

# INDICADORES SOCIOECONÔMICOS DOS PAÍSES DO CONTINENTE AMERICANO

## *SOCIOECONOMIC INDICATORS OF COUNTRIES OF THE AMERICAS*

*Paloma Santana Moreira Pais*<sup>(\*)</sup>

*Felipe Figueiredo da Silva*<sup>(\*\*)</sup>

*João Eustáquio de Lima*<sup>(\*\*\*)</sup>

**Resumo:** Este trabalho objetiva avaliar os indicadores socioeconômicos dos países americanos a partir da análise multivariada. As técnicas utilizadas foram a análise de componentes principais e de *cluster*. Os resultados obtidos destacam, de um lado, Canadá e Estados Unidos, por apresentarem melhores indicadores e, de outro, Haiti, com o pior escore e precárias condições socioeconômicas. O Brasil assumiu a oitava posição na classificação geral e situou-se em um grupo com países como Uruguai, Chile e Argentina. Quanto às condições do mercado de trabalho, os resultados não foram muito conclusivos, pela ausência de conhecimento das especificidades desses países.

**Palavras-chave:** Indicadores socioeconômicos. América. Análise de componentes principais. *Cluster*.

**Abstract:** This paper aims to evaluate the socioeconomic indicators of countries in the Americas from a multivariate analysis. The techniques used were the principal component analysis and cluster analysis. The results obtained out, of one side, Canada and United States, for presenting the best indicators and, on the other, Haiti, with the worst score and poor socioeconomic conditions. Brazil took the eighth position overall and stood in a group of with countries like Uruguay, Chile and Argentina. As to the conditions of labor market, the results were not very conclusive, due to lack of knowledge about the specifics of these countries.

**Key-words:** Socioeconomic indicators. America. Principal component analysis. Cluster.

---

(\*) Bacharel em Gestão do Agronegócio pela Universidade Federal de Viçosa; Mestre e Doutoranda do Programa de Pós-Graduação em Economia Aplicada, do Departamento de Economia Rural da Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brasil. *E-mail:* <palomapais@hotmail.com>.

(\*\*) Bacharel em Economia pela Universidade Estadual de Feira de Santana; Mestre e Doutorando do Programa de Pós-Graduação em Economia Aplicada do Departamento de Economia Rural da Universidade Federal de Viçosa. Viçosa, Minas Gerais, Brasil. *E-mail:* <fsilva.f@hotmail.com>.

(\*\*\*) Eng.-Agr., 1969; M.S. Economia Rural, 1972, UFV; Ph.D. Economia Rural, 1977, Michigan State University (USA); Pós-Doctor Métodos Quantitativos, 1996, University of Florida (USA); Professor Titular (Econometria) pelo Departamento de Economia Rural da Universidade Federal de Viçosa. Viçosa, Minas Gerais, Brasil. *E-mail:* <jelima@ufv.br>. Recebido em 23.08.2012 e aceito em 15.10.2012.

## 1. INTRODUÇÃO

No continente americano, as desigualdades de renda, emprego, acesso à educação e saúde são enormes. Para relatar tais disparidades, tem-se, por exemplo, que o PIB *per capita* de 2009 para os Estados Unidos foi de US\$ 44.872,00, enquanto o mesmo dado para o Haiti foi de apenas US\$ 626,00 (IBGE, 2011). Analisando apenas a América Latina, a qual não inclui apenas cinco países do continente americano como um todo, que são Canadá, Estados Unidos, Guiana, Suriname e Belize, tem-se que esta é considerada a região do mundo com a maior desigualdade (PRADOS DE LA ESCOSURA, 2005). Apenas algumas sociedades da África têm desigualdade mais elevada do que em alguns países latino-americanos (RIBEIRO, 2006).

As taxas elevadas de desigualdade têm altos custos para a região. Para Ribeiro (2006), a desigualdade aumenta a pobreza, diminui os impactos do crescimento econômico na redução da pobreza, dificulta o crescimento econômico (especialmente quando associado à educação e acesso ao crédito) e aumenta as chances de conflitos sociais.

Considerando as divisões territoriais do continente, as desigualdades de renda entre os países tornam-se menores. Na América do Norte, o PIB *per capita* dos Estados Unidos em 2009 possuiu valor próximo ao do Canadá, e o México possuiu um valor inferior aos demais, US\$ 7.956,00, porém, muito superior ao do Haiti, quando se analisa o continente americano como um todo (IBGE, 2011).

No que tange à América Central, o maior PIB *per capita* em 2009 pertenceu às Bahamas, US\$ 20.711,00, e o menor valor foi o do Haiti (IBGE, 2011). Nessa porção da América, apenas cinco países, que são Antígua e Barbuda, Bahamas, Barbados, São Cristóvão e Nevis e Trinidad e Tobago, possuíram PIB *per capita* superior a US\$ 10.000,00 em 2009, entre os vinte países que constituem a região.

Em relação à América do Sul, o maior PIB *per capita* em 2009 pertenceu à Venezuela, US\$ 11.404,00, e o menor valor desse indicador coube à Bolívia, US\$ 1.758,00 (IBGE, 2011). Entre os doze países que compõem a região, apenas a Venezuela possuiu PIB *per capita* superior a US\$ 10.000,00 naquele ano.

Com base no exposto, este artigo busca analisar as diferenças em alguns indicadores sociais, econômicos e de condições básicas dos países do continente americano e separar os países em grupos com características mais homogêneas. Além disso, pretende-se classificar os países com base nos indicadores socioeconômicos, a fim de identificar aqueles países que possuem piores resultados e que, portanto, necessitam de maior assistência internacional.

Este trabalho está estruturado em quatro seções, além desta introdução. Na segunda seção, expõe-se uma revisão da literatura sobre países e suas condições sociais e econômicas. Na seção três são apresentados os procedimentos metodológicos da análise multivariada, bem como a descrição das variáveis utilizadas. Na quarta, os resultados obtidos são analisados e discutidos, além de serem comparados com as estatísticas descritivas das variáveis empregadas. Finalmente, na última seção, são apresentadas algumas considerações sobre o estudo.

## 2. REVISÃO DE LITERATURA

Estudos que analisam as condições sociais e econômicas dos países são amplamente utilizados na literatura. O presente trabalho baseou-se em três estudos principais, os quais também analisaram e compararam países com base em diferenças nesses indicadores. Inicialmente, têm-se os artigos de Curi (1991; 1993), em que o autor buscou agrupar os países do mundo com base em indicadores básicos e econômicos e indicadores de padrão de vida, e no artigo de Leite, Brigatte e Aguilar (2009), os quais agruparam os países do G-20 a partir de indicadores socioeconômicos.

Em seu estudo, Curi (1991) encontrou dois componentes principais para agrupar os 36 países a partir da análise dos indicadores básicos e econômicos, fazendo uma análise individual para cada região geográfica. O autor nomeou o primeiro componente como “grau de desenvolvimento”, a partir do qual formou quatro grupos com características semelhantes e ordenou os 25 países americanos. Nesta ordenação, Estados Unidos e Canadá foram os grupos que apresentaram melhores condições, sendo acompanhados por Cuba, Trinidad e Tobago, Uruguai e Chile. Haiti e Bolívia ocuparam o extremo oposto, apresentando os piores resultados.

Ao analisar os indicadores de padrão de vida, Curi (1993) considerou 125 países. Neste trabalho, o autor não realizou análises distintas para cada continente, mas seus resultados corroboram a sua análise feita anteriormente, embora tenha utilizado indicadores diferentes. Para os países americanos, tem-se que Estados Unidos e Canadá novamente apresentaram os melhores indicadores, acompanhados por Cuba, Argentina, Chile e Trinidad e Tobago. Bolívia e Haiti também apresentaram as piores classificações para o continente americano.

Quanto ao trabalho de Leite, Brigatte e Aguilar (2009), foram considerados os países do G-20, exceto Cuba, Equador e Peru, para os quais os autores não obtiveram dados. Em sua análise, os autores encontraram três componentes principais, chamados de componentes sociais, de comércio internacional e de emprego. Em relação ao primeiro componente, os países que apresentaram os melhores resultados foram Argentina, Brasil, Chile, México e Uruguai. Esses países apresentaram superioridade social em indicadores como *idh* e PIB *per capita*. A Bolívia foi o país americano com pior resultado para esse componente, assim como nos resultados de Curi (1991; 1993).

A partir das análises realizadas por esses autores, o presente artigo irá comparar os resultados a serem estimados com aqueles obtidos no passado, a fim de observar se o mesmo padrão permanece ou se houve mudanças estruturais capazes de alterar os resultados.

## 3. METODOLOGIA

Com o intuito de estudar os indicadores socioeconômicos dos países do continente americano, esse estudo utiliza dois procedimentos: análise dos componentes principais e análise de agrupamentos, também conhecido por análise de *cluster*. Assim, a seguir, serão apresentadas a metodologia e as variáveis utilizadas, para posteriormente demons-

trar e discutir os resultados. Mingoti (2005) aponta que existem dois tipos de técnicas na análise multivariada: a primeira refere-se às técnicas de simplificação dos dados, e a segunda, à inferência estatística. As duas técnicas empregadas neste estudo dizem respeito à simplificação dos dados.

### 3.1. ANÁLISE DE COMPONENTES PRINCIPAIS

Essa técnica é muito utilizada em estudos semelhantes a este, de análise de indicadores, tais como indicadores sociais e econômicos. Leite, Brigatte e Aguilar (2009) realizaram estudo semelhante<sup>(1)</sup> para os países do G-20. Mingoti (2005, p. 59) afirma que, com o intuito de reduzir o número de variáveis, o “objetivo principal deste método é explicar a variância e covariância de um vetor aleatório, composto de  $p$ -variáveis aleatórias, através da construção de combinações lineares das variáveis originais”. Essas combinações lineares são chamadas de componentes principais e são não correlacionadas entre si.

Dessa forma, segundo Mingoti (2005), assumindo que  $X' = [X_1, X_2, \dots, X_p]$  seja um vetor de  $p$  variáveis com matriz de variâncias e covariâncias  $\Sigma$ , a técnica possibilita a construção de componentes principais que são combinações lineares das variáveis  $X$ 's, conforme as Equações (1), (2) e (3):

$$Y_1 = a'_1 x = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \quad (1)$$

$$Y_2 = a'_2 x = a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \quad (2)$$

$$Y_p = a'_p x = a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p \quad (3)$$

A Análise de Componentes Principais consiste em encontrar essas combinações lineares das variáveis, chamadas de componentes principais, de forma que o primeiro componente tenha a maior variância, o segundo tenha a segunda maior variância e assim por diante. Porém, é importante considerar duas condições essenciais: a de ortogonalidade dos vetores e a de não correlação entre os mesmos.

Assim, dado o componente principal  $Y_i = a'_i x$ , o problema de maximização é dado pela Equação (4), sujeito às restrições impostas pela Equação (5):

$$\text{Max } \text{Var}(Y_i) = a'_i \Sigma a_i \quad (4)$$

$$\text{sujeito a } \begin{cases} a'_i a_i = 1 & (\text{vetor unitário}) \\ a'_i a_k = 0 & (\text{ortogonalidade}) \end{cases} \quad (5)$$

Nesse caso, o problema principal resume-se a encontrar as raízes características (autovalores) e os vetores característicos (autovetores), que são os coeficientes dos componentes principais. É claro que considerando as suas condições expostas na equação anterior.

(1) Esse estudo distingue-se do proposto neste trabalho quanto à amostra de países considerados e quanto às variáveis utilizadas.

Dentre as principais características dessa técnica, destacam-se a ausência de correlação entre os componentes, o que significa que a covariância de dois componentes seja igual a zero, e o fato de que a variância total das variáveis originais deve ser igual à variância total dos componentes principais. Mingoti (2005) aponta que a padronização das variáveis é uma transformação usual e, segundo Leite, Brigatte e Aguilar (2009), essa transformação torna-se relevante quando utilizada a análise de *cluster*.

Essa técnica disponibiliza a variância explicada por componente. Supondo, por exemplo, que haja  $p$  variáveis, se incluirmos  $p$  componentes, a variância explicada pelos componentes utilizados será total. Dessa forma, à medida que aumentamos o número de componentes incluídos, aumentamos a variância explicada em que, normalmente, o primeiro componente apresenta a maior variância explicada. Vale ressaltar que a técnica é utilizada, na maioria das vezes, com o objetivo de reduzir o número de variáveis, sem perdas na interpretação dos dados, ou seja, deve-se assumir algum critério para determinar o número de componentes a ser utilizado. Dentre os diversos critérios, existe, *a priori*, o critério de autovalores superiores à unidade e de variância total explicada superior a 70% (MINGOTI, 2005). O presente trabalho adotará ambos os critérios.

Muitos estudos, como o de Leite, Brigatte e Aguilar (2009), procuram apenas construir as combinações lineares das variáveis, sem classificar os objetos. No presente estudo, os países serão ordenados de acordo com os escores dos componentes, que são obtidos a partir dos componentes principais de cada observação e podem ser utilizados para elaborar um índice, análise de *cluster*, ordenar observações, entre outras ações.

### 3.2 ANÁLISE DE AGRUPAMENTOS

A análise de agrupamentos, segundo Mingoti (2005, p. 155), “tem como objetivo dividir os elementos da amostra (ou população) em grupos, de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si com respeito às variáveis (características) que neles foram medidas, e os elementos em grupos diferentes sejam heterogêneos em relação a estas mesmas características”.

Para a aplicação do método cabe, inicialmente, definir a medida de similaridade ou dissimilaridade utilizada para agrupar os países. O presente estudo utiliza a medida de dissimilaridade mais empregada, a Distância Euclidiana. Para ilustrar, assumindo a existência de dois pontos, A e B, a distância entre eles será dada pela Equação (6):

$$\text{Distância entre A e B} = D_{AB} = \sqrt{\sum_{j=1}^P (x_{ja} - x_{jb})^2} \quad (6)$$

E, em termos matriciais, essa distância é dada pela Equação (7):

$$D_{AB} = \sqrt{(X_a - X_b)'(X_a - X_b)} \quad (7)$$

Com base nessa medida de dissimilaridade, agrupam-se os países do continente americano, ilustrados nas equações como pontos A e B. Dada essa medida, existem diversos tipos de métodos a serem utilizados, como o de ligação simples e de centroides, conhecido como método de *Ward*. No presente estudo, esse foi o método utilizado, com

o objetivo de encontrar grupos mais heterogêneos, de forma que os objetos (países) dentro de cada grupo sejam homogêneos (MINGOTI, 2005). O método de *Ward* também é chamado de *mínima variância*.

Esse procedimento é classificado como hierárquico aglomerativo, o qual parte do princípio de que, no início, cada elemento é considerado um conglomerado, de forma que uma vez que haja o agrupamento de dois objetos (países, no presente estudo), esses dois continuem juntos até o fim do processo aglomerativo. Essa característica possibilita a construção de um diagrama de árvore denominado *dendrograma*. Vale ressaltar que no eixo horizontal estão presentes os conglomerados iniciais e, no eixo vertical, a altura correspondente ao nível em que os elementos foram agrupados, ou seja, as distâncias. Então, pode-se afirmar que quanto maior a altura, maior a heterogeneidade dos grupos unidos.

Para uma melhor utilização desse método, o presente trabalho realizou alguns testes com o objetivo de determinar o número exato de grupos a ser considerado na análise. Dentre os mais importantes, foram realizados dois testes: o Pseudo-F de Calinski-Harabasz e o Índice de Duda-Hart. Em ambos os testes, a indicação do número de grupos a ser adotado é dado pelo maior valor do índice encontrado.

Neste trabalho, a utilização da análise de *cluster* possibilitará agrupar, num mesmo conglomerado, os países mais homogêneos quanto às características analisadas, de forma que os grupos construídos sejam heterogêneos entre si. Isso possibilitará corroborar ou não trabalhos anteriores, além de identificar e analisar as diferenças entre os países de um mesmo continente.

### 3.3. VARIÁVEIS UTILIZADAS

O objetivo central do presente estudo é analisar os indicadores socioeconômicos dos países do continente americano. E para isso são utilizadas 11 variáveis, sendo que oito destas representam indicadores sociais e três referem-se às variáveis econômicas. Os dados foram coletados no Instituto Brasileiro de Geografia e Estatística (IBGE, 2011) e podem ser descritos como:

#### a) Indicadores sociais:

1. idh: índice de desenvolvimento humano, varia entre zero e um;
2. expv: expectativa de vida ao nascer, medida em anos;
3. subn: população subnutrida, em termos percentuais;
4. kcal: calorias consumidas, em kcal/dia;
5. aguap: domicílios com acesso à água potável, em termos percentuais;
6. sanb: domicílios com acesso a saneamento básico, também em percentual;
7. txmat: taxa bruta de matrícula para todos os níveis de ensino, também em percentual;
8. net: usuários com acesso à internet, número de usuários por 100 habitantes.

b) Indicadores Econômicos:

1. ppc: Produto Interno Bruto *per capita*, em dólares (US\$);
2. pea: população de 15 anos ou mais, economicamente ativa, em termos percentuais;
3. mpea: mulheres de 15 anos ou mais, economicamente ativas, em termos percentuais.

Dessa forma, o estudo procura analisar a formação de agrupamentos de países do continente americano com base em semelhanças nessas variáveis. Alguns países não foram incluídos em razão da ausência de dados, como, por exemplo, Cuba e Peru. Os países abordados no estudo são, ao todo, vinte e quatro: Argentina, Bolívia, Brasil, Canadá, Chile, Colômbia, Costa Rica, El Salvador, Equador, Estados Unidos, Guatemala, Guiana, Haiti, Honduras, Jamaica, México, Nicarágua, Panamá, Paraguai, República Dominicana, Suriname, Trinidad e Tobago, Uruguai e Venezuela.

#### 4. RESULTADOS E DISCUSSÃO

Os resultados apresentados a seguir serão expostos de forma individualizada. Inicialmente, porém, procedeu-se à análise das estatísticas descritivas e da correlação entre as variáveis para posteriormente proceder-se à análise multivariada. A primeira técnica empregada foi a análise de componentes principais e, em seguida, a análise de agrupamentos. O *software* utilizado para a realização dessas técnicas foi o STATA 10.0.

##### 4.1. ESTATÍSTICAS DESCRITIVAS DAS VARIÁVEIS

Com o intuito de investigar os dados a partir da análise multivariada, buscou-se encontrar e analisar, primeiramente, as estatísticas descritivas destes e a correlação entre as variáveis. A estatística descritiva é relevante em estudos como este porque apresenta claramente as dessemelhanças entre os países em relação às variáveis em que serão comparados. Por exemplo, a variável *idh* apresenta uma média relativamente alta para os vinte e quatro países em estudo, porém, quando analisados o mínimo (Haiti – 0.40) e o máximo (Estados Unidos – 0.90), constata-se a presença no continente americano de países bem distintos.

Tabela 1 – Estatística descritiva das variáveis

|              | MÉDIA    | DESVIO PADRÃO | MÍNIMO | MÁXIMO |
|--------------|----------|---------------|--------|--------|
| <b>Idh</b>   | 0.689208 | 0.10493       | 0.404  | 0.902  |
| <b>Expv</b>  | 73.3875  | 4.446133      | 61.7   | 81     |
| <b>Subnu</b> | 12.29167 | 11.31939      | 5      | 58     |
| <b>Kcal</b>  | 2708.083 | 447.3217      | 1835   | 3825   |
| <b>Aguap</b> | 90.75    | 9.246621      | 58     | 100    |

|              | MÉDIA    | DESVIO PADRÃO | MÍNIMO | MÁXIMO |
|--------------|----------|---------------|--------|--------|
| <b>Sanb</b>  | 79.125   | 19.29477      | 19     | 100    |
| <b>Txmat</b> | 78.70833 | 10.15133      | 52.1   | 99.3   |
| <b>Ppc</b>   | 8699     | 10990.46      | 626    | 44872  |
| <b>Pea</b>   | 64.42375 | 5.783624      | 51.63  | 78.01  |
| <b>Mpea</b>  | 51.26167 | 9.922764      | 36.1   | 71.79  |
| <b>Telef</b> | 22.1325  | 17.93595      | 1.08   | 72.4   |
| <b>Int</b>   | 32.2475  | 19.85877      | 3.48   | 80.3   |

**Fonte:** Resultados da pesquisa, a partir dos dados do IBGE (2011).

Assim como a variável *idh*, o PIB *per capita*, representado pela variável *ppc*, também assinala a heterogeneidade entre os países americanos. Da mesma forma, apesar desta variável possuir uma média alta, deve-se ressaltar que, dos vinte e quatro países da amostra, dezoito apresentaram um PIB *per capita* inferior ao valor médio. Apenas Estados Unidos, Canadá, Trinidad Tobago, Venezuela, Chile e Uruguai apresentaram *ppc* superior à média.

Duas outras variáveis que também marcam significativamente essa disparidade entre os países são *sanb* e *aguap*. Tais indicadores apresentam médias relativamente altas, acima de 70%, porém, quando analisados os valores máximo e mínimo dos mesmos, percebe-se o grande fosso existente entre os países americanos. De acordo com o IBGE (2011), os países que apresentam os piores valores para essas variáveis são Nicarágua (79% e 48%) e Haiti (58% e 19%). Por outro lado, o Canadá destaca-se por apresentar 100% de domicílios com acesso à água potável e saneamento básico.

A existência de uma maior correlação entre as variáveis ajuda a fundamentar a análise multivariada, cujo objetivo é a redução do número de variáveis a um conjunto de componentes que não prejudique a interpretação dos dados (MINGOTI, 2005). A partir matriz de correlações entre as variáveis tem-se, inicialmente, um indício de como se dará a formação dos componentes utilizados na análise de *clusters*. Assim, como essa ferramenta permite identificar o grau de associação entre as variáveis utilizadas, ela constitui um método acessório à análise de componentes principais.

**Tabela 2 – Matriz de correlação entre as variáveis**

|      | IDH  | EXPV  | PPC  | PEA  | MPEA | TXMAT | SUBN | NET | KCAL | SANB | AGUAP |
|------|------|-------|------|------|------|-------|------|-----|------|------|-------|
| Idh  | 1.00 |       |      |      |      |       |      |     |      |      |       |
| Expv | 0.83 | 1.00  |      |      |      |       |      |     |      |      |       |
| Ppc  | 0.76 | 0.57  | 1.00 |      |      |       |      |     |      |      |       |
| Pea  | 0.06 | -0.10 | 0.07 | 1.00 |      |       |      |     |      |      |       |
| Mpea | 0.34 | 0.07  | 0.31 | 0.90 | 1.00 |       |      |     |      |      |       |

|       | IDH   | EXPV  | PPC   | PEA   | MPEA  | TXMAT | SUBN  | NET  | KCAL | SANB | AGUAP |
|-------|-------|-------|-------|-------|-------|-------|-------|------|------|------|-------|
| Txmat | 0.74  | 0.63  | 0.55  | 0.19  | 0.37  | 1.00  |       |      |      |      |       |
| Subn  | -0.73 | -0.66 | -0.34 | 0.01  | -0.19 | -0.64 | 1.00  |      |      |      |       |
| Net   | 0.79  | 0.58  | 0.83  | 0.07  | 0.40  | 0.57  | -0.51 | 1.00 |      |      |       |
| Kcal  | 0.82  | 0.68  | 0.77  | -0.06 | 0.19  | 0.67  | -0.72 | 0.79 | 1.00 |      |       |
| Sanb  | 0.79  | 0.73  | 0.52  | -0.11 | 0.15  | 0.55  | -0.85 | 0.65 | 0.68 | 1.00 |       |
| Aguap | 0.78  | 0.69  | 0.45  | -0.06 | 0.19  | 0.65  | -0.81 | 0.61 | 0.61 | 0.88 | 1.00  |

**Fonte:** Resultados da pesquisa, a partir dos dados do IBGE (2011).

A Tabela 2 apresenta a matriz de correlações entre as variáveis. A sua análise permite identificar um elevado grau de associação linear entre as variáveis *idh*, *expv*, *txmat*, *net*, *subn*, *ppc*, *sanb* e *aguap*, visto que essas variáveis refletem, em sua maioria, características sociais dos países americanos. Tal fato pode ser um indício da construção de um grupo formado por esses indicadores. Da mesma forma, as variáveis *pea* e *mpea* possuem uma correlação de 0,90, o que pode indicar a formação de um grupo com características relativas às condições de emprego nesses países.

A partir dessa análise preliminar será apresentada, a seguir, a metodologia utilizada para construir os componentes referentes aos possíveis grupos de variáveis aqui mencionados. Também será realizada a análise de *cluster*, a qual possibilitará a construção dos grupos de países a partir desses componentes encontrados.

#### 4.2. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS

A análise de componentes principais busca encontrar os autovalores e autovetores, a fim de determinar, respectivamente, o número de componentes principais a ser utilizado e os coeficientes das combinações lineares. Dessa forma, cabe analisar, primeiramente, os autovalores encontrados. Para a aplicação dessa técnica foram utilizadas as 11 variáveis descritas anteriormente, para os vinte e quatro países em estudo.

A Tabela 3 apresenta todos os componentes que explicam a variância total das variáveis originais. No entanto, a partir dos critérios de autovalor superior à unidade e de variância acumulada superior a 70%, determinou-se a utilização de apenas dois componentes principais. Contata-se, como esperado, que o primeiro componente responde pela maior variância explicada pelos dados (59,58%), seguido pelo segundo (17,92%). A variância total explicada a partir da utilização desses dois componentes é igual a 77,5%, mostrando que esses dois componentes conseguem explicar bem a variabilidade dos dados.

**Tabela 3 – Autovalores, variância total explicada e variância acumulada**

|               | AUTOVALORES | % VARIÂNCIA TOTAL | % ACUMULADA |
|---------------|-------------|-------------------|-------------|
| Componente 1  |             |                   |             |
| Componente 2  |             |                   |             |
| Componente 3  | 0.985317    | 0.0896            | 0.8646      |
| Componente 4  | 0.472913    | 0.043             | 0.9076      |
| Componente 5  | 0.362332    | 0.0329            | 0.9405      |
| Componente 6  | 0.290294    | 0.0264            | 0.9669      |
| Componente 7  | 0.135799    | 0.0123            | 0.9793      |
| Componente 8  | 0.0883954   | 0.008             | 0.9873      |
| Componente 9  | 0.0803317   | 0.0073            | 0.9946      |
| Componente 10 | 0.0351712   | 0.0032            | 0.9978      |
| Componente 11 | 0.0243246   | 0.0022            | 1           |

**Fonte:** Dados da pesquisa.

Após definido o número de componentes principais a ser utilizado, cabe analisar os coeficientes de correlação entre os componentes e as variáveis. As correlações em negrito na Tabela 4 indicam uma associação mais forte entre os componentes e as variáveis, possibilitando uma melhor caracterização de cada componente principal.

**Tabela 4 – Correlação entre os componentes e as variáveis utilizadas**

| VARIÁVEIS | COMPONENTE 1 | COMPONENTE 2 |
|-----------|--------------|--------------|
| IDH       |              | 0.011        |
| Expv      |              | -0.1545      |
| Ppc       |              | 0.0888       |
| Txmat     |              | 0.1125       |
| Subnu     |              | 0.1096       |
| Net       |              | 0.0895       |
| Kcal      |              | -0.0587      |
| Sanb      |              | -0.1605      |
| Aguap     |              | -0.1162      |
| Peã       | 0.0236       |              |
| Mpea      | 0.134        |              |

**Fonte:** Dados da pesquisa.

A análise da Tabela 4 mostra que o Componente 1 apresenta correlação mais forte com as variáveis *idh*, *expv*, *txmat*, *ppc*, *net*, *subn*, *aguap*, *sanb* e *kcal*. Isso permite caracterizá-lo como componente de condições socioeconômicas. Das nove variáveis destacadas nesse componente, oito apresentaram sinais positivos, indicando que aumentos nessas variáveis, como, por exemplo, no *ppc*, levam a uma melhora nas condições socioeconômicas do país. A variável *subn*, no entanto, apresenta sinal negativo, como era esperado, uma vez que ela representa população subnutrida. Assim, um aumento nessa variável leva a uma redução nas condições socioeconômicas, determinada pelo Componente 1.

As variáveis que mais explicam a variância do primeiro componente são o *idh*, *kcal*, *aguap* e *sanb*. A primeira variável, Índice de Desenvolvimento Humano, apresenta forte correlação por representar uma síntese de indicadores socioeconômicos, enquanto que a segunda refere-se às calorias consumidas por dia, importante indicador relativo à saúde/alimentação. A terceira e a quarta variáveis referem-se ao acesso dos domicílios à água potável e saneamento básico, importantes indicadores de condições básicas, essenciais ao desenvolvimento econômico.

Quanto ao Componente 2, nota-se uma relação mais forte deste com as variáveis *mpea* e *pea*, o que permite a sua caracterização como condições de mercado de trabalho. O relacionamento positivo entre essas variáveis e o Componente 2 indica que os países com elevados valores percentuais de população economicamente ativa e mulheres economicamente ativas possuem melhores condições de trabalho. Por se tratar de um componente explicado por apenas duas variáveis, constata-se, conforme esperado, que as duas apresentam sinais iguais e magnitude semelhante. A variável *pea* destaca-se por sua correlação, uma vez que retrata o mercado de trabalho como um todo, enquanto que a variável *mpea* representa uma fração desse mercado.

#### 4.3. RESULTADOS DA ANÁLISE DE AGRUPAMENTOS

Dada a composição dos componentes acima descritos, podem-se agrupar os países do continente americano de acordo com cada componente, objetivando constituir grupos formados por países com características mais homogêneas. A análise seguinte está dividida em duas partes. A primeira aborda os agrupamentos gerados a partir do componente de condições socioeconômicas e a segunda trata do componente das condições do mercado de trabalho.

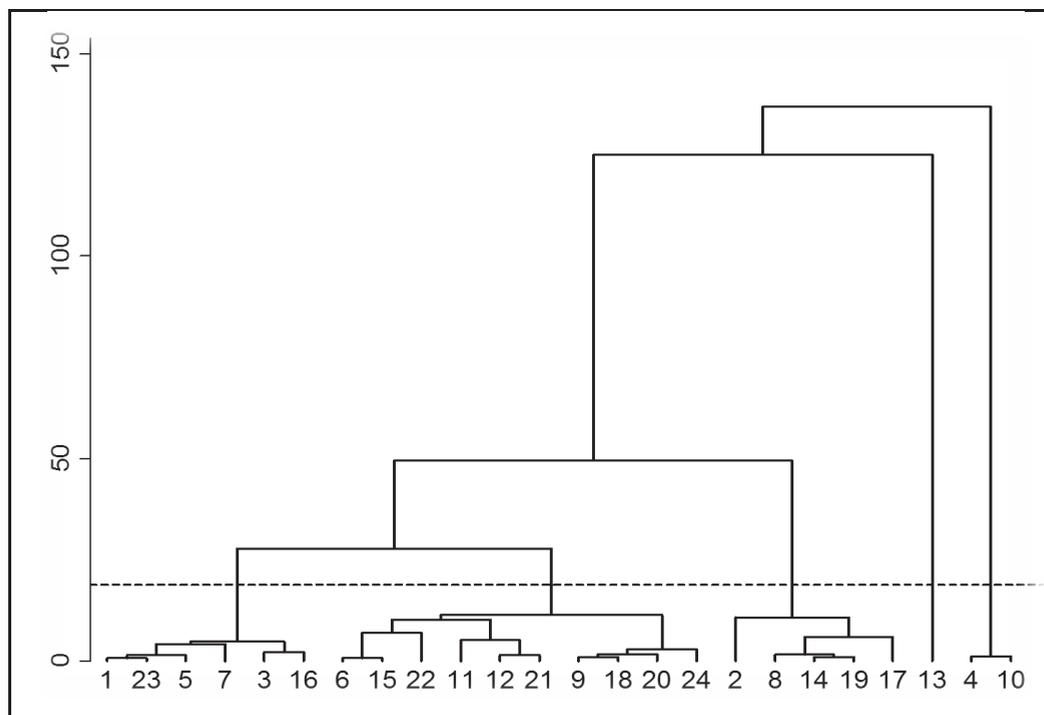
Ambas as análises serão realizadas por meio da aplicação do método de Ward, que é um método hierárquico aglomerativo. Essa técnica possibilita a construção de um diagrama de árvore, conhecido como dendrograma na análise de *cluster*, no qual o eixo horizontal representa os países e o eixo vertical mede a distância (altura) em que os países foram agrupados. Assim, quanto maior a altura menor é a semelhança entre os grupos.

##### 4.3.1. Análise das condições socioeconômicas

A Figura 1 apresenta o dendrograma resultante da análise de agrupamentos elaborada a partir do componente de condições sociais. Para a definição do número de grupos, adotou-se os testes Pseudo-F de Calisnki-Harabsz e o índice de Duda-Hart. Ambos os

testes apontaram a formação de cinco grupos, indicados pela linha tracejada na Figura 1. Os grupos e os países que os compõem foram apresentados na Tabela 5.

**Figura 1 – Dendrograma do processo de agrupamento do componente de condições socioeconômicas**



**Fonte:** Dados da pesquisa.

A Tabela 5, além de apresentar os grupos, classifica os países de acordo com as condições socioeconômicas dos mesmos. A técnica permitiu a formação de cinco grupos. Destacam-se na formação destes, os grupos A e E. O primeiro refere-se aos países com melhores condições socioeconômicas, demonstrado pelos escores do Componente 1 da tabela abaixo. Isso é corroborado pela análise descritiva feita anteriormente. Enquanto que o segundo grupo, formado apenas pelo Haiti, apresenta o pior escore para o mesmo componente, o que indica que este país tem as piores condições socioeconômicas. Os demais grupos, de modo geral, são mais heterogêneos, já que neles encontram-se países de diferentes porções territoriais e a variação dos escores do Componente 1 no mesmo grupo é maior.

O presente trabalho utilizou os escores do Componente 1 para ordenar os países segundo condições socioeconômicas. Isso está demonstrado na última coluna da Tabela 5. Como esperado, o Canadá e os Estados Unidos apresentaram as melhores condições socioeconômicas, graças à superioridade dos valores de suas variáveis. Os estudos já elaborados, como o de Curi (1991), apresentaram resultados semelhantes, indicando que esses dois países já apresentavam as melhores condições socioeconômicas naquela

época. No entanto, houve uma modificação na classificação dos países em relação ao trabalho anterior, já que, naquele, Estados Unidos estava mais bem classificado que Canadá, e no presente trabalho, essa classificação se inverte.

**Tabela 5 – Agrupamento e classificação dos países a partir do Componente 1 (condições socioeconômicas)**

| PAÍS                 | COMPONENTE 1 | GRUPOS | CLASSIFICAÇÃO |
|----------------------|--------------|--------|---------------|
| Canadá               | 5.258111     | A      | 1             |
| Estados Unidos       | 5.188612     | A      | 2             |
| Uruguai              | 2.150618     | B      | 3             |
| Chile                | 1.647176     | B      | 4             |
| Argentina            | 1.588244     | B      | 5             |
| México               | 1.095198     | B      | 6             |
| Costa Rica           | 1.019293     | B      | 7             |
| Brasil               | 0.952606     | B      | 8             |
| Jamaica              | 0.6849374    | C      | 9             |
| Venezuela            | 0.4926518    | C      | 10            |
| Colômbia             | 0.4910766    | C      | 11            |
| Trinidad e Tobago    | 0.2977076    | C      | 12            |
| Panamá               | -0.1218405   | C      | 13            |
| Equador              | -0.2303403   | C      | 14            |
| Guiana               | -0.4570865   | C      | 15            |
| Suriname             | -0.6567787   | C      | 16            |
| República Dominicana | -0.823658    | C      | 17            |
| Guatemala            | -1.569837    | C      | 20            |
| El Salvador          | -1.091296    | D      | 18            |
| Paraguai             | -1.245697    | D      | 19            |
| Honduras             | -1.817888    | D      | 21            |
| Bolívia              | -2.226197    | D      | 22            |
| Nicarágua            | -2.986948    | D      | 23            |
| Haiti                | -7.638663    | E      | 24            |

**Fonte:** Dados da pesquisa.

Analisando o extremo oposto, tem-se que o Haiti assumiu a última colocação dos países americanos, apresentando as piores condições socioeconômicas. Esse resultado também foi encontrado por Curi (1991). Esse mesmo autor destacou que os países com piores resultados eram, além deste, a Bolívia. Tal fato também é comprovado nesse estudo, já que a Bolívia apresentou a vigésima segunda posição na classificação geral.

Em seus trabalhos, Curi (1991, 1993) apontou que Trinidad e Tobago apresentava condições socioeconômicas e de saúde próximas às do Chile, Uruguai e Argentina. No entanto, os resultados apontam um distanciamento de Trinidad e Tobago desses países, o que o posicionou em grupo distinto. Dessa forma, pode-se inferir que esse país apresentou uma retração dessas condições em relação aos outros países.

Ao analisar o trabalho de Leite, Brigatte e Aguilar (2009), tem-se que os autores encontraram que Uruguai, Brasil, México, Argentina e Chile formavam um grupo por apresentarem superioridade nos padrões socioeconômicos em relação aos países que compõem o G-20. Isso foi corroborado neste trabalho, assim como o fato de a Bolívia ser o pior país americano a compor o G-20.

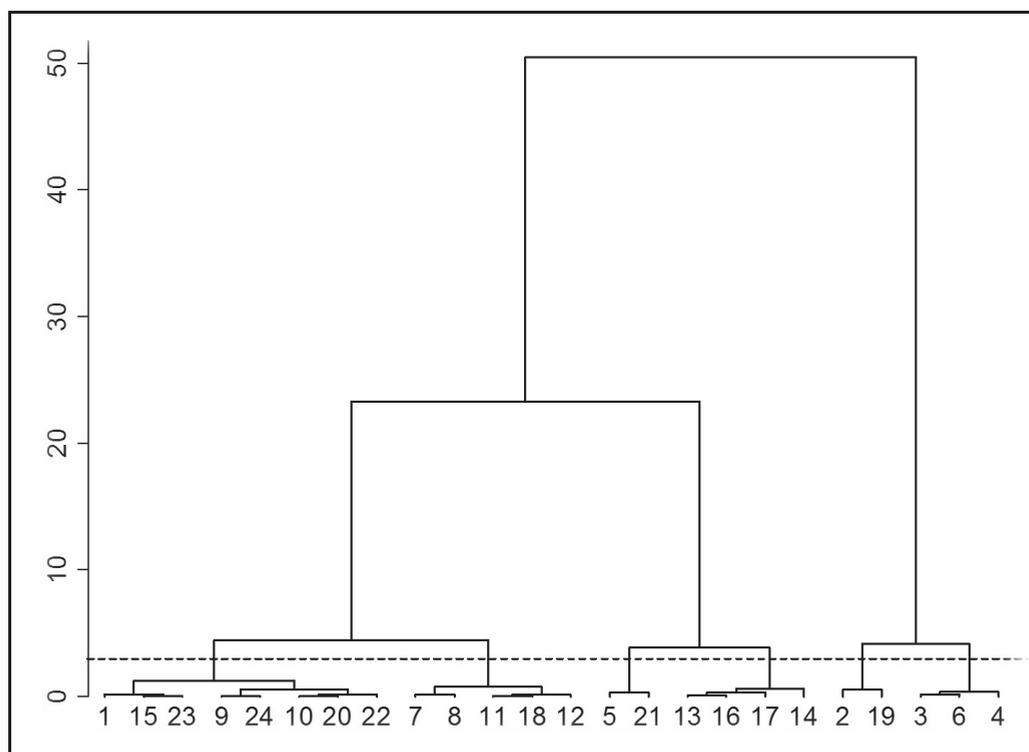
O Brasil, situado em oitavo lugar na classificação geral, o que corresponde à última colocação num grupo formado por Uruguai, Chile, Argentina, México e Costa Rica, pouco se destaca quanto aos indicadores socioeconômicos quando comparado com os demais países do grupo B, apresentando valor superior apenas em calorias consumidas por dia. Porém, quando comparado à Venezuela, à Colômbia e ao Equador (entre outros), o Brasil apresenta resultados melhores em indicadores como *idh*, população subnutrida, calorias consumidas por dia e taxa bruta de matrículas. Apesar disso, em outras variáveis o Brasil apresenta resultados inferiores aos países em questão.

Após essa análise, tem-se que a aplicação eficiente da técnica adotada nesta pesquisa é corroborada em três pontos. Em primeiro lugar, os resultados aqui apresentados corroboram as análises anteriores de Curi (1991, 1993) e Leite, Brigatte e Aguilar (2009), além de oferecer novos resultados. Em segundo lugar, percebe-se que há uma concordância entre as estatísticas descritivas das variáveis e a técnica utilizada. Por último, constata-se a complementaridade entre os resultados obtidos com o agrupamento dos países e a classificação obtida a partir dos escores do Componente 1, o que indica um bom ajustamento da técnica aplicada aos dados.

#### 4.3.2 Análise das condições de mercado de trabalho

A análise de *cluster*, quando aplicada ao Componente 2, aponta para a formação de seis grupos, conforme fica evidenciado pela linha tracejada na Figura 2. O número de grupos foi determinado pelos testes Pseudo-F de Calinski-Harabsz e índice de Duda-Hart.

**Figura 2 – Dendrograma do processo de agrupamento do componente de condições de mercado de trabalho**



**Fonte:** Dados da pesquisa.

Os grupos formados, apesar de homogêneos em relação às condições do mercado de trabalho, são bastante heterogêneos no que concerne ao desenvolvimento socioeconômico e territorial. Isso fica evidenciado a seguir:

- Grupo A: Argentina, Equador, Estados Unidos, Jamaica, República Dominicana, Trinidad Tobago, Uruguai e Venezuela;
- Grupo B: Costa Rica, El Salvador, Guatemala Guiana e Panamá;
- Grupo C: Bolívia e Paraguai;
- Grupo D: Chile e Suriname;
- Grupo E: Haiti, México, Nicarágua e Honduras;
- Grupo F: Brasil, Colômbia e Canadá.

Leite, Brigatte e Aguilar (2009) apontam que, como os grupos foram formados somente com base nas porcentagens das variáveis *pea* e *mpea*, uma análise mais profunda deveria estar fundamentada no conhecimento de cada mercado de trabalho, mas isso iria requerer o conhecimento de aspectos sociais e culturais de cada país. Entretanto, os resultados, assim como para o componente anterior, apresentam complementaridade às estatísticas descritivas das variáveis.

## 5. CONSIDERAÇÕES FINAIS

Este trabalho objetivou analisar indicadores socioeconômicos dos países do continente americano a partir da análise multivariada, especificamente, de componentes principais e de agrupamento. A aplicação dessas técnicas foi eficiente, visto que os resultados foram corroborados por trabalhos anteriores e houve uma complementaridade entre as estatísticas descritivas das variáveis e os resultados encontrados.

A contribuição deste estudo decorre da inserção de novos indicadores sociais diante de outros trabalhos, como população subnutrida e domicílios com acesso à água potável, da ordenação/classificação dos países americanos segundo condições socioeconômicas e do recorte geográfico dado, nunca antes analisado.

A partir dos resultados encontrados, pode-se inferir que os Estados Unidos e o Canadá são os países que possuem as melhores condições socioeconômicas. Isso é corroborado pela literatura, bem como pelos dados, já que esses países apresentaram, na estatística descritiva, valores superiores para as variáveis PIB *per capita*, expectativa de vida, domicílios com acesso à água potável e saneamento básico, calorias consumidas por dia e taxa bruta de matrícula.

Em contrapartida, o Haiti apresentou as piores condições socioeconômicas, assumindo a última posição na classificação dos países americanos em estudo. Em algumas variáveis, a inferioridade dos valores dos indicadores sociais do Haiti levou a um grande distanciamento da média das variáveis, o que fez com que esse país formasse um grupo único na análise de *cluster* do Componente 1. Tal fato, aliado ao escore obtido com o componente condições socioeconômicas, indica o distanciamento e a situação precária desse país em relação aos demais. Essas conclusões são corroboradas por outros trabalhos.

Os países da América Latina apresentaram resultados díspares, além de estarem distribuídos em grupos distintos. Destaca-se, negativamente, a Bolívia, em vigésimo terceiro lugar, e o Paraguai, na décima nona posição. Positivamente, os destaques foram para Uruguai, Chile, Argentina e Brasil, situados no mesmo grupo, indicando características socioeconômicas semelhantes. O Brasil, último país desse grupo, assume classificação inferior a esses países latinos por apresentar quase todos indicadores inferiores a pelo menos um país, excetuando a variável relativa às calorias consumidas por dia.

Os resultados obtidos com o componente das condições do mercado de trabalho, apesar de coerentes, impossibilitam uma análise mais profunda, principalmente no que tange à análise de agrupamento. Isso porque exige uma investigação mais profunda de aspectos culturais e sociais de cada país. Porém, no que concerne às condições socioeconômicas, a aplicação da análise multivariada corrobora a ideia de que existem grandes diferenças entre os países do continente americano; ideia confirmada pela construção de agrupamentos e escores do primeiro componente.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

CURI, Paulo Roberto. Agrupamentos de países segundo indicadores básicos e econômicos. *Revista Brasileira de Estudos da População*. Campinas, v. 8, n. 1/2, p. 112-124, 1991.

CURI, Paulo Roberto. Agrupamentos de países segundo indicadores de padrão de vida. *Revista Saúde Pública*. São Paulo, v. 27, n. 2, p. 127-134, 1993.

IBGE, Banco de Dados. Disponível em: <<http://www.ibge.gov.br/paisesat/main.php>> Acesso em: 5 abr./2011.

LEITE, Gil Bracarense; BRIGATTE, Henrique; AGUILAR, Eder Barbosa. Análise multivariada de indicadores socioeconômicos dos países do G-20. *Revista de Economia Mackenzie*, v. 7, n. 1, p. 125-147.

MINGOTI, Sueli Aparecida. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte, Ed. UMG, 2005.

PRADOS DE LA ESCOSURA, Leandro. Growth, inequality and poverty in Latin America: historical evidence, controlled conjectures. Universidad Carlos III de Madrid, Working Paper 05-41, n. 4, Economic History and Instituciones Series, June 2005. Disponível em: <<http://docubib.uc3m.es/WORKINGPAPERS/WH/wh054104.pdf>> Acesso em: 10 abr./2011.

RIBEIRO, Carlos Antônio Costa. *Um panorama das desigualdades na América Latina. Análise de Conjuntura do Observatório Político Sul-Americano*. Instituto Universitário de Pesquisas do Rio de Janeiro, n. 6, junho de 2006. Disponível em: <[http://www.opsa.com.br/images/pdf/analise/19\\_analises\\_Um\\_panorama\\_das\\_desigualdades\\_na\\_America\\_Latina.pdf](http://www.opsa.com.br/images/pdf/analise/19_analises_Um_panorama_das_desigualdades_na_America_Latina.pdf)> Acesso em: 20 abr./2011.