

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**Instituto de Matemática**  
**Departamento de Estatística**

**APLICAÇÃO DE MÉTODOS MULTIVARIADOS  
NO ESTUDO DE TELEFACILIDADES**

**Autor: Denis Altieri de Oliveira Moraes**

**Orientador: Álvaro Vigo**

**Monografia apresentada para a obtenção  
do título de Bacharel em Estatística**

**Porto Alegre, Abril de 2001**

## ***AGRADECIMENTOS***

À CRT Brasil Telecom, por gentilmente ceder os recursos para que este trabalho fosse realizado, e aos colegas da Departamento de Análise de Mercado pela experiência e aprendizado no trabalho de equipe.

Aos vários colegas do curso de Bacharelado em Estatística que perseveraram comigo em momentos difíceis, enfrentando as dificuldades até o final.

A todos os professores do Departamento de Estatística, especialmente ao Professor Álvaro Vigo, pela competência e orientação durante este trabalho, bem como durante grande parte do curso, sempre disposto a conceder explicações detalhadas e auxílio nas dúvidas.

Aos grandes e valiosos amigos que encontrei na faculdade durante estes anos e que, certamente, me acompanharão ainda durante muito tempo, em especial a Carolina Cabrales e Francisca Cabrales.

Aos meus pais, Delci e Eneli, pela dedicação e esforço que tiveram de várias formas e em todos os momentos de minha vida e de meus irmãos Úrsula e Bruno, com muito amor.

A Deus, sustentador de todas as coisas com insondável compaixão e sabedoria.

## ÍNDICE

<b>1 Introdução .....</b>	<b>4</b>
<b>2 Apresentação dos Dados .....</b>	<b>6</b>
2.1 O Conjunto de Dados .....	6
2.2 Aspectos Computacionais .....	9
<b>3 Aspectos Metodológicos .....</b>	<b>11</b>
3.1 Análise Fatorial .....	11
3.2 Análise de Agrupamento .....	17
<b>4 Estudo das Telefácilidades .....</b>	<b>22</b>
4.1 Faturamento Total .....	22
4.2 Divisão em Estratos .....	26
4.2.1 Estrato I: Faturamento Alto .....	28
4.2.2 Estrato II: Faturamento Baixo .....	30
<b>5 Considerações Finais .....</b>	<b>39</b>
<b>6 Anexos .....</b>	<b>42</b>
6.1 Anexo 1 – Programas .....	43
6.2 Anexo 2 - Diagramas de Árvore .....	45
6.3 Anexo 3 – Gráficos e Tabelas .....	53
<b>7 Referências Bibliográficas .....</b>	<b>56</b>

## 1 INTRODUÇÃO

A motivação para esta pesquisa surgiu durante a realização do estágio obrigatório do curso de Bacharelado em Estatística na Companhia Riograndense de Telecomunicações – CRT Brasil Telecom. A CRT presta serviços na área de telefonia fixa no estado do Rio Grande do Sul e oferece aos seus assinantes diversas opções de serviços, objetivando a satisfação dos seus clientes e sua confirmação no mercado, que desde julho de 1999 foi aberto à concorrência.

O estágio foi realizado na Superintendência de Marketing e Comercial, Gerência de Análise de Mercado. Esta gerência estuda mensalmente bancos de dados contendo todo o tráfego e faturamento da empresa. Com base nestas análises são produzidos informes de periodicidade mensal, e outros que são solicitados pelos demais setores para auxílio na tomada de decisões estratégicas.

Contudo, a maior parte do trabalho desenvolvido até o momento abrange somente técnicas descritivas univariadas, em parte devido ao desconhecimento da aplicabilidade das técnicas multivariadas disponíveis. Particularmente no que tange a alguns tipos de serviços, tais como telefacilidades, poucos trabalhos foram feitos até agora.

Telefacilidades são os serviços especiais incorporados ao telefone do assinante, tais como Siga-me, Teleconferência, Atendimento Simultâneo e outros.

Devido ao desconhecimento do perfil dos assinantes que possuem as telefacilidades, decidiu-se realizar um trabalho descrevendo estes usuários através de uma amostra obtida de uma pesquisa prévia sobre satisfação dos clientes quanto a estes serviços. Nesta pesquisa, estimou-se que apenas 20% dos usuários que possuem telefacilidades realmente utilizavam os serviços.

O objetivo deste trabalho é encontrar fatores subjacentes que expliquem o comportamento dos usuários, reduzindo a dimensão das variáveis através das técnicas estatísticas de Análise Multivariada, particularmente Análise Fatorial, e posteriormente classificar os usuários através da Análise de Agrupamento. Ou seja, deseja-se encontrar um pequeno número de fatores que expliquem o consumo dos assinantes de forma mais simplificada e posteriormente verificar se existem grupos de usuários gerados pelo uso das diferentes Telefácilidades em relação aos diferentes fatores. Estas técnicas estão amplamente discutidas na literatura, destacando-se Johnson & Wichern (1988) e Mardia, Kent and Bibby (1979), entre outros.

Devido ao faturamento da empresa ter caráter confidencial, foi realizada uma transformação nas variáveis, preservando-se, assim, o sigilo das informações.

Inicialmente são apresentadas no Capítulo 2 as variáveis do banco de dados, referente ao mês de Maio de 2000, o mesmo mês do qual foi extraída a amostra para a pesquisa sobre satisfação dos clientes em relação às telefácilidades. No Capítulo 3, são tratados dos aspectos metodológicos das técnicas de Análise Fatorial e de Agrupamento. No Capítulo 4 é feita a aplicação destas técnicas no banco de dados, e uma análise descritiva mais detalhada dos grupos de interesse. No Capítulo 5 estão as considerações finais sobre os resultados obtidos. Os programas utilizados para gerar os resultados e outros gráficos e tabelas maiores serão apresentadas nos Anexos.

## 2 APRESENTAÇÃO DOS DADOS

Para uma melhor compreensão das variáveis estudadas neste trabalho, a seguir é apresentado um breve resumo sobre o sistema de faturamento da empresa. Como já foi mencionado, com o objetivo de conservar o caráter confidencial destas informações, os valores monetários sofreram uma transformação linear do tipo  $a.X + b$ , onde  $a$  e  $b$  são constantes arbitrárias.

### 2.1 O CONJUNTO DE DADOS

Todas as ligações telefônicas ou chamadas realizadas são tarifadas de acordo com a distância, a hora, o tipo de dia, a duração e o tipo de destino. Após o sistema de tarifação valorar cada chamada, elas são classificadas segundo o tipo de chamada e anexadas ao total da fatura do telefone. Os tipos de chamada existentes para o sistema de faturamento são:

- *Chamadas Locais*: o destino pertence à mesma cidade;
- *Chamadas Intra-rede*: o destino é outra cidade qualquer dentro da área de cobertura da CRT, ou seja, dentro do Rio Grande do Sul;
- *Chamadas Inter-rede*: o destino está fora da área de cobertura da CRT, ou seja, ligações para outros estados do Brasil ou para a operadora CTMR;
- *Chamadas Internacionais*: o destino é outro país;

- *Chamadas Fixo-móvel locais*: o destino é um telefone celular da mesma área de numeração fechada;
- *Chamadas Fixo-móvel intra-rede*: o destino é um telefone celular de diferente área de numeração fechada dentro do estado; e,
- *Chamadas Fixo-móvel inter-rede*: o destino é um telefone celular de fora do estado do Rio Grande do Sul.

Os tipos de telefones podem ser classificados de várias maneiras. Este trabalho, entretanto, limita-se a observar as três categorias principais, que são: *não-residencial*, *residencial* e *tronco*. Geograficamente, os telefones são distribuídos de acordo com o código de área no Estado, que são: 051, 053, 054 e 055.

Ao total faturado por cada telefone devido às ligações realizadas, também é acrescido um valor que corresponde à assinatura básica, consertos e outras despesas como anúncios de publicidade e taxas relativas à disponibilização das telefácilidades, para o telefone, as quais são o interesse principal.

As informações sobre as telefácilidades originalmente eram compostas por nove variáveis separadas, indicadas por um número de 1 a 9 correspondente a cada possível serviço oferecido ao telefone, ou zero caso não houvesse nenhum. Como um trabalho envolvendo estas variáveis deste banco de dados nunca havia sido realizado, foram descobertas inconsistências que depois de um intenso estudo foram corrigidas e as telefácilidades foram identificadas por uma única variável.

Desta variável, surgiram 233 combinações de telefácilidades diferentes. Estas combinações foram reagrupadas de acordo com critérios de freqüência e

importância, restando assim 14 combinações denominadas de *pacotes*. Os pacotes **A**, **B**, **C** e **D** são formados pelas combinações que tiveram maiores freqüências, enquanto que o pacote **O** é formado pelas combinações restantes, com freqüências menores. Finalmente, os pacotes rotulados de **1** a **9** identificam as nove telefacilidades, individualmente. O pacote **Z** é apenas uma denominação para telefones sem telefacilidades.

A Tabela 2.1 a seguir, apresenta a distribuição de freqüências das telefacilidades, identificadas pelos pacotes, dos clientes da CRT em maio de 2000. Na Tabela 2.2 são apresentadas as variáveis de interesse no presente estudo.

**Tabela 2.1 – Pacotes de telefacilidades dos clientes em maio de 2000**

Pacote	Descrição	Freqüência	% Telefacilidades	% Total
1	Atendimento Simultâneo	25.228	12,14%	
2	Conferência	5.362	2,58%	
3	Siga-me	19.132	9,21%	
4	Discagem Abreviada	354	0,17%	
5	Linha Executiva	277	0,13%	
6	Não Perturbe	481	0,23%	
7	Transferência	83	0,04%	
8	Transferência Caso Não Responde	2.979	1,43%	13,18%
9	Transferência Caso Ocupado	4.772	2,30%	
A	Atendimento Simultâneo & Siga-me	125.275	60,30%	
B	Atendim. Simultâneo & Conferência & Transf. N. Resp.	10.660	5,13%	
C	Atendim. Simultâneo & Conferência & Siga-me	2.878	1,39%	
D	Atendim. Simultâneo & Conferência	1.982	0,95%	
O	Outras Combinações Menores	8.280	3,99%	
<b>Total com telefacilidade:</b>		<b>207.743</b>	<b>100%</b>	-
Z	Não possui telefacilidade	1.367.864	-	86,82%
<b>Total:</b>		<b>1.575.607</b>	-	<b>100%</b>



**Tabela 2.2 - Descrição das variáveis do estudo**

Nome da Variável	Descrição
COD_DDD	Código de área (DDD)
TELEF	Número do telefone
TOT_ING	Total de ingresso faturado no mês
VAL_ME	Valor devido às chamadas locais
PULSOS	Quantidade de pulsos (chamadas locais)
VAL_IUA	Valor das chamadas intra-rede
QTDE_IUA	Quantidade de minutos intra-rede
VAL_IUR	Valor das chamadas inter-rede
QTDE_IUR	Quantidade de minutos inter-rede
VAL_INT	Valor das chamadas internacionais
QTDE_INT	Quantidade de minutos internacionais
VAL_FML	Valor das chamadas fixo-móvel locais
QTDE_FML	Quantidade de minutos fixo-móvel locais
VAL_FMA	Valor das chamadas fixo-móvel intra-rede
QTDE_FMA	Quantidade de minutos fixo-móvel intra-rede
VAL_FMR	Valor das chamadas fixo-móvel inter-rede
QTDE_FMR	Quantidade de minutos fixo-móvel inter-rede
VAL_OUT	Valor de outros serviços (assinatura básica, concertos, etc.)
CATE3	Categoria do telefone
PACOTE	Grupo de telefacilidades

Convém mencionar que as chamadas internacionais deixaram de ser faturadas pela CRT após a entrada do CSP (Código de Seleção de Prestadora), em julho de 1999. Contudo, restaram telefones com contas pendentes nesta abertura durante alguns meses, mas a maioria dos valores no banco de dados do mês de maio de 2000 estão zerados para chamadas internacionais.

## **2.2 ASPECTOS COMPUTACIONAIS**

O tamanho do banco de dados foi uma das dificuldades encontradas neste estudo, pois a quantidade imensa de registros, mais de um milhão e meio, demanda uma grande capacidade de armazenamento. Cada mês faturado tem em média 400 Mb (quatrocentos Mega Bytes), precisando ser compactados e transferidos periodicamente para unidades de fita DAT afim de liberar espaço em disco para processamento de outras tarefas. Geralmente, quando é necessário

trabalhar com arquivos de faturamento e fazer alterações nos dados originais, são feitas cópias e divisões deste arquivo, chegando a demandar até 2 Gb (dois Giga Bytes) de espaço no disco rígido do computador, somente para os dados.

Outro problema, decorrente do tamanho dos arquivos de dados, é o tempo de processamento. Para que seja executada uma rotina de cálculo de estatísticas descritivas de uma única variável são necessários cerca de três minutos em um computador com processador AMD K6-II<sup>®</sup> 500Mhz e 128 Mb de memória RAM. Isto considerando-se que não exista nenhuma interferência de outro programa no processador. Uma rotina de importação de dados pode levar 15 minutos, enquanto que procedimentos mais sofisticados, como os que utilizam cálculos matriciais na Análise Fatorial, chegam a demandar mais de 30 minutos de processamento. Vale ressaltar que a maior parte deste trabalho foi realizada com um processador Pentium<sup>®</sup> 200 MMX, bem mais lento que o citado acima, utilizado somente na fase final do trabalho.

Além disto, a escolha do software para o processamento das informações é de crucial importância. No início das pesquisas, o departamento de Análise de Mercado da CRT utilizou o software **SPSS<sup>®</sup>**. Com o tempo este software foi substituído pelo sistema **SAS<sup>®</sup>** (Statistical Analysis System), por apresentar um sistema próprio de compactação de dados que mantém os arquivos em um formato que utiliza menos espaço em disco, podendo fazer as análises dos dados com este sistema de compactação ativo, e também pela melhor performance para a obtenção dos resultados. Todas as análises realizadas neste trabalho foram feitas no sistema **SAS**.

### 3 ASPECTOS METODOLÓGICOS

Neste trabalho utilizam-se duas técnicas da Análise Multivariada, a Análise Fatorial e a Análise de Agrupamento, juntamente com outros procedimentos da análise descritiva e exploratória de dados. Nas próximas seções são brevemente descritas as técnicas multivariadas mencionadas.

#### 3.1 ANÁLISE FATORIAL

O objetivo básico da Análise Fatorial é tentar descrever a estrutura de correlação de um grupo grande de variáveis, através de um número menor de fatores subjacentes, não observáveis diretamente. Basicamente, os fatores subjacentes são formados pelos grupos de variáveis que possuem correlação forte entre si e devem explicar a maior parte da variação total dos indivíduos. Não é objetivo deste trabalho entrar em detalhes matemáticos sobre o assunto, pois estes podem ser encontrados em uma ampla bibliografia, como por exemplo, em Mardia, Kent and Bibby (1979, p. 255), Johnson & Wichern (1988, p. 378) e Fachel (1976). Contudo, descreveremos em linhas gerais os conceitos básicos do modelo k-fatorial.

Dado um vetor aleatório  $X = (X_1, X_2, X_3, \dots, X_p)'$  com vetor de médias  $\mu$  e matriz de covariância  $\Sigma$ , o modelo k-fatorial é dado por

$$X_i = \sum_{j=1}^k \lambda_{ij} f_j + e_i + \mu_i; i = 1, \dots, p$$

onde

- $k$  é o número de fatores comuns do modelo ( $k < p$ );

- $\lambda_{ij}$  é a carga fatorial da  $i$ -ésima variável no  $j$ -ésimo fator comum, refletindo a importância do  $j$ -ésimo fator na composição da  $i$ -ésima variável;
- $f_j$  é o  $j$ -ésimo fator comum, que está presente em todas as variáveis  $X_i$  (o fator  $f_j$  explica uma parte da variabilidade de  $X_i$  e  $f_j$  explica uma parte da variabilidade de  $X_i$  que não foi explicada pelo fator  $f_j$ );
- $e_j$  são os fatores únicos ou específicos que descrevem a variação residual específica da  $i$ -ésima variável; ou seja, é a parte da variabilidade da variável  $X_i$  que não é partilhada com as outras variáveis através dos fatores comuns; e,
- $\mu_i$  é a média da  $i$ -ésima variável.

Usando a notação matricial, onde caracteres grifados em negrito representam uma matriz ou um vetor, o modelo  $k$ -fatorial pode ser escrito como

$$\mathbf{X} = \mathbf{\Lambda} + \mathbf{e} + \boldsymbol{\mu}$$

onde

$$\mathbf{X}_{p \times 1} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \longrightarrow \text{vetor das variáveis aleatórias originais;}$$

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1k} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pk} \end{bmatrix} \longrightarrow \text{matriz de cargas fatoriais;}$$

$$\mathbf{f}_{k \times 1} = \begin{bmatrix} f_1 \\ \vdots \\ f_k \end{bmatrix} \longrightarrow \text{vetor dos fatores comuns;}$$

$$\mathbf{e}_{p \times 1} = \begin{bmatrix} e_1 \\ \vdots \\ e_p \end{bmatrix} \rightarrow \text{vetor dos fatores específicos ou erros aleatórios;}$$

$$\boldsymbol{\mu}_{p \times 1} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} \rightarrow \text{vetor de médias do vetor aleatório } \mathbf{X} = (X_1, X_2, \dots, X_p)'$$

O modelo k-fatorial possui as seguintes propriedades:

$E(\mathbf{f}) = \mathbf{0}$ ;  $Cov(\mathbf{f}) = \mathbf{I}$ ;  $E(\mathbf{e}) = \mathbf{0}$ ;  $Cov(e_i, e_j) = 0$ ;  $\forall i \neq j$  e  $Cov(\mathbf{f}, \mathbf{e}) = \mathbf{0}$ . Assim, todos os fatores são não correlacionados entre si e os fatores comuns são padronizados de tal forma que têm variância igual a um. Por sua vez, a matriz de covariâncias do vetor de fatores específicos  $\mathbf{e}$  pode ser escrita como

$$Cov(\mathbf{e}) = diag \{ \Psi_{11}, \dots, \Psi_{pp} \}.$$

Dessa forma, no modelo k-fatorial as variâncias das variáveis observadas são escritas por

$$\begin{aligned} Var(X_i) = \sigma_i^2 &= \lambda_{i1}^2 + \lambda_{i2}^2 + \dots + \lambda_{ik}^2 + Var(e_i) \\ &= \sum_{j=1}^k \lambda_{ij}^2 + \Psi_{ii}; i = 1, \dots, p \end{aligned}$$

Assim, a variância da variável aleatória  $X_i$  pode ser dividida em duas partes: uma primeira parte, chamada de *comunalidade*, é determinada por  $\sum_{j=1}^k \lambda_{ij}^2$ . Ela representa a variância da variável  $X_i$  que é partilhada com as outras variáveis através dos fatores comuns. A segunda parte, denominada *especificidade*, é a variância específica ou única  $\Psi_{ii}$ , devida ao fator único  $e_{ii}$ , e representa a

variabilidade de  $X_i$  que não partilhada com as outras variáveis através dos fatores comuns.

O modelo fatorial especifica também que a covariância entre duas variáveis quaisquer, digamos  $X_i$  e  $X_j$ , pode ser expressada por  $Cov(X_i, X_j) = \sum \lambda_{ij} \lambda_{kj}$ ;  $i \neq j$ , enquanto que a covariância entre a variável  $X_i$  e o fator comum  $f_j$  é  $Cov(X_i, f_j) = \lambda_{ij}^2$ , podendo ser interpretada como o grau em que  $X_i$  depende do  $j$ -ésimo fator comum.

A validade do modelo  $k$ -fatorial pode ser expressada em termos de uma condição da matriz de covariâncias  $\Sigma$ , na forma  $\Sigma = \Lambda \Lambda' + \Psi$ ; isto é, se a matriz  $\Sigma$  pode ser decomposta desta forma, então o modelo  $k$ -fatorial está satisfeito para o vetor aleatório  $X$ . Veja Mardia, Kent and Bibby (1979, p. 257).

No entanto, como a análise é invariante quanto à escala, isto é, não é afetada por transformações na escala das variáveis, na estimação dos parâmetros pode-se decompor a matriz de correlação amostral  $R$  ao invés da matriz de covariância amostral  $S$ .

Um dos métodos de extração dos fatores freqüentemente utilizado é o denominado Análise do Fator Principal, o qual possui grande similaridade com a técnica da Análise de Componentes Principais. A diferença básica está nas estimativas iniciais das comunalidades: na ACP decompõe-se a matriz de correlação amostral  $R$  (ou, em certas situações especiais, a matriz de covariâncias  $S$ ) enquanto no modelo fatorial decompõe-se a matriz de correlação reduzida  $R - \hat{\Psi}$ . Nesta matriz, os elementos da diagonal principal são as estimativas iniciais das comunalidades, enquanto que os demais elementos são as correlações entre as variáveis. Assim, se no método de extração dos fatores pela Análise do Fator Principal as comunalidades iniciais são iguais a 1, então o

modelo fatorial não possui fatores específicos e, conseqüentemente, é equivalente à Análise de Componentes Principais.

No procedimento de estimação, em uma primeira etapa obtêm-se as estimativas preliminares das comunalidades e, na etapa seguinte, estima-se a matriz de cargas dos fatores  $\Lambda$ . Os procedimentos computacionais disponíveis já possuem implementadas diferentes técnicas pelas quais as comunalidades podem ser obtidas, cabendo destacar:

- i) O quadrado do coeficiente de correlação múltipla da  $i$ -ésima variável com todas as outras variáveis;
- ii) O maior coeficiente de correlação entre a  $i$ -ésima e cada um das outras variáveis.

No presente trabalho, foi aplicado o segundo método de estimação das comunalidades iniciais, obtido através da opção PRIORS=MAX no procedimento **FACTOR** do sistema SAS (veja os programas no Anexo 1). Finalmente, na terceira etapa obtêm-se a estimativa da matriz de variâncias específicas  $\Psi$ .

Outro método de extração dos fatores é o Método da Máxima Verossimilhança, o qual exige que o vetor aleatório  $\underline{X}$  tenha uma distribuição normal multivariada. Como é relativamente difícil o atendimento desta exigência, o método da Análise do Fator Principal é mais utilizado na prática.

É importante destacar que podem ocorrer soluções impróprias (*heywood case*, em inglês) para o modelo  $k$ -fatorial, tanto na estimação pelo método do fator principal como pelo método da máxima verossimilhança, ocorrendo com maior freqüência neste último.

A apresentação detalhada dos métodos de estimação dos parâmetros está fora dos objetivos deste trabalho, e os mesmos encontram-se amplamente descritos e discutidos na literatura. Veja, por exemplo, Mardia, Kent and Bibby (1979, p. 261 – 267), Johnson & Wichern (1988, p. 384 – 400), Morrison (1976), entre outros.

Outro aspecto importante do modelo k-fatorial é a não unicidade das cargas dos fatores, que pode ser resolvida mediante o uso de restrições matemáticas apropriadas. Para melhorar a interpretação dos fatores, em algumas situações pode ser conveniente realizar uma rotação dos fatores, em geral ortogonal, sem afetar a validade do modelo.

Vários métodos para rotação dos fatores estão descritos na literatura e implementados nos procedimentos computacionais e nenhum deles pode ser considerado melhor do que outros sob o ponto de vista estatístico. Um dos mais comuns e utilizados é o método de rotação ortogonal denominado Varimax, cuja descrição teórica pode ser encontrada em Mardia, Kent and Bibby (1979, p. 269).

Após a estimação das cargas fatoriais  $\lambda_{ij}$  e das variâncias específicas  $\psi_{ii}$ , pode ser importante estimar os valores dos fatores comuns para cada indivíduo, chamados escores fatoriais. Os escores fatoriais são usados como variáveis de entrada em análises subseqüentes, como, por exemplo, para a Análise de Agrupamento. Para estimar os escores fatoriais, alguns métodos têm sido referidos, destacando-se o *Método de Bartlett* e o *Método de Regressão* ou *Método de Thompson*. Estes métodos encontram-se descritos em Mardia, Kent and Bibby (1979, p. 274), Johnson & Wichern (1988, p. 409) e Possoli (1992, p. 45), por exemplo.



### **3.2 ANÁLISE DE AGRUPAMENTO**

A Análise de Agrupamento é uma técnica de Análise Multivariada extremamente útil para encontrar agrupamentos naturais ou convenientes dos indivíduos, simplificando a descrição de um grande conjunto de dados ou para gerar hipóteses a serem testadas em outros estudos.

A informação básica para esta técnica é um conjunto de  $N$  indivíduos sobre os quais são registradas  $p$  medidas. A escolha deste conjunto inicial de medidas, usadas para descrever cada indivíduo, deve refletir o julgamento de investigador quanto a sua importância para a obtenção de um esquema de classificação. Desta forma, a primeira questão a ser considerada sobre as variáveis é se foram eleitas corretamente, ou seja, se são adequadas para o tipo de classificação desejado.

A próxima questão que deve ser considerada diz respeito ao número de variáveis em uma aplicação prática das técnicas de Análise de Agrupamento. Isto é importante porque a similaridade entre os indivíduos depende não somente dos atributos medidos mas também do seu número. Assim como na escolha das variáveis, não há uma base teórica para a determinação do número de variáveis mais adequado, devendo ser feita empiricamente. Em muitas situações o pesquisador tende a manter um número grande de variáveis, o que poderá causar problemas nos métodos de agrupamento nos quais a quantidade de cálculos aumenta drasticamente com um aumento no número de variáveis. Assim, nestas situações usualmente recomenda-se o uso da técnica de Análise Fatorial ou de Componentes Principais antes de proceder a Análise de Agrupamento, onde os

escores fatoriais (ou das componentes) são usadas como variáveis de entrada no processo de agrupamento.

O objetivo básico da Análise de Agrupamento é, dado um número de unidades ou indivíduos, descritos por um conjunto de variáveis, obter um esquema de classificação para agrupá-los de tal forma que os indivíduos dentro dos grupos são similares em algum aspecto, mas diferentes dos elementos dos outros grupos. Maiores detalhes sobre as idéias envolvidas na Análise de Agrupamento são encontradas em Anderberg (1973), Everitt (1982) e Hartigan (1975), entre outros.

Os métodos de Análise de Agrupamento que têm sido referidos na literatura podem ser classificados em aproximadamente cinco tipos principais. Estes métodos encontram-se amplamente discutidos na literatura, veja por exemplo, Mardia, Kent and Bibby (1979, p. 360), Johnson & Wichern (1988, p. 543), Everitt (1982, p. 23), Anderberg (1973, p. 131). No entanto, é útil apresentar brevemente os princípios básicos de dois métodos de agrupamento utilizados no trabalho, denominados de *Métodos Hierárquicos* e *Métodos Não Hierárquicos*.

O agrupamento obtido através dos métodos hierárquicos utiliza dois tipos de algoritmos: os aglomerativos e os divisivos. Nos *métodos aglomerativos hierárquicos*, o número inicial de grupos é igual ao número de indivíduos. Os indivíduos ou objetos mais similares são agrupados primeiro e esses grupos iniciais são unidos de acordo com suas similaridades. No final do processo, todos os subgrupos são unidos em um único grupo.

Os *métodos divisivos hierárquicos* funcionam de modo inverso: o processo inicia com todos os indivíduos em um único grupo, que é dividido em dois subgrupos tais que os indivíduos pertencentes a um subgrupo são diferentes, em

algum sentido, dos indivíduos do outro subgrupo. Esses subgrupos são então divididos em subgrupos dissimilares e o processo continua até que existam tantos grupos quanto o número de indivíduos.

É importante destacar que a avaliação de quão próximo ou similar um indivíduo (ou grupo) está em relação a outro, depende da definição de medidas de distância e similaridade apropriadas. Uma ampla discussão sobre essas medidas pode ser encontrada em Mardia, Kent and Bibby (1979, p. 375) e Johnson & Wichern (1988, p. 545).

Outro aspecto importante é a escolha de um ponto de parada ou de corte do processo de agrupamento, o qual determina o número de grupos resultantes. Deve-se observar que não existem métodos satisfatórios para a determinação do número de grupos existentes na população, qualquer que seja o método de agrupamento usado, sendo, assim, uma escolha subjetiva. No entanto, alguns métodos para estimar o número de grupos têm sido sugeridos na literatura. Veja Mardia, Kent and Bibby (1979, p. 365), Everitt (1982, p. 64) ou SAS Institute Inc. (1989, p. 97).

O método de agrupamento aglomerativo hierárquico utilizado neste trabalho é o *Método de Mínima Variância de Ward*, descrito em Everitt (1982, p. 31) e Anderberg (1973, p. 142). Neste método, a distância entre dois grupos é dada pela soma dos quadrados entre os grupos. A cada etapa do processo, são combinados os dois indivíduos ou grupos cuja união apresentar o menor aumento na soma de quadrados dentro dos grupos. Assim, a perda de informação resultante da união dos indivíduos em grupos pode ser medida pela divisão da soma de quadrados dentro dos grupos pela soma de quadrados total, ou seja;

$$\% \text{ perda de informação} = \frac{\text{SQDentro}}{\text{SQTotal}} \times 100$$

No sistema **SAS** está disponível o método de mínima variância de Ward no procedimento **PROC CLUSTER**. Aspectos metodológicos deste procedimento também são apresentados em SAS Institute Inc. (1989, p. 535).

Diferentemente dos métodos hierárquicos, os métodos não hierárquicos são designados para agrupar indivíduos em uma única classificação de  $k$  grupos, onde  $k$  pode ser especificado *a priori* ou determinado como parte do procedimento de agrupamento. A idéia central destes métodos é escolher alguma partição inicial dos indivíduos e então movê-los entre os grupos até obter-se a melhor partição. Os métodos não hierárquicos são procedimentos mais ágeis para agrupar grandes conjuntos de dados do que os métodos hierárquicos, pois não necessitam do cálculo e armazenamento da matriz de similaridades, podendo ser aplicados a conjuntos de dados maiores. Veja Anderberg (1973, p. 156).

No sistema **SAS** está disponível o procedimento não hierárquico denominado **FASTCLUS**, que se destina ao uso em conjuntos de dados de, aproximadamente, 100 até 100.000 observações. O procedimento **FASTCLUS** foi diretamente inspirado em um algoritmo proposto por Hartigan (1975, p. 74), denominado "*leader algorithm*" e no algoritmo denominado "*k-means algorithm*". Veja SAS Institute Inc. (1989, p. 824). O método utilizado pelo procedimento **FASTCLUS** é chamado por Anderberg (1973, p. 160) de "*nearest centroid sorting*" e segue o seguinte princípio: um conjunto de pontos denominados *sementes dos grupos* (*cluster seeds*, em inglês) é selecionado como uma estimativa da média dos grupos. Cada observação é designada à semente mais próxima para formar grupos temporários. As sementes são então substituídas pela média dos grupos

temporários e o processo é repetido até que não ocorram novas mudanças nos grupos. Detalhes destes e de outros algoritmos similares encontram-se em Anderberg (1973) e Hartigan (1975). Este procedimento foi aplicado ao conjunto de dados deste trabalho, com aproximadamente 1.500.000 observações e apresentou resultados bastante satisfatórios.

Nos próximos capítulos são apresentados e discutidos os resultados das técnicas de análise multivariada abordadas neste capítulo, aplicadas ao conjunto de dados apresentados no Capítulo 2.

## 4 ESTUDO DAS TELEFACILIDADES

### 4.1 FATURAMENTO TOTAL

Na primeira fase do estudo, todas as variáveis quantitativas descritas na Tabela 2.2 são utilizadas na Análise Fatorial. O objetivo é tentar reduzir a dimensão dos dados, detectando possíveis fatores subjacentes que possam ser úteis para a criação de grupos homogêneos de usuários, mediante a técnica de Análise de Agrupamento. Com base nos grupos, deseja-se descrever o perfil dos usuários quanto às telefácilidades.

Os primeiros resultados da análise sugerem que a variável *Total de Ingresso* (TOT\_ING) pode ser removida, pois ela é uma combinação linear das variáveis VAL\_ME, VAL\_IUA, VAL\_IUR, VAL\_INT, VAL\_FML, VAL\_FMA, VAL\_FMR e VAL\_OUT. Uma implicação matemática da inclusão da variável TOT\_ING é a não singularidade da matriz de correlação, inviabilizando a utilização do modelo fatorial.

Também constatou-se que as variáveis relativas ao faturamento e aos minutos internacionais (VAL\_INT e QTDE\_INT) não contribuem substancialmente às análises, pois somente 108 telefones não estão zerados. Como seria esperado, constatou-se, também, que as quantidades de pulsos e minutos (PULSOS, QTDE\_IUA, QTDE\_IUR, QTDE\_FML, QTDE\_FMA, QTDE\_FMR) estão fortemente correlacionadas com os seus respectivos valores faturados, isto é, a quantidade de pulsos medidos está fortemente correlacionada com o valor cobrado em ligações locais; a quantidade de minutos intra-rede está fortemente correlacionada com o valor faturado em ligações intra-rede, e assim por diante.

Isso sugere que a informação relativa às quantidades de pulsos e minutos está contida nos valores faturados, de tal modo que pode-se remover estas variáveis das análises, mantendo somente as variáveis correspondentes aos tipos de faturamento. Estas remoções também contribuem para um aumento da medida de adequacidade amostral de Kaiser, como é mostrado nos quadros 4.1 e 4.2. Apesar dos valores não serem muito altos, somente valores abaixo de 0,5 são considerados inaceitáveis e requerem ações reparadoras, que pode ser tanto a retirada quanto a inclusão de variáveis ao banco de dados. Veja SAS Institute Inc. (1989, p. 812).

**Quadro 4.1 – Medida de adequacidade amostral de Kaiser antes da remoção das variáveis PULSOS, QTDE\_IUA, QTDE\_IUR, VAL\_INT, QTDE\_INT, QTDE\_FML, QTDE\_FMA e QTDE\_FMR**

Kaiser's Measure of Sampling Adequacy: Over-all MSA = 0.58807357							
VAL_ME	PULSOS	VAL_IUA	QTDE_IUA	VAL_IUR	QTDE_IUR	VAL_INT	QTDE_INT
0.579724	0.580389	0.656222	0.513064	0.593676	0.319275	0.499102	0.499299
VAL_FML	QTDE_FML	VAL_FMA	QTDE_FMA	VAL_FMR	QTDE_FMR	VAL_OUT	
0.622163	0.611480	0.697627	0.696244	0.556247	0.535401	0.879435	

**Quadro 4.2 – Medida de adequacidade amostral de Kaiser após a remoção das variáveis PULSOS, QTDE\_IUA, QTDE\_IUR, VAL\_INT, QTDE\_INT, QTDE\_FML, QTDE\_FMA e QTDE\_FMR**

Kaiser's Measure of Sampling Adequacy: Over-all MSA = 0.64057674						
VAL_ME	VAL_IUA	VAL_IUR	VAL_FML	VAL_FMA	VAL_FMR	VAL_OUT
0.491606	0.713774	0.530032	0.750157	0.744118	0.530620	0.842840

Como estimativas preliminares para as comunalidades foi usado o critério da máxima correlação absoluta da *i*-ésima variável com todas as outras variáveis, o qual pode ser obtido no pacote **SAS** com a inclusão da opção **PRIORS=MAX** no procedimento **FACTOR**. Veja SAS Institute Inc. (1989, p.453).

Na definição do número de fatores retidos na análise, foi utilizado o critério que especifica que o número de fatores é igual ao número de autovalores da matriz de correlação maiores que 1. Desta forma foram mantidos 2 fatores, como mostra a Tabela 4.1.

**Tabela 4.1 - Autovalores, percentual da variância explicada e percentual acumulado da variância total explicada para os 2 fatores selecionados**

FATORES	AUTOVALORES	% DA VARIÂNCIA EXPLICADA PELOS FATORES	% ACUMULADO DA VARIÂNCIA TOTAL EXPLICADA
1	2,6889	0,5624	0,5624
2	1,6490	0,3449	0,9073

Os dois primeiros fatores explicam aproximadamente 91% da variação total das 7 variáveis mantidas na análise. O Quadro 4.3 mostra a composição dos fatores. Para permitir uma interpretação mais clara, foi utilizado o método de rotação ortogonal VARIMAX do pacote **SAS**. Veja SAS Institute Inc. (1989, p.453).

**Quadro 4.3 – Matriz de cargas fatoriais com rotação VARIMAX para o Faturamento Total**

Rotated Factor Pattern		
	Factor1	Factor2
VAL_ME	0.27798	-0.02414
VAL_IUA	0.87840	0.16482
VAL_IUR	0.02866	0.98567
VAL_FML	0.85561	0.09868
VAL_FMA	0.82895	0.26003
VAL_FMR	0.03822	0.98653
VAL_OUT	-0.04422	-0.12527

O **FATOR 1** pode ser interpretado como um indicador do faturamento gerado por ligações para a área de cobertura da CRT, incluindo o tipo de chamada intra-rede, fixo-móvel local, fixo-móvel intra-rede e uma pequena participação positiva do faturamento local (valor medido). Por sua vez, o **FATOR 2** pode ser interpretado como um indicador do faturamento gerado por ligações para



fora da área de cobertura da CRT, pois é formado pelo tipo de chamada inter-rede e fixo-móvel inter-rede. A variável correspondente ao faturamento gerado por outros serviços não ficou bem caracterizada em nenhum dos dois fatores.

A partir destes fatores, foram estimados os escores fatoriais de cada um dos telefones associados aos dois fatores subjacentes identificados na Análise Fatorial, gerando uma nova matriz de dados, chamada de matriz de escores fatoriais, utilizada como entrada na técnica de Análise de Agrupamento. Baseado nestes escores, foram gerados 50 grupos iniciais pelo método não-hierárquico do procedimento **FASTCLUS**, disponível do pacote estatístico **SAS**. Então, mediante o método hierárquico de mínima variância de Ward, constante no procedimento **CLUSTER**, foram estimadas as distâncias dos escores dos telefones à semente destes grupos iniciais e representados graficamente pelo procedimento **TREE**, onde foi estabelecido o ponto de corte para a formação dos grupos finais (veja os programas no Anexo 1).

Mesmo considerando um valor muito baixo do r-quadrado semi-parcial para estabelecer o ponto de corte que gera os grupos na técnica de Análise de Agrupamento (veja Anexo 2.1), um dos grupos é composto por 1.575.496 telefones. Um segundo grupo é composto por 105 telefones, enquanto que ao último grupo foram alocados apenas 6 telefones. Estes resultados não permitem extrair informações relevantes dos grupos, pois apenas destacam alguns telefones com valores altíssimos de faturamento. Um procedimento alternativo é separar o banco de dados em dois segmentos, de acordo com o faturamento, como descrito na próxima seção.

## **4.2 DIVISÃO EM ESTRATOS**

Uma suposição inicial é que as telefácilidades deveriam incrementar de alguma maneira o consumo e, conseqüentemente, o faturamento da empresa.

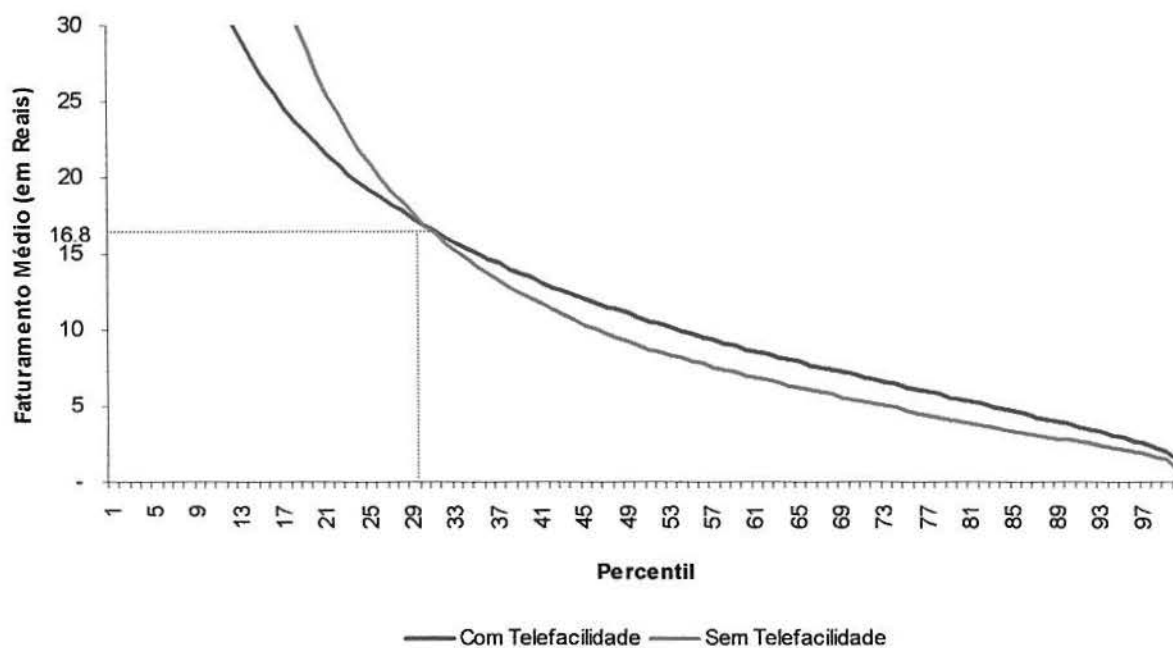
Deste modo, foi adotado o procedimento de dividir o faturamento total em percentis para averiguar se a presença de telefácilidades aumentam o consumo e a partir de que ponto isto seria verdadeiro.

Inicialmente os telefones foram identificados e rotulados quanto à utilização ou não de telefácilidades. Para estes dois grupos, foram calculados os percentis de faturamento e o faturamento médio dentro de cada percentil.

Observou-se que, a partir do trigésimo percentil, o grupo de telefones com telefácilidades apresentou faturamento médio superior ao grupo sem telefácilidades, mesmo considerando-se que no grupo de telefones com telefácilidades apenas 20% deles realmente as utilizam, conforme a pesquisa sobre satisfação de serviços previamente realizada concluiu. Isto significa que até o trigésimo percentil, gerado por telefones com conta média acima de R\$ 16,84, é acumulado 30% do faturamento total da empresa, que neste trabalho considera-se como faturamento alto, e 70% do faturamento total é gerado por telefones com conta média igual ou abaixo de R\$ 16,84, considerado aqui como faturamento baixo.

Também vale ressaltar que estes 30% do faturamento total da empresa são realizados por apenas 5,4% dos telefones, enquanto que os 70% do faturamento restante são gerados por 94,6% dos telefones. A Figura 4.1 mostra claramente a mudança de comportamento no faturamento médio dos telefones com e sem telefácilidades. É conveniente salientar que o gráfico da Figura 4.1

está truncado no eixo vertical em R\$ 30,00, para que seja possível visualizar a região onde ocorre a mudança de comportamento da curva.



**Figura 4.1 – Faturamento médio por percentil dos telefones com e sem telefacilidades**

Desta forma, a população foi dividida em dois estratos, os quais são analisados separadamente. O primeiro estrato é formado por 5,4% dos telefones que possuem faturamento mais elevado e, o segundo, por 94,6% dos telefones com faturamento abaixo de R\$ 16,84. A seguir são apresentados os resultados da análise estatística de cada estrato.

#### 4.2.1 ESTRATO I: FATURAMENTO ALTO

A análise dos dados do estrato com faturamento alto foi realizada mediante o modelo fatorial, onde as variáveis são as mesmas da primeira etapa do trabalho. O Quadro 4.4 mostra as medidas de adequacidade amostral de Kaiser, destacando-se que a medida geral é aproximadamente igual à obtida anteriormente.

**Quadro 4.4 – Medida de adequacidade amostral de Kaiser para o Estrato I**

Kaiser's Measure of Sampling Adequacy: Over-all MSA = 0.64005145						
VAL_ME	VAL_IUA	VAL_IUR	VAL_FML	VAL_FMA	VAL_FMR	VAL_OUT
0.463563	0.711238	0.529397	0.754093	0.741977	0.529963	0.874682

A Tabela 4.2 apresenta os autovalores da matriz de correlação reduzida, onde as comunalidades iniciais foram estimadas através da máxima correlação absoluta de cada variável com todas as outras. Veja SAS Institute Inc. (1989, p.453). Os dois primeiros fatores explicam aproximadamente 91% da variação total.

**Tabela 4.2 - Autovalores, percentual da variância explicada e percentual acumulado da variância total explicada para os 2 fatores selecionados no Estrato I**

FATORES	AUTOVALORES	% DA VARIÂNCIA EXPLICADA PELOS FATORES	% ACUMULADO DA VARIÂNCIA TOTAL EXPLICADA
1	2,6994	0,5638	0,5638
2	1,6517	0,3450	0,9088

O Quadro 4.5 mostra as cargas dos fatores gerados pela rotação VARIMAX. O **FATOR 1** pode ser interpretado como o faturamento gerado por ligações do tipo local e intra-rede (fixo e móvel), embora o faturamento local (valor medido) tenha uma contribuição pequena. O **FATOR 2**, por sua vez, pode ser

interpretado como o faturamento gerado por ligações para outras redes, tanto para telefones fixos quanto móveis.

**Quadro 4.5 – Matriz de cargas fatoriais com rotação VARIMAX para o Estrato I**

Rotated Factor Pattern		
	FACTOR1	FACTOR2
VAL_ME	0.25211	-0.02292
VAL_IUA	0.88147	0.17175
VAL_IUR	0.01985	0.98732
VAL_FML	0.85392	0.10506
VAL_FMA	0.83475	0.26691
VAL_FMR	0.03008	0.98819
VAL_OUT	-0.05966	-0.12659

Os escores fatoriais foram então utilizados para gerar os 50 grupos iniciais de clientes através método não-hierárquico, disponível no procedimento **FASTCLUS** e, posteriormente, os grupos finais através do método de mínima variância de Ward, disponível no procedimento **CLUSTER**, bem como a representação gráfica destas distâncias através do procedimento **TREE** para o estabelecimento do ponto de parada no agrupamento. Em linhas gerais, observou-se aproximadamente o mesmo comportamento encontrado no banco de dados completo, com a identificação de três grupos (veja o Anexo 2.2). É conveniente observar que um dos grupos foi formado pela maioria dos casos, 87.915 telefones, enquanto outros dois grupos contiveram uma parcela mínima de casos, com 54 e 6 telefones.

A interpretação destes grupos parece pouco informativa, pois o **GRUPO 2**, que contém quase a totalidade dos casos, apresenta escores fatoriais em torno de zero para os dois fatores, como mostra a Figura 4.2. Seria pouco frutífero extrair os valores extremos do **GRUPO 1** e do **GRUPO 3** para refazer todo o

processo somente com o **GRUPO 2**, pois possivelmente a mesma situação se repetiria neste grupo restante devido à presença de valores extremos.

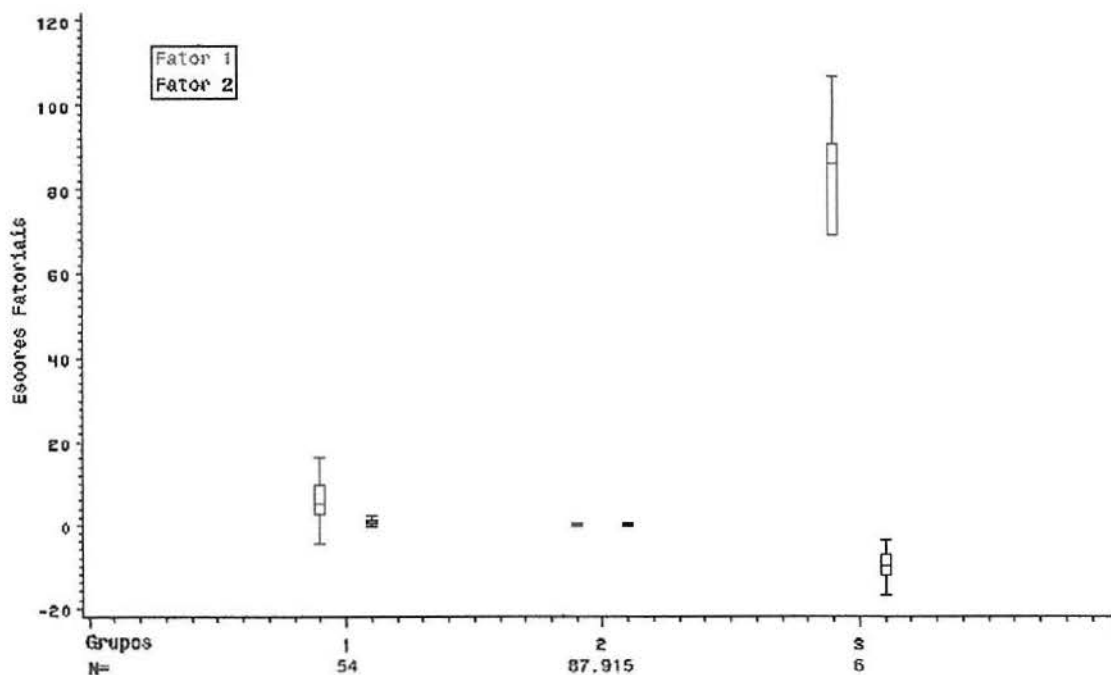


Figura 4.2 – Escores fatoriais dos dois fatores para os três grupos

#### 4.2.2 ESTRATO II: FATURAMENTO BAIXO

A segunda parcela da população, formada por telefones com faturamento abaixo de R\$ 16,84, apresentou uma medida de adequabilidade amostral inferior à encontrada para o Estrato I, conforme mostra o Quadro 4.6. Contudo, valores acima de 0,5 ainda são aceitáveis. Veja SAS Institute Inc. (1989, p. 812).

Quadro 4.6 – Medida de adequabilidade amostral de Kaiser para o Estrato II

Kaiser's Measure of Sampling Adequacy: Over-all MSA = 0.53357074						
VAL_ME	VAL_IUA	VAL_IUR	VAL_FML	VAL_FMA	VAL_FMR	VAL_OUT
0.522307	0.554307	0.535348	0.528497	0.534795	0.523078	0.667757

Para o Estrato II foram selecionados três fatores, diferentemente do que foi realizado no Faturamento Total e no Estrato I. O critério para definir o número de fatores selecionados foi baseado na interpretação subjetiva dos fatores, já que os autovalores da matriz de correlação, apresentados na Tabela 4.3, são todos inferiores a um. Isto pode ser devido à baixa correlação entre as variáveis originais, mas pode ser compensado pelo grande número de observações que fornece mais consistência aos resultados. Nota-se que 97,3% da variância total das variáveis originais é explicada por estes três fatores comuns.

**Tabela 4.3 - Autovalores, percentual da variância explicada e percentual acumulado da variância total explicada para os 3 fatores selecionados no Estrato II**

FATORES	AUTOVALORES	% DA VARIÂNCIA EXPLICADA PELOS FATORES	% ACUMULADO DA VARIÂNCIA TOTAL EXPLICADA
1	0,8308	0,5214	0,5214
2	0,4379	0,2749	0,7963
3	0,2815	0,1767	0,9730

As cargas fatoriais que dão interpretação aos fatores comuns são mostradas no Quadro 4.7, onde pode-se notar que o **FATOR 1** parece estar relacionado ao faturamento devido a ligações locais (fixo e móvel); o **FATOR 2** pode estar relacionado como faturamento devido às ligações do tipo intra-rede (fixo e móvel) e o **FATOR 3** ao faturamento devido a ligações inter-rede (fixo e móvel). Por esta composição, os telefones do Estrato II com escores fatoriais altos em qualquer um dos três fatores podem ser caracterizados por um interesse maior tanto em ligações do tipo fixo-fixo quanto fixo-móvel daquele tipo. Nota-se também uma pequena participação do faturamento devido a outros serviços (VAL\_OUT) no **FATOR 1**.

**Quadro 4.7 – Matriz de cargas fatoriais com rotação VARIMAX para o Estrato II**

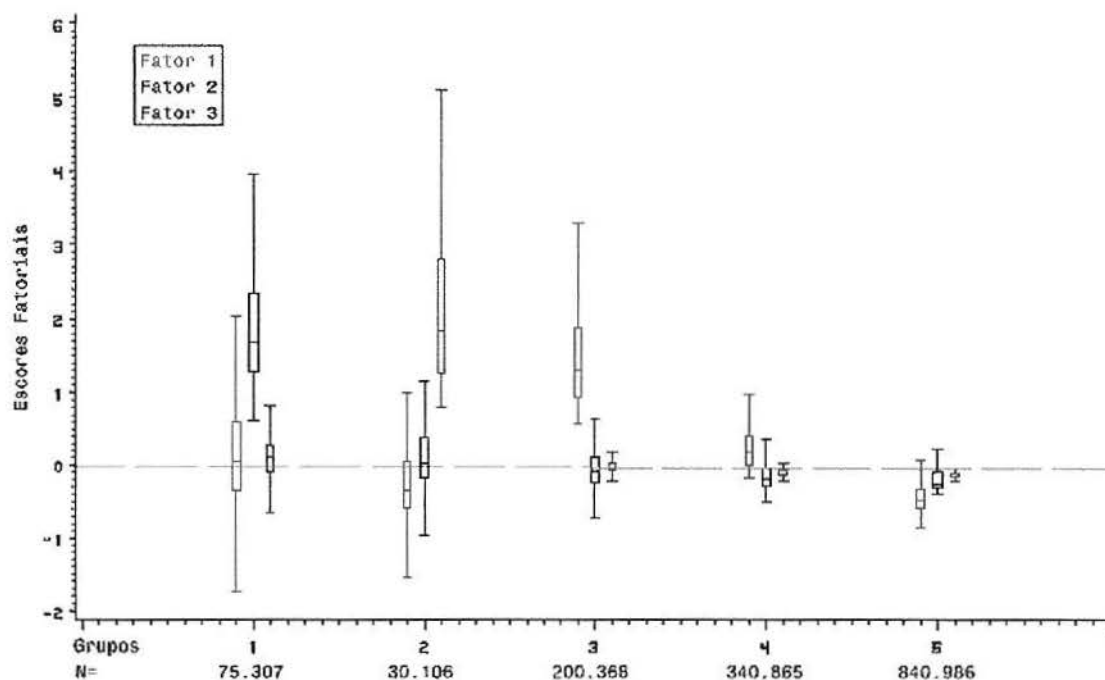
Rotated Factor Pattern			
	FACTOR1	FACTOR2	FACTOR3
VAL_ME	0.61069	0.02389	0.04695
VAL_IUA	0.08529	0.45787	0.10673
VAL_IUR	0.00763	0.07649	0.40020
VAL_FML	0.59724	0.14525	0.05814
VAL_FMA	-0.00128	0.46888	0.00823
VAL_FMR	-0.01206	0.01457	0.39946
VAL_OUT	0.13564	-0.00556	-0.02572

Na Análise de Agrupamento realizada após a extração e interpretação dos fatores, novamente utilizou-se o método de agrupamento não-hierárquico do procedimento **FASTCLUS** para gerar 50 grupos iniciais e, posteriormente, o método de agrupamento hierárquico do procedimento **CLUSTER**, que então foi representada graficamente pelo procedimento **TREE**, onde sugeriu-se a formação de cinco grupos de clientes (Veja o Anexo 2.3).

Estes grupos, com características homogêneas dentro dos grupos e heterogêneas entre os grupos, apresentam frequências maiores do que as observadas no Faturamento Total e no Estrato I.

A Figura 4.3 ilustra o comportamento dos escores fatoriais dos três fatores associados aos telefones dos cinco grupos.





**Figura 4.3 – Escores fatoriais dos três fatores para os cinco grupos**

Por esta figura, pode-se relacionar o **FATOR 1** basicamente ao **GRUPO 3** pois a totalidade dos telefones neste grupo tem escores fatoriais positivos e altos para este fator. O **GRUPO 4** apresenta escores positivos no **FATOR 1**, porém apenas um pouco acima da média, em aproximadamente 75% dos telefones. A seguir vemos o **GRUPO 1** com escores fatoriais positivos e negativos aproximadamente na mesma proporção no primeiro fator e o **GRUPO 2** e o **GRUPO 5** com uma predominância de escores fatoriais negativos.

O **FATOR 2** está bem caracterizado pelo **GRUPO 1**, com a totalidade de telefones apresentando escores fatoriais positivos e altos neste fator. O **GRUPO 2** e o **GRUPO 3** apresentam escores em torno de zero aproximadamente na mesma proporção e os demais grupos apresentam escores negativos para a maioria dos telefones.

Quanto ao **FATOR 3**, percebe-se que ele é fortemente associado ao **GRUPO 2** com escores positivos e altos, enquanto que para os demais grupos os escores do terceiro fator estão aproximadamente centrados em torno de zero.

Para auxiliar a interpretação dos grupos formados, é útil verificar a composição de cada grupo com relação às variáveis de interesse descritas no Capítulo 2.

Primeiramente, verificou-se o percentual de telefones com e sem telefacilidades em cada grupo, e conforme mostra a Tabela 4.4, o maior percentual de telefones com telefacilidades em relação ao grupo encontra-se no **GRUPO 3**, caracterizado por escores fatoriais altos no **FATOR 1**. Também, em relação ao total de telefacilidades, conforme observa-se na Tabela 4.5, o **GRUPO 3** contém 22,25% dos serviços instalados.

**Tabela 4.4 - Percentual de telefones com e sem telefacilidades por grupo**

<i>Telefacilidade</i>	<i>Grupo</i>				
	1	2	3	4	5
Sem	84,21	87,08	78,99	84,57	90,64
Com	15,79	12,92	21,01	15,43	9,36
Total	100	100	100	100	100

**Tabela 4.5 - Percentual de telefones com telefacilidades por grupo**

<i>Grupo</i>	1	2	3	4	5	Total
<i>% Telefacilidade</i>	6,28	2,06	22,25	27,80	41,61	100

O valor 22,25% encontrado neste grupo é particularmente relevante, pois está bem próximo à estimativa inicial obtida pela pesquisa de satisfação dos clientes em relação aos serviços prestados, onde constatou-se que apenas 20% dos clientes que possuem telefacilidades instaladas em seus telefones realmente as utilizam.

Após o levantamento destes dados, foram realizadas discussões com especialistas do Departamento de Análise de Mercado da CRT, e constatou-se que é razoável relacionar as telefacilidades ao **GRUPO 3**, embora não seja possível afirmar que estes sejam exatamente os telefones que utilizam os serviços, que é caracterizado por altos escores no **FATOR 1**, que representa o faturamento gerado por ligações locais. Deste modo, seria conveniente associar estes telefones com telefacilidades aos 20% citados inicialmente, pois o faturamento local é justamente onde espera-se que as telefacilidades causem um maior impacto. Isto também é reforçado pelo próprio método de criação do Estrato II, que leva em conta um nível de corte no faturamento total onde os telefones com telefacilidades apresentam uma conta média superior aos telefones sem estes serviços. Para confirmar esta hipótese, de que esta parcela de telefones seja formada pela maioria daqueles que utilizam os serviços, seria necessário realizar uma pesquisa por amostragem, mas está fora do alcance do trabalho, neste momento. Deste modo, partiu-se para uma breve descrição dos telefones com telefacilidades no **GRUPO 3**. Um outro possível grupo ao qual poderiam ser relacionadas as telefacilidades é o **GRUPO 4**, pois este apresenta escores fatoriais um pouco acima da média do Estrato II para o **FATOR 1** em aproximadamente 75% dos telefones.

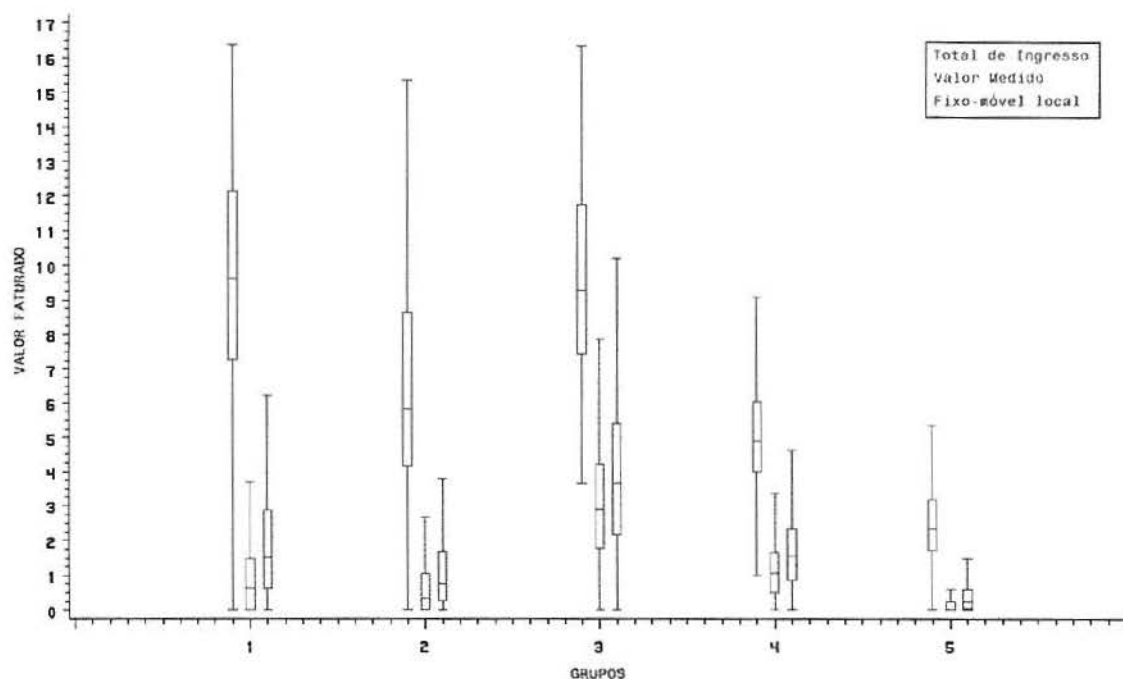
Todavia, os demais grupos não serão analisados detalhadamente, uma vez que o alvo principal deste trabalho são os telefones que utilizam telefacilidades relacionados ao **GRUPO 3**. Cabe ressaltar que o **GRUPO 1** caracteriza-se por telefones com alto faturamento do tipo intra-rede (**FATOR 2**) e o **GRUPO 2** por telefones com alto faturamento do tipo inter-rede (**FATOR3**). Contudo, este tipo de faturamento elevado dificilmente pode ser atribuído às

telefacilidades. As análises destes grupos poderiam ser realizadas em pesquisas futuras, utilizando-se estes grupos reduzidos para tentar identificar quais os fatores que levaram tais telefones a um faturamento maior deste tipo. O **GRUPO 4** é basicamente composto por telefones que apresentam escores fatoriais positivos e baixos para o **FATOR 1**. Aproximadamente 57% dos telefones do Estrato II estão concentrados no **GRUPO 5**, que apresenta escores fatoriais negativos para os três fatores, indicando um faturamento abaixo da média geral para os três tipos de faturamento: local, intra-rede e inter-rede. De fato, este também é um comportamento esperado, pois o faturamento de uma grande parcela dos telefones apresenta valores muito próximos da tarifa básica mensal.

Um refinamento da análise das telefacilidades pode ser realizado retornando às variáveis originais, em particular as variáveis total de ingresso (TOT\_ING), valor medido (VAL\_ME) e fixo-móvel local (VAL\_FML). A Tabela 4.6 contém as médias e desvios padrões das variáveis total de ingresso, valor medido e fixo-móvel local para os cinco grupos. O comportamento simultâneo destas variáveis pode ser visualizado na Figura 4.4.

**Tabela 4.6 – Média e desvio padrão do total de ingresso, valor medido e fixo-móvel local por grupo**

Nome da Variável	Grupo 1		Grupo 2		Grupo 3		Grupo 4		Grupo 5	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
<i>Total de ingresso</i>	9,80	3,15	6,74	3,37	9,78	2,78	5,27	1,77	2,62	1,47
<i>Valor medido</i>	0,95	1,08	0,74	1,04	3,17	1,99	1,12	0,81	0,17	0,29
<i>Fixo-móvel local</i>	1,96	1,70	1,22	1,37	3,96	2,41	1,69	1,09	0,39	0,47



**Figura 4.4 – Distribuição das variáveis total de ingresso, valor medido e fixo-móvel local, por grupo**

A Tabela 4.7 mostra os valores médios faturados e desvios padrões das variáveis total de ingresso (TOT\_ING), valor medido (VAL\_ME) e fixo-móvel local (VAL\_FML) dos telefones com e sem telefacilidades nos cinco diferentes grupos.

**Tabela 4.7 – Média e desvio padrão do total de ingresso, valor medido e fixo-móvel local dos telefones com e sem telefacilidades por grupo**

Nome da Variável	Grupo 1		Grupo 2		Grupo 3		Grupo 4		Grupo 5	
	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio	Média	Desvio
<b>C</b> <i>Total de ingresso</i>	10,43	3,03	7,83	3,51	10,16	2,76	5,66	1,87	3,12	1,46
<b>O</b> <i>Valor medido</i>	1,17	1,14	0,98	1,19	3,17	1,92	1,11	0,79	0,20	0,31
<b>M</b> <i>Fixo-móvel local</i>	2,13	1,68	1,50	1,51	4,06	2,32	1,73	1,05	0,47	0,48
<b>S</b> <i>Total de ingresso</i>	9,69	3,16	6,58	3,32	9,68	2,78	5,20	1,74	2,57	1,46
<b>E</b> <i>Valor medido</i>	0,91	1,06	0,70	1,00	3,16	2,01	1,12	0,81	0,17	0,29
<b>M</b> <i>Fixo-móvel local</i>	1,93	1,71	1,17	1,34	3,93	2,43	1,68	1,10	0,39	0,46

A comparação dos valores da Tabela 4.7 sugere que em geral os telefones com telefacilidades apresentam médias de faturamento levemente superiores aos telefones sem os serviços. Contudo, estas diferenças poderiam ser atribuídas a

outros fatores que não o uso das telefacilidades. Isto sugere que, apesar da indicação de que o faturamento gerado por telefones com telefacilidades é superior àqueles que não as possuem, esta superioridade é relativamente pequena e não pode ser especificada com precisão. É importante salientar, também, que as menores diferenças entre telefones com e sem os serviços são encontradas justamente no **GRUPO 3** e no **GRUPO 4**, mas isto é devido ao processo de formação dos grupos, que agrupou aqueles telefones com características similares de faturamento, independentemente de possuírem ou não as telefacilidades.

Apesar disto, considerando-se que os clientes do **GRUPO 3** possivelmente são os usuários dos serviços adicionais incorporados ao telefone, as figuras e as tabelas do Anexo 3 mostram o faturamento médio destes telefones por pacote, servindo como base para uma possível estimativa do faturamento médio real dos telefones com telefacilidades. As variáveis de interesse são o total de ingresso e aquelas que compõem o **FATOR 1**, valor medido e fixo-móvel local. As figuras mostram o faturamento médio dos telefones aberto por código de área e categoria, os pacotes estão em ordem decrescente segundo o total de ingresso.

Estas análises gráficas são usualmente utilizadas em informes gerenciais, servindo de suporte para decisões no setor de marketing da empresa. Em geral, elas são úteis em vários setores para se ter idéia do comportamento dos clientes em relação a algum fator de interesse, neste caso as telefacilidades, e são feitas mensalmente para que se torne possível acompanhar a evolução dos resultados obtidos.

## 5 CONSIDERAÇÕES FINAIS

O estudo aprofundado das telefacilidades, através de técnicas da Análise Multivariada apresentou resultados bastante interessantes do ponto de vista prático, pois permitem um melhor entendimento do fenômeno. Contudo, estes resultados devem ser considerados com cautela, já que referem-se somente ao mês de maio de 2000, no qual, de acordo com uma pesquisa realizada sobre satisfação dos clientes em relação aos serviços prestados, estima-se que apenas 20% dos clientes que possuem telefones com telefacilidades as utilizam.

A abordagem sobre o faturamento geral proporcionou novas informações sobre o relacionamento entre os tipos de faturamento e as telefacilidades, tais como a redução do número de variáveis, podendo-se concentrar os esforços em um número menor de fatores, que não todas as variáveis quantitativas originais.

Através da análise do faturamento em percentis, a divisão em estratos parece necessária para isolar os casos de faturamento extremamente elevados, já que a contribuição das telefacilidades parece ser relativamente baixa nestes casos.

A Análise Multivariada no Estrato I (faturamento alto) não produziu grupos interpretáveis, sendo que este estrato contém os valores mais extremos do faturamento, repetindo-se a situação do faturamento geral, onde as telefacilidades não parecem influenciar na identificação de grupos similares.

No Estrato II, formado por telefones com faturamento mais baixo, onde os telefones com telefacilidades apresentaram faturamento médio superior aos demais, as análises mostraram resultados mais eficientes. Foram mantidos três fatores comuns, explicando aproximadamente 97% da variação total das variáveis

originais. Estes três fatores foram interpretados respectivamente como: o faturamento gerado por ligações locais (fixo e móvel), o faturamento gerado por ligações intra-rede (fixo e móvel), e o faturamento correspondente a ligações do tipo inter-rede (fixo e móvel).

A Análise de Agrupamento gerou 5 grupos neste estrato, onde o **GRUPO 1** é caracterizado pelo faturamento elevado devido ao tipo de ligações intra-rede, o **GRUPO 2** pelo faturamento inter-rede, o **GRUPO 3** pelo faturamento local, o **GRUPO 4** também pelo faturamento do tipo local, contudo inferior ao terceiro grupo, e o **GRUPO 5** pelo faturamento abaixo da média para os três tipos de ligações, contendo aproximadamente 57% dos telefones deste estrato.

O terceiro grupo possivelmente está associado às telefacilidades, pois espera-se um maior impacto dos serviços adicionais no faturamento do tipo local, e também porque identificou-se um percentual de telefones com telefacilidades próximo de 20%. Isto reforça a suspeita de que estes telefones são aqueles que utilizam os serviços, já que pela própria formação do Estrato II eram esperados resultados favoráveis neste sentido. Entretanto, este resultado só pode ser efetivamente comprovado mediante uma pesquisa futura com estes clientes, que venha a confirmar que estes são os que realmente utilizam os serviços incorporados aos seus telefones.

A análise descritiva dos telefones com telefacilidades, tratadas como *pacotes* de telefacilidades, apresentou um faturamento superior àqueles sem os serviços, mesmo entre telefones do terceiro grupo, que tem um comportamento semelhante em relação ao tipo de faturamento realizado.

Os gráficos apresentados nos anexos, mostram uma estimativa do possível faturamento médio real das variáveis total de ingresso, valor medido e fixo-móvel



local dos telefones com pacotes de telefacilidades, segundo a sua distribuição regional e por categoria, afim de que possam ser utilizados como suporte para futuras análises.

Além dos resultados da análise estatística, é importante mencionar também as contribuições do trabalho aos aspectos computacionais. Em particular, destaca-se a necessidade de padronização da rotina de extração das informações das bases de dados, para evitar vícios na formação das variáveis.

Ao contrário de esgotar o assunto, o trabalho apresenta resultados que podem ser extremamente valiosos do ponto de vista gerencial. Por exemplo, como continuidade, sugere-se que as técnicas estatísticas tradicionais utilizadas sejam ampliadas através de novas tecnologias, como o *data mining*, e que em conjunto com o Departamento de Análise de Mercado, seja realizado um refinamento das composições dos grupos, permitindo o uso destas informações no plano estratégico da empresa.

## **6 ANEXOS**

## ANEXO 1 PROGRAMAS

### 1.1 Procedimento do software SAS para a Análise Fatorial

```
libname crt 'c:\estatistica\monografia\dados';  
options nonumber nodate;  
  
proc factor  
  data=crt.fat0500 simple method=principal out=crt.scores priors=max  
  corr scree score msa eigenvectors preplot rotate=varimax plot;  
  var val_me val_iua val_iur val_fml val_fma val_fmr val_out;  
run;
```

Este programa, primeiramente define a *library crt* que aponta o diretório onde se encontra o banco de dados, já em formato SAS, chamado fat0500.sd2 e algumas opções de sistema.

Então, a *procedure factor* é executada sobre este banco de dados com as opções desejadas. Após a análise dos resultados obtidos, o procedimento é repetido, incluindo a opção *nfactor=2*, que define o número de fatores a serem extraídos, para o caso do Faturamento Total. Para a aplicação nos estratos I e II são gerados novos bancos de dados a partir do original e o procedimento é repetido com a opção *nfactor=2* para o Estrato I e *nfactor=3* para o Estrato II.

## 1.2 Procedimento do software SAS para a Análise de Agrupamento

A seguir, é apresentado o procedimento **FASTCLUS** para a criação dos 50 grupos iniciais com o método não hierárquico.

```
proc fastclus
  data=crt.scores drift mean=crt.cmeans
  out=crt.cluster maxclusters=50 maxiter=20;
  id ddd_tel;
  var factor1 factor2 factor3;
run;
```

O procedimento **CLUSTER** a seguir executa a criação dos grupos pelo método hierárquico de mínima variância de Ward.

```
proc cluster
  data=crt.cmeans outtree=crt.tree method=ward;
  var factor1 factor2 factor3;
run;
```

O procedimento **TREE** cria o diagrama de árvore no qual é estipulada a linha de corte para a formação dos grupos finais.

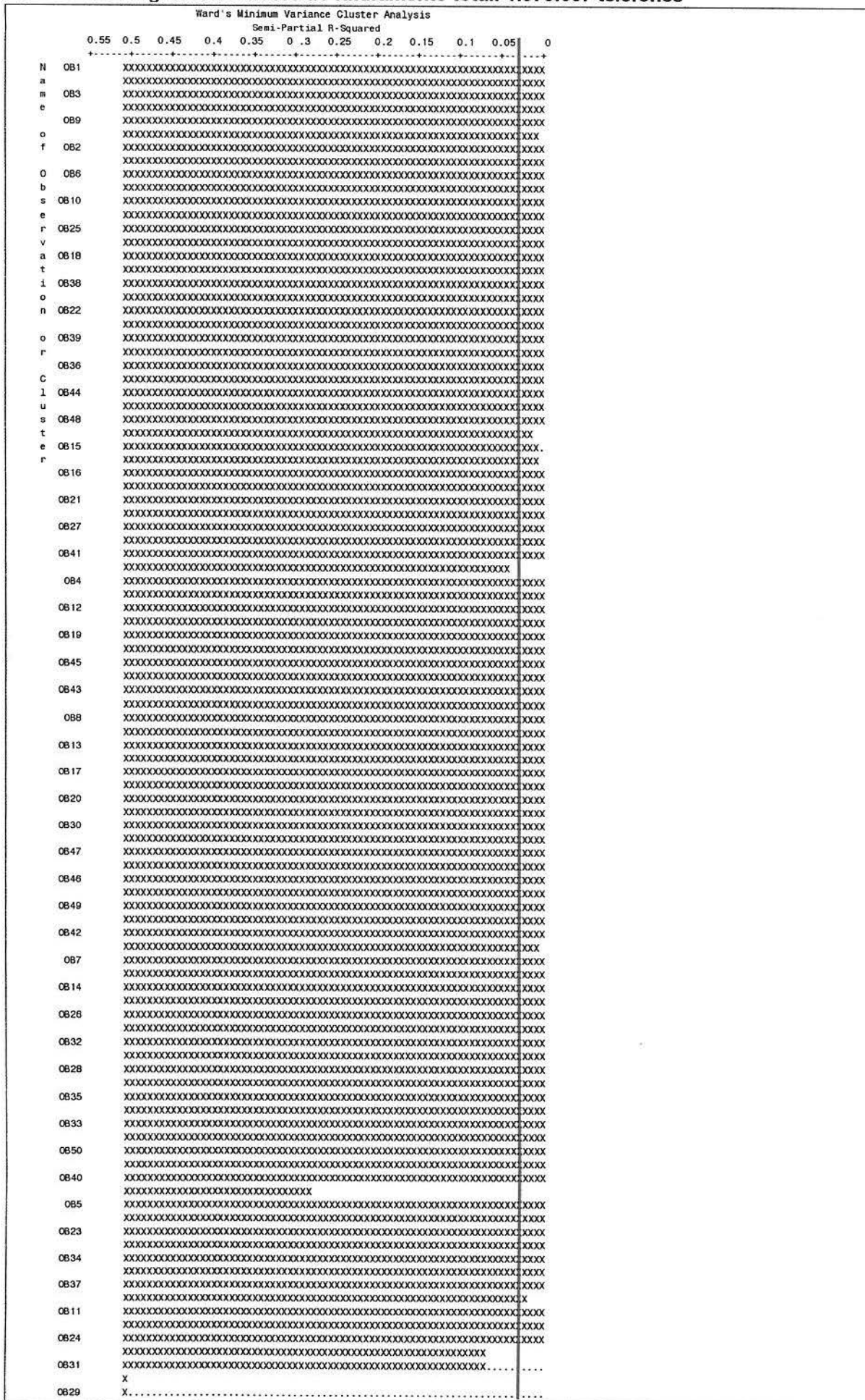
```
proc tree
  data=crt.tree out=crt.tree nclusters=5 horizontal spaces=1;
run;
```

O *data step* a seguir cria no arquivo cluster.sd2 os grupos delimitados no diagrama de árvore criado anteriormente.

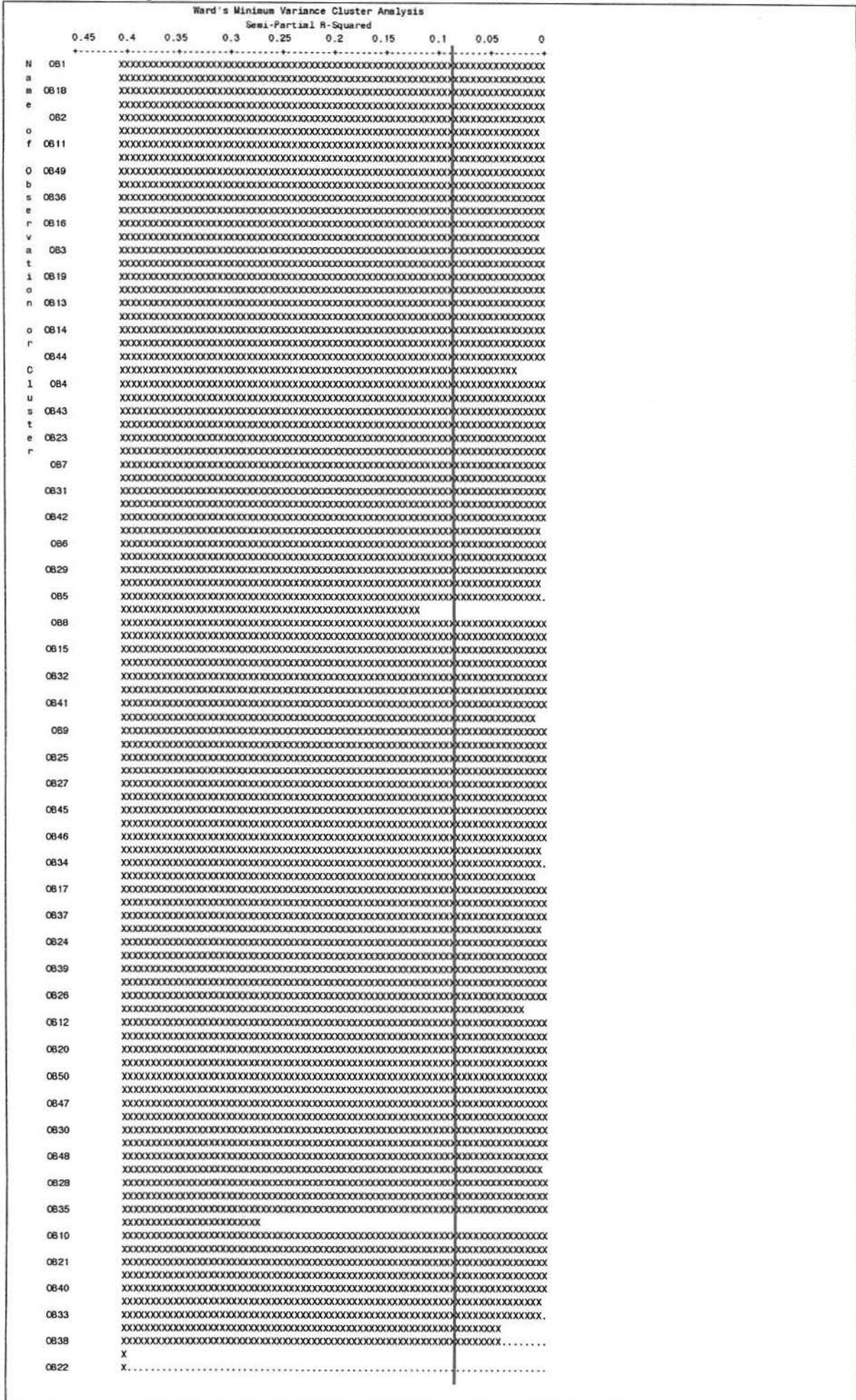
```
data crt.cluster;
  set crt.cluster;
  length grupo 3.;
  if cluster in (1,31,43,28,2,33,16,29,19,32,44) then grupo = 1;
  else if cluster in (5,17,37,34) then grupo = 3;
  else if cluster in (6,30,45) then grupo = 4;
  else if cluster in (18,20,26,36,39) then grupo = 5;
  else grupo = 2;
run;
```

# ANEXO 2 DIAGRAMAS DE ÁRVORE

## Quadro 2.1 Diagrama de árvore do faturamento total: 1.575.607 telefones



Quadro 2.2 Diagrama de árvore do Estrato I: 87.975 telefones



Quadro 2.3 Diagrama de árvore de Estrato II: 1.487.632 telefones



## ANEXO 3 GRÁFICOS E TABELAS

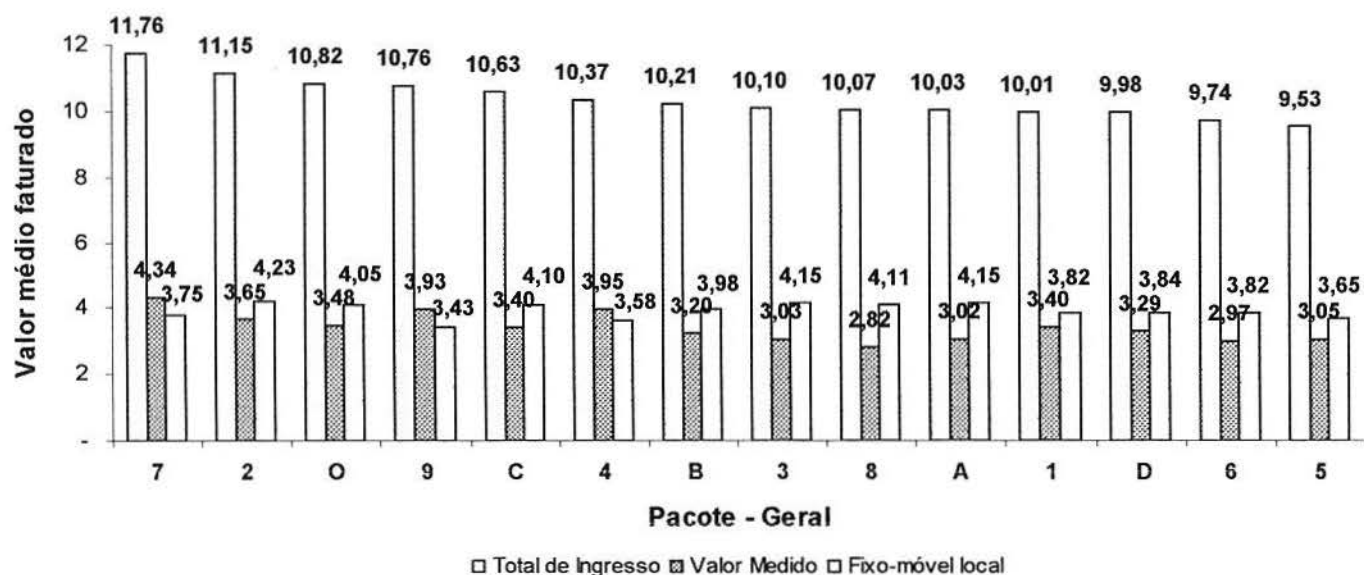


Figura 3.1 Faturamento médio geral do total de ingresso, valor medido e fixo-móvel local dos pacotes de telefácilidades no GRUPO 3

Tabela 3.1 Faturamento médio geral do total de ingresso, valor medido, fixo-móvel local e freqüências dos pacotes de telefácilidades no GRUPO 3

Pacote	Freqüência	Total de ingresso	Valor medido	Fixo-móvel local
7	22	11,76	4,34	3,75
2	1.307	11,15	3,65	4,23
O	2.443	10,82	3,48	4,05
9	1.466	10,76	3,93	3,43
C	840	10,63	3,40	4,10
4	61	10,37	3,95	3,58
B	2.264	10,21	3,20	3,98
3	3.814	10,10	3,03	4,15
8	576	10,07	2,82	4,11
A	23.030	10,03	3,02	4,15
1	5.697	10,01	3,40	3,82
D	451	9,98	3,29	3,84
6	96	9,74	2,97	3,82
5	31	9,53	3,05	3,65



## PACOTE POR CÓDIGO DE ÁREA

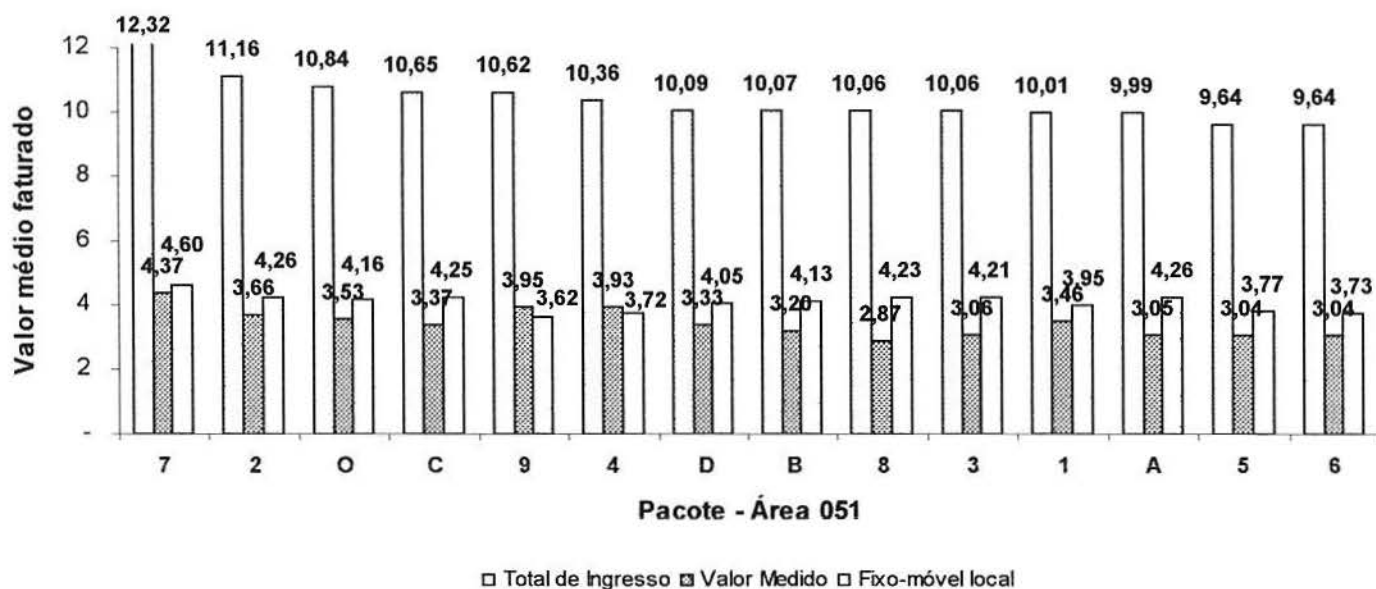


Figura 3.2 Faturamento médio da área 051 do total de ingresso, valor medido e fixo-móvel local dos pacotes de telefácilidades no GRUPO 3

Tabela 3.2 Faturamento médio da área 051 do total de ingresso, valor medido, fixo-móvel local e freqüências dos pacotes de telefácilidades no GRUPO 3

Área 051				
Pacote	Freqüência	Total de ingresso	Valor medido	Fixo-móvel local
7	13	12,32	4,37	4,60
2	1.214	11,16	3,66	4,26
O	2.024	10,84	3,53	4,16
C	718	10,65	3,37	4,25
9	897	10,62	3,95	3,62
4	53	10,36	3,93	3,72
D	314	10,09	3,33	4,05
B	1.622	10,07	3,20	4,13
8	402	10,06	2,87	4,23
3	3.033	10,06	3,06	4,21
1	4.252	10,01	3,46	3,95
A	17.491	9,99	3,05	4,26
5	19	9,64	3,04	3,77
6	70	9,64	3,04	3,73

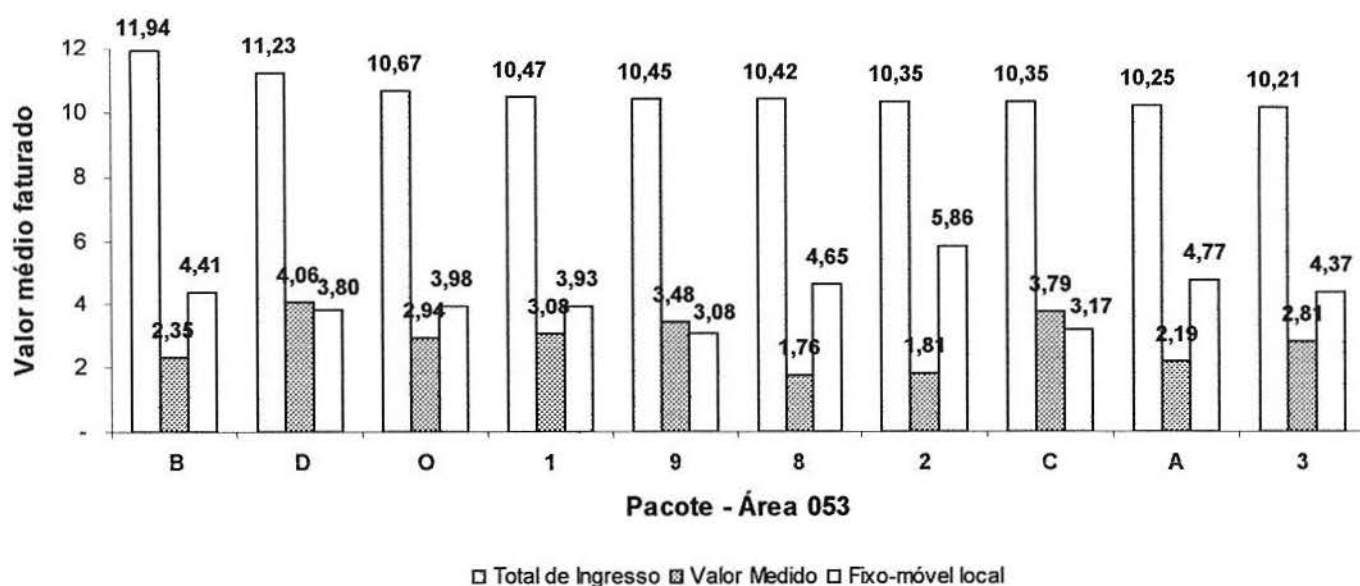
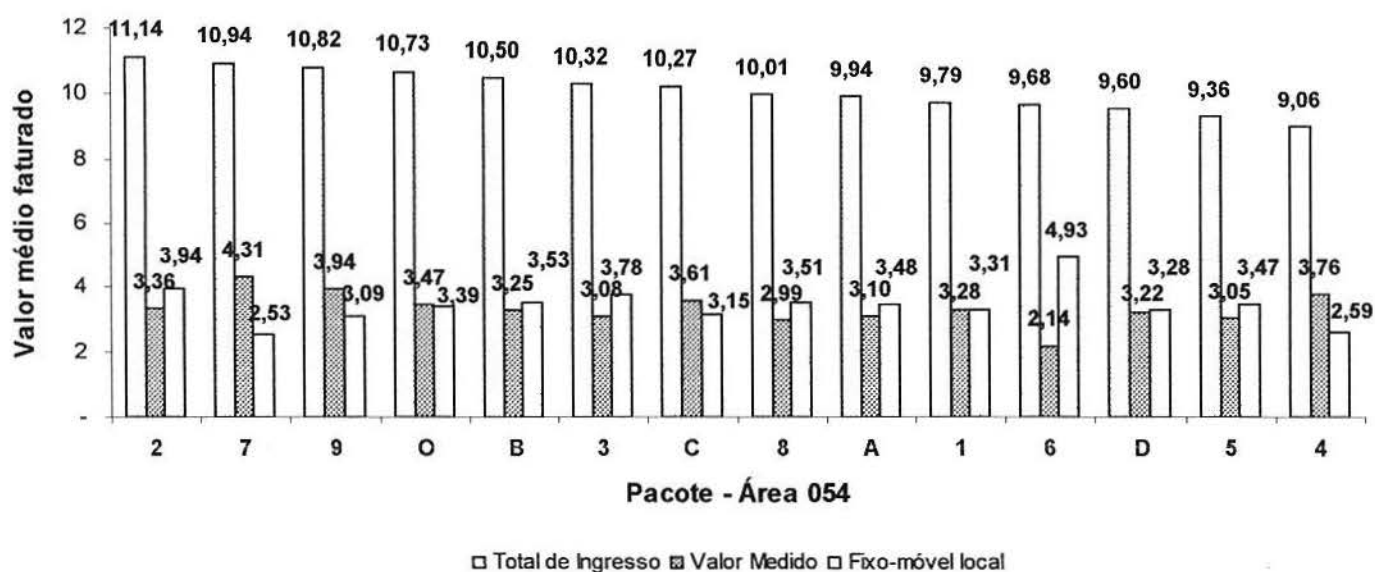


Figura 3.3 Faturamento médio da área 053 do total de ingresso, valor medido e fixo-móvel local dos pacotes de telefácilidades no GRUPO 3

Tabela 3.3 Faturamento médio da área 053 do total de ingresso, valor medido, fixo-móvel local e freqüências dos pacotes de telefácilidades no GRUPO 3

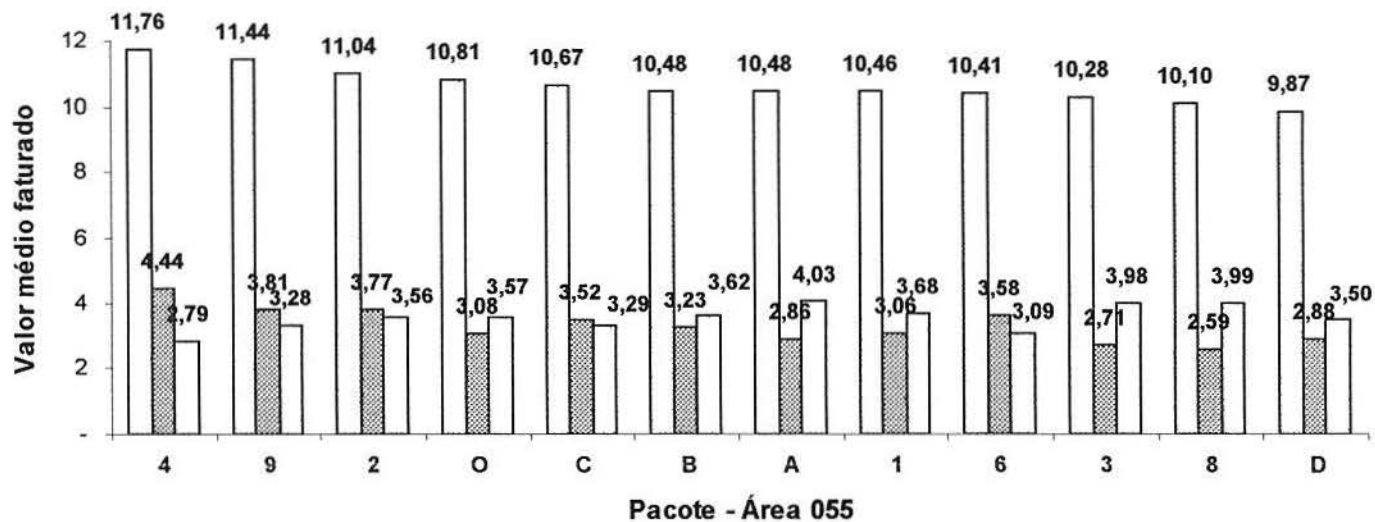
Área 053					
Pacote	Freqüência	Total de ingresso	Valor medido	Fixo-móvel local	
B	29	11,94	2,35	4,41	
D	7	11,23	4,06	3,80	
O	44	10,67	2,94	3,98	
1	70	10,47	3,08	3,93	
9	10	10,45	3,48	3,08	
8	15	10,42	1,76	4,65	
2	5	10,35	1,81	5,86	
C	17	10,35	3,79	3,17	
A	538	10,25	2,19	4,77	
3	38	10,21	2,81	4,37	



**Figura 3.4 Faturamento médio da área 054 do total de ingresso, valor medido e fixo-móvel local dos pacotes de telefácilidades no GRUPO 3**

**Tabela 3.4 Faturamento médio da área 054 do total de ingresso, valor medido, fixo-móvel local e freqüências dos pacotes de telefácilidades no GRUPO 3**

Área 054				
Pacote	Freqüência	Total de Ingresso	Valor Medido	Fixo-móvel local
2	48	11,14	3,36	3,94
7	9	10,94	4,31	2,53
9	411	10,82	3,94	3,09
O	211	10,73	3,47	3,39
B	431	10,50	3,25	3,53
3	360	10,32	3,08	3,78
C	43	10,27	3,61	3,15
8	84	10,01	2,99	3,51
A	2.906	9,94	3,10	3,48
1	988	9,79	3,28	3,31
6	14	9,68	2,14	4,93
D	108	9,60	3,22	3,28
5	12	9,36	3,05	3,47
4	4	9,06	3,76	2,59



□ Total de Ingresso ■ Valor Medido □ Fixo-móvel local

Figura 3.5 Faturamento médio da área 055 do total de ingresso, valor medido e fixo-móvel local dos pacotes de telefácilidades no GRUPO 3

Tabela 3.5 Faturamento médio da área 055 do total de ingresso, valor medido, fixo-móvel local e freqüências dos pacotes de telefácilidades no GRUPO 3

Área 055					
Pacote	Freqüência	Total de Ingresso	Valor Medido	Fixo-móvel local	
4	4	11,76	4,44	2,79	
9	148	11,44	3,81	3,28	
2	40	11,04	3,77	3,56	
O	164	10,81	3,08	3,57	
C	62	10,67	3,52	3,29	
B	182	10,48	3,23	3,62	
A	2.095	10,48	2,86	4,03	
1	387	10,46	3,06	3,68	
6	12	10,41	3,58	3,09	
3	383	10,28	2,71	3,98	
8	75	10,10	2,59	3,99	
D	22	9,87	2,88	3,50	

### PACOTE POR CATEGORIA

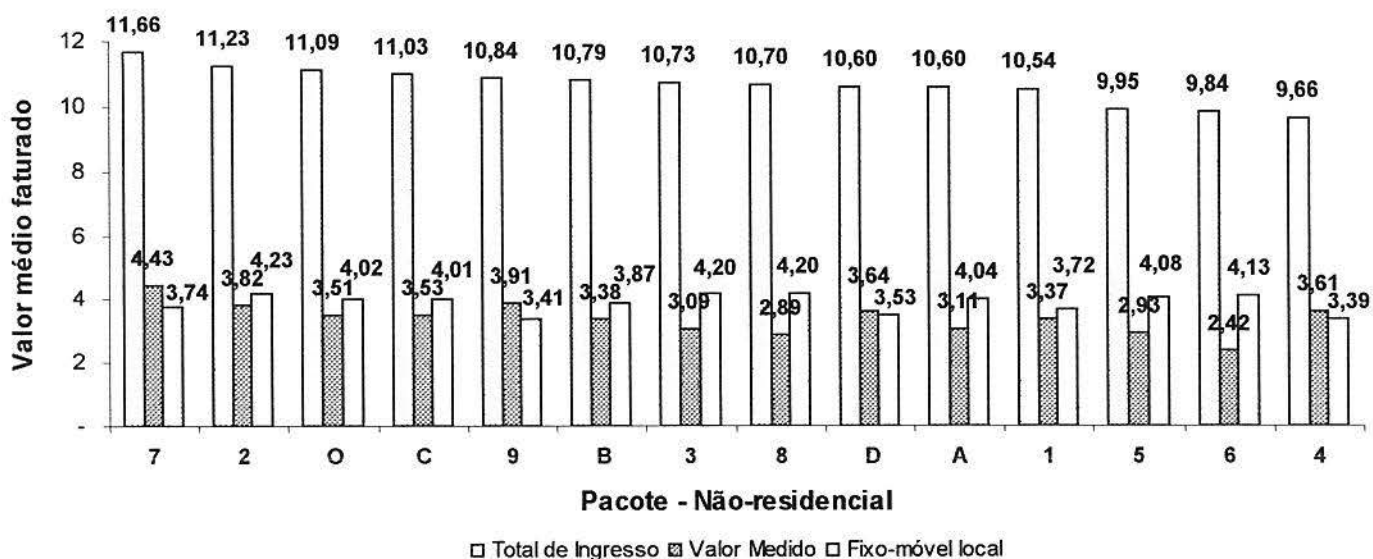


Figura 3.6 Faturamento médio da categoria não-residencial do total de ingresso, valor medido e fixo-móvel local dos pacotes de telefonicidades no GRUPO 3

Tabela 3.6 Faturamento médio da categoria não-residencial do total de ingresso, valor medido, fixo-móvel local e frequências dos pacotes de telefonicidades no GRUPO 3

Categoria Não-residencial					
Pacote	Frequência	Total de Ingresso	Valor Medido	Fixo-móvel local	
7	14	11,66	4,43	3,74	
2	679	11,23	3,82	4,23	
O	1.378	11,09	3,51	4,02	
C	422	11,03	3,53	4,01	
9	1.072	10,84	3,91	3,41	
B	1.133	10,79	3,38	3,87	
3	1.500	10,73	3,09	4,20	
8	275	10,70	2,89	4,20	
D	160	10,60	3,64	3,53	
A	8.747	10,60	3,11	4,04	
1	1.989	10,54	3,37	3,72	
5	9	9,95	2,93	4,08	
6	20	9,84	2,42	4,13	
4	22	9,66	3,61	3,39	

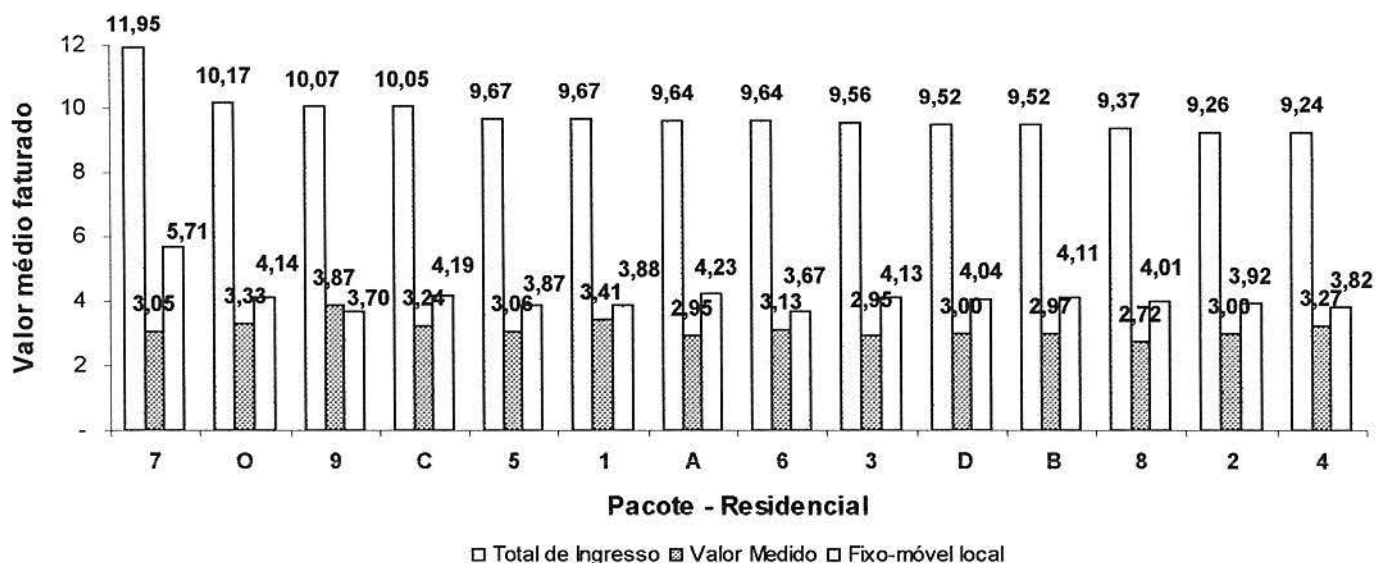
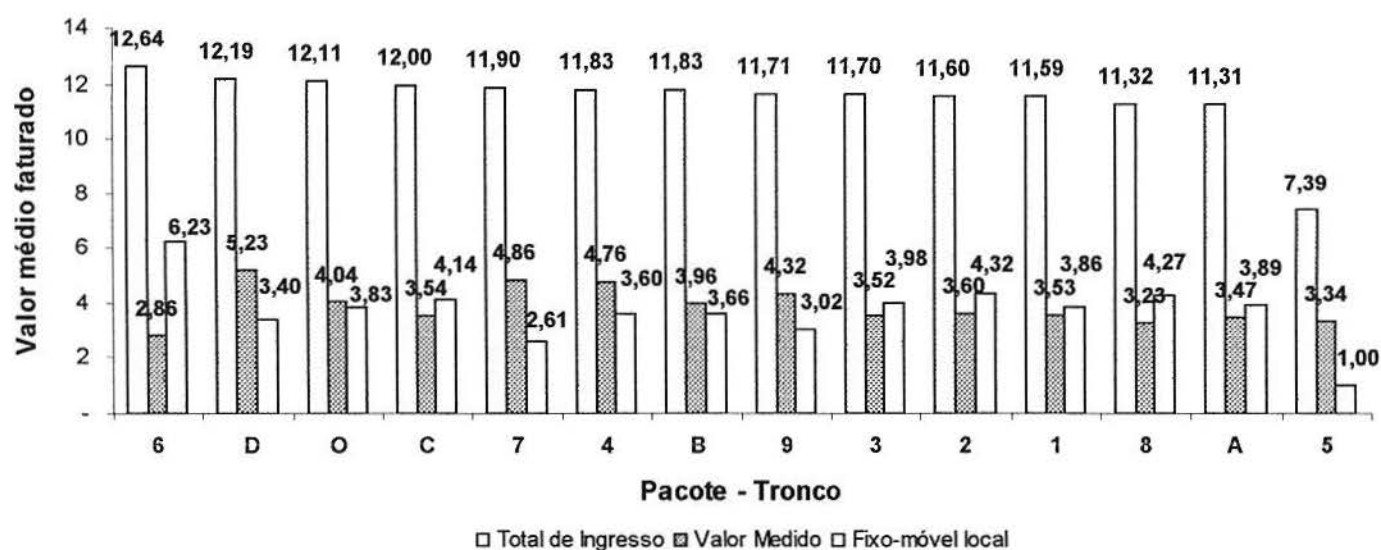


Figura 3.7 Faturamento médio da categoria residencial do total de ingresso, valor medido e fixo-móvel local dos pacotes de telefonicidades no GRUPO 3

Tabela 3.7 Faturamento médio da categoria residencial do total de ingresso, valor medido, fixo-móvel local e frequências dos pacotes de telefonicidades no GRUPO 3

Categoria Residencial					
Pacote	Frequência	Total de Ingresso	Valor Medido	Fixo-móvel local	
7	3	11,95	3,05	5,71	
O	897	10,17	3,33	4,14	
9	284	10,07	3,87	3,70	
C	380	10,05	3,24	4,19	
5	19	9,67	3,06	3,87	
1	3.591	9,67	3,41	3,88	
A	13.870	9,64	2,95	4,23	
6	74	9,64	3,13	3,67	
3	2.162	9,56	2,95	4,13	
D	278	9,52	3,00	4,04	
B	1.076	9,52	2,97	4,11	
8	282	9,37	2,72	4,01	
2	143	9,26	3,00	3,92	
4	16	9,24	3,27	3,82	



**Figura 3.8 Faturamento médio da categoria tronco do total de ingresso, valor medido e fixo-móvel local dos pacotes de telefacilidades no GRUPO 3**

**Tabela 3.8 Faturamento médio da categoria tronco do total de ingresso, valor medido, fixo-móvel local e frequências dos pacotes de telefacilidades no GRUPO 3**

Categoria Tronco					
Pacote	Frequência	Total de Ingresso	Valor Medido	Fixo-móvel local	
6	2	12,64	2,86	6,23	
D	13	12,19	5,23	3,40	
O	168	12,11	4,04	3,83	
C	38	12,00	3,54	4,14	
7	5	11,90	4,86	2,61	
4	23	11,83	4,76	3,60	
B	55	11,83	3,96	3,66	
9	110	11,71	4,32	3,02	
3	152	11,70	3,52	3,98	
2	485	11,60	3,60	4,32	
1	117	11,59	3,53	3,86	
8	19	11,32	3,23	4,27	
A	413	11,31	3,47	3,89	
5	3	7,39	3,34	1,00	

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

1. ANDERBERG, M.R. (1973) – **Cluster Analysis for Applications**. New York.
2. EVERITT, B. (1982) – **Cluster Analysis**. Hasteld Press.
3. FACHEL, J.M.G. (1976) – **Análise Fatorial**. Dissertação de Mestrado, Universidade de São Paulo.
4. HARTIGAN, J.A. (1975) – **Clustering Algorithms**. New York: John Wiley & Sons, Inc.
5. JOHNSON, R.A. & WICHERN, D.W. (1988) – **Applied Multivariate Statistical Analysis**. 2<sup>nd</sup> Edition – New Jersey: Prentice Hall.
6. MARDIA, K.V.; KENT, J.T. and BIBBY, J.M. (1979) – **Multivariate Analysis**. London: Academic Press.
7. MORRISON, D.F. (1976) – **Multivariate Statistical Methods**. McGraw Hill.
8. POSSOLI, S. (1992) – **Análise Multivariada**. Cadernos de Matemática e Estatística, Série B, nº 10. Porto Alegre: Universidade Federal do Rio Grande do Sul.
9. SAS Institute Inc. (1989) – **SAS/STAT User's Guide**. Version 6. Fourth Edition, Vol 1, Cary, NC: SAS Institute, Inc.