

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
FEA - Faculdade de Economia e Administração
Programa de Estudos Pós-Graduados em Administração

METODOS QUANTITATIVOS

TRABALHO FINAL

Análise de indicadores econômicos em comparação com Índice de Desenvolvimento Humano (HDI) e Índice de desigualdade (GINI)

Disciplina: Métodos Quantitativos

Professor: Dr. Arnaldo Jose de Hoyos

RODRIGO GURIAN BORTOLOTO

Trabalho final: 08/06/2011

1. INTRODUÇÃO

O presente trabalho tem por objetivo efetuar uma avaliação de países através de dados referentes a indicadores econômicos internacionais obtidos junto ao Banco Mundial e às Nações Unidas. Para tal, na primeira parte realiza-se a análise exploratória dos dados, incluindo a definição dos indivíduos e das variáveis, os significados e unidades de medida. Após esta análise faz-se o estudo utilizando vários métodos de classificação das variáveis (análise discriminante, regressão logística e árvores de classificação). O software estatístico utilizado é o **MINITAB** e **XLSTAT**.

2. ENTENDENDO OS DADOS

Os dados utilizados neste trabalho foram coletados a partir de informações disponibilizadas pelo Banco Mundial (<http://databank.worldbank.org/ddp/home.do>). Somente os valores de HDI foram coletados no site das Nações Unidas (<http://hdr.undp.org/en/statistics/>) e agrupados aos dados do Banco Mundial.

2.1 Os Indivíduos

Dos mais de 200 países com informações disponíveis, a partir das análises exploratórias e classificatórias, os indivíduos foram reduzidos a 50 países. No processo de avaliação os países considerados *outliers* foram sendo retirados para permitir melhor análise.

2.2 As Variáveis

As variáveis utilizadas neste trabalho foram escolhidas para contemplar uma visão de indicadores econômicos e de desenvolvimento (HDI) e desigualdade. Sabe-se, por estudos anteriores, que tanto o Índice de Desenvolvimento Humano (HDI) quanto o Índice de Desigualdade (GINI) são os indicadores que mais representam as condições de desenvolvimento social de um país. Na tabela abaixo são apresentadas as variáveis utilizadas:

Variável	Significado	Unid. Medida	Tipo
Crescimento PIB per capita (CresPIBpercapita)	Crescimento do Produto interno Bruto em 2009	% (anual)	Contínua
Inflação (INF)	Inflação dos produtos ao consumidor	% (anual)	Contínua
Desemprego (UNEM)	Desemprego total, como percentual da força de trabalho	%	Contínua
GINI	Índice de desigualdade	Índice	Contínua
HDI	Índice de Desenvolvimento Humano	Índice	Contínua
Crescimento do PIB (CresPIB)	Crescimento do PIB em 2009	%	Contínua
Produto Interno Bruto per capita (PIB)	Produto Interno Bruto dos países em 2009 com valores atualizados	USD	Contínua

Tabela 1: Variáveis

2.3 A Tabela de Dados

A tabela original dos dados, por ser muito grande não está apresentada neste documento, mas pode ser disponibilizada. Lembrando que para as análises, países sem informação ou com informação inadequada foram excluídos da análise, totalizando 50 países após a depuração.

3. ANÁLISE DAS VARIÁVEIS

3.1. ANÁLISE DO CONJUNTO DAS VARIÁVEIS

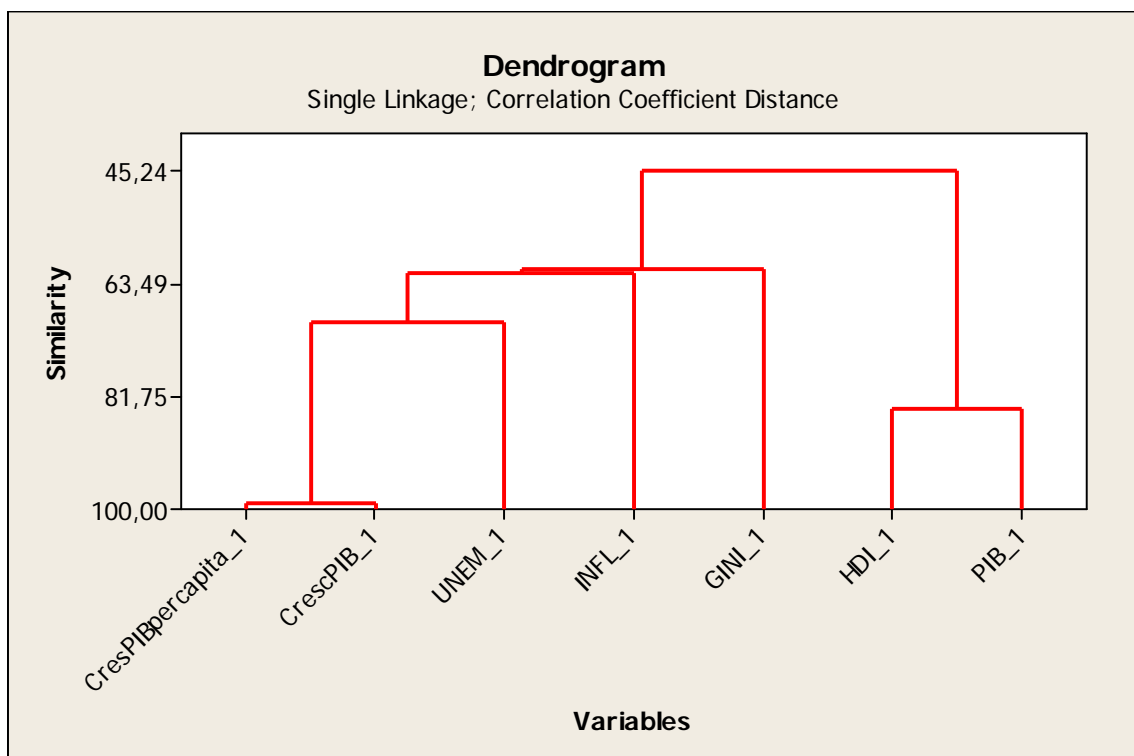
Antes de iniciar a análise propriamente, fez-se a análise das correlações das variáveis utilizando-se o método dos Componentes Principais para identificar a possibilidade de se reduzir o número de variáveis, simplificando assim o modelo de dados e possibilitando uma maior assertividade nos resultados.

A seguir o resultado das análises realizadas no MINITAB.

Correlations: CresPIBperca; INFL_1; UNEM_1; GINI_1; HDI_1; CrescPIB_1; PIB_1

	CresPIBpercapita	INFL_1	UNEM_1	GINI_1	HDI_1	CrescPIB_1
INFL_1	0,220 0,124					
UNEM_1	0,396 0,004	0,026 0,857				
GINI_1	0,143 0,324	0,221 0,122	-0,093 0,522			
HDI_1	-0,434 0,002	-0,363 0,010	-0,095 0,511	-0,358 0,011		
CrescPIB_1	0,978 0,000	0,235 0,101	0,330 0,019	0,194 0,177	-0,522 0,000	
PIB_1	-0,316 0,025	-0,409 0,003	-0,298 0,036	-0,501 0,000	0,673 0,000	-0,313 0,027

Cell Contents: Pearson correlation
P-Value



A partir da tabela de correlações e do Dendograma, pode-se depreender que as variáveis de Crescimento do PIB e Crescimento do PIB per capita têm uma alta correlação (conforme indicado na tabela acima em amarelo 0,978) e portanto, se consideradas conjuntamente, não apresentarão muitas diferenças em seu comportamento em conjunto com as demais variáveis, podendo ser utilizada apenas uma na sequência do trabalho.

Situação semelhante acontece com as variáveis HDI e PIB (conforme indicado na tabela acima em amarelo 0,673), porém como neste caso a correlação não é tão grande, as duas variáveis permanecerão na análise.

Principal Component Analysis: CresPIBperca; INFL_1; UNEM_1; GINI_1; HDI_1; Cres

Eigenanalysis of the Correlation Matrix

Eigenvalue	3,1400	1,4530	0,8586	0,7781	0,5519	0,2072	0,0113
Proportion	0,449	0,208	0,123	0,111	0,079	0,030	0,002
Cumulative	0,449	0,656	0,779	0,890	0,969	0,998	1,000

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
CresPIBpercapita_1	0,454	0,414	0,297	-0,011	0,134	-0,234	-0,679
INFL_1	0,284	-0,307	-0,049	0,835	0,337	0,105	-0,002
UNEM_1	0,226	0,444	-0,761	-0,086	0,206	0,349	0,012
GINI_1	0,265	-0,498	0,119	-0,527	0,566	0,260	-0,044
HDI_1	-0,441	0,224	-0,054	0,019	0,682	-0,520	0,131
CrescPIB_1	0,467	0,360	0,370	-0,025	0,062	-0,021	0,714
PIB_1	-0,422	0,330	0,419	0,125	0,193	0,689	-0,099

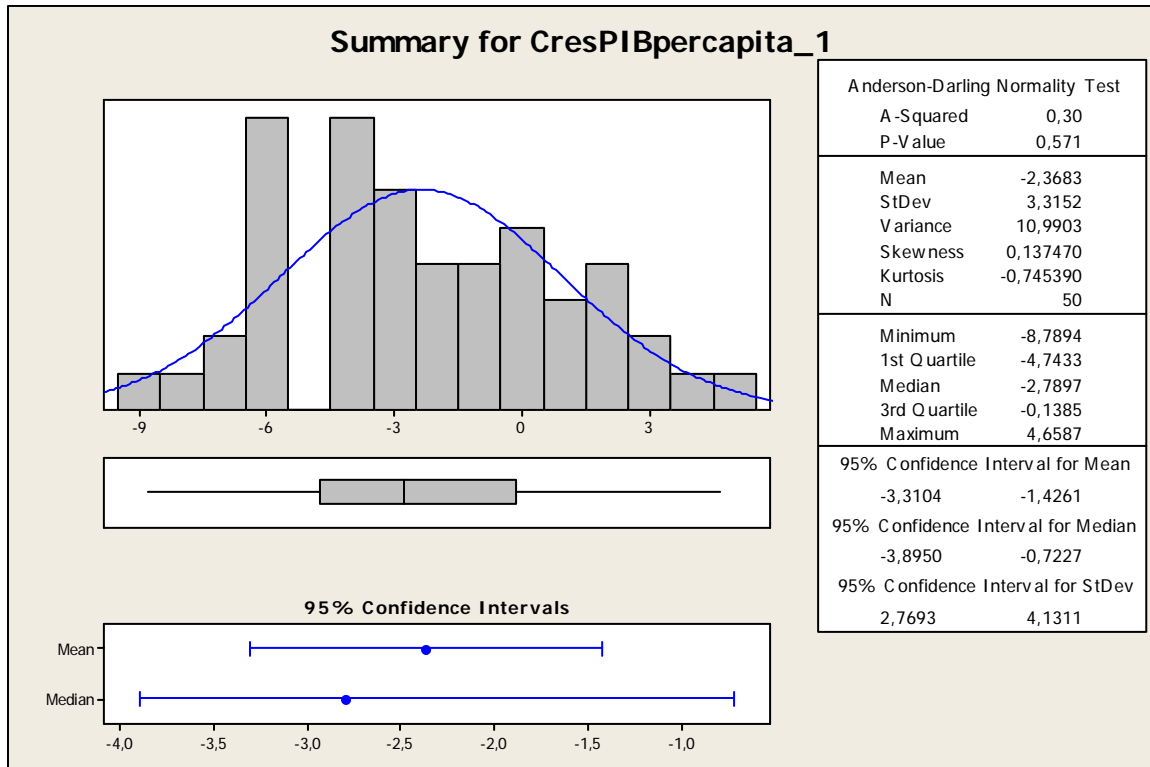
Apesar do objetivo não ser identificar um índice a partir das variáveis apresentadas que possa representar uma nova avaliação, a forma como os componentes de PC4 são agrupados representa claramente um contraste entre o Índice de Inflação e o Índice de Desigualdade. No caso do PC1, pode-se entender como contraste entre o crescimento da economia e o grau de desenvolvimento humano.

A relação entre estas variáveis deve ser analisada com mais profundidade na sequência do trabalho.

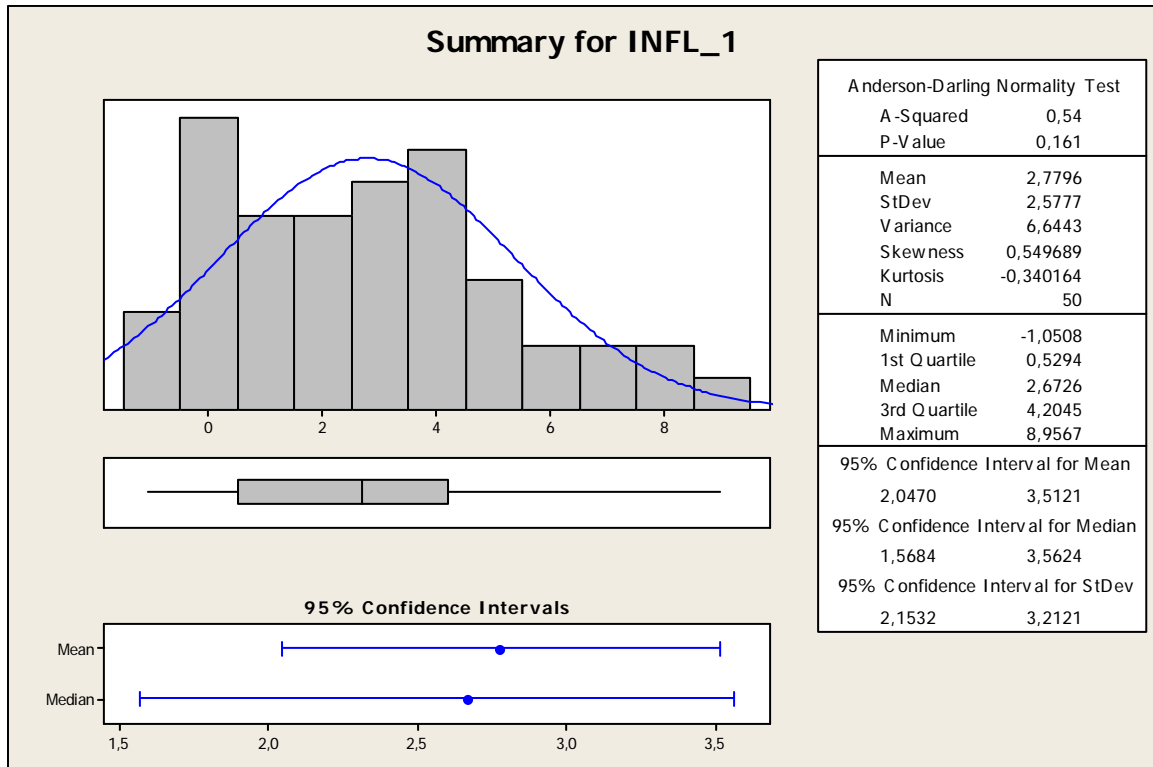
Assim, para a sequência do trabalho, será excluída somente a variável Crescimento PIB, por já estar bem representada pelo Crescimento PIB per capita.

3.2. ANÁLISE EXPLORATÓRIA

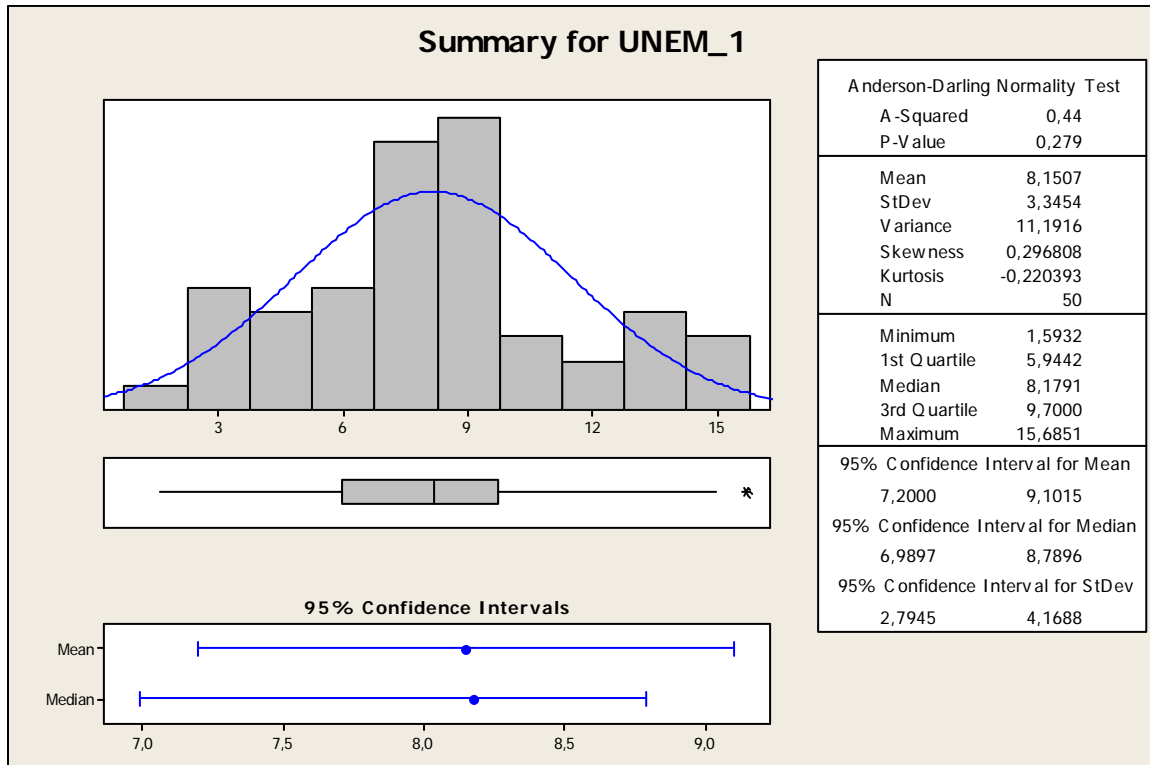
A seguir, são apresentados os gráficos resumo das estatísticas de cada variável do estudo e um breve resumo das suas condições.



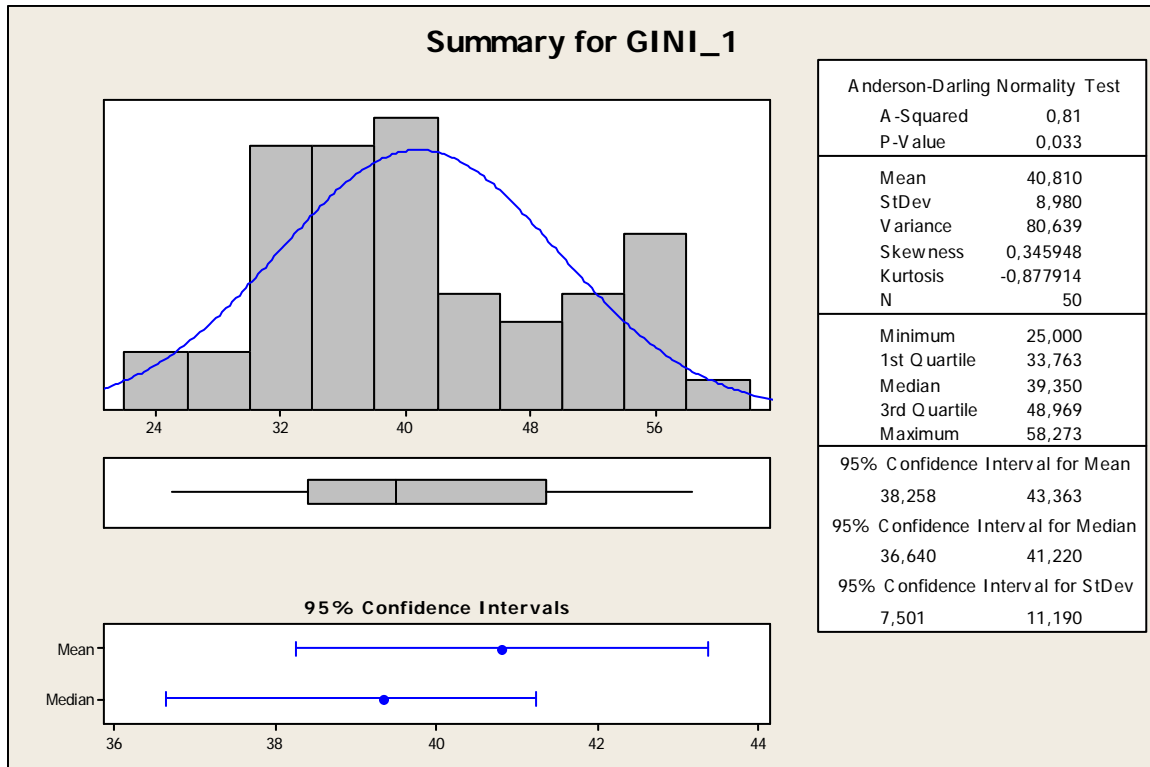
- Forma: O Histograma e os indicadores de Anderson Darling permitem verificar que não se trata de uma distribuição normal. A curva representativa da distribuição apresenta uma leve assimetria para a esquerda, que também pode ser verificada na distância entre os valores de média e mediana.
- Valores Atípicos: Não há valores atípicos.
- Centro e Dispersão: A mediana nos indica que metade dos países tem Taxa de Crescimento do PIB per capita menor do que -2,7897% e metade maior do que este valor. A média da amostra é de -2,3683%, com desvio-padrão (medida de dispersão) 3,3152. A mediana neste caso é um pouco menor que a média, conforme já mencionado, indicando a leve assimetria que também pode ser verificada na curva.



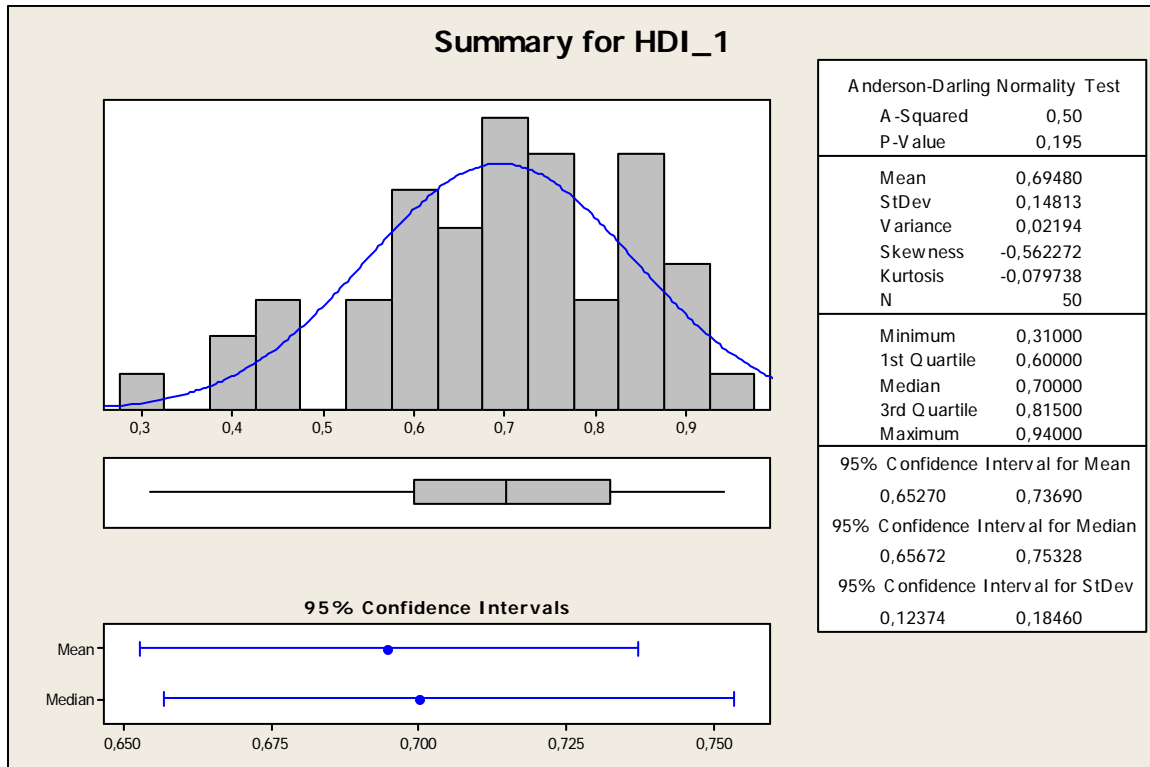
- **Forma:** O Histograma e os indicadores de Anderson Darling permitem verificar que não se trata de uma distribuição normal. A curva representativa da distribuição apresenta-se assimétrica para a esquerda. Neste caso a distância dos valores de média e mediana não indicam a assimetria pela alta concentração dos valores no lado esquerdo do histograma.
- **Valores Atípicos:** Não há valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que metade dos países tem Taxa de Inflação menor do que 2,6726% e metade maior do que este valor. A média da amostra é de 2,7796%, com desvio-padrão (medida de dispersão) 2,5777. A mediana neste caso está bastante próxima da média.



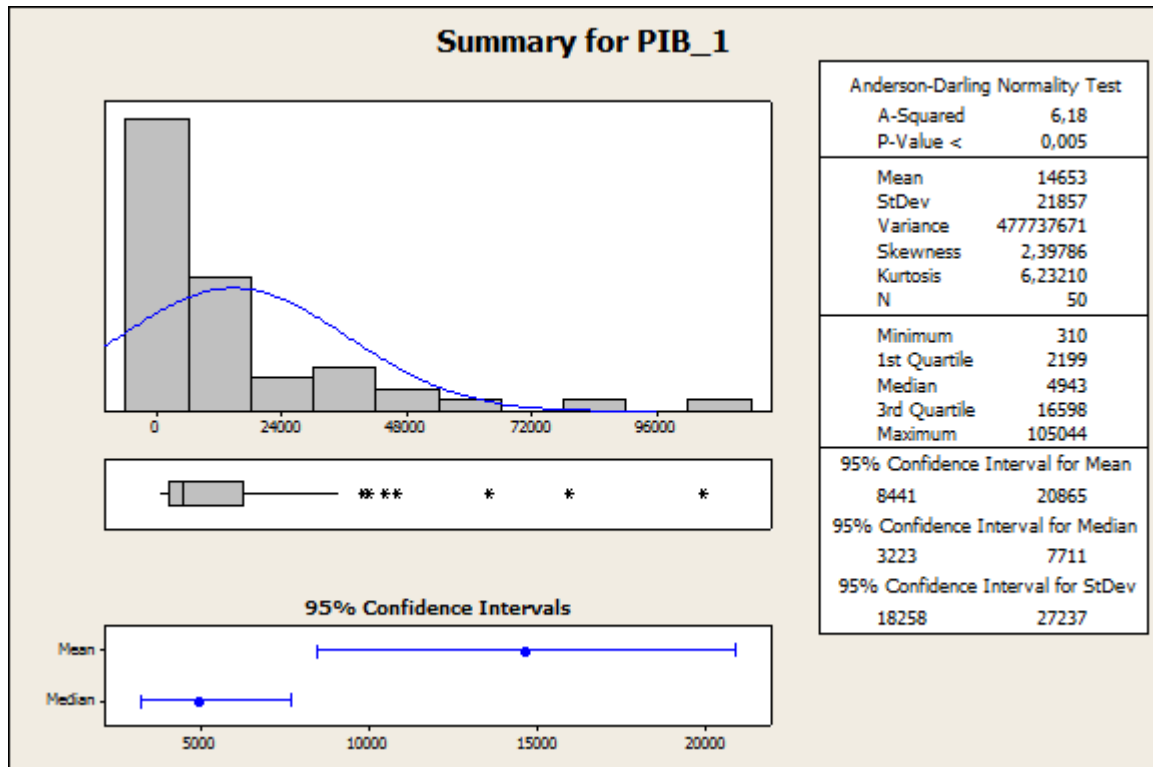
- **Forma:** O Histograma e os indicadores de Anderson Darling permitem verificar que não se trata de uma distribuição normal. A curva representativa da distribuição apresenta uma característica simétrica, também corroborada pela proximidade dos valores de média e mediana.
- **Valores Atípicos:** Existe um valor atípico (outlier) no lado direito do gráfico. Este valor de Desemprego (15,6851) é da República do Yemen.
- **Centro e Dispersão:** A mediana nos indica que metade dos países tem Taxa de Desemprego menor do que 8,1791% e metade maior do que este valor. A média da amostra é de 8,1507%, com desvio-padrão (medida de dispersão) 3,3454. A mediana neste caso está bastante próxima da média, conforme já mencionado, indicando uma simetria que também pode ser verificada na curva.



- Forma: O Histograma e os indicadores de Anderson Darling permitem verificar que não se trata de uma distribuição normal. A curva representativa da distribuição apresenta leve assimetria para a esquerda. Neste caso a distância do valor de média e mediana não indicam a assimetria pela alta concentração dos valores no lado esquerdo do histograma.
- Valores Atípicos: Não há valores atípicos.
- Centro e Dispersão: A mediana nos indica que metade dos países tem Índice de Desigualdade menor do que 39,350 e metade maior do que este valor. A média da amostra é de 40,810, com desvio-padrão (medida de dispersão) 8,980. A mediana neste caso está relativamente distante da média.



- Forma: O Histograma e os indicadores de Anderson Darling permitem verificar que não se trata de uma distribuição normal. A curva representativa da distribuição apresenta uma assimetria com deslocamento para a esquerda.
- Valores Atípicos: Não há valores atípicos.
- Centro e Dispersão: A mediana nos indica que metade dos países tem Índice de Desenvolvimento Humano menor do que 0,7000 e metade maior do que este valor. A média da amostra é de 0,6948, com desvio-padrão (medida de dispersão) 0,14183. A mediana neste caso está bastante próxima da média.



- Forma: O Histograma e os indicadores de Anderson Darling permitem verificar que se trata da variável com o maior grau de normalidade deste estudo. A curva representativa da distribuição apresenta-se achatada e bastante assimétrica para direita. Este comportamento também pode ser percebido pela distância entre os valores de média e mediana.
- Valores Atípicos: Existem valores atípicos (outlier) do lado esquerdo do gráfico, indicando países com maior Produto Interno Bruto. Os países são: Canadá, Suíça, Alemanha, Luxemburgo, Noruega, Suécia e Estados Unidos.
- Centro e Dispersão: A mediana nos indica que metade dos países tem Produto Interno Bruto menor do que USD 4.943 trilhões e metade maior do que este valor. A média da amostra é de USD 14.653 trilhões, com desvio-padrão (medida de dispersão) USD 21.857 trilhões. Este valor de desvio-padrão é extremamente alto, indicando a grande dispersão também percebida no histograma. A mediana neste caso está relativamente distante da média e pode ser também indicativo da assimetria da curva para a direita.

3.3. ANÁLISE DE CONGLOMERADOS – CLASSIFICAÇÃO EM GRUPOS (CLUSTERS)

Abaixo são apresentados dois Dendogramas: um com o agrupamento dos países utilizando-se as variáveis Crescimento PIB per capita, Inflação, Desemprego, GINI, HDI e outro acrescentando-se o PIB ao grupo anterior.

Cluster Analysis of Observations: CresPIBperca; INFL_1; UNEM_1; GINI_1; HDI_1

Euclidean Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	49	94,9646	1,74235	25	26	25	2
2	48	94,6226	1,86071	23	30	23	2
3	47	93,9167	2,10497	1	5	1	2
4	46	92,9819	2,42843	27	34	27	2
5	45	92,3467	2,64822	36	49	36	2
6	44	91,5994	2,90681	21	35	21	2
7	43	90,8427	3,16862	22	27	22	3
8	42	90,7528	3,19973	14	16	14	2
9	41	90,2189	3,38450	22	29	22	4
10	40	90,0843	3,43107	7	37	7	2
11	39	89,8756	3,50326	10	14	10	3
12	38	89,7450	3,54847	48	50	48	2
13	37	89,6215	3,59120	9	10	9	4
14	36	89,3672	3,67920	28	44	28	2
15	35	89,2824	3,70855	23	32	23	3
16	34	89,2206	3,72990	9	11	9	5
17	33	89,1031	3,77056	20	42	20	2
18	32	89,0575	3,78634	9	15	9	6
19	31	88,9609	3,81980	22	46	22	5
20	30	88,9254	3,83205	9	23	9	9
21	29	88,6477	3,92815	9	22	9	14
22	28	88,4264	4,00472	36	47	36	3
23	27	88,2962	4,04979	4	31	4	2
24	26	88,2694	4,05905	3	39	3	2
25	25	88,2387	4,06968	41	48	41	3
26	24	88,0165	4,14658	1	6	1	3
27	23	87,9475	4,17043	7	8	7	3
28	22	87,9465	4,17078	9	28	9	16
29	21	87,8047	4,21985	1	3	1	5
30	20	87,7838	4,22708	9	43	9	17
31	19	87,7203	4,24906	2	7	2	4
32	18	87,5658	4,30253	9	21	9	19
33	17	87,4763	4,33348	12	40	12	2
34	16	87,2955	4,39605	1	2	1	9
35	15	87,0230	4,49034	13	45	13	2
36	14	86,9692	4,50894	9	24	9	20
37	13	86,5878	4,64094	4	33	4	3
38	12	86,5032	4,67020	13	20	13	4
39	11	86,4151	4,70067	9	13	9	24
40	10	86,3272	4,73110	17	18	17	2
41	9	86,2402	4,76121	1	4	1	12
42	8	86,1993	4,77537	9	41	9	27
43	7	86,1454	4,79400	1	17	1	14
44	6	85,9625	4,85730	9	19	9	28
45	5	85,8067	4,91121	9	38	9	29
46	4	85,5649	4,99487	12	36	12	5
47	3	85,5008	5,01707	9	25	9	31
48	2	85,1366	5,14309	9	12	9	36
49	1	80,7867	6,64824	1	9	1	50

Final Partition

Number of clusters: 2

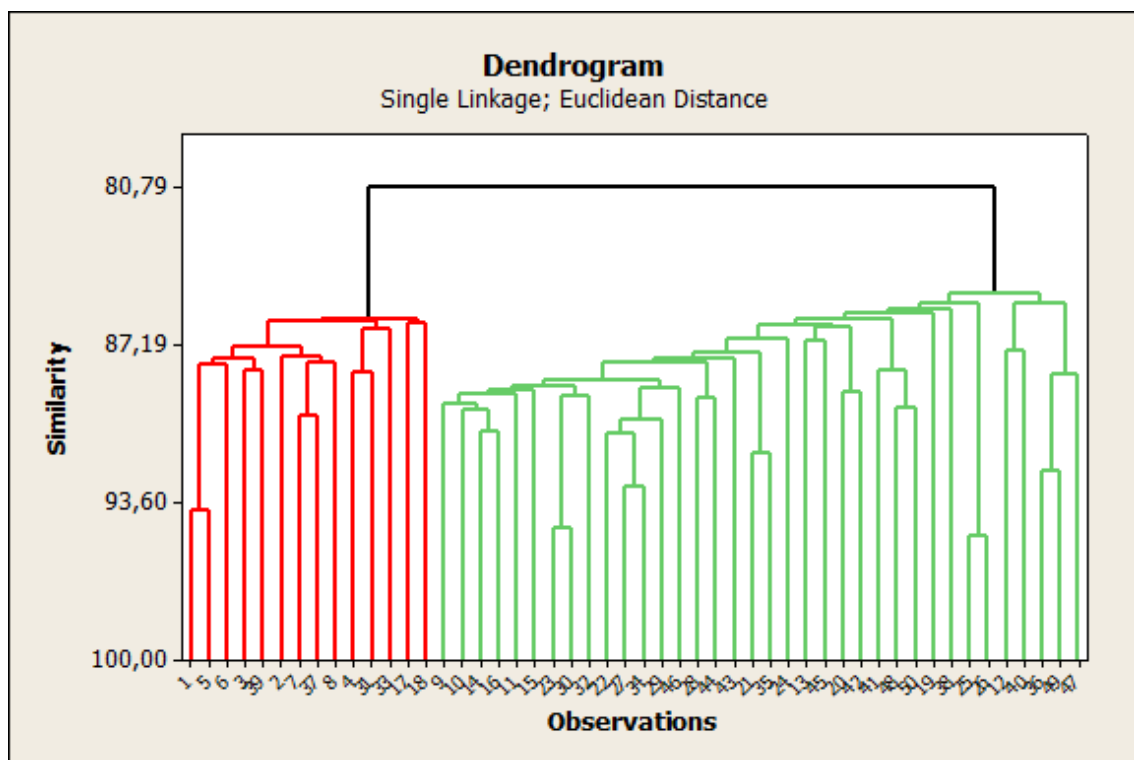
	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	14	417,33	5,20494	8,4275
Cluster2	36	2010,98	7,07014	12,2335

Cluster Centroids

Variable	Cluster1	Cluster2	Grand centroid
CresPIBpercapita_1	-3,1023	-2,0828	-2,3683
INFL_1	4,0267	2,2946	2,7796
UNEM_1	7,0646	8,5731	8,1507
GINI_1	52,9656	36,0834	40,8104
HDI_1	0,6643	0,7067	0,6948

Distances Between Cluster Centroids

	Cluster1	Cluster2
Cluster1	0,0000	17,0682
Cluster2	17,0682	0,0000



Os países são separados em dois grupos (clusters):

Grupo 1 (vermelho no gráfico e na tabela com 14 integrantes) os países de menor grau de desenvolvimento econômico e humano (principalmente países da América Latina e Central) e maior grau de desigualdade e;

Grupo 2 (verde no gráfico e na tabela com 36 integrantes) os países com maior grau de desenvolvimento econômico e humano (principalmente países da América do Norte, Europa e Ásia) e menor grau de desigualdade.

A tabela abaixo identifica cada um dos países e seu grupo correspondente.

País	Grupo		País	Grupo
Brazil	1		Israel	2
Chile	1		Italy	2
Colombia	1		Jordan	2
Costa Rica	1		Kazakhstan	2
Ecuador	1		Kyrgyz Republic	2
Peru	1		Sri Lanka	2
Paraguay	1		Luxembourg	2
El Salvador	1		Morocco	2
Guatemala	1		Moldova	2
Honduras	1		Maldives	2
Madagascar	1		Mali	2
Mexico	1		Malawi	2
Nicaragua	1		Malaysia	2
Panama	1		Norway	2
Bulgaria	2		Philippines	2
Canada	2		Poland	2
Switzerland	2		Romania	2
Camerron	2		Senegal	2
Germany	2		Slovenia	2
Spain	2		Sweden	2
Georgia	2		Syrian Arab Republic	2
Greece	2		Thailand	2
Croatia	2		Tunisia	2
Hungary	2		United States	2
Indonesia	2		Yemen, Rep.	2

Abaixo é apresentado o resultado da análise de conglomerados com a inclusão do PIB.

Com dois grupos:

Cluster Analysis of Observations: CresPIBperca; INFL_1; UNEM_1; GINI_1; ...

Euclidean Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	49	99,9787	22,3	12	50	12
2	48	99,9776	23,5	5	24	5
3	47	99,9753	25,9	15	46	15
4	46	99,9613	40,5	4	9	4
5	45	99,9540	48,2	37	43	37
6	44	99,9502	52,1	12	37	12
7	43	99,9160	88,0	1	33	1
8	42	99,9043	100,2	15	21	15
9	41	99,9030	101,6	47	48	47
10	40	99,8993	105,5	25	39	25
11	39	99,8962	108,8	7	15	7

12	38	99,8799	125,8	25	36	25	3
13	37	99,8771	128,7	31	35	31	2
14	36	99,8560	150,8	18	27	18	2
15	35	99,8555	151,3	17	29	17	2
16	34	99,8444	163,0	12	26	12	5
17	33	99,8417	165,8	18	40	18	3
18	32	99,8386	169,0	12	34	12	6
19	31	99,8325	175,4	7	18	7	7
20	30	99,8204	188,1	7	17	7	9
21	29	99,7738	236,9	7	30	7	10
22	28	99,7686	242,4	25	42	25	4
23	27	99,7589	252,5	5	6	5	3
24	26	99,7573	254,2	12	31	12	8
25	25	99,7211	292,1	5	32	5	4
26	24	99,7045	309,5	5	47	5	6
27	23	99,6503	366,3	3	5	3	7
28	22	99,6480	368,6	3	8	3	8
29	21	99,6377	379,4	7	12	7	18
30	20	99,4209	606,5	4	25	4	6
31	19	99,4149	612,8	3	7	3	26
32	18	99,3855	643,6	1	4	1	8
33	17	98,9778	1070,6	10	13	10	2
34	16	98,7971	1259,8	1	3	1	34
35	15	98,7065	1354,7	19	20	19	2
36	14	98,6498	1414,2	1	2	1	35
37	13	98,4776	1594,4	19	41	19	3
38	12	98,4446	1629,0	1	19	1	38
39	11	97,7700	2335,5	45	49	45	2
40	10	97,5837	2530,7	22	44	22	2
41	9	97,5808	2533,8	14	16	14	2
42	8	97,1511	2983,7	14	22	14	4
43	7	97,1509	2984,0	10	45	10	4
44	6	96,8398	3309,8	14	23	14	5
45	5	95,6887	4515,4	10	14	10	9
46	4	90,9263	9503,3	1	10	1	47
47	3	85,2383	15460,5	11	38	11	2
48	2	83,1578	17639,5	1	11	1	49
49	1	75,2186	25954,5	1	28	1	50

Final Partition

Number of clusters: 2

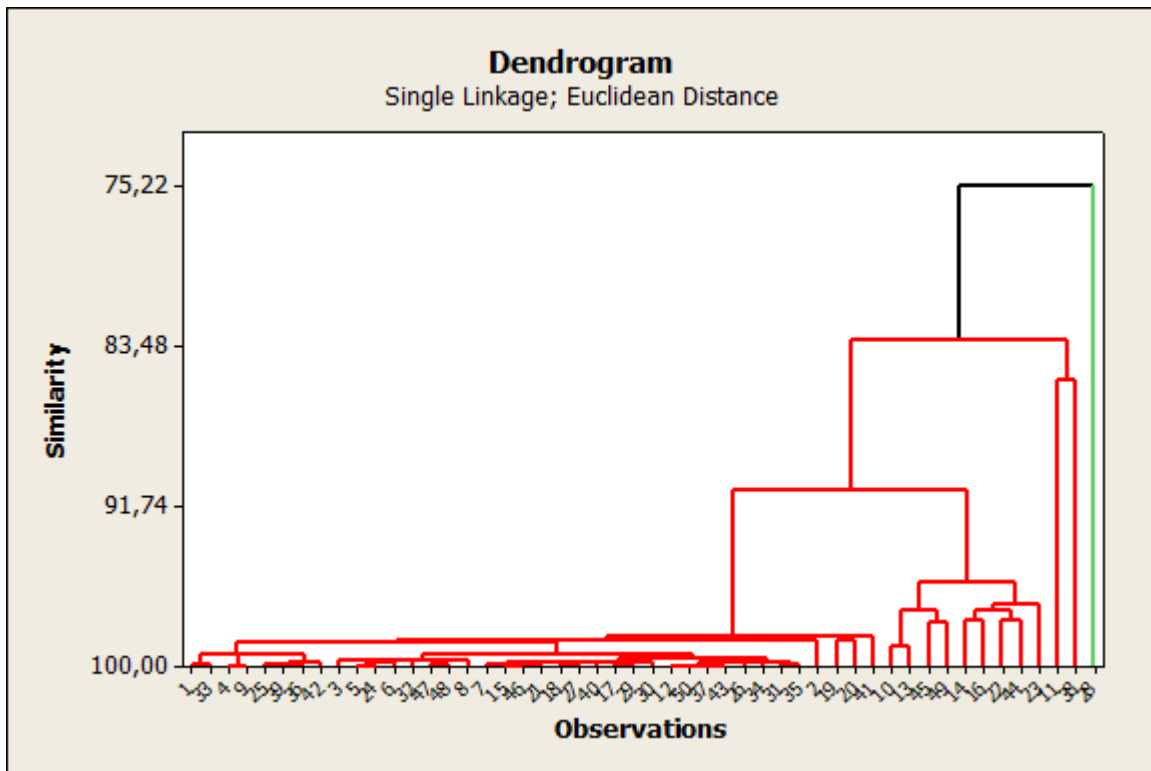
	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	49	1,50719E+10	13032,3	66280,8
Cluster2	1	0,00000E+00	0,0	0,0

Cluster Centroids

Variable	Cluster1	Cluster2	Grand centroid
CresPIBpercapita_1	-2,3	-6	-2,4
INFL_1	2,8	0	2,8
UNEM_1	8,2	4	8,2
GINI_1	41,0	31	40,8
HDI_1	0,7	1	0,7
PIB_1	12808,3	105044	14653,0

Distances Between Cluster Centroids

	Cluster1	Cluster2
Cluster1	0,0	92235,3
Cluster2	92235,3	0,0



Com cinco grupos:

Cluster Analysis of Observations: CresPIBperca; INFL_1; UNEM_1; GINI_1; ...

Euclidean Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	49	99,9787	22,3	12 50	12	2
2	48	99,9776	23,5	5 24	5	2
3	47	99,9753	25,9	15 46	15	2
4	46	99,9613	40,5	4 9	4	2
5	45	99,9540	48,2	37 43	37	2
6	44	99,9502	52,1	12 37	12	4
7	43	99,9160	88,0	1 33	1	2
8	42	99,9043	100,2	15 21	15	3
9	41	99,9030	101,6	47 48	47	2
10	40	99,8993	105,5	25 39	25	2
11	39	99,8962	108,8	7 15	7	4
12	38	99,8799	125,8	25 36	25	3
13	37	99,8771	128,7	31 35	31	2
14	36	99,8560	150,8	18 27	18	2
15	35	99,8555	151,3	17 29	17	2
16	34	99,8444	163,0	12 26	12	5
17	33	99,8417	165,8	18 40	18	3
18	32	99,8386	169,0	12 34	12	6
19	31	99,8325	175,4	7 18	7	7
20	30	99,8204	188,1	7 17	7	9
21	29	99,7738	236,9	7 30	7	10
22	28	99,7686	242,4	25 42	25	4
23	27	99,7589	252,5	5 6	5	3

24	26	99,7573	254,2	12	31	12	8
25	25	99,7211	292,1	5	32	5	4
26	24	99,7045	309,5	5	47	5	6
27	23	99,6503	366,3	3	5	3	7
28	22	99,6480	368,6	3	8	3	8
29	21	99,6377	379,4	7	12	7	18
30	20	99,4209	606,5	4	25	4	6
31	19	99,4149	612,8	3	7	3	26
32	18	99,3855	643,6	1	4	1	8
33	17	98,9778	1070,6	10	13	10	2
34	16	98,7971	1259,8	1	3	1	34
35	15	98,7065	1354,7	19	20	19	2
36	14	98,6498	1414,2	1	2	1	35
37	13	98,4776	1594,4	19	41	19	3
38	12	98,4446	1629,0	1	19	1	38
39	11	97,7700	2335,5	45	49	45	2
40	10	97,5837	2530,7	22	44	22	2
41	9	97,5808	2533,8	14	16	14	2
42	8	97,1511	2983,7	14	22	14	4
43	7	97,1509	2984,0	10	45	10	4
44	6	96,8398	3309,8	14	23	14	5
45	5	95,6887	4515,4	10	14	10	9
46	4	90,9263	9503,3	1	10	1	47
47	3	85,2383	15460,5	11	38	11	2
48	2	83,1578	17639,5	1	11	1	49
49	1	75,2186	25954,5	1	28	1	50

Final Partition

Number of clusters: 5

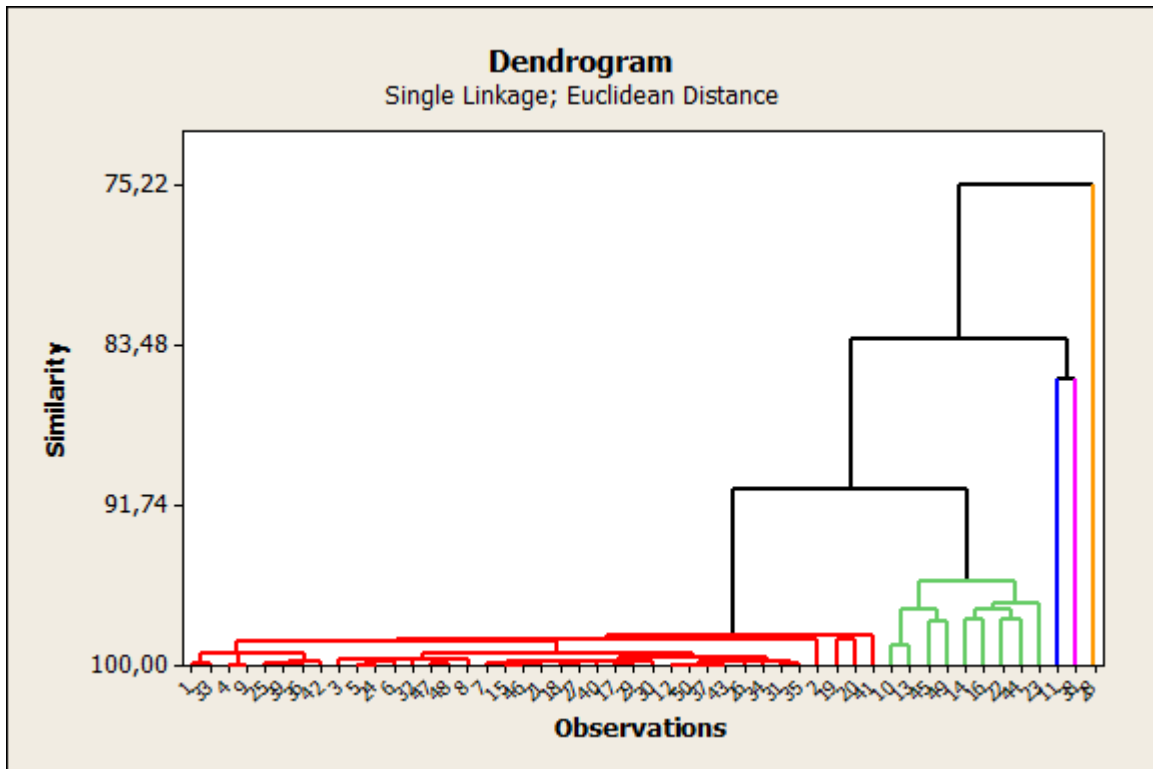
	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	38	465655105	2832,48	9777,7
Cluster2	9	495990973	6549,15	11384,5
Cluster3	1	0	0,00	0,0
Cluster4	1	0	0,00	0,0
Cluster5	1	0	0,00	0,0

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Grand centroid
CresPIBpercapita_1	-1,76	-4,4	-3,0	-6	-2,8	-2,4
INFL_1	3,45	0,6	-0,5	0	2,2	2,8
UNEM_1	8,55	8,0	3,6	4	3,6	8,2
GINI_1	43,40	33,4	33,7	31	25,8	40,8
HDI_1	0,64	0,9	0,9	1	0,9	0,7
PIB_1	4444,70	35110,1	63628,7	105044	79089,1	14653,0

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	0	30665,4	59184,0	100599	74644,4
Cluster2	30665	0,0	28518,5	69934	43979,0
Cluster3	59184	28518,5	0,0	41415	15460,5
Cluster4	100599	69933,5	41415,0	0	25954,5
Cluster5	74644	43979,0	15460,5	25955	0,0



Percebe-se que com a inclusão do PIB, a classificação em dois grupos passa a não ser adequada. Nem mesmo em 5 grupos é possível fazer um agrupamento consistente dos países, uma vez que as distâncias entre os centróides das variáveis no grupos são muito pequenas. Portanto, na sequência do trabalho, utiliza-se o agrupamento resultado da análise de conglomerados com as variáveis Crescimento PIB per capita, Inflação, Desemprego, GINI, HDI, sem a inclusão do PIB.

3.4. COMPARAÇÕES ANOVA

Para identificar as variáveis que mais caracterizam a formação dos grupos utilizam-se as comparações ANOVA de cada variável que são apresentadas a seguir.

One-way ANOVA: CresPIBpercapita_1 versus Cluster_1

Source	DF	SS	MS	F	P
Cluster_1	1	10,5	10,5	0,95	0,334
Error	48	528,0	11,0		
Total	49	538,5			

S = 3,317 R-Sq = 1,95% R-Sq(adj) = 0,00%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
1	14	-3,102	2,624
2	36	-2,083	3,540

-----+-----+-----+-----+-----+
 (-----*-----)
 (-----*-----)
 -----+-----+-----+-----+-----+

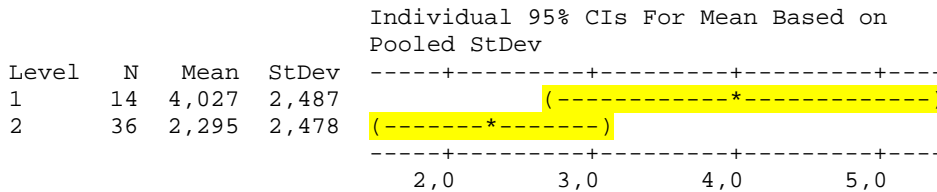
-4,0 -3,0 -2,0 -1,0

Pooled StDev = 3,317

One-way ANOVA: INFL_1 versus Cluster_1

Source	DF	SS	MS	F	P
Cluster_1	1	30,24	30,24	4,92	0,031
Error	48	295,33	6,15		
Total	49	325,57			

S = 2,480 R-Sq = 9,29% R-Sq(adj) = 7,40%

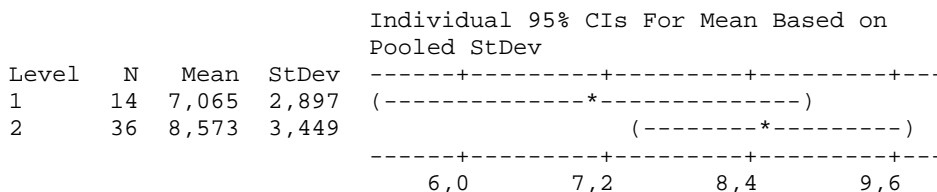


Pooled StDev = 2,480

One-way ANOVA: UNEM_1 versus Cluster_1

Source	DF	SS	MS	F	P
Cluster_1	1	22,9	22,9	2,10	0,154
Error	48	525,5	10,9		
Total	49	548,4			

S = 3,309 R-Sq = 4,18% R-Sq(adj) = 2,19%



Pooled StDev = 3,309

Como no caso desta variável o valor de P ficou perto do limite indicado para segmentação de grupos (0,100), utiliza-se o teste T para confirmar se de fato existe a relação configurada de maior e menor entre os cluster.

Two-Sample T-Test and CI: UNEM_1; Cluster_1

Two-sample T for UNEM_1

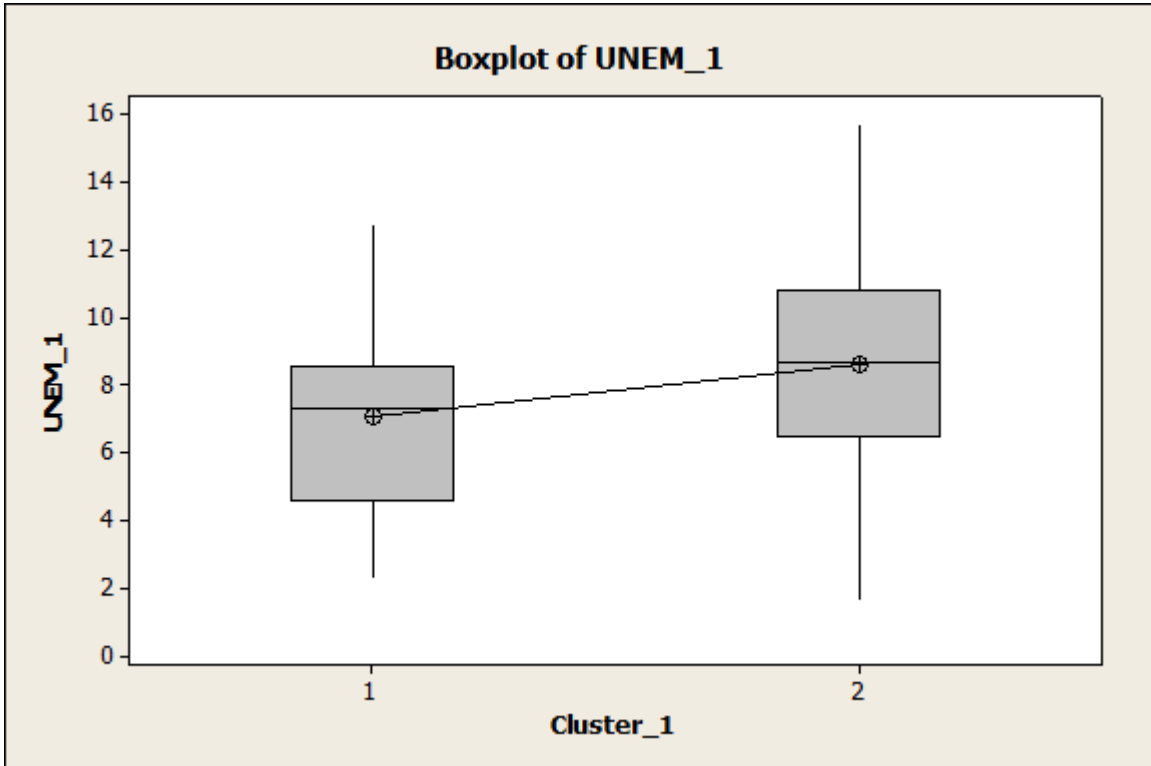
Cluster_1	N	Mean	StDev	SE Mean
1	14	7,06	2,90	0,77
2	36	8,57	3,45	0,57

Difference = mu (1) - mu (2)

Estimate for difference: -1,508

95% upper bound for difference: 0,132

T-Test of difference = 0 (vs <): T-Value = -1,56 P-Value = 0,064 DF = 28

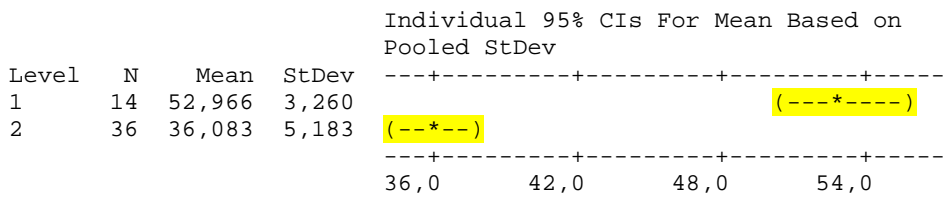


Percebe-se que no teste T a condição de que o Cluster 1 é menor que o 2 se configura e pode ser percebido pelo P-value = 0,064 (menos da metade do valor de P na análise ANOVA padrão).

One-way ANOVA: GINI_1 versus Cluster_1

Source	DF	SS	MS	F	P
Cluster_1	1	2872,9	2872,9	127,87	0,000
Error	48	1078,4	22,5		
Total	49	3951,3			

S = 4,740 R-Sq = 72,71% R-Sq(adj) = 72,14%



Pooled StDev = 4,740

One-way ANOVA: HDI_1 versus Cluster_1

Source	DF	SS	MS	F	P
Cluster_1	1	0,0181	0,0181	0,82	0,369
Error	48	1,0571	0,0220		
Total	49	1,0752			

S = 0,1484 R-Sq = 1,68% R-Sq(adj) = 0,00%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
1	14	0,6643	0,0942
2	36	0,7067	0,1640

-----+-----+-----+-----+-----
 (-----*-----)
 (-----*-----)
 -----+-----+-----+-----+-----
 0,600 0,650 0,700 0,750

Pooled StDev = 0,1484

Observando as características das análises ANOVA, percebe-se que as variáveis com maior possibilidade de representar o agrupamento dos países são Inflação e GINI (ambos marcados em amarelo nas representações acima), pois ambos apresentaram indicador P muito próximo ou igual a 0.

No caso da variável Taxa de Desemprego (UNEMP), quando complementado o resultado da análise ANOVA com o teste T, também conclui-se que esta variável segrega os indivíduos dos grupos, sendo incluída na sequência do trabalho.

Confirmando as primeiras percepções, estes dois indicadores, um econômico e outro de desigualdade, representam grande peso no agrupamento dos países. Fato este que será confirmado ou não nas próximas etapas do trabalho.

3.4. ANÁLISE DISCRIMINANTE

Para confirmar ou não a representatividade das variáveis Inflação, Desemprego e GINI em relação aos países em análise, será utilizado o técnica de Análise Discriminante com os métodos Linear e Quadrático.

Para permitir uma comparação, utilizaremos a análise em três etapas: primeiro as três variáveis juntas e depois somente Inflação e GINI e depois as variáveis independentemente.

Todas as variáveis (INFL, UNEMP e GINI)

Discriminant Analysis: Cluster_1 versus GINI_1; INFL_1; UNEM_1

Linear Method for Response: Cluster_1

Predictors: GINI_1; INFL_1; UNEM_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	14	2
2	0	34
Total N	14	36
N correct	14	34
Proportion	1,000	0,944

N = 50

N Correct = 48

Proportion Correct = **0,960**

Squared Distance Between Groups

	1	2
1	0,0000	14,9664
2	14,9664	0,0000

Linear Discriminant Function for Groups

	1	2
Constant	-65,547	-30,889
GINI_1	2,395	1,586
INFL_1	1,005	0,559
UNEM_1	0,025	0,380

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
Camarões 12**	2	1	1	3,485	0,714
			2	5,311	0,286
Tailandia47**	2	1	1	8,439	0,738
			2	10,511	0,262

Discriminant Analysis: Cluster_1 versus GINI_1; INFL_1; UNEM_1

Quadratic Method for Response: Cluster_1

Predictors: GINI_1; INFL_1; UNEM_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	14	0
2	0	36
Total N	14	36
N correct	14	36
Proportion	1,000	1,000

N = 50 N Correct = 50 Proportion Correct = 1,000

From Generalized Squared Distance to Group

Group	1	2
1	5,857	19,506
2	52,784	7,531

Somente Inflação e GINI (INFL e GINI)**Discriminant Analysis: Cluster_1 versus INFL_1; GINI_1**

Linear Method for Response: Cluster_1

Predictors: INFL_1; GINI_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	14	1
2	0	35
Total N	14	36
N correct	14	35
Proportion	1,000	0,972

N = 50 N Correct = 49 Proportion Correct = 0,980

Squared Distance Between Groups

	1	2
1	0,0000	13,6390
2	13,6390	0,0000

Linear Discriminant Function for Groups

	1	2
Constant	-65,543	-30,128
INFL_1	1,009	0,614
GINI_1	2,398	1,631

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
Filipinas40**	2	1	1	3,125	0,572
			2	3,707	0,428

Discriminant Analysis: Cluster_1 versus INFL_1; GINI_1

Quadratic Method for Response: Cluster_1

Predictors: INFL_1; GINI_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	14	0
2	0	36
Total N	14	36
N correct	14	36
Proportion	1,000	1,000

N = 50 N Correct = 50 Proportion Correct = 1,000

From Generalized Squared Distance to Group

Group	1	2
1	4,043	16,257
2	38,568	5,106

Desemprego (UNEMP)

Discriminant Analysis: Cluster_1 versus UNEM_1

Linear Method for Response: Cluster_1

Predictors: UNEM_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	9	15
2	5	21
Total N	14	36
N correct	9	21
Proportion	0,643	0,583

N = 50 N Correct = 30 Proportion Correct = 0,600

Squared Distance Between Groups

	1	2
1	0,000000	0,207865
2	0,207865	0,000000

Linear Discriminant Function for Groups

	1	2
Constant	-2,2796	-3,3570
UNEM_1	0,6454	0,7831

Summary of Misclassified Observations

Observation	True Group	Pred Group	Squared Distance	Probability
1**	1	2	1 0,241691	0,470
			2 0,001274	0,530
2**	1	2	1 0,184425	0,477
			2 0,000701	0,523
3**	1	2	1 2,964	0,336
			2 1,602	0,664
5**	1	2	1 0,164228	0,480
			2 0,002568	0,520
10**	2	1	1 0,000467	0,528
			2 0,228047	0,472
11**	2	1	1 1,097	0,641
			2 2,260	0,359
12**	2	1	1 0,3246	0,590
			2 1,0519	0,410
20**	2	1	1 0,000724	0,529
			2 0,233128	0,471
27**	2	1	1 0,01012	0,515
			2 0,12627	0,485
28**	2	1	1 0,9133	0,632
			2 1,9927	0,368
30**	2	1	1 0,002911	0,532
			2 0,259978	0,468
35**	2	1	1 0,04940	0,501

			2	0,05460	0,499
36**	2	1	1	1,219	0,647
			2	2,434	0,353
38**	2	1	1	1,104	0,642
			2	2,269	0,358
39**	1	2	1	1,5551	0,386
			2	0,6259	0,614
42**	2	1	1	0,000612	0,529
			2	0,231029	0,471
44**	2	1	1	0,1107	0,564
			2	0,6220	0,436
45**	2	1	1	0,04651	0,550
			2	0,45103	0,450
47**	2	1	1	2,735	0,702
			2	4,450	0,298
49**	2	1	1	0,2125	0,578
			2	0,8408	0,422

Discriminant Analysis: Cluster_1 versus UNEM_1

Quadratic Method for Response: Cluster_1

Predictors: UNEM_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	12	19
2	2	17
Total N	14	36
N correct	12	17
Proportion	0,857	0,472

N = 50 N Correct = 29 Proportion Correct = **0,580**

From Generalized Squared Distance to Group

Group	1	2
1	2,127	2,667
2	2,398	2,476

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
3**	1	2	1	5,993	0,265
			2	3,950	0,735
10**	2	1	1	2,128	0,569
			2	2,686	0,431
11**	2	1	1	3,558	0,622
			2	4,556	0,378
12**	2	1	1	2,551	0,610
			2	3,444	0,390
20**	2	1	1	2,128	0,570
			2	2,691	0,430
23**	2	1	1	2,269	0,528
			2	2,491	0,472
25**	2	1	1	2,412	0,508

			2	2,476	0,492
27**	2	1	1	2,141	0,556
			2	2,592	0,444
28**	2	1	1	3,319	0,621
			2	4,310	0,379
30**	2	1	1	2,131	0,573
			2	2,715	0,427
32**	2	1	1	2,282	0,526
			2	2,488	0,474
34**	2	1	1	2,468	0,501
			2	2,479	0,499
35**	2	1	1	2,192	0,542
			2	2,526	0,458
36**	2	1	1	3,717	0,622
			2	4,716	0,378
38**	2	1	1	3,567	0,622
			2	4,565	0,378
39**	1	2	1	4,156	0,365
			2	3,052	0,635
42**	2	1	1	2,128	0,570
			2	2,689	0,430
44**	2	1	1	2,272	0,596
			2	3,049	0,404
45**	2	1	1	2,188	0,587
			2	2,891	0,413
47**	2	1	1	5,694	0,608
			2	6,572	0,392
49**	2	1	1	2,405	0,604
			2	3,250	0,396

Inflação (INFL)

Discriminant Analysis: Cluster_1 versus INFL_1

Linear Method for Response: Cluster_1

Predictors: INFL_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	8	13
2	6	23
Total N	14	36
N correct	8	23
Proportion	0,571	0,639

N = 50 N Correct = 31 Proportion Correct = 0,620

Squared Distance Between Groups

	1	2
1	0,000000	0,487644
2	0,487644	0,000000

Linear Discriminant Function for Groups

1 2

Constant -1,3177 -0,4279
 INFL_1 0,6545 0,3729

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
2**	1	2	1	2,6353	0,291
			2	0,8557	0,709
6**	1	2	1	0,19327	0,484
			2	0,06692	0,516
7**	1	2	1	0,33458	0,460
			2	0,01437	0,540
8**	1	2	1	1,4344	0,356
			2	0,2493	0,644
17**	1	2	1	0,76598	0,409
			2	0,03129	0,591
20**	2	1	1	0,005407	0,573
			2	0,595750	0,427
21**	2	1	1	0,9010	0,712
			2	2,7144	0,288
22**	2	1	1	0,08003	0,512
			2	0,17257	0,488
25**	2	1	1	1,748	0,763
			2	4,082	0,237
26**	2	1	1	1,307	0,739
			2	3,391	0,261
27**	2	1	1	0,04366	0,524
			2	0,23947	0,476
32**	2	1	1	0,000306	0,558
			2	0,463514	0,442
35**	2	1	1	3,140	0,815
			2	6,102	0,185
39**	1	2	1	0,425312	0,447
			2	0,002130	0,553
40**	2	1	1	0,1043	0,505
			2	0,1409	0,495
41**	2	1	1	0,006549	0,547
			2	0,381170	0,453
42**	2	1	1	0,3960	0,664
			2	1,7626	0,336
48**	2	1	1	0,006008	0,547
			2	0,385396	0,453
50**	2	1	1	0,02035	0,536
			2	0,30876	0,464

Discriminant Analysis: Cluster_1 versus INFL_1

Quadratic Method for Response: Cluster_1

Predictors: INFL_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	8	13
2	6	23
Total N	14	36
N correct	8	23

Proportion 0,571 0,639

N = 50

N Correct = 31

Proportion Correct = 0,620

From Generalized Squared Distance to Group

Group	1	2
1	1,822	2,303
2	2,307	1,815

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
2**	1	2	1	4,443	0,292
			2	2,672	0,708
6**	1	2	1	2,015	0,483
			2	1,882	0,517
7**	1	2	1	2,155	0,459
			2	1,829	0,541
8**	1	2	1	3,249	0,356
			2	2,065	0,644
17**	1	2	1	2,584	0,409
			2	1,846	0,591
20**	2	1	1	1,828	0,573
			2	2,412	0,427
21**	2	1	1	2,718	0,713
			2	4,535	0,287
22**	2	1	1	1,902	0,511
			2	1,988	0,489
25**	2	1	1	3,561	0,764
			2	5,906	0,236
26**	2	1	1	3,122	0,740
			2	5,213	0,260
27**	2	1	1	1,866	0,524
			2	2,055	0,476
32**	2	1	1	1,823	0,557
			2	2,279	0,443
35**	2	1	1	4,945	0,816
			2	7,930	0,184
39**	1	2	1	2,245	0,447
			2	1,817	0,553
40**	2	1	1	1,926	0,504
			2	1,956	0,496
41**	2	1	1	1,829	0,546
			2	2,197	0,454
42**	2	1	1	2,216	0,664
			2	3,581	0,336
48**	2	1	1	1,828	0,546
			2	2,201	0,454
50**	2	1	1	1,843	0,535
			2	2,124	0,465

GINI

Discriminant Analysis: Cluster_1 versus GINI_1

Linear Method for Response: Cluster_1

Predictors: GINI_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	14	3
2	0	33
Total N	14	36
N correct	14	33
Proportion	1,000	0,917

N = 50 N Correct = 47 Proportion Correct = 0,940

Squared Distance Between Groups

	1	2
1	0,0000	12,6856
2	12,6856	0,0000

Linear Discriminant Function for Groups

	1	2
Constant	-62,433	-28,976
GINI_1	2,357	1,606

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
Camarões 12**	2	1	1	3,145	0,507
			2	3,198	0,493
Filipinas40**	2	1	1	2,917	0,565
			2	3,436	0,435
Tailandia47**	2	1	1	2,627	0,639
			2	3,767	0,361

Discriminant Analysis: Cluster_1 versus GINI_1

Quadratic Method for Response: Cluster_1

Predictors: GINI_1

Group	1	2
Count	14	36

Summary of classification

Put into Group	True Group	
	1	2
1	14	0
2	0	36
Total N	14	36
N correct	14	36
Proportion	1,000	1,000

N = 50 N Correct = 50 Proportion Correct = 1,000

From Generalized Squared Distance to Group

Group	1	2
1	2,363	13,900
2	29,181	3,291

A seguir apresenta-se a tabela com o resumo com os resultados das análises discriminantes.

Variáveis	Método	Proporção de correção
Todas	Linear	0,960
	Quadrática	1,000
INFL e GINI	Linear	0,980
	Quadrática	1,000
UNEMP	Linear	0,600
	Quadrática	0,580
INFL	Linear	0,620
	Quadrática	0,620
GINI	Linear	0,940
	Quadrática	1,000

Pela técnica de Análise Discriminante método Linear com as variáveis Inflação e GINI, encontrou-se uma proporção de correção de 0,980, ou seja, somente para um dos países (Filipinas) o modelo não conseguiu prever o agrupamento corretamente.

Para as mesmas variáveis utilizando-se do método Quadrático, a proporção de correção foi 1,000, ou seja, 100% de acerto.

Entretanto, quando colocado ao lado dos resultados da análise discriminante que utilizou apenas a variável GINI, encontra-se uma diferença mínima com a máxima simplificação do modelo, ou seja, apenas com uma variável consegue-se determinar os agrupamentos com 94% de certeza.

3.4. REGRESSÃO LOGÍSTICA

Para permitir a comparação entre as duas técnicas e também direcionar a conclusão, utiliza-se a técnica de Regressão Logística com as mesmas variáveis da Análise Discriminante.

Como do agrupamento da Análise de Conglomerados resultou dois grupos, utiliza-se o método de Regressão Logística Binária.

Binary Logistic Regression: Cluster_1 versus INFL_1; GINI_1

```
* WARNING * Algorithm has not converged after 20 iterations.
* WARNING * Convergence has not been reached for the parameter estimates
            criterion.
* WARNING * The results may not be reliable.
* WARNING * Try increasing the maximum number of iterations.
```

Link Function: Logit

Response Information

Variable	Value	Count	
Cluster_1	2	36	(Event)
	1	14	
	Total	50	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	350,790	25298,3	0,01	0,989			
INFL_1	-2,58846	512,725	-0,01	0,996	0,08	0,00	*
GINI_1	-7,28019	527,867	-0,01	0,989	0,00	0,00	*

Log-Likelihood = -0,000

Test that all slopes are zero: G = 59,295, DF = 2, P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0,0000003	47	1,000
Deviance	0,0000005	47	1,000
Hosmer-Lemeshow	0,0000000	8	1,000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
2												
Obs	0	0	1	5	5	5	5	5	5	5	36	
Exp	0,0	0,0	1,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0		
1												
Obs	5	5	4	0	0	0	0	0	0	0	14	
Exp	5,0	5,0	4,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0		
Total	5	5	5	5	5	5	5	5	5	5	50	

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	504	100,0	Somers' D 1,00
Discordant	0	0,0	Goodman-Kruskal Gamma 1,00
Ties	0	0,0	Kendall's Tau-a 0,41
Total	504	100,0	

Binary Logistic Regression: Cluster_1 versus INFL_1; GINI_1

- * WARNING * Algorithm has not converged after 50 iterations.
- * WARNING * Convergence has not been reached for the parameter estimates criterion.
- * WARNING * The results may not be reliable.
- * WARNING * Try increasing the maximum number of iterations.

Link Function: Logit

Response Information

Variable	Value	Count
Cluster_1	2	36 (Event)
	1	14
Total		50

Logistic Regression Table

Odds 95% CI

Predictor	Coef	SE Coef	Z	P	Ratio	Lower	Upper
Constant	513,160	237088	0,00	0,998			
INFL_1	-3,73308	25567,8	-0,00	1,000	0,02	0,00	*
GINI_1	-10,6451	5224,87	-0,00	0,998	0,00	0,00	*

Log-Likelihood = -0,000

Test that all slopes are zero: G = 59,295, DF = 2, P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0,0000000	47	1,000
Deviance	0,0000000	47	1,000
Hosmer-Lemeshow	0,0000000	8	1,000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
2												
Obs	0	0	1	5	5	5	5	5	5	5	36	
Exp	0,0	0,0	1,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0		
1												
Obs	5	5	4	0	0	0	0	0	0	0	14	
Exp	5,0	5,0	4,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0		
Total	5	5	5	5	5	5	5	5	5	5	50	

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	504	100,0	Somers' D 1,00
Discordant	0	0,0	Goodman-Kruskal Gamma 1,00
Ties	0	0,0	Kendall's Tau-a 0,41
Total	504	100,0	

Logo no início foram apresentadas mensagens de aviso relativamente ao número de iterações. Na primeira execução foram utilizadas 20 iterações e na segunda 50 iterações, porém não houve alteração e a mensagem continuava a ser apresentada.

Com relação aos resultados das análises, ambas apresentaram percentual de concordância de 100%. Apesar deste percentual, existe a indicação de que os resultados podem não ser confiáveis em razão da não convergência nos critérios de estimativa de parâmetros (mensagem no início das informações disponibilizadas pelo sistema).

Esta mensagem é apresentada em todas as tentativas de combinação e número de iterações que foram utilizadas.

Da mesma forma como foi realizado na Análise Discriminante, a seguir apresenta-se a Regressão Logística apenas para a variável GINI (que no caso da Análise Discriminante demonstrou-se capaz de explicar o agrupamento com 94% de certeza)

Binary Logistic Regression: Cluster_1 versus GINI_1

* WARNING * Algorithm has not converged after 50 iterations.

* WARNING * Convergence has not been reached for the parameter estimates criterion.

* WARNING * The results may not be reliable.
 * WARNING * Try increasing the maximum number of iterations.

Link Function: Logit

Response Information

Variable	Value	Count	
Cluster_1	2	36	(Event)
	1	14	
	Total	50	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	954,494	223892	0,00	0,997			
GINI_1	-20,6067	4823,52	-0,00	0,997	0,00	0,00	*

Log-Likelihood = -0,000

Test that all slopes are zero: G = 59,295, DF = 1, P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0,0000000	47	1,000
Deviance	0,0000000	47	1,000
Hosmer-Lemeshow	0,0000000	8	1,000

Table of Observed and Expected Frequencies:

(See Hosmer-Lemeshow Test for the Pearson Chi-Square Statistic)

Value	Group										Total	
	1	2	3	4	5	6	7	8	9	10		
2												
Obs	0	0	1	5	5	5	5	5	5	5		36
Exp	0,0	0,0	1,0	5,0	5,0	5,0	5,0	5,0	5,0	5,0		
1												
Obs	5	5	4	0	0	0	0	0	0	0		14
Exp	5,0	5,0	4,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0		
Total	5	5	5	5	5	5	5	5	5	5		50

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	504	100,0	Somers' D 1,00
Discordant	0	0,0	Goodman-Kruskal Gamma 1,00
Ties	0	0,0	Kendall's Tau-a 0,41
Total	504	100,0	

Novamente encontra-se resultado igual ao resultado apresentado pela Regressão com a combinação das variáveis. A variável GINI sozinha consegue explicar com 100% de certeza o agrupamento. Também na análise da variável GINI sozinha o sistema apresentou a mensagem de possibilidade de que os resultados podem não ser confiáveis em razão da não convergência nos critérios de estimativa de parâmetros

3.5. ANÁLISE DE CORRESPONDÊNCIA

O método de Análise de Correspondência permite identificar a intensidade das relações entre os países e as variáveis. Para esta análise foram selecionados do universo de países do trabalho os 8 países da América Latina (grupo 1) e mais 8 países do grupo 2 (seleção aleatória), apresentados na tabela abaixo, com seus respectivos códigos. Também são apresentados na tabela abaixo os códigos das variáveis utilizados nos gráficos resultantes das análises.

Não foi possível a utilização da variável de Crescimento do PIB per capita, pois esta apresenta valores negativos não suportados pelas funções utilizadas neste modelo.

País	Código	Variável	Código
Brasil	BRA	Inflação	INFL
Chile	CHL	Desemprego	UNEMP
Colômbia	COL	GINI	GINI
Costa Rica	CRI	HDI	HDI
Equador	ECU		
Peru	PER		
Paraguai	PRY		
El Salvador	SLV		
Malasia	MYS		
Estados Unidos	USA		
Hungria	HUN		
Sri Lanka	LKA		
Luxemburgo	LUX		
Croacia	HRV		
Tunísia	TUN		
Itália	ITA		

O primeiro gráfico a seguir representa a análise de todas as variáveis com todos os países. A segunda representa somente as variáveis Inflação, GINI e HDI. Na tentativa de se realizar a análise apenas com Inflação e GINI, novamente por restrições das suas funções, o modelo apresentou erro, não sendo possível gerar o gráfico.

Simple Correspondence Analysis: INFL_1_1; UNEM_1_1; GINI_1_1; HDI_1_1

Analysis of Contingency Table

Axis	Inertia	Proportion	Cumulative	Histogram
1	0,0283	0,5935	0,5935	*****
2	0,0183	0,3848	0,9784	*****
3	0,0010	0,0216	1,0000	*
Total	0,0476			

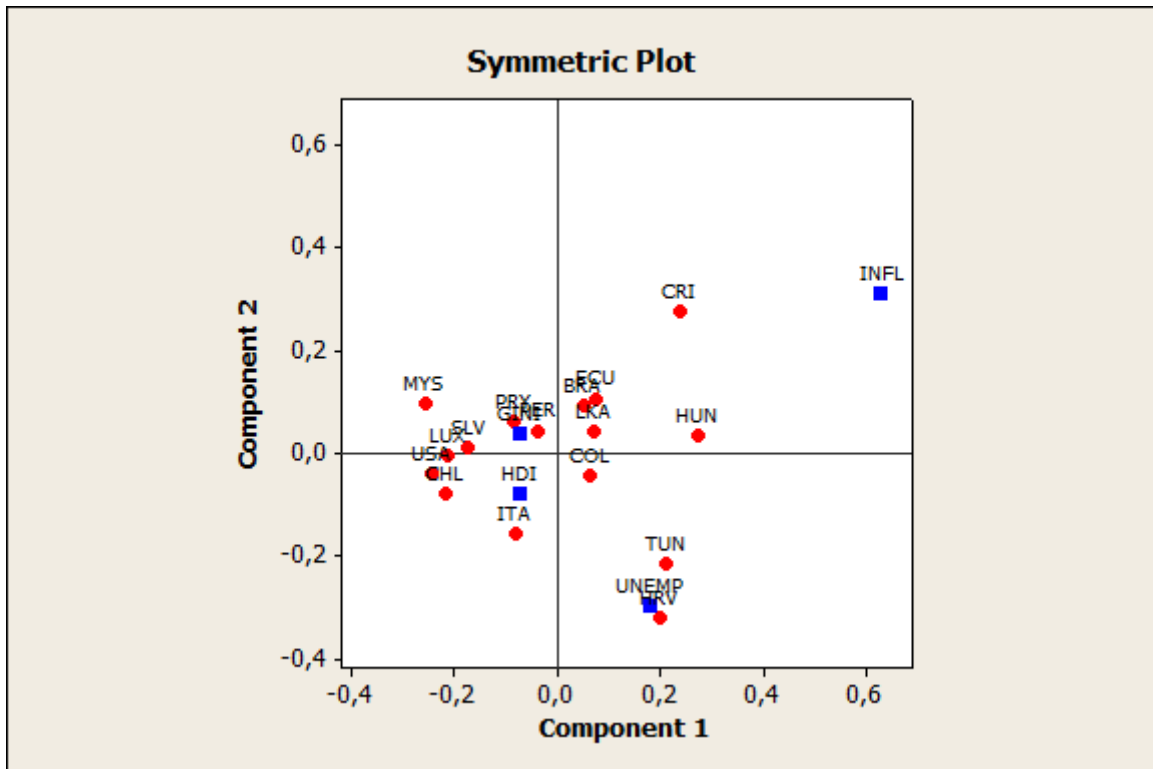
Row Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	BRA	0,966	0,078	0,019	0,053	0,242	0,008	0,092	0,724	0,036
2	CHL	0,991	0,070	0,078	-0,216	0,872	0,115	-0,080	0,119	0,024
3	COL	0,831	0,084	0,013	0,064	0,547	0,012	-0,046	0,283	0,010
4	CRI	0,998	0,070	0,195	0,239	0,430	0,141	0,274	0,568	0,288
5	ECU	0,986	0,076	0,028	0,077	0,340	0,016	0,105	0,646	0,046

6	PER	0,945	0,069	0,005	-0,037	0,410	0,003	0,043	0,535	0,007
7	PRY	0,930	0,072	0,018	-0,084	0,609	0,018	0,061	0,320	0,015
8	SLV	0,981	0,065	0,041	-0,172	0,978	0,068	0,009	0,003	0,000
9	MYS	0,996	0,052	0,082	-0,257	0,877	0,121	0,095	0,119	0,025
10	USA	0,978	0,052	0,068	-0,243	0,949	0,109	-0,042	0,028	0,005
11	HUN	0,929	0,045	0,077	0,274	0,917	0,119	0,032	0,013	0,003
12	LKA	1,000	0,058	0,009	0,074	0,779	0,011	0,039	0,220	0,005
13	LUX	0,870	0,040	0,043	-0,213	0,870	0,064	-0,004	0,000	0,000
14	HRV	0,998	0,052	0,160	0,201	0,278	0,075	-0,323	0,720	0,299
15	TUN	0,996	0,066	0,129	0,213	0,492	0,107	-0,216	0,504	0,169
16	ITA	0,966	0,051	0,034	-0,081	0,206	0,012	-0,156	0,760	0,068

Column Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	INFL	1,000	0,049	0,510	0,632	0,807	0,693	0,309	0,193	0,256
2	UNEMP	1,000	0,143	0,362	0,181	0,272	0,166	-0,297	0,728	0,685
3	GINI	0,998	0,795	0,104	-0,070	0,795	0,139	0,035	0,202	0,055
4	HDI	0,129	0,013	0,024	-0,071	0,057	0,002	-0,079	0,072	0,005



Simple Correspondence Analysis: INFL_1_1; GINI_1_1; HDI_1_1

Analysis of Contingency Table

Axis	Inertia	Proportion	Cumulative	Histogram
1	0,0300	0,9559	0,9559	*****
2	0,0014	0,0441	1,0000	*
Total	0,0314			

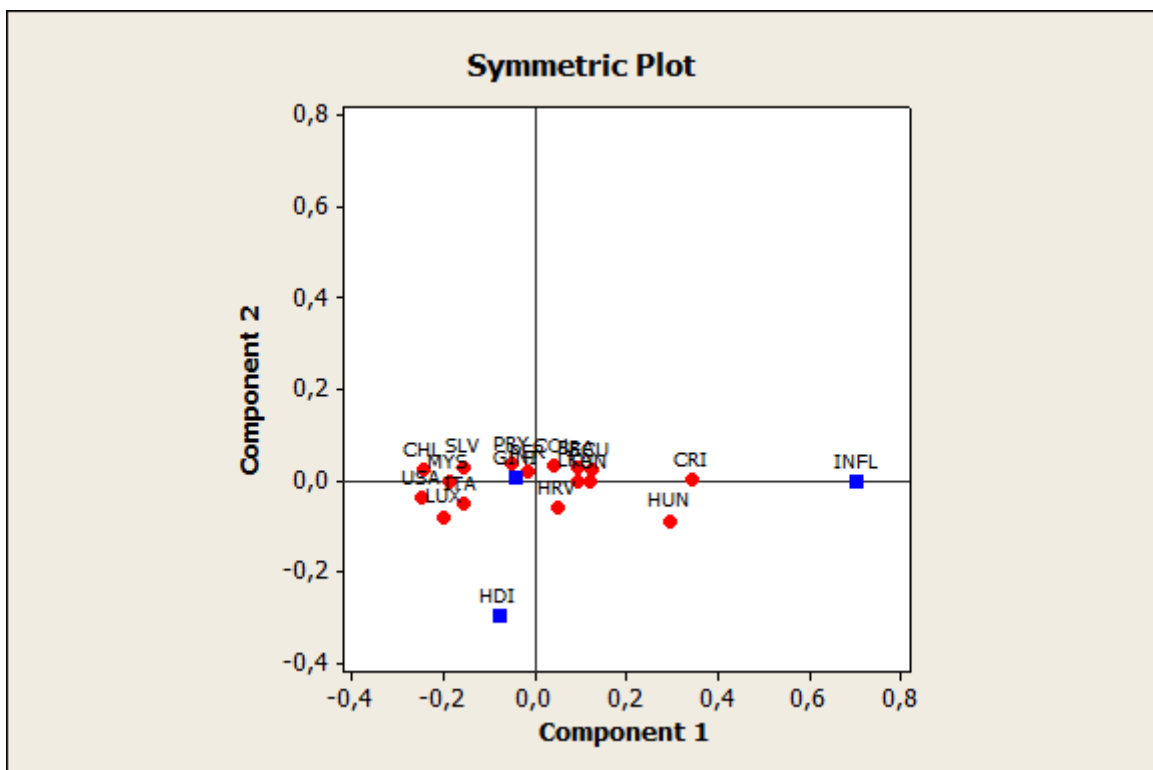
Row Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	BRA	1,000	0,080	0,024	0,093	0,917	0,023	0,028	0,083	0,045
2	CHL	1,000	0,070	0,136	-0,246	0,992	0,141	0,022	0,008	0,025

3	COL	1,000	0,082	0,007	0,041	0,598	0,005	0,034	0,402	0,066
4	CRI	1,000	0,074	0,276	0,342	1,000	0,289	0,002	0,000	0,000
5	ECU	1,000	0,078	0,038	0,121	0,963	0,038	0,024	0,037	0,032
6	PER	1,000	0,071	0,001	-0,016	0,406	0,001	0,019	0,594	0,019
7	PRY	1,000	0,074	0,010	-0,051	0,655	0,007	0,037	0,345	0,074
8	SLV	1,000	0,067	0,055	-0,158	0,965	0,055	0,030	0,035	0,044
9	MYS	1,000	0,056	0,063	-0,189	1,000	0,066	-0,003	0,000	0,000
10	USA	1,000	0,054	0,107	-0,247	0,979	0,110	-0,037	0,021	0,052
11	HUN	1,000	0,043	0,129	0,293	0,915	0,123	-0,089	0,085	0,247
12	LKA	1,000	0,058	0,015	0,091	1,000	0,016	-0,000	0,000	0,000
13	LUX	1,000	0,041	0,061	-0,199	0,860	0,055	-0,080	0,140	0,191
14	HRV	1,000	0,044	0,009	0,049	0,400	0,004	-0,060	0,600	0,116
15	TUN	1,000	0,059	0,026	0,118	0,999	0,027	-0,004	0,001	0,001
16	ITA	1,000	0,049	0,043	-0,159	0,911	0,041	-0,050	0,089	0,087

Column Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	INFL	1,000	0,057	0,901	0,703	1,000	0,942	-0,004	0,000	0,001
2	GINI	1,000	0,927	0,053	-0,042	0,985	0,055	0,005	0,015	0,018
3	HDI	1,000	0,015	0,046	-0,080	0,067	0,003	-0,298	0,933	0,981



Em ambos os gráficos e nas tabelas resultantes da análise percebe-se que o Componente 1 (eixo horizontal) é fortemente influenciado pela variável Inflação (INFL). No primeiro gráfico, analisando a posição da variável INFL, mais distante do eixo horizontal, indica a menor influência no Componente 1 em comparação ao segundo gráfico.

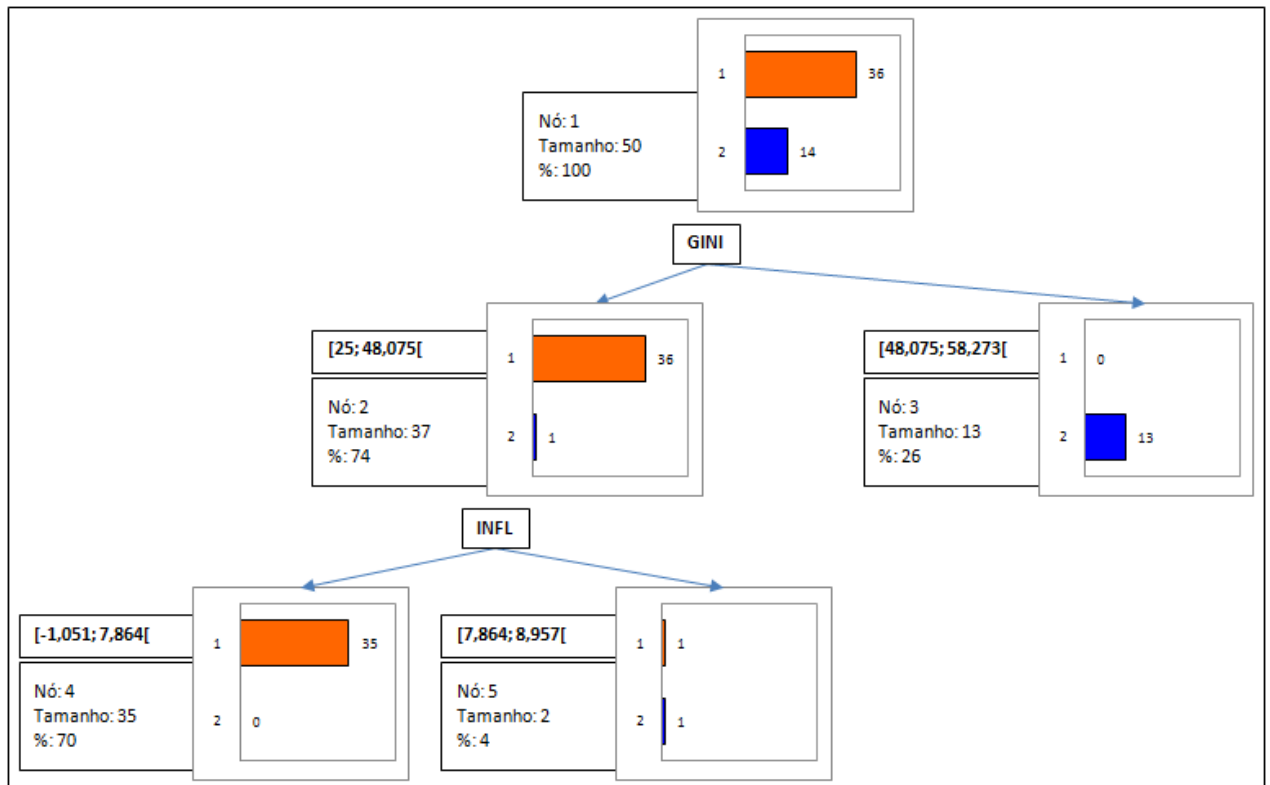
No primeiro gráfico percebe-se uma maior polarização entre INFL (do lado direito do gráfico) e HDI e GINI (mais do lado esquerdo) ambos bem próximos no eixo horizontal. Entretanto a variável Desemprego (UNEMP) tem alta influência no Componente 2 (eixo vertical), causando uma grande dispersão dos países ao longo do gráfico, demonstrando a sua influência nos países.

No segundo gráfico, com a retirada da variável Desemprego (UNEMP), percebe-se um reforço na polarização entre INFL e GINI, sendo que HDI se desloca ligeiramente para baixo. Também percebe-se a quase totalidade da influência no primeiro eixo, concentrando o peso nas variáveis INFL e GINI, conforme visto nas demais análises até agora.

Nota-se que no segundo gráfico existe uma proximidade maior em relação ao eixo horizontal, demonstrando que a influência da variável Desemprego (UNEMP) afeta alguns dos países de maneira mais significativa. Este comportamento pode ser identificado pela variação da posição dos países de um gráfico para o outro. Por exemplo, Tunísia (TUN) e Croácia (HRV) (com maior desemprego) “desce” no primeiro gráfico em relação ao segundo gráfico. Já Costa Rica (CRI) (com menor desemprego), “sobem” no primeiro gráfico em relação ao segundo gráfico.

3.4. ÁRVORE DE CLASSIFICAÇÃO

Mais um método de classificação, o de Árvore de Classificação, será utilizado a seguir para avaliar com outro enfoque os dados apresentados até aqui.



Novamente as variáveis Inflação e GINI aparecem como as principais responsáveis pela classificação e agrupamento dos países conforme demonstrado na figura acima.

4. CONCLUSÃO

Após a utilização dos vários métodos de classificação apresentados nas seções anteriores deste trabalho, conclui-se que as variáveis Inflação e Índice GINI de desigualdade são as mais representativas na classificação e agrupamento dos países analisados neste trabalho.

Percebe-se que quanto menor o nível de Inflação, menor o Índice GINI, relação não encontrada tão fortemente com outras variáveis.

Outra conclusão tem relação com a variável Desemprego, que isoladamente tem capacidade semelhante a variável Inflação para classificar os países, demonstra influência grande em alguns países quando se analisa o gráfico da análise de Correspondência.

O modelo mais simples, ou seja, o modelo parcimonioso, de representação e classificação dos países deste estudo é a Análise Discriminante método Linear utilizando-se a variável GINI como classificadora. Nesta análise consegue-se 94% de precisão na classificação (apenas 3 países são incorretamente classificados).

Atualmente, estudiosos têm utilizado os índices de desigualdade (por exemplo GINI) como principal indicador de desenvolvimento social e parâmetro de medição do sucesso de ações de combate a desigualdade social.

Assim os esforços dos governos dos países devem ser bastante focalizados na manutenção dos índices de inflação em níveis baixos, o que conseqüentemente leva a níveis de desenvolvimento maiores e principalmente menores níveis de desigualdade.