

TT34

**USO DE CRITÉRIOS TÉCNICOS PARA AGRUPAMENTO DE BAIROS DE  
QUALIDADE DE LOCALIZAÇÃO SIMILARES -CLUSTER**

**MAURO CELSO VICENTE COELHO**

ENG CIVIL (MACKENZIE/SP-1.981), MESTRANDO NA POLI-USP EM REAL ESTATE, SÓCIO DIRETOR DA  
PERITENG ENG DE AVALIAÇÕES, ATUANDO, DESDE O ANO DE 1983, NO RAMO DE PERÍCIAS E  
AVALIAÇÕES DE IMÓVEIS E ECONÔMICA DE EMPREENDIMENTOS. CREDENCIADO CEF E BB

## **XIV COBREAP – CONGRESSO BRASILEIRO DE ENGENHARIA DE AVALIAÇÕES E PERÍCIAS. IBAPE/BA**

### **NATUREZA DO TRABALHO – PROFISSIONAL**

#### **Resumo**

*O presente trabalho se propõe a apresentar uma metodologia científica que pode ser utilizada para classificação de bairros ou regiões de uma cidade, possibilitando a estratificação da mesma em zonas de qualidade de localização similares, que pode ser entendido como sendo regiões de equivalência “geo-economica”, importando numa drástica redução no número de modelos necessários para modelagem do mercado imobiliário, um procedimento que tem como resultante a utilização de modelagem em campos amostrais mais homogêneos, sendo de grande valia para os profissionais que necessitam proceder a avaliações em massa, ou estudos macroeconômicos do mercado.*

**Palavras-Chave:** *Aglomeração; Avaliação em massa; Região geo-economica, Cluster, Inferência estatística.*

## INDICE

1) Introdução .....	2
2) O que é Cluster?.....	4
3) Objetivos da Análise de Cluster.....	6
4) Conceitos Associados a Análise de Cluster .....	6
5) A Escolha da Variáveis para Aglomeração de Bairros.....	9
6) Estudo de caso.....	11
7) Aplicação da técnica numa modelagem.....	18
8) Conclusões finais – recomendações.....	21
9) Bibliografia .....	22
Referenciada.....	22
Consultada.....	23
Link das Tabelas.....	23
10) Anexo.....	24

### 1) Introdução

Tem-se como princípio básico no processo de avaliação de bens que essa valoração seja feita por comparação com outros bens de características similares, sendo desejado que essa similaridade seja a mais abrangente possível.

No caso de bens imóveis um dos aspectos mais importantes na sua caracterização diz respeito especificamente a sua localização, por muitos considerados como um dos principais fatores de influencia no processo de precificação de um imóvel.

Com efeito, Pascale (2005) discorrendo sobre as motivações para mudança de domicilio destaca a importância da localização nessa escolha, atribuindo à localização e seus atributos o principal papel na escolha final da aquisição.

Nesse sentido também se manifesta Hermann (2003) ao justificar a necessidade de um embasamento teórico para elaboração de modelo hedônico para habitação, ou vejamos:

*“Um dos motivos é o papel determinante da localização na formação de preços de imóveis e aluguéis”.*

Assim é que na utilização do método comparativo de dados de mercado o conceito geral dos trabalhos avaliatórios preconiza que a coleta de dados seja restrita a uma mesma região dita: “geo-economica”.

De fato Dantas(1998) define esse método como sendo;

*“aquele em que o valor do bem é estimado através da comparação com dados de mercado assemelhados quanto às características intrínsecas e extrínsecas”.*

Nas avaliações isoladas por metodologia tradicional ou por inferência estatística considera-se suficiente a delimitação do campo amostral da pesquisa comparativa ao entorno do imóvel avaliando, porém nas avaliações em massa onde se pretende definir equações que simulem o comportamento do mercado imobiliário possibilitando a avaliação de um espectro de imóveis, torna-se imperativo uma definição mais precisa com relação ao que se entende como “mesma região geo-economica”, e que essa possa ser feita por metodologia científica.

Dividir a cidade por regiões (exemplo: norte, sul, leste, oeste) não se mostra um procedimento dos mais adequados, em função da heterogeneidade que pode existir dentro dessas regiões sob vários outros aspectos de atributos, não se podendo garantir que a simples proximidade geográfica importe em equivalência de qualidade de localização.

Nesse sentido, Fávero (2003) apud SARTORIS NETO (1996), confere a questão de regionalização como de pouca importância, ou vejamos:

*“poder-se-ia pensar em uma divisão geográfica, porém a mesma ‘seria descabida, pois seria supor que a zona sul de São Paulo teria, por si só, algum grau de homogeneidade’ ”.*

Diante desse contexto e considerando o crescimento do interesse pela realização de trabalhos de avaliação em massa, cada vez mais difundidos entre os profissionais de avaliação, a delimitação de campo amostral se apresenta como uma questão que merece melhor definição para que a mesma seja feita por metodologia científica.

O que anteriormente era, praticamente, de interesse das empresas ou departamentos de prefeitura designados para elaboração de plantas de valores dos municípios, agora se apresenta como um procedimento corriqueiro aos profissionais que necessitam realizar grandes números de avaliações de imóveis, de tipologia padronizadas, onde a modelagem matemática do mercado imobiliário se mostra como a ferramenta mais prática e adequada.

Com efeito, basta citar que a Caixa Econômica Federal, a partir de seu edital de 2.006, passou a exigir a realização de suas avaliações de garantia de financiamentos através de modelos de inferência, o que torna o serviço de avaliação muito trabalhoso se o profissional pretender montar um modelo específico para cada avaliação a ser realizada.

A grande vantagem na utilização da inferência estatística é exatamente simplificar o comportamento do mercado, simulando-o através de modelos matemáticos em função de certas variáveis, portanto em tese não haveria restrições quanto a região amostral, desde que as diferenças estivessem contempladas no modelo.

De qualquer forma mesmo se admitindo que a realização dessa inferência fosse abrangente (uma equação única para todo município), seria necessário fazer uma distinção das regiões da cidade, onde poderia ser feita uma leitura por bairros, o que exigiria a consideração de grande número de variáveis dicotômicas<sup>1</sup>, sem falar na grande quantidade de elementos comparativos que seriam necessários em número suficiente em todos os bairros da cidade.

Especificamente ao aspecto econômico do campo amostral, se depreende que a pesquisa deveria ficar adstrita a regiões de características econômicas similares, o que confere um certo aspecto de que essa estratificação deveria observar uma distribuição sócio-econômica dos moradores da região.

No entanto, outras questões se mostram importantes em termos de atributo de localização que simplesmente seriam desprezadas com a simples divisão de regiões similares economicamente.

É assim que se mostra cabível a elaboração de uma metodologia de classificação dos bairros ou regiões da cidade de forma que se possam fazer a

---

<sup>1</sup> Nesse caso a dicotomia seria, estar no bairro ou não, assumindo valores 1 e 0, respectivamente.

definição de campos amostrais mais homogêneos e com a vantagem de se obter um numero menor de grupos que apresente entre si certa similaridade, de forma a que possam ser consideradas como uma região mais homogênea possível sob diversos aspectos, procedimento que se mostra viável através da utilização da estatística de “CLUSTER”, que é o tema do presente trabalho.

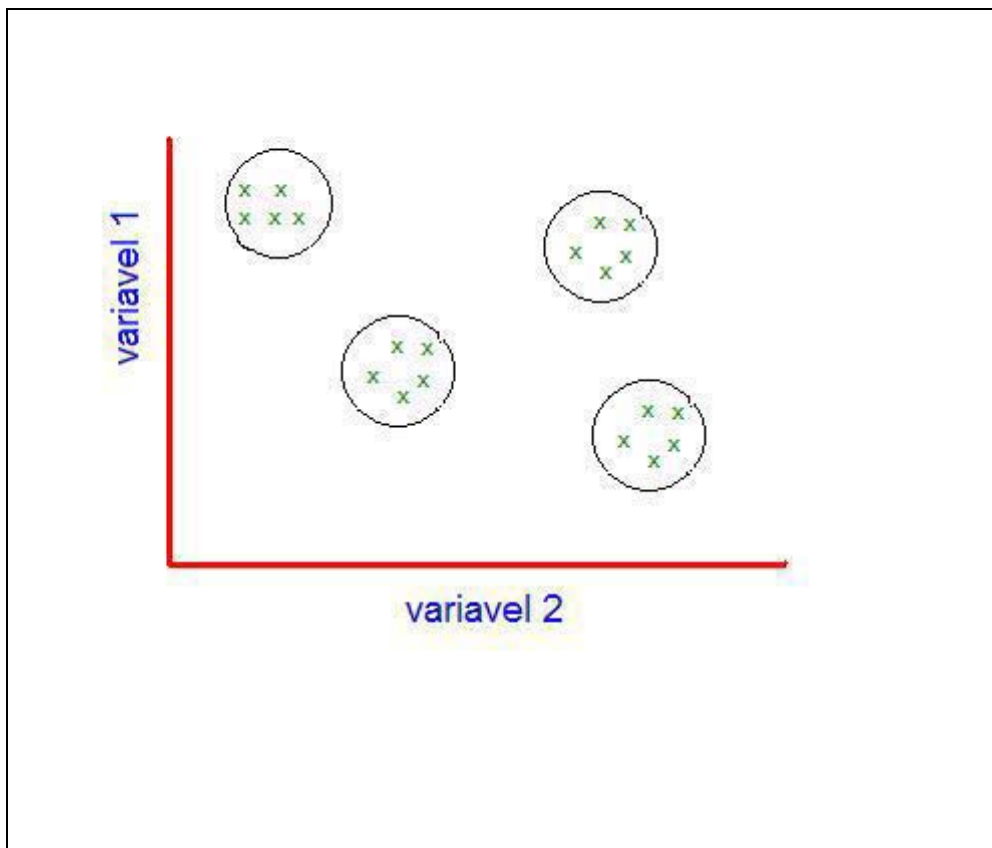
## 2) O que é Cluster?

Cluster por definição é um aglomerado de coisa ou objetos com algum grau de semelhança entre si dentro de determinadas variáveis consideradas.

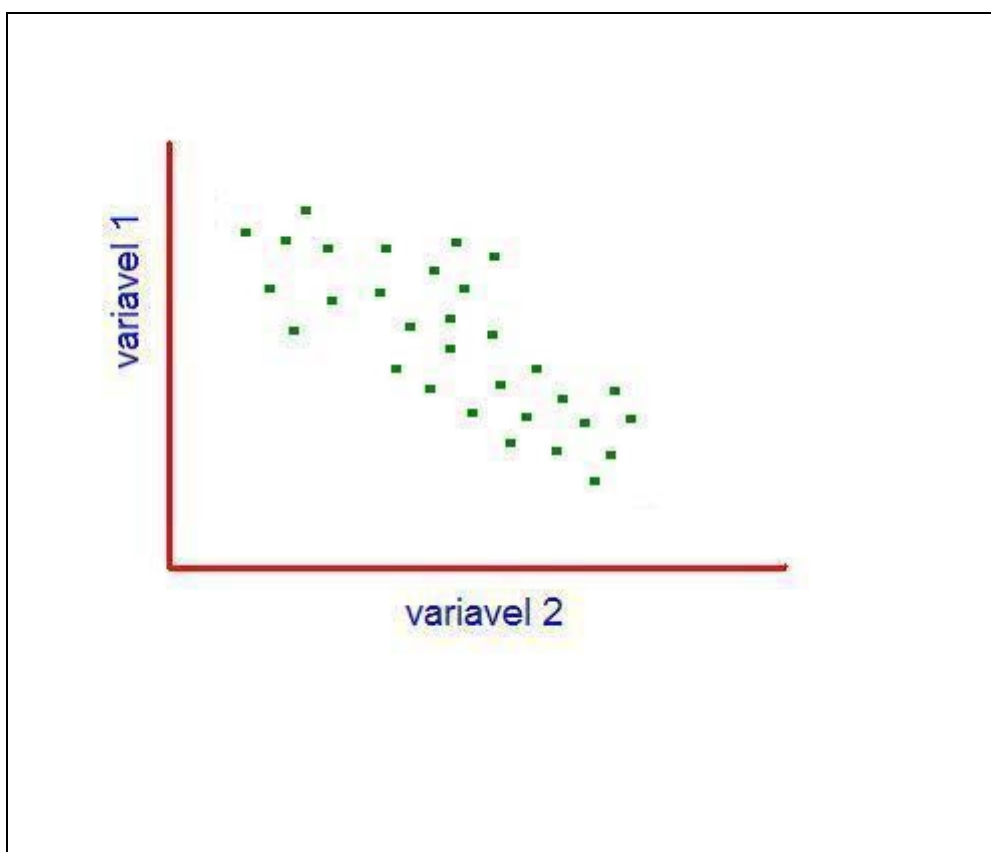
Malhotra (2006) explica que essa análise também pode ser chamada de “Análise de Classificação” ou “Taxonomia numérica”.

A montagem dos cluster dependem da escolha das variáveis que deverão ser objeto de análise e que definirão as semelhanças, ou seja, as regras de classificação. Em resumo, os objetos ou coisas são semelhantes em alguns critérios que os caracterizam.

No exemplo seguinte mostramos um conjunto de elementos que são caracterizados pelas variáveis V1 e V2, mostrada na figura abaixo.



Na pratica, porém os clusters não se mostram tão evidentes e bem definidos como no exemplo acima, apresentando certa dificuldade para definição das fronteiras e do numero de clusters, como se mostra na figura abaixo.



Na cidade de São Paulo, o SECOVI-SP<sup>2</sup> realiza mensalmente uma pesquisa de alugueis de apartamentos residenciais diferenciado por numero de dormitórios e regiões da cidade, que é assim estratificada:

- Centro: Barra Funda, Bom Retiro, Cambuci, Centro, Liberdade, Pari, Santa Cecília;
- Leste-A: Alto da Moóca, Belém, Moóca, Tatuapé;
- Leste-B: Artur Alvim, Brás, Cangaíba, Ermelino Matarazzo, Itaim Paulista, Itaquera, Jardim Aricanduva, Penha, São Mateus, São Miguel Paulista, Sapopemba, Vila Carrão, Vila Formosa, Vila Matilde, Vila Prudente;
- Norte: Brasilândia, Casa Verde, Freguesia do Ó, Limão, Mandaqui, Pq Edu Chaves, Santana, Tremembé, Tucuruvi, Vila Guilherme, Vila Mazzei, Vila Nova Cachoeirinha;
- Oeste-A: Alto da Lapa, Alto de Pinheiros, Lapa, Perdizes, Pinheiros, Pompéia, Sumaré, Vila Leopoldina, Vila Madalena;
- Oeste-B: Butantã, Jaguaré, Perus, Pirituba;
- Sul-A: Aclimação, Bosque da Saúde, Chácara Santo Antônio, Higienópolis, Itaim, Jardim da Saúde, Jardins, Moema, Morumbi, Paraíso, Real Parque, Saúde, Vila Mariana, Vila Olímpia;
- Sul-B: Campo Limpo, Cidade Ademar, Interlagos, Ipiranga, Moinho Velho, Pedreira, Sacomã, São João Clímaco, Vila das Mercês, Vila Gumercindo.

<sup>2</sup> SECOVI-SP – Sindicato das Empresas de Compra, Venda, Locação e Administração de Imóveis Residenciais e Comerciais de São Paulo

Essas divisões efetivamente se constituem de “clusters” de bairros de padrão de valor locativo considerados equivalentes ou que apresentam níveis similares de valor, o problema nesse caso é que, salvo melhor juízo, esse parcelamento foi realizado de forma empírica e baseada em prática do mercado, portanto carece de fundamento técnico que justifique cada agrupamento.

### **3) Objetivos da Análise de Cluster**

Diferente da inferência estatística, na análise de “cluster” não se faz distinção entre variáveis dependente e independente já que o objetivo principal é a classificação dos objetos em grupos relativamente homogêneos sob o aspecto de determinadas variáveis.

Como foi acima exposto, basicamente ao juntarmos os elementos da amostra em grupos, pretende-se principalmente reduzir e simplificar a quantidade de campos de amostragem, conseqüentemente a quantidade de modelagens que se pretende produzir para simular o mercado imobiliário.

Em São Paulo- SP, por exemplo, a prefeitura apresenta dados estatísticos referentes a 96 distritos em 13 regionais. Assim, sem a classificação criteriosa dos grupos de distritos que apresentam características similares entre si, para estudar esse mercado dever-se-ia montar um número de, no mínimo, 96 modelos de econometria, pois se considerar que dentro de mesmo bairro pode ocorrer condições econômicas diversas esse número de simulações ainda aumentaria.

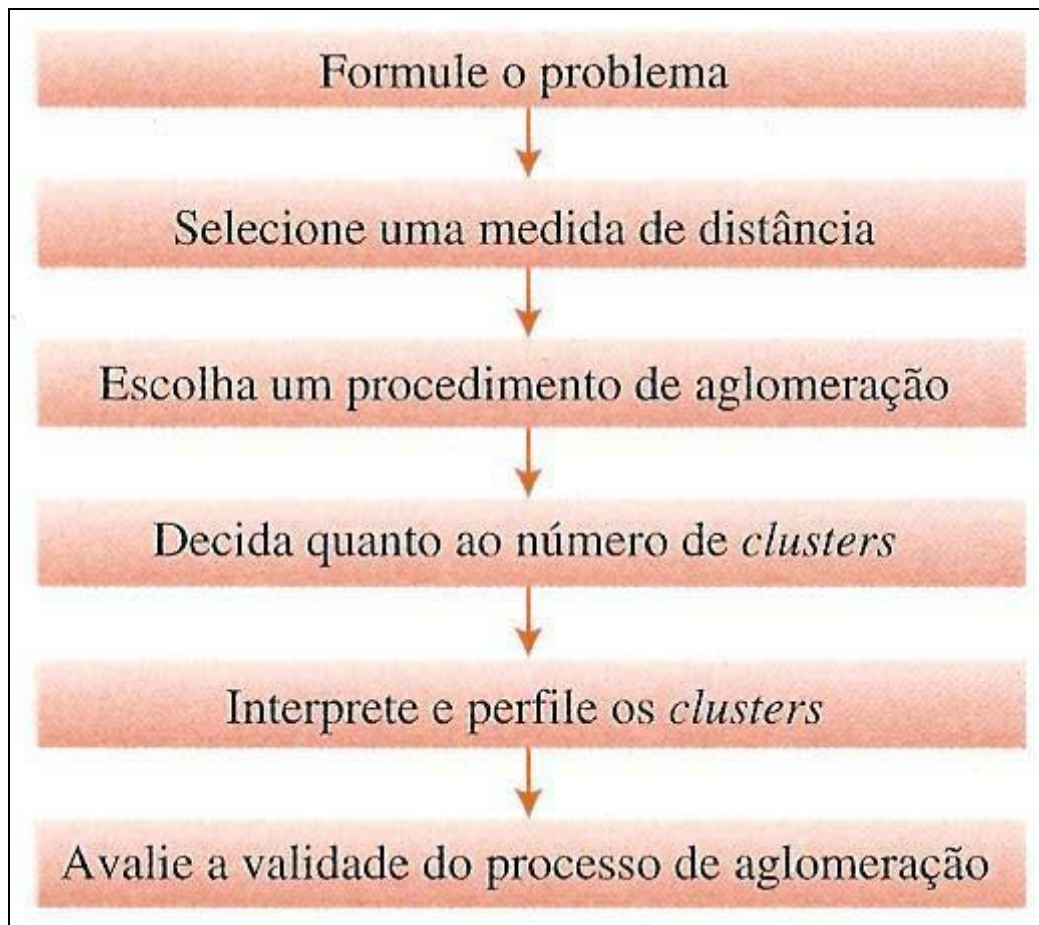
Com a estatística de cluster é possível se aglomerar bairros ou regiões de características similares entre si, dentro de determinados critérios tecnicamente comprovados, de forma que se possa estudar o mercado do município através de um número bem menor de conjuntos.

É importante consignar que o fato de proceder a aglomeração de bairros ou regiões para delimitação de um campo amostral homogêneo não significa a eliminação do uso de outras variáveis extrínsecas que podem e devem integrar a montagem da inferência estatística, em aspectos mais específicos e particulares, tais como: distância a pólos, qualidade ambiental, etc, fatores que evidentemente não foram objeto de critério de classificação dos clusters, que assim.

### **4) Conceitos Associados a Análise de Cluster**

Serão apresentados alguns conceitos mais importantes para conhecimento básico da metodologia a fim de possibilitar a compreensão dos cálculos e procedimentos que serão mostrados no estudo de caso apresentado mais adiante.

Recomenda-se que a análise de cluster seja feita observando as seguintes etapas, como ensina Malhotra (2006)



Obs: extraído de Malhotra (2006) – fls 575

As etapas ensejam as seguintes explicações:

- **Formulação do problema:** definir as variáveis que irão se definir a aglomeração;
- **Selecione uma medida de distancia:** As medidas de cada variável devem ser equivalentes entre si, tais como apresentar medidas em escala “Likert”<sup>3</sup> com numero de pontos constantes; Caso não seja possível deve-se normalizar as variáveis para que tenham a mesma média (exemplo, 0 ou 10) e desvio padrão 1; A não verificação dessa recomendação pode comprometer totalmente o trabalho, pois as medidas de distancia entre os pontos normalmente é feita por calculo euclidiano<sup>4</sup>;
- **Procedimento de aglomeração:** São vários os procedimentos de aglomeração, classificados em 2 grandes grupos, hierárquicos e não-hierárquicos; dos métodos existentes um dos mais utilizados é o método de “Ward”, que faz parte do grupo “Hierárquico”, consiste num método de variância que procura minimizar o quadrado da distancia euclidiana às médias dos aglomerados;

<sup>3</sup> Tipo de escala qualitativa ou código alocado, do tipo: 1 = satisfeito; 2 = não insatisfeito; 3 = pouco satisfeito; 4 = insatisfeito; 5 = muito insatisfeito.

<sup>4</sup> Distancia euclidiana = raiz quadrada da soma dos quadrados das diferenças de valores para cada variável.



- **Decidir sobre o numero de clusters:** trata-se de uma questão que não apresenta regra fixa, cabendo ao analista um prévio conhecimento da quantidade de aglomerados que espera ter em média. Malhotra (2006), no entanto faz algumas recomendações de algumas diretrizes que podem ser seguidas para essa determinação:
  - *Considerações teóricas ou práticas podem sugerir um certo numero de clusters;*
  - *Nas aglomerações hierárquicas as distancias em que os clusters são combinados pode ser utilizado como critério limitador de um conjunto;*
  - *No cluster não-hierarquico, a razão da variancia entre grupos pode ser grafada em função do numero de clusters;*
  - *Tamanho relativo da quantidade de elementos em cada agrupamento, simplesmente contando-se a freqüência de objetos em cada um, procurando-se uma conformação onde haja equilíbrio entre as quantidades, não tendo, em principio, que um ou mais clusters fique por exemplo com um só elemento;*
- **Interpretar e perfilar os resultados:** Interpretar e traçar o perfil de cada aglomerado em função do exame do seu centróide<sup>5</sup>, podendo servir de critério para rotulação dos grupos<sup>6</sup>.

No estudo de caso que se pretende ilustrar a aplicação dessa metodologia, será utilizado um procedimento hierárquico pelo método de “Ward”, que basicamente funciona da seguinte maneira:

- Digamos que temos um conjunto de 20 objetos;
- Calcula-se a distancia euclidiana entre os objetos e se agrupa os dois com menor distancia, passando a ter nesse momento 2 clusters, um contendo esses dois objetos de menor distancia e os demais objetos;
- Num segundo estagio compara-se novamente as distancias euclidianas, lembrando que agora já temos um cluster formado na etapa anterior que agora será representado pelo seu centróide (valor médio dentro do grupo);
- Procede-se desta forma até que todos os elementos formem a quantidade de cluster pré-determinada e todos os elementos sejam analisados e agrupados. No caso do exemplo com 20 objetos, realiza-se, portanto 19 etapas de aglomeração.

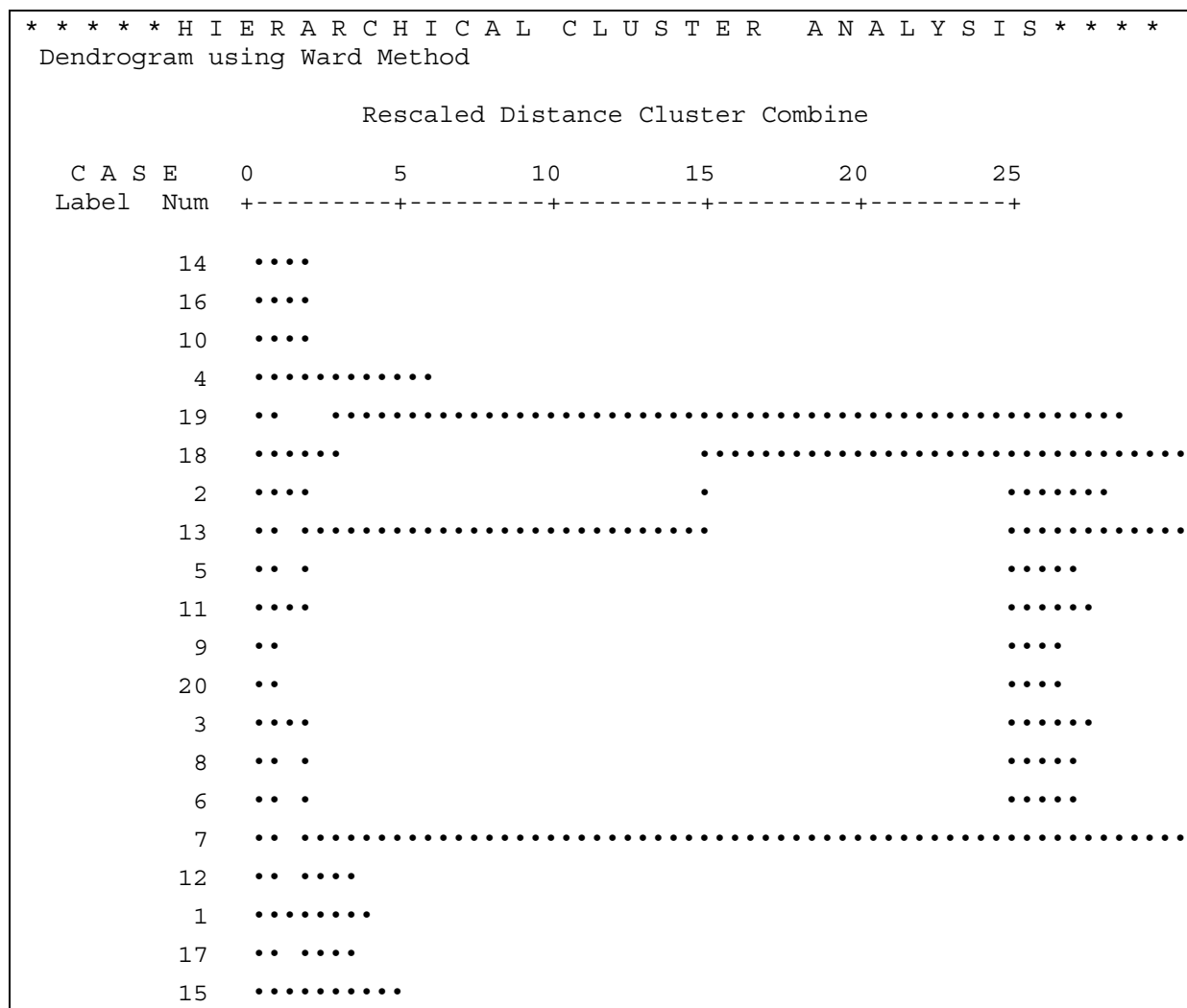
Um dispositivo gráfico muito utilizado e que ajuda também a enxergar a quantidade de clusters mais adequada é o dendograma, que deve ser lido da esquerda para a direita, com as linhas verticais representa os cluster que são unidos e na reta a escala da distancia da aglomeração.

Um exemplo de dendograma é mostrado na figura abaixo:

---

<sup>5</sup> Centróides representam os valores médios dos objetos contidos no cluster.

<sup>6</sup> Exemplo de rotulação: bairro de luxo; bairro médio; bairro simples; bairro popular.



OBS: tela de saída do SPSS<sup>7</sup>

## 5) A Escolha das Variáveis para Aglomeração de Bairros

Para podermos entender como se deve proceder para a escolha dos critérios ou variáveis que serão utilizadas na classificação dos bairros ou regiões, devemos reconhecer o funcionamento do mercado no que diz respeito a qualidade de localização dos imóveis.

De maneira geral é interessante como se pode constatar que quando da montagem de modelagem de mercado por estatística inferencial, se ocupa com muita ênfase no detalhamento das qualidades intrínsecas dos imóveis (área, padrão, idade, topografia, etc), relegando às características extrínsecas para um segundo plano, o que representa verdadeira inversão de valores, pois como foi dito na introdução deste trabalho, em geral a localização é o fator principal na escolha dos imóveis por parte dos compradores.

Com efeito, Pascale (2005) destaca a importância da localização como vantagem competitiva para os empreendimentos, reiterando que as prioridades dos

<sup>7</sup> SPSS (acrônimo de Statistical Package for the Social Sciences - pacote estatístico para as ciências sociais) – fonte - <http://pt.wikipedia.org/wiki/SPSS>

potenciais compradores de imóveis pela escolha de melhores localizações são feitas mesmo em detrimento de perda de qualidade das características intrínsecas dos imóveis.

Em resumo, o que pode se reconhecer é que o público consumidor de imóveis, em regra geral, prefere adquirir um imóvel mais modesto, porém com uma localização de melhor qualidade.

Pascale (2005) apresenta então uma matriz que procura traduzir os grupos de atributos que devem ser utilizados como critério para a realização de uma análise de classificação das localizações residenciais, estabelecido por cinco conjuntos de atributos, quais sejam:

- Qualidade Ambiental e Vizinhança: aspectos como existência de parques, centros esportivos, níveis de poluição, dentre outros;
- Acessibilidade, sistema viário e transporte: facilidades de acesso da região aos locais de trabalho e lazer relacionados diretamente com o uso residencial;
- Comércio e Serviços: qualidade e quantidade desse tipo de serviço no entorno;
- Infra-estrutura urbana: presença e qualidade dos melhoramentos públicos;
- Aspectos Sócio – econômicos: nível social da vizinhança, presença de favelas e índices de violência.

Assim, os atributos acima poderiam ser utilizados como variáveis na montagem dos “clusters”, porém há de ser ressaltar que em princípio o presente trabalho não tem como objetivo indicar quais seriam as variáveis a serem consideradas para agrupamento das regiões ou bairros, mas sim divulgar a técnica da estatística de “cluster” aplicada com essa finalidade.

Desta forma não entraremos no mérito de quais atributos acima devem compor as variáveis de classificação, podendo ser essas acima indicadas ou outras que os avaliadores entenderem cabíveis para cada município com suas especificidades.

Poderíamos sugerir uma relação de variáveis a serem consideradas como listada abaixo:

- Espectro de renda dos moradores do bairro;
- Melhoramentos públicos;
- Quantidade e tipo de comércio e serviços disponíveis no bairro;
- Distância ao centro;
- Região da cidade;
- Qualidade do transporte coletivo;
- Existência de escolas;
- Densidade ocupacional;
- Conceito do bairro (perigoso, tranquilo, seguro, etc);
- Proximidade com parques;
- Proximidade com shopping center;
- Hospitais e postos de saúde.

Desta forma o estudo de caso apresentado a seguir tem uma finalidade meramente ilustrativa de como se pode ser aplicada essa metodologia estatística de análise de cluster.

## 6) Estudo de caso

A escolha das variáveis para aplicação da estatística de cluster deve ser feita de maneira criteriosa e pode ser feita em vários níveis, porquanto em termos de localização residencial os critérios de qualidade são muito diferenciados para diversos níveis sociais.

Enquanto é importante a existência de comércio farto no entorno de residências de classe média baixa, nos bairros mais nobres tem-se preferência pela inexistência de qualquer tipo de atividade comercial ou serviço, pois nessa classe de renda as compras são normalmente realizadas em locais distantes mediante utilização de automóvel.

Nesse sentido Fávero (2003) realizando estudo sobre preços hedônicos de apartamentos na cidade de São Paulo conclui que para cada faixa de renda existem diferentes variáveis que agregam valor na questão das características extrínsecas, como se mostra nos quadros abaixo:

TABELA 4.5.2 – DEFINIÇÃO DAS VARIÁVEIS HEDÔNICAS

VARIÁVEL	DEFINIÇÃO DA VARIÁVEL
COLP	Colégio Particular no Distrito = 1 ; Não = 0
SHOP	Shopping Center no Distrito = 1; Não = 0
METR	Estação de Metrô no Distrito = 1; Não = 0
HOSP	Hospital Particular no Distrito = 1; Não = 0
PARQ	Parques e Áreas Verdes no Distrito = 1; Não = 0
FATP	Fator Principal do Distrito

O quadro 5.4.2 a seguir possibilita a elaboração da análise da importância relativa das variáveis hedônicas, para cada faixa de renda e entre as mesmas, uma vez que os coeficientes apresentados estão padronizados:

QUADRO 5.4.2 – COEFICIENTES PADRONIZADOS PRA CADA VARIÁVEL HEDÔNICA<sup>28</sup>

Variável	Renda Baixa	Renda Média	Renda Alta
COLP	0,055		
SHOP		0,039	
METR	0,056	0,209	
HOSP	-0,080		
PARQ		0,071	0,198
FATP	0,249	0,161	0,162

OBS: quadros extraídos de Fávero 2003.

Assim não se mostra adequado realizar uma análise de cluster para diferentes níveis sociais, pois as variáveis de importância são distintas para cada caso. Realizamos então uma estratificação que denominamos Cluster – Nível 1, onde os bairros são agrupados por um único critério de classe de renda.

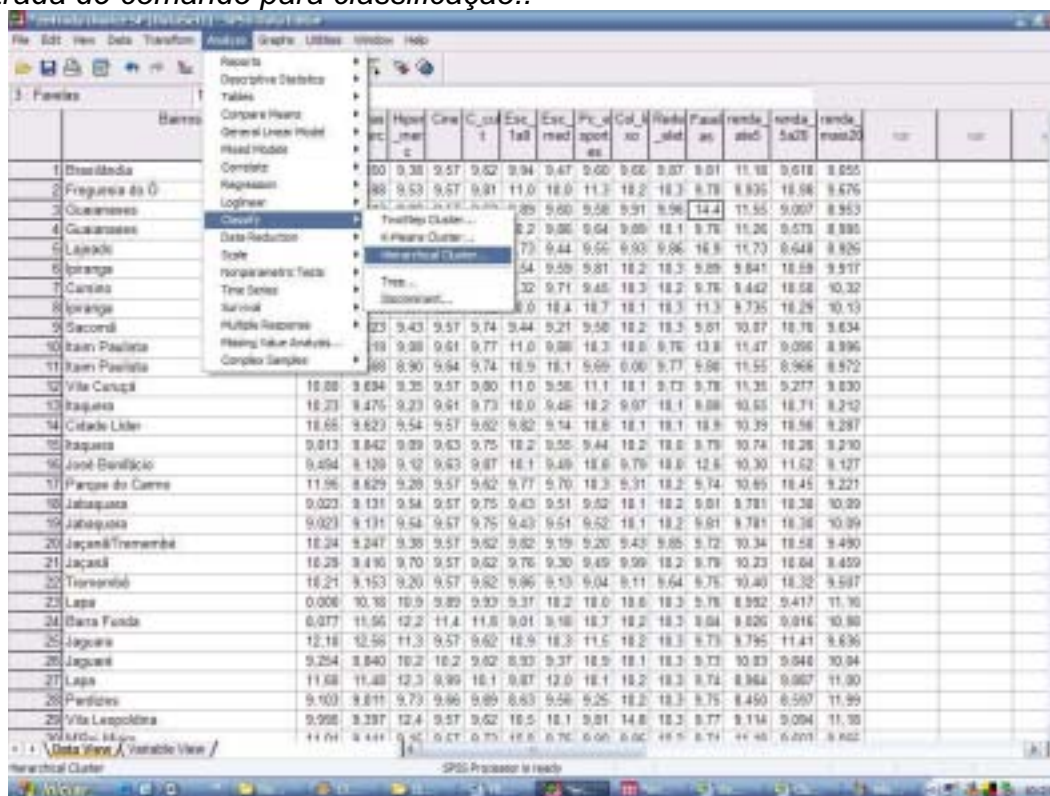
Tomamos os dados constantes do site da prefeitura de São Paulo<sup>8</sup> que indica a distribuição da renda nos 96 distritos<sup>9</sup> ou bairros que compõem o município, e que foram assim considerados em termos de variáveis<sup>10</sup>:

- Variável\_1: porcentagem de moradores com renda até 5 salários mínimos;
- Variável\_2: porcentagem de moradores com renda entre 5 a 20 salários mínimos;
- Variável\_3: porcentagem de moradores com renda maior que 20 salários mínimos.

As variáveis foram normalizadas para não resultarem em distorção de medidas, de forma que os conjuntos de dados representem uma média normalizada de 10 e um desvio padrão 1.

Utilizamos o pacote estatístico SPSS para classificação através da técnica discutida no capítulo anterior, como mostramos na seqüência de telas abaixo:

*Mostra-se a tela com as variáveis e objetos (bairros) que compõem o estudo e a entrada de comando para classificação..*

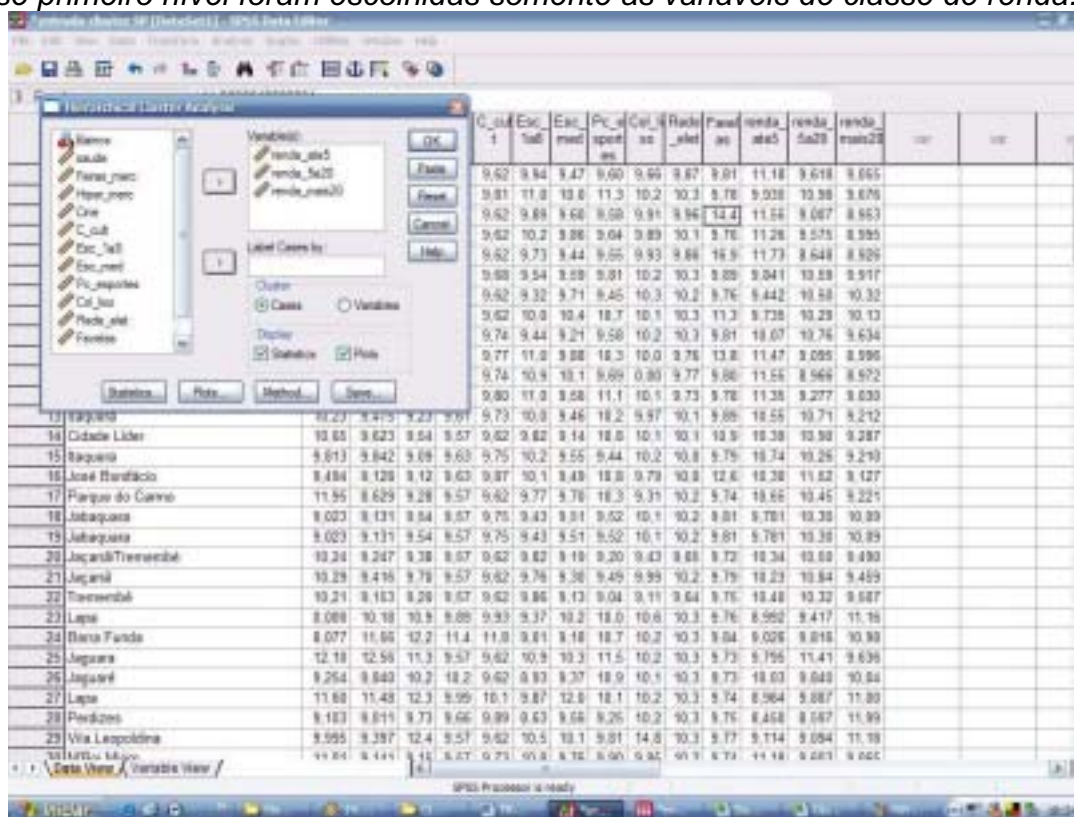


<sup>8</sup> www.prefeitura.sp.gov.br

<sup>9</sup> Vide mapa em anexo

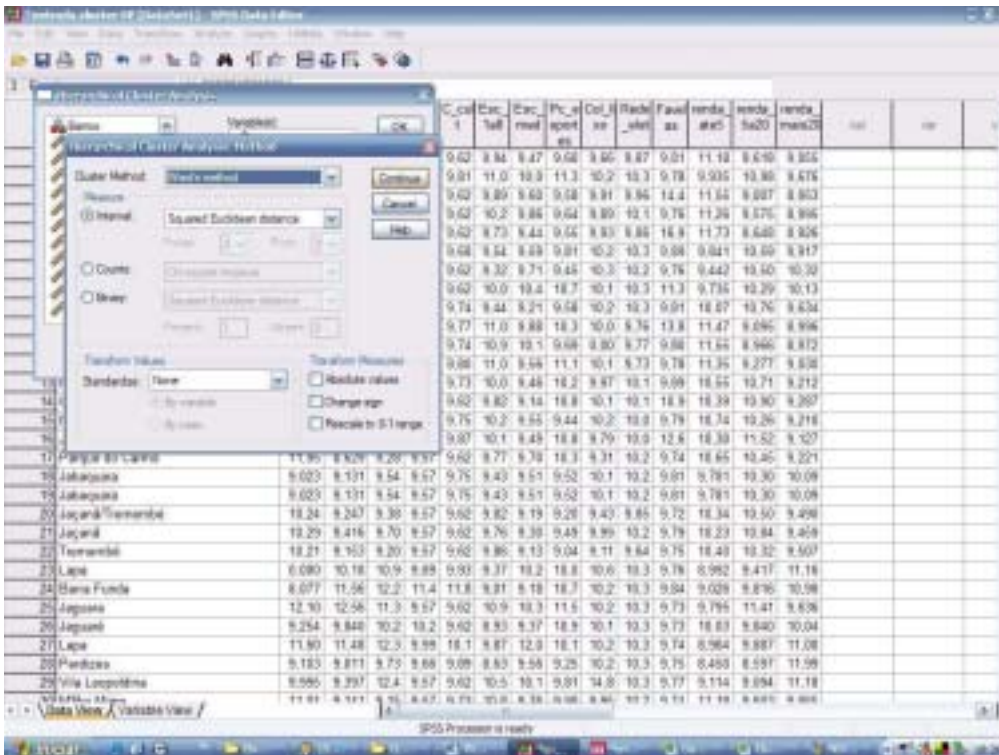
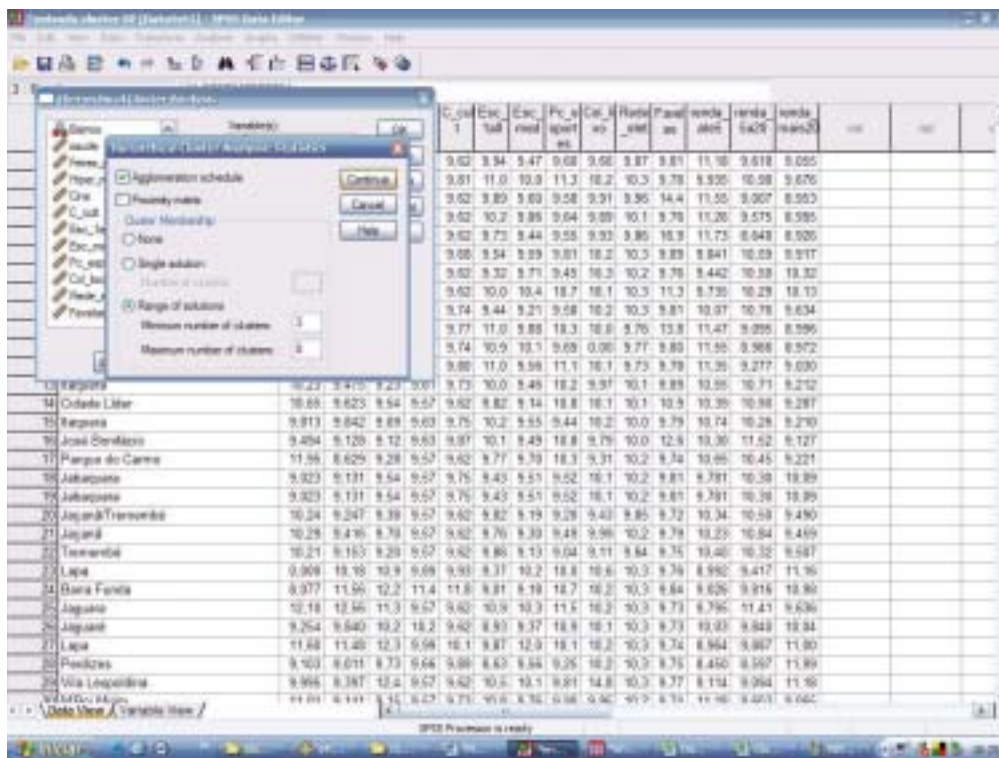
<sup>10</sup> A tabela de dados pode ser obtida por download no link 1

Faz-se a seleção das variáveis que irão balizar o agrupamento, no caso desse primeiro nível foram escolhidas somente as variáveis de classe de renda.



Nos próximos quadros solicitamos que a saída seja feita por uma faixa de clusters, em numero de 3 a 8, para posteriormente podermos optar pelo numero que melhor seja adequado e eficiente para o agrupamento e é feita a escolha do método "Ward"





Vista parcial do relatório de saída do agrupamento mostrando os estágios de agrupamento.

SPSS Output: Ward Linkage

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	12	13	.950	0	0	15
2	3	31	.900	0	0	5
3	1	33	.900	0	0	13
4	8	95	.901	0	0	44
5	84	83	.901	0	0	38
6	3	11	.903	2	0	49
7	27	81	.904	0	0	17
8	8	52	.905	8	0	88
9	84	88	.907	8	0	70
10	86	87	.908	0	0	47
11	48	50	.910	0	0	40
12	13	70	.912	0	0	53
13	1	75	.914	3	0	40
14	27	83	.915	0	0	52
15	21	52	.910	0	0	47
16	8	78	.921	0	1	72
17	27	65	.924	7	0	86
18	28	45	.928	0	0	35
19	55	75	.931	0	0	52
20	19	88	.934	0	0	49
21	2	94	.936	0	0	52
22	48	88	.940	0	0	40
23	34	35	.947	0	0	55
24	24	27	.952	0	0	43

SPSS Output: Cluster Membership

Case	0 Clusters	1 Cluster	2 Clusters	3 Clusters	4 Clusters	5 Clusters
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	1	1	1	1
4	1	1	1	1	1	1
5	5	3	1	1	1	1
6	2	2	2	2	2	2
7	4	4	3	2	2	2
8	2	2	2	2	2	2
9	2	2	2	2	2	2
10	3	3	1	1	1	1
11	3	3	1	1	1	1
12	1	1	1	1	1	1
13	5	5	4	3	2	2
14	5	5	4	3	2	2
15	5	5	4	3	2	2
16	6	2	2	2	2	2
17	5	5	4	3	2	2
18	2	2	2	2	2	2
19	2	2	2	2	2	2
20	5	5	4	3	2	2
21	6	6	4	3	2	2
22	5	5	4	3	2	2
23	7	8	5	4	3	3
24	7	8	5	4	3	3
25	6	2	2	2	2	2
26	2	2	2	2	2	2
27	7	6	5	4	3	3
28	6	7	5	5	4	3
29	7	6	5	4	3	3
30	1	1	1	1	1	1

Nessa tela se apresentam os agrupamentos em função do numero de cluster. Por exemplo o case 7 (bairro 7) se classifica no grupo 4 se forem considerados a



*divisão em 7 e 8 cluster, no grupo 3 se divisão em 6 cluster e no grupo 2 se o numero de cluster for entre 3 e 5.*

Dessa ultima tabela pode-se reconhecer que o nível 1 se mostrou mais equilibrado no caso de divisão em 4 clusters, ficando assim distribuídos:

- GRUPO 1 do NÍVEL 1: Brasilândia; Guaianases; Lajeado; Itaim Paulista; Vila Curuçá; M'Boi Mirim; Jardim Ângela; Parelheiros; Marsilac; Perus; Anhanguera; Perus; São Mateus; Iguatemi; São Rafael; São Miguel; Jardim Helena; Vila Jacuí;
- GRUPO 2 do NÍVEL 1: Freguesia do Ó; Ipiranga; Cursino; Ipiranga; Sacomã; Itaquera; Cidade Líder; Itaquera; José Bonifácio; Parque do Carmo; Jabaquara; Jaçanã; Tremembé; Jaguará; Jaguaré; Jardim São Luís; Moóca Água Rasa; Belém; Brás; Pari; Penha; Artur Alvim; Cangaíba; Penha; Vila Matilde; Pirituba; Jaraguá; Pirituba; São Domingos; Mandaqui; Tucuruvi; São Mateus; São Miguel; Sé; Bela Vista; Bom Retiro; Cambuci; República; Santa Cecília; Vila Guilherme; Vila Maria; Vila Medeiros; São Lucas; Sapopemba; Vila Prudente;
- GRUPO 3 do NÍVEL 1: Lapa; Barra Funda; Lapa; Vila Leopoldina; Tatuapé; Santana; Santo Amaro; Campo Grande; Consolação; Liberdade Saúde;
- GRUPO 4 do NÍVEL 1: Perdizes; Pinheiros; Alto de Pinheiros; Itaim Bibi Jardim Paulista; Pinheiros; Campo Belo; Vila Mariana; Moema; Vila Mariana;

Os clusters costumam ser denominados em função da característica média do grupo, que nesse nível 1, em que somente se fez classificação por nível sócio-econômico, poderíamos assim denominar os grupos: GRUPO 1 = bairros populares; GRUPO 2 = bairros simples; GRUPO 3 = bairros de classe média; GRUPO 4 = bairros de classe média- alta.

O estudo de nível 2 será feito em cima do GRUPO 2, simplesmente por apresentar um numero maior de objetos podendo ilustrar melhor a nova subdivisão.

Temos então 42 bairros com características similares em termos de classe de renda, podendo agora serem os mesmos diferenciados entre si através de outros critérios mais apurados, tendo sido escolhidas as variáveis básicas referente aos níveis de:

- Saúde (quantidade de leitos por 20.00 habitantes);
- Escola (quantidade de vagas para cada grupo de 20.000 habitantes);
- Alimentação (quantidade de feiras, mercados e sacolões para grupo de 20.000 habitantes)

Foi realizado procedimento similar ao mostrado para o agrupamento de NÍVEL 1, resultando então nos seguintes clusters mostrados na tabela abaixo<sup>11</sup>:

---

<sup>11</sup> As tabelas com cluster de nível 1 e nível 2 podem ser obtidas por download no link 2

CLASSIFICAÇÃO EM CLUSTER - NÍVEL 2				
Critério: SAÚDE - ESCOLA - ALIMENTAÇÃO				
		CLUSTER4	CLUSTER3	CLUSTER2
bairro 02	Freguesia do Ó	1	1	1
bairro 07	Cursino	1	1	1
bairro 08	Ipiranga	1	1	1
bairro 09	Sacomã	1	1	1
bairro 14	Cidade Líder	1	1	1
bairro 15	Itaquera	1	1	1
bairro 16	José Bonifácio	1	1	1
bairro 17	Parque do Carmo	1	1	1
bairro 19	Jabaquara	1	1	1
bairro 21	Jaçanã	1	1	1
bairro 22	Tremembé	1	1	1
bairro 25	Jaguara	1	1	1
bairro 26	Jaguaré	1	1	1
bairro 32	Jardim São Luís	1	1	1
bairro 33	Moóca	1	1	1
bairro 34	Água Rasa	1	1	1
bairro 35	Belém	1	1	1
bairro 36	Brás	1	1	1
bairro 38	Pari	1	1	1
bairro 44	Artur Alvim	1	1	1
bairro 45	Cangaíba	1	1	1
bairro 46	Penha	1	1	1
bairro 47	Vila Matilde	1	1	1
bairro 57	Jaraguá	1	1	1
bairro 58	Pirituba	1	1	1
bairro 59	São Domingos	1	1	1
bairro 61	Mandaqui	1	1	1
bairro 63	Tucuruvi	1	1	1
bairro 70	São Mateus	1	1	1
bairro 74	São Miguel	3	1	1
bairro 77	Bela Vista	3	1	1
bairro 78	Bom Retiro	3	1	1
bairro 79	Cambuci	2	2	1
bairro 82	República	2	2	1
bairro 83	Santa Cecília	2	2	1
bairro 84	Sé	2	2	1
bairro 86	Vila Guilherme	2	2	1
bairro 87	Vila Maria	4	3	2
bairro 88	Vila Medeiros	4	3	2
bairro 94	São Lucas	4	3	2
bairro 95	Sapopemba	4	3	2
bairro 96	Vila Prudente	4	3	2

OBS: tabela de própria elaboração.

Adotando-se então a quantidade de 3 clusters teremos os bairros do cluster 2 do nível 1 ora subdividido em mais 3 clusters de nível 2.

O que significa que cada agrupamento acima se mostra similar nos aspectos sócio-econômico (objeto do nível 1) e de características de nível de qualidade com relação à saúde, escolas e alimentação.

Em princípio cada um dos agrupamentos acima poderiam compor uma amostra para estudo de simulação de seus mercados imobiliários, devido ao alto grau de similaridade com relação aos aspectos de qualidade de seus atributos de localização e formariam, assim, uma denominada “mesma região geo-econômica”, tão preconizada nos meios da engenharia de avaliação.

Poder-se ia ainda subdividir novos grupamentos em função de localização geográfica de regiões da cidade (norte, sul, leste, oeste), se for entendimento do avaliador que cada região apresente características diferenciadas entre si no que diz respeito ao mercado imobiliário.

Lembramos, no entanto, o já exposto no capítulo introdutório com relação a pouca importância da regionalização geográfica, vide Fávero (2003) apud SARTORIS NETO (1996):

*“poder-se-ia pensar em uma divisão geográfica, porém a mesma ‘seria descabida, pois seria supor que a zona sul de São Paulo teria, por si só, algum grau de homogeneidade’ ”.*

Na aplicação do cluster acima em uma modelagem de mercado no entanto essa regionalização será observada, sendo então selecionados os bairros do cluster 1 do nível 2 apenas para os bairros posicionados no vetor leste do município.

## **7) Aplicação da técnica numa modelagem**

O que se obtém na aplicação da estatística de “cluster” é que o conjunto de objetos analisados, no caso os bairros de uma cidade, apresentem características equivalentes dentro das variáveis consideradas.

Não se pode garantir que a equivalência de qualidade importe em equivalência de valores de venda dos imóveis, mas em tese se o agrupamento tiver sido feito com os critérios enxergados pelo mercado como sendo as variáveis importantes para formação do valor de mercado espera-se que essa equivalência de preço acabe ocorrendo, e esse melhor ajuste pode ser notado em uma melhor correlação dos dados quando da modelagem.

Fique claro ainda que o fato de termos agrupado bairros de características gerais similares não significa que a modelagem por inferência estatística fique isenta de outras variáveis de localização, que podem expressar detalhes não observados no critério geral.

Assim, para verificação da eficiência da montagem do agrupamento realizado no estudo de caso, iremos aplicá-lo numa modelagem de apartamentos a venda na cidade de São Paulo.

Os elementos foram extraídos do banco de dados que a Caixa Econômica Federal gerencia com os terceirizados, denominado de “BR-Elementos”, tendo sido utilizado o banco do mês de janeiro de 2007.

Tomamos inicialmente os dados em sua totalidade, apenas limitando a faixa de estudo para imóveis com preço entre R\$ 40.000,00 a R\$ 150.000,00.

Com isso resultou numa amostra contendo 1.244 dados, propiciando a montagem de uma modelagem com as seguintes variáveis:

- Dependente: preço em R\$;
- Independentes:
  - Área construída (útil dos apartamentos em m<sup>2</sup>);
  - Vagas: numero de vagas de garagem;
  - Banho: numero de banheiros completos;
  - Padrão: proxy associando o padrão da construção às constantes de tipologia e padrão de IBAPE (2005);
  - Estado: conservação como código alocado: 1 = novo, 2 = bom estado; 3 = conservação regular.

DADOS DE ENTRADA DA INFERENCIA						
Região amostral - São Paulo - Apartamentos - Jan/2007						
Dado	Area Construida	Vagas	Banho	Padrão	Estado	Preço
1	65,00	1	2	1,81	3	R\$ 150.000,00
2	77,00	1	1	1,81	3	R\$ 150.000,00
3	72,42	0	1	1,41	3	R\$ 150.000,00
4	60,00	0	2	1,81	2	R\$ 150.000,00
5	81,00	1	1	1,81	3	R\$ 150.000,00
...	...	...	...	...	...	...
...	...	...	...	...	...	...
1236	47,22	1	1	1,05	3	R\$ 40.000,00
1237	56,00	1	1	1,81	3	R\$ 40.000,00
1238	51,00	1	1	1,41	3	R\$ 40.000,00
1239	51,34	0	1	1,41	3	R\$ 40.000,00
1240	51,34	1	1	1,41	3	R\$ 40.000,00
1241	35,00	0	1	1,41	3	R\$ 40.000,00
1242	41,00	0	1	1,81	2	R\$ 40.000,00
1243	48,99	0	1	1,41	3	R\$ 40.000,00
1244	51,00	1	1	1,05	3	R\$ 40.000,00

OBS: A tabela completa pode ser obtida via internet no Link 3.

Foi utilizado o programa TS-SISREG para análise do modelo, apresentando os seguintes resultados mostrados abaixo:

#### **Resultados Estatísticos:**

##### Linear

- Coeficiente de correlação: 0,655363
- Coeficiente de determinação: 0,429500
- Coeficiente de determinação ajustado: 0,427196
- Fisher-Snedecor: 186,41
- Significância: 0,01

##### Não-Linear

- Coeficiente de correlação: 0,646940
- Coeficiente de determinação: 0,418531

#### **Normalidade dos resíduos**

- 67% dos resíduos situados entre -1 e +1 s
- 89% dos resíduos situados entre -1,64 e +1,64 s
- 95% dos resíduos situados entre -1,96 e +1,96 s

### Outliers do modelo: 52

#### Equação

Regressores	Equação	T-Observado	Significância
• Area Construida	$\ln(x)$	18,99	0,01
• Vagas	$x$	6,68	0,01
• Banho	$x$	8,97	0,01
• Padrão	$\ln(x)$	10,42	0,01
• Estado	$x$	-4,24	0,01
• Preço	$\ln(y)$		

A questão do ajuste pode ser melhor visualizada no gráfico abaixo:

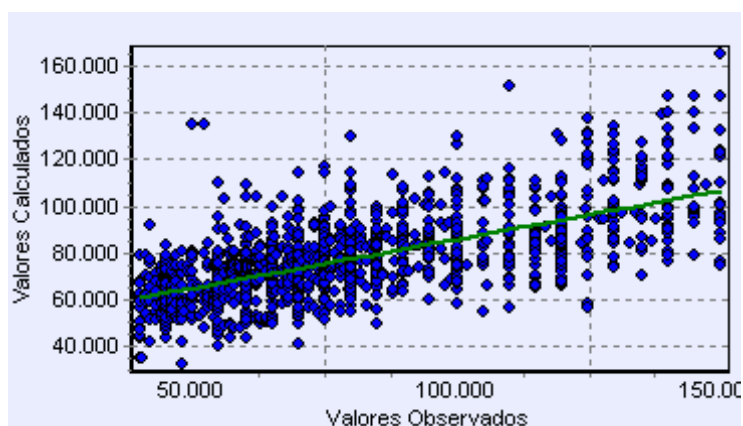


Figura 7.1 – distribuição dos resíduos com a modelagem de 1244 apartamentos a venda em SP.

Do conjunto de elementos acima, foram selecionados aqueles que se agruparam dentro dos seguintes critérios:

- Nível 1 – características sócio-econômicas – distribuição de classe de renda dos moradores dos bairros;
- Nível 2 – Condições básicas de escola / saúde / alimentação;
- Nível 3 – Região da cidade – selecionado os bairros da zona leste.

Do conjunto de elementos foram então selecionados 173 dados que formaram o grupo que vai ser modelado em seu mercado imobiliário, seguindo-se os mesmos princípios acima<sup>12</sup>, resultando no seguinte:

#### Resultados Estatísticos:

##### Linear

- Coeficiente de correlação: 0,802671
- Coeficiente de determinação: 0,644281
- Coeficiente de determinação ajustado: 0,633631
- Fisher-Snedecor: 60,49
- Significância: 0,01

##### Não-Linear

- Coeficiente de correlação: 0,775151

<sup>12</sup> Dados de entrada e saída da inferência podem ser obtidas por download no link3.

- Coeficiente de determinação: 0,600859

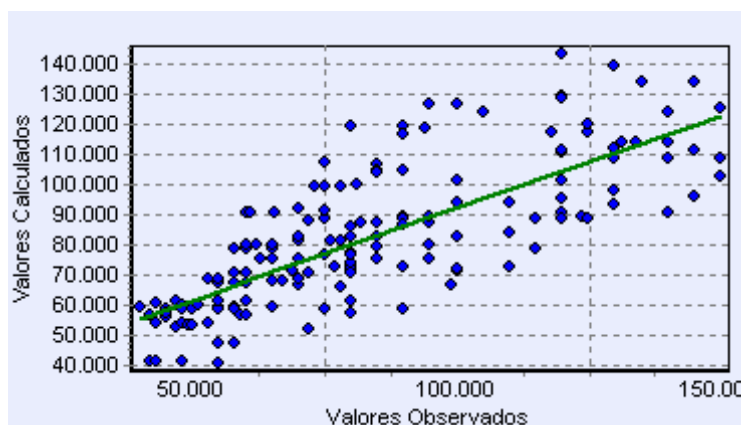
#### **Normalidade dos resíduos**

- 62% dos resíduos situados entre -1 e +1 s
- 92% dos resíduos situados entre -1,64 e +1,64 s
- 97% dos resíduos situados entre -1,96 e +1,96 s

#### **Outliers do modelo: 2**

#### **Equação**

Regressores	Equação	T-Observado	Significância
• Area Construída	$\ln(x)$	9,81	0,01
• Vagas	x	2,20	2,86
• Banho	x	4,43	0,01
• Padrão	x	4,37	0,01
• Estado	x	-4,05	0,01
• Preço	$\ln(y)$		



Nota-se portanto uma sensível melhora na homogeneização do conjunto de dados em cluster, como mostram os coeficientes de correlação, a saber:

- Correlação da amostra completa: 0,6553;
- Correlação do grupo classificado: 0,8026

Também se denota a grande quantidade de “outliers” na primeira modelagem (52 em 1244 ou 4,2% da amostra), o que na modelagem com cluster é reduzido de forma substancial (2 dados em 173 ou 1,2%).

Assim, nesse caso em particular ficou demonstrada a eficiência de se agrupar os bairros com critérios técnicos de estatística de cluster.

## **8) Conclusões finais – recomendações**

O intuito do presente trabalho foi o de demonstrar a utilização da matemática de agrupamento para determinação de regiões de um município que apresentam uma certa homogeneidade em termos de parâmetros de qualidade de localização,

com intuito de parametrizar o campo de amostras para que dele se possa inferir um modelo de simulação de mercado com maior grau de qualidade.

Como foi dito no decorrer do trabalho, a aplicação dessa técnica não exige a utilização de outras variáveis de localização dentro de cada modelo, já que mesmo dentro de mesmos bairros podemos encontrar particularidades que também exercem influencia na precificação dos imóveis (exemplos: proximidade de favelas, proximidade de shopping center; distancia para corredores, dentre outras).

O estudo de caso realizado foi elaborado com simples intuito de demonstrar a utilização da estatística envolvida e não foi realizada para mapeamento de regiões na cidade de São Paulo, que exigiria uma abordagem mais meticulosa, não só com a compilação simples de dados, mas também com aspectos locais que exige do avaliador o conhecimento pessoal das regiões.

Pode-se recomendar também que em modelagens mais apuradas e sofisticadas mesmo com a formação do cluster nos critérios acima, poder-se-ia verificar com o próprio conjunto de dados comparativos se a utilização de (n-1) variáveis dicotômicas de local não resultaria em rejeição do teste de hipótese nula nos níveis de significância preconizados por norma, o que indicaria que a região não se mostraria efetivamente homogênea.

Pode-se também utilizar essa técnica em modelos mais abrangentes, eventualmente contendo como campo amostral uma cidade inteira (para municípios de menor porte), servindo o agrupamento para justificar a introdução de variáveis dicotômicas que assim expressariam a qualidade de localização de cada agrupamento (exemplo: bairros de classe alta – sim/não; bairros de classe média – sim/não; bairros de classe de renda baixa – sim/não; ausência em todos poderia ser o centro)

Enfim a técnica ora apresentada tem por objetivo ser utilizada para possibilitar que os agrupamentos de bairros possam ser elaborados com um embasamento teórico e não por simples opinião de conhecedores do mercado, em critérios pessoais e não isentos de subjetividade.

## **9) Bibliografia**

### ***Referenciada***

DANTAS (1998) – ENGENHARIA DE AVALIAÇÕES – Uma introdução a metodologia científica – 1.998

FAVERO (2.003) - PREÇOS HEDONICOS – FAVERO (Modelos de Preços Hedônicos Aplicados a Imóveis Residenciais em Lançamento no Município de São Paulo – dissertação de mestrado FEA-USP – Luis Paulo Lopes Fávero – 2.003

HERMANN (2003) – ESTIMANDO O PREÇO IMPLÍCITO DE AMENIDADES URBANAS: Evidências para o município de São Paulo – Bruno Martins Hermann – Dissertação de mestrado – FEA-USP – 2003.

IBAPE (2005) – Valores de venda de imóveis urbanos em São Paulo – estudo publicado

MALHOTRA (2006) – PESQUISA DE MARKETING – Uma orientação Aplicada – 4ª edição – Malhotra, Naresh K.

PASCALE (2005) - LOCALIZAÇÃO RESIDENCIAL – Atributos que Configuram Qualidade às Localizações Residenciais: Uma Matriz para Clientes de

Mercado na Cidade de São Paulo- Andréa Pascale – dissertação de mestrado POIL-USP – 2005)

PREFEITURA DE SÃO PAULO – [www.prefeitura.sp.gov.br](http://www.prefeitura.sp.gov.br)

SARTORI NETO, A – Estimação de modelos de preços hedônicos: um estudo para residências na cidade de São Paulo. São Paulo, 1996 . Dissertação de Mestrado FEA - USP

SECOVI-SP – [www.secovi.com.br](http://www.secovi.com.br)

SPSS – Pacote estatístico para estudo de ciências sociais ([www.spss.com](http://www.spss.com))

TS-SISREG – pacote de calculo por estatística inferencial para avaliação de imóveis ([www.tecsysengenharia.com.br](http://www.tecsysengenharia.com.br))

### **Consultada**

ABNT- NBR 14.653-2 – Avaliação de imóveis

BUSSAB, W. O; MORETTIN, P. A. – ESTATISTICA BÁSICA- 5ª ed – SP 2003

CERQUEIRA, Lucila Lopes J – 2004 – DM – A qualidade de localização intrametropolitana nos espaços de escritório: evidências teóricas e sua evolução recente no mercado de São Paulo. Aborda questões referentes a importância de atributos de localização de imóveis comerciais.

CEZARINO, Luciana Oranges (DM) – Um estudo sobre Clusters e arranjos produtivos locais: Uma aplicação em micro e pequenas empresas; 2005;

CHAPIN, F. Stuart (1962) – Urban Growth dynamics in a regional cluster of a cities – ed Wiley – 484 pg;.

ENGEL, J. F.; Blackwell, R. D; Miniard, P. W. – 2000 – Comportamento do consumidor;

FISHER, Walter Dummer – 1969 – Clustering and agregation in economics – 195 pg;

MALHOTRA, Naresh K. – 2001 – Pesquisa de marketing: Uma orientação aplicada

REA, L. M.; Parker, R. A. (2000) – Metodologia de pesquisa – do planejamento a execução;

SANTOVITO, Rogério F. – 2004 – DM – A dinâmica do mercado de edifícios de escritórios e a produção de indicadores de comportamento: Uma simulação do índice de preços de locação em regiões de ocupação típica na cidade de São Paulo;

TELLES, Luciana Oliveira – (DM) – São Paulo – 2002 - Cluster e a Industria Ligada à Área da Saúde em Ribeirão Preto

TRISTÃO, Helcio Martins – 2000 – Cluster e a Cadeia produtiva de calçados de Franca – 93 pg;

TRIVELLONI, P – CARLOS A.- HOCHHEIM, NORBERTO – Valor de Localização dos Imóveis: Determinação por Métodos de Analise Espacial – XXIII COBREAP – 2.006

VIROLI NETO, Amadeu (DM) – Estudo dos Clusters Comerciais na Cidade de São Paulo – 2003;

### **Link das Tabelas**

1) Dados SP: <http://www.divshare.com/download/1896627-4bf>

2) Cluster e sub-cluster: <http://www.divshare.com/download/1896637-a10>

3) Arquivo das inferências: <http://www.divshare.com/download/1896616-e59>



# 10) Anexo

Mapa dos distritos do município de São Paulo - <http://atlasambiental.prefeitura.sp.gov.br/mapas/102.jpg>

