

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CAMPUS SÃO CARLOS  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
ENGENHARIA DE PRODUÇÃO

SIDNEY CARLOS FERRARI

**ABORDAGENS DE MODELOS DE FILAS COM ABANDONO PARA  
ANÁLISE DE CONGESTÃO EM *CALL CENTERS***

São Carlos  
2016

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CAMPUS SÃO CARLOS  
PROGRAMA DE PÓS-GRADUAÇÃO EM  
ENGENHARIA DE PRODUÇÃO

SIDNEY CARLOS FERRARI

**ABORDAGENS DE MODELOS DE FILAS COM ABANDONO PARA  
ANÁLISE DE CONGESTÃO EM *CALL CENTERS***

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção, para obtenção do título de doutor em Engenharia de Produção.

Orientação: Prof. Dr. Reinaldo Morábito

São Carlos  
2016

Ficha catalográfica elaborada pelo DePT da Biblioteca Comunitária UFSCar  
Processamento Técnico  
com os dados fornecidos pelo(a) autor(a)

F375a Ferrari, Sidney Carlos  
Abordagens de modelos de filas com abandono para  
análise de congestão em Call Centers / Sidney Carlos  
Ferrari. -- São Carlos : UFSCar, 2016.  
240 p.

Tese (Doutorado) -- Universidade Federal de São  
Carlos, 2016.

1. Call Center. 2. Contact Center. 3. Clientes  
impacientes. 4. Modelos de fila com abandono. 5.  
Distribuições mistas. I. Título.



**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

Centro de Ciências Exatas e de Tecnologia  
Programa de Pós-Graduação em Engenharia de Produção

---

**Folha de Aprovação**

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Sidney Carlos Ferrari, realizada em 22/06/2016:

---

Prof. Dr. Reinaldo Morabito Neto  
UFSCar

---

Prof. Dr. Marcelo Seido Nagano  
USP

---

Prof. Dr. Marcos Nascimento Magalhães  
USP

---

Prof. Dr. Pedro Carlos Oprime  
UFSCar

---

Prof. Dr. Roberto Antonio Martins  
UFSCar



## **DEDICATÓRIA**

A minha esposa Jane e ao meu filho Bruno meus  
maiores incentivadores

## **AGRADECIMENTO**

Ao meu orientador, Reinaldo Morábito, pela inestimável ajuda no direcionamento desta pesquisa, com muita paciência, respeito e dedicação.

Aos demais integrantes da banca de defesa da minha tese, professores Roberto Martins, Pedro Oprime, Marcelo Nagano e Marcos Magalhães, pelas valiosas contribuições proporcionadas.

Agradecimento especial e carinhoso faço a minha esposa Jane e ao meu filho Bruno pela compreensão que tiveram durante esses três anos, suportando as ansiedades, angústias e ausências inevitáveis.

Agradeço aos professores, colegas de Curso e amigos que colaboraram com seus conselhos para a realização deste trabalho e que com certeza trilhamos juntos uma etapa importante das nossas vidas.

## RESUMO

Este trabalho trata da análise de sistemas de filas para *Call Center* considerando a possibilidade do cliente abandonar o sistema antes de ser servido, devido a sua impaciência. Os *Call Centers* são organizações de serviço que predominantemente servem os clientes via chamada telefônica e são considerados um tipo particular dos Centros de Contato, que atendem seus clientes por meio de telefone, fax, e-mail, *chat*, dispositivos móveis e outros canais de comunicação. Do ponto de vista dos negócios, a principal preocupação na gestão de um *Call Center* é oferecer serviço de qualidade com mínimo custo. A percepção da qualidade dos serviços oferecidos é manifestada, pelo cliente, por exemplo, por meio do abandono da fila de espera, que é considerado uma das medidas operacionais mais importantes para avaliar o desempenho de um *Call Center*. O objetivo deste trabalho é apresentar e aplicar modelos analíticos de filas com abandono, representado por distribuições genéricas (particularmente distribuições mistas), como uma abordagem de análise efetiva para representar o problema de congestão em sistemas de *Call Centers* e apoiar decisões de dimensionamento e operação nesses sistemas. Dois estudos foram conduzidos com dados extraídos do *Call Center* de uma empresa localizada no interior do Estado de São Paulo e de outro localizado no exterior, em Israel. Os parâmetros (por exemplo, taxa de chegada, taxa de serviço e taxa de abandono) e algumas medidas de desempenho (por exemplo, tempo médio de espera, probabilidade de esperar, probabilidade de abandonar e intensidade de tráfego) são determinadas com esses dados. Essas medidas amostrais são comparadas com as mesmas medidas obtidas por meio dos modelos analíticos de fila  $M/M/c+G$ ,  $M/G^c/I+G$  e  $M/G/c+G$  considerados nesse estudo, que utilizam os parâmetros obtidos empiricamente e as distribuições mistas, Exponencial, *Fatigue Life*, Normal e Lognormal para representar o abandono (paciência) dos usuários. Observou-se como resultado, que em alguns casos, dependendo da medida de desempenho considerada, os modelos de fila com distribuição mista para o abandono têm apresentado resultados melhores (desvios menores em relação aos dados reais) que os seus correspondentes com distribuições não mistas. Observou-se, também, que independentemente das medidas de desempenho analisadas, os *Call Centers* podem ser representados por um modelo analítico de fila com uma distribuição mista para os tempos de abandono, que se comporta melhor do que os outros. Não houve evidências que as distribuições mistas para representar os tempos de abandono foram as melhores em todas as medidas de desempenho analisadas, mas foram sempre competitivas. Utilizou-se, também, um modelo experimental de simulação discreta, que representa apropriadamente o *Call Center*, para verificar os resultados dos modelos analíticos e explorar cenários alternativos. Os mesmos cenários foram considerados com o modelo teórico e as medidas de desempenho obtidas pela simulação e pelo modelo teórico foram comparadas, mostrando o potencial do uso de abordagens baseadas em modelos analíticos com abandono para análise de *Call Centers*.

Palavras-chave: *Call Center*. *Contact Center*. *Clientes impacientes*. *Modelos de fila com abandono*. *Distribuições mistas*. *Análise de congestão*. *Simulação*.

## ABSTRACT

This work deals with the analysis of queueing systems for Call Centers regarding the possibility of the customer abandon the system before being served, due to his/her impatience. The Call Centers are service organizations that predominantly serve customers by phone calls and are considered a particular kind of Contact Centers that serve their costumers through phone, fax, e-mail, chat, mobile devices and other communication channels. From a business point of view, the main concern on the management of the Call Center is the offer quality service with minimum cost. The perception of the quality of services offered is expressed, by customer, for example, through the abandonment of queue, which is considered one of the most important operational measures to evaluate the performance of a Call Center. This work aims to present and apply analytical queueing models with abandonment, represented by generic probability distributions (particularly mixed distributions), as an effective analysis approach to represent the problem of congestion in Call Center systems and support decisions of dimensioning and operations on these systems. Two studies were conducted with Call Center extracted data of a company located in the countryside of São Paulo State, Brazil and another located overseas, in Israel. The parameters (e.g, arrival rate, service rate, abandonment rate) and some measures of performance (e.g, average waiting time, waiting probability, abandonment probability and traffic intensity) were established based on these data. These sampling measures were equated with the same measures achieved by the analytical queueing models  $M/M/c+G$ ,  $M/Gc/1+G$  and  $M/G/c+G$  considered in this research, using the parameters obtained empirically and mixed, Exponential, Fatigue Life, Normal and Lognormal distributions to represent the abandonment (patience) of users. It was observed that in some cases, depending on the considered performance measure, the queueing models with mixed distributions for the abandonment have better results (minor deviations, compared to the real data) than their corresponding with non-mixed distributions. It was observed, also, that independently of the analyzed performance measures, the Call Centers may be represented by a analytical queueing model with a mixed distribution for abandonment times, which behaves better than the others. There were no evidences that the mixed distributions to represent the abandonment times were the best in all analyzed performance measures, but they were always competitive. It was used, also, an experimental model of discrete simulation that properly represents the Call Center in order to check the results of the analytical models and explore alternative scenarios. The same scenarios were considered with the theoretical model and the performance measures achieved by the simulation and by the theoretical models were compared, showing the potential of the use of the approaches based on analytical models with abandonment for the Call Centers analysis.

**Keywords:** *Call Center. Contact Center. Impatient customers. Queueing models with Abandonment. Mixed distributions. Congestion Analysis. Simulation.*

## LISTA DE FIGURAS

Gráfico 1.1	Faturamento bruto <i>Call Center</i> – Brasil.....	19
Gráfico 1.2	Posições de atendimento em <i>Call Center</i> – Brasil.....	20
Figura 1.1	<i>Call Center</i> dos casos estudados.....	28
Figura 1.2	Estrutura da tese.....	31
Figura 2.1	Diagrama esquemático das operações de um <i>Call Center</i> .....	36
Figura 2.2	Tecnologia em um <i>Call Center</i> .....	37
Gráfico 2.1	Fração de clientes na fila M/M/c vs. M/M/c+M.....	44
Figura 2.3	Dependência do desempenho sobre a distribuição da paciência.....	45
Figura 3.1	Diagrama de transição do modelo Erlang-A.....	55
Quadro 3.1	Principais modelos de fila.....	87
Figura 4.1	Fluxo das chamadas do <i>Call Center</i> do banco.....	90
Figura 4.2	Variáveis do <i>Call Center</i> do banco.....	94
Figura 4.3	Curvas de sobrevivência.....	101
Quadro 4.1	Sumário das distribuições de probabilidades - <i>Call Center</i> Banco.....	104
Gráfico 4.1	Warm up – Caso Banco – 11-12 horas.....	112
Gráfico 4.2	Warm up – Caso Banco – 12-13 horas.....	112
Gráfico 4.3	Warm up – Caso Banco – 15-16 horas.....	113
Figura 4.5	Representação gráfica da simulação do <i>Call Center</i> do banco e empresa.....	113
Figura 5.1	Fluxo das chamadas do <i>Call Center</i> da empresa.....	120
Gráfico 5.1	Distribuição do tempo entre chegadas – <i>Call Center</i> empresa.....	127
Gráfico 5.2	Distribuição do tempo de serviço – <i>Call Center</i> empresa.....	129
Gráfico 5.3	Distribuição do tempo de abandono – <i>Call Center</i> empresa.....	130
Gráfico 5.4	Estimativa da função de sobrevivência – <i>Call Center</i> empresa.....	132
Quadro 5.1	Sumário das distribuições de probabilidades - Call Center Empresa.....	134
Gráfico 5.5	Warm up – caso empresa.....	135
Gráfico 6.1	Comparação do desempenho dos modelos – <i>Call Center</i> bando 11-12 h.....	142
Gráfico 6.2	Comparação do desempenho dos modelos – <i>Call Center</i> banco 12-13 h.....	143
Gráfico 6.3	Comparação do desempenho dos modelos – <i>Call Center</i> banco 15-16 h.....	143
Gráfico 6.4	Comparação do desempenho dos modelos – probabilidade de abandonar <i>Call Center</i> do banco – 11-12 horas.....	145
Gráfico 6.5	Comparação do desempenho dos modelos – probabilidade de abandonar <i>Call Center</i> do banco – 12-13 horas.....	146

Gráfico 6.6	Comparação do desempenho dos modelos – probabilidade de abandonar <i>Call Center</i> do banco – 15-16 horas.....	146
Gráfico 6.7	Comparação do desempenho dos modelos – intensidade de tráfego <i>Call Center</i> do banco – 11-12 horas.....	147
Gráfico 6.8	Comparação do desempenho dos modelos – intensidade de tráfego <i>Call Center</i> do banco – 12-13 horas.....	148
Gráfico 6.9	Comparação do desempenho dos modelos – intensidade de tráfego <i>Call Center</i> do banco – 15-16 horas.....	148
Gráfico 6.10	Comparação do desempenho dos modelos – probabilidade de esperar <i>Call Center</i> do banco – 11-12 horas.....	149
Gráfico 6.11	Comparação do desempenho dos modelos – probabilidade de esperar <i>Call Center</i> do banco – 12-13 horas.....	150
Gráfico 6.12	Comparação do desempenho dos modelos – probabilidade de esperar <i>Call Center</i> do banco – 15-16 horas.....	150
Gráfico 6.13	Comparação do desempenho dos modelos – tempo médio espera <i>Call Center</i> do banco – 11-12 horas.....	151
Gráfico 6.14	Comparação do desempenho dos modelos – tempo médio espera <i>Call Center</i> do banco – 12-13 horas.....	152
Gráfico 6.15	Comparação do desempenho dos modelos – tempo médio espera <i>Call Center</i> do banco – 15-16 horas.....	152
Quadro 6.1	Modelo analítico de fila mais efetivo para representar o <i>Call Center</i> do Banco.....	155
Quadro 6.2	Modelo analítico de fila mais apropriado para medidas de desempenho <i>Call Center</i> do Banco.....	156
Gráfico 6.16	Comparação do desempenho dos modelos – caso empresa.....	158
Gráfico 6.17	Comparação do desempenho dos modelos – probabilidade de abandonar caso empresa.....	159
Gráfico 6.18	Comparação do desempenho dos modelos – intensidade de tráfego caso empresa.....	160
Gráfico 6.19	Comparação do desempenho dos modelos – probabilidade de esperar caso empresa.....	160
Gráfico 6.20	Comparação do desempenho dos modelos – tempo médio de espera caso empresa.....	161
Quadro 6.3	Modelo analítico de fila mais efetivo para representar o <i>Call Center</i>	

	da Empresa.....	163
Quadro 6.4	Modelo analítico de fila mais apropriado para medidas de desempenho	
	<i>Call Center</i> do Banco.....	163

## LISTA DE TABELAS

Tabela 2.1	Comparando os resultados para modelos com e sem abandono.....	45
Tabela 3.1	Verificação do modelo $M/M/c+G$ -exponencial com o modelo $M/M/c+M$ .....	84
Tabela 3.2	Verificação do modelo $M/M/c+G$ - <i>Fatigue Life</i> (alta paciência média e baixa variância).....	84
Tabela 3.3	Verificação do modelo $M/M/c+G$ - <i>Fatigue Life</i> (baixa paciência média e baixa variância).....	85
Tabela 3.4	Verificação do modelo $M/G^c/1+G$ -exponencial com o modelo $M/M/c+M$ .....	85
Tabela 3.5	Verificação do modelo $M/G/c+G$ -exponencial com o modelo $M/M/c+M$ .....	86
Tabela 4.1	Tempo entre chegadas (min).....	97
Tabela 4.2	Tempo de espera para abandono (min).....	99
Tabela 4.3	Tempo de paciência.....	100
Tabela 4.4	Tempo de serviço (min).....	102
Tabela 4.5	Sumário das estatísticas descritivas – <i>Call Center</i> Banco.....	103
Tabela 4.6	Comparação dos modelos pelo tempo médio de espera (11-12 horas).....	105
Tabela 4.7	Comparação dos modelos pela probabilidade de esperar (11-12 horas).....	106
Tabela 4.8	Comparação dos modelos pela probabilidade de abandonar (11-12 horas)....	106
Tabela 4.9	Comparação dos modelos pela intensidade de tráfego (11-12 horas).....	106
Tabela 4.10	Comparação dos modelos pelo tempo médio de espera (12-13 horas).....	107
Tabela 4.11	Comparação dos modelos pela probabilidade de esperar (12-13 horas).....	107
Tabela 4.12	Comparação dos modelos pela probabilidade de abandonar (12-13 horas)....	108
Tabela 4.13	Comparação dos modelos pela intensidade de tráfego (12-13 horas).....	108
Tabela 4.14	Comparação dos modelos pelo tempo médio de espera (15-16 horas).....	109
Tabela 4.15	Comparação dos modelos pela probabilidade de esperar (15-16 horas).....	109
Tabela 4.16	Comparação dos modelos pela probabilidade de abandonar (15-16 horas)....	109
Tabela 4.17	Comparação dos modelos pela intensidade de tráfego (15-16 horas).....	110
Tabela 4.18	Verificação dos modelos de fila – tempo médio espera e probabilidade de esperar – <i>Call Center</i> banco - 15-16 horas.....	117
Tabela 4.19	Verificação dos modelos de fila – probabilidade de abandonar e intensidade de tráfego – <i>Call Center</i> banco – 15-16 horas.....	118
Tabela 5.1	Atendimentos realizados no ano 2009 – <i>Call Center</i> empresa.....	123



Tabela 5.2	Atendimentos diários – <i>Call Center</i> empresa.....	124
Tabela 5.3	Atendimento por intervalo de hora – <i>Call Center</i> empresa.....	125
Tabela 5.4	Abandono por intervalo de hora – <i>Call Center</i> empresa.....	125
Tabela 5.5	Tempo entre chegadas (min) – <i>Call Center</i> empresa.....	127
Tabela 5.6	Tempo de serviço (min) – <i>Call Center</i> empresa.....	128
Tabela 5.7	Tempo de abandono (min) – <i>Call Center</i> empresa.....	130
Tabela 5.8	Tempo de paciência (min) – <i>Call Center</i> empresa.....	131
Tabela 5.9	Resumo das estatísticas – <i>Call Center</i> empresa.....	132
Tabela 5.10	Medidas de desempenho – <i>Call Center</i> empresa.....	133
Tabela 5.11	Verificação dos modelos de fila – tempo médio espera e probabilidade de esperar – <i>Call Center</i> empresa.....	136
Tabela 5.12	Verificação dos modelos de fila – probabilidade de abandonar e intensidade de tráfego – <i>Call Center</i> empresa.....	137
Tabela 6.1	Comparando abandono misto com não misto – caso banco – 11-12 horas.....	154
Tabela 6.2	Comparando abandono misto com não misto – caso banco – 12-13 horas.....	154
Tabela 6.3	Comparando abandono misto com não misto – caso banco – 15-16 horas.....	155
Tabela 6.4	Comparando abandono misto com não misto – caso empresa.....	162
Tabela 6.5	Probabilidade de abandonar versus quantidade de atendentes <i>Call Center</i> empresa.....	167
Tabela 6.6	Cenário 3: variação da demanda – tempo médio de espera e probabilidade de esperar – <i>Call Center</i> banco – 11-12 horas.....	169
Tabela 6.7	Cenário 3: variação da demanda – probabilidade de abandonar e intensidade de tráfego – <i>Call Center</i> banco – 11-12 horas.....	169
Tabela 6.8	Cenário 3: variação da demanda – tempo médio de espera e probabilidade de esperar – <i>Call Center</i> banco – 12-13 horas.....	171
Tabela 6.9	Cenário 3: variação da demanda – probabilidade de abandonar e intensidade de tráfego – <i>Call Center</i> banco – 12-13 horas.....	171
Tabela 6.10	Cenário 3: variação da demanda – tempo médio de espera e probabilidade de esperar – <i>Call Center</i> banco – 15-16 horas.....	173
Tabela 6.11	Cenário 3: variação da demanda – probabilidade de abandonar e intensidade de tráfego – <i>Call Center</i> banco – 15-16 horas.....	173

## LISTA DE ABREVIATURAS E SIGLAS

$\lambda$	taxa de chegada em um sistema de fila;
$\mu$	taxa de serviço em um sistema de fila;
$\theta$	taxa de abandono individual em um sistema de fila;
$c$	quantidade de servidores (canais) em um sistema de fila;
$M/M/c$	sistema de fila com tempos entre chegadas distribuídos exponencialmente; tempos de serviço com distribuição exponencial e $c$ canais de atendimento;
$M/M/c/K$	sistema de fila com tempos entre chegadas distribuídos exponencialmente; tempos de serviço com distribuição exponencial e $c$ canais de atendimento e com capacidade limitada em $K$ usuários;
$M/G/c$	sistema de fila com tempos entre chegadas distribuídos exponencialmente; tempos de serviço com distribuição genérica e $c$ canais de atendimento;
$M/M/c+M$	sistema de fila com tempos entre chegadas distribuídos exponencialmente; tempos de serviço com distribuição exponencial ; $c$ canais de atendimento e distribuição exponencial para o abandono;
$M/M/c+G$	sistema de fila com tempos entre chegadas distribuídos exponencialmente; tempos de serviço com distribuição exponencial; $c$ canais de atendimento e distribuição genérica para o abandono;
$M/G/c+G$	sistema de fila com tempos entre chegadas distribuídos exponencialmente; tempos de serviço com distribuição genérica; $c$ canais de atendimento e distribuição genérica para o abandono;
$p_n$	probabilidade dos estados do sistema;
$p_0$	probabilidade do sistema estar vazio (ocioso);
$\rho$	intensidade de tráfego;
$L_q$	número médio de usuários na fila;
$W_q$	tempo médio de espera na fila;
$W$	tempo médio gasto no sistema;
$L$	número médio de usuários no sistema;
$p_d$	probabilidade de esperar (ou de formar fila);
$p_e$	probabilidade de espera na fila;
$p_b$	probabilidade de bloqueio;

$V$	tempo de espera oferecido pelo sistema;
$W'q$	tempo de espera de um usuário impaciente;
$L(t)$	número total de usuários no sistema no instante $t$ , em serviço e na fila;
$Ws$	tempo de espera dos usuários até serem atendidos por um agente;
$Wr$	tempo de espera dos usuários até abandonarem o sistema (tempo de paciência);
$Wt$	tempo de espera de todos os usuários (os que abandonaram e os que foram servidos);
$P_A = P\{A\}$	probabilidade de abandonar.

## SUMÁRIO

<b>1. INTRODUÇÃO</b>	19
1.1 O problema de pesquisa.....	23
1.2 Objetivos.....	26
1.3 Os casos de estudo.....	26
1.4 Métodos e técnicas.....	28
1.5 Contribuição e organização do texto.....	29
<b>2 O SISTEMA CALL CENTER</b>	33
2.1 História do Call Center.....	33
2.2 Definição e descrição de Call Center .....	34
2.3 Tecnologia.....	36
2.4 Tipos de Call Center.....	38
2.5 Níveis de serviço.....	39
2.6 Dimensionamento.....	41
2.7 Abandono.....	42
<b>3 MODELAGEM DO CALL CENTER COMO SISTEMA DE FILA</b>	47
3.1 O sistema M/M/c.....	49
3.2 O sistema M/M/c/K.....	51
3.3 O sistema M/M/c+M.....	54
3.3.1 Modelagem do sistema M/M/c+M.....	54
3.3.2 Convergência da soma infinita em $p_0$ .....	58
3.4 O sistema M/M/c+G.....	65
3.4.1 Modelagem do sistema M/M/c+G.....	65
3.5 O sistema M/G/c+G.....	68
3.5.1 O sistema de fila M/G/1+G.....	69
3.5.2 O sistema de fila M/G/c+G.....	73
3.6 Resumo de modelos de filas para <i>Call Centers</i> .....	80
3.7 Verificação dos algoritmos computacionais.....	81

<b>4 O CASO DO CALL CENTER DO BANCO</b>	<b>89</b>
4.1 <b>Descrição do Call Center do banco</b> .....	89
4.2 <b>Análise do Call Center do banco</b> .....	91
4.2.1 Descrição dos dados.....	91
4.2.2 Análise dos dados.....	94
4.3 <b>Análise dos processos de chegada, abandono e serviço</b> .....	96
4.3.1 O processo de chegada.....	96
4.3.2 O processo de abandono.....	97
4.3.3 O processo de serviço.....	102
4.4 <b>Aplicação dos modelos M/M/c+M e M/M/c+G</b> .....	104
4.5 <b>Modelo de simulação</b> .....	110
4.5.1 Verificação dos modelos – intervalo 15-16 horas.....	115
<b>5 O CASO DO CALL CENTER DA EMPRESA</b>	<b>119</b>
5.1 <b>Descrição do Call Center da empresa</b> .....	119
5.2 <b>Análise estatística do Call Center da empresa</b> .....	120
5.2.1 Descrição dos dados.....	121
5.2.2 Análise dos dados.....	122
5.3 <b>Análise dos processos de chegada, serviço e abandono</b> .....	126
5.3.1 O processo de chegada.....	126
5.3.2 O processo de serviço.....	128
5.3.3 O processo de abandono.....	129
5.4 <b>Modelo de simulação</b> .....	134
<b>6 ANÁLISE DOS RESULTADOS</b>	<b>139</b>
6.1 <b>Resultados do Call Center do banco</b> .....	140
6.2 <b>Resultados do Call Center da empresa</b> .....	157
6.3 <b>Cenários alternativos</b> .....	165
6.3.1 Cenário 1: quantidade de atendentes.....	167
6.3.2 Cenário 2: nível de serviço.....	168
6.3.3 Cenário 3: variação da demanda.....	168

<b>7 CONCLUSÕES E PESQUISAS FUTURAS</b>	<b>175</b>
7.1 <b>Conclusões</b> .....	175
7.2 <b>Pesquisas futuras</b> .....	178
<b>REFERÊNCIAS</b>	<b>181</b>
<b>APÊNDICE A – Análise dos dados do <i>Call Center</i> do banco</b> .....	<b>187</b>
<b>APÊNDICE B – Comparando os modelos com e sem abandono</b> .....	<b>195</b>
<b>APÊNDICE C – Análise do processo de chegada, abandono e serviço do caso <i>Call Center</i> do banco</b> .....	<b>201</b>
<b>APÊNDICE D – Verificação dos modelos – intervalo 11-12 horas e 12-13 horas</b> .....	<b>213</b>
<b>APÊNDICE E – Razão entre a mediana e média da Distribuição Exponencial</b> .....	<b>219</b>
<b>APÊNDICE F – Distribuição Fatigue Life</b> .....	<b>221</b>
<b>APÊNDICE G – Distribuição Fréchet</b> .....	<b>223</b>
<b>APÊNDICE H– Distribuições mistas</b> .....	<b>225</b>
<b>APÊNDICE I – Regra do trapézio</b> .....	<b>231</b>
<b>ANEXO A – Amostra dos dados coletados do <i>Call Center</i> do banco</b> .....	<b>237</b>
<b>ANEXO B – Amostra dos dados coletado do <i>Call Center</i> da empresa</b> .....	<b>239</b>

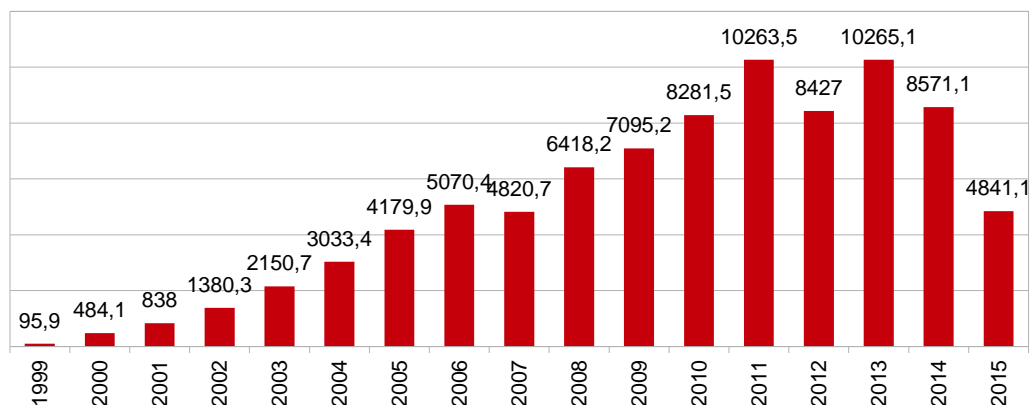
## 1 INTRODUÇÃO

Os *Call Centers*, também denominados de Centros de Contato (*Contact Centers*), têm sido atualmente um dos principais canais de comunicação entre as empresas e seus clientes em diversos setores. Conhecer o cliente para oferecer o que ele precisa torna-se uma exigência básica para que as operações sejam cada vez mais customizadas, satisfazendo as suas expectativas. Os *Call Centers* são organizações de prestação de serviço que predominantemente servem os clientes via chamadas telefônicas e são considerados um tipo particular dos Centros de Contato que atendem seus clientes por meio de telefone, fax, e-mail, *chat*, dispositivos móveis e outros canais de comunicação.

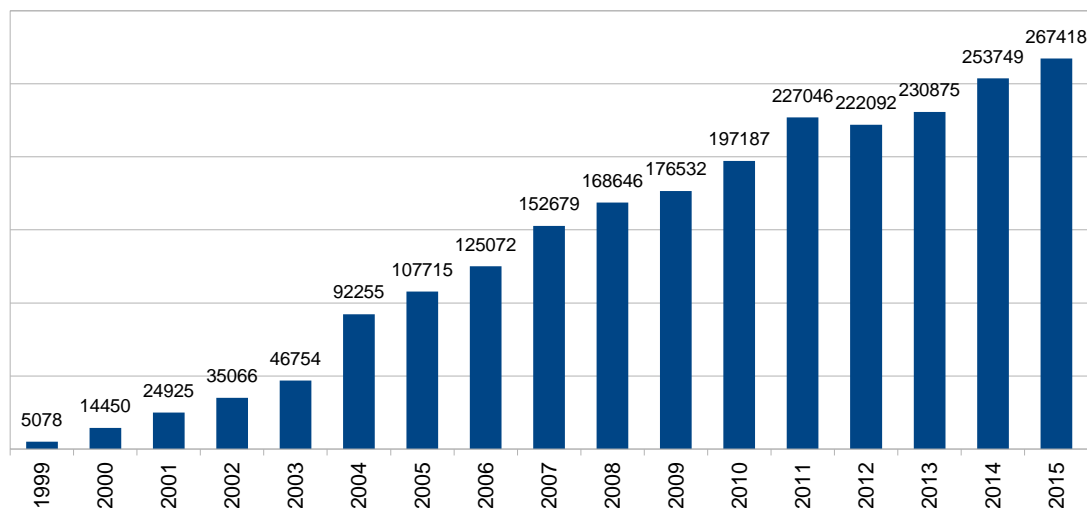
A indústria de *Call Centers* tem se expandido rapidamente e o mercado brasileiro desses serviços é considerado um dos mais importantes da economia. Este setor é fundamental para os avanços das empresas na gestão dos clientes, além de ser um dos maiores empregadores, oferecendo muitas vezes, a oportunidade do primeiro emprego. A comprovação desses fatos pode ser observada nos Gráficos 1.1 e 1.2 que mostram a evolução histórica da expansão dessa indústria de serviços em termos econômicos e da força de trabalho.

GRÁFICO 1.1 - Faturamento bruto *Call Center* – Brasil

### Faturamento bruto (milhões R\$)



Fonte: callcenter.inf, 2015

GRÁFICO 1.2 - Posições atendimento *Call Center* – Brasil**Posições de atendimento**

Fonte: callcenter.inf, 2015

Observando-se os dois gráficos, nota-se um crescimento contínuo tanto no faturamento bruto quanto nas posições de atendimento. O mercado brasileiro de *Call Centers* teve um crescimento exponencial no seu faturamento até 2011, chegando a movimentar 10,2 bilhões de reais. Houve uma retração do faturamento em 2012 motivada pela desaceleração da economia, segundo os analistas do setor, que se recuperou em 2013 com o reposicionamento do faturamento no patamar de 2011. A diminuição do faturamento em 2014 e 2015 é explicada, pelos especialistas, pelos eventos que deixaram o ano mais curto (Copa do Mundo e Eleições) e pelas incertezas econômicas e políticas ocorridas durante o ano 2015. Apesar disso, segundo eles, as empresas de *Call Centers* devem apresentar crescimento em 2015, buscando constante inovação e eficiência (CALLCENTER.INF, 2015).

O número de posições de atendimento próprias instaladas no Brasil cresceu 9,91%, em 2014, chegando a 253,7 mil posições (Gráfico 1.2). Foram criados 45 mil novos postos de trabalho, representando um crescimento de 13,35% em 2014. Esse crescimento continua em 2015 atingindo 267,4 mil posições de atendimento, representando um aumento de 5,4% (CALL CENTER.INF, 2015). O motivo desse crescimento justifica-se pelo aumento da demanda por serviços de terceirização de *Call Centers* e também pelo aumento da oferta de serviços por parte dos provedores de serviços. Esse crescimento exigiu da indústria de *Call Center* investimentos em novas tecnologias e na capacitação dos profissionais do setor, visando melhoria na qualidade



e eficiência dos serviços. Um novo conceito em operação de *Call Center* surgiu, trazendo modificações na gestão de processos em busca da melhoria contínua dos serviços prestados com redução de custos, profissionais mais capacitados, aumento da produtividade e flexibilidade e agilidade no atendimento (BRAZ, 2010).

Essas mudanças podem ser percebidas na maneira mais eficiente dos *Call Centers* se relacionarem com os clientes. Eles estão antecipando as chamadas e não ficam, como costumeiramente, apenas aguardando a ligação dos clientes para solucionar os seus problemas ou obter alguma informação. Ampliam, também, a possibilidade do autoatendimento por meio das tecnologias móveis dos *smatphones* e *tablets*, reduzindo as chamadas nas centrais de atendimento. Com a necessidade de se antecipar às chamadas e interagir no universo multimídia para melhor se relacionar com os clientes, os Centros de Contatos se traduzem no conceito mais moderno de centrais de atendimento, reunindo em um único local o atendimento por telefone, fax, *e-mail*, *website* e *chat*, permitindo o melhor contato entre o cliente e a empresa. Com a implementação de bancos de dados complexos e a internet, que facilita e agiliza o trânsito das informações, os Centros de Contatos aprimoram o gerenciamento de todas as relações da empresa com a sociedade onde atua, traçando um perfil dos seus consumidores. Assim, a gestão de um *Call Center* torna-se uma atividade complexa e os seus gestores, segundo Lima (2012), devem selecionar um conjunto de métricas mais adequadas que darão os subsídios para a tomada de decisões estratégicas e operacionais.

Como o propósito deste trabalho é analisar *Call Centers* sob o ponto de vista quantitativo, decidiu-se utilizar como métricas para as análises a serem realizadas, entre aquelas consideradas importantes para os gestores, o abandono dos usuários, medido pela probabilidade de abandonar, que é um indicador existente nos contratos mantidos entre os *Call Centers* e as empresas e que ocorre quando o usuário desliga o telefone antes de ser atendido por um agente (LIMA, 2012). Além desta métrica, também são utilizadas neste trabalho o tempo de espera, medido por sua média, que indica o tempo que o usuário espera para ser atendido, a probabilidade de esperar e a intensidade de tráfego que mede o congestionamento do sistema. São comparados os valores dessas métricas extraídas dos dados observados de um *Call Center* com aqueles obtidos por meio dos modelos matemáticos e de simulação.

A escolha dessas métricas foi motivada considerando que a principal preocupação na gestão de um *Call Center* é conseguir um nível de serviço de qualidade, sujeito a um orçamento especificado. Essa qualidade pode ser medida de duas maneiras: qualitativa e quantitativa (MANDELBAUM e ZELTYN, 2005). Do ponto de vista quantitativo, a qualidade dos serviços é percebida pelos usuários principalmente pelo tempo de espera na fila até ser atendido pelos

agentes. Quando esse tempo de espera é longo, o usuário se cansa e pode abandonar a fila, ficando sem resposta para a sua solicitação. O abandono torna-se uma medida de desempenho operacional importante e é por meio dele que os usuários manifestam a sua percepção subjetiva sobre os serviços oferecidos. Aqueles que abandonam declaram desta maneira que não vale a pena esperar pelos serviços oferecidos. Por outro lado, os usuários que não abandonam o sistema somente manifestam a sua percepção sobre os serviços oferecidos se forem questionados após o término do atendimento.

O fenômeno do abandono pode ser provocado pelo dimensionamento incorreto da quantidade de agentes, fazendo com que aumente o tempo de espera na fila. Por isso, ignorar o abandono produz distorções nas informações do sistema, que são importantes para as decisões a serem tomadas pelos gestores. Considerar explicitamente o abandono nos modelos analíticos para dar suporte em *Call Center* pode deixar o sistema mais preciso, em função da capacidade que se aloca no *Call Center*, diminuindo o tempo de espera na fila. A inclusão do abandono nos modelos analíticos que representam o *Call Center* permite aos gestores interferir nas decisões, explorar cenários e melhorar o nível de serviço dos usuários. (MANDELBAUM e ZELTYN, 2005).

Desse modo, o abandono, medido pela paciência do usuário, pode ser considerado como uma das medidas operacionais mais importantes, senão a mais importante, para avaliar o desempenho de um *Call Center*. Certamente, com menor tempo de espera, poucos clientes abandonam devido à impaciência e, em geral, essa menor espera implica em um melhor serviço. O abandono se diferencia das outras medidas de desempenho porque desta maneira os clientes informam ao sistema sua percepção subjetiva do serviço oferecido, ou que estão recebendo. Outras medidas como a taxa de retorno após o abandono, ou depois de ser servido, não fornecem essa informação. O mesmo acontece com as medidas mais utilizadas para avaliar o desempenho, como o tempo de espera, que são objetivas para o nível do sistema, mas não relatam as experiências dos serviços dos clientes.

Algumas tentativas têm sido realizadas para ajustar o comportamento do abandono por diferentes modelos clássicos de distribuição de probabilidade (BACCELLI e HEBUTERNE, 1981; MANDELBAUM e ZELTYN, 2005; MANDELBAUM e ZELTYN, 2012). Entretanto, devido às características distintas na população de usuários, em geral, pode ser interessante o uso de distribuições mistas de probabilidade (Apêndice H) para absorver melhor essas nuances populacionais. Entende-se como distribuições mistas (ou misturas), segundo Forbes (2011), uma distribuição de probabilidades que é uma combinação de duas ou mais distribuições de

probabilidades específicas, como por exemplo, uma distribuição normal e uma distribuição exponencial. Essas distribuições são chamadas de componentes mistas.

Esses modelos de mistura conseguem representar a presença de subpopulações dentro de uma população geral (MANDELBAUM e ZELTYN, 2012; OLIVEIRA, 2009). Neste estudo, as subpopulações são representadas pelos grupos de clientes que possuem seu próprio padrão para o tempo de paciência, formando uma população com a mistura desses grupos com características diferentes. Essas considerações justificam a preocupação em considerar o abandono na modelagem matemática dos *Call Centers* e analisar os padrões do tempo de paciência dos clientes e sua implicação nas medidas de desempenho do sistema *Call Center*, como um tema a ser aqui estudado e desenvolvido.

### 1.1 O problema de pesquisa

As operações e a gestão do desempenho de um *Call Center* são baseadas em princípios científicos. É possível medir a qualidade dos serviços, fazendo uso dos modelos estocásticos da teoria de filas. Esta teoria foi concebida por A. K. Erlang no início do Século XX e se desenvolveu como um tema central de pesquisa da Pesquisa Operacional (WOLFF, 1989). Em um modelo de fila de um *Call Center*, os usuários são as pessoas que telefonam para obter informações ou resolver um problema particular, os servidores são os agentes do sistema que atendem os usuários, ou são os equipamentos de comunicação, e as filas são os usuários que aguardam para serem servidos.

Um modelo de fila amplamente usado em *Call Centers* para quantificar o nível de serviço operacional em termos de algumas medidas de desempenho e congestionamento é o bem conhecido modelo M/M/c. Neste modelo, o processo de chegada dos usuários no sistema é considerado como um processo de Poisson (indicado pelo primeiro M na notação) e a duração do tempo de serviço é suposta distribuída exponencialmente (indicado pelo segundo M na notação), com  $c$  servidores idênticos operando em paralelo ( $c$  na notação), atendendo a uma única fila de espera de usuários. Este sistema de fila não reconhece o comportamento do abandono dos usuários e também as várias classes de usuários, ficando restrito para absorver as características do processo de chegada de usuários dos *Call Centers* (WHITT, 1993).

Para suprir essa falha, vários modelos de filas que incorporam o abandono têm sido estudados na literatura (BACCELLI, HEBUTERNE, 1981; MANDELBAUM, SAKOV, ZELTYN, 2001; BROWN *et al.*, 2002; MANDELBAUM, ZELTYN, 2005). Esses modelos

consideram uma única classe de família de distribuições paramétricas para modelar a paciência, sendo as mais comuns a distribuição exponencial, Weibull e Erlang. Entretanto, é comum encontrar em *Call Center* uma população de clientes heterogênea, normalmente constituída de uma mistura de grupos com características diferentes. Cada um deles tem o seu próprio padrão para o tempo de paciência, sendo uns mais pacientes que outros. Por esse motivo, em geral, não é razoável modelar a paciência por meio de uma única classe de distribuição, mas ajustá-lo com uma distribuição mista de probabilidade (mistura de distribuições).

Com base nessas considerações, investigar e utilizar distribuições mistas de probabilidades que melhor aderem aos tempos de paciência é um tópico interessante de pesquisa, por exemplo, para melhorar os modelos de filas  $M/M/c + G$  e  $M/G/c + G$ , utilizados no dimensionamento dos *Call Centers*. Na notação da fila  $M/M/c+G$ , as três primeiras letras possuem o mesmo significado que no modelo  $M/M/c$  e o  $G$  indica que os tempos de paciência dos clientes obedecem uma distribuição qualquer ou genérica. Analogamente, na fila  $M/G/c+G$ , o primeiro  $G$  indica que a duração de cada serviço tem distribuição genérica e o último  $G$ , que os tempos de paciência também possuem distribuição genérica.

Nesta tese, uma questão de pesquisa a ser investigada é: modelos analíticos de filas, incorporando o abandono e baseados em distribuições genéricas (particularmente mistas) para modelar a paciência, são efetivos para representar o problema de congestão em *Call Center*? Algumas hipóteses podem, inicialmente, ser sugeridas, como:

- Os modelos de filas que utilizam distribuições genéricas (particularmente mistas) para modelar a paciência, em geral, captam melhor as diferentes características dos vários tipos de clientes e descrevem melhor o desempenho de um sistema;
- Os modelos de filas que utilizam distribuições genéricas (particularmente mistas) para modelar a paciência não apresentam diferenças significativas no desempenho do sistema, quando comparados com os modelos que utilizam uma única classe de distribuição.

Outras questões também são consideradas no estudo do problema desta pesquisa:

- i) Os modelos de fila disponíveis capturam bem a realidade do *Call Center*?
- ii) Os modelos que utilizam distribuições mistas para representar a paciência são melhores que os modelos que utilizam uma única classe de distribuição paramétrica para modelar o abandono?

iii) Existe um modelo que se comporta melhor que qualquer outro em todas as medidas, utilizando ou não distribuições mistas para modelar a paciência?

Observa-se que algumas misturas de distribuições foram utilizadas em estudos da literatura para ajustar empiricamente o tempo de paciência em *Call Center*, como em Mandelbaum e Zeltyn (2012). Contudo, não foram inseridas nos modelos de filas para observar o seu impacto nas medidas de desempenho. Em Oliveira (2009), foi analisada a relação entre o tempo médio de espera na fila e a probabilidade de abandono dos clientes para distribuições de paciência do tipo mista no modelo de fila  $M/M/c+G$ , aplicado a *Call Center*. Foi considerado um ambiente em que mensagens eletrônicas são transmitidas aos clientes que aguardam atendimento, informando a sua posição e o seu tempo de espera na fila. Observou-se se essas gravações eletrônicas podem interferir na paciência dos clientes enquanto esperam na fila. Usando simulação e distribuição da paciência do tipo mista, mostrou-se que a exponencial mista e a uniforme mista apresentaram menores tempos de espera na fila, com dados coletados de *Call Centers*, quando comparadas com as distribuições exponencial e uniforme de mesmos parâmetros.

Este trabalho não trata particularmente do desenvolvimento formal de novos modelos matemáticos de congestão em sistemas estocásticos, que já estão relativamente bem desenvolvidos pela teoria de filas. A preocupação maior é estudar aprofundadamente este tema e realizar um complemento empírico dessa teoria, área reservada à ciência de congestão em fila. Pretende-se validar e calibrar modelos teóricos de filas via análise científica baseada em dados reais, por meio de casos estudados em *Call Centers* de uma fábrica de *software* e de um banco. Raros estudos de aplicações foram encontrados na literatura (BROWN, et al., 2002), em que ocorre um esforço sistemático para validar resultados teóricos de filas aplicados a *Call Centers*, principalmente abordando o fenômeno do abandono.

Também é utilizado um modelo experimental de simulação discreta, que também represente apropriadamente o sistema de *Call Center* analisado, para comparar e verificar os modelos analíticos utilizados, e explorar junto com os modelos analíticos, diferentes cenários alternativos contemplando possíveis aumentos na demanda, taxa crescente da paciência dos clientes, diminuição na capacidade de serviço, entre outras. As medidas de desempenho obtidas pela simulação e via modelos analíticos de filas são comparadas com o propósito de também avaliar a sensibilidade dos modelos. Esses modelos e resultados poderão ser úteis para os gestores melhor configurarem seus sistemas e gerenciarem suas operações, assim como planejarem mais efetivamente o futuro.

## 1.2 Objetivos

O objetivo desta tese é analisar a aplicação de modelos analíticos de filas com abandono, representados por distribuições genéricas (particularmente mistas), como uma abordagem de análise efetiva para representar o problema de congestão em sistemas de *Call Centers* e apoiar decisões de dimensionamento e operação nesses sistemas. São considerados como objetivos específicos:

- Verificar a aderência de distribuições mistas, por meio do teste *Kolmogorov-Smirnov*, para representarem o comportamento da paciência na fila de espera dos *Call Centers*;
- Analisar a aplicação desses modelos de filas com abandono em dois casos reais de *Call Centers*, com dados de uma empresa no Brasil e outra no exterior;
- Comparar os desvios das medidas de desempenho computadas pelos modelos analíticos com as obtidas por simulação discreta e as observadas com os dados reais dos casos;
- Explorar cenários alternativos para a configuração do sistema, com a intenção de mostrar o potencial das abordagens e melhorar as decisões dos gestores.

## 1.3 Os Casos de estudo

Dois *Call Centers* são analisados neste trabalho. Um deles é o *Call Center* de uma empresa do Estado de São Paulo, denominado neste estudo de *Call Center* da Empresa. O outro é o *Call Center* de um banco de Israel, denominado de *Call Center* do Banco.

O *Call Center* da Empresa interage com os clientes de uma empresa com sede no interior do Estado de São Paulo que, desde 1992, desenvolve softwares especializados no gerenciamento de farmácias e drogarias. A primeira versão do software foi desenvolvida e implantada em algumas drogarias da cidade. Com a evolução do sistema e a satisfação dos clientes, a empresa logo conseguiu centenas de clientes no interior do Estado de São Paulo e no norte do Estado do Paraná. Atualmente está presente nos estados de São Paulo, Paraná, Mato Grosso do Sul, Mato Grosso, Rio Grande do Sul, Piauí, Minas Gerais e Goiás.

Os clientes, quando acessam o *Call Center* da empresa por meio de uma chamada telefônica, passam por uma triagem na Unidade Interativa de Voz (IVR), algumas vezes denominada de Unidade de Resposta por Voz (VRU). Esse equipamento é essencial para o *Call Center* porque permite a identificação do usuário por meio do seu número de telefone e também

oferece um menu e, em função da opção selecionada, transfere automaticamente a chamada para o atendente. Além disso, provê serviços automáticos aos usuários, fornecendo informações extraídas de um banco de dados, tais como saldo ou pedidos de compras para que ocorra o autoatendimento. Durante esta fase, o usuário se identifica e escolhe uma das quatro opções fornecidas para solucionar os seus problemas, sendo elas:

Opção 1: suporte técnico, que oferece serviços relacionados aos problemas com os *softwares*. O atendimento nessa opção se divide em atendimento telefônico, quando o cliente liga para o *Call Center*, e o atendimento *on-line*, em que o cliente é atendido via internet por meio de um *chat*.

Opção 2: Loja virtual, que atende os clientes que desejam comprar produtos da empresa.

Opção 3: Administração, que atende os clientes com problemas em contratos.

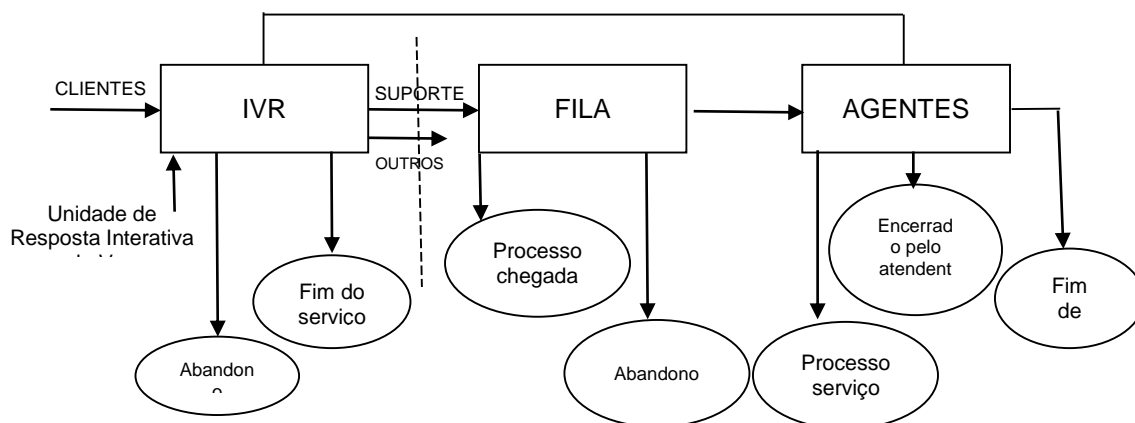
Opção 4: Financeiro, que atende os serviços relacionados com mensalidades, boletos e outros serviços financeiros.

As opções 2, 3 e 4 possuem baixo fluxo de ligações e com pouca regularidade, sendo então, por simplicidade, desconsideradas neste estudo. É possível também neste estágio o usuário resolver o seu problema e, então, completar o seu serviço ou, ainda, abandonar o sistema a partir da IVR. Aqueles usuários que desejam conversar com um atendente possuem duas opções nesse ponto: começar o seu serviço imediatamente, se houver um atendente livre, ou juntar-se à fila de espera e aguardar o atendimento. Enquanto aguardam na fila, os usuários impacientes podem julgar que não vale a pena esperar mais tempo para serem atendidos e abandonar a fila, deixando o sistema. Os usuários que são atendidos, no diálogo com o atendente, podem completar o seu serviço desejado e deixar o sistema, ou podem ter seu diálogo encerrado pelo próprio atendente, sem um motivo justificável. Uma representação básica deste *Call Center* pode ser vista na Figura 1.1.

O *Call Center* do banco serve os clientes de um banco de Israel, que oferece vários tipos de serviços, como informações para clientes atuais e futuros, transações de contas correntes e contas de poupança, negociação de ações e suporte técnico para os usuários de internet do site do banco. O cliente acessa o *Call Center* por meio de uma chamada telefônica, conectando-se a uma IVR. Neste ponto, o cliente se identifica e escolhe uma das opções disponíveis para realizar o seu serviço. A partir da IVR, o cliente pode abandonar o sistema sem solucionar o seu problema ou resolver o seu problema e deixar o sistema ou, ainda, conectar-se com um atendente, se houver algum livre. Caso não encontre atendentes livres, o cliente aguarda na fila para ser atendido. Nesta etapa, ele pode abandonar o sistema, depois de um longo tempo de espera. Aqueles que são atendidos podem completar o seu serviço, após o diálogo com o

atendente, ou podem ser desconectados pelo próprio atendente. A mesma Figura 1.1 também pode ser usada para representar o fluxo das chamadas deste *Call Center*.

FIGURA 1.1 - *Call Center* dos casos estudados



Fonte: elaborada pelo autor

Neste trabalho, os processos de chegada, abandono e serviço são analisados a partir do instante em que o usuário ingressa na fila de espera, após a sua triagem na IVR. Portanto, a chegada, o abandono e o serviço completado na IVR são desconsiderados.

#### 1.4 Métodos e técnicas

Neste estudo, adotou-se a concepção indutivista, partindo da observação dos fatos ocorridos num fenômeno de interesse, que é o tempo de paciência dos clientes na fila de espera de *Call Centers* nos casos de estudos propostos (DEMO, 2000). A abordagem quantitativa foi utilizada, justificada pelo fato de se ter uma variável bem definida, como, por exemplo, o tempo de paciência na fila de espera do *Call Center*, e por fazer uso de modelos matemáticos da teoria de filas. Estão presentes a mensuralidade por meio da coleta de dados nos casos estudados, a causalidade em procurar explicar o comportamento da paciência, a generalização ao adaptar e aplicar modelos de fila existentes e a replicação possibilitando a oportunidade em se reproduzir a pesquisa. A modelagem foi o método utilizado da abordagem quantitativa, porque existem nos casos estudados (os *Call Centers*) um problema a ser estudado, representado pelo abandono dos usuários da fila de espera.



A pesquisa quantitativa empírica descritiva (BERTAND E FRANSOO, 2002; MORABITO E PUREZA, 2010) foi a técnica de pesquisa utilizada com a finalidade de analisar duas situações ainda pouco exploradas na literatura: o uso de modelos de filas com distribuições genéricas (particularmente mistas) representando a paciência e a validação desses modelos de fila com análise baseada em dados extraídos de *Call Centers*. A abordagem quantitativa empírica se justifica pelos estudos e análises realizados nos *Call Centers*, observando que as distribuições da paciência são genéricas, com forte potencial para aplicações de distribuições mistas. Além disso, os modelos analíticos de filas existentes na literatura são adaptados a essa realidade, com base nas observações e dados desses casos estudados, e utilizados para analisar esses *Call Centers*.

Diferentes parâmetros de interesse foram extraídos desses dados, como taxas e distribuições de chegada, paciência e serviço, que são consideradas nos modelos de filas, utilizando distribuições genéricas (particularmente mistas) para representar a paciência. A análise dos dados foi realizada por meio da estatística descritiva, por testes de aderência, como *Kolmogorov-Smirnov* e outros, para ajustar as distribuições de probabilidades aos dados observados nos casos estudados, e também pelo teste de *Kaplan-Meier*, aplicado aos dados censurados para inferir o tempo médio de paciência. Analisando e comparando algumas medidas de desempenho obtidas pelos modelos de filas com as mesmas medidas extraídas dos dados reais amostrados, pretende-se verificar qual o modelo de fila mais adequado para representar apropriadamente os *Call Centers*.

Diferentes cenários também são estudados e avaliados, usando também um modelo experimental de simulação discreta, que represente detalhadamente e acuradamente os casos de estudo, para analisar a corretude e a sensibilidade dos modelos teóricos de filas diante de modificações provocadas na demanda, capacidade de serviço e impaciência dos clientes, entre outras. A simulação também é utilizada para analisar e validar modelos analíticos de fila com distribuições genéricas (particularmente mistas) para a paciência. O diagrama da Figura 1.2 resume essas discussões.

## **1.5 Contribuições e organização do texto**

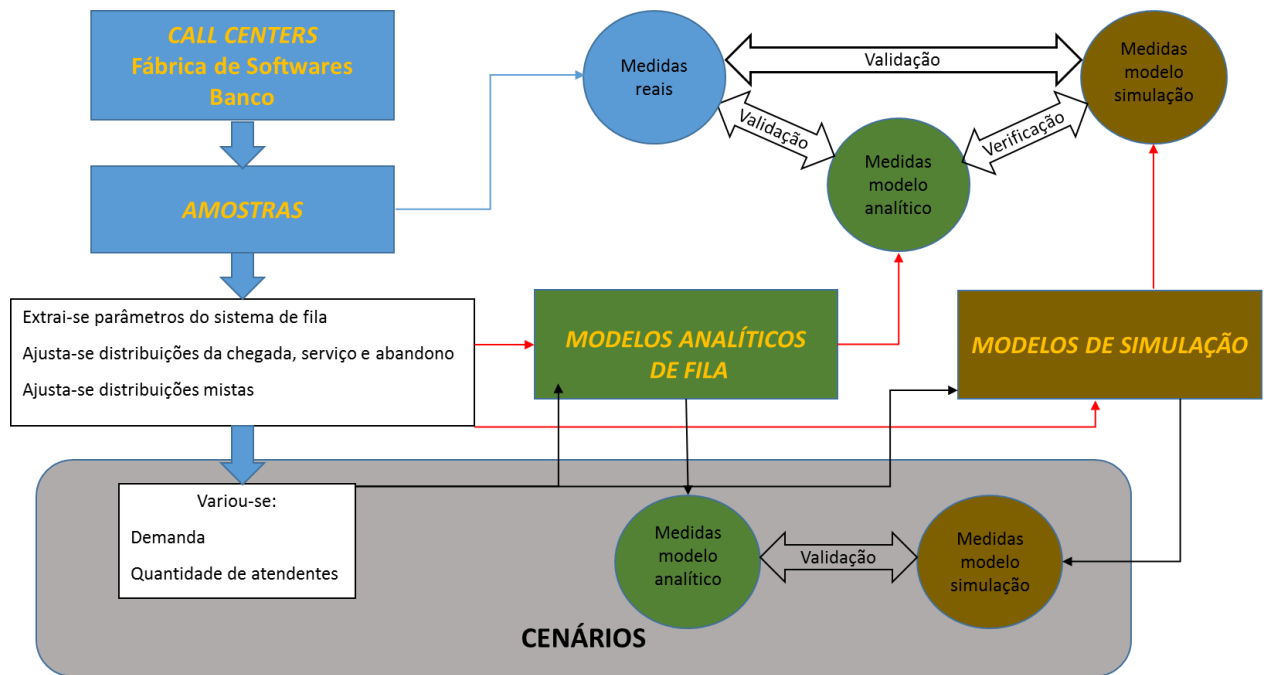
A relevância deste estudo está em estudar e aplicar modelos analíticos de filas, desenvolver e propor métodos de análises de desempenho de *Call Center*, e em considerar distribuições genéricas (particularmente mistas) de probabilidade para ajustar o tempo de

paciência dos usuários nas filas de *Call Center*, incorporando essas distribuições nos modelos de filas com abandono. Acredita-se que confrontando os resultados obtidos destes modelos com os extraídos empiricamente, obter-se-á boas aproximações da realidade. Esse tipo de análise foi pouco explorado na literatura e entende-se que poderá trazer um avanço do conhecimento por meio da avaliação e validação de modelos de fila com abandono para aplicação em *Call Centers*. O abandono na literatura existente, em geral, tem sido modelado por meio de uma única classe de distribuição de probabilidade. Ele tem sido pouco discutido do ponto de vista das distribuições mistas de probabilidade, em geral mais adequadas para captar a heterogeneidade do tempo de paciência de cada cliente.

O conhecimento gerado neste trabalho poderá produzir melhorias na compreensão e gestão desses sistemas, deixando-os mais ajustados em função das capacidades alocadas, e permitindo aos gestores interferirem nas decisões, explorando cenários que otimizem a qualidade do nível de serviço. Espera-se que sistemas mais calibrados melhorem a satisfação dos clientes com relação aos serviços prestados, em função da diminuição dos tempos de espera e conseqüente redução das frações do abandono. Espera-se que os métodos de análise propostos nesta tese ajudem as empresas a melhor dimensionarem as suas equipes de atendentes, melhorando os desempenhos dos sistemas e reduzindo seus custos.

Este estudo está estruturado do seguinte modo: no Capítulo 2, descreve-se resumidamente o sistema *Call Center* desde a sua história, contemplando a sua estrutura e funcionamento. No Capítulo 3, apresenta-se uma revisão da literatura sobre os modelos de filas para *Call Center*. No Capítulo 4, descreve-se o caso do *Call Center* do Banco, apresentando em detalhes o sistema, as análises estatísticas, os resultados dos modelos  $M/M/c$  versus  $M/M/c+M$ , as análises dos processos de chegada dos dados coletados, paciência e serviço e aplicação dos modelos  $M/M/c+M$  e  $M/G/c+G$ . No Capítulo 5, apresenta-se o caso do *Call Center* da Empresa, de maneira similar ao Capítulo 4. No Capítulo 6, encontra-se a análise dos resultados obtidos e apresentação e avaliação de cenários alternativos. Finalmente, no Capítulo 7, as conclusões do estudo são apresentadas.

FIGURA 1.2 - Estrutura da tese



Fonte: elaborada pelo autor



## 2 O SISTEMA *CALL CENTER*

Neste capítulo, descreve-se resumidamente os sistemas de *Call Center*, com a finalidade de apresentar suas estruturas, funcionamento, preocupações com a qualidade dos serviços oferecidos e suas evoluções até os atuais Centros de Contato.

### 2.1 História do *Call Center*

O *Call Center* como entende-se hoje, constituído por um sistema sofisticado, envolvendo alta tecnologia e em que se deve aguardar algum tempo para se solucionar um problema, nem sempre foi assim. Segundo Mancini (2001), esse tipo de interação com os elementos da sociedade teve sua origem nos serviços específicos de corpo de bombeiros, emergência médica e policiamento urbano. O poder de relacionamento nesses serviços foi rapidamente percebido pelas empresas, que passaram a utilizá-lo com os seus clientes, tratando cada um deles como únicos.

Em 1880, ocorreu o início da era *Call Center*, quando um vendedor de doces resolveu utilizar o telefone para vender o seu produto. Esse episódio é considerado o mais antigo desse tipo de serviço, denominado atualmente por televendas. O vendedor formou uma equipe com mais de cem pessoas, que entravam em contato com os clientes, por telefone, para oferecer os seus produtos. Após esse ponto de partida, houve uma expansão do uso do telefone abrangendo os serviços de cobrança e supervisão. Nos anúncios de propaganda, foram incorporados os números de telefones para que as pessoas pudessem fazer os seus pedidos de produtos e serviços, iniciando uma nova fase de cadastro e *mailing*. Baseado no sucesso obtido por essas atividades, ocorreu uma rápida disseminação dessa estratégia com a Ford, investindo em uma campanha maciça de marketing por telefone. Na década de 70, era frequente as pessoas receberem telefonemas propondo vendas ou pesquisas, o que culminou na década de 80 na criação oficial do termo “telemarketing”.

Com a evolução da informática nos anos 90, o conceito de telemarketing evoluiu para o modelo de *Call Center*, apresentando uma visão mais ampla de como impulsionar novas vendas, antecipar as necessidades e manter a marca da empresa sempre na mente do consumidor. Abraçando os diferentes canais de comunicação como o telefone, *chat*, *e-mail*,

internet, fax, entre outros, um novo conceito de *Call Center* apareceu, o Centro de Contato, que centraliza o gerenciamento das relações entre a empresa e seus clientes (MANCINI, 2001).

Alguns conceitos foram evoluindo juntamente com a evolução histórica do *Call Center*. Destaca-se o de *televendas* que simplesmente usava o telefone para apresentar produtos e serviços com o objetivo de efetivar vendas. Com o *telemarketing*, o telefone deixa de ser utilizado somente para realizar vendas e passa a ser uma ferramenta de marketing, agregando outras funções como a de observar o mercado, interagir com o cliente, realizar pesquisas e propagandas. O *call center* vai além do marketing e se preocupa mais com o cliente, procurando conhecê-lo melhor e atender as necessidades do mercado. O Centro de Contato ampliou a maneira da empresa relacionar-se com o cliente, utilizando-se de todos os veículos de comunicação.

Inicialmente o telefone era utilizado para apoiar vendas e cobranças. Com o avanço da telefonia, surgiram as centrais para vendas a distância espalhadas geograficamente. Posteriormente o sistema passou a ser focado no cliente e atualmente com o *Call Center*, ou o Centro de Contato, consegue-se traçar o perfil de cada consumidor. Os *Call Centers* estudados, nesta tese, enquadram-se na categoria de Centro de Contato, relacionando-se com os clientes da empresa por telefone, *chat*, *e-mail* ou internet, por meio de uma única central de atendimento. Embora possuam várias possibilidades de comunicação com os clientes, neste estudo são consideradas nas análises apenas os contatos por telefones.

## **2.2 Definição e descrição de *Call Center***

Basicamente, um *Call Center* é um sistema composto por recursos humanos, computadores e equipamentos de telecomunicação que presta serviços à distância. É um importante meio de comunicação entre clientes e a empresa para prover serviços solicitados pelos clientes para a empresa, como também serviços requisitados pela empresa para os clientes, como nos casos de crédito e cobrança (KOOLE e MANDELBAUM, 2002). Sob o aspecto econômico, o *Call Center* pode ser considerado um sistema capaz de unir recursos de voz e dados permitindo às empresas reduzirem custos, aumentarem os rendimentos e estabelecerem boas estratégias de negócios. As atividades de um *Call Center* são principalmente direcionadas para vendas, ações de marketing, suporte técnico e serviços de atendimento aos clientes (DOMBACKER, 2010).

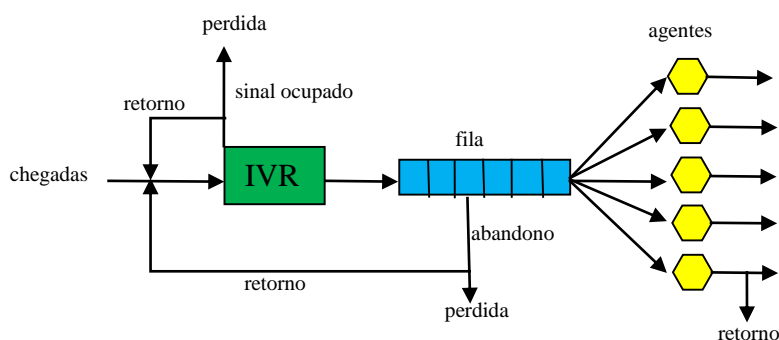
Tradicionalmente, o ambiente de trabalho de um *Call Center* é formado por corredores que comportam as baias com atendimentos individuais, onde o atendente trabalha sozinho com o cliente invisível, sem a colaboração dos demais, utilizando terminais computadorizados. Um novo modelo nessa linha de produção foi, então, proposto por meio das células de atendimento, onde as baias são colocadas em círculos, possibilitando um trabalho de atendimento em equipe. Essas células de atendimento são especializadas por classes de clientes, tipo de produtos e de empresas, proporcionando o atendimento dos clientes de forma mais especializada, melhorando a qualidade dos serviços. Os usuários dos *Call Centers* não são mais atendidos por qualquer agente, pois as chamadas são direcionadas para essas células de atendimento especializadas (BRAZ, 2010).

Com essa nova forma de organizar o trabalho, os gestores do *Call Center* passaram a ter melhor controle do processo por meio da visualização direta do comportamento e do desempenho dos integrantes da equipe. Desse modo, é possível gerenciar e garantir a qualidade dos serviços, atingir as metas preestabelecidas por equipe e solucionar os problemas dos clientes de forma mais dinâmica (BRAZ, 2010). Como vantagens dessa nova organização da linha de produção, inclui-se o comprometimento, envolvimento e participação dos agentes, melhorando o ambiente de trabalho, trazendo autonomia na tomada de decisão e maior flexibilidade nas relações entre o cliente e o agente (BRAZ, 2010). A distribuição dos atendentes no caso do *Call Center* da Empresa e no caso do *Call Center* do Banco estudados neste trabalho é em linha, e raramente os atendentes consultam seus colegas. No *Call Center* da Empresa, a consulta de uma dúvida de atendimento é feita diretamente ao gerente, mas há planos para transformar este *Call Center* em células de atendimento.

O funcionamento de um *Call Center* pode ser sumariamente descrito do seguinte modo: as chamadas que chegam são primeiramente atendidas por uma unidade de resposta interativa por voz (IVR - *Interactive Voice Response*), que são “máquinas de respostas automáticas”, capazes de interagir com o cliente procurando solucionar o seu problema. Se o problema for solucionado, o cliente deixa o sistema. Caso contrário, a chamada é direcionada para um agente qualificado e que esteja disponível para atendê-lo. Não havendo agentes qualificados e disponíveis, a chamada é inserida em uma fila virtual (invisível) de espera. Durante a espera, o cliente pode abandonar o sistema devido a sua impaciência, ou permanecer na fila até ser atendido por um dos agentes e deixar o sistema depois do atendimento ter-se completado. Em algumas raras situações, o cliente pode, ao efetuar a chamada, encontrar o sistema ocupado e ser desconectado sem completar o seu serviço. As perdas de chamadas também podem ocorrer se, no instante da sua chamada, o cliente souber antecipadamente a quantidade de tempo que

demorará para ser atendido, causando o desencorajamento e recusar-se a entrar no sistema. Em todos esses casos de perda de chamada, existe a possibilidade de o cliente retornar ao sistema por meio de uma nova ligação em um tempo posterior. O esquema operacional de um *Call Center* básico pode ser visualizado na Figura 2.1.

FIGURA 2.1 - Diagrama esquemático das operações de um Call Center



Fonte: elaborado pelo autor

Conforme mencionado, os *Call Centers* modernos recebem chamadas de múltiplos tipos vindos de diferentes canais de comunicação, como o telefone, *e-mail*, internet, *messenger*, *chat*, fax, telefones celulares, entre outros, ocasionando mudança na sua denominação para Centros de Contatos. Essa tendência para os Centros de Contatos é justificada pela nova maneira de comunicação trazida pela internet, com os clientes demandando uma variedade de canais para comunicar-se e, também, para se obter ganhos em qualidade e eficiência (KOOLE; MANDELBAUM, 2002).

### 2.3 Tecnologia

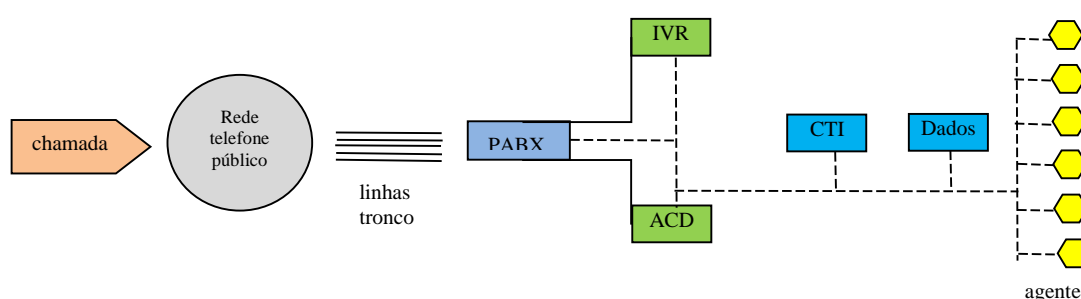
Os avanços tecnológicos na área das Tecnologias da Informação e Comunicação (TIC) têm contribuído para acelerar a evolução dos *Call Centers*. Ter uma tecnologia adequada ajuda na realocação de agentes para atendimento dos clientes em momentos de pico, bem como a gestão dos tempos ociosos, a diminuição do tempo de atendimento, a gestão da fila, entre outros recursos. Por isso, os *Call Centers* tem buscado o suporte de uma boa solução tecnológica possível de promover ganhos significativos à operação (LIMA, 2012).

Tudo começou, segundo Koole e Mandelbaum (2002), com os PABX's (*Private Automatic Branch Exchanges*), que conectaram a rede de telefones públicos com os telefones



dentro do *Call Center*, por meio de uma quantidade de linhas de telefones denominadas de linhas tronco. Entre o PABX e os agentes existia o IVR, que interagiu com os clientes consultando sobre suas necessidades e o ACD (*Automatic Call Distributor*), que fazia o roteamento das chamadas conectando o cliente com o agente, além de arquivar os dados operacionais. Outra importante tecnologia era a CTI (*Computer Telephony Integration*), que acessava automaticamente o arquivo do cliente, exibindo o seu histórico, detalhando informações importantes, bem como apontando prováveis respostas para o seu problema (Figura 2.2).

FIGURA 2.2 - Tecnologia de um Call Center



Fonte: Koole; Mandelbaum, 2003

Os *Call Centers* e os Centros de Contato estão cada vez mais aderindo à tecnologia da computação em nuvem (*cloud computing*), dando origem ao *Call Center* virtual, também conhecido como *Cloud Call Center* ou *Hosted Call Center*. Esse novo sistema permite uma redução dos custos, pois, não exige uma equipe de tecnologia da informação própria e altamente especializada, além de baixo investimento inicial. Como adicional, o modelo permite operar com estações de trabalhos domésticas (*home office*), recomendadas para as pequenas centrais de atendimento e para aquelas que necessitam de recursos adicionais de atendimento (LIMA, 2012).

Com o propósito de eliminar a IVR e as filas de espera, surge um novo aplicativo que tem a finalidade de atuar como intermediário entre o cliente e o *Call Center*. Usando qualquer dispositivo móvel, a chamada do cliente é registrada no *Call Center* e depois é direcionada para um operador, que oportunamente faz a ligação para o cliente, concluindo a operação (CALL CENTER Info, 2013). Neste estudo a configuração da tecnologia utilizada no caso do *Call Center* da Empresa tem como base um PABX-IP e um link E1, que recebe as chamadas com 400 contatos e infinitos canais. Usam um IVR com um *software* Elasticos (linux). No ACD

possuem um outro *software* de gerenciamento das distribuições das chamadas, onde tem a base de dados dos clientes. Este *software* é próprio (desenvolvido pela empresa) e recebe o nome de CRM. O caso do *Call Center* do Banco utiliza 6 IVRs, cada uma com 16 linhas.

## 2.4 Tipos de *Call Centers*

Os *Call Centers* são classificados, de acordo com Koole e Mandelbaum (2002), em diversas dimensões: *funcionalidade*, agrupando as várias funções exercidas como de *help desk*, serviços de emergência, telemarketing, provedores de informação, entre outras; *tamanho*, que considera a quantidade de agentes que pode variar de alguns poucos até milhares deles; *geografia*, variando de um único local até múltiplos locais espalhados em diversas regiões ou países; *características dos agentes*, que podem ter um baixo ou alto treinamento e também uma variação no grau de habilidades para atender um ou mais tipos de chamadas, como, por exemplo, atender chamadas em diversas línguas por telefone, *e-mail* ou *chat*.

De acordo com o tráfego de chamadas, os *Call Centers* são categorizados em receptivos ou *inbound*, e ativos ou *outbound*. No primeiro caso, os agentes atendem as chamadas de clientes solicitando serviços como o de solicitação de informações, sugestões, suporte técnico, reclamações, entre outros. No segundo caso, são os agentes que ligam para os clientes para vender produtos, efetuar cobranças, fazer propagandas e pesquisas. Os *Call Centers* que possuem os dois tipos de tráfegos, ativo e receptivo, são considerados mistos ou *blended*. Uma atividade importante dos *Call Centers* mistos é a utilização do tempo ocioso dos agentes para chamar os clientes que deixaram pedidos no IVR, para serem contatados, ou que abandonaram para checar seus desejos (KOOLE; MANDELBAUM, 2002). Os dois casos de *Call Centers* objetos deste estudo são do tipo *inbound*, que atendem as chamadas dos clientes que necessitam de suporte técnico dos produtos adquiridos, ou que desejam comprar esses produtos, ou ainda que desejam resolver os seus problemas relacionados com sua movimentação bancária.

Os *Call Centers* modernos possuem uma arquitetura organizacional, segundo os mesmos autores, que varia desde pequenos ambientes contendo essencialmente os agentes recebendo as chamadas externas, até aqueles formados por várias camadas, cada uma delas especializadas em determinados tipos de serviços, em que os clientes são transferidos de uma para outra até serem completamente servidos. Existem ainda arquiteturas mais complexas formadas por vários *Call Centers* conectados e dispersos geograficamente, com tamanhos

variados, desde aqueles formados por um único agente trabalhando em sua própria casa, até os de grande porte distribuídos em alguns continentes.

Mancini (2001), observando as funcionalidades, estabeleceu vários critérios para classificar os *Call Centers* e os Centros de Contato, tais como: setor econômico atendido, serviços oferecidos, constituição da empresa, sistema adotado e origem do contato. Em relação ao setor econômico atendido, considera que o *Call Center* é útil tanto para o setor público quanto para as empresas privadas. No setor público, estabelece o contato entre a sociedade e os órgãos públicos e no setor privado é utilizado para criar diferenciais competitivos e solidificar a imagem da empresa. Quanto ao critério serviços oferecidos, destaca-se alguns dos serviços mais utilizados, como o SAC (serviço de atendimento ao consumidor), que funciona como um canal de fidelização de clientes. Também o serviço 0800 que disponibiliza um número para ligações gratuitas, serviços de pesquisas à distância, serviços de propaganda, entre outros.

O critério constituição da empresa auxilia a classificar os *Call Centers* conforme o seu tamanho em pequeno, médio e grande porte, considerando não apenas o seu faturamento, mas também a quantidade de agentes existentes. O critério sistema adotado separa os centros de contato em sistemas internos, formados pelos departamentos da própria empresa, e em sistemas terceirizados, quando se contrata uma empresa especializada para realizar o atendimento ao cliente e, ainda, os sistemas mistos. O critério origem do contato categoriza os *Call Centers* em receptivos (*inbound*), quando as chamadas são originadas pelos clientes, e em ativos (*outbound*), quando as chamadas partem da empresa.

Nesse ambiente de *Call Center*, existe uma equipe de profissionais exercendo múltiplas funções. As funções variam para cada empresa e devem ser determinadas considerando a melhor relação custo-benefício. Entre essas funções destacam-se as de atendente, psicólogo, roteirista, técnico de banco de dados, analista de sistema e tráfego, *controller*, assessor de logística, gerente de projetos e facilitador (MANCINI, 2001).

## 2.5 Níveis de serviço

A maior preocupação em um *Call Center* é oferecer serviços de qualidade respeitando um orçamento especificado. Koole e Mandelbaum (2002) consideram duas dimensões para medir a qualidade dos serviços: qualitativa, que se preocupa com os aspectos psicológicos do sistema e como o serviço é percebido pelo cliente, e a quantitativa, que se preocupa com a acessibilidade do serviço, como o tempo de espera para ser atendido. Qualquer nível de serviço

estipulado requer um custo associado e a meta é minimizar o custo total. Desse modo, a gestão de um *Call Center* está sempre preocupada com o *trade-off* entre o nível de serviço e os custos.

Nesse aspecto, três regimes operacionais são considerados, cada um dos quais relacionados a diferentes economias de escala. Halfin e Whitt (1981) foram os pioneiros na formalização dos regimes operacionais. O primeiro caso é o QD (Quality-Driven), cujo foco está na qualidade dos serviços e oferece um nível de serviço de alta qualidade contra um custo da baixa eficiência (utilização) dos agentes. O nível de pessoal (quantidade de agentes) aumenta linearmente com a carga oferecida, não havendo economia de escala. Nos *Call Centers* de grande porte que operam neste regime, a maioria das chamadas é servida sem espera. No outro extremo, o regime ED (*Efficiency-Driven*), o foco está na eficiência dos agentes ou na alta utilização dos agentes com uma significativa fração de abandono. Entre os dois extremos encontra-se o regime QED (*Quality-and-Efficiency Driven*), em que o foco está em um cuidadoso balanço entre qualidade e eficiência, combinando um alto nível de eficiência (utilização dos agentes) e alta qualidade de serviço, com pequenas esperas na fila e escassos abandonos. Um novo regime é proposto por Mandelbaum e Zeltyn (2009) combinando os regimes ED e QED, formando o regime operacional ED+QED. O trabalho de Halfin e Whitt (1981) foi estendido por Garnett et al. (2002), posteriormente refinado por Zhang et al. (2012) e generalizado por Zeltyn e Mandelbaum (2005).

Procurando diminuir custos, os *Call Centers* estão utilizando dos recursos tecnológicos disponíveis para realizar serviços que não necessitam de atendimento humano (sem agentes). Com o torpedo SMS é possível disparar mensagens de texto para os clientes com custos reduzidos de telefonia. Na mesma dinâmica, pode-se enviar mensagens de voz ou vocalizada por meio de *software*, com o torpedo de voz, com a possibilidade dessas mensagens serem enviadas simultaneamente, ou seja, disparadas todas ao mesmo tempo. Com mais interatividade pode-se utilizar a URA (Unidade de Resposta Audível) ativa, que permite maior interação automática com o cliente por meio de um sistema de vocalização que estabelece um diálogo com o usuário

Três noções de qualidade de serviços são apresentadas por Gans et al. (2003) e que são frequentemente usadas pelos gestores dos *Call Centers*. A primeira noção está relacionada com a acessibilidade dos agentes, preocupando-se com o tempo de espera na fila e com a quantidade de abandonos. A segunda relaciona-se com a efetividade do serviço, preocupando-se em saber se o problema do cliente foi totalmente resolvido, evitando o retrabalho. A terceira noção de qualidade de serviço está relacionada com a interação dos agentes com os clientes, se preocupando com o tratamento dado ao cliente, inclusive o tom de voz utilizado. Neste estudo,

é adotada a primeira noção de qualidade dos serviços porque acredita-se que diminuindo-se a quantidade de abandonos, melhora-se o nível de qualidade dos serviços recebidos. Aqueles que abandonam de certa forma informam aos gestores do sistema que não vale a pena esperar pelos serviços que irão receber.

## 2.6 Dimensionamento

O dimensionamento é uma atividade importante em um *Call Center* e define como serão as suas operações a médio e longo prazo, imprescindíveis para uma visão estratégica e controle de custos. O dimensionamento de um *Call Center* consiste em alocar recursos de maneira correta, com menor custo, para atingir metas estabelecidas pelos gestores. Do ponto de vista administrativo, a quantidade de postos de atendimentos (agentes) é determinada considerando-se alguns indicadores, que são ferramentas de controle que permitem comparar os resultados da operação com suas metas. Nos *Call Centers inbound* (receptivos), são importantes os indicadores: a previsão de chamadas a cada intervalo, geralmente, de 30 minutos; o nível de serviço contratado que fornece a quantidade de chamadas (em porcentagem) por um intervalo de tempo (em segundos) e os tempos de pausa que indicam os tempos de afastamento do agente quando é recrutado pela própria empresa para atender a atividades de treinamento e reuniões. Os *Call Centers* do tipo *outbound* (ativos) utilizam como indicadores: o volume de *mailing* que representa a quantidade de contratos previstos; o contato com a pessoa certa (CPC) que considera o percentual, dentre as chamadas atendidas, no qual se conseguiu falar com a pessoa que era o objetivo da ligação, e o discador que pode ser manual, progressivo ou preditivo; o absentéismo que mede o percentual de agentes ausentes dentre todos os que foram escalados para trabalhar (PIERRE, 2012).

Sob o ponto de vista quantitativo, o dimensionamento de um *Call Center* se baseia na conhecida fórmula de Erlang do modelo clássico de fila  $M/M/c$ , discutido no Capítulo 3. Considera-se como indicador importante o tempo de espera (*wait time*), que é o tempo que o cliente espera para ter sua chamada atendida. É um indicador imprescindível para medir o desempenho do sistema porque quanto menor for o tempo de espera, mais eficiente é a operação e, ainda, é por meio do tempo de espera que o cliente percebe a qualidade dos serviços prestados. Também o tempo pós-atendimento (*wrap time*), considerado o tempo que o agente gasta entre uma chamada atendida e o início da próxima. Este indicador avalia o tempo de ocupação do agente ou sua produtividade, sendo que quanto maior for o *wrap time*, menor será

a sua produtividade. Além disso, o tempo de atendimento (*talk time*), que é o tempo que o agente gasta na conversação com o cliente para atender as suas necessidades de serviço e também a taxa de abandono, que indica o percentual de clientes que abandonam o sistema devido a longa espera na fila. A definição dos indicadores e os cálculos matemáticos e estatísticos necessários são realizados pela equipe gestora, formada por pessoal capacitado, que também deve considerar o histórico das chamadas ocorridas no *Call Center* (PIERRI, 2012). Os *Call Centers*, casos de estudo nesta tese, tem os seus dimensionamentos analisados principalmente sob o ponto de vista quantitativo, ajustando os tamanhos das equipes de atendentes para melhorar os níveis de serviços oferecidos aos seus clientes.

## 2.7 Abandono

A impaciência do cliente, manifestada por meio do abandono, está presente em vários sistemas de serviços, tais como nos atendimentos telefônicos enquanto se aguarda por uma resposta de um atendente, nos serviços de compras prestados pela internet, nos atendimentos médicos de emergência quando se espera pelo médico ou, ainda, nos atendimentos bancários enquanto se aguarda para ser atendido. O principal objetivo do gestor de um *Call Center* é obter um bom nível de serviço que é atingido com chamadas bem-sucedidas, atendendo o cliente e a empresa. O dimensionamento inadequado pode atrapalhar esse objetivo, fazendo o cliente abandonar o sistema depois de um longo tempo de espera na fila. Desse modo, o abandono torna-se uma medida de desempenho operacional importante. É por meio dele que os clientes manifestam as suas impaciências, informando a percepção sobre o serviço que estão recebendo.

Existem alguns fatores que afetam a impaciência e que são classificados por Mandelbaum e Zeltyn (2012) nas seguintes categorias: a categoria de clientes (cultura, necessidades, sexo, idade etc.); a categoria da chamada (tipo de serviço, hora do dia etc.); experiência global (experiência acumulada, qualidade do serviço prestado no passado etc.) e a experiência local (anúncio automático da espera na fila, histórico de recentes retornos etc.). Outros fatores que afetam a impaciência dos usuários em filas de espera foram devidamente comprovadas por Diekmann et al. (1996), que mostraram que o *status* social do usuário na fila está positivamente correlacionado com a tendência em direção a um comportamento agressivo (impaciência). Thierry (1994) conduziu um estudo relacionando a importância subjetiva da finalidade da fila (serviço) com as reações dos usuários desta fila, e concluiu que o tempo gasto na fila é mais notável (o usuário fica impaciente) quando a finalidade da fila é menos atrativa.

Estes fatores não foram considerados neste estudo pela falta de informações adicionais nas bases de dados dos *Call Centers* aqui analisados. Dados como sexo, idade e outros não foram disponibilizados, por serem informações confidenciais.

A impaciência pode ser capturada por meio de cinco “tempos de espera” divididos em dois grupos: antecipados e percebidos, que estão relacionados com os aspectos psicológicos da espera (MANDELBAUM; ZELTYN, 2012). São eles:

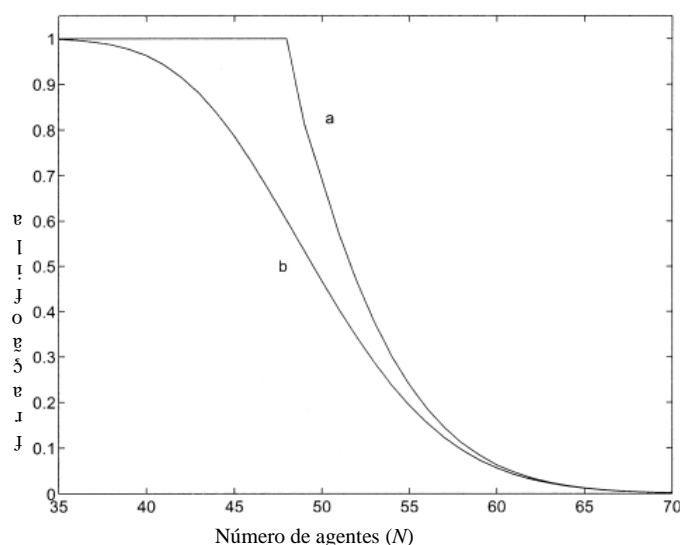
- *Tempo antecipado*: formado pelas experiências acumuladas individualmente ou por meio de uma concepção geral extraída dos serviços de *Call Center*. É o tempo que o cliente antecipa da sua espera na fila;
- *Tempo de paciência ou impaciência*: tempo que o cliente está disposto a aguardar na fila de espera. Caso o tempo de paciência termine e o cliente ainda esteja na fila, então ele abandona o sistema. Geralmente, o tempo de paciência é considerado a partir do instante da chegada do cliente na fila do *Call Center*, podendo ser reavaliado quando surgem eventos externos durante a sua espera, como o anúncio automático da sua posição na fila;
- *Tempo oferecido*: tempo que o cliente precisa aguardar para ser servido (cliente com paciência infinita). Esse tempo depende de diversos fatores, como a quantidade de agentes, prioridade dos clientes;
- *Tempo de espera*: tempo que o cliente atualmente espera, definido como o menor entre o tempo de paciência e o tempo oferecido;
- *Tempo percebido*: tempo com o qual o cliente percebe a espera. Está relacionado com a experiência do serviço do *Call Center*. Esse tempo afeta o tempo antecipado em uma visita posterior.

Ignorar o abandono (impaciência) gera distorções ou falhas nas informações que são importantes para os gestores do *Call Center*. Pode causar um sub ou super dimensionamento de agentes, afetando o nível de serviço das operações. Em muitos *Call Centers*, é fixado um índice de abandono que varia de 2% a 4%, como ocorre nas concessionárias de distribuição de energia elétrica, que consideram para esse índice valores menores ou iguais a 4% (ANAEEL, 2009). Em geral esse índice é obtido dividindo-se o total de chamadas abandonadas pelo total das chamadas recebidas (compostas pelas chamadas abandonadas e chamadas atendidas) dentro de um intervalo de tempo estipulado. Em muitos casos, esse índice não é atingido ou completamente ignorado, justificado pela falta de compreensão do fenômeno do abandono, ou pela escassez de modelos de análise que o incorporam com precisão. O efeito imediato de um

abandono é a diminuição do tempo de espera daqueles que estão em posições atrás do abandono, ou para as futuras chegadas, e pode induzir um sub dimensionamento da quantidade de agentes. Nas situações em que o abandono é ignorado, pode ocorrer um superdimensionamento de agentes, quando uma quantidade menor seria necessária.

Garnett, Mandelbaum e Reiman (2002) mostram a importância do abandono por meio de exemplos numéricos reproduzidos neste texto pelo Gráfico 2.1 e Tabela 2.1.

GRÁFICO 2.1: Fração de clientes na fila –  $M(48)/M(1)/N$  (a) vs.  $M(48)/M(1)/N+M(0,5)$  (b)



Fonte: GARNETT, MANDELBAUM E REIMAN, 2002

Nota: 48 chamadas por minuto, 1 minuto de tempo médio de serviço,  
2 minutos de tempo médio de paciência (0,5 abandono por minuto)

Vale observar que a fração de clientes na fila é bem diferente nos dois modelos do Gráfico 2.1 quando a quantidade de agentes  $N$  situa-se entre 40 e 50. Para grandes valores de  $N$ , o sistema fica subcarregado e a fração de clientes na fila coincide nos dois modelos porque os abandonos são desprezíveis. O desempenho do sistema em tráfego intenso (pesado) é bastante sensível em relação ao tamanho da equipe de agentes, porque ao adicionar-se poucos agentes, o modelo sem abandono passa a ter desempenho similar aquele com abandono. Como os custos com pessoal representam cerca de setenta por cento dos custos totais, essa pequena redução do número de agentes nos modelos com abandono passa a ser significativa na economia do sistema (GARNETT, MANDELBAUM e REIMAN, 2002). Essa análise evidencia a vantagem de se adicionar o abandono nos modelos de fila, deixando-os mais representativos e robustos e oferecendo informações importantes para os gestores de *Call Center*. Desse modo,



na administração de grandes *Call Centers* em tráfego intenso (pesado), deve-se considerar o efeito do abandono dos clientes no nível de serviço.

Ao adicionar-se o abandono, diminui-se a congestão do sistema porque nem toda chegada necessita de serviço e, como consequência, reduz-se o tamanho da fila e o tempo de espera. Esses resultados podem ser observados na Tabela 2.1 pelas diferenças significativas nos valores do tempo de espera e tamanho da fila, que são bem menores quando o abandono é considerado no modelo. Esses dados também indicam que é possível obter, ao mesmo tempo, uma alta eficiência (utilização dos agentes próximo de 100%) e um bom serviço (pequenos tempos de espera e abandono).

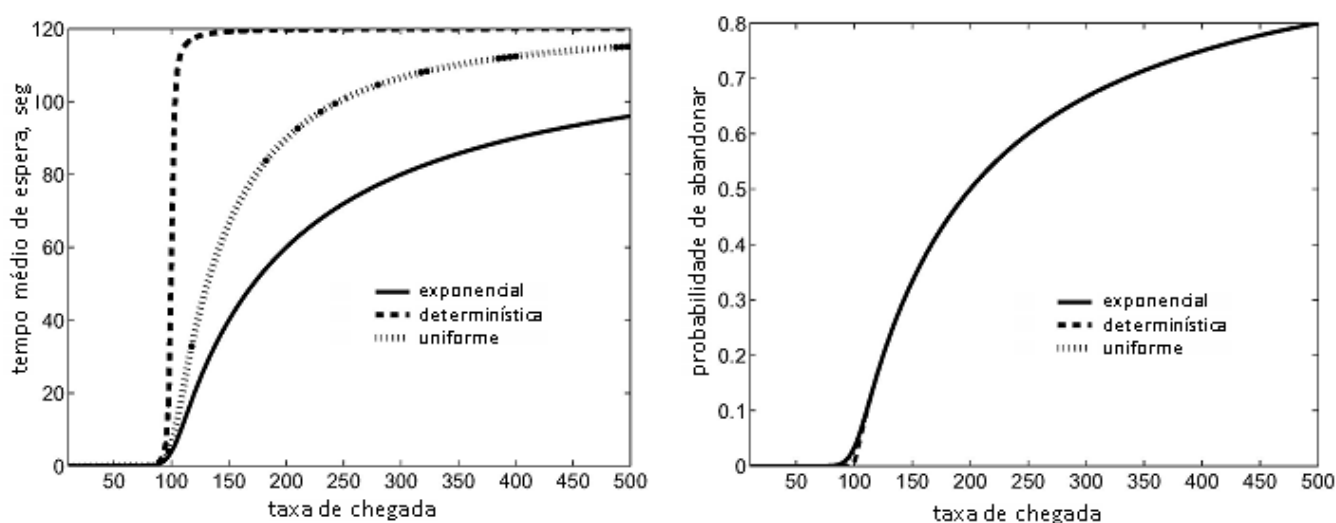
TABELA 2.1: Comparando resultados para modelos com e sem abandono

Medidas	M/M/N	M/M/N+M
Fração do abandono	-	3,1 %
Tempo médio de resposta	20,8 seg.	3,6 seg.
Percentil-90 do tempo de espera	58,1 seg.	12,5 seg.
Tamanho médio da fila	17	3
Utilização dos agentes	96 %	93 %

Fonte: GARNETT, MANDELBAUM E REIMAN, 2002

Nota: 50 agentes, 48 chamadas por minuto, tempo médio de serviço de 1 min, tempo médio de paciência de 2 min.

FIGURA 2.3: Dependência do desempenho sobre a distribuição da paciência



Fonte: ZELTYN E MANDELBAUM, 2005

Outro aspecto da importância do abandono é observado em Zeltyn e Mandelbaum (2005), que compararam três distribuições de probabilidade para a paciência com o mesmo

tempo médio de paciência: uma exponencial com média 2 minutos, uma constante com média 2 e uma uniforme no intervalo  $[0, 4]$ . O experimento foi realizado com o modelo de fila  $M/M/c+G$  (discutido no Capítulo 3), com 100 servidores (agentes) e taxa de serviço unitária. Esses autores escolheram duas medidas de desempenho, o tempo médio de espera e a probabilidade de abandono, para realizar as comparações. Eles variaram a taxa de chegada de 10 a 500, com intervalos de 2,5, e os resultados estão exibidos na Figura 2.3.

Observando os dois gráficos da Figura 2.3, deduz-se que as três curvas para o tempo médio são muito diferentes, exceto para pequenos valores da taxa de chegada quando a espera é desprezível. Por outro lado, as curvas para a probabilidade de abandono são quase idênticas, havendo uma pequena diferença entre a paciência determinística e as outras duas distribuições nas proximidades da taxa de chegada igual a 100.

Com essas observações, esses autores concluem que:

- A distribuição da paciência pode afetar significativamente o desempenho da fila  $M/M/c+G$ ;
- O efeito da distribuição paciência depende fortemente da medida de desempenho considerada;
- O efeito da distribuição da paciência depende da carga que o sistema está operando.

Desse modo, o abandono das tele filas tem um impacto grande nas operações de *Call Center*. Existem algumas pesquisas teóricas sobre modelos de filas que incorporam o abandono e que podem ser potencialmente aplicados para análise dessas operações. Entretanto, há um hiato entre a oportunidade de aplicação desses modelos teóricos e as escassas pesquisas sobre os padrões da paciência dos clientes nas tele filas. Essa tese pretende contribuir para reduzir esse hiato.

### 3 MODELAGEM DO *CALL CENTER* COMO SISTEMA DE FILA

Alguns modelos analíticos de filas são revisados neste capítulo, devido às suas importâncias na modelagem dos sistemas de Call Centers. São revistos os modelos de fila com e sem o abandono dos usuários, bem como aqueles que consideram a distribuição exponencial e não exponencial para os tempos do abandono e serviço. Modelos de filas são amplamente utilizados para analisar e dimensionar Call Centers. A literatura é vasta e os pesquisadores têm encontrado dificuldades em modelar alguns aspectos do processo de abandono e de serviço.

Um modelo de fila amplamente utilizado em *Call Centers* é o conhecido sistema M/M/c, também denominado de modelo Erlang-C. Neste modelo, dada uma taxa de chegada  $\lambda$ , uma taxa de serviço  $\mu$  e  $c$  servidores trabalhando em paralelo, a expressão Erlang-C,  $C(\lambda, \mu, n)$  descreve a fração do tempo que todos os  $c$  servidores estão simultaneamente ocupados, ou a fração dos usuários que aguardam em fila antes de serem servidos (WOLFF, 1989). De acordo com Brown (2002), o modelo Erlang-C é bastante restritivo, assumindo que o processo de chegada das chamadas telefônicas obedece ao processo de Poisson, que as durações dos tempos de serviços sejam exponencialmente distribuídas e que usuários e servidores são estatisticamente idênticos e independentes um do outro. Além disso, não reconhece o comportamento do abandono dos usuários, a dependência das taxas de chegada e de serviço em relação ao tempo e a heterogeneidade dos clientes (GARNETT & MANDELBAUN, 1999).

Com a preocupação em desenvolver modelos que consideram esses efeitos, Garnett et al. (2002) propôs um modelo de abandono, no qual a paciência (ou o tempo até abandonar) dos usuários é exponencialmente distribuída e a capacidade de espera do sistema é ilimitada (M/M/c+M). Com o comportamento assintótico, ele deduziu aproximações para as medidas de desempenho e propôs regras para projetar grandes *Call Centers*. Zohar et al. (2002) propõem um modelo baseado no modelo M/M/c que incorpora o comportamento adaptativo dos clientes, partindo da premissa que os usuários adaptam sua paciência (modelada pela distribuição do tempo até abandonar) em relação a sua expectativa de serviço, em particular para o seu tempo de espera antecipado.

Motivados pela prática dos *Call Centers*, Mandelbaum e Zeltyn (2009) modelaram um *Call Center* com uma fila M/M/c+G, que é caracterizada por chegadas Poisson, tempos de serviços exponenciais,  $c$  servidores e o tempo de paciência dos usuários com distribuição genérica. Determinaram uma equipe de atendentes mínima  $n$  que satisfaz a um dado custo que

pode incorporar a fração do abandono, o tempo médio de espera e as probabilidades de aguardar. São apresentados vários regimes operacionais de custos e os autores demonstram, com uso de experimentos numéricos, que o regime QED (*Quality and Efficiency-Driven*) resulta em uma boa aproximação para as medidas de desempenho do modelo M/M/c+G.

Os tempos até abandonar são dados censurados e podem ser estimados por meio de Análise de Sobrevivência (COX e OAKES, 1990). A taxa de risco dos clientes pacientes pode ser estimada usando um modelo de riscos competitivos, conforme Palm (1953), que postulou que a taxa de risco do tempo disposto para aguardar é proporcional à irritação do cliente devido à espera. Neste sentido, Aalen e Gjessing (2001) comentam a interpretação dinâmica da taxa de risco, mas advertem com a possibilidade da taxa de risco da população necessariamente não representar a dos indivíduos. Examina-se a validade de uma regra da teoria de filas que relaciona o tempo médio de espera na fila com a probabilidade do abandono. Esta “regra” está deduzida em Baccelli and Hebuterne (1981) e em Zohar et al. (2002). Ela se aplica em modelos com paciência distribuída exponencialmente. Essa regra se mantém para os dados analisados cuja distribuição da paciência é não exponencial. Pesquisa-se uma explicação teórica dessas observações empíricas.

É também possível verificar a aplicabilidade da generalização da fórmula clássica de Pollaczek-Khintchine da teoria de filas para sistemas M/G/1, conforme Whitt (2002) e Fendick, Saksena e Whitt (1989). Essa fórmula não resulta em um modelo razoável para *Call Center*, pois ignora o comportamento de abandono dos clientes. Modelos que explicitamente consideraram o abandono obtiveram visível melhora no desempenho (BACCELLI e HEBUTERNE, 1981; GARNETT, MANDELBAUM e REIMAN, 2002).

Analisando os estudos relatados, percebe-se um interesse na literatura pelo abandono dos clientes das filas de espera dos *Call Centers*. Sempre que considerado o abandono, melhorou-se a avaliação de desempenho do *Call Center*. Entretanto, o comportamento do abandono foi modelado preferencialmente pela distribuição exponencial, ou ainda por uma distribuição geral, determinística ou Weibull. Por esse motivo, propõe-se nesta tese que outra distribuição de probabilidade também seja utilizada na modelagem dos tempos de paciência nas filas de espera dos *Call Centers*, formada por uma mistura de distribuições (Apêndice H), capaz de captar melhor possíveis características distintas na população de clientes, como os diferentes tipos de chamadas. O processo do abandono, descrito por meio de distribuições mistas, foi estudado na literatura, mostrando-se vantajoso em relação aqueles que não foram descritos

pelas distribuições mistas. Oliveira (2009), usando simulação, mostra que ocorre uma diminuição do tempo médio de espera quando a paciência é modelada por distribuições mistas. Mandelbaum e Zeltyn (2012) utilizam várias distribuições mistas para modelar o processo do abandono e obtém melhor ajuste para o tempo de paciência, do que as distribuições não mistas. Porém, não se tem conhecimento de trabalhos na literatura incorporando essas distribuições mistas em modelos analíticos de filas com abandono, e avaliando esses modelos com os dados reais extraídos de *Call Centers*.

A seguir apresenta-se um resumo dos resultados importantes de sistemas de filas, contendo as características e as medidas de desempenho dos modelos M/M/• e M/G/• que incorporam o tempo de paciência (abandono) nas suas equações, e que são abordados nesta tese.

### 3.1 O Sistema M/M/c

O clássico sistema de fila M/M/c é aqui brevemente revisado porque constitui a base para a elaboração dos correspondentes sistemas com abandono M/M/c+M e M/M/c+G. Este sistema é formado por  $c$  canais de serviço estatisticamente idênticos, operando em paralelo e independentemente um dos outros, atendendo a uma única fila de espera. O processo de chegada ocorre de acordo com o processo de Poisson com tempos entre chegadas de usuários distribuídos exponencialmente com taxa de chegada constante e igual a  $\lambda_n = \lambda$ . Os tempos de serviços dos usuários são distribuídos exponencialmente com taxa de serviços

$$\mu_n = \begin{cases} n\mu, & 1 \leq n < c \\ c\mu, & n \geq c \end{cases}$$

Os clientes (ou chamadas, no caso dos *Call Centers*), que chegam e encontram servidores (agentes, para os *Call Centers*) livres, ocupam esses servidores e recebem os serviços. Se ao chegarem encontrarem todos os servidores ocupados, aguardam em uma fila de espera para serem atendidos, em geral na ordem de chegada (disciplina FIFO – *First In First Out*). Nas centrais de atendimento, essa fila é invisível porque os clientes não podem enxergar uns aos outros, mesmo que a fila exista fisicamente. Substituindo-se as taxas de chegada e de serviços nas equações do Processo de Nascimento e Morte (GROSS E HARRIS, 1985), dadas por:

$$p_n = \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i} p_0 \quad (3.1)$$

$$p_0 = \left[ 1 + \sum_{n=1}^{\infty} \prod_{i=1}^n \frac{\lambda_{i-1}}{\mu_i} \right]^{-1} \quad (3.2)$$

obtém-se a equação que fornece a probabilidade dos estados do sistema ( $p_n$ ) e a probabilidade do sistema estar vazio ( $p_0$ ), do sistema de fila M/M/C:

$$p_n = \begin{cases} \frac{\lambda^n}{n! \mu^n} p_0, & 1 \leq n \leq c \\ \frac{\lambda^n}{c^{n-c} c! \mu^n} p_0, & n \geq c \end{cases} \quad (3.3)$$

$$p_0 = \left[ \sum_{n=0}^{c-1} \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n + \frac{1}{c!} \left( \frac{\lambda}{\mu} \right)^c \left( \frac{c\mu}{c\mu - \lambda} \right) \right]^{-1} \quad (3.4)$$

A intensidade de tráfego para este sistema é definida como  $\rho = \frac{\lambda}{c\mu}$  e, como não existe nenhuma restrição imposta para a capacidade do sistema, o equilíbrio estatístico é atingido sempre que  $\rho < 1$ . Usando as probabilidades dos estados com o sistema em equilíbrio e a fórmula de Little (GROSS E HARRIS, 1985), algumas medidas de desempenho como o número médio de usuários na fila ( $L_q$ ), o tempo médio de espera na fila ( $W_q$ ), o tempo médio de espera no sistema ( $W$ ) e o número médio de usuários no sistema ( $L$ ), podem ser expressas por:

$$L_q = \left[ \frac{\left( \frac{\lambda}{\mu} \right)^c \lambda \mu}{(c-1)(c\mu - \lambda)^2} \right] p_0 \quad (3.5)$$

$$W_q = \frac{L_q}{\lambda} \quad (3.6)$$

$$W = W_q + \frac{1}{\mu} \quad (3.7)$$

$$L = \lambda W \quad (3.8)$$

Uma medida útil desse sistema é a probabilidade de espera (ou de formar fila),  $p_d$ , conhecida como fórmula Erlang-C, que fornece a probabilidade de nenhum servidor estar

disponível no instante da chegada de um cliente no sistema. Ela é denotada por  $C(c, \lambda/\mu)$  ou por  $E_{2,c}(\lambda/\mu)$ , expressa por (GROSS E HARRIS, 1985):

$$p_d = \sum_{n=c}^{\infty} \frac{p_0 \rho^n}{c! c^{n-c}} \quad (3.9)$$

Devido a essa equação, o sistema de fila M/M/c é também conhecido como sistema Erlang-C. Vale observar que esse sistema não considera os casos em que os clientes abandonam a fila depois de terem aguardado um certo tempo. Após ingressarem na fila, permanecem nela até serem servidos. Por isso, são considerados com paciência infinita. Também não incorpora a possibilidade de os clientes pertencerem a uma população heterogênea, formada por classes distintas de clientes com características próprias, como o nível cultural, social e outros, e o tipo de serviços que necessitam. Mais detalhes sobre esse sistema podem ser vistos em Gross e Harris (1985) ou em Arenales et al. (2007).

### 3.2 O Sistema M/M/c/K

Outro sistema básico de fila é o M/M/c/K, cuja equação de bloqueio é utilizada na obtenção das equações dos sistemas com abandono. Este sistema difere do anterior M/M/c pela limitação  $K$  da capacidade do sistema formada pelo número de clientes na fila e em serviço. As chegadas ocorrem segundo um processo de Poisson, com tempos entre chegadas distribuídos exponencialmente com taxa de chegada:

$$\lambda_n = \begin{cases} \lambda, & 0 \leq n < K \\ 0, & n \geq K \end{cases}$$

Os tempos de serviços também possuem uma distribuição exponencial com taxa:

$$\mu_n = \begin{cases} n\mu, & 1 \leq n < c \\ c\mu, & c \leq n \leq K \\ 0, & n > K \end{cases}$$

Os clientes ao chegarem e encontrarem servidores livres, ocupam esses servidores. Se nas suas chegadas todos os servidores estão ocupados, eles aguardam na fila de espera de tamanho máximo  $K-c$  ( $K > c$ ). Caso ao chegarem a fila tenha atingido seu tamanho máximo, eles são impedidos de entrar no sistema e, então, são perdidos. Aqueles que se posicionam na fila

de espera são atendidos pela disciplina FIFO. Como anteriormente, substituindo-se as taxas de chegada e de serviço nas equações do Processo de Nascimento e Morte, obtém-se as probabilidades dos estados do sistema ( $p_n$ ), em que  $\rho = \frac{\lambda}{\mu}$ , conforme Dombacher (2010):

$$p_n = \begin{cases} \frac{\rho^n}{n!} p_0, & 1 \leq n \leq c \\ \frac{\rho^n}{c!c^{n-c}} p_0 & c < n \leq K \\ 0, & n > K \end{cases} \quad (3.10)$$

com a probabilidade do sistema estar vazio ( $p_0$ ) dada por:

$$p_0 = \begin{cases} \left[ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \cdot \frac{1 - \left(\frac{\rho}{c}\right)^{K-c+1}}{1 - \frac{\rho}{c}} \right]^{-1}, & \frac{\rho}{c} \neq 1 \\ \left[ \sum_{n=0}^{c-1} \frac{c^n}{n!} + \frac{c^c}{c!} \cdot (K - c + 1) \right]^{-1}, & \frac{\rho}{c} = 1 \end{cases} \quad (3.11)$$

Essas duas últimas equações mostram que não há mais necessidade que  $\rho < 1$  para que o sistema esteja em equilíbrio estatístico. Uma medida de desempenho importante para esse sistema é a probabilidade de espera na fila (*delay*),  $p_e$ , definida por:

$$p_e = \begin{cases} \frac{\rho^c}{c!} \cdot \frac{1 - \left(\frac{\rho}{c}\right)^{K-c}}{1 - \frac{\rho}{c}} p_0, & \frac{\rho}{c} \neq 1 \\ \frac{c^c}{c!} \cdot (K - c) p_0, & \frac{\rho}{c} = 1 \end{cases} \quad (3.12)$$

Vale observar que ao se considerar  $K = c$ , o sistema fica sem sala de espera e nesse caso, toda chegada é bloqueada e nenhuma espera ocorre. O sistema passa a ser denotado por M/M/c/c e a probabilidade de bloqueio é calculada por:



$$p_b = p_c = p_K = \frac{\frac{\rho^c}{c!}}{\sum_{n=0}^c \frac{\rho^n}{n!}} \quad (3.13)$$

também conhecida como a fórmula de perda de Erlang ou Erlang-B. O modelo é então denominado por modelo Erlang-B. Embora esta fórmula de perda tenha surgido em um modelo com tempos de serviço exponenciais, ela também é válida nos modelos em que o tempo de serviço não é exponencial, como o modelo M/G/c/c (GROSS e HARRIS, 1985). Por esse motivo, toda distribuição de tempo de serviço fica reduzida apenas ao seu tempo médio de serviço. Um recurso que facilita o manuseio da fórmula de perda é torná-la recursiva:

$$p_c = \frac{\rho p_{c-1}}{c + \rho p_{c-1}}, \quad p_0 = 1 \quad (3.14)$$

Fazendo  $\varepsilon_c = \frac{1}{p_c}$  na expressão anterior, obtemos uma fórmula recorrente

$$\varepsilon_c = 1 + \frac{c}{\rho} \varepsilon_{c-1}, \quad \varepsilon_0 = 1 \quad (3.15)$$

para o cálculo da probabilidade de perda. Como no sistema anterior, usando as probabilidades dos estados do sistema e a fórmula de Little, obtém-se algumas medidas de desempenho: (GROSS E HARRIS, 1985)

$$L_q = \begin{cases} \frac{p_0 \rho^{c+1}}{c! c (1-u)^2} (1 - u^{K-c+1} - (1-u)(K-c+1)u^{K-c}), & \frac{\rho}{c} \neq 1 \quad \text{com } u = \frac{\rho}{c} \\ \frac{p_0 c^c}{c!} \cdot \frac{(K-c)(K-c+1)}{2}, & \frac{\rho}{c} = 1 \end{cases} \quad (3.16)$$

$$W_q = \frac{1}{\lambda(1-p_K)} L_q \quad (3.17)$$

$$W = W_q + \frac{1}{\mu} \quad (3.18)$$

$$L = \lambda(1-p_K)W \quad (3.19)$$

Mais informações sobre esse sistema podem ser obtidas em Dombacher (2010), Gross e Harris (1985) e em Arenales et al (2007).

### 3.3 O Sistema M/M/c + M

O modelo com abandono M/M/c + M foi primeiramente idealizado por Palm (1957), que introduziu um modo simples de modelar o abandono. Ele sugeriu melhorar o modelo Erlang-C (M/M/c) do seguinte modo: associou com cada chamada que chega um tempo de paciência distribuído exponencialmente com média  $1/\theta$ . Um cliente que chega encontra um tempo de espera oferecido, que é definido como o tempo que esse cliente teria que esperar, caso sua paciência fosse infinita. Se essa espera oferecida excede o tempo de paciência do cliente, o cliente abandona o sistema. Caso contrário, o cliente aguarda até ser servido. O parâmetro de paciência  $\theta$  é denominado de taxa de abandono individual. O modelo é denominado de Erlang-A (A de abandono) e também pelo fato do modelo situar-se entre os modelos Erlang-C e Erlang-B.

#### 3.3.1 Modelagem do Sistema M/M/c + M

Os clientes chegam ao sistema M/M/c + M por meio de um processo de chegada Poisson com taxa de chegada  $\lambda$  e tempos entre chegadas distribuídos exponencialmente. Possuem um tempo de paciência  $\tau$  assumido ter distribuição exponencial com taxa  $\theta$ , a sua taxa individual de abandono. São servidos por  $c$  canais de serviços estatisticamente idênticos, operando em paralelo e independentemente uns dos outros, atendendo a uma única fila de espera. O tempo de serviço é considerado exponencialmente distribuído com taxa  $\mu$ . O processo de serviço, de chegada e de paciência são considerados mutuamente independentes.

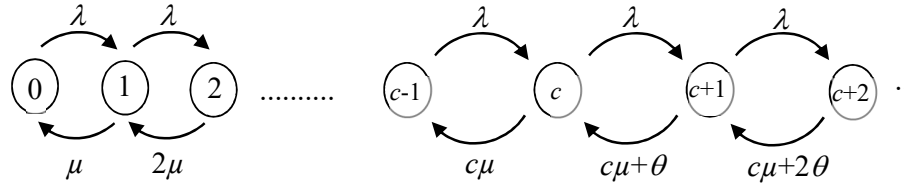
Nesse sistema, para um determinado cliente, o tempo de paciência  $\tau$  é definido como o tempo que um cliente suporta aguardar para ser servido e considera-se que uma espera que atinja  $\tau$  resulte em abandono. Define-se também o tempo de espera oferecido pelo sistema,  $V$ , como sendo o tempo que um cliente com paciência infinita aguarda até ser servido. Desse modo, o tempo de espera de um cliente impaciente, ou o tempo de espera na fila até iniciar atendimento ou abandonar, o que ocorrer primeiro, é igual a  $W' = \min\{V, \tau\}$ , em que  $V$  é o tempo de espera na fila até iniciar atendimento, se não ocorrer abandono antes.

Seguindo as formulações de Mandelbaum e Zeltyn (2005), o número total de clientes no sistema no instante  $t$ , em serviço e na fila, é denotado por  $L(t)$  e o processo  $L = \{L(t), T \geq 0\}$  é markoviano de nascimento e morte, com taxa de chegada  $\lambda_n = \lambda$  e taxa de serviço:

$$\mu_n = \begin{cases} n\mu, & 1 \leq n \leq c \\ [c\mu + (n-c)\theta], & n > c \end{cases}$$

e com o diagrama de transição entre os estados do sistema apresentado na Figura 3.1:

FIGURA 3.1 - diagrama de transição do modelo Erlang-A



Fonte: elaborada pelo autor

Admitindo que o sistema esteja em equilíbrio e considerando a lei de conservação de fluxos, o conjunto de equações pode ser obtido por:

Para  $n \geq c+1$ :

$$\lambda p_{n-1} + (c\mu + 2\theta)p_{n+1} = \lambda p_n + (c\mu + \theta)p_n \quad (3.20)$$

Generalizando:

$$\lambda p_{n-1} + [c\mu + (n-c)\theta]p_{n+1} = \lambda p_n + [c\mu + (n-c)\theta]p_n \quad (3.21)$$

$$\lambda + [c\mu + (n-c)\theta]p_n = [c\mu + (n-c)\theta]p_{n+1} + \lambda p_{n-1} \quad (3.22)$$

Isolando-se  $p_{n+1}$ :

$$p_{n+1} = \frac{\lambda + [c\mu + (n-c)\theta]}{[c\mu + (n-c)\theta]} p_n - \frac{\lambda}{[c\mu + (n-c)\theta]} p_{n-1} \quad (3.23)$$

$$p_{n+1} = \left(1 + \frac{\lambda}{c\mu + (n-c)\theta}\right) p_n - \frac{\lambda}{c\mu + (n-c)\theta} p_{n-1}, \quad \text{para } n > c \quad (3.24)$$

Para  $0 \leq n \leq c$ :

$$\lambda p_{n-1} + (n+1)\mu p_{n+1} = \lambda p_n + n\mu p_n \quad (3.25)$$

$$\lambda p_{n-1} + (n+1)\mu p_{n+1} = (\lambda + n\mu)p_n \quad (3.26)$$

Isolando-se  $p_{n+1}$ :

$$p_{n+1} = \frac{\lambda + n\mu}{(n+1)\mu} p_n - \frac{\lambda}{(n+1)\mu} p_{n-1}, \quad \text{para } 0 \leq n \leq c \quad (3.27)$$

Em geral, as equações para o sistema em equilíbrio podem ser escritas como:

$$\begin{cases} p_{n+1} = \frac{\lambda + n\mu}{(n+1)\mu} p_n - \frac{\lambda}{(n+1)\mu} p_{n-1}, & \text{para } 0 \leq n \leq c \\ p_{n+1} = \left(1 + \frac{\lambda}{c\mu + (n-c)\theta}\right) p_n - \frac{\lambda}{c\mu + (n-c)\theta} p_{n-1}, & \text{para } n > c \end{cases} \quad (3.28)$$

Observando-se o diagrama de transição no estado  $(c-1)$  e no estado  $c$ , em que a fila começa, as equações (3.28) podem ser reescritas como:

$$\begin{cases} \lambda p_n = (n+1)\mu p_{n+1}, & 0 \leq n \leq (c-1) \\ \lambda p_n = (c\mu + (n+1-c)\theta) p_{n+1}, & n \geq c \end{cases} \quad (3.29)$$

Resolvendo as equações (3.29), obtém-se as probabilidades dos estados do sistema de fila adotado. Para essa finalidade, novamente consideram-se os dois casos:

Para  $0 \leq n \leq c$ :

Neste caso, o sistema M/M/c+M é equivalente ao sistema M/M/c/c e a solução para a equação em equilíbrio correspondente é igual a (ver 3.10):

$$p_n = \frac{\left(\frac{\lambda}{\mu}\right)^n}{n!} p_0, \quad \text{para } 0 \leq n \leq c \quad (3.30)$$

Para  $n > c$ :

Considerando a equação  $\lambda p_n = [c\mu + (n+1-c)\theta] p_{n+1}$ , tem-se:

se  $n = c$ :

$$\lambda p_c = [c\mu + (c+1-c)\theta] p_{c+1} \quad (3.31)$$

$$p_{c+1} = \frac{\lambda}{c\mu + \theta} p_c \quad (3.32)$$

$$p_{c+1} = \frac{\lambda}{c\mu + \theta} \cdot \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!} p_0 \quad (3.33)$$

se  $n = c+1$ :

$$\lambda p_{c+1} = [c\mu + (c+2-c)\theta] p_{c+2} \quad (3.34)$$

$$p_{c+2} = \frac{\lambda}{c\mu + 2\theta} p_{c+1} \quad (3.35)$$

$$p_{c+2} = \frac{\lambda}{c\mu + 2\theta} \cdot \frac{\lambda}{c\mu + \theta} \cdot \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!} p_0 \quad (3.36)$$

se  $n = c+2$ :

$$\lambda p_{c+2} = [c\mu + (c+3-c)\theta] p_{c+3} \quad (3.37)$$

$$p_{c+3} = \frac{\lambda}{c\mu + 3\theta} p_{c+2} \quad (3.38)$$

$$p_{c+3} = \frac{\lambda}{c\mu + 3\theta} \cdot \frac{\lambda}{c\mu + 2\theta} \cdot \frac{\lambda}{c\mu + \theta} \cdot \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!} p_0 \quad (3.39)$$

Por indução matemática, pode-se escrever:

$$p_n = \prod_{i=c+1}^n \frac{\lambda}{c\mu + (i-c)\theta} \cdot \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!} p_0, \quad \text{para } n \geq c+1 \quad (3.40)$$

Vale notar que a equação acima pode ser escrita do seguinte modo:

$$p_n = \frac{\lambda^{n-c}}{\prod_{i=c+1}^n [c\mu + (i-c)\theta]} \cdot \frac{\left(\frac{\lambda}{\mu}\right)^c}{c!} p_0, \quad \text{para } n \geq c+1 \quad (3.41)$$

Em resumo, as soluções para o sistema de fila com abandono M/M/c+M são dadas por:

$$p_n = \begin{cases} \frac{(\lambda/\mu)^n}{n!} p_0, & \text{para } 0 \leq n \leq c \\ \prod_{i=c+1}^n \frac{\lambda}{c\mu + (i-c)\theta} \cdot \frac{(\lambda/\mu)^c}{c!} p_0, & \text{para } n \geq c+1 \end{cases} \quad (3.42)$$

Obtendo  $p_0$ :

Considerando que:

$$p_0 + p_1 + p_2 + \dots + p_{c-1} + p_c + p_{c+1} + \dots = 1 \quad (3.43)$$

e utilizando as equações (3.42), obtidas para  $p_n$ , podemos escrever:

$$\sum_{i=0}^c \frac{(\lambda/\mu)^i}{i!} p_0 + \sum_{i=c+1}^{\infty} \prod_{k=c+1}^i \frac{\lambda}{c\mu + (k-c)\theta} \cdot \frac{(\lambda/\mu)^c}{c!} p_0 = 1 \quad (3.44)$$

$$\left[ \sum_{i=0}^c \frac{(\lambda/\mu)^i}{i!} + \sum_{i=c+1}^{\infty} \prod_{k=c+1}^i \frac{\lambda}{c\mu + (k-c)\theta} \cdot \frac{(\lambda/\mu)^c}{c!} \right] p_0 = 1 \quad (3.45)$$

$$p_0 = \left[ \sum_{i=0}^c \frac{(\lambda/\mu)^i}{i!} + \sum_{i=c+1}^{\infty} \prod_{k=c+1}^i \frac{\lambda}{c\mu + (k-c)\theta} \cdot \frac{(\lambda/\mu)^c}{c!} \right]^{-1} \quad (3.46)$$

### 3.3.2 Convergência da soma infinita em $p_0$ .

A existência de  $p_0$  depende da convergência da soma infinita que aparece na segunda parcela da soma de  $p_0$  (equação (3.46)). Para tanto, vamos considerar duas situações:

Quando  $\mu = \theta$ :

Nesse caso, o sistema de fila com abandono M/M/c+M fica equivalente ao sistema M/M/ $\infty$  (isto é M/M/c+M  $\approx$  M/M/ $\infty$ ) e a probabilidade dos estados do sistema é dada por:

$$p_n = \frac{(\lambda/\mu)^n}{n!} p_0 \quad (3.47)$$

Como a soma das probabilidades de todos os estados é unitária (ver 3.43), tem-se:

$$\sum_{n=0}^{\infty} p_n = 1 \quad (3.48)$$

$$\sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!} p_0 = 1 \quad (3.49)$$

Então, obtém-se a expressão para determinar  $p_0$  no sistema M/M/ $\infty$ :

$$p_0 = \frac{1}{\sum_{n=0}^{\infty} \frac{(\lambda/\mu)^n}{n!}} \quad (3.50)$$

$$p_0 = \frac{1}{e^{\lambda/\mu}} \Rightarrow p_0 = e^{-\lambda/\mu} \quad (3.51)$$

Quando  $\mu \neq \theta$ :

Considera-se  $d_j$  como a taxa de morte do estado  $j$  com  $0 \leq j \leq \infty$ . Então, pode-se escrever:

$$j \cdot \min(\mu, \theta) \leq d_j \leq j \cdot \max(\mu, \theta) \quad (3.52)$$

Os limites  $j \cdot \min(\mu, \theta)$  e  $j \cdot \max(\mu, \theta)$  correspondem às taxas de morte da fila M/M/ $\infty$ , e  $\min(\mu, \theta)$  e  $\max(\mu, \theta)$  as suas taxas de serviço. É possível provar-se via ordem estocástica entre distribuições que essas duas filas M/M/ $\infty$  estabelecem os limites estocásticos superior e inferior para o sistema M/M/c+M (Erlang-A). A soma infinita em  $p_0$  converge como consequência do limite inferior acima e obtém-se a condição:

$$\sum_{i=0}^c \frac{(\lambda/\mu)^i}{i!} + \sum_{i=c+1}^{\infty} \prod_{k=c+1}^i \frac{\lambda}{c\mu + (k-c)\theta} \cdot \frac{(\lambda/\mu)^c}{c!} \leq \sum_{i=0}^{\infty} \frac{(\lambda/\min(\mu, \theta))^i}{i!} = e^{\frac{-\lambda}{\min(\mu, \theta)}} \quad (3.53)$$

Para evitar dificuldades numéricas causadas pela soma infinita, Palm (1957) reescreveu a distribuição de equilíbrio do sistema Erlang-A utilizando a função gama:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt, \quad x > 0 \quad (3.54)$$

e a função gama incompleta:

$$\gamma(x, y) = \int_0^y t^{x-1} e^{-t} dt, \quad x > 0, y > 0 \quad (3.55)$$

Definiu, ainda, a função auxiliar:

$$A(x, y) = \frac{x e^y}{y^x} \gamma(x, y) = 1 + \sum_{n=1}^{\infty} \frac{y^n}{\prod_{k=1}^n (x+k)}, \quad x > 0, y > 0 \quad (3.56)$$

e usou a probabilidade de bloqueio do sistema Erlang-B (M/M/c/c), reescrita como:

$$E_{1,c} = \frac{(\lambda/\mu)^c}{\sum_{n=0}^c \frac{(\lambda/\mu)^n}{n!}} \quad (3.57)$$

calculada de modo mais simples com uma fórmula recursiva encontrada ao substituir  $c$  por 0, 1, 2, ....

Se  $c = 0$ , obtém-se:

$$E_{1,0} = \frac{(\lambda/\mu)^0}{\sum_{i=0}^0 \frac{(\lambda/\mu)^i}{i!}} = \frac{1}{1} = 1 \quad (3.58)$$

Se  $c = 1$ , tem-se:

$$E_{1,1} = \frac{\frac{(\lambda/\mu)^1}{1!}}{\sum_{i=0}^1 \frac{(\lambda/\mu)^i}{i!}} = \frac{\frac{\lambda}{\mu} \cdot 1}{1 + \frac{\lambda}{\mu} \cdot 1} = \frac{\frac{\lambda}{1 \cdot \mu} \cdot 1}{1 + \frac{\lambda}{1 \cdot \mu} \cdot 1} = \frac{\rho \cdot E_{1,0}}{1 + \rho \cdot E_{1,0}}, \quad \text{com } \rho = \frac{\lambda}{c\mu} \quad (3.59)$$

Continuando o processo para os próximos valores de  $c$ , pode-se generalizar uma expressão recursiva para a probabilidade de bloqueio:

$$E_{1,c} = \frac{\rho \cdot E_{1,c-1}}{1 + \rho \cdot E_{1,c-1}}, \quad E_{1,0} = 1, \quad \rho = \lambda/c\mu, \quad c \geq 1 \quad (3.60)$$

Aplicando as expressões de  $A(x, y)$  e  $E_{1,n}$ , deduziu-se as soluções para a distribuição em equilíbrio sem a presença da soma infinita. A dedução dessas soluções é apresentada a seguir. Iniciando com a expressão de  $p_0$  obtida anteriormente, tem-se:

$$p_0^{-1} = \sum_{i=0}^c \frac{(\lambda/\mu)^i}{i!} + \sum_{i=c+1}^{\infty} \prod_{k=c+1}^i \frac{\lambda}{c\mu + (k-c)\theta} \cdot \frac{(\lambda/\mu)^c}{c!} \quad (3.61)$$

Colocando-se em evidência o termo  $\frac{(\lambda/\mu)^c}{c!}$ :



$$p_0^{-1} = \frac{(\lambda/\mu)^c}{c!} \cdot \left[ \frac{\sum_{i=0}^c \frac{(\lambda/\mu)^i}{i!}}{\frac{(\lambda/\mu)^c}{c!}} + \sum_{i=c+1}^{\infty} \prod_{k=c+1}^i \frac{\lambda}{c\mu + (k-c)\theta} \right] \quad (3.62)$$

Considerando-se que  $\frac{\sum_{i=0}^c \frac{(\lambda/\mu)^i}{i!}}{\frac{(\lambda/\mu)^c}{c!}} = \frac{1}{E_{1,c}}$  e mudando-se o valor do índice do somatório e produtório, tem-se uma expressão mais simples para  $p_0^{-1}$ :

$$p_0^{-1} = \frac{(\lambda/\mu)^c}{c!} \cdot \left[ \frac{1}{E_{1,c}} + \sum_{i=1}^{\infty} \prod_{k=1}^i \frac{\lambda}{c\mu + k\theta} \right] \quad (3.63)$$

Dividindo-se por  $\theta$  o segundo termo da soma, na equação (63), tem-se:

$$p_0^{-1} = \frac{(\lambda/\mu)^c}{c!} \cdot \left[ \frac{1}{E_{1,c}} + \sum_{i=1}^{\infty} \prod_{k=1}^i \frac{\lambda/\theta}{c\mu/\theta + k\theta/\theta} \right] \quad (3.64)$$

$$p_0^{-1} = \frac{(\lambda/\mu)^c}{c!} \cdot \left[ \frac{1}{E_{1,c}} + \sum_{i=1}^{\infty} \frac{(\lambda/\theta)^i}{\prod_{k=1}^i (c\mu/\theta + k)} \right] \quad (3.65)$$

A soma infinita que aparece na última igualdade (equação (3.65)) pode ser substituída pela função  $A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1$ , definida anteriormente com  $x = \frac{c\mu}{\theta}$  e  $y = \frac{\lambda}{\theta}$ . Reescrevendo:

$$p_0^{-1} = \frac{(\lambda/\mu)^c}{c!} \cdot \left[ \frac{1}{E_{1,c}} + A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1 \right] \quad (3.66)$$

$$p_0^{-1} = \frac{(\lambda/\mu)^c}{c!} \cdot \left[ \frac{1 + \left[ A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1 \right] \cdot E_{1,c}}{E_{1,c}} \right] \quad (3.67)$$

Finalmente, a expressão para determinar o valor de  $p_0$  é dada por:

$$p_0 = \frac{E_{1,c}}{1 + \left[ A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1 \right] \cdot E_{1,c}} \cdot \frac{c!}{\left(\frac{\lambda}{\mu}\right)^c} \quad (3.68)$$

Nessa igualdade, não ocorre a soma infinita, pois a função  $A(x, y)$  depende da função gama incompleta, cujos valores podem ser encontrados facilmente. As soluções para as equações de equilíbrio do sistema, considerando essa nova expressão para  $p_0$  (equação (3.68)), são obtidas. Separando-se os casos para  $n \leq c$  e para  $n > c$ .

Para  $1 \leq n \leq c$ :

$$p_n = \frac{(\lambda/\mu)^n}{n!} p_0 \quad (3.69)$$

$$p_n = \frac{(\lambda/\mu)^n}{n!} \cdot \frac{E_{1,c}}{1 + \left[ A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1 \right] \cdot E_{1,c}} \cdot \frac{c!}{\left(\frac{\lambda}{\mu}\right)^c} \quad (3.70)$$

$$p_n = \frac{E_{1,c}}{1 + \left[ A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1 \right] \cdot E_{1,c}} \cdot \frac{c!}{n! \left(\frac{\lambda}{\mu}\right)^{c-n}} \quad (3.71)$$

especificamente para  $n = c$ , temos que:

$$p_c = \frac{E_{1,c}}{1 + \left[ A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1 \right] \cdot E_{1,c}} \cdot \frac{c!}{c! \left(\frac{\lambda}{\mu}\right)^0} \quad (3.72)$$

$$p_c = \frac{E_{1,c}}{1 + \left[ A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1 \right] \cdot E_{1,c}} \quad (3.73)$$

Para  $n > c$ :

$$p_n = \prod_{i=c+1}^n \frac{\lambda}{c\mu + (i-c)\theta} \cdot \frac{(\lambda/\mu)^c}{c!} p_0 \quad (3.74)$$

Substituindo  $p_0$ :

$$p_n = \frac{\lambda^{n-c}}{\prod_{i=1}^{n-c} (c\mu + i\theta)} \cdot \frac{(\lambda/\mu)^c}{c!} \cdot \frac{E_{1,c}}{1 + \left[ A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1 \right] \cdot E_{1,c}} \cdot \frac{c!}{(\lambda/\mu)^c} \quad (3.75)$$

$$p_n = \frac{\lambda^{n-c}}{\prod_{i=1}^{n-c} (c\mu + i\theta)} \cdot \frac{E_{1,c}}{1 + \left[ A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1 \right] \cdot E_{1,c}} \quad (3.76)$$

Dividindo-se por  $\theta$  o primeiro fator do produto obtém-se:

$$p_n = \frac{(\lambda/\theta)^{n-c}}{\prod_{i=1}^{n-c} \left(\frac{c\mu}{\theta}, i\right)} \cdot \frac{E_{1,c}}{1 + \left[A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1\right] \cdot E_{1,c}} \quad (3.77)$$

Em resumo, as soluções para o sistema de equações do modelo de fila com abandono M/M/c+M são dadas por (MANDELBAUM e ZELTYN, 2005):

$$p_n = \begin{cases} \frac{c!}{n! \left(\frac{\lambda}{\mu}\right)^{c-n}} p_c, & \text{para } 0 \leq n \leq c \\ \frac{(\lambda/\theta)^{n-c}}{\prod_{i=1}^{n-c} \left(\frac{c\mu}{\theta} + i\right)} p_c, & \text{para } n \geq c + 1 \end{cases} \quad (3.78)$$

com

$$p_c = \frac{E_{1,c}}{1 + \left[A\left(\frac{c\mu}{\theta}, \frac{\lambda}{\theta}\right) - 1\right] \cdot E_{1,c}} \quad (3.79)$$

A partir dessas equações, algumas medidas de desempenho do sistema são deduzidas:

*Probabilidade de espera no sistema:* representa a fração de clientes que aguardam na fila e sua expressão é dada por:

$$P\{W > 0\} = \sum_{j=c}^{\infty} \pi_j = \frac{A(c\mu/\theta, \lambda/\theta) E_{1,c}}{1 + [A(c\mu/\theta, \lambda/\mu) - 1] E_{1,c}} \quad (3.80)$$

Trata-se de uma medida importante porque identifica o foco operacional da organização.

*Fração de abandono:* probabilidade de abandonar o serviço, dado que todos os servidores estão ocupados e  $n$  clientes estão na fila no instante da sua chegada, dada por:

$$P_n\{Ab\} = 1 - P_n\{Sr\} = \frac{(n+1)\theta}{c\mu + (n+1)\theta}, \quad n \geq 0 \quad (3.81)$$

*Probabilidade de abandonar devido a impaciência:* a fração de abandono  $P\{Ab\}$  pode ser entendida simplesmente como o produto entre a probabilidade de abandonar devido a impaciência  $P\{Ab | W > 0\}$  e a probabilidade de espera no sistema  $P\{W > 0\}$ , ou seja:

$$P\{Ab\} = P\{Ab | W > 0\} \times P\{W > 0\} \quad (3.82)$$

Então, segue que:

$$P\{Ab | W > 0\} = \frac{\sum_{n=c}^{\infty} \pi_n P_{n-c}\{Ab\}}{P\{W > 0\}} = \frac{1}{\rho A(c\mu/\theta, \lambda/\mu)} + 1 - \frac{1}{\rho} \quad (3.83)$$

*Tempo médio de espera dos clientes que estão esperando:* explorando a relação linear entre a fração do abandono  $P\{Ab\}$  e o tempo médio de espera  $E\{W\}$  que ocorre não somente no sistema Erlang-A, mas em outros modelos com paciência exponencial e expressa por  $P\{Ab\} = \theta \cdot E(W)$ , encontra-se:

$$E[W | W > 0] = \frac{1}{\theta} \left[ \frac{1}{\rho A(c\mu/\theta, \lambda/\theta)} + 1 - \frac{1}{\rho} \right] \quad (3.84)$$

Outras medidas de desempenho são obtidas de maneira tradicional utilizando-se a fórmula de Little:

$$\begin{aligned} L_q &= \frac{\lambda}{\theta} P\{Ab\} \\ W_q &= \frac{P\{Ab\}}{\theta} \\ W &= W_q + \frac{1}{\mu} = \frac{P\{Ab\}}{\theta} + \frac{1}{\mu} \\ L &= \lambda W = \frac{\lambda P\{Ab\}}{\theta} + \rho \end{aligned} \quad (3.85)$$

Mais detalhes sobre esse sistema de fila, que incorpora o abandono, são encontrados em Manddelbaum e Zeltyn (2005).

### 3.4 O Sistema M/M/c + G

O sistema com abandono M/M/c + G tem um processo de chegada Poisson com taxa de chegada  $\lambda$  e tempos entre chegadas distribuídos exponencialmente, assim como o sistema anterior. No entanto, possui um tempo de paciência  $\tau$  assumido ter uma distribuição genérica de paciência  $G$ . É servido por  $c$  canais de serviços idênticos, operando em paralelo e independentemente uns dos outros, atendendo a uma única fila de espera. O tempo de serviço é considerado exponencialmente distribuído com taxa  $\mu$ . O processo de serviço, de chegada e a paciência são mutuamente independentes.

#### 3.4.1 Modelagem do sistema M/M/c+G

Admite-se, ainda, que  $\bar{G}$  é a função de sobrevivência do tempo de paciência  $\tau$ , ou seja,  $\bar{G}(x) = 1 - G(x) = P\{\tau > x\}$ ,  $x \geq 0$ . Estando o sistema em equilíbrio estatístico, admite-se que o cliente que chega, encontra um tempo virtual  $V$  de espera na fila. Esse tempo pode ser entendido como sendo o tempo que um cliente com paciência infinita teria que aguardar na fila. Desse modo, o tempo que o cliente efetivamente espera na fila é  $W'_q = \min(V, \tau)$ .

Nesse sistema, Baccelli e Hebuterne (1981) admitem que os clientes que chegam podem calcular o tempo virtual de espera (ou tempo de espera oferecido)  $V$  no instante da sua chegada, e se esse tempo for maior que o seu tempo de paciência, o cliente abandona o sistema imediatamente e não se junta à fila. Desse modo, consideram o processo de Markov  $\{(N(t), \eta(t)), t \geq 0\}$ , em que  $N(t)$  é o número de agentes ocupados no instante  $t$  e  $\eta(t)$  é o tempo virtual de espera do cliente que chegou no instante  $t$ . Estando o sistema em equilíbrio estatístico, definem a função de densidade de probabilidade do tempo virtual de espera  $v(t)$  e a distribuição estacionária  $\pi_n$  :

$$\begin{cases} \nu(t) = \lim_{t \rightarrow \infty} \lim_{\varepsilon \rightarrow 0} \frac{P\{N(t) = c, x < \eta(t) \leq x + \varepsilon\}}{\varepsilon}, & x \geq 0 \\ \pi_n = \lim_{t \rightarrow \infty} P\{N(t) = n, \eta(t) = 0\}, & 0 \leq n \leq c-1 \end{cases} \quad (3.86)$$

Conforme esses mesmos autores, a única solução das equações acima, estando o sistema em equilíbrio estatístico, é dada por:

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} \pi_0, \quad 0 \leq n \leq c-1 \quad (3.87)$$

e

$$\nu(x) = \lambda \pi_{c-1} \exp\left\{\lambda \int_0^x \bar{G}(u) du - c\mu x\right\} \quad (3.88)$$

em que

$$\pi_0 = \left[1 + \frac{\lambda}{\mu} + \dots + \left(\frac{\lambda}{\mu}\right)^{c-1} \frac{1}{(c-1)!} (1 + \lambda J)\right]^{-1} \quad (3.89)$$

em que  $J$  (em(3.89)) é dado por:

$$J = \int_0^{\infty} \exp\left\{\lambda \int_0^x \bar{G}(u) du - c\mu x\right\} dx \quad (3.90)$$

A probabilidade do cliente abandonar o sistema se seu tempo de paciência for menor que o tempo virtual de espera (probabilidade de abandono) é (BACELLI E HEBUTERNE, 1981):

$$P\{Ab\} = \left(1 - \frac{c\mu}{\lambda}\right) \left(1 - \sum_{n=0}^{c-1} \pi_n\right) + \pi_{c-1} \quad (3.91)$$

As fórmulas exatas das medidas de desempenho do sistema M/M/c+G, baseadas em Baccelli e Hebuterne (1981) e apresentadas na tese de Zeltyn (2004), são reproduzidas a seguir.

Primeiramente é definida a função  $H(x) = \int_0^x \bar{G}(u) du$  e observa que  $H(\infty) = \bar{\tau}$ , em que  $\tau$  é o

tempo médio de paciência. Introduz-se as seguintes integrais definidas conforme abaixo, que serão usadas para o cálculo de medidas de desempenho:

$$J = \int_0^{\infty} \exp \{ \lambda H(x) - c\mu x \} dx \quad (3.92)$$

$$J_H = \int_0^{\infty} H(x) \exp \{ \lambda H(x) - c\mu x \} dx \quad (3.93)$$

Define-se ainda:

$$\varepsilon = \frac{\sum_{n=0}^{c-1} \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n}{(c-1)! \left( \frac{\lambda}{\mu} \right)^{c-1}} = \int_0^{\infty} e^{-t} (1 + t\mu/\lambda)^{c-1} dt \quad (3.94)$$

Um conveniente modo de calcular  $\varepsilon$  é utilizar a sua expressão recursiva:

$$\varepsilon_k = \frac{\sum_{n=0}^k \frac{1}{n!} \left( \frac{\lambda}{\mu} \right)^n}{\frac{1}{k!} \left( \frac{\lambda}{\mu} \right)^k}, \quad k \geq 0 \quad (3.95)$$

e usar:

$$\varepsilon_0 = 1; \quad \varepsilon_k = 1 + \frac{k\mu}{\lambda} \varepsilon_{k-1}; \quad 1 \leq k \leq c-1; \quad \varepsilon = \varepsilon_{c-1} \quad (3.96)$$

Várias medidas de desempenho importantes para o sistema podem ser expressas pelo bloco de definições acima e a distribuição da paciência  $G$ . Algumas delas são:

$$\text{Fração do tempo virtual de espera: } P\{V > 0\} = \frac{\lambda J}{\varepsilon + \lambda J} \quad (3.97)$$

$$\text{Fração do tempo de espera no sistema: } P\{W > 0\} = \frac{\lambda J}{\varepsilon + \lambda J} \bar{G}(0) \quad (3.98)$$

$$\text{Probabilidade de abandono: } P\{Ab\} = \frac{1 + (\lambda - c\mu)J}{\varepsilon + \lambda J} \quad (3.99)$$

$$\text{Tempo médio de espera no sistema: } E(W) = \frac{\lambda J_H}{\varepsilon + \lambda J} \quad (3.100)$$

$$\text{Tamanho médio da fila: } E(Lq) = \frac{\lambda^2 J_H}{\varepsilon + \lambda J} \quad (3.101)$$

Outras medidas de desempenho e demais informações sobre o sistema M/M/c+G podem ser obtidas em Zeltyn (2004).

### 3.5 O Sistema M/G/c+G

Esse sistema com abandono multiservidor, com  $c$  servidores idênticos operando em paralelo e independentemente uns dos outros, recebe os clientes que chegam de acordo com o processo de Poisson, com taxa de chegada  $\lambda$ , e tempos entre serviços distribuídos conforme uma distribuição genérica ( $G$ ). No instante da sua chegada, se o cliente encontrar pelo menos um dos servidores livres, ele é servido imediatamente. Caso contrário, ele aguarda em uma fila de espera. O tempo de serviço é descrito por uma variável aleatória  $B$  com função de distribuição geral de probabilidade  $B(x)$ . Enquanto aguarda na fila, o cliente pode abandonar o sistema com uma função de distribuição geral de probabilidade  $R(x)$ . Admite-se que as variáveis  $B$  e  $R$  são independentes uma da outra e também do processo de chegada.

O modelo M/G/c+G não permite uma solução analítica tratável e por isso seguiu-se, neste estudo, as mesmas formulações adotadas por Iravani e Balcioglu (2008), que utilizaram o sistema M/G/1+G para obter as aproximações para o sistema M/G/c+G. Esses autores consideram o sistema com um único servidor com o mesmo processo de chegada Poisson e a mesma distribuição do tempo de abandono que o sistema multiservidor. Preservaram, também, a distribuição do tempo de serviço do sistema com único servidor como a do sistema multiservidor, mas sua taxa é aumentada em  $c$  vezes (o número de servidores), ficando então  $c\mu_B$ . Denotando-se o sistema por M/G<sup>c</sup>/1+G e, conseqüentemente, a intensidade de tráfego para ambos os sistemas fica igual a:

$$\rho_{MIG/c+G} = \rho_{MIG^c/1+G} = \frac{\lambda}{c\mu_B} \quad (3.102)$$

Considerando que o sistema de fila M/G/1+G foi utilizado para se obter as aproximações do sistema de fila multicanal M/G/c+G, é importante revisá-lo, como é feito a seguir.



### 3.5.1 Sistema de fila M/G/1+G

Neste sistema de fila com único canal de atendimento e com a possibilidade do usuário abandonar a fila antes de ser atendido, admite-se que as chegadas ao sistema ocorram de acordo com um processo de Poisson com taxa de chegada  $\lambda$  e com disciplina de atendimento FIFO. A variável aleatória  $B$  representa os tempos de serviços, obedecendo uma função de distribuição genérica  $B(x)$ . Do mesmo modo, a variável aleatória  $R$  denota os tempos para abandonar (paciência), seguindo uma distribuição genérica  $R(x)$ . Considera-se que ambas as variáveis aleatórias são independentes e identicamente distribuídas (iid) e, também, que são independentes uma da outra e do processo de chegada.

Considera-se, ainda, as variáveis aleatórias  $W_S$  para descrever o tempo de espera dos usuários até serem atendidos por um agente,  $W_R$  para descrever o tempo de espera dos usuários até abandonar o sistema (tempo de paciência), e  $W_T$  para descrever o tempo de espera de todos os usuários (os que abandonaram e os que foram servidos). Usando os teoremas 4, 5 e 6 de Stanford (1979) e definindo a variável aleatória  $V$  como o tempo virtual de espera com  $F(x)$  e  $f(x)$ , as suas respectivas função de distribuição e função de densidade (assumindo que existem para  $x > 0$ ), para a fila G/G/1+G, Iravani e Balcioglu (2008) definem:

$$\text{Probabilidade do sistema estar vazio: } P_0 = F(0) \quad (3.103)$$

$$\text{Probabilidade de abandono: } P_A = 1 - \frac{1 - P_0}{\rho} \quad (3.104)$$

$$\text{Função distribuição do tempo de espera dos servidos: } W_S(x) = \int_0^x \frac{dF(t)\bar{R}dt}{1 - P_A} \quad (3.105)$$

$$\text{Probabilidade do tempo de espera dos servidos ser igual a zero: } P(W_S = 0) = \frac{P_0}{1 - P_A} \quad (3.106)$$

$$\text{Função distribuição do tempo de espera para abandonar: } W_R(x) = \int_0^x \frac{\bar{F}(t)R(t)dt}{P_A} \quad (3.107)$$

*Função distribuição do tempo de espera de todos os usuários:*

$$W_T(x) = (1 - P_A)W_S(x) + P_A W_R(x) \quad (3.108)$$

em que  $\rho = \frac{\lambda}{\mu}$  sendo  $\mu$  a taxa de serviço e  $\bar{F}(t)$  e  $\bar{R}(t)$  as respectivas funções de distribuição complementar do tempo virtual de espera e do tempo para abandonar.

A função de densidade do tempo virtual de espera da fila M/G/1+G foi estabelecida por Iravani e Balcioglu (2008) utilizando o método *level crossing*, desenvolvido originalmente por Brill (1979), Cohen (1977), Cohen e Rubinovitch (1977) e Brill (2008). Este método garante que, para um certo nível  $x > 0$  e estando o sistema em equilíbrio, a taxa de passagem acima do nível é igual à taxa de passagem abaixo do nível, ou seja:

$$\text{downcrossing rate}(x) = \text{upcrossing rate}(x)$$

Usando a lei estabelecida acima Iravani e Balcioglu (2008), deduzem a equação integral para o tempo virtual de espera:

$$f(x) = \lambda P_0 \bar{B}(x) + \lambda \int_0^x \bar{B}(x-t) \bar{R}(t) f(t) dt \quad (3.109)$$

com a equação de normalização igual a:  $P_0 + \int_0^{\infty} f(t) dt = 1$  (3.110)

A equação (3.109) pode ser entendida do seguinte modo: o lado esquerdo é a taxa de passagem para baixo do nível  $x$  e o lado direito, a taxa de passagem para cima do nível  $x$ , consistindo de duas partes. A primeira considera que se um usuário, ao chegar, encontra o sistema vazio, ele causa uma passagem para cima do nível se seu tempo de serviço é maior que  $x$ . A segunda parte considera que se um usuário, ao chegar, encontra uma quantidade  $t$  de trabalhos no sistema, ele causa uma passagem para cima do nível se seu tempo de serviço excede  $(x-t)$ , com probabilidade  $1 - B(x-t)$ , e não abandona o sistema com probabilidade  $(1 - R(t))$ .

A equação (3.109) é uma equação integral de Volterra de segunda espécie, que exige uma solução numérica na determinação de resultados aproximados dessas equações. Iravani e Balcioglu (2008) empregaram a regra do trapézio (Apêndice I) para obter as aproximações para  $f(x)$  e obtiveram:

$$f(\Delta) = \lambda P_0 \bar{B}(\Delta) + \lambda \Delta \left[ \frac{1}{2} \bar{B}(\Delta) \bar{R}(0) f(0) + \frac{1}{2} \bar{B}(0) \bar{R}(\Delta) f(\Delta) \right] \quad (3.111)$$

Fazendo  $\Delta = 0$  na equação (3.111), obtém-se:

$$f(0) = \lambda P_0 \quad (3.112)$$

Eliminando-se os colchetes da equação (3.111) e substituindo (3.112), obtém-se:

$$f(\Delta) = \lambda P_0 \bar{B}(\Delta) + \lambda \Delta \frac{1}{2} \bar{B}(\Delta) \bar{R}(0) \lambda P_0 + \lambda \Delta \frac{1}{2} \bar{B}(0) \bar{R}(\Delta) f(\Delta)$$

Considerando que  $\bar{R}(0) = 1$  e  $\bar{B}(0) = 1$  tem-se que:

$$f(\Delta) = \lambda P_0 \bar{B}(\Delta) + \lambda \Delta \frac{1}{2} \bar{B}(\Delta) \lambda P_0 + \lambda \Delta \frac{1}{2} \bar{R}(\Delta) f(\Delta)$$

Transpondo os termos com  $f(\Delta)$  para o lado esquerdo da igualdade acima e fatorando em  $P_0$ , obtém-se:

$$f(\Delta) - \lambda \Delta \frac{1}{2} \bar{R}(\Delta) f(\Delta) = P_0 [\lambda \bar{B}(\Delta) + \lambda \Delta \frac{1}{2} \bar{B}(\Delta) \lambda]$$

Fatorando o lado esquerdo da equação, encontra-se uma expressão para  $f(\Delta)$ :

$$f(\Delta) [1 - \lambda \Delta \frac{1}{2} \bar{R}(\Delta)] = P_0 [\lambda \bar{B}(\Delta) + \lambda \Delta \frac{1}{2} \bar{B}(\Delta) \lambda]$$

$$f(\Delta) = P_0 \frac{\lambda \bar{B}(\Delta) + \lambda \Delta \frac{1}{2} \bar{B}(\Delta) \lambda}{1 - \lambda \Delta \frac{1}{2} \bar{R}(\Delta)}$$

Chamando de  $f_1(\Delta)$  o segundo fator do lado direito da última equação, tem-se:

$$f_1(\Delta) = \frac{\lambda \bar{B}(\Delta) + \lambda \Delta \frac{1}{2} \bar{B}(\Delta) \lambda}{1 - \lambda \Delta \frac{1}{2} \bar{R}(\Delta)} \quad (3.113)$$

Se  $\Delta = 0$  em (3.113), obtém o resultado:

$$f_1(0) = \lambda$$

que substituído na equação (3.112) reescreve-se:

$$f(0) = P_0 f_1(0) \quad (3.114)$$

Da equação (3.114) pode-se obter:

$$f(\Delta) = P_0 f_1(\Delta) \quad (3.115)$$

Substituindo a equação (3.113) na equação (3.115) obtém-se:

$$f(\Delta) = P_0 \frac{\lambda \bar{B}(\Delta) + \frac{\lambda \Delta \bar{B}(\Delta) f_1(0)}{2}}{1 - \frac{\lambda \Delta \bar{R}(\Delta)}{2}} \quad (3.116)$$

Aplicando a regra do trapézio (Apêndice I) na equação (3.116) sucessivamente,  $n$  vezes, obtém-se uma aproximação para  $f(x)$ , o tempo virtual de espera, dada em (3.109):

$$f(n\Delta) = P_0 f_1(n\Delta)$$

$$f(n\Delta) = P_0 \frac{\lambda \bar{B}(n\Delta) + \lambda \Delta \left[ \frac{\bar{B}(n\Delta)f_1(0)}{2} + \sum_{i=1}^{n-1} \bar{B}((n-i)\Delta) \bar{R}(i\Delta) f_1(i\Delta) \right]}{1 + \frac{\lambda \Delta \bar{R}(n\Delta)}{2}} \quad (3.117)$$

Essa aplicação sucessiva da regra do trapézio (recursão) termina para um valor  $MAX$ , tal que para  $n \geq MAX$  tem-se que  $f(n\Delta) = 0$ . Aplicando a regra do trapézio em (3.110), Iravani e Balcioglu (2008) obtiveram uma equação para determinar o valor da probabilidade do sistema estar vazio:

$$P_0 \left\{ 1 + \Delta \left[ \frac{f_1(0) + f_1(MAX\Delta)}{2} + \sum_{n=1}^{MAX-1} f_1(n\Delta) \right] \right\} = 1 \quad (3.118)$$

A probabilidade de abandono  $P_A$  pode agora ser calculada usando as equações (3.104) e (3.118). A função de distribuição do tempo médio de espera dos usuários que são servidos pode ser determinada empregando a regra do trapézio na equação (3.105), considerando o número inteiro  $n_x$  satisfazendo a igualdade  $n_x \Delta = x$ :

$$P(0 < W_s \leq x) = \frac{\Delta P_0}{1 - P_A} \left\{ \frac{f_1(\Delta) \bar{R}(\Delta) + f_1(n_x \Delta) \bar{R}(n_x \Delta)}{2} + \sum_{n=2}^{n_x-1} f_1(n\Delta) \bar{R}(n\Delta) \right\} \quad (3.119)$$

Outra medida de desempenho obtida por Iravani e Balcioglu (2008) foi o  $k$ -ésimo momento do tempo de espera dos que são servidos. Para tanto, primeiramente derivaram a equação (3.109) e posteriormente aplicaram a regra do trapézio e obtiveram:

$$E[W_s^k] = \frac{\Delta P_0}{1 - P_A} \left\{ \frac{\Delta^k f_1(\Delta) \bar{R}(\Delta) + (MAX\Delta)^k f_1(MAX\Delta) \bar{R}(MAX\Delta)}{2} + \sum_{n=2}^{MAX-1} (n\Delta)^k f_1(n\Delta) \bar{R}(n\Delta) \right\} \quad (3.120)$$

### 3.5.2 Sistema de fila M/G/c + G

Voltando ao sistema M/G/c+G com  $c$  servidores, Iravani e Balcioglu (2008) argumentaram que o possível paralelismo entre os modelos de fila analiticamente tratáveis

M/M<sup>c</sup>/1+G e M/M/c+G seria um indicativo de um possível paralelismo entre os modelos de fila M/G<sup>c</sup>/1+G e M/G/c+G. Como consequência imediata deste fato, tem-se a igualdade das intensidades de tráfego dos dois últimos sistemas, como descrito na equação (3.102).

Uma razão para a compreensão desse paralelismo é que quando todos os servidores estão ocupados em ambos os sistemas, os usuários que aguardam na fila de espera serão atendidos com a mesma taxa de serviço  $c\mu$ , deixando os sistemas equivalentes. Desse modo, considerando que esses usuários aguardam para serem atendidos ou abandonam o sistema, Iravani e Balcioglu (2008) propõe a seguinte aproximação para a probabilidade de espera:

$$P_{M/G/C+G}(W > x | W > 0) \cong P_{M/G^c/1+G}(W > x | W > 0), \quad x > 0$$

Escrevendo separadamente essas probabilidades para os usuários que são atendidos (antes de abandonar) e para todos os usuários, temos:

$$P_{M/G/C+G}(W_S > x | W_S > 0) \cong P_{M/G^c/1+G}(W_S > x | W_S > 0)$$

$$P_{M/G/C+G}(W_T > x | W_T > 0) \cong P_{M/G^c/1+G}(W_T > x | W_T > 0)$$

Usando a definição de probabilidade condicional, das duas últimas igualdades segue que:

$$\begin{aligned} \frac{P_{M/G/C+G}(W_S > x, W_S > 0)}{P_{M/G/C+G}(W_S > 0)} &\cong \frac{P_{M/G^c/1+G}(W_S > x, W_S > 0)}{P_{M/G^c/1+G}(W_S > 0)} \\ 1 - \frac{P_{M/G/C+G}(0 < W_S \leq x)}{P_{M/G/C+G}(W_S > 0)} &\cong 1 - \frac{P_{M/G^c/1+G}(0 < W_S \leq 0)}{P_{M/G^c/1+G}(W_S > 0)} \\ \frac{P_{M/G/C+G}(0 < W_S \leq x)}{P_{M/G/C+G}(W_S > 0)} &\cong \frac{P_{M/G^c/1+G}(0 < W_S \leq x)}{P_{M/G^c/1+G}(W_S > 0)} \end{aligned} \quad (3.121)$$

e, ainda:

$$\begin{aligned}
\frac{P_{M/G/C+G}(W_T > x, W_T > 0)}{P_{M/G/C+G}(W_T > 0)} &\cong \frac{P_{M/G^C/1+G}(W_T > x, W_T > 0)}{P_{M/G^C/1+G}(W_T > 0)} \\
1 - \frac{P_{M/G/C+G}(0 < W_T \leq x)}{P_{M/G/C+G}(W_T > 0)} &\cong 1 - \frac{P_{M/G^C/1+G}(0 < W_T \leq 0)}{P_{M/G^C/1+G}(W_T > 0)} \\
\frac{P_{M/G/C+G}(0 < W_T \leq x)}{P_{M/G/C+G}(W_T > 0)} &\cong \frac{P_{M/G^C/1+G}(0 < W_T \leq x)}{P_{M/G^C/1+G}(W_T > 0)} \quad (3.122)
\end{aligned}$$

Lembrando que:

i) na fila com um único servidor, neste caso  $M/G^C/1+G$ , tem-se  $P_{M/G^C/1+G}(W_s > 0) = 1 - P_{M/G^C/1+G}(W_s = 0)$  e, usando a equação (3.106), se reescreve:

$$P_{M/G^C/1+G}(W_s > 0) = 1 - \frac{P_0}{1 - P_A} \quad (3.123)$$

e, ainda,  $P_{M/G^C/1+G}(W_T > 0) = 1 - P_{M/G^C/1+G}(W_T = 0)$ . Considerando que  $P_{M/G^C/1+G}(W_T = 0) = P_0$ , tem-se:

$$P_{M/G^C/1+G}(W_T > 0) = 1 - P_0 \quad (3.124)$$

ii) na fila com múltiplos servidores, ou seja  $M/G/c+G$ , pode-se escrever:

$$P_{M/G/c+G}(W_s > 0) = 1 - P_{M/G/c+G}(W_s = 0) \quad (3.125)$$

e

$$P_{M/G/C+G}(W_T > 0) = 1 - P_{M/G/C+G}(W_T = 0) \quad (3.126)$$

Substituindo-se (3.123) a (3.126) nas equações (3.121) e (3.122), obtém-se:

$$\frac{P_{M/G/C+G}(0 < W_S \leq x)}{1 - P_{M/G/C+G}(W_S = 0)} \cong \frac{P_{M/G^C/1+G}(0 < W_S \leq x)}{1 - \frac{P_0}{1 - P_A}} \quad (3.127)$$

e

$$\frac{P_{M/G/C+G}(0 < W_T \leq x)}{1 - P_{M/G/C+G}(W_T = 0)} \cong \frac{P_{M/G^C/1+G}(0 < W_T \leq x)}{1 - P_0} \quad (3.128)$$

Isolando-se o numerador da fração do lado esquerdo da equação (3.127), tem-se:

$$P_{M/G/C+G}(0 < W_S \leq x) \cong (1 - P_{M/G/C+G}(W_S = 0)) \frac{P_{M/G^C/1+G}(0 < W_S \leq x)}{1 - \frac{P_0}{1 - P_A}} \quad (3.129)$$

e, considerando-se que  $P_{M/G/C+G}(0 < W_S \leq x) = P_{M/G/C+G}(W_S \leq x) - P_{M/G/C+G}(W_S = 0)$ , a equação (3.129) é reescrita como:

$$P_{M/G/C+G}(W_S \leq x) \cong P_{M/G/C+G}(W_S = 0) + (1 - P_{M/G/C+G}(W_S = 0)) \frac{P_{M/G^C/1+G}(0 < W_S \leq x)}{1 - \frac{P_0}{1 - P_A}} \quad (3.130)$$

Aplicando-se o mesmo procedimento na equação (3.128), obtém-se:



$$P_{M/G/C+G}(W_T \leq x) \cong P_{M/G/C+G}(W_T = 0) + (1 - P_{M/G/C+G}(W_T = 0)) \frac{P_{M/G^C/1+G}(0 < W_T \leq x)}{1 - P_0} \quad (3.131)$$

Usando probabilidade condicional, Iravani e Balcioglu (2008) obtiveram a probabilidade do tempo de espera ser zero dos usuários que são servidos antes de abandonar. Com esse resultado, deduziram o  $k$ -ésimo momento do tempo de espera dos usuários que são servidos e de todos os usuários, como segue:

$$P_{M/G/C+G}(W_S = 0) = \frac{P_{M/G/C+G}(W_T = 0)}{1 - P_{M/G/C+G}(A)} \quad (3.132)$$

$$E_{M/G/C+G}[W_S^k] \cong [1 - P_{M/G/C+G}(W_S = 0)] \frac{E_{M/G^C/1+G}[W_S^K]}{1 - \frac{P_0}{1 - P_A}} \quad (3.133)$$

$$E_{M/G/C+G}[W_T^K] \cong [1 - P_{M/G/C+G}(W_T = 0)] \frac{E_{M/G^C/1+G}[W_T^K]}{1 - P_0} \quad (3.134)$$

Nas equações (3.130) a (3.134), todos os termos necessários nos seus cálculos podem ser obtidos da fila  $M/G^C/1+G$ , exceto para as probabilidades  $P_{M/G/C+G}(W_T = 0)$  e  $P_{M/G/C+G}(A)$ , que não possuem solução. Para resolver este problema, Iravani e Balcioglu (2008) usaram uma aproximação substituindo-as pelas probabilidades  $P_{M/M/C+G}(W_T = 0)$  e  $P_{M/M/C+G}(A)$ , cujos cálculos são conhecidos na literatura. Os sistemas de filas utilizados nesta substituição são idênticos exceto na distribuição dos tempos de serviços. Desse modo, a equação (3.132) pode ser reescrita como:

$$P_{M/G/C+G}(W_S = 0) = \frac{P_{M/M/C+G}(W_T = 0)}{1 - P_{M/M/C+G}(A)} \quad (3.135)$$

Substituindo (3.135) e  $P_{M/M/C+G}(W_T = 0)$  nas equações (3.130), (3.131), (3.133) e (3.134), Iravani e Balcioglu (2008) propõem a primeira aproximação, denominada de *aproximações primárias por extensão* (*primay scaling approximation – PSA*), para o modelo de fila M/G/c+G, apresentada pelas equações (3.136 – 3.139):

$$P_{M/G/C+G}(W_S \leq x) \cong \frac{P_{M/M/C+G}(W_T = 0)}{1 - P_{M/M/C+G}(A)} + \left[ 1 - \frac{P_{M/M/C+G}(W_T = 0)}{1 - P_{M/M/C+G}(A)} \right] \frac{P_{M/G^c/1+G}(0 < W_S \leq x)}{1 - \frac{P_0}{1 - P_A}} \quad (3.136)$$

$$P_{M/G/C+G}(W_T \leq x) \cong P_{M/M/C+G}(W_T = 0) + [1 - P_{M/M/C+G}(W_T = 0)] \frac{P_{M/G^c/1+G}(0 < W_T \leq x)}{1 - P_0} \quad (3.137)$$

$$E_{M/G/C+G}[W_S^K] \cong \left[ 1 - \frac{P_{M/M/C+G}(W_T = 0)}{1 - P_{M/M/C+G}(A)} \right] \frac{E_{M/G^c/1+G}[W_S^K]}{1 - \frac{P_0}{1 - P_A}} \quad (3.138)$$

$$E_{M/G/C+G}[W_T^K] \cong [1 - P_{M/M/C+G}(W_T = 0)] \frac{E_{M/G^c/1+G}[W_T^K]}{1 - P_0} \quad (3.139)$$

Os experimentos numéricos de Iravani e Balcioglu (2008) mostraram que a PSA (equações (3.136) – (3.139)) apresenta desempenho insatisfatório, principalmente com alta variabilidade na distribuição do tempo de serviço. Para melhorar o desempenho, esses autores argumentam que o paralelismo entre os modelos de fila M/M<sup>c</sup>/1+G e M/M/c+G sugere o possível paralelismo entre o modelo de fila analiticamente tratável M/G<sup>c</sup>/1+G, e o modelo de fila M/G/c+G e propõem as aproximações:

$$\frac{P_{M/G/c+G}(W_S > x)}{P_{M/G^c/1+G}(W_S > x)} \cong \frac{P_{M/M/c+G}(W_S > x)}{P_{M/M^c/1+G}(W_S > x)} \quad (3.140)$$

$$\frac{P_{M/G/c+G}(W_T > x)}{P_{M/G^c/1+G}(W_T > x)} \cong \frac{P_{M/M/c+G}(W_T > x)}{P_{M/M^c/1+G}(W_T > x)} \quad (3.141)$$

Usando essas últimas expressões e a equação (3.132), deduzem:

$$P_{M/G/c+G}(W_S = 0) \cong 1 - \left(1 - \frac{P_0}{1 - P_A}\right) \frac{1 - P_{M/M/c+G}(W_S = 0)}{1 - P_{M/M^c/1+G}(W_S = 0)} \quad (3.142)$$

e

$$P_{M/G/c+G}(W_T = 0) \cong 1 - (1 - P_0) \frac{1 - P_{M/M/c+G}(W_T = 0)}{1 - P_{M/M^c/1+G}(W_T = 0)} \quad (3.143)$$

Substituindo-se (3.142) e (3.143) nas equações (3.130), (3.131), (3.133) e (3.134), Iravani e Balcioglu (2008) propõem uma segunda aproximação para o sistema M/G/c+G, que denominam de aproximações modificadas por extensão (*modified scaling approximation - MSA*), apresentadas pelas equações (3.144 -3.147).

$$\begin{aligned} P_{M/G/c+G}(W_S \leq x) \cong & 1 - \left(1 - \frac{P_0}{1 - P_A}\right) \frac{1 - P_{M/M/c+G}(W_S = 0)}{1 - P_{M/M^c/1+G}(W_S = 0)} + \\ & + \left(1 - \frac{P_0}{1 - P_A}\right) \frac{1 - P_{M/M/c+G}(W_S = 0)}{1 - P_{M/M^c/1+G}(W_S = 0)} \frac{P_{M/G^c/1+G}(0 < W_S \leq x)}{1 - \frac{P_0}{1 - P_A}} \end{aligned} \quad (3.144)$$

$$\begin{aligned} P_{M/G/c+G}(W_T \leq x) \cong & 1 - (1 - P_0) \frac{1 - P_{M/M/c+G}(W_T = 0)}{1 - P_{M/M^c/1+G}(W_T = 0)} + \\ & + (1 - P_0) \frac{1 - P_{M/M/c+G}(W_T = 0)}{1 - P_{M/M^c/1+G}(W_T = 0)} \frac{P_{M/G^c/1+G}(0 < W_T \leq x)}{1 - P_0} \end{aligned} \quad (3.145)$$

$$E_{M/G/c+G}[W_S^K] \cong \left(1 - \frac{P_0}{1 - P_A}\right) \frac{1 - P_{M/M/c+G}(W_S = 0)}{1 - P_{M/M^c/1+G}(W_S = 0)} \frac{E_{M/G^c/1+G}[W_S^K]}{1 - \frac{P_0}{1 - P_A}} \quad (3.146)$$

$$E_{M/G/C+G}[W_T^K] \cong (1-P_0) \frac{1-P_{M/M/C+G}(W_T=0)}{1-P_{M/M^C/1+G}(W_T=0)} \frac{E_{M/G^C/1+G}[W_T^K]}{1-P_0} \quad (3.147)$$

Experimentos numéricos realizadas por Iravani e Balcioglu (2008) indicam que as aproximações MSA produzem melhor desempenho do que a aproximação anterior quando a congestão do sistema de fila aumenta, resultando em uma alta taxa de abandono e uma baixa probabilidade de um usuário encontrar servidores disponíveis.

### 3.6 Resumo de modelos para Call Centers

O sistema de fila sem abandono M/M/c, revisado na Secção 3.1, é o mais utilizado para gerenciar os *Call Centers* e os Centros de Contato, motivado pela sua simplicidade matemática, permitindo aos gestores manusear facilmente as suas equações e obter resultados numéricos que indicam o desempenho da central de atendimento. Com base nesses resultados, as decisões sobre a qualidade e eficiência são tomadas. Entretanto, os *Call Centers* possuem características peculiares que o sistema M/M/c não considera nas suas formulações, como o fenômeno do abandono da fila de espera do cliente sem ter sido atendido e os diversos tipos de chamadas, exigindo que os agentes sejam especialistas e recebam treinamento adequado. Essas características podem deixar inadequado a configuração e o uso desse sistema no gerenciamento dos *Call Centers*, porque o modelo pode fornecer informações distorcidas sobre o desempenho do sistema, além de recomendar sub ou superdimensionar o tamanho da equipe de agentes, atingindo um nível de serviço operacional não condizente com a realidade.

Ao se incorporar o abandono da fila de espera no sistema M/M/c, obtém-se o sistema M/M/c+M, descrito na Secção 3.3, que se ajusta mais à realidade das centrais de atendimento, embora considere na modelagem da paciência dos clientes a distribuição exponencial de probabilidade. Esse sistema ainda pode não ser suficiente para captar com suficiente precisão os diversos problemas que ocorrem no abandono, como a interferência externa provocada pelo anúncio do tempo de espera e as variações do tempo de paciência individual de cada chamada. Desempenhos melhores são obtidos ajustando o tempo de abandono com uma distribuição não

exponencial de probabilidade, como nos sistemas  $M/M/c+G$ , descrito na Secção 3.4, que adotam distribuição geral de probabilidade para incorporar o tempo de paciência dos clientes.

Outro aspecto a ser considerado é o fato do tempo de atendimento em alguns *Call Centers* não se ajustar perfeitamente ao modelo exponencial, levando ao sistema mais geral  $M/G/c+G$ , analisado na secção 3.5, com distribuição geral de probabilidade também para os tempos de serviço, que pode oferecer melhor desempenho e aproximar-se mais da realidade das centrais de atendimento.

A partir dessa revisão da literatura sobre o uso de modelos de filas em *Call Center*, observa-se que existem poucas aplicações que validam os resultados teóricos de filas por meio de uma análise científica baseada em dados, principalmente quando a paciência e o abandono dos usuários são considerados. Até onde temos conhecimento, pouco esforço foi realizado na literatura para verificar se um modelo existente, utilizando distribuições genéricas e particularmente mistas para representar o abandono, é adequado para uma aplicação em particular. Percebe-se uma grande preocupação com o desenvolvimento teórico de modelos de filas para *Call Center* e o uso de simulação para a comprovação da adequação desses modelos.

Um resumo contendo as diferenças entre os modelos de filas abordados nesta revisão da literatura é apresentado no Quadro 3.1. No desenvolvimento deste estudo utilizou-se o modelo de fila  $M/G/c+G$  para representar o caso do *Call Center* do banco. A decisão por este modelo de fila justifica-se pelo melhor ajuste de distribuições não exponencial para representarem os tempos de serviço e os tempos de paciência (abandono), obtidas por meio da análise estatística aplicada aos dados do objeto de estudo considerado.

### **3.7 Verificação dos algoritmos computacionais**

Os algoritmos computacionais dos modelos matemáticos de filas descritos neste Capítulo 3 para representarem os *Call Centers*, precisam ser verificados quanto às suas implementações computacionais e os resultados obtidos das medidas de desempenho por eles calculadas. Isso requer que eles sejam verificados e comparados entre si. As implementações computacionais foram realizadas no *software* Mathematica (2013) e executadas no *ultrabook* Samsung com processador Intel® Core™ I7, de 2,4 GHz e 8,00 GB de RAM.

Desse modo, o modelo de fila  $M/M/c+G$  com distribuição exponencial para a paciência ( $G \sim$ exponencial) foi aqui verificado, comparando os seus resultados com o modelo correspondente de fila  $M/M/c+M$ , que considera distribuição exponencial para a paciência com fórmulas fechadas e que já foi bastante explorado na literatura. Os resultados da Tabela 3.1 indicam que o algoritmo computacional do modelo  $M/M/c+G$  aqui implementado está consistente para este caso, por apresentar resultados idênticos ao modelo  $M/M/c+M$ , como seria esperado.

Outra verificação é que o modelo de fila  $M/M/c+G$  com distribuição *Fatigue Life* para a paciência ( $G \sim$ *Fatigue Life*), com um tempo médio de paciência relativamente alto (100) e com variância relativamente baixa (0,036), deve se comportar aproximadamente como um modelo sem abandono  $M/M/c$ , porque desta maneira teremos usuários muito pacientes que não abandonarão o sistema. A coincidência dos resultados da Tabela 3.2 indicam que o algoritmo computacional aqui implementado do modelo  $M/M/c+G$ , com distribuição *Fatigue Life*, para modelar o comportamento do abandono dos usuários, tem resultados consistentes com o modelo  $M/M/c$ , como seria esperado. Nesta verificação, por ilustração, utilizou-se para parâmetros de entrada (taxas de chegada, serviço e número de canais), no modelo  $M/M/c$ , valores extraídos de Jain (1991, p.532).

Considerando novamente o modelo de fila  $M/Mc+G$  com distribuição *Fatigue Life* para a paciência ( $G \sim$ *Fatigue Life*) mas, com tempo médio de paciência relativamente baixo (0,005) e com variância relativamente baixa (0,0001), caracterizando usuários altamente impacientes, este modelo deve ter um comportamento similar ao modelo  $M/M/c/c$  da Seção 3.2. Isso se justifica devido a pouca paciência dos usuários, que não ingressariam na fila sabendo que deveriam aguardar para serem atendidos. Os dados da Tabela 3.3 também indicam que a implementação do algoritmo computacional para o modelo  $M/M/c+G$  com distribuição *Fatigue Life* para o abandono é consistente com o modelo  $M/M/c/c$  para esse caso, como seria esperado.

Os modelos analíticos de fila com distribuição de probabilidade genérica para os tempos de serviço e para os tempos de paciência foram comparados com o modelo analítico de fila  $M/M/c+M$ , que considera a distribuição Exponencial de probabilidade para os tempos de serviço e para os tempos de paciência. A Tabela 3.4 mostra os resultados da comparação do modelo analítico de fila  $M/M/c+M$  com o modelo analítico de fila  $M/G^c/1+G$ , com distribuição de probabilidade Exponencial para representar os tempos de serviço e para representar os

tempos de paciência. Os resultados indicam que o modelo analítico de fila  $M/G^c/1+G$  é consistente com o modelo analítico de fila  $M/M/c+M$  nas medidas de desempenho tempo médio de espera e probabilidade de esperar, apresentando diferenças significativas nas medidas de desempenho probabilidade de abandonar e intensidade de tráfego.

Ainda comparando os modelos analíticos de fila com distribuição de probabilidade genérica para os tempos de serviço e para os tempos de paciência, o modelo analítico de fila  $M/G/c+G$ , com distribuição de probabilidade Exponencial para representar os tempos de serviço e para os tempos de paciência foi comparado com o modelo analítico de fila  $M/M/c+M$ . Os resultados da Tabela 3.5 indicam que o modelo analítico de fila  $M/G/c+G$  é consistente com o modelo analítico de fila  $M/M/c+M$  nas medidas de desempenho probabilidade de abandonar e intensidade de tráfego. Em relação às medidas de desempenho tempo médio de espera e probabilidade de esperar, o modelo analítico de fila  $M/G/c+G$  revela diferenças significativas, principalmente com a medida tempo médio de espera.

Observa-se das comparações realizadas que os modelos analíticos de fila  $M/G^c/1+G$  e  $M/G/c+G$ , com distribuições genéricas para os tempos de serviço e também para os tempos de paciência, foram os únicos que mostraram inconsistência em algumas medidas de desempenho. Uma justificativa para as diferenças apresentadas é que esses modelos analíticos de fila com distribuições genéricas são modelos aproximados, e não apresentam fórmulas fechadas para o cálculo das suas medidas de desempenho.

TABELA 3.1 - Verificação do modelo M/M/c+G~Exponencial com o modelo M/M/c + M

Especificações	M/M/c+G~Exponencial <=> M/M/c+M		Desvio absoluto
	M/M/c+G~Exponencial	M/M/c+M	
Taxa chegada	0.16667	0.16667	0
Taxa serviço	0.05	0.05	0
Taxa abandono	0,02	0,02	0
Número canais	5	5	0
Paciência média = 1/θ	-----	-----	-----
Variância paciência	-----	-----	-----
Tempo médio espera	1,93506	1,93517	0,00011
Probabilidade esperar	0,274554	0,274567	0,000013
Prob. abandonar	0,0387011	0,03870	0,0000011
Intensidade tráfego	0,640867	0,640877	0,00001

Fonte: elaborado pelo autor

TABELA 3.2 - Verificação do modelo M/M/c+G~*Fatigue Life* (alta paciência média e baixa variância)

Especificações	M/M/c+G~ <i>Fatigue Life</i> <=> M/M/c		Desvio absoluto
	M/M/c+G~ <i>Fatigue Life</i>	M/M/c	
Taxa chegada	0.16667	0.16667	0
Taxa serviço	0.05	0.05	0
Tx. abandono = 1/Pac.média	0,01	-----	-----
Número canais	5	5	0
Pac. média = 1/Taxa aband.	100 (alta)	-----	-----
Variância paciência	0,036 (baixa)	-----	-----
Tempo médio espera	3,91368	3,92042	0,00674
Probabilidade esperar	0,326633	0,32668 (prob. formar fila)	0,000047
Prob. abandonar	0,00004162	-----	-----
Intensidade tráfego	0,666652	0,66668	0,000028

Fonte: elaborado pelo autor



TABELA 3.3 - Verificação do modelo M/M/c+G~*Fatigue Life* (baixa paciência média e baixa variância)

Especificações	M/M/c+G~ <i>Fatigue Life</i> <=> M/M/c/c		Desvio absoluto
	M/M/c+G~ <i>Fatigue Life</i>	M/M/c/c	
Taxa chegada	0.16667	0.16667	0
Taxa serviço	0.05	0.05	0
Tx. abandono = 1/Pac.média	200	-----	-----
Número canais	5	5	0
Pac. média = 1/Taxa aband.	0,005 (baixa)	-----	-----
Variância paciência	0,0001 (baixa)	-----	-----
Tempo médio espera	0,000002159	0 (não forma fila)	0,000002159
Probabilidade esperar	0,00005214	0 (prob. formar fila)	0,00005214
Prob. abandonar	0,139212	0,139212 (prob. de perda)	0
Intensidade tráfego	0,57387	0,57387	0

Fonte: elaborado pelo autor

TABELA 3.4 - Verificação do modelo M/G<sup>c</sup>/1+G~Exponencial com o modelo M/M/c + M

Especificações	M/G <sup>c</sup> /1+G~Exponencial <=> M/M/c+M		Desvio absoluto
	M/G <sup>c</sup> /1+G~Exponenci	M/M/c+M	
Taxa chegada	0,16667	0,16667	0
Taxa serviço	0,05x5=0,25	0.05	0,20
Taxa abandono	0,02	0,02	0
Número canais	5	5	0
Paciência média = 1/θ	-----	-----	-----
Variância paciência	-----	-----	-----
Tempo médio espera	1,92480	1,935170	0,01037
Probabilidade esperar	0,27583	0,274567	0,00126
Prob. abandonar	0,15616	0,038700	0,11746
Intensidade tráfego	0,52528	0,640877	0,11560

Fonte: elaborado pelo autor

TABELA 3.5 - Verificação do modelo M/G/c+G~Exponencial com o modelo M/M/c + M

Especificações	M/G/c+G~Exponencial <=> M/M/c+M		Desvio absoluto
	M/G/c+G~Exponencial	M/M/c+M	
Taxa chegada	0,16667	0,16667	0
Taxa serviço	0,05	0,05	0
Taxa abandono	0,02	0,02	0
Número canais	5	5	0
Paciência média = 1/θ	-----	-----	-----
Variância paciência	-----	-----	-----
Tempo médio espera	4,7943	1,935170	2,85913
Probabilidade esperar	0,7468	0,274567	0,47223
Prob. abandonar	0,0387	0,038700	0,00000
Intensidade tráfego	0,5626	0,640877	0,07828

Fonte: elaborado pelo autor

QUADRO 3.1 - Principais modelos de filas					
Características	M/M/c Erlang-C	M/M/c/K Erlang-B	M/M/c + M Erlang-A	M/M/c + G	M/G/c + G
Processo de chegada	Processo de Poisson, com tempos entre chegadas distribuídos exponencialmente	Processo de Poisson, com tempos entre chegadas distribuídos exponencialmente	Processo de Poisson, com tempos entre chegadas distribuídos exponencialmente	Processo de Poisson, com tempos entre chegadas distribuídos exponencialmente	Processo de Poisson, com tempos entre chegadas distribuídos exponencialmente
Processo de serviço	Tempos de serviços distribuídos exponencialmente	Tempos de serviços distribuídos exponencialmente	Tempos de serviços distribuídos exponencialmente	Tempos de serviços distribuídos exponencialmente	Tempos de serviços possuem distribuição geral
Canais de serviço	Possui $c$ canais idênticos, operando em paralelo e independentes uns dos outros	Possui $c$ canais idênticos, operando em paralelo e independentes uns dos outros	Possui $c$ canais idênticos, operando em paralelo e independentes uns dos outros	Possui $c$ canais idênticos, operando em paralelo e independentes uns dos outros	Possui $c$ canais idênticos, operando em paralelo e independentes uns dos outros
Capacidade do sistema	Ilimitada	Limitada em $K$ usuários	Ilimitada	Ilimitada	Ilimitada
População	Homogênea com infinitos usuários	Homogênea com infinitos usuários	Homogênea com infinitos usuários	Homogênea com infinitos usuários	Homogênea com infinitos usuários
Disciplina atendimento	FIFO	FIFO	FIFO	FIFO	FIFO
Abandono	Não considera	Não considera	Considera o abandono. O usuário ao chegar no sistema encontra um tempo de espera virtual (oferecido)	Considera o abandono. O usuário ao chegar no sistema encontra um tempo de espera virtual (oferecido)	Considera o abandono
Paciência dos usuários	Infinita	Infinita	Tempo de paciência do usuário é distribuído exponencialmente	Tempo de paciência do usuário possui distribuição geral	Tempo de paciência do usuário possui distribuição geral
Bloqueio dos usuários	Não considera	Os usuários são impedidos de entrar no sistema quando a fila atinge seu limite máximo	Não considera	Não considera	Não considera
Referências	GROSS e HARRIS, 1985 ARENALES et al, 2007	GROSS e HARRIS, 1985 ARENALES et al, 2007 DOMBACHER, 2010	PALM, 1957 MANDELBAUM e ZELTYN, 2005	BACCELLI e HEBUTERNE, 1981 ZELTYN, 2004	IRAVANI e BALCIOGLU, 2008



## 4 O CASO DO *CALL CENTER* DO BANCO

Neste capítulo, inicialmente, é feita uma descrição da estrutura de funcionamento do caso *Call Center* do Banco, apresentado na secção 1.3. Apresenta-se uma descrição e análise dos dados coletados, do comportamento estatístico dos processos de chegada, serviço e abandono e, uma explanação e verificação do modelo simulado deste *Call Center*.

Ainda reproduz-se os resultados de um estudo realizado por Mandelbaum et al. (2001) e Brown et al. (2002), com o modelo  $M/M/c+M$ , tendo como fonte de dados um pequeno *Call Center* de um Banco de Israel, denominado por eles de Banco Anônimo. Todos os dados deste *Call Center* foram gentilmente disponibilizados pelos autores. O interesse aqui é absorver o conhecimento existente e aplicar outros modelos de fila revisados no Capítulo 3, comparando os resultados aqui obtidos com aqueles reportados na literatura, para analisar a precisão dos modelos.

A abordagem adotada no estudo conduzido por Mandelbaum et al. (2001) e Brown et al. (2002) considerou o modelo de fila Erlang-A ( $M/M/c+M$ ), para representar o *Call Center* do Banco, procurando os parâmetros do modelo que melhor se ajustam aos dados reais. Esses autores justificam o uso do modelo Erlang-A por ter um tratamento matemático amigável, embora os tempos de serviço e os tempos de paciência (abandono) dos usuários possam ser ajustados por outras distribuições de probabilidade diferentes da distribuição exponencial, como comprovado pela análise empírica dos dados.

Observa-se que a abordagem conduzida nesta pesquisa considera vários modelos de fila para representar o *Call Center* do Banco. Nesses modelos, são inseridos os parâmetros (por exemplo, taxas de chegada, serviço e abandono) extraídos dos dados reais e as medidas de desempenho geradas pelos modelos são comparadas com os dados reais. Uma questão interessante de pesquisa é qual o modelo que produz medidas de desempenho com menor desvio em relação aos dados reais observados, para representar o *Call Center* do banco.

### 4.1 Descrição do *Call Center* do banco

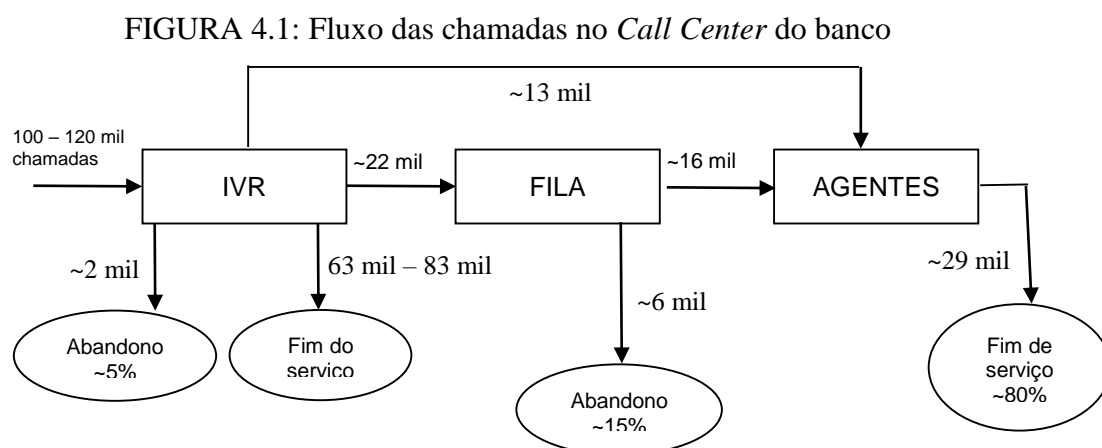
O *Call Center* do banco oferece vários tipos de serviços listados a seguir, com suas respectivas codificações:

- i) informações para clientes atuais (PS) e futuros (NW);

- ii) transações de cheques e contas de poupança (PS);
- iii) negociação de ações (NE);
- iv) suporte técnico para os usuários de internet do site do banco (IN).

O *Call Center* possuía 8 posições regulares para agentes, 5 posições para os agentes atenderem as solicitações relacionadas à internet e uma posição para o supervisor de turnos. Trabalhava em dias úteis de domingo a quinta-feira das 7 horas até meia-noite, fechando às 14 horas de sexta-feira e reabrindo às 8 horas no sábado. O fluxo das chamadas iniciava-se com um usuário discando para um dos números de telefones disponíveis, dependendo do tipo de serviço solicitado, raramente encontrando um sinal de ocupado. Conectava-se à unidade de resposta por voz (VRU – *Voice Responce Unit*), algumas vezes também chamada de resposta interativa por voz (IVR – *Interactive Voice Responce*), fornecendo dados para sua identificação.

Existe um total de 100 a 120 mil chamadas mensais no banco de dados. Dessas chamadas, aproximadamente 65 mil (65%) são resolvidas pela VRU e finalizam o serviço. O restante das chamadas, aproximadamente 35 mil (35%) desejam dialogar com um dos agentes. Nesse momento, três possibilidades podem ocorrer: os usuários (aproximadamente 13 mil) encontram um agente livre e inicia o seu serviço; outros usuários (aproximadamente 22 mil) juntam-se à fila de espera e desses últimos, aproximadamente 6 mil, abandonam o sistema. Os dados analisados pelos autores consistem das chamadas que desejam falar com um agente e correspondem às chamadas mensais de janeiro a dezembro do ano de 1999. Essas chamadas são originadas pelo usuário (*incoming calls*). Aquelas originadas pelo *Call Center* (*outgoing calls*), regularmente feitas para os usuários com alta prioridade que abandonam o sistema, foram desprezadas nas análises. Os fluxos dessas chamadas estão sumarizados na Figura 4.1, cujos valores indicam as chamadas mensais.



Fonte: Mandelbaum, Sakov e Zeltyn (2001)

Os usuários que necessitam aguardar na fila para serem servidos juntam-se no final da fila, no instante da sua chegada, e aqueles que possuem alta prioridade avançam na fila em 1,5 minuto e todos são atendidos pela disciplina FIFO. Enquanto aguardam na fila, os usuários recebem informações sobre a sua localização na fila e a quantidade esperada de tempo que levará para serem atendidos. O anúncio é repetido a cada 60 segundos, intercalado com música, notícias ou comerciais. A seguir, descreve-se as variáveis consideradas neste estudo, prosseguindo-se com as suas análises empíricas e estatísticas.

## 4.2 Análise estatística do Call Center do banco

A fonte de dados utilizada neste estudo é a mesma utilizada por Mandelbaum et al. (2001) e Brown, et al. (2002). Os dados coletados foram organizados em uma planilha, em que cada linha corresponde ao registro de cada chamada, contendo a sua identificação, as informações dos usuários e as informações sobre o fluxo das chamadas. Uma amostra desses dados encontra-se no Anexo A. Uma descrição detalhada das informações contidas em cada registro é dada no tópico seguinte, com a finalidade de oferecer uma melhor compreensão nas análises realizadas.

### 4.2.1 Descrição dos dados

A descrição contida neste item refere-se ao conteúdo das variáveis extraídas dos dados do *Call Center* do Banco, gentilmente disponibilizados por Mandelbaum et al. (2001). A intenção é inicialmente reproduzir os resultados com o modelo analítico de fila  $M/M/c+M$ , e em seguida também gerar novos resultados com outros modelos analíticos. Esses dados correspondem a todas as chamadas atendidas pelo *Call Center* do banco no período dos 12 meses de janeiro a dezembro de 1999. Existem entre 20 mil a 30 mil chamadas por mês e cada registro dessas chamadas contém 17 campos (variáveis), que são descritos em detalhes.

1. VRU+LINE (6 dígitos): cada chamada que entra no sistema é roteada por meio da VRU. Existem 6 classificações na VRU de AA01 a AA06. Cada VRU possui muitas linhas

identificadas de 1 a 16. Há no total 65 linhas. A cada chamada é atribuída um número para a VRU e um número para a linha. Cada registro dos dados coletados começa com AA01 e, por isso, foi omitido no referido campo.

2. CALL\_ID (5 dígitos): a cada chamada que chega é atribuído um identificador ID da chamada. Os ID's não são necessariamente consecutivos devido a terem sido atribuídos por diferentes VRU's. Todos os ID's dos dados coletados começam com 44, portanto, foram omitidos.

3. CUSTOMER\_ID (0 a 12 dígitos): número de identificação do usuário que o identifica unicamente. O ID é zero se o usuário não é identificado pelo sistema, como no caso dos clientes potenciais. Devido a um erro no sistema, o ID não foi registrado para usuários que não aguardaram na fila, isto é, abandonaram da VRU, ou foi atendido pelo agente diretamente da VRU.

4. PRIORITY (1 dígito): a prioridade de um usuário é feita de um arquivo de fora da fila. Há dois tipos de usuários: com (alta) prioridade e regular. A prioridade 0 e 1 indica usuários não identificados ou usuários regulares e 2 é para usuários com prioridade. Os usuários são servidos na ordem do seu “tempo de fila”; entretanto, os usuários com prioridade são avançados na fila por um minuto e meio. Os usuários não são informados sobre a existência de prioridades. Os usuários que não aguardam na fila, e cujo ID é registrado como 0 (zero) tem o valor 0 (zero) registrado como sua prioridade, mesmo se esta não é sua verdadeira prioridade.

5. TYPE (2 dígitos): tipo de serviço solicitado pelo usuário (pela chamada). Existem 6 tipos diferentes de pedidos: atividade regular (código PS-Peilut Shotefet), atividade regular em inglês (código PE-èilut English), suporte técnico de internet (código IN-Internet), mercado de ações (código NE-Niarot Erech), e usuários potenciais recebendo informações (código NW). Todas essas chamadas são “*inbound*” (solicitadas pelos usuários). Os usuários podem deixar seu número de telefone para ser chamado pelo *Call Center*. Usuários com alta prioridade que abandonam são automaticamente chamados pelo *Call Center*. Estas são as chamadas “*outbound*” (feitas pelo Banco) e codificadas por TT. As discagem das chamadas TT são realizadas pelo agente ou pelo computador. O *Call Center* possui muitos números de telefones, cada um deles está associado com um tipo diferente de serviço. O sistema registra o tipo de serviço de acordo com o número discado. Quando um usuário chama o número “PS”, a atividade é registrada como um “PS”, mesmo se alguma outra atividade é realizada, por



exemplo, mercado de ações. Enquanto o conteúdo exato do serviço não é conhecido, manteve-se a chamada para o primeiro tipo, considerando que a mudança para outro tipo é rara.

6. DATE (6 dígitos): data da chamada no formato ano-mês-dia. Na amostra dos dados foi omitido o ano (1999) e o mês (09) de todas as chamadas. Uma entrada típica é 990901.

7. VRU\_ENTRY (6 dígitos): tempo em que a chamada entra (chega) no *Call Center*. Mais especificamente é o tempo (instante) que a chamada entra (chega) na VRU e recebe um ID do usuário. Todos os tempos são registrados no formato hh:mm:ss (horas, minutos, segundos).

8. VRU\_EXIT (6 dígitos): tempo (instante) de saída da VRU, isto é, se vai para a fila, ou diretamente para ser servido, ou deixa o sistema (abandona a VRU).

9. VRU\_TIME (1 a 3 dígitos): tempo (em segundos) gasto na VRU, calculado como a diferença entre os dois campos VRU\_EXIT e VRU\_ENTRY.

10. Q\_START (6 dígitos): tempo em que se junta à fila (colocado em espera). Esta entrada é 00:00:00, para os usuários que não atingiram a fila (abandonaram a VRU).

11. Q\_EXIT (6 dígitos): tempo (instante) de saída da fila por ter recebido serviço ou por ter abandonado o sistema.

12. Q\_TIME (1 a 3 dígitos): tempo gasto na fila, calculado como a diferença entre os dois últimos campos Q\_EXIT e Q\_START.

13. OUTCOME (4, 5 ou 7 dígitos): existem 3 possíveis resultados para cada chamada AGENT quando o usuário recebe serviço; HANG quando o usuário desliga (abandona) e PHANTOM é, de acordo com a equipe do *Call Center*, o abandono da VRU. Entretanto os dados não apoiam isso, porque mais que 95% das chamadas PHANTOM tem tempo de serviço positivo. É uma chamada virtual ignorada e também por ter poucos casos desses.

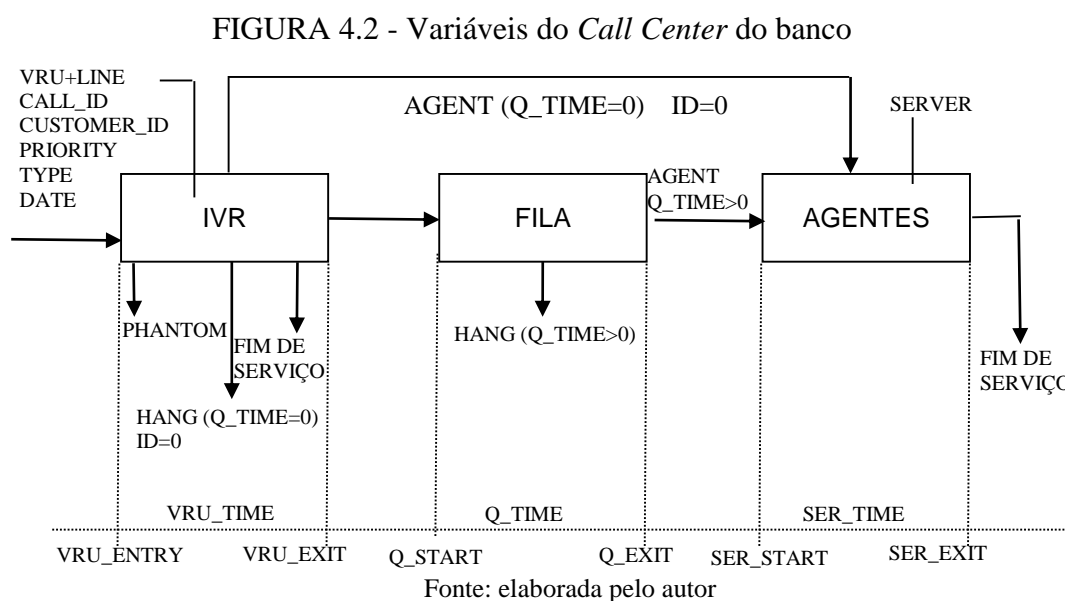
14. SER\_START (6 dígitos): tempo (instante) do início de serviço pelo agente.

15. SER\_EXIT (6 dígitos): tempo (instante) do final do serviço pelo agente.

16. SER\_TIME (1 a 3 dígitos): duração do serviço, calculado como a diferença entre os dois últimos campos SER\_EXIT e SER\_START, em segundos.

17. SERVER (texto): o nome do agente que serviu o usuário. Se nenhum serviço foi realizado, este campo é NO\_SERVER.

O diagrama da Figura 4.2 ilustra o posicionamento dessas variáveis.



#### 4.2.2 Análise dos dados

Alguns cálculos básicos sobre as chamadas do *Call Center* do Banco, segmentadas por diferentes variáveis (por exemplo: chamadas mensais, resultados por espera e outras), foram realizados por Mandelbaum et al. (2001). A intenção foi inicialmente quantificar as ligações que ocorreram em cada mês e também mensurar as chamadas do tipo *HANG*, *AGENT* e *PHANTOM*. Depois verificar possíveis inconsistências nos dados. Esses cálculos daquela pesquisa foram resumidamente reproduzidos nesta pesquisa e são apresentados no Apêndice A. Utilizou-se aqui o software MS-Excel ® para efetuar os cálculos. Conforme mencionado, os dados referentes a todas as chamadas de todos os meses do ano de 1999 foram gentilmente disponibilizados por Mandelbaum et al. (2001).

Nessas análises, utilizou-se as frequências absolutas e relativas (em porcentagens) das quantidades de chamadas ocorridas em cada mês, estratificadas com outras variáveis. Também usou-se o cálculo de médias, medianas e desvios padrão os tempos de espera na fila, de serviço e de abandono para inferir sobre o aspecto da distribuição desses tempos. Alguns resultados merecem destaque, como o número médio de 37.037 chamadas mensais recebidas pelo *Call Center* do banco. Observou-se que a maioria das chamadas (79,5%) dialoga com os atendentes, contra 19,5% delas que abandona o sistema. Uma quantidade pequena das chamadas (0,8%) abandona o sistema diretamente da IVR.

Com relação ao processo de espera, verificou-se que 59,2% das chamadas tiveram que aguardar uma certa quantidade positiva de tempo na fila para serem atendidas pelos atendentes e o restante (40,8%) encontrou o sistema vazio e iniciou o seu serviço imediatamente, após ter encerrado as suas atividades na IVR. Destaca-se também que 14,31% abandonaram a fila depois de esperarem uma quantidade positiva de tempo. Observando-se as estatísticas do tempo de espera de todas as chamadas que aguardam na fila (servidas ou abandonadas), percebe-se que a sua distribuição é assimétrica para a direita, porque possuem média maior que a mediana (Apêndice A). Analisando-se separadamente os tempos de espera das chamadas que abandonaram o sistema e o tempo de espera daquelas que foram servidas, nota-se que a distribuição desses dois tempos de espera também é assimétrica para a direita, pelo mesmo motivo. A mesma assimetria também aparece nos tempos de serviço.

Com a finalidade de mostrar a significância do abandono, por meio de exemplos numéricos, Mandelbaum et al. (2001) efetuaram uma comparação dos modelos de filas com abandono Erlang-A ( $M/M/c+M$ ) e o modelo sem abandono Erlang-C ( $M/M/c$ ) e notaram que quando a intensidade de tráfego é alta, existe uma diferença significativa entre o tempo médio de espera em ambos os modelos. Observaram que o modelo Erlang-A ( $M/M/c+M$ ), que incorpora o abandono, possui tempos de espera sempre inferiores ao Erlang-C ( $M/M/c$ ).

Desse modo, ao considerar-se o abandono dos usuários, reduz-se o tempo de espera daqueles que estão na fila e, também, dos que entrarão no sistema futuramente. Vale ressaltar que se o abandono não for considerado, pode-se chegar a decisões incorretas na gestão de um *Call Center*, como por exemplo, superdimensionar a equipe de atendentes. Todos esses resultados estão disponibilizados no Apêndice B.

### 4.3 Análise dos processos de chegada, abandono e serviço

Fez-se uma análise para verificar o comportamento da distribuição dos tempos de chegada, abandono e serviço dos dados obtidos, nos três intervalos de hora considerados (11-12 horas, 12-13 horas e 15-16 horas). Os resultados dessa análise são importantes para definir o modelo de fila que será utilizado para representar o *Call Center* do Banco.

#### 4.3.1 O processo de chegada

O processo de chegada registra os instantes em que as chamadas telefônicas chegam na fila de espera, depois de terem visitado a VRU. Ao deixarem a VRU, os usuários movem-se para a fila e se acontece de um atendente estar desocupado, o serviço começa imediatamente e seu tempo de permanência na fila é zero. Por outro lado, se não houver atendentes disponíveis, o usuário aguarda uma quantidade positiva de tempo na fila, até iniciar o seu serviço, ou perde a paciência e abandona o *Call Center*.

Este processo pode ser descrito em dois diferentes níveis de detalhes, considerando-se a quantidade de chamadas por unidade de tempo, ou por meio do tempo entre as chegadas consecutivas das chamadas telefônicas na fila. Este estudo toma por base o tempo entre as chegadas das chamadas, que é avaliado por meio das informações da estatística descritiva e também por sua variabilidade estocástica. Esta seção analisa os tempos entre as chegadas telefônicas do *Call Center* do Banco citado na seção 4.2.2. Os mesmos três intervalos de tempos analisados anteriormente foram aqui considerados nesta análise.

Em cada um dos intervalos de hora (11-12 horas, 12-13 horas, 15-16 horas), considerou-se os instantes de chegada na fila, registrados na variável  $Q\_Start$ , das chamadas telefônicas que buscavam por serviços do tipo NE, NW e PS, que iniciaram serviço ou que abandonaram o sistema ( $Q\_time > 0$ ). Incluiu-se também os instantes de chegada na fila das chamadas telefônicas que imediatamente iniciaram serviço, porque encontraram, no momento da sua chegada, pelo menos um atendente desocupado ( $Q\_time = 0$ ). Desprezou-se as chamadas que abandonaram na VRU ( $Customer\_ID < 0$ ).

Em seguida, selecionou-se as informações somente do dia 09 de novembro por ser um dia típico do *Call Center* (MANDELBAUM et al., 2001) e estimou-se os tempos entre as chegadas das chamadas telefônicas. Após essa etapa, procedeu-se com a análise estatística dos tempos entre as chegadas das chamadas telefônicas, calculando-se a sua média, mediana e

desvio padrão. Ajustou-se, também, uma distribuição teórica de probabilidade para esses tempos entre chegadas que caracterize o padrão das chegadas dos usuários ao *Call Center*. As estatísticas dos tempos entre chegadas das chamadas telefônicas estão sumarizadas na Tabela 4.1. Observa-se que a mediana é significativamente menor que a média (Mediana < Média) em todos os intervalos de horas, indicando que a distribuição dos tempos entre chegadas é assimétrica à direita, sugerindo a aderência da distribuição Exponencial de probabilidade aos dados reais.

TABELA 4.1 – Tempo entre chegadas (min)

Dados amostra	11 – 12 h	12 – 13 h	15 – 16 h
Média	0,518	0,577	0,476
Mediana	0,483	0,400	0,350
Desvio padrão	0,384	0,495	0,409
Coef. Variação	0,741	0,858	0,859
Taxa chegada (ch/min)	1,931	1,733	2,101

Fonte: elaborada pelo autor

Outras análises foram realizadas por Mandelbaum et al. (2001), utilizando o histograma, razão entre a mediana e a média e, também, o uso do gráfico P-P plot. Os resultados dessas análises estão no Apêndice C.

O teste Kolmogorov-Smirnov, aplicado neste estudo, também confirma a aproximação com a distribuição exponencial com *p-value* 0,08012; 0,66578 e 0,69484 para os intervalos de tempos analisados das 11 às 12 horas, das 12 às 13 horas e das 15 às 16 horas, respectivamente. Nesta análise estatística dos tempos entre chegadas das chamadas telefônicas, utilizou-se o *software EasyFit*®, versão 5.5 (2010), que ordena as distribuições mais aderentes.

#### 4.3.2 O processo de abandono

Este processo ocorre quando os usuários ingressam na fila de espera, aguardam uma quantidade positiva de tempo, perdem a paciência e abandonam o sistema. Essa atitude dos usuários sinaliza, para os gestores do *Call Center*, se vale a pena esperar pelo serviço prestado. Esse nível de serviço oferecido pelo sistema pode ser avaliado pela quantidade de abandonos que ocorrem em um determinado intervalo de tempo. Um aumento da quantidade de abandonos pode apontar um alto tempo de espera e, como consequência, um baixo nível de serviço. Assim sendo, o comportamento do abandono e o tempo de espera estão profundamente relacionados,

pois, todos os usuários que abandonam a fila, esperaram. Além disso, o tempo necessário para um cliente ser servido não é observado se ele abandona a fila. Então, esse tempo necessário para ser servido e o tempo de paciência são baseados em dados censurados.

Para melhor analisar esses tempos, Brown *et al.* (2002) faz três distinções. A primeira é a diferença entre o tempo de fila e o tempo de espera, desprezando a espera zero. Considera o tempo de espera (isto é, somente com esperas positivas) como sendo mais relevante para os gestores do *Call Center*, principalmente quando analisado juntamente com a fração dos que esperam. A segunda, a distinção entre os tempos de espera dos usuários que foram servidos no sistema e daqueles que abandonaram o sistema. Brown *et al.* (2002) afirmam que o tempo apenas daqueles que abandonaram (tempo de paciência) não descreve a paciência de todos os usuários. A terceira, distinção é entre o tempo que um usuário necessita esperar antes de ser atendido por um agente (tempo virtual de espera) e o tempo que um usuário está disposto a esperar antes de abandonar o sistema (paciência).

O tempo virtual de espera pode ser entendido como a quantidade de tempo que um usuário (virtual) equipado com infinita paciência tem que aguardar até ser atendido, enquanto que o tempo de paciência refere-se à medida operacional da paciência (ou impaciência) do usuário. Entretanto, nenhuma dessas medidas são diretamente observáveis e devem ser estimadas.

A caracterização da paciência e do tempo do tempo virtual de espera está baseada em dados censurados de acordo com Mandelbaum *et al.* (2001). Consideram que para os usuários que são pacientes o suficiente para esperar um agente, os seus tempos de espera são uma amostra do tempo necessário para aguardar (tempo virtual de espera). Neste caso, afirmam que a paciência dos usuários (tempo disposto para esperar) é censurado pelo tempo de espera para o serviço (tempo virtual de espera), em que apenas o tempo de espera para o serviço é observável. Por outro lado, para os usuários que abandonam, o tempo de paciência (tempo disposto para esperar) censura o tempo necessário para esperar (tempo virtual de espera) e, neste caso, o tempo de paciência é observável.

Para estimar esses tempos que são censurados e, portanto, não observáveis como a paciência dos usuários, utiliza-se o estimador Kaplan-Meier que é uma ferramenta da Análise de Sobrevivência, um ramo da estatística que se preocupa em analisar dados censurados (KALBFLEISCH & PRENTICE; 1980). Esta seção analisa as experiências que os usuários tiveram na fila até abandonar o sistema do *Call Center* do Banco analisado na seção 4.2.2. Os mesmos três intervalos de tempos, analisados anteriormente, foram aqui considerados nesta análise.

Procedeu-se à análise estatística dos tempos de espera até abandonar (tempo de paciência) das chamadas telefônicas, calculando-se a sua média, mediana e desvio padrão. Ajustou-se, também, uma distribuição teórica de probabilidade para esses tempos de paciência, para caracterizar o padrão da paciência dos usuários ao *Call Center*. Estimou-se o tempo de paciência dos usuários do *Call Center* por meio do estimador Kaplan-Meier para dados censurados.

Na análise estatística de cada um dos três intervalos de hora (11-12 horas, 12-13 horas, 15-16 horas), considerou-se os tempos de espera até abandonar (tempos de paciência), registrados na variável *Q\_Time* ( $Q\_Time > 0$ ), das chamadas telefônicas que buscavam por serviços do tipo NE, NW e PS, e que abandonaram o sistema. O abandono foi identificado pelo registro *HANG* na variável *Outcome*. Desprezou-se as chamadas que abandonaram na VRU ( $Customer\_ID \diamond 0$ ). Com os valores da variável *Q\_Time*, procedeu-se à análise estatística dos tempos de espera até abandonar (paciência) as chamadas telefônicas, calculando-se a sua média, mediana e desvio padrão. A Tabela 4.2 sumariza essas estatísticas.

TABELA 4.2 – Tempo de espera para abandono (paciência) (min)

Dados amostra	11 – 12 h	12 – 13 h	15 – 16 h
Média (seg)	65,859	68,390	67,661
Mediana (seg)	43,000	45,500	42,500
Desvio padrão (seg)	85,431	77,565	78,586
Coef. Variação	1,297	1,134	1,161
Média (min)	1,098	1,139	1,128
Taxa abandono(ch/min)	0,911	0,877	0,887

Fonte: elaborado pelo autor

Outras análises foram realizadas por Mandelbaum et al. (2001), utilizando o histograma, razão entre a mediana e a média e, também, o uso do gráfico P-P plot. Os resultados dessas análises estão no Apêndice C.

O *p-value* para o teste Kolmogorov-Smirnov (K-S) para a exponencialidade foi 0,01538; 0,02587 e 0,000019 para os intervalos de tempos analisados das 11 às 12 horas, das 12 às 13 horas e das 15 às 16 horas, respectivamente. Esses resultados indicam a aproximação com a distribuição exponencial dos tempos de paciência, nos dois primeiros horários, com  $\alpha = 1\%$  e  $\alpha = 2\%$ , respectivamente, e rejeita a exponencialidade para o terceiro horário. Além da exponencial, foi verificado o ajuste de outras distribuições. A distribuição *Fatigue Life* (ver Apêndice F) foi a que teve melhor ajuste com os tempos de paciência, fornecendo os mais altos *p-value* para o teste Kolmogorov-Smirnov 0,23208; 0,11255 e 0,00725, para os mesmos três intervalos de tempo, respectivamente. A distribuição Lognormal também não é rejeitada em

relação aos tempos de paciência, nos dois primeiros horários, com *p-value* 0,04495 e 0,04254, respectivamente, e é rejeitada no terceiro horário, com *p-value* 0,0000. Pelo teste Kolmogorov-Smirnov, todas as distribuições foram rejeitadas no horário das 15-16 horas. Diante das possibilidades oferecidas pelo teste K-S, optou-se pelas distribuições Exponencial, Fatigue Life e Lognormal para modelar os tempos de paciência dos usuários, nos três horários, mesmo considerando que todas as distribuições foram rejeitadas no horário das 15-16 horas. Nesta análise estatística dos tempos de paciência das chamadas telefônicas, utilizou-se o *software EasyFit*®, versão 5.5 (2010).

Como o interesse é estimar a paciência dos usuários do *Call Center*, ou seja, quanto tempo um usuário está disposto de aguardar antes de abandonar, então deve-se considerar o tempo necessário para um usuário atingir um agente como uma observação censurada. De fato, se ele atingiu o agente, é porque o tempo que desejava esperar era maior.

Para estimar a paciência dos usuários em cada um dos intervalos de hora, considerou-se os tempos de espera na fila registrados na variável *Q\_Time* ( $Q\_Time > 0$ ), das chamadas telefônicas que buscavam por serviços do tipo NE, NW e PS, que abandonaram o sistema ou foram atendidos por um agente. Essas chamadas foram identificadas pelo registro *HANG* ou *AGENT* na variável *Outcome*. Desprezou-se as chamadas que abandonaram na VRU ( $Customer\_ID \leq 0$ ). Considerando-se as chamadas que atingiram o agente como observações censuradas na variável *Q\_Time*, procedeu-se a estimação da paciência dos usuários, usando-se o estimador Kaplan-Meier, por meio do *software SPSS*®.

A Figura 4.3 mostra as estimativas Kaplan-Meier da função de sobrevivência da paciência dos usuários (tempo que ele está disposto a aguardar antes de abandonar) dos três intervalos de tempo analisados. A Tabela 4.3 reporta as médias e os desvios padrões da estimativa de Kaplan-Meier para a distribuição da paciência dos usuários em cada intervalo de hora. Esses resultados foram obtidos pelos procedimentos convencionais do *software SPSS*® ao estimar a função de sobrevivência.

TABELA 4.3 – Tempo de paciência estimado pelo Kaplan-Meier

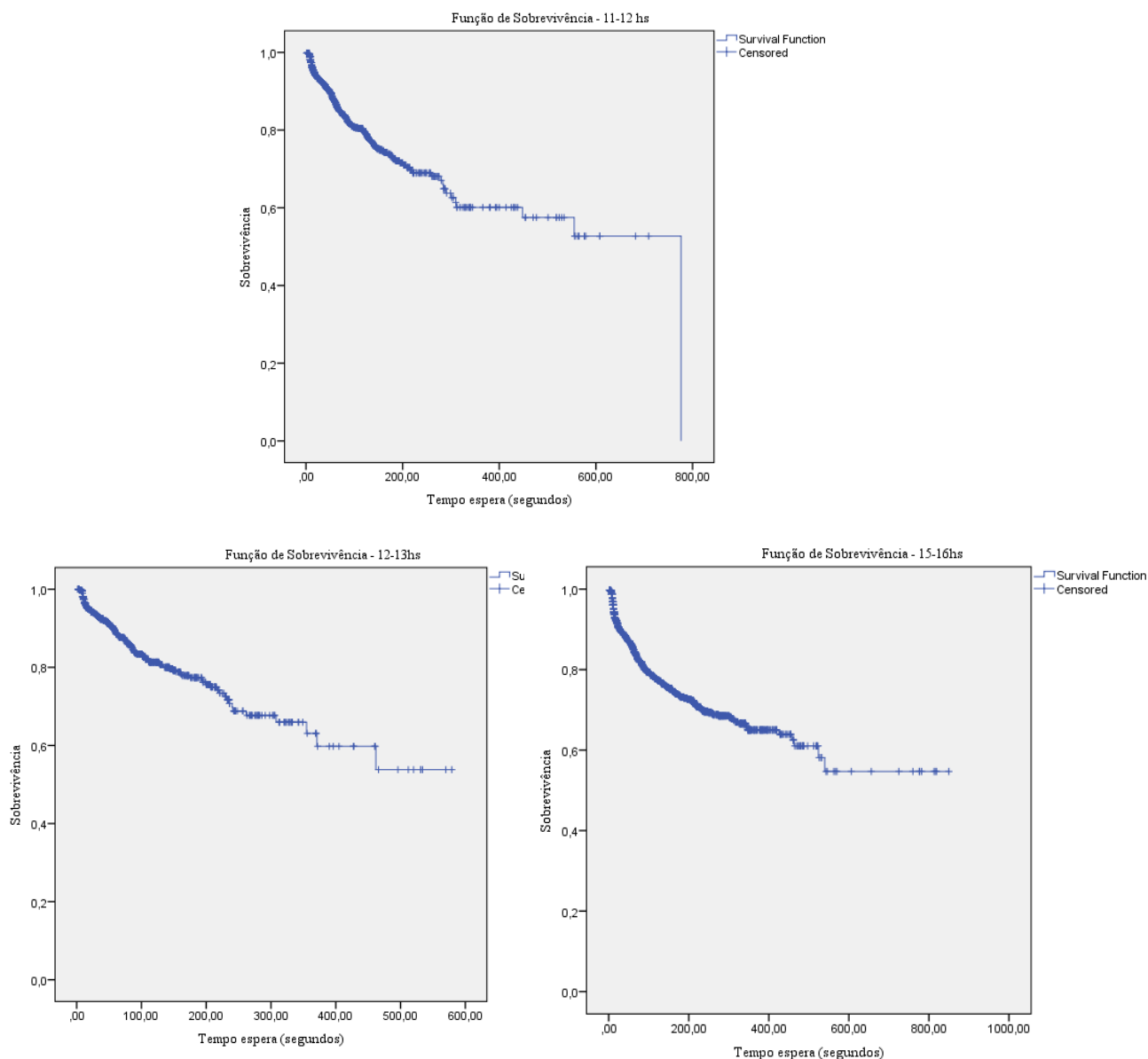
Medidas	11 – 12 h	12 – 13 h	15 – 16 h
Média (seg)	500,851	404,520	555,350
Desvio padrão (seg)	22,343	16,308	19,084
Média (min)	8,348	6,742	9,256
Taxa abandono (ch/min)	0,119	0,148	0,108

Fonte: elaborada pelo autor

As médias da Tabela 4.3 indicam que os usuários do horário das 15 às 16 horas são mais pacientes (suportam aguardar mais) que aqueles dos outros horários.



FIGURA 4.3: Curvas de Sobrevivência



Fonte: elaborada pelo autor

A razão disso é que, nesse horário (das 15 às 16 horas), os usuários têm mais urgência dos serviços que necessitam e maior confiança de que o sistema irá fornecê-lo. É o horário que ocorre o fechamento do mercado de ações (BROWN *et al.*, 2002). Observa-se, ainda, da Tabela 4.5, que a paciência média e as correspondentes taxas de abandono em cada intervalo de hora estão próximos dos valores ajustados por Brown *et al.* (2002), ao utilizarem para paciência média 540 segundos. A última paciência média foi inserida nas equações do modelo  $M/M/c+M$  e obtiveram medidas de desempenho próximas com os dados reais observados.

### 4.3.3 O processo de serviço

O atendimento pelo agente é a última etapa da visita de um usuário ao *Call Center*. Por isso, o conhecimento desse atendimento revela informações sobre o nível de serviço que está sendo prestado. O serviço pode ser avaliado quantitativamente, por meio de informações estatísticas sobre a duração do serviço, ou qualitativamente, quando o usuário revela sua satisfação sobre o serviço recebido, respondendo a um breve questionário antes de deixar o sistema.

A preocupação nesta seção é com a avaliação quantitativa do serviço prestado pelo *Call Center* do Banco citado na seção 4.2.2. Considerando-se os mesmos três intervalos de tempo analisados anteriormente, fez-se, inicialmente, uma análise estatística da duração dos tempos de serviço (tempo de atendimento), calculando-se a sua média, mediana e desvio padrão. Após essa etapa, analisou-se o ajuste de uma distribuição teórica de probabilidades aos tempos de serviço, para definir o modelo teórico de fila que melhor represente o *Call Center*. A tabela 4.4 sumariza as estatísticas dos tempos de serviço. Em cada um dos três intervalos de hora analisados, considerou-se, como nas análises anteriores, os tipos de serviço NE, NW e PS e ignorou-se os registros de tempos de serviço igual a zero.

TABELA 4.4 – Tempo de serviço (min)

Dados amostra	11 – 12 h	12 – 13 h	15 – 16 h
Média	3,184	3,021	3,479
Mediana	2,050	2,050	2,391
Desvio padrão	4,225	3,339	3,545
Coef. Variação	1,327	1,105	1,019
Taxa chegada (ch/min)	0,314	0,331	0,2887

Fonte: elaborada pelo autor

Outras análises foram realizadas por Mandelbaum et al. (2001), utilizando o histograma, razão entre a mediana e a média e, também, o uso do gráfico P-P plot. Os resultados dessas análises estão no Apêndice C e sugerem a aderência da distribuição Lognormal para os tempos de serviço.

O teste Kolmogorov-Smirnov confirma esse ajuste da Lognormal com *p-value* 0,01646; 0,30402 e 0,44411 para os intervalos de tempos analisados das 11 às 12 horas, das 12 às 13 horas e das 15 às 16 horas, respectivamente. Nesta análise estatística dos tempos de serviços, utilizou-se o *software EasyFit*®, versão 5.5 (2010).

Entretanto, a distribuição Fréchet (Apêndice G) se ajusta melhor aos dados dos tempos de serviço, conforme o teste K-S, com *p-value* de 0,87816 e 0,69548, para os dois primeiros

horários, sendo rejeita para os dados do terceiro horário. Por isso, optou-se para representar os tempos de serviço a distribuição Fréchet para os horários das 11-12 horas e 12-13 horas e a distribuição Lognormal para o horário das 15-16 horas.

Com o propósito de fornecer uma visão geral das estatísticas descritivas dos processos de chegada, serviço e abandono, em cada horário analisado, um sumário delas é exibido na Tabela 4.5.

TABELA 4.5 - Sumário das estatísticas descritivas - *Call Center* Banco

Estatísticas		11 – 12 h	12 – 13 h	15 – 16 h
Processo chegada	Média	0,518	0,577	0,476
	Mediana	0,483	0,400	0,350
	Desvio padrão	0,384	0,495	0,409
	Coef. Variação	0,741	0,858	0,859
	Taxa chegada (ch/min)	1,931	1,733	2,101
Processo serviço	Média	3,184	3,021	3,479
	Mediana	2,050	2,050	2,391
	Desvio padrão	4,225	3,339	3,545
	Coef. Variação	1,327	1,105	1,019
	Taxa chegada (ch/min)	0,314	0,331	0,2887
Processo abandono	Média (min)	8,348	6,742	9,256
	Mediana	-	-	-
	Desvio padrão (seg)	22,343	16,308	19,084
	Coef. Variação	-	-	-
	Taxa chegada (ch/min)	0,119	0,148	0,108

Fonte: elaborada pelo autor

O Quadro 4.1 resume as distribuições de probabilidade que mais se aderem aos dados dos processos de chegada, serviço e abandono, do *Call Center* do Banco, em cada um dos intervalos analisados (11-12 h, 12-13 h e 15-16 h). Essas distribuições foram utilizadas nas análises deste trabalho.

QUADRO 4.1 - Sumário das distribuições de probabilidades - *Call Center* Banco

Processos	Distribuição ajustada	<i>p-value</i>	<i>p-value</i>	<i>p-value</i>
		11 – 12 h	12 – 13 h	15 – 16 h
Chegada	Exponencial	0,08012	0,66578	0,69484
Serviço	Lognormal	0,01646	0,30402	0,45920
	Fréchet	0,87816	0,69548	0,0000
Abandono	Fatigue Life	0,23208	0,11255	0,00725
	Lognormal	0,04495	0,04254	0,0000
	Exponencial	0,01538	0,02587	0,0000

Fonte: elaborado pelo autor

#### 4.4 Aplicação dos modelos M/M/c+M e M/M/c+G

O modelo analítico de fila M/M/c+M, com distribuição Exponencial para representar os tempos de paciência, foi amplamente utilizado por Mandelbaum et al. (2001) e Brown et al. (2002) nas análises do caso *Call Center* do Banco. Porém, poderiam ter usado o modelo analítico de fila M/M/c+G, que admite distribuição genérica para representar o abandono, por exemplo utilizar a distribuição Fatigue Life (G~Fatigue Life), para tentar obter melhores resultados, o que vai ser investigado nesta secção.

Para atingir esse objetivo, utilizou-se os dados reais dos três intervalos de hora (11-12 horas, 12-13 horas e 15-16 horas) coletados no *Call Center* do Banco. Utilizou-se também os resultados já obtidos pelo teste de aderência *Kolmogorov-Smirnov*, com o *software EasyFit*®, versão 5.5 (2010), para as distribuições de probabilidade da paciência dos usuários. Incluiu-se essas distribuições nos modelos teóricos de fila, dando origem a várias possibilidades diferentes de modelos para representar o *Call Center* do Banco.

Nesses vários modelos de filas aplicados, inseriu-se as taxas e distribuições de chegada, serviço e abandono dos diversos horários e obteve-se, pelos seus respectivos algoritmos computacionais programados no *software Mathematica* (Wolfram, 2013), as seguintes medidas de desempenho: tempo médio de espera, probabilidade de esperar, probabilidade de abandonar e intensidade de tráfego. Essas medidas de desempenho calculadas pelos modelos de filas foram comparadas, neste capítulo, com as mesmas medidas obtidas diretamente dos dados reais, descritas na análise estatística dos processos de chegada, abandono e de serviço.

Com a finalidade de facilitar a compreensão das análises, a partir dessa secção, os modelos de filas foram representados com uma notação específica. Identificou-se as distribuições genéricas no processo de serviço e no processo de abandono com um índice

numérico para se distinguir uma das outras. Quando a distribuição é do tipo mista (Apêndice H), acrescentou-se a letra “*m*” junto ao índice numérico para representá-la. Assim,  $G_1$  identifica uma única distribuição paramétrica ( $G_1 \sim \text{LogN}(1,2)$ ) uma distribuição Lognormal com parâmetros 1 e 2) e  $G_{1m}$  representa uma distribuição mista ( $G_{1m} \sim 0,5\text{LogN}(1,2) + 0,5\text{LogN}(3,4)$ ) uma distribuição mista com duas componentes Lognormal e peso  $p = 0,5$ ).

Considerando o intervalo das 11 às 12 horas, os parâmetros de entrada extraídos dos dados reais foram: taxa de chegada igual a 1,933 chamadas por minuto, taxa de serviço de 0,314 chamadas por minuto, taxa de abandono de 0,119 chamadas por minuto, com 6 agentes servindo os usuários. Esses parâmetros foram inseridos nos vários modelos de fila considerados. O modelo  $M/M_1/c+G_2$  com distribuição *Fatigue Life* para a paciência apresentou as medidas de desempenho tempo médio de espera, probabilidade de esperar e intensidade de tráfego mais próximas das medidas de desempenho dos dados reais, mas teve a medida de desempenho probabilidade de abandonar mais afastada desses dados reais, quando comparada com a mesma medida do modelo  $M/M_1/c+M_4$ , com distribuição Exponencial para a paciência, conforme resultados das Tabelas 4.6, 4.7, 4.8 e 4.9.

TABELA 4.6 - Comparação de modelos pelo tempo médio de espera – 11-12horas

Modelos de fila	Tempo médio de espera		
	Valores	Diferença absoluta	Diferença relativa
Caso real	1,043 min		
$M/M_1/c+M_4$	1,1203 min	0,0773	7,41%
$M/M_1/c+G_2$	1,1143 min	0,0713	6,84%

Distribuições: M ~ Exponencial (1,933)  $M_1$  ~ Exponencial (0,314)  
 $M_4$  ~ Exponencial (0,119)  $G_2$  ~ *Fatigue Life* (0,7095; 6,6691)

Fonte: elaborada pelo autor

TABELA 4.7 - Comparação de modelos pela probabilidade de esperar – 11-12horas

Modelos de fila	Probabilidade de Esperar		
	Valores	Diferença absoluta	Diferença relativa
Caso real	71,9%		
M/M <sub>1</sub> /c+M <sub>4</sub>	68,8%	3,1	4,31%
M/M <sub>1</sub> /c+G <sub>2</sub>	70,5%	1,4	1,95%
Distribuições: M ~ Exponencial (1,933) M <sub>1</sub> ~ Exponencial (0,314)			
M <sub>4</sub> ~ Exponencial (0,119) G <sub>2</sub> ~ Fatigue Life (0,7095; 6,6691)			
Fonte: elaborada pelo autor			

TABELA 4.8 - Comparação de modelos pela probabilidade de abandonar – 11-12horas

Modelos de fila	Probabilidade de Abandonar		
	Valores	Diferença absoluta	Diferença relativa
Caso real	13,9%		
M/M <sub>1</sub> /c+M <sub>4</sub>	13,3%	0,6	4,32%
M/M <sub>1</sub> /c+G <sub>2</sub>	12,7%	1,2	8,63%
Distribuições: M ~ Exponencial (1,933) M <sub>1</sub> ~ Exponencial (0,314)			
M <sub>4</sub> ~ Exponencial (0,119) G <sub>2</sub> ~ Fatigue Life (0,7095; 6,6691)			
Fonte: elaborada pelo autor			

TABELA 4.9: Comparação de modelos pela intensidade de tráfego – 11-12horas

Modelos de fila	Intensidade de Tráfego		
	Valores	Diferença absoluta	Diferença relativa
Caso real	90,6%		
M/M <sub>1</sub> /c+M <sub>4</sub>	88,9%	1,7	1,87%
M/M <sub>1</sub> /c+G <sub>2</sub>	89,5%	1,1	1,24%
Distribuições: M ~ Exponencial (1,933) M <sub>1</sub> ~ Exponencial (0,314)			
M <sub>4</sub> ~ Exponencial (0,119) G <sub>2</sub> ~ Fatigue Life (0,7095; 6,6691)			
Fonte: elaborada pelo autor			

No intervalo de hora das 12 às 13 horas, as taxas de chegada, serviço e abandono, em chamadas por minuto, foram, respectivamente: 1,717; 0,328; 0,148. O número de agentes para servir os usuários é, também, igual a 6. Procedeu-se de modo semelhante ao intervalo de hora anterior e novamente o modelo  $M/M_2/c+G_{10}$ , com distribuição *Fatigue Life* para a paciência, apresentou menor desvio em relação aos dados reais, para as medidas de desempenho tempo médio de espera, probabilidade de esperar e intensidade de tráfego. Para a medida de desempenho probabilidade de abandonar, o desvio foi maior em relação a esses dados reais, quando comparada com a mesma medida do modelo  $M/M_2/c+M_5$ , com distribuição exponencial para a paciência, conforme resultados apresentados nas Tabelas 4.10, 4.11, 4.12 e 4.13.

TABELA 4.10 - Comparação de modelos pelo tempo médio de espera – 12-13 horas

Modelos de fila	Tempo médio de espera		
	Valores	Diferença absoluta	Diferença relativa
Caso real	0,813 min		
$M/M_2/c+M_5$	0,569 min	0,24	29,52%
$M/M_2/c+G_{10}$	0,761 min	0,052	6,39%
Distribuições: M ~ Exponencial (1,717) M <sub>2</sub> ~ Exponencial (0,328)			
M <sub>5</sub> ~ Exponencial (0,148) G <sub>10</sub> ~ Fatigue Life (0,5069; 5,9746)			

Fonte: elaborada pelo autor

TABELA 4.11 - Comparação de modelos pela probabilidade de esperar – 12-13 horas

Modelos de fila	Probabilidade de esperar		
	Valores	Diferença absoluta	Diferença relativa
Caso real	58,3%		
$M/M_2/c+M_5$	48,8%	9,5	16,29%
$M/M_2/c+G_{10}$	53,9%	4,4	7,55%
Distribuições: M ~ Exponencial (1,717) M <sub>2</sub> ~ Exponencial (0,328)			
M <sub>5</sub> ~ Exponencial (0,148) G <sub>10</sub> ~ Fatigue Life (0,5069; 5,9746)			

Fonte: elaborada pelo autor

TABELA 4.12 - Comparação de modelos pela probabilidade de abandonar – 12-13 horas

Modelos de fila	Probabilidade de abandonar		
	Valores	Diferença absoluta	Diferença relativa
Caso real	10,2%		
M/M <sub>2</sub> /c+M <sub>5</sub>	8,4%	1,8	17,65%
M/M <sub>2</sub> /c+G <sub>10</sub>	6,11%	4,09	66,93%
Distribuições: M ~ Exponencial (1,717) M <sub>2</sub> ~ Exponencial (0,328)			
M <sub>5</sub> ~ Exponencial (0,148) G <sub>10</sub> ~ Fatigue Life (0,5069; 5,9746)			
Fonte: elaborada pelo autor			

TABELA 4.13 - Comparação de modelos pela intensidade de tráfego – 12-13 horas

Modelos de fila	Intensidade de tráfego		
	Valores	Diferença absoluta	Diferença relativa
Caso real	82,8%		
M/M <sub>2</sub> /c+M <sub>5</sub>	79,9%	2,9	3,50%
M/M <sub>2</sub> /c+G <sub>10</sub>	81,9%	0,9	1,10%
Distribuições: M ~ Exponencial (1,717) M <sub>2</sub> ~ Exponencial (0,328)			
M <sub>5</sub> ~ Exponencial (0,148) G <sub>10</sub> ~ Fatigue Life (0,5069; 5,9746)			
Fonte: elaborada pelo autor			

Considerando-se o último intervalo de hora das 15 às 16 horas, as taxas de chegada, serviço e abandono, em chamadas por minuto, foram, respectivamente: 2,1; 0,291; 0,108. O número de agentes para servir os usuários é, também, igual a 6. Procedeu-se de modo semelhante aos intervalos anteriores e os modelos M/M<sub>3</sub>/c+G<sub>17</sub> com distribuição Fatigue Life para a paciência e M/M<sub>3</sub>/c+M<sub>6</sub>, com distribuição Exponencial para a paciência, apresentaram o mesmo desvio em relação aos dados reais para as medidas de desempenho probabilidade de esperar, probabilidade de abandonar e intensidade de tráfego. Para a medida de desempenho tempo médio de espera o modelo M/M<sub>3</sub>/c+M<sub>6</sub>, com distribuição Exponencial para a paciência apresentou menor desvio em relação aos dados reais, conforme resultados apresentados nas Tabelas 4.14, 4.15, 4.16 e 4.17. Uma possível explicação para esse comportamento das duas distribuições é o aumento da intensidade de tráfego nesse último horário, motivado pela proximidade do fechamento das negociações com ações.



TABELA 4.14 - Comparação de modelos pelo tempo médio de espera – 15-16 horas

Modelos de fila	Tempo médio de espera		
	Valores	Diferença absoluta	Diferença relativa
Caso real	1,98 min		
M/M <sub>3</sub> /c+M <sub>6</sub>	1,93 min	0,05	2,53%
M/M <sub>3</sub> /c+G <sub>17</sub>	1,65 min	0,33	16,67%
Distribuições: M ~ Exponencial (2,1) M <sub>3</sub> ~ Exponencial (0,291)			
M <sub>6</sub> ~ Exponencial (0,108) G <sub>17</sub> ~ Fatigue Life (0,7484; 7,2309)			
Fonte: elaborada pelo autor			

TABELA 4.15 - Comparação de modelos pela probabilidade de esperar – 15-16 horas

Modelos de fila	Probabilidade de esperar		
	Valores	Diferença absoluta	Diferença relativa
Caso real	90,3% min		
M/M <sub>3</sub> /c+M <sub>6</sub>	85,1%	5,2	5,76%
M/M <sub>3</sub> /c+G <sub>17</sub>	84,9%	5,4	5,98%
Distribuições: M ~ Exponencial (2,1) M <sub>3</sub> ~ Exponencial (0,291)			
M <sub>6</sub> ~ Exponencial (0,108) G <sub>17</sub> ~ Fatigue Life (0,7484; 7,2309)			
Fonte: elaborada pelo autor			

TABELA 4.16 - Comparação de modelos pela probabilidade de abandonar – 15-16 horas

Modelos de fila	Probabilidade de abandonar		
	Valores	Diferença absoluta	Diferença relativa
Caso real	23,4%		
M/M <sub>3</sub> /c+M <sub>6</sub>	20,9%	2,5	10,68%
M/M <sub>3</sub> /c+G <sub>17</sub>	20,9%	2,5	10,68%
Distribuições: M ~ Exponencial (2,1) M <sub>3</sub> ~ Exponencial (0,291)			
M <sub>6</sub> ~ Exponencial (0,108) G <sub>17</sub> ~ Fatigue Life (0,7484; 7,2309)			
Fonte: elaborada pelo autor			

TABELA 4.17 - Comparação de modelos pela intensidade de tráfego – 15-16 horas

Modelos de fila	Intensidade de tráfego		
	Valores	Diferença absoluta	Diferença relativa
Caso real	95,9%		
M/M <sub>3</sub> /c+M <sub>6</sub>	95,2%	0,7	0,73%
M/M <sub>3</sub> /c+G <sub>17</sub>	95,1%	0,8	0,83%

Distribuições: M ~ Exponencial (2,1) M<sub>3</sub> ~ Exponencial (0,291)  
M<sub>6</sub> ~ Exponencial (0,108) G<sub>17</sub> ~ Fatigue Life (0,7484; 7,2309)

Fonte: elaborada pelo autor

Os resultados obtidos das análises desses três horários (11-12 horas, 12-13 horas e 15-16 horas), sugerem que o modelo do tipo M/M/c+G, com distribuição Fatigue Life para a paciência, em geral, tem suas medidas de congestão (tempo médio de espera, probabilidade de esperar e intensidade de tráfego) mais próximas das medidas dos dados reais, enquanto que o modelo do tipo M/M/c+M, com distribuição Exponencial para a paciência, em geral, possui mais proximidade com os dados reais nas medidas de abandono (probabilidade de abandonar), para esses dados analisados. Isso sugere uma questão para ser investigada: uma distribuição mista para representar o abandono poderia ser melhor que ambas, Fatigue Life e Exponencial? Essa questão é retomada adiante.

Esperava-se que o modelo analítico de fila do tipo M/M/c+G, com distribuição de probabilidade Fatigue Life para representar a paciência fosse melhor que o modelo analítico de fila do tipo M/M/c+M, que possui distribuição de probabilidade Exponencial para representar a paciência nos três horários, principalmente no terceiro horário (15-16 horas), onde a distribuição de probabilidade Exponencial foi rejeitada no teste de aderência *Kolmogorov-Smirnov*, e a distribuição de probabilidade Fatigue Life foi aceita no mesmo teste. Porém, não foi o que ocorreu já que o modelo analítico de fila com distribuição Fatigue Life para a paciência foi, em geral, equivalente ao modelo analítico de fila com distribuição Exponencial para a paciência nos três horários, mostrando-se melhor em algumas medidas de desempenho, mas sendo ineficiente em outras.

#### 4.5 Modelo de simulação

No Capítulo 3, realizou-se a verificação dos modelos aproximados de fila M/M/c + G, do M/G<sup>c</sup>/1 + G e do M/G/c + G, comparando esses modelos com os modelos exatos M/M/c e

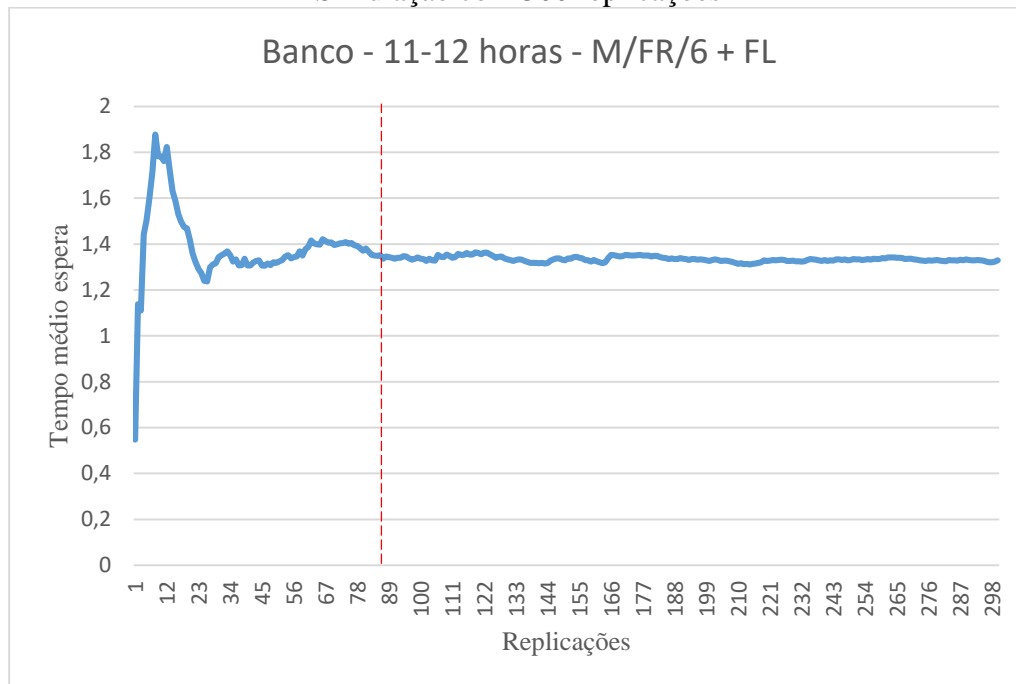
M/M/c + M. Constatou-se dessa comparação que os algoritmos computacionais aqui implementados dos modelos aproximados estão consistentes por apresentar resultados idênticos aos dos algoritmos computacionais dos modelos exatos (ver item 3.7). Desse modo, fica verificada a corretude dessas implementações e que os algoritmos computacionais dos modelos M/M/c + G, do M/G<sup>c</sup>/1 + G e do M/G/c + G podem ser utilizados para avaliar o *Call Center* do Banco, objeto deste estudo. Entretanto, os modelos de fila que utilizam distribuições mistas para representar o processo de abandono não podem ser verificados comparando-os com o modelo M/M/c + M, que não comporta distribuições mistas no processo de abandono. Por esse motivo, a verificação desses modelos de fila que utilizam distribuições mistas no processo de abandono é realizada por meio de simulação discreta, a seguir descrita nos seus detalhes.

A implementação da simulação foi realizada com o *software* Arena ®, por meio do sistema de simulação *Arena Contact Center Edition*, especialmente desenvolvido para os gestores de *Call Centers* construírem modelos de simulação e analisarem os resultados produzidos por esses modelos. Esse sistema utiliza conceitos gráficos para gerar o fluxo das chamadas telefônicas, desde a sua chegada, direcionando-as através do centro até encontrar um atendente. Para criar um modelo de simulação utilizando esse sistema, é necessário descrever a sequência de eventos que ocorrem no fluxo das chamadas, por meio dos módulos: contatos, padrão de chegada, capacidade das linhas tronco, descrição das rotas e agentes.

Os modelos de fila utilizados neste estudo foram simulados com 300 replicações e cada replicação com duração de um dia, gerando no final de todas as rodadas uma média de 18 mil chamadas no *Call Center* do Banco. As estatísticas foram coletadas a cada intervalo de 60 minutos e informadas em um relatório produzido pelo *software*.

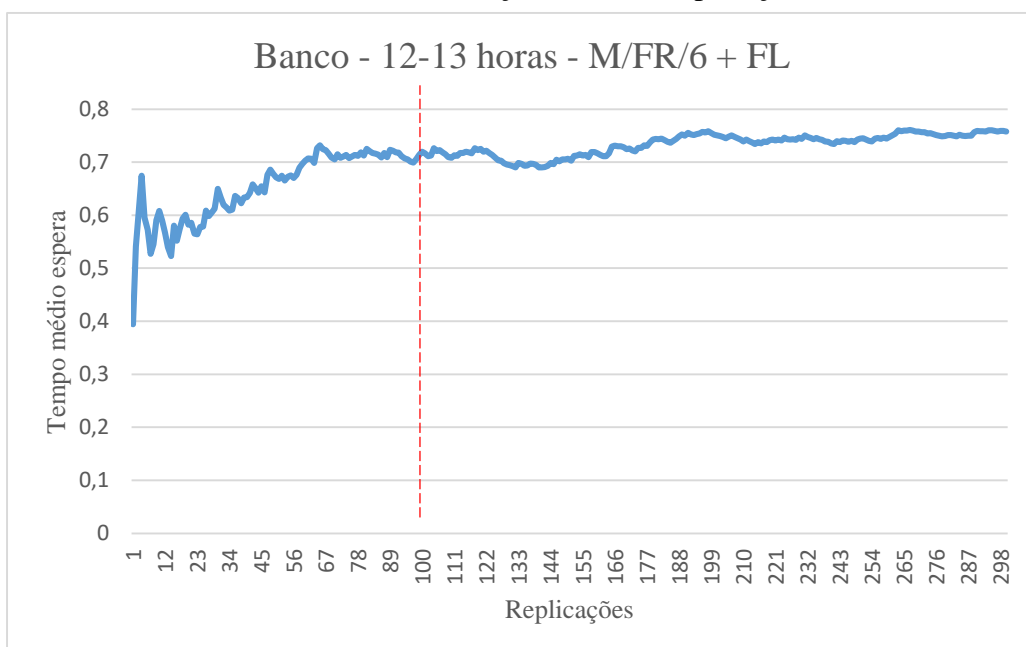
Utilizou-se um período de *warm up* de 144 mil minutos que ocorre na 100<sup>a</sup> replicação. A partir desse período o sistema passa a operar em equilíbrio estatístico, conforme mostram visualmente os Gráficos 4.1, 4.2 e 4.3, como exemplos, e desse momento em diante as estatísticas foram contadas. Todas as simulações foram realizadas em um *ultrabook* Samsung com processador Intel® Core™ I7, de 2,4 GHz e 8,00 GB de RAM e levaram em média 6,00 minutos para estimar as medidas de desempenho em cada modelo de fila.

GRÁFICO 4.1 - *Warm up* – Caso Banco – Modelo de fila M/FR/6 + FL  
Simulação com 300 replicações



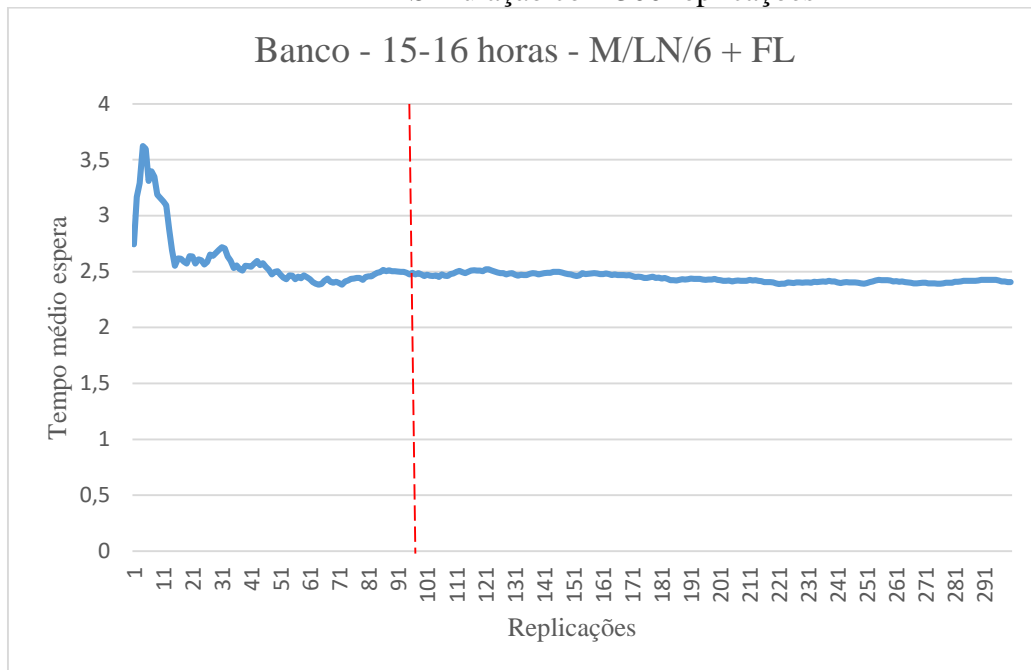
Fonte: elaborado pelo autor

GRÁFICO 4.2 - *Warm up* – Caso Banco – Modelo de fila M/FR/6 + FL  
Simulação com 300 replicações



Fonte: elaborado pelo autor

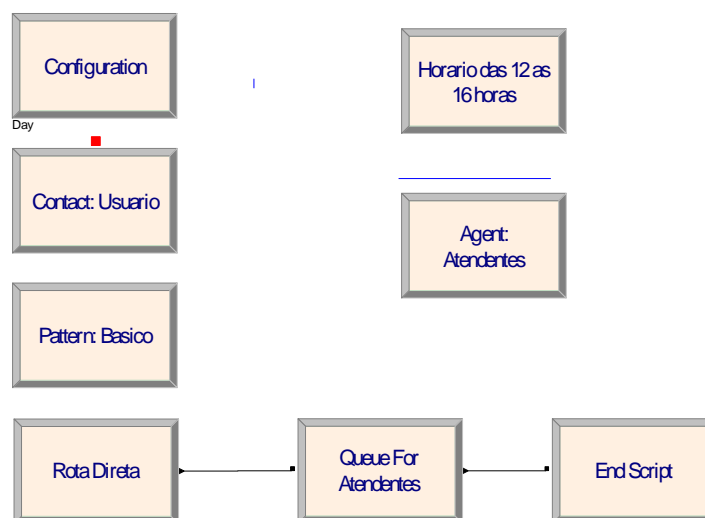
GRÁFICO 4.3 - Warm up – Caso Banco – Modelo de fila M/LN/6 + FL  
Simulação com 300 replicações



Fonte: elaborado pelo autor

A modelagem das simulações foi realizada com a implementação de 8 módulos do software Arena®, conforme Figura 4.5.

FIGURA 4.5: Representação gráfica da simulação do Call Center do Banco e da Empresa, no Arena.



Fonte: elaborado pelo autor

No módulo “*Configuration*” é especificado o esboço da simulação do *Call Center*, definindo-se o horizonte do planejamento (o dia, neste estudo), a quantidade de troncos, o tipo do *Call Center* (*inbound*, neste estudo), o número de replicações e a quantidade máxima de agentes. O módulo “*Schedule*”, redefinido como “*Horário das 12 às 16 horas*”, define como os agentes podem ser designados aos usuários, com base no horizonte de planejamento e nos intervalos de 60 minutos (*timeslot*). Informa-se também se os agentes estão trabalhando ou estão em outra atividade (reuniões, pausas, lanches). O módulo “*Pattern*” contém as informações de como os usuários chegam ao *Call Center*, tendo como base o horizonte de planejamento considerado. As chegadas sempre ocorrem obedecendo a um Processo de Poisson e a taxa de chegada é informada neste módulo. O módulo “*Agent*” define os agentes do *Call Center*, informando se eles fazem parte de um só grupo ou se existem vários grupos de agentes trabalhando em conjunto. Além da quantidade de agentes em cada grupo, outras informações são dadas como o nome, habilidades, custo, entre outras.

No módulo “*Contact*” descreve-se o comportamento do usuário dentro do *Call Center*, incluindo a duração da conversa com o agente (tempo de serviço), a propensão dele abandonar, o seu possível retorno ao sistema. Neste módulo são informados a distribuição e taxa de serviço bem como a distribuição e a taxa de abandono. O módulo “*Begin*” redefinido como “*Rota Direta*” tem como propósito iniciar as chegadas dos usuários no *Call Center*, definindo para cada um deles qual tronco deverá utilizar para fazer sua trajetória no sistema. O módulo “*Queue for Agent*” coloca em contato o usuário com o grupo de agentes, obedecendo a disciplina de atendimento FIFO, ou recoloca-o na fila de espera de acordo com sua prioridade (não considerada neste estudo). O módulo “*End Script*” tem como propósito identificar o final do atendimento do usuário, não necessitando informações adicionais. Os três últimos módulos são usados para estabelecer o fluxo do usuário no *Call Center*, simulando suas ações, atividades e o estado que cada contato se submete até atingir um agente.

Ao final das replicações, o *software* Arena gera um relatório contendo várias informações de cada uma das rodadas e também um relatório síntese (*Siman Summary Report*) que contém os resultados de todas as replicações. Entre estes resultados estão as medidas de desempenho, como o tempo médio de espera, a probabilidade de abandono, a intensidade de tráfego e a probabilidade de esperar, que é calculada pela divisão entre o número de usuários que foram atendidos (excluídos os que abandonaram) e o número de usuários criados na simulação. Além do resultado dessas medidas, também é informado, no relatório, o semi-intervalo (h) do intervalo de confiança.

A verificação dos modelos de fila utilizados neste estudo foi realizada comparando-se os modelos analíticos com seus respectivos modelos simulados. Os resultados dos modelos analíticos foram obtidos por meio de um algoritmo computacional que o represente, programado no *software* Mathematica (Wolfram, 2013). As comparações dos resultados entre os modelos foram obtidas para cada uma das medidas de desempenho consideradas neste estudo, ou seja, o tempo médio de espera  $E(W)$ , a probabilidade de esperar  $P(W)$ , a probabilidade de abandonar  $P(A)$  e a intensidade de tráfego ( $\rho$ ). Estas medidas foram exibidas em tabelas separadas, para cada intervalo de hora estudado, contendo, além dos resultados obtidos, os intervalos de confiança de 95% de probabilidade.

A análise da verificação para constatar se os modelos de fila utilizados neste estudo são consistentes para representar o *Call Center* do Banco é descrita, em detalhes, nos tópicos seguintes. Para essa análise é considerado o intervalo de confiança de 95% de probabilidade, obtidos com a simulação.

#### 4.5.1 Verificação dos modelos - intervalo 15-16 horas

A intenção desta seção é verificar se os modelos de simulação implementados, relativos ao intervalo das 15-16 horas, foram desenvolvidos corretamente. Essa verificação é feita comparando-se os valores obtidos com o modelo simulado de cada uma das medidas de desempenho com os correspondentes valores obtidos com os modelos analíticos de fila. Essa comparação é realizada em cada um dos horários, do *Call Center* do Banco, considerados neste estudo.

Nessa análise utilizou-se os intervalos de confiança, com  $\gamma = 95\%$  de probabilidade, calculados para cada uma das medidas de desempenho obtidas em cada um dos modelos de fila simulados. Foi considerado verificado o modelo de simulação cujo intervalo de confiança contém o resultado da medida de desempenho obtido pelo correspondente modelo analítico de fila. Segundo Magalhães e Lima (2002), obtendo-se várias amostras de mesmo tamanho e se para cada uma calcula-se os correspondentes intervalos de confiança com coeficiente de confiança  $\gamma$ , espera-se que a proporção de intervalos que contenham o valor do parâmetro  $\theta$  seja igual a  $\gamma$ . Com base neste conceito, considerando-se coeficiente de confiança (ou probabilidade de cobertura teórica)  $\gamma = 0,95$ , é esperado que em aproximadamente 95% das réplicas extraídas das simulações, os intervalos correspondentes contenham o valor da medida de desempenho obtida por um dos modelos analíticos de fila utilizado neste estudo.

Os resultados dos modelos simulados, os intervalos de confiança correspondentes e os resultados dos modelos analíticos estão discriminados em tabelas, para cada uma das medidas de desempenho consideradas neste estudo, em cada um dos horários estabelecidos. Os resultados dessas tabelas foram organizados de acordo com os tipos de modelos de filas (analíticos ou simulados) e, também, de acordo com a distribuição de probabilidade da paciência. Com relação aos modelos de filas a ordem estabelecida foi considerar primeiro os modelos de fila do tipo  $M/M/c+G$ , seguido dos modelos de fila  $M/G^c/1+G$  e depois dos modelos de fila  $M/G/c+G$ , com distribuição não mista para representar a paciência e, em seguida, os mesmos modelos com as respectivas distribuições mistas para a paciência. Com a intenção de facilitar a apresentação das tabelas, separou-se com uma linha tracejada os modelos de filas agrupados de acordo com a ordem estabelecida.

Desse modo, observando-se as Tabelas 4.18 e 4.19 que mostram os resultados das medidas de desempenho obtidos com os modelos analíticos e de simulação, relativos ao intervalo das 15 às 16 horas, nota-se que a consistência (verificação) dos modelos de fila  $M/M/c + G$ ,  $M/G^c/1 + G$  e  $M/G/c + G$  varia conforme a medida de desempenho considerada. Para facilitar a análise, marcou-se com asterisco (\*) os valores da medida de desempenho cujo intervalo de confiança contém o valor da mesma medida obtida por meio do modelo analítico de fila. Assim, observa-se que apenas cinco resultados simulados não foram considerados verificados para a probabilidade de abandonar e três resultados simulados não são verificados para a intensidade de tráfego, porque os seus intervalos de confiança não contêm os valores calculados pelo correspondente modelo analítico (Tabela 4.19). Quanto ao tempo médio de espera, apenas três resultados simulados são considerados verificados e oito resultados simulados são verificados para a probabilidade de esperar (Tabela 4.18).



TABELA 4.18: Verificação dos modelos de fila – tempo médio de espera e probabilidade de esperar *Call Center* Banco – Intervalo 15-16 horas

Modelo	Tempo médio espera			Probabilidade de esperar		
	Simulação	Intervalo	Confiança Analítico	Simulação	Intervalo	Confiança Analítico
M/M <sub>3</sub> /c + G <sub>17</sub>	2,3366	(2,2442; 2,4290)	1,6526	0,8048	(0,7219; 0,8877)	0,8492*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>17</sub>	1,9139	(1,7708; 2,0570)	1,9787*	0,7279	(0,5848; 0,8710)	0,9067
M/G <sub>16</sub> /c + G <sub>17</sub>	2,4052	(2,2626; 2,5478)	2,0812	0,7961	(0,7036; 0,8886)	0,9541
M/M <sub>3</sub> /c + G <sub>19m</sub>	2,0637	(1,9918; 2,1356)	1,2437	0,8086	(0,7279; 0,8893)	0,7986*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>19m</sub>	1,6256	(1,5220; 1,7292)	1,9843	0,7075	(0,4575; 0,9575)	0,9028*
M/G <sub>16</sub> /c + G <sub>19m</sub>	2,0220	(1,9118; 2,1322)	2,1069*	0,7865	(0,6304; 0,9426)	0,9593
M/M <sub>3</sub> /c + G <sub>20</sub>	1,5778	(1,4930; 1,6626)	1,0074	0,7954	(0,7228; 0,8680)	0,7492*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>20</sub>	1,6564	(1,5329; 1,7799)	1,9765	0,6942	(0,5368; 0,8516)	0,9036
M/G <sub>16</sub> /c + G <sub>20</sub>	1,7354	(1,6407; 1,8301)	2,0562	0,7687	(0,6487; 0,8887)	0,9409
M/M <sub>3</sub> /c + G <sub>21m</sub>	1,2207	(0,8435; 1,5979)	0,5428	0,7938	(0,7469; 0,8407)	0,6470
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>21m</sub>	1,0183	(0,9644; 1,0722)	1,4659	0,6695	(0,4942; 0,8448)	0,5687*
M/G <sub>16</sub> /c + G <sub>21m</sub>	1,3430	(1,2851; 1,4009)	1,5161	0,7774	(0,6952; 0,8596)	0,5895
M/M <sub>3</sub> /c + G <sub>22</sub>	2,3899	(2,2883; 2,4915)	1,7694	0,8135	(0,7314; 0,8956)	0,8590*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>22</sub>	1,9090	(1,7591; 2,0589)	1,9813*	0,7422	(0,4198; 1,0646)	0,9015*
M/G <sub>16</sub> /c + G <sub>22</sub>	2,4891	(2,3904; 2,5878)	2,0859	0,7958	(0,7114; 0,8802)	0,9495
M/M <sub>3</sub> /c + G <sub>23m</sub>	2,1465	(2,0505; 2,2425)	1,4366	0,8178	(0,5490; 1,0866)	0,8037*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>23m</sub>	1,7057	(1,5801; 1,8313)	1,9815	0,7263	(0,5570; 0,8956)	0,9029
M/G <sub>16</sub> /c + G <sub>23m</sub>	2,3541	(2,2437; 2,4645)	2,0757	0,7893	(0,6428; 0,9358)	0,9464
Processo chegada: Poisson – taxa $\lambda = 2,1$ – média $E(\tau) = 0,4762$ Distribuições: M ~ Exponencial(2,1)			Processo serviço: Taxa $\mu = 0,291$ – média $E(S) = 3,4364$ Distribuições: G <sub>16</sub> ~ Lognormal (0,9415; 0,8377) G <sub>18</sub> ~ Lognormal (0,5727; 4,2996) M <sub>3</sub> ~ Exponencial (0,291)			
Processo abandono: taxa 0,108 – média $E(R) = 9,2593$ Distribuições: G <sub>17</sub> ~ Fatigue Life (0,7484; 7,2309)      G <sub>20</sub> ~ Exponencial (0,108)      G <sub>22</sub> ~ Lognormal (1,9962; 0,6769) G <sub>19m</sub> ~ 0,9215×Fatigue Life (0,7086; 6,5378) + 0,0785×Fatigue Life (0,1503; 0,8598) G <sub>21m</sub> ~ 0,4926×Exponencial (0,2193) + 0,5074×Exponencial (0,2129) G <sub>23m</sub> ~ 0,95×Lognormal (1,9353; 0,6909) + 0,05×Lognormal (-2,0382; 1,5923)						

Fonte: elaborada pelo autor

TABELA 4.19: Verificação dos modelos de fila – probabilidade abandonar e intensidade de tráfego *Call Center* Banco – Intervalo 15-16 horas

Modelo	Probabilidade de abandonar			Intensidade de tráfego		
	Simulação	Intervalo Confiança	Análítico	Simulação	Intervalo Confiança	Análítico
M/M <sub>3</sub> /c + G <sub>17</sub>	0,1962	(0,1133; 0,2791)	0,2091*	0,9759	(0,9251; 1,0267)	0,9513*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>17</sub>	0,2755	(0,1324; 0,4186)	0,1086	0,8489	(0,5311; 1,1667)	1,0721*
M/G <sub>16</sub> /c + G <sub>17</sub>	0,2056	(0,1131; 0,2981)	0,2091*	1,0018	(0,9405; 1,0631)	1,0721
M/M <sub>3</sub> /c + G <sub>19m</sub>	0,1922	(0,1115; 0,2729)	0,2227*	0,9766	(0,8878; 1,0654)	0,9349*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>19m</sub>	0,2901	(0,0401; 0,5401)	0,2074*	0,892	(0,5245; 1,2595)	0,9533*
M/G <sub>16</sub> /c + G <sub>19m</sub>	0,2157	(0,0596; 0,3718)	0,2227*	0,9958	(0,8744; 1,1172)	0,9533*
M/M <sub>3</sub> /c + G <sub>20</sub>	0,2020	(0,1294; 0,2746)	0,2338*	0,9627	(0,8918; 1,0336)	0,9216*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>20</sub>	0,3058	(0,1484; 0,4632)	0,1045	0,8773	(0,5236; 1,2310)	1,0771*
M/G <sub>16</sub> /c + G <sub>20</sub>	0,2364	(0,1164; 0,3564)	0,2338*	0,9885	(0,9324; 1,0446)	1,0771
M/M <sub>3</sub> /c + G <sub>21m</sub>	0,2076	(0,1607; 0,2545)	0,2596	0,9579	(0,8865; 1,0293)	0,8905*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>21m</sub>	0,3326	(0,1573; 0,5079)	0,1023	0,8274	(0,5055; 1,1493)	1,0797*
M/G <sub>16</sub> /c + G <sub>21m</sub>	0,2225	(0,1403; 0,3047)	0,2596*	0,9869	(0,9330; 1,0408)	1,0797
M/M <sub>3</sub> /c + G <sub>22</sub>	0,1868	(0,1047; 0,2689)	0,2064*	0,9782	(0,8947; 1,0617)	0,9544*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>22</sub>	0,2629	(0,0000; 0,5853)	0,1089*	0,8866	(0,5054; 1,2678)	1,0717*
M/G <sub>16</sub> /c + G <sub>22</sub>	0,2028	(0,1184; 0,2872)	0,2064*	1,0080	(0,9425; 1,0735)	1,0717*
M/M <sub>3</sub> /c + G <sub>23m</sub>	0,1916	(0,0000; 0,4604)	0,2165*	0,9835	(0,9017; 1,0653)	0,9423*
M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>23m</sub>	0,2789	(0,1096; 0,4482)	0,1071	0,8935	(0,4109; 1,3761)	1,0739*
M/G <sub>16</sub> /c + G <sub>23m</sub>	0,2117	(0,0652; 0,3582)	0,2165*	1,0083	(0,9032; 1,11340)	1,0739*
Processo chegada: Poisson – taxa $\lambda = 2,1$ – média $E(\tau) = 0,4762$ Distribuições: M ~ Exponencial(2,1)			Processo serviço: Taxa $\mu = 0,291$ – média $E(S) = 3,4364$ Distribuições: G <sub>16</sub> ~ Lognormal (0,9415; 0,8377) G <sub>18</sub> ~ Lognormal (0,5727; 4,2996) M <sub>3</sub> ~ Exponencial (0,291)			
Processo abandono: taxa 0,108 – média $E(R) = 9,2593$ Distribuições: G <sub>17</sub> ~ Fatigue Life (0,7484; 7,2309)      G <sub>20</sub> ~ Exponencial (0,108)      G <sub>22</sub> ~ Lognormal (1,9962; 0,6769) G <sub>19m</sub> ~ 0,9215×Fatigue Life (0,7086; 6,5378) + 0,0785×Fatigue Life (0,1503; 0,8598) G <sub>21m</sub> ~ 0,4926×Exponencial (0,2193) + 0,5074×Exponencial (0,2129) G <sub>23m</sub> ~ 0,95×Lognormal (1,9353; 0,6909) + 0,05×Lognormal (-2,0382; 1,5923)						

Fonte: elaborada pelo autor

No Apêndice D encontra-se a verificação dos modelos de simulação implementados nos horários das 11-12 horas e das 12-13 horas.

## 5 O CASO DO CALL CENTER DA EMPRESA

Este capítulo se dedica à análise dos dados do caso do *Call Center* da Empresa, descrita na Seção 1.3. Inicialmente é feita uma descrição da estrutura de funcionamento desse *Call Center*, destacando-se a quantidade de chamadas em cada ponto do sistema. A seguir apresenta-se uma descrição e análise dos dados coletados, também, uma descrição do comportamento estatístico dos processos de chegada, serviço e abandono e, finalizando com uma explanação e verificação do modelo de simulação desenvolvido para analisar esse *Call Center*.

### 5.1 Descrição do Call Center da empresa

O *Call Center* considerado neste estudo é de uma empresa que desenvolve *softwares* especializados para o gerenciamento de farmácias e drogarias, localizada no interior do Estado de São Paulo. Este *Call Center* oferece aos seus usuários os seguintes tipos de serviços:

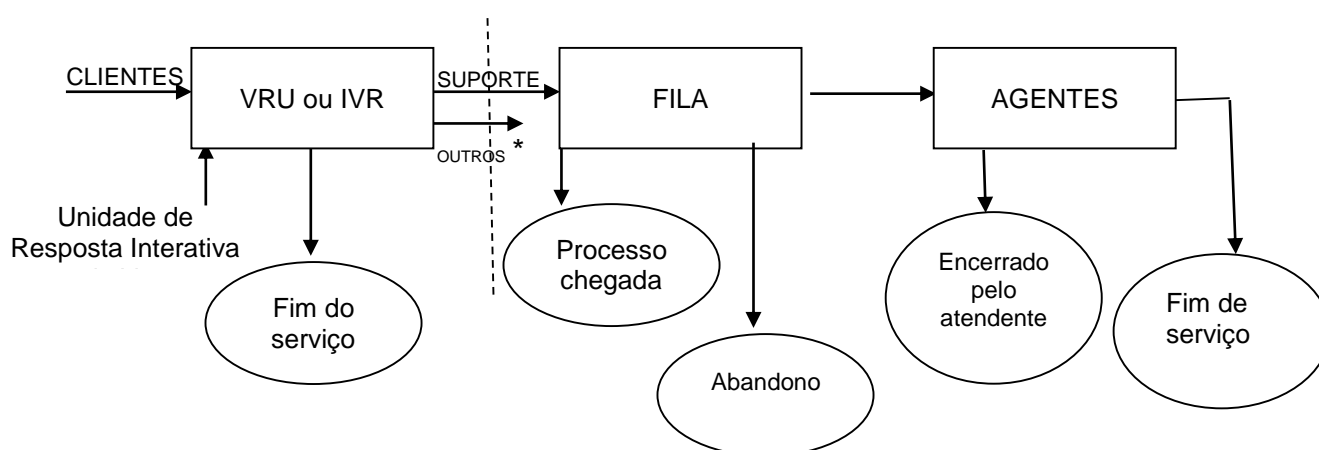
- i) suporte técnico para os clientes solucionarem problemas relacionados com o software;
- ii) venda virtual para os usuários que desejam comprar produtos da empresa;
- iii) atendimento dos clientes com problemas em contrato;
- iv) serviços relacionados com mensalidades, boletos e outros serviços financeiros.

Neste estudo, foi considerado apenas o serviço de suporte técnico. Os demais tipos de serviços possuem baixo fluxo de ligações e pouca regularidade, sendo desconsiderados. O serviço de suporte técnico se divide em atendimento telefônico, quando o cliente liga para o *Call Center*, e o atendimento *on line*, em que o cliente é atendido pela internet, por meio de um *chat*. O *Call Center* possui 15 posições de atendimento, sendo 14 delas destinadas ao suporte técnico, trabalhando no horário comercial de segunda a sábado. Os usuários quando acessam o *Call Center* da Empresa conectam-se com a unidade de resposta por voz (IVR), fornecendo seus dados para identificação e escolhem uma das opções de serviço.

Existem, em média, 4.832 chamadas mensais que escolhem a opção de suporte técnico para solucionar os seus problemas. Esses clientes juntam-se à fila de espera para dialogar com os atendentes e, em média, 434 (8,98%) deles abandonam o sistema depois de aguardar uma quantidade positiva de tempo na fila. Os restantes 4.398 (91,01%) solucionam os seus

problemas com os atendentes e deixam o sistema, sendo que 0,11% (5 clientes) deles têm os seus serviços encerrados pelo próprio atendente, sem solucionarem seus problemas. Os que aguardam na fila recebem informações sobre sua posição na fila e a quantidade estimada de tempo que levarão para serem atendidos. Essas mensagens são repetidas a cada 1 minuto, intercaladas com músicas e comerciais dos produtos vendidos pela empresa. A Figura 5.1 resume os fluxos das chamadas relativas ao serviço de suporte técnico, cujos valores representam as quantidades mínimas, médias e máximas de chamadas mensais.

FIGURA 5.1 - Fluxo das chamadas do *Call Center* da empresa



Fonte: elaborada pelo autor

## 5.2 Análise estatística do Call Center da empresa

Como fonte de dados, considera-se os registros de cada chamada, contendo a sua identificação e os instantes de tempo ocorridos em cada ponto do fluxo das chamadas. Esses dados já foram coletados e referem-se ao período de 02 de janeiro a 31 de dezembro de 2009. Foram extraídos do sistema gerenciador do *Call Center* que registra os detalhes de cada chamada e foram organizados em uma planilha, em que cada linha corresponde ao registro de cada chamada. Uma análise crítica foi realizada para detectar e corrigir alguma incoerência ocorrida. No Anexo B, encontra-se um extrato dos dados obtidos no período informado.

### 5.2.1 Descrição dos dados

Para uma melhor compreensão das informações contidas nesses dados, uma breve descrição do conteúdo das variáveis extraídas do *Call Center* da empresa é feita a seguir. Esses dados correspondem a todas as chamadas atendidas no período de 02 de janeiro a 31 de dezembro de 2009, referentes ao tipo de serviço “Suporte técnico”. Cada registro dessas chamadas contém 20 campos (variáveis).

**ID ATENDIMENTO:** número sequencial que identifica a chamada

**ID TIPO ATENDIMENTO:** são os números que identificam os filtros realizados pela URA. Foram consideradas neste estudo apenas as solicitações de suporte técnico porque possuem maior regularidade.

**TIPO DE ATENDIMENTO:** os atendimentos dividem-se em atendimento telefônico, que são as ligações telefônicas originadas pelo cliente ou aquelas em que a empresa liga para o cliente, e o atendimento *online* em que o cliente aguarda numa fila para ser atendido via internet (*chat online*).

**CÓDIGO DO CLIENTE:** variável numérica que identifica o cliente por meio do seu código.

**NÚMERO DA LOJA:** cada cliente pode ter uma ou mais lojas e essa variável registra essa quantidade de lojas.

**NOME DO USUÁRIO:** variável que registra o nome do usuário.

**ASSUNTO DO ATENDIMENTO:** variável que recebe o assunto do atendimento.

**DATA ENTRADA:** variável que registra a data (dia, mês, ano) da chegada da chamada.

**HORA ENTRADA:** variável que registra a hora (hora, minuto, segundo) da chegada da chamada.

**POSIÇÃO FILA:** variável numérica que indica a posição do cliente na fila.

**DATA INÍCIO:** variável que registra a data (dia, mês, ano) do início do atendimento.

**HORA INÍCIO:** variável que registra a hora (hora, minuto, segundo) do início do atendimento.

**DATA TÉRMINO:** variável que registra a data (dia, mês, ano) do encerramento do atendimento.

**HORA TÉRMINO:** variável que registra a hora (hora, minuto, segundo) do encerramento do atendimento.

**STATUS:** variável que classifica a condição do atendimento nas categorias D: quando houve uma desistência e por isso os registros com esse status não possuem data e nem hora de início do atendimento; E: quando o atendimento foi encerrado pelo atendente ou quando o cliente

encerrou, mas não houve uma finalização do assunto e, ainda, pode ter ocorrido um encerramento por perda de conexão; F: quando a atendimento foi finalizado e considerado concluído por ambas as partes.

CONCLUSÃO: essa variável registra a condição que o cliente classificou o atendimento na pesquisa de satisfação realizada automaticamente pelo sistema, assumindo valores A: “aberto”, indicando que o cliente não julgou se o seu problema foi resolvido; N: indicando a condição “não resolvido” e S: classificando o atendimento em “resolvido”.

ID CATEGORIA: variável que indica uma subdivisão da variável sistema, indicando o assunto por meio de um número. Por exemplo, 392 significa Categoria Relatórios no sistema M.

SISTEMA: essa variável indica em qual sistema o atendimento foi prestado, ou seja, sobre qual sistema recai a dúvida ou principal problema, sendo F: FrontFarma; O: OuroFarma e M: OuroFormulas (manipulação, por isso o M).

CATEGORIA ATENDIMENTO: variável que registra o assunto do atendimento.

MOTIVO DESISTÊNCIA: essa variável armazena o motivo pelo qual o cliente desistiu da espera pelo atendimento. No período da coleta de dados, tratava-se de um novo item que ainda não era requerido dos atendentes do *Call Center*.

Outras variáveis são adicionadas por meio de cálculos realizados com as variáveis existentes para expressar o tempo entre chegadas das chamadas, o tempo de serviço exigido em cada chamada, o tempo de abandono e outras que sejam importantes para o estudo a ser realizado.

### 5.2.2 Análise dos dados

A análise dos dados reais do *Call Center* da Empresa foi conduzida determinando-se as estatísticas descritivas de interesse, tais como a média do tempo entre chegadas das chamadas, o tempo médio de duração de cada chamada (serviço), o tempo médio para abandonar que mede a impaciência do cliente. Assim, é possível determinar as taxas de chegada, serviço e abandono e algumas medidas de desempenho do sistema, como o tempo médio de espera na fila, a probabilidade de espera, a probabilidade de abandono e também a intensidade de tráfego. Para essa finalidade utilizou-se do software MS-Excel® para efetuá-las.

Com os dados coletados referente ao ano de 2009, identificou-se em qual momento o *Call Center* operou mais congestionado para realizar as análises estatísticas. Observando-se a Tabela 5.1, verifica-se que o mês de abril possui o maior número de atendimentos no ano (5.569

atendimentos) e a maior quantidade de abandonos (723 abandonos). Diante dessas informações, considerou-se o mês de abril para realizar as análises estatísticas descritivas do *Call Center* da Empresa.

TABELA 5.1 - atendimentos realizados no ano 2009 – *Call Center* Empresa

Mês	Abandono	Encerrada atendente	Finalizada	Total
Janeiro	373	2	3968	4343
Fevereiro	421	3	3490	3914
Março	547	18	4864	5429
Abril	723	10	4836	5569
Mai	486	1	4197	4684
Junho	468	4	4180	4652
Julho	501	13	4555	5069
Agosto	292	3	4364	4659
Setembro	284	1	4547	4832
Outubro	317	1	4547	4865
Novembro	374	1	4413	4788
Dezembro	423	2	4750	5175
<b>Total</b>	<b>5209</b>	<b>59</b>	<b>52711</b>	<b>57979</b>

Fonte: elaborada pelo autor

Em seguida, identificou-se em qual dia do mês de abril ocorreu maior fluxo de ligações no *Call Center*. A Tabela 5.2 mostra que o dia 01 de abril possui o maior número de chamadas (459) atendidas ou abandonadas e também a maior quantidade de abandonos (71). Portanto, o dia 01 de abril de 2009 foi escolhido para as análises do sistema *Call Center*.

TABELA 5.2 - Atendimentos diários – *Call Center* Empresa

<b>Dia</b>	<b>Abandono</b>	<b>Encerrado atendente</b>	<b>Finalizada</b>	<b>Total</b>
01/04/2009	71		388	459
02/04/2009	28	1	269	298
03/04/2009	20	1	240	261
04/04/2009	3		77	80
06/04/2009	40		225	265
07/04/2009	22		189	211
08/04/2009	23		210	233
09/04/2009	29		210	239
11/04/2009	8		64	72
13/04/2009	34		183	217
14/04/2009	57	1	239	297
15/04/2009	55		245	300
16/04/2009	57		254	311
17/04/2009	30	3	236	269
18/04/2009	5	1	70	76
20/04/2009	39	1	191	231
22/04/2009	24		236	260
23/04/2009	29	1	223	253
24/04/2009	37	1	222	260
25/04/2009	3		66	69
27/04/2009	19		196	215
28/04/2009	32		197	229
29/04/2009	35		213	248
30/04/2009	23		193	216
<b>Total</b>	<b>723</b>	<b>10</b>	<b>4836</b>	<b>5569</b>

Fonte: elaborada pelo autor

Depois de conhecer o mês e o dia do mês que estarão sujeitos às análises estatísticas descritivas, é necessário identificar o horário mais conveniente do dia 01 de abril para realizar essas análises. O fluxo de chamadas atendidas ou abandonadas em cada hora de funcionamento do *Call Center*, no dia 01 de abril, é mostrado na Tabela 5.3. Conforme cálculos realizados, os horários das 8-9 horas, 9-10 horas e 10-11 horas apresentaram intensidade de tráfego muito alta, com  $\rho > 1$  ou muito próxima da unidade. Por isso, não foram considerados neste estudo. Entretanto, o intervalo mais amplo das 12-16 horas apresenta um número de chamadas próximas um do outro e com uma intensidade de tráfego inferior à unidade e, foi escolhido para as análises estatísticas. Nesse intervalo mais amplo ocorreram 212 chamadas, correspondendo a 46,2% do total das chamadas ocorridas no dia 01 de abril com 35 abandonos, representando 49,3% dos abandonos do dia considerado (Tabela 5.4).



TABELA 5.3 - Atendimento por intervalo de horas – *Call Center* Empresa

<b>Horário</b>	<b>Suporte Online</b>	<b>Suporte telefônico</b>	<b>Total</b>
8 horas	28	36	64
9 horas	41	30	71
10 horas	34	23	57
11 horas	17	16	33
12 horas	23	14	37
13 horas	25	16	41
14 horas	25	25	50
15 horas	18	24	42
16 horas	19	23	42
17 horas	7	15	22
<b>Total</b>	<b>237</b>	<b>222</b>	<b>459</b>

Fonte: elaborada pelo autor

TABELA 5.4 - Abandono por intervalo de horas  
*Call Center* Empresa

<b>Horário</b>	<b>Suporte Online</b>
8 horas	3
9 horas	9
10 horas	17
11 horas	3
12 horas	2
13 horas	7
14 horas	12
15 horas	7
16 horas	7
17 horas	4
<b>Total</b>	<b>71</b>

Fonte: elaborada pelo autor

Assim, os dados que representam o Call Center da Empresa nas análises estatísticas, nos modelos analíticos e de simulação de filas, são os do horário das 12-16 horas, do dia 01 de abril de 2009.

### 5.3 Análise dos processos de chegada, serviço e abandono

Como os processos de chegada, serviço e abandono são aleatórios, é necessário fazer uma análise para verificar o comportamento da distribuição dos tempos entre chegadas, dos tempos de serviços e dos tempos de paciência que ocorrem no *Call Center* da Empresa. Essa investigação é realizada nas próximas seções, em detalhes. Dessa análise é definido o modelo analítico de fila a ser utilizado para representar esse *Call Center*.

#### 5.3.1 O processo de chegada

O processo de chegada do *Call Center* da Empresa é semelhante ao processo de chegada do *Call Center* do Banco descrito no Capítulo 4. Especificamente no caso Empresa, esse processo registra os instantes em que as chamadas *online* chegam na fila de espera, depois de terem visitado a VRU. Depois de deixarem a VRU, os usuários ingressam na fila de espera e se ocorrer de encontrar um atendente desocupado, seu serviço inicia-se imediatamente. Neste caso sendo zero o seu tempo de permanência na fila. Se não houver atendentes disponíveis, o usuário aguarda uma quantidade positiva de tempo na fila até iniciar o seu serviço, ou até abandonar o sistema, devido à sua impaciência.

Este processo é avaliado por meio das informações da estatística descritiva e também por sua variabilidade estocástica. Adotou-se o tempo entre as chegadas das chamadas *online* para avaliar o processo. Para essa finalidade, considerou-se os instantes de chegada na fila, registrados na variável *Hora entrada*, referentes às chamadas *online* que aguardaram até iniciar o serviço ou que abandonaram o sistema. Em seguida, calculou-se os tempos entre chegadas das chamadas *online* do dia 01 de abril de 2009, no horário das 12-16 horas. Após essa etapa, procedeu-se a análise estatística descritiva desses tempos entre chegadas das chamadas *online*, obtendo-se as medidas mais características, a sua média, mediana e desvio padrão, que estão sumarizadas na Tabela 5.5.

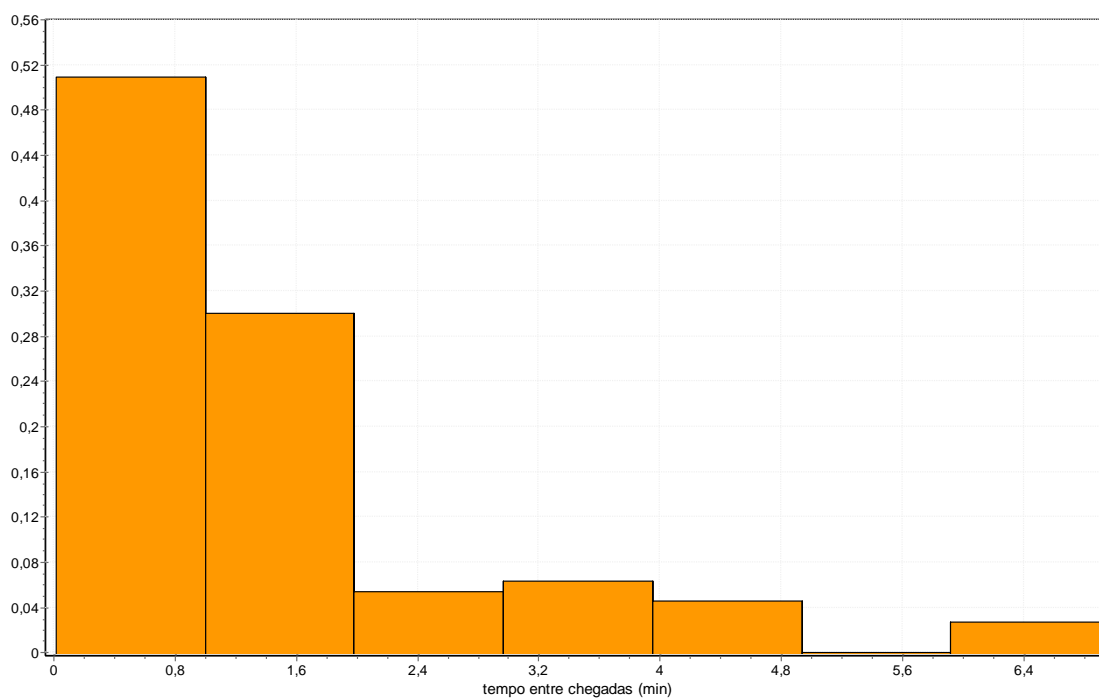
TABELA 5.5 - Tempo entre chegadas (min)  
*Call Center* Empresa – 12-16 horas

<b>Estatística</b>	<b>Valor</b>
Média	1,3665
Mediana	1
Desvio padrão	1,4584
Coef. variação	1,0673
Taxa chegada (ch/min)	0,7318

Fonte: elaborada pelo autor

Observa-se que a mediana é significativamente menor que a média, indicando que a distribuição dos tempos entre as chegadas das chamadas *online* é assimétrica à direita. O coeficiente de variação é próximo da unidade e essas duas informações sugerem que a distribuição dos tempos entre chegadas se aproxima da Exponencial. Outra comprovação dessa aproximação com a Exponencial é o formato do histograma (Gráfico 5.1) desses tempos entre chegadas das chamadas *online*.

GRÁFICO 5.1 - Distribuição do tempo entre chegadas – *Call Center* Empresa – 12-16 horas



Fonte: elaborado pelo autor

O teste Kolmogorov-Smirnov também confirmou a aproximação dos tempos entre chegada das chamadas *online* com a distribuição Exponencial de probabilidade, com *p-value* de 0,43675 ( $\alpha = 5\%$ ), conforme resultados obtidos pelo *software EasyFit*®, que ordena as distribuições mais aderentes aos dados.

### 5.3.2 O processo de serviço

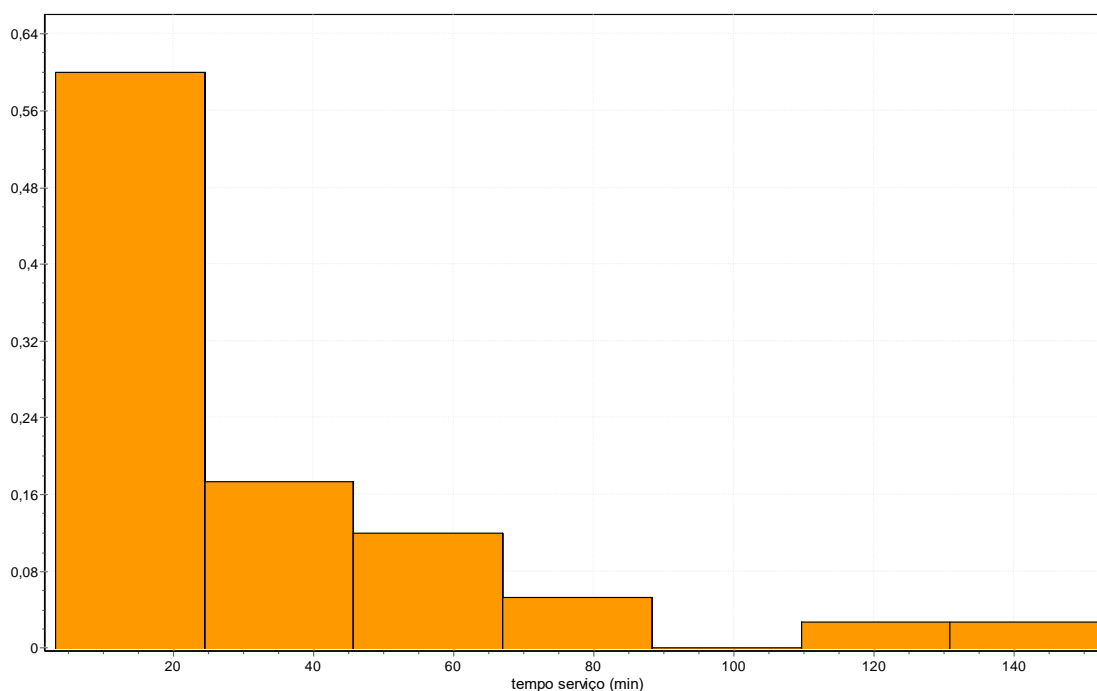
O processo de serviço do *Call Center* da Empresa é caracterizado pelo tempo gasto pelo atendente para realizar cada uma das chamadas *online*. Esse processo foi investigado quantitativamente e os tempos de serviço (tempo de atendimento) foram obtidos dos dados reais, calculando-se a diferença entre os valores das variáveis *hora início* e *hora término*, que registram os instantes inicial e final de cada atendimento. Obteve-se, em seguida, a média, mediana e desvio padrão desses tempos de serviço, informados na Tabela 5.6.

TABELA 5.6: Tempo de serviço (min)  
*Call Center* Empresa – 12-16 horas

<b>Estatística</b>	<b>Valor</b>
Média	30,9238
Mediana	12,5167
Desvio padrão	31,3813
Coef. variação	1,0148
Taxa serviço (ch/min)	0,0323

Fonte: elaborada pelo autor

Os resultados da Tabela 5.6 sugerem que a distribuição Exponencial deve se ajustar bem aos tempos de serviço deste *Call Center*, porque possui a mediana significativamente menor que a média e o coeficiente de variação próximo da unidade. O formato do histograma (Gráfico 5.2) dos tempos de serviço também parece confirmar esse suposto ajuste, devido a sua forma geométrica. Entretanto, o teste Kolmogorov-Smirnov aplicado a esses tempos de serviço, com o auxílio do *software EasyFit*®, versão 5.5 (2010), que ordena as distribuições mais aderentes, rejeitou a distribuição Exponencial (*p-value* 0,0016) e também rejeitou todas as demais 65 distribuições analisadas por esse *software*. Assim, optou-se pela distribuição Weibull, com 3 parâmetros e *p-value* 0,0034, representar os tempos de serviço do *Call Center* da Empresa, considerando que é a primeira da lista ordenada pelo *software*.

GRÁFICO 5.2 - Distribuição do tempo de serviço – *Call Center* Empresa – 12-16 horas

Fonte: elaborado pelo autor

### 5.3.3 O processo de abandono

O processo de abandono acontece quando os usuários das chamadas *online* do *Call Center* da Empresa que estão na fila de espera, aguardam uma quantidade positiva de tempo e depois abandonam o sistema, devido às suas impaciências. As considerações feitas na secção 4.4.2 quanto as três distinções de Brown *et al* (2002) para o tempo de espera, o fato desses tempos estarem baseados em dados censurados, conforme Mandelbaum, Sakov e Zeltyn (2001), foram respeitadas nas análises desta secção.

Na análise da estatística descritiva, considerou-se os tempos de paciência dos usuários das chamadas *online* do *Call Center* da Empresa, calculado com a diferença entre os valores registrados nas variáveis *hora entrada* e *hora término*, se a variável *hora início* apresenta nenhum registro de dados. A variável *hora entrada* assume os valores do instante de chegada da chamada *online* na fila de espera. Os valores dessas diferenças ficaram registrados na variável *T.AB*, que foi utilizada para calcular a média, mediana e desvio padrão, sumarizados na Tabela 5.7.

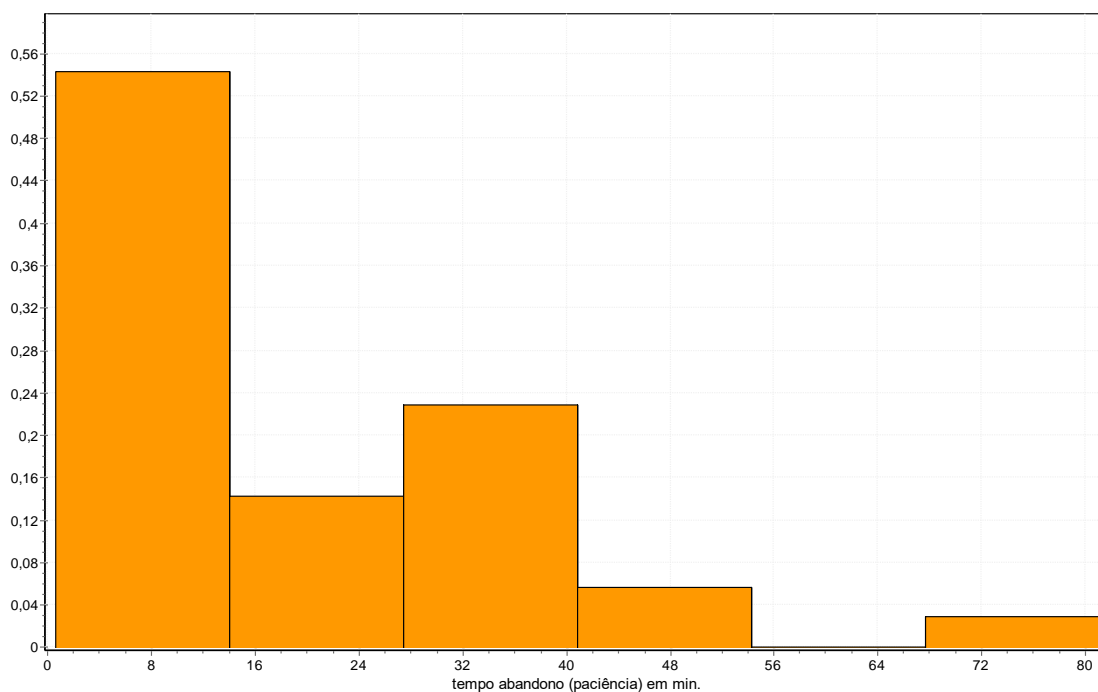
TABELA 5.7 - Tempo de paciência (min)  
*Call Center* Empresa – 12-16 horas

<b>Estatística</b>	<b>Valor</b>
Média	18,5038
Mediana	12,0333
Desvio padrão	18,0332
Coef. variação	0,9999
Taxa abandono (ch/min)	0,0540

Fonte: elaborada pelo autor

A distribuição Exponencial pode ser sugerida para representar os tempos de paciência porque possui uma mediana menor que a média e um coeficiente de variação muito próximo da unidade (Tabela 5.7). A forma do histograma (Gráfico 5.3) dos tempos de abandono pode deixar dúvidas quanto ao ajuste da distribuição Exponencial.

GRÁFICO 5.3 - Distribuição do tempo de abandono (paciência) – *Call Center* Empresa – 12-16 horas



Fonte: elaborado pelo autor

Conforme o teste Kolmogorov-Smirnov aplicado aos dados dos tempos de abandono, com o software *EasyFit*®, versão 5.5 (2010), a distribuição Exponencial não foi rejeitada ao nível de 5% de probabilidade, com *p-value* 0,8248. Entre outras distribuições aderentes a esses

dados e que aparecem na lista ordenada do *software*, estão a Fatigue Life com *p-value* 0,5538 e a distribuição Normal com *p-value* 0,1886. Essas três distribuições foram adotadas para representar os tempos de paciência das chamadas online do *Call Center* da Empresa.

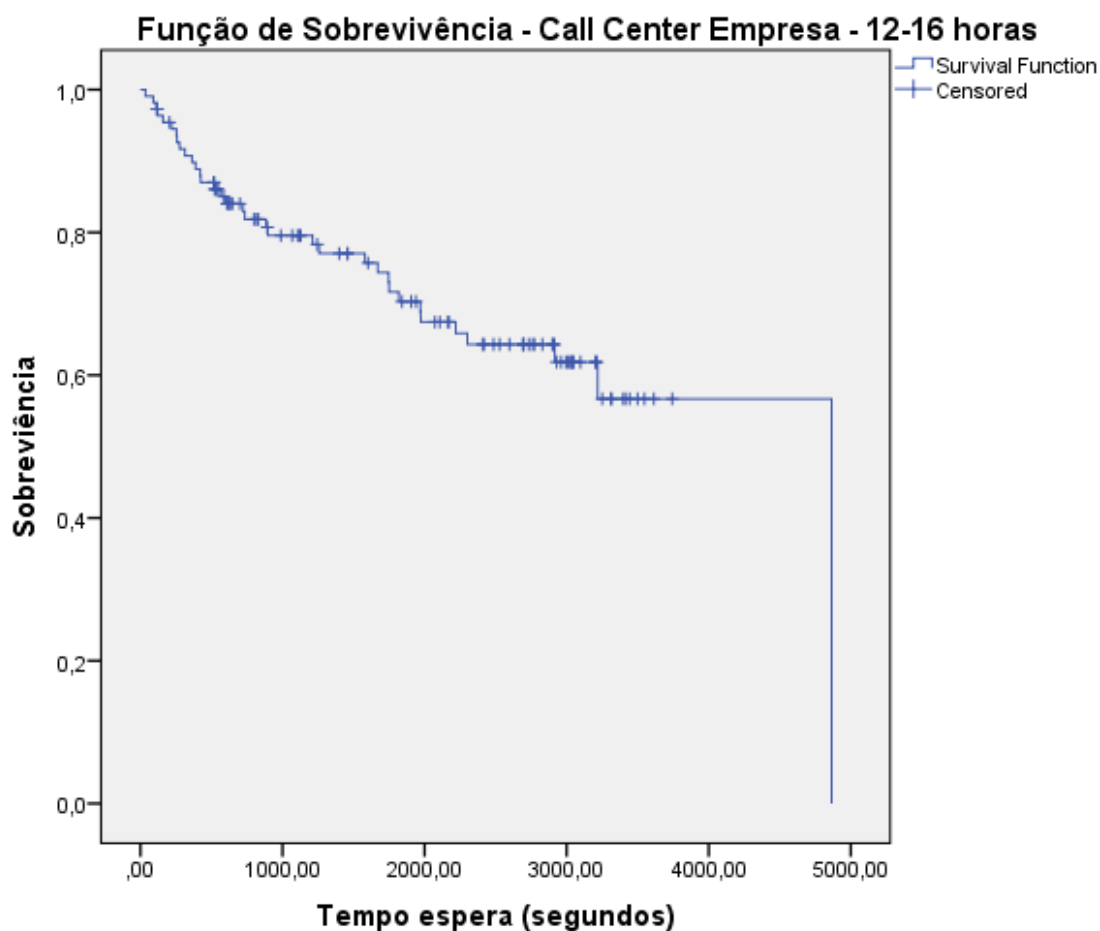
Como o interesse aqui é estimar a paciência dos usuários do *Call Center* da Empresa, ou seja, quanto tempo ele está disposto a aguardar antes de abandonar, então deve-se considerar o tempo necessário para um usuário atingir um agente como uma observação censurada. De fato, se ele atingiu o atendente é porque o tempo que estava disposto a esperar era maior.

Considerando-se as chamadas *online* que atingiram o atendente como observações censuradas, procedeu-se a estimação da paciência dos usuários por meio do estimador Kaplan-Meier, utilizando o *software* SPSS ®. A média, mediana e o desvio padrão das estimativas do tempo de paciência para as chamadas *online* do *Call Center* da Empresa, estão reportadas na Tabela 5.8 e o Gráfico 5.4 mostra as estimativas Kaplan-Meier da função de sobrevivência da paciência dos usuários do *Call Center* da Empresa.

TABELA 5.8 - Tempo de paciência (min). Estimativa K.M.  
*Call Center* Empresa – 12-16 horas

<b>Estatística</b>	<b>Valor</b>
Média	55,9089
Mediana	81,0833
Desvio padrão	21,1027
Coef. Variação	0,3774
Taxa abandono (ch/min)	0,0179

Fonte: elaborada pelo autor

GRÁFICO 5.4 - Estimativa da função de sobrevivência – *Call Center* Empresa – 12-16 horas

Fonte: elaborado pelo autor

Apenas para facilitar o entendimento dessa secção, um sumário das estatísticas descritivas e das taxas dos processos de chegada, serviço e abandono é mostrado na Tabela 5.9.

TABELA 5.9 - Resumo das estatísticas  
*Call Center* Empresa – 12-16 horas

<b>Estatística</b>	<b>Processo chegada</b>	<b>Processo Serviço</b>	<b>Processo Abandono</b>
Média	1,3605	30,9238	55,9089
Mediana	1	12,5167	81,0833
Desvio padrão	1,4584	31,3813	21,1027
Coef. Variação	1,0673	1,0148	0,3774
Taxas (ch/min)	0,7318	0,0323	0,0179

Fonte: elaborada pelo autor



Conhecidas as estatísticas descritivas, taxas e distribuições de probabilidades dos processos de chegada, serviço e abandono das chamadas *online* do *Call Center* da Empresa, obteve-se, por meio dos dados reais, suas medidas de desempenho: tempo médio de espera, probabilidade de esperar, probabilidade de abandonar e intensidade de tráfego. Essas medidas são comparadas com os resultados das mesmas medidas obtidas por meio dos modelos analíticos e de simulação de fila que representam o *Call Center* da Empresa, neste estudo.

O tempo médio de espera foi obtido calculando-se a média dos tempos de espera na fila das chamadas *online* que foram servidas por um atendente ou que abandonaram o sistema. A probabilidade de esperar foi calculada com o quociente entre a soma dos tempos das chamadas *online* que esperaram na fila e foram atendidas por um agente e a soma dos tempos das chamadas *online* que esperam na fila e são atendidas ou abandonam.

A probabilidade de abandonar também foi obtida por meio de um quociente, entre a quantidade de chamadas online que esperaram na fila e abandonaram e a quantidade de chamadas online que esperaram na fila e foram atendidas ou abandonadas. Obteve-se a intensidade de tráfego dividindo-se o número médio de atendentes provendo serviço pela quantidade de postos de atendimentos. Os resultados de todas essas medidas de desempenho, obtidas por meio dos dados reais, estão exibidos na Tabela 5.10.

TABELA 5.10: Medidas de desempenho  
*Call Center* Empresa – 12-16 horas

<b>Medidas desempenho</b>	<b>Valor</b>
Tempo médio espera	29,0418 (min)
Probabilidade esperar	0,7973
Probabilidade abandonar	0,3182
Intensidade tráfego	0,9333

Fonte: elaborada pelo autor

Com o propósito de fornecer uma visão geral das distribuições de probabilidade que mais se aderem aos dados dos processos de chegada, serviço e abandono, do *Call Center* da Empresa, no intervalo analisado das 12-16 horas, um sumário delas é exibido no Quadro 5.1. Essas distribuições foram utilizadas nas análises deste trabalho.

QUADRO 5.1 - Sumário das distribuições de probabilidades  
*Call Center* Empresa

<b>Processos</b>	<b>Distribuição ajustada</b>	<b><i>p-value</i> 12 – 16 h</b>
Chegada	Exponencial	0,43675
Serviço	Weibull	0,00340
	Exponencial	0,00160
Abandono	Fatigue Life	0,55380
	Normal	0,18860
	Exponencial	0,82480

Fonte: elaborada pelo autor

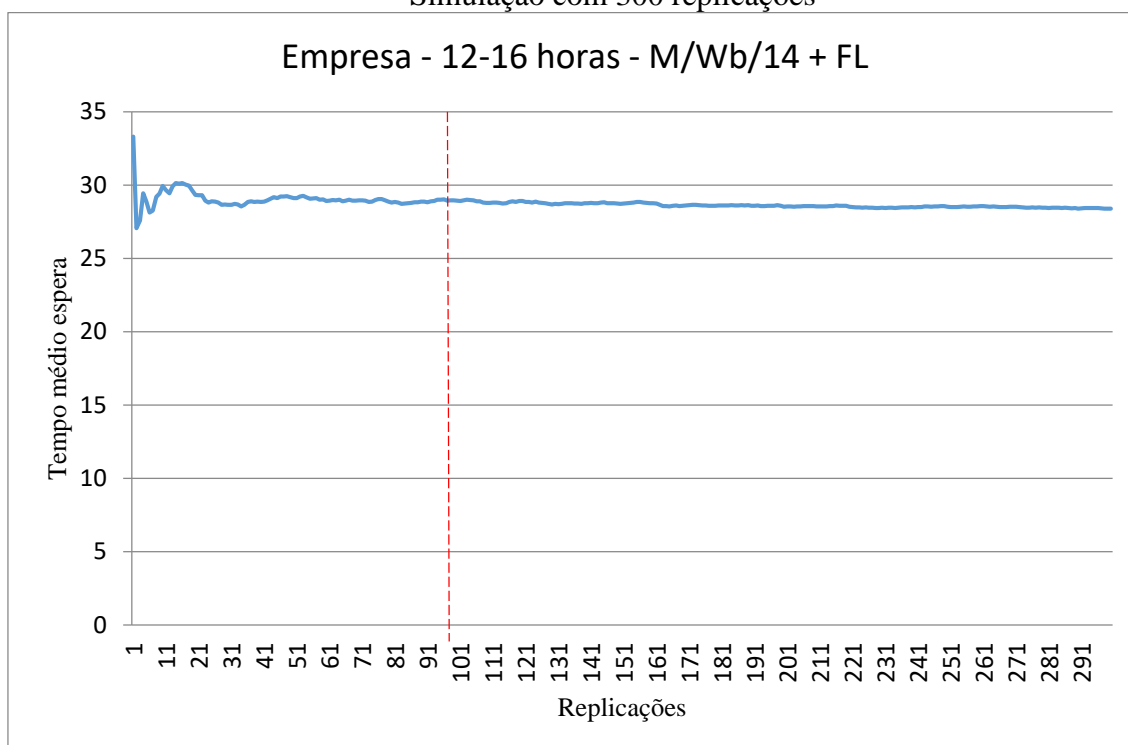
#### 5.4 Modelo de simulação

Os mesmos modelos aproximados de fila  $M/M/c + G$ ,  $M/G^c/1 + G$  e  $M/G/c + G$ , que foram considerados para representar o *Call Center* do Banco (ver item 3.7), também foram utilizados aqui para representar o *Call Center* da Empresa. No caso Empresa, adotou-se o mesmo procedimento descrito na seção 4.5 para realizar a verificação da consistência dos modelos de fila  $M/M/c + G$ ,  $M/G^c/1 + G$  e  $M/G/c + G$ , que utilizam ou não distribuição mista para representar o processo de abandono, usando simulação discreta.

A implementação da simulação é idêntica ao caso do Banco, no Capítulo 4, considerando-se também 300 replicações e cada replicação com duração de um dia, gerando no final de todas as rodadas uma média de 33 mil chamadas no *chat* do *Call Center* da Empresa. As estatísticas também foram coletadas a cada intervalo de 60 minutos e informadas em um relatório produzido pelo *software* Arena. Considerou-se as 100 primeiras rodadas da simulação como período de *warm up*, porque a partir desse período o sistema passou a operar em equilíbrio estatístico, conforme mostra o Gráfico 5.5, como exemplo.

No caso da Empresa, as simulações também foram realizadas em um *ultrabook* Samsung com processador Intel® Core™ I7 de 2,4 GHz e 8,00 GB de RAM e levaram em média 8,00 minutos para estimar as medidas de desempenho em cada modelo de fila. Com a finalidade de fazer a verificação dos modelos analíticos de fila utilizados para representar o *Call Center* da Empresa, efetuou-se a comparação destes modelos com os respectivos resultados simulados. Os resultados dos modelos analíticos também foram obtidos por meio de um algoritmo computacional, programado no *software* Mathematica (Wolfram, 2013). As comparações dos resultados entre os modelos foi obtida para cada uma das medidas de desempenho consideradas neste estudo, ou seja,  $E(W)$ ,  $P(W)$ ,  $P(A)$  e  $\rho$  e exibidas em tabelas separadas, contendo além dos resultados obtidos os intervalos de confiança de 95% obtidos com a simulação.

GRÁFICO 5.5 - *Warm up* – Caso Empresa – Modelo de fila M/Wb/14 + FL  
Simulação com 300 replicações



Fonte: elaborado pelo autor

Utilizou-se, no caso da Empresa, a mesma notação específica para representar os modelos de fila, usado no caso do Banco (ver seção 4.5). Os resultados das medidas de desempenho obtidos com os modelos analíticos (aproximados) e de simulação, referentes ao horário das 12 às 16 horas, do *Call Center* da Empresa, estão exibidos nas Tabelas 5.11 e 5.12.

Seguindo os mesmos procedimentos de análise adotados na seção 4.5.1, observa-se que cinco resultados simulados não foram considerados verificados para as medidas de desempenho probabilidade de abandonar e seis resultados simulados não foram considerados simulados para a medida intensidade de tráfego (Tabela 5.12). Nenhum dos 18 resultados simulados são verificados para a medida tempo médio de espera e onze resultados simulados são considerados verificados para a medida de desempenho probabilidade de esperar (Tabela 5.11). Essas análises indicam que existem modelos analíticos de fila que produzem melhores resultados em algumas medidas de desempenho do que em outras.

TABELA 5.11 - Verificação dos modelos de fila - tempo médio de espera e probabilidade de esperar *Call Center* Empresa – Intervalo 12-16 horas

Modelo	Tempo médio espera			Probabilidade de esperar		
	Simulação	Intervalo Confiança	Analítico	Simulação	Intervalo Confiança	Analítico
M/M <sub>4</sub> /c + G <sub>25</sub>	17,0870	(16,6100; 17,5640)	9,9227	0,6423	(0,5255; 0,7591)	0,9684
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>25</sub>	26,4110	(25,8166; 27,0054)	25,7044	0,6146	(0,5374; 0,6918)	0,6817*
M/G <sub>24</sub> /c + G <sub>25</sub>	19,1870	(18,7010; 19,6730)	25,7798	0,6878	(0,6276; 0,7480)	0,6836*
M/M <sub>4</sub> /c + G <sub>31m</sub>	15,9090	(15,3916; 16,4264)	8,8489	0,6554	(0,4908; 0,8200)	0,9566
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>31m</sub>	19,728	(19,2871; 20,1689)	24,3863	0,611	(0,5401; 0,6819)	0,6813*
M/G <sup>c</sup> <sub>24</sub> /c + G <sub>31m</sub>	14,3620	(13,8786; 14,8454)	24,4642	0,6907	(0,6274; 0,7540)	0,6834*
M/M <sub>4</sub> /c + G <sub>27</sub>	28,7120	(28,1766; 29,2474)	26,2853	0,6529	(0,5911; 0,7147)	0,9998
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>27</sub>	35,444	(35,0009; 35,8871)	31,8777	0,6161	(0,4629; 0,7693)	0,6759*
M/G <sub>24</sub> /c + G <sub>27</sub>	28,4010	(27,9996; 28,8024)	31,9469	0,6888	(0,6344; 0,7432)	0,6774*
M/M <sub>4</sub> /c + G <sub>29m</sub>	22,5540	(22,0223; 23,0857)	15,5145	0,6502	(0,4203; 0,8801)	0,9924
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>29m</sub>	27,1080	(26,7113; 27,5047)	26,0368	0,6210	(0,5337; 0,7083)	0,6819*
M/G <sub>24</sub> /c + G <sub>29m</sub>	21,9300	(21,5413; 22,3187)	26,1401	0,6975	(0,5971; 0,7979)	0,6846*
M/M <sub>4</sub> /c + G <sub>28</sub>	15,1780	(14,6764; 15,6796)	5,6283	0,6490	(0,5446; 0,7534)	0,4609
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>28</sub>	19,1400	(18,6708; 19,6092)	29,1182	0,6103	(0,5315; 0,6891)	0,6891*
M/G <sub>24</sub> /c + G <sub>28</sub>	14,1200	(13,7183; 14,5217)	29,2688	0,6800	(0,5953; 0,7647)	0,6927*
M/M <sub>4</sub> /c + G <sub>30m</sub>	16,1790	(15,8882; 16,4698)	4,3933	0,6555	(0,6073; 0,7037)	0,4493
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>30m</sub>	19,7490	(19,4957; 20,0023)	29,0288	0,6168	(0,5867; 0,6469)	0,6877
M/G <sub>24</sub> /c + G <sub>30m</sub>	15,9040	(15,5572; 16,2508)	29,1700	0,6827	(0,6066; 0,7588)	0,6911*
<b>Processo chegada:</b> Poisson – taxa $\lambda = 0,7318$ – média $E(\tau) = 1,3665$			<b>Processo serviço:</b> Taxa $\mu = 0,0323$ – média $E(S) = 30,9598$			
<b>Distribuições:</b> M ~ Exponencial(1,3665)			<b>Distribuições:</b> G <sub>24</sub> ~ Weibull (0,9276; 26,004; 3,1333) G <sub>26</sub> ~ Weibull (1,3021; 2,3952; 0) M <sub>4</sub> ~ Exponencial (30,9598)			
<b>Processo abandono:</b> taxa 0,01789 – média $E(R) = 55,9089$						
<b>Distribuições:</b> G <sub>25</sub> ~ Exponencial (55,9089)      G <sub>27</sub> ~ Fatigue Life (0,3726; 52,2792)      G <sub>28</sub> ~ Normal (55,9089; 445,3239) G <sub>29m</sub> ~ 0,8418×Fatigue Life (0,4055; 43,4875) + 0,1582×Fatigue Life (0,9434; 6,1219) G <sub>30m</sub> ~ 0,5221×Normal (29,1926; 232,5241) + 0,4779×Normal (26,7163; 212,7998) G <sub>31m</sub> ~ 0,9673×Exponencial (54,0799) + 0,0327×Exponencial (1,8289)						

Fonte: elaborada pelo autor

TABELA 5.12: Verificação dos modelos de fila - probabilidade abandonar e intensidade de tráfego *Call Center* Empresa – Intervalo 12-16 horas

Modelo	Probabilidade de abandonar			Intensidade de tráfego		
	Simulação	Intervalo Confiança	Análítico	Simulação	Intervalo Confiança	Análítico
M/M <sub>4</sub> /c + G <sub>25</sub>	0,3625	(0,2457; 0,4793)	0,3847*	1,0578	(1,0283; 1,0873)	0,9958
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>25</sub>	0,3867	(0,3095; 0,4639)	0,3832*	0,9936	(0,9800; 1,0072)	0,9982*
M/G <sub>24</sub> /c + G <sub>25</sub>	0,3132	(0,2530; 0,3734)	0,3847	1,0332	(0,9937; 1,0727)	0,9982*
M/M <sub>4</sub> /c + G <sub>31m</sub>	0,3492	(0,1846; 0,5138)	0,3856*	1,0558	(1,0263; 1,0853)	0,9943
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>31m</sub>	0,3908	(0,3199; 0,4617)	0,3833*	0,9942	(0,9727; 1,0157)	0,9980*
M/G <sup>c</sup> <sub>24</sub> /c + G <sub>31m</sub>	0,3094	(0,2461; 0,3727)	0,3856	1,0339	(0,9841; 1,0837)	0,9981*
M/M <sub>4</sub> /c + G <sub>27</sub>	0,3467	(0,3087; 0,3847)	0,3821*	1,0608	(1,0212; 1,1004)	0,9999
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>27</sub>	0,3808	(0,2736; 0,4880)	0,3829*	0,9978	(0,9877; 1,0079)	0,9987*
M/G <sub>24</sub> /c + G <sub>27</sub>	0,3123	(0,2665; 0,3581)	0,3821	1,0351	(1,0047; 1,0655)	0,9987
M/M <sub>4</sub> /c + G <sub>29m</sub>	0,3547	(0,1915; 0,5179)	0,3827*	1,0550	(1,0279; 1,0821)	0,9989
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>29m</sub>	0,3785	(0,3208; 0,4362)	0,3836*	0,9959	(0,9800; 1,0118)	0,9976*
M/G <sub>24</sub> /c + G <sub>29m</sub>	0,3147	(0,2249; 0,4045)	0,3827*	1,0303	(0,9942; 1,0664)	0,9975*
M/M <sub>4</sub> /c + G <sub>28</sub>	0,3521	(0,2477; 0,4565)	0,3962*	1,0476	(0,9761; 1,1191)	0,9771*
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>28</sub>	0,3909	(0,3121; 0,4697)	0,3840*	0,9931	(0,9679; 1,0183)	0,9968*
M/G <sub>24</sub> /c + G <sub>28</sub>	0,3226	(0,2379; 0,4073)	0,3962*	1,029	(0,9691; 1,0889)	0,9968*
M/M <sub>4</sub> /c + G <sub>30m</sub>	0,3445	(0,2963; 0,3927)	0,3981	1,0570	(1,0076; 1,1064)	0,9741
M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>30m</sub>	0,3830	(0,3529; 0,4131)	0,3839*	0,9970	(0,9843; 1,0097)	0,9970*
M/G <sub>24</sub> /c + G <sub>30m</sub>	0,3188	(0,2427; 0,3949)	0,3981	1,0329	(0,9959; 1,0699)	0,9970*
<b>Processo chegada:</b> Poisson – taxa $\lambda = 0,7318$ – média $E(\tau) = 1,3665$			<b>Processo serviço:</b> Taxa $\mu = 0,0323$ – média $E(S) = 30,9598$			
<b>Distribuições:</b> M ~ Exponencial(1,3665)			<b>Distribuições:</b> G <sub>24</sub> ~ Weibull (0,9276; 26,004; 3,1333) G <sub>26</sub> ~ Weibull (1,3021; 2,3952; 0) M <sub>4</sub> ~ Exponencial (30,9598)			
<b>Processo abandono:</b> taxa 0,01789 – média $E(R) = 55,9089$						
<b>Distribuições:</b> G <sub>25</sub> ~ Exponencial (55,9089)      G <sub>27</sub> ~ Fatigue Life (0,3726; 52,2792)      G <sub>28</sub> ~ Normal (55,9089; 445,3239) G <sub>29m</sub> ~ 0,8418×Fatigue Life (0,4055; 43,4875) + 0,1582×Fatigue Life (0,9434; 6,1219) G <sub>30m</sub> ~ 0,5221×Normal (29,1926; 232,5241) + 0,4779×Normal (26,7163; 212,7998) G <sub>31m</sub> ~ 0,9673×Exponencial (54,0799) + 0,0327×Exponencial (1,8289)						

Fonte: elaborada pelo autor



## 6 ANÁLISE DOS RESULTADOS

Este trabalho analisa a aplicação de modelos de fila com abandono para análise de congestão em *Call Centers* e investiga a utilização de distribuições genéricas (particularmente mistas) de probabilidades que melhor se ajustam aos tempos de paciência dos usuários de *Call Centers*. O interesse maior é considerar as distribuições mistas (Apêndice H) nos modelos analíticos de filas com abandono  $M/M/c+G$  e  $M/G/c+G$ , para serem aplicados nos problemas de congestão de *Call Centers*, obtendo medidas de desempenho que melhor expressem e representem a realidade.

Vários modelos de filas que incorporam o abandono têm sido estudados na literatura, mas consideram uma única classe de distribuições paramétricas para modelar a paciência, sendo as mais comuns a Exponencial, Weibull e Erlang. Entretanto, é comum encontrar em *Call Centers* uma população de usuários formada por grupos com características diferentes, cada um deles tendo comportamentos diferentes para o tempo de paciência, sendo uns mais pacientes que outros. Por esse motivo, em geral, parece mais razoável modelar a paciência por meio de distribuições mistas de probabilidade, que são mais sensíveis para captar essas características das subpopulações de usuários.

Observa-se que algumas distribuições mistas de probabilidade foram utilizadas em estudos da literatura para ajustar o tempo de paciência em *Call Centers*, mas não foram consideradas para validar na prática modelos analíticos de filas, aplicados a *Call Centers*, com dados extraídos da realidade. Neste estudo, analisa-se a aplicação desses modelos de filas, também considerando distribuições mistas para modelar a paciência dos usuários, em casos reais de *Call Centers*, com dados de uma empresa no Brasil e no exterior (Israel). A validação foi feita comparando-se os desvios das medidas de desempenho descritas nos Capítulos 4 e 5, obtidas por meio dos modelos analíticos de filas, com os dados reais e também com os resultados de simulação discreta. Para atingir esse objetivo, estudou-se e aplicou-se diferentes modelos analíticos de fila, encontrados na literatura, que admitiam distribuições exponenciais e genéricas no processo de serviço e de abandono.

Entre as medidas de desempenho consideradas importantes para os gestores, decidiu-se utilizar neste estudo a probabilidade de abandono,  $P(Ab)$ , para avaliar e comparar os modelos analíticos de fila que melhor represente o *Call Center*. O que motivou a escolha desta métrica é que por meio dela os usuários informam, subjetivamente, a sua percepção sobre os serviços recebidos. Outras medidas de desempenho também foram consideradas nas análises realizadas, como o tempo médio de espera dos usuários na fila,  $E(W)$ , a probabilidade de esperar,  $P(W)$ , e

a intensidade de tráfego,  $\rho$ , conforme Capítulos 4 e 5. Os resultados obtidos das análises do *Call Center* da Empresa e do *Call Center* do Banco estão descritos nas seções a seguir.

## 6.1 Resultados do caso do Call Center do banco

Conforme apresentado no Capítulo 4, após a análise estatística dos dados do *Call Center* do Banco, realizados com uso do software EasyFit® (2010), ao nível de 5% de probabilidade, obteve-se como resultado que a distribuição Exponencial de probabilidade não foi rejeitada para modelar o processo de chegada do *Call Center* do Banco, nos três horários considerados 11-12 horas, 12-13 horas e 15-16 horas, com *p-value* 0,08012; 0,66578 e 0,69484, respectivamente.

No entanto, conforme visto no Capítulo 4, no processo de serviço a distribuição Exponencial de probabilidade foi rejeitada nos três horários considerados, com *p-value* 0,000. A distribuição de probabilidade Fréchet mostrou-se mais ajustada para representar os tempos de serviço nos dois primeiros horários 11-12 horas e 12-13 horas, com *p-value* 0,87816 e 0,69548, respectivamente, sendo rejeitada no horário as 15-16 horas (*p-value* 0,000). Neste último horário, entre as distribuições ajustadas, optou-se pela Lognormal com *p-value* 0,45920.

Ainda conforme visto no Capítulo 4, no processo de abandono várias distribuições de probabilidade se ajustaram aos tempos de paciência dos usuários do *Call Center* do Banco, nos horários das 11-12 horas e 12-13 horas. Entre as possíveis optou-se aqui pela *Fatigue Life* (*p-value* 0,2321), Lognormal (*p-value* 0,04495,  $\alpha = 2\%$ ) e Exponencial (*p-value* 0,01538,  $\alpha = 1\%$ ) para representar os tempos de paciência no horário das 11-12 horas e *Fatigue Life* (*p-value* 0,1126), Lognormal (*p-value* 0,0425,  $\alpha = 2\%$ ), Exponencial (*p-value* 0,0258) para o horário das 12-13 horas. Todas as distribuições de probabilidade foram rejeitadas no horário das 15-16 horas. Mesmo assim, escolheu-se as distribuições *Fatigue Life* (*p-value* 0,0073), Lognormal (*p-value* 0,000) e Exponencial (*p-value* 0,000), para representar os tempos de paciência deste *Call Center*.

Distribuições mistas foram consideradas na elaboração desses modelos analíticos, para modelar o processo de abandono. Nos três horários, combinou-se a mistura de duas componentes, duas *Fatigue Life* (FL+FL), duas Lognormais (LogN+LogN) e duas Exponenciais (Ex+Ex). Embora existam métodos para determinar a quantidade de componentes das distribuições mistas, optou-se, neste estudo, pelo uso de duas componentes apenas como uma abordagem inicial do problema de pesquisa desta tese. Os parâmetros dessas distribuições



mistas, bem como o peso de cada uma delas, foram obtidos pelo método dos estimadores de máxima verossimilhança, calculados com o uso do software Mathematica (2013) (Apêndice H). Esses parâmetros foram recalculados utilizando-se do valor da média e da variância estimadas pelo estimador Kaplan-Meier. Esses valores estimados foram inseridos nas expressões da média e variância das componentes das distribuições mistas para obtenção dos novos parâmetros. Este procedimento foi necessário porque os tempos de paciência são dados censurados. Essas distribuições mistas com esses novos parâmetros é que foram utilizadas nos modelos analíticos de fila considerados neste estudo, para se obter as medidas de desempenho e posterior análise dos resultados.

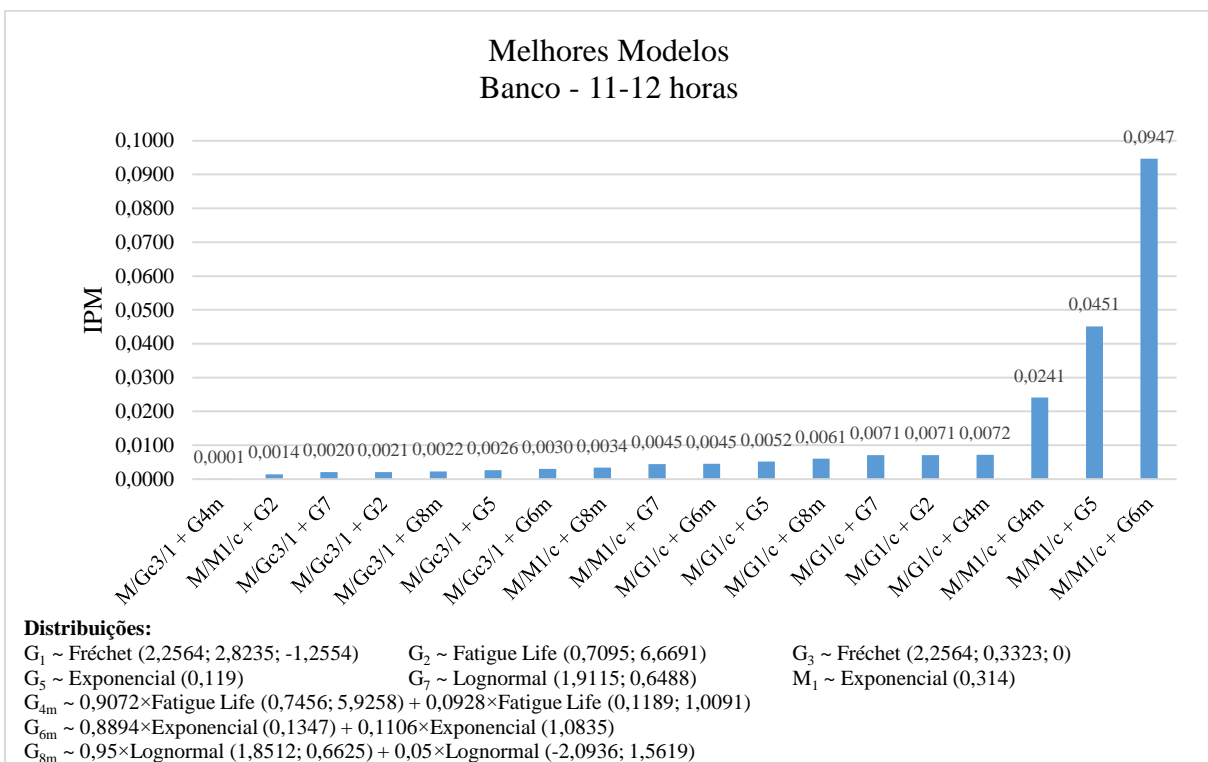
Desta análise estatística (Capítulo 4), também se extraiu as taxas de chegada, serviço e abandono em três horários (11 às 12 horas, 12 às 13 horas, 15 às 16 horas) e, obteve-se por meio dos dados, o tempo médio de espera, a probabilidade de espera, a probabilidade de abandono e a intensidade de tráfego. Esses parâmetros (taxas) foram inseridos nos modelos analíticos de fila considerados neste estudo e as medidas de desempenho foram calculadas e comparadas com aquelas extraídas dos dados reais.

Nessas comparações, extraiu-se um modelo analítico de fila considerado efetivo para representar, em cada um dos horários (11-12h, 12-13h e 15-16h), o *Call Center* do Banco. Para esta finalidade, calculou-se o desvio entre os valores das medidas de desempenho extraídas dos modelos analíticos de fila e dos dados reais. Em seguida, obteve-se o Índice de Precisão do Modelo (IPM) desses desvios e o modelo analítico de fila que possui o menor IPM foi considerado o mais eficiente para representar o *Call Center*, porque os resultados obtidos das medidas de desempenho deste modelo têm menores desvios em relação às mesmas medidas extraídas dos dados reais. A estatística IPM é a soma dos quadrados das diferenças entre o valor estimado ( $\hat{y}_i$ ) e o valor real dos dados ( $y_i$ ), ponderados pelo número de termos ( $n$ ) e pelo peso  $\alpha_i$  (eq. 6.1). Os pesos  $\alpha_i$  são atribuídos conforme a importância da medida de desempenho que ele representa. Neste estudo, atribuiu-se pesos iguais a 1 ( $\alpha_i = 1$ ) para todos os desvios obtidos, considerando-se que as medidas de desempenho possuem o mesmo nível de importância. O modelo que apresenta menor IPM é o melhor.

$$IPM = \frac{\sum_{i=1}^n \alpha_i (y_i - \hat{y}_i)^2}{n} \quad (6.1)$$

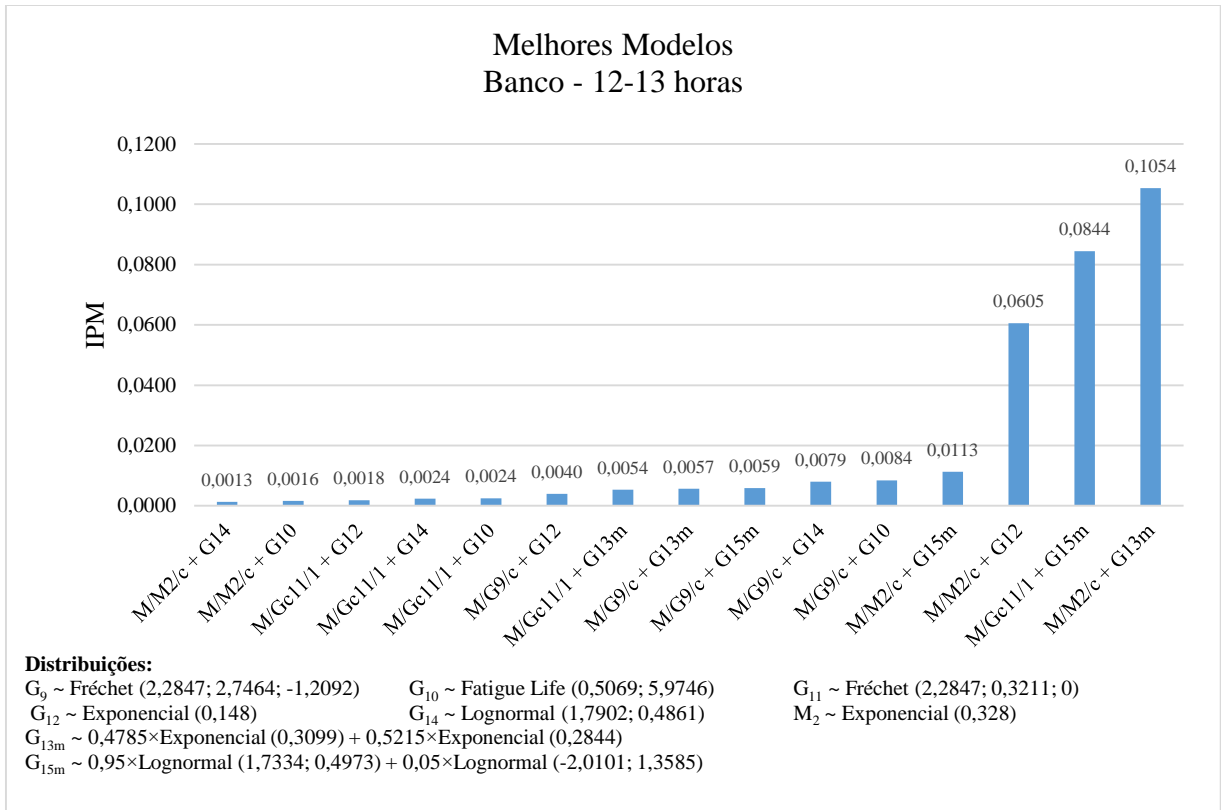
Os gráficos 6.1, 6.2 e 6.3 mostram os IPM dos modelos analíticos de fila utilizados em cada um dos horários considerados neste estudo.

GRÁFICO 6.1 - Comparação do desempenho dos modelos – 11-12 horas



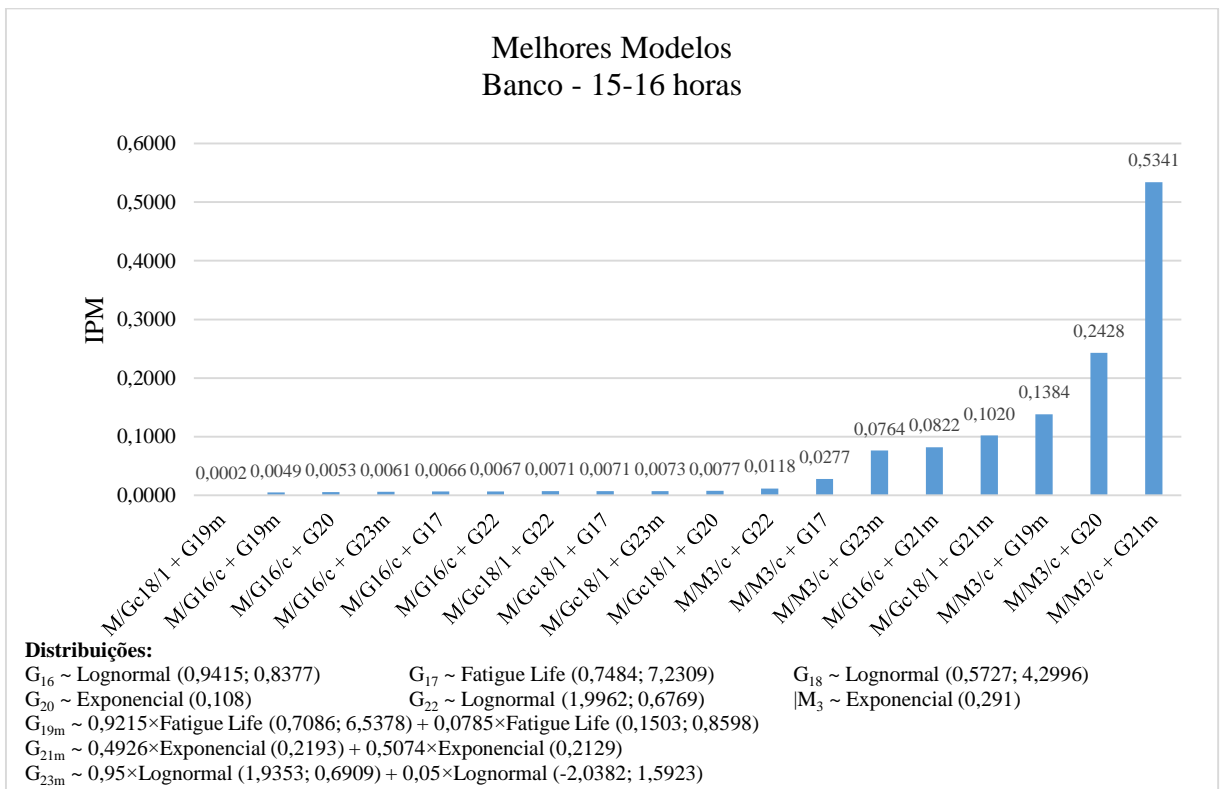
Fonte: elaborado pelo autor

GRÁFICO 6.2 - Comparação do desempenho dos modelos – 12-13 horas



Fonte: elaborado pelo autor

GRÁFICO 6.3 - Comparação do desempenho dos modelos – 15-16 horas



Fonte: elaborado pelo autor

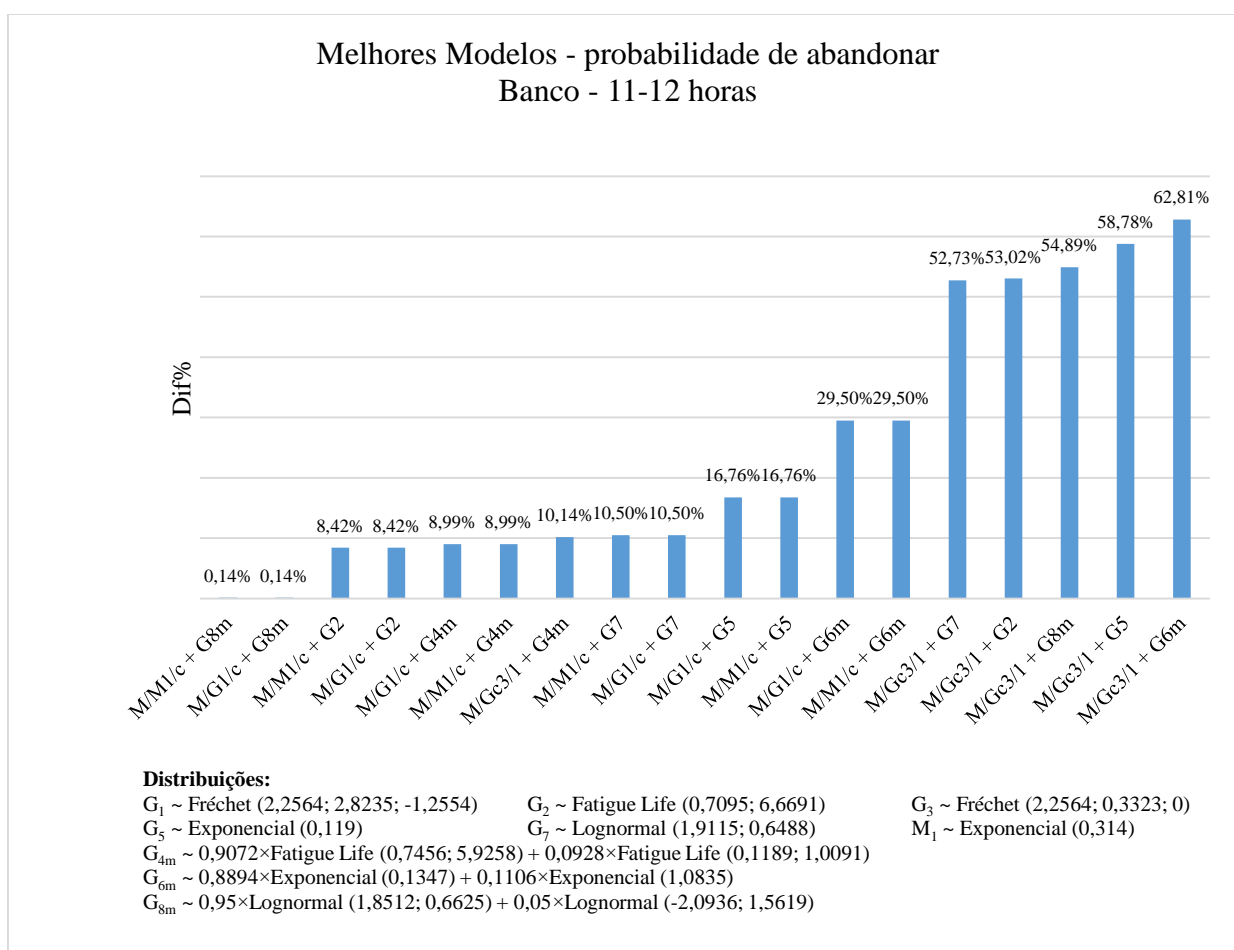
Conforme esses gráficos, cada horário possui em modelo analítico de fila mais eficiente para representar o *Call Center* do Banco. No horário das 11-12 horas, o modelo  $M/G^c_3/1+G_{4m}$ , com distribuição Fréchet para o serviço e distribuição mista *Fatigue Life* para a paciência, é o mais indicado. No horário das 12-13 horas, o modelo analítico de filas mais apropriado é o  $M/M_2/c+G_{14}$ , com serviço Exponencial e a paciência modelada com uma única distribuição Lognormal de probabilidade. O modelo analítico de fila que melhor representa o horário das 15-16 horas é o  $M/G^c_{18}/1+G_{9m}$ , com serviço Lognormal e a paciência com distribuição mista *Fatigue Life*. Nestes casos, as distribuições mistas mostraram-se eficientes para representar os tempos de paciência.

Existem modelos analíticos de fila que produzem resultados melhores em algumas medidas de desempenho do que em outras. Esse fato motivou encontrar o modelo analítico de fila mais apropriado para representar o *Call Center*, para cada uma das medidas de desempenho em cada um dos horários analisados. Considerou-se nesta análise a métrica “Diferença Percentual” (*Dif%*) definida pela Equação (6.2), em que  $V$  é o valor da medida de desempenho obtido pelo modelo analítico de fila e  $R$ , é o valor da medida de desempenho real:

$$Dif\% = \frac{|V - R|}{R} \quad (6.2)$$

Inicialmente, considerou-se a probabilidade de abandonar  $P(A)$  como a medida de desempenho com maior importância para os gestores. Essas diferenças percentuais dessas probabilidades estão exibidas nos Gráficos 6.4, 6.5 e 6.6 para cada um dos horários considerados neste estudo.

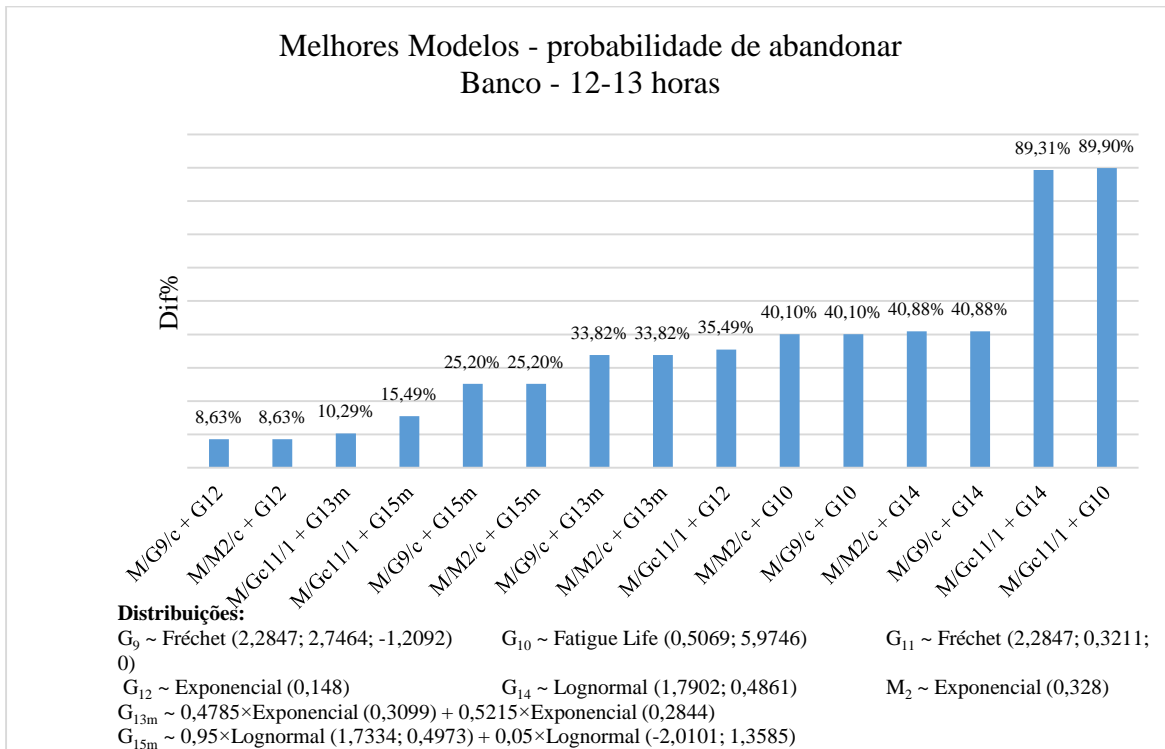
GRÁFICO 6.4 - Comparação do desempenho dos modelos – 11-12 horas



Fonte: elaborado pelo autor

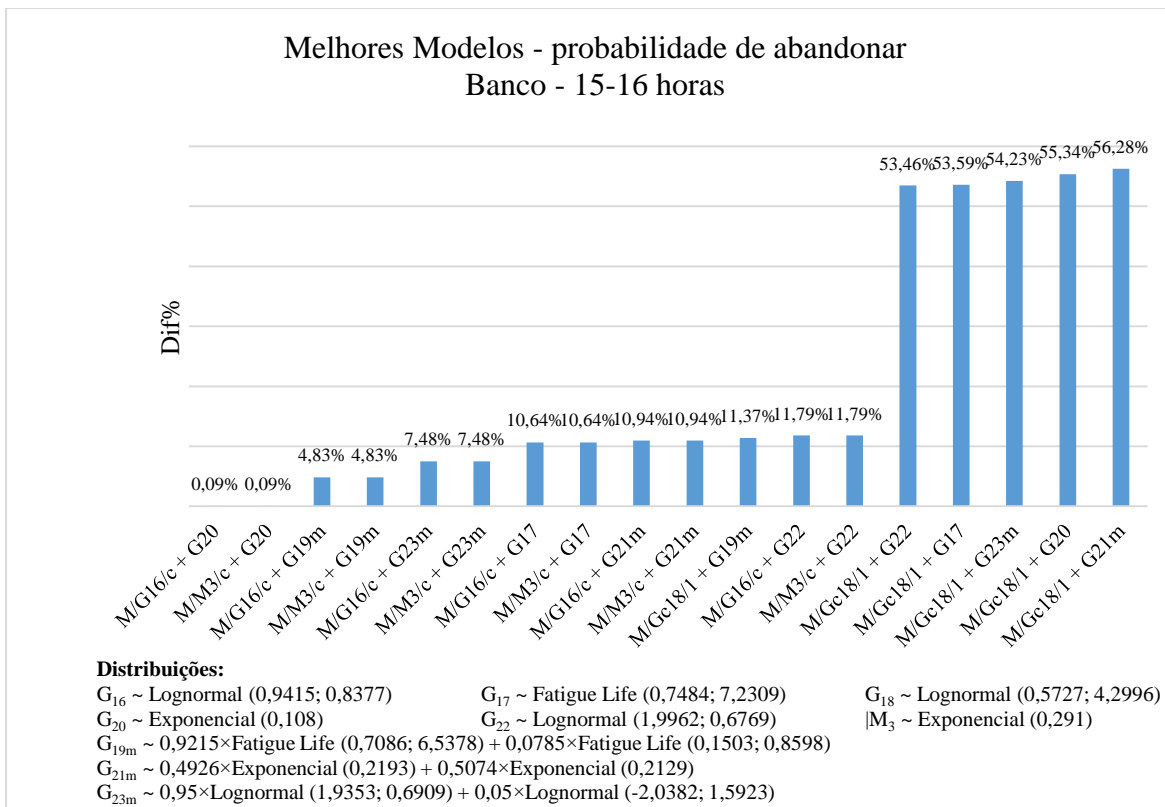
Em cada horário existe um modelo analítico de fila que melhor representa o *Call Center*. No horário das 11-12 horas, o modelo  $M/M_1/c+G_{8m}$ , com distribuição mista Lognormal para a paciência e distribuição Exponencial para os tempos de serviço, e o modelo  $M/G_1/c+G_{8m}$ , com a mesma distribuição mista para a paciência e distribuição Fréchet para o serviço, são os mais acurados. Para o horário 12-13 horas, os modelos  $M/G_9/c+G_{12}$ , com distribuição Fréchet para serviço e com uma única distribuição Exponencial para a paciência, e o modelo  $M/M_2/c+G_{12}$ , também com uma única distribuição Exponencial para a paciência e serviço, são adequados para representar o *Call Center*.

GRÁFICO 6.5 - Comparação do desempenho dos modelos – 12-13 horas



Fonte: elaborado pelo autor

GRÁFICO 6.6 - Comparação do desempenho dos modelos – 15-16 horas

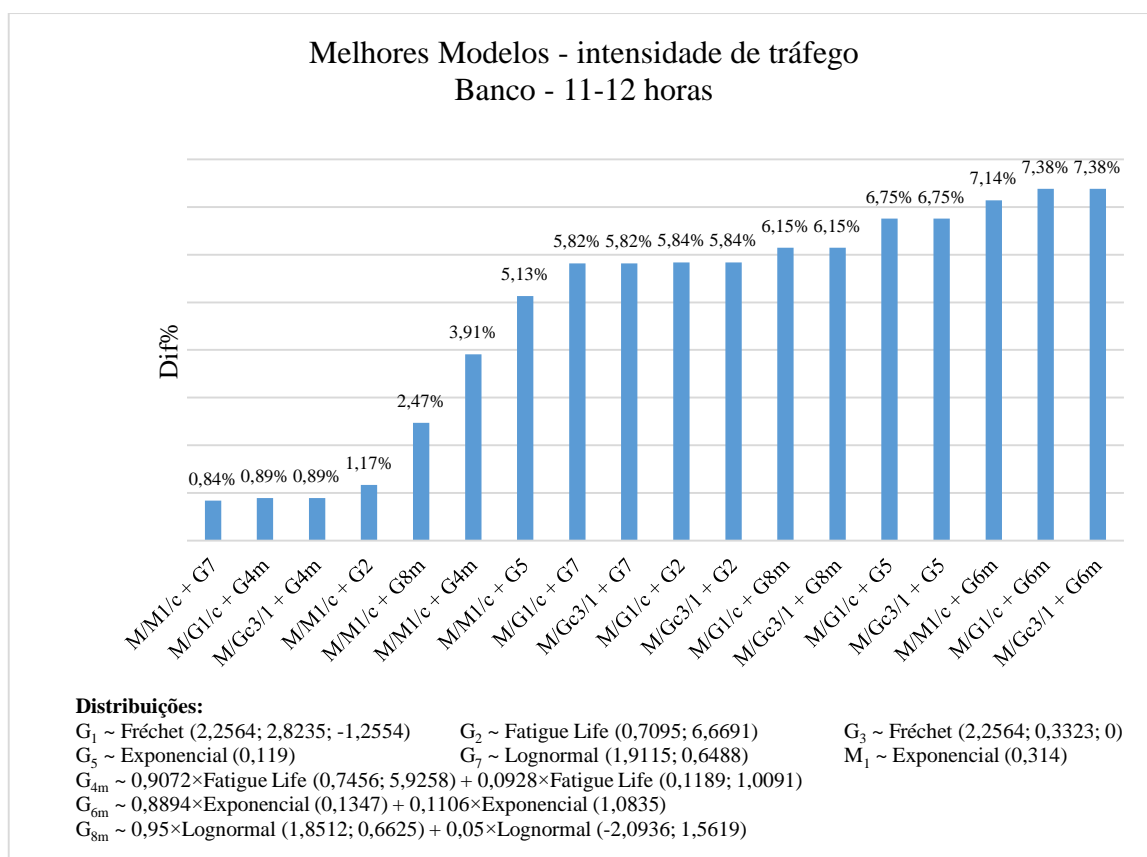


Fonte: elaborado pelo autor

Considerando o horário das 15-16 horas, o modelo com serviço Lognormal e com paciência Exponencial,  $M/G_{16}/c+G_{20}$ , e o modelo com serviço e paciência Exponenciais,  $M/M_3/c+G_{20}$ , representam efetivamente o *Call Center* do Banco. Destaca-se desses resultados que a distribuição mista para a paciência é mais adequada do que as distribuições não mistas para representar a probabilidade de abandonar somente no horário das 11-12 horas.

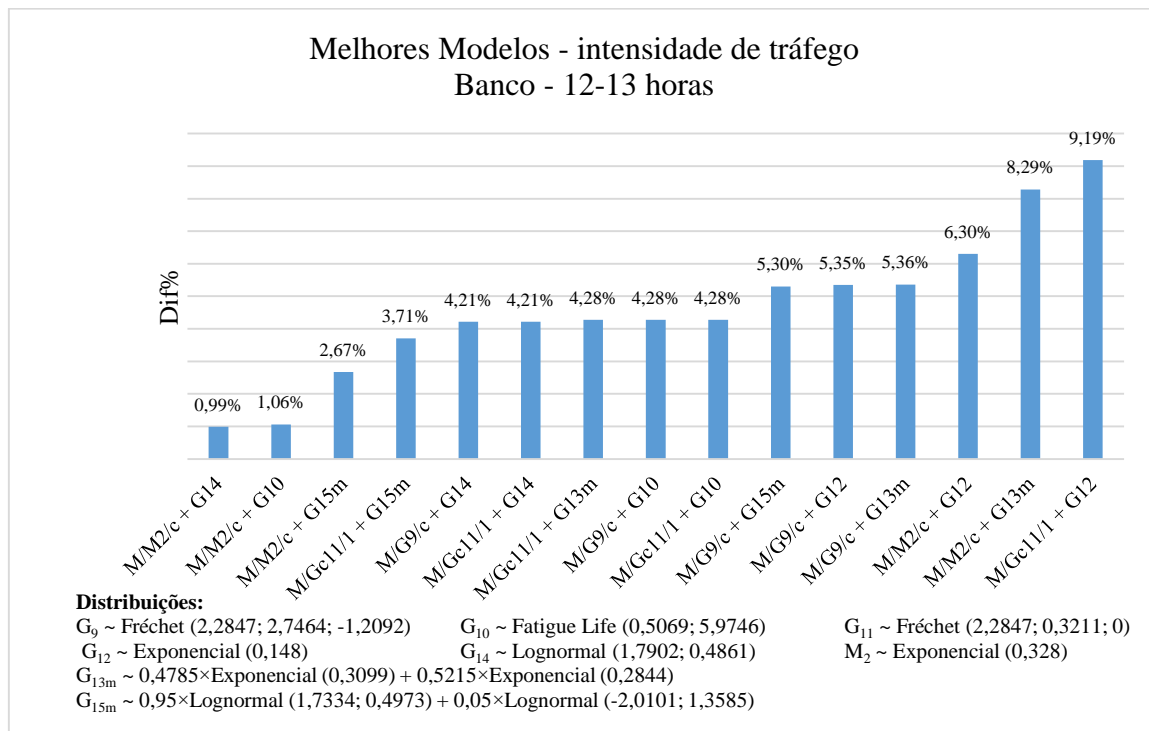
Considerando a medida de desempenho intensidade de tráfego, os Gráficos 6.7, 6.8 e 6.9 mostram que nos horários considerados o modelo que melhor representa esta medida é o  $M/M_1/c+G_7$ , caracterizado pela distribuição Exponencial para os tempos de serviço e uma única distribuição paramétrica Lognormal para modelar os tempos de paciência.

GRÁFICO 6.7 - Comparação do desempenho dos modelos – 11-12 horas



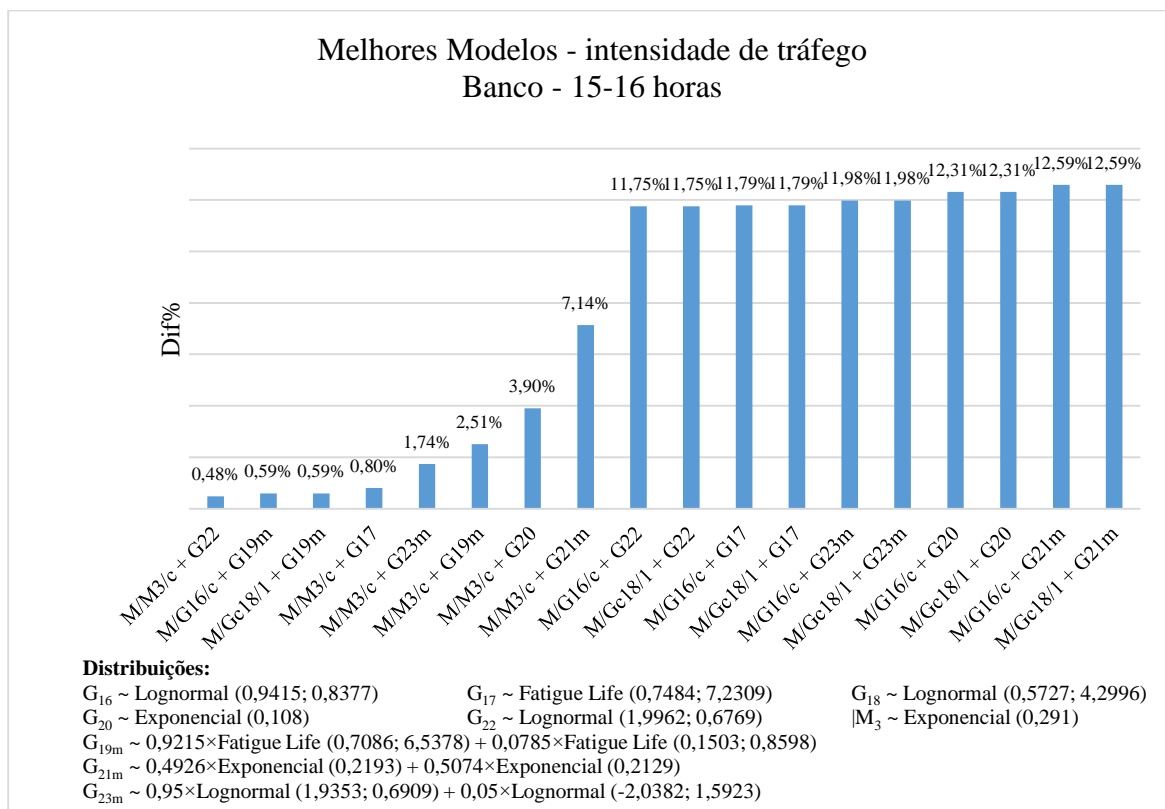
Fonte: elaborado pelo autor

GRÁFICO 6.8 - Comparação do desempenho dos modelos – 12-13 horas



Fonte: elaborado pelo autor

GRÁFICO 6.9 - Comparação do desempenho dos modelos – 15-16 horas

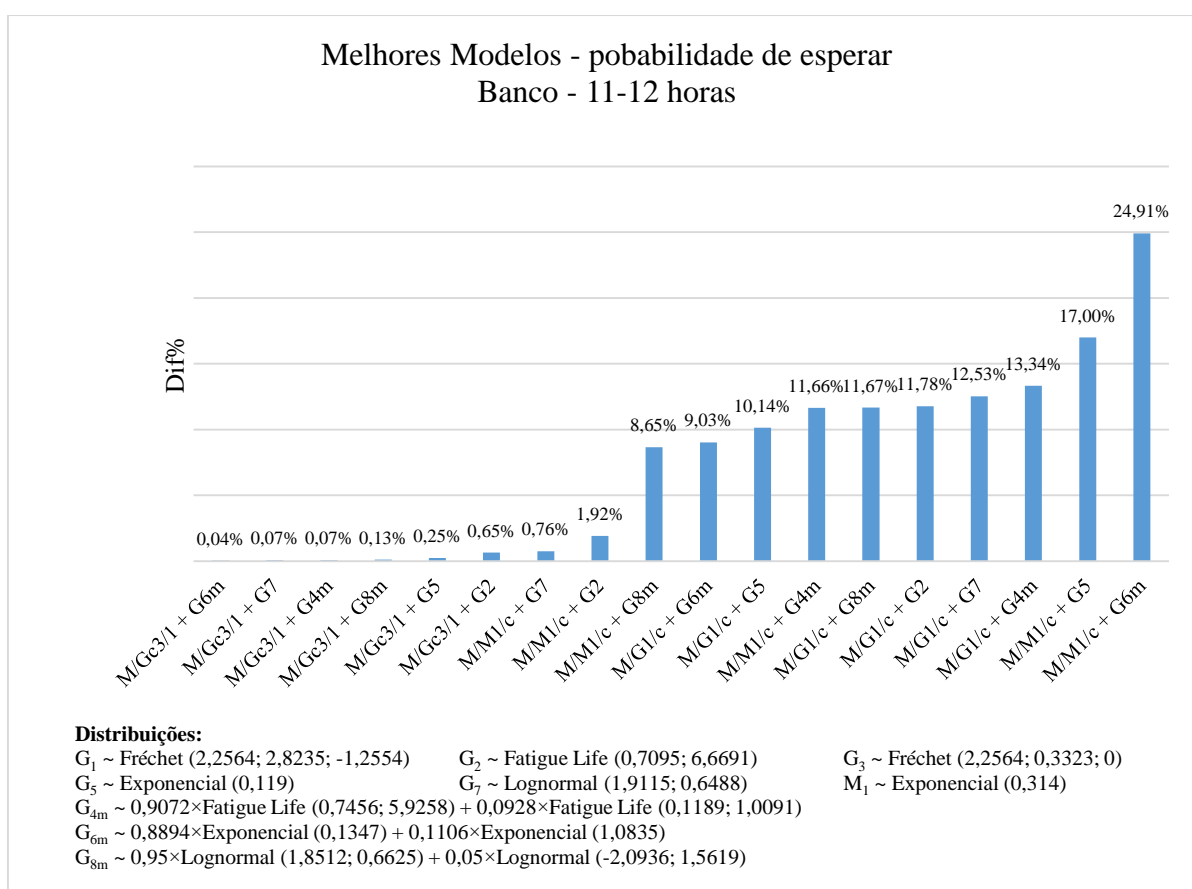


Fonte: elaborado pelo autor



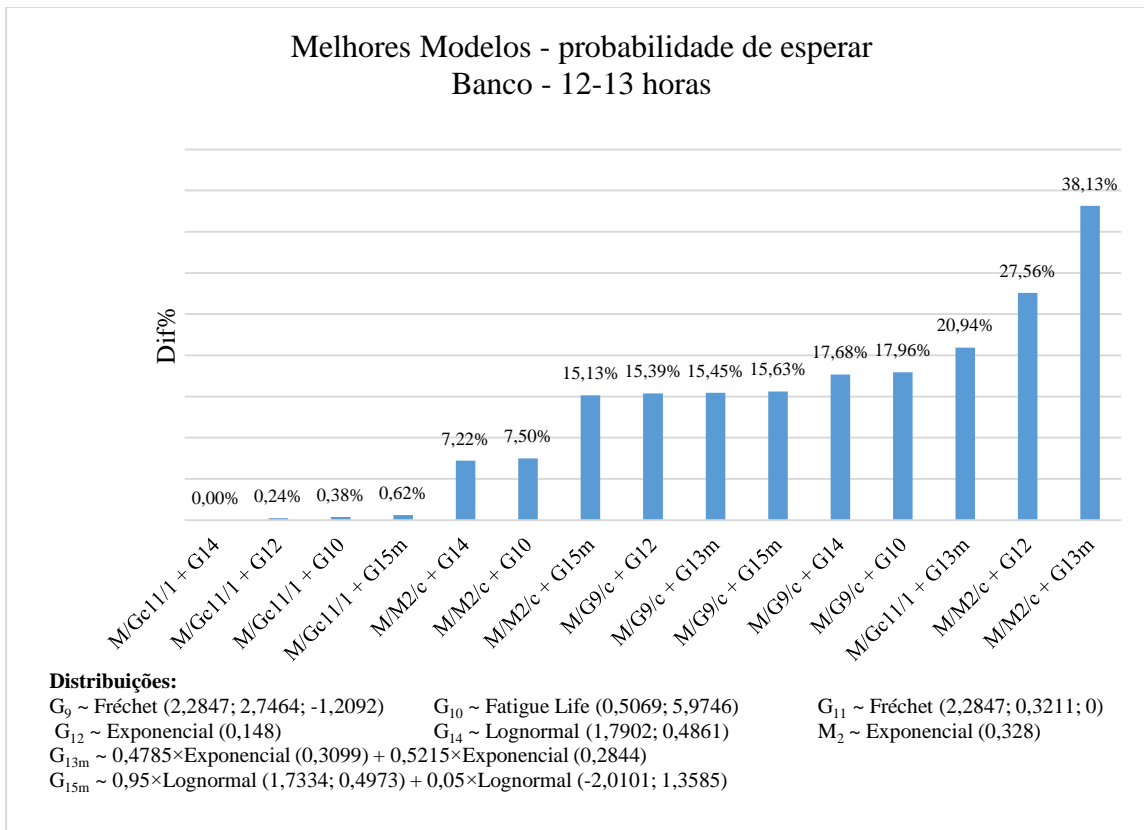
A probabilidade de esperar é melhor estimada no horário das 11-12 horas pelo modelo  $M/G^c_3/1+G_{6m}$ , formado por uma distribuição Exponencial mista para representar os tempos de paciência e por uma distribuição Fréchet para os tempos de serviço. Uma distribuição não mista Lognormal para representar os tempos de paciência no modelo  $M/G^c_{11}/1+G_{14}$ , com distribuição Fréchet para os tempos de serviço, é o mais preciso para modelar a probabilidade de esperar no horário das 12-13 horas. No último horário, das 15-16 horas, a probabilidade de esperar fica melhor estimada pelo modelo  $M/G^c_1/1+G_{23m}$ , com distribuição Lognormal mista para a paciência e também distribuição Lognormal para os tempos de serviço, conforme mostram os gráficos 6.10, 6.11 e 6.12.

GRÁFICO 6.10 - Comparação do desempenho dos modelos – 11-12 horas



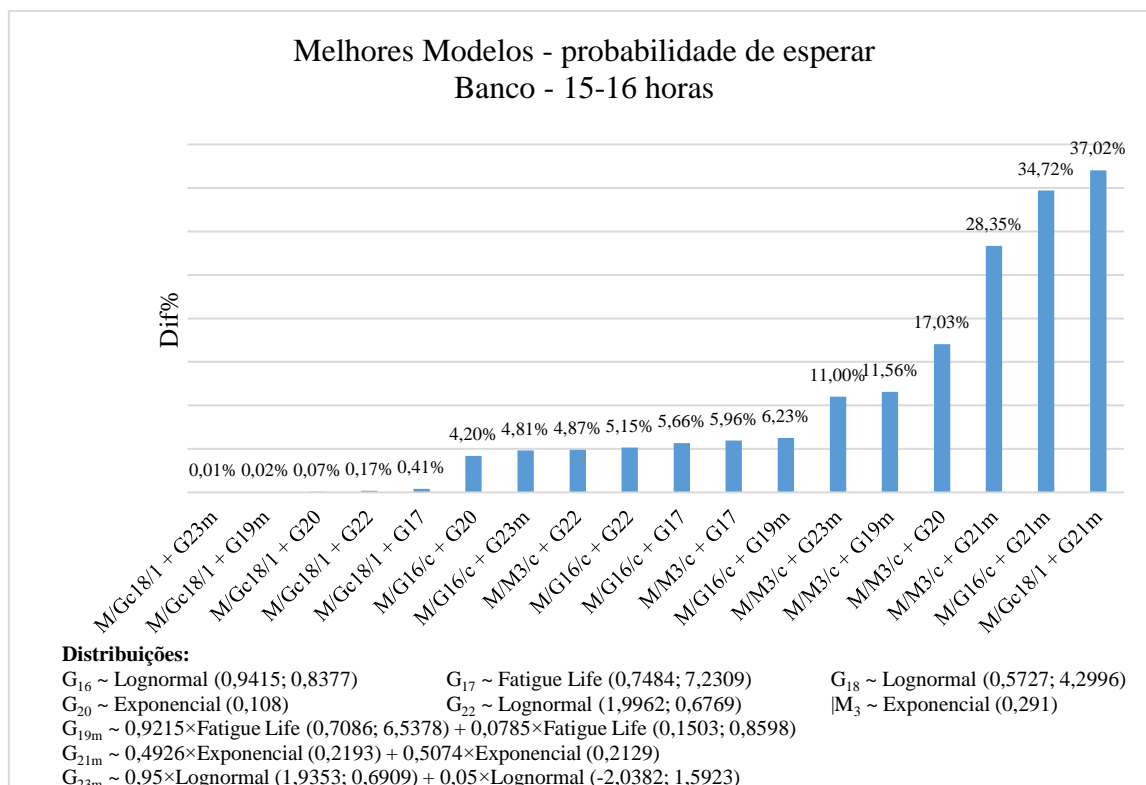
Fonte: elaborado pelo autor

GRÁFICO 6.11 - Comparação do desempenho dos modelos – 12-13 horas



Fonte: elaborado pelo autor

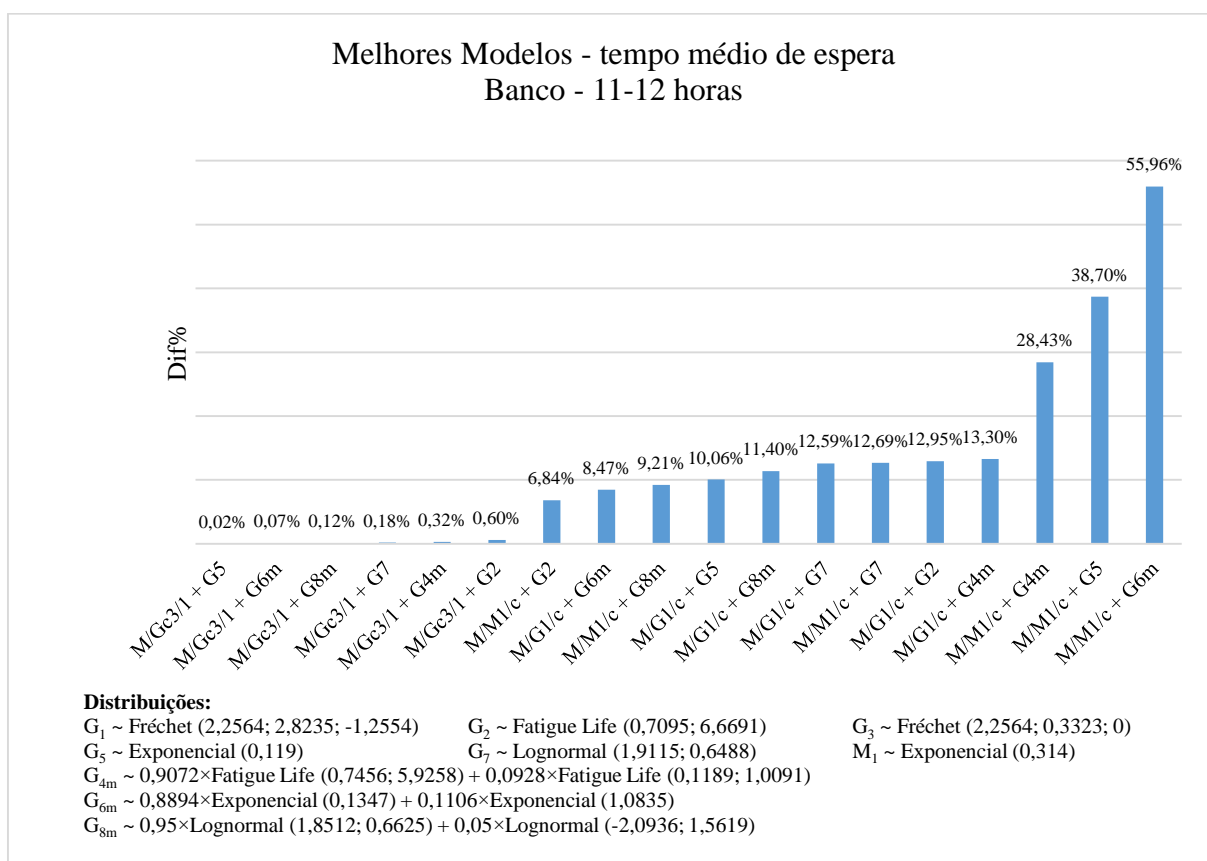
GRÁFICO 6.12 - Comparação do desempenho dos modelos – 15-16 horas



Fonte: elaborado pelo autor

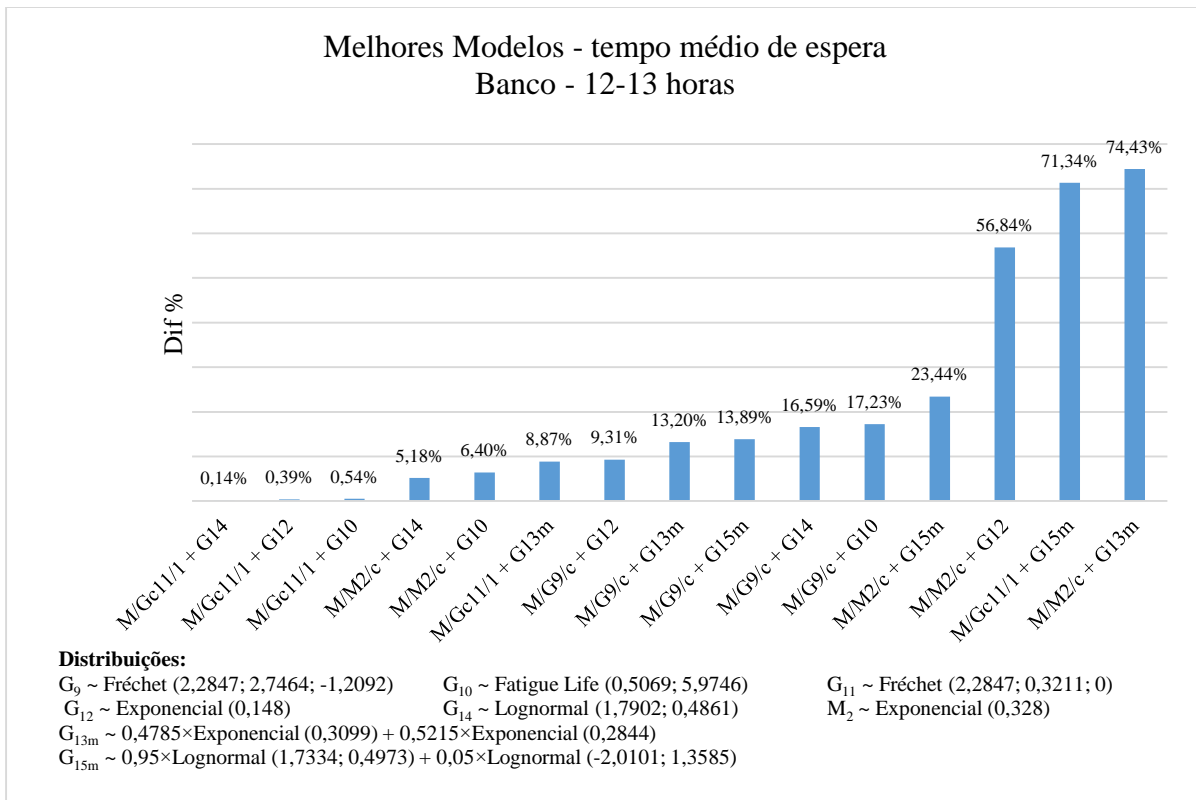
Os modelos analíticos de fila que melhor se adequam para estimar o tempo de espera do *Call Center* do Banco não apresentam distribuição mista para representar os tempos de paciência e são todos do mesmo tipo, diferenciando-se apenas pelas distribuições que representam os processos de serviço e abandono. O modelo  $M/G^c_3/1+G_5$ , com distribuição Exponencial para a paciência e Fréchet para o serviço, é o mais indicado para o horário das 11-12 horas. O modelo  $M/G^c_{11}/1+G_{14}$ , com distribuição Fréchet e Lognormal para representar os tempos de serviço e paciência, respectivamente, é o mais representativo para o horário das 12-13 horas. No último horário, das 15-16 horas, o modelo  $M/G^c_{18}/1+G_{22}$ , com distribuição Lognormal para o processo de serviço e paciência, é o mais indicado. Esses resultados podem ser observados nos Gráficos 6.13, 6.14 e 6.15.

GRÁFICO 6.13 - Comparação do desempenho dos modelos – 11-12 horas



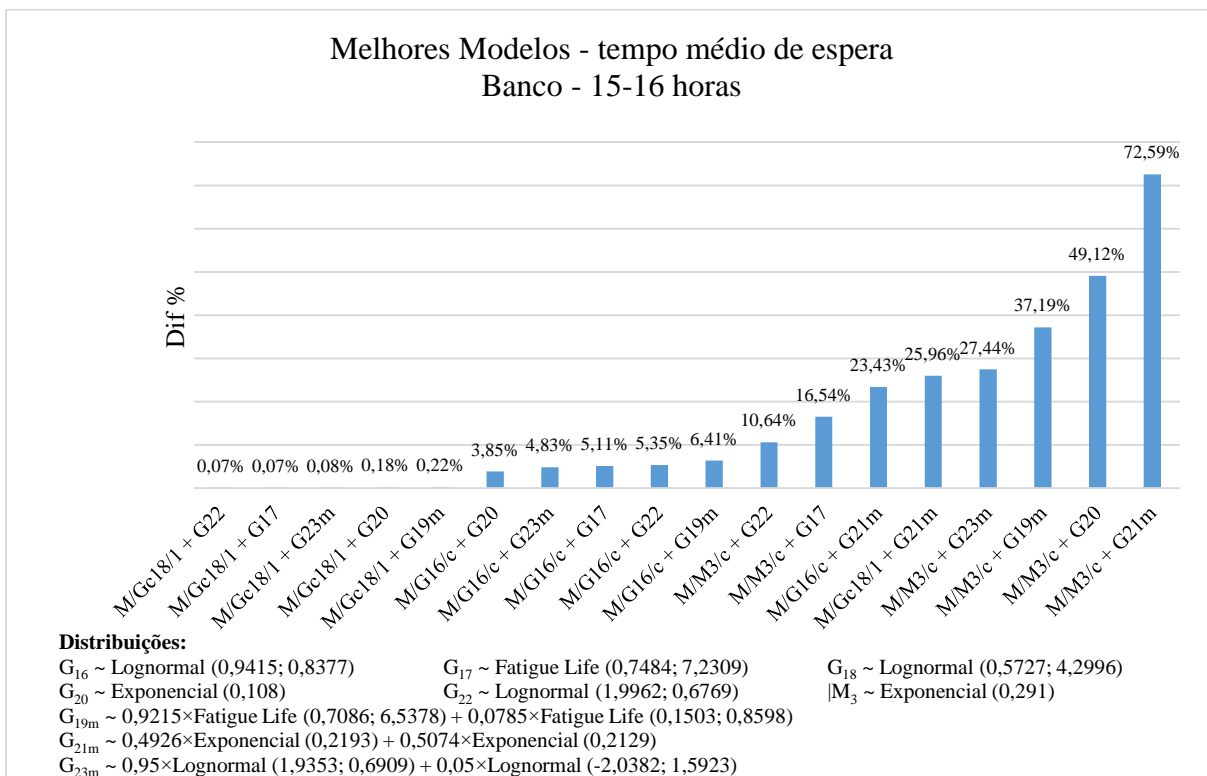
Fonte: elaborado pelo autor

GRÁFICO 6.14 - Comparação do desempenho dos modelos – 13-14 horas



Fonte: elaborado pelo autor

GRÁFICO 6.15: Comparação do desempenho dos modelos – 15-16 horas



Fonte: elaborado pelo autor

Ainda como resultado, os modelos analíticos de fila considerados adequados para representar cada uma das medidas de desempenho do *Call Center* do Banco, possuem diferenças percentuais inferiores a 1% e, portanto, são também considerados válidos para a sua finalidade. Como outros resultados, observa-se que os modelos analíticos de fila com distribuição mista para a paciência, em geral, se mostraram eficientes para representar a probabilidade de abandonar no horário das 11-12 horas e a probabilidade de esperar nos horários 11-12 horas e 15-16 horas. Além disso, esses modelos com distribuições mistas para a paciência também são efetivos para representar as medidas intensidade de tráfego e tempo médio de espera, embora não sejam os mais precisos.

Dependendo do tipo de modelo analítico de fila adotado para representar o *Call Center* do Banco, em cada um dos horários adotados neste estudo e independentemente de qualquer medida de desempenho, os modelos que utilizam distribuições mistas para a paciência, em alguns casos, são melhores que os seus correspondentes modelos com uma única classe de distribuição paramétrica. No horário das 11-12 horas, a Tabela 6.1 mostra que os modelos analíticos de fila  $M/G/c+G$ , com distribuição mista Exponencial e com distribuição mista Lognormal para os tempos de paciência, são melhores que os seus correspondentes modelos com distribuições não mistas (menores IPM). No caso dos modelos  $M/G^c/1+G$ , apenas aquele com distribuição Fatigue Life mista para a paciência é melhor que o seu correspondente não misto. O modelo  $M/M/c+G$ , com distribuição mista Lognormal, comporta-se melhor que o seu correspondente não misto para os tempos de paciência.

Observando-se a Tabela 6.2, no horário das 12-13 horas, nos modelos  $M/G/c+G$ , somente aquele com distribuição mista Lognormal para a paciência é melhor que seu correspondente não misto. Nos modelos  $M/G^c/1+G$  e  $M/M/c+G$ , com distribuição mista para a paciência, nenhum se comportou melhor que o seu correspondente não misto. Analisando-se o horário das 15-16 horas (Tabela 6.3) os modelos  $M/G/c+G$  com distribuição mista Lognormal e com distribuição mista Fatigue Life para a paciência se comportam melhores que os seus correspondentes não mistos. Apenas o modelo  $M/G^c/1+G$  com distribuição mista Fatigue Life para os tempos de paciência é melhor que seu correspondente não misto e nenhum dos modelos  $M/M/c+G$  com distribuição mista para a paciência se comporta melhor que seus correspondentes não mistos.

TABELA 6.1 - Comparando paciência com distribuição mista e não mista – Caso Banco-11-12 horas

Modelos		Modelos		Modelos	
M/G/c+G	IPM	M/G <sup>c</sup> /1+G	IPM	M/M/c+G	IPM
M/G <sub>1</sub> /c + G <sub>6m</sub>	0,0045	M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>4m</sub>	0,0001	M/M <sub>1</sub> /c + G <sub>2</sub>	0,0014
M/G <sub>1</sub> /c + G <sub>5</sub>	0,0052	M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>7</sub>	0,0020	M/M <sub>1</sub> /c + G <sub>8m</sub>	0,0034
M/G <sub>1</sub> /c + G <sub>8m</sub>	0,0061	M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>2</sub>	0,0021	M/M <sub>1</sub> /c + G <sub>7</sub>	0,0045
M/G <sub>1</sub> /c + G <sub>7</sub>	0,0071	M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>8m</sub>	0,0022	M/M <sub>1</sub> /c + G <sub>4m</sub>	0,0241
M/G <sub>1</sub> /c + G <sub>2</sub>	0,0071	M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>5</sub>	0,0026	M/M <sub>1</sub> /c + G <sub>5</sub>	0,0451
M/G <sub>1</sub> /c + G <sub>4m</sub>	0,0072	M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>6m</sub>	0,0030	M/M <sub>1</sub> /c + G <sub>6m</sub>	0,0947
<b>Processo chegada:</b> Poisson – taxa $\lambda = 1,933$ – média $E(\tau) = 0,5173$ <b>Distribuições:</b> M ~ Exponencial(1,933)			<b>Processo serviço:</b> Taxa $\mu = 0,314$ – média $E(S) = 3,1847$ <b>Distribuições:</b> G <sub>1</sub> ~ Fréchet (2,2564; 2,8235; -1,2554) G <sub>3</sub> ~ Fréchet (2,2564; 0,3323; 0) M <sub>1</sub> ~ Exponencial (0,314)		
<b>Processo abandono:</b> taxa 0,119 – média $E(R) = 8,4034$ <b>Distribuições:</b> G <sub>2</sub> ~ Fatigue Life (0,7095; 6,6691)    G <sub>5</sub> ~ Exponencial (0,119)    G <sub>7</sub> ~ Lognormal (1,9115; 0,6488) G <sub>4m</sub> ~ 0,9072×Fatigue Life (0,7456; 5,9258) + 0,0928×Fatigue Life (0,1189; 1,0091) G <sub>6m</sub> ~ 0,8894×Exponencial (0,1347) + 0,1106×Exponencial (1,0835) G <sub>8m</sub> ~ 0,95×Lognormal (1,8512; 0,6625) + 0,05×Lognormal (-2,0936; 1,5619)					

Fonte: elaborado pelo autor

TABELA 6.2 - Comparando paciência com distribuição mista e não mista – Caso Banco-12-13 horas

Modelos		Modelos		Modelos	
M/G/c+G	IPM	M/G <sup>c</sup> /1+G	IPM	M/M/c+G	IPM
M/G <sub>9</sub> /c + G <sub>12</sub>	0,0040	M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>12</sub>	0,0018	M/M <sub>2</sub> /c + G <sub>14</sub>	0,0013
M/G <sub>9</sub> /c + G <sub>13m</sub>	0,0057	M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>14</sub>	0,0024	M/M <sub>2</sub> /c + G <sub>10</sub>	0,0016
M/G <sub>9</sub> /c + G <sub>15m</sub>	0,0059	M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>10</sub>	0,0024	M/M <sub>2</sub> /c + G <sub>15m</sub>	0,0113
M/G <sub>9</sub> /c + G <sub>14</sub>	0,0079	M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>13m</sub>	0,0054	M/M <sub>2</sub> /c + G <sub>12</sub>	0,0605
M/G <sub>9</sub> /c + G <sub>10</sub>	0,0084	M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>15m</sub>	0,0844	M/M <sub>2</sub> /c + G <sub>13m</sub>	0,1054
<b>Processo chegada:</b> Poisson – taxa $\lambda = 1,717$ – média $E(\tau) = 0,5824$ <b>Distribuições:</b> M ~ Exponencial(1,717)			<b>Processo serviço:</b> Taxa $\mu = 0,328$ – média $E(S) = 3,0488$ <b>Distribuições:</b> G <sub>9</sub> ~ Fréchet (2,2847; 2,7464; -1,2092) G <sub>11</sub> ~ Fréchet (2,2847; 0,3211; 0) M <sub>2</sub> ~ Exponencial (0,328)		
<b>Processo abandono:</b> taxa 0,148 – média $E(R) = 6,7568$ <b>Distribuições:</b> G <sub>10</sub> ~ Fatigue Life (0,5069; 5,9746)    G <sub>12</sub> ~ Exponencial (0,148)    G <sub>14</sub> ~ Lognormal (1,7902; 0,4861) G <sub>13m</sub> ~ 0,4785×Exponencial (0,3099) + 0,5215×Exponencial (0,2844) G <sub>15m</sub> ~ 0,95×Lognormal (1,7334; 0,4973) + 0,05×Lognormal (-2,0101; 1,3585)					

Fonte: elaborado pelo autor

TABELA 6.3 - Comparando paciência com distribuição mista e não mista – Caso Banco-15-16 horas

Modelos		Modelos		Modelos	
M/G/c+G	IPM	M/G <sup>c</sup> /1+G	IPM	M/M/c+G	IPM
M/G <sub>16</sub> /c + G <sub>19m</sub>	0,0049	M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>19m</sub>	0,0002	M/M <sub>3</sub> /c + G <sub>22</sub>	0,0118
M/G <sub>16</sub> /c + G <sub>20</sub>	0,0053	M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>22</sub>	0,0071	M/M <sub>3</sub> /c + G <sub>17</sub>	0,0277
M/G <sub>16</sub> /c + G <sub>23m</sub>	0,0061	M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>17</sub>	0,0071	M/M <sub>3</sub> /c + G <sub>23m</sub>	0,0764
M/G <sub>16</sub> /c + G <sub>17</sub>	0,0066	M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>23m</sub>	0,0073	M/M <sub>3</sub> /c + G <sub>19m</sub>	0,1384
M/G <sub>16</sub> /c + G <sub>22</sub>	0,0067	M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>20</sub>	0,0077	M/M <sub>3</sub> /c + G <sub>20</sub>	0,2428
M/G <sub>16</sub> /c + G <sub>21m</sub>	0,0822	M/G <sup>c</sup> <sub>18</sub> /1 + G <sub>21m</sub>	0,1020	M/M <sub>3</sub> /c + G <sub>21m</sub>	0,5341
<b>Processo chegada:</b> Poisson – taxa $\lambda = 2,1$ – média $E(\tau) = 0,4762$			<b>Processo serviço:</b> Taxa $\mu = 0,291$ – média $E(S) = 3,4364$		
<b>Distribuições:</b> M ~ Exponencial(2,1)			<b>Distribuições:</b> G <sub>16</sub> ~ Lognormal (0,9415; 0,8377) G <sub>18</sub> ~ Lognormal (0,5727; 4,2996) M <sub>3</sub> ~ Exponencial (0,291)		
<b>Processo abandono:</b> taxa 0,108 – média $E(R) = 9,2593$					
<b>Distribuições:</b> G <sub>17</sub> ~ Fatigue Life (0,7484; 7,2309)      G <sub>20</sub> ~ Exponencial (0,108)      G <sub>22</sub> ~ Lognormal (1,9962; 0,6769) G <sub>19m</sub> ~ 0,9215×Fatigue Life (0,7086; 6,5378) + 0,0785×Fatigue Life (0,1503; 0,8598) G <sub>21m</sub> ~ 0,4926×Exponencial (0,2193) + 0,5074×Exponencial (0,2129) G <sub>23m</sub> ~ 0,95×Lognormal (1,9353; 0,6909) + 0,05×Lognormal (-2,0382; 1,5923)					

Fonte: elaborado pelo autor

Em resumo, considerando o sistema *Call Center* do Banco, independentemente de qualquer medida de desempenho, os modelos analíticos de fila com distribuições mistas para representar os tempos de paciência são mais efetivos nos horários das 11-12 horas e das 15-16 horas. No horário intermediário das 12-13 horas, as distribuições mistas para a paciência não foram as mais adequadas (Gráficos 6.1, 6.2 e 6.3).

Com o propósito de fornecer uma visão geral dos modelos analíticos de fila que são mais eficientes para representar o *Call Center* do Banco, organizou-se um sumário desses modelos exibidos nos Quadros 6.1 e 6.2. O Quadro 6.1 destaca o modelo analítico mais efetivo, em cada um dos horários analisados (11-12h, 12-13h e 15-16h), para representar o *Call Center* do Banco.

QUADRO 6.1 - Modelo analítico de fila mais efetivo para representar o *Call Center* do Banco

<b>CALL CENTER DO BANCO</b>	<b>Distribuições de probabilidade:</b>
Horário 11-12 horas: M/G <sub>3</sub> <sup>c</sup> /1+G <sub>4m</sub>	G <sub>3</sub> ~ Fréchet (2,2564; 0,3323; 0) G <sub>14</sub> ~ Lognormal (1,7902; 0,4861)
Horário 12-13 horas: M/M <sub>2</sub> /c+G <sub>14</sub>	G <sub>18</sub> ~ Lognormal (0,5727; 4,2996) M <sub>2</sub> ~ Exponencial (0,328)
Horário 15-16 horas: M/G <sub>18</sub> <sup>c</sup> /1+G <sub>19m</sub>	G <sub>4m</sub> ~ 0,9072×Fatigue Life (0,7456; 5,9258) + 0,0928×Fatigue Life (0,1189; 1,0091) G <sub>19m</sub> ~ 0,9215×Fatigue Life (0,7086; 6,5378) + 0,0785×Fatigue Life (0,1503; 0,8598)

Fonte: elaborado pelo autor

O Quadro 6.2 mostra os dois modelos analíticos de fila mais apropriados para representar o *Call Center* do Banco, para cada uma das medidas de desempenho, em cada um dos horários analisados.

QUADRO 6.2-Modelo analítico de fila mais apropriado para medidas de desempenho  
*Call Center* Banco

<b>TEMPO MÉDIO DE ESPERA:</b>	<b>Distribuições de probabilidade:</b>
Horário 11-12 horas: $M/G_3^c/1+G_5$ $M/G_3^c/1+G_{6m}$	$G_3 \sim$ Fréchet (2,2564; 0,3323; 0) $G_5 \sim$ Exponencial (0,119) $G_{11} \sim$ Fréchet (2,2847; 0,3211; 0) $G_{12} \sim$ Exponencial (0,148)
Horário 12-13 horas: $M/G_{11}^c/1+G_{14}$ $M/G_{11}^c/1+G_{12}$	$G_{14} \sim$ Lognormal (1,7902; 0,4861) $G_{17} \sim$ Fatigue Life (0,7484; 7,2309) $G_{18} \sim$ Lognormal (0,5727; 4,2996)
Horário 15-16 horas: $M/G_{18}^c/1+G_{22}$ $M/G_{18}^c/1+G_{17}$	$G_{22} \sim$ Lognormal (1,9962; 0,6769) $G_{6m} \sim$ 0,8894×Exponencial (0,1347) + 0,1106×Exponencial (1,0835)
<b>PROBABILIDADE ESPERAR:</b>	<b>Distribuições de probabilidade:</b>
Horário 11-12 horas: $M/G_3^c/1+G_{6m}$ $M/G_3^c/1+G_7$	$G_3 \sim$ Fréchet (2,2564; 0,3323; 0) $G_7 \sim$ Lognormal (1,9115; 0,6488) $G_{11} \sim$ Fréchet (2,2847; 0,3211; 0) $G_{12} \sim$ Exponencial (0,148)
Horário 12-13 horas: $M/G_{11}^c/1+G_{14}$ $M/G_{11}^c/1+G_{12}$	$G_{14} \sim$ Lognormal (1,7902; 0,4861) $G_{18} \sim$ Lognormal (0,5727; 4,2996) $G_{6m} \sim$ 0,8894×Exponencial (0,1347) + 0,1106×Exponencial (1,0835)
Horário 15-16 horas: $M/G_{18}^c/1+G_{23m}$ $M/G_{18}^c/1+G_{19m}$	$G_{19m} \sim$ 0,9215×Fatigue Life (0,7086; 6,5378) + 0,0785×Fatigue Life (0,1503; 0,8598) $G_{23m} \sim$ 0,95×Lognormal (1,9353; 0,6909) + 0,05×Lognormal (-
<b>PROBABILIDADE ABANDONAR:</b>	<b>Distribuições de probabilidade:</b>
Horário 11-12 horas: $M/M_1/c+G_{8m}$ $M/G_1/c+G_{8m}$	$G_1 \sim$ Fréchet (2,2564; 2,8235; -1,2554) $G_9 \sim$ Fréchet (2,2847; 2,7464; -1,2092) $G_{12} \sim$ Exponencial (0,148)
Horário 12-13 horas: $M/G_9/c+G_{12}$ $M/M_2/c+G_{12}$	$G_{16} \sim$ Lognormal (0,9415; 0,8377) $G_{20} \sim$ Exponencial (0,108) $M_1 \sim$ Exponencial (0,314) $M_2 \sim$ Exponencial (0,328)
Horário 15-16 horas: $M/G_{16}/c+G_{20}$ $M/M_3/c+G_{20}$	$M_3 \sim$ Exponencial (0,291) $G_{8m} \sim$ 0,95×Lognormal (1,8512; 0,6625) + 0,05×Lognormal (-2,0936; 1,5619)
<b>INTESIDADE DE TRÁFEGO:</b>	<b>Distribuições de probabilidade:</b>
Horário 11-12 horas: $M/M_1/c+G_7$ $M/G_1/c+G_{4m}$	$G_1 \sim$ Fréchet (2,2564; 2,8235; -1,2554) $G_7 \sim$ Lognormal (1,9115; 0,6488) $G_{10} \sim$ Fatigue Life (0,5069; 5,9746) $G_{14} \sim$ Lognormal (1,7902; 0,4861)
Horário 12-13 horas: $M/M_2/c+G_{14}$ $M/M_2/c+G_{10}$	$G_{16} \sim$ Lognormal (0,9415; 0,8377) $G_{22} \sim$ Lognormal (1,9962; 0,6769) $M_1 \sim$ Exponencial (0,314) $M_2 \sim$ Exponencial (0,328)
Horário 15-16 horas: $M/M_3/c+G_{22}$ $M/G_{16}/c+G_{19m}$	$M_3 \sim$ Exponencial (0,291) $G_{4m} \sim$ 0,9072×Fatigue Life (0,7456; 5,9258) + 0,0928×Fatigue Life (0,1189; 1,0091) $G_{19m} \sim$ 0,9215×Fatigue Life (0,7086; 6,5378) + 0,0785×Fatigue



## 6.2 Resultados do caso do Call Center da empresa

Conforme o Capítulo 5, após a aplicação dos testes de aderência aos dados do *Call Center* da Empresa, realizados com o uso do software EasyFit® (2010), ao nível de 5% de probabilidade, obteve-se como resultado que a distribuição Exponencial de probabilidade se mostra mais ajustada para representar o processo de chegada desse *Call Center*, com  $p$ -value 0,43675. Com relação ao processo de serviço, todas as distribuições analisadas foram rejeitadas. Entretanto, a distribuição Weibull (3p) mostrou-se a mais ajustada entre elas para representar esse processo, com  $p = 0,00342$ . Várias distribuições de probabilidade se mostraram aderentes aos tempos de paciência do *Call Center*. Entre essas possibilidades, optou-se pela distribuição Exponencial de probabilidade ( $p = 0,82479$ ), pela distribuição Normal ( $p = 0,18864$ ) e pela distribuição *Fatigue Life* ( $p = 0,55383$ ) na elaboração dos modelos analíticos de filas para representar o *Call Center* da empresa. Com base nesses resultados, considerou-se que o modelo analítico de filas M/G/c + G, com distribuições genéricas para o processo de serviço e de abandono, é o mais indicado para representar o *Call Center* da empresa. O modelo analítico de fila M/M/c + G, apesar de considerar a distribuição Exponencial para modelar os tempos de serviço, também foi utilizado para representar o mesmo *Call Center*, por ser um modelo muito comum na literatura e usado na modelagem das centrais de atendimento.

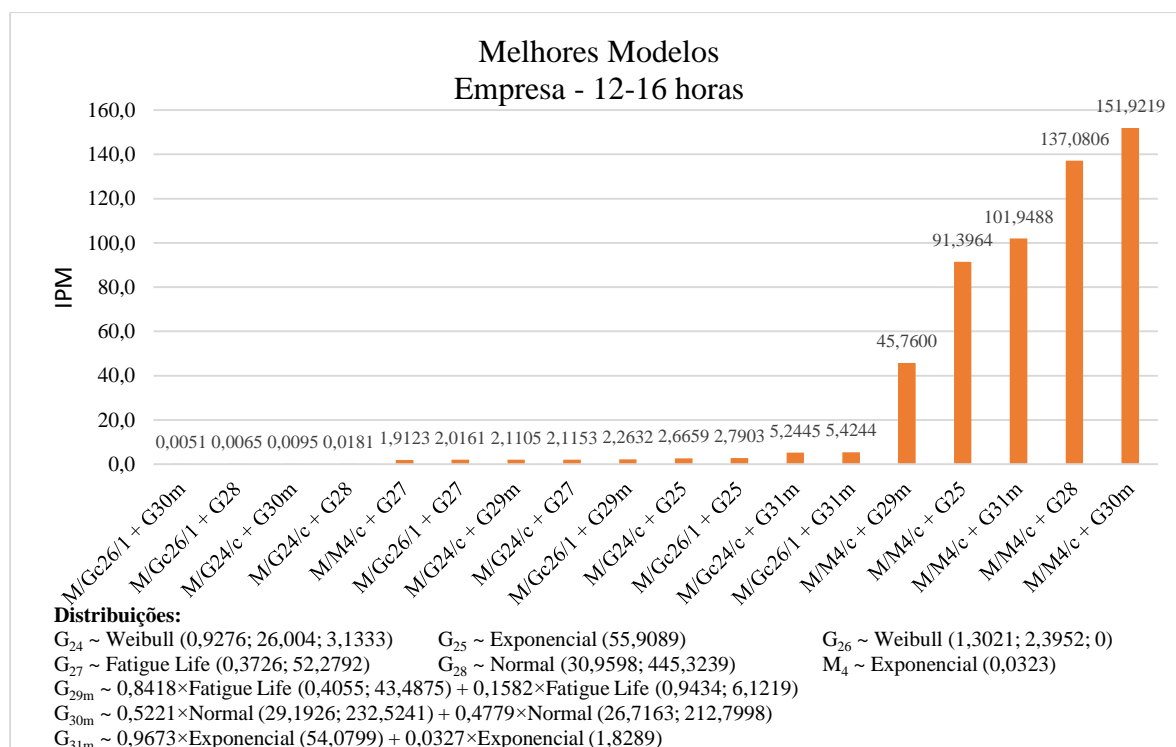
Várias possibilidades de distribuições mistas também foram aqui consideradas na elaboração desses modelos analíticos, combinando a mistura de duas componentes Exponenciais (Ex+Ex), duas Normais (N+N) e duas *Fatigue Life* (FL+FL). Os parâmetros dessas distribuições mistas, bem como o peso de cada uma delas, foram obtidos pelo método dos estimadores de máxima verossimilhança, calculados pelo software Mathematica (2013). As taxas de chegada, serviço e abandono, obtidas pela análise descritiva dos dados (Capítulo 5), foram inseridas nesses modelos e as medidas de desempenho foram calculadas. Convém lembrar que a paciência dos usuários são valores censurados que foram estimados, usando o estimador Kaplan-Meier, constituindo uma nova taxa de abandono que foi utilizada para recalculer os parâmetros das distribuições de probabilidade envolvidas no processo de abandono, como já descrita na seção 6.1.

As equações 6.1 e 6.2 foram consideradas nas análises do *Call Center* da Empresa para obtenção do modelo analítico de fila que represente o *Call Center* e para obtenção do modelo analítico de fila que seja mais adequado para representar cada uma das medidas de desempenho. Comparando-se as medidas de desempenho obtidas pelos modelos analíticos de fila com

aquelas extraídas dos dados reais, investigou-se a existência de um único modelo analítico de fila adequado para representar o *Call Center* da Empresa. O Gráfico 6.16 mostra que o modelo procurado é o  $M/G_{26}^c/1+G_{30m}$ , com distribuição Normal mista para os tempos de paciência e distribuição Weibull para os tempos de serviço.

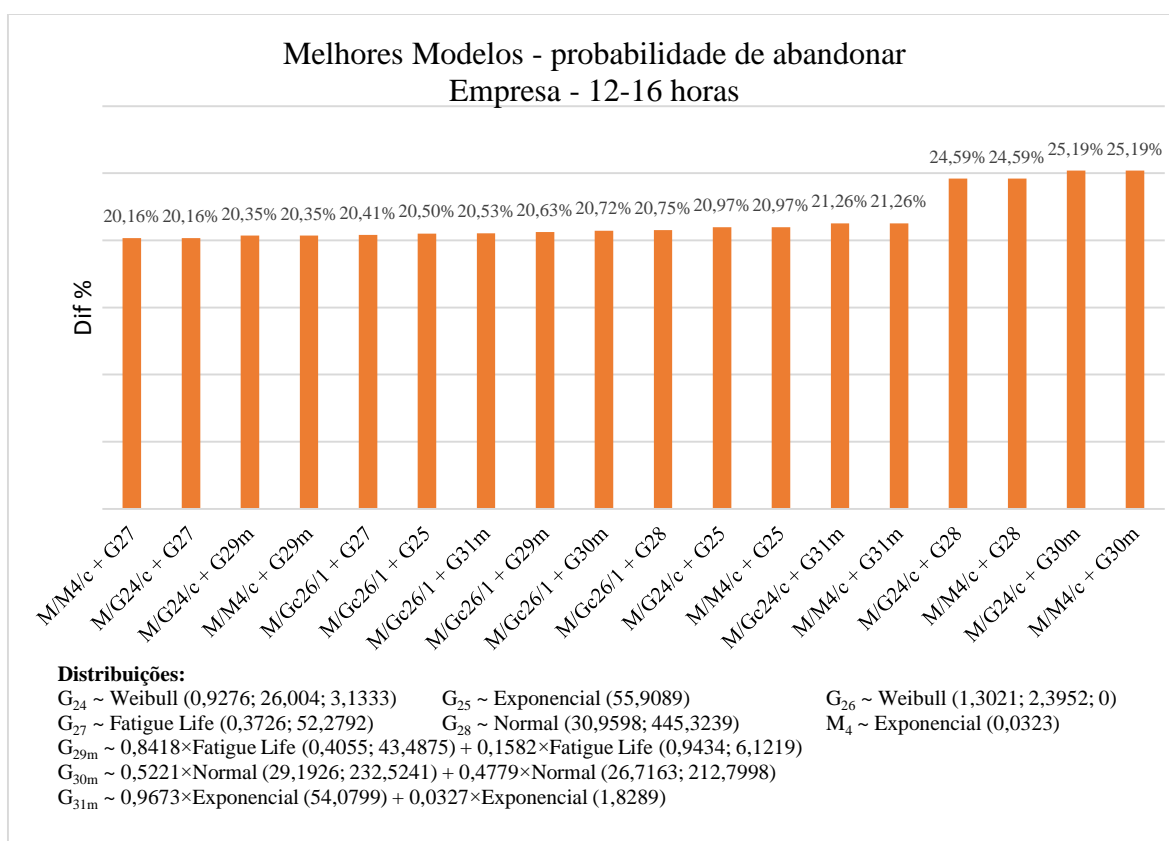
Buscou-se também o modelo analítico de fila que represente adequadamente cada uma das medidas de desempenho consideradas neste estudo. A métrica utilizada nesta análise é a mesma usada anteriormente, no *Call Center* do Banco, ou seja, a “diferença percentual” fornecida pela Equação 6.2. Desse modo, a medida de desempenho probabilidade de abandonar, como mostra o Gráfico 6.17, é adequadamente representada pelo modelo analítico de fila  $M/G_{24}/c+G_{27}$ , que possui distribuição Fatigue Life para os tempos de paciência e distribuição Weibull para os tempos de serviço, ou pelo modelo analítico de fila  $M/M_4/c+G_{27}$ , com a mesma

GRÁFICO 6.16 - Comparação do desempenho dos modelos – 12-16 horas



Fonte: elaborado pelo autor

GRÁFICO 6.17 - Comparação do desempenho dos modelos – 12-16 horas

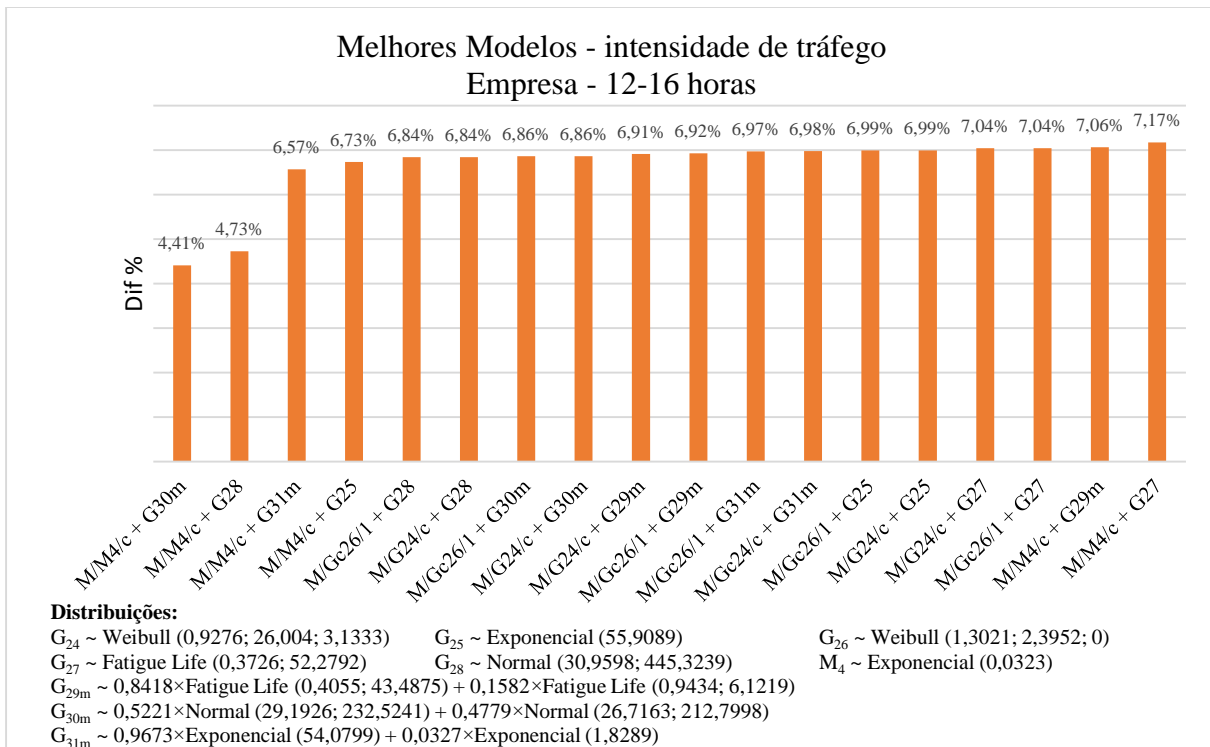


Fonte: elaborado pelo autor

distribuição Fatigue Life para a paciência, mas com distribuição Exponencial para os tempos de serviços. Modelos que também se mostraram adequados para estimar a probabilidade de abandonar, por possuírem diferença percentual (*Dif%*) muito próxima dos outros dois modelos, e que possuem distribuição mista Fatigue Life para os tempos de paciência, são o  $M/G_{24}/c+G_{29m}$ , com distribuição Weibull para os tempos de serviço, e o modelo  $M/M_4/c+G_{29m}$ , com distribuição Exponencial para o serviço.

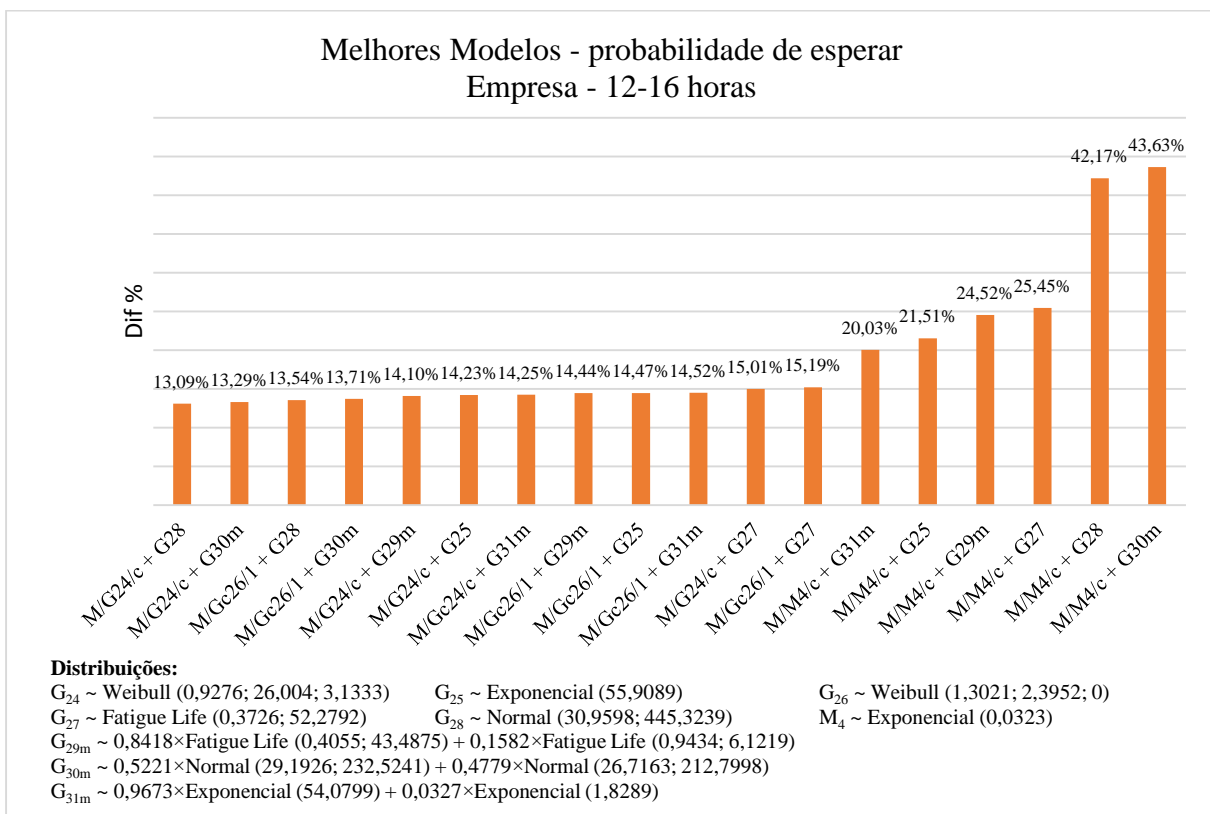
Considerando a medida de desempenho intensidade de tráfego, o Gráfico 6.18 mostra que o modelo  $M/M_4/c+G_{30m}$ , com distribuição Normal mista para os tempos de paciência e distribuição Exponencial para os tempos de serviço é o mais adequado. A probabilidade de esperar pode ser representada adequadamente pelo modelo analítico de fila  $M/G_{24}/c+G_{28}$  com distribuição Weibull para os tempos de serviço uma única distribuição Normal para a paciência, ou também pelo modelo  $M/G_{24}/c+G_{30m}$ , com a mesma distribuição para os tempos de serviço e distribuição Normal mista para a paciência. Esses dois modelos possuem diferenças percentuais muito próximas, conforme mostra o Gráfico 6.19.

GRÁFICO 6.18 - Comparação do desempenho dos modelos – 12-16 horas



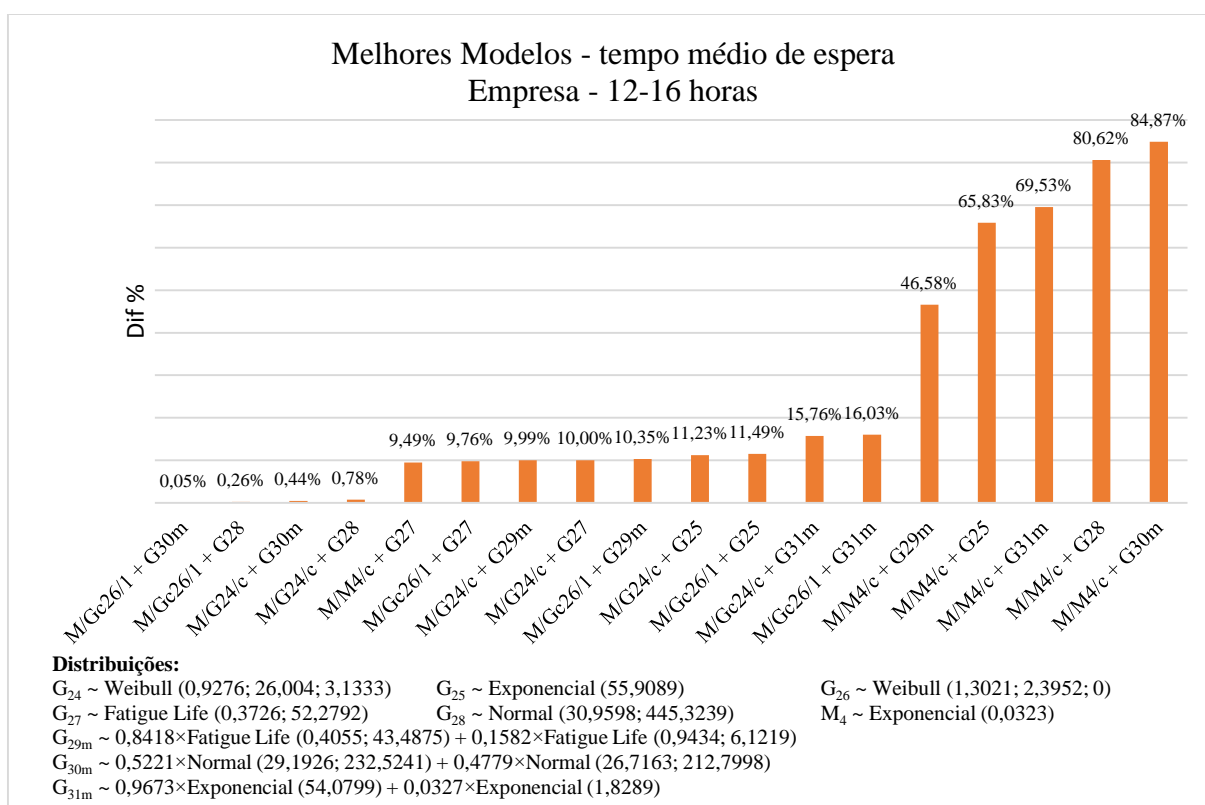
Fonte: elaborado pelo autor

GRÁFICO 6.19 - Comparação do desempenho dos modelos – 12-16 horas



Fonte: elaborado pelo autor

GRÁFICO 6.20 - Comparação do desempenho dos modelos – 12-16 horas



Fonte: elaborado pelo autor

O modelo analítico de fila  $M/G^c_{26}/1+G_{30m}$ , que possui distribuição Normal mista para os tempos de paciência e distribuição Weibull para os serviços, é o que mais se ajusta para a medida de desempenho tempo médio de espera, como indica o Gráfico 6.20.

Com os resultados obtidos, observa-se que os modelos analíticos de fila que são adequados para representar a medida de desempenho intensidade de tráfego, com diferenças percentuais inferiores a 5%, e os modelos analíticos de fila considerados adequados para representar o tempo médio de espera, com diferenças percentuais menores que 1%, são considerados válidos para essas medidas. Os modelos analíticos de fila considerados adequados para representar as medidas de desempenho probabilidade de esperar e probabilidade de abandonar, embora apresentam diferenças percentuais maiores, 13,09% (Gráfico 6.19) e 20,16% (Gráfico 6.17), respectivamente, também foram considerados válidos para essas medidas. Para as duas últimas medidas, recomenda-se também o uso de modelos de simulação para melhor estimá-las, porque possuem diferenças percentuais menores que as dos modelos analíticos.

Destaca-se ainda que os modelos analíticos de fila que possuem distribuição mista para os tempos de paciência são efetivos para representar a intensidade de tráfego e o tempo médio

de espera do *Call Center* da Empresa. Nas outras medidas de desempenho, probabilidade de abandonar e probabilidade de esperar, as distribuições mistas para a paciência também ocorrem nos modelos analíticos de fila, mas estes modelos não são os mais adequados para representá-las.

Conforme o tipo de modelo analítico de fila adotado para representar o *Call Center* da Empresa, independentemente de qualquer medida de desempenho, os modelos que utilizam distribuições mistas para a paciência em alguns casos são melhores que o seu correspondente modelo com uma única classe de distribuição paramétrica para modelar o paciência. Observando a Tabela 6.4, o modelo analítico de fila do tipo  $M/G/c+G$ , as distribuições mistas Normal e Fatigue Life para os tempos de paciência são melhores que as suas correspondentes não mistas (menores IPM). No modelo do tipo  $M/G^c/1+G$ , apenas a paciência com distribuição Normal mista é melhor que a sua correspondente não mista. No modelo do tipo  $M/M/c+G$ , as distribuições mistas para a paciência não são melhores que as correspondentes não mistas. Outro resultado que se obtém é que o *Call Center* da Empresa, independentemente de qualquer medida de desempenho, pode ser representado por um modelo analítico de fila com uma distribuição mista para os tempos de paciência (Gráfico 6.16).

TABELA 6.4 - Comparando paciência mista com não mista – Caso Empresa

Modelos		Modelos		Modelos	
M/G/c+G	IPM	M/G <sup>c</sup> /1+G	IPM	M/M/c+G	IPM
M/G <sub>24</sub> /c + G <sub>30m</sub>	0,0095	M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>30m</sub>	0,0051	M/M <sub>4</sub> /c + G <sub>27</sub>	1,9123
M/G <sub>24</sub> /c + G <sub>28</sub>	0,0181	M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>28</sub>	0,0065	M/M <sub>4</sub> /c + G <sub>29m</sub>	45,7600
M/G <sub>24</sub> /c + G <sub>29m</sub>	2,1105	M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>27</sub>	2,0161	M/M <sub>4</sub> /c + G <sub>25</sub>	91,3964
M/G <sub>24</sub> /c + G <sub>27</sub>	2,1153	M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>29m</sub>	2,2632	M/M <sub>4</sub> /c + G <sub>31m</sub>	101,9488
M/G <sub>24</sub> /c + G <sub>25</sub>	2,6659	M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>25</sub>	2,7903	M/M <sub>4</sub> /c + G <sub>28</sub>	137,0806
M/G <sup>c</sup> <sub>24</sub> /c + G <sub>31m</sub>	5,2445	M/G <sup>c</sup> <sub>26</sub> /1 + G <sub>31m</sub>	5,4244	M/M <sub>4</sub> /c + G <sub>30m</sub>	151,9219
<b>Processo chegada:</b> Poisson – taxa $\lambda = 0,7318$ – média $E(\tau) = 1,3665$			<b>Processo serviço:</b> Taxa $\mu = 0,0323$ – média $E(S) = 30,9598$		
<b>Distribuições:</b> M ~ Exponencial(1,3665)			<b>Distribuições:</b> G <sub>24</sub> ~ Weibull (0,9276; 26,004; 3,1333) G <sub>26</sub> ~ Weibull (1,3021; 2,3952; 0) M <sub>4</sub> ~ Exponencial (30,9598)		
<b>Processo abandono:</b> taxa 0,01789 – média $E(R) = 55,9089$					
<b>Distribuições:</b> G <sub>25</sub> ~ Exponencial (55,9089)    G <sub>27</sub> ~ Fatigue Life (0,3726; 52,2792)    G <sub>28</sub> ~ Normal (30,9598; 445,3239) G <sub>29m</sub> ~ 0,8418×Fatigue Life (0,4055; 43,4875) + 0,1582×Fatigue Life (0,9434; 6,1219) G <sub>30m</sub> ~ 0,5221×Normal (29,1926; 232,5241) + 0,4779×Normal (26,7163; 212,7998) G <sub>31m</sub> ~ 0,9673×Exponencial (54,0799) + 0,0327×Exponencial (1,8289)					

Fonte: elaborado pelo autor

Uma visão geral dos modelos analíticos de fila que são mais eficientes para representar o *Call Center* da Empresa, está sumarizada nos Quadros 6.3 e 6.4. O Quadro 6.3 destaca o modelo analítico mais efetivo, no horário analisado das 12-16 horas, para representar o *Call Center* da Empresa.

QUADRO 6.3 - Modelo analítico de fila mais efetivo para representar o *Call Center* da Empresa

<p><b>CALL CENTER DA EMPRESA:</b> Horário 12-16 horas: <math>M/G_{26}^c/1+G_{30m}</math></p>	<p><b>Distribuições de probabilidade:</b> <math>G_{26} \sim \text{Weibull} (1,3021; 2,3952; 0)</math> <math>G_{30m} \sim 0,5221 \times \text{Normal} (29,1926; 232,5241) + 0,4779 \times \text{Normal} (26,7163; 212,7998)</math></p>
--	---

Fonte: elaborado pelo autor

O Quadro 6.4 mostra os dois modelos analíticos de fila mais apropriados para representar o *Call Center* da Empresa, para cada uma das medidas de desempenho, no horário analisado das 12-16 horas.

QUADRO 6.4 - Modelo analítico de fila mais apropriado para medidas de desempenho-C.C.Empresa

<p><b>TEMPO MÉDIO DE ESPERA:</b> Horário 12-16 horas: <math>M/G_{26}^c/1+G_{30m}</math> <math>M/G_{26}^c/1+G_{28}</math></p>	<p><b>Distribuições de probabilidade:</b> <math>G_{26} \sim \text{Weibull} (1,3021; 2,3952; 0)</math> <math>G_{28} \sim \text{Normal} (30,9598; 445,3239)</math> <math>G_{30m} \sim 0,5221 \times \text{Normal} (29,1926; 232,5241) + 0,4779 \times \text{Normal} (26,7163; 212,7998)</math></p>
<p><b>PROBABILIDADE ESPERAR:</b> Horário 12-16 horas: <math>M/G_{24}/c+G_{28}</math> <math>M/G_{24}/c+G_{30m}</math></p>	<p><b>Distribuições de probabilidade:</b> <math>G_{24} \sim \text{Weibull} (0,9276; 26,004; 3,1333)</math> <math>G_{28} \sim \text{Normal} (30,9598; 445,3239)</math> <math>G_{30m} \sim 0,5221 \times \text{Normal} (29,1926; 232,5241) + 0,4779 \times \text{Normal} (26,7163; 212,7998)</math></p>
<p><b>PROBABILIDADE ABANDONAR:</b> Horário 12-16 horas: <math>M/M_4/c+G_{27}</math> <math>M/G_{24}/c+G_{27}</math></p>	<p><b>Distribuições de probabilidade:</b> <math>G_{24} \sim \text{Weibull} (0,9276; 26,004; 3,1333)</math> <math>G_{27} \sim \text{Fatigue Life} (0,3726; 52,2792)</math> <math>M_4 \sim \text{Exponencial} (0,0323)</math></p>
<p><b>INTENSIDADE DE TRÁFEGO:</b> Horário 12-16 horas: <math>M/M_4/c+G_{30m}</math> <math>M/M_4/c+G_{28}</math></p>	<p><b>Distribuições de probabilidade:</b> <math>G_{28} \sim \text{Normal} (30,9598; 445,3239)</math> <math>M_4 \sim \text{Exponencial} (0,0323)</math> <math>G_{30m} \sim 0,5221 \times \text{Normal} (29,1926; 232,5241) + 0,4779 \times \text{Normal} (26,7163; 212,7998)</math></p>

Fonte: elaborado pelo autor

Em resumo, nos dois casos analisados, o *Call Center* do Banco e o *Call Center* da Empresa, recomenda-se utilizar os modelos analíticos de fila  $M/G/c+G$ , com distribuição genérica para os tempos de serviço e também para os tempos de paciência, porque apresentaram melhores resultados que os modelos  $M/M/c+G$ , com distribuição Exponencial para os tempos de serviço e genérica para os tempos de paciência (Gráficos 6.1, 6.2, 6.3 e 6.16), exceto no horário das 12-13 horas do *Call Center* do Banco, em que os modelos analíticos de fila  $M/M/c+G$  foram mais representativos.

Ainda, considerando-se os dois casos analisados, entre os modelos que possuem distribuições genéricas para os tempos de serviço e, também, para os tempos de paciência, é recomendável utilizar os modelos analíticos de fila  $M/G^c/1+G$  para representar os dois *Call Centers*, porque em geral, tem mostrado resultados mais efetivos que os modelos analíticos de fila  $M/G/c+G$  (Gráficos 6.1, 6.2, 6.3 e 6.16).

Observa-se também que, entre os dois melhores modelos analíticos de fila recomendados para representar os *Call Centers* analisados estão, em geral, um com distribuição mista de probabilidade para a paciência e outro com distribuição não mista de probabilidade para a paciência, sendo que, os que possuem distribuição mista de probabilidade para representar os tempos de paciência são os melhores. Esse resultado não se aplica no segundo horário (12-13 horas) do *Call Center* do Banco, em que os dois melhores modelos possuem distribuição não mista de probabilidade para os tempos de paciência. De fato, no caso do *Call Center* do Banco, no horário das 11-12 horas o modelo analítico de fila mais indicado para representá-lo é o  $M/G^c_3/1+G_{4m}$ , com distribuição de probabilidade Fréchet para os tempos de serviço e com distribuição mista de probabilidade Fatigue Life para os tempos de paciência e o outro é o modelo analítico de fila  $M/M_1/c+G_2$ , com distribuição Exponencial para os tempos de serviço e distribuição não mista de probabilidade Fatigue Life para os tempos de paciência, conforme Gráfico 6.1. No horário das 12-13 horas, os dois melhores modelos analíticos de fila recomendados para representar o *Call Center* possuem distribuição Exponencial para os tempos de serviço, sendo o mais indicado com distribuição não mista de probabilidade Lognormal para os tempos de paciência ( $M/M_2/c+G_{14}$ ) e o outro com distribuição não mista de probabilidade Fatigue Life para os tempos de paciência ( $M/M_2/c+G_{10}$ ), conforme o Gráfico 6.2. No último horário, das 15-16 horas, os dois melhores modelos de fila recomendados para representar o *Call Center* do Banco possuem distribuição de probabilidade Lognormal para os tempos de serviço e distribuição mista de probabilidade Fatigue Life para os tempos de paciência, sendo que o mais preciso é o modelo analítico de fila  $M/G^c_{18}/1+G_{19m}$ , e o outro o modelo de fila  $M/G^c_{16}/c+G_{19m}$ , conforme Gráfico 6.3.



Ainda com relação aos dois melhores modelos analíticos de fila para o *Call Center* da Empresa, recomenda-se como o mais efetivo o  $M/G_{26}^c/1+G_{30m}$ , com distribuição de probabilidade Lognormal para os tempos de serviço e com distribuição mista de probabilidade Normal para os tempos de paciência. O outro modelo analítico de fila recomendado é o  $M/G_{26}^c/1+G_{28}$ , com distribuição de probabilidade Lognormal para os tempos de serviço e com distribuição não mista de probabilidade Normal para os tempos de paciência, conforme o Gráfico 6.16.

Nos casos analisados, os modelos analíticos de fila com distribuição mista para os tempos de paciência foram mais precisos e, portanto, recomendados para representar os dois *Call Centers*, exceto no segundo horário (12-13 horas) do *Call Center* do Banco. Considerando-se o caso *Call Center* do Banco os modelos analíticos de fila recomendados para representá-lo são os  $M/G_3^c/1+G_{4m}$ ,  $M/M_2/c+G_{14}$  e  $M/G_{18}^c/1+G_{19m}$ , nos horários das 11-12 horas, das 12-13 horas e das 15-16 horas, respectivamente, conforme Gráficos 6.1, 6.2 e 6.3. Para o caso *Call Center* da Empresa recomenda-se o modelo analítico de fila  $M/G_{26}^c/1+G_{30m}$  para representá-lo, conforme comprova o Gráfico 6.16. Esses modelos analíticos de fila são recomendados para representar os dois casos de *Call Centers* analisados em todas as medidas de desempenho estudadas.

### 6.3 Cenários alternativos

Uma das aplicações importantes dos modelos analíticos de fila, que representem adequadamente um *Call Center*, é permitir aos seus gestores tomarem decisões com base em diversos cenários e realizar previsões sobre as operações desses sistemas. Em particular, uma das preocupações para quem administra esses sistemas é o abandono dos seus usuários, que pode ser quantificado pela probabilidade de abandonar. O aumento do valor dessa probabilidade sugere ao gestor um baixo nível de serviço oferecido pelo *Call Center*. Ao abandonar, os usuários estão indiretamente informando aos gestores que não vale a pena esperar pelo serviço oferecido. Manter um nível de serviço adequado implica em reduzir a probabilidade de abandono, que depende da quantidade de atendentes em operação. É prática comum nestes sistemas manter a probabilidade de abandono próxima de 2% (BOUZADA, 2009). Assim, o gestor do *Call Center* pode dimensionar o tamanho da equipe calculando a quantidade de atendentes necessária para manter o nível de serviço desejado. Este cenário, denominado de

Cenário 1, foi analisado com o modelo analítico de fila e os resultados verificados com o correspondente modelo de simulação de fila que representam o caso *Call Center* da Empresa.

A probabilidade de abandono é uma medida subjetiva do nível dos serviços oferecidos. Mais precisamente, o nível de serviço pode ser medido pelo percentual das chamadas atendidas em até uma quantidade estabelecida de tempo entre aquelas que aguardaram na fila até serem atendidas pelos operadores. Dessa forma, o nível de serviço é a probabilidade do usuário aguardar na fila de espera, no máximo, uma quantidade de tempo para ser atendido, calculado pela Equação (6.3):

$$NS = P(W \leq x) = \frac{\text{número de chamadas atendidas até o tempo } x}{\text{número de chamadas atendidas pelo operador}} \quad (6.3)$$

Nas operações de *Call Centers*, normalmente, se estabelece um nível mínimo de serviço de 75% das chamadas que devem ser atendidas em até 10 segundos (BOUZADA, 2009). Este cenário, denominado de Cenário 2, também foi analisado com o modelo analítico de fila e os resultados verificados com o correspondente modelo de simulação de fila que representam o caso *Call Center* da Empresa.

A análise de cenários também pode ser utilizada para saber o que aconteceria com as medidas de desempenho do *Call Center* se houvesse um aumento na demanda das chamadas. Para essa análise, variaram-se as taxas de chegadas de 0,5 chamada por minuto até 3,0 chamadas por minuto, com intervalo de 0,5 minuto, que foram substituídas nos modelos analíticos e nos modelos de simulação de fila, para obter os valores das medidas de desempenho tempo médio de espera, probabilidade de esperar, probabilidade de abandonar e intensidade de tráfego. Com os valores dessas medidas de desempenho pode-se avaliar o impacto que elas provocam no desempenho geral do *Call Center* e, também, até que aumento na demanda das chamadas o *Call Center* suporta. Este cenário, denominado de Cenário 3, foi analisado com o modelo analítico de fila e os resultados verificados com o correspondente modelo de simulação de fila que representam o *Call Center* do Banco, em cada um dos três horários analisados (11-12 horas, 12-13 horas e 15-16 horas).

Os resultados obtidos desses cenários são projeções realizadas das medidas de desempenho que poderão ser utilizadas pelos gestores para análise futura do congestionamento do *Call Center*. Por esse motivo, não possuem um valor correspondente nos dados reais dos *Call Centers* para verificar a precisão obtida. Essa precisão foi verificada comparando-se os

resultados dos cenários com os resultados obtidos dos modelos de simulação discreta descritos e verificados nas seções 4.5 e 5.4. Cada um desses cenários é analisado em detalhes nas seções seguintes.

### 6.3.1 Cenário 1: quantidade de atendentes

Considerou-se o caso do *Call Center* da Empresa que pode ser representado pelo modelo de fila com distribuição mista para a paciência,  $M/G_{26}^c/1+G_{30m}$ , conforme mostrado na seção 6.2 (Gráfico 6.16). Apesar de se considerar um único servidor, a taxa de serviço desse modelo é multiplicada por uma quantidade  $c$  de servidores, para ficar próximo ao sistema multiservidor  $M/G_{26}/c+G_{30m}$ , com a mesma distribuição para os tempos de serviço e tempos de paciência que o anterior. Desse modo, o modelo analítico de fila  $M/G_{26}^c/1+G_{30m}$  opera aproximadamente como o modelo analítico de fila  $M/G_{26}/c+G_{30m}$ . Utilizando-se do modelo  $M/G_{26}^c/1+G_{30m}$ , com distribuição Weibull para os tempos de serviço e com distribuição Normal mista para os tempos de paciência, calculou-se a quantidade de atendentes necessários para que a probabilidade de abandonar, do *Call Center* da Empresa, diminuísse de 31,8% (dados reais) para próximo de 2%.

O *Call Center* da Empresa opera normalmente com uma equipe de 14 atendentes. Para diminuir a probabilidade de abandono, aumentou-se gradativamente a quantidade de atendentes, substituindo-se essa quantidade nas equações do modelo analítico  $M/G_{26}^c/1+G_{30m}$  e, também, no seu respectivo modelo simulado. Os resultados obtidos estão na Tabela 6.5.

TABELA 6.5 - Probabilidade de abandono versus quantidade atendentes  
*Call Center* Empresa – 12-16 horas

Atendentes	Modelo de fila $M/G_{26}^c/1+G_{30m}$	
	Analítico	Simulado
$c = 16$	31,2%	28,9%
$c = 18$	22,7%	21,7%
$c = 20$	14,1%	14,4%
$c = 22$	5,5%	9,4%
$c = 24$	5,4%	5,3%
$c = 26$	2,7%	3,6%
$c = 28$	2,1%	2,4%
$c = 29$	2,0%	1,3%

**Distribuições:**  
 $G_{26} \sim$  Weibull (1,3021; 2,3952; 0)  
 $G_{30m} \sim 0,5221 \times \text{Normal} (29,1926; 232,5241) + 0,4779 \times \text{Normal} (26,7163; 212,7998)$

Fonte: elaborada pelo autor

Primeiramente, observa-se que os valores das probabilidades de abandono estão bastante próximos em ambos os modelos analíticos e simulados, exceto para  $c = 22$ , evidenciando que o modelo de fila  $M/G_{26}^c/1+G_{30m}$  é efetivo para representar o *Call Center* da Empresa e que o mesmo pode ser utilizado para apoiar as decisões de dimensionamento e operação desse sistema. Como resultado desse cenário alternativo, o gestor do *Call Center* deveria manter uma equipe de 29 atendentes, o dobro do que utiliza tradicionalmente, para que a probabilidade de abandono fique próxima de 2%, melhorando o seu nível de serviço.

### 6.3.2 Cenário 2: nível de serviço

Nas condições do Cenário 1, analisado com 29 atendentes e simulando o modelo de fila  $M/G_{26}^c/1+G_{30m}$  que representa o *Call Center* da Empresa, obteve-se 32.842 usuários atendidos pelos operadores, dos quais 10.555 foram atendidos em até 10 segundos. Desse modo, o nível de serviço para esse cenário é de:

$$NS = \frac{10555}{32842} = 0,3214 \text{ ou } 32,14\%$$

Esse valor para o nível de serviço é menor que a condição de 75% estabelecida. Assim, de acordo com esse cenário, o *Call Center* da Empresa com uma equipe de 29 atendentes é suficiente para garantir a meta de 2% para a probabilidade de abandono, mas insuficiente para atingir o nível de serviço estabelecido em 75% das chamadas atendidas em até 10 segundos.

### 6.3.3 Cenário 3: variação da demanda

O *Call Center* do Banco, em cada um dos três horários considerados neste estudo (11-12 horas, 12-13 horas e 15-16 horas), foi utilizado para as análises do Cenário 3. Para essa finalidade utilizou-se os modelos analíticos de fila estudados e que melhor representam o *Call Center* do Banco nesses horários.

O modelo analítico de fila  $M/G_3^c/1+G_{4m}$  é o mais preciso para representar o *Call Center* do Banco no horário das 11-12 horas, conforme mostrado na seção 6.1 (Gráfico 6.1). Nesse

modelo analítico de fila, os tempos de serviço são representados pela distribuição de probabilidade Fréchet e os tempos de paciência pela distribuição mista de probabilidade com duas componentes Fatigue Life. Variando-se a taxa de chegada de 0,5 chamada por minuto até 3,0 chamadas por minuto e inserindo-se esses valores no modelo analítico e no modelo de simulação da fila  $M/G^c_3/1+G_{4m}$ , obteve-se os resultados para as medidas de desempenho que se encontram nas Tabelas 6.6 e 6.7.

O *Call Center* do Banco realiza suas operações com seis atendentes e possui uma taxa de chegada de 1,933 chamadas por minuto (dados reais), no horário das 11-12 horas. Se houver um aumento de 29,3% na demanda, a nova taxa de chegada seria de 2,5 chamadas por minuto. Com essa nova taxa de chegada a probabilidade de abandonar varia de 0,139 (dados reais) para 0,266 (Tabela 6.7), implicando um aumento de 91,4%, revelando um grande impacto nessa medida de desempenho e prejudicando o nível dos serviços prestados.

TABELA 6.6 - Cenário 3 – Variação da demanda – *Call Center* Banco – 11-12 horas

Taxa chegada (ch/min)	Tempo médio espera			Probabilidade de esperar		
	Simulação	Intervalo Confiança	Analítico	Simulação	Intervalo Confiança	Analítico
0,5	0,1386	(0,1231; 0,1541)	0,0889	0,9868	(0,9517; 1,0219)	0,9826
1,0	0,3941	(0,3659; 0,4223)	0,2307	0,9618	(0,9198; 1,0038)	0,9663
1,5	0,8621	(0,8155; 0,9087)	0,5428	0,9147	(0,8363; 0,9931)	0,9152
2,0	1,5491	(1,4735; 1,6247)	1,4942	0,8389	(0,6646; 1,0132)	0,8360
2,5	2,5974	(2,5195; 2,6753)	2,5083	0,7292	(0,5676; 0,8908)	0,7245
3,0	3,6289	(3,5341; 3,7237)	3,2059	0,6318	(0,5546; 0,7090)	0,6308

Fonte: elaborada pelo autor

TABELA 6.7 - Cenário 3 – Variação da demanda – *Call Center* Banco – 11-12 horas

Taxa chegada (ch/min)	Probabilidade de abandonar			Intensidade de tráfego		
	Simulação	Intervalo Confiança	Analítico	Simulação	Intervalo Confiança	Analítico
0,5	0,0132	(0,0000; 0,0483)	0,0129	0,2572	(0,1907; 0,3237)	0,2619
1,0	0,0382	(0,0000; 0,0802)	0,0297	0,5098	(0,3982; 0,6214)	0,5150
1,5	0,0853	(0,0069; 0,1637)	0,0359	0,7473	(0,6052; 0,8894)	0,7676
2,0	0,1610	(0,0000; 0,3353)	0,1158	0,9206	(0,5736; 1,2676)	0,9386
2,5	0,2708	(0,1092; 0,4324)	0,2655	0,9775	(0,9107; 1,0443)	0,9747
3,0	0,3682	(0,2910; 0,4454)	0,3815	0,9984	(0,9747; 1,0221)	0,9848

Fonte: elaborada pelo autor

Conforme esse Cenário 3, a medida de desempenho intensidade de tráfego atinge o valor 0,985 (Tabela 6.7), próximo do máximo admissível que é 1, com uma taxa de 3,0 chegadas por minuto, que corresponde a um aumento de 55,5% na demanda do *Call Center* do Banco.

Portanto, esse é o limite de aumento suportado na demanda do *Call Center* do Banco, deixando o sistema altamente congestionado e comprometendo as suas operações.

Considerando-se o aumento de 29,3% (de 1,933 chamadas por minuto para 2,5 chamadas por minuto) na demanda do *Call Center* do Banco, o tempo médio de espera varia de 1,043 minutos para 2,51 minutos, sofrendo um aumento de 139,7%. A probabilidade de esperar mantém-se praticamente inalterada, variando de 71,9% para 72,5%, com esse aumento na demanda das chamadas, conforme Tabela 6.6. O valor da Tabela considerado em todos esses cálculos refere-se aquele do modelo analítico de fila e o valor discriminado como dados reais encontram-se na seção 4.3.1.

Assim, a análise desse Cenário 3 mostrou que com o efetivo original de seis atendentes, um aumento razoável na quantidade de chamadas que chegam no *Call Center* do Banco, no horário das 11-12 horas, provoca um forte impacto no desempenho global do sistema.

No horário das 12-13 horas, o modelo analítico de fila mais indicado para representar o *Call Center* do Banco é o  $M/M_2/c+G_{14}$ , com distribuição de probabilidade Exponencial para os tempos de serviços, e com distribuição não mista de probabilidade Lognormal para os tempos de paciência, conforme mostrado no Gráfico 6.2 da seção 6.1. Nesse horário o *Call Center* do Banco também possui uma equipe de 6 atendentes, com uma taxa de chegada de 1,717 chamadas por minuto (dados reais). Adotando-se o mesmo intervalo de variação para a taxa de chegada, de 0,5 até 3,0 chamadas por minuto, e substituindo-se cada um desses valores no modelo analítico de fila  $M/M_2/c+G_{14}$ , obteve-se as medidas de desempenho das Tabelas 6.8 e 6.9.

Supondo um aumento de 16,5% na demanda das chamadas desse *Call Center*, a nova taxa de chegada seria de 2,0 chamadas por minuto. Com esse aumento, a probabilidade de abandonar varia de 10,2% (dados reais) para 11,4% (Tabela 6.9), aumentando em 11,8% e a intensidade de tráfego, varia de 82,8% (dados reais) para 90,0% (Tabela 6.9), com um aumento de 8,7%.

Com o mesmo aumento na demanda (16,5%) mudanças mais significativas ocorrem no tempo médio de espera, que varia de 0,813 minutos (dados reais) para 1,2065 minutos (Tabela 6.8), sofrendo um aumento de 48,4%, e na probabilidade de esperar que varia de 58,3% (dados reais) para 72,0% (Tabela 6.8), aumentando o seu valor em 13,5%.

A intensidade de tráfego que mede o congestionamento do sistema atinge o valor de 99,4% quando a taxa de chegada atingir 3,0 chamadas por minuto. Isso ocorre se houver um aumento de 74,7% (de 1,717 ch/min para 3,0 ch/min) na demanda das chamadas, sendo este o limite de aumento suportado na demanda do *Call Center* do Banco, no horário das 12-13 horas.

Em todos esses cálculos considerou-se como valor da Tabela aquele obtido com o modelo analítico de fila, e o valor discriminado como dados reais encontram-se na secção 4.3.1.

O Cenário 3, aplicado no horário das 12-13 horas do *Call Center* do Banco, mostrou que um aumento na demanda das chamadas do sistema também provoca impacto no seu desempenho global, mas inferiores àqueles obtidos no horário anterior.

TABELA 6.8 - Cenário 3 – Variação da demanda – *Call Center* Banco – 12-13 horas

Taxa chegada (ch/min)	Tempo médio espera			Probabilidade de esperar		
	Simulação	Intervalo Confiança	Analítico	Simulação	Intervalo Confiança	Analítico
0,5	0,0039	(0,0000; 0,0080)	0,0033	0,9998	(0,9977; 1,0019)	0,0056
1,0	0,0854	(0,0702; 0,1006)	0,0915	0,9963	(0,9515; 1,0411)	0,1021
1,5	0,4581	(0,4012; 0,5150)	0,4821	0,9765	(0,9111; 1,0419)	0,3875
2,0	1,4201	(1,3292; 1,5110)	1,2065	0,9098	(0,8131; 1,0065)	0,7201
2,5	2,5984	(2,5085; 2,6883)	1,9127	0,7688	(0,6834; 0,8542)	0,912
3,0	3,4773	(3,4053; 3,5493)	2,3721	0,6654	(0,5240; 0,8068)	0,9781

Fonte: elaborada pelo autor

TABELA 6.9 - Cenário 3 – Variação da demanda – *Call Center* Banco – 12-13 horas

Taxa chegada (ch/min)	Probabilidade de abandonar			Intensidade de tráfego		
	Simulação	Intervalo Confiança	Analítico	Simulação	Intervalo Confiança	Analítico
0,5	0,0002	(0,0000; 0,0023)	0,00008	0,2530	(0,1451; 0,3609)	0,2541
1,0	0,0037	(0,0000; 0,0485)	0,0037	0,5026	(0,3620; 0,6432)	0,5062
1,5	0,0235	(0,0000; 0,089)	0,0318	0,7403	(0,6700; 0,8106)	0,7379
2,0	0,0902	(0,0000; 0,1869)	0,1143	0,9166	(0,7821; 1,0511)	0,9001
2,5	0,2310	(0,1456; 0,3164)	0,2345	0,9993	(0,9189; 1,0797)	0,9724
3,0	0,3346	(0,1932; 0,476)	0,3481	1,0189	(0,9740; 1,0638)	0,9938

Fonte: elaborada pelo autor

No último horário, das 15-16 horas, o *Call Center* do Banco é devidamente representado pelo modelo analítico de fila  $M/G^c_{18}/1+G_{19m}$ , com distribuição de probabilidade Lognormal para representar os tempos de serviço e com distribuição mista de probabilidade, formada por duas componentes Fatigue Life, para representar os tempos de paciência, conforme Gráfico 6.3 da secção 6.1. Neste horário das 15-16 horas, o *Call Center* do Banco também utiliza seis atendentes nas suas operações e possui uma taxa de chegada de 2,1 chamadas por minuto, extraída dos dados reais. Utilizando-se a mesma variação na taxa de chegada adotada no horário anterior, de 0,5 chamada por minuto até 3,0 chamadas por minuto, que depois de inseridas no

modelo analítico  $M/G^c_{18}/1+G_{19m}$ , obteve-se os valores das medidas de desempenho das Tabelas 6.10 e 6.11.

Uma variação na taxa de chegada de 2,1 chamadas por minuto para 2,5 chamadas por minuto, provoca um aumento de 19,1% na demanda do *Call Center*. Como consequência a probabilidade de abandono varia de 23,4% (dados reais) a 33,1% (Tabela 6.11), aumentando em 41,5%, comprometendo o nível de serviço do sistema. O tempo médio de espera varia de 1,98 minutos (dados reais) para 2,12 minutos (Tabela 6.10), também aumentando, mas em menor valor, de 6,9%. A probabilidade de esperar sofre uma variação negativa de 90,3% (dados reais) para 66,1% (Tabela 6.10), alterando-se em -26,8%. A intensidade de tráfego também varia negativamente passando de 95,9% para 93,8%, reduzindo em -2,2%.

A pequena variação ocorrida no tempo médio de espera e as variações negativas da probabilidade de esperar e da intensidade de tráfego são justificadas pelo possível aumento significativo na probabilidade de abandono, se ocorresse o aumento de 19,1% na demanda. Um abandono acarreta uma diminuição no tempo de espera dos usuários que chegam depois daquele que abandonou e dos que ainda irão chegar. Desse modo, o aumentando significativo da quantidade de abandonos provoca uma diminuição dos tempos médios de espera dos usuários, uma diminuição na probabilidade de abandonar e também na intensidade de tráfego do sistema.

Ainda com relação a intensidade de tráfego observa-se que o *Call Center* do Banco, nesse horário das 15-16 horas, está sobrecarregado com 95,9% (dados reais) de ocupação dos atendentes. Entretanto, se a variação da demanda for de 48,2%, mudando de 2,1 chamadas por minuto (dados reais) para 3,0 chamadas por minuto (Tabela 6.11), a intensidade de tráfego atinge o valor de 97,1%, indicando uma ocupação ainda maior do sistema e bastante próximo do máximo 100%. Assim, 48,2% é o limite máximo permitido para uma variação na demanda nas chamadas do *Call Center* do Banco, nesse horário das 15-16 horas.

Em todos esses cálculos considerou-se como valor da Tabela aquele obtido com o modelo analítico de fila e o valor discriminados como dados reais encontram-se na seção 4.3.1.

Nesse horário das 15-16 horas o Cenário 3 mostrou que um aumento na demanda das chamadas no *Call Center* do Banco provoca um comportamento no desempenho global do sistema diferente daqueles ocorridos nos dois primeiros horários, diminuindo o valor do desempenho, em alguns casos.



TABELA 6.10 - Cenário 3 – Variação da demanda – *Call Center* Banco – 15-16 horas

Taxa chegada (ch/min)	Tempo médio espera			Probabilidade de esperar				
	Simulação	Intervalo	Confiança Analítico	Simulação	Intervalo	Confiança Analítico		
0,5	0,2744	(0,2286;	0,3202)	0,0989	0,9397	(0,7271;	1,1523)	0,9365
1,0	0,6499	(0,5834;	0,7164)	0,2294	0,8596	(0,6484;	1,0708)	0,8589
1,5	1,0489	(0,9562;	1,1416)	0,4451	0,8071	(0,6333;	0,9809)	0,8053
2,0	1,5870	(1,4706;	1,7034)	0,8374	0,7459	(0,4406;	1,0512)	0,7472
2,5	2,1071	(2,0040;	2,2102)	2,1159	0,6692	(0,4613;	0,8771)	0,6602
3,0	2,6328	(2,4867;	2,7789)	2,6241	0,5944	(0,3692;	0,8196)	0,5935

Fonte: elaborada pelo autor

TABELA 6.11 - Cenário 3 – Variação da demanda – *Call Center* Banco – 15-16 horas

Taxa chegada (ch/min)	Probabilidade de abandonar			Intensidade de tráfego				
	Simulação	Intervalo	Confiança Analítico	Simulação	Intervalo	Confiança Analítico		
0,5	0,0603	(0,0000;	0,2729)	0,0589	0,2519	(0,0000;	0,5088)	0,2695
1,0	0,1404	(0,0000;	0,3516)	0,1398	0,4801	(0,3328;	0,6274)	0,4927
1,5	0,1929	(0,0191;	0,3667)	0,1922	0,6923	(0,4522;	0,9324)	0,6941
2,0	0,2541	(0,0000;	0,5594)	0,2589	0,8274	(0,4796;	1,1752)	0,8488
2,5	0,3308	(0,1229;	0,5387)	0,3307	1,1962	(0,0000;	3,5739)	0,9384
3,0	0,4056	(0,1804;	0,6308)	0,4349	1,0405	(0,5146;	1,5664)	0,9709

Fonte: elaborada pelo autor

Os três cenários analisados serviram para revelar o potencial do uso dos modelos analíticos de fila com abandono no dimensionamento da quantidade de atendentes, para adequar o sistema ao nível de serviço desejado. Outro potencial revelado foi a possibilidade de apoiar os gestores em suas decisões diante de um possível aumento na quantidade de chamadas do *Call Center*. Os modelos de simulação de filas também foram utilizados para produzir os resultados de cada um dos cenários e mostrar a consistência dos modelos analíticos de fila, pela comparação dos resultados obtidos pelos dois tipos de modelagem.



## 7 CONCLUSÕES

### 7.1 Conclusões

Este trabalho apresentou e aplicou modelos analíticos de filas com abandono, representado por distribuições genéricas, como uma abordagem de análise efetiva para representar o problema de congestão em *Call Centers*. Em particular, investigou a utilização de distribuições mistas de probabilidades que melhor se ajustam aos tempos de paciência dos usuários de *Call Centers*. O interesse maior foi considerar as distribuições mistas nos modelos analíticos de filas  $M/M/c+G$  e  $M/G/c+G$ , para serem aplicados nos problemas de congestão de *Call Centers*, obtendo medidas de desempenho que melhor expressem e representem a realidade. O modelo analítico de fila  $MG^c/1+G$  também foi considerado neste estudo porque constitui a base para atingir as aproximações obtidas para o modelo  $M/G/c+G$  (Capítulo 3).

Vários modelos de filas que incorporam o abandono têm sido estudados na literatura, mas todos consideram uma única classe de distribuições paramétricas para modelar a paciência, sendo as mais comuns a Exponencial, Weibull e Erlang. Entretanto, é comum encontrar em *Call Center* uma população de usuários formada por grupos com características diferentes, cada um deles tendo comportamentos diferentes para o tempo de paciência, sendo uns mais pacientes que outros. Por esse motivo, em geral, é mais razoável modelar o abandono por meio de distribuições mistas de probabilidade, que são mais sensíveis para captar essas características das subpopulações de usuários.

Algumas distribuições mistas de probabilidade, tais como a Lognormal mista, a Exponencial mista e a Uniforme mista, foram utilizadas em estudos da literatura para ajustar o tempo de paciência em *Call Center*, mas nestes estudos elas não foram consideradas para validar na prática modelos analíticos de filas aplicados a *Call Center*, com dados extraídos da realidade. No presente estudo, analisa-se a aplicação desses modelos de filas, considerando distribuições genéricas e mistas para modelar a paciência dos usuários, em dois casos reais de *Call Centers*, com dados de uma empresa no Brasil e no exterior. A validação foi feita comparando-se os desvios das medidas de desempenho, obtidos por meio dos modelos analíticos de filas, e os observados com os dados reais e também estimados com os resultados de modelos de simulação discreta. Para atingir esse objetivo, estudou-se diferentes modelos analíticos de fila com abandono encontrados na literatura, que admitem distribuições de probabilidades exponenciais e genéricas no processo de serviço e de abandono.

Após a análise estatística dos dados do caso do *Call Center* do Banco, obteve-se como resultado dos testes de aderência aplicados (Capítulo 4) que a distribuição Exponencial de probabilidade se ajusta aos dados do processo de chegada. Os mesmos testes foram aplicados aos tempos de serviços dos usuários do *Call Center* e entre as várias distribuições que se ajustaram a esses dados, optou-se pelas distribuições de Fréchet e Lognormal como as mais apropriadas para representar o processo de serviço. Do mesmo modo, utilizou-se distribuições mistas e não mistas baseadas em Fatigue Life, Lognormal e Exponencial para modelar o processo de abandono, entre as possíveis distribuições ajustadas pelos testes a esses dados. Desta análise estatística, também extraiu-se as taxas de chegada, serviço e abandono em três horários de operação do Banco (11 às 12 horas, 12 às 13 horas, 15 às 16 horas). Além disso, obteve-se por meio dos dados, estimativas para medidas de desempenho do sistema, tais como: o tempo médio de espera, a probabilidade de esperar, a probabilidade de abandonar e a intensidade de tráfego.

Uma análise estatística também foi realizada com os dados do caso do Call Center da Empresa e obteve-se como resultado dos testes de aderência (Capítulo 5) que no processo de chegada a distribuição Exponencial de probabilidade se ajusta bem aos dados. Para modelar o processo de serviço, optou-se pelas distribuições Weibull e Exponencial e o processo de abandono foi modelado por mistas e não mistas baseadas em distribuições Fatigue Life, Normal e Exponencial, entre as oferecidas pelos testes. Obteve-se, também, as taxas de chegada, serviço e abandono e extraiu-se dos dados estimativas para o tempo médio de espera, a probabilidade de esperar, a probabilidade de abandonar e a intensidade de tráfego.

Esses parâmetros (taxas) foram inseridos nos modelos analíticos de fila com abandono considerados neste estudo e essas medidas de desempenho foram calculadas. Comparando-se essas medidas com aquelas extraídas dos dados reais, pode-se afirmar que alguns modelos analíticos de fila considerados nesse estudo produziram resultados bem próximos daqueles extraídos dos dados reais, com desvios menores que 1%, portanto, capturando bem a realidade dos *Call Centers*. Entretanto, nota-se que existem modelos que produziram resultados melhores em algumas medidas de desempenho do que em outras. Em alguns casos, dependendo da medida de desempenho considerada, os modelos de fila com distribuições mistas para a paciência apresentaram resultados melhores (desvios menores em relação aos dados reais) do que seus correspondentes com distribuições não mistas.

Os sistemas dos Call Centers considerados neste estudo, independentemente de qualquer medida de desempenho, puderam ser representados por modelos analíticos de fila com distribuições mistas para os tempos de paciência, que se comportaram melhor que os outros

modelos. No Call Center do Banco, o modelo que se comportou melhor, no horário das 11-12 horas, foi o  $M/Gc3/1+G4m$  (serviço Fréchet e paciência Fatigue Life mista) e no horário das 15-16 horas, foi o modelo  $M/Gc18/1+G19m$  (serviço Lognormal e paciência Fatigue Life mista). No horário das 12-13 horas, os modelos com as distribuições mistas para a paciência não foram os melhores e o que se comportou melhor foi o  $M/M2/c+G14$  (serviço Exponencial e paciência Lognormal). Considerando o Call Center da Empresa, o modelo analítico de fila que se comportou melhor foi o  $M/Gc26/1+G30m$  com serviço Weibull e paciência modelada pela distribuição Normal mista.

Em geral, pode-se concluir que o uso de modelos analíticos de fila com abandono, explorando distribuições mais elaboradas para modelar os tempos de serviço e distribuições mistas para os tempos de paciência, se mostram mais efetivos para avaliar Call Centers. Conclui-se, também, que os modelos com distribuições mistas para representar os tempos de paciência nem sempre foram os melhores em todas as medidas de desempenho, mas sempre foram bem competitivos. Quanto as aproximações disponíveis para os modelos analíticos de fila  $M/G/c+G$  e  $M/Gc/1+G$ , conclui-se que estão próximas dos valores observados na prática e também dos valores simulados e dentro dos intervalos de confiança (95%) obtidos na simulação. Por outro lado, o uso de abordagens analíticas de fila mais simples, usual na prática, baseadas em modelos  $M/M/c+M$  e  $M/G/c+M$ , não necessariamente produzem boas aproximações.

Espera-se que esta pesquisa tenha contribuído para melhor compreender a análise da paciência do usuário na fila de espera dos Call Centers ao utilizar distribuições mistas de probabilidade para modelar a paciência. Quando comparada com a abordagem de Brown et al. (2002) e Mandelbaum e Zeltyn (2012) esta tese se diferencia ao modelar a paciência dos usuários com distribuições genéricas de probabilidade, particularmente distribuições mistas, incorporando essas distribuições mistas em modelos analíticos de filas com abandono e aplicando esses modelos aos dados reais dos dois casos de Call Centers analisados. Assim, acredita-se que esta tese produziu uma expansão do conhecimento prático na área ao validar os resultados obtidos por esses modelos analíticos de fila com abandono, que incorporam distribuições mistas para modelar a paciência dos usuários, comparando-os com os dados reais extraídos de Call Center.

Este estudo apresentou limitações na elaboração das distribuições mistas para representar o fenômeno do abandono. Apenas por simplicidade, optou-se neste trabalho pela utilização de duas componentes da mesma família paramétrica na composição da distribuição mista, para representar os tempos de paciência dos usuários. Nenhum método foi utilizado para

determinar a quantidade de componentes nas distribuições mistas utilizadas neste estudo, considerando que o objetivo é analisar se a aplicação de distribuições mistas para ajustar os tempos de paciência, quando inseridos nos modelos analíticos de fila os deixa mais efetivos para representar os Call Centers dos casos analisados.

Observa-se também, como limitação, que não foram incorporados na composição dessas distribuições mistas de probabilidade fatores que afetam a impaciência dos usuários como, por exemplo, sexo, idade, tipo de serviço, níveis de paciência, entre outros, que caracterizam as subpopulações de usuários do Call Center.

O uso de modelos analíticos de fila com abandono para representar o problema de congestão em Call Centers se torna vantajoso na elaboração de projetos novos e nos casos de otimização das operações desses sistemas. Todas essas considerações mostram que os objetivos propostos nesta tese foram atingidos, apresentando e aplicando modelos analíticos de filas com abandono, representado por distribuições genéricas, particularmente mistas, como uma abordagem de análise efetiva para representar o problema de congestão em sistemas de Call Centers e apoiar decisões de dimensionamento e operação nesses sistemas.

## 7.2 Pesquisas futuras

Esta pesquisa explorou dados reais de dois *Call Centers* e utilizou modelos analíticos de filas com distribuições mistas e não mistas para modelar o tempo de paciência dos seus usuários. Como sugestão para estudos futuros, propõe-se que outras investigações sejam realizadas, envolvendo novos dados e novas arquiteturas de *Call Centers*, para reforçar a validade dos resultados aqui obtidos em outros casos.

Outra pesquisa a ser desenvolvida seria avaliar a possibilidade de modelar os tempos de atendimento também por meio de distribuições mistas de probabilidade, considerando que esses tempos devem ser influenciados pelos tipos de serviços solicitados pelos usuários, caracterizando uma possível existência de subpopulações nesse processo. Como sugestão, propõe-se utilizar o modelo desenvolvido por Cantisano (2009) para várias classes de usuários, incorporando as distribuições mistas na modelagem dos tempos de atendimento.

Ampliar os estudos dos modelos analíticos de fila  $M/G/c+G$ , buscando outras aproximações mais elaboradas e adequadas para representar com mais eficiência os *Call Centers*, considerando as distribuições mistas de probabilidades na modelagem dos tempos de paciência dos usuários, também seria um tópico interessante para pesquisas futuras.

Outras pesquisas promissoras seriam explorar o uso dos modelos analíticos de fila que comportam as distribuições mistas no processo de abandono em outras arquiteturas de *Call Centers*, especificamente com aqueles que operam em série. Ainda seguindo este raciocínio, utilizar estes modelos com filas paralelas ou modelos hipercubos de filas, provavelmente presentes nos *Call Centers* espacialmente distribuídos.

Também como outra pesquisa promissora seria utilizar outros métodos para estimar as distribuições mistas na modelagem dos tempos de paciência dos usuários, procurando captar com mais precisão as diversas características desses usuários presentes na fila de espera. Além disso também verificar qual a interferência desses modelos mais refinados no desempenho desses sistemas.

Outra linha de pesquisa interessante seria desenvolver procedimentos de otimização, baseados nos modelos analíticos de fila aqui propostos, para apoiar decisões de configuração ótima do sistema. Por exemplo, alocação mínima de capacidade no sistema de maneira a atingir valores de medidas de desempenho desejadas pelo gestor, como a probabilidade máxima de abandono ou o tempo máximo de espera na fila.





## REFERÊNCIAS

- AALEN, O. O., GJESSIN, H. Understanding the shape of the hazard rate: A process point of view. **Statistical Science**, 16(2001), 1–22.
- ANAEEL. Resolução Normativa nº 363 de 22.04.2009.
- ARENA, v.14.70. Rockwell Automation Technologies, Inc. 2014
- ARENALES, M; ARMENTANO, V; MORABITO, R; YAMASSE, H. **Pesquisa Operacional para cursos de engenharia**. Rio de Janeiro: Elsevier, 2007.
- BACELLI, F.,HEBUTERNE, G., On queues with impatient customers. In: INTERNATIONAL SYMPOSIUM ON COMPUTER PERFORMANCE, (1981), pp. 159–179.
- BARROSO, L.C.; BARROSO, M. M. A; CAMPOS. F. F; CARVALHO, M. L. B; MAIA, M. L. **Cálculo Numérico (com aplicações)**. São Paulo: Harbra, 1987.
- BERTRAND, J.W.M.; FRANSOO, J.C. Modelling and simulation: operations management research methodologies using quantitative modeling. **International Journal of Operations & Production Management**, v.22, n.3, p. 241-264, 2002.
- BOUZADA, M. A. C; Análise de cenários em um *call center* usando simulação. In: SIMPOI, 2009. **Anais**.
- BRAZ, Stan. **A Evolução do Setor**. Disponível em: < <http://www.callcenter.inf.br/tecnologia/49302/cadre-lanca-app-para-evitar-filas-de-espera/ler.aspx> >. Acesso em: 15 set. 2010.
- BRILL, P.H. An embedded level crossing technique for dams and queues. **Journal Applied Probability**, 16(1), 174 – 186, 1979.
- BRILL, Precy H. **Level Crossing Methods in Stochastic Models**. New York: Springer, 2008.
- BROWN, L.; SHEN, H. Analysis of service times for a bank call center data. **Technical Report, University of Pennsylvania**, (2002).
- BROWN, L.; GANS, N.; MANDELBAUM, A.; SAKOV, A.; SHEN, H.; ZELTYN, S.; ZHAO, L. Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective. **Working Paper**, 2002, 1-61.
- CALLCENTER.INF. **Radiografia do Mercado**. Resumo. Disponível em: < <http://www.callcenter.inf.br/> >. Acesso em: 13 abr.2015
- CALLCENTER.INF. **Cadre lança APP para evitar fila de espera**. Disponível em: <<http://www.callcenter.inf.br/tecnologia/49302/cadre-lanca-app-para-evitar-filas-de-espera/ler.aspx>>. Acesso em: 24 fev. 2013

CAMARGO, C.C. **O Futuro do Call Center.** Disponível em: <http://www.callcenter.inf.br/artigos/default.asp?pag=matintegra&matID=39857>. Acessado em: 15.09.10.

CANTISANO, G. **Aproximações para a fila M/G/s/r + G.** Dissertação mestrado. Instituto de Matemática e Estatística. Dissertação Mestrado. USP, São Paulo, 2009.

COHEN, J.M. On up and downcrossings. **Journal Applied Probability**, 14(2), 405 – 410, 1977.

COHEN, J.M.; RUBINOVITCH, M. On level crossing and cycles in dam processes. **Mathematic Operations Research**, 2(4), 297 – 310, 1977.

COX, D.R.; OAKES, D. **Analysis of Survival Data.** London: Chapman and Hall, 1990.

DEMO, P. **Metodologia do conhecimento científico.** São Paulo: Atlas, 2000.

DIEKMANN, A; JUNGBAUER-GANS, M.; KRASSNIG, H.; LORENZ, S. Social status and aggression: A field study analyzed by survival analysis. **J. Social Psych.**, 1996, 136 761-768.

DOMBACKER, C. **Queueing Models for Call Centers.** Thesis, 2010.

FELGUEIRAS, M.M. **Modelação Estatísticas com Misturas e Pseudos-Misturas.** Tese. Universidade de Lisboa, 2009.

FENDICK, K., SAKSENA, V., WHITT, W. Dependence in packet queues. **IEEE Transactions on Communications**, 37, (1989), 1173–1183.

FORBES, C; EVANS, M; HASTINGS, N; PEACOCK, B. **Statistical Distributions.** New Jersey: John Wiley and Sons, Inc, 2011.

FRÉCHET, M. Sur la loi de probabilité de l'écart maximum. **Ann. Soc. Polon. Math.** 6, 93, 1927.

GANS, N; KOOLE, G; MANDELBAUM, A. Telephone Call Centers: Tutorial, Review and Research Prospects. **Manufacturing & Service Operations Management**, vol. 5, n. 2, 2003, pp. 79-141.

GARNETT, O., MANDELBAUM, A. An Introduction to Skills-Based Routing and its Operational Complexities. **Teaching note sponsored by the Fraunhofer IAO Institute**, Stuttgart, Germany, (1999)

GARNETT, O., MANDELBAUM, A., REIMAN, M. Designing a call-center with impatient customer. **Manufacturing & Service Operatios Management**. 3, (2002), 208-227.

GROSS, D; HARRIS, C.M. **Fundamentals of Queueing Theory.** Wiley, 1985.

HALFIN, S.; WHITT, W. Heavy-Traffic Limits for Queues with Many Exponential Servers. **Operations Research**, 29, (1981), 576-587.

IBM CORPORATION. *SPSS Statistics*. Version 20.0, 2011.

IRAVANI, F; BALCIOGLU, B. Approximations for the M/G/N+G type call center. **Queueing Systems**, 58, 137-153, 2008.

JAIN, Raj. **The art of computer systems performance analysis: techniques for experimental design, measurement, simulation and modeling**. USA: John Wiley & Sons, Inc, 1991.

KALBFLEISCH, J. D.; PRENTICE, R. L. **The Statistical Analysis of Failure Time Data**. John Wiley & Sons, 1980.

KOOLE, G.; MANDELBAUM, A. Queueing Models of Call Centers: An Introduction. **Annals of Operations Research**, 113 (2002), 41-59.

LAWLESS, J.F. **Statistical Models and Methods for Lifetime Data**. USA: John Wiley & Sons, Inc, 2003.

LIMA, H. **Campanha nas nuvens: como funciona o cloud call center**. Disponível em: <http://blog.teclan.com.br/campanhas-na-nuvem-como-funciona-o-cloud-call-center/>. Acesso em: 18.12.12.

LIMA, H. **Call Center em números: métricas e como medir o desempenho**. Disponível em: <http://blog.teclan.com.br/call-center-em-numeros-metricas-e-como-medir-o-desempenho/> Acesso em: 18.12.12.

LIMA, H. **Por que desligam o telefone: funcionamento e redução das chamadas desistentes**. Disponível em: <<http://blog.teclan.com.br/por-que-desligam-o-telefone-funcionamento-e-reducao-das-chamadas-desistentes/#more-127> > Acesso em: 18 dez. 1012.

MAGALHÃES, M. N.; LIMA, A.C.P. **Noções de Probabilidade e Estatística**. São Paulo: Ed. USP, 2002.

MANCINI, L. **Call Center: estratégia para vencer**. Sumus Editorial, 2001.

MANDELBAUM, A; ZELTYN, S. Data-Stoies about (Im)Patient Customers in Tele-Queues. *Submetido para publicação*, 2012.

MANDELBAUM, A; ZELTYN, S. Staffing Many-Server Queues with Impatient Customers: Constraint Satisfaction in Call Centers. **Operations Research**, 57 (5), 2009, 1189-1205.

MANDELBAUM, A.; ZELTYN, S. The Palm/Erlang-A Queue with Applications to Call Centers. **Working paper**, 2005.

MANDELBAUM, A.; SAKOV, A; ZELTYN, S. Empirical Analysis of a Call Center. **Technical Report**, Technion, Israel Institute of Tecnology, 2001.

MARTINS, Roberto A. Princípios da pesquisa científica. In: MIGUEL, Paulo A.M (org.). **Metodologia de pesquisa em engenharia de produção e gestão de operações**. Rio de Janeiro: Elsevier, 2010.

MATHEMATICA 9. Wolfram Research, Inc. v.9.0.1.0, 2013.

MATHWAVE TECHNOLOGIES. EasyFit Professional. Version 5.5. 2010

MICROSOFT CORPORATION. Excel 2013

MORABITO, R.; PUREZA, V. Modelagem e Simulação. In: MIGUEL, P.A.M (Org.). **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações**. Rio de Janeiro: Elsevier, 2010. p.165-194.

NADARAJAH, S; KOTZ, S. **Extreme Value Distributions: Theory and Applications. The exponentiated type distribution**. London: Imperial College Press, 2000.

OLIVEIRA, Camila Cardoso. **Espera e Abandono na fila M/M/n+G e variantes**. Instituto de Matemática e Estatística. Dissertação Mestrado. USP, São Paulo, 2009.

PALM, C., Methods of judging the annoyance caused by congestion. **Tele**, 4, (1953), 189–208.

PALM, C. Research on telephone traffic carried by full availability groups. **Tele**, 1, 107 pp, 1957.

PIERRI, K. **Hora de planejar: a importância do dimensionamento do Call Center**. Disponível em: <http://blog.teclan.com.br/hora-de-planejar-a-importancia-do-dimensionamento-do-call-center/> . Acesso em: 18.12.12.

PIERRI, K. **Métricas de tempo no Call Center**. Disponível em: <http://blog.teclan.com.br/metricas-de-tempo-no-call-center/#more-302> > Acesso em: 18 dez. 2012.

STANFORD, R.E. Reneging phenomenon in single channel queues. **Mathematics Operations Research**, 4(2), 162 – 178, 1979.

THIERRY, M. Subjective importance of goal and reactions to waiting in line. **J. Social Psych.** 819-827, 1994.

TITTERINGTON, D.M.; SMITH, A.F.M.; MAKOV, U.E. **Statistical Analysis of Finite Mixture Distributions**. John Wiley & Sons, 1985.

ZHANG, B.; LEEUWAARDEN, L.S.H; ZWART. Staffing Call Center with Impatient Customers: Refinements to Many-Server Asymptotics. **Operations Research**, 60, 461-474, 2012.

ZELTYN, S; MANDELBAUM, A. Call Centers with Impatient Customers: Many-Server Asymptotics of the M/M/n+G Queue. **Queueing Systems**, 51, 361-402, 2005.

ZELTYN, S. **Call Center with Impatient Customer: exact analysis and many server asymptotics of M/M/n+G queue**. Thesis, 2004.

ZOHAR, E., MANDELBAUM, A., SHIMKIN, N., Adaptive Behavior of Impatient Customers in Tele-queues: Theory and Empirical Support. **Management Science**, 48(4), (2002), 556–583.

WHITT, W. Approximations for the GI/G/m queue. **Production and Operations Management**, 2, (1993),114–161.

\_\_\_\_\_ **Stochastic-Process Limits**, Springer-Verlag, (2002)

WOLFF, Ronald W. **Stochastic Modeling and the Theory of Queues**. Prentice Hall, Inc,1989.



## APÊNDICE - A

### Análise dos dados do *Call Center* do Banco

Este apêndice apresenta em detalhes alguns cálculos realizados sobre os dados do *Call Center* do banco, analisando as chamadas mensais estratificadas pelo tipo de chamada (*Agent*, *Hang* ou *Phantom*), tempo de espera na fila, tempo de serviço e tempo de IVR. Nesta análise utilizou-se as frequências absolutas e relativas. Descreve-se também, os processos de chegada, abandono e de serviço por meio de sua média, mediana e desvio padrão. Uma distribuição das chamadas em cada um dos meses do ano de 1999 é apresentada na Tabela 1.

TABELA 1 - Quantidade chamadas mensais

Mês	Chamadas	Porcentagem
Janeiro	31599	7,11%
Fevereiro	33344	7,50%
Março	38801	8,73%
Abril	32036	7,21%
Maió	39553	8,90%
Junho	37935	8,54%
Julho	39022	8,78%
Agosto	42078	9,47%
Setembro	31371	7,06%
Outubro	34625	7,79%
Novembrc	41019	9,23%
Dezembro	43065	9,69%
Total	444448	100,00%

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

As chamadas também foram estratificadas de acordo com os resultados possíveis (Tabela 2). Aquelas chamadas que atingiram o agente (codificado como AGENT), fizeram diretamente da IVR ou depois de ter aguardado na fila. As chamadas que abandonaram o *Call Center* (codificado como HANG), fizeram diretamente da IVR ou depois de aguardarem uma quantidade positiva de tempo na fila. As chamadas que não foram identificadas e abandonaram diretamente da IVR, foram classificadas como PHANTOM. Observa-se que a quantidade de chamadas do último tipo é bastante pequena (0,8%) e, portanto, foram consideradas

desprezíveis na análise dos dados. A maioria das chamadas (79,5%) dialogam com os agentes, contra 19,5% delas que abandonam o sistema.

TABELA 2 - Distribuição das chamadas por resultados (Agent – Hang - Phantom)

Mês	Agent	Hang	Phantom	Total
Janeiro	27060	4299	240	31599
	85,64%	13,60%	0,76%	100,00%
Fevereiro	27162	5904	278	33344
	81,46%	17,71%	0,83%	100,00%
Março	27847	10547	407	38801
	71,77%	27,18%	1,05%	100,00%
Abril	23583	8148	305	32036
	73,61%	25,43%	0,95%	100,00%
Maiο	29923	9212	418	39553
	75,65%	23,29%	1,06%	100,00%
Junho	31193	6420	322	37935
	82,23%	16,92%	0,85%	100,00%
Julho	29596	9048	378	39022
	75,84%	23,19%	0,97%	100,00%
Agosto	31402	10366	310	42078
	74,63%	24,64%	0,74%	100,00%
Setembro	27062	4146	163	31371
	86,26%	13,22%	0,52%	100,00%
Outubro	30861	3617	147	34625
	89,13%	10,45%	0,42%	100,00%
Novembro	33371	7351	297	41019
	81,35%	17,92%	0,72%	100,00%
Dezembro	34083	8648	334	43065
	79,14%	20,08%	0,78%	100,00%
<b>Total</b>	<b>353143</b>	<b>87706</b>	<b>3599</b>	<b>444448</b>
	<b>79,46%</b>	<b>19,73%</b>	<b>0,81%</b>	<b>100,00%</b>

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

No *Call Center* analisado, 59,2% das chamadas tiveram que aguardar uma certa quantidade positiva de tempo na fila ( $Q\_TIME > 0$ ) para ser atendida pelo agente. As demais chamadas (40,8%) conseguiram realizar o seu serviço sem a necessidade de aguardar na fila de espera ( $Q\_TIME = 0$ ). Nesses casos, o usuário após ter encerrado as suas atividades na VRU,



encontrou o sistema vazio e iniciou o seu atendimento imediatamente. A quantidade de chamadas por *status* de espera é informada na Tabela 3.

TABELA 3 - Quantidade de chamadas por *status* de espera

<b>Mês</b>	<b>Q-time = 0</b>	<b>Q_time &gt; 0</b>	<b>Total</b>
Janeiro	14694 46,50%	16905 53,50%	31599 100,00%
Fevereiro	15038 45,10%	18306 54,90%	33344 100,00%
Março	10635 27,41%	28166 72,59%	38801 100,00%
Abril	10691 33,37%	21345 66,63%	32036 100,00%
Maiο	13031 32,95%	26522 67,05%	39553 100,00%
Junho	15942 42,02%	21993 57,98%	37935 100,00%
Julho	11975 30,69%	27047 69,31%	39022 100,00%
Agosto	16116 38,30%	25962 61,70%	42078 100,00%
Setembro	17704 56,43%	13667 43,57%	31371 100,00%
Outubro	22578 65,21%	12047 34,79%	34625 100,00%
Novembro	16643 40,57%	24376 59,43%	41019 100,00%
Dezembro	16176 37,56%	26889 62,44%	43065 100,00%
<b>Total</b>	<b>181223</b> <b>40,77%</b>	<b>263225</b> <b>59,23%</b>	<b>444448</b> <b>100,00%</b>

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

A Tabela 4 sumariza os resultados das chamadas que atingiram o agente diretamente da VRU (AGENT e Q\_TIME = 0), as que atingiram o agente e esperaram na fila (AGENT e Q\_TIME > 0), as chamadas que abandonaram na VRU (HANG e Q\_TIME = 0), as que abandonaram da fila (HANG e Q\_TIME > 0) e também as chamadas identificadas como PHANTOM.

TABELA 4 - Resultados por espera (A – AGENT, H – HANG, P – PHANTOM e  
Q\_TIME(Q)=0, Q\_TIME(Q)>0)

Mês	A, Q=0	A, Q>0	H, Q=0	H, Q>0	Phantom	Total
Janeiro	13475 42,64%	13585 42,99%	1219 3,86%	3080 9,75%	240 0,76%	31599 100,00%
Fevereiro	13066 39,19%	14096 42,27%	1972 5,91%	3932 11,79%	278 0,83%	33344 100,00%
Março	8669 22,34%	19178 49,43%	1966 5,07%	8581 22,12%	407 1,05%	38801 100,00%
Abril	8202 25,60%	15381 48,01%	2489 7,77%	5659 17,66%	305 0,95%	32036 100,00%
Mai	10892 27,54%	19031 48,12%	2139 5,41%	7073 17,88%	418 1,06%	39553 100,00%
Junho	14157 37,32%	17036 44,91%	1785 4,71%	4635 12,22%	322 0,85%	37935 100,00%
Julho	9714 24,89%	19882 50,95%	2261 5,79%	6787 17,39%	378 0,97%	39022 100,00%
Agosto	13010 30,92%	18392 43,71%	3106 7,38%	7260 17,25%	310 0,74%	42078 100,00%
Setembro	16113 51,36%	10949 34,90%	1591 5,07%	2555 8,14%	163 0,52%	31371 100,00%
Outubro	21050 60,79%	9811 28,34%	1528 4,41%	2089 6,03%	147 0,42%	34625 100,00%
Novembro	14787 36,05%	18584 45,31%	1856 4,52%	5495 13,40%	297 0,72%	41019 100,00%
Dezembro	14001 32,51%	20082 46,63%	2175 5,05%	6473 15,03%	334 0,78%	43065 100,00%
<b>Total</b>	<b>157136</b> <b>35,36%</b>	<b>196007</b> <b>44,10%</b>	<b>24087</b> <b>5,42%</b>	<b>63619</b> <b>14,31%</b>	<b>3599</b> <b>0,81%</b>	<b>444448</b> <b>100,00%</b>

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

As três fontes que contribuem para o tempo total gasto no sistema, VRU\_TIME, Q\_TIME e SER\_TIME, estão combinadas na Tabela 5. Os tempos negativos para VRU (VRU\_TIME < 0) são considerados dados inconsistentes e foram desprezados devido a sua baixa ocorrência 0,02% e 0,06%.

TABELA 5 - Combinando VRU, fila e tempo de serviço (VRU\_TIME, Q - Q\_TIME e S - SER\_TIME)

Mês	VRU<0		VRU=0				VRU>0				Total
	Q>0		Q=0		Q>0		Q=0		Q>0		
	S=0	S>0	S=0	S>0	S=0	S>0	S=0	S>0	S=0	S>0	
Janeiro	11	22	328	2	127	189	1064	13300	2978	13578	31599
Fevereiro	2	23	355	2	55	255	1653	13028	3828	14143	33344
Março	8	19	608	0	107	419	1380	8647	8420	19193	38801
Abril	6	31	600	2	44	224	1902	8187	5561	15479	32036
Maiο	30	48	558	2	48	268	1601	10870	6913	19215	39553
Junho	4	23	433	1	47	264	1396	14112	4563	17092	37935
Julho	12	34	598	2	55	264	1692	9683	6738	19944	39022
Agosto	4	15	547	1	61	124	2602	12966	7151	18607	42078
Setembro	4	10	279	0	29	81	1370	16055	2529	11014	31371
Outubro	2	2	186	2	19	63	1370	21020	2039	9922	34625
Novembro	2	18	408	1	100	174	1463	14771	5338	18744	41019
Dezembro	3	17	490	0	93	181	1708	13978	6338	20257	43065
<b>Total</b>	<b>88</b>	<b>262</b>	<b>5390</b>	<b>15</b>	<b>785</b>	<b>2506</b>	<b>19201</b>	<b>156617</b>	<b>62396</b>	<b>197188</b>	<b>444448</b>
	<b>0,02%</b>	<b>0,06%</b>	<b>1,21%</b>	<b>0,00%</b>	<b>0,18%</b>	<b>0,56%</b>	<b>4,32%</b>	<b>35,24%</b>	<b>14,04%</b>	<b>44,37%</b>	<b>100,00%</b>

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

## O Processo de Chegada

O processo de chegada registra os instantes em que as chamadas telefônicas chegam no *Call Center*. Ele pode ser descrito com diferentes níveis de detalhes, mas fez-se apenas uma descrição determinística sem considerar a variabilidade estocástica. O interesse foi analisar a distribuição das chamadas de acordo com os diferentes tipos de serviço, como apresentado na Tabela 6. Existem duas chamadas em janeiro, uma em abril, uma em julho e uma em setembro, todas classificadas como tipo AA, um tipo de serviço desconhecido (e inexistente), que foram desconsideradas nas análises e atribuídas a erros do sistema. Por isso, existem somente 444443 chamadas no total.

Observa-se nos dados da Tabela 6 um aumento da quantidade de chamadas do tipo IN (suporte técnico da internet) durante todo o ano, principalmente um salto nos meses de julho e agosto. A explicação dada por Mandelbaum et al. (2001) foi a expansão dos serviços de internet feita pelo Banco, que afetaram os usuários, exigindo mais suporte técnico.

TABELA 6 - Distribuição das chamadas por tipo de serviço e por mês

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
<b>PS</b>	22998 72,8%	21460 64,4%	25910 66,8%	22551 70,4%	28482 72%	26608 70,1%	27269 69,9%	27312 64,9%	20810 66,3%	23145 66,8%	27156 66,2%	28821 66,9%	302522 68,1%
<b>NW</b>	5582 17,7%	7604 22,8%	6732 17,4%	4760 14,9%	5126 13%	5832 15,4%	5723 14,7%	6923 16,5%	4340 13,8%	5222 15,1%	5104 12,4%	4780 11,1%	67728 15,2%
<b>NE</b>	1586 5%	2750 8,2%	3490 9%	2776 8,7%	4011 10,1%	3776 10%	3536 9,1%	3381 8%	2635 8,4%	2815 8,1%	4418 10,8%	4168 9,7%	39342 8,9%
<b>IN</b>	295 0,9%	293 0,9%	503 1,3%	451 1,4%	492 1,2%	483 1,3%	1202 3,1%	3563 8,5%	3139 10%	2987 8,6%	3257 7,9%	4067 9,4%	20732 4,7%
<b>TT</b>	972 3,1%	1088 3,3%	2003 5,2%	1380 4,3%	1296 3,3%	1080 2,8%	1131 2,9%	718 1,7%	333 1,1%	315 0,9%	914 2,2%	1065 2,5%	12295 2,8%
<b>PE</b>	164 0,5%	149 0,4%	163 0,4%	117 0,4%	146 0,4%	156 0,4%	160 0,4%	181 0,4%	113 0,4%	141 0,4%	170 0,4%	164 0,4%	1824 0,4%
<b>Total</b>	<b>31597</b>	<b>33344</b>	<b>38801</b>	<b>32035</b>	<b>39553</b>	<b>37935</b>	<b>39021</b>	<b>42078</b>	<b>31370</b>	<b>34625</b>	<b>41019</b>	<b>43065</b>	<b>444443</b>

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

### Processo de espera para serviço ou abandono

Segundo Mandelbaum et al. (2001), o tempo de espera é uma medida de grande interesse no *Call Center*. Se o tempo de espera na fila for igual a zero, indica um sucesso operacional dos serviços prestados pelo *Call Center*. Mas, o tempo de espera positivo (os usuários aguardam na fila) é a manifestação de um descompasso entre a demanda (solicitações de serviços pelos usuários) e a disponibilidade dos agentes. Esse descompasso pode causar o abandono dos usuários que pode ser percebido como uma falha do sistema.

As estatísticas do tempo de espera estão sumarizadas na Tabela 7, que consideram todas as chamadas que aguardaram na fila ( $Q\_TIME > 0$ ) e que foram servidas ou abandonaram o sistema. Observa-se que a distribuição desses tempos de espera é assimétrica para a direita (média > mediana). As mesmas estatísticas foram calculadas para o tempo de espera dos que abandonam (Tabela 8) e para o tempo de espera dos usuários que foram servidos (Tabela 9). A distribuição desses tempos de espera também é assimétrica para a direita (média > mediana), em ambos os casos.

TABELA 7 - Sumário das estatísticas para tempo de espera ( $Q\_TIME > 0$ )

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Anual
<b>Média</b>	68	76	120	110	97	88	106	116	85	76	102	113	100
<b>Med.</b>	46	50	75	72	62	55	69	72	54	45	63	71	62
<b>DP</b>	75	78	130	113	104	271	110	130	93	247	114	125	142
<b>Total</b>	<b>16905</b>	<b>18306</b>	<b>28166</b>	<b>21345</b>	<b>26522</b>	<b>21993</b>	<b>27047</b>	<b>25962</b>	<b>13667</b>	<b>12047</b>	<b>24376</b>	<b>26889</b>	<b>263225</b>

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

TABELA 8 - Sumário das estatísticas para tempo de espera dos que abandonam

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Anual
<b>Média</b>	61	66	85	85	72	71	82	90	76	79	80	86	79
<b>Med.</b>	39	43	56	56	43	47	53	54	51	49	52	56	51
<b>DP</b>	95	73	101	104	91	80	103	129	101	112	100	120	104
<b>Total</b>	<b>3080</b>	<b>3932</b>	<b>8581</b>	<b>5659</b>	<b>7073</b>	<b>4635</b>	<b>6787</b>	<b>7260</b>	<b>2555</b>	<b>2089</b>	<b>5495</b>	<b>6473</b>	<b>63619</b>

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

TABELA 9 - Sumário das estatísticas para tempo de espera dos servidos

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Anual
<b>Média</b>	70	79	135	118	105	89	113	126	86	73	107	120	105
<b>Med.</b>	48	52	91	83	71	58	78	83	55	44	68	80	67
<b>DP</b>	67	79	138	114	105	91	110	128	91	87	116	125	111
<b>Total</b>	<b>13585</b>	<b>14096</b>	<b>19178</b>	<b>15381</b>	<b>19031</b>	<b>17036</b>	<b>19882</b>	<b>18392</b>	<b>10949</b>	<b>9811</b>	<b>18584</b>	<b>20082</b>	<b>196007</b>

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

Várias ramificações dos tempos de espera dos usuários são descritas por Mandelbaum et al. (2001). Primeiramente eles consideram a distinção entre o tempo de fila e o tempo de espera. O tempo de espera não considera as esperas zero e é mais relevante para os gestores, especialmente quando considerado junto com a fração dos usuários que esperaram. A segunda distinção, e mais fundamental, é entre o tempo de espera dos usuários que foram servidos e aqueles que abandonaram o sistema. Neste caso, é importante reconhecer que o abandono não

descreve a paciência dos usuários. A terceira distinção está entre o tempo virtual de espera e o tempo de paciência do usuário. O tempo virtual de espera é o tempo que um usuário necessita aguardar antes de ser atendido por um agente. É, então, a quantidade de tempo que um usuário (virtual), equipado com infinita paciência, teria que aguardar até ser servido. O tempo de paciência do usuário é o tempo que um usuário suporta aguardar antes de abandonar o sistema. Serve como medida operacional da paciência do usuário. São duas medidas de grande importância, mas que não são diretamente observáveis e devem ser estimadas.

Para os usuários que são pacientes o suficiente para atingir um agente, o tempo de espera é uma amostra do tempo necessário para aguardar, mas é apenas um limite inferior do tempo que os usuários suportam aguardar. Neste caso, diz-se que a paciência dos usuários (tempo que suporta aguardar) é censurada pelo tempo de espera para serviço (tempo necessário para aguardar), onde somente o último (tempo de serviço) é diretamente observável. Para os usuários que abandonam, neste caso, observável é o tempo que eles suportam aguardar, que censura o tempo necessário para aguardar. A estimativa da paciência dos usuários, que é um dado censurado, é feita pelas técnicas da Análise de Sobrevivência.

### **Tempo de serviço: o processo de serviço**

Esta é a última fase de uma chamada no *Call Center*. A Tabela 10 fornece as estatísticas dos tempos de serviço, ignorando registros com tempo de serviço zero. Também neste caso, a distribuição desses tempos é assimétrica à direita (média > mediana).

TABELA 10 - Estatísticas para tempo de serviço

	<b>Jan</b>	<b>Fev</b>	<b>Mar</b>	<b>Abr</b>	<b>Mai</b>	<b>Jun</b>	<b>Jul</b>	<b>Ago</b>	<b>Set</b>	<b>Out</b>	<b>Nov</b>	<b>Dez</b>	<b>Anual</b>
<b>Média</b>	184	175	184	211	197	189	182	183	191	181	196	207	190
<b>Med.</b>	113	106	110	123	117	115	112	110	111	105	117	128	114
<b>DP</b>	230	217	239	365	259	313	301	362	296	437	272	273	304
<b>Total</b>	<b>27091</b>	<b>27451</b>	<b>28278</b>	<b>23923</b>	<b>30403</b>	<b>31492</b>	<b>29927</b>	<b>31713</b>	<b>27160</b>	<b>31009</b>	<b>33708</b>	<b>34433</b>	<b>356588</b>

Fonte: dados de Mandelbaum, Sakov e Zeltyn (2001), que foram reproduzidos pelo autor.

## APÊNDICE - B

### Comparando os modelos com e sem abandono

O propósito da comparação dos modelos de filas com abandono Erlang-A ( $M/M/c+M$ ) e o modelo sem abandono Erlang-C ( $M/M/c$ ) é mostrar a significância do abandono, por meio de exemplos numéricos. Em ambos os modelos, considerou-se como parâmetros os dados da análise empírica discutidos na seção 4.2.2 e no Apêndice 1. Os três intervalos de horas também foram mantidos com suas respectivas taxas de serviço e quantidade de agentes. Variou-se também a taxa de chegada para esses dois modelos, de 0,1 até 5,0 chamadas por minuto, com intervalos de 0,1 por exemplo, a taxa de chegada observada no intervalo das 11 – 12 horas foi de 1,933 chamadas por minuto, destacada no Gráfico 4.4. No modelo Erlang-A ( $M/M/c+M$ ), a taxa de abandono de 0,111 chamadas por minuto foi considerada, como anteriormente, em cada um dos três intervalos de horas. Essas informações estão resumidas na Tabela 4.20.

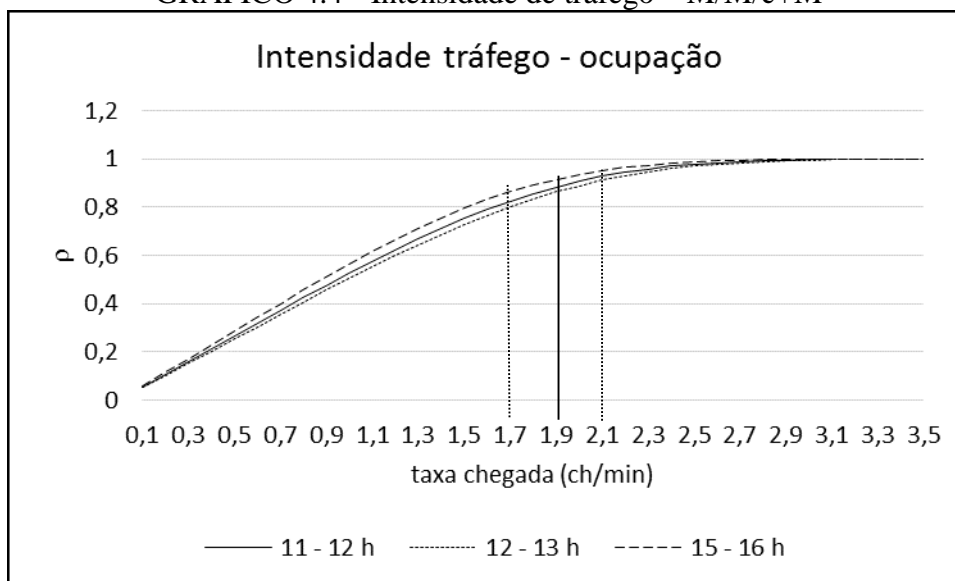
TABELA 4.20 - Dados utilizados na comparação dos modelos Erlang-A e Erlang-C

	Erlang-A ( $M/M/c+M$ ) e Erlang-C ( $M/M/c$ )		
	Dados 11 – 12 h	Dados 12 – 13 h	Dados 15 – 16 h
Taxa serviço ( $\mu$ )	0,314 ch/min	0,328 ch/min	0,291 ch/min
Quantidade servidores( $c$ )	6	6	6
Taxa abandono ( $\theta$ )	0,111 ch/min	0,111 ch/min	0,111 ch/min

Fonte: elaborado pelo autor

Analisando inicialmente o modelo  $M/M/c+M$ , nota-se no Gráfico 4.4 que o sistema mantém-se sempre estável. No intervalo das 15 às 16 horas, o *Call Center* analisado está mais congestionado que nos outros dois horários. Considerando-se as taxas de chegada em cada um dos três horários (1,933 ch/min, 1,717 ch/min e 2,1 ch/min, respectivamente), o sistema opera com um tráfego variando de 0,80727 a 0,95094, permitindo classificar o horário das 12 às 13 horas como o de congestionamento moderado, o das 11 às 12 horas como relativamente alto e o das 15 às 16 horas como alto.

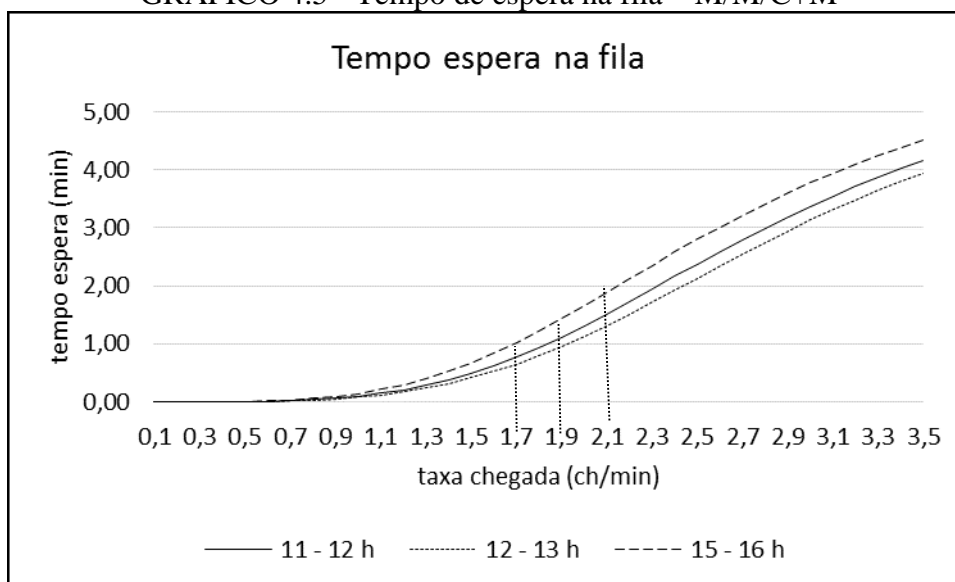
GRÁFICO 4.4 - Intensidade de tráfego – M/M/c+M



Fonte: adaptado de MANDELBAUM, SAKOV E ZELTYN, 2001

Os Gráficos 4.5 e 4.6 mostram a dependência do tempo de espera e da probabilidade de esperar em relação a ocupação do modelo M/M/c+M. Observe que, quanto maior for a taxa de chegada (maior ocupação), maior também é o tempo médio de espera e maior é a probabilidade do usuário abandonar o sistema. Este fato ocorre nos três horários analisados.

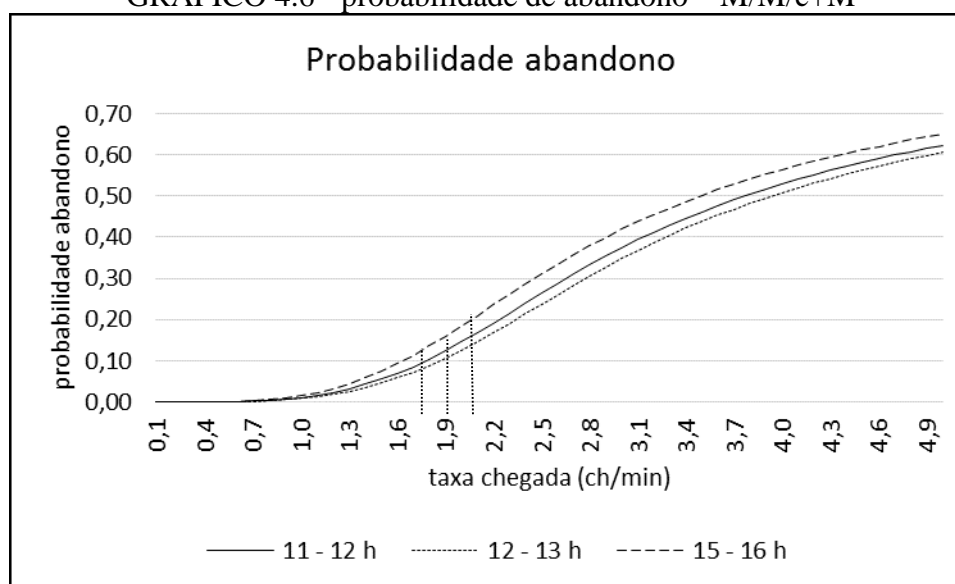
GRÁFICO 4.5 - Tempo de espera na fila – M/M/C+M



Fonte: adaptado de MANDELBAUM, SAKOV E ZELTYN, 2001



GRÁFICO 4.6 - probabilidade de abandono – M/M/c+M



Fonte: adaptado de MANDELBAUM, SAKOV E ZELTYN, 2001

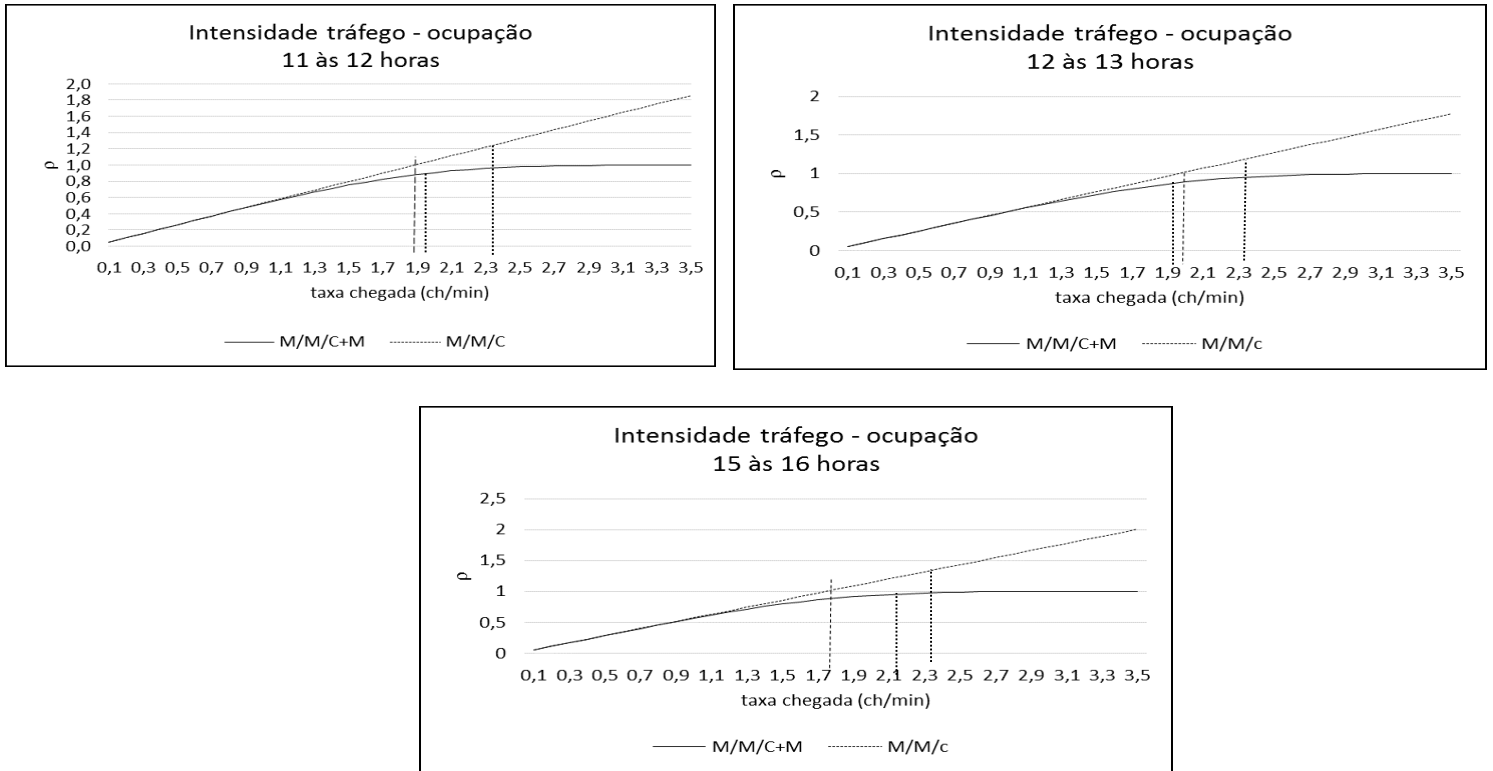
Esses dois últimos gráficos (Gráficos 4.5 e 4.6) fornecem uma informação empírica importante para a gestão do *Call Center* do banco: diminuindo-se o tempo de espera na fila, ocorrem poucos abandonos e como consequência há uma melhora no nível de serviço percebido pelos usuários. A comparação dos modelos com abandono (M/M/c+M) e sem abandono (M/M/c) inicia-se nos três gráficos da Figura 4.6, que mostra a intensidade de tráfego, quando calculada com o modelo Erlang-C (M/M/c) e o seu correspondente Erlang-A (M/M/c+M). Em ambos os modelos utilizou-se os dados constantes da Tabela 4.20. As duas curvas, em cada um dos três intervalos de horas, mostram que o modelo Erlang-C fica instável ( $\rho > 1$ ) a partir de uma determinada taxa de chegada, enquanto o modelo Erlang-A permanece sempre estável. No intervalo das 11 às 12 horas, o modelo Erlang-C fica instável a partir de uma taxa de chegada de 1,9 chamadas por minuto ( $\rho = 1,0085$ ), enquanto que no intervalo das 12 às 13 horas, ele fica instável a partir de 2 chamadas por minuto ( $\rho = 1,0163$ ), e no intervalo das 15 às 16 horas, a partir de 1,8 chamadas por minuto ( $\rho = 1,0309$ ). Essas taxas estão assinaladas com um traço vertical nos gráficos que compõe a Figura 4.6.

A importância em considerar-se o abandono nos modelos analíticos de filas é demonstrada nos três gráficos da Figura 4.7. Nela, compara-se o tempo médio de espera na fila obtido por meio do modelo Erlang-C (M/M/c) e o correspondente Erlang-A (M/M/c+M), que incorpora o abandono nas suas equações. Nos dois modelos também foram utilizados os dados da Tabela 4.20. Nota-se que as duas curvas têm comportamentos semelhantes em cada um dos três horários analisados. Em cada um dos três gráficos da Figura 4.7, ocorre uma sobreposição

das duas curvas para taxas de chegada aproximadamente inferiores a 1,2 chamadas por minuto, que corresponde nos dois modelos a uma intensidade de tráfego moderada ( $\rho \cong 0,6$ ). Desse modo, deduz-se que não há diferença entre os tempos de espera do modelo que considera o abandono (Erlang-A) com o modelo que não considera (Erlang-C), quando a intensidade de tráfego é moderada ou baixa. Portanto, é vantajoso utilizar o modelo Erlang-C (M/M/c), que não considera o abandono dos usuários, para representar *Call Centers* que possuem intensidades de tráfego baixas ou moderadas, devido a sua simplicidade.

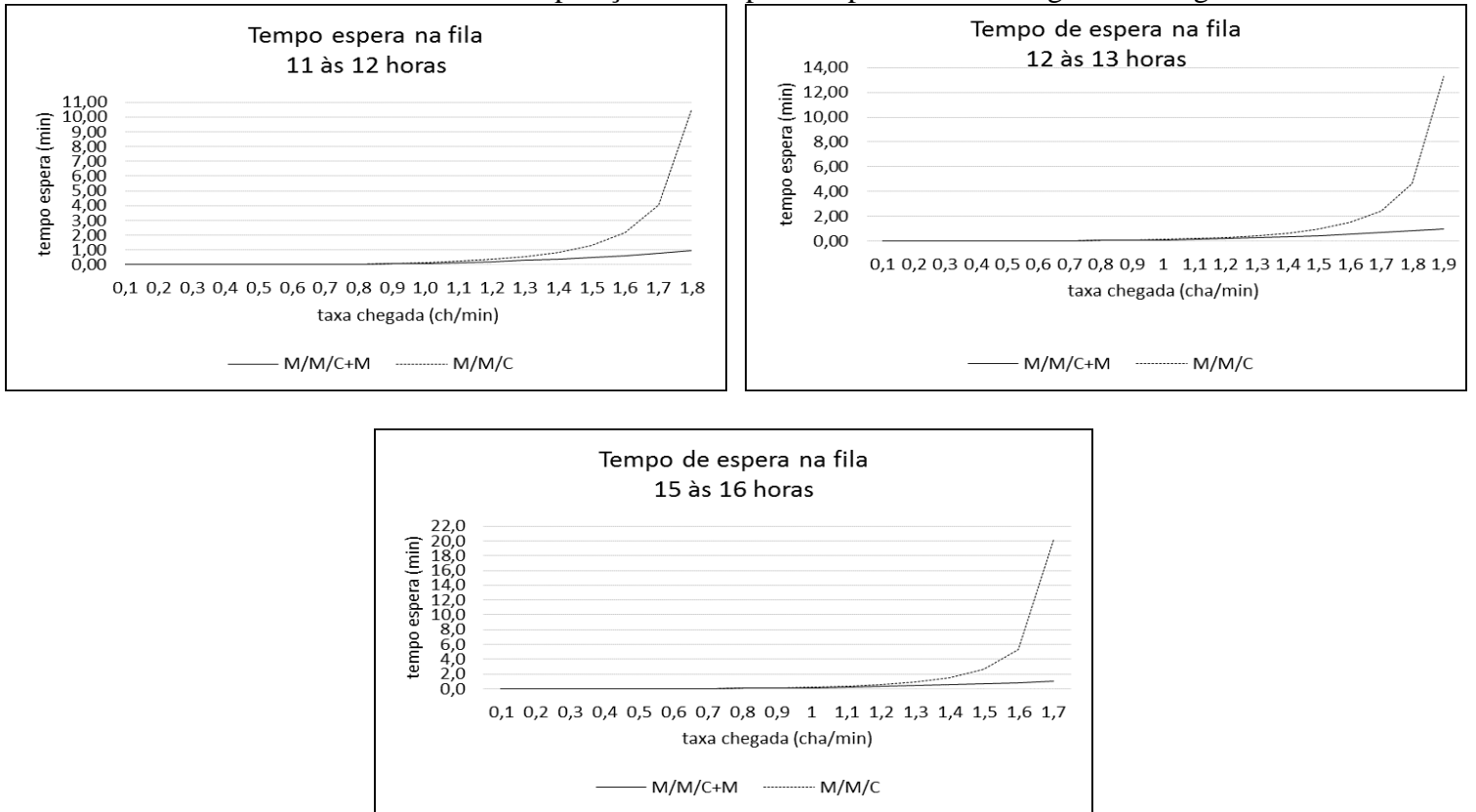
Quando a intensidade de tráfego é alta (região da Figura 4.7 com taxa de chegada maior que 1,2 chamadas por minuto), existe uma diferença significativa entre o tempo médio de espera em ambos os modelos. Claramente se observa que o modelo Erlang-A (M/M/c+M), que incorpora o abandono, possui tempos de espera sempre inferiores ao Erlang-C (M/M/c), e a diferença vertical entre as duas curvas torna-se maior quanto mais elevada for a taxa de chegada (maior intensidade de tráfego). Limitou-se, nos gráficos da Figura 4.7, a taxa de chegada próxima de 1,8 chamadas por minuto, porque acima deste valor o modelo Erlang-C (M/M/c) torna-se inconsistente com os dados. Desse modo, ao considerar-se o abandono dos usuários, reduz-se o tempo de espera daqueles que estão na fila e, também, dos que entrarão no sistema futuramente. A Figura 4.7 demonstra a significância quantitativa do abandono e que o modelo Erlang-A (M/M/c+M) deve ser preferido, principalmente para modelar situações em que a intensidade de tráfego é relativamente alta, o que é comum em diversos *Call Centers*. É importante ressaltar que se o abandono não for considerado, pode-se chegar a decisões incorretas na gestão de um Call Center, como por exemplo, superdimensionar a equipe de atendentes (por exemplo, considerando-se atendentes para “servir” os usuários que abandonam o sistema).

FIGURA 4.6 - Comparação da intensidade de tráfego entre Erlang-A e Erlang-C



Fonte: elaborada pelo autor

FIGURA 4.7 - Comparação do tempo de espera entre Erlang-A e Erlang-C



Fonte: elaborada pelo autor

Outra análise da significância do abandono foi realizada com os modelos Erlang-A e Erlang-C, comparando os resultados obtidos, com os dados da análise empírica do *Call Center* do banco discutidos na secção 4.2.2. Considerou-se nesta análise os parâmetros do período das 12 às 13 horas porque, neste período, o modelo Erlang-C (M/M/c) encontra-se estável, conforme mostrado nas figuras anteriores. Os resultados estão exibidos na Tabela 4.21.

TABELA 4.21: Comparação dos modelos Erlang-A e Erlang-C, com redução da taxa de chegada

Dados 12 – 13 hs	M/M/c+M	M/M/c	M/M/c $\lambda \downarrow 7,5\%$
Taxa chegada (ch/min)	1,717	1,717	1,588
Taxa serviço (ch/min)	0,328	0,328	0,328
Taxa abandono (ch/min)	0,111	---	---
Quant. Agentes	6	6	6
Tempo médio espera (min)	0,673	2,689	1,399
Probabilidade abandono	7,5%	---	---
Ocupação agentes	80,7%	87,3%	80,7%

Fonte: elaborado pelo autor

Observa-se na Tabela 4.21 que a taxa de chegada de 1,717 chamadas por minuto induz um abandono de 7,5% dos usuários do *Call Center*. Se for descontado esses abandonos do modelo Erlang-C (M/M/c), isto é, diminuir a sua taxa de chegada em 7,5%, não há redução significativa da intensidade de tráfego, que continua sendo alta (80,7%), igualando-se à intensidade de tráfego do modelo Erlang-A (ver última coluna da Tabela 4.21). O tempo de espera na fila sofre uma redução para 1,39 minutos, no modelo Erlang-C, mas continua muito distante quando comparado com os 0,673 minutos do modelo Erlang-A. Como consequência, pode-se afirmar que o alto nível de desempenho do modelo Erlang-A não é atingido quando a taxa de chegada do modelo Erlang-C for diminuída de 7,5%.

## APÊNDICE - C

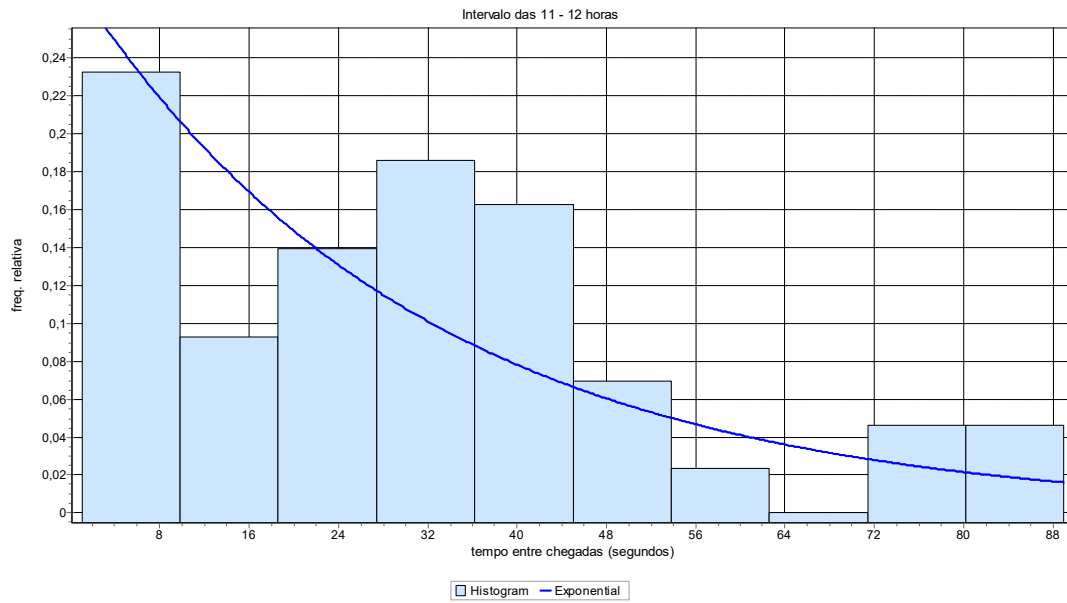
### **Análise do processo de chegada, abandono e serviço do caso Call Center do Banco**

Neste apêndice se reproduz os resultados de um estudo realizado por Mandelbaum et al. (2001) e Brown et al. (2002), com relação aos processos de chegada, abandono e serviço, tendo como fonte de dados do *Call Center* de um Banco. Todos os dados deste *Call Center* foram gentilmente disponibilizados pelos autores.

#### **O processo de chegada**

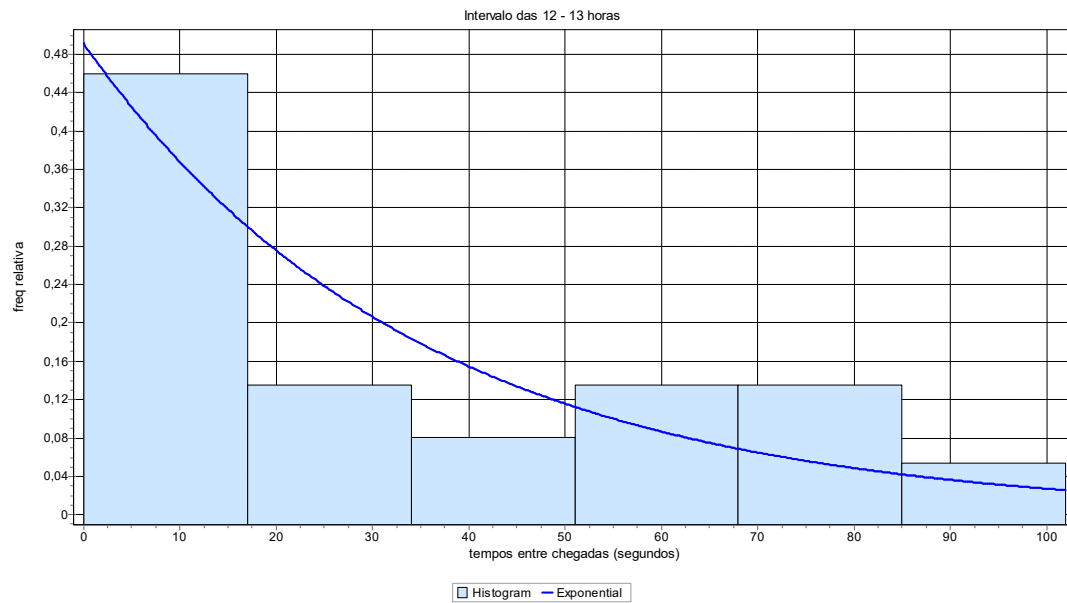
A distribuição dos tempos entre chegadas das chamadas telefônicas se aproxima da distribuição exponencial. Isto é consistente com os valores dos coeficientes de variação que são próximos da unidade (a distribuição exponencial tem seu coeficiente de variação igual a 1), principalmente nos dois últimos horários. Outra comprovação desta aproximação é dada pela razão entre a mediana e a média que, em cada um dos intervalos de hora é 0,93: 0,69 e 0,73, respectivamente. Na distribuição exponencial esta razão é igual a  $\ln 2$  ( $\ln 2 = 0,69$ ) (Apêndice E). Os histogramas e densidades dos Gráficos 4.7, 4.8 e 4.9 tem a configuração da distribuição exponencial de probabilidades, em cada um dos intervalos de horas analisados. Este fato deixa evidente que os tempos entre chegadas das chamadas telefônicas no *Call Center* do banco podem ser aproximados pela distribuição exponencial, o que foi considerado em ambos os modelos M/M/c e M/M/c+M.

GRÁFICO 4.7 - Distribuição do tempo entre chegadas – 11 – 12 horas



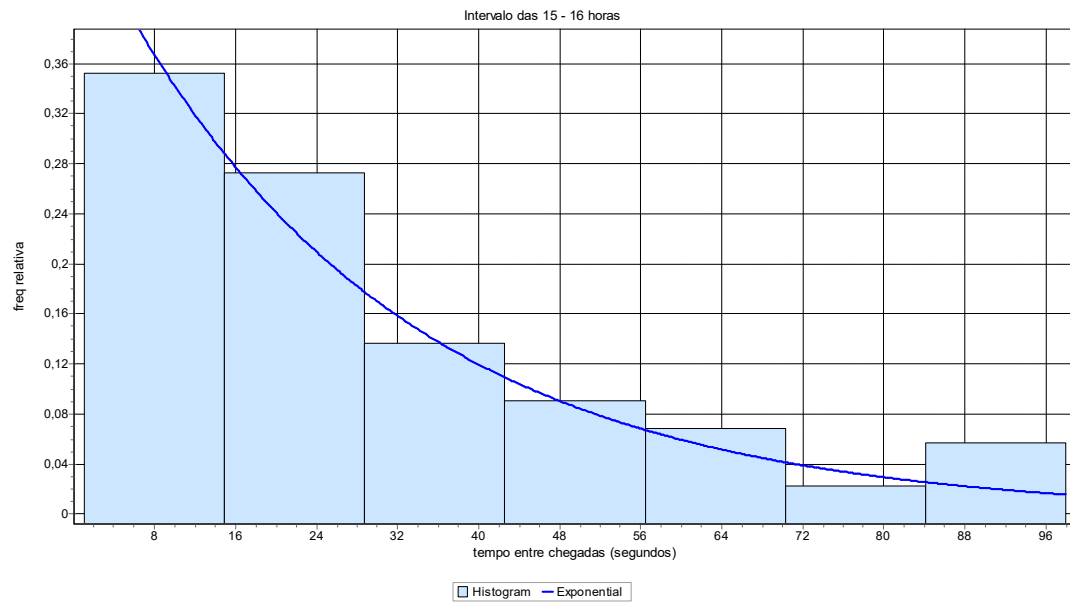
Fonte: elaborado pelo autor

GRÁFICO 4.8 - Distribuição do tempo entre chegadas – 12 – 13 horas



Fonte: elaborado pelo autor

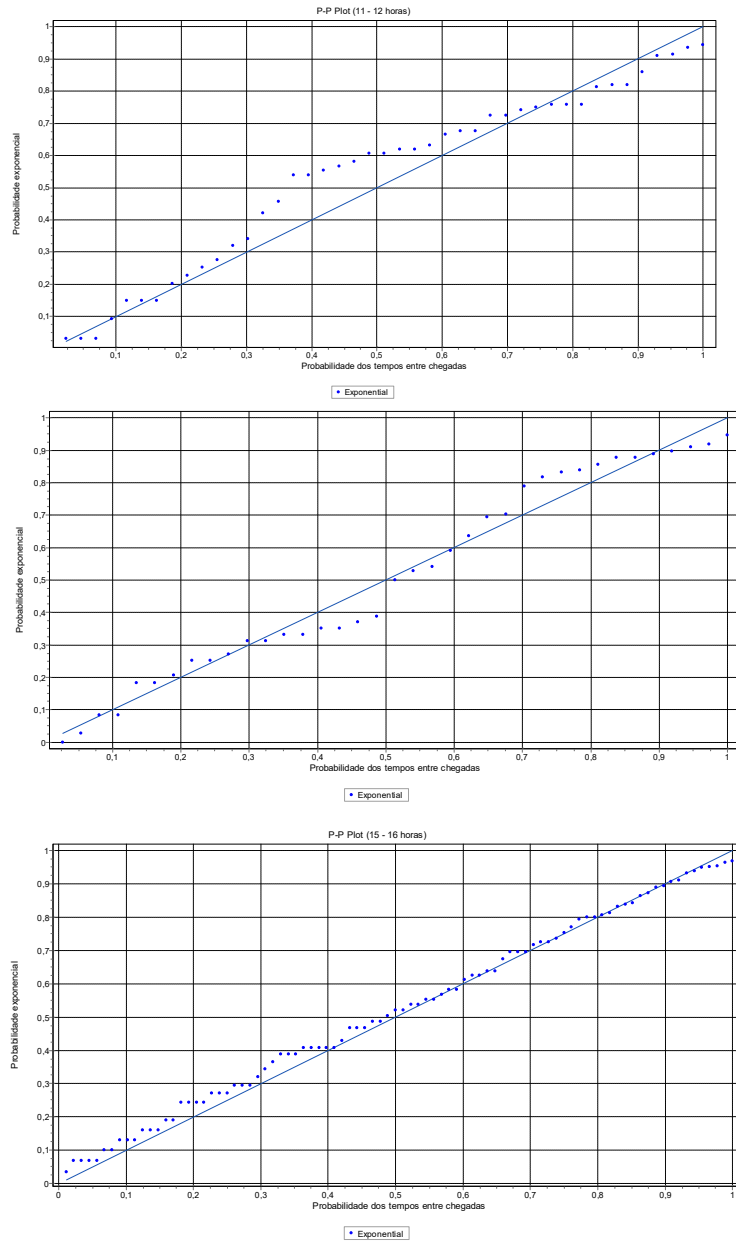
GRÁFICO 4.9 - Distribuição do tempo entre chegadas – 15 – 16 horas



Fonte: elaborado pelo autor

Outra maneira de verificar se os tempos entre chegadas se ajustam a uma distribuição exponencial é utilizar-se dos gráficos P-P plots. Estes gráficos comparam a distribuição teórica (no caso a exponencial) com a função de probabilidade acumulada observada nos dados. Se os dados seguirem a distribuição teórica esperada (a exponencial), os pares de pontos estarão sobre uma reta ou próximos dela. Observando-se os três gráficos da Figura 4.8, percebe-se que os pares de pontos estão muito próximos da reta bissetriz dos eixos coordenados, confirmando a aproximação com a distribuição exponencial de probabilidade.

FIGURA 4.8 - P-P plot dos tempos entre chegadas - exponencial



Fonte: elaborada pelo autor

## O processo de abandono

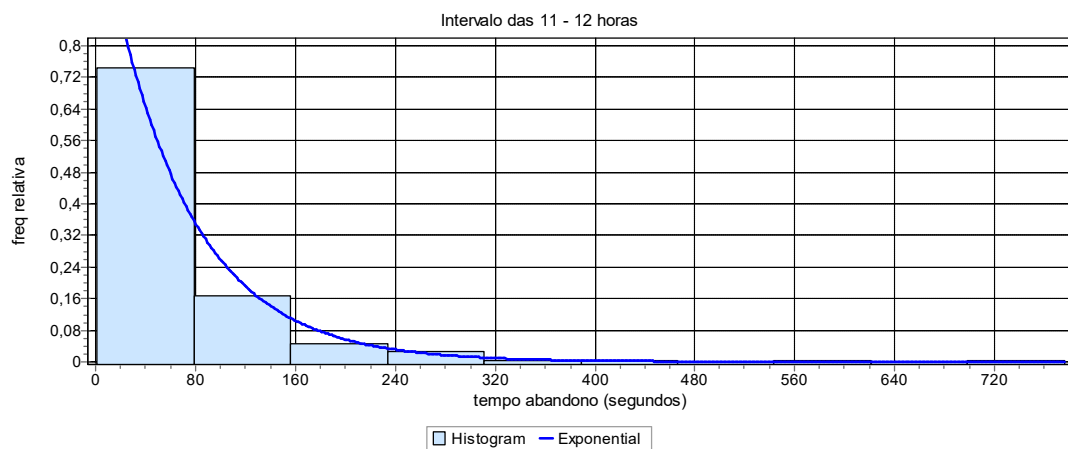
Cada uma das médias da Tabela 4.2 está próxima do correspondente desvio padrão. Isto sugere uma distribuição exponencial para cada estrato de hora analisado. Os coeficientes de variação estão próximos da unidade, principalmente nos dois últimos horários, e a razão entre a mediana e a média se aproxima do  $\ln 2$  ( $\ln 2 = 0,69$ , ver Apêndice E) em cada um dos três



intervalos de hora observado (0,653; 0,665; 0,628, respectivamente), reforçando a semelhança com a exponencial. Observa-se também que a mediana é significativamente menor que a média (mediana < média) em todos os intervalos de horas, indicando que a distribuição dos tempos de espera para abandono é assimétrica à direita.

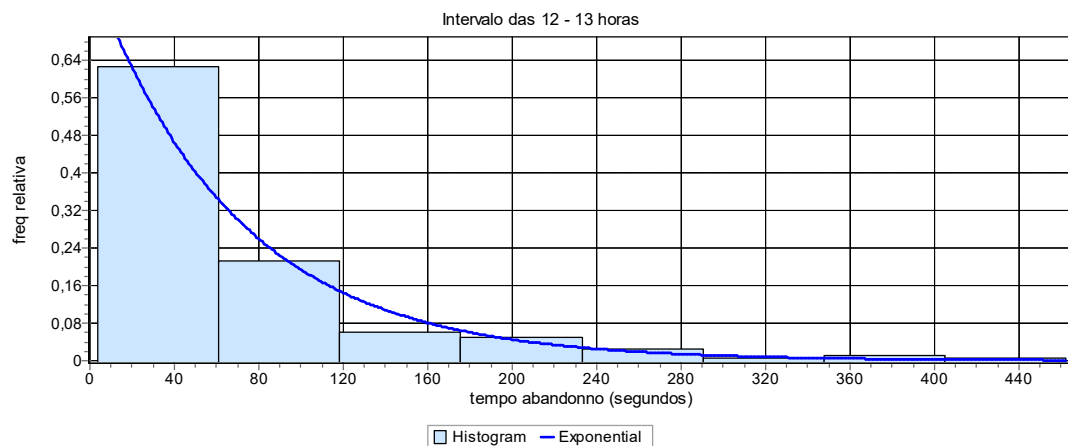
A exponencialidade dos tempos de espera para abandono também é evidenciada pelo histograma e respectiva densidade mostrada nos Gráficos 4.10, 4.11 e 4.12, em cada um dos três intervalos de hora analisados.

GRÁFICO 4.10: Distribuição do tempo de espera até abandonar – 11 – 12 horas



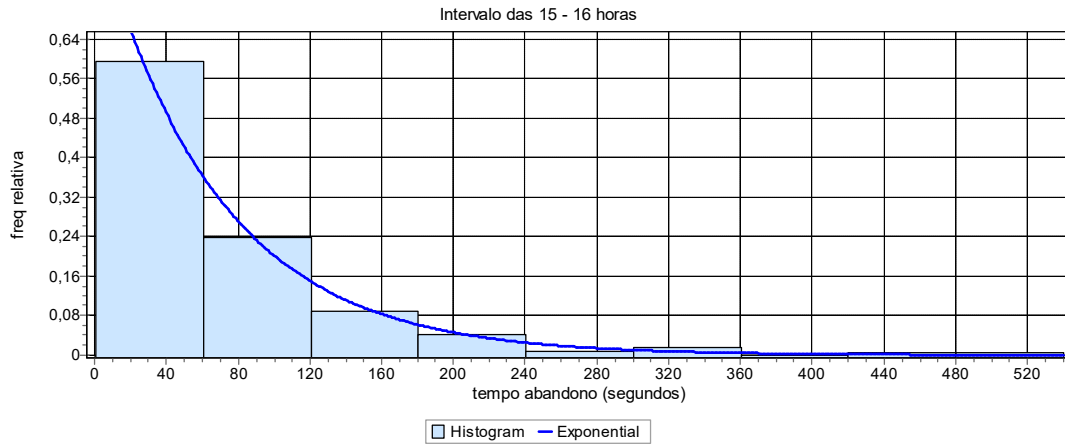
Fonte: elaborado pelo autor

GRÁFICO 4.11 - Distribuição do tempo de espera até abandonar – 12 – 13 horas



Fonte: elaborado pelo autor

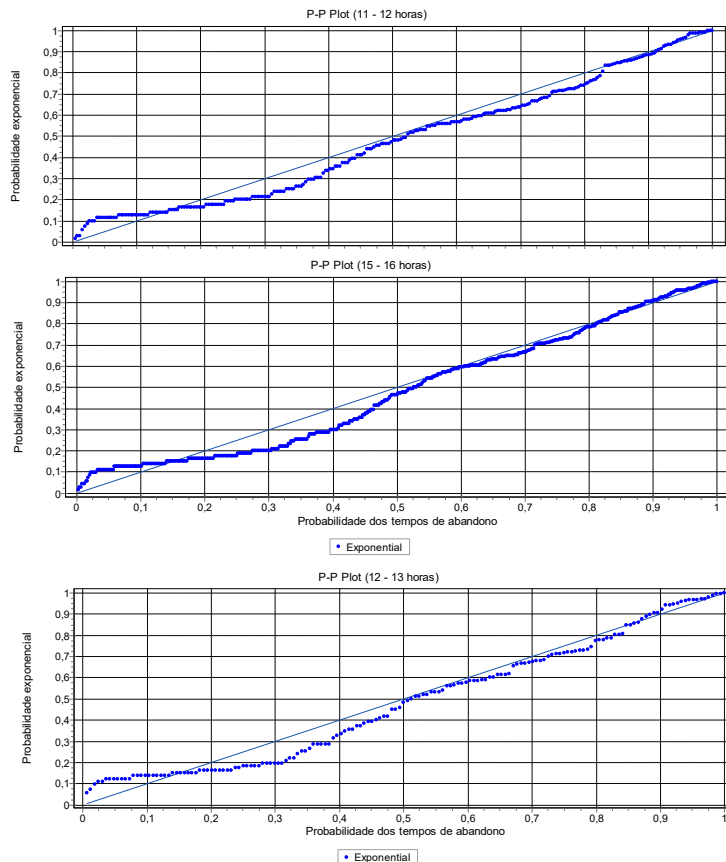
GRÁFICO 4.12 - Distribuição do tempo de espera até abandonar – 15 – 16 horas



Fonte: elaborado pelo autor

A Figura 4.9 exibe os gráficos P-P plots de cada um dos três intervalos de hora. Em cada um deles os pares de pontos estão muito próximos da reta bissetriz dos eixos coordenados, confirmando a aproximação dos tempos de espera para abandono com a distribuição exponencial de probabilidade.

FIGURA 4.9: P-P plot dos tempos de espera para abandono - exponencial



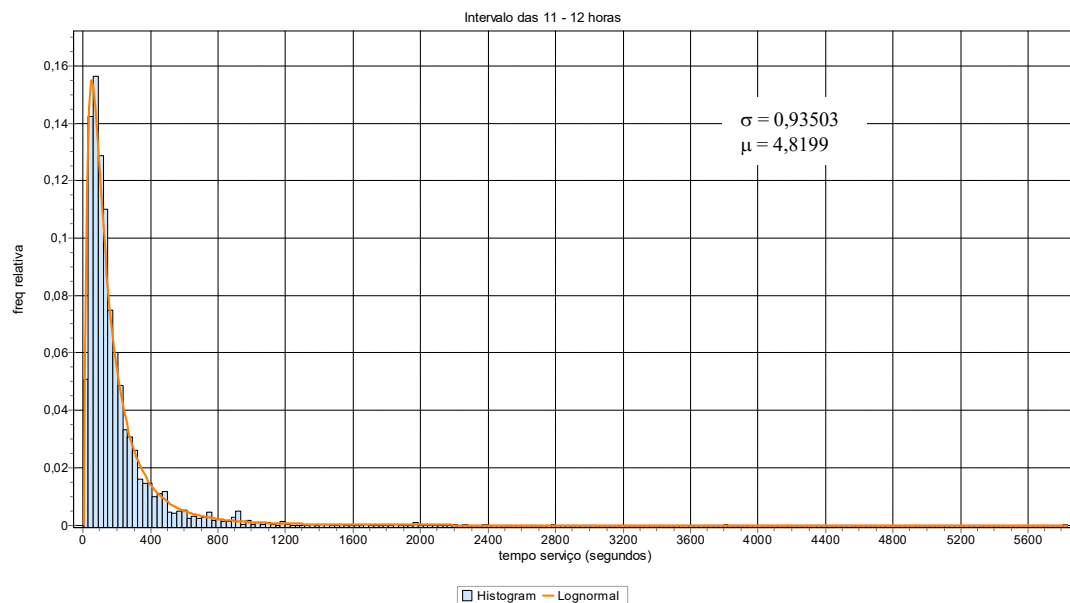
Fonte: elaborado pelo autor

## O processo de serviço

A distribuição dos tempos de serviço é assimétrica à direita em cada um dos horários analisados, comprovado pela comparação entre a mediana e a média (Mediana < média). Observa-se, também, na Tabela 4.4, que a distribuição dos tempos de serviço não se ajusta à distribuição exponencial. Isto é consistente com o valor do coeficiente de variação maior que 1 nos horários analisados. Uma outra verificação da possível exponencialidade é obtida pela propriedade que afirma que a razão entre a mediana e a média é igual a  $\ln 2$  ( $\ln 2 = 0,69$ ) (Apêndice E). Esta razão é igual a 0,64; 0,67 e 0,68 em cada um dos horários, respectivamente.

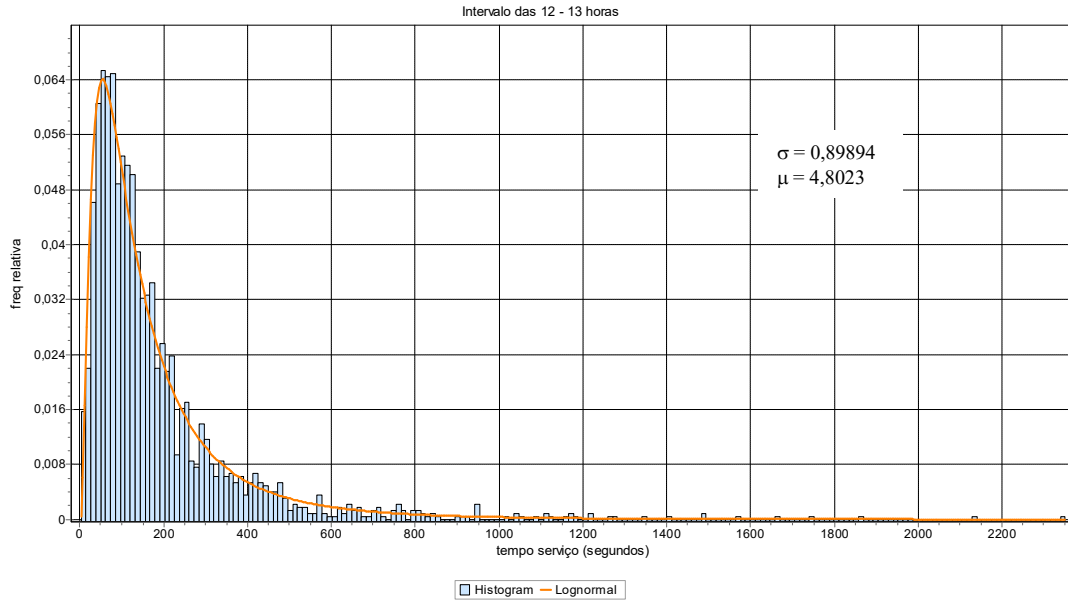
Nos dois últimos horários, os valores da razão entre a mediana e a média sugerem que a distribuição dos tempos de serviços se assemelha a distribuição exponencial. Entretanto, fica claro que a distribuição dos tempos de serviço não é exponencial nos três horários analisados, como mostram os gráficos 4.13, 4.14 e 4.15. Os seus histogramas e densidades não tem a configuração da exponencial, o que questiona o uso de modelo  $M/M/c+M$  para análise desse *Call Center*.

GRÁFICO 4.13 - Distribuição do tempo de serviço – 11 – 12 horas



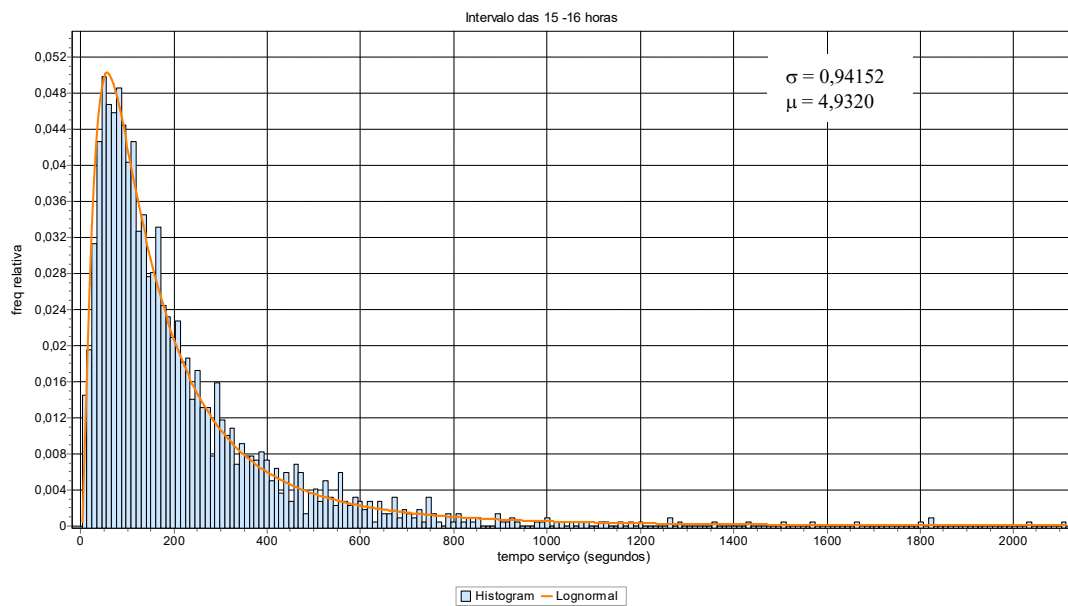
Fonte: elaborado pelo autor

GRÁFICO 4.14 - Distribuição do tempo de serviço – 12 – 13 horas



Fonte: elaborado pelo autor

GRÁFICO 4.15 - Distribuição do tempo de serviço – 15 – 16 horas

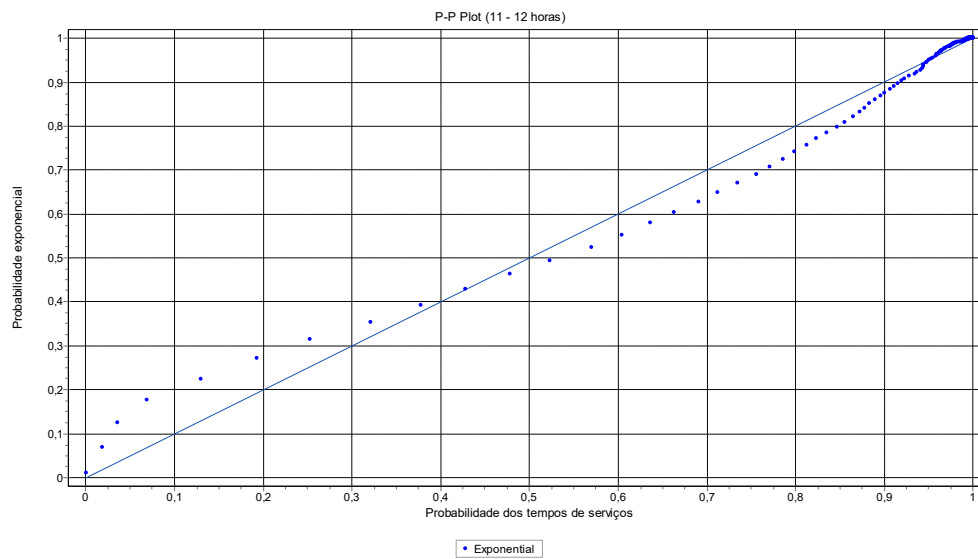


Fonte: elaborado pelo autor

Utilizando-se dos três Gráficos P-P plot 4.16, 4.17 e 4.18, nota-se que os tempos de serviços não provém de uma população com distribuição exponencial de probabilidade. Entretanto, os Gráficos 4.19, 4.20 e 4.21 sugerem que os tempos de serviços se ajustam à distribuição lognormal.

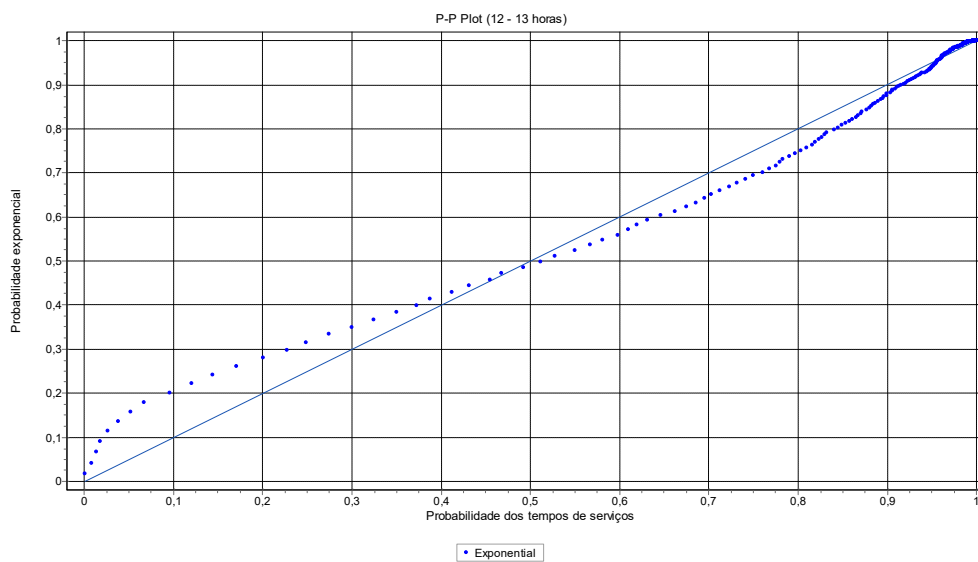
Mandelbaum et al. (2001) e Brown et al (2002) afirmam que a lognormalidade aparece e mantém-se para todos os tipos e prioridades de clientes, para servidores individuais, para diferentes dias da semana e também quando condicionou-se a hora do dia.

GRÁFICO 4.16 - P-P plot dos tempos de serviços – exponencial



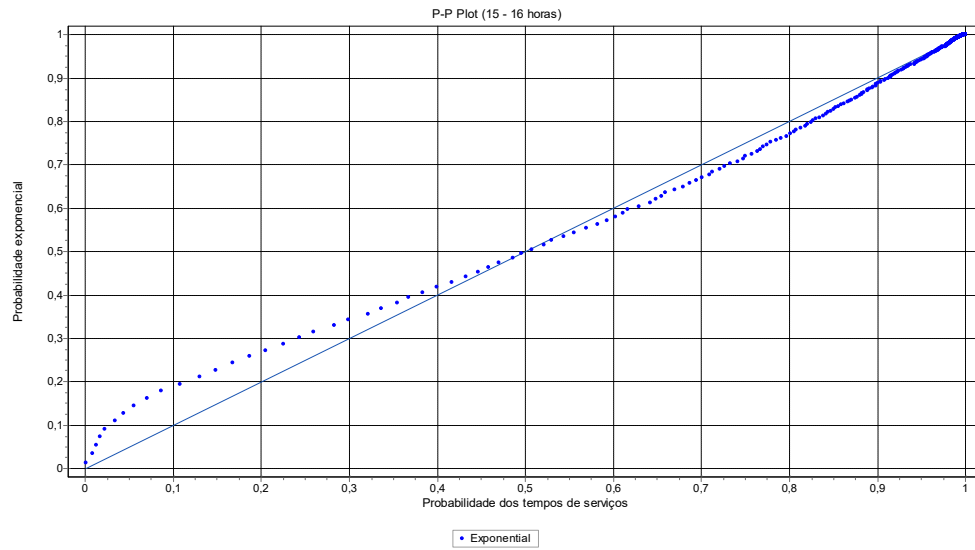
Fonte: elaborado pelo autor

GRÁFICO 4.17 - P-P plot dos tempos de serviços - exponencial



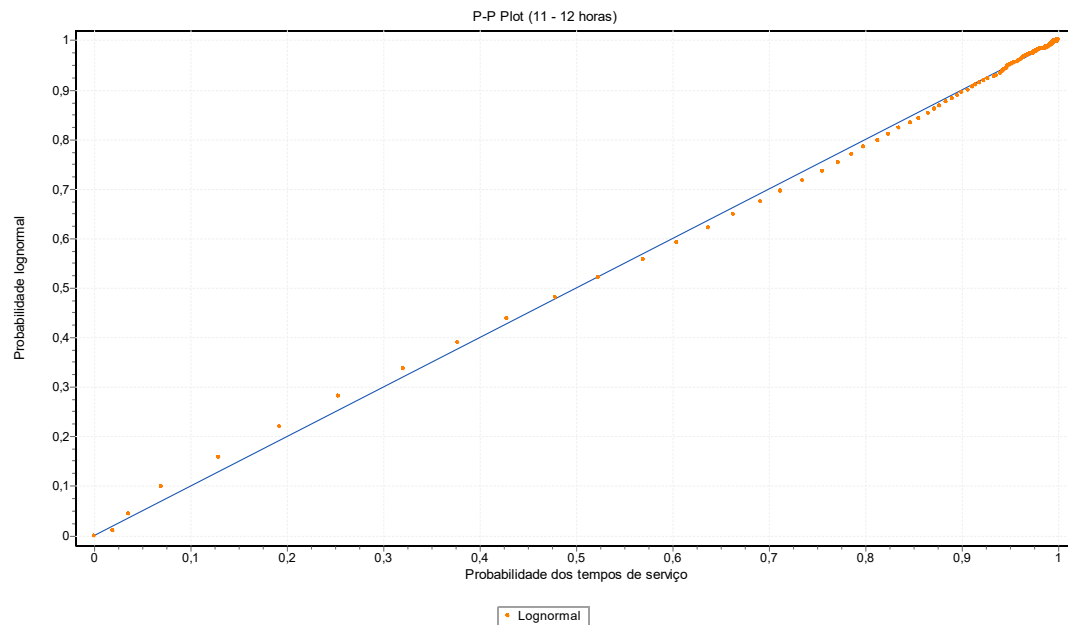
Fonte: elaborado pelo autor

GRÁFICO 4.18 - P-P plot dos tempos de serviços - exponencial



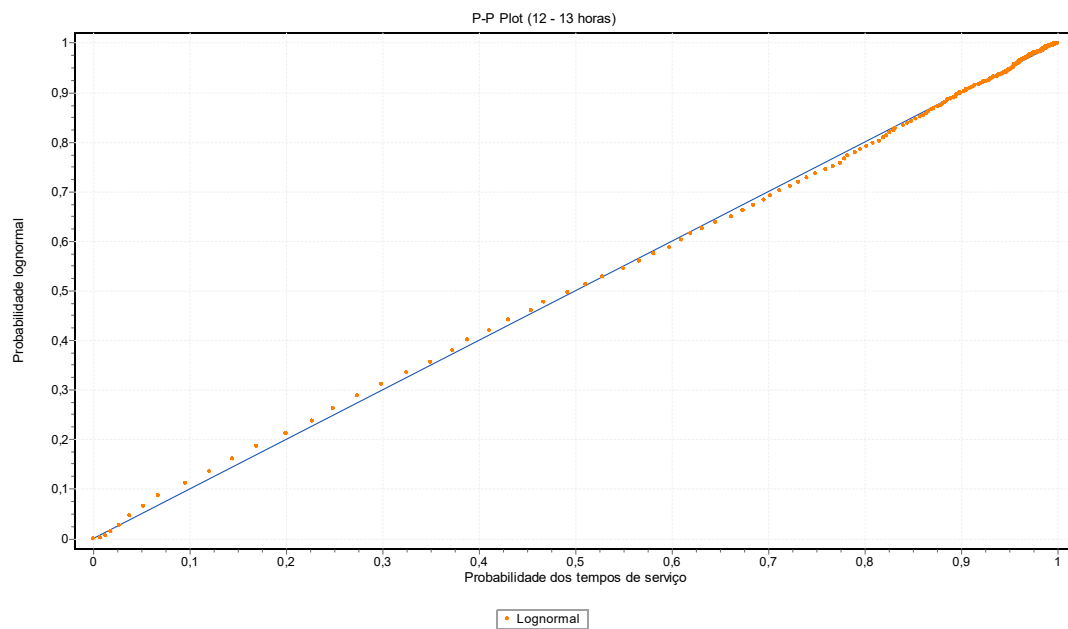
Fonte: elaborado pelo autor

GRÁFICO 4.19 - P-P plot dos tempos de serviços - lognormal



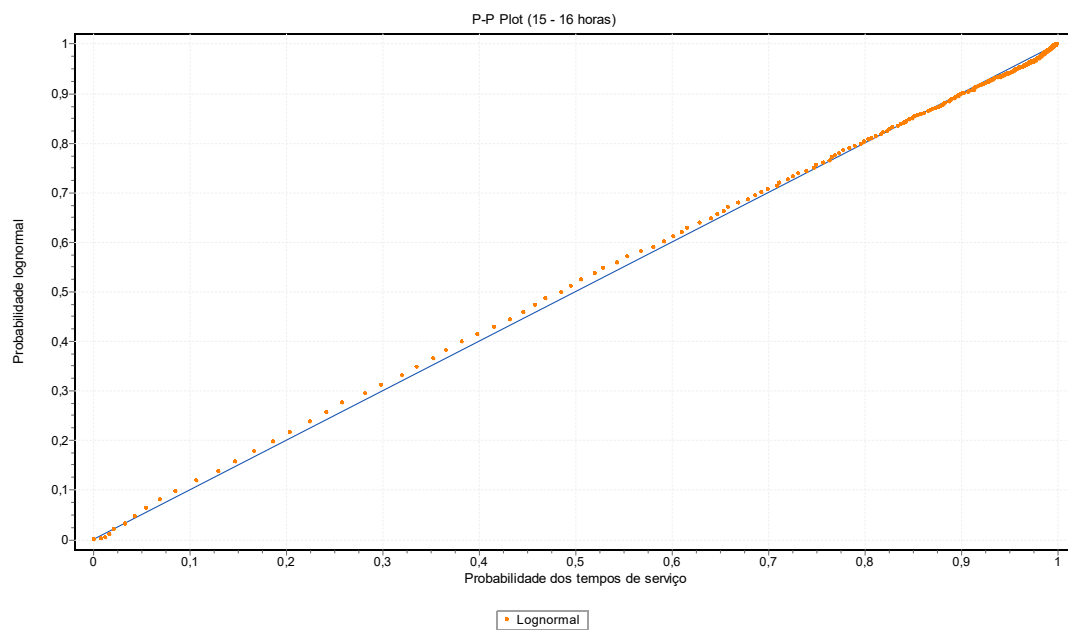
Fonte: elaborado pelo autor

GRÁFICO 4.20 - P-P plot dos tempos de serviços – lognormal



Fonte: elaborado pelo autor

GRÁFICO 4.21 - P-P plot dos tempos de serviços - lognormal



Fonte: elaborado pelo autor





## APÊNDICE - D

### Verificação dos modelos - intervalo 11-12 horas

As medidas de desempenho obtidas com os modelos analíticos (aproximados) e simulados, referentes ao horário das 11 às 12 horas, do *Call Center* do Banco, estão mostrados nas Tabelas 4.22 e 4.23.

Seguindo os mesmos procedimentos de análise adotados na secção 4.61, deduz-se que apenas dois resultados simulados não foram considerados verificados para a medida de desempenho probabilidade de abandonar e, também, dois resultados simulados não foram considerados verificados para a medida intensidade de tráfego (Tabela 4.23). Somente um resultado simulado é verificado para a medida tempo médio de espera e sete resultados simulados são considerados verificados para a medida probabilidade de esperar (Tabela 4.22).

TABELA 4.22 - Verificação dos modelos de fila - tempo médio de espera e probabilidade de esperar *Call Center* Banco – Intervalo 11-12 horas

Modelo	Tempo médio espera			Probabilidade de esperar		
	Simulação	Intervalo Confiança	Análítico	Simulação	Intervalo Confiança	Análítico
M/M <sub>1</sub> /c + G <sub>2</sub>	1,4261	(1,3303; 1,5219)	1,1143	0,8895	(0,7716; 1,0074)	0,7052
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>2</sub>	1,6820	(1,6166; 1,7474)	1,0493	0,8730	(0,7888; 0,9572)	0,7143
M/G <sub>1</sub> /c + G <sub>2</sub>	1,3292	(1,2494; 1,4090)	1,1781	0,8884	(0,7833; 0,9935)	0,8037*
M/M <sub>1</sub> /c + G <sub>4m</sub>	1,2008	(1,1096; 1,2920)	0,7465	0,8744	(0,8045; 0,9443)	0,6352
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>4m</sub>	1,4600	(1,3688; 1,5512)	1,0463	0,8576	(0,7130; 1,0022)	0,7195*
M/G <sub>1</sub> /c + G <sub>4m</sub>	1,0832	(0,9953; 1,1711)	1,1817	0,889	(0,7902; 0,9878)	0,8149*
M/M <sub>1</sub> /c + G <sub>5</sub>	0,8927	(0,8463; 0,9391)	0,6394	0,8668	(0,7778; 0,9558)	0,5968
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>5</sub>	1,1576	(1,1090; 1,2062)	1,0432	0,8346	(0,7925; 0,8767)	0,7172
M/G <sub>1</sub> /c + G <sub>5</sub>	0,7949	(0,7352; 0,8546)	1,1479	0,8784	(0,6954; 1,0614)	0,7919*
M/M <sub>1</sub> /c + G <sub>6m</sub>	0,8365	(0,7958; 0,8772)	0,4593	0,8573	(0,7660; 0,9486)	0,5399
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>6m</sub>	1,0694	(1,0149; 1,1239)	1,0423*	0,8234	(0,7357; 0,9111)	0,7193
M/G <sub>1</sub> /c + G <sub>6m</sub>	0,7336	(0,6884; 0,7788)	1,1313	0,8691	(0,7765; 0,9617)	0,7839*
M/M <sub>1</sub> /c + G <sub>7</sub>	1,3666	(1,2957; 1,4375)	1,1754	0,9012	(0,8284; 0,9740)	0,7135
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>7</sub>	1,7719	(1,6786; 1,8652)	1,0449	0,8808	(0,8057; 0,9559)	0,7185
M/G <sub>1</sub> /c + G <sub>7</sub>	1,3285	(1,2343; 1,4227)	1,1743	0,8937	(0,6780; 1,1094)	0,8091*
M/M <sub>1</sub> /c + G <sub>8m</sub>	1,3538	(1,2697; 1,4379)	0,9469	0,8907	(0,8112; 0,9702)	0,6568
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>8m</sub>	1,6638	(1,6012; 1,7264)	1,0418	0,8704	(0,7748; 0,9660)	0,7181
M/G <sub>1</sub> /c + G <sub>8m</sub>	1,2435	(1,1622; 1,3248)	1,1619	0,8945	(0,7756; 1,0134)	0,8029*
<b>Processo chegada:</b> Poisson – taxa $\lambda = 1,933$ – média $E(\tau) = 0,5173$			<b>Processo serviço:</b> Taxa $\mu = 0,314$ – média $E(S) = 3,1847$			
<b>Distribuições:</b> M ~ Exponencial(1,933)			<b>Distribuições:</b> G <sub>1</sub> ~ Fréchet (2,2564; 2,8235; -1,2554) G <sub>3</sub> ~ Fréchet (2,2564; 0,3323; 0) M <sub>1</sub> ~ Exponencial (0,314)			
<b>Processo abandono:</b> taxa 0,119 – média $E(R) = 8,4034$						
<b>Distribuições:</b> G <sub>2</sub> ~ Fatigue Life (0,7095; 6,6691)      G <sub>5</sub> ~ Exponencial (0,119)      G <sub>7</sub> ~ Lognormal (1,9115; 0,6488) G <sub>4m</sub> ~ 0,9072×Fatigue Life (0,7456; 5,9258) + 0,0928×Fatigue Life (0,1189; 1,0091) G <sub>6m</sub> ~ 0,8894×Exponencial (0,1347) + 0,1106×Exponencial (1,0835) G <sub>8m</sub> ~ 0,95×Lognormal (1,8512; 0,6625) + 0,05×Lognormal (-2,0936; 1,5619)						

Fonte: elaborada pelo autor

TABELA 4.23 - Verificação dos modelos de fila – probabilidade abandonar e intensidade de tráfego *Call Center* Banco – Intervalo 11-12 horas

Modelo	Probabilidade de abandonar			Intensidade de tráfego		
	Simulação	Intervalo Confiança	Analítico	Simulação	Intervalo Confiança	Analítico
M/M <sub>1</sub> /c + G <sub>2</sub>	0,1097	(0,0000; 0,2276)	0,1273*	0,9189	(0,8100; 1,0278)	0,8954*
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>2</sub>	0,1252	(0,0410; 0,2094)	0,0653*	0,8936	(0,7805; 1,0067)	0,9589*
M/G <sub>1</sub> /c + G <sub>2</sub>	0,1140	(0,0090; 0,2190)	0,1273*	0,9338	(0,8144; 1,0532)	0,9589*
M/M <sub>1</sub> /c + G <sub>4m</sub>	0,1272	(0,0573; 0,1971)	0,1515*	0,9202	(0,8429; 0,9975)	0,8706*
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>4m</sub>	0,1461	(0,0015; 0,2907)	0,1249*	0,8914	(0,7543; 1,0285)	0,8979*
M/G <sub>1</sub> /c + G <sub>4m</sub>	0,1092	(0,0104; 0,2080)	0,1515*	0,9257	(0,8075; 1,0439)	0,8979*
M/M <sub>1</sub> /c + G <sub>5</sub>	0,1343	(0,0453; 0,2233)	0,1623*	0,8965	(0,7674; 1,0256)	0,8595*
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>5</sub>	0,1657	(0,1236; 0,2078)	0,0573	0,8582	(0,7634; 0,9530)	0,9672
M/G <sub>1</sub> /c + G <sub>5</sub>	0,1239	(0,0000; 0,3069)	0,1623*	0,9036	(0,8098; 0,9974)	0,9672*
M/M <sub>1</sub> /c + G <sub>6m</sub>	0,1429	(0,0516; 0,2342)	0,1800*	0,8962	(0,7450; 1,0474)	0,8413*
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>6m</sub>	0,1802	(0,0925; 0,2679)	0,0517	0,8564	(0,7782; 0,9346)	0,9729
M/G <sub>1</sub> /c + G <sub>6m</sub>	0,1342	(0,0416; 0,2268)	0,1800*	0,9021	(0,8247; 0,9795)	0,9729*
M/M <sub>1</sub> /c + G <sub>7</sub>	0,1012	(0,0284; 0,1740)	0,1244*	0,9223	(0,7253; 1,1193)	0,8984*
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>7</sub>	0,1192	(0,0441; 0,1943)	0,0657*	0,8902	(0,8029; 0,9775)	0,9587*
M/G <sub>1</sub> /c + G <sub>7</sub>	0,1104	(0,0000; 0,3261)	0,1244*	0,9400	(0,8312; 1,0488)	0,9587*
M/M <sub>1</sub> /c + G <sub>8m</sub>	0,1074	(0,0279; 0,1869)	0,1388*	0,9154	(0,8162; 1,0146)	0,8836*
M/G <sup>c</sup> <sub>3</sub> /1 + G <sub>8m</sub>	0,1283	(0,0327; 0,2239)	0,0627*	0,9016	(0,8209; 0,9823)	0,9617*
M/G <sub>1</sub> /c + G <sub>8m</sub>	0,1061	(0,0000; 0,2250)	0,1388*	0,9446	(0,8213; 1,0679)	0,9617*
<b>Processo chegada:</b> Poisson – taxa $\lambda = 1,933$ – média $E(\tau) = 0,5173$			<b>Processo serviço:</b> Taxa $\mu = 0,314$ – média $E(S) = 3,1847$			
<b>Distribuições:</b> M ~ Exponencial(1,933)			<b>Distribuições:</b> G <sub>1</sub> ~ Fréchet (2,2564; 2,8235; -1,2554) G <sub>3</sub> ~ Fréchet (2,2564; 0,3323; 0) M <sub>1</sub> ~ Exponencial (0,314)			
<b>Processo abandono:</b> taxa 0,119 – média $E(R) = 8,4034$						
<b>Distribuições:</b> G <sub>2</sub> ~ Fatigue Life (0,7095; 6,6691)      G <sub>5</sub> ~ Exponencial (0,119)      G <sub>7</sub> ~ Lognormal (1,9115; 0,6488) G <sub>4m</sub> ~ 0,9072×Fatigue Life (0,7456; 5,9258) + 0,0928×Fatigue Life (0,1189; 1,0091) G <sub>6m</sub> ~ 0,8894×Exponencial (0,1347) + 0,1106×Exponencial (1,0835) G <sub>8m</sub> ~ 0,95×Lognormal (1,8512; 0,6625) + 0,05×Lognormal (-2,0936; 1,5619)						

Fonte: elaborada pelo autor

### **Verificação dos modelos - intervalo 12-13 horas**

Os resultados das medidas de desempenho obtidos com os modelos analíticos (aproximados) e simulados, referentes ao horário das 12 às 13 horas, do *Call Center* do Banco, estão exibidos nas Tabelas 4.24 e 4.25.

Utilizando-se o mesmo procedimento de análise adotado na secção 4.61, nota-se que todos os resultados simulados são considerados verificados para as medidas de desempenho probabilidade de abandonar e intensidade de tráfego (Tabela 4.25) e apenas dois resultados simulados são verificados para a medida tempo médio de espera. Nenhum resultado simulado foi verificado para a medida probabilidade de esperar (Tabela 4.24).

TABELA 4.24 - Verificação dos modelos de fila – tempo médio de espera e probabilidade de esperar *Call Center* Banco – Intervalo 12-13 horas

Modelo	Tempo médio espera			Probabilidade de esperar		
	Simulação	Intervalo Confiança	Analítico	Simulação	Intervalo Confiança	Analítico
M/M <sub>2</sub> /c + G <sub>10</sub>	0,8102	(0,7318; 0,8886)	0,7610*	0,9465	(0,8728; 1,0202)	0,5393
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>10</sub>	1,1483	(1,0892; 1,2074)	0,8174	0,9293	(0,8349; 1,0237)	0,5852
M/G <sub>9</sub> /c + G <sub>10</sub>	0,7576	(0,6989; 0,8163)	0,9531	0,9526	(0,8395; 1,0657)	0,6877
M/M <sub>2</sub> /c + G <sub>12</sub>	0,4315	(0,4039; 0,4591)	0,3509	0,9184	(0,8507; 0,9861)	0,4223
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>12</sub>	0,6812	(0,6377; 0,7247)	0,8098	0,8777	(0,8118; 0,9436)	0,5844
M/G <sub>9</sub> /c + G <sub>12</sub>	0,4069	(0,3725; 0,4413)	0,8887	0,9216	(0,8732; 0,9700)	0,6727
M/M <sub>2</sub> /c + G <sub>13m</sub>	0,3865	(0,3573; 0,4157)	0,2079	0,9171	(0,8593; 0,9749)	0,3607
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>13m</sub>	0,6244	(0,5996; 0,6492)	0,7409	0,8674	(0,8163; 0,9185)	0,4609
M/G <sub>9</sub> /c + G <sub>13m</sub>	0,3512	(0,3223; 0,3801)	0,7057	0,9196	(0,8556; 0,9836)	0,6731
M/M <sub>2</sub> /c + G <sub>14</sub>	0,8217	(0,7540; 0,8894)	0,7709*	0,9522	(0,8353; 1,0691)	0,5409
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>14</sub>	1,1724	(1,0903; 1,2545)	0,8119	0,9359	(0,8437; 1,0281)	0,5830
M/G <sub>9</sub> /c + G <sub>14</sub>	0,6975	(0,6446; 0,7504)	0,9479	0,9606	(0,9067; 1,0145)	0,6861
M/M <sub>2</sub> /c + G <sub>15m</sub>	0,7764	(0,7184; 0,8344)	0,6224	0,9528	(0,8812; 1,0244)	0,4948
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>15m</sub>	1,1238	(1,0674; 1,1802)	1,3930	0,9299	(0,8482; 1,0116)	0,5794
M/G <sub>9</sub> /c + G <sub>15m</sub>	0,7386	(0,6787; 0,7985)	0,9259	0,9528	(0,8964; 1,0092)	0,6741
<b>Processo chegada:</b> Poisson – taxa $\lambda = 1,717$ – média $E(\tau) = 0,5824$			<b>Processo serviço:</b> Taxa $\mu = 0,328$ – média $E(S) = 3,0488$			
<b>Distribuições:</b> M ~ Exponencial(1,717)			<b>Distribuições:</b> G <sub>9</sub> ~ Fréchet (2,2847; 2,7464; -1,2092) G <sub>11</sub> ~ Fréchet (2,2847; 0,3211; 0) M <sub>2</sub> ~ Exponencial (0,328)			
<b>Processo abandono:</b> taxa 0,148 – média $E(R) = 6,7568$						
<b>Distribuições:</b> G <sub>10</sub> ~ Fatigue Life (0,5069; 5,9746)      G <sub>12</sub> ~ Exponencial (0,148)      G <sub>14</sub> ~ Lognormal (1,7902; 0,4861) G <sub>13m</sub> ~ 0,4785×Exponencial (0,3099) + 0,5215×Exponencial (0,2844) G <sub>15m</sub> ~ 0,95×Lognormal (1,7334; 0,4973) + 0,05×Lognormal (-2,0101; 1,3585)						

Fonte: elaborada pelo autor

TABELA 4.25 - Verificação dos modelos de fila – probabilidade abandonar e intensidade de tráfego *Call Center* Banco – Intervalo 12-13 horas

Modelo	Probabilidade de abandonar			Intensidade de tráfego		
	Simulação	Intervalo Confiança	Analítico	Simulação	Intervalo Confiança	Analítico
M/M <sub>2</sub> /c + G <sub>10</sub>	0,0543	(0,0000; 0,1280)	0,0611*	0,8241	(0,6708; 0,9774)	0,8192*
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>10</sub>	0,0670	(0,0000; 0,1614)	0,0103*	0,8078	(0,7459; 0,8697)	0,8634*
M/G <sub>9</sub> /c + G <sub>10</sub>	0,0426	(0,0000; 0,1558)	0,0611*	0,8314	(0,7093; 0,9535)	0,8634*
M/M <sub>2</sub> /c + G <sub>12</sub>	0,0827	(0,0150; 0,1504)	0,1108*	0,7907	(0,7057; 0,8757)	0,7758*
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>12</sub>	0,1217	(0,0558; 0,1876)	0,1382*	0,7469	(0,6740; 0,8198)	0,7519*
M/G <sub>9</sub> /c + G <sub>12</sub>	0,0799	(0,0315; 0,1283)	0,1108*	0,8189	(0,6475; 0,9903)	0,8723*
M/M <sub>2</sub> /c + G <sub>13m</sub>	0,0818	(0,0240; 0,1396)	0,1365*	0,7990	(0,7424; 0,8556)	0,7594*
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>13m</sub>	0,1337	(0,0826; 0,1848)	0,0915*	0,7601	(0,6927; 0,8275)	0,7926*
M/G <sub>9</sub> /c + G <sub>13m</sub>	0,0827	(0,0187; 0,1467)	0,1365*	0,8144	(0,6981; 0,9307)	0,8724*
M/M <sub>2</sub> /c + G <sub>14</sub>	0,0506	(0,0000; 0,1674)	0,0603*	0,8301	(0,7419; 0,9183)	0,8198*
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>14</sub>	0,0668	(0,0000; 0,1590)	0,0109*	0,8214	(0,6921; 0,9507)	0,8629*
M/G <sub>9</sub> /c + G <sub>14</sub>	0,0406	(0,0000; 0,0945)	0,0603*	0,8469	(0,7133; 0,9805)	0,8629*
M/M <sub>2</sub> /c + G <sub>15m</sub>	0,0485	(0,0000; 0,1201)	0,0763*	0,8202	(0,6837; 0,9567)	0,8059*
M/G <sup>c</sup> <sub>11</sub> /1 + G <sub>15m</sub>	0,0674	(0,0000; 0,1491)	0,0862*	0,8096	(0,7308; 0,88840)	0,7973*
M/G <sub>9</sub> /c + G <sub>15m</sub>	0,0452	(0,0000; 0,1016)	0,0763*	0,8363	(0,7549; 0,9177)	0,8719*
<b>Processo chegada:</b> Poisson – taxa $\lambda = 1,717$ – média $E(\tau) = 0,5824$ <b>Distribuições:</b> M ~ Exponencial(1,717)			<b>Processo serviço:</b> Taxa $\mu = 0,328$ – média $E(S) = 3,0488$ <b>Distribuições:</b> G <sub>9</sub> ~ Fréchet (2,2847; 2,7464; -1,2092) G <sub>11</sub> ~ Fréchet (2,2847; 0,3211; 0) M <sub>2</sub> ~ Exponencial (0,328)			
<b>Processo abandono:</b> taxa 0,148 – média $E(R) = 6,7568$ <b>Distribuições:</b> G <sub>10</sub> ~ Fatigue Life (0,5069; 5,9746)      G <sub>12</sub> ~ Exponencial (0,148)      G <sub>14</sub> ~ Lognormal (1,7902; 0,4861) G <sub>13m</sub> ~ 0,4785×Exponencial (0,3099) + 0,5215×Exponencial (0,2844) G <sub>15m</sub> ~ 0,95×Lognormal (1,7334; 0,4973) + 0,05×Lognormal (-2,0101; 1,3585)						

Fonte: elaborada pelo autor

## APÊNDICE - E

### Razão entre mediana e média da Distribuição Exponencial

Uma variável aleatória contínua  $X$  com distribuição exponencial de probabilidade de parâmetro  $\lambda$ , tem função de distribuição acumulada dada por:

$$F(x) = P(X \leq x) = 1 - e^{-\lambda x}$$

Considerando que a mediana corresponde ao valor de  $x$  onde  $P(X \leq x) = 0,5$  e usando a igualdade anterior, temos:

$$\begin{aligned} 0,5 &= 1 - e^{-\lambda x} \\ 0,5 - 1 &= -e^{-\lambda x} \\ -0,5 &= -e^{-\lambda x} \\ 0,5 &= e^{-\lambda x} \end{aligned}$$

Aplicando logaritmo em ambos os lados da última igualdade, obtemos:

$$\begin{aligned} \ln(0,5) &= \ln e^{-\lambda x} \\ \ln(0,5) &= -\lambda x \ln e \\ \ln(0,5) &= -\lambda x \\ x &= \frac{-\ln(0,5)}{\lambda} \end{aligned}$$

Portanto, a mediana da distribuição exponencial de probabilidade é igual a:

$$Md = \frac{-\ln(0,5)}{\lambda}$$

Assim, obtém-se a razão  $r$  entre a mediana e a média da distribuição exponencial de probabilidade:

$$\begin{aligned} r &= \frac{Md}{x} \\ &= \frac{-\ln(0,5)}{\lambda} \\ &= \frac{\lambda}{\lambda} \\ r &= -\ln(0,5) = 0,6931 \end{aligned}$$





## APÊNDICE - F

### Distribuição *Fatigue Life*

A distribuição Birnbaum-Saunders (BS), também conhecida como a distribuição *Fatigue Life*, foi desenvolvida por Birnbaum e Saunders (1949) com o propósito de modelar os tempos de falhas de materiais e também dados de tempos de vida em muitas áreas aplicadas. Desmonde (1985) mostrou que a distribuição BS descreve o tempo total que passou até que algum tipo de dano tenha acumulado, produzido pelo desenvolvimento e pelo crescimento de uma trinca dominante, ultrapassando um limite, e causando falha no material.

A distribuição BS, denotada por  $T \sim \text{BS}(\alpha, \beta)$ , é definida em termos da distribuição normal, por meio da variável aleatória

$$T = \beta \left[ \frac{\alpha}{2} Z + \sqrt{\left(\frac{\alpha}{2}\right)^2 + 1} \right]^2$$

onde  $Z \sim N(0,1)$ ,  $\alpha > 0$  é o parâmetro forma e  $\beta > 0$  é o parâmetro escala e a mediana da distribuição.

Assim, se  $T$  é o tempo total até a ocorrência da falha, então a função de distribuição de  $T$  é dada por:

$$F(t, \alpha, \beta) = P(T \leq t) = \Phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \right], t > 0$$

e sua função de densidade:

$$f(t, \alpha, \beta) = \frac{\sqrt{\frac{t}{\beta}} + \sqrt{\frac{\beta}{t}}}{2\alpha\beta} \cdot \phi \left[ \frac{1}{\alpha} \left( \sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right) \right], t > 0$$

onde  $\phi$  é a função de densidade de probabilidade da distribuição normal padrão e  $\Phi$  a função de distribuição acumulada da distribuição normal padrão.

Conforme Meeker e Escobar (1998), a média e a variância da distribuição *Fatigue Life* são dadas por:

$$E(T) = \beta \left( 1 + \frac{\alpha^2}{2} \right)$$

$$\text{Var}(T) = (\alpha\beta)^2 \left( 1 + \frac{5\alpha^2}{4} \right)$$



## APÊNDICE - G

### Distribuição Fréchet

A distribuição de probabilidade Fréchet é um caso especial da distribuição valor extremo generalizada. Tem sido utilizada com grande frequência em estudos de fenômenos ambientais, principalmente para solucionar problemas de precipitação fluvial e outros relacionados às áreas de Engenharia.

O trabalho mais relevante foi publicado por Fréchet (1927) que considerou uma distribuição assintótica de grandes valores.

Sua função de distribuição, com parâmetro forma  $\alpha > 0$ , parâmetro locação  $m$  (o mínimo) e parâmetro escala  $s > 0$ , é da forma:

$$F(x) = P(X \leq x) = e^{-\frac{(x-m)^{-\alpha}}{s}}, \text{ se } x > m$$

A média e variância desta distribuição é dada por:

$$E(X) = \begin{cases} m + s\Gamma\left(1 - \frac{1}{\alpha}\right), & \text{para } \alpha > 1 \\ \infty, & \text{caso contrário} \end{cases}$$

$$\text{Var}(X) = \begin{cases} s^2 \left( \Gamma\left(1 - \frac{2}{\alpha}\right) - \left( \Gamma\left(1 - \frac{1}{\alpha}\right) \right)^2 \right), & \text{para } \alpha > 2 \\ \infty, & \text{caso contrário} \end{cases}$$

Outros detalhes e aplicações sobre a distribuição de probabilidade Fréchet podem ser encontrados em Nadarajah e Kotz (2000).



## APÊNDICE - H

### Distribuições mistas

A mistura finita de distribuições (distribuições mistas) tem sido uma abordagem útil na modelagem estatística de uma grande variedade de fenômenos aleatórios. São geralmente utilizadas em situações que se suspeita que a população é heterogênea, formada por duas ou mais subpopulações independentes, mas não é possível distinguir entre os indivíduos dos diferentes tipos. Nesses casos os modelos formados por uma única classe de distribuição de probabilidade, muitas vezes não são suficientes para captar a presença dessas subpopulações de forma adequada. Um modelo formado por uma mistura finita de modelos componentes, muitas vezes da mesma família paramétrica (do mesmo tipo), pode ser mais apropriado para captar essas subpopulações. (LAWLESS, 2003; FELGUEIRAS, 2009).

Nos Call Centers é provável que o fenômeno do abandono, caracterizado pela paciência do usuário, seja formado por classes distintas de usuários que abandonam o sistema. Nesses casos, temos subpopulações distintas e um modelo analítico de fila com uma distribuição mista (mistura finita) para a paciência poderá captar melhor essas características diferentes, que um modelo analítico de fila formado por uma única distribuição para modelar a paciência.

Provavelmente, o modelo  $M$  é uma mistura ponderada de modelos componentes  $M_i$ , com  $M = \sum_{i=1}^k w_i M_i$ , onde  $w_i$  representa a proporção dos dados (indivíduos dos diferentes tipos) “explicados” pelo modelo componente  $M_i$ .

Assim, Lawless (2009) considera que modelos de misturas discretas surgem quando uma população é formada por indivíduos de  $k$  tipos diferentes, com uma proporção  $p_i$  ( $i = 1, \dots, k$ ), onde  $0 \leq p_i \leq 1$  e  $\sum_{i=1}^k p_i = 1$ . Uma amostra aleatória selecionada dessa população tem função de distribuição dada por:

$$F(x) = \sum_{i=1}^k p_i F_i(x) \quad (3.102)$$

onde  $F_i(x)$  é a função distribuição componente da mistura e  $p_i$  é o peso da subpopulação  $i$ . Em muitas situações a função de distribuição componente pertence à mesma família paramétrica. A densidade de  $X$ , se houver, não é mais que uma soma ponderada (combinação linear complexa) de outras densidades. Os pesos  $p_i$  são determinísticos, podendo ser considerados como parâmetros desconhecidos ou não.

Felgueiras (2009) admite que se  $X$  é uma mistura finita de variáveis aleatórias independentes  $X_1, \dots, X_N$  e se existirem os  $k$ -ésimos momentos populacionais, então os momentos centrados na média podem ser determinados por:

$$\mu_k = \sum_{j=1}^N p_j \mu_{j,k}$$

As misturas finitas possuem diversas aplicações classificadas em diretas ou indiretas, conforme Felgueiras (2009). As diretas ocorrem quando cada observação pertence a uma subpopulação ou categoria, apesar de raramente se saber a qual delas se vincula. Cada subpopulação é descrita pela sua densidade e os pesos da mistura são as probabilidades de cada observação pertencer a essa subpopulação. As aplicações indiretas ocorrem quando não existe a divisão dos dados em subpopulações, sendo o ajustamento da mistura feito por permitir uma grande flexibilidade, como multimodalidade.

Uma variável aleatória  $X$  é uma mistura infinita se:

$$F_X(x) = \int_{-\infty}^{+\infty} h(\theta) F_Y(x/\theta) d\theta$$

onde  $F$  representa uma função de distribuição e  $h$  uma função de densidade que fornece as várias interpretações para a variável aleatória  $X$ . Por meio da densidade  $h$  é possível calcular expressões para os momentos e para a função característica (FELGUEIRAS, 2009).

Entre os métodos para estimar os parâmetros das distribuições mistas, Titterington et al. (1985) sugere o método gráfico, dos momentos, da máxima verossimilhança e bayesianos. Neste trabalho utilizou-se o método da máxima verossimilhança, desenvolvido por Fisher e que consiste em encontrar o vetor de parâmetros  $\theta$  com maior probabilidade de ter produzido a amostra observada  $x$  (FELGUEIRAS, 2009).

Desse modo, considerando a função de verossimilhança:

$$L(\theta | x) = \prod_{i=1}^n f(x_i)$$

a estimativa da máxima verossimilhança para o vetor de parâmetros  $\theta$  é o ponto  $\hat{\theta}$  tal que:

$$\max_{\theta \in \Theta} L(\theta | x) = L(\hat{\theta} | x), \quad \forall \theta \in \Theta$$

Os parâmetros podem ser encontrados determinando-se os zeros da derivada do logaritmo da função de verossimilhança. Assim, se  $L(\theta | x)$  é uma função de verossimilhança, para cada parâmetro  $\theta_j, j = 1, \dots, p$ , as equações de verossimilhança são definidas por:

$$\frac{\partial}{\partial \theta_j} \ln L(\theta | x) = 0$$

Neste trabalho, todas as distribuições mistas foram obtidas pelo método da máxima verossimilhança, com os cálculos realizados por meio da função `EstimatedDistribution[data, dist]`, do *software* Mathematica (2013), que estima uma distribuição paramétrica a partir dos dados informados.

Apenas por simplicidade, optou-se neste trabalho pela utilização de duas componentes da mesma família paramétrica na composição da distribuição mista, para representar os tempos de paciência dos usuários. Nenhum método foi utilizado para determinar a quantidade de componentes nas distribuições mistas utilizadas neste estudo, considerando que o objetivo é verificar se o uso de distribuições mistas para ajustar os tempos de paciência, quando inseridos nos modelos analíticos de fila os deixa mais efetivos para representar os *Call Centers* dos casos analisados.

Como um exemplo, a distribuição mista de probabilidade para o caso do *Call Center* da Empresa, formada por duas componentes Normais foi estimada, a partir dos dados reais, por meio da função (MATHEMATICA, 2013):

```
EstimatedDistribution[data,MixtureDistribution[{p,1-p},{NormalDistribution[a,b],NormalDistribution[c,d]}]]
```

A distribuição mista de probabilidade que resultou dessa estimativa possui pesos  $p = 0,5221$  e  $(1-p) = 0,4779$  e componentes Normal (30,3035; 17,4153) e Normal (5,61038; 3,36321), com parâmetros média e desvio padrão, ou seja:

```
MixtureDistribution[{0.522146,0.477854},{NormalDistribution[30.3035,17.4153], NormalDistribution[5.61038, 3.36321]}]
```

Esta distribuição mista foi ajustada a partir dos dados reais dos tempos de paciência observados no *Call Center* da Empresa.

Considerando-se que os tempos de paciência são dados censurados, os parâmetros desta distribuição mista foram recalculados utilizando-se da média e da variância estimadas pelo

estimador Kaplan-Meier para dados censurados. Desse modo, a nova média e o novo desvio padrão, de acordo com Felgueiras (2009), foram obtidos por:

a) Média estimada pelo Kaplan-Meier: 55,9089 minutos

$$\text{Média da primeira componente: } a = 0,522146 \times 55,9089 = 29,1926$$

$$\text{Média da segunda componente: } c = 0,477854 \times 55,9089 = 26,7163$$

b) Variância estimada pelo Kaplan-Meier: 445,3239

$$\text{Desvio padrão da primeira componente: } b = \sqrt{0,522146 \times 445,3239} = 15,2487$$

$$\text{Desvio padrão da segunda componente: } d = \sqrt{0,477854 \times 445,3239} = 14,5877$$

Assim, a distribuição mista que se ajusta aos tempos de paciência censurados do *Call Center* da Empresa, e que foi utilizada nos modelos analíticos, com pesos  $p = 0,5221$  e  $(1-p) = 0,4779$  e componentes Normal (29,1926; 15,2487) e Normal (26,7163; 14,5877), com parâmetros média e desvio padrão, recalculados conforme a estimativa Kaplan-Meier, é a:

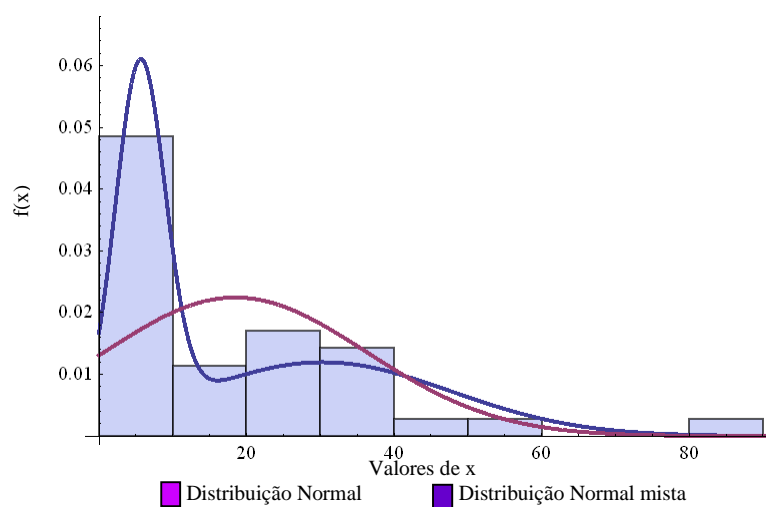
$$F(x) = 0,5221 * N(29,1926; 15,2487) + 0,4779 * N(26,7163; 14,5877)$$

Este procedimento foi usado para ajustar os parâmetros de todas as distribuições mistas utilizadas neste trabalho.

Apenas para comparação o Gráfico da distribuição mista de probabilidade, estimada pelo Mathematica (2013) e o gráfico da correspondente distribuição Normal não mista, estimada pelo Mathematica (2013), obtida com os mesmos dados reais, são:



GRÁFICO - Distribuição Normal e Normal mista – Call Center Empresa – 12-16 horas



Fonte: elaborado pelo autor

Observando o Gráfico percebe-se que o ajuste aos dados reais dos tempos de paciência da distribuição de probabilidade Normal mista é mais preciso do que o ajuste da distribuição de probabilidade Normal não mista.



## APÊNDICE - I

### Regra do Trapézio

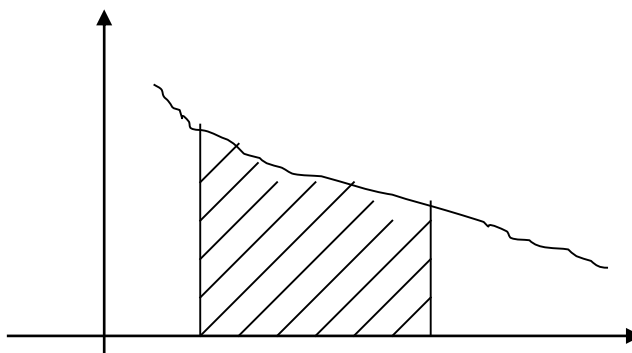
A partir do cálculo diferencial e integral sabe-se que, dada uma função  $f(x)$  contínua em um intervalo  $[a,b]$ , tem-se (BARROSO et al., 1987):

$$\int_a^b f(x)dx = F(b) - F(a) \quad (1)$$

onde

$$F'(x) = f(x)$$

Graficamente a interpretação da integral é a área sob o gráfico da função.



Em muitas situações pode ser difícil ou mesmo impossível a obtenção de  $F(x)$ . Também podem existir aplicações em que a função  $f(x)$  é conhecida apenas para valores tabelados em um intervalo  $[a,b]$ . Nestas situações fica inviabilizada a solução da integral. A saída é a utilização de métodos numéricos. A ideia básica da integração numérica é substituir a função  $f(x)$  por um polinômio que a aproxime razoavelmente no intervalo  $[a,b]$  e integrar o polinômio, ou seja:

$$\int_a^b f(x)dx = \int_a^b P_n(x)dx = A_0f(x_0) + A_1f(x_1) + A_2f(x_2) + \dots + A_nf(x_n)$$

onde  $x_i \in [a,b]$ ,  $i = 1, \dots, n$ .

Existem vários métodos para se obter o polinômio. Neste trabalho foi utilizado o método da Regra do Trapézio.

Nesta regra, a função a ser integrada será aproximada por um polinômio interpolador é de ordem 1. Portanto, necessita-se de dois pontos para a interpolação, ou seja  $[x_0, x_1] = [a, b]$ . Tem-se a expressão:

$$A = \int_a^b f(x)dx \approx \int_a^b P_1(x)dx$$

A partir da fórmula de Gregory-Newton, tem-se:

$$P_1(x) = f(x_0) + \frac{(x-x_0)}{h} \Delta f(x_0)$$

Resultando em:

$$A = \int_a^b f(x)dx \approx \int_a^b P_1(x)dx = \int_{x_0}^{x_1} \left[ f(x_0) + \frac{(x-x_0)}{h} \Delta f(x_0) \right] dx$$

Para facilitar esta integração, faz-se uma mudança de variáveis:

$$y_0 = f(x_0) \quad (2)$$

$$z = \frac{(x-x_0)}{h} \quad (3)$$

$$\Delta y_0 = \Delta f(x_0) = f(x_1) - f(x_0) \quad (4)$$

Resultando em:

$$A \approx \int_a^b [y_0 + z \Delta y_0] dx$$

Como  $z = \frac{(x-x_0)}{h}$  e derivando em relação a x, tem:

Mudando os limites  $\frac{dz}{dx} = \frac{1}{h} \Rightarrow dx = h dz$  de integração, temos:

$$a = x_0 \Rightarrow z = \frac{(x_0-x_0)}{h} = 0$$

$$b = x_1 \Rightarrow z = \frac{(x_1-x_0)}{h} = \frac{h}{h} = 1$$

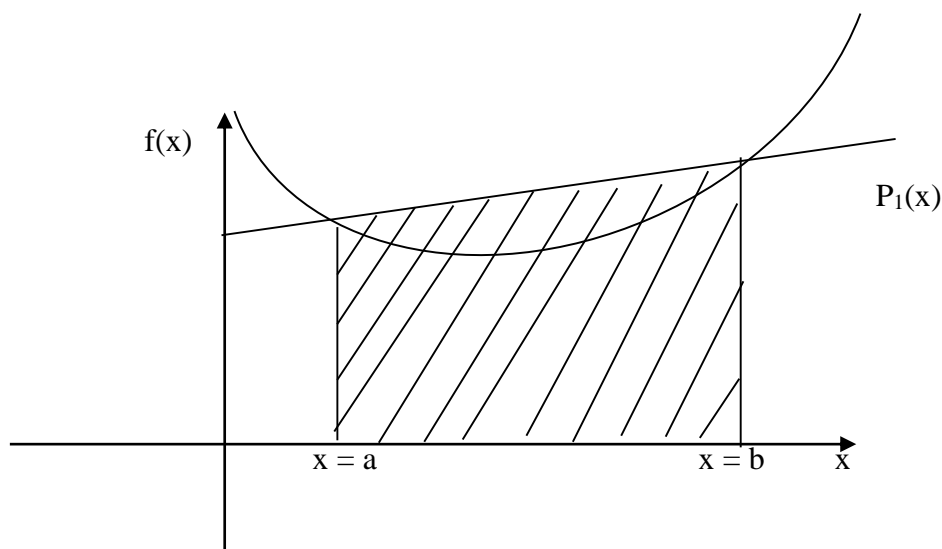
Integrando a expressão  $A \approx \int_0^1 [y_0 + z \Delta y_0] \cdot h \cdot dz$  acima, tem-se:

$$A \approx \int_0^1 [y_0 + z \Delta y_0] \cdot h \cdot dz = h \left[ zy_0 + \frac{z^2}{2} \Delta y_0 \right]_0^1 = h \left[ y_0 + \frac{\Delta y_0}{2} \right] \quad (5)$$

Substituindo-se as expressões (2) e (4) em (5), tem-se:

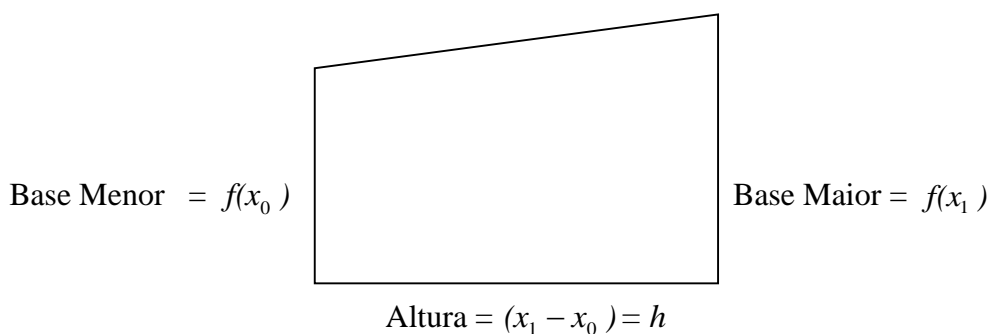
$$A = \int_a^b f(x) dx \approx h \left[ f(x_0) + \frac{f(x_1) - f(x_0)}{2} \right] = \frac{h}{2} [f(x_0) + f(x_1)] \quad (6)$$

A expressão (6) pode ser interpretada graficamente, do seguinte modo:



Observe que a área da região destacada representa um trapézio. A integral é aproximada pela área do trapézio:

$$A_T = \frac{(base\ maior + base\ menor) \times altura}{2} = \frac{h}{2} [f(x_0) + f(x_1)]$$



O erro cometido pela integração pela Regra do Trapézio é dado pela expressão (7). Este erro pode ser chamado de erro de truncamento, pois se truncou o a ordem do polinômio interpolador em 1:

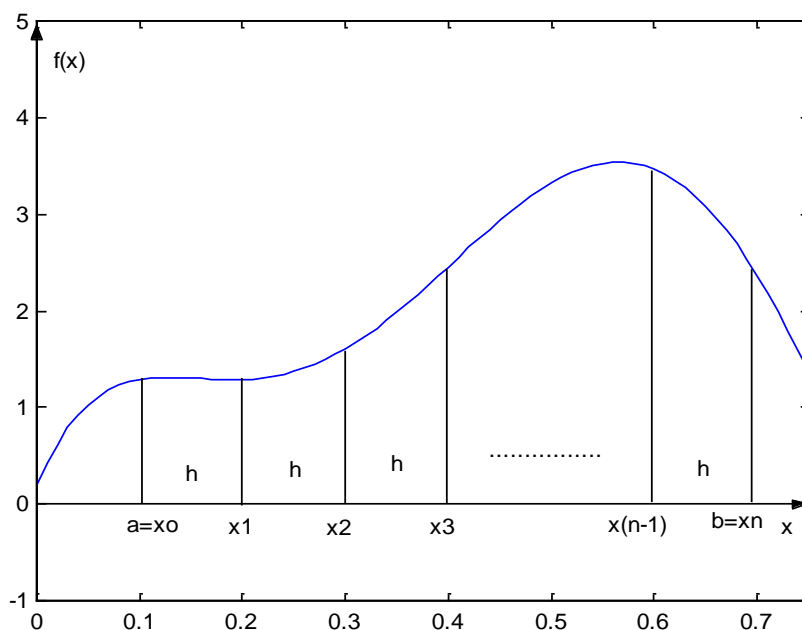
$$E_T = \frac{-h^3}{12} f''(c) \quad c \in [x_0, x_1] \quad (7)$$

Como o ponto  $c$  não é um ponto conhecido, a expressão (7) pode ser aproximada pela expressão (9):

$$E_i \leq \frac{h^3}{12} \cdot \max |f''(x)| \quad x \in [x_0, x_1] \quad (9)$$

### Regra do Trapézio Composta

Para melhorar a aproximação da integral obtida pela Regra do Trapézio, pode-se subdividir o intervalo  $[a,b]$  de integração em  $n$  subintervalos de amplitude  $h$  e aplicar a Regra do Trapézio em cada subintervalos, como pode ser visto na figura:



Aplicando-se novamente a Regra do Trapézio a cada um dos subintervalos, tem-se para a aproximação da integral:

$$A = \int_a^b f(x) dx \approx \frac{h}{2} [f(x_0) + f(x_1)] + \frac{h}{2} [f(x_1) + f(x_2)] + \dots + \frac{h}{2} [f(x_{n-1}) + f(x_n)]$$

Resultando em:

$$A = \int_a^b f(x) dx \approx \frac{h}{2} [f(x_0) + 2f(x_1) + 2f(x_2) + \dots + 2f(x_{n-1}) + f(x_n)]$$

O erro resultante é a soma dos erros cometidos na aplicação da Regra do Trapézio em cada subintervalo.

$$E_T = E_0 + E_1 + E_2 + \dots + E_{n-1} + E_n = \frac{-\left(\frac{b-a}{n}\right)^3 \cdot n}{12} f''(c) = \frac{-(b-a)^3}{12n^2} f''(c) \quad a \leq c \leq b$$

O erro cometido pode ser aproximado por:

$$E_T \leq \left| \frac{(b-a)^3}{12n^2} \right| \max |f''(x)| \quad a \leq x \leq b$$





## ANEXO - A

Amostra dos dados coletados do *Call Center* do Banco

call_id	ustomer_ic	priority	type	date	vru_entry	vru_exit	vru_time	q_start	q_exit	q_time	outcome	ser_start	ser_exit	ser_time	server
47529	0	0	PS	991101	06:31:36	06:31:47	11	00:00:00	00:00:00	0	HANG	00:00:00	00:00:00	0	NO_SERVER
47530	17878737	2	PE	991101	07:14:55	07:14:59	4	07:14:59	07:15:17	18	AGENT	07:15:16	07:23:16	480	YITZ
47531	0	0	PS	991101	07:42:41	07:42:51	10	00:00:00	00:00:00	0	AGENT	07:42:50	07:43:30	40	TOVA
47532	0	0	NW	991101	08:03:36	08:03:46	10	08:03:46	08:05:33	107	AGENT	08:05:31	08:07:38	127	YITZ
47533	4345955	2	PS	991101	08:17:40	08:17:46	6	08:17:46	08:18:01	15	AGENT	08:18:00	08:18:35	35	KAZAV
47534	0	0	PS	991101	08:27:47	08:27:56	9	00:00:00	00:00:00	0	AGENT	08:27:55	08:30:46	171	KAZAV
47535	0	0	PS	991101	08:41:11	08:41:21	10	00:00:00	00:00:00	0	AGENT	08:41:20	08:43:50	150	TOVA
47536	69438695	1	PS	991101	08:50:34	08:50:40	6	08:50:40	08:51:59	79	HANG	00:00:00	00:00:00	0	NO_SERVER
47537	0	0	PS	991101	08:58:25	08:58:50	25	00:00:00	00:00:00	0	HANG	00:00:00	00:00:00	0	NO_SERVER
47538	0	0	PS	991101	09:05:19	09:05:28	9	00:00:00	00:00:00	0	AGENT	09:05:27	09:08:09	162	ZIPI
47539	57281792	2	PS	991101	09:12:47	09:12:54	7	09:12:54	09:13:31	37	AGENT	09:13:30	09:15:16	106	ZIPI
47540	55279970	2	NE	991101	09:18:40	09:18:46	6	09:18:46	09:18:58	12	AGENT	09:18:57	09:25:12	375	VICKY
47541	25219700	2	PS	991101	09:25:52	09:25:58	6	09:25:58	09:26:17	19	AGENT	09:26:16	09:27:37	81	ZOHARI
47542	2057255	2	PS	991101	09:32:20	09:32:25	5	09:32:25	09:32:43	18	AGENT	09:32:43	09:33:47	64	VICKY
47543	65338329	2	PS	991101	09:37:47	09:37:52	5	09:37:52	09:38:03	11	AGENT	09:38:02	09:47:09	547	ZIPI
47544	2018299	2	PS	991101	09:44:16	09:44:21	5	09:44:21	09:44:53	32	AGENT	09:44:53	09:52:31	458	TOVA
47545	69687275	2	PS	991101	09:51:04	09:51:10	6	09:51:10	09:52:27	77	AGENT	09:52:26	09:55:18	172	ROTH
47546	57379505	2	PS	991101	09:56:01	09:56:07	6	09:56:07	09:56:52	45	HANG	00:00:00	00:00:00	0	NO_SERVER
47547	24286130	2	PS	991101	09:59:35	09:59:40	5	09:59:40	10:01:29	109	AGENT	10:01:29	10:02:17	48	TOVA
47548	52031556	2	PS	991101	10:07:20	10:07:25	5	10:07:25	10:07:29	4	AGENT	10:07:29	10:08:41	72	YITZ
47549	3968047	2	PS	991101	10:12:35	10:12:40	5	10:12:40	10:12:58	18	HANG	00:00:00	00:00:00	0	NO_SERVER
47550	51458594	2	PS	991101	10:18:55	10:19:00	5	10:19:00	10:19:16	16	AGENT	10:19:15	10:19:41	26	GOLDIE
47551	7191646	2	NE	991101	10:24:48	10:24:54	6	10:24:54	10:25:38	44	AGENT	10:25:37	10:54:16	1719	ELI
47552	63730469	1	PS	991101	10:29:42	10:29:47	5	10:29:47	10:30:03	16	AGENT	10:30:02	10:30:53	51	VICKY
47553	23053663	2	PS	991101	10:37:27	10:37:32	5	10:37:32	10:38:06	34	AGENT	10:38:04	10:40:24	140	TOVA
47554	27317791	1	PS	991101	10:43:20	10:43:25	5	10:43:25	10:44:53	88	HANG	00:00:00	00:00:00	0	NO_SERVER
47555	55476253	2	NE	991101	10:49:29	10:49:35	6	10:49:35	10:50:31	56	AGENT	10:50:30	10:51:06	36	VICKY
47556	42450676	2	PS	991101	10:55:19	10:55:25	6	10:55:25	10:56:53	88	HANG	00:00:00	00:00:00	0	NO_SERVER
47557	25022625	2	PS	991101	11:04:32	11:04:38	6	11:04:38	11:04:44	6	AGENT	11:04:44	11:05:10	26	KAZAV
47558	74780917	2	NE	991101	11:09:57	11:10:03	6	11:10:03	11:11:17	74	AGENT	11:11:15	11:16:41	326	KAZAV
47559	30663462	2	PS	991101	11:20:30	11:20:35	5	11:20:35	11:21:50	75	AGENT	11:21:49	11:28:03	374	GOLDIE
47560	50571827	2	PS	991101	11:28:09	11:28:14	5	11:28:14	11:29:17	63	AGENT	11:29:16	11:30:28	72	TOVA
47561	33052234	1	PS	991101	11:33:06	11:33:12	6	11:33:12	11:33:17	5	HANG	00:00:00	00:00:00	0	NO_SERVER
47562	0	0	PS	991101	11:44:11	11:44:29	18	00:00:00	00:00:00	0	AGENT	11:44:28	11:47:40	192	ZOHARI
47563	0	0	NE	991101	11:54:37	11:54:49	12	00:00:00	00:00:00	0	AGENT	11:54:48	11:57:56	188	ELI
47564	0	0	NE	991101	12:09:37	12:09:51	14	00:00:00	00:00:00	0	AGENT	12:09:50	12:11:44	114	VICKY
47565	0	0	PS	991101	12:21:27	12:21:35	8	00:00:00	00:00:00	0	AGENT	12:21:35	12:22:31	56	ZIPI
47566	0	0	PS	991101	12:34:45	12:34:55	10	00:00:00	00:00:00	0	AGENT	12:34:53	12:37:59	186	AVNI
47567	0	0	NW	991101	12:45:32	12:45:47	15	00:00:00	00:00:00	0	AGENT	12:45:45	12:46:12	27	GOLDIE



## ANEXO B

### Amostra dos dados coletados do *Call Center* da Empresa

ID ATENDIMENTO	ID TIPO ATENDIMENTO	TIPO ATENDIMENTO	CODIGO CLIENTE	NUMERO LOJA	NOME USUARIO	ASSUNTO DO ATENDIMENTO	DATA ENTRADA	HORA ENTRADA	POSICAO FILA
25782	4	SUPORTE TELEFÔNICO	598	1	NEILON	ECF	02/01/2009	08:04:10	1
25783	4	SUPORTE TELEFÔNICO	581	1	KALIL	MANUTENÇÃO/CONFIGURAÇÕES	02/01/2009	08:04:47	1
25784	4	SUPORTE TELEFÔNICO	1149	1	NILSON	ESTOQUE	02/01/2009	08:10:57	1
25785	4	SUPORTE TELEFÔNICO	142	1	OTACILIO	MANUTENÇÃO/CONFIGURAÇÕES	02/01/2009	08:11:41	1
25786	4	SUPORTE TELEFÔNICO	519	1	MARCIO	MANUTENÇÃO/CONFIGURAÇÕES	02/01/2009	08:13:42	1
25787	4	SUPORTE TELEFÔNICO	1463	1	ANTONIO	MANUTENÇÃO/CONFIGURAÇÕES	02/01/2009	08:18:52	1
25788	4	SUPORTE TELEFÔNICO	581	1	KALIL	SINTEGRA	02/01/2009	08:20:37	1
25789	4	SUPORTE TELEFÔNICO	333	1	AGHATA	INSTALAÇÃO DE SERVIDOR/TERMINAL	02/01/2009	08:22:05	1
25790	4	SUPORTE TELEFÔNICO	22	1	RAFAEL	INSTALAÇÃO DE SERVIDOR/TERMINAL	02/01/2009	08:29:11	1
25791	4	SUPORTE TELEFÔNICO	1474	1	EDISERGIO	MANUTENÇÃO/CONFIGURAÇÕES	02/01/2009	08:38:45	1
25792	4	SUPORTE TELEFÔNICO	772	1	JULIANO	MANUTENÇÃO/CONFIGURAÇÕES	02/01/2009	08:39:49	1
25793	4	SUPORTE TELEFÔNICO	581	1	KALIL	PBM	02/01/2009	08:45:40	1
25794	4	SUPORTE TELEFÔNICO	453	1	ANDREA	PBM	02/01/2009	08:46:31	1

**Amostra dos dados coletados do *Call Center* da Empresa - continuação**

DATA INICIO	HORA INICIO	DATA TERMINO	HORA TERMINO	STATUS	CONCLUSAO	ID CATEGORIA	SISTEMA	CATEGORIA ATENDIMENTO	MOTIVO DESISTENCIA
02/01/2009	08:05:18	02/01/2009	08:08:30	F	S	23	O	ECF	
02/01/2009	08:04:56	02/01/2009	08:06:16	F	S	9	O	MANUTENÇÃO/CONFIGURA	
02/01/2009	08:12:05	02/01/2009	08:15:15	F	S	19	O	ESTOQUE	
02/01/2009	08:12:58	02/01/2009	08:14:16	F	S	9	O	MANUTENÇÃO/CONFIGURA	
02/01/2009	08:14:01	02/01/2009	08:15:11	F	S	9	O	MANUTENÇÃO/CONFIGURA	
02/01/2009	08:20:01	02/01/2009	08:20:07	F	S	9	O	MANUTENÇÃO/CONFIGURA	
02/01/2009	08:21:46	02/01/2009	08:31:55	F	S	41	O	SINTEGRA	
02/01/2009	08:23:35	02/01/2009	08:28:19	F	S	31	O	INSTALAÇÃO DE SERVID	
02/01/2009	08:30:40	02/01/2009	08:37:03	F	S	31	O	INSTALAÇÃO DE SERVID	
02/01/2009	08:38:59	02/01/2009	08:39:54	F	S	9	O	MANUTENÇÃO/CONFIGURA	
02/01/2009	08:40:58	02/01/2009	08:41:03	F	S	9	O	MANUTENÇÃO/CONFIGURA	
02/01/2009	08:45:51	02/01/2009	08:47:10	F	S	24	O	PBM	
02/01/2009	08:48:00	02/01/2009	08:49:10	F	S	24	O	PBM	