



**UNIVERSIDADE FEDERAL DO CEARÁ  
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA DE ECONOMIA PROFISSIONAL – PEP  
MESTRADO PROFISSIONAL EM ECONOMIA DO SETOR PÚBLICO – MESP**

**FELIPE GONDIM MELO**

**DETERMINANTES DA ECONOMIA INFORMAL: UMA ANÁLISE DE  
CLUSTERIZAÇÃO E COMPONENTES PRINCIPAIS EM PAÍSES  
DESENVOLVIDOS E EM DESENVOLVIMENTO**

**FORTALEZA**

**2024**

FELIPE GONDIM MELO

DETERMINANTES DA ECONOMIA INFORMAL: UMA ANÁLISE DE  
CLUSTERIZAÇÃO E COMPONENTES PRINCIPAIS EM PAÍSES DESENVOLVIDOS E  
EM DESENVOLVIMENTO

Dissertação apresentada ao Mestrado Profissional em Economia do Setor Público da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Mestre. Área de Concentração: Economia do setor público.

Orientador: Prof. Dr. Márcio Veras Corrêa.

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

M485d Melo, Felipe Gondim.  
Determinantes da Economia Informal: Uma Análise de Clusterização e Componentes Principais em Países Desenvolvidos e em Desenvolvimento / Felipe Gondim Melo. – 2024.  
42 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Mestrado Profissional em Economia do Setor Público, Fortaleza, 2024.  
Orientação: Prof. Dr. Márcio Veras Corrêa.

1. Economia Informal. 2. Clusterização. 3. Análise de Componentes Principais. 4. Desenvolvimento Econômico. 5. Políticas Públicas. I. Título.

CDD 330

---

FELIPE GONDIM MELO

DETERMINANTES DA ECONOMIA INFORMAL: UMA ANÁLISE DE  
CLUSTERIZAÇÃO E COMPONENTES PRINCIPAIS EM PAÍSES DESENVOLVIDOS E  
EM DESENVOLVIMENTO

Dissertação apresentada ao Mestrado  
Profissional em Economia do Setor Público da  
Universidade Federal do Ceará, como requisito  
parcial à obtenção do título de Mestre. Área de  
Concentração: Economia do setor público.

Orientador: Prof. Dr. Márcio Veras Corrêa

Aprovada em 12/07/2024.

BANCA EXAMINADORA

---

Márcio Veras Corrêa (Orientador)  
Universidade Federal do Ceará – UFC

---

Marcos Renan Vasconcelos Magalhães  
Secretaria de Saúde do Ceará – SESA

---

Maurício Benegas  
Universidade Federal do Ceará – UFC

## **AGRADECIMENTOS**

A Deus, nosso Pai, em primeiro lugar, por tudo que o Senhor tem realizado em minha vida para honra e glória de seu filho, nosso Salvador Jesus Cristo.

À Nossa Senhora das Graças (da Medalha Milagrosa), minha maior intercessora, mãe do Santíssimo e nossa mãe, por todas as preces atendidas.

À minha mãe, Eliane Pires Gondim, por sempre ter me amado incondicionalmente e sempre ter acreditado em mim.

À minha amada avó, Francisca Pires Gondim, que deixa saudades há 15 anos em minha família.

À minha namorada e futura esposa, Ana Cecilia Amaral Falcão, por ser a luz da minha vida e por retribuir todo o meu amor.

Por fim, a todos os envolvidos na construção desse trabalho.

“Não fui eu que ordenei a você? Seja forte e corajoso! Não se apavore nem desanime, pois o Senhor, o seu Deus, estará com você por onde você andar”. (Josué 1:9).

## RESUMO

Este estudo investiga a diversidade e os determinantes da informalidade em diferentes países, utilizando métodos de clusterização e Análise de Componentes Principais (PCA). Os resultados da clusterização revelaram dois grupos principais de países com características socioeconômicas distintas. O cluster 1 é composto majoritariamente por países desenvolvidos, caracterizados por economias mais formalizadas, elevados níveis de capital humano, maior integração econômica global, maior liberdade econômica e maior atração de investimento estrangeiro direto. Em contraste, o cluster 2 inclui principalmente países em desenvolvimento, que enfrentam desafios significativos relacionados à informalidade, como baixos níveis de capital humano, alta percepção de corrupção, maior desigualdade de gênero, maior burocracia e taxas de desemprego mais altas. O PCA identificou os principais fatores que contribuem para essas diferenças, destacando a importância da globalização econômica, liberdade econômica, corrupção e capital humano no CP1, e das dificuldades burocráticas e condições do mercado de trabalho no CP2. Este estudo oferece uma base sólida para a formulação de políticas públicas específicas e contextualizadas, visando mitigar os problemas da informalidade e promover o desenvolvimento socioeconômico de maneira equitativa e sustentável.

**Palavras-chave:** economia informal; clusterização; análise de componentes principais; desenvolvimento econômico; políticas públicas.

**Códigos JEL:** E26; O17; C38; J46; O57.

## **ABSTRACT**

This study investigates the diversity and determinants of informality in different countries, using clustering methods and Principal Component Analysis (PCA). The clustering results revealed two main groups of countries with distinct socioeconomic characteristics. Cluster 1 is mainly composed of developed countries, characterized by more formalized economies, high levels of human capital, greater global economic integration, greater economic freedom, and higher attraction of foreign direct investment. In contrast, Cluster 2 mainly includes developing countries, which face significant challenges related to informality, such as low levels of human capital, high perception of corruption, greater gender inequality, more bureaucracy, and higher unemployment rates. PCA identified the main factors contributing to these differences, highlighting the importance of economic globalization, economic freedom, corruption, and human capital in CP1, and bureaucratic difficulties and labor market conditions in CP2. This study provides a solid basis for the formulation of specific and contextualized public policies aimed at mitigating the problems of informality and promoting socioeconomic development in an equitable and sustainable manner.

**Keywords:** informal economy; clustering; principal component analysis; economic development; public policies.

**JEL Codes:** E26; O17; C38; J46; O57.



## LISTA DE FIGURAS

Figura 1 – Matriz de Correlação.....	24
Figura 2 – Estatísticas das clusterizações. ....	26
Figura 3 – CP1 × CP2 sobre as variáveis selecionadas. ....	32

## LISTA DE TABELAS

Tabela 1 – Variáveis selecionadas.....	22
Tabela 2 – Estatísticas descritivas das variáveis selecionadas dos 140 países.....	25
Tabela 3 – Resultados das métricas de validação interna dos clusters.....	27
Tabela 4 – Médias das variáveis selecionadas para os clusters.....	28
Tabela 5 – Análise dos Componentes Principais: autovalores e variância acumulada.....	30
Tabela 6 – Coeficientes de ponderação dos dois componentes principais.....	31
Tabela A1 – Países do cluster 1 em ordem decrescente para o primeiro componente principal. .....	36
Tabela A2 – Países do cluster 2 em ordem decrescente para o segundo componente principal. .....	39

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>10</b>
<b>2</b>	<b>REVISÃO DE LITERATURA.....</b>	<b>11</b>
<b>3</b>	<b>METODOLOGIA.....</b>	<b>13</b>
<b>4</b>	<b>BASE DE DADOS.....</b>	<b>21</b>
<b>5</b>	<b>RESULTADOS.....</b>	<b>26</b>
<b>6</b>	<b>CONCLUSÃO.....</b>	<b>34</b>
	<b>REFERÊNCIAS.....</b>	<b>36</b>
	<b>APÊNDICE A – PAÍSES DO CLUSTER 1 E DO CLUSTER 2.....</b>	<b>38</b>

## 1 INTRODUÇÃO

A economia informal, apesar de sua onipresença global e impacto substancial no desenvolvimento socioeconômico, permanece um fenômeno complexo e multifacetado. Estudos como os de La Porta e Shleifer (2014) demonstram que a informalidade varia consideravelmente entre países e regiões, refletindo a diversidade de contextos históricos, culturais, políticos e econômicos. Além disso, a informalidade se manifesta em diferentes níveis, abrangendo tanto empresas quanto trabalhadores, como apontado por Ulyssea (2018).

A importância de compreender a heterogeneidade da informalidade para a formulação de políticas eficazes é destacada por Yusuf (2007) e Rodgers et al (2016). Adicionalmente, Ulyssea (2018) revela a coexistência de empresas formais e informais em setores específicos, com sobreposição nas distribuições de produtividade e tamanho, evidenciando a natureza heterogênea da informalidade mesmo dentro de um mesmo país.

Diante desse cenário, o presente estudo visa identificar grupos distintos de países com base em suas características de informalidade, utilizando dados relacionados à economia informal compilados pelo Banco Mundial, por meio de análise de clusterização.

Este estudo busca contribuir para a literatura ao expandir o conhecimento sobre a economia informal, fornecendo uma nova perspectiva sobre sua diversidade e determinantes. A classificação dos países em grupos distintos com base em suas características de informalidade e a análise dos diferentes tipos de informalidade e seus determinantes podem auxiliar na formulação de políticas públicas mais eficazes para lidar com esse desafio. Em suma, esta pesquisa visa aprofundar a compreensão da economia informal, identificando padrões e relações entre suas diferentes dimensões e desenvolvendo tipologias que possam informar políticas e enriquecer o debate acadêmico sobre o tema.

Os objetivos específicos incluem:

- a) investigar padrões e relações entre diferentes dimensões da informalidade, como tamanho do setor informal, tipo de emprego informal, setores de atividade e determinantes socioeconômicos;
- b) comparar e contrastar grupos de países em relação às suas características de informalidade, identificando semelhanças e diferenças;
- c) segmentar os países em grupos distintos com base em suas características de informalidade, oferecendo uma nova perspectiva sobre a diversidade dessa característica.

## 2 REVISÃO DE LITERATURA

A economia informal, setor não regulamentado e majoritariamente não tributado, tem sido objeto de estudo de diversos autores devido à sua relevância para o desenvolvimento econômico e social. La Porta e Shleifer (2014) destacam que a informalidade varia significativamente entre países, refletindo diferentes contextos históricos, culturais e institucionais. Essa heterogeneidade, observada tanto em economias desenvolvidas quanto em desenvolvimento, evidencia a complexidade do fenômeno e a necessidade de abordagens personalizadas para sua compreensão e enfrentamento. Nesse sentido, Lehmann e Pan (2016) corroboram essa perspectiva, enfatizando a importância de considerar as nuances da informalidade para a formulação de políticas públicas eficazes, visto que as mesmas estratégias podem gerar resultados distintos em diferentes contextos.

Indo além da mera constatação da heterogeneidade, Ulyssea (2018) aprofunda a análise ao examinar a coexistência de empresas formais e informais, mesmo em setores específicos, e a sobreposição nas distribuições de produtividade e tamanho. Essa coexistência e sobreposição revelam a complexidade do fenômeno, indicando que a informalidade não é um problema homogêneo, mas sim multifacetado, com diferentes causas e consequências em diferentes contextos. A informalidade varia em intensidade e características não apenas entre países, mas também dentro de um mesmo país, o que demanda uma análise mais granular para sua compreensão, como apontado por Felix et al. (2024).

Diversos fatores têm sido apontados como determinantes da informalidade. Djankov et al. (2002) destacam os custos de formalização, como tempo e burocracia para registrar empresas e propriedades, como um dos principais obstáculos à formalização. Ulyssea (2018) complementa essa perspectiva ao argumentar que a rigidez das regulamentações trabalhistas também pode incentivar a informalidade, especialmente em países com alto desemprego, onde as empresas buscam reduzir custos e flexibilizar as relações de trabalho. Para Felix et al. (2024), esses custos de formalização podem variar significativamente entre países, impactando a decisão das empresas de operar na informalidade.

O nível de capital humano também desempenha um papel crucial na informalidade. Gennaioli et al. (2013) e La Porta e Shleifer (2008) discutem como o baixo nível de educação e habilidades pode levar à informalidade, seja pela falta de oportunidades no setor formal ou pela percepção de que a educação formal não é essencial para o sucesso no setor informal. Essa relação entre capital humano e informalidade é explorada por Felix et al. (2024), que destacam

a importância de considerar o nível de educação e habilidades da força de trabalho na análise da informalidade.

A globalização, por sua vez, apresenta um impacto ambivalente na informalidade. Canh e Dinh Thanh (2020) e Elbahnasawy (2021) sugerem que a maior concorrência gerada pela globalização pode pressionar as empresas a reduzir custos, levando-as à informalidade. No entanto, a globalização também pode abrir novas oportunidades de mercado e impulsionar o crescimento econômico, o que poderia, em teoria, reduzir a necessidade de atividades informais. Felix et al. (2024) também exploram a relação entre globalização e informalidade, analisando o impacto da integração do país na economia global e do fluxo de investimento estrangeiro direto.

Por fim, características socioculturais, como a desigualdade de gênero, também podem influenciar a informalidade. A discriminação enfrentada pelas mulheres no mercado de trabalho formal pode levá-las a buscar alternativas no setor informal, onde a flexibilidade e a menor rigidez podem ser atrativas. A influência de fatores socioculturais na informalidade é também abordada por Felix et al. (2024), que incluem a desigualdade de gênero como uma das variáveis em sua análise.

Medina e Schneider (2018) aprofundam a análise das causas e consequências da economia informal em diversos países, utilizando modelos MIMIC para estimar seu tamanho e investigar a influência da percepção de corrupção. Seus achados reforçam a complexidade do fenômeno e a necessidade de abordagens multidisciplinares para compreendê-lo e enfrentá-lo, alinhando-se com a perspectiva de Felix et al. (2024) sobre a importância de considerar múltiplos determinantes na análise da informalidade.

### 3 METODOLOGIA

A presente pesquisa busca aprofundar a compreensão da economia informal, um fenômeno complexo e heterogêneo que varia significativamente entre países e regiões, refletindo diferentes contextos históricos, culturais, políticos e econômicos (La Porta; Shleifer, 2014). Para tanto, será adotada a abordagem K-P (Anaraki et al., 2021), alinhada à metodologia de Wu (2023), que combina a clusterização k-means com a Análise de Componentes Principais (PCA) para identificar padrões e relações entre diferentes dimensões da informalidade.

A análise se desdobrará nas seguintes etapas:

- a) coleta de dados: Serão utilizados dados da base de dados da economia informal do Banco Mundial, complementados por outras fontes relevantes, abrangendo variáveis como tamanho do setor informal, tipo de emprego informal (firma informal ou trabalhador informal em firma formal), setores de atividade e determinantes socioeconômicos;
- b) tratamento e padronização dos dados: Os dados serão tratados para lidar com valores ausentes e outliers, e em seguida, padronizados para garantir a comparabilidade entre as variáveis e evitar que diferentes escalas influenciem a análise;
- c) clusterização (k-means): O algoritmo k-means será aplicado aos dados padronizados para agrupar os países em clusters com base em suas características de informalidade. O número ideal de clusters será determinado através de métodos como o do Cotovelo, da Silhueta e Gap Estatístico, buscando maximizar a similaridade intragrupo e a dissimilaridade intergrupo. Adicionalmente, serão calculados o Índice de Silhueta, o Índice de Davies-Bouldin (DBI) e o Índice de Calinski-Harabasz (CHI) para avaliar a qualidade dos clusters e garantir a robustez dos resultados, conforme recomenda Milligan e Cooper (1985);
- d) análise de componentes principais (PCA): O PCA será aplicado aos dados para reduzir a dimensionalidade e identificar as variáveis que mais contribuem para a formação dos clusters. A seleção das variáveis para essa metodologia será baseada na matriz de correlação, priorizando variáveis com alta correlação (acima de 0,3) e excluindo aquelas com baixa correlação, visando evitar a multicolinearidade e garantir a robustez da análise;

e) análise dos resultados: Os clusters formados serão analisados em termos de suas características de informalidade, buscando identificar semelhanças e diferenças entre os grupos de países. As variáveis mais importantes em cada cluster, reveladas pela PCA, serão interpretadas para entender os principais fatores que impulsionam a informalidade em cada grupo.

Espera-se que esta estratégia empírica permita identificar tipologias de países com base em suas características de informalidade, oferecendo uma nova perspectiva sobre a diversidade do fenômeno e seus determinantes. As informações obtidas contribuirão para a formulação de políticas públicas mais eficazes e personalizadas para lidar com a informalidade em diferentes contextos, além de enriquecer a literatura acadêmica sobre o tema.

A clusterização é uma técnica de mineração de dados que visa agrupar observações em conjuntos distintos (clusters) com base em suas similaridades. O objetivo é maximizar a homogeneidade dentro dos clusters e a heterogeneidade entre eles, revelando padrões e estruturas subjacentes aos dados (Jain; Dubes, 1988; Hair et al., 2009).

No presente estudo, o algoritmo k-means, utilizando a distância euclidiana, será empregado para agrupar países com base em suas características de informalidade (Favero; Belfiore, 2021). O k-means busca minimizar a variação total intragrupo, ou seja, a soma dos quadrados das distâncias entre as observações e o centroide de seu respectivo cluster. O algoritmo Hartigan-Wong (1979) será utilizado, definindo a variação total intra-cluster como:

$$\sum_{k=1}^K W(C_k) = \sum_{x \in C_k} (x_i - \mu_k)^2 \quad (1)$$

onde,  $x_i$  representa uma dada variável da unidade de análise  $i$  pertencente ao cluster  $C_k$ ;  $\mu_k$  é a média de uma dada variável  $x$  pertencente ao cluster  $C_k$ ; e  $K$  é o total de clusters.

Cada observação ( $x_i$ ) é atribuída a um cluster específico de forma que a soma das distâncias ao quadrado da observação até os centros de cluster atribuídos  $\mu_k$  seja mínima. Define-se a variação total dentro do cluster da seguinte forma:

$$\sum_{k=1}^K W(C_k) = \sum_{k=1}^K \sum_{x \in C_k} (x_i - \mu_k)^2 \quad (2)$$

Para determinar o número ideal de clusters no algoritmo k-means, diversos métodos estatísticos podem ser empregados, como o método do Cotovelo, o método da Silhueta e o Gap Estatístico. Entre eles, o método do Cotovelo é um dos mais utilizados (Favero; Belfiore, 2021).



O método do Cotovelo consiste em executar o algoritmo k-means repetidamente, variando o número de clusters ( $k$ ). Em seguida, plota-se um gráfico da variância explicada ou da soma dos quadrados intra-cluster (WCSS) em relação ao número de clusters. O ponto da curva em que ocorre uma mudança abrupta, assemelhando-se a um “cotovelo”, indica o número ótimo de clusters, pois a partir desse ponto, adicionar mais clusters não reduz significativamente o WCSS (Equação 3).

$$WCSS = \sum d(P_1, C_1) + \sum d(P_2, C_2) + \sum d(P_3, C_3) + \dots + \sum d(P_i, C_k) \quad (3)$$

Da equação anterior,  $C_k$  representa o centro de cada cluster  $k$ , enquanto  $P_i$  representa os elementos que estão agrupados juntos dentro de um cluster específico  $k$ .

O método da Silhueta, proposto por Rousseeuw (1987), avalia a adequação de cada observação ao seu respectivo cluster, sendo que uma maior média da largura da silhueta indica um agrupamento mais coeso. O cálculo da silhueta média envolve a análise de diversas observações e diferentes valores de  $k$  (número de clusters), buscando maximizar a silhueta média dentro de um intervalo de valores possíveis para  $i$  (elementos do cluster). Batool e Henning (2021) destacam a utilização da silhueta média como uma medida objetiva, intuitiva e robusta a valores atípicos. O cálculo dessa métrica inicia-se com a determinação do coeficiente de silhueta para cada elemento  $i$ , conforme a equação:

$$s(i) = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (4)$$

Nesta equação,  $a_i$  representa a média das distâncias entre o elemento  $i$  e todos os outros elementos no mesmo cluster, enquanto  $b_i$  é a menor média das distâncias entre o elemento  $i$  e os elementos de qualquer outro cluster, distintos do próprio cluster do elemento  $i$ . O coeficiente de silhueta  $s(i)$  varia de  $-1$  a  $1$ , onde valores próximos de  $1$  indicam que a observação está bem alocada no seu cluster e distante de outros clusters, valores próximos de  $0$  indicam sobreposição ou ambiguidade, e valores próximos de  $-1$  indicam que a observação foi alocada ao cluster errado. Assim, o método da silhueta oferece uma abordagem analítica para determinar o número ideal de clusters, considerando a adequação das unidades de análise dentro de seus clusters e o grau de separação entre os clusters.

O método do “gap statistic”, proposto por Tibshirani et al. (2001), compara a estrutura dos dados reais com a estrutura de dados gerados aleatoriamente sob uma hipótese nula de não haver clusterização. A premissa é que, se o número escolhido de clusters for

adequado, a estrutura dos dados reais será substancialmente mais proeminente do que a estrutura aleatória. O método calcula o “gap statistic”, que é a diferença entre a dispersão intracluster observada nos dados reais e a esperada em dados aleatórios. O número ideal de clusters é aquele que maximiza o “gap statistic”, indicando que a estrutura dos dados reais é significativamente diferente da estrutura aleatória e, portanto, a clusterização é justificada.

No presente estudo, os três métodos foram aplicados. O método do Cotovelo e o método da Silhueta indicaram três clusters, enquanto o “gap statistic”, indicou dois. Optou-se por dois clusters, como será detalhado posteriormente. É importante ressaltar que a análise de agrupamentos não garante similaridades intracluster significativas e deve ser combinada com o conhecimento do problema (Fawcett; Provost, 2016). Para isso foi utilizado métricas de validação interna dos agrupamentos para atestar a qualidade dos resultados.

A validação interna de clusters constitui um procedimento fundamental na análise de agrupamentos, visando avaliar a qualidade das partições geradas por algoritmos de clusterização sem o uso de informações externas ou rótulos prévios. Segundo Liu et al. (2009), a validação interna busca responder à seguinte questão: “Quão bem os clusters refletem a estrutura inerente aos dados?”. A importância da validação interna reside em diversos fatores. Primeiramente, ela garante que os clusters não sejam meros artefatos do algoritmo, mas sim representações significativas de agrupamentos nos dados. Além disso, permite a comparação objetiva entre diferentes algoritmos de clusterização, possibilitando a escolha do mais adequado para um determinado conjunto de dados (Milligan; Cooper, 1985). Por fim, auxilia na otimização dos parâmetros dos algoritmos, como o número de clusters, visando obter a melhor solução possível (Halkidi et al., 2001).

As métricas de validação interna geralmente se baseiam em duas propriedades desejáveis dos clusters: coesão e separação. A coesão mede a similaridade entre os objetos dentro de um mesmo cluster, enquanto a separação mede a dissimilaridade entre objetos de clusters diferentes. Para o presente estudo serão utilizados o coeficiente de Silhueta, que mede o quão similar um objeto é ao seu próprio cluster em comparação com outros clusters, variando de  $-1$  a  $1$ , onde valores próximos a  $1$  indicam boa qualidade do agrupamento (Rousseeuw, 1987); o índice Davies-Bouldin, por sua vez, calcula a similaridade média entre cada cluster e o seu cluster mais similar: valores mais baixos indicam melhor qualidade (Davies; Bouldin, 1979). E, por fim índice Calinski-Harabasz, a soma dos quadrados total (SST), a soma dos quadrados dentro dos grupos (SSW) e a soma dos quadrados entre grupos (SSB), com valores maiores indicando clusters mais densos e bem separados (Calinski; Harabasz, 1974).

Há ainda uma classe de indicadores para a validação externa de clusters, que busca avaliar a qualidade dos agrupamentos gerados por algoritmos de clusterização utilizando informações externas, como rótulos de classe conhecidos ou partições de referência. Segundo Halkidi et al. (2001), a validação externa visa responder à questão: “Quão bem os clusters correspondem a uma estrutura externa predefinida?”. Como importância da validação externa reside na sua capacidade de fornecer uma medida objetiva da qualidade dos clusters em relação a um padrão externo. E como nenhum trabalho desse foi desenvolvido, essa classe de validação não tem razão de ser utilizada para o presente estudo.

De outro lado, utiliza-se o PCA numa segunda etapa do trabalho. Para a obtenção dos componentes principais, realiza-se a diagonalização de matrizes simétricas e semi-definidas positivas formadas pelos dados em questão. Este processo permite o cálculo direto desses componentes, tornando o PCA uma ferramenta versátil e aplicável em diversas áreas científicas. A ampla disponibilidade de softwares e bibliotecas computacionais, capazes de realizar operações matriciais complexas, como a diagonalização, facilita a aplicação dessa metodologia em diferentes contextos (Jolliffe, 2002).

O PCA apresenta vantagens significativas, como a eliminação da multicolinearidade, transformando variáveis correlacionadas em componentes principais não correlacionados. Essa técnica também reduz a dimensionalidade dos dados, projetando-os em eixos ortogonais que representam a variabilidade em ordem decrescente, facilitando a interpretação e visualização (Hongyu, 2015; Regazzi, 2000; Jolliffe, 2002).

No entanto, o PCA possui limitações. É sensível a valores extremos (outliers), inadequado para dados com muitos valores ausentes ou quando o número de variáveis excede o de amostras. A redução de dimensionalidade implica perda de informação, embora a parte retida seja o padrão de resposta e a descartada, o ruído (erro e redundância). A eficácia do PCA varia: se as variáveis originais têm pouca correlação, a redução pode não simplificar o problema. No caso extremo de uma matriz de correlação identidade, os componentes principais são as próprias variáveis originais (Hongyu, 2015; Regazzi, 2000).

Formalmente, segundo Hongyu, Sandanielo e Oliveira Junior (2016), sejam as variáveis  $X_1, X_2, \dots, X_k$  em cada uma das  $n$  observações. Este conjunto de  $n \times k$  medidas originais forma uma matriz de dados  $\mathbf{X}_{(n \times k)}$ :

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \quad (5)$$

Para obter os componentes principais de uma forma geral, seja um conjunto de  $k$  variáveis  $X_1, X_2, \dots, X_k$  com médias  $\mu_1, \mu_2, \dots, \mu_k$  e variância  $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ , respectivamente. Tais variáveis são dependentes entre si e, portanto, possuem covariância entre a  $i$ -ésima e  $j$ -ésima variável definida por  $\sigma_{ij}$ , para  $i \neq j = 1, 2, \dots, k$ . Desse modo, as  $k$  variáveis podem ser expressas em forma vetorial por:  $X = [X_1, X_2, \dots, X_k]'$ , com vetor de médias  $\mu = [\mu_1, \mu_2, \dots, \mu_k]'$  e matriz de covariância  $\Sigma$ .

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1k}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_{nk}^2 \end{bmatrix} \quad (6)$$

Então, os pares de autovalores e autovetores  $(\lambda_1, v_1), (\lambda_2, v_2), \dots, (\lambda_k, v_k)$  são encontrados, de modo que se pode definir o  $i$ -ésimo componente principal a partir dos autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ , associados à matriz  $\Sigma$  (Johnson; Wichern, 1998; Hongyu, 2015):

$$Z_i = \mathbf{v}'_i \mathbf{X} = v_{i1}X_1 + v_{i2}X_2 + \cdots + v_{ik}X_k, \quad \forall i = 1, \dots, k. \quad (7)$$

O objetivo será determinar a variável  $Z_i$  a partir da  $k$  variáveis contidas na matriz  $X$ . A ideia envolve a transformação dos pontos de coordenadas originais em um plano, maximizando a distância entre eles, o que se traduz em aumentar ao máximo a variabilidade de  $Z_i$ . A variância de  $Z_i$  é dada por:

$$Var(\mathbf{Z}_i) = Var(\mathbf{v}'_i \mathbf{X}) = \mathbf{v}'_i Var(\mathbf{X}) \mathbf{v}_i = \mathbf{v}'_i \Sigma \mathbf{v}_i \quad (8)$$

em que  $i = 1, \dots, k$ .

Utilizando a decomposição espectral<sup>1</sup> da matriz  $\Sigma$ , tem-se que  $\Sigma = \Lambda P'$ , em que  $P$  é a matriz composta de autovetores de  $\Sigma$  em suas colunas, enquanto  $\Lambda$  representa a matriz diagonal formada pelos autovalores de  $\Sigma$ , ou seja:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} \quad (9)$$

<sup>1</sup> Seja  $A$  uma matriz simétrica  $n \times n$ , então  $A$  tem uma decomposição espectral  $A = CDC'$ , onde  $C$  é uma matriz  $n \times n$  nas quais suas colunas são os autovetores  $C_1, C_2, \dots, C_n$  que correspondem aos autovalores  $1, 2, \dots, n$  de  $A$ ; enquanto  $D$  é uma matriz diagonal  $n \times n$  que possui como diagonal principal  $1, 2, \dots, n$  (Wansbeek; Kapteyn, 1982).

e

$$\mathbf{P} = [v'_1 \quad v'_2 \quad \dots \quad v'_k] \quad (10)$$

Deste modo:

$$tr(\mathbf{\Sigma}) = tr(\mathbf{P}\mathbf{\Lambda}\mathbf{P}') = tr(\mathbf{\Lambda}\mathbf{P}\mathbf{P}') = tr(\mathbf{\Lambda}\mathbf{I}) = tr(\mathbf{\Lambda}) = \sum_{i=1}^k \lambda_i \quad (11)$$

Ou seja,  $tr(\mathbf{\Sigma})$  é dado pela soma dos elementos da diagonal principal:

$$tr(\mathbf{\Sigma}) = \sum_{i=1}^k \sigma_{ii} = \sum_{i=1}^k \lambda_i \quad (12)$$

Dessa forma, a quantidade total de variação presente nas variáveis originais equivale à quantidade total de variação presente nos componentes principais (Johnson; Wichern, 1998). A contribuição de cada componente principal ( $Z_i$ ) é expressa como uma porcentagem, e é possível calcular a explicação individual de cada componente, como por exemplo, para o  $k$ -ésimo componente principal, a proporção da explicação é determinada por:

$$C_k = \frac{Var(\mathbf{Z}_i)}{\sum_{i=1}^k Var(\mathbf{Z}_i)} = \frac{\lambda_i}{\sum_{i=1}^k \lambda_i} = \frac{\lambda_i}{tr(\mathbf{\Sigma})} \quad (13)$$

A análise da proporção da variância total explicada por cada componente principal auxilia na decisão de quantos componentes reter na Análise de Componentes Principais (PCA). Geralmente, priorizam-se os componentes que explicam a maior parte da variabilidade dos dados, ou seja, aqueles com maior variância (Hongyu et al., 2016; Manly, 1986). O primeiro componente principal é o mais importante, pois captura a maior variância, seguido pelo segundo componente, e assim por diante. Os últimos componentes, por outro lado, representam direções com menor variabilidade, podendo estar associados a relações lineares quase constantes entre as variáveis originais (Johnson; Wichern, 1998; Anderson, 2003). Esses componentes podem ser interpretados como ruído ou redundância nos dados e, portanto, podem ser descartados sem comprometer a qualidade do modelo.

A determinação do número ideal de componentes principais a serem retidos na Análise de Componentes Principais (PCA) é crucial para a interpretabilidade e eficácia do modelo. A avaliação da proporção da variância total explicada por cada componente oferece um ponto de partida. Geralmente, busca-se identificar componentes que expliquem a maior parte da variabilidade dos dados, priorizando aqueles com maior variância (Hongyu et al., 2016;

Manly, 1986). Componentes com variância menor podem ser considerados menos importantes ou até mesmo representar ruído nos dados (Johnson & Wichern, 1998; Anderson, 2003).

Diversos critérios podem auxiliar na seleção do número de componentes, cada um com suas vantagens e desvantagens:

- a) critério de Kaiser: Retém apenas os componentes com autovalores (variância explicada) maiores que 1, sob a premissa de que componentes com variância menor que uma variável original não contribuem significativamente para a explicação dos dados (Kaiser, 1960);
- b) scree plot (Jolliffe): Analisa visualmente o gráfico dos autovalores em ordem decrescente, buscando um ponto de inflexão ("cotovelo") que indique a partir de qual componente a variância explicada se torna marginal (Cattell, 1966);
- c) percentual de variância acumulada (Pearson): Define um percentual mínimo de variância total a ser explicada pelos componentes retidos, geralmente entre 70% e 90%, buscando um equilíbrio entre parcimônia e poder explicativo;
- d) teste de significância (Polegar): Utiliza testes estatísticos para avaliar a significância dos autovalores, retendo apenas os componentes com autovalores estatisticamente significativos (Horn, 1965).

A escolha do critério mais adequado depende dos objetivos da análise, das características dos dados e do conhecimento prévio sobre o problema em questão. É fundamental analisar a matriz de correlação das variáveis originais, buscando alta correlação entre elas (geralmente acima de 0,3) para garantir a aplicabilidade do PCA (Hongyu et al., 2016). No presente estudo, optou-se pela utilização do Critério de Kaiser (1960), retendo apenas os componentes principais com autovalores superiores a 1. Essa escolha se baseia na simplicidade e praticidade do critério, especialmente em análises exploratórias, e na busca por componentes que representem uma contribuição significativa para a variância total dos dados.

#### 4 BASE DE DADOS

A seleção de variáveis para a análise da informalidade neste estudo foi fundamentada em uma revisão bibliográfica abrangente, abarcando tanto estudos teóricos quanto empíricos sobre o tema (Felix et al., 2024). Essa revisão permitiu identificar as variáveis mais relevantes e frequentemente associadas à informalidade, guiando a seleção de dados e *proxies* adequadas para a análise. Para assegurar a robustez do modelo e mitigar problemas de multicolinearidade, a matriz de correlação de Pearson foi empregada para avaliar a relação entre as variáveis selecionadas. Conforme recomendado por Field (2009), as variáveis que apresentaram correlação inferior a 0,3 foram excluídas da análise, pois correlações baixas indicam relações fracas ou inexistentes, podendo comprometer a precisão e interpretabilidade dos resultados.

A escolha das variáveis se baseou em cinco categorias principais de determinantes da informalidade:

- a) custos de formalização: Representados pelas variáveis *tempo\_reg\_prop* (tempo para registrar uma propriedade) e *tempo\_inic\_neg* (tempo para iniciar um negócio), que capturam a burocracia e os custos associados à formalização de um negócio, um fator crucial na decisão de operar na informalidade (Djankov et al., 2002; Ulyssea, 2018; de Mel; McKenzie; Woodruff, 2013);
- b) regulamentações trabalhistas: A taxa de desemprego foi utilizada como um indicador indireto da rigidez do mercado de trabalho, sob a hipótese de que mercados de trabalho mais rígidos podem incentivar a informalidade como meio de reduzir custos (Ulyssea, 2018);
- c) baixo nível de capital humano: A variável *hc* (índice de capital humano) reflete o nível de educação e habilidades da força de trabalho. Um baixo nível de capital humano pode levar à informalidade por conta da falta de oportunidades no setor formal ou à percepção de que a educação formal não é necessária para o sucesso no setor informal (Gennaioli et al., 2013; La Porta; Shleifer, 2008);
- d) globalização: As variáveis *global\_econ* (globalização da economia) e *fdi\_pib* (fluxo de investimento estrangeiro direto em relação ao PIB) representam a integração do país na economia global. A globalização pode aumentar a concorrência e pressionar as empresas a reduzir custos, o que pode levar à informalidade (Canh; Dinh Thanh, 2020; Elbahnasawy, 2021);

e) características socioculturais: A variável *genderIne* (desigualdade de gênero) captura a disparidade entre homens e mulheres em termos de oportunidades e participação no mercado de trabalho. Essa desigualdade pode levar as mulheres a buscar empregos no setor informal, onde a discriminação pode ser menor.

De outro lado, as variáveis *mimic\_p* (estimativa do tamanho da economia informal por meio de modelos MIMIC) e *corrup* (corrupção) foram incluídas como controles importantes, refletindo a influência da percepção de corrupção e do tamanho estimado da economia informal (Medina; Schneider, 2018). As variáveis *idh* (Índice de Desenvolvimento Humano), *lib\_econ* (liberdade econômica), *inflação* (inflação) e *pib\_pc* (PIB per capita) foram incorporadas como controles para o nível de desenvolvimento, a qualidade institucional e as condições macroeconômicas do país, que podem influenciar a informalidade de maneira indireta. É importante ressaltar que a relação entre essas variáveis e a informalidade pode ser complexa e não linear (Felix et al., 2024). Os dados utilizados neste estudo são referentes a 2019, devido à disponibilidade limitada de algumas variáveis para anos posteriores.

Tabela 1 – Variáveis selecionadas.

Variável	Definição	Fonte
dge_p	Estimativas do modelo de equilíbrio geral dinâmico (DGE) da produção informal (% do PIB oficial)	Elgin, C., Kose, M. A., Ohnsorge, F., e Yu, S. 2021. “Understanding Informality”, C.E.P.R. Discussion Paper 16497, Centre for Economic Policy Research, Londres
mimic_p	Estimativas do modelo de múltiplos indicadores e múltiplas causas (MIMIC) da produção informal (% do PIB oficial)	Elgin, C., Kose, M. A., Ohnsorge, F., e Yu, S. 2021. “Understanding Informality”, C.E.P.R. Discussion Paper 16497, Centre for Economic Policy Research, Londres
hc	Índice de capital humano, baseado em anos de escolaridade e retornos à educação	Penn World Table (10.01)
emp	Número de pessoas empregadas (em milhões)	Penn World Table (10.01)
corrup	Índice de Percepção de Corrupção (0 a 10, corrupção máxima = 0)	Transparência Internacional
global_econ	Globalização da economia local do Banco Mundial (0 a 100, economia globalizada = 100)	Quality of Government - Basic Dataset
genderIne	Desigualdade de gênero (0 a 1, disparidade total = 1)	Quality of Government - Basic Dataset
idh	Índice de Desenvolvimento Humano (0 a 1)	Relatório de Desenvolvimento Humano (2024)



lib_econ	Liberdade econômica do Banco Mundial (0 a 10, liberdade máxima = 10)	Banco Mundial
fdi_pib	Investimento direto estrangeiro, entradas líquidas (% do PIB)	Fundo Monetário Internacional, Banco Mundial, OECD
inflacao	Inflação, deflator do PIB (anual %)	Banco Mundial, dados de contas nacionais e dados de contas nacionais da OECD
tempo_reg_prop	Tempo necessário para registrar propriedade (dias)	Banco Mundial, projeto Doing Business
tempo_inic_neg	Tempo necessário para iniciar um negócio (dias)	Banco Mundial, projeto Doing Business
desemprego	Desemprego total (% da força de trabalho total)	Organização Internacional do Trabalho, ILOSTAT, 2024
pib_pc	PIB per capita (em paridade de poder de compra e em US\$ de 2017)	Penn World Table (10.01)

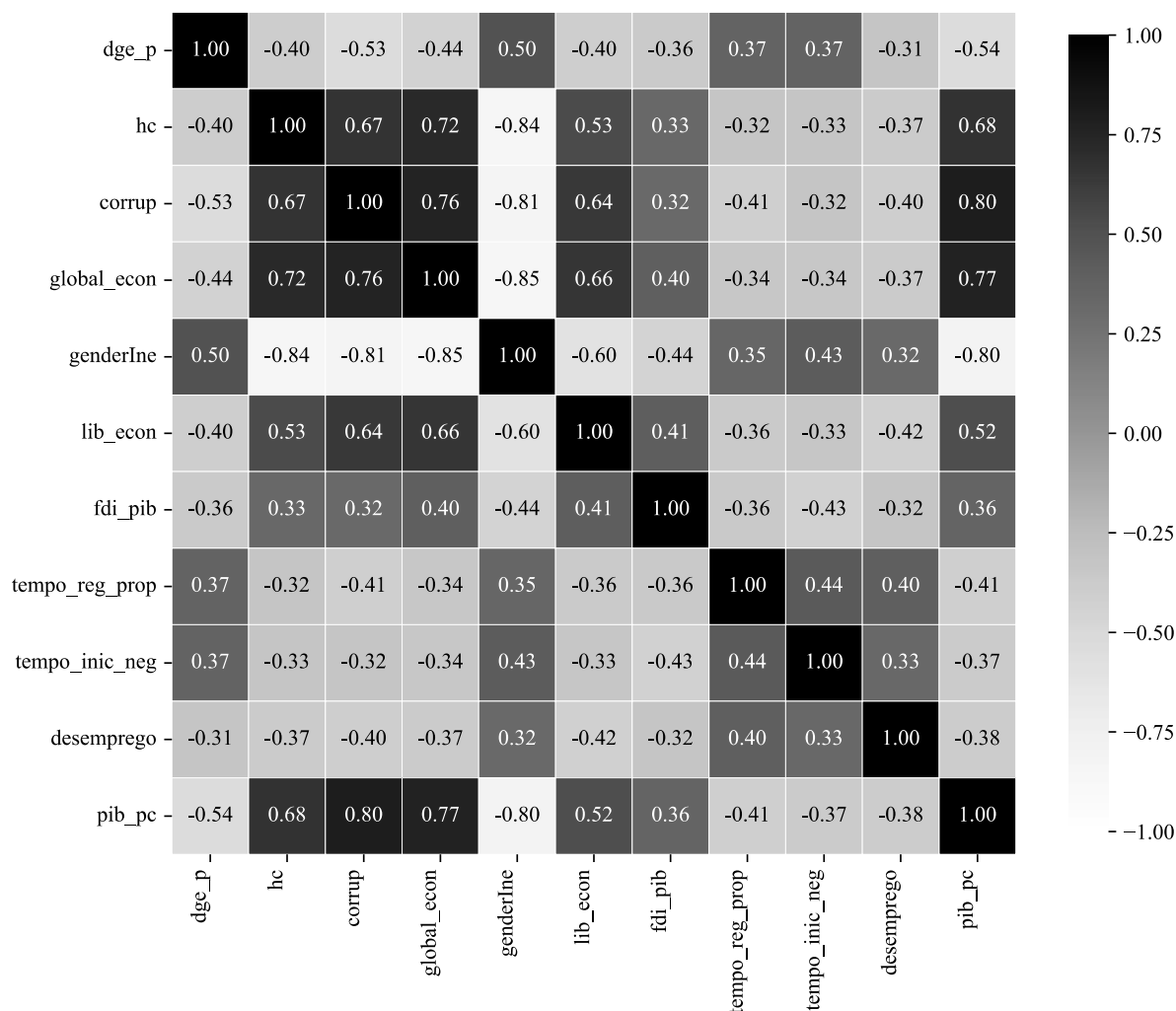
Fonte: Elaboração própria.

Após a remoção dos valores faltantes, a base de dados foi reduzida de 194 para 140 observações. Em seguida, uma matriz de correlação foi construída para identificar e lidar com potenciais problemas de multicolinearidade. O objetivo foi o de manter variáveis com correlação moderada, entre 0,3 e 0,85, conforme recomendado por Field (2009).

A matriz de correlação revelou relações importantes entre as variáveis. Destacam-se as altas correlações positivas entre *dge\_p* e *mimic\_p* (0,8); *dge\_p*, *tempo\_reg\_prop* e *tempo\_inic\_neg* (0,6); e *dge\_p* e *desemprego* (0,8). Essas correlações sugerem que o tamanho da economia informal está fortemente associado a outras estimativas do mesmo fenômeno, à burocracia nos processos de registro de propriedade e abertura de negócios, e às taxas de desemprego. Correlações moderadas foram observadas entre *dge\_p* e *hc* (-0,4) e *dge\_p* e *corrup* (-0,5), indicando que maior capital humano e menor percepção de corrupção tendem a estar associados com menores economias informais. A maioria das outras variáveis apresentou baixa correlação entre si.

Com base na análise da matriz de correlação, algumas variáveis foram excluídas. *mimic\_p* foi removida devido à alta correlação com *dge\_p*, evitando problemas de multicolinearidade em futuras análises de regressão. *emp* foi excluída por não apresentar correlação significativa com as demais variáveis. *idh* e *inflacao* também foram excluídas devido à baixa correlação, sugerindo que essas variáveis não estão fortemente associadas aos outros fatores em análise. As demais variáveis, mesmo com baixa correlação, foram mantidas para evitar a perda de informações potencialmente relevantes e permitir a investigação de relações não lineares ou interações complexas com a informalidade.

Figura 1 – Matriz de Correlação.



Fonte: Elaboração própria.

A análise descritiva das variáveis é fundamental para compreender a distribuição, variabilidade e tendências centrais dos dados, que pode ser observado na Tabela 2. Observa-se uma grande heterogeneidade no tamanho da economia informal (*dge\_p*) entre os países da amostra, com valores variando de 3,22% a 60,52% e média de 26,88%. O PIB per capita varia de 233,94 a 107.477, com média de 22.037, o que evidencia a desigualdade de renda entre os países da amostra.

O índice de capital humano (*hc*) apresenta uma média de 2,69, com valores entre 1,22 e 4,35, indicando diferenças significativas no nível de capital humano entre os países, entretanto a variável é a que apresenta a menor variabilidade, mensurada pelo coeficiente de variação, dentre as variáveis selecionadas. A percepção de corrupção (*corrup*), por sua vez, varia de 13 a 87, com média de 44,42, o que sugere que a percepção de corrupção é um problema presente em diversos países da amostra, porém com intensidade variável.

Ato contínuo, a taxa de desemprego apresenta média de 6,74%, mas com amplitude que vai de 0,1% a 25,54%, o que indica diferentes condições do mercado de trabalho entre os países. O nível de globalização de uma economia também sugere diferenças significativas entre os países. Em 2019, por exemplo, os países com maior e menor níveis de globalização da série foram, respectivamente, Singapura (94,24) e Irã (28,03).

Do mesmo modo, tem-se a variável que capta o nível de desigualdade entre gêneros: *genderIne*. Quanto mais próximo da unidade, maior o nível de desigualdade; menor será esse nível, quanto mais próximo de zero. A média dessa variável orbita em torno de 0,34.

O tempo necessário para registrar uma propriedade (*tempo\_reg\_prop*) e iniciar um negócio (*tempo\_inic\_neg*) varia consideravelmente entre os países, o que pode refletir diferenças na burocracia e eficiência dos sistemas regulatórios. Essas duas variáveis juntas com o investimento estrangeiro entre os países, principalmente, são as que apresentam a maior variabilidade da base.

Tabela 2 – Estatísticas descritivas das variáveis selecionadas dos 140 países.

Variáveis	Média	Mínimo	Mediana	Máximo	Desvio-Padrão	Coefficiente de Variação
dge_p	26,88	3,22	26,74	60,52	11,89	0,4423
hc	2,69	1,22	2,76	4,35	0,71	0,2639
corrup	44,42	13	40	87	19,14	0,4309
global_econ	59,02	28,03	58,06	94,24	17,18	0,2911
genderIne	0,34	0,02	0,36	0,79	0,2	0,5882
lib_econ	6,88	0	7,14	8,81	1,16	0,1686
fdi_pib	6,55	-11,69	2,36	234,25	26,76	4,0855
tempo_reg_prop	39,88	1	29	319	47,75	1,1973
tempo_inic_neg	18,7	0,5	11	230	27,32	1,4610
desemprego	6,74	0,1	4,97	25,54	5,12	0,7596
pib_pc	22.038	233,94	12.983,10	107.477	22.147,90	1,0050

Fonte: Elaboração própria.

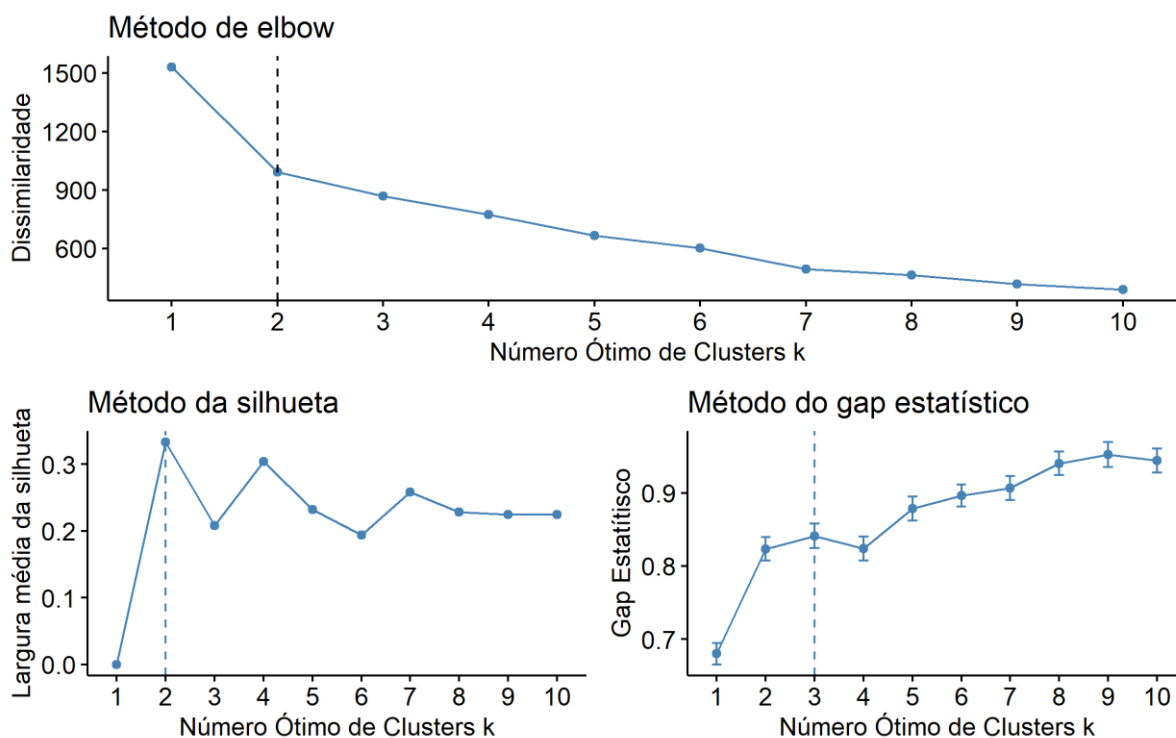
Nota: O coeficiente de variação é dado pela razão entre o desvio-padrão e a média.

## 5 RESULTADOS

Com as variáveis selecionadas para análise, procedeu-se à padronização dos dados, visando garantir que todas as variáveis tivessem a mesma escala e, assim, evitar que variáveis com maior magnitude dominassem o processo de agrupamento. A padronização consistiu em subtrair a média e dividir pelo desvio padrão de cada variável, resultando em dados com média zero e desvio padrão unitário. Em seguida, foram aplicados três métodos para determinar o número ideal de clusters: o método do cotovelo (*elbow*), o método da silhueta e o método do gap estatístico.

A Figura 2 exibe o método do cotovelo (primeiro gráfico) que explora a relação entre o número de clusters ( $k$ ) e a soma dos quadrados das distâncias intra-cluster (dissimilaridade). Observa-se uma queda acentuada na dissimilaridade até  $k = 2$ , seguida de uma diminuição mais suave. A partir de  $k = 2$ , a curva se assemelha a um “cotovelo”, sugerindo que dois seria um número adequado de clusters.

Figura 2 – Estatísticas das clusterizações.



Fonte: Elaboração própria.

Já o método da silhueta (segundo gráfico) apresenta a largura média da silhueta para diferentes valores de  $k$ . A largura da silhueta mede a qualidade da atribuição de cada ponto a

um cluster, sendo valores mais próximos da unidade indicativos de melhor agrupamento. O gráfico mostra que a largura média da silhueta é máxima para  $k = 2$ , corroborando a indicação do método do cotovelo.

Por fim, o gráfico do método do gap estatístico (terceiro gráfico) compara a variação total intra-cluster observada nos dados com a esperada em dados com distribuição uniforme. O número ideal de clusters é aquele que maximiza o gap estatístico, ou seja, a diferença entre a variação observada e a esperada. No gráfico, o maior gap ocorre em  $k = 3$ , sugerindo que três clusters poderiam ser uma escolha adequada.

Dessa feita, os resultados dos três métodos apontam para diferentes números de clusters. O método do cotovelo e da silhueta sugerem dois clusters, enquanto o método do gap estatístico indica três clusters. A decisão sobre o número de clusters a ser utilizado depende de uma análise mais aprofundada dos dados, conforme explanado na seção de metodologia. Dessa feita serão utilizados teste de validação interna de forma que seja possível elencar o número ótimo de agrupamento. A tabela abaixo apresenta os resultados dos testes de validação interna para dois e três clusters.

Tabela 3 – Resultados das métricas de validação interna dos clusters

Métricas	2 clusters	3 clusters
Índice de Silhueta	0,333	0,340
Índice de Davies- Bouldin	74,64	52,48
Índice de Calinski-Harabasz	1,32	0,97

Fonte: Elaboração própria.

Os resultados da validação interna dos clusters fornecem informações importantes sobre a qualidade dos agrupamentos obtidos. Para dois clusters, o índice da Silhueta de 0,333 indica uma separação moderada entre os grupos, enquanto o índice Davies-Bouldin de 74,64 sugere uma boa coesão interna dos clusters. No entanto, o índice Calinski-Harabasz de 1,32 aponta para uma sobreposição moderada entre os clusters. Ao aumentar o número de clusters para três, o Silhouette Score melhora ligeiramente para 0,340, indicando uma separação ligeiramente melhor entre os grupos. O índice Davies-Bouldin diminui para 52,48, sugerindo uma menor coesão interna dos clusters em comparação com a solução de dois clusters. Assim, foi escolhido o número ótimo de dois clusters.

A análise de clusters revelou dois grupos distintos de países com características socioeconômicas significativamente divergentes. Os dois grupos de países, juntamente com a análise dos componentes principais, são apresentados no Apêndice deste trabalho. O cluster 1,

composto majoritariamente por países desenvolvidos como Estados Unidos, nações da Europa Ocidental e Austrália, apresenta uma economia mais formalizada, com a produção informal representando apenas 19,5% do PIB oficial. Esse grupo também se destaca pelos elevados níveis de capital humano, com um índice de 3,28, refletindo altos níveis de educação e habilidades na força de trabalho. Além disso, esses países possuem maior integração à economia global, com uma pontuação de 76,20 no índice de globalização, e maior liberdade econômica, com uma média de 7,75, comparado a 6,30 no cluster 2. A maior atração de investimento estrangeiro direto é outro diferencial, com 12,21% do PIB, comparado a apenas 2,77% no cluster 2.

Tabela 4 – Médias das variáveis selecionadas para os clusters.

Variáveis	Cluster 1	Cluster 2
dge_p	19,5	31,8
hc	3,28	2,3
corrup	62,66	32,26
global_econ	76,2	47,56
genderIne	0,13	0,48
lib_econ	7,75	6,3
fdi_pib	12,21	2,77
tempo_reg_prop	25,9	49,2
tempo_inic_neg	9,97	24,51
desemprego	5,67	7,45
pib_pc	42.714,59	8.253,27

Fonte: Elaboração própria.

Adicionalmente, o cluster 1 apresenta tempos significativamente menores para registrar propriedades e iniciar negócios, com médias de 25,9 dias e 9,97 dias, respectivamente, indicando um ambiente de negócios mais eficiente e menos burocrático. A taxa de desemprego é relativamente baixa, com uma média de 5,67%, sinalizando um mercado de trabalho mais robusto. O PIB per capita é substancialmente maior, com média de \$42.714,59, refletindo um nível de prosperidade econômica elevado.

Em contraste, o cluster 2, formado principalmente por países em desenvolvimento como Brasil, México, China e Índia, enfrenta desafios significativos relacionados à informalidade. A produção informal nesse grupo representa 31,8% do PIB oficial, em média, possivelmente impulsionada por baixos níveis de capital humano, com um índice de 2,30, e elevada percepção de corrupção, com uma média de 32,26. Esses fatores corroboram estudos como o de Elgin et al. (2021), que destacam a influência negativa do capital humano e da corrupção na informalidade. A teoria da dualidade do mercado de trabalho (La Porta; Shleifer,

2014) sugere que a informalidade pode ser uma resposta à rigidez do mercado de trabalho formal e aos altos custos de formalização, agravada pela percepção de corrupção (Schneider; Enste, 2000).

Além da maior percepção de corrupção, o Cluster 2 apresenta maior desigualdade de gênero, com um índice de 0,48, indicando uma participação desigual no mercado de trabalho. O tempo necessário para registrar propriedades e iniciar negócios é significativamente maior, com médias de 49,2 dias e 24,51 dias, respectivamente, refletindo maior burocracia e ineficiência nos sistemas regulatórios. A taxa de desemprego é mais alta, com média de 7,45%, indicando um mercado de trabalho mais desafiador. O PIB per capita é significativamente menor, com média de \$8.253,27, o que reflete um nível de prosperidade econômica mais baixo.

Essas diferenças destacam a diversidade das economias informais e os diferentes desafios enfrentados pelos países de cada cluster. Enquanto o primeiro cluster beneficia-se de altos níveis de capital humano, integração econômica global e práticas regulatórias eficientes, o segundo luta com desafios relacionados à educação, governança e burocracia. A análise sugere que políticas públicas específicas e contextualizadas são necessárias para abordar as particularidades de cada grupo, promovendo melhorias na formalização da economia, na educação, na redução da corrupção e na inclusão econômica. Essa abordagem personalizada pode ajudar a mitigar os problemas de informalidade e promover o desenvolvimento socioeconômico de maneira mais eficaz.

Ato contínuo, a Análise de Componentes Principais (PCA) aprofunda a compreensão da clusterização, permitindo identificar os principais fatores que contribuem para essas diferenças. Esta foi conduzida para reduzir a dimensionalidade dos dados e identificar as variáveis mais influentes na formação dos clusters.

A Tabela 5 mostra os autovalores dos componentes principais e a variância explicada por cada componente. O primeiro componente principal ( $CP_1$ ) apresenta um autovalor de 5,2374, explicando 47,27% da variância total dos dados. O segundo componente principal ( $CP_2$ ) tem um autovalor de 1,2026, explicando mais 10,85% da variância. Juntos, os dois primeiros componentes principais explicam 58,13% da variância total dos dados. Este resultado é significativo, pois indica que mais da metade da variabilidade nos dados pode ser capturada por apenas dois componentes principais, facilitando a interpretação e visualização dos padrões subjacentes.

Para a escolha do número de componentes principais a serem retidos, foi utilizado o critério de Kaiser. Este critério recomenda reter apenas os componentes principais com autovalores superiores a 1. Como  $CP_1$  e  $CP_2$  são os únicos componentes que atendem a este

critério, a escolha por dois componentes principais é justificada. Este critério é amplamente aceito em análises exploratórias devido à sua simplicidade e eficácia em identificar os componentes mais significativos.

Sobre os demais componentes principais, vê-se que estes explicam proporções menores da variância, com o terceiro componente explicando 8,85% e o quarto 8,54%. A partir do quinto componente, por exemplo, a variância adicional explicada é inferior a 8%, sugerindo que os primeiros dois componentes principais são suficientes para capturar a maior parte da informação relevante nos dados.

Tabela 5 – Análise dos Componentes Principais: autovalores e variância acumulada.

Componentes	Autovalores	Diferença entre autovalores	Variância explicada	
			Proporção	Acumulada
1	5,2374	-	0,4727	0,4727
2	1,2026	-4,0349	0,1085	0,5813
3	0,9802	-0,2224	0,0885	0,6697
4	0,9457	-0,0345	0,0854	0,7551
5	0,8271	-0,1186	0,0747	0,8298
6	0,6200	-0,2071	0,0560	0,8857
7	0,4809	-0,1391	0,0434	0,9291
8	0,3088	-0,1721	0,0279	0,9570
9	0,2137	-0,0951	0,0193	0,9763
10	0,1687	-0,0451	0,0152	0,9915
11	0,0939	-0,0748	0,0085	1,0000

Fonte: Elaboração própria.

A Tabela 6 apresenta os coeficientes de ponderação dos dois componentes principais, mostrando a contribuição de cada variável para a formação dos componentes. A variável  $dge_p$ , que representa a produção informal como uma porcentagem do PIB oficial, apresenta um coeficiente negativo significativo no  $CP_1$  (-0,5043). Isso indica que, à medida que a produção informal aumenta, o valor do  $CP_1$  diminui. O segundo componente principal, por sua vez, apresenta um valor positivo (0,1492), de modo que sua interpretação segue contrário à anterior. Portanto, altos níveis de produção informal estão associados a baixos níveis de desenvolvimento econômico e institucional. Este resultado é coerente com a literatura que sugere que a informalidade está frequentemente associada a economias menos desenvolvidas e instituições mais fracas. A inclusão da variável  $dge_p$  reforça a importância da informalidade como um fator crítico que afeta o desenvolvimento econômico e institucional dos países.



O primeiro componente ( $CP_1$ ) é fortemente influenciado por variáveis como *global\_econ* (0,3979), *corrup* (0,3918), *hc* (0,3642) e *pib\_pc* (0,3842). Estas variáveis refletem características econômicas e institucionais, indicando que o  $CP_1$  pode ser interpretado como um componente que captura o nível de desenvolvimento econômico e institucional dos países. A variável *global\_econ* sugere que países mais integrados na economia global tendem a ter melhores indicadores econômicos e institucionais. A variável *corrup* indica que menores níveis de corrupção estão associados a um desenvolvimento econômico mais robusto. O índice de capital humano (*hc*) mostra a importância da educação e das habilidades da força de trabalho para o desenvolvimento econômico. Finalmente, o PIB per capita (*pib\_pc*) reflete diretamente a prosperidade econômica dos países.

Por outro lado, o segundo componente ( $CP_2$ ) é mais influenciado pelas variáveis *tempo\_inic\_neg* (0,5457) e *desemprego* (0,5179), sugerindo que este componente está relacionado às dificuldades burocráticas e às condições do mercado de trabalho. A contribuição negativa da variável *lib\_econ* (-0,2275) no  $CP_2$  também indica que países com maior liberdade econômica enfrentam menos barreiras burocráticas e menores taxas de desemprego. A variável *tempo\_inic\_neg* destaca as dificuldades enfrentadas para iniciar negócios, enquanto a variável *desemprego* reflete as condições desafiadoras do mercado de trabalho. Além disso, a variável *tempo\_reg\_prop* (0,2977) no  $CP_2$  indica que o tempo necessário para registrar propriedades também afeta as condições do mercado de trabalho e a eficiência burocrática.

Tabela 6 – Coeficientes de ponderação dos dois componentes principais.

Variáveis	$CP_1$	$CP_2$
dge_p	-0,5043	0,1492
hc	0,3642	0,0941
corrup	0,3918	0,0595
global_econ	0,3979	-0,0044
genderIne	-0,4097	-0,0544
lib_econ	0,3199	-0,2275
fdi_pib	0,1082	0,1017
tempo_reg_prop	-0,1722	0,2977
tempo_inic_neg	-0,1601	0,5457
desemprego	-0,0929	0,5179
pib_pc	0,3842	0,1163

Fonte: Elaboração própria.



corroborando a interpretação do  $CP_1$  como um indicador de desenvolvimento econômico e institucional.

O eixo vertical ( $CP_2$ ), conforme já explanado, distingue os países com base em suas condições burocráticas e de mercado de trabalho. Países com altos valores de  $CP_2$ , como Venezuela (VEN) e Haiti (HTI), enfrentam maiores dificuldades para iniciar negócios e apresentam taxas de desemprego mais altas. A posição da Venezuela, marcada por severas crises econômicas e instabilidade institucional, faz sentido, dado seu elevado  $CP_2$ . Da mesma forma, o Haiti, com problemas crônicos de governança e desafios burocráticos, reflete um  $CP_2$  ainda maior que o da Venezuela, destacando a gravidade das dificuldades enfrentadas pelo país.

## 6 CONCLUSÃO

Este estudo explorou a diversidade e os determinantes da informalidade em diferentes países, utilizando uma abordagem de clusterização e de Análise de Componentes Principais (PCA). Os resultados evidenciam a complexidade do fenômeno da informalidade e destacam a necessidade de políticas públicas específicas e contextualizadas para abordar as particularidades de cada grupo de países.

A análise de clusters revelou dois grupos principais de países com características socioeconômicas distintas. O cluster 1, composto majoritariamente por países desenvolvidos, destaca-se por uma economia mais formalizada, elevados níveis de capital humano, maior integração econômica global, maior liberdade econômica e maior atração de investimento estrangeiro direto. Esses países apresentam também um ambiente de negócios mais eficiente, refletido em tempos menores para registrar propriedades e iniciar negócios, além de uma taxa de desemprego relativamente baixa e um PIB per capita substancialmente maior.

Por outro lado, o cluster 2, formado principalmente por países em desenvolvimento, enfrenta desafios significativos relacionados à informalidade. Este grupo apresenta níveis mais baixos de capital humano, maior percepção de corrupção, maior desigualdade de gênero, maior burocracia para registrar propriedades e iniciar negócios, taxas de desemprego mais altas e um PIB per capita significativamente menor. A produção informal é mais prevalente neste grupo, representando uma parcela maior do PIB oficial.

O PCA, por sua vez, permitiu identificar os principais fatores que contribuem para essas diferenças. As variáveis mais influentes no primeiro componente principal ( $CP_1$ ) foram a globalização econômica, a liberdade econômica, a corrupção e o capital humano, enquanto o segundo componente principal ( $CP_2$ ) destacou as variáveis relacionadas à burocracia e ao mercado de trabalho, como o tempo para iniciar negócios e a taxa de desemprego. Além disso, a visualização dos componentes principais através da Figura 3 proporcionou compreensões adicionais sobre a distribuição dos países.

Países no segundo cluster demandam políticas focadas na melhoria do capital humano, no enfrentamento da corrupção, na simplificação dos processos burocráticos e na promoção da igualdade entre homens e mulheres. Já os países do primeiro podem se beneficiar de políticas que continuem a fomentar a integração econômica e a liberdade de mercado, aproximando-se de países como a Singapura, por exemplo.

Em resumo, este estudo contribui para a compreensão da economia informal ao identificar padrões e relações entre suas diferentes dimensões, oferecendo uma base sólida para

a formulação de políticas públicas mais eficazes e direcionadas. As análises apresentadas reforçam a necessidade de intervenções específicas para cada grupo de países, visando mitigar os problemas da informalidade e promover o desenvolvimento socioeconômico de maneira mais equitativa e sustentável.

## REFERÊNCIAS

- ABDI, H.; WILLIAMS, L. J. Principal component analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 2, n. 4, p. 433-459, 2010.
- ACEMOGLU, D.; JOHNSON, S.; ROBINSON, J. A. The colonial origins of comparative development: An empirical investigation. **American Economic Review**, v. 91, n. 5, p. 1369-1401, 2001.
- ANARAKI, K. H.; IBRAHIM, S.; PRASAD, R. A new methodology for solving the customer clustering problem using the K-P algorithm. **Expert Systems with Applications**, v. 168, p. 114223, 2021.
- BATOOL, F.; HENNING, J. **Determining the optimal number of clusters using the elbow and average silhouette methods**: A study on publicly available datasets. 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021.
- CALINSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. **Communications in Statistics-theory and Methods**, v. 3, n. 1, p. 1-27, 1974.
- CANH, N. P.; DINH THANH, S. Shadow economy and its determinants: Evidence from Vietnam. **Journal of Risk and Financial Management**, v. 13, n. 11, p. 269, 2020.
- DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, n. 2, p. 224-227, 1979.
- DE MEL, S.; MCKENZIE, D.; WOODRUFF, C. The demand for, and consequences of, formalization among informal firms in Sri Lanka. **American Economic Journal: Applied Economics**, v. 5, n. 2, p. 122-150, 2013.
- DJANKOV, S.; LA PORTA, R.; LOPEZ-DE-SILANES, F.; SHLEIFER, A. The regulation of entry. **The Quarterly Journal of Economics**, v. 117, n. 1, p. 1-37, 2002.
- ELBAHNASAWY, N. G. Globalization and the size of the informal economy: Empirical evidence from Egypt. **Theoretical and Applied Economics**, v. 28, n. 1, p. 147-162, 2021.
- ELGIN, C.; KOSE, M. A.; OHNSORGE, F.; YU, S. Understanding informality. C.E.P.R. Discussion Paper 16497, **Centre for Economic Policy Research**, London, 2021.
- FAWCETT, T.; PROVOST, F. **Data Science para Negócios**: O que você precisa saber sobre mineração de dados e pensamento analítico de dados. Alta Books Editora, 2016.
- FELIX, A.; SANDANIEL, A.; DE OLIVEIRA JUNIOR, S. **The determinants of informality**: A cross-country analysis. *Journal of Economic Studies*, 2024.
- FIELD, A. **Discovering statistics using SPSS** (3rd ed.). Sage, 2009.
- GENNAIOLI, N.; LA PORTA, R.; LOPEZ-DE-SILANES, F.; SHLEIFER, A. Human capital and regional development. **The Quarterly Journal of Economics**, v. 128, n. 1, p. 105-164, 2013.

HALKIDI, M.; BATISTAKIS, Y.; VAZIRGIANNIS, M. On clustering validation techniques. **Journal of Intelligent Information Systems**, v. 17, n. 2, p. 107-145, 2001.

JOLLIFFE, I. T. **Principal component analysis** (2nd ed.). Springer, 2002.

LA PORTA, R.; SHLEIFER, A. The unofficial economy and economic development. **Brookings Papers on Economic Activity**, v. 2008, n. 2, p. 275-363, 2008.

LA PORTA, R.; SHLEIFER, A. Informality and development. **Journal of Economic Perspectives**, v. 28, n. 3, p. 109-126, 2014.

LEHMANN, H.; PAN, J. The informal sector in developing countries: A literature review. **IZA Journal of Labor & Development**, v. 5, n. 1, p. 1-22, 2016.

LIU, Y.; LI, Z.; XIONG, H.; GAO, X.; WU, J. Understanding of internal clustering validation measures. In 2009 **IEEE International Conference on Data Mining**, 2009.

MEDINA, L.; SCHNEIDER, F. **Shadow economies around the world: What did we learn over the last 20 years?**. IMF Working Papers, 2018.

MILLIGAN, G. W.; COOPER, M. C. An examination of procedures for determining the number of clusters in a data set. **Psychometrika**, v. 50, n. 2, p. 159-179, 1985.

RODGERS, G. **Urban poverty and the labour market: Access to jobs and incomes in Asian and Latin American cities**. International Labour Office, 1989.

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53-65, 1987.

SCHNEIDER, F.; ENSTE, D. H. Shadow economies: Size, causes, and consequences. **Journal of Economic Literature**, v. 38, n. 1, p. 77-114, 2000.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 63, n. 2, p. 411-423, 2001.

ULYSSEA, G. Firms, informality, and development: Theory and evidence from Brazil. **American Economic Review**, v. 108, n. 8, p. 2015-2047, 2018.

WU, J. A novel K-means clustering algorithm based on principal component analysis. **Journal of Physics: Conference Series**, v. 2478, n. 3, p. 032013, 2023.

YUSUF, S. **Informal workers in the labor market**. The World Bank, 2007.

TRICHES, D.; BERTUSSI, L. A. S. Multicointegração e sustentabilidade da política fiscal no Brasil com regime de quebras estruturais (1997-2015). **Revista Brasileira de Economia**, v. 71, n. 3, p. 379-394, 14 jul. 2017.

## APÊNDICE A – PAÍSES DO CLUSTER 1 E DO CLUSTER 2

Tabela A1 – Países do cluster 1 em ordem decrescente para o primeiro componente principal.

País	CP1	CP2
Singapura (SGP)	5,4256	0,1885
Luxemburgo (LUX)	5,3301	1,8522
Suíça (CHE)	4,5810	0,5291
Irlanda (IRL)	4,0546	0,4766
Dinamarca (DNK)	3,9667	-0,0433
Noruega (NOR)	3,9638	0,0077
Holanda (NLD)	3,9362	-0,0479
Suécia (SWE)	3,6031	0,2808
Catar (QAT)	3,5255	-0,0195
Reino Unido (GBR)	3,4627	0,0856
Áustria (AUT)	3,4213	0,6570
Alemanha (DEU)	3,3922	0,2510
Canadá (CAN)	3,3687	0,0796
Finlândia (FIN)	3,3508	0,7087
Nova Zelândia (NZL)	3,3081	-0,1530
Austrália (AUS)	3,2634	0,0656
Estados Unidos (USA)	3,2427	0,2243
Emirados Árabes Unidos (ARE)	3,0604	-0,4904
Islândia (ISL)	3,0565	-0,0090
Estônia (EST)	3,0114	-0,5760
Japão (JPN)	2,9483	0,0816
Bélgica (BEL)	2,9424	0,0879
Chipre (CYP)	2,9026	0,3916
França (FRA)	2,6322	0,6103
República Tcheca (CZE)	2,6113	0,0633
Israel (ISR)	2,4741	0,0475
República da Coreia (KOR)	2,3589	-0,3620
Lituânia (LTU)	2,3528	-0,5958
Malta (MLT)	2,3327	-0,2585
Eslovênia (SVN)	2,2529	-0,0891
Espanha (ESP)	2,2355	0,7553
Eslováquia (SVK)	2,1647	0,3490
Portugal (PRT)	2,0053	-0,3052
Letônia (LVA)	1,9596	-0,4529
Hungria (HUN)	1,9441	-0,2942
Itália (ITA)	1,8947	0,1475
Chile (CHL)	1,7933	0,1278
Sérvia (SRB)	1,6696	1,2970
Bahrein (BHR)	1,6320	-0,5209
Croácia (HRV)	1,5767	-0,0135
Polônia (POL)	1,4158	0,9477
Malásia (MYS)	1,4090	-0,5786
Maurício (MUS)	1,3749	-0,4107
Arábia Saudita (SAU)	1,3431	0,1576



Grécia (GRC)	1,2663	0,7756
Romênia (ROU)	1,2140	-0,4799
Bulgária (BGR)	1,1882	-0,4524
Kuwait (KWT)	1,1569	0,0691
Costa Rica (CRI)	0,7819	0,4345
Brunei Darussalam (BRN)	0,7635	1,4854
Armênia (ARM)	0,7201	-0,4707
Albânia (ALB)	0,6463	-0,2248
Cazaquistão (KAZ)	0,6215	-1,1302
Uruguai (URY)	0,5698	-0,5044
Mongólia (MNG)	0,5359	-0,0323
Barbados (BRB)	0,4172	0,3045

Fonte: Elaboração própria.

Tabela A2 – Países do cluster 2 em ordem decrescente para o segundo componente principal.

País	CP1	CP2
Haiti (HTI)	-3,6753	4,8306
Venezuela (VEN)	-4,2604	4,6557
Laos (LAO)	-2,1453	2,7494
África do Sul (ZAF)	-0,7050	2,3429
Iraque (IRQ)	-1,8275	2,1226
Guiana (GUY)	-0,4025	2,0779
Irã (IRN)	-2,2637	2,0698
Namíbia (NAM)	-1,1228	2,0302
Botsuana (BWA)	-0,1519	1,7733
Congo (COG)	-3,3036	1,6719
Angola (AGO)	-3,5285	1,5760
Sudão (SDN)	-3,4216	1,4251
Maldivas (MDV)	-2,0340	1,2518
Síria (SYR)	-2,3565	1,2440
Iémen (YEM)	-3,2169	1,2340
Bangladesh (BGD)	-3,1020	1,1009
Jordânia (JOR)	0,2913	1,0456
Eswatini (SWZ)	-2,2059	0,8955
Argélia (DZA)	-2,0503	0,8077
Lesoto (LSO)	-1,7534	0,8073
Argentina (ARG)	-0,3536	0,7996
Tunísia (TUN)	-0,6430	0,6969
Gabão (GAB)	-2,2021	0,6968
Turquia (TUR)	-0,1025	0,3974
Índia (IND)	-1,2034	0,3191
Camboja (KHM)	-1,6783	0,2773
China (CHN)	0,2025	0,2584
Vietnã (VNM)	-0,0204	0,2172
Egito (EGY)	-1,3632	0,1942
Brasil (BRA)	-0,9387	0,1408
Equador (ECU)	-1,0025	0,1159
Etiópia (ETH)	-2,4766	-0,0567

Paraguai (PRY)	-1,1949	-0,1634
Colômbia (COL)	-0,7394	-0,1817
Nepal (NPL)	-2,0405	-0,1821
Paquistão (PAK)	-2,5599	-0,1997
Trinidad e Tobago (TTO)	0,2328	-0,2346
Indonésia (IDN)	-0,5136	-0,2443
Marrocos (MAR)	-0,9113	-0,2887
Mauritânia (MRT)	-2,3319	-0,3139
Quênia (KEN)	-1,6127	-0,3188
Ruanda (RWA)	-0,6509	-0,3366
Camarões (CMR)	-2,4999	-0,3433
República Dominicana (DOM)	-0,5063	-0,3827
Malawi (MWI)	-2,5629	-0,4454
Mianmar (MMR)	-1,9905	-0,5983
Burquina Faso (BFA)	-2,3847	-0,6139
Moçambique (MOZ)	-1,8490	-0,6215
República Centro-Africana (CAF)	-3,6757	-0,6574
Filipinas (PHL)	-0,6858	-0,6793
México (MEX)	-0,0942	-0,7016
Honduras (HND)	-1,4656	-0,7170
Sri Lanka (LKA)	-0,7334	-0,7544
Bolívia (BOL)	-1,9454	-0,8302
Uganda (UGA)	-1,4542	-0,8715
Ucrânia (UKR)	-0,2010	-0,9168
Tanzânia (TZA)	-2,4597	-0,9440
Nicarágua (NIC)	-1,3310	-0,9717
Madagascar (MDG)	-2,4200	-0,9861
Federação Russa (RUS)	0,2191	-0,9930
Tajiquistão (TJK)	-1,1381	-0,9992
Togo (TGO)	-2,1460	-1,0305
Quirguistão (KGZ)	-0,0239	-1,0370
Gana (GHA)	-1,1676	-1,0439
Zimbábue (ZWE)	-2,9741	-1,0505
Zâmbia (ZMB)	-1,2251	-1,0583
Serra Leoa (SLE)	-2,6592	-1,0831
Mali (MLI)	-2,7140	-1,1066
Libéria (LBR)	-2,5931	-1,1072
Jamaica (JAM)	-0,0541	-1,1080
Benim (BEN)	-2,3936	-1,1692
Gâmbia (GMB)	-2,0610	-1,2196
El Salvador (SLV)	-0,8326	-1,2286
República Democrática do Congo (COD)	-3,2061	-1,2425
Senegal (SEN)	-1,7632	-1,2605
Panamá (PAN)	0,2691	-1,2814
Peru (PER)	-0,3754	-1,4017
Costa do Marfim (CIV)	-2,2790	-1,4038
Níger (NER)	-2,8078	-1,5400
Burundi (BDI)	-2,8472	-1,5800
Nigéria (NGA)	-2,7753	-1,6695

República da Moldávia (MDA)	0,3750	-1,8041
Tailândia (THA)	-0,0815	-1,9164
Guatemala (GTM)	-1,5794	-1,9602

---

Fonte: Elaboração própria.