

Análise de agrupamento envolvendo o índice de percepção da corrupção mundial: Uma abordagem aplicada a América do Sul

Cluster analysis involving the world corruption perception index: An approach applied to South America

Análisis de conglomerados que involucra el índice mundial de percepción de la corrupción: Un enfoque aplicado a América del Sur

Recebido: 13/05/2020 | Revisado: 31/05/2020 | Aceito: 01/06/2020 | Publicado: 16/06/2020

Lucas Silva do Amaral

ORCID: <https://orcid.org/0000-0003-1096-3241>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: lucasihua@gmail.com

Diego Alves Gomes

ORCID: <https://orcid.org/0000-0003-2108-4672>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: diegoalvesestatistica@gmail.com

Augusto César Ferreira de Miranda Oliveira

ORCID: <https://orcid.org/0000-0001-6353-2096>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: auguustocesarf@gmail.com

Marília Gabriela Ferreira de Miranda Oliveira

ORCID: <https://orcid.org/0000-0003-1938-4772>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: marilia_gabri@hotmail.com

José Edvaldo de Oliveira Nunes

ORCID: <https://orcid.org/0000-0003-4417-1875>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: edvaldooliveiranunes@outlook.com

Maria Marciele de Lima Silva

ORCID: <https://orcid.org/0000-0001-5185-2711>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: marciele.liima@gmail.com

Natália Moraes Cordeiro

ORCID: <https://orcid.org/0000-0001-5294-1353>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: natalia_mcordeiro@hotmail.com

Edgo Jackson Pinto Santiago

ORCID: <https://orcid.org/0000-0003-3981-7635>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: edgoj@hotmail.com

André Luiz Pinto dos Santos

ORCID: <https://orcid.org/0000-0002-7703-2102>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: andredefensor@hotmail.com

Guilherme Rocha Moreira

ORCID: <https://orcid.org/0000-0001-6344-1151>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: guirocham@gmail.com

Resumo

A corrupção é um problema de âmbito mundial, proporcionando o agravamento dos cofres públicos, mas também prejuízos em aspectos sociais e de qualidade da população. Analisando este cenário, a ONG - Transparência Internacional, desde 1995, divulga anualmente o Índice de Percepção da Corrupção (IPC), principal indicador mundial da corrupção. Neste sentido, esta pesquisa buscou através de uma análise quantitativa e exploratória, verificar indicadores dos 12(doze) países sul-americanos: Índice de desenvolvimento humano (IDH), PIB per capita, Taxa de alfabetização, Índice de Gini, Índice de qualidade e acesso de serviços de saúde (QASS), Expectativa de Vida e Índice de Liberdade Econômica e compará-los com os dados do IPC do ano de 2019, com o intuito de calcular a correlação de Pearson, posteriormente realizou-se a análise de agrupamentos, com as variáveis que tiveram correlação significativa. Para tanto, utilizou-se as seguintes técnicas estatísticas multivariadas: Distância Euclidiana, para gerar uma matriz de distância entre os países, e o método Aglomerativo Hierárquico Average para a obtenção do dendrograma. Os resultados apontaram a formação de três grupos com diferentes características, em relação à corrupção. Portanto, pode-se afirmar que os indicadores analisados podem ter influência nos dados sobre a corrupção na América do Sul.

Palavras-chave: Países; distância euclidiana; dendrograma.

Abstract:

The Corruption is a global problem, causing the public coffers to worsen, but it also damages social and quality aspects of the population. Analyzing this scenario, the NGO - Transparency International, since 1995, annually discloses the Corruption Perception Index (IPC), the main global indicator of corruption. In this sense, this research sought, by means of a quantitative and exploratory analysis, to verify indicators of the 12 (twelve) countries of South America: Human Development Index (HDI), GDP per in habitant, Literacy rate, Gini Index, Index Gini's Index of Quality and Access to Health Services (QASS), life expectancy and economic freedom index and compare them with the IPC data for 2019, in order to calculate Pearson's correlation; subsequently, the cluster analysis was performed, with the variables that showed significant correlation. For this, the following multivariate statistical techniques were used: Euclidean distance, to generate a distance matrix between countries, and the Average Hierarchical Agglomerate method to obtain the dendogram. The results indicated the formation of three groups with different characteristics in relation to corruption. Therefore, it can be said that the analyzed indicators can influence data on corruption in South America.

Keywords: Countries; euclidean distance; dendrogram.

Resumen:

La corrupción es un problema global que empeora las arcas públicas, pero también las pérdidas en los aspectos sociales y de calidad de la población. Analizando este escenario, la ONG - Transparencia Internacional, desde 1995, publica anualmente el Índice de Percepción de la Corrupción (IPC), el principal indicador global de corrupción. En este sentido, esta investigación buscó, a través de un análisis cuantitativo y exploratorio, verificar los indicadores de los 12 (doce) países de América del Sur: Índice de Desarrollo Humano (IDH), PIB per cápita, tasa de alfabetización, índice de Gini, calidad y acceso a servicios de salud (QASS), índice de esperanza de vida y libertad económica y compárelos con los datos del IPC para el año 2019, para calcular la correlación de Pearson y luego se realizó el análisis de conglomerados, con las variables que mostraron una correlación significativa. Para esto, se utilizaron las siguientes técnicas estadísticas multivariadas: distancia euclidiana, para generar una matriz de distancia entre países, y el método de aglomeración jerárquica promedio para obtener el dendograma. Los resultados indicaron la formación de tres grupos con diferentes características en relación con la corrupción. Por lo tanto, se puede decir que los indicadores analizados pueden influir en los datos de corrupción en América del Sur.

Palabras clave: Países; distancia euclidiana; dendrograma.

1. Introdução

A corrupção é um dos principais problemas que influenciam negativamente o desenvolvimento da sociedade, ela não tem limites, é uma realidade transcultural, quanto mais comum a prática estiver na cultura de um país, mais fértil tende a ser a sua definição social (Sousa, 2016).

Com o intuito de fiscalizar e nortear a população sobre o quão corrupto é um país, a ONG Transparência internacional¹ desde 1995 divulga anualmente o IPC (Índice de Percepção de Corrupção), que estuda 180 países e territórios e os avalia em uma escala de 0 a 100, onde 0 significa que o país é percebido como altamente corrupto e 100 significa que o país é percebido com uma alta integridade.

Os países com altas pontuações no IPC têm altos níveis de desenvolvimento econômico (Produto Interno Bruto, PIB per capita), democracia liberal de longa data, imprensa livre, compartilhamento de poder, transparência da informação, estabilidade sociopolítica, forte cultura ética de respeito às leis e baixa corrupção (Sorensen ,2002; Treisman, 2007). Em contrapartida em países com baixos índices, as pessoas dão a corrupção como garantida. Os limitados recursos dificultam a legalidade no ganho do dinheiro, eles não têm nada a perder (liberdade, dignidade, integridade e reputação), todavia muita tentação de cobiçar alguma coisa (Tang et al, 2018; Drori et al., 2006; Vynoslavaska et al.,2005).

Recentemente alguns estudos analisam os indicadores dos países, calculando a correlação entre eles e o IPC (Cajaiba & Silva, 2018; Fernandes, 2017).

Cajaiba & Silva (2018) analisaram as seguintes variáveis: IDH, expectativa de vida ao nascer, anos de escolaridade esperados, PIB Per Capita, taxa de crescimento do PIB per Capita, rendimento nacional bruto (Gini esperado), liberdade econômica, liberdade fiscal, taxa de inflação e taxa de desemprego. Tendo como variáveis com correlação significativa: Gini esperado, IDH, liberdade econômica e expectativa de vida ao nascer.

Outros trabalhos utilizaram diferentes ferramentas estatísticas para estudar o comportamento da corrupção, abordando a estatística multivariada. (Carraro et al, 2012; Araújo, 2014).

¹ NGO Tansparency internacional: <https://beta.transparency.org/en/>

Ibid (2014) fez uso da análise de agrupamentos para obter-se a segregação de grupos semelhantes em relação à percepção dos tipos de corrupção no ambiente de trabalho entre profissionais de contabilidade.

Carraro et al.(2015) utilizaram a análise de componentes principais, para identificação da corrupção nos estados brasileiros.

No presente estudo buscou avaliar os indicadores a partir da correlação de Pearson e posteriormente a análise de agrupamento dos países sul-americanos, e sua influência com o resultado do índice de percepção da corrupção.

2. Metodologia

Os dados foram obtidos dos países sul-americanos, continente que contém uma extensão de cerca de 17,8 milhões de km², e comporta 6% da população mundial, distribuídos em 12 países (Argentina, Bolívia, Brasil, Chile, Colômbia, Equador, Guiana, Paraguai, Peru, Suriname, Uruguai e Venezuela).

Os dados coletados foram índices sociais e econômicos dos países de natureza quantitativa: Expectativa de vida, índice de desenvolvimento humano (IDH), Taxa de alfabetização, coeficiente de Gini, índice de liberdade econômica, índice de qualidade e acesso de serviços de saúde (QASS), PIB per capita, afim de realizar uma abordagem exploratória, segundo Pereira et al.(2018) essa abordagem tem uma forma muito comum de investigação que é o estudo de caso , que tem o intuito compreender um acontecimento ou um fenômeno em estudo, que neste caso é a corrupção.

A expectativa de vida é a estimação média de anos que se espera que um indivíduo possa viver em um determinado lugar. O índice de desenvolvimento humano (IDH) é um indicador que tem o propósito de medir, o grau de desenvolvimento de um determinado território tendo como campos analisados a educação, saúde e renda.

A taxa de alfabetização é uma porcentagem da população de um local que apresenta o quantitativo de indivíduos que sabem ler e escrever. O coeficiente de Gini esperado é uma medida que tem o intuito de aferir a desigualdade na distribuição da renda. O índice de liberdade econômica o grau de liberdade econômica, no que se refere a quão livre é a população em suas relações de comércio e o direito à propriedade privada.

O índice de qualidade e acesso de serviços de saúde (QASS), foi fundamentada por Barber et al.(2017), onde avaliaram o desempenho dos países no aspecto estrutural e sistemático no serviço de saúde oferecido aos cidadãos. O PIB per capita é a renda média por

habitante, de acordo com o resultado do Produto Interno Bruto (PIB) de um país.

Seguidamente, foi realizado o cálculo da correlação de Pearson (equação 1), segundo Figueiredo & Silva(2009) ela é uma medida da variância compartilhada entre duas variáveis a partir da associação linear. Com relação, a pesquisa foi utilizado para comparar o índice de percepção da corrupção Mundial aos indicadores para selecionar os mais correlatos.

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2][\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad (1)$$

No qual x_i é a observação do índice de percepção da corrupção de um determinado país, \bar{x} é a média aritmética do índice de corrupção de todos os países, y_i é a observação em um índice em específico de um determinado país, e \bar{y} é a média aritmética das observações dos países nesse mesmo índice.

Posteriormente, tomado os índices mais correlatos, a um nível de significância de 5% , realizou-se a análise de agrupamento, inicialmente será analisado as medidas de dissimilaridade para gerar a matriz de distâncias. As medidas de dissimilaridade consideram que se dois indivíduos são similares, eles estão próximos um do outro dessa forma eles são comuns ao conjunto de variáveis e vice-versa (Vicini, 2005).

Considerando que as variáveis presentes no estudo tem diferentes escalas então os resultados são distintos, o uso dos dados originais pode tornar muito difícil a interpretação do conceito de homogeneidade, por esta razão se faz necessário a padronização (Bussab,1990).

Após a extração da matriz, ocorreu a utilização de métodos aglomerativos, Hair et al., (2005) afirmam que os agrupamentos resultantes tendem a ter elevada homogeneidade interna e uma considerável heterogeneidade entre os agrupamentos).

O presente estudo optou pela utilização dos métodos aglomerativos hierárquicos. Os métodos hierárquicos não exigem que o usuário especifique previamente o número de agrupamentos e outros parâmetros adicionais. Outra vantagem significativa desses métodos é a de possibilitar a visualização do resultado da classificação por meio de um dendrograma, que ilustra de forma hierárquica o grau de semelhança entre os agrupamentos que são formados por divisões em cada estágio sucessivo da análise (Gonçalves,2008).

Os métodos de dissimilaridade e aglomerativo hierárquico utilizados serão definidos pelo coeficiente de correlação cofenética (CCC). Segundo afirma Valentin (2000), um método tem um desempenho melhor que do que o outro quando o dendrograma fornece uma imagem menos distorcida da realidade. É possível avaliar o grau de deformação provocado pela

construção do dendrograma, obtendo-se o coeficiente de correlação cofenética. Para as análises estatísticas, foi utilizado o Software livre R, com o auxílio do pacote “nbClust” e “Mvar.pt”.

3. Resultados

Na Tabela 1 estão presentes os resultados decorrentes da correlação de Pearson entre as variáveis e o índice de percepção da corrupção.

Tabela 1: Correlação entre o coeficiente de corrupção e os outros índices.

*Significativo ao nível de significância de 5%.

Índices	Correlação	Valor-p
Liberdade Econômica	0,697*	0,012
Taxa de Alfabetização	0,740*	0,006
Expectativa de Vida	0,353	0,261
IDH	0,667*	0,018
PIB per Capita	0,767*	0,004
QASS	0,511	0,089
Coefficiente de Gini	-0,271	0,394

Fonte: Elaboração dos próprios autores.

Os índices que tiveram correlações significativas foram: Índice de liberdade econômica, Taxa de Alfabetização, IDH e PIB per capita. Essa variáveis foram usadas para análise de agrupamentos

A Tabela 2, apresenta os coeficientes de correlação cofenética, que definiu os método de dissimilaridade e aglomerativos abordados.

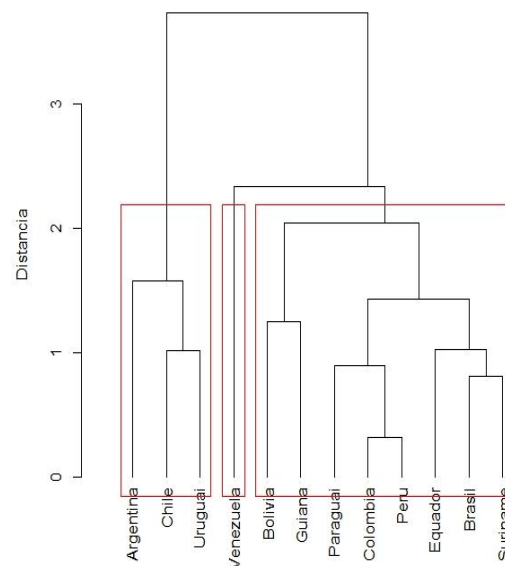
Tabela 2: Correlação cofenética entre os métodos de dissimilaridade e os métodos aglomerativos

	Ligação Completa	Ligação Simples	Average	Ward	Mcquitty
Mahalanobis	0.607552	0.606328	0.7025968	0.60855	0.6988723
Manhatham	0.871125	0.866619	0.8751416	0.86900	0.8718345
Euclidiana	0.871315	0.8668742	0.8753288	0.86916	0.8720239
Camberra	0.715147	0.710307	0.7603021	0.68956	0.7207619

Fonte: Elaboração dos próprios autores

A combinação que obteve o maior coeficiente de correlação cofenética, teve como medidas de dissimilaridade: A distância de Euclidiana e o método aglomerativo: Average, com o valor do coeficiente dado por 0,87532, o que implica em um bom ajustamento dos dados na análise de agrupamento. Dado esse valor extraímos o dendrograma (Figura 2):

Figura 2: Dendrograma gerado a partir do método aglomerativo



Fonte: Software R Pacote :“Mvar.pt”

De acordo com ele, temos a formação de três grupos: o Grupo 1 composto pela Argentina, Chile e Uruguai, o Grupo 2 é formado pela Bolívia, Guiana, Paraguai, Colômbia, Peru, Equador, Brasil e Suriname e por último o Grupo 3, com uma única nação pertencente, a Venezuela.

4. Discussão

No Grupo 1 estão presentes os países sul-americanos com menor Índice de Percepção corrupção, o Uruguai ocupa a 21ª posição em âmbito mundial, o Chile e a Argentina respectivamente na 26ª e a 66ª colocações. O que corrobora com o que afirma, Pires et al (2012), em que o Uruguai e Chile, desde a pós democratização nos anos 1985 e 1990, priorizaram a transparência, a prestação de contas e o acesso à informação como eixos fundamentais das políticas públicas de combate à corrupção e este acesso permite ao cidadão avaliar das ações praticadas por seus governantes dando-lhes capacidade de avaliar o desempenho do governo e

debater sobre suas ações.

A Argentina não tem um histórico favorável de luta a corrupção, o que leva a uma instabilidade social e econômica, contudo no último levantamento subiu levemente sua posição no ranqueamento.

No segundo grupo estão presentes nações com posições intermediárias ou baixas: Suriname em 70^a, Guiana em 85^a, Equador em 93^o, Colômbia em 96^o, Peru em 101^o, Brasil em 106^o, Bolívia em 123^o e Paraguai em 137^o. Fukuyama(2014) apud Nascimento(2018) argumenta que os altos graus de corrupção de alguns países da América Latina são consequência da introdução da democracia na região antes da construção dos seus Estados. A capacidade estatal desestabilizada propiciou o clientelismo, onde aqueles que chegavam ao poder por meio de eleições tinham na máquina estatal uma oportunidade de fonte de renda.

No último grupo, está presente a Venezuela que segundo o índice ocupa a 173^a posição. Estando entre os nove países com menor índice no levantamento. Segundo Gamboa (2016), a Venezuela atravessa, a mais grave crise econômica da sua história, com a inflação a disparar. A escassez de alimentos e medicamentos tornou-se comum. As taxas de homicídios alcançaram níveis históricos, além disso, o governo perseguiu líderes da oposição mais importantes sob acusações falsas, permitindo o uso da violência contra seus oponentes durante as campanhas eleitorais, o que acaba favorecendo a falta de transparência e evidencia um cenário propício à corrupção.

5. Considerações Finais

O objetivo do estudo foi alcançado visto que, as variáveis estudadas, influenciaram na observação dos resultados do índice de percepção da corrupção na América do Sul.

A aplicação da correlação de Pearson, possibilitou identificar os indicadores mais relacionados a corrupção, posteriormente com estas variáveis, partir da análise de agrupamentos, foi possível realizar um comparativo com o resultado do índice de percepção da corrupção, o que pode ser relevante na busca do avanço ao combate à corrupção nas nações sul-americanas.

Como perspectiva para novas pesquisas, pode-se futuramente trabalhar com a corrupção no âmbito mundial, assim como analisar outros indicadores que possam estar fortemente ligados a corrupção dos países.

Agradecimentos

A Universidade Federal Rural de Pernambuco (UFRPE) e ao PPGBEA(Programa de Pós Graduação em Biometria e Estatística Aplicada(PPGBEA). Este trabalho foi realizado com apoio da Coordenação ao de Aperfeiçoamento de Pessoal de nível Superior - Brasil (CAPES) - Código de Financiamento 001

Referências

Araújo R. F. (2014). *A percepção de diferentes tipos de corrupção na ótica dos profissionais da contabilidade*. UnB-Universidade de Brasília. Recuperado de: https://repositorio.unb.br/bitstream/10482/16292/1/2014_RodrigoFonsecadeAraujo.pdf

Barber R. M. et al.(2017). Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990–2015: a novel analysis from the Global Burden of Disease Study 2015. *The Lancet*, 390(10091): 231-66. doi: [https://doi.org/10.1016/S0140-6736\(17\)30818-8](https://doi.org/10.1016/S0140-6736(17)30818-8)

Bussab, W. de O., Miazaki, E. S., & Andrade, D. F. (1990). *Introdução a análise de agrupamentos*. São Paulo: Ime-Usp.

Cajaiba, K.S, & Silva, R. R. (2018). Determinantes socioeconômicos da percepção de corrupção nos países da América do Sul. *Ciência & Desenvolvimento-Revista Eletrônica da FAINOR*, 11(3), 608-632.

Carraro, A., Machado, I. B., Canever, M. D., & Boll, J. L. S. (2015). Proposta para a estimação da corrupção regional no Brasil. *Política & Sociedade*, 14(31), 326-352. doi: <http://dx.doi.org/10.5007/2175-7984.2015v14n31p326>

Drori, G. S., Meyer, J. W., & Hwang, H. (2006). *Globalization and organization: World society and organizational change*. Oxford University Press.

Fernandes, C. M. (2017). *Uma análise do índice de percepção da corrupção da transparência internacional (2000-2016)*. UFGD.

Figueiredo Filho, D. B., & Silva Júnior, J. A. D. (2009). Desvendando os Mistérios do Coeficiente de Correlação de Pearson (r). *Revista Política Hoje*, 18, 115-146.

Fukuyama, F. (2014). *Political order and political decay: From the industrial revolution to the globalization of democracy*. New York: Farrar, Straus and Giroux.

Gamboa, L. (2016). Venezuela: Aprofundamento do autoritarismo ou transição para a democracia?. *Relações Internacionais (R: I)*, (52), 55-66.

Gonçalves, M. L., de Andrade Netto, M. L., Zullo Jr, J., & Costa, J. A. F. (2008). Classificação não-supervisionada de imagens de sensores remotos utilizando redes neurais auto-organizáveis e métodos de agrupamentos hierárquicos. *Revista brasileira de cartografia*, 1(60).

Nascimento, L. N. (2018). Mais capacidade estatal, menos corrupção? Uma análise para a América Latina (1996-2015). *Revista de Sociologia e Política*, 26(68), 95-120. doi: <https://doi.org/10.1590/1678987318266806>

Pires, M. S., de Freitas Rodrigues, D., & Soares, A. G.(2012). Instituições de controle e accountability no Brasil, Chile e Uruguai: estabilidade política importa?. Disponível em: <https://cienciapolitica.org.br/index.php/system/files/documentos/eventos/2017/02/instituicoes-controle-e-accountability-brasil-chile-e.pdf>

Pereira, A.S., Shitsuka, D.M., Parreira, F.J. & Shitsuka, R. (2018). *Metodologia da pesquisa científica*. [e-book]. Santa Maria. Ed. UAB/NTE/UFSM. Disponível em: https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf?sequence=1.

Sørensen, J. B. (2002). The strength of corporate culture and the reliability of firm performance. *Administrative science quarterly*, 47(1), 70-91.

Sousa, L.de (2016). *Corrupção*. Lisbon. Fundação Francisco Manuel dos Santos.

Tang, T. L. P., Sutarso, T., Ansari, M. A., Lim, V. K., Teo, T. S., Arias-Galicia, F., ... & Vlerick, P. (2018). Monetary Intelligence and Behavioral Economics: The Enron Effect—Love of money, corporate ethical values, Corruption Perceptions Index (CPI), and dishonesty across 31 geopolitical entities. *Journal of Business Ethics*, 148(4), 919-937. doi: <https://doi.org/10.1007/s10551-015-2942-4>

Valentim, J. L. (2000). Ecologia numérica: Uma introdução à análise multivariada de dados. *Editora Interciências, Rio de Janeiro*.

Vicini, L. (2005). Análise multivariada: da teoria à prática. Santa Maria, Universidade Federal de Santa Maria, 215p. (Monografia de Especialização)

Vynoslavska, O., McKinney, J. A., Moore, C. W., & Longenecker, J. G. (2005). Transition ethics: A comparison of Ukrainian and United States business professionals. *Journal of Business Ethics*, 61(3), 283-299.

Porcentagem de contribuição de cada autor no manuscrito

Lucas Silva do Amaral – 50 %

Diego Alves Gomes – 8%

Augusto César Ferreira de Miranda – 7%

Marília Gabriela Ferreira de Miranda Oliveira – 5%

José Edvaldo de Oliveira Nunes – 6%

Natália Moraes Cordeiro – 5%

Maria Marciele de Lima Silva – 5%

Edgo Jackson Pinto Santiago – 4%

André Luiz Pinto dos Santos – 5%

Guilherme Rocha Moreira – 5%