

Análise de homogeneidade de localidades da cidade do Rio de Janeiro quanto a preços de locação: um estudo de caso

Homogeneity analysis of localities in the city of Rio de Janeiro in relation to rental prices: a case study

Giovani Glaucio de Oliveira Costa (UFRRJ – RJ/Brasil) - giovaniglaucio@ufrj.br
• R. Governador Roberto Silveira, s/n, Posse, Nova Iguaçu-RJ

RESUMO Este artigo tem o objetivo de apresentar a aplicação da Análise de Homogeneidade como uma alternativa à análise de dados qualitativos. A Análise de Homogeneidade é uma técnica multivariada que tem duas finalidades básicas: redução de dimensionalidade e análise estrutural. O trabalho analisou valores em reais de aluguéis médios residenciais de 24 locais ou bairros da cidade do Rio de Janeiro em 2010. O estudo visa o agrupamento das localidades em função da similaridade dos seus aluguéis médios. O perfil 1, identificado, é formado pelos bairros com os aluguéis nas faixas mais baixas. A maioria dos bairros deste perfil é da zona oeste e zona norte. O perfil 2 é formado pelos bairros com os aluguéis nas faixas intermediárias. Contemplam os bairros da zona sul e barra. O perfil 3 é formado pelos bairros com os aluguéis nas faixas mais altas. Contemplam somente bairros da zona sul. Estes resultados são importantes tanto para ressaltar a importância da Análise de Homogeneidade quanto para as tomadas de decisão de administradores da área imobiliária.

Palavras-chave Análise de dados qualitativos. Análise de homogeneidade. Localidades das cidades do Rio de Janeiro.

ABSTRACT *This article aims to present the application of Homogeneity analysis as an alternative to the analysis of qualitative data. Homogeneity analysis is a multivariate technique that has two basic purposes: dimensionality reduction and structural analysis. The work examined actual values of average rents for residential sites in 24 districts of the city of Rio de Janeiro in December 2010 for living room and bedroom, living room and two bedrooms and living room and three bedrooms. The study aimed at the grouping of localities due to similarity in their average rents. As the result of the analysis, we identified 3 homogeneous groups, coexisting in the same space. Profile 1 is formed by neighborhoods with rents in lower ranges. Most of the neighborhoods in this profile are on the westside and northside. Profile 2 is formed by neighborhoods with rents in the intermediate ranges. They include the south zone and Barra. Profile 3 is formed by neighborhoods with rents in higher ranges. They include only neighborhoods in the South side. These results are important to highlight the importance of Homogeneity analysis as to decisions by real estate administrators.*

Keywords *Analysis of qualitative data. Analysis of homogeneity. Homogeneity of localities in the city of Rio de Janeiro.*

1. INTRODUÇÃO

Na elaboração de um plano de pesquisa que planeja uma coleta extensa de informações, a seleção dos métodos de análise de dados é uma tarefa (entre outras) de importância decisiva, quaisquer que seja o campo em estudo. É essencial garantir que os instrumentos que venham a ser selecionados permitam operacionalizar as orientações sistematizadas no modelo de análise e sejam consistentes com a natureza da informação escolhida.

Frequentemente, quando o pesquisador elabora esse plano de pesquisa quantitativa, utilizando o método estatístico, ele se depara com questionários onde se não todas, a maioria das variáveis são qualitativas nominais, isto é, as questões que envolvem o questionário têm como respostas, muitas das vezes estimuladas, a forma categorial. As técnicas tradicionais de Análise Multivariada pressupõem que as variáveis sejam quantitativas e que sigam à Curva Normal. O analista, então, tem utilizado, como alternativa, o recurso analítico trivial do exame de tabelas de frequências, tabelas de contingências com testes de independência do qui-quadrado. A Análise de Homogeneidade ou Análise de Correspondência Múltipla (GREENACRE, 1984) permite que os pesquisadores ao utilizar questionários que envolvam perguntas com respostas nominais ou ordinais, saiam do campo meramente descritivo e desenvolvam técnicas multivariadas para o tratamento avançado dos múltiplos indicadores envolvidos na pesquisa (PÉREZ, 2001). A Análise de Homogeneidade trata múltiplas variáveis categoriais e realiza análises estruturais (PÉREZ, 2001).

A Análise de Homogeneidade busca averiguar, por via das interações estabelecidas entre as múltiplas categorias, posicionadas num espaço definido em função do cruzamento das dimensões, se são definidos grupos distintos (CARVALHO, 2004). Em caso afirmativo, pretende-se averiguar como se configuram e como se posicionam os grupos uns com os outros. Categorias próximas à origem indicam que elas não são importantes para a definição das dimensões do plano de análise (CARVALHO, 2004). Categorias próximas podem indicar uma configuração existente no espaço de análise.

A Análise de Homogeneidade ou HOMALS (*Homogeneity Alternating Least Squares*) é indicada para a reprodução num espaço de menores dimensões a multiplicidade que sustenta o espaço original (CARVALHO, 2000). Com isso, a complexidade que se estabelece no espaço de partida é reduzida e simplificada num espaço de dimensionalidade menor.

A Análise de Homogeneidade é uma técnica multivariada que tem dois objetivos básicos: redução de dimensionalidade e análise estrutural.

O referido artigo tem o seguinte problema de pesquisa: Quais os subperfis ou configurações de bairros do Rio de Janeiro quanto a preços de locação ou aluguel?

A hipótese que se estabelece é que há configurações latentes de bairros em função de valores de locação que coexistem na cidade do Rio de Janeiro.

Este artigo tem o objetivo diminuir a dimensão do espaço de análise dos preços de aluguéis de bairros do Rio de Janeiro e assim, de forma mais simplificada e racional, efetuar uma análise estrutural.

Os resultados deste estudo podem dar suporte às tomadas de decisão de administradores imobiliários sobre as localidades, ou bairros, com perfis comuns de preços de locação. Os planos de gerência de locação poderiam ser direcionadas a cada perfil de localidades evidenciadas pela análise e assim aumentar as locações de imóveis, satisfazer e fidelizar inquilinos. Podem também subsidiar informações a visitantes e turistas quanto aos preços de aluguéis de acordo com os bairros e características dos imóveis que almejam.

2. REVISÃO DE LITERATURA

O método de Análise de Homogeneidade (HOMALS-*Homogeneity Analysis by Means of Least Square*), também conhecida como Análise de Correspondência Múltipla (CARVALHO, 2008), permite resumir um grande número de variáveis qualitativas (ou tratadas como tal) de uma base de dados de estudo em um pequeno número de variáveis quantitativas, facilitando o estudo das relações entre as diversas características existentes num determinado espaço de análise (CARVALHO, 2004).

A Análise de Homogeneidade integra os métodos de “*Optimal Scaling*” (GIFI, 1996). Estes métodos consistem, na prática, numa intervenção sobre as categorias com o objetivo de proceder a sua quantificação (GIFI, 1996).

A HOMALS viabiliza uma Análise de Homogeneidade por via de um método algébrico do tipo “*Alternating Least Squares*” (GIFI, 1996). O algoritmo da HOMALS é, então, do tipo ALS, o que significa que em cada uma das iterações vão sendo determinadas, alternativamente, estimativas quantitativas para os objetos e categorias da base de dados (GIFI, 1996).

Este processo é um procedimento de transformação ótima porque vai sendo minimizada uma função perda, cuja convergência corresponderá às quantificações das categorias e dos objetos (GIFI, 1996). Estas quantificações têm a propriedade de guardarem e resumirem as características essenciais das categorias e objetos de análise da matriz de dados de partida, mantendo a variação de informação do espaço original intacta no espaço gerado de menores dimensões (GIFI, 1996).

Em sentido restrito, a Análise de Homogeneidade é um método de análise quantitativa para variáveis nominais e tem por objetivo atingir a solução ótima na quantificação de dados qualitativos (GREENACRE, 1996). Em sentido amplo, a Análise de Homogeneidade refere-se a um conjunto de critérios para analisar dados multivariados, com o objetivo de aperfeiçoar a homogeneidade das variáveis (GREENACRE, 1996).

A HOMALS constitui, portanto, num conjunto de critérios para analisar dados multivariados, com o objetivo de se atingir a homogeneidade das variáveis (GREENACRE, 1996). A técnica trata múltiplas variáveis categoriais e desenvolve análises estruturais (GREENACRE, 1996). É um método de análise quantitativa para variáveis qualitativas e tem por objetivo atingir a solução ótima na quantificação de dados qualitativos ou categoriais (GREENACRE, 1996).

Na Análise de Homogeneidade, procura-se visualizar através da disposição relativa dos indivíduos ou das categorias em planos, as relações estatísticas delineadas entre as múltiplas variáveis em análise (GREENACRE, 1996). A partir das configurações definidas pelas categorias das variáveis, espera-se conhecer de forma aproximada a estrutura que sustenta o espaço em análise (GREENACRE, 1996).

As configurações nos planos ao exibirem diferentes combinações das propriedades envolvidas refletem a presença de grupos de indivíduos relativamente homogêneos (CARVALHO, 2004). Seja, então, qual for a natureza das unidades de análise, o que determina a opção pela Análise de Homogeneidade é o propósito de realizar uma abordagem relacional sobre as múltiplas variáveis que caracterizam os objetos, podendo vir a serem definidos diferentes grupos (CARVALHO, 2004).

A descrição desses grupos pode contemplar dois vetores analíticos (CARVALHO, 2004):

- Identificação do perfil de cada grupo definido;
- Observação do posicionamento relativo dos vários grupos que permita detectar a existência de relação de associação ou de oposição.

Os resultados ou output da HOMALS incluem (CARVALHO, 2004):

- 1) Estatísticas – valores próprios (variâncias explicadas de eixos estruturantes), medidas de discriminação ou importância das variáveis, quantificações das categorias e escores de objetos;
- 2) Imagens gráficas (planos bidimensionais) que permitem a visualização da informação contida na matriz de entrada num espaço de menores dimensões.

Para validar os resultados da Análise de Homogeneidade, pode-se proceder a uma Análise de *Cluster* a partir dos resultados encontrados para as novas variáveis do espaço de análise.

Na articulação HOMALS/Análise de Homogeneidade, a Análise de *Cluster* toma como variáveis de *input*, para a classificação, as dimensões que sustentam o plano em estudo na Análise de Homogeneidade (CARVALHO, 2004). Quando se faz referência às dimensões enquanto novas variáveis, isso já pressupõe que depois de realizada a HOMALS e tomada a decisão sobre quantas dimensões interpretar, foram salvos na base de dados os escores de objetos por dimensão (CARVALHO, 2004). Com este procedimento, serão acrescentadas à base de dados novas variáveis, em número igual ao indicado para as dimensões (CARVALHO, 2004).

As dimensões são transformadas em novas variáveis com características particularmente interessantes: variáveis compostas e contínuas, que têm enorme potencial analítico (CARVALHO, 2004). As dimensões são constituídas pelos escores de objetos e estes, algebricamente, são determinadas a partir das múltiplas variáveis qualitativas que figuram na matriz de *input* e que definem o perfil de cada indivíduo (CARVALHO, 2004). Daí afirmar-se que as dimensões são variáveis compostas, garantindo que a multidimensionalidade foi preservada (CARVALHO, 2004). Quanto à sua natureza, e apesar do *input* ser qualitativas, as dimensões são variáveis contínuas (CARVALHO, 2004).

Para realizar a Análise de *Cluster* começa-se por recorrer-se ao Método *K-Means Clusters* (CARVALHO, 2004). Assim, será utilizada como variáveis de *input* as dimensões consideradas para a análise e solicitar o número de grupos que quer se formar correspondente ao número de configurações reveladas no gráfico das categorias da Análise de Homogeneidade (CARVALHO, 2004).

Para garantir que a partição dos indivíduos nos *clusters* gerados pela análise fique definida como nova variável é necessário salvar esta variável na base de dados (CARVALHO, 2004). Em seguida, vem à caracterização dos grupos obtidos, a fim de efetuar a correspondência destes com os perfis previamente definidos via HOMALS (CARVALHO, 2004).

Na prática, a variável de classificação, salva na base de dados, classifica cada objeto em um dos *clusters* gerados (CARVALHO, 2004). Contudo, o método de *cluster* só diz quem é cada indivíduo, a que grupo pertence, mas não informa como eles são (CARVALHO, 2004). A associação com a informação da HOMALS vai justamente traçar o perfil de todos os *cluster* gerado pela Análise de *Cluster* no Método *K-Means* (CARVALHO, 2004).

Usa-se o Método *K-Means*, um método não hierárquico, pois geralmente já se sabe, através da análise do plano da análise da HOMALS, o número de grupos que a Análise de *Cluster* deve segmentar o espaço de análise original (CARVALHO, 2004).

Caso não se tenha a informação prévia de quantos grupos deve a Análise de *Cluster* segmentar o espaço de análise ou para validar se aquela que se considera a melhor partição é consistente com as conclusões retiradas do plano em estudo da HOMALS, sugere-se, antes, usar um método hierárquico de aglomeração, que indique o número de *cluster* ideal para a classificação do espaço de análise (CARVALHO, 2004).

Para se concluir quanto ao número de *clusters*, pode-se usar a leitura do dendograma, estatística gerada pelo método hierárquico. Em casos em que o número de objetos da base de dados é alto, sugere-se usar um gráfico com as distâncias dos coeficientes de fusão (Screen Plots), estatística fornecida pelo output da análise (CARVALHO, 2004). Este gráfico relaciona números de *cluster* em função das distâncias ou coeficientes de fusão (CARVALHO, 2004). Observam-se neste gráfico a partir de quantos *clusters* as distâncias deixam de ser tão significativas. São considerados os pontos até a linha tornar-se horizontal. Os resultados da realização de um método de classificação hierárquicos permitem, assim, validar a solução sugerida pela HOMALS (CARVALHO, 2004).

Classificando os casos, é, então, feita a correspondência entre a solução dos *clusters* determinados e os perfis configurados pelo plano da HOMALS (CARVALHO, 2004). Este procedimento pode ser conseguido pelo cruzamento de cada uma das variáveis usadas na definição dos perfis, via HOMALS, com a variável na base de dados que define os *clusters* (CARVALHO, 2004).

Para tanto, são retomadas as variáveis de *input* da HOMALS que mais discriminam nas dimensões retidas e efetuar-se o cruzamento com a variável de *cluster* salva na base de dados (CARVALHO, 2004). A partir das percentagens mais elevadas para cada um dos *clusters*, pode-se fazer uma síntese interpretativa de suas principais características (CARVALHO, 2004).

Para validar graficamente a consistência desta classificação, pode ainda proceder-se à projeção dos *clusters* no plano das categorias (CARVALHO, 2004). Do ponto de vista técnico, consiste numa representação gráfica na qual estão simultaneamente representadas as variáveis que configuram os perfis, via HOMALS, como variáveis ativas, e a variável que identifica a tipologia de segmentação, tendo esta o status de variável passiva (CARVALHO, 2004). O objetivo é observar a disposição dos *clusters* formados pela Análise de *Cluster* com a realização da Análise de Homogeneidade (CARVALHO, 2004).

A representação gráfica deve evidenciar a tipologia de segmentação coerente com os resultados desde logo obtidos com a realização da Análise de Homogeneidade (CARVALHO, 2004).

A exemplificação sistematizada evidencia como a gestão complementar entre a HOMALS e a Análise de *Cluster* pode de fato, desempenhar um papel importante na operacionalização de grupos, na perspectiva de definir tipologias (CARVALHO, 2004).

Assim, num primeiro momento, afere-se sobre a configuração topológica do espaço de análise, o que implica a operar com múltiplas variáveis qualitativas que interagem de forma sistêmica (CARVALHO, 2004). Este processo analítico resulta na identificação de diferentes perfis. Num segundo momento, parte-se para o ensaio de tipologias dos objetos do espaço de análise. De cada tipo gerado pelo método não hierárquico de aglomeração, é traçado um perfil coerente com o evidenciado pela Análise de Homogeneidade (CARVALHO, 2004). É este encadeamento operatório que se considera ter uma concretização possível da articulação da Análise de Homogeneidade com a Análise de *Cluster* (CARVALHO, 2004) e que será desenvolvido neste estudo.

3. MATERIAL E MÉTODO DE PESQUISA

Os parágrafos seguintes consistirão do desenvolvimento das fases descritas acima da HOMALS, com vistas a redução de dimensionalidade e busca de estruturas internas.

3.1. Caracterização do Objeto de Estudo

Este trabalho tem como objeto de estudo a cidade do Rio de Janeiro. Os dados da Tabela 1 constituem a base de dados alvo das análises propostas neste estudo e se referem aos valores em reais de aluguéis médios residenciais de 24 locais ou bairros da cidade em dezembro de 2010 para sala e quarto (SQ2), sala e dois quartos (S2Q2) e sala e três quartos (S3Q2).

O objetivo deste artigo é realizar o agrupamento das localidades em função da similaridade dos seus aluguéis médios. Para atingir a finalidade da proposta, irá se aplicar a Análise da Homogeneidade com articulação com a Análise de *Cluster*.

Tabela 1 – Localidades da cidade do Rio de Janeiro em termos de valores médios em reais de aluguéis residenciais.

Local	Zona	SQ2(reais)	S2Q2(reais)	S3Q2(reais)
Bangu	Zona Oeste	350-750SQ	450-1450S2Q	550-2550S3Q
Barra da Tijuca	Barra da Tijuca	1150-1950SQ	1450-2450S2Q	2550-4550S3Q
Botafogo	Zona Sul	350-750SQ	1450-2450S2Q	550-2550S3Q
Jardim Botânico	Zona Sul	1150-1950SQ	1450-2450S2Q	550-2550S3Q
Centro	Centro	350-750SQ	450-1450S2Q	550-2550S3Q
Copacabana	Zona Sul	350-750SQ	1450-2450S2Q	2550-4550S3Q
Ilha do Governador	Zona Norte	350-750SQ	450-1450S2Q	550-2550S3Q
Inhaúma	Zona Norte	350-750SQ	450-1450S2Q	550-2550S3Q
Irajá	Zona Norte	350-750SQ	450-1450S2Q	550-2550S3Q
Jacarepaguá	Zona Oeste	350-750SQ	450-1450S2Q	550-2550S3Q
Lagoa	Zona Sul	1950-2750SQ	2450-3450S2Q	2550-4550S3Q
Madureira	Zona Oeste	350-750SQ	450-1450S2Q	550-2550S3Q
Méier	Zona Oeste	350-750SQ	450-1450S2Q	550-2550S3Q
Ramos	Zona Norte	350-750SQ	450-1450S2Q	550-2550S3Q
Santa Teresa	Zona Sul	350-750SQ	450-1450S2Q	Sem Dado
São Cristóvão	Zona Oeste	350-750SQ	450-1450S2Q	550-2550S3Q
Tijuca	Zona Oeste	350-750SQ	450-1450S2Q	550-2550S3Q
Vila Isabel	Zona Oeste	350-750SQ	450-1450S2Q	550-2550S3Q
Ipanema	Zona Sul	1150-1950SQ	2450-3450S2Q	4550-6550S3Q
Recreio dos Bandeirantes	Barra da Tijuca	1150-1950SQ	450-1450S2Q	550-2550S3Q
Flamengo	Zona Sul	350-750SQ	450-1450S2Q	550-2550S3Q
Leblon	Zona Sul	1150-1950SQ	2450-3450S2Q	4550-6550S3Q
Gávea	Zona Sul	1150-1950SQ	1450-2450S2Q	2550-4550S3Q
Laranjeiras	Zona Sul	350-750SQ	1450-2450S2Q	550-2550S3Q

Nota: SQ2 (apartamento sala, um quarto), S2Q2(apartamento sala, dois quartos) e S3Q2(apartamento sala, três quartos). Os resultados das variáveis foram transformados em faixas no objetivo de transformar as variáveis inicialmente quantitativas em qualitativas.

A Tabela 2 apresenta a matriz de *input* para inserção no SPSS e realização da HOMALS.

Para a realização da HOMALS, serão incluídas no modelo as variáveis Zona, SQ2, S2Q2 e S3Q2. A variável Local (Bangu, Barra da Tijuca, Botafogo, etc.) será utilizada para identificar os objetos no “Gráfico dos Objetos”.

Tabela 2 – Matriz de *input* para HOMALS.

Local	Zona	SQ2	S2Q2	S3Q2
1	4	1	1	1
2	5	2	2	2
3	2	1	2	1
4	2	2	2	1
5	1	1	1	1
6	2	1	2	2
7	3	1	1	1
8	3	1	1	1
9	3	1	1	1
10	4	1	1	1
11	2	3	3	2
12	4	1	1	1
13	4	1	1	1
14	3	1	1	1
15	2	1	1	-1
16	4	1	1	1
17	4	1	1	1
18	4	1	1	1
19	2	2	3	3
20	5	2	1	1
21	2	1	1	1
22	2	2	3	3
23	2	2	2	2
24	2	1	2	1

Fonte: Jornal O Globo (2012).

3.2. Caracterização da Pesquisa

A finalidade da pesquisa é realizar a segmentação da base de dados em grupos de localidade ou bairros de acordo com suas similaridades de aluguéis médios na cidade do Rio de Janeiro. Para tanto se utilizará a análise multivariada de dados qualitativos, denominada Análise de Homogeneidade, de acordo com as etapas apresentadas em parágrafos seguintes.

3.3. Método de Coleta e Análise de Dados

Os dados foram obtidos através de fontes secundárias disponíveis em jornais e revistas de grande circulação nacional.

As etapas da análise de dados são:

- 1ª) Identificação das Dimensões mais Importantes;
- 2ª) Identificação das Variáveis mais Importantes;
- 3ª) Identificação das Categorias mais Importantes;
- 4ª) Estruturação do Espaço de Análise;
- 5ª) Articulação da HOMALS com a Análise de *Cluster*.

A seguir se irá desenvolver as etapas do método para a segmentação da base de dados, com apresentação dos resultados.

4. RESULTADOS DA ANÁLISE DE DADOS

1º) Identificação das dimensões mais importantes

Nesta seção, serão selecionadas as dimensões mais importantes para a estruturação do espaço de análise.

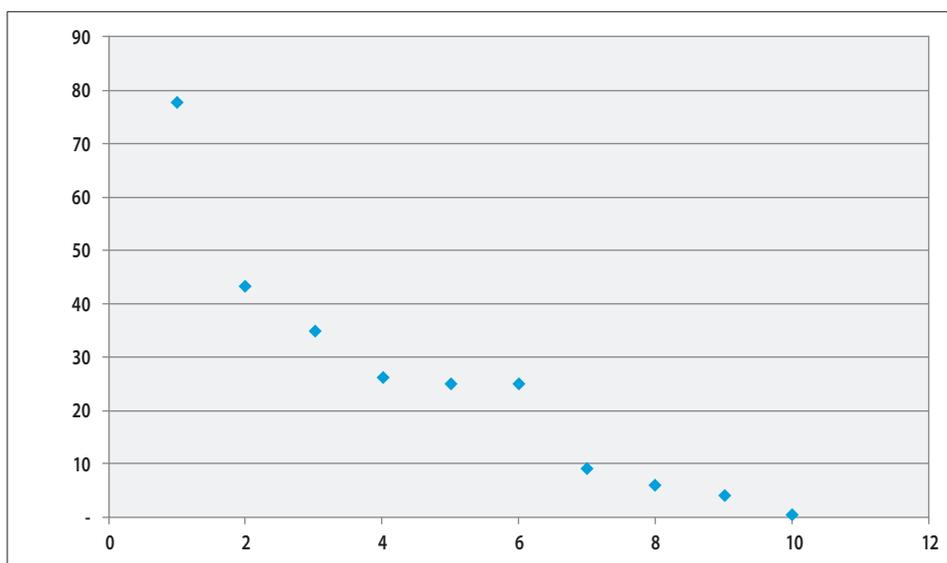
Apresentam-se os valores próprios para o espaço dos aluguéis. Foi usado um número elevado de dimensões, a fim de ser possível analisar o comportamento dos valores próprios. Solicitou-se uma solução com 11 dimensões, número de categorias (15) menos o número de variáveis (4), que corresponde ao número máximo que é possível solicitar no pacote estatístico utilizado neste estudo de caso, o SPSS 15.0. No entanto, o SPSS somente forneceu por output os valores próprios de 10 dimensões, o que não prejudica a análise em foco.

Tabela 3 – Distribuição dos valores próprios.

Dimensão	Valor Próprio
1	0,778
2	0,433
3	0,348
4	0,263
5	0,250
6	0,250
7	0,091
8	0,060
9	0,040
10	0,005

Fonte: Pesquisador (2012).

Gráfico 1 – Representação das variâncias das dimensões.



Fonte: Pesquisador (2012).

Pelo que se pode observar da Tabela 3 e Gráfico 1, é muito evidente a superioridade das duas primeiras dimensões na explicação do espaço de partida. As duas dimensões juntas representam 1,213 da variação total, o que representa aproximadamente 48,0% de explicação em relação à variação total. Portanto, trabalhar-se-á, neste estudo, com o usual, duas dimensões retidas.

2º) Identificação das dimensões mais importantes

Neste momento, analisarão as variáveis mais importantes em cada uma das dimensões mais importantes. Este procedimento permitirá nomear as dimensões e interpretá-las.

Tabela 4 – Medidas de discriminação.

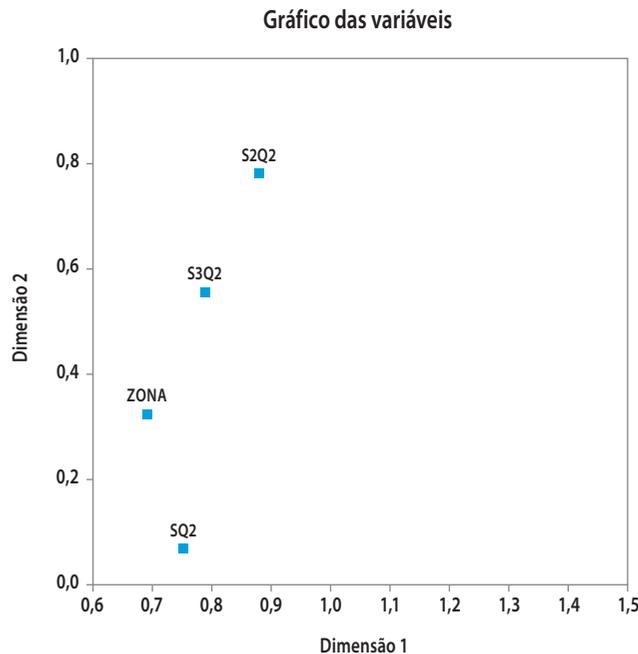
Variáveis	Dimensão	
	1	2
Zona	0,692	0,326
SQ2	0,752	0,071
S2Q2	0,878	0,780
S3Q2	0,789	0,557

Fonte: Pesquisador (2012).

A Tabela 4 apresenta as medidas de discriminação das dimensões retidas para o problema na fase anterior. As variáveis mais importantes em cada uma das dimensões retidas é aquela que possui medida de discriminação maior ou igual a 0,3 em cada uma das dimensões mais importantes. Sendo assim, foram sombreadas em cada uma das dimensões da Tabela 4, as suas variáveis mais importantes. Na dimensão 1, todas as variáveis são importantes. Na dimensão 2, todas as variáveis são relevantes, com exceção da variável sala e quarto (SQ2).

Pelo Gráfico 2, observa-se que as variáveis não se concentram na direção de nenhuma dimensão mais acentuadamente, se concentrando mais na diagonal, o que ratifica as variáveis são importantes igualmente para definição das duas dimensões retidas. Neste estudo não se tem dimensões puras e sim complexas.

Gráfico 2 – Gráfico das variáveis.



Fonte: Pesquisador (2012).

Quase todas as variáveis importantes para a dimensão 1, também são para a dimensão 2. O que significa que não se podem nomear as dimensões ou interpretar o seu significado neste problema, dada a complexidade das configurações das novas variáveis do espaço. O que se pode fazer é estudar como elas agem separadamente sobre cada variável na definição das configurações, o que pode ser obtido pela análise interna das dimensões, o que será feito na próxima etapa da HOMALS.

3º) Identificação das categorias mais importantes

A análise das quantificações das categorias permite avaliar as categorias mais relevantes, necessárias e diferenciadoras dos objetos dentre as variáveis mais significativas.

Será realizada uma análise interna, mais detalhada do significado das dimensões. As dimensões, em função de suas variáveis mais relevantes, estratificam os objetos em análise em grupos distintos. Esta análise procurará evidenciar como cada dimensão organiza, então, em grupos (associados e opostos) os elementos da base de dados. As associações são analisadas a partir da presença de categorias que detenham quantificações com mesmo sinal e as oposições pela presença de categorias que possuem sinais contrários.

Para descrever a dimensão 1, pode-se proceder à análise da Tabela 5. Primeiro, considera-se todas as variáveis importantes para esta dimensão. Somente vão-se considerar as quantificações das categorias destes indicadores. Segundo, consideram-se na coluna da dimensão 1, em cada va-

riável assinalada anteriormente, as categorias afastadas de zero, separando-as em dois blocos: as de sinal negativos, em negrito, vão para o da dimensão $1 < 0$, e as de sinal positivo, para a dimensão $1 > 0$. Raciocínio análogo é feito para se discriminar as categorias mais importantes na dimensão 2. Os resultados desta operação estão registrados nos Quadros 1 e 2. Portanto, somente serão consideradas para a análise as categorias com quantificação altas, isto é, as mais importantes, as que promovem a discriminação dos grupos.

Tabela 5 – Quantificação das categorias.

Zona

Categorias	Freqüências	Quantificação das Categorias	
		1	2
Centro	1	0,834	0,248
Zona Sul	10	-0,838	0,082
Zona Norte	4	0,834	0,248
Zona Oeste	7	0,834	0,248
Barra da Tijuca	2	-0,792	-1,872

Fonte: Pesquisador (2012).

SQ2

Categorias	Freqüências	Quantificação das Categorias	
		1	2
350-750SQ	17	0,552	0,047
1150-1450SQ	6	-1,236	-,301
1950-2750SQ	1	-1,927	1,054

Fonte: Pesquisador (2012).

S2Q2

Categorias	Freqüências	Quantificação das Categorias	
		1	2
450-1450S2Q	15	0,680	0,145
1450-2450S2Q	6	-0,746	-1,232
2450-3450S2Q	3	-1,895	1,759

Fonte: Pesquisador (2012).

S3Q2

Categorias	Freqüências	Quantificação das Categorias	
		1	2
550-2550S3Q2	17	0,528	-0,010
2550-4550S3Q2	4	-1,336	-1,053
4550-6550S3Q2	2	-1,879	2,112

Fonte: Pesquisador (2012).

Abaixo, apresentar-se-ão os quadros resumos da análise interna das dimensões retidas com as categorias mais importantes.

Quadro 1 – Dimensão 1: tipos de discriminação.

Variáveis	Quantificação das Categorias (QC)	
	QC<0	QC>0
Zona SQ2 S2Q2 S3Q2	Zona Sul Barra da Tijuca 1150-1450SQ 1950-2750SQ 1450-2450S2Q 2450-3450S2Q 2550-4550S3Q 4550-6550S3Q	Centro Zona Norte Zona Oeste 350-750SQ 450-1450S2Q 550-2550S3Q

Fonte: Pesquisador (2012).

Observando o Quadro 1, relativamente à primeira dimensão, pode-se dizer que ela divide os bairros em dois grupos opostos. No primeiro grupo ($QC<0$), estão os bairros da zona sul e barra, onde os valores médios de aluguéis residências estão na faixas intermediárias e altas para sala e quarto, sala e dois quartos e sala e três quartos. O segundo grupos ($QC>0$) estão os bairros do centro, zona norte e zona oeste, onde os valores de aluguéis estão nas faixas mais baixas para sala e quarto, sala e dois quartos e sala e três quartos.

Quadro 2 – Dimensão 2: tipos de discriminação.

Variáveis	Quantificação das Categorias (QC)	
	QC<0	QC>0
Zona S2Q2 S3Q2	1450-2450S2Q 2550-4550S3Q	Centro Zona Norte Zona Oeste 2450-3450S2Q 4550-6550S3Q

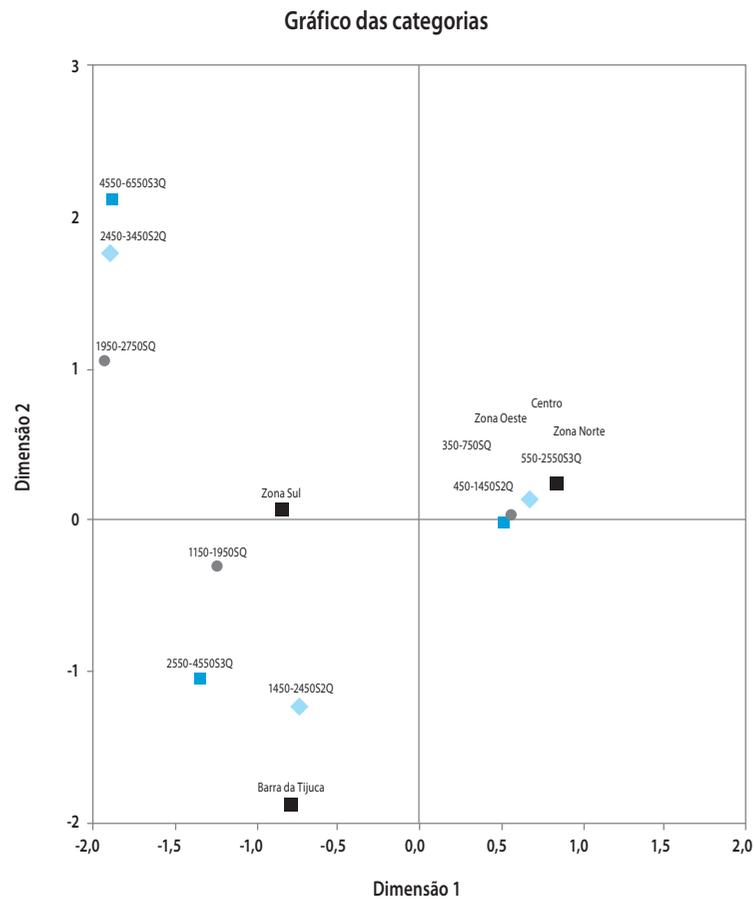
Fonte: Pesquisador (2012).

Pela análise do Quadro 2, a dimensão 2 separa os bairros em dois grupos opostos. O primeiro grupo ($QC<0$) inclui os imóveis da barra com preços na faixa intermediária para sala e dois quartos e sala e três quartos. Em oposição, está o segundo grupo, que inclui os imóveis do centro, zona norte e zona oeste com preços nas faixas mais caras para sala e dois quartos e sala e três quartos.

4º) Estruturação do espaço de análise

Previamente se investigou as dimensões mais importantes, as variáveis mais importantes e as categorias mais importantes para a topologia do espaço de análise. Os próximos passos da análise consistirão da estruturação do espaço dos aluguéis no Rio de Janeiro, no sentido que se procurará evidências de perfis e configurações latentes.

Gráfico 3 – Gráfico das categorias.



Fonte: Pesquisador (2012).

Após analisar o Gráfico 3, observa-se a presença de três configurações ou perfis ou configurações.

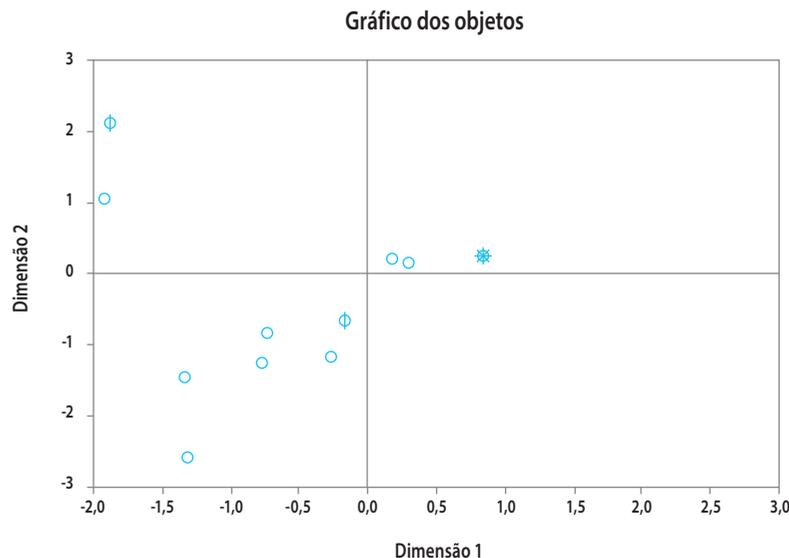
O Perfil 1 ou Configuração 1, localizado no 1º quadrante, é formado pelos bairros com os aluguéis nas faixas mais baixas: 350-1150 reais para sala e quarto, 450-1450 reais para sala e dois quartos e 550-2550 reais para sala e três quartos. A grande maioria dos bairros deste perfil é da zona oeste e zona norte, mas existem também bairros da zona sul neste perfil.

O Perfil 2 ou Configuração 2 observado, localizado no 3º quadrante, é formado pelos bairros com os aluguéis nas faixas intermediárias: 1150-1950 reais para sala e quarto, 1450-2450 reais para sala e dois quartos e 2550-4550 reais para sala e três quartos. Contemplam os bairros da zona sul e barra.

O Perfil 3 ou Configuração 3 observado, localizado no 2º quadrante, é formado pelos bairros com os aluguéis nas faixas mais altas: 1950-2750 reais para sala e quarto, 2450-3450 reais para sala e dois quartos e 4550-6550 reais para sala e três quartos. Contemplam os somente bairros da zona sul.

Com essa etapa, concluiu-se a estruturação do espaço de aluguéis no Rio de Janeiro. Na próxima etapa do processo analítico, observar-se-á as formações de configurações no espaço de análise em função das proximidades dos objetos em análise. Com isso além de se conhecer as características dos grupos identificados, pode também ter a informação preliminar de quem são os objetos em cada uma dessas estruturadas evidenciadas.

Gráfico 4 – Distribuição dos bairros no espaço de Aluguéis.



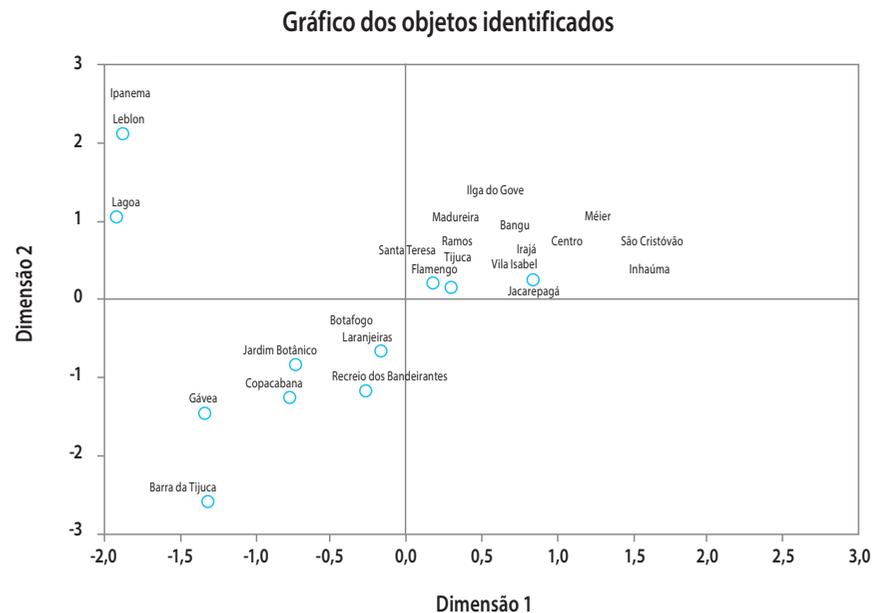
Fonte: Pesquisador (2012).

Pelo Gráfico 4, pode-se visualizar a presença de três configurações. A primeira formada pelos objetos presentes no 1º quadrante. A segunda pelos objetos presentes no 3º quadrante e a última a formada pelos objetos do 2º quadrante.

Pelo Gráfico 5, já se pode identificar quais são os bairros que compreendem cada perfil ou configuração detectada.

O Perfil 1 ou a Configuração 1 é formada pelos bairros de Bangu, Centro, Flamengo, Inhaúma, Ilha do Governador, Irajá, Jacarepaguá, Madureira, Méier, Ramos, Santa Teresa, São Cristóvão, Tijuca, Vila Isabel. O Perfil 2 ou Configuração 2 compreende os bairros de Botafogo, Barra da Tijuca, Copacabana, Gávea, Jardim Botânico, Laranjeiras e Recreio dos Bandeirantes. O Perfil 3 ou a Configuração 3 é formada pelos bairros Lagoa, Leblon e Ipanema.

Gráfico 5 – Distribuição dos bairros no espaço de análise.



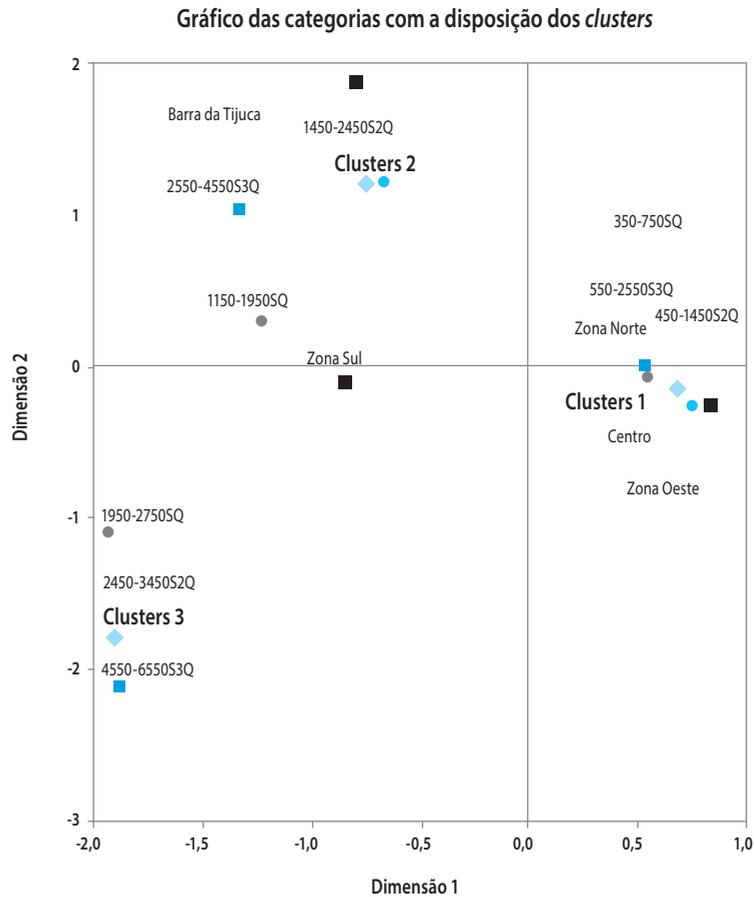
Fonte: Pesquisador (2012).

5º) Articulação análise de homogeneidade versus análise de *cluster*

O objetivo desta seção é associar as configurações identificadas na HOMALS com os *clusters* que uma Análise de *Cluster* geraria, para validar os resultados da Análise de Homogeneidade e também classificar cada bairro da base de dados nos grupos caracterizados na articulação Análise de Homogeneidade/Análise de *Cluster*. Para tanto foi realizada uma análise classificatória das empresas aplicando a Análise de *Cluster*, em dois estágios:

- Método Hierárquico, para validar o número de *clusters* a ser segmentada a base de dados;
- Método não hierárquico *K-Means*, para a classificação propriamente dita.

O resultado da articulação da Análise de Homogeneidade e da Análise de *Cluster* encontra-se no Gráfico 2.6.

Gráfico 6 – Disposição dos *clusters* no espaço de alugueis.

Fonte: Pesquisador (2012).

Tem-se então:

Cluster 1 com 58,3% é formado pelos bairros da zona norte e zona oeste, com alugueis na faixa de 350-750 reais para sala e quarto, 450-1450 reais para sala e dois quartos, 550-2550 reais para sala e três quartos. *Cluster 2* com 29,2% é formado pelos bairros da zona sul e barra da tijuca, na faixa de 350-750 reais para sala e quarto, na faixa 1150-1950 reais para sala e quarto, 1450-2450 reais para sala e dois quartos, 550-2550 reais para sala e três quartos e 2550-4550 reais para sala e três quartos. *Cluster 3* com 12,5% é formado pelos bairros da zona sul, na faixa 1150-1950 reais para sala e quarto, 1950-2750 reais para sala e quarto, 2450-3450 reais para sala e dois quartos, 2550-4550 reais para sala e três quartos e também na faixa de 4550-6550 reais para sala e três quartos.

Estes perfis de *clusters* são muito parecidos com os perfis das configurações da HOMALS. Nota-se que os perfis dos *clusters* são mais dispersos, mais abrangentes que os perfis da HOMALS, principalmente nos *clusters 2* e *3*.

Do ponto de vista técnico, consiste numa representação gráfica na qual estão simultaneamente representadas as variáveis que configuram os perfis (via HOMALS), como variáveis ativas, e a variável que identifica a tipologia de segmentação, tendo esta o status de variável passiva. O objetivo é observar a disposição dos quatro *clusters*. O resultado da operação se encontra no Gráfico 2.1. Percebe-se que as categorias da variável passiva (*cluster 1*, *cluster 2* e *cluster 3*) estão exatamente juntas com as configurações da HOMALS.

5. CONCLUSÃO

Este artigo teve como objetivo apresentar a aplicação da Análise de Homogeneidade como uma alternativa para segmentação de bases de dados com observações multivariadas qualitativas ou tratadas como tal. Pela análise desenvolvida, pôde-se, num espaço de menores dimensões (duas novas variáveis), estruturar o espaço de análise em três perfis distintos de bairros, em função de seus preços de aluguéis médios na cidade do Rio de Janeiro, alcançando o objetivo do trabalho de pesquisa.

A opção pela técnica da Análise de Homogeneidade à amostra selecionada se confirmou principalmente pela constatação da existência de uma alta dimensionalidade de indicadores inter-relacionados, de configuração complexa, uns de natureza originalmente categorial e outros de natureza inicial quantitativa (tratados como qualitativos para padronização do espaço de partida), coexistentes num mesmo universo e que seria relevante reduzir a dimensionalidade, para que se realizasse uma análise estrutural pertinente, com mais nitidez, simplicidade e racionalidade.

A investigação é meramente preliminar, o ideal seria uma amostra com maior representatividade, mas abre caminhos para estudos mais generalizáveis sobre a temática estudada.

Os resultados a que se chegou deixam o campo aberto para outras pesquisas na mesma área, mas que envolvam outras cidades importantes de serem estruturadas em nível de valores de locação, como por exemplo, São Paulo, Salvador e Recife, estas últimas importantes de serem segmentadas em função de suas importâncias turísticas.

Concluindo, está proposta uma alternativa consistente para pesquisas em engenharia de produção, que se precisem realizar análises estruturantes e relacionais de variáveis em escala nominal de mensuração e/ou tratadas como tal.

A nossa expectativa é que estudos futuros sejam realizados para que se estendam os conhecimentos sobre grupos homogêneos em localidades de todo o Brasil.

REFERÊNCIAS

- CARVALHO, H. **Análise Multivariada de Dados Qualitativos** – Utilização da HOMALS com SPSS. Lisboa: Edições Silabo, 2004.
- CARVALHO, H. **Análise Multivariada de Dados Qualitativos** – Utilização da análise de Correspondência Múltipla com SPSS. Lisboa: Edições Silabo, 2008.
- CARVALHO, H. **Homogeneidade e Correspondências Múltiplas**: Comparação de Dois Métodos de Análise. Lisboa: Edições Silabo, 2000.
- GIFI, A. **Nonlinear Multivariate Analysis**. England: John Wiley & Sons, 1996.
- GREENACRE, M. **Theory and Applications of Correspondence Analysis**. London: Academic Press, 1984.
- PÉREZ, C. **Técnicas Estatística Multivariadas com SPSS**. Madrid: Pearson Educación, 2001.

