1. INTRODUÇÃO

Ao se iniciar as atividades de estágio no Atacadão dos Eletros, logo apareceu a dúvida de que tipo de problema seria escolhido para fazer parte do trabalho de conclusão de curso (T.C.C.), já que inúmeras tarefas foram realizadas, mas são de extremo sigilo, pois são informações estratégicas. Com este problema a ser solucionado, foram feitas algumas propostas, até que se definiu por estudar o comportamento do faturamento das filiais do Atacadão dos Eletros. Para tanto foram levantadas as seguintes variáveis:

- Localização da loja.
- Área da loja.
- Gasto com marketing.
- Quantidade de marcas em exposição.
- Tempo de funcionamento da filial (tempo de vida).
- Tempo de experiência como gerente no Atacadão dos Eletros, de cada gerente.
- Quantidade de concorrentes.
- Quantidade de vendedores.

Com base nestas variáveis, deseja-se classificar e agrupar as diversas filiais, onde exista uma homogeneidade dentro dos grupos e heterogeneidade entre eles, através da técnica estatística de análise de agrupamento. Também estudaremos através da análise fatorial (componentes principais) para avaliar a importância das diversas variáveis dentro dos fatores e que auxiliam na compreensão do faturamento das filiais.

2. OBJETIVOS

2.1.Objetivo Geral

Caracterizar as variáveis que são fatores preponderantes no comportamento do faturamento e separar as filiais em grupos distintos, cujos componentes são semelhantes dentro do grupo.

2.2.Objetivos Específicos

- Enumerar e descrever que variáveis são consideradas importantes para explicar o faturamento das filiais.
- Classificar e separar filiais por grupos.
- Analisar o perfil dos grupos encontrados.

3. REFERENCIAL TEÓRICO

3.1. Descrição das variáveis.

As variáveis coletadas nas 19 filiais do Atacadão dos Eletros, foram selecionadas por critérios de importância nos resultados de desempenho das filiais, tendo como período base o ano de 2004. O fornecimento das informações se deu pelos setores:

- Marketing: Área da loja e gasto com marketing.
- Recursos humanos: Tempo de gerência e número de vendedores.
- Contabilidade: Tempo de vida das filiais.
- Gerência comercial: Localização, quantidade de marcas e quantidade de concorrentes.

A variável localização da loja foi coletada junto à gerência comercial que respondeu um formulário, que tinha uma escala de *Likert*, a qual trazia a pergunta de como estava situada a localização da filial a respeito sobre o ponto de venda, a escala compreendia desde uma localização péssima até muito boa, contendo também o ponto de neutralidade.

A área da loja foi medida por metro quadrado.

O gasto com marketing foi obtido como sendo o total de gastos por filial.

A quantidade de marcas em exposição foi resultado do total de marcas de cada filial.

Tempo de funcionamento de cada filial (tempo de vida), foi coletado em meses.

Tempo de experiência como gerente no Atacadão dos Eletros (de cada gerente), também foi coletado em meses, na empresa.

A quantidade de concorrentes foi obtida como sendo o total de concorrentes num raio de 800 metros, conforme avaliação da gerência comercial e marketing, sobre quanto um consumidor percorreria para realizar uma pesquisa e efetuar suas compras.

3.2. Análise multivariada.

As técnicas de análise multivariada são de importância grande no ramo da estatística, Moreira (1985) cita que o objetivo da análise multivariada é analisar dados obtidos quando várias variáveis são medidas (ou avaliadas) num mesmo elemento. O termo multivariado ou multidimensional é usado em virtude de, sob o ponto de vista matemático, cada variável ser vista como uma dimensão.

A maioria das técnicas multivariadas é baseada no simples conceito de distância, por mais formidável que isso possa parecer. O conceito de distância euclidiana deve ser familiar para a maioria. Se for considerado um ponto P=(x1, x2) no plano cartesiano, a distância deste ponto P da origem O=(0,0), definida por d(O,P), é dada pelo teorema de Pitágoras por:

$$d(O,P)=\sqrt{x_i^2 + x_2^2} \qquad (1)$$

Em geral, se o ponto P tem p coordenadas, de tal forma que podemos dizer que P= (x1, x2, ... xp), a distância de P da origem O=(0,0, ...,0), pode ser generalizada por:

$$d(O,P)= \sqrt{x_i^2 + x_2^2 + ... + x_p^2} \qquad (2)$$

É desejável que as *p* respostas multivariadas sejam representadas por uma notação concisa. Os dados multivariados podem ser dispostos convenientemente como um arranjo de números. Em geral, um arranjo retangular destes números, com *n* linhas e *p* colunas, por exemplo, é chamada de matriz de dimensões *n* x *p*. Se por outro lado, o arranjo consiste em *n* mensurações em apenas 1 variável, ou ainda, de uma observação multivariada em *p* variáveis, esses arranjos são denominados de vetores (FERREIRA, 1996). Com esse arranjo bidimensional, não só, a notação fica mais concisa, mas os muitos resultados matemáticos de álgebra vetorial e matricial facilitam a derivação e exposição dos métodos estatísticos multivariados, os elementos de álgebra vetorial e matricial, serão considerados como conhecidos.

Dois vetores não nulos são denominados ortogonais, se o cosseno do ângulo entre eles for zero. Isto indica que:

$$\underset{\sim}{X}.\underset{\sim}{Y} = 0 \qquad (3)$$

Muitas vezes é desejável (em sistemas de equações lineares) construir uma base ortonormal de vetores, isto é, cada vetor da base possui comprimento unitário $\left(\underset{\sim}{x}_i \cdot \underset{\sim}{x}_i = 1 \right)$ (FERREIRA, 1996). Uma observação multivariada é uma coleção de medidas em p variáveis tomadas na mesma unidade experimental

Unidades amostrais ou experimentais	Variáveis			
	1	2 ...	k ...	p
1	X_{11}	X_{12}...	X_{1k}...	X_{1p}
2	X_{21}	X_{22}...	X_{2k}...	X_{2p}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
j	X_{j1}	X_{j2}...	X_{jk}...	X_{jp}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
n	X_{n1}	X_{n2}...	X_{nk}...	X_{np}

As n observações obtidas são dispostas em um arranjo (Matriz) **X**, em que cada linha de **X** representa uma observação multivariada. Desde que conjunto todo de mensurações é muitas vezes uma particular realização de variáveis aleatórias, diz-se que os dados representam uma amostra de tamanho n de uma população p variada.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2k} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{j1} & x_{j2} & \cdots & x_{jk} & \cdots & x_{jp} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} & \cdots & x_{np} \end{bmatrix} \quad (4)$$

Os dados podem ser plotados por um gráfico com p coordenadas. As colunas de **X** representam n pontos no espaço p dimensional. Esse tipo de gráfico fornece informações de locação dos pontos e de variabilidade. Se os pontos pertencem a uma esfera, o vetor de médias, $\underset{\sim}{\mu}$, é o centro de balanço ou de massa (FERREIRA, 1996). Se a variabilidade ocorre em mais de uma direção, pode-se detectar pela matriz de covariância (ou de dispersão), Σ. Uma medida numérica única de variabilidade é fornecida pelo determinante da matriz de covariância. Onde:

$$\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_p \end{bmatrix}, \text{ (vetor de média)} \qquad (5)$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix} \text{ (vetor de dispersão).} \qquad (6)$$

Os métodos de análise multivariada de dados visam à obtenção de informações da matriz de dados e a apresentação delas numa forma que permita uma fácil compreensão dos resultados. Estas técnicas podem ser classificadas de acordo com os objetivos da pesquisa e a maneira pela qual os dados foram obtidos. Ferreira (1996) diz que os métodos de análise multivariada podem dividir-se em dois grandes grupos: Técnicas de análise fatorial e Técnicas de classificação.

Por outro lado, as técnicas de análise fatorial podem ser divididas em três subgrupos, de acordo com a forma em que se apresenta a matriz de dados e pelos objetivos da análise. Esta divisão dos três subgrupos, assim como as técnicas utilizadas, são dadas a seguir como subgrupo 1,2 e 3.

No subgrupo 1 encontram-se as técnicas para o estudo de matrizes de dados nas quais não existe uma partição prévia, nem do conjunto de elementos *i*, nem do conjunto de variáveis *j*, também chamadas de métodos de interdependência. Neste subgrupo as três técnicas mais importantes são:

análise fatorial em fatores comuns, análise dos componentes principais e análise fatorial de correspondência. De uma maneira geral, o objetivo de todas elas é encontrar novas variáveis, em menor número que o das varáveis iniciais e que são combinações lineares das mesmas.

O subgrupo 2 consiste das técnicas para o estudo de matrizes de dados nas quais existe uma partição prévia do conjunto de elementos *i*. Neste caso destacam-se as técnicas de análise discriminante que podem ser utilizadas para fins descritivos ou para tomadas de decisão, quando se pretende uma descrição dos grupos (ou parte deles).

O subgrupo 3 é composto por técnicas para matrizes de dados nas quais o conjunto de variáveis *j* está subdividido em subconjunto, os quais são denominados de métodos de dependência.

As técnicas de classificação, que chamaremos de análise de agrupamento, são técnicas que tem por objetivo proporcionar uma ou várias partições do conjunto de elementos *i*, ou do conjunto de variáveis *j*, desde que na matriz de dados não exista partição prévia nem no conjunto *i* e nem no conjunto *j*.

Moreira (1985) cita que as técnicas de agrupamento (ou classificação) podem ser divididas em dois grandes subgrupos: Técnicas de agrupamento hierárquico e técnicas de classificação não hierárquicas.

As técnicas de agrupamento hierárquicas mais utilizadas são aquelas que partindo de um número de grupos (*clusters*) igual ao número de elementos, procedem à agregação dos mesmos, por etapas conduzindo a elaboração de um gráfico de árvore, denominado de dendograma, cuja análise ajuda a constituição de grupos e subgrupos. Os métodos utilizados nesta técnica são: método de ligação simples, ligação completa, método do centróide, método da variância mínima (*Ward*) dentre outros.

Também chamado *gráfico de árvore*, o dendograma é um dispositivo gráfico para apresentar os resultados de aglomeração. As linhas verticais representam conglomerados unidos. A posição da reta na escala indica as distâncias às quais os conglomerados foram unidos. O dendograma é lido da direita para a esquerda (MALHOTRA, 2001).

As técnicas de agrupamento não hierárquicos são caracterizadas pela partição do conjunto de elementos em agrupamentos e a maioria das técnicas de partição admitem conhecido a priori o número de agrupamentos desejados.

3.2.1 Análise de agrupamento

As técnicas aqui apresentadas têm seus fundamentos discutidos e foram desenvolvidas para análise de dados multivariados, podendo ser entendidas como um capítulo da análise multivariada. A formulação geral dos problemas em que podem ser utilizadas é a seguinte:

Dado um conjunto E de n elementos, onde cada um é representado por um conjunto de medidas de p variáveis, deseja-se determinar os sub-conjuntos ou agrupamentos de elementos em que pode ser decomposto o conjunto E, com base nas p variáveis.

Talvez a parte mais importante da formulação de um problema de conglomerado seja a escolha das variáveis sobre as quais se baseará o processo de aglomeração. A simples inclusão de uma, ou duas, variáveis sem importância pode distorcer uma solução que, de outra forma, se revelaria útil. Basicamente, o conjunto de variáveis escolhidas deve descrever a semelhança entre objetos (MALHOTRA, 2001).

Apesar da atualidade do assunto não é o mesmo, de forma alguma novo. As primeiras técnicas de remontam ao passado da humanidade. As grandes civilizações sempre se preocuparam com problemas de classificação (GAMA, 1980). Em sua forma mais moderna a análise de agrupamento tem suas origens nos trabalhos de Pearson (1901) e de Spearman (1904), onde Spearman publicou um trabalho famoso, considerado ponto de partida, trabalho esse, que examinou os escores obtidos em provas de uma escola inglesa (NASCIMENTO, 2005). Gama (1980) também cita este trabalho onde ele relata que foi feito um estudo visando a construção de um algoritmo denominado V-Análise (análise de agrupamento de variáveis) (GAMA, 1980).

A análise de agrupamento teve seu desenvolvimento pela necessidade de técnicas alternativas a análise fatorial e a análise de componentes principais.

A análise de agrupamento é uma técnica usada para classificar objetos ou casos em grupos relativamente homogêneos chamados *conglomerados*. Os objetos em cada conglomerado tendem a ser semelhantes entre si, mas diferentes de objetos em outros agrupamentos (MALHOTRA, 2001).

A análise de agrupamento é um instrumento que, em algumas situações, serve de técnica alternativa as utilizadas pela analisa estatística clássica. Alguns procedimentos de análise de dados podem ser substituídos pela análise de agrupamento.

Ao contrário, a maioria dos métodos de conglomeração é heurística, baseada em algoritmos. Assim, a análise de conglomerados contrasta fortemente com a análise se variância, a regressão, a análise discriminante e a análise fatorial, que se baseiam em um rigoroso raciocínio estatístico. Embora muitos métodos de aglomeração tenham importantes propriedades estatísticas, não se pode deixar de reconhecer a simplicidade de fundamentos de tais métodos (MALHOTRA, 2001).

Na análise de agrupamento, não há preocupação com o comportamento probabilístico das variáveis, isto é, não é exigido o conhecimento das distribuições de probabilidade das variáveis.

Admitindo que se tenha um conjunto *E* de *n* elementos pertencentes a grupos ou populações distintas, admitindo-se ainda que em cada um dos elementos são realizadas *p* medidas referentes a *p* variáveis ou características presentes, os elementos de *E* poderão ser caracterizados por um vetor, ponto de espaço *p*-dimensional da forma:

$$\underset{\sim}{x}_i = (x_{i1}, x_{i2}, ..., x_{ip}) \text{, i=1,...,n} \tag{7}$$

onde x_{iq}, (q=1,...,q) é a medida da característica *q*-ésima no *i*-ésimo elemento.

Seja $E = \{E_1, E_2, ... E_n,\}$, o conjunto de *n* elementos. A cada um dos elementos de *E* pode-se associar um vetor $\underset{\sim}{x}_i$, podendo o conjunto de observações ser representado por uma matriz $X_{(nxp)}$ (GAMA, 1980).

É importante notar que os valores de $\underset{\sim}{x}_i$ formam um conjunto de *n* pontos em um espaço Euclidiano *p*-dimensional (I_p).

Considerando um inteiro *g*, tal que, $g \leq n$, o problema de determinar agrupamentos associados ao conjunto de *n* elementos é o da determinação de *g* grupos ou subconjuntos em *E*, de forma que cada um de seus elementos pertença a um e somente um subconjunto. Os elementos pertencentes a o mesmo subconjunto são ditos similares, e em conjuntos distintos não similares ou dissimilares (GAMA, 1980).

A construção dos *g* agrupamentos pode ser vista como a determinação de uma partição de espaço I_p em regiões R_k (K=1,...,*g*), onde a interseção é exceção a regra geral:

$R_I \cap R_J = \emptyset$, para $i \neq j$ tal que, se um ponto de (7) pertence a uma região R_k, não pertencerá a nenhuma outra. A partição do espaço I_p obedece $\bigcup_{K=1}^{g} R_k = I_p$ (GAMA, 1980).

Logo os elementos pertencentes ao subconjunto R_k constituirão o que denominamos de agrupamentos. A determinação da partição de I_p é a solução do problema, e é obtida pela aplicação de um critério de agrupamento de cada elemento de *E* a um particular conjunto R_k de I_p.

Ainda na atualidade não existe uma definição formal para agrupamento que possa ser aceita sem discussões. Uma definição de agrupamentos deverá conter a idéia de espaço *p*-dimensional, e poderá ser feira através da conceituação de partição deste espaço em regiões contíguas.

A solução de um problema de agrupamento é obtida pela determinação de uma partição do espaço I_p, onde uma função dos valores dos vetores representativos dos elementos de *E*, para os quais quer-se determinar uma medida de similaridade ou distância. Várias considerações são necessárias para definir as diversas funções possíveis de serem utilizadas como medidas de similaridade ou distância.

Como o objetivo do conglomerado é agrupar objetos semelhantes, torna-se necessária alguma medida para avaliar quão semelhantes, ou quão diferentes, são os objetos. A abordagem mais comum consiste em avaliar a semelhança entre os termos de distância entre pares de objetos. Os objetos com menor distância entre si são mais semelhantes uns dos outros, do que objetos com maior distância. Há varias maneiras de calcular distância entre objetos (MALHOTRA, 2001).

A fórmula de distância a ser utilizada será a distância Euclidiana quadrática. Distância esta que a métrica de maior utilização como função de agrupamento e apresenta enorme facilidade de cálculo.

Sejam os vetores $\underset{\sim}{x}_i$ e $\underset{\sim}{x}_j$, pontos do espaço I_p. Define-se como distância euclidiana quadrática entre os elementos E_i e E_J a função:

$$d_{ij} = \left[\sqrt{\sum_{q=1}^{p} \left(x_{iq} - x_{jq} \right)^2} \right]^2 \tag{8}$$

onde d_{ij} é a distância e i para j.

Existe um grande número de medidas que podem ser utilizadas como função de agrupamento, todas as medidas descritas quantificam à distância, ou a similaridade entre dois elementos quaisquer do conjunto E de elementos.

A quantificação da distância entre agrupamentos assume algumas particularidades, principalmente no que se refere à dependência do número de elementos em cada grupo, até obter-se o número de agrupamentos desejados, ou até que todos os elementos tenham sido alocados em u único agrupamento, formado por todos os elementos de E (GAMA, 1980).

Malhotra (2001) relata que a medida de semelhança mais comumente utilizada é a distância Euclidiana ou o seu quadrado. A distância Euclidiana é a raiz quadrada da soma dos quadrados da diferença dos valores para cada variável

Também é necessário, a uma análise completa, uma medida de homogeneidade para cada um dos agrupamentos obtidos, bem como uma medida de diferença entre dois agrupamentos.

Do ponto de vista estatístico o problema de agrupamento pode ser posto nos seguintes termos:

Um conjunto E, com n elementos é observado em relação a p variáveis, e deseja-se dividi-lo em agrupamentos tal que $g \leq n$.. O problema básico é o de determinar o melhor número de agrupamentos em deva ser dividido o conjunto de n elementos. Uma consideração estatística importante na procura da solução é aquela que diz respeito à dispersão dos agrupamentos a serem formados Um razoável critério para determinar o número de agrupamentos g seria aquele que torne a dispersão entre os agrupamentos máxima e dentro dos agrupamentos mínima (GAMA, 1980).

A determinação do número de agrupamentos é um dos mais difíceis problemas na análise de agrupamentos, pois sempre se têm as perguntas: Quantos agrupamentos devem ser escolhidos? Quais métodos são adequados ao experimento e dentre eles, qual escolher?

Em algumas situações o número de g agrupamentos é pré-fixado, geralmente, pela própria natureza do problema a ser analisado. A determinação do número de agrupamentos é ainda um problema na análise de agrupamentos.

Um outro problema é o fato de que as medidas de distância ou de similaridade podem não apresentar uma significância, uma vez que a arbitrariedade que envolve a escala numérica de algumas variáveis não justifica sempre a reunião dos dois ou mais elementos de E em um agrupamento. Na maioria dos casos, em particular nas variáveis qualitativas, os valores são escolhidos pelo seu significado ordinal. Gama (1980) cita que seria razoável o estudo que, aceitando transformações monótonas, como, por exemplo, logaritmo ou raiz quadrada, fizessem com que os resultados fossem invariantes.

O método de agrupamento a ser utilizado é método de ligação completa, ou de encadeamento completo, ou ainda de vizinho mais distante (GAMA, 1980). É uma técnica de hierarquização aglomerativa de maior aplicação e não exige a fixação a priori de número de *g* de agrupamentos.

Gama (1980) diz que dados *n* elementos e admitindo-se conhecidos os $\frac{n(n-1)}{2}$ valores de uma função de agrupamento, d_{ij} apresentados em forma matricial, o método de ligação completa pode ser sintetizado nas seguintes etapas:

Determina-se com base no conjunto de *n* vetores $\underset{\sim}{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})$ Isto é, na matriz $|X|$ de dados, o conjunto de valores de uma função de agrupamento, constituindo-se os elementos da matriz em medidas de distância ou similaridade, entre os elementos de *E*;

Determina-se o valor de $\min\{d_{ij}\} > 0$, reunindo-se os elementos E_i e E_j, formando um primeiro agrupamento, passa-se então a ter (n-1) agrupamentos distintos;

Com base na matriz de distância d_1, determina-se à distância entre cada um dos (n-2) elementos e o novo agrupamento formado pelos elementos (E_i E_j). Esta distância é calculada pela relação:

$$d_{(ij)\ell} = MAX\{d_{i\ell}, d_{j\ell}\},\ \ell = 1,...,n-2 \text{ e } \ell \neq i \neq j \quad (9)$$

Determina-se no novo conjunto de valores da função de agrupamento, isto é, na nova matriz de distância d_2, o menor valor não nulo, agrupando-se os elementos correspondentes, dando origem a um novo agrupamento, obtendo-se então (n-2) agrupamentos;

Compõe-se novo conjunto de valores de distâncias da interação anterior, entre o novo agrupamento e os demais. A distância entre dois agrupamentos A e B quaisquer é dada por:

$$D_{A,B}=\text{MAX}\, d\left(\underset{\sim}{x}_{v}, \underset{\sim}{x}_{\Gamma} \right), \text{ onde } v = 1,\ldots,n_A \text{ e } r = 1,\ldots,n_B \qquad (10)$$

Como vimos, a combinação de atributos (classes) que definem os agrupamentos é o fator preponderante na determinação do número de agrupamentos. Caso os agrupamentos sejam definidos a priori, o problema é de alocação.

No caso de utilização para construir-se os agrupamentos com base na matriz de dados, é de fundamental importância o estudo descritivo de cada uma das variáveis, em particular no número de classes me que serão divididas (GAMA, 1980).

O motivo dá escolha pelo método de ligação completa é baseada na informação do quadro resumo de métodos de agrupamentos descrito por Gama (1980).

Outro aspecto ligado à utilização das variáveis, é se as mesmas devem ou não ser padronizadas, isto é, reduzidas à escala (0,1). Gama (1980) diz que processos que utilizam como função de agrupamento à distância Euclidiana, são variantes para padronização das variáveis e que se deve evitar variáveis padronizadas quando da utilização de técnicas que envolvam distância Euclidiana.

Um outro problema associado as variáveis é o da ponderação das mesmas, uma vez que nem todas têm a mesma influência no problema. Algumas medidas de distância como, por exemplo, a distância Euclidiana, permite este procedimento. Cabe observar que há influência de aspectos subjetivos, uma vez que a ponderação é feita com base no conhecimento do analista.

3.2.2 Análise fatorial

A análise fatorial é um nome genérico que denota uma classe de processos utilizados essencialmente para redução e sumarização dos dados. Estudam-se as relações entre conjuntos de muitas variáveis inter-relacionadas representando-as em termos de alguns fatores fundamentais (MALHOTRA, 2001).

Barroso (2003) diz que a análise fatorial é uma técnica estatística que tem como objetivo descrever a estrutura de dependência de um conjunto de variáveis através da criação de fatores, que são as variáveis que, supostamente, medem aspectos comuns.

Em análise de variância, regressão múltipla e análise discriminante, uma variável é considerada como dependente, ou variável critério, e as outras, como variáveis independentes. Em análise fatorial, entretanto, não se faz tal distinção. Ao contrário, a análise fatorial é uma técnica de interdependência, no sentido de que examinamos todo um conjunto de relações interdependentes (MALHOTRA, 2001).

Historicamente, a origem das técnicas de análise fatorial está ligada a estudos da área de psicologia. Sua criação data do início do século quando foi desenvolvido um método para a criação de um índice geral de inteligência (fator *g*) com base nos resultados de vários testes que refletiriam essa aptidão. Barroso (2003) relata que em 1940 Lawley fez o primeiro trabalho com um maior rigor matemático, o que fez que se aumentasse a aceitação dessas técnicas.

Matematicamente, a análise fatorial é algo semelhante à análise de regressão múltipla, pelo fato de cada variável ser expressa como uma combinação linear de fatores subjacentes. A quantidade de variância que uma variável compartilha com todas as outras variáveis incluídas na análise é chamada comunalidade. A covariação entre as variáveis é descrita em termos de um pequeno número de fatores comuns, mais um fator único (ou exclusivo) para cada variável esses fatores não são observados abertamente. Se as variáveis são padronizadas, o modelo fatorial pode ser representado como:

$$X_i = A_{i1} F_1 + A_{i2} F_2 + A_{i3} F_3 + \ldots + A_{im} F_m + V_i U_i \qquad (11)$$

onde

X_i = i-ésima variável.

A_{ij} = coeficiente de regressão múltipla da variável *i* sobre o fator comum *j*.

F = fator comum.

V_i = coeficiente de regressão múltipla da variável *i* sobre o fator único *i*.

U_i = o fator único para a variável *i*.

m = número de fatores comuns.

Os modelos de análise fatorial buscam explicar o comportamento das variáveis observadas em relação ao comportamento de um conjunto de variáveis não observadas.

Uma suposição que forma a base do modelo de análise fatorial é que cada variável X_I, que é componente do vetor $\underset{\sim}{X}$ é influenciada por uma parte comum a todas as variáveis de $\underset{\sim}{X}$, chamadas de fator comum (NASCIMENTO, 2005).

Os fatores únicos não são correlacionados uns com os outros e com os fatores comuns. Os fatores comuns podem, eles próprios, ser expresso como combinações lineares de variáveis observáveis.

$$F_i = W_{i1} X_1 + W_{i2} X_2 + W_{i3} X_3 + \ldots + W_{ik} X_k \qquad (12)$$

Onde

Fi = estimativa do i-ésimo fator.

W_i = peso ou coeficiente do escore fatorial.

k = número de variáveis.

É possível escolher pesos ou coeficientes de escore do fator de modo que o primeiro fator explique a maior parte da variância total.

Para se usar a análise de fatores, ou seja, análise fatorial é preciso que o conjunto de dados se adequem a esta técnica através de estatísticas como:

- Teste de esfericidade ou de esfera de Bartlett : Uma estatística de teste usada para examinar a hipótese de que as variáveis não sejam correlacionadas na população. Em outras palavras, a matriz de correlação da população é uma matriz de identidade; cada variável se relaciona perfeitamente com ela própria (r=1), mas não apresenta correlação com as outras variáveis (r=0) (MALHOTRA, 2001). Nascimento (2005) diz que o teste de esfericidade de Bartlett objetiva a informação de que as correlações são baixas demais para o modelo de análise fatorial e que Bartlett desenvolveu este teste de hipóteses, que permite decidir se as correlações fora da diagonal da matriz de correlações são nulas, ou seja, $H_o: \rho_{ij} = 0$, com $i \neq j$, contra $H_1: \rho_{ij} \neq 0$, com $i \neq j$. Se H_o for aceita a matriz de correlações populacional será a matriz identidade de ordem p (I_P), e como a forma quadrática associada a I_P representa um sólido geométrico de esfera de dimensão p, daí o nome *teste de esfera*. Se o valor do nível descritivo (*p-valor*) for inferior ou igual ao nível de significância que for estipulado diremos que a matriz de correlações populacional será diferente da matriz identidade de ordem *p.* A estatística de teste da esfericidade se baseia em uma transformação qui-quadrada do determinante da matriz de correlação.

- Medida de adequacidade da amostra de Kaiser-Meyer-Olkin (KMO): É um índice que compara a grandeza das correlações da matriz de variáveis e as correlações da matriz de correlações parciais. Se o modelo de análise fatorial estiver adequado as correlações parciais deveriam ser muito pequenas, então Kaizer, Meyer e Olkin idealizaram uma medida que quando está próxima de 1(um) reflete o fato de que estas correlações parciais estariam próximas de 0(zero) confirmando um bom casamento dos dados com o modelo de análise fatorial. Pequenos valores para a medida de KMO indicam que análise fatorial pode não ser uma boa idéia, pois as correlações entre os pares de variáveis não podem ser explicadas pelos fatores comuns (NASCIMENTO, 2005).

Barroso (2003) explica que existe na literatura clássica da estatística uma tabela que auxilia na interpretação do KMO. Esta tabela foi originalmente criada por Kaiser e Rice em 1974.

Tabela 01: Interpretação do KMO.

KMO	Interpretação
0,80 - 1,00	Excelente
0,70 - 0,80	Ótimo
0,60 - 0,70	Bom
0,50 - 0,60	Regular
0,00 - 0,50	Insuficiente

Fonte: *Barroso e Artes(2003, p 97).*

A formulação do problema envolve várias fases. Em primeiro lugar, devemos identificar os objetivos da análise fatorial. As variáveis a serem incluídas na análise fatorial devem ser especificadas com base em pesquisas anteriores, na teoria e no julgamento do pesquisador. É importante medir adequadamente as variáveis em uma escala de intervalo ou de razão (MALHOTRA, 2001).

Uma vez assegurado que a análise fatorial é uma técnica adequada para analisar os dados, devemos selecionar um método apropriado. A abordagem usada para deduzir os pesos, ou coeficientes dos escores dos fatores, diferencia os diversos métodos de análise fatorial.

O método selecionado foi o de análise de componentes principais, método esse que leva-se em conta a variância total dos dados. A diagonal da matriz de correlação consiste de unidades, e a variância plena é introduzida na matriz de fatores. Recomenda-se a análise de componentes principais quando a preocupação maior é determinar o número mínimo de fatores que respondem pela máxima variância nos dados para a utilização em análises multivariadas subseqüentes (MALHOTRA, 2001), motivo esse da escolha do método. Os fatores são chamados de *componentes principais*.

Existem ainda outros métodos para estimar os fatores comuns, incluindo o método de análise comum, dos mínimos quadrados não-ponderados, dos mínimos quadrados generalizados, a máxima verossimilhança, alpha, e o fatoramento da imagem.

Na análise fatorial aparecem resultados que irão auxiliar no entendimento e resolução do problema, alguns dos resultados que são utilizados nesta técnica serão descritos, abaixo, pois só serão utilizados alguns deles:

- Matriz de correlação: O triângulo inferior da matriz que exibe as correlações simples *r*, entre todos os pares possíveis de variáveis incluídas na análise. Os elementos da diagonal, que são todos igual a 1, em geral são omitidos. Nascimento (2005) cita que um método difere do outro na forma como se altera a matriz de correlação antes de obter uma solução final e me todos os métodos, se substitui a diagonal principal da matriz de correlação pelas comunalidades estimadas antes de se chegar à estimação final dos fatores comuns que é obtida de forma interativa.

- Comunalidade: Porção da variância em que uma variável compartilha com todas as outras variáveis consideradas. É também a proporção de variância explicada pelos fatores comuns.

- Autovalores (*Eigenvalues*): Representa a variância total explicada por cada fator.

- Cargas dos fatores: Correlações simples entre as variáveis e os fatores.

- Matriz de fatores: Contém as cargas dos fatores de todas as variáveis em todos os fatores extraídos.

- Escores fatoriais: Escores compostos estimados para cada entrevistado nos fatores derivados.

- Resíduos: Diferenças entre as correlações observadas, dadas na matriz de correlação de entrada (*input*) e as correlações reproduzidas, conforme estimadas pela matriz de fatores.

- Gráfico de declive (*Scree plot*): Gráfico dos autovalores versus número de fatores por ordem extração.

O processo analítico se baseia em uma matriz de correlação entre as variáveis. Um exame dessa matriz permite uma boa observação. Para que a análise fatorial seja apropriada, as variáveis devem ser correlacionadas (MALHOTRA, 2001).

Um fato importante na análise fatorial é a matriz de fatores, que contém os coeficientes utilizados para expressar as variáveis em termos dos fatores. Esses coeficientes (*cargas fatoriais)* representam as correlações entre os fatores e as variáveis. Um coeficiente com valor absoluto grande indica que o fator e a variável estão altamente relacionados. Com isso podemos utilizar os coeficientes da matriz de fatores para interpretar os fatores.

Embora a matriz inicial (não rotada) de fatores indique a relação entre os fatores e as variáveis individuais, ela raramente resulta em fatores que possam ser interpretados, porque os fatores são correlacionados com muitas variáveis.

Quando rotamos os fatores, tentamos fazer que cada fator tivesse cargas ou coeficientes diferentes de 0(zero) para apenas algumas das variáveis. Quando se faz a rotação as comunalidades não são afetadas nem mesmo a percentagem explicada da variância total. Contudo a variância explicada por cada fator individual é redistribuída por rotação. Logo, diferentes métodos de rotação podem resultar na identificação de diferentes fatores.

A rotação é chamada de ortogonal se os eixos são mantidos e ângulo reto. O método de rotação mais comumente utilizado é o processo varimax. Trata-se de um método ortogonal de rotação que minimiza o número de variáveis, com altas cargas sobre um fator, reforçando assim a interpretabilidade dos fatores. A rotação ortogonal tem como resultado fatores não-correlacionados (MALHOTRA, 2001). Com base nesta informação iremos utilizar a rotação ortogonal com processo varimax.

O número de componentes principais pode ser igual ao número de variáveis, mas com isso nada se ganha em economia. Para resumir as informações contidas nas variáveis originais, deve-se extrair um número menor de fatores. Para determinar o número de fatores, são sugeridos vários processos. Entre eles podemos citar a determinação a priori e abordagens baseadas em autovalores (*eingenvalues*) e gráfico de declive (*scree plot*).

Quando se usa a extração a priori ela acaba quando se atinge o número desejado de fatores.

Quando se utiliza a extração por autovalores apenas são considerados os fatores com autovalores superiores a 1(um).

Quando se usa a extração pelo gráfico de declive, ele apresenta uma acentuada interrupção entre o acentuado declive dos fatores com grandes autovalores e uma gradual redução relacionada com o restante dos fatores.

A interpretação dos fatores é facilitada pela identificação das variáveis que apresentam cargas altas sobre o mesmo fator. Para ajudar a interpretação ainda podemos utilizar o gráfico de declive. As variáveis próximas da origem têm pequenas cargas sobre os fatores. As variáveis que não estão próximas de nenhum dos eixos estão relacionadas a ambos aos fatores.

Em seguida à interpretação, pode-se calcular os escores fatoriais, se necessário. A análise fatorial tem valor por si só. Todavia, se o objetivo da análise fatorial é reduzir o conjunto de variáveis originais a um conjunto menor de variáveis compostas (fatores) para uso em uma análise multivariada subseqüente, é conveniente calcular escores fatoriais para cada entrevistado. Um fator nada mais é do que uma combinação linear das variáveis originais (MALHOTRA, 2001).

O escore para o i-ésimo fator pode ser estimado como segue:

$$F_i = W_{i1} X_1 + W_{i2} X_2 + W_{i3} X_3 + \ldots + W_{ik} X_k \qquad (13)$$

Malhotra (2001) comenta que somente no caso de análise de componentes principais é que podemos exatamente calcular os escores fatoriais. Além disso, em análise de componentes principais, esses escores não são correlacionados. Em análise fatorial comum, obtêm-se estimativas desses escores, mas não garantia de que os fatores sejam não-correlacionados uns com os outros.

A última etapa da análise fatorial é a de determinar o ajuste do modelo. Temos que supor que a correlação observada entre as variáveis pode ser atribuída a fatores comuns. Com isso as correlações entre as variáveis podem ser reproduzidas das correlações estimadas entre as variáveis e os fatores. Pode-se examinar as diferenças entre as correlações observadas e as correlações reproduzidas a fim de se determinar o ajuste do modelo. Chamamos essas diferenças de resíduos. Se há muitos resíduos grandes (50% ou mais), o modelo fatorial não dá um bom ajuste aos dados, e deve ser reconsiderado.

4. RESULTADOS

Na seqüência, de acordo com o tema de estudo proposto, verificou-se através da análise estatística dos dados, o resultado dos objetivos gerais e específicos como já definidos.

Para caracterizar as filiais, em grupos distintos, iremos observar os resultados a seguir.

Tabela 02: Esquema de aglomeração dos conglomerados e seus coeficientes de distância dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

	Conglomerado combinado			Estágio em que o conglomerado aparece pela primeira vez		
Estágio	Conglomerado 1	Conglomerado 2	Coeficientes	Conglomerado 1	Conglomerado 2	Próximo estágio
1	4	6	147.940.946,41	0	0	13
2	7	9	1.262.107.977,86	0	0	6
3	5	10	3.312.616.056,07	0	0	12
4	14	16	5.577.444.880,56	0	0	11
5	11	12	10.908.885.022,34	0	0	10
6	7	15	14.479.221.891,14	2	0	11
7	3	13	23.357.976.175,10	0	0	12
8	8	18	57.045.790.715,75	0	0	14
9	2	19	83.434.360.676,12	0	0	13
10	1	11	85.054.573.154,60	0	5	16
11	7	14	123.802.160.902,26	6	4	14
12	3	5	129.435.407.235,11	7	3	15
13	2	4	443.349.383.433,44	9	1	18
14	7	8	575.037.807.731,58	11	8	16
15	3	17	612.743.982.821,68	12	0	17
16	1	7	1.720.014.309.650,00	10	14	17
17	1	3	5.230.134.720.314,14	16	15	18
18	1	2	12.808.559.120.447,60	17	13	0

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.***

Na tabela 02, que é o esquema de aglomeração podemos verificar que a primeira coluna (estágio) é o total de casos (n) menos 1(um) (no nosso caso n= 19 filiais), na primeira linha observamos que o estágio 1, com 19 conglomerados. Nesse estágio são combinadas os casos 4 (filial D) e 6 (filial F), conforme mostram as colunas de conglomerados combinados. O quadrado da distância euclidiana entre os dois casos é dada na coluna de coeficientes. O estágio em que o conglomerado aparece pela primeira vez é composto pelas colunas de conglomerado 1, conglomerado 2 e próximo estágio, indica primeiramente (conglomerado 1 e conglomerado 2) se o caso já apareceu em um conglomerado combinado, enquanto que a coluna próximo estágio mostra em que estágio (primeira coluna) aparecerá um dos casos.

O gráfico 01 mostra o gráfico de árvore ou dendograma (observa-se da esquerda para direita) como os casos (filiais) se agrupam em relação as variáveis de estudo, podemos perceber se dividirmos o conjunto de conglomerados em três blocos, temos:

- **Grupo I**: Formado pelas filiais B, D, F e S.
- **Grupo II**: Formado pelas filiais C, E, J, M e Q.
- **Grupo III**: Formado pelas filiais A, G, H, I, K, L, N, O, P e R.

Nos conglomerados estabelecidos as filiais ficam separadas de tal forma que no primeiro grupo ficaram as filiais de menor desempenho, no segundo as de maior desempenho e no terceiro as de desempenho intermediário, quando comparado com o *ranking* das filiais quanto ao faturamento.

Gráfico 01: Gráfico de árvore ou dendograma dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004.*

Tabela 03: Centróide dos grupos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

Ranks

	Centróide dos grupos							
Grupos	Área Construída	Quantidade de Vendedores	Tempo de Vida da loja	Tempo de Experiência	Quantidade de Marcas	Gasto com Marketing	Quantidade de Concorrentes	Localização
I	944,75	3,50	75,75	34,75	74,00	14.740,50	5,00	5,75
II	1.190,20	9,80	65,00	40,00	108,80	68.638,00	5,80	6,60
III	581,60	7,80	53,50	38,70	97,50	47.568,20	5,50	6,20

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004*

Os valores que constam na tabela 03 correspondem às médias das variáveis dentro de cada grupo. Observa-se que em relação a variável quantidade de vendedores o grupo I apresenta média de 3,50 vendedores por loja, para o grupo II a média foi de 9,80 vendedores e para o grupo III de 7,80. A variável quantidade de concorrentes teve a sua maior média no grupo II (5,80), tendo a segunda maior média no grupo II (5,50) e conseqüentemente a menor média no grupo III (5,00).

Para descrever que variáveis são consideradas importantes para explicar o faturamento das filiais, iremos observar os resultados a seguir.

Tabela 04: Matriz de correlação dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

		Localização	Área da loja	Quantidade de vendedores	Tempo de vida da loja	Tempo de gerência	Quantidade de marcas	Gasto com Marketing	Quantidade de concorrentes
Correlação	Localização	1,000	,683	,495	-,285	,158	,491	,529	-,259
	Área da loja	,683	1,000	,405	-,235	-,107	,291	,322	-,333
	Quantidade de vendedores	,495	,405	1,000	-,134	-,086	,739	,754	-,198
	Tempo de vida da loja	-,285	-,235	-,134	1,000	-,155	-,261	-,167	-,124
	Tempo de gerência	,158	-,107	-,086	-,155	1,000	,147	,217	,254
	Quantidade de marcas	,491	,291	,739	-,261	,147	1,000	,828	,158
	Gasto com Marketing	,529	,322	,754	-,167	,217	,828	1,000	,151
	Quantidade de concorrentes	-,259	-,333	-,198	-,124	,254	,158	,151	1,000
Nível descritivo (p-valor)	Localização		,001	,016	,119	,259	,016	,010	,143
	Área da loja	,001		,043	,167	,331	,114	,089	,082
	Quantidade de vendedores	,016	,043		,292	,364	,000	,000	,208
	Tempo de vida da loja	,119	,167	,292		,263	,140	,248	,307
	Tempo de gerência	,259	,331	,364	,263		,274	,186	,147
	Quantidade de marcas	,016	,114	,000	,140	,274		,000	,259
	Gasto com Marketing	,010	,089	,000	,248	,186	,000		,268
	Quantidade de concorrentes	,143	,082	,208	,307	,147	,259	,268	

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004.*

A tabela 04 mostra a matriz de correlações dos dados. Podemos observar que na diagonal principal os valores são iguais a 1(um), pois a correlação de uma variável com ela mesma é 1(um) e que o triângulo inferior da matriz é igual ao triângulo superior da matriz, pois a correlação de A com B, é igual a correlação de B com A. Ao detalhar mais a tabela 04 vemos as correlações entre as variáveis, além do nível descritivo (p-valor) sobre as correlações. Se o nível descritivo for inferior ao nível de significância (no nosso caso adotamos $\alpha = 0{,}05$), isso indica que existe correlação significativa ao nível de significância de 5%. As variáveis área da loja, quantidade de vendedores, quantidade de marcas e gasto com marketing, tem correlações significativas com a localização; já quantidade de vendedores tem correlação significativa com quantidade de marcas e gasto com marketing e por último vemos que quantidade de marcas tem correlação significativa com gasto com marketing.

Tabela 05: Teste de adequação de KMO e teste de esfericidade de Bartett's dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

Teste de Kaiser-Meyer-Olkin		,691
Teste de esfericidade de Bartlett's	Aproximação por. Qui-Quadrado	60,512
	Graus de Liberade	28
	Nível descritivo (p-valor)	,00035

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004.*

Pela tabela 05 verificamos que o teste de adequação de KMO, indica que existe adequação dos dados para a análise fatorial, já que o valor é considerado como bom (Tabela 01). Já o teste de esfericidade de Bartett's resulta na rejeição de que a matriz de correlação é uma matriz identidade, pois nível descritivo é menor que nível de significância (0,00035<0,05).

Tabela 06: Comunalidades dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

	Inicial	Extração
Localização	1,000	,789
Área da loja	1,000	,762
Quantidade de vendedores	1,000	,858
Tempo de vida da loja	1,000	,606
Tempo de gerência	1,000	,576
Quantidade de marcas	1,000	,877
Gasto com Marketing	1,000	,902
Quantidade de concorrentes	1,000	,718

Método de Extração: Componentes principais.

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004.*

As comunalidades iniciais e com a extração são mostradas na tabela 06, onde percebemos que a priori cada variável explicaria a variância como mesmo peso (1,000) e que com a extração as comunalidades ficam com um poder de explicação menor que 1(um). As comunalidades após a extração ainda tem poder de explicação da variância bom, com exceção do tempo de gerência que podemos considerar satisfatório.

Tabela 07: Total de variância explicada dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

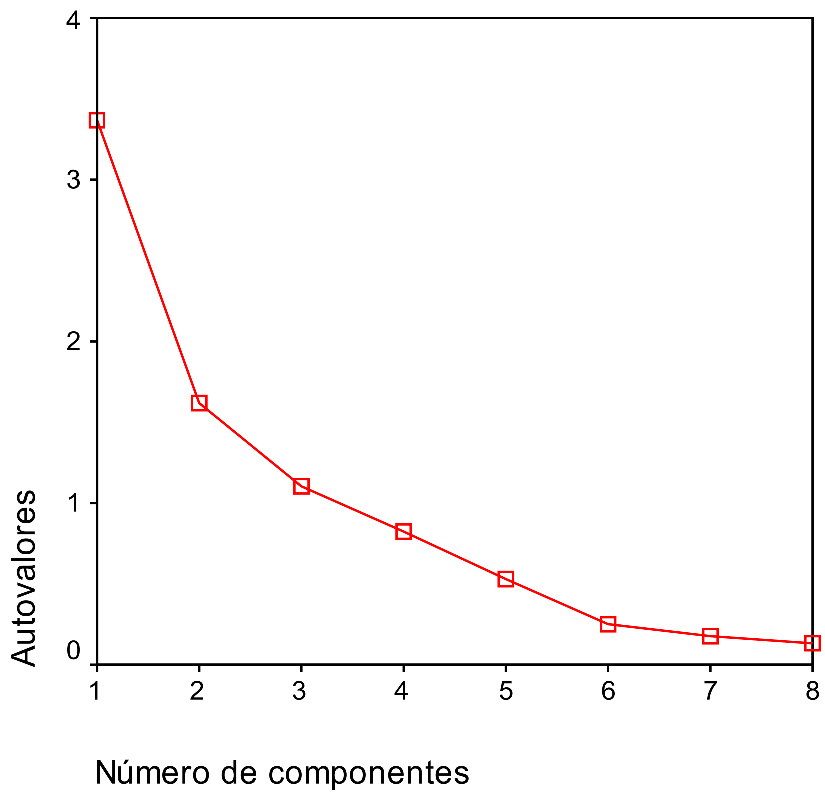
	Autovalores iniciais			EXtração de fatores		
Componentes	Total	% da Variância	Variância acumulada %	Total	% da Variância	Variância acumulada %
1	3,365	42,060	42,060	3,365	42,060	42,060
2	1,622	20,274	62,334	1,622	20,274	62,334
3	1,100	13,748	76,082	1,100	13,748	76,082
4	,820	10,253	86,335			
5	,534	6,675	93,009			
6	,253	3,160	96,169			
7	,171	2,133	98,302			
8	,136	1,698	100,000			

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004.*

O total da variância explicada por cada componente ou fator é mostrada na tabela 07, onde na primeira coluna estão os fatores, na 2ª, 3ª e 4ª colunas estão respectivamente o total da variância explicada por cada fator (a soma dessa coluna é igual a 8 (número de variáveis)); na 3ª coluna estão os valores em percentual da explicação das variáveis; na 4ª coluna estão os valores acumulados da variância. Na 5ª, 6ª e 7ª colunas estão o total da variância explicada por cada fator, os valores em percentual da explicação das variáveis e os valores acumulados da variância dos fatores com explicação maior que1(um). Foram extraídos 3(três) componentes por terem autovalores maior que 1(um).

O gráfico 02, mostra o declive comumente chamado de *scree plot*, o qual ilustra uma diminuição do grau de declive após o terceiro fator. Após o 3° fator a explicação que cada fator fornece torna-se pequena e como queremos um número de fatores que expliquem satisfatoriamente bem os dados, ficaremos com 3 primeiros fatores, que contribuem como 76,08% da variabilidade dos dados, como apresentado na tabela 06.

Gráfico 02: Gráfico de declive dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.



Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004.*

A tabela 08 indica os coeficientes que relacionam as variáveis aos 3 fatores não rotados. No fator 1, os maiores coeficientes são os das variáveis gasto com marketing, quantidade de marcas, quantidade de vendedores, localização e área da loja; no fator 2 estão quantidade de concorrentes e tempo de gerência e no fator 3 temos tempo de vida das filiais.

Tabela 08: Matriz de componentes dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

	Componentes		
	1	2	3
Gasto com marketing	,8610	,2981	,2678
Quantidade de marcas	,8488	,3045	,2516
Quantidade de vendedores	,8390	-,0987	,3797
Localização	,7929	-,2360	-,3227
Área da loja	,6387	-,5030	-,3187
Quantidade de concorrentes	-,1009	,8396	,0510
Tempo de gerência	,1385	,6112	-,4281
Tempo de vida da loja	-,3647	-,2086	,6551

Método de extração: Componentes principais.

a. 3 componentes foram extraídos

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004.*

As correlações reproduzidas e os resíduos são mostrados na tabela 09. Podemos verificar que na diagonal principal os valores foram substituídos pelas comunalidades após a extração, valores esses mostrados na tabela 06. Nos resíduos percebemos que existem 11 (39%) valores com valor absoluto superior a 0,05.

Tabela 09: Matriz de correlações reproduzidas dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

		Localização	Área da loja	Quantidade de vendedores	Tempo de vida da loja	Tempo de gerência	Quantidade de marcas	Gasto com Marketing	Quantidade de concorrentes
Correlações reproduzidas	Localização	,7885 [b]	,7280	,5660	-,4513	,1037	,5200	,5259	-,2946
	Área da loja	,7280	,7625 [b]	,4645	-,3368	-,0825	,3088	,3147	-,5030
	Quantidade de vendedores	,5660	,4645	,8578 [b]	-,0367	-,1066	,7777	,7947	-,1481
	Tempo de vida da loja	-,4513	-,3368	-,0367	,6056 [b]	-,4584	-,2083	-,2008	-,1049
	Tempo de gerência	,1037	-,0825	-,1066	-,4584	,5760 [b]	,1960	,1868	,4773
	Quantidade de marcas	,5200	,3088	,7777	-,2083	,1960	,8765 [b]	,8890	,1829
	Gasto com Marketing	,5259	,3147	,7947	-,2008	,1868	,8890	,9020 [b]	,1771
	Quantidade de concorrentes	-,2946	-,5030	-,1481	-,1049	,4773	,1829	,1771	,7177 [b]
Resíduos [a]	Localização		-,0446	-,0714	,1666	,0542	-,0294	,0029	,0360
	Área da loja	-,0446		-,0594	,1023	-,0249	-,0181	,0078	,1701
	Quantidade de vendedores	-,0714	-,0594		-,0973	,0209	-,0388	-,0409	-,0497
	Tempo de vida da loja	,1666	,1023	-,0973		,3033	-,0532	,0341	-,0189
	Tempo de gerência	,0542	-,0249	,0209	,3033		-,0488	,0304	-,2237
	Quantidade de marcas	-,0294	-,0181	-,0388	-,0532	-,0488		-,0609	-,0245
	Gasto com Marketing	,0029	,0078	-,0409	,0341	,0304	-,0609		-,0257
	Quantidade de concorrentes	,0360	,1701	-,0497	-,0189	-,2237	-,0245	-,0257	

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004.*

Tabela 10: Matriz de componentes rotados dos dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de 2004.

	Componentes		
	1	2	3
Gasto com marketing	,9308	,0664	,1763
Quantidade de marcas	,9146	,0631	,1897
Quantidade de vendedores	,8613	,3033	-,1545
Localização	,2628	,8284	,0850
Área da loja	,4607	,7033	,2857
Quantidade de concorrentes	,1557	-,7025	,4472
Tempo de gerência	,0769	-,1988	,7284
Tempo de vida da loja	-,0588	-,3151	-,7091

Método de extração: Componentes principais e método de rotação Varimax.

Fonte: ***Dados calculados do levantamento feito sobre variáveis que explicam o faturamento das filiais do Atacadão dos Eletros, no ano de*** *2004.*

A tabela 10 mostra a matriz de componentes rotadas, pelo método de rotação varimax. A tabela 10 indica os coeficientes que relacionam as variáveis aos 3 fatores rotados. No fator 1, os maiores coeficientes são os das variáveis gasto com marketing, quantidade de marcas e quantidade de vendedores; no fator 2 estão localização, área da loja e quantidade de concorrentes e no fator 3 temos tempo de vida e tempo de gerência.

5. DISCUSSÃO

Utilizando a técnica de agrupamentos, tivemos primeiro a formação dos grupos e a separação dos casos (filiais) em três grupos distintos. Onde no primeiro grupo ficaram as filiais B, D, F e S, no segundo C, E, J, M e Q e no terceiro A, G, H, I, K, L, N, O, P e R. Com a análise da média dos centróides (tabela 03), caracterizamos os grupos da seguinte forma:

- **Grupo I**: Formado pelas filiais de localização menos privilegiadas, menor investimento com marketing, um número menor de variedades nas marcas, menor número de vendedores, menor tempo de experiência do gerente no Atacadão dos Eletros e menor número de concorrentes.
- **Grupo II**: Formado pelas filiais de melhor localização, maior investimento com marketing, um número maior de variedades nas marcas, maior número de vendedores, maior tempo de experiência do gerente no Atacadão dos Eletros e maior número de concorrentes.
- **Grupo III**: Formado pelas filiais que estão justamente, com médias de centróides entre o grupo I e grupo II.

Essa caracterização foi comparada com o *ranking* das filais em relação ao faturamento no ano de 2004. Quando comparadas verificou-se que o grupo I é o grupo formado pelas filiais de menor faturamento, as formadas no grupo II, com o maior faturamento, já as do terceiro grupo são as consideradas intermediárias.

A análise fatorial fornece a informação, que com base nas correlações, pode-se dizer que as variáveis área da loja, quantidade de vendedores, quantidade de marcas e gasto com marketing, tem correlações significativas com a localização; já quantidade de vendedores tem correlação significativa com quantidade de marcas e gasto com marketing e por último vemos que quantidade de marcas tem correlação significativa com gasto com marketing.

O teste de esfericidade de Bartett's resulta na rejeição de que a matriz de correlação é uma matriz identidade e o teste de adequação de KMO, indica que existe adequação dos dados para a análise fatorial.

As comunalidades após a extração ainda tem um bom poder de explicação da variância.

Ao se observar o total de variância explicada pelos autovalores iniciais e pelo gráfico de declive, foram extraídos 3(três) componentes por terem autovalores maiores que 1(um).

Os resíduos causados pela substituição das correlações iniciais pelas correlações reproduzidas apresentaram 11 resíduos não redundantes (39%) com valor absoluto superior a 0,05, que é menor que o número "*máximo*" permitido (50%), ou seja, mais um argumento que fundamenta a utilização da análise fatorial.

Além disso podemos dizer que os coeficientes que relacionam as variáveis aos 3 fatores rotados são: no fator 1, os maiores coeficientes são referentes as variáveis: gasto com marketing, quantidade de marcas e quantidade de vendedores; no fator 2, estão as variáveis: localização, área da loja e quantidade de concorrentes e no fator 3, temos tempo de vida e tempo de gerência.

6. CONSIDERAÇÕES FINAIS

Com a conclusão do estudo podemos relatar que a separação das filiais, em 3(três) grupos distintos ocorreu de modo que um grupo ficou com as 4(quatro) filiais, que estavam em 2004, com o menor desempenho em relação ao faturamento; em outro grupo, as 5(cinco) filiais com maior desempenho, e no terceiro grupo, as 10(dez) filiais que tiveram um desempenho intermediárias.

Também foi verificado pelos componentes principais, que com 3(três) fatores a variação explicada atinge 76,08% (o qual é considerada uma boa explicação), sendo 42,06% no primeiro fator; 20,27% no segundo fator e 13,75% no terceiro. Observamos que no primeiro fator as variáveis que mais contribuem para explicação da variabilidade dos dados são: investimento com marketing, variedade das marcas em exposição e quantidade de vendedores. O segundo fator tem maior contribuição das variáveis: quantidade de concorrentes, localização e área da loja. E no terceiro fator as variáveis: tempo de gerência e tempo de vida da loja são as que têm maiores cargas, com isso podemos constatar que as variáveis do terceiro fator são as que menos contribuem para explicação da variabilidade do conjunto de dados.

7. REFERÊNCIAS

BARROSO, L. P., Artes, R. Análise Multivariada. In Simpósio de Estatística Aplicada à Experimentação Agronômica, 1996. Lavras.

FURTADO, D. F. **Análise Multivariada.** Lavras: Departamento de Ciências Exatas, 1996. Paginação irregular.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, **NBR 6023:** Informação e documentação: referências: elaboração. Rio de Janeiro, 2002.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, **NBR 6027:** Informação e documentação: sumário: apresentação. Rio de Janeiro, 2003.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, **NBR 10520:** Informação e documentação: citações em documentos: apresentação. Rio de Janeiro, 2002.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, **NBR 14724:** Informação e documentação: trabalhos acadêmicos: apresentação. Rio de Janeiro, 2002.

GAMA, M. P. **Base da análise de grupamentos**. 1980. Dissertação (Mestrado em Estatística) – Departamento de Estatística, Universidade de Brasília, Brasília, 1980.

IBGE. **Normas de apresentação tabular**. 3ºed. Rio de Janeiro, 1993.

MOREIRA, A. M. **Metodologia para definir padrões pluviométricos**. 1985. Dissertação (Mestrado em Estatística) – Departamento de Estatística, Universidade de Brasília, Brasília,1985.

NASCIMENTO, J. A. **Análise Fatorial**: notas de aulas, 2005**.** João Pessoa: Departamento de Estatística, 2005.