

UM EXEMPLO DE APLICAÇÃO DE TÉCNICAS DE ESTATÍSTICA MULTIVARIADA NA CONSTRUÇÃO DE ÍNDICES DE PREÇOS

Sueli Aparecida Mingoti¹
Aldy Fernandes da Silva²

RESUMO

Este artigo consiste na ilustração de como as técnicas de Estatística Multivariada podem ser empregadas na construção de índices econômicos e na comparação de estabelecimentos comerciais no que se refere aos preços de produtos praticados. Como exemplo, apresentamos uma análise estatística dos preços praticados pelos diversos hipermercados de Belo Horizonte, Minas Gerais, no que se refere à comercialização de produtos de primeira necessidade, como os gêneros alimentícios, de higiene e de limpeza. Sete dos principais hipermercados de Belo Horizonte foram analisados. A amostra de preços foi coletada no período de 16 de março a 23 de abril de 1994. A partir da construção de índices de preços pela Técnica de Análise de Componentes Principais, foi possível identificar os hipermercados que estavam comercializando seus produtos a preços mais baratos e aqueles com preços mais caros. Uma outra comparação foi efetuada pela Técnica Alternativa de Análise de Conglomerados. O *software* estatístico utilizado para análise foi o SAS.

1 Ph.D. em Estatística – Profa. Adjunto do Departamento de Estatística da UFMG.

2 Mestre em Estatística – Prof. Assistente do Departamento de Estatística da UFMG.

1 INTRODUÇÃO

A procura dos consumidores por produtos de baixo preço tornou-se uma constante no mercado brasileiro. A despesa com o consumo de produtos de primeira necessidade, como os gêneros alimentícios, de higiene e de limpeza, possui grande peso no orçamento familiar surgindo, daí, a necessidade da pesquisa de preços pelo consumidor em vários estabelecimentos. Uma forma de reduzir a despesa com tais produtos está na procura destes em estabelecimentos comerciais de grande porte como os hipermercados. A característica essencial destes estabelecimentos está em seu tamanho e em sua capacidade de compra e estocagem de produtos, fatores que poderiam estar relacionados com a oferta de produtos mais baratos. Um dos problemas encontrados pelos consumidores é que, devido à grande variação de produtos e marcas existentes, torna-se difícil saber onde comprar com menor despesa global. Outro ponto de destaque está na propaganda “enganosa” de alguns destes estabelecimentos que, de uma maneira sutil, escondem dos consumidores os verdadeiros preços de seus produtos, através da promoção de ofertas irrealistas. Este trabalho tem como objetivo apresentar indicadores econômicos, construídos a partir de técnicas estatísticas multivariadas, e que permitem identificar quais hipermercados estão praticando o menor preço global numa determinada época de interesse. Os gêneros considerados foram aqueles classificados como de primeira necessidade. Desse modo, uma amostra dos preços de alguns produtos básicos comercializados por sete dos maiores hipermercados de Belo Horizonte foi analisada e a comparação dos estabelecimentos pesquisados foi efetuada através de índices de preços, elaborados utilizando-se a Técnica de Análise de Componentes Principais (Johnson, Wichern, 1992). Uma comparação dos hipermercados também foi efetuada pela Técnica de Análise de Conglomerados (Everitt, 1993), utilizando-se os preços observados como base para o agrupamento.

2 DEFINIÇÃO DAS VARIÁVEIS E CAMPO DE ATUAÇÃO DA PESQUISA

Belo Horizonte possui cinco importantes hipermercados localizados nas cinco regiões da cidade e duas grandes redes de supermercados. Para proceder-se a este trabalho, uma amostra de preços de produtos básicos foi coletada em sete estabelecimentos de interesse. O termo “hipermercado” foi atribuído aos estabelecimentos analisados para facilitar a linguagem, embora estes apresentem características diferentes entre si. A escolha destes

estabelecimentos deveu-se ao fato de que, além de serem os principais focos de concorrência e atração de consumidores na Região de Belo Horizonte, também apresentavam uma grande variedade de produtos a preços bastante competitivos. A potencialidade de vendas em termos da localização, variedade de produtos comercializados e competitividade de preços dos estabelecimentos foram fatores determinantes na escolha dos pontos de coleta de dados. Por questões de ética, os nomes dos hipermercados pesquisados serão mantidos em sigilo e serão aqui representados pelas letras A-G.

A pesquisa de preços foi realizada no período de 16 de março a 23 de abril de 1994. Os produtos pesquisados foram aqueles considerados como de maior consumo familiar dentro da classe dos gêneros alimentícios, de higiene e de limpeza. Devido à grande quantidade de marcas existentes no mercado, decidiu-se coletar os preços de todas as marcas de produtos encontradas nos hipermercados pesquisados, com exceção daquelas marcas exclusivas e de propriedade do próprio hipermercado. Essas marcas não foram consideradas por serem encontradas somente nos estabelecimentos responsáveis pela produção e comercialização destes produtos e que, naturalmente, podiam comercializá-los a preços mais baixos. Para proceder-se a coleta de dados dividiu-se os 7 hipermercados em 4 grupos e em cada grupo a pesquisa foi realizada em dias diferentes da semana. Os preços dos produtos foram observados na unidade monetária da época, ou seja, em Cruzeiros Reais que, posteriormente, foram convertidos para a URV (Unidade Real de Valor). Devido à grande falta de estabilidade da moeda na época em que esta pesquisa foi conduzida, problema que gerava constantes remarcações e promoções, a coleta de preços para cada hipermercado foi feita durante 6 semanas, em dias alternados, de forma que, no final da coleta, cada produto de cada estabelecimento tivesse o preço pesquisado nos 6 dias da semana (segunda a sábado). A medida utilizada na análise de dados para cada produto foi a média dos preços das marcas encontradas do produto durante o período de coleta da amostra de preços. Desta forma, o número de variáveis foi reduzido evitando-se, assim, trabalhar com marcas específicas de produtos. Uma análise descritiva inicial dos dados indicou que algumas marcas pesquisadas deveriam ser excluídas da análise por não serem representativas do mercado consumidor padrão. De um modo geral, as marcas que foram utilizadas na construção do índice final de preços foram aquelas mais comumente encontradas em todos os hipermercados analisados. As variáveis trabalhadas e seus significados estão apresentadas no Quadro 1.

Quadro 1

VARIÁVEIS UTILIZADAS NA ANÁLISE DE DADOS

Variáveis	Significado
ARROZT1	Preço Médio de Arroz Tipo I – Pac. 5 kg
FEIJAO	Preço Médio de Feijão – Pac. 1 kg
ACREFIN	Preço Médio de Açúcar Refinado – Pac. 1 kg
ACUCRIST	Preço Médio de Açúcar Cristal – Pac. 5 kg
OLEOSOJA	Preço Médio de Óleo de Soja – Lata 900 ml
CAFEEMPO	Preço Médio de Café em Pó – Pac. 500 g
SALREFIN	Preço Médio de Sal Refinado – Pac. 1 kg
FTESPEC	Preço Médio de Farinha de Trigo Especial – Pac. 1 kg
FMCRUAT1	Preço Médio de Farinha de Mandioca Crua Tipo I – Pac. 1 kg
FMTORT1	Preço Médio de Farinha de Mandioca Torrada Tipo I – Pac. 1 kg
FUBA	Preço Médio de Fubá – Pac. 1 kg
EXTTOMAT	Preço Médio de Extrato de Tomate – Lata 270 g
MAIONESE	Preço Médio de Maionese – Vidro 250 g
VINABRAN	Preço Médio de Vinagre Branco – Embalagem de 750 ml
ACHOCOLA	Preço Médio de Achocolatado em Pó – Emb. 500 g
LINTEGLV	Preço Médio de Leite Integral Longa Vida – Emb. 1 litro
OVOSEMB	Preço Médio de Ovos Embalados – 1 dúzia
PAOFORMA	Preço Médio de Pão de Forma Comum – Pac. 500 g
BAYMORES	Preço Médio de Biscoito Aymoré (Maizena Maria e Cream Cracker)
FRANGOR	Preço Médio de Frango Resfriado – Emb. 1 kg
SABBRANC	Preço Médio de Sabonete Branco – Emb. 90 g
PAPHIGFD	Preço Médio de Papel Higiêncio Folha Dupla – Emb. 4 rolos
CREMEDEN	Preço Médio de Creme Dental – Emb. 90 g
FOSFPIN	Preço de Fósforo Pinheiro – Caixa c/ 10
SABBARRA	Preço Médio de Sabão em Barra – Emb. c/ 5
DETERGL	Preço Médio de Detergente Líquido – Emb. 500 ml
DESINFET	Preço Médio de Desinfetante – Emb. 500 ml

3 COMPARAÇÃO DE PREÇOS DOS HIPERMERCADOS ATRAVÉS DE ANÁLISE DE COMPONENTES PRINCIPAIS

O índice de preços foi elaborado através da técnica estatística de Análise de Componentes Principais, aplicada à matriz de correlação amostral das 27 variáveis observadas. Esta técnica estatística foi utilizada por permitir a criação de combinações lineares dos preços dos vários produtos coletados, sendo, portanto, uma ferramenta estatística importante na criação de um índice de preços que nos permitisse a comparação dos hipermercados pesquisados, comparação não só para cada produto individualmente, mas globalmente.

3.1 Análise de componentes principais

Brevemente, a técnica de Análise de Componentes Principais (Dillon, Goldstein, 1984) pode ser descrita da seguinte forma:

Seja $X = (X_1, X_2, \dots, X_p)$ o vetor aleatório contendo as p variáveis de interesse observadas em n elementos amostrais. As componentes principais denotadas por Y_j , $j = 1, 2, \dots, p$ são definidas como:

$$Y_j = \sum_{k=1}^p c_{jk} X_k$$

onde os coeficientes do vetor $c_j = (c_{j1}, c_{j2}, \dots, c_{jp})$ devem ser escolhidos de forma a satisfazerem as seguintes condições:

- i) a $\text{Var}(Y_j)$ é máxima;
- ii) a $\text{Cov}(Y_j, Y_l) = 0$, para qualquer $l \neq j$;
- iii) $c_j' c_j = 1$.

onde $\text{Var}(\cdot)$ denota variância e $\text{Cov}(\cdot, \cdot)$ denota covariância.

A solução do modelo descrito envolve na realidade o conhecimento dos autovalores e autovetores da matriz de covariância (ou de correlação) do vetor aleatório X . Pode ser mostrado que o vetor c_j que satisfaz as condições i) – iii) é simplesmente o autovetor normalizado (ou seja de comprimento igual a 1) correspondente ao autovalor λ_j , da matriz de covariâncias (ou correlação) do vetor X , $j = 1, 2, \dots, p$. Neste caso, as seguintes afirmações são válidas:

- a) a variância da j -ésima componente principal é igual a λ_j ;

$$b) V.T. (X) = \sum_{i=1}^p Var (X_i) = \sum_{j=1}^p Var (Y_j) = V.T (Y),$$

onde $V.T. (.)$ denota variância total, e $Y = (Y_1, Y_2, \dots, Y_p)$ é o vetor contendo as p -componentes principais. Pode ser mos-

$$trado que $V.T. (X) = \sum_{j=1}^p \lambda_j$;$$

c) $Corr (Y_j, Y_l) = 0$, para qualquer $j \neq l$, onde $Corr(.)$ denota o Coeficiente de Correlação Linear de Pearson.

Deste modo, temos que a estrutura de variância e covariância das p variáveis aleatórias do vetor aleatório X é explicada por um conjunto de p combinações lineares destas variáveis, combinações estas que são não correlacionadas e, portanto, mais fáceis de serem entendidas conjuntamente. Na prática a matriz de covariâncias (ou correlação) do vetor X é estimada pela matriz de covariâncias amostral (ou correlação amostral), da qual os autovalores e autovetores normalizados correspondentes são obtidos para a construção das p componentes principais. A razão da variância de cada componente em relação à variância total, é uma medida que descreve a importância de cada componente na explicação da estrutura de variância e covariância das p variáveis aleatórias do vetor X . Quanto maior for o valor desta razão mais importante é a componente.

Para cada componente j , a soma dos coeficientes do vetor c_j não necessariamente é igual a 1, propriedade em geral desejada para os índices, principalmente quando se trata de índices de preços. Para obter tal propriedade basta dividir-se os coeficientes de cada componente principal y_j pela correspondente soma dos coeficientes em c_j . Na análise que se segue preferimos trabalhar com as componentes principais, sem efetuar tal padronização.

3.2 Índice de Preços Global

O Índice de Preços (IP) global obtido utilizando-se a metodologia descrita foi:

$$\begin{aligned} IP = & + 0.231ARROZT1^* + 0.156FEIJAO^* + 0.236ACREFIN^* + 0.193ACUCRIST^* \\ & + 0.263OLEOJOJA^* + 0.215CAFEEMPO^* + 0.279SALREFIN^* + 0.266FTESPEC^* \\ & + 0.089FMCRUAT1^* + 0.146FMTORT1^* + 0.210FUBA^* + 0.079ESTTOMAT^* \\ & + 0.96MAIONESE^* + 0.240VINABRAN^* + 0.239ACHOCOLA^* + 0.050LINTEGLV^* \\ & + 0.075OVOSEMB^* + 0.102PAOFORMA^* + 0.059BAYMORES^* + 0.268FRANGOR^* \\ & + 0.187SABBRANC^* + 0.170PAPHIGFD^* + 0.165CREMEDEN^* + 0.291FOSFPIN^* \\ & + 0.243SABBARRA^* + 0.101DETERGL^* + 0.135DESINFET^* \end{aligned}$$

O símbolo * indica que as variáveis estão padronizadas pela média e desvio padrão amostral. O valor do índice de preços global IP, obtido para cada hipermercado, está apresentado no Quadro 2. De acordo com este índice o hipermercado A foi o que praticou os melhores preços, enquanto o hipermercado G foi o que vendeu mais caro dentre os 7 hipermercados pesquisados. Outros índices de preços específicos para os gêneros dos produtos avaliados foram obtidos. A classificação dos hipermercados de acordo com os valores dos índices de preços global, de produtos alimentícios, de higiene e de limpeza está apresentada no Quadro 2. Pode ser observado que a classificação dos hipermercados não é exatamente a mesma encontrada no índice global. Entretanto, todos os índices construídos indicam que o hipermercado A é aquele que tem preços mais baixos.

Quadro 2

VALOR OBSERVADO DOS ÍNDICES DE PREÇOS ESPECÍFICOS PARA OS HIPERMERCADOS

Hipermercado	Índice Global de Preços	Índice dos produtos alimentícios	Índice dos produtos de higiene	Índice dos produtos de limpeza
A	-4.606	-3.058	-2.967	-2.178
B	-1.643	-1.791	-0.545	-0.761
C	-1.250	-1.161	-0.486	-0.298
D	-0.915	-0.798	0.312	-0.054
E	0.545	-0.752	1.035	0.545
F	2.128	2.176	1.196	1.194
G	5.743	5.384	1.454	1.553

4 COMPARAÇÃO DE PREÇOS DOS HIPERMERCADOS ATRAVÉS DE ANÁLISE DE CONGLOMERADOS

O comportamento dos preços dos hipermercados pesquisados foi também analisado através do Método de Agrupamento das K-médias (Johnson, Wichern, 1992). Os produtos utilizados para formação de grupos foram os mesmos utilizados para a obtenção do índice de preços global tratado na Seção 3.

4.1 Descrição de análise de conglomerados

A análise de conglomerados, também conhecida como análise de grupamentos, tem como objetivo detectar semelhanças e diferenças entre elementos amostrais, dado o conhecimento de um certo número de variáveis medidas em cada elemento amostral. De uma forma geral, as semelhanças e diferenças entre elementos amostrais são representadas pelas distâncias entre eles no espaço das variáveis medidas. Uma distância comumente utilizada é a euclidiana. Brevemente, o método de análise de conglomerados (Anderberg, 1973) pode ser descrito da seguinte forma:

Seja $X = (X_1, X_2, \dots, X_p)$ o conjunto contendo as p variáveis medidas em cada um dos n elementos amostrais, aqui representados por: (E_1, E_2, \dots, E_n) . Com base no conjunto X os n elementos amostrais são agrupados em grupos g_i de modo que:

- i) se E_l e $E_k \in g_i \Rightarrow E_l$ e E_k são semelhantes;
- ii) se E_l e $E_k \notin g_i \Rightarrow E_l$ e E_k são distintos.

O número total de grupos é denotado por g . Existem basicamente dois métodos que podem ser utilizados para obter uma partição dos n elementos amostrais que satisfaça as condições i) e ii). Os métodos aglomerativos hierárquicos são úteis quando o pesquisador está interessado numa análise exploratória dos dados amostrados e não tem, a princípio, uma noção maior de qual é o número total de grupos g que deve ser utilizado para o agrupamento dos n elementos amostrais. Deste modo, a aplicação deste método permite o descobrimento do número de grupos g que resulta na melhor partição dos n elementos amostrais, ou ainda o valor de g mais próximo àquele relacionado com a partição natural dos dados. A segunda classe a que cabe referência é a dos chamados métodos aglomerativos não-hierárquicos, que permitem que o pesquisador escolha *a priori* o número de grupos g . Em outras palavras, o método permite a escolha da melhor partição para um valor de g fixado inicialmente. Dentre os procedimentos não-hierárquicos está o algoritmo das K-Médias, onde K denota o número total de grupos em que se deseja dividir os elementos amostrais e “Médias” vem do fato de que as diferenças entre as médias dos conglomerados são avaliadas ao longo da execução do algoritmo do agrupamento.

4.2 Resultado da análise de conglomerados para os hipermercados avaliados

No Quadro 3 são apresentados os quatro conglomerados formados, a partir dos 27 produtos avaliados. A distância utilizada para o agrupa-

mento dos hipermercados foi a euclidiana. Como pode ser observado, a divisão é semelhante àquela obtida no Quadro 3, a seguir. Os hipermercados que pertencem a um mesmo conglomerado praticam aproximadamente a mesma política de preços. No Quadro 3 os conglomerados estão apresentados na ordem relacionada com a grandeza dos preços praticados globalmente. Deste modo, pode-se observar que os hipermercados que vendem mais barato e mais caro são, respectivamente, os hipermercados A e G. É importante observar também, que a partição dos hipermercados em 4 conglomerados usando apenas os preços de produtos como determinantes para a formação dos grupamentos, reflete claramente a classificação obtida no Quadro 2, quando apenas o índice de preço global foi levado em consideração.

Quadro 3

AGRUPAMENTO DE HIPERMERCADOS

		Conglomerados		
1	2	3	4	
A	B	F	G	
	C			
	D			

5 COMENTÁRIOS FINAIS

Os resultados apresentados servem para ilustrar a importância prática da utilização das técnicas de Estatística Multivariada. Muitas vezes, são realizadas e publicadas em jornais pesquisas de preços em hipermercados com o objetivo de apresentar à população uma escolha de compra entre os diversos estabelecimentos pesquisados. Entretanto, em geral, estas pesquisas não levam em consideração a utilização de técnicas estatísticas que permitem identificar mais globalmente a política de preços praticada pelos hipermercados. As Técnicas de Componentes Principais e Análise de Conglomerados são de fácil utilização e podem levar a comparações com resultados mais conclusivos e precisos.

É importante notar que na comparação dos preços dos hipermercados as duas técnicas, embora essencialmente diferentes, conduziram à mesma conclusão final e refletem, claramente, a política de preços praticada

pelos hipermercados na época em que a pesquisa foi realizada. Na construção dos índices de preços a aplicação da Técnica de Análise de Componentes Principais é uma ferramenta bastante poderosa, pois permite que os próprios dados coletados indiquem os valores dos “pesos” (coeficientes) dos produtos que compõem os índices. Os pesos, assim calculados, caracterizam de uma certa forma a política global de preços que está sendo praticada pelo mercado na época em que os dados estão sendo coletados. Finalmente, vale ressaltar que as duas técnicas tratadas neste artigo não dependem de qualquer suposição sobre distribuição de probabilidade das variáveis observadas, e podem ser encontradas em vários *softwares* estatísticos, como *SAS*, *SPSS*, *MINITAB* etc.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ANDERBERG, M. R. *Cluster analysis for applications*. Academic Press, 1973.

DILLON, W. R., GOLDSTEIN, M. *Multivariate analysis methods and application*. New York : John Wiley, 1984.

EVERITT, B. *Cluster analysis*. 3. ed. London : Heinemann Educational Books, 1993.

JOHNSON, R. A., WICHERN, D. W. *Applied multivariate statistical analysis*. 3. ed. New Jersey : Prentice Hall Inc., 1992.