

ANÁLISE MULTIVARIADA DE INDICADORES SOCIOECONÔMICOS DOS PAÍSES DO G-20

*MULTIVARIATE ANALYSIS FOR SOCIO-ECONOMIC
INDICATORS OF THE G-20 COUNTRIES*

Gil Bracarense Leite

Mestrando em Economia Aplicada pela Universidade Federal de Viçosa (UFV).
Bacharel em Ciências Econômicas pela Universidade Federal de Juiz de Fora (UFJF).
Departamento de Economia Rural, Campus Universitário – Viçosa – MG – CEP 36570-000
E-mail: gilbracarense@yahoo.com.br

Henrique Brigatte

Mestrando em Economia Aplicada pela Universidade Federal de Viçosa (UFV).
Bacharel em Ciências Econômicas pela Universidade Federal de Juiz de Fora (UFJF).
Universidade Federal de Viçosa, Departamento de Economia Rural,
Campus Universitário – Viçosa – MG – CEP 36570-000
E-mail: hbrigatte@yahoo.com.br

Eder Barbosa de Aguiar

Mestre em Economia pela Universidade Federal do Espírito Santo (UFES).
Bacharel em Ciências Econômicas pela Universidade Federal de Juiz de Fora (UFJF).
Rua Humberto de Almeida Francklin, 257, Bairro Universitário,
São Mateus – ES – CEP 29933-415
E-mail: edereconomia@yahoo.com.br

Resumo

O objetivo deste trabalho é apresentar uma análise dos indicadores socioeconômicos dos países do G-20, por análises de componentes principais e de agrupamentos, a fim de detectar associações ou semelhanças entre estas nações. Os resultados obtidos com relação às variáveis que caracterizam aspectos sociais foram de acordo com o esperado, possibilitando a formação de grupos internamente homogêneos, mas heterogêneos entre si. Os demais resultados – relacionados às variáveis econômicas – apresentaram agrupamentos surpreendentes, reunindo nações mais distintas. De modo geral, os países do G-20 confirmam que não é estritamente necessária a condição de que nações que se reúnam num grupo de interesses comuns sejam totalmente homogêneas quanto a seus indicadores socioeconômicos.

Palavras-chave: G-20; Indicadores socioeconômicos; Análise multivariada.

Abstract

This study presents an analysis on the socio-economic indicators of the G-20 countries by principal components and cluster analysis, in order to detect associations or similarities between the nations. The results concerning variables that characterize the social aspects were in line with expectations, allowing the formation of groups internally homogeneous, but heterogeneous among themselves. The other results – related to economic variables – showed unexpected groups, gathering nations with very different characteristics among themselves. Generally, the countries of the G-20 confirm that a lot of dissimilarities in socio-economic indicators may exist between nations meeting a set of common interests in an international organization.

Keywords: G-20; Socio-economic indicators; Multivariate analysis.

1

INTRODUÇÃO

A formação de blocos econômicos por países envolve, via de regra, a semelhança de interesses existente entre os componentes, dada pela predisposição comum a eles de eliminar vicissitudes contextuais que possam lhes trazer prejuízos de grande impacto. É em movimentos desse tipo que se basearam os casos mais conhecidos de agrupamento de nações, que buscaram, na conjunção de esforços, algumas soluções para problemas que dificilmente poderiam ser solucionados se cada um destes Estados decidisse agir por conta própria.

No cenário econômico atual, há diversos blocos econômicos e órgãos multilaterais internacionais espalhados pelo mundo, compostos por nações intencionadas ao cumprimento dos mais diversos objetivos comuns. Dentre os objetivos mais importantes, é possível citar a flexibilização das regras de comércio internacional, as tomadas de decisão de perfil geopolítico, entre outros (KRUGMAN; OBSTFELD, 2005).

Normalmente, pode-se pensar que tais organizações reúnem – ou ao menos deveriam reunir – países com características socioeconômicas similares, sem distorções fortes o bastante a ponto de serem capazes de classificá-los como nações dotadas de realidades equivalentes entre si. Todavia, há diversos exemplos que contradizem essa noção, mostrando que nações marcadas por diferenças entre si podem se tornar parceiras em grupos de interesses econômicos comuns.

Como casos de maior relevância, destacam-se grupos como o Nafta e a União Européia (UE), formados com vistas à liberalização do comércio entre seus países membros. O primeiro – formado por Canadá, Estados Unidos e México – é um exemplo nítido de que países com fortes diferenças entre si têm plenas condições de participar de um acordo mútuo, uma vez que não restam dúvidas acerca das profundas distinções socioeconômicas entre tais parceiros. O segundo também é marcado pelos mesmos fatos, dado que, nos últimos anos, diversos países do Leste Europeu – de menor desenvolvimento que seus parceiros da Europa Ocidental – têm aderido ao acordo. Outro exemplo – talvez o mais importante – é a Organização Mundial do Comércio (OMC), composta por mais de 150 países e destinada à supervisão de acordos comerciais estabelecidos entre estas nações.

Foi no âmbito da quinta conferência ministerial da OMC, realizada em setembro de 2003, em Cancun (México), que nasceu o G-20: um grupo de países em desenvolvimento que concentra sua atenção no tema da agricultura,

o principal na agenda de desenvolvimento da Rodada de Doha¹. Esse bloco é caracterizado pela relevante abrangência geográfica, dado que sua composição reúne vinte e três nações de três continentes diferentes: cinco da África (África do Sul, Egito, Nigéria, Tanzânia e Zimbábue), seis da Ásia (China, Filipinas, Índia, Indonésia, Paquistão e Tailândia) e doze da América Latina (Argentina, Bolívia, Brasil, Chile, Cuba, Equador, Guatemala, México, Paraguai, Peru, Uruguai e Venezuela).

O G-20 foi concebido com o objetivo de abrir espaço para as negociações relacionadas à agricultura no encontro de Cancun, de tal sorte que o processo negociador resultante refletisse o nível de ambição do mandato de Doha e os interesses dos países componentes do G-20. O grupo rapidamente consolidou-se como relevante interlocutor nas discussões ligadas à liberalização do comércio de produtos agrícolas. Tal legitimidade pode ser explicada pela importância considerável dos países membros na economia e no setor da agricultura mundial: 60% da população do planeta está concentrada nestas nações, as quais detêm, ainda, 70% da população rural em todo o mundo e 26% das exportações agrícolas mundiais (G-20, 2008).

Assim como nos outros grupos de nações existentes, o G-20 também guarda semelhanças e diferenças entre os países que o compõem. A análise de tais fenômenos é o objetivo maior deste trabalho. O que se pretende, mais especificamente, é averiguar a existência de grandes diferenças na realidade socioeconômica das nações do G-20, de modo que seja possível separá-las em grupos com características semelhantes. Estudos deste tipo podem ser considerados importantes na medida em que apontam o grau de assimetria envolvendo os parceiros do grupo – o que permite, de certa forma, identificar com qual importância cada nação irá participar da cooperação e que influências esta poderá exercer, dadas suas condições socioeconômicas.

Como foi dito anteriormente, a junção de países com níveis de desenvolvimento distinto não é impeditiva dentro de um mesmo bloco, mas é plausível considerar a possibilidade de que, dentro de um grupo composto por países em processo de desenvolvimento, aqueles dotados de características sociais e econômicas mais favoráveis possam ter um nível mais relevante de representatividade no transcorrer das negociações.

Apesar de contar com a participação apenas de países subdesenvolvidos ou em desenvolvimento, é de esperar que haja distinções de grandes proporções

1 Rodada de negociações comerciais da OMC lançada em novembro de 2001. Concentra-se na diminuição das barreiras comerciais existentes entre os países, com foco voltado para a promoção do livre comércio para as nações em desenvolvimento.

entre alguns países do G-20, principalmente no que se refere às estatísticas sociais. A presença de Estados africanos na organização leva, *a priori*, a se admitir que estes possam formar um grupo homogêneo nos dados sociais; é igualmente admissível que os países do grupo que possuem maior robustez em suas economias, como Brasil, China, Índia e México, guardem maior semelhança entre si, na magnitude dos dados econômicos, do que em relação às demais nações. Os resultados aqui demonstrados posteriormente podem confirmar ou não tais afirmações. Assim, este artigo busca analisar como os países do G-20 estão distribuídos entre si no que tange às suas características socioeconômicas.

As próximas seções do presente trabalho são, então, apresentadas da seguinte forma: em primeiro lugar, são expostas as descrições das variáveis utilizadas para a análise, assim como suas estatísticas descritivas. Em seguida, demonstram-se os procedimentos metodológicos de análise multivariada escolhidos para a obtenção dos resultados. A seguir, são mostrados os resultados das técnicas empreendidas e, por fim, na última seção, estão as principais conclusões do artigo.

2

DADOS E ANÁLISE PRELIMINAR

Com o intuito de estudar o comportamento socioeconômico dos países integrantes do G-20, foram selecionadas 12 variáveis – 7 das quais descrevendo aspectos sociais e 5 representando características econômicas. Entre os indicadores sociais, estão:

- índice de desenvolvimento humano (IDH), com valores entre 0 e 1;
- esperança de vida ao nascer (ESPV), medida em anos;
- calorias consumidas (CALC), medidas em calorias por dia;
- domicílios com acesso à rede sanitária (DARS), medidos em termos percentuais;
- taxa bruta de matrícula para todos os níveis de ensino (TXM), também em termos percentuais;
- usuários com acesso à internet (PANET), indicados pelo número de usuários em cada 100 habitantes;
- linhas telefônicas (PTEL), indicadas pelo número de linhas para cada 100 habitantes.

Os indicadores econômicos, por sua vez, incluem:

- PIB *per capita* (PIBPC), avaliado em US\$;
- população de 15 anos de idade, ou mais, economicamente ativa (PEA), representada em termos percentuais;
- mulheres de 15 anos de idade, ou mais, economicamente ativas (MPEA), também em termos percentuais;
- total de importações (M), em milhões de US\$;
- total de exportações (X), também em milhões de US\$.

Todas essas informações foram obtidas da base de dados do Instituto Brasileiro de Geografia e Estatística (IBGE).

Por fim, os objetos da análise, ou seja, os países do G-20 que foram estudados são: Argentina (ARG), Bolívia (BOL), Brasil (BRA), Chile (CHL), China (CHN), Egito (EGT), Guatemala (GUA), Índia (IND), Indonésia (INS), México (MEX), Nigéria (NIG), Paquistão (PAQ), Paraguai (PAR), Filipinas (FIL), África do Sul (AFS), Tanzânia (TAN), Tailândia (TAI), Uruguai (URU), Venezuela (VEN) e Zimbábue (ZIM)².

Apresentados as variáveis e os objetos, é exposta, a seguir, a análise das estatísticas descritivas e da matriz de correlação, que possibilitarão a extração de importantes conclusões preliminares antes de aplicar-se a metodologia da análise multivariada. A Tabela 1 apresenta as médias, os desvios padrão, os valores máximos e mínimos das variáveis em questão. O número entre parênteses ao lado de cada valor máximo ou mínimo indica o país correspondente.

Ao observarmos as colunas referentes aos valores máximo e mínimo dos indicadores sociais, os destaques positivos são para três países sul-americanos: Argentina, Chile e Uruguai, que, de fato, são nações historicamente reconhecidas como possuidoras de boa qualidade de vida. O único indicador social em que o valor máximo não pertence a um dos países acima citados é “Calorias consumidas”, em que o destaque é o Egito. Nos valores mínimos, os indicadores sociais destacam em quatro ocasiões a Tanzânia, além de Zimbábue, Índia e Paquistão, reforçando o conhecimento que se tem das precárias condições sociais em que se vive na África e em certas partes da Ásia. No caso indiano, a alta densidade populacional pode ser fator de explicação para o baixo valor de domicílios com acesso à rede sanitária.

2 Por insuficiência de dados, foram retirados da análise Cuba, Equador e Peru.

Tabela 1

Média, desvio padrão, valores máximo e mínimo das variáveis socioeconômicas observadas nos países integrantes do G-20

VARIÁVEIS	MÉDIA	DESVIO PADRÃO	VALOR MÁXIMO	VALOR MÍNIMO
IDH	0.70	0.13	0.86 (1 e 4)	0.43 (16)
ESPV	64.82	11.63	77.9 (4)	37.2 (20)
CALC	2633.00	394.45	3350 (6)	1960 (16)
DARS	67.85	19.85	100 (18)	33 (8)
TXM	70.95	14.14	89 (1 e 18)	38 (12)
PANET	10.48	7.14	28.93 (4)	1 (16)
PTEL	11.81	9.23	30.95 (18)	0.4 (16)
PIBPC	3161.60	2710.58	8857 (4)	133 (20)
PEA	66.46	9.39	87.7 (16)	46.8 (6)
MPEA	52.15	15.61	85.6 (16)	20.3 (6)
M	79330.05	148472.99	659953 (5)	2341.35 (2)
X	85263.40	168201.76	761953 (5)	1480.49 (16)

(1) ARG, (2) BOL, (3) BRA, (4) CHL, (5) CHN, (6) EGT, (7) GUA, (8) IND, (9) INS, (10) MEX, (11) NIG, (12) PAQ, (13) PAR, (14) FIL, (15) AFS, (16) TAN, (17) TAI, (18) URU, (19) VEN, (20) ZIM.

Fonte dos dados: IBGE.

Já em relação aos valores máximo e mínimo dos indicadores econômicos, tem-se um comportamento mais dissipado. O Chile se destaca como o maior “PIB *per capita*” dos países, enquanto o menor valor está no Zimbábue, uma diferença relevante já que o PIB *per capita* do país africano representa apenas 1,5% do chileno. Nas variáveis PEA e MPEA, os valores máximos estão a cargo da Tanzânia e os mínimos com o Egito. Em relação às importações e às exportações, os maiores valores pertencem à China – o que é plausível, já que o país é a maior economia entre os componentes do G-20. Já os menores valores, também de acordo com o esperado, ficaram reservados a países de economias menores: Bolívia (“Importações”) e Tanzânia (“Exportações”).

Observando a coluna dos valores médios e dos desvios padrão, destacam-se, entre os indicadores econômicos, as variáveis referentes ao comércio internacional, em que os desvios representam proporcionalmente mais do que 100% da média; o que pode ser explicado pela grande diferença entre o tamanho de economias como Brasil, China, Índia e México em comparação a nações de pequeno porte (populacional e econômico) como Bolívia, Guatemala, Tanzânia e

Zimbábue. Pelas mesmas razões relacionadas ao tamanho das economias envolvidas, o “PIB *per capita*” também apresenta considerável variação, com o desvio representando quase 86% do valor da média. Sobre os indicadores sociais, as variáveis com maior variação foram “Usuários com acesso à internet” (PANET) e “Linhas telefônicas” (PTEL), com os desvios representando, respectivamente, 68% e 78% do valor da média.

Com o objetivo de verificar as associações existentes entre as 12 variáveis escolhidas, a Tabela 2 apresenta uma matriz com as correlações entre elas. A matriz confirma que todos os indicadores sociais são altamente correlacionados entre si. O maior valor, existente para a associação entre as variáveis IDH e ESPV, não surpreende, já que a segunda entra no cálculo da primeira. No entanto, as variáveis econômicas apresentam um comportamento mais individual, com destaque apenas para a alta correlação existente entre X e M – ambas relacionadas ao comércio internacional –, e entre PEA e MEA – ambas variáveis relacionadas com nível de emprego.

Tabela 2

Matriz de correlação das variáveis observadas nos países do G-20

VARIÁVEIS	IDH	ESPV	CALC	DARS	TXM	PANET	PTEL	PIBPC	PEA	MPEA	M	X
IDH	1											
ESPV	0.901	1										
CALC	0.557	0.473	1									
DARS	0.695	0.590	0.289	1								
TXM	0.837	0.633	0.520	0.538	1							
PANET	0.69	0.534	0.493	0.643	0.572	1						
PTEL	0.781	0.665	0.630	0.549	0.668	0.786	1					
PIBPC	0.719	0.565	0.515	0.631	0.591	0.893	0.718	1				
PEA	-0.222	-0.300	-0.544	-0.231	-0.145	-0.362	-0.218	-0.317	1			
MPEA	-0.059	-0.193	-0.439	-0.077	0.018	-0.181	-0.044	-0.152	0.975	1		
M	0.219	0.224	0.315	-0.246	0.027	0.020	0.395	0.044	0.106	0.130	1	
X	0.223	0.218	0.316	-0.242	0.043	0.040	0.422	0.053	0.148	0.177	0.991	1

Obs.: Em negrito estão as correlações significativamente diferentes de zero ao nível de significância de 5%.

Fonte: Resultados da pesquisa.

Intercalando os dois conjuntos de variáveis, há de se destacar a alta associação existente entre o “PIB *per capita*” e todos os indicadores sociais. Outro destaque é a correlação negativa entre “Calorias consumidas” e a variável PEA. Como entre os países com maior nível de população economicamente ativa estão algumas das nações mais pobres do G-20, pode-se supor que esses empregos sejam de baixa qualificação e remuneração, dificultando um maior acesso aos alimentos e justificando a correlação negativa com a variável “Calorias consumidas”.

3 METODOLOGIA

Nesta seção – baseada em Barroso e Artes (2003), Johnson e Wichern (1992) e Anderberg (1973) – contempla-se a metodologia utilizada no presente trabalho, tais como as técnicas de estatística multivariada. Tais técnicas de estatística multivariada têm, como maior objetivo, simplificar e facilitar a interpretação de grandes conjuntos de dados. De forma geral, buscam a redução da base de dados para uma dimensão mais facilmente interpretável, bem como a ordenação, a classificação e o agrupamento das observações.

Neste trabalho, serão utilizadas duas das abordagens de análise multivariada mais comuns: a análise de componentes principais, baseada na matriz de correlações das variáveis em estudo, e a análise de agrupamentos (*clusters*), relacionada com a matriz de distâncias observada entre as variáveis. Tal estratégia é amplamente utilizada em estudos aplicados de análise multivariada; a pesquisa de Kubrusly e Sabóia (2006), por exemplo, lança mão dessa metodologia ao investigar características relacionadas à população ocupada nas regiões metropolitanas brasileiras. As duas próximas subseções trazem breves esclarecimentos acerca das metodologias.

■ 3.1 Análise de componentes principais

A análise de componentes principais se propõe a formar combinações lineares das variáveis originais, de tal forma que estas novas variáveis formem um conjunto de dados mais resumido. Assim, seja o vetor de p variáveis $\mathbf{x} = (X_1, X_2, \dots, X_p)$ com matriz de covariâncias $\text{cov}(\mathbf{x}) = \mathbf{S}$; a técnica será capaz de formar p combinações lineares como descritas a seguir:

$$\begin{aligned}
 Y_1 &= \mathbf{a}_1' \mathbf{x} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\
 Y_2 &= \mathbf{a}_2' \mathbf{x} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\
 &\quad \vdots \\
 Y_p &= \mathbf{a}_p' \mathbf{x} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p
 \end{aligned}$$

Essas p combinações lineares são as componentes principais. A idéia intuitiva por trás deste método é que as componentes principais sejam obtidas de modo que a primeira tenha a maior variância, a segunda tenha a segunda maior variância, seja ortogonal e não correlacionada em relação à primeira e assim sucessivamente. Formalmente, o que se pretende é encontrar a resolução do seguinte problema de maximização condicionada da variância para uma dada i -ésima componente principal representada por $Y_i = \mathbf{a}_i' \mathbf{x} = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$, $i = (1, 2, \dots, p)$:

$$\begin{aligned}
 \max \quad & \text{var}(Y_i) = \mathbf{a}_i' \mathbf{S} \mathbf{a}_i \\
 \text{s.a.} \quad & \mathbf{a}_i' \mathbf{a}_i = 1 \\
 & \mathbf{a}_i' \mathbf{a}_k = 0
 \end{aligned}$$

A solução desse problema envolve a obtenção dos autovalores e autovetores da matriz \mathbf{S} , obedecida a restrição de ortonormalidade destes autovetores.

Usualmente, recomenda-se que as variáveis originais sejam consideradas na forma padronizada, se o objetivo é a utilização das componentes para a construção de *clusters*. Essa padronização confere o mesmo peso para as variáveis. Nesses casos, pode-se usar a matriz de correlações \mathbf{R} na abordagem, no lugar da matriz de covariâncias \mathbf{S} .

É importante que se destaquem algumas propriedades vigentes na análise de componentes principais. Uma delas diz que a variância de uma i -ésima componente principal pode ser dada diretamente pelo i -ésimo autovalor da matriz de correlações. Além disso, os elementos dos autovetores obtidos por meio dos autovalores são exatamente os pesos que formam as combinações lineares das variáveis originais. Citam-se, ainda, outras duas proposições relevantes: a covariância entre as componentes Y_i e Y_k , com $i \neq k$ deve ser igual a zero, indicando que as componentes principais formadas devem ser não correlacionadas; e a variância total das variáveis originais deve ser igual à variância total das componentes principais.

Frequentemente, a maior parte da variância total é explicada pelas primeiras componentes principais. Como a variância do conjunto depende inerentemente das correlações entre as variáveis originais, tem-se que, quanto mais forte for a estrutura dessas correlações, tanto menos componentes serão necessárias para descrever grande parte da variância total.

■ 3.2 Análise de agrupamentos

Esta técnica de análise multivariada busca formar grupos homogêneos com base em características diversas – as variáveis – relacionadas às unidades amostrais, ou seja, aos objetos. Baseia-se no cálculo de medidas de distância, que, por sua vez, atuam como índices de similaridade ou dissimilaridade entre os objetos. Uma das formas mais conhecidas de calcular essa distância é a medida de distância euclidiana. A ilustração da fórmula dessa medida, para o caso geral de p variáveis, considerando-se os indivíduos A e B , é dada a seguir:

$$D_{AB} = \sqrt{\sum_{j=1}^p (X_{jA} - X_{jB})^2}$$

Em termos matriciais, essa fórmula é representada da seguinte forma:

$$D_{AB} = \sqrt{(\mathbf{X}_A - \mathbf{X}_B)'(\mathbf{X}_A - \mathbf{X}_B)}$$

Com base nas distâncias observadas, forma-se uma matriz de distâncias entre os objetos. Neste trabalho, os objetos são os países do G-20; o objetivo da análise, assim, é agrupar estas nações, por meio do método hierárquico aglomerativo de Ward³, de acordo com possíveis semelhanças existentes entre elas, segundo as variáveis em estudo.

A solução dessa análise de agrupamento é, então, apresentada em um diagrama em árvore denominado dendrograma. Nesse diagrama, é possível que se identifiquem os grupos de objetos semelhantes entre si (se eles existirem), associados à referida medida de distância ou similaridade.

3 De acordo com Barroso e Artes (2003), o método de Ward busca, a cada etapa, unir objetos de modo que os grupos formados tenham a maior homogeneidade possível. A medida de homogeneidade utilizada baseia-se na partição da soma de quadrados total de uma análise de variância.

4

RESULTADOS E DISCUSSÃO

A seguir, apresentam-se os resultados obtidos das análises de componentes principais e de agrupamentos. Primeiramente, são mostrados os argumentos extraídos da análise de componentes principais, por meio da exposição da proporção da variância explicada por cada componente e, em seguida, por meio da demonstração das correlações das componentes principais mais relevantes com as variáveis em estudo. Posteriormente, a análise de agrupamentos mostra, por meio dos dendrogramas construídos para as componentes principais mais relevantes, as semelhanças e diferenças existentes entre os países do G-20. Estes resultados foram obtidos utilizando-se os *softwares* Eviews 5.0 e SPSS 11.0.

■ 4.1 Resultados da análise de componentes principais

Os resultados da análise de componentes principais foram extraídos considerando-se todas as doze variáveis e os vinte países do G-20, para a obtenção das conclusões. Além disso, é válido ressaltar, a título de recordação, que as variâncias de cada uma das componentes principais obtidas correspondem exatamente aos autovalores da matriz de correlações equivalentes a cada componente. Ademais, a proporção da variância total que é explicada em cada uma dessas componentes é simplesmente a divisão da variância correspondente à componente em questão pela variância total.

Assim, apresentam-se, na Tabela 3, os resultados relacionados à variância de cada componente principal, bem como sua contribuição para a variância total dos dados. Como se pode perceber pela coluna correspondente à porcentagem acumulada da variância explicada pelas componentes, a análise recomenda que se considerem apenas as três primeiras componentes principais, ou seja, CP1, CP2 e CP3. Isto se justifica pelo fato de que elas contribuem, conjuntamente, com mais de 80% de explicação da variância total (83,642%, mais especificamente); a proporção superior a 80% de explicação da variância como parâmetro de escolha do número de componentes principais a ser considerado pela análise é utilizada como referência na literatura.

Há duas outras explanações fornecidas pela teoria da análise estatística multivariada que justificam a escolha pelas três primeiras componentes:

1. Observa-se, pela terceira coluna da Tabela 3, que a porcentagem da variância total explicada pelas demais componentes é bem inferior à proporção de explicação fornecida pelas componentes escolhidas.

2. De acordo com uma regra teórica que direciona a escolha das componentes àquelas cujos autovalores correspondentes sejam superiores à unidade, vê-se, pela segunda coluna da Tabela 3, que, da quarta componente em diante, tal condição não pode ser satisfeita.

Dessa forma, com apoio na teoria da análise multivariada, a escolha recaiu sobre as três primeiras componentes principais, considerando-se, portanto, que estas são suficientes para uma descrição adequada dos dados.

Tabela 3

Variância das componentes principais e sua participação percentual sobre a variância total dos dados das variáveis observadas nos países do G-20

COMPONENTES	VARIÂNCIA DO CP	% DA VARIÂNCIA TOTAL	% ACUMULADA
CP1	5.731	47.762	47.762
CP2	2.441	20.343	68.105
CP3	1.865	15.538	83.642
CP4	0.702	5.849	89.491
CP5	0.515	4.292	93.783
CP6	0.299	2.490	96.273
CP7	0.202	1.684	97.957
CP8	0.151	1.261	99.218
CP9	0.068	0.568	99.786
CP10	0.016	0.136	99.922
CP11	0.006	0.050	99.972
CP12	0.003	0.028	100.000

Fonte: Resultados da pesquisa.

A Tabela 4, em contrapartida, fornece as correlações detectadas entre os escores das componentes selecionadas e as variáveis de análise. As correlações destacadas em negrito são as de maior magnitude para cada uma das componentes, e servem para orientar uma caracterização mais precisa das referidas combinações lineares, que será importante na formulação da análise de agrupamentos apresentada na próxima seção.

Tabela 4

Correlação entre as componentes principais escolhidas e as variáveis observadas nos países do G-20

VARIÁVEIS	CP1	CP2	CP3
IDH	-0.925	0.112	0.194
ESPV	-0.821	0.074	0.044
CALC	-0.722	0.044	-0.450
DARS	-0.706	-0.322	0.407
TXM	-0.782	0.018	0.286
PANET	-0.846	-0.133	0.131
PTEL	-0.885	0.274	0.031
PIBPC	-0.846	-0.104	0.149
PEA	0.442	0.563	0.684
MPEA	0.266	0.578	0.756
M	-0.211	0.873	-0.404
X	-0.216	0.895	-0.363

Fonte: Resultados da pesquisa.

Como fica claro pela Tabela 4, a análise de componentes principais resulta na formação de três combinações lineares de distintos significados socioeconômicos. A primeira componente (CP1) reúne todo o conjunto de variáveis representando características sociais (IDH, ESPV, CALC, DARS, TXM, PANET e PTEL) e, ainda, uma variável representando uma característica econômica, dada por PIBPC. Todavia, é razoável afirmar que o PIB *per capita* também enfatiza, em seu bojo, um dado social, uma vez que ele acaba por fornecer a renda média da população de um país e, em última instância, sua própria condição de subsistência.

Os sinais dos coeficientes da CP1 – todos negativos – indicam que não há qualquer oposição entre as variáveis que a compõem, ou seja, todas elas caminham no mesmo sentido, dada determinada variação em uma delas. Isso apenas confirma os dados fornecidos anteriormente pela Tabela 2, de modo que as variáveis da CP1 possuem forte correlação positiva entre si. Tal componente pode, assim, ser caracterizada como a componente social da presente análise, diferenciando os países de melhores indicadores sociais das nações com dados sociais mais desfavoráveis.

As outras duas componentes são as componentes econômicas da análise. A segunda (CP2), por reunir as variáveis de comércio internacional (M e X), pode ser caracterizada como a componente de comércio internacional; da mesma forma que a CP1, as correlações também confirmam as informações disponibilizadas pela Tabela 2, de modo que há relevante associação positiva entre o volume de importações e o volume de exportações das nações. Isso permite afirmar que os países com números generosos relacionados à importação devem ser participantes ativos do comércio internacional, efetuando, portanto, níveis igualmente generosos de exportação.

A terceira componente (CP3) pode ser denominada como a componente de emprego, uma vez que é uma combinação linear das variáveis relacionadas às proporções da população economicamente ativa geral (PEA) e feminina (MPEA) dos países do G-20. Novamente, os sinais dos coeficientes da CP3 corroboram as correlações encontradas na Tabela 2. Isso significa que há relevante associação positiva entre a proporção da população economicamente ativa de um país do G-20 e entre a proporção da população feminina economicamente ativa deste país – o que, igualmente, não deixa de ser algo plausível.

■ 4.2 Resultados da análise de agrupamento

A análise de componentes principais desenvolvida na subseção anterior evidenciou a existência de três conjuntos de variáveis que fornecem as componentes que representam a base de dados analisada: componente social, componente de comércio internacional e componente de emprego. Desse modo, mantendo tal divisão na aplicação do método da análise de agrupamento, obtêm-se resultados para cada um dos três conjuntos de variáveis. O agrupamento baseado na primeira componente indica países semelhantes quanto aos indicadores sociais; a segunda análise mostra as nações que se assemelham quanto às variáveis de comércio internacional, o que também ocorre para as variáveis que representam o emprego. As figuras 1, 2 e 3 mostram os dendrogramas, que servem como soluções das análises. O eixo horizontal fornece as distâncias, que, conforme já ressaltado, são interpretadas como as medidas de similaridade entre os países.

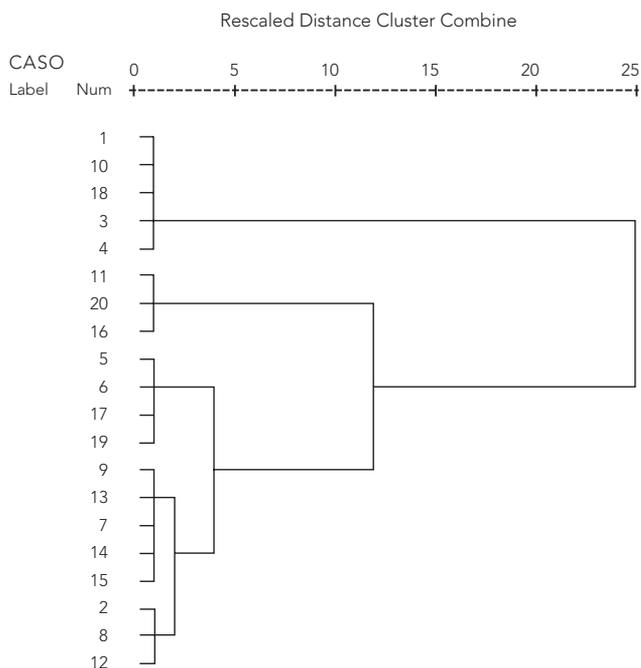
Análise 1 – variáveis sociais

A Figura 1 apresenta o dendrograma construído com base nas semelhanças referentes às variáveis sociais. Percebe-se a formação de três grupos:

- Primeiro grupo, que representa as nações do G-20 com melhores índices sociais e é composto por alguns países da América Latina (Argentina, Brasil, Chile, México e Uruguai).
- Segundo grupo, que se destaca quanto à homogeneidade dos indicadores sociais, sendo formado por três nações africanas que apresentam os piores índices (Nigéria, Tanzânia e Zimbábue).
- Terceiro grupo, que se caracteriza por ser maior e mais heterogêneo quanto à região dos países que a compõem, incluindo as nações asiáticas (China, Índia, Indonésia, Paquistão, Filipinas e Tailândia), nações latino-americanas menos desenvolvidas que aquelas do primeiro grupo (Bolívia, Paraguai, Guatemala e Venezuela), além de dois países africanos de economia mais forte que as nações do segundo grupo (África do Sul e Egito).

Figura 1

Análise de agrupamento das variáveis sociais nos países do G-20



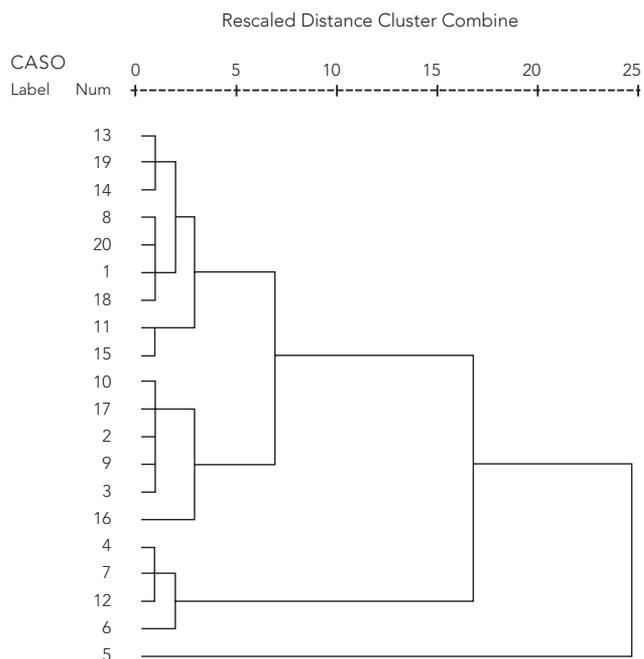
Fonte: Resultados da Pesquisa.

Análise 2 – variáveis de comércio internacional

A Figura 2 apresenta o resultado da análise de agrupamento usando como critério a componente de comércio internacional. O dendrogama indica a formação de quatro grupos. Percebe-se uma forte distinção entre a China e os demais, pois este país forma um grupo isolado e indica o tamanho de sua economia e sua participação evidentemente mais relevante no comércio internacional em relação aos demais países em foco. A formação dos três grupos restantes parece não seguir qualquer padrão regional ou econômico, já que nações de grande disparidade socioeconômica estão juntas, fazendo com que a construção dos agrupamentos não seja tão clara quanto aquela verificada nos indicadores sociais.

Figura 2

Análise de agrupamento das variáveis de comércio internacional nos países do G-20



Fonte: Resultados da Pesquisa.

Dessa forma, um segundo grupo é formado por Chile, Egito, Guatemala e Paquistão. Outro grupo indica semelhanças quanto ao comércio internacional nestes países: Bolívia, Brasil, Indonésia, México, Tanzânia e Tailândia. Por último, um grupo mais extenso inclui as seguintes nações: Argentina, Índia, Nigéria, Paraguai, Filipinas, África do Sul, Uruguai, Venezuela e Zimbábue.

Essa composição dos grupos indica a dificuldade de uma análise mais precisa, uma vez que estes agrupamentos reúnem economias bastante diferentes. Este resultado também se mostra de difícil interpretação no sentido de que os grupos são formados por países com diferenças consideráveis não só nos volumes de importação e exportação, mas também nos próprios saldos de suas balanças comerciais – países deficitários e superavitários dividem os mesmos grupos – ocasionando, assim, certa limitação na tentativa de obter conclusões mais precisas acerca da análise.

Análise 3 – variáveis de emprego

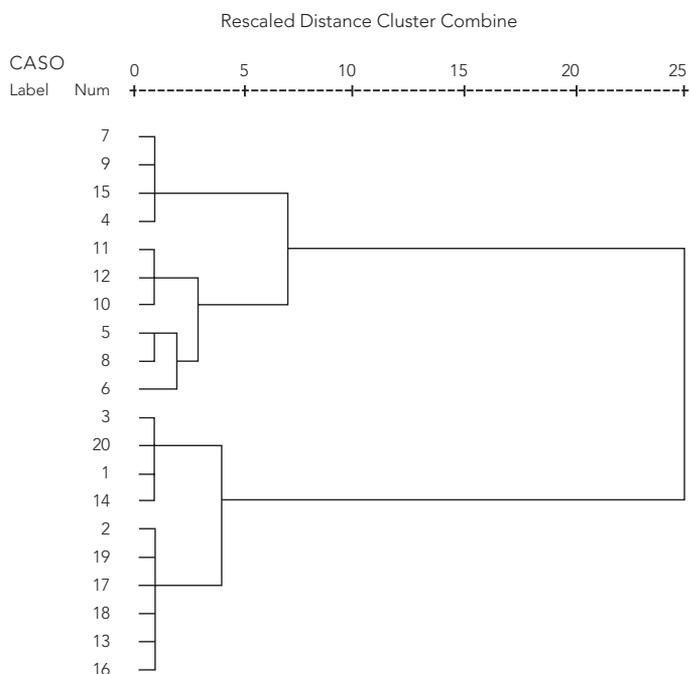
A Figura 3 mostra o dendrograma formado de acordo com as similaridades que os países apresentam no que tange à componente de emprego. Assim como nas variáveis de comércio, não é possível extrair uma análise tão clara quanto aquela vista nos indicadores sociais, nos quais os grupos tinham nações das mesmas regiões e/ou com níveis de desenvolvimento semelhantes.

Além disso, como os agrupamentos foram formados tendo como base somente as porcentagens das variáveis PEA e MPEA, uma análise mais profunda teria que estar pautada no conhecimento de cada mercado de trabalho – o que é difícil, pois envolveria características econômicas próprias de cada nação, inclusive aspectos sociais, culturais e religiosos. De qualquer forma, o dendrograma indica a formação de três grupos:

- Primeiro grupo, que indica a semelhança quanto à componente de emprego das nações Chile, Guatemala, Indonésia e África do Sul.
- Segundo grupo, que inclui China, Egito, Índia, México, Nigéria e Paquistão.
- Terceiro grupo, que é mais extenso e formado por Brasil, Argentina, Bolívia, Paraguai, Filipinas, Tanzânia, Tailândia, Uruguai, Venezuela e Zimbábue.

Figura 3

Análise de agrupamento das variáveis de emprego nos países do G-20



Fonte: Resultados da Pesquisa.

■ 4.3 Interpretação global dos resultados

A análise multivariada aplicada aos dados estudados permitiu a obtenção de alguns resultados interessantes, do ponto de vista social, e relativamente ambíguos, quanto às variáveis econômicas. A princípio, o mecanismo utilizado – a técnica de componentes principais – levou à separação das variáveis em três grupos, cada um dos quais representando diferentes características socioeconômicas. A primeira componente principal relaciona-se às variáveis sociais, enquanto as outras duas referem-se aos dados econômicos dos países observados no presente estudo.

Com relação à análise de agrupamentos realizada para a primeira componente principal, pode-se afirmar que os resultados encontrados são os esperados. De acordo com a análise empreendida, os três grupos constituídos refletem similaridades bastante plausíveis entre os elementos que os compõem. O grupo formado por Argentina, Brasil, Chile, México e Uruguai reúne as nações com melhores indicadores sociais, dentre aquelas que estão dentro do G-20. Nesses países, observa-se que praticamente todas as variáveis que denotam aspectos relacionados à qualidade de vida da população possuem valores que representam superioridade social em comparação com as demais nações, como IDH, PIB *per capita*, entre outras⁴.

As nações africanas que formam um grupo homogêneo no que tange aos dados sociais – Nigéria, Tanzânia e Zimbábue – são, efetivamente, países cuja qualidade de vida de suas populações é notoriamente baixa. Nelas, observam-se os piores índices sociais, o que já era algo esperado acerca da presente análise.

O outro grupo é nitidamente mais heterogêneo do que os dois anteriormente citados, reunindo países com características distintas entre si. Não obstante, é possível visualizar determinadas tendências que justifiquem a presença dessas nações nesse agrupamento. Como exemplo mais claro, são apontados todos os países asiáticos, que, apesar de possuírem economias fortes, ainda não demonstram números tão favoráveis nas variáveis sociais, o que talvez seja reflexo do fato de serem bastante populosos.

Os outros Estados desse agrupamento também podem ser analisados pelo mesmo argumento: os representantes latino-americanos apresentam economias mais fracas do que seus vizinhos do primeiro agrupamento analisado; da mesma forma, é fácil observar que África do Sul e Egito são países africanos com economias evidentemente mais desenvolvidas do que os componentes do segundo agrupamento. A análise relacionada a esse grupo permite, de certa forma, classificar tais nações numa posição de desenvolvimento intermediário, ou seja, entre aquelas com melhores indicadores sociais e aquelas com dados de qualidade de vida desfavoráveis, dentro do G-20.

Quanto aos demais agrupamentos obtidos – relacionados aos dados econômicos –, estes não transmitem, com tanta clareza, justificativas referentes à formação dos *clusters*. No agrupamento extraído para a componente de comércio internacional, destaca-se, como já esperado, a China, como um grupo isolado. Isso denota a grande disparidade da corrente de comércio chinesa,

4 Kuczynski e Williamson (2004), ao argumentarem que, apesar do decepcionante crescimento econômico da América Latina, observam um grande progresso, ao longo das últimas duas décadas, na melhoria dos padrões de longevidade, nutrição, saúde pública e alfabetização, mantendo, inclusive, suas médias acima das da Ásia Oriental.

em comparação com a dos demais países do G-20. Os outros três grupos relacionados ao comércio internacional reúnem países com características similares e dissimilares entre si.

O agrupamento observado para a componente principal de emprego tem um perfil semelhante ao agrupamento de países feito com relação ao comércio internacional, ou seja, também não possibilita justificativas de visualização tão evidente para a constituição dos *clusters*.

Nesses conjuntos, são observadas nações que guardam, entre si, diversas diferenças de ordem regional, econômica e cultural. Isso pode ser visto de forma mais nítida no grupo que agrega Brasil e Zimbábue – países que, a despeito de suas características internas se distinguem de forma marcante, têm números relativamente próximos quanto à composição de suas populações economicamente ativas. Tal semelhança quanto aos dados sobre o mercado de trabalho, todavia, pode estar ligada a fatores inatos à realidade de cada um desses países, fatores estes não contemplados na análise desenvolvida no presente trabalho⁵.

Para contornar tais limitações, procedeu-se à técnica de agrupamento não hierárquica das k-Médias⁶. O intuito principal dessa ação foi propiciar grupos que reunissem países com características ainda mais homogêneas, em comparação com os obtidos pelo critério de Ward. A consecução desse objetivo, porém, não pôde ser efetivada, dado que os grupos formados para cada uma das componentes principais, utilizando-se o método das k-Médias, mostraram-se praticamente idênticos aos fornecidos pelo método de Ward. Assim, optou-se por desenvolver a análise final seguindo esta última técnica.

5 CONCLUSÕES

Este artigo buscou analisar os países do G-20 considerando um conjunto de doze variáveis socioeconômicas, para constatar possíveis semelhanças e diferenças entre tais nações, em relação às características representadas por

5 Um estudo publicado por Jones (2000) ressalta esta conclusão: ao comparar diversas estatísticas de crescimento e desenvolvimento econômico para vários países, o autor constata que a taxa de participação da mão-de-obra nos Estados Unidos e em Uganda é a mesma, de 49%.

6 Esta técnica é uma das mais conhecidas e utilizadas em problemas práticos. Segundo Mingoti (2005), nela cada elemento amostral é alocado ao *cluster* cujo centróide (vetor de médias amostral) é o mais próximo do vetor de valores observados para o respectivo elemento.

estas variáveis. As técnicas de análise multivariada utilizadas para a obtenção das conclusões foram as de componentes principais – destinadas a formar combinações lineares das variáveis em estudo para cada unidade, ou seja, para cada nação do G-20 –, seguidas da análise de agrupamentos, que procurou formar grupos com base nos escores das componentes principais constituídas na análise precedente, com países de características similares entre si.

Os resultados para a componente principal representando características sociais corresponderam às expectativas, reunindo países em grupos bastante homogêneos não só quanto aos seus aspectos sociais, mas também quanto aos aspectos regionais e econômicos. Em contrapartida, os grupos formados para as componentes de comércio internacional e de emprego foram relativamente mais heterogêneos, agregando nações com características regionais e econômicas notoriamente mais diferenciadas em comparação com o agrupamento realizado para a componente social.

A constituição dos grupos na presente análise corrobora a afirmação de que existem diferenças relevantes entre países que compõem um mesmo grupo de interesse econômico. Fica claro que a existência de tais diferenças – profundamente marcantes entre as nações estudadas – não impede que estes países atuem de forma mútua em torno de interesses comuns no cenário econômico internacional. Pode-se dizer, por fim, que essas distinções são ainda mais relativizadas no caso dos países do G-20, dado que elas pertencem, em geral, ao mundo em desenvolvimento.

Referências

- ANDERBERG, M. R. *Cluster analysis for applications*. New York: Academic, 1973.
- BARROSO, L. P.; ARTES, R. *Análise multivariada*. Lavras: UFLA, 2003.
- G-20. *Histórico do G-20*. Disponível em: <http://www.g-20.mre.gov.br/history_port.asp>. Acesso em: 20 mar. 2008.
- IBGE. Banco de dados. Disponível em: <<http://www.ibge.gov.br/paisesat/>>. Acesso em: 19 mar. 2008.
- JOHNSON, R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. New Jersey: Prentice-Hall, 1992.
- JONES, C. I. *Teoria do crescimento econômico*. Rio de Janeiro: Campus, 2000.
- KRUGMAN, P.; OBSTFELD, M. *Economia internacional: teoria e política*. São Paulo: Makron Books, 2005.

- *Análise multivariada de indicadores socioeconômicos dos países do G-20*, Gil Bracarense Leite, Henrique Brigatte, Eder Barbosa de Aguiar

KUBRUSLY, M. S.; SABÓIA, J. Uma análise multivariada da população ocupada nas regiões metropolitanas brasileiras. *Ensaio FEE*, Porto Alegre, v. 27, n. 2, p. 411-436, 2006.

KUCZYNSKI, P. P.; WILLIAMSON, J. *Depois do consenso de Washington*: retomando o crescimento e a reforma na América Latina. São Paulo: Saraiva, 2004.

MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada*: uma abordagem aplicada. Belo Horizonte: Ed. UFMG, 2005.

SPSS, *Base 11.0*: user's guide. Chicago: SPSS Inc., 2002.