de chamadas recebidas e abandonadas. As variáveis 'número de chamadas recebidas' e 'tempo médio de fila' darão uma idéia sobre os tempos entre chegadas e o tempo de serviço, respectivamente. As demais variáveis são variáveis de controle, as quais futuramente compararemos com os resultados das simulações.



a) números de operadores mensais



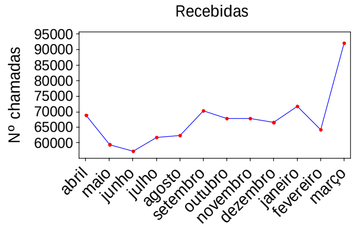
b) *box-plot* do número de operadores



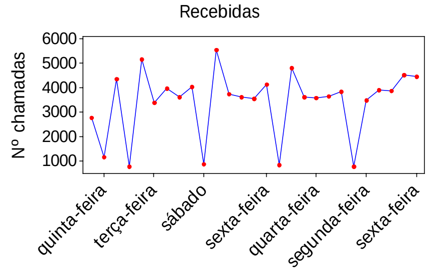
c) iteração entre as principais variáveis

Figura 1: Análise da variável número de operadores

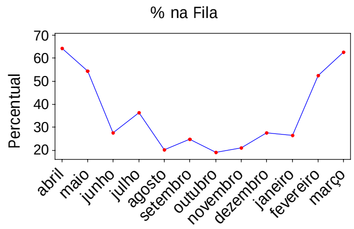
Antes de iniciarmos a análise mensal, é importante informar que o número de operadores no *call center* de abril até março não é fixo, por isso, faremos uma breve análise descritiva [2,3] destes dados. Na Figura 1-a, pode ser observada a série temporal do número de operadores. Perceba que, entre outubro e fevereiro, há uma variabilidade marcante. Esta variação de número de operadores está relacionada com tentativas da empresa em diminuir a fila máxima, sua maior preocupação. Já na Figura 1-b, note a presença de dois pontos atípicos [3] no gráfico. Estes pontos representam os meses de novembro e janeiro. Por fim, na Figura 1-c, perceba que não existe nenhuma relação clara entre o número de operadores e as variáveis ‘chamadas recebidas’, ‘percentual na fila’, ‘fila máxima’, ‘tempo máximo de fila’, ‘tempo médio de fila’ e ‘eficiência’. Isto indica que a variação no número de operadores pode não estar influenciando de maneira significativa o sistema.



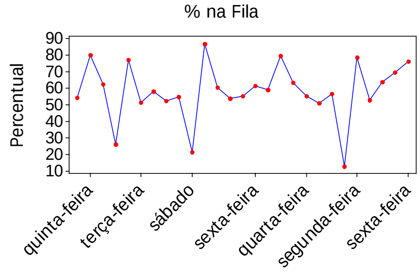
a) chamadas recebidas mensais



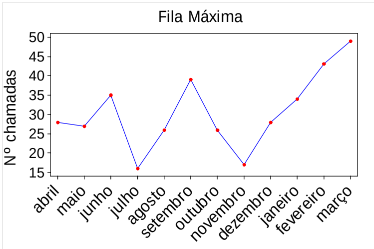
b) chamadas recebidas diárias



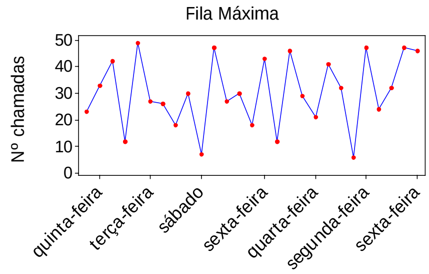
c) percentuais na fila mensais



d) percentuais na fila diários



e) filas máximas mensais



f) filas máximas diárias

Figura 2: Séries temporais mensais e diárias de variáveis de interesse

Na Figura 2, vemos as principais variáveis de interesse no nível estratégico (periodicidade mensal) e no nível tático (periodicidade diária), e notamos a diferença significativa entre os perfis. Isto justifica portanto a adoção de algum tipo de estratificação, como a que aqui será proposta.

A análise descritiva feita nesta seção mostrou que o *call center* possui como pontos importantes uma tendência de aumento no número de chamadas recebidas, um crescente percentual de chamadas na fila, um aumento no tempo médio de fila e queda na eficiência. Investigaremos a seguir se a tendência de crescimento no número de chamadas é significativa e se março é um mês com grande variabilidade.

2.1. TESTES DE HIPÓTESES NÃO-PARAMÉTRICOS ENTRE MESES

Após a análise descritiva inicial, é interessante testar se existe diferença entre o semestre de abril a setembro (semestre 1) e o semestre de outubro a março (semestre 2) em número de chamadas recebidas. Se não houver diferença, o ajuste de um único modelo de filas é suficiente. Para testar esta hipótese, usaremos o teste não paramétrico de Mann-Whitney[1] [2,4,5], para comparação de dois grupos independentes, já que um semestre é independente do outro. O uso de um teste não-paramétrico justifica-se pelo fato de termos

[1] O teste de Mann-Whitney é um teste que utiliza os postos das observações. É testado se as funções de distribuições dos grupos envolvidos são iguais ou não.

poucas observações mensais e de os dados não seguirem a distribuição normal. Além da independência, esse teste tem como suposição variabilidades entre grupos aproximadamente iguais. Pelo teste de Levene [6] para igualdade entre variâncias, notamos um valor-*p* de 0,699 e assumimos que as variabilidades dos dois semestres são iguais. Do teste de Mann-Whitney. Vemos que o valor-*p* para o teste é 0,1282, indicando que se adotarmos um nível de significância α=0,05, não rejeitaremos a hipótese de igualdade entre as distribuições do número de chamadas recebidas por semestres. Desta forma, o teste mostrou que não há diferenças significativas entre estas distribuições. Logo, os dois semestres tendem a receber um mesmo número de chamadas e ajustar um único modelo é suficiente para o sistema.

Agora, é necessário testar se o número de chamadas recebidas em março é significativamente diferente dos outros meses. Para isso, utilizaremos o teste não-paramétrico de Kruskal-Wallis[2] [2,4,5] para vários grupos independentes, pois como já dito, não existe a suposição de normalidade dos dados. Do teste de Kruskal-Wallis, vemos que o valor-*p* é menor que 0,000. Isto é, com um nível de significância α=0,05, rejeitaremos a hipótese de igualdade entre as distribuições do número de chamadas recebidas por meses, ou seja, existe pelo menos um mês com a distribuição do número de chamadas recebidas diferente. Identificar onde estão as diferenças é importante. Para isso, usaremos um teste não-paramétrico para comparações múltiplas[3] [7].

Podemos observar na Tabela 1 os valores-*p* referentes às comparações mês a mês. Se adotarmos um nível de significância α=0,05, os valores-*p* em negrito indicam quais os meses têm distribuições diferentes. Perceba que março é igual a janeiro e fevereiro, porém, diferente dos outros meses. Estes resultados não ficam evidentes exclusivamente pela análise gráfica.

Tabela 3: Valores-*p* do teste para comparações múltiplas entre meses

Mês	Outubro	Novembro	Dezembro	Janeiro	Fevereiro	Março
Outubro		1,000000	1,000000	1,000000	1,000000	**0,008141**
Novembro	1,000000		1,000000	1,000000	0,797060	**0,005463**
Dezembro	1,000000	1,000000		1,000000	0,112978	**0,000228**
Janeiro	1,000000	1,000000	1,000000		1,000000	0,070420
Fevereiro	1,000000	0,797060	0,112978	1,000000		1,000000
Março	**0,008141**	**0,005463**	**0,000228**	0,070420	1,000000	

Como o teste de comparações múltiplas mostrou, não existe diferença entre as distribuições dos meses de janeiro, fevereiro e março, ou seja, estes meses tendem a receber um mesmo número de chamadas. Na análise seguinte apresentaremos o mês de março. A justificativa por esta escolha é o fato de março possuir maior variabilidade que os outros dois meses.

2.2. TESTES DE HIPÓTESES NÃO-PARAMÉTRICOS ENTRE DIAS

Até agora, apenas especulamos sobre as diferenças entre dias. Através de análises descritivas, podemos chegar à pergunta: em número de chamadas recebidas, segunda-feira é diferente dos ouros dias? Para respondê-la, usaremos novamente o teste não-paramétrico de Kruskal-Wallis para grupos independentes, já que um dia é independente do outro e como não podemos assumir a normalidade dos dados. Neste teste usaremos dados diários de janeiro até março, pois estes foram os meses que se mostraram significativamente iguais.

Observamos que o valor-*p* associado ao teste de Levene para os dados é 0,169, que nos permite supor variabilidade constante entre meses. Dos resultados do teste de Kruskal-Wallis, observamos que o valor-*p* associado ao teste é menor que 0,000, o que indica que com um nível α=0,05, devemos rejeitar a hipótese de igualdade entre as distribuições do número de

[2] O teste de Kruskal-Wallis utiliza os postos das observações. É testado se as funções de distribuições dos vários grupos envolvidos são iguais ou contra se pelo menos uma é diferente.

[3] O teste para comparações múltiplas utiliza postos para fazer as comparações. As comparações são feitas por pares e são testadas se as funções de distribuição são iguais ou não.

chamadas recebidas por dias. Com a hipótese de igualdade entre distribuições rejeitada, devemos encontrar as diferenças. Na Tabela 2, podemos ver o teste de comparações múltiplas para dias. Note que se adotarmos um nível de significância α=0,05, os valores-*p* em negrito, indicam quais dias possuem distribuições diferentes umas das outras. Perceba que sábado é diferente de todos os demais dias. Note também que não há diferenças entre as distribuições nos dias de segunda-feira à sexta-feira. Portanto, nossas suposições iniciais se confirmaram parcialmente, ou seja, sábado possui o menor número de chamadas recebidas e segunda-feira, possui um número igual aos outros dias. Com isso, o ajuste de um modelo baseado no número máximo de chamadas recebidas de segunda-feira à sexta-feira, possivelmente descreverá bem a demanda diária do *call center*.

Tabela 2: Valores-*p* do teste para comparações múltiplas entre dias

Dia	Segunda-feira	Terça-feira	Quarta-feira	Quinta-feira	Sexta-feira	Sábado
Segunda-feira		1,000000	0,440644	0,247573	1,000000	**0,000000**
Terça-feira	1,000000		1,000000	1,000000	1,000000	**0,001232**
Quarta-feira	0,440644	1,000000		1,000000	1,000000	**0,004795**
Quinta-feira	0,247573	1,000000	1,000000		1,000000	**0,010712**
Sexta-feira	1,000000	1,000000	1,000000	1,000000		**0,000019**
Sábado	**0,000000**	**0,001232**	**0,004795**	**0,010712**	**0,000019**	

Como os dias de segunda à sexta-feira tendem a receber o mesmo número de chamadas, na análise por horas, é recomendável adotar o dia com maior variabilidade, ou seja, segunda-feira. Utilizaremos todas as segundas-feiras do mês de março estratificadas por horas. O próximo objetivo é encontrar qual turno e hora possuem o maior número de chamadas recebidas.

2.3. Testes de Hipóteses Não-paramétricos entre Horas

Testaremos agora hipóteses tais como: existe diferença entre turnos em relação ao número de chamadas recebidas? O horário de 19 horas é o ponto de máximo no segundo turno para chamadas recebidas? Para responder estas hipóteses usaremos novamente os testes não-paramétricos de Mann-Whitney e Kruskal-Wallis. Nestes testes, usaremos dados horários de todas as segundas, quartas e sextas de março, dias considerados significativamente iguais. O teste de Levene mostrou que a suposição de variabilidades aproximadamente iguais entre turnos é respeitada. Um valor-*p* de 0,925 nos garante isto. Além disso, o valor-*p* para o teste de igualdade entre as distribuições do número de chamadas recebidas por turnos para os dados transformados é menor que 0,000. Isto é, se adotarmos um nível de significância α=0,05, rejeitaremos a hipótese de igualdade entre as distribuições. Ou seja, concluímos que o segundo turno tende a receber um maior número de chamadas.

Também desejamos saber se existe diferença, em número de chamadas recebidas, entre vários horários do dia, aqui chamados de faixas. A primeira faixa vai das 8 horas até às 12 horas. A segunda faixa começa às 12 horas termina às 17 horas. E a terceira faixa vai das 17 horas até às 22 horas. Diferenças entre estas faixas, possivelmente indicam que alocações diferentes no número de operadores entre turnos podem otimizar o atendimento no c*all center*. O teste de Kruskal-Wallis para igualdade entre as distribuições do número de chamadas recebidas por faixas para os dados transformados indica que o valor-*p* é menor que 0,000. Isto nos leva a rejeição da hipótese da igualdade entre as distribuições.

Na Tabela 3, o teste para comparações entre faixas indica, para um nível de significância α=0,05, que a distribuição do número de chamadas recebidas na terceira faixa para os dados transformados é diferente das distribuições na primeira e segunda faixas. Logo, a terceira faixa tende a receber um maior número de chamadas que as outras duas.

Tabela 3: Valores-*p* do teste para comparações múltiplas entre faixas

Faixas	1	2	3
1		0,196062	**0,000022**
2	0,196062		**0,000000**
3	**0,000022**	**0,000000**	

Finalmente, dentro da terceira faixa, é importante testar se o horário de 19 horas recebe o maior número de chamadas. Para isto, usaremos mais uma vez o teste de Kruskal-Wallis. Ele testará se há diferenças significativas entre as distribuições do número de chamadas recebidas por horas.

Na Tabela 4, os valores-*p* do teste para comparações múltiplas entre horas mostraram que a distribuição para o horário de 19 horas é igual a distribuição de 18 horas e de 20 horas, se adotarmos um nível de significância de 0,05. Ou seja, estes horários tendem a receber um número igual de chamadas. Em nosso estudo usaremos o horário de 19 horas devido o fato deste horário possuir a maior variabilidade.

Tabela 4: Valores-*p* do teste para comparações múltiplas entre horas

Horas	17	18	19	20	21
17		0,197123	**0,002656**	1,000000	0,165722
18	0,197123		1,000000	1,000000	**0,000023**
19	**0,002656**	1,000000		0,173818	**0,000000**
20	1,000000	1,000000	0,173818		**0,002481**
21	0,165722	**0,000023**	**0,000000**	**0,002481**	

2.4. OBSERVAÇÕES

Após análises descritivas e testes de hipóteses, concluímos que os valores usados como base para a estimação dos parâmetros do modelo de filas *M/G/c/k* podem ser aqueles referentes ao mês de março. Podemos utilizar a segunda-feira como dia base. O turno de trabalho poderá ser o segundo e o horário escolhido, o de 19 horas. A Tabela 5 mostra os valores recomendados na estimação e controle.

Tabela 12: Valores utilizados para estimação e controle de ajuste do modelo

Referência	Recebidas	Tempo Méd. Fila	% na Fila	Fila Máxima	Tempo Máx. Fila	Eficiência	Nº. Operadores
Março	92062	1,19	62,69	49	6,54	82,33	44
Segunda-feira	5545	1,28	86,67	49	6,37	73,69	44
2º Turno	802	3,13	91,03	49	6,37	99,3	30
19 horas	802	3,13	100	49	6,34	52,07	30

3. CONCLUSÕES E OBSERVAÇÕES FINAIS

Analisamos neste artigo um banco de dados original de um *call center* de porte médio. Nossa análise foi orientada para o ajuste de um modelo de filas finitas gerais. Os dados nos permitiram caracterizar importantes variáveis de desempenho para o ajuste posterior do modelo de filas. A análise dos dados também possibilitou a utilização de métodos estatísticos não-paramétricos modernos na estimação dos parâmetros do modelo de filas,s endo importante lembrar que o uso de testes de hipóteses na análise foi uma ferramenta importante para a rápida obtenção de resultados realistas e relevantes ao estudo.

Finalmente, notamos que nossa análise acarretou um número de novas questões. Propostas de trabalho futuro incluem a estimativa dos tempos entre chegadas e tempos de serviços compatíveis com os valores encontrados pela análise descritiva. O ajuste de alguns modelos e a comparação dos resultados simulados com os reais, também são interessantes tópicos para futuras pesquisas na área [8].

4. REFERÊNCIAS BIBLIOGRÁFICAS

[1] Kendall, D. G. (1953), Stochastic Process occurring in the theory of queues and their analysis by the method of imbedded Markov chains, Annals of Mathematical Statistics, 24:338-354.

[2] Triola, M. F. (1999), Introdução à Estatística, 7ª ed. Editora LCT.

[3] Toscano, E. M. M. (2002), Estatística Aplicada (Apostila). UFMG.

[4] Sprent, P. & Smeeton, N.C. (2001), Applied Nonparametric Statistical Methods. New York: Chapman Hall.

[5] Lehman, E. L. (1975), Nonparametric: Statistical Methods Based on Ranks. San Francisco: Holden Day.

[6] Levene, H. (1960), Contributions to Probability and Statistics. Stanford University Press, CA.

[7] Siegel, S. & Castellan, N. J. (1988), Nonparametric Statistics for the Behavioral Sciences, 2ª ed. New York: McGraw-Hill.

[8] Smith, J. MacGregor & Cruz, F. R. B. (2005), The buffer allocation problem for general finite buffer queueing networks, IIE Transactions on Design & Manufacturing, 37(4): 343-365.