

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

TRABALHO FINAL

Environmental Performance Index (EPI)

“Índice de Desempenho Ambiental”

Yale Center for Environmental Law and Policy

Yale University

Disciplina: Métodos Quantitativos

Professor: Dr. Arnaldo Jose de Hoyos

KARINE SETTERVALL MORAES

1. INTRODUÇÃO

O presente trabalho tem por objetivo efetuar uma análise dos dados obtidos na Universidade de Yale, na Yale Center for Environmental Law and Policy (YCELP), que estuda o índice de performance do Meio Ambiente. Em 2010, o *Environmental Performance Index* (EPI) classificou os países em 25 indicadores de desempenho monitorados em dez categorias, que abrangem tanto a política pública de saúde ambiental e vitalidade de ecossistemas. Estes indicadores fornecem uma medida em uma escala de governo nacional do quão perto estão os países das metas de política ambiental. A metodologia de EPI's apresenta proximidade em relação ao objetivo de facilitar comparações entre países, bem como análise de como a comunidade global está fazendo coletivamente em cada questão política específica. O software estatístico utilizado é o **MINITAB**.

A pesquisa encontra-se disponível em: <http://epi.yale.edu>

Também consideraremos o GDP e o HDI, obtidos através da pesquisa da International Human development indicators, em <http://www.undp.org/>

Para tanto, iniciaremos com o entendimento dos dados, incluindo a definição dos indivíduos e das variáveis, suas classificações em variáveis categóricas ou quantitativas, os significados e unidades de medida, além da apresentação da tabela de dados. Na sequência, analisamos cada uma das variáveis separadamente quanto a sua forma de distribuição, os valores atípicos, medidas de centro e dispersão. Para tal contamos com o auxílio de gráficos (*pie chart*, histogramas, box-plot) e de medidas numéricas (média, mediana, quartis, desvio-padrão, variância, intervalo de confiança e teste de normalidade de Anderson-Darling). No final, buscamos comparar as análises efetuadas para cada variável.

2. ENTENDENDO OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são 233 países do mundo. Para efeito de análise, desconsideraremos os países que não apresentaram dados em alguma variável. Logo, nosso estudo contará com 154 países. Os dados analisados são as variáveis que descrevemos a seguir.

2.2 As Variáveis

São 7 as variáveis desta pesquisa, tendo a amostra 166 indivíduos. As mesmas são melhores explicadas na Tabela 1. Ressaltamos que todos os dados desta pesquisa são referentes ao ano de 2010, exceto GDP per capita que é referente ao ano de 2008.

Tabela 1. As Variáveis

Variável	Código	Significado	Unidade de Medida
Enviromental Burde of Disease	DALY	A Organização Mundial da Saúde capta o impacto ambiental sobre a saúde humana através de anos de vida ajustados por incapacidade (DALYS). DALY é a soma do número de anos de vida perdidos devido à mortalidade prematura causada pela doença influenciada pelo ambiente e os anos de vida saudável perdidos por incapacidade causada por essa doença.	Número (10 DALYs per 1,000 population) Menor, melhor

Air Pollution (effects on nature)	AIR_E	Compostos reativos tais como o ozônio (O3), benzeno (C6H6), dióxido de enxofre (SO2), óxidos de azoto (NOx) e compostos orgânicos voláteis (VOCs) que agem negativamente no impacto do crescimento das plantas	Gg/sq km Menor, melhor
Water (effects on nature)	WATER_E	Índice de qualidade do ar, índice de estresse hídrico, índices Escassez de Água	mg/l Menor, melhor
Biodiversity & Habitat	BIODIVERSITY	Proteção do Bioma, habitat crítico, proteção marinha.	% área protegida Maior, melhor
Climate Change	CLIMATE	Índice composto por: emissões de gases com efeito de estufa per capita, incluindo as emissões da mudança do uso da terra; as emissões de dióxido de carbono por unidade de geração de eletricidade, e; em terceiro lugar, a intensidade de emissões de gases com efeito de estufa industrial por unidade de PPP gerado.	g CO2 per kWh Menor melhor
Human Development Index	HDI	Um índice composto que mede a realização média em três dimensões básicas da vida humana, o desenvolvimento de uma vida longa e saudável, conhecimento e um padrão de vida decente	Número Maior, melhor
GDP per capita (2008 PPP US\$)	GDP	Soma do valor adicionado por todos os produtores residentes na economia mais quaisquer impostos sobre os produtos (menos subsídios) não incluídos na valoração da produção, calculado sem deduções por depreciação de ativos de capital fabricados ou para esgotamento e a degradação dos recursos naturais. Valor adicionado é a saída líquida de uma indústria depois de somar todas as saídas e subtrair insumos intermediários. Quando expressa em termos de EUA \$, ele é convertido pela taxa de câmbio média oficial relatada pelo Fundo Monetário Internacional. Um fator de conversão alternativa é aplicado se a taxa de câmbio oficial é julgada diverge por uma margem excepcionalmente grande da taxa efetivamente aplicada às operações em moeda estrangeira e dos produtos comercializados. Quando expresso em paridade de poder aquisitivo (PPP) EUA \$ termos, ele é convertido para dólares internacionais usando taxas de PPP. Um dólar internacional tem o mesmo poder de compra em relação ao PIB que o dólar dos EUA nos Estados Unidos.	Número Maior, melhor

2.3 A Tabela de Dados

A tabela de dados utilizada no presente trabalho encontra-se em:

<http://epi.yale.edu/Files>

Para o GDP e HDI, a tabela encontra-se em:

<http://hdrstats.undp.org/en/indicators>

3. ANÁLISE DAS VARIÁVEIS

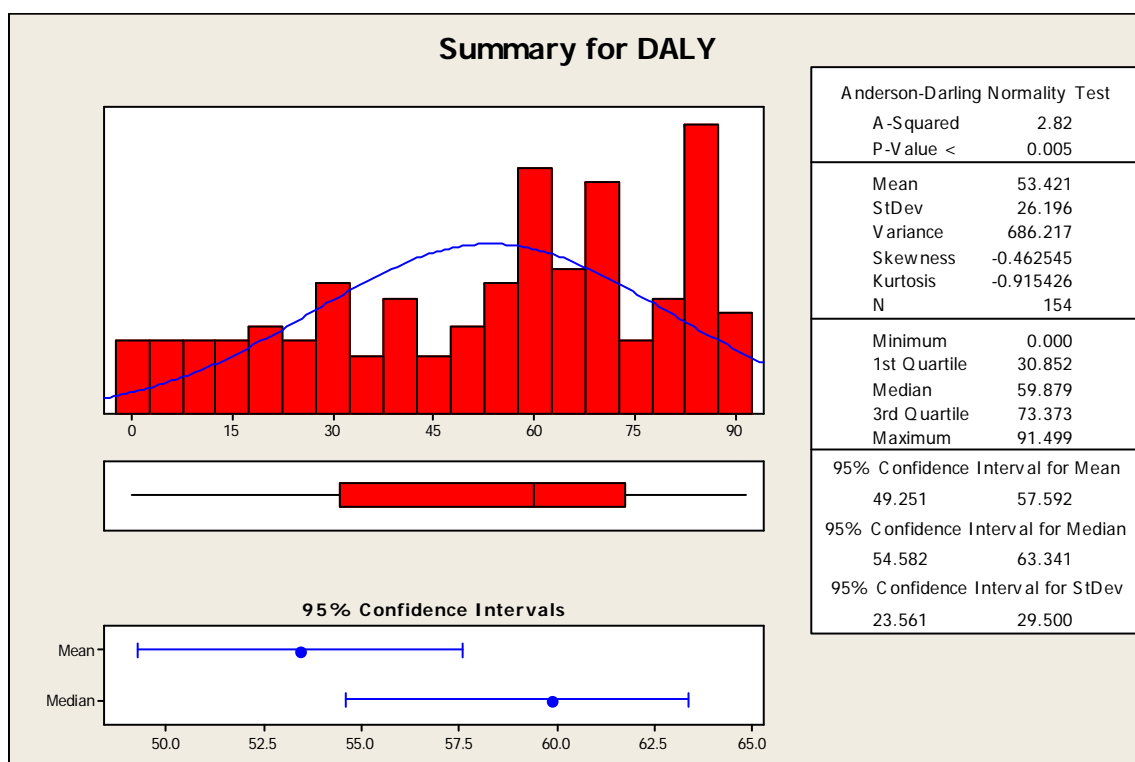
3.1. Caracterização da amostra

Variáveis Quantitativas

A análise deste tipo de variável permite a utilização de uma maior gama de ferramentas de análise como histogramas, curvas de densidade, gráfico de ramos, box-plot e dot-plot, além de informações numéricas como média, desvio-padrão, mediana, quartis, 5 números, intervalo de confiança e teste de normalidade de Anderson-Darling.

3.1.1 Variável: DALY

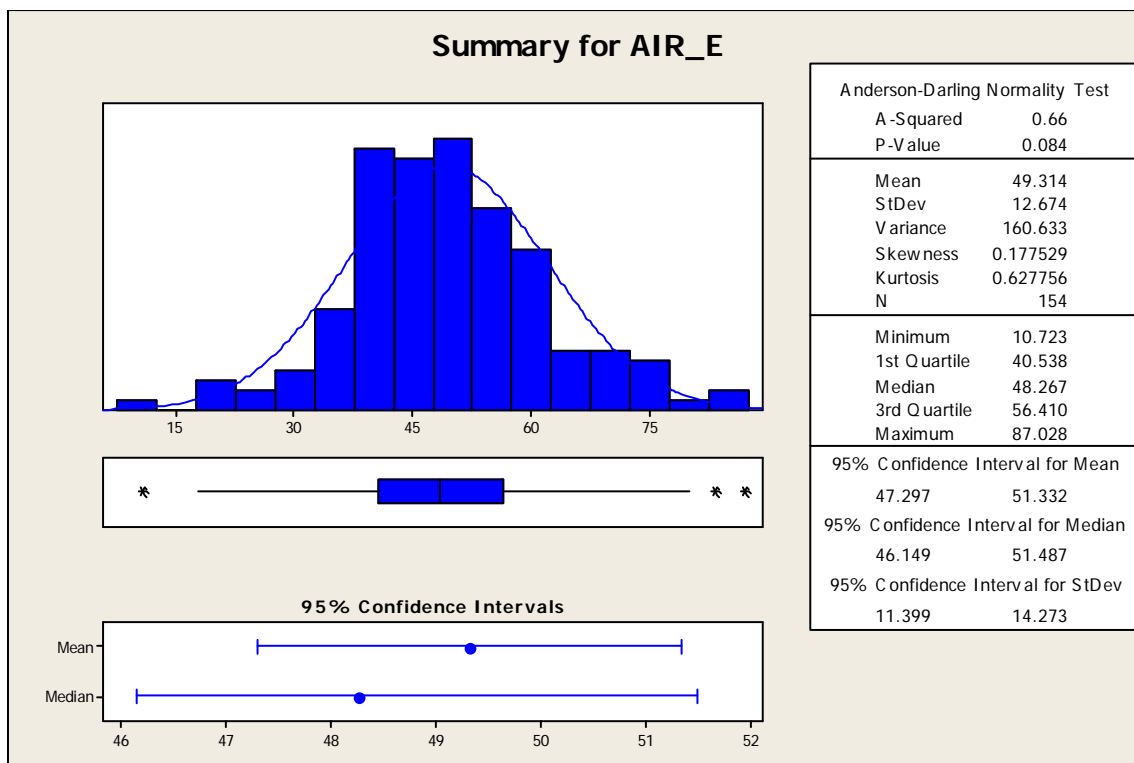
Segue abaixo quadro contendo Histograma, Curva de Densidade, Box-Plot, Intervalo de confiança da média e mediana, além das medidas numéricas como média, desvio-padrão, variância, quantidade de observações, valores mínimos, máximos, informações dos quartis e o teste de normalidade de Anderson-Darling (A-Squared e P-Value), para a variável RLV.



- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição assimétrica a esquerda, o que é confirmado pelo P-Value do teste de Anderson DarLing, que nos indica não tratar-se de uma distribuição Normal.
- **Valores Atípicos:** Não foi identificado valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos Países está abaixo de 59,8 e metade maior do que este valor. A média da amostra é de 53,4, com

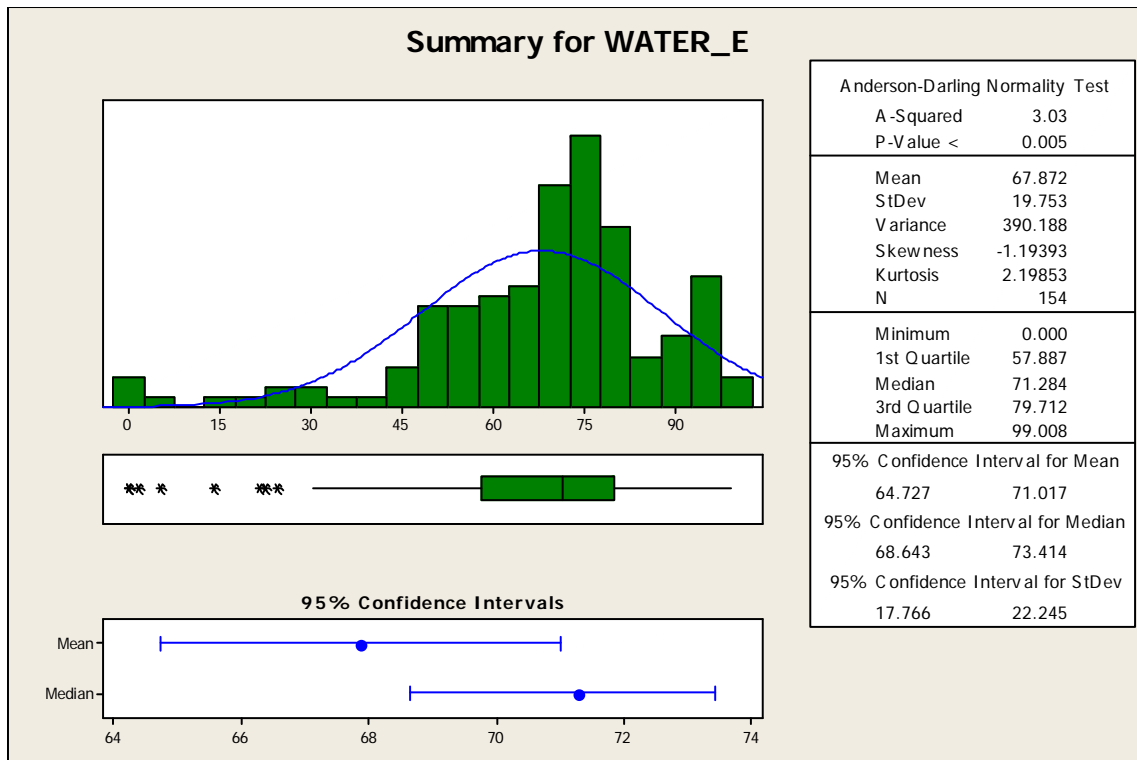
desvio-padrão (medida de dispersão) bastante elevado de 26,19. O valor mínimo foi zero (Angola, Serra Leoa e Nigéria) e o máximo 91,49 (Israel e Islândia).

3.1.2 Variável: Air E



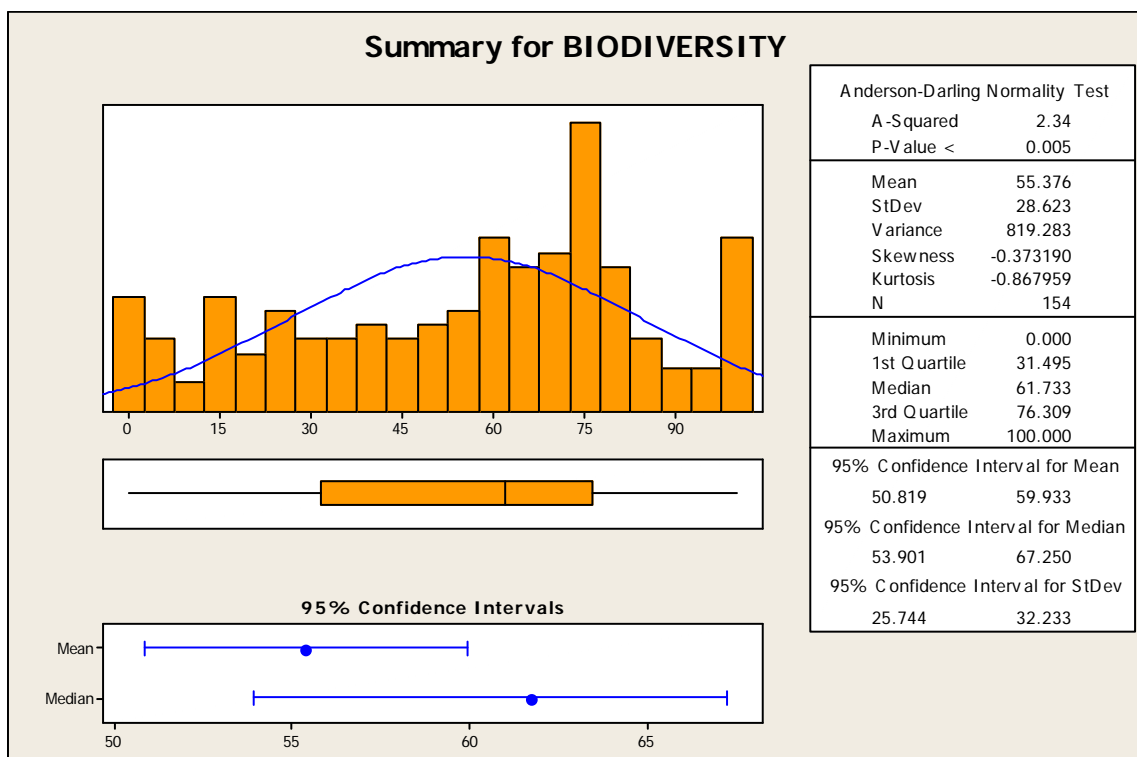
- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição assimétrica a direita, o que é confirmado pelo P-Value do teste de Anderson DarLing, que nos indica não tratar-se de uma distribuição Normal.
- **Valores Atípicos:** Os valores atípicos são: Singapura: 10,72; Kazaquistão 83,2 e Ilhas Salomão 87,0.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos Países está abaixo de 48,26 e metade maior do que este valor. A média da amostra é de 49,3, com desvio-padrão (medida de dispersão) bastante elevado de 12,67. O valor mínimo foi 10,72 (Singapura) e o máximo 87,0 (Ilhas Salomão).

3.1.3 Variável: Water E



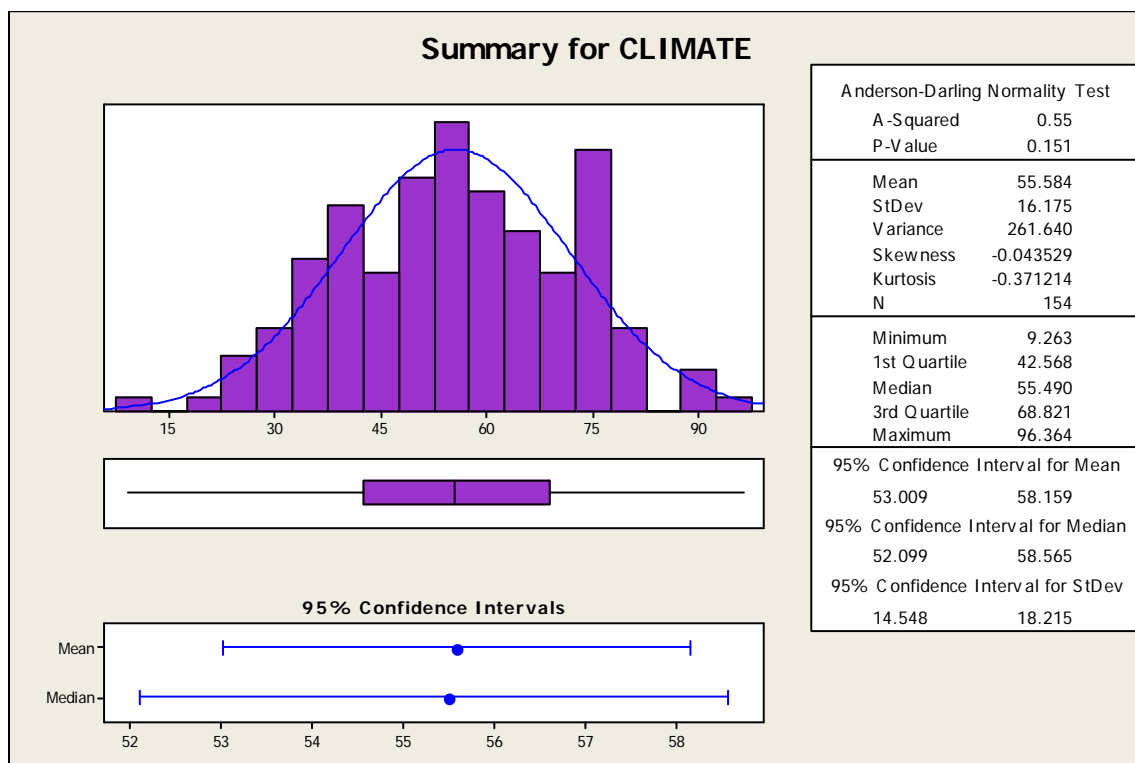
- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição assimétrica a esquerda, o que é confirmado pelo P-Value do teste de Anderson DarLing, que nos indica não tratar-se de uma distribuição Normal.
- **Valores Atípicos:** OS valores atípicos foram Kuwait, Bahrain, Yemen, Emirados Árabes, Catar, Usbequistão, Arábia Saudita e Jordânia, todos abaixo de 30.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos Países está abaixo de 71,28 e metade maior do que este valor. A média da amostra é de 67,87, com desvio-padrão (medida de dispersão) bastante elevado de 19,75. O valor mínimo foi zero (Kuwait e Bahrain) e o máximo 91,49 (Singapura).

3.1.4 Variável: Biodiversidade



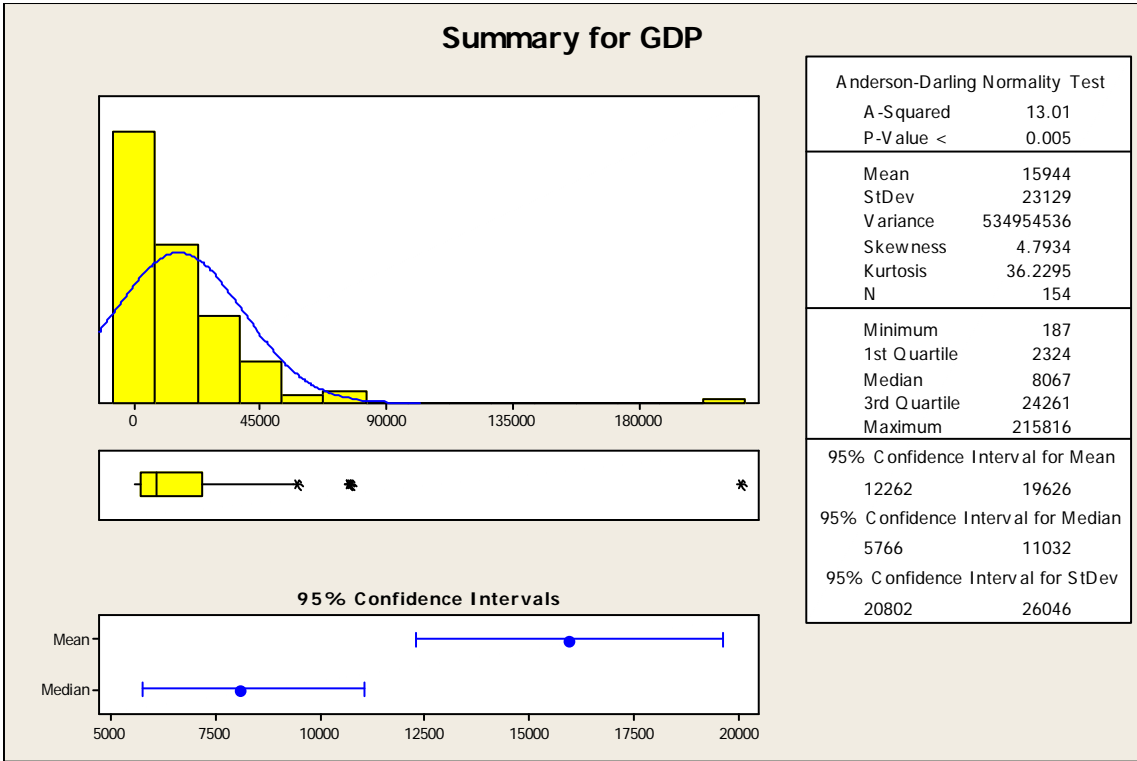
- Forma: O Histograma nos permite verificar que trata-se de uma distribuição assimétrica a esquerda, o que é confirmado pelo P-Value do teste de Anderson DarLing, que nos indica não tratar-se de uma distribuição Normal.
- Valores Atípicos: Não foram identificados valores atípicos na amostra.
- Centro e Dispersão: A mediana nos indica que aproximadamente metade dos Países está abaixo de 61,73 e metade maior do que este valor. A média da amostra é de 55,37, com desvio-padrão (medida de dispersão) bastante elevado de 28,62. O valor mínimo foi zero (São Tomé e Príncipe) e o máximo 91,49 (Bostswana, Burkina, República da África Central, Zâmbia, Alemanha, República Tcheca, Luxemburgo, Laos, Suíça, Eslováquia e Áustria).

3.1.5 Variável: Mudança de Clima



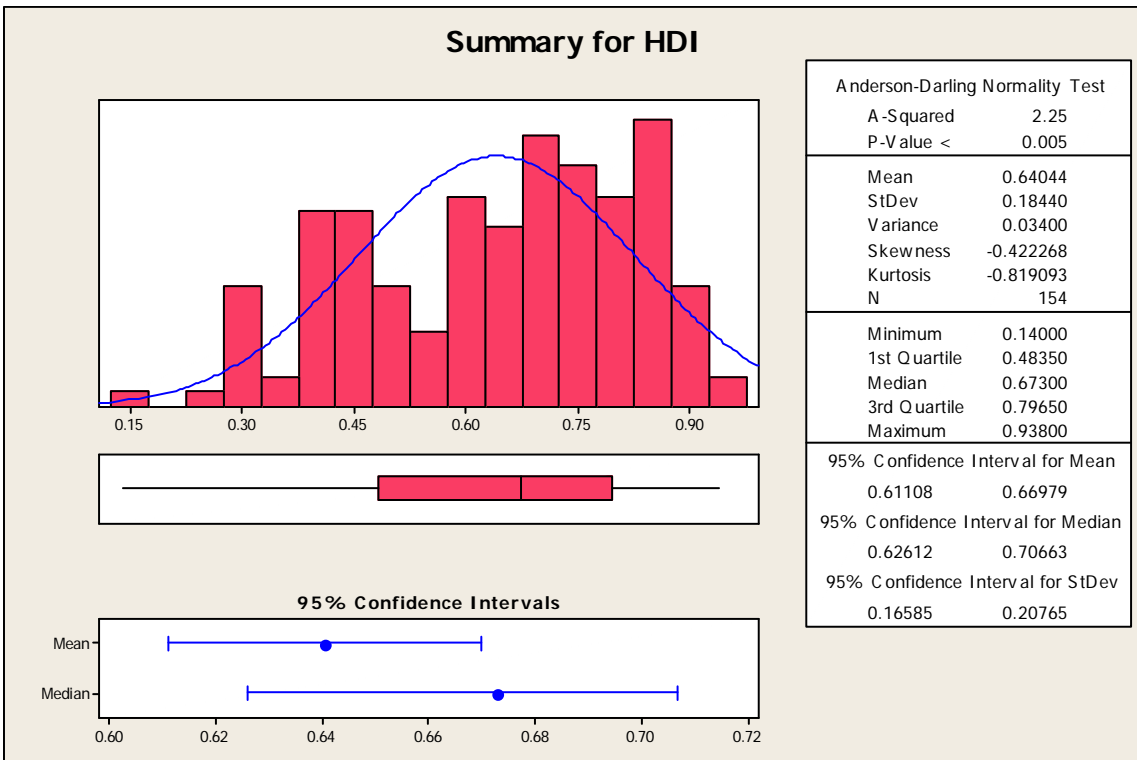
- Forma: O Histograma nos permite verificar que trata-se de uma distribuição levemente assimétrica a direita, o que é confirmado pelo P-Value do teste de Anderson DarLing, que nos indica não tratar-se de uma distribuição Normal.
- Valores Atípicos: Não foram identificados valores atípicos na amostra.
- Centro e Dispersão: A mediana nos indica que aproximadamente metade dos Países está abaixo de 55,49 e metade maior do que este valor. A média da amostra é de 55,58, com desvio-padrão (medida de dispersão) bastante elevado de 16,17. O valor mínimo foi 9,26 (Cipros) e o máximo 96,36 (Nepal).

3.1.6 Variável: GDP



- **Forma:** O Histograma nos permite verificar que trata-se de uma distribuição assimétrica a esquerda, o que é confirmado pelo P-Value do teste de Anderson DarLing, que nos indica não tratar-se de uma distribuição Normal.
- **Valores Atípicos:** Foram identificados quatro valores atípicos: Luxemburgo, Macedônia, Catar e Malta.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos Países está abaixo de 8.067 e metade maior do que este valor. A média da amostra é de 15944, com desvio-padrão (medida de dispersão) bastante elevado de 23129. O valor mínimo foi 187 (Zimbábue) e o máximo 215816 (Malta).

3.1.7 Variável: HDI



- Forma: O Histograma nos permite verificar que trata-se de uma distribuição assimétrica a esquerda, o que é confirmado pelo P-Value do teste de Anderson DarLing, que nos indica não tratar-se de uma distribuição Normal.
- Valores Atípicos: Não foram identificados valores atípicos.
- Centro e Dispersão: A mediana nos indica que aproximadamente metade dos Países está abaixo de 0,67 e metade maior do que este valor. A média da amostra é de 0,64, com desvio-padrão (medida de dispersão) 0,18. O valor mínimo foi 0,14 (Zimbábue) e o máximo 0,93 (Noruega).

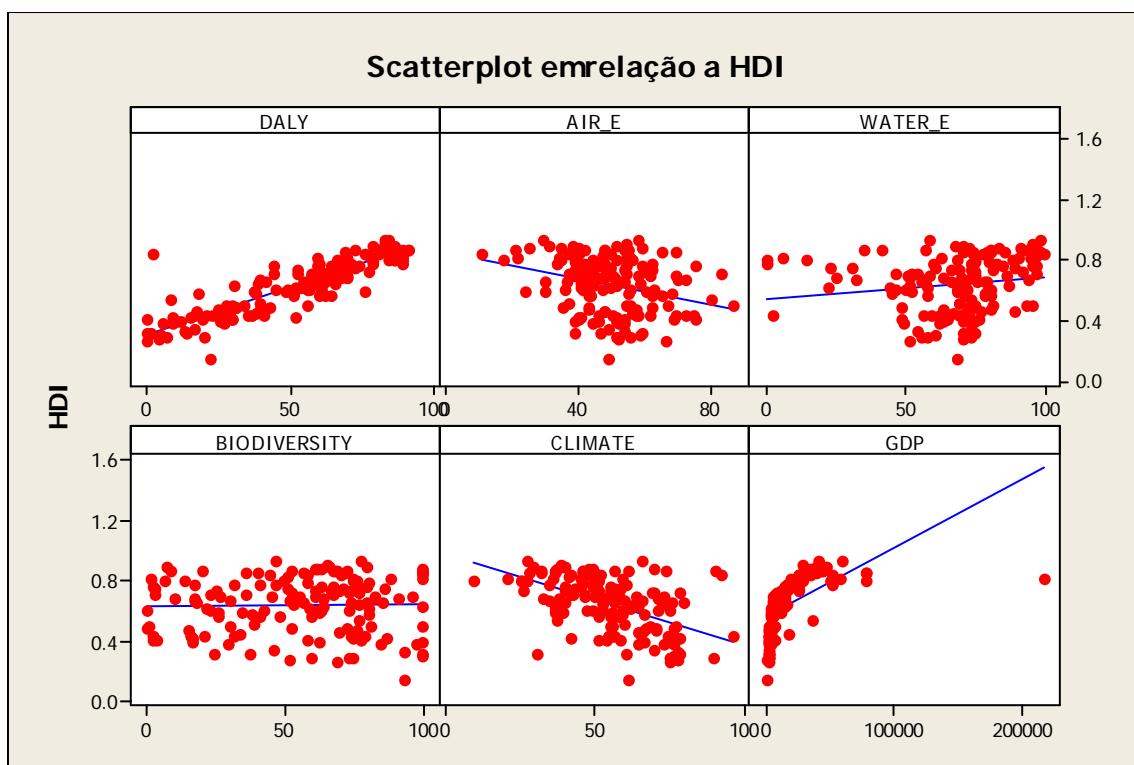
4. CORRELAÇÃO ENTRE AS VARIÁVEIS

Correlations: DALY, AIR_E, WATER_E, BIODIVERSITY, CLIMATE, GDP, HDI					
	DALY	AIR_E	WATER_E	BIODIVERSITY	
AIR_E	-0.348 0.000				
WATER_E	0.089 0.272	0.040 0.623			
BIODIVERSITY	-0.090 0.269	-0.139 0.086	0.305 0.000		
CLIMATE	-0.535 0.000	0.280 0.000	0.111 0.171	-0.005 0.946	
GDP	0.544 0.000	-0.304 0.000	0.018 0.822	0.053 0.511	
HDI	0.888 0.000	-0.296 0.000	0.162 0.045	0.032 0.690	
	CLIMATE	GDP			
GDP	-0.348 0.000				
HDI	-0.524 0.000	0.576 0.000			

Cell Contents: Pearson correlation
P-Value

Observamos que a maior correlação se dá entre as variáveis HDI e DALY. Também estão correlacionadas as variáveis HDI e Ar e, GDP e DALY.

A seguir, veremos como se comportam as variáveis através do gráfico Scatterplot, se comprova a correlação acima.



Tendo em vista os valores estarem mais agrupados em DALY, verificamos que as variáveis HDI e DALY são as que mais se correlacionam.

5. REGRESSÕES MÚLTIPLAS E CORRELAÇÃO ENTRE SEUS COMPORTAMENTOS

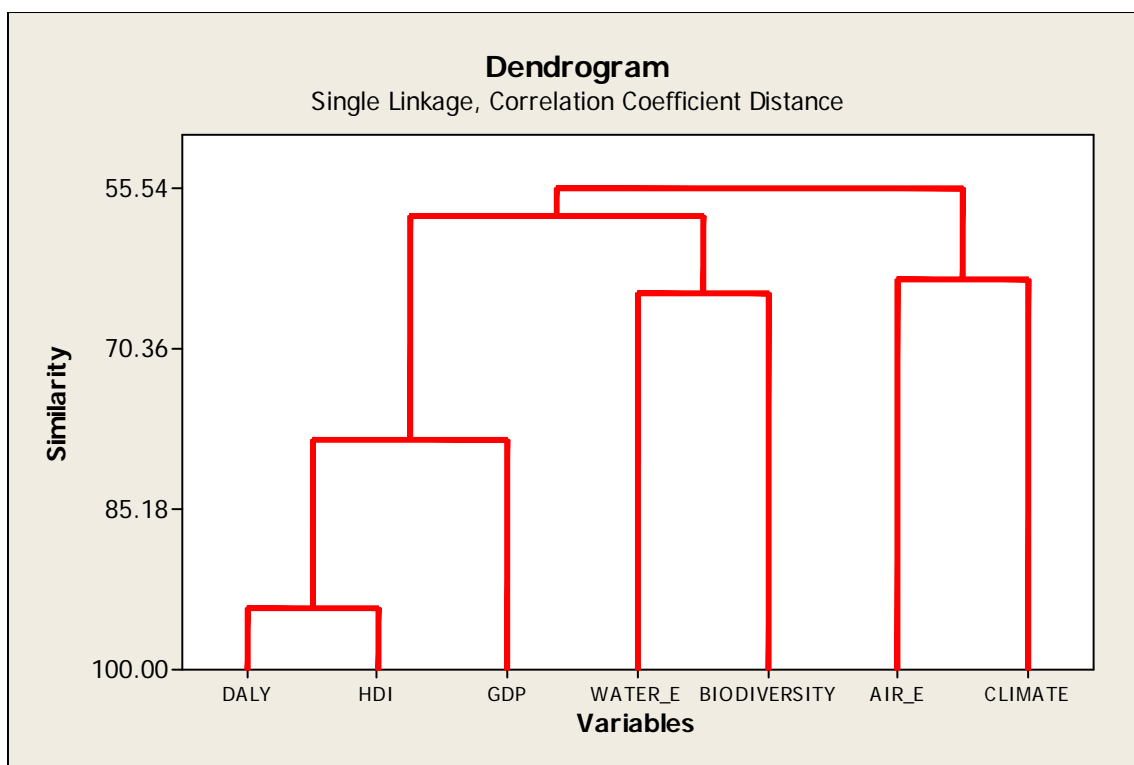
A matriz de correlação abaixo nos ajuda tirar conclusões mais precisas sobre a associação entre as variáveis.

Cluster Analysis of Variables: DALY, AIR_E, WATER_E, BIODIVERSITY, CLIMATE, ...

Correlation Coefficient Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	6	94.4099	0.111802	1	7	2
2	5	78.8203	0.423593	1	6	3
3	4	65.2286	0.695428	3	4	2
4	3	63.9823	0.720353	2	5	2
5	2	58.1036	0.837928	1	3	5
6	1	55.5413	0.889174	1	2	7

Dendrogram



Com base na análise do dendrograma acima, podemos verificar que existe maior correlação entre HDI e DALY, logo em seguida a melhor correlação ocorre com GDP, Água e Biodiversidade e por último, com uma correlação bem mais baixa, com Ar e Clima.

Primeiramente em relação ao HDI:

Regression Analysis: HDI versus DALY, AIR_E, ...

The regression equation is

$$\text{HDI} = 0.262 + 0.00562 \text{ DALY} + 0.000728 \text{ AIR_E} + 0.000651 \text{ WATER_E} \\ + 0.000532 \text{ BIODIVERSITY} - 0.000850 \text{ CLIMATE} + 0.000001 \text{ GDP}$$

Predictor	Coef	SE Coef	T	P
Constant	0.26183	0.05171	5.06	0.000
DALY	0.0056230	0.0003471	16.20	0.000
AIR_E	0.0007275	0.0005636	1.29	0.199
WATER_E	0.0006507	0.0003575	1.82	0.071
BIODIVERSITY	0.0005324	0.0002474	2.15	0.033
CLIMATE	-0.0008497	0.0004871	-1.74	0.083
GDP	0.00000100	0.00000034	2.94	0.004

S = 0.0798348 **R-Sq = 82.0%** R-Sq(adj) = 81.3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	4.26534	0.71089	111.54	0.000
Residual Error	147	0.93692	0.00637		
Total	153	5.20226			

Source	DF	Seq SS
DALY	1	4.10404
AIR_E	1	0.00108
WATER_E	1	0.03534
BIODIVERSITY	1	0.04688
CLIMATE	1	0.02274
GDP	1	0.05526

Unusual Observations

Obs	DALY	HDI	Fit	SE Fit	Residual	St Resid
29	2.0	0.84100	0.33452	0.02130	0.50648	6.58R
78	76.0	0.59700	0.75473	0.01912	-0.15773	-2.03R
92	86.9	0.81500	1.01164	0.06335	-0.19664	-4.05RX
114	52.0	0.42300	0.59974	0.01562	-0.17674	-2.26R
154	22.0	0.14000	0.46350	0.01321	-0.32350	-4.11R

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large leverage.

Verificamos que o comportamento do HDI está 82,0% sendo explicado pelo comportamento das demais variáveis, ou seja, está sendo consideravelmente explicado por elas.

Este comportamento está sendo explicado através das seguintes equações:

$$\text{HDI} = 0.262 + 0.00562 \text{ DALY} + 0.000728 \text{ AIR_E} + 0.000651 \text{ WATER_E} \\ + 0.000532 \text{ BIODIVERSITY} - 0.000850 \text{ CLIMATE} + 0.000001 \text{ GDP}$$

Outra possível análise que podemos fazer em relação ao comportamento do HDI, onde República do Congo obteve um valor acima do esperado e Coréia do Sul, Malta, Coréia do Norte e Zimbábue apresentaram valores abaixo do esperado.

Stepwise Regression: HDI versus DALY, AIR_E, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is HDI on 6 predictors, with N = 154

Step	1	2	3	4	5
Constant	0.3064	0.2623	0.2762	0.2473	0.3007
DALY	0.00625	0.00632	0.00587	0.00579	0.00553
T-Value	23.83	24.68	19.50	19.16	16.25
P-Value	0.000	0.000	0.000	0.000	0.000
BIODIVERSITY		0.00073	0.00065	0.00052	0.00047
T-Value		3.10	2.81	2.14	1.94
P-Value		0.002	0.006	0.034	0.055
GDP			0.00000	0.00000	0.00000
T-Value			2.75	2.89	2.81
P-Value			0.007	0.004	0.006
WATER_E				0.00058	0.00070
T-Value				1.65	1.98
P-Value				0.100	0.050
CLIMATE					-0.00080
T-Value					-1.65
P-Value					0.101
S	0.0850	0.0827	0.0809	0.0805	0.0800
R-Sq	78.89	80.16	81.11	81.45	81.79
R-Sq(adj)	78.75	79.89	80.73	80.95	81.17
Mallows Cp	22.3	14.0	8.2	7.4	6.7

Com 78,89% DALY já explicaria o comportamento da variável.HDI.

Em relação ao GDP:

Regression Analysis: GDP versus DALY, AIR_E, ...

The regression equation is

$$\text{GDP} = -9879 + 99 \text{ DALY} - 218 \text{ AIR_E} - 87.1 \text{ WATER_E} + 44.5 \text{ BIODIVERSITY} \\ - 18 \text{ CLIMATE} + 55803 \text{ HDI}$$

Predictor	Coef	SE Coef	T	P
Constant	-9879	13225	-0.75	0.456
DALY	99.2	136.7	0.73	0.469
AIR_E	-218.0	132.8	-1.64	0.103
WATER_E	-87.09	85.18	-1.02	0.308
BIODIVERSITY	44.52	59.30	0.75	0.454
CLIMATE	-18.3	116.4	-0.16	0.876
HDI	55803	18952	2.94	0.004

S = 18877.7 R-Sq = 36.0% R-Sq(adj) = 33.4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	29462086615	4910347769	13.78	0.000
Residual Error	147	52385957412	356367057		
Total	153	81848044027			

Source	DF	Seq SS
DALY	1	24251304600
AIR_E	1	1227813802
WATER_E	1	36241868
BIODIVERSITY	1	726956890
CLIMATE	1	130198484
HDI	1	3089570970

Unusual Observations

Obs	DALY	GDP	Fit	SE Fit	Residual	St Resid
29	2.0	24419	20829	11084	3590	0.23 X
84	82.8	76440	28197	5224	48243	2.66R
90	69.0	76440	23459	4471	52981	2.89R
92	86.9	215816	36337	4469	179479	9.79R
117	89.1	77178	34693	5410	42485	2.35R

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large leverage.

Verificamos que o comportamento do GDP está 36,0% sendo explicado pelo comportamento das demais variáveis, ou seja, está sendo levemente explicado por elas. Este comportamento está sendo explicado através das seguintes equações:

$$\text{GDP} = -9879 + 99 \text{ DALY} - 218 \text{ AIR_E} - 87.1 \text{ WATER_E} + 44.5 \text{ BIODIVERSITY} \\ - 18 \text{ CLIMATE} + 55803 \text{ HDI}$$

Outra possível análise que podemos fazer em relação ao comportamento do HDI, onde República do Congo, Luxemburgo, Macedônia, Malta e Catar apresentaram valores abaixo do esperado.

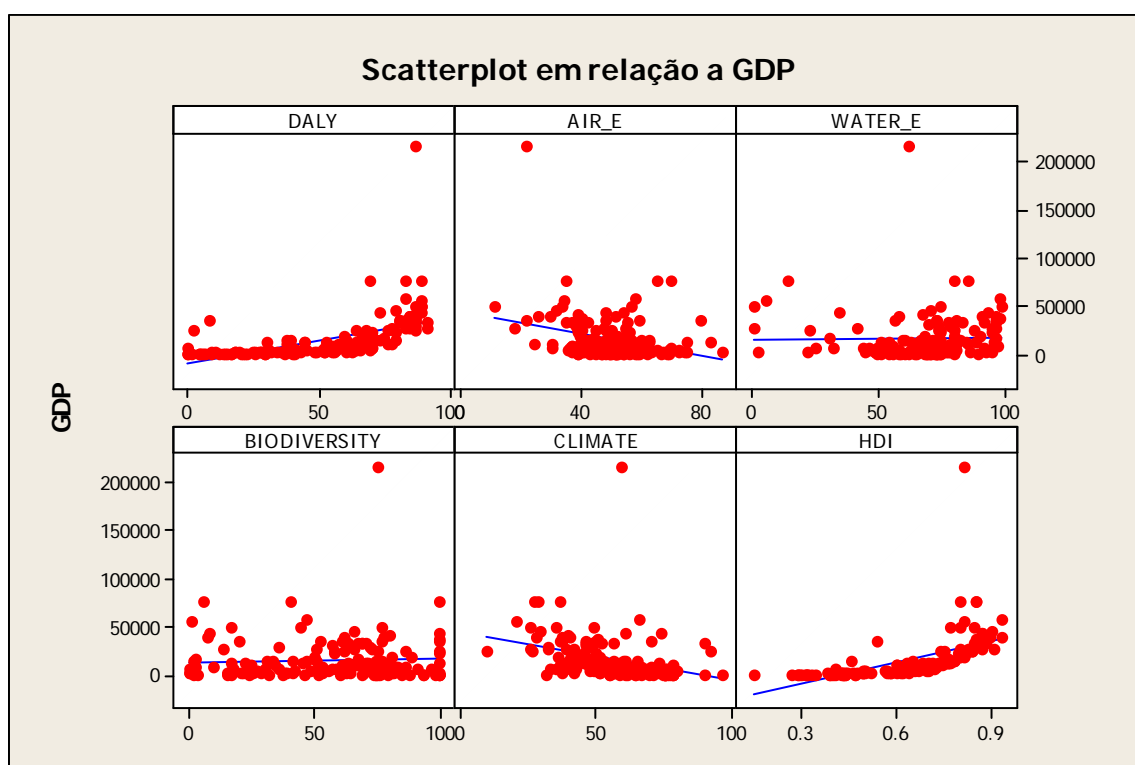
Stepwise Regression: GDP versus DALY, AIR_E, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is GDP on 6 predictors, with N = 154

Step	1	2
Constant	-30359	-13671
HDI	72300	66857
T-Value	8.70	7.77
P-Value	0.000	0.000
AIR_E		-268
T-Value		-2.14
P-Value		0.034
S	18962	18743
R-Sq	33.22	35.19
R-Sq(adj)	32.79	34.33
Mallows Cp	3.4	0.9

Com apenas 33,22% DALY explica o comportamento da variável.HDI. Valor muito baixo.



Tendo em vista os gráficos de Daly e HDI apresentarem os dados mais agrupados, verificamos que GDP está mais correlacionado com estas variáveis, HDI e GDP.

Em relação a DALY:

Regression Analysis: DALY versus AIR_E, WATER_E, ...

The regression equation is

$$\text{DALY} = 2.65 - 0.204 \text{ AIR_E} + 0.0162 \text{ WATER_E} - 0.124 \text{ BIODIVERSITY} - 0.126 \text{ CLIMATE} + 0.000036 \text{ GDP} + 114 \text{ HDI}$$

Predictor	Coef	SE Coef	T	P
Constant	2.655	7.975	0.33	0.740
AIR_E	-0.20405	0.07892	-2.59	0.011
WATER_E	0.01616	0.05145	0.31	0.754
BIODIVERSITY	-0.12373	0.03429	-3.61	0.000
CLIMATE	-0.12582	0.06930	-1.82	0.071
GDP	0.00003596	0.00004957	0.73	0.469
HDI	113.992	7.036	16.20	0.000

S = 11.3670 R-Sq = 81.9% R-Sq(adj) = 81.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	85998	14333	110.93	0.000
Residual Error	147	18994	129		
Total	153	104991			

Source	DF	Seq SS
AIR_E	1	12730
WATER_E	1	1115
BIODIVERSITY	1	3428
CLIMATE	1	23014
GDP	1	11800
HDI	1	33910

Unusual Observations

Obs	AIR_E	DALY	Fit	SE Fit	Residual	St Resid
1	40.1	0.000	27.662	1.994	-27.662	-2.47R
29	49.9	1.977	71.602	3.405	-69.625	-6.42R
56	80.0	8.392	35.844	3.619	-27.452	-2.55R
92	21.8	86.863	83.168	9.306	3.695	0.57 X
154	49.2	22.012	-9.630	3.341	31.643	2.91R

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large leverage.

Verificamos que o comportamento do DALY está 81,9% sendo explicado pelo comportamento das demais variáveis, ou seja, está sendo consideravelmente explicado por elas.

Este comportamento está sendo explicado através das seguintes equações:

$$\text{DALY} = 2.65 - 0.204 \text{ AIR_E} + 0.0162 \text{ WATER_E} - 0.124 \text{ BIODIVERSITY} - 0.126 \text{ CLIMATE} + 0.000036 \text{ GDP} + 114 \text{ HDI}$$

Outra possível análise que podemos fazer em relação ao comportamento do HDI, onde República do Congo, Angola e Guiné Equatorial apresentaram valores abaixo do esperado, enquanto que Malta e Zimbábue, abaixo do esperado.

Stepwise Regression: DALY versus AIR_E, WATER_E, ...

Alpha-to-Enter: 0.15 Alpha-to-Remove: 0.15

Response is DALY on 6 predictors, with N = 154

Step	1	2	3	4
------	---	---	---	---

Constant	-27.389	-21.729	-6.536	2.434
HDI	126.2	126.7	122.1	116.9
T-Value	23.83	24.68	23.33	19.73
P-Value	0.000	0.000	0.000	0.000
BIODIVERSITY		-0.109	-0.122	-0.120
T-Value		-3.28	-3.75	-3.71
P-Value		0.001	0.000	0.000
AIR_E			-0.233	-0.210
T-Value			-3.03	-2.73
P-Value			0.003	0.007
CLIMATE				-0.123
T-Value				-1.83
P-Value				0.069
S	12.1	11.7	11.4	11.3
R-Sq	78.89	80.29	81.43	81.84
R-Sq(adj)	78.75	80.03	81.06	81.35
Mallows Cp	21.5	12.1	4.9	3.6

Com apenas 78,89% HDI explica o comportamento da variável DALY.

O alto valor de R-Quadrado encontrado demonstra uma similaridade de comportamento entre HDI e as demais variáveis e entre DALY e demais variáveis.

Correlations: DALY, AIR_E, WATER_E, BIODIVERSITY, CLIMATE, GDP, HDI

	DALY	AIR_E	WATER_E	BIODIVERSITY
AIR_E	-0.348 0.000			
WATER_E	0.089 0.272	0.040 0.623		
BIODIVERSITY	-0.090 0.269	-0.139 0.086	0.305 0.000	
CLIMATE	-0.535 0.000	0.280 0.000	0.111 0.171	-0.005 0.946
GDP	0.544 0.000	-0.304 0.000	0.018 0.822	0.053 0.511
HDI	0.888 0.000	-0.296 0.000	0.162 0.045	0.032 0.690
	CLIMATE	GDP		
GDP	-0.348 0.000			
HDI	-0.524 0.000	0.576 0.000		

Cell Contents: Pearson correlation
P-Value

De acordo com os percentuais encontrados existe uma forte correlação entre as variáveis independentes HDI e DALY. As variáveis HDI e Ar, GDP e Daly também estão correlacionadas.

6. ANÁLISE DOS COMPONENTES PRINCIPAIS

O objetivo deste tópico é, através da análise dos componentes principais, tentarmos reduzir o número de variáveis, ou seja, percebermos as relações entre as variáveis e a possibilidade de agruparmos as mesmas. Certamente a análise de correlações e dendogramas acima já nos dão uma idéia de que a possibilidade de agrupamento é grande pelos elevados índices de correlação entre algumas variáveis

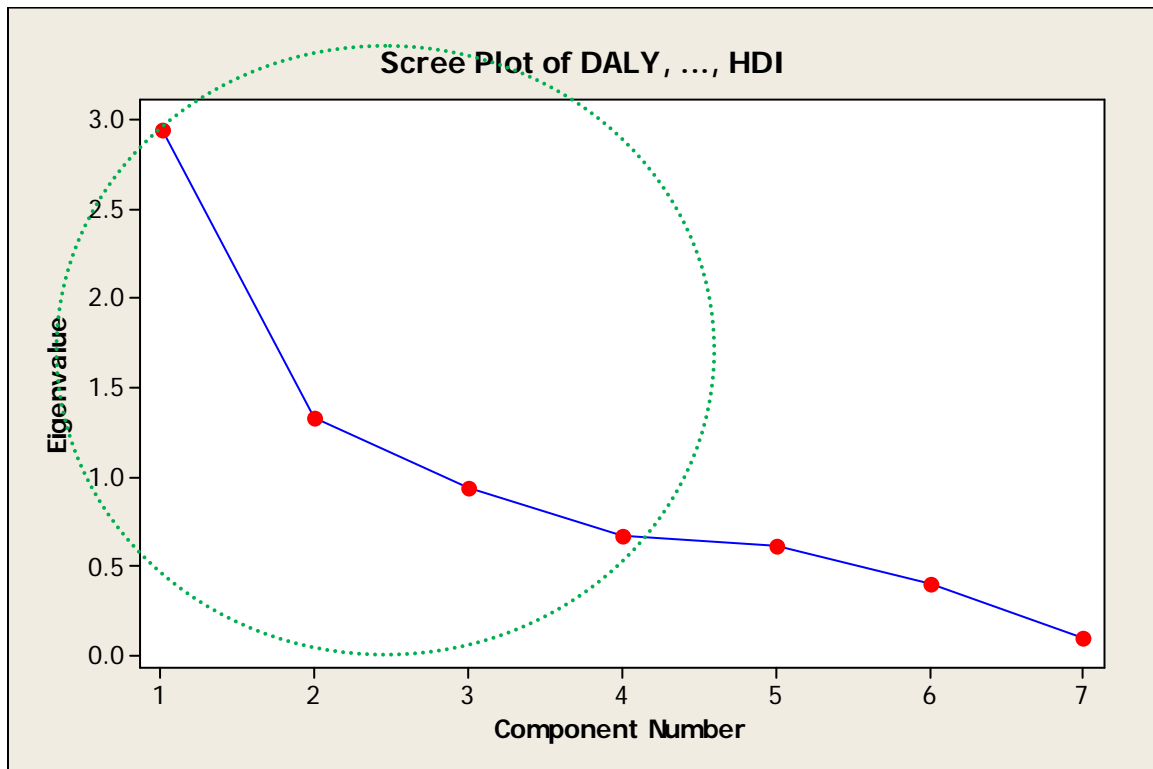
Principal Component Analysis: DALY, AIR_E, WATER_E, BIODIVERSITY, CLIMATE, GDP,

Eigenanalysis of the Correlation Matrix

Eigenvalue	2.9402	1.3297	0.9420	0.6725	0.6129	0.4037	0.0988
Proportion	0.420	0.190	0.135	0.096	0.088	0.058	0.014
Cumulative	0.420	0.610	0.745	0.841	0.928	0.986	1.000

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
DALY	0.529	-0.058	-0.232	-0.062	-0.131	0.389	0.700 Menor
AIR_E	-0.302	-0.056	-0.698	0.518	0.374	0.072	0.068 Menor
WATER_E	0.047	0.698	-0.444	-0.204	-0.329	-0.404	0.021 Menor
BIODIVERSITY	0.026	0.688	0.425	0.417	0.264	0.303	0.096 Maior
CLIMATE	-0.406	0.173	-0.129	-0.676	0.299	0.492	0.001 Menor
GDP	0.427	0.014	0.008	-0.233	0.758	-0.432	0.043 Maior
HDI	0.529	0.052	-0.254	0.024	-0.021	0.399	-0.702 Maior

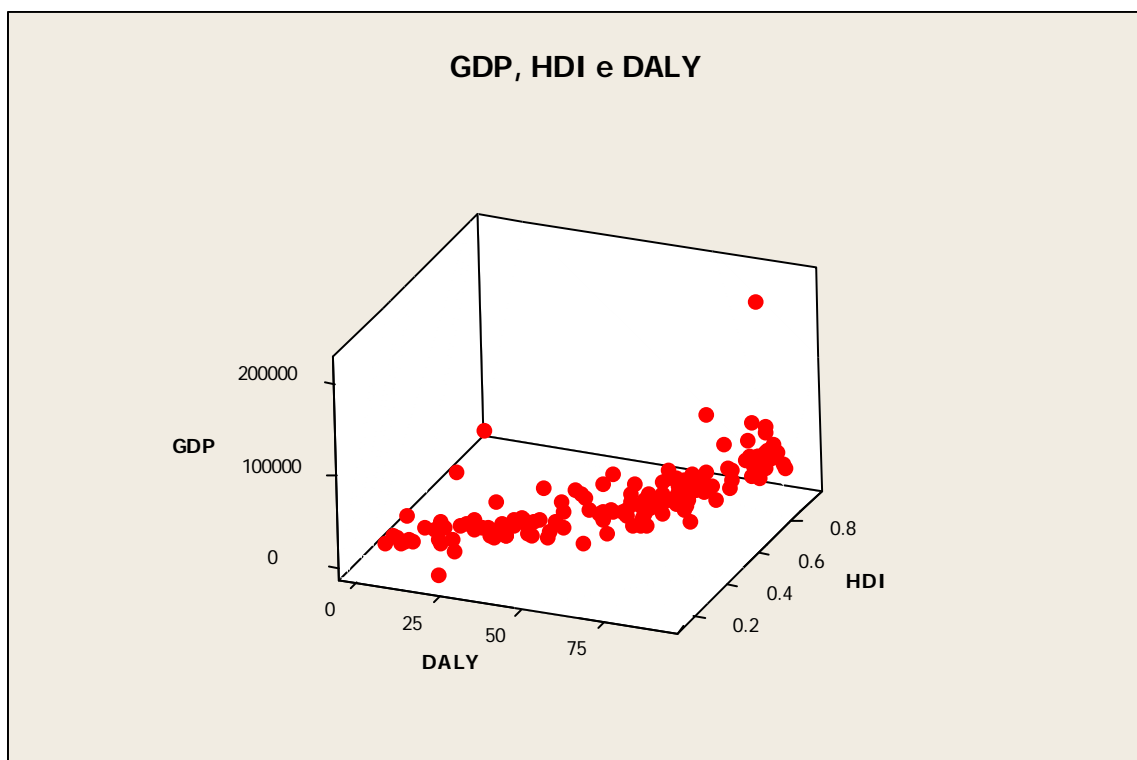
Scree Plot of DALY, ..., HDI



Pela análise dos detalhes e gráfico acima percebemos que se juntarmos as 7 variáveis em apenas 3 (PC3) teremos um proporção de 74,5%. Se juntarmos mais uma variável, em PC4, alcançaremos 84,1%. Isto é algo extremamente significativo, pois ao invés de trabalharmos com 7 variáveis poderíamos trabalhar com o índice PC4, que já explica 84,1% das variáveis., ou em PC5, com 92,8%.

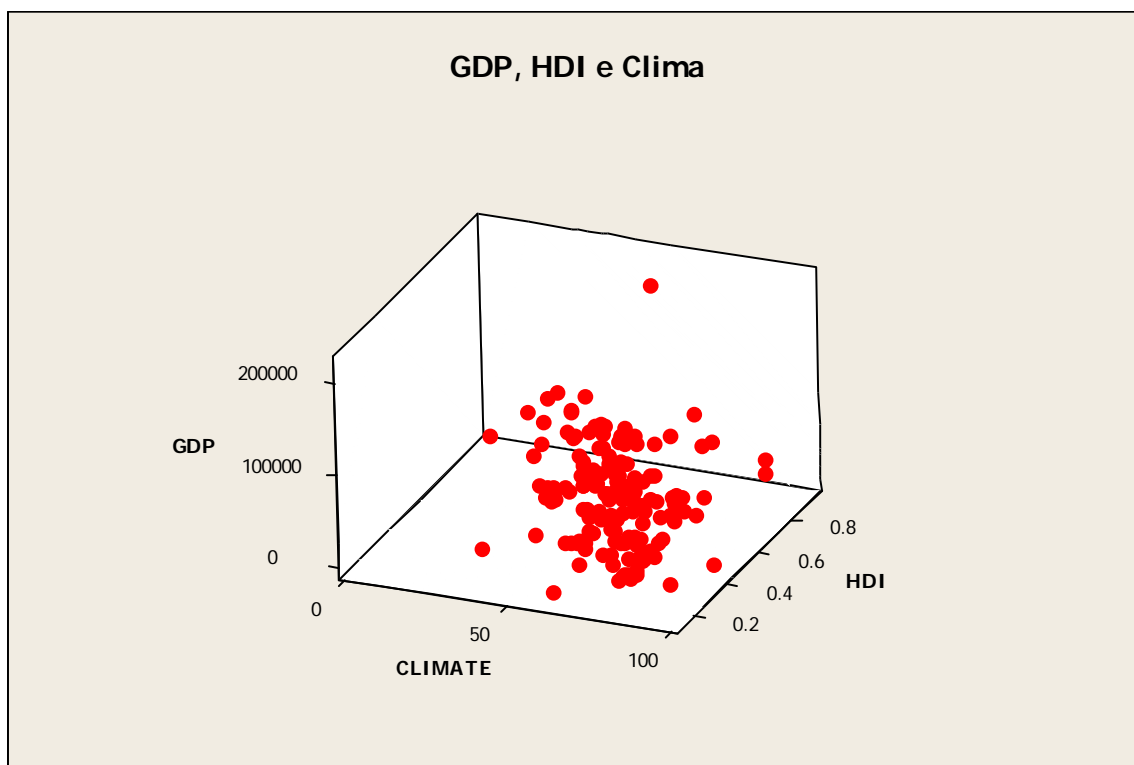
A seguir, gráficos que mostram o comportamento das variáveis mais correlacionadas:

3D Scatterplot of GDP vs HDI vs DALY



Percebe-se um agrupamento dos resultados das variáveis, revelando um comportamento mais próximo a Daly.

3D Scatterplot of GDP vs HDI vs CLIMATE



Percebe-se um agrupamento dos resultados das variáveis, revelando um comportamento mais próximo a HDI.

De acordo com todas as análises acima, percebemos claramente que o agrupamento de variáveis é bastante pertinente no caso dos indicadores obtidos pela pesquisa da universidade de Yale, sobre os indicadores que compõem o EPI.

Isto pôde ser observado inicialmente pelas matrizes de correlação e dendogramas e depois comprovados pela análise dos componentes principais.

Assim, ao invés de trabalharmos com um grupo grande de variáveis (7) poderíamos utilizar apenas quatro ou cinco índices (PC4 ou PC5) que as represente satisfatoriamente.

ESTUDO DOS BRICS EM RELAÇÃO AO EPI

Os BRIC's são compostos pelos países: Brasil, Rússia, Índia, China e África do Sul. Realizaremos um estudo com estes 5 países, que atualmente possuem forte característica em desenvolvimento, conhecidos com países emergente, a fim de verificar o comportamento deles em relação às variáveis que compõem o EPI, bem como seu posicionamento em relação aos demais países.

Seguem os dados dos 5 países a serem comparados nas amostras:

	Country	DALY	AIR_E	WATER_E	BIODIVERSITY	CLIMATE	GDP	HDI
46	Brazil	58.50	39.31	85.63	61.31	46.44	10847	0.699
47	China	62.31	30.19	65.95	57.22	40.18	7206	0.663
48	India	39.35	37.08	68.35	38.65	60.25	3354	0.519
49	Russia	44.18	54.63	84.51	80.26	45.28	12910	0.767
50	South Africa	37.42	30.36	68.10	62.39	39.48	10140	0.597

7. ANÁLISE CORRESPONDENTE DOS BRIC'S

Ao realizarmos uma análise correspondente, verificamos o seguinte gráfico, relacionando os BRIC's com as variáveis. Primeiramente, considerando apenas DALY, Água, Ar, Clima e Biodiversidade.

Simple Correspondence Analysis: DALY, AIR_E, WATER_E, BIODIVERSITY, CLIMATE

Analysis of Contingency Table

Axis	Inertia	Proportion	Cumulative	Histogram
1	0.0124	0.5048	0.5048	*****
2	0.0101	0.4112	0.9160	*****
3	0.0014	0.0587	0.9747	***
4	0.0006	0.0253	1.0000	*
Total	0.0246			

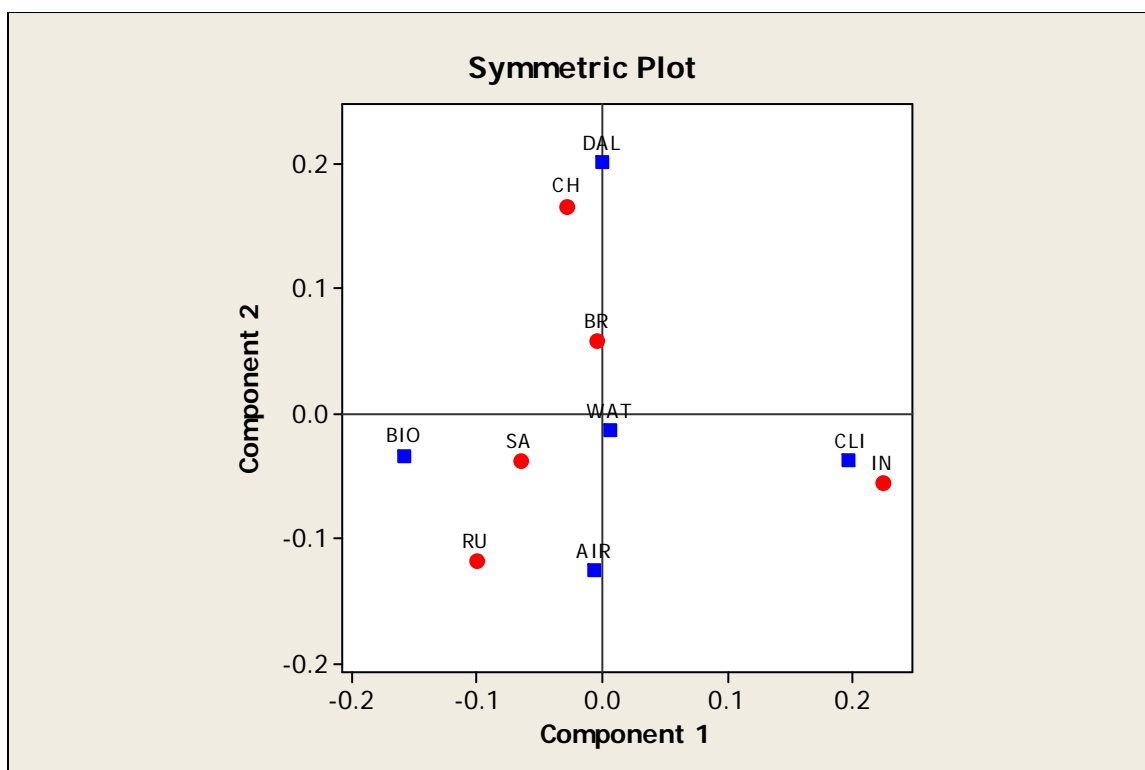
Row Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	BR	0.615	0.218	0.049	-0.004	0.002	0.000	0.058	0.613	0.073
2	CH	0.967	0.191	0.228	-0.028	0.026	0.012	0.166	0.941	0.523
3	IN	0.999	0.182	0.397	0.224	0.941	0.739	-0.056	0.058	0.056
4	RU	0.941	0.231	0.241	-0.100	0.391	0.187	-0.119	0.550	0.322
5	SA	0.494	0.178	0.085	-0.066	0.368	0.062	-0.039	0.126	0.026

Column Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	DAL	0.982	0.181	0.301	0.001	0.000	0.000	0.200	0.982	0.719
2	AIR	0.716	0.143	0.130	-0.006	0.001	0.000	-0.126	0.715	0.225
3	WAT	0.126	0.279	0.022	0.006	0.021	0.001	-0.014	0.105	0.006
4	BIO	0.950	0.224	0.250	-0.158	0.908	0.450	-0.034	0.042	0.026
5	CLI	0.967	0.173	0.297	0.198	0.933	0.549	-0.038	0.034	0.025

Symmetric Plot



A análise da tabela de contingência mostra uma decomposição da inércia (χ^2/n). Do total da inércia da matriz de dados, 50,48% é contabilizada no primeiro componente, 41,12% é contabilizada no segundo componente e assim por diante. O indicador Água se encontra mais perto do primeiro componente, seguido de Biodiversidade e Clima. No segundo componente, se encontra mais próximo Água juntamente com Ar e Daly.

No Symmetric Plot observa-se:

- I. A Índia está próxima do clima, e África do Sul e Rússia próximos a Biodiversidade e ar.
- II. Brasil e África do Sul são os mais próximos de Água e China ficou bem próxima de DALY.

Considerando agora o GINI, HDI e demais variáveis da análise anterior:

Para os valores de GINI, serão considerados: Brasil: 56,4 – Rússia: 37,5 – Índia: 36,8 – China: 41,5; África do Sul: 57,8

Simple Correspondence Analysis: DALY, AIR_E, WATER_E, BIODIVERSITY, CLIMATE, HD

Analysis of Contingency Table

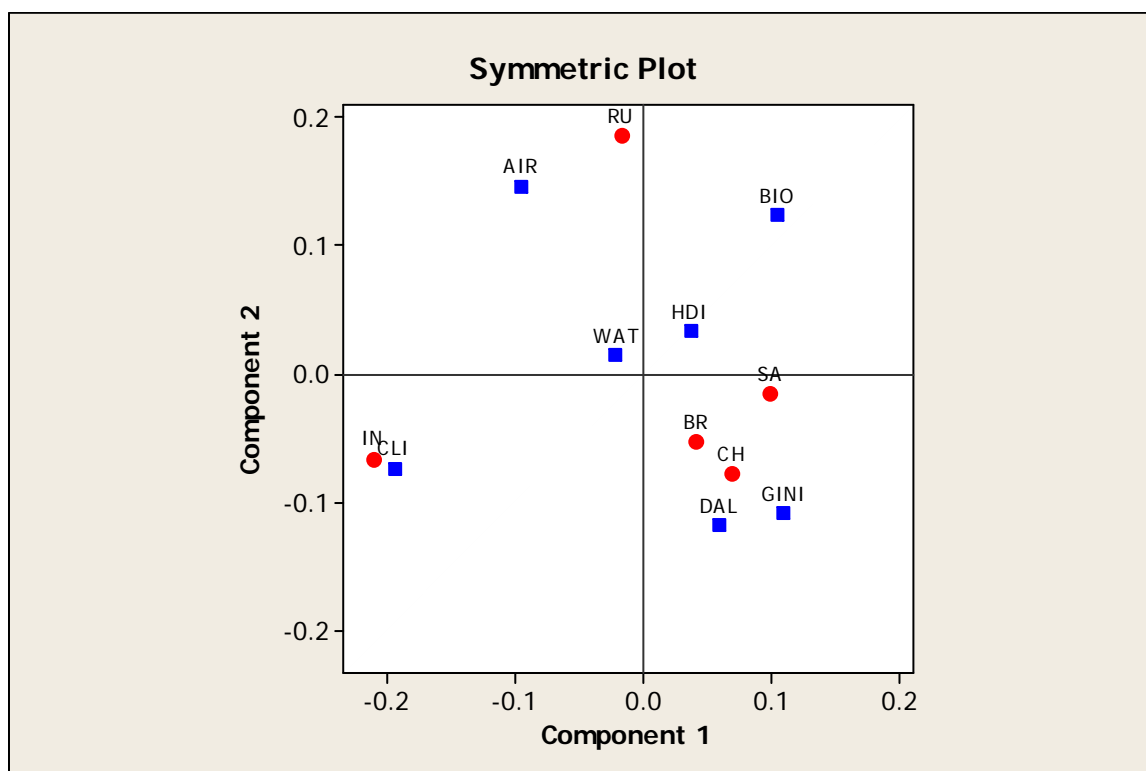
Axis	Inertia	Proportion	Cumulative	Histogram
1	0.0111	0.3938	0.3938	*****
2	0.0102	0.3625	0.7563	*****
3	0.0063	0.2220	0.9782	*****
4	0.0006	0.0218	1.0000	*
Total	0.0283			

Row Contributions

ID	Name	Qual	Mass	Inert	Component 1			Component 2		
					Coord	Corr	Contr	Coord	Corr	Contr
1	BR	0.704	0.222	0.050	0.042	0.270	0.035	-0.053	0.433	0.060
2	CH	0.425	0.190	0.173	0.070	0.189	0.083	-0.078	0.236	0.113
3	IN	0.985	0.179	0.311	-0.210	0.897	0.709	-0.066	0.089	0.076

4	RU	0.979	0.221	0.279	-0.017	0.008	0.006	0.186	0.971	0.747
5	SA	0.363	0.189	0.186	0.100	0.355	0.168	-0.015	0.008	0.004
Column Contributions										
					Component 1			Component 2		
ID	Name	Qual	Mass	Inert	Coord	Corr	Contr	Coord	Corr	Contr
1	DAL	0.431	0.154	0.220	0.059	0.087	0.049	-0.118	0.344	0.209
2	AIR	0.974	0.122	0.135	-0.095	0.291	0.100	0.146	0.682	0.254
3	WAT	0.393	0.237	0.013	-0.021	0.267	0.009	0.014	0.126	0.005
4	BIO	0.964	0.191	0.184	0.104	0.401	0.187	0.124	0.563	0.286
5	CLI	0.956	0.147	0.235	-0.194	0.835	0.498	-0.074	0.121	0.078
6	HDI	0.572	0.002	0.000	0.038	0.324	0.000	0.033	0.248	0.000
7	GINI	0.577	0.146	0.213	0.109	0.290	0.157	-0.108	0.286	0.168

Symmetric Plot



A análise da tabela de contingência mostra uma decomposição da inércia (χ^2/n). Do total da inércia da matriz de dados, 39,38% é contabilizada no primeiro componente, 36,25% é contabilizada no segundo componente, 22,20% no terceiro e assim por diante. Água está mais próxima do primeiro componente, seguido de HDI. Água também se encontra próxima do segundo componente também juntamente com HDI.

No Symmetric Plot observa-se:

- I. A Índia está bem próxima do clima, China próxima a DALY e GINI, seguido por Brasil e África do Sul.
- II. Brasil está mais próximo à Água e HDI.
- III. Rússia se aproxima mais de Ar e logo em seguida, Biodiversidade.

8. CLASSIFICAÇÃO EM TRÊS AMOSTRAS DISTINTAS

Efeturemos três amostras de 50 países a fim de verificar o posicionamento dos BRIC's (Brasil, Rússia, Índia, China e África do Sul) no contexto dos indicadores observados em relação aos demais países.

AMOSTRA 1	AMOSTRA 2	AMOSTRA 3
Sierra Leone	Mongolia	Algeria
Tanzania	Dominican Republic	Djibouti
Belize	Burkina Faso	Sri Lanka
Norway	Thailand	Mauritius
Azerbaijan	Greece	Morocco
Australia	Belgium	New Zealand
Congo	Mexico	Pakistan
Iran	Niger	Mauritania
Switzerland	Benin	Bosnia and Herzegovina
Myanmar	Maldives	Uganda
Saudi Arabia	Zimbabwe	Dominican Republic
Kyrgyzstan	Mauritius	Poland
Slovakia	Italy	Viet Nam
Uganda	Argentina	Kuwait
Malawi	Kyrgyzstan	Turkmenistan
Japan	Egypt	Ghana
Egypt	Tanzania	Mali
Hungary	USA	Nepal
Greece	Bosnia and Herzegovina	Canada
Denmark	Azerbaijan	Jamaica
Senegal	Libyan Arab Jamahiriya	Namibia
Mexico	Spain	Sudan
Rwanda	Chile	Greece
Singapore	Mali	Malta
Israel	Mozambique	Sweden
Namibia	Algeria	Belgium
Niger	Uruguay	South Korea
Sweden	Macedonia	Chad
Chad	Georgia	Bahrain
Macedonia	Malawi	United Kingdom
Turkmenistan	Netherlands	Belarus
Peru	Iran	Malaysia
Georgia	Namibia	Austria
Papua New Guinea	Iceland	Macedonia
Czech Republic	Israel	Kenya
Ecuador	Hungary	Botswana
Guinea	Qatar	Norway
Croatia	Sao Tome and Principe	Philippines
Cyprus	Ethiopia	Peru
Mauritius	Japan	Spain
Turkey	Congo	Laos
Central African Republic	Yemen	Indonesia
Nicaragua	Cameroon	Papua New Guinea
Guatemala	Togo	Uruguay
Uruguay	Zambia	Hungary
Brazil	Brazil	Brazil
China	China	China
India	India	India

Russia	Russia	Russia
South Africa	South Africa	South Africa

AMOSTRA 1

Neste primeiro momento, realizaremos uma análise de conglomerados, para ver como a amostra 1 se divide, a fim de possibilitar o tratamento da amostra, para futura classificação.

Cluster Analysis of Observations: DALY1, AIR_E1, WATER_E1, BIODIVERSITY, ...

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	99.9483	39.1	3	26	3
2	48	99.9110	67.4	12	34	12
3	47	99.8987	76.7	23	37	23
4	46	99.8913	82.4	14	29	14
5	45	99.8776	92.8	1	42	1
6	44	99.8637	103.3	1	15	1
7	43	99.8543	110.4	1	27	1
8	42	99.8061	146.9	1	23	1
9	41	99.8015	150.4	1	14	1
10	40	99.7630	179.5	22	45	22
11	39	99.7602	181.7	7	44	7
12	38	99.7490	190.1	7	33	7
13	37	99.7399	197.0	7	8	7
14	36	99.7380	198.5	40	49	40
15	35	99.7103	219.5	11	35	11
16	34	99.6891	235.5	2	21	2
17	33	99.6795	242.8	12	43	12
18	32	99.6561	260.5	40	41	40
19	31	99.6397	273.0	5	50	5
20	30	99.6360	275.8	2	12	2
21	29	99.5770	320.5	19	25	19
22	28	99.4664	404.3	20	28	20
23	27	99.4381	425.7	31	47	31
24	26	99.3398	500.2	1	2	1
25	25	99.2766	548.1	31	36	31
26	24	99.1790	622.0	3	17	3
27	23	99.1241	663.6	22	40	22
28	22	99.0660	707.6	5	46	5
29	21	99.0456	723.1	1	48	1
30	20	99.0322	733.2	3	31	3
31	19	98.8823	846.8	3	32	3
32	18	98.8720	854.6	3	5	3
33	17	98.8184	895.2	3	7	3
34	16	98.6616	1014.0	10	38	10
35	15	98.6266	1040.6	1	3	1
36	14	98.2594	1318.7	11	39	11
37	13	97.6551	1776.5	10	22	10
38	12	97.5338	1868.5	11	13	11
39	11	97.3518	2006.3	10	18	10
40	10	97.2768	2063.2	1	10	1
41	9	97.2453	2087.1	16	20	16
42	8	96.6269	2555.5	11	19	11
43	7	96.2729	2823.8	6	9	6
44	6	95.5711	3355.5	1	11	1
45	5	94.5256	4147.5	6	16	6
46	4	93.3462	5041.1	1	6	1
47	3	90.5530	7157.4	1	24	1
48	2	89.4247	8012.2	1	4	1
49	1	76.0278	18162.0	1	30	1

Final Partition

Number of clusters: 5

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	42	2874786751	6843.62	19180.5
Cluster2	1	0	0.00	0.0
Cluster3	5	58618522	3131.12	5325.3
Cluster4	1	0	0.00	0.0
Cluster5	1	0	0.00	0.0

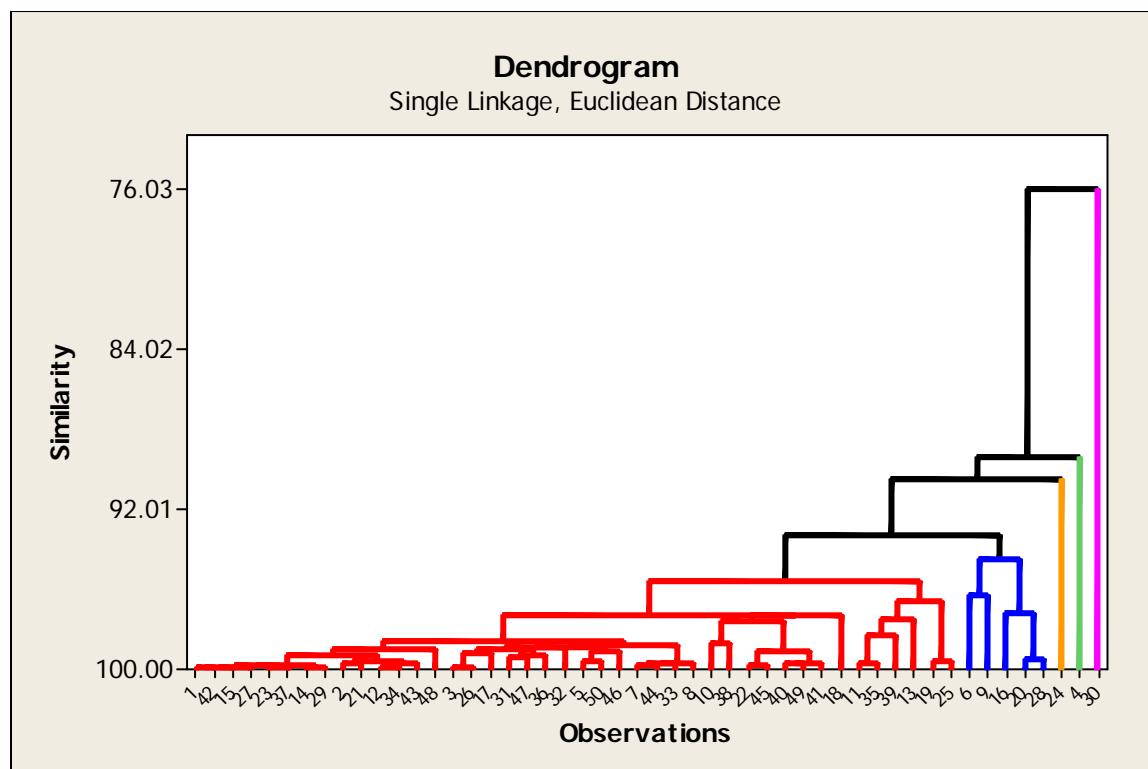
Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Grand centroid
DALY1	48.23	82.8	85.7	89.1	69.0	53.9
AIR_E1	49.23	58.1	43.9	10.7	69.8	48.5
WATER_E1	66.52	97.5	82.1	99.0	79.8	69.6
BIODIVERSITY1	60.61	46.6	71.0	44.2	40.2	60.6
CLIMATE1	54.37	65.7	53.2	48.6	36.5	54.0
GDP1	9427.52	58278.0	37783.8	50266.0	76440.0	15397.2
HDI1	0.61	0.9	0.9	0.8	0.9	0.7

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	0.0	48850.5	28356.3	40838.5	67012.5
Cluster2	48850.5	0.0	20494.2	8012.2	18162.0
Cluster3	28356.3	20494.2	0.0	12482.3	38656.2
Cluster4	40838.5	8012.2	12482.3	0.0	26174.1
Cluster5	67012.5	18162.0	38656.2	26174.1	0.0

Dendrogram



Países Outliers: 4-Noruega; 24-Singapura; 30-Macedônia. Todos os Brics localizados no cluster 1.

Faremos agora uma nova análise com o tratamento dos dados.

NOVA ANÁLISE DE CONGLOMERADOS

Cluster Analysis of Observations: DALY1, AIR_E1, WATER_E1, BIODIVERSITY, ...

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	46	99.9078	39.13	3	24	3
2	45	99.8410	67.45	11	31	11
3	44	99.8192	76.71	22	34	22
4	43	99.8059	82.38	13	27	13
5	42	99.7814	92.75	1	39	1
6	41	99.7566	103.30	1	14	1
7	40	99.7398	110.40	1	25	1
8	39	99.6538	146.88	1	22	1
9	38	99.6456	150.37	1	13	1
10	37	99.5769	179.54	21	42	21
11	36	99.5718	181.70	6	41	6
12	35	99.5519	190.14	6	30	6
13	34	99.5356	197.04	6	7	6
14	33	99.5322	198.48	37	46	37
15	32	99.4828	219.46	10	32	10
16	31	99.4449	235.53	2	20	2
17	30	99.4278	242.79	11	40	11
18	29	99.3860	260.52	37	38	37
19	28	99.3566	273.01	4	47	4
20	27	99.3501	275.75	2	11	2
21	26	99.2447	320.48	18	23	18
22	25	99.0472	404.30	19	26	19
23	24	98.9967	425.72	28	44	28
24	23	98.8212	500.20	1	2	1
25	22	98.7083	548.10	28	33	28
26	21	98.5342	621.98	3	16	3
27	20	98.4361	663.61	21	37	21
28	19	98.3323	707.62	4	43	4
29	18	98.2958	723.12	1	45	1
30	17	98.2720	733.22	3	28	3
31	16	98.0044	846.79	3	29	3
32	15	97.9859	854.61	3	4	3
33	14	97.8903	895.21	3	6	3
34	13	97.6102	1014.04	9	35	9
35	12	97.5477	1040.56	1	3	1
36	11	96.8922	1318.71	10	36	10
37	10	95.8132	1776.54	9	21	9
38	9	95.5966	1868.47	10	12	10
39	8	95.2716	2006.34	9	17	9
40	7	95.1377	2063.19	1	9	1
41	6	95.0814	2087.08	15	19	15
42	5	93.9773	2555.54	10	18	10
43	4	93.3453	2823.75	5	8	5
44	3	92.0921	3355.49	1	10	1
45	2	90.2255	4147.54	5	15	5
46	1	88.1196	5041.10	1	5	1

Final Partition

Number of clusters: 5

Cluster	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	36	1013467492	4523.40	12254.1
Cluster2	1	0	0.00	0.0
Cluster3	1	0	0.00	0.0
Cluster4	6	30275793	1945.80	3260.1
Cluster5	3	3573578	1017.27	1525.7

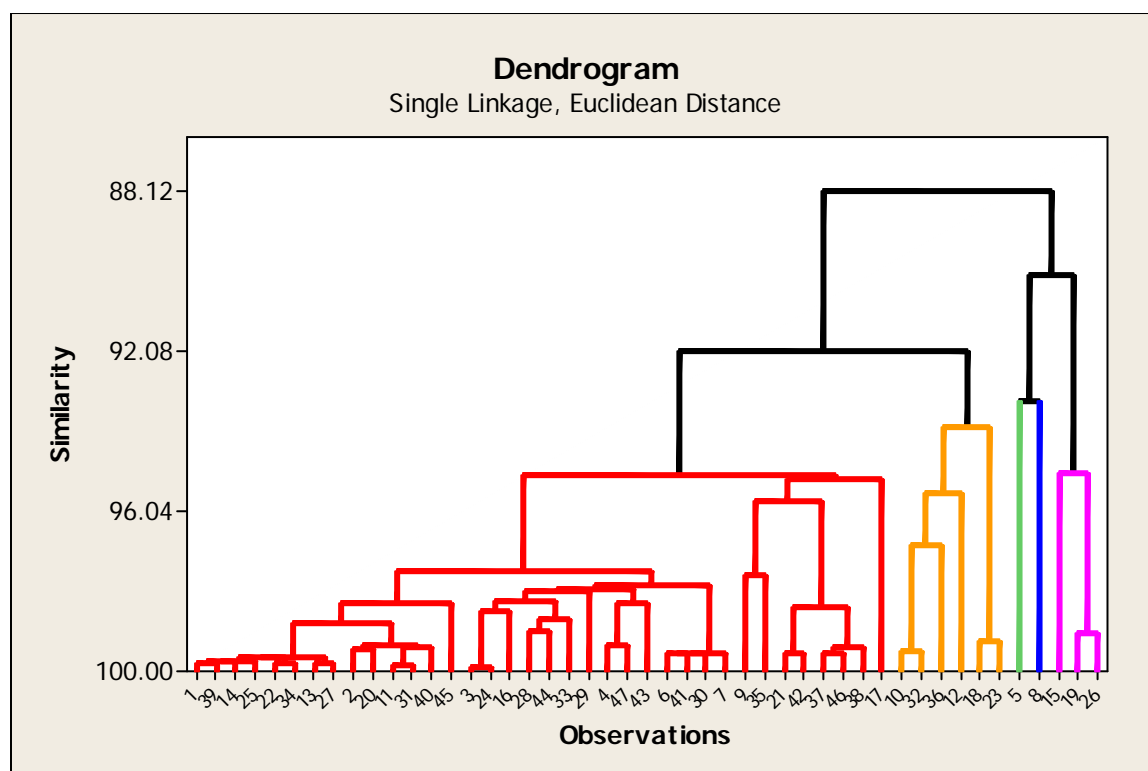
Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Grand centroid
DALY1	43.16	84.8	89.1	78.6	84.9	52.2
AIR_E1	50.05	29.5	47.8	44.3	47.4	48.7
WATER_E1	66.39	58.0	93.5	67.3	86.4	68.2
BIODIVERSITY1	57.66	77.9	100.0	78.3	59.0	61.7
CLIMATE1	57.84	27.6	73.8	33.6	54.8	54.2
GDP1	6731.97	40286.0	43109.0	25600.8	35174.7	12444.1
HDI1	0.57	0.9	0.9	0.8	0.9	0.6

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	0.0	33554.1	36377.1	18868.9	28442.7
Cluster2	33554.1	0.0	2823.8	14685.2	5111.6
Cluster3	36377.1	2823.8	0.0	17508.2	7934.5
Cluster4	18868.9	14685.2	17508.2	0.0	9573.9
Cluster5	28442.7	5111.6	7934.5	9573.9	0.0

Dendrogram

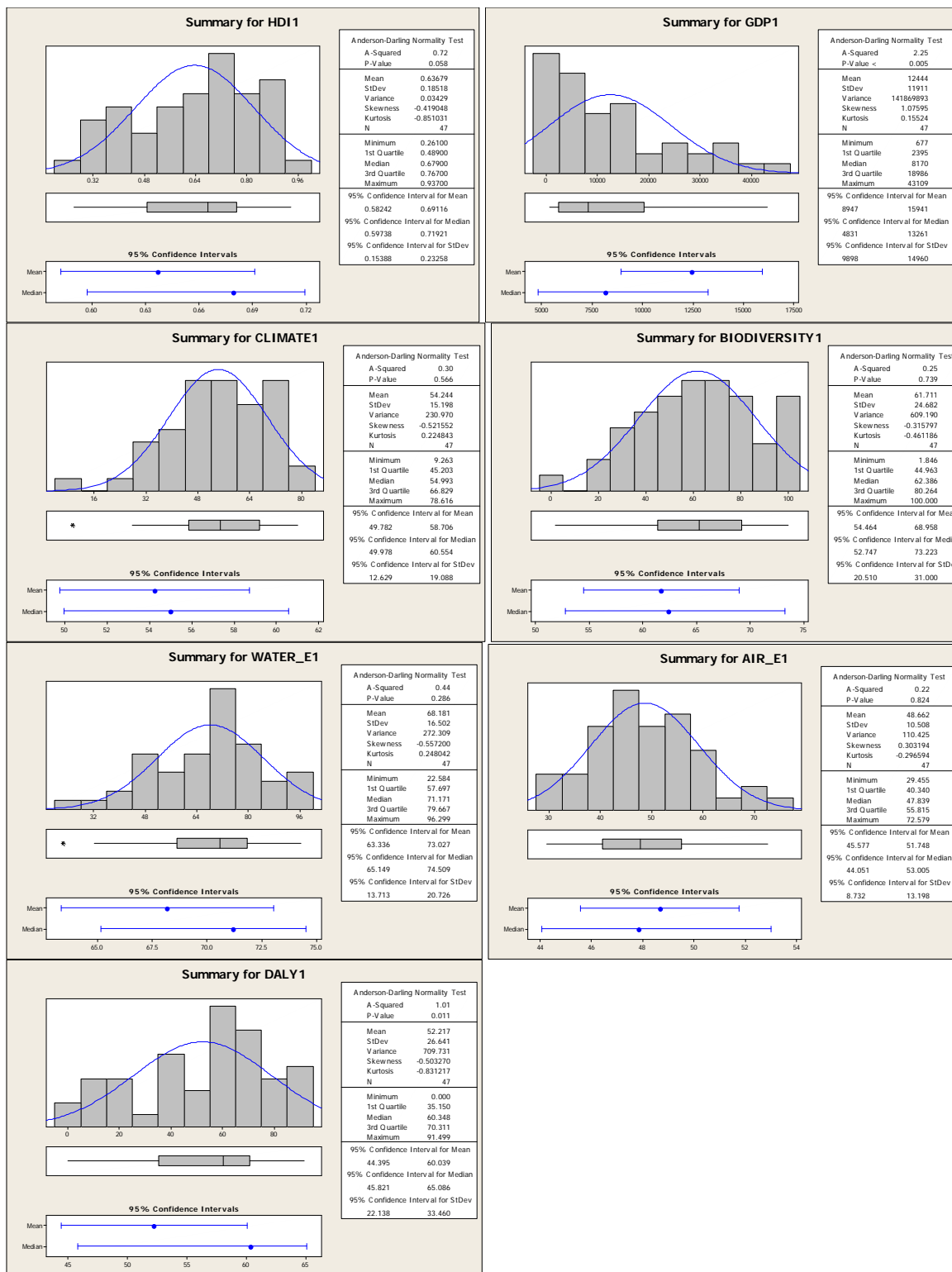


Como as menores distâncias foram obtidas entre os clusters 2, 3, 4 e 5, reuniremos todos em apenas 1 cluster.

Na análise cluster x cluster observamos que a maior distância entre os centroides se dá entre o cluster 1 e 3, e a menor, entre o cluster 2 e 3.

O grande centroide do indicador DALY foi 52,2, AR 48,7, ÁGUA 68,2, CLIMA 54,2, BIODIVERSIDADE 61,7, HDI 0.6 e GDP 12444,1. No cluster 1 obtivemos 36 observações. OS Bric's estão todos localizados no cluster 1.

Primeiramente realizaremos uma análise exploratória da amostra 1, com a base tratada:



Baseado na análise do One-Way ANOVA, é possível identificar as variáveis com medianas mais distantes, que melhor servirão como base na classificação da amostra 1.

One-way ANOVA: DALY1 versus C10

Source	DF	SS	MS	F	P
C10	1	12614	12614	28.33	0.000
Error	45	20033	445		
Total	46	32648			

S = 21.10 R-Sq = 38.64% R-Sq(adj) = 37.27%

Individual 95% CIs For Mean Based on Pooled StDev

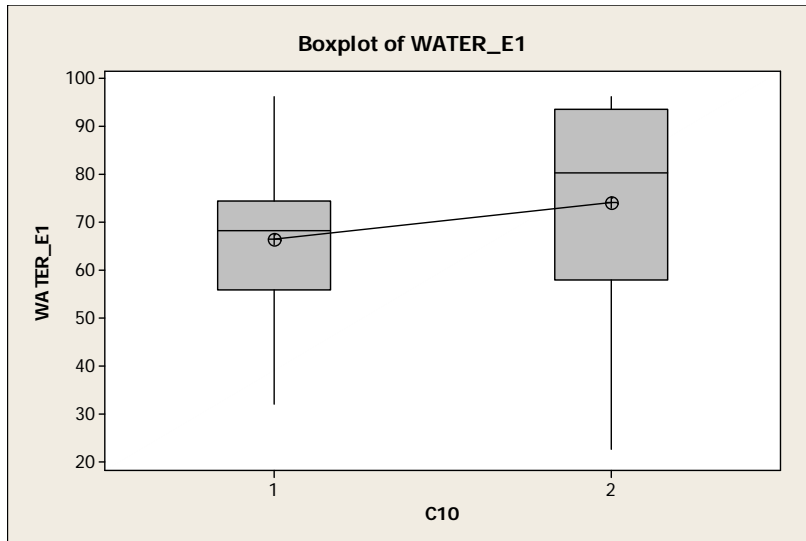
Level	N	Mean	StDev	-----+-----+-----+-----+-----+-----+-----
-------	---	------	-------	---

Total 46 12526
 S = 16.35 R-Sq = 3.96% R-Sq(adj) = 1.83%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
1	36	66.39	13.51
2	11	74.06	23.75

Pooled StDev = 16.35



One-way ANOVA: BIODIVERSITY1 versus C10

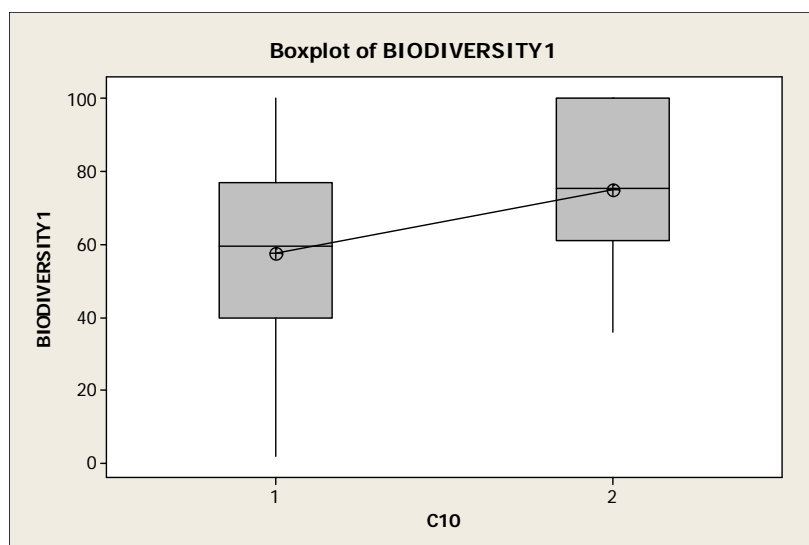
Source	DF	SS	MS	F	P
C10	1	2527	2527	4.46	0.040
Error	45	25496	567		
Total	46	28023			

S = 23.80 R-Sq = 9.02% R-Sq(adj) = 7.00%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
1	36	57.66	24.55
2	11	74.98	20.97

Pooled StDev = 23.80



One-way ANOVA: CLIMATE1 versus C10

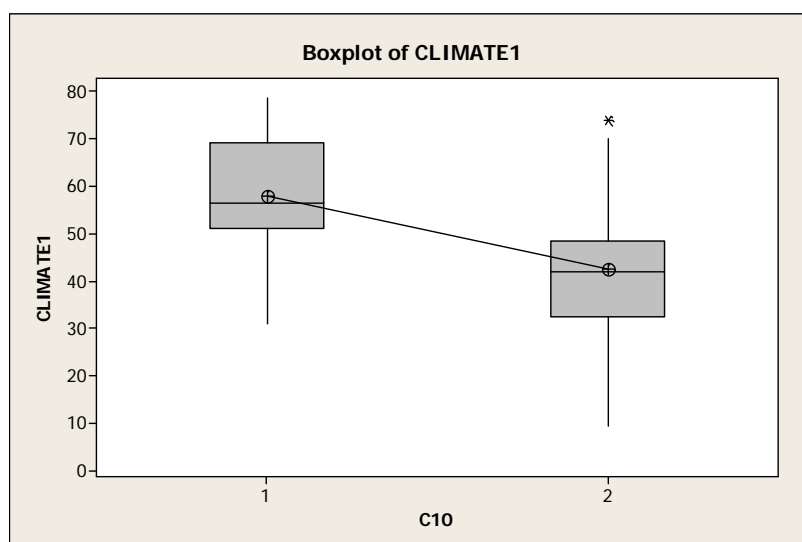
Source	DF	SS	MS	F	P
C10	1	1986	1986	10.35	0.002
Error	45	8638	192		
Total	46	10625			

S = 13.86 R-Sq = 18.70% R-Sq(adj) = 16.89%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
1	36	57.84	12.31	40.0	64.0
2	11	42.48	18.26	24.0	61.0

Pooled StDev = 13.86



One-way ANOVA: GDP1 versus C10

Source	DF	SS	MS	F	P
C10	1	5018927734	5018927734	149.86	0.000
Error	45	1507087336	33490830		
Total	46	6526015070			

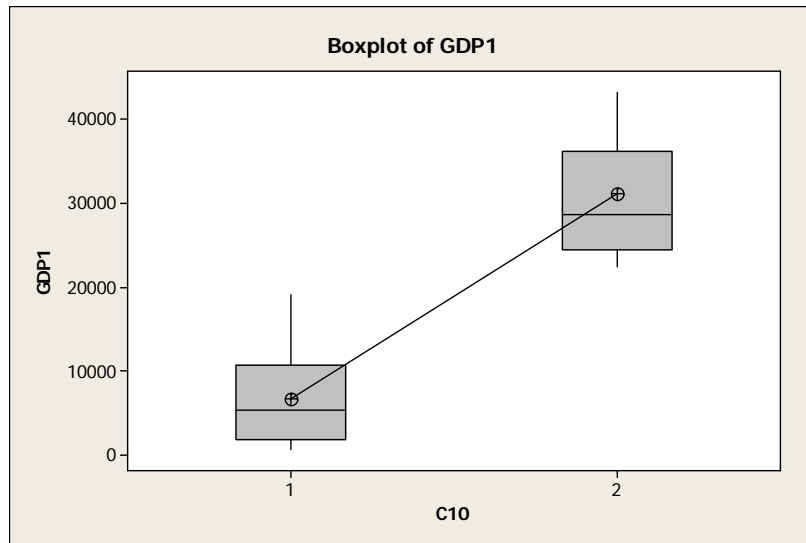
S = 5787 R-Sq = 76.91% R-Sq(adj) = 76.39%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	
1	36	6732	5381	(---*---)
2	11	31139	7026	(-----*-----)

8000 16000 24000 32000

Pooled StDev = 5787



One-way ANOVA: HDI1 versus C10

Source	DF	SS	MS	F	P
C10	1	0.6776	0.6776	33.89	0.000
Error	45	0.8998	0.0200		
Total	46	1.5774			

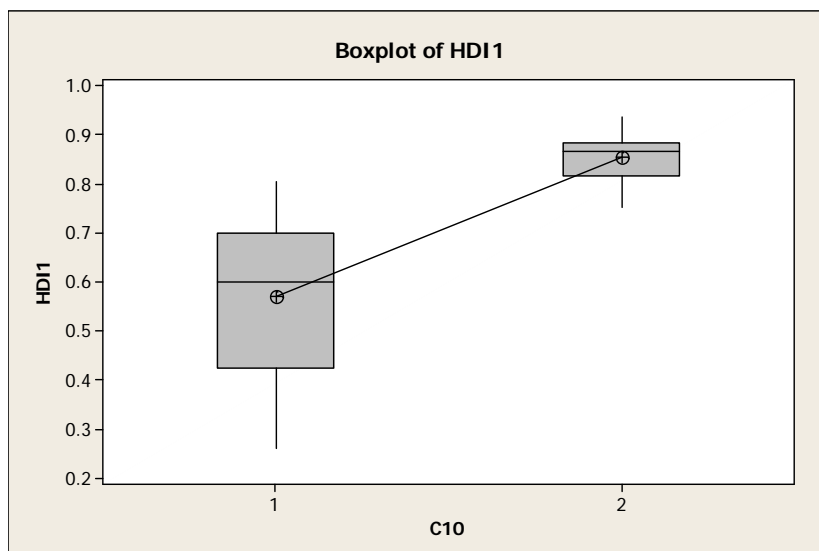
S = 0.1414 R-Sq = 42.96% R-Sq(adj) = 41.69%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	
1	36	0.5704	0.1582	(---*---)
2	11	0.8540	0.0484	(-----*-----)

0.60 0.72 0.84 0.96

Pooled StDev = 0.1414



Com P-value maior que 0,05, com intervalo de confiança de 95%, conclui-se que as médias de ÁGUA E AR podem ser as mesmas em relação aos clusters. A seguir, veremos a análise do 2-Sample t:

Two-Sample T-Test and CI: DALY1, GDP1

Two-sample T for DALY1 vs GDP1

	N	Mean	StDev	SE Mean
DALY1	47	52.2	26.6	3.9
GDP1	47	12444	11911	1737

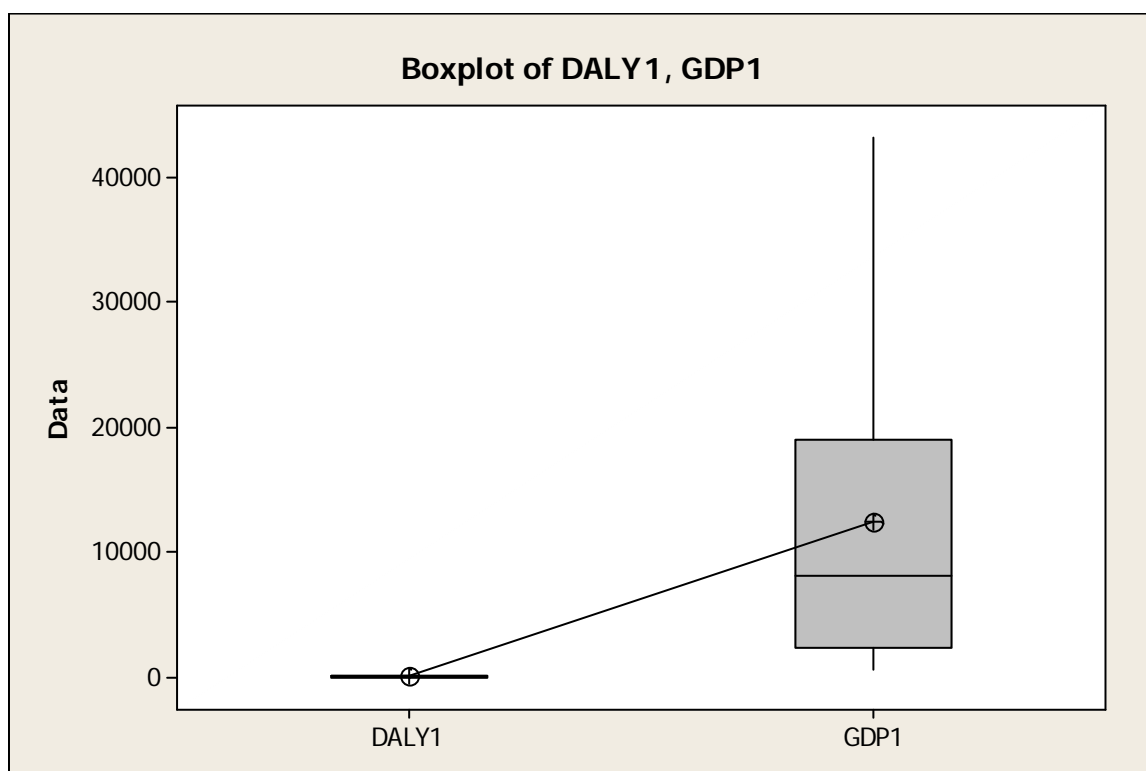
Difference = mu (DALY1) - mu (GDP1)

Estimate for difference: -12392

95% CI for difference: (-15889, -8895)

T-Test of difference = 0 (vs not =): T-Value = -7.13 P-Value = 0.000 DF = 46

Boxplot of DALY1, GDP1



O P-value igual a zero demonstra que a chance de serem iguais é igual a zero.

Os melhores indicadores pela análise de ANOVA ONE WAY foram Daly, GDP e HDI, sobre os quais realizaremos A Análise de Classificação através dos modelos: Análise Discriminante, Regressão Logística e Árvores de Classificação.

9a. ANÁLISE DISCRIMINANTE – AMOSTRA 1

Considerando primeiramente todas as variáveis:

Discriminant Analysis: C10 versus DALY1, AIR_E1, ...

Linear Method for Response: C10

Predictors: DALY1, AIR_E1, WATER_E1, BIODIVERSITY1, CLIMATE1, GDP1, HDI1

Group	1	2
Count	36	11

Summary of classification

Put into Group	True Group	
	1	2
1	36	0
2	0	11
Total N	36	11
N correct	36	11
Proportion	1.000	1.000

N = 47 N Correct = 47

Proportion Correct = 1.000

Squared Distance Between Groups

	1	2
1	0.0000	27.8231
2	27.8231	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-51.614	-60.886
DALY1	-0.189	-0.116
AIR_E1	0.483	0.595
WATER_E1	0.225	0.172
BIODIVERSITY1	0.113	0.232
CLIMATE1	0.346	0.187
GDP1	-0.001	0.000
HDI1	88.953	61.543

Através da análise linear, considerando todas as variáveis, foi possível obter um acerto de 100%. Atendendo ao critério parcimonioso, veremos quais variáveis é possível obter um maior número de acerto.

Discriminant Analysis: C10 versus DALY1, GDP1, HDI1

Linear Method for Response: C10

Predictors: DALY1, GDP1, HDI1

Group	1	2
Count	36	11

Summary of classification

Put into Group	True Group	
	1	2
1	35	0
2	1	11
Total N	36	11
N correct	35	11
Proportion	0.972	1.000

N = 47 N Correct = 46

Proportion Correct = 0.979

Squared Distance Between Groups

	1	2
1	0.0000	19.4026
2	19.4026	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-16.563	-26.011
DALY1	-0.368	-0.384
GDP1	-0.000	0.000
HDI1	91.749	81.375

Summary of Misclassified Observations

Observation	True Group	Pred Group	Squared Distance	Probability
9**	1	2	10.78	0.426
			10.19	0.574

Miamar

Através da análise linear, estas variáveis se mostraram melhor ajustadas, já que alcançaram 97,9% de acerto. Veremos com menos variáveis:

Discriminant Analysis: C10 versus DALY1, GDP1

Linear Method for Response: C10

Predictors: DALY1, GDP1

Group	1	2
Count	36	11

Summary of classification

Put into Group	True Group	
	1	2
1	36	0
2	0	11
Total N	36	11
N correct	36	11
Proportion	1.000	1.000

N = 47 N Correct = 47

Proportion Correct = 1.000

Squared Distance Between Groups

	1	2
1	0.0000	19.0330
2	19.0330	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-2.109	-14.642
DALY1	0.104	0.035
GDP1	-0.000	0.001

Apenas com DALY e GDP, também foi possível obter 100% de acerto neste modelo. Veremos agora com apenas 1 variável, se este modelo fica mais perfeito.

Discriminant Analysis: C10 versus DALY1

Linear Method for Response: C10

Predictors: DALY1

Group	1	2
Count	36	11

Summary of classification

	True Group	
Put into Group	1	2
1	29	0
2	7	11
Total N	36	11
N correct	29	11
Proportion	0.806	1.000

N = 47 N Correct = 40

Proportion Correct = 0.851

Squared Distance Between Groups

	1	2
1	0.00000	3.36297
2	3.36297	0.00000

Linear Discriminant Function for Groups

	1	2
Constant	-2.0922	-7.5250
DALY1	0.0970	0.1839

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability	
17**	1	2	1	1.2382	0.411	Hungria
			2	0.5200	0.589	
21**	1	2	1	2.0017	0.286	México
			2	0.1756	0.714	
30**	1	2	1	1.5044	0.362	Georgia
			2	0.3688	0.638	
35**	1	2	1	2.1995	0.261	Croácia
			2	0.1230	0.739	
37**	1	2	1	1.6558	0.337	Mauricius
			2	0.2993	0.663	
38**	1	2	1	1.1212	0.435	Turquia
			2	0.6006	0.565	
42**	1	2	1	1.6558	0.337	Uruguai
			2	0.2993	0.663	

Considerando apenas o DALY, 85,1% no modelo linear.

Discriminant Analysis: C10 versus GDP1

Linear Method for Response: C10

Predictors: GDP1

Group	1	2
Count	36	11

Summary of classification

Put into Group	True Group	
	1	2
1	35	0
2	1	11
Total N	36	11
N correct	35	11
Proportion	0.972	1.000

N = 47 N Correct = 46

Proportion Correct = 0.979

Squared Distance Between Groups

	1	2
1	0.0000	17.7864
2	17.7864	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-0.677	-14.476
GDP1	0.000	0.001

Summary of Misclassified Observations

Observation	True Group	Pred Group	Squared Distance	Probability
17**	1	2	4.484	0.491
			4.410	0.509

Considerando apenas o GDP, obtivemos 97,9% de acerto, no modelo linear.

O melhor modelo de análise discriminante ocorreu considerando GDP e DALY, através da função linear, onde foi possível identificar 100% de acerto, considerando o melhor ajuste que atenda o critério parcimonioso.

10a. REGRESSÃO LOGÍSTICA – AMOSTRA 1

Não foi possível realizar a Análise de Regressão Logística neste modelo.

Binary Logistic Regression: C10 versus DALY1, GDP1

* 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
C10	2	11 (Event)
	1	36
Total		47

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds	95% CI	
					Ratio	Lower	Upper

Constant	-145.648	13541.5	-0.01	0.991				
DALY1	-0.967789	275.377	-0.00	0.997	0.38	0.00	9.57316E+233	
GDP1	0.0102572	0.741927	0.01	0.989	1.01	0.24	4.32	

Log-Likelihood = -0.000

Test that all slopes are zero: G = 51.147, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0.0000004	44	1.000
Deviance	0.0000008	44	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	0	0	0	0	0	1	5	5	11	
Exp	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	5.0	5.0		
1												
Obs	4	5	5	4	5	5	4	4	0	0	36	
Exp	4.0	5.0	5.0	4.0	5.0	5.0	4.0	4.0	0.0	0.0		
Total	4	5	5	4	5	5	4	5	5	5	47	

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	396	100.0	Somers' D 1.00
Discordant	0	0.0	Goodman-Kruskal Gamma 1.00
Ties	0	0.0	Kendall's Tau-a 0.37
Total	396	100.0	

11a. ÁRVORES DE CLASSIFICAÇÃO – AMOSTRA 1

Estatísticas descritivas:								
Variável	Categorias	Frequências		%				
nova	1		36					76.596
	2		11					23.404
Variável	Observações	Obs. com dados faltantes	Obs. sem dados faltantes	Mínimo	Máximo	Média	Desvio padrão	
DALY1	47	0	47	0.000	91.499	52.217	26.641	
AIR_E1	47	0	47	29.455	72.579	48.662	10.508	
WATER_E1	47	0	47	22.584	96.299	68.181	16.502	
BIODIVERSITY1	47	0	47	1.846	100.000	61.711	24.682	
CLIMATE1	47	0	47	9.263	78.616	54.244	15.198	
GDP1	47	0	47	677.000	43109.000	12444.149	11910.915	
HDI1	47	0	47	0.261	0.937	0.637	0.185	

Matriz de correlação:

Variáveis	DALY1	AIR_E1	WATER_E1	BIODIVERSITY1	CLIMATE1	GDP1	HDI1
DALY1	1.000	-0.245	0.210	-0.089	-0.462	0.785	0.938
AIR_E1	-0.245	1.000	-0.134	-0.247	0.277	-0.284	-0.240
WATER_E1	0.210	-0.134	1.000	0.135	-0.036	0.293	0.240
BIODIVERSITY1	-0.089	-0.247	0.135	1.000	0.022	0.145	0.023
CLIMATE1	-0.462	0.277	-0.036	0.022	1.000	-0.395	-0.475
GDP1	0.785	-0.284	0.293	0.145	-0.395	1.000	0.826
HDI1	0.938	-0.240	0.240	0.023	-0.475	0.826	1.000

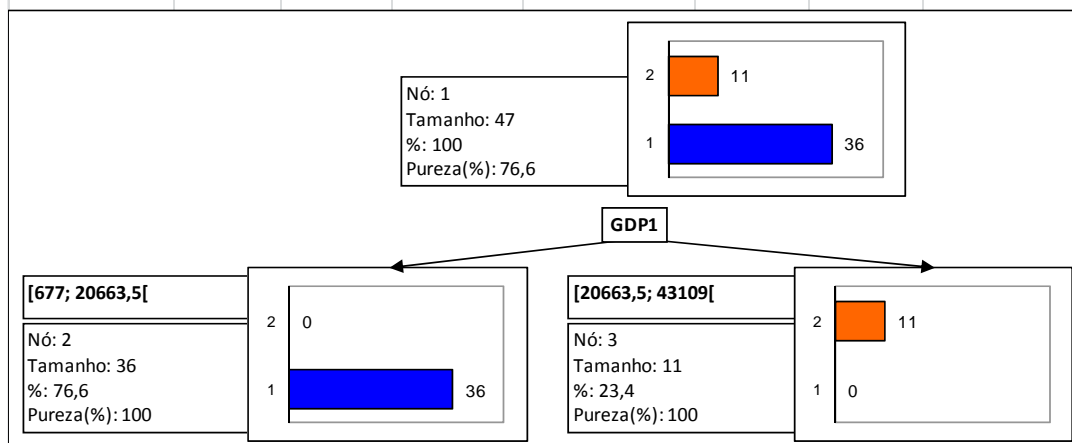
Estrutura da árvore:

Nó	p-valor	Objetos	%	Nó pai	Filhos	Vel de sepa	Valores	Pureza
1	1.000	47	100.00%		2; 3			76.60%
2	0.000	36	76.60%	1		GDP1	[677; 20663,5[100.00%
3	0.000	11	23.40%	1		GDP1	[20663,5; 43109[100.00%

Legenda:

2	
1	

Árvore de classificação:



Réguas:

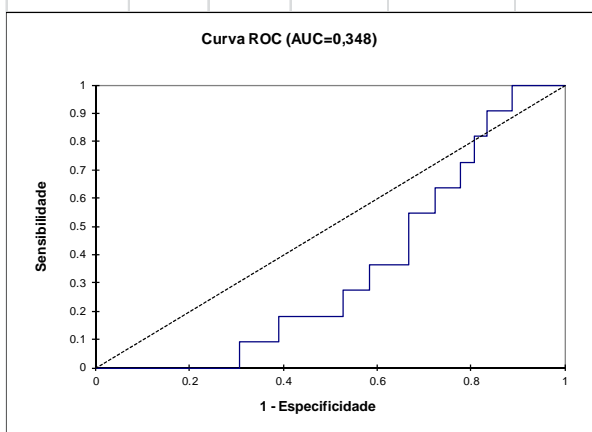
Nó	Pred(nova)	Frequência	Pureza	Réguas
Nó1	1.000	36	76.60%	
Nó2	1.000	36	100.00%	Se GDP1 em [677; 20663,5[então nova = 1 em 100% dos casos
Nó3	2.000	11	100.00%	Se GDP1 em [20663,5; 43109[então nova = 2 em 100% dos casos

Resultados por objeto:				
Observação	A priori	\ posterior	Pr(1)	Pr(2)
Obs1	1	1	1.000	0.000
Obs2	1	1	1.000	0.000
Obs3	1	1	1.000	0.000
Obs4	1	1	1.000	0.000
Obs5	2	2	0.000	1.000
Obs6	1	1	1.000	0.000
Obs7	1	1	1.000	0.000
Obs8	2	2	0.000	1.000
Obs9	1	1	1.000	0.000
Obs10	2	2	0.000	1.000
Obs11	1	1	1.000	0.000
Obs12	2	2	0.000	1.000
Obs13	1	1	1.000	0.000
Obs14	1	1	1.000	0.000
Obs15	2	2	0.000	1.000
Obs16	1	1	1.000	0.000
Obs17	1	1	1.000	0.000
Obs18	2	2	0.000	1.000
Obs19	2	2	0.000	1.000
Obs20	1	1	1.000	0.000
Obs21	1	1	1.000	0.000
Obs22	1	1	1.000	0.000
Obs23	2	2	0.000	1.000
Obs24	1	1	1.000	0.000
Obs25	1	1	1.000	0.000
Obs26	2	2	0.000	1.000
Obs27	1	1	1.000	0.000
Obs28	1	1	1.000	0.000
Obs29	1	1	1.000	0.000
Obs30	1	1	1.000	0.000
Obs31	1	1	1.000	0.000
Obs32	2	2	0.000	1.000
Obs33	1	1	1.000	0.000
Obs34	1	1	1.000	0.000
Obs35	1	1	1.000	0.000
Obs36	2	2	0.000	1.000
Obs37	1	1	1.000	0.000
Obs38	1	1	1.000	0.000
Obs39	1	1	1.000	0.000
Obs40	1	1	1.000	0.000
Obs41	1	1	1.000	0.000
Obs42	1	1	1.000	0.000
Obs43	1	1	1.000	0.000
Obs44	1	1	1.000	0.000
Obs45	1	1	1.000	0.000
Obs46	1	1	1.000	0.000
Obs47	1	1	1.000	0.000

Matriz de confusão para a amostra de estimação:

de \ a	1	2	Total	% correto
1	36	0	36	100.00%
2	0	11	11	100.00%
Total	36	11	47	100.00%

Tabela de classificação para a amostra de validação:



Área sob a curva: 0.348

Realizada a árvore de classificação, foi possível observar que nesta amostra 1, tanto pelo aplicativo Minitab (pela Análise Discriminante, equação quadrática, pois a Regressão Logística não deu certo) quanto pelo aplicativo XLSTAT (Árvore de classificação e Regressão), a variável que apresenta maior importância na separação dos grupos foi o GDP, sendo que na análise discriminante, o melhor modelo ocorreu juntamente com o indicador DALY.

AMOSTRA 2

Neste primeiro momento, realizaremos uma análise de conglomerados, para ver como a amostra 2 se divide, a fim de possibilitar o tratamento da amostra, para futura classificação.

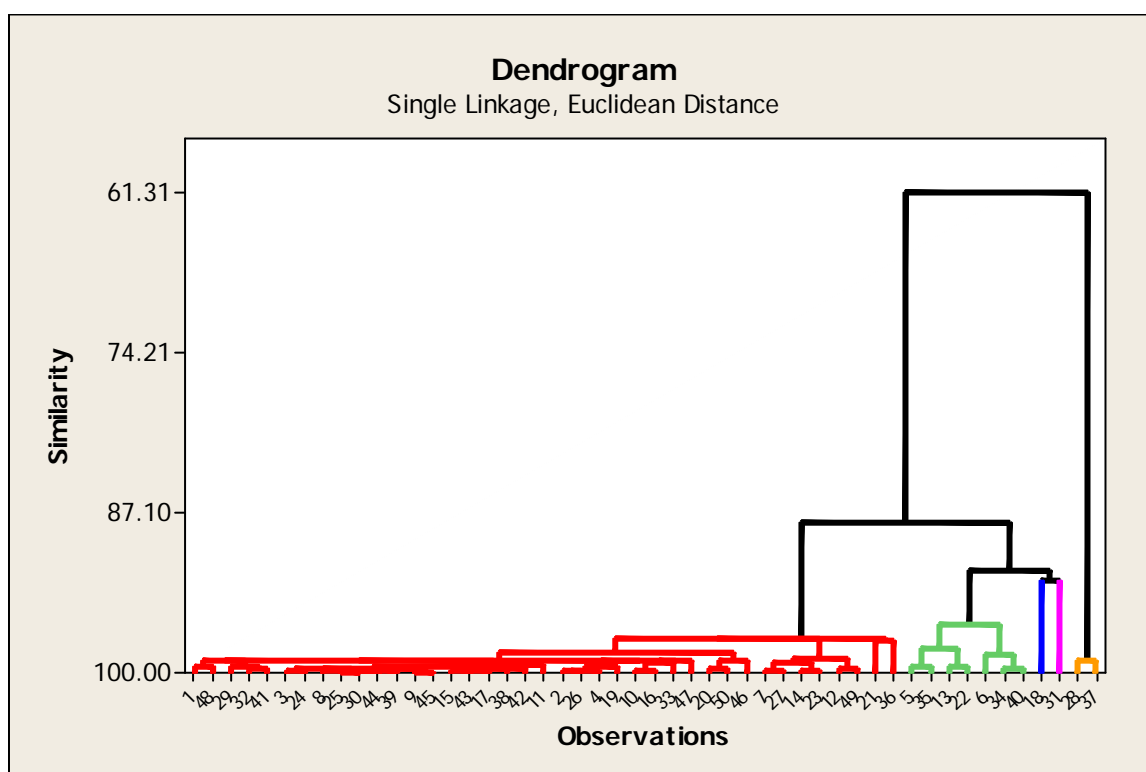
Cluster Analysis of Observations: DALY2, AIR_E2, WATER_E2, BIODIVERSITY, ...

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	99.9581	32.2	9	45	9	2
2	48	99.9347	50.2	25	30	25	2
3	47	99.9096	69.6	25	44	25	3
4	46	99.9064	72.0	25	39	25	4
5	45	99.8992	77.6	3	24	3	2
6	44	99.8329	128.6	15	43	15	2
7	43	99.8231	136.2	10	16	10	2
8	42	99.8133	143.7	2	26	2	2
9	41	99.8017	152.6	14	23	14	2
10	40	99.7999	154.1	2	4	2	3
11	39	99.7993	154.5	15	17	15	3
12	38	99.7738	174.2	8	25	8	5
13	37	99.7668	179.5	7	27	7	2
14	36	99.7441	197.0	32	41	32	2
15	35	99.7422	198.5	12	49	12	2
16	34	99.7308	207.2	15	38	15	4
17	33	99.7060	226.4	3	8	3	7
18	32	99.6995	231.4	34	40	34	2
19	31	99.6489	270.3	15	42	15	5
20	30	99.6454	273.0	20	50	20	2
21	29	99.6349	281.1	3	9	3	9
22	28	99.5837	320.5	5	35	5	2
23	27	99.5314	360.7	1	48	1	2
24	26	99.5193	370.1	29	32	29	3
25	25	99.5157	372.9	2	19	2	4
26	24	99.5097	377.5	3	15	3	14
27	23	99.4865	395.3	13	22	13	2
28	22	99.3611	491.9	3	11	3	15
29	21	99.2361	588.1	7	14	7	4
30	20	99.1761	634.3	10	33	10	3
31	19	99.1111	684.4	1	29	1	5
32	18	99.0809	707.6	20	46	20	3
33	17	99.0477	733.2	10	47	10	4
34	16	99.0351	742.9	28	37	28	2
35	15	99.0120	760.6	2	10	2	8
36	14	99.0085	763.4	1	3	1	20
37	13	98.9912	776.7	1	2	1	28
38	12	98.8018	922.5	7	12	7	6
39	11	98.5890	1086.3	6	34	6	3
40	10	98.3710	1254.1	1	20	1	31
41	9	98.0874	1472.6	5	13	5	4
42	8	97.4180	1987.9	21	36	21	2
43	7	97.3202	2063.2	1	7	1	37
44	6	97.3131	2068.6	1	21	1	39
45	5	96.1717	2947.5	5	6	5	7
46	4	92.6627	5649.0	18	31	18	2
47	3	91.8558	6270.3	5	18	5	9
48	2	87.9127	9306.1	1	5	1	48

49	1	61.3111	29787.0	1	28	1	50
Final Partition							
Number of clusters: 5							
				Average	Maximum		
	Number of	Within cluster		distance	distance		
	observations	sum of squares		from	from		
				centroid	centroid		
Cluster1	39	1089618942		4525.88	12538.9		
Cluster2	7	40280488		2239.19	3411.4		
Cluster3	1	0		0.00	0.0		
Cluster4	2	275915		371.43	371.4		
Cluster5	1	0		0.00	0.0		
Cluster Centroids							
Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Grand	centroid
DALY2	43.34	86.5	79.2	79.1	86.9		52.4
AIR_E2	48.22	35.5	31.6	52.4	38.6		46.1
WATER_E2	62.99	71.2	70.2	47.0	67.4		63.7
BIODIVERSITY2	54.69	55.7	65.9	23.2	80.1		54.3
CLIMATE2	60.75	47.7	29.4	31.9	39.2		56.7
GDP2	6447.10	31322.9	46653.0	76809.0	41004.0		14239.4
HDI2	0.57	0.9	0.9	0.8	0.9		0.6
Distances Between Cluster Centroids							
	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5		
Cluster1	0.0	24875.8	40205.9	70361.9	34556.9		
Cluster2	24875.8	0.0	15330.2	45486.2	9681.2		
Cluster3	40205.9	15330.2	0.0	30156.0	5649.0		
Cluster4	70361.9	45486.2	30156.0	0.0	35805.1		
Cluster5	34556.9	9681.2	5649.0	35805.1	0.0		

Dendrogram



Países outliers 18: USA 31: Holanda 28: Macedonia 37: Catar

NOVA ANÁLISE DE CONGLOMERADOS

Cluster Analysis of Observations: DALY2, AIR_E2, WATER_E2, BIODIVERSITY, ...

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	45	99.9067	32.24	9	41	9
2	44	99.8546	50.24	24	28	24
3	43	99.7985	69.61	24	40	24
4	42	99.7915	72.04	24	35	24
5	41	99.7753	77.64	3	23	3
6	40	99.6276	128.64	15	39	15
7	39	99.6057	136.21	10	16	10
8	38	99.5839	143.73	2	25	2
9	37	99.5582	152.65	14	22	14
10	36	99.5540	154.09	2	4	2
11	35	99.5527	154.54	15	17	15
12	34	99.4958	174.19	8	24	8
13	33	99.4803	179.54	7	26	7
14	32	99.4296	197.04	29	37	29
15	31	99.4255	198.48	12	45	12
16	30	99.4001	207.24	15	34	15
17	29	99.3448	226.37	3	8	3
18	28	99.3302	231.40	31	36	31
19	27	99.2175	270.33	15	38	15
20	26	99.2097	273.01	19	46	19
21	25	99.1863	281.13	3	9	3
22	24	99.0723	320.48	5	32	5
23	23	98.9558	360.74	1	44	1
24	22	98.9287	370.11	27	29	27
25	21	98.9206	372.90	2	18	2
26	20	98.9072	377.52	3	15	3
27	19	98.8557	395.32	13	21	13
28	18	98.5761	491.92	3	11	3
29	17	98.2976	588.14	7	14	7
30	16	98.1639	634.32	10	30	10
31	15	98.0189	684.41	1	27	1
32	14	97.9517	707.62	19	42	19
33	13	97.8776	733.22	10	43	10
34	12	97.7983	760.64	2	10	2
35	11	97.7903	763.37	1	3	1
36	10	97.7519	776.66	1	2	1
37	9	97.3298	922.48	7	12	7
38	8	96.8555	1086.32	6	31	6
39	7	96.3698	1254.15	1	19	1
40	6	95.7376	1472.55	5	13	5
41	5	94.2457	1987.93	20	33	20
42	4	94.0279	2063.19	1	7	1
43	3	94.0121	2068.64	1	20	1
44	2	91.4682	2947.48	5	6	5
45	1	73.0625	9306.13	1	5	1

Final Partition

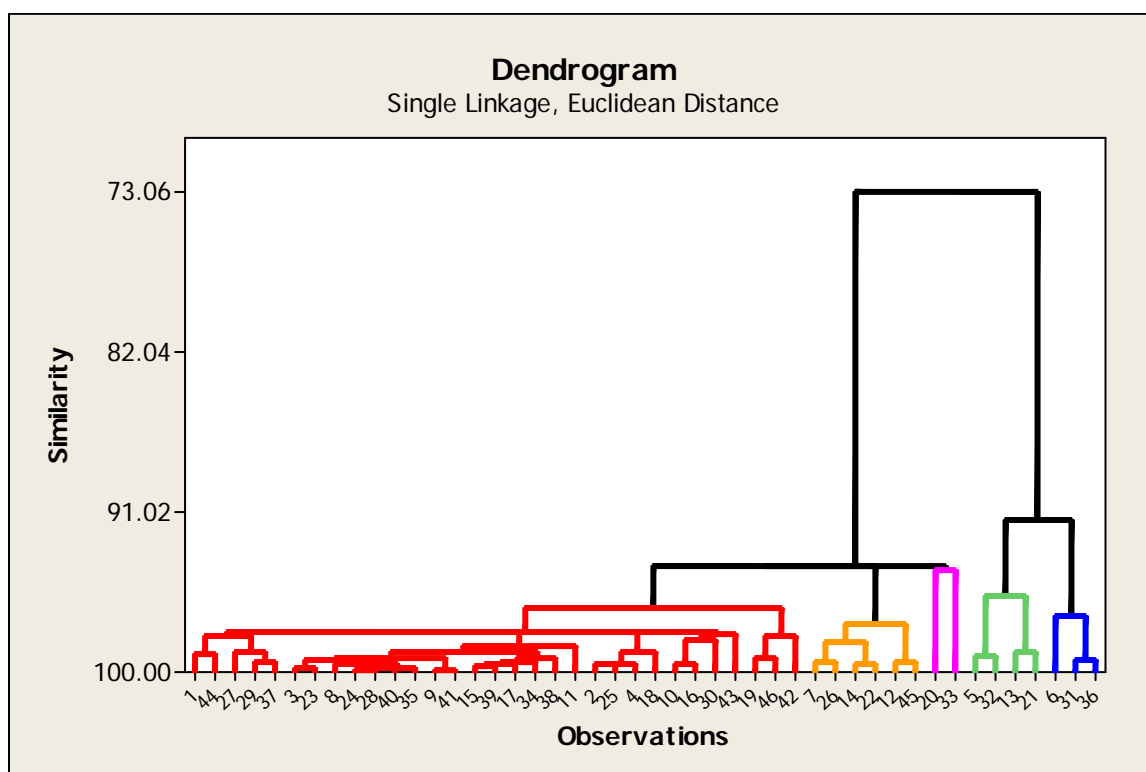
Number of clusters: 5

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	31	325539600	2840.07	6604.65
Cluster2	4	3469723	914.04	1111.28
Cluster3	3	987400	533.81	800.26
Cluster4	6	3513521	659.86	1080.43
Cluster5	2	1975934	993.97	993.97

Cluster Centroids						
Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Grand centroid
DALY2	36.93	86.5	86.4	68.1	68.5	49.9
AIR_E2	48.08	38.4	31.5	48.1	50.8	46.3
WATER_E2	62.13	65.7	78.4	72.1	49.0	64.2
BIODIVERSITY2	59.02	59.5	50.7	41.7	26.4	54.8
CLIMATE2	62.58	39.7	58.4	55.7	47.4	58.8
GDP2	4242.45	29363.8	33935.0	13989.3	17992.5	10232.5
HDI2	0.52	0.9	0.9	0.8	0.8	0.6

Distances Between Cluster Centroids					
	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	0.0	25121.4	29692.6	9747.0	13750.1
Cluster2	25121.4	0.0	4571.3	15374.5	11371.3
Cluster3	29692.6	4571.3	0.0	19945.7	15942.6
Cluster4	9747.0	15374.5	19945.7	0.0	4003.3
Cluster5	13750.1	11371.3	15942.6	4003.3	0.0

Dendrogram



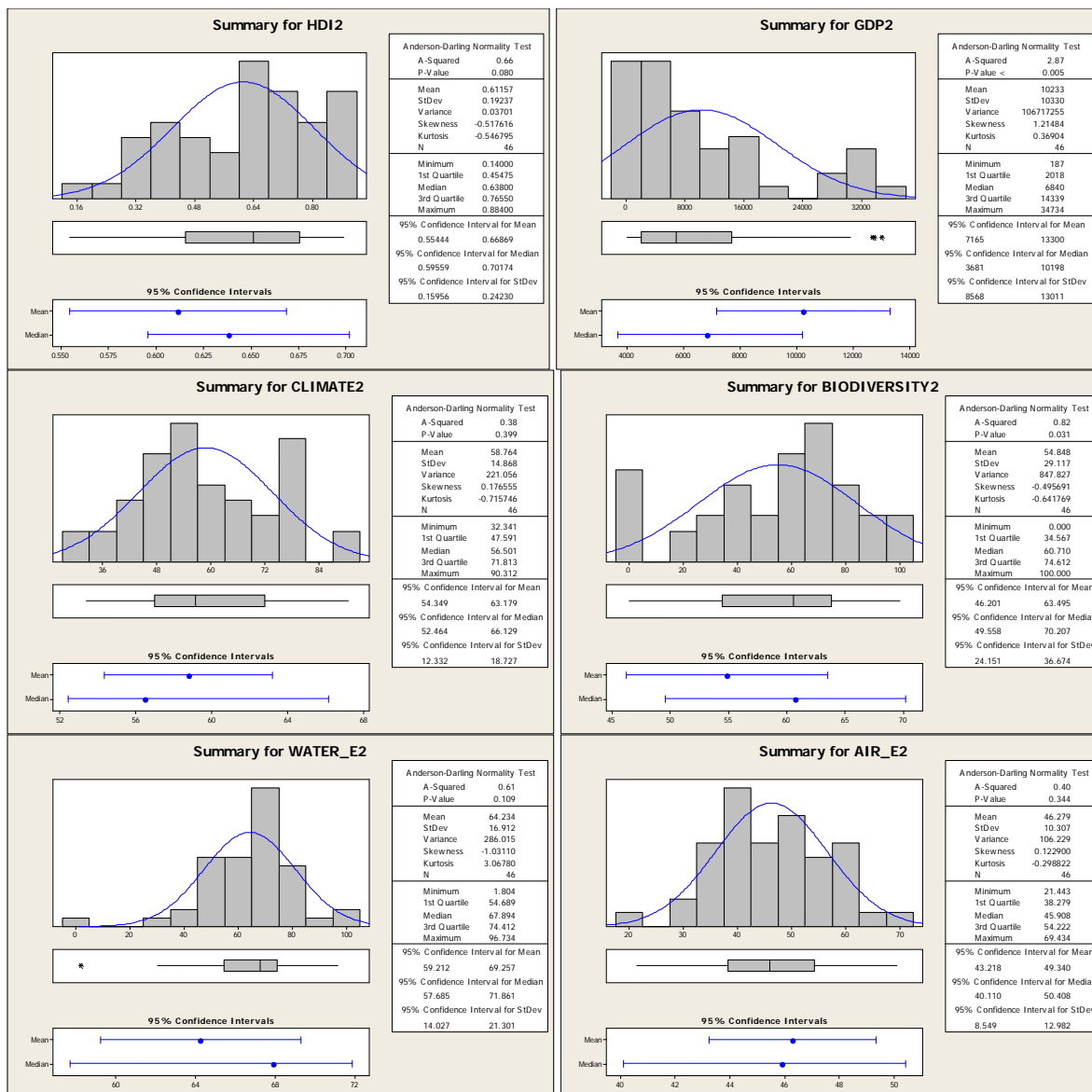
Na análise cluster x cluster observamos que a maior distância entre os centroides se dá entre o cluster 1 e 3, e a menor, entre o cluster 4 e 5.

O grande centroide do indicador DALY foi 49,9, AR 46,3, ÁGUA 64,2, CLIMA 58,8, BIODIVERSIDADE 54,8, HDI 0,6 e GDP 10232,5. No cluster 1 obtivemos 31 observações.

Os Bric's estão todos localizados no cluster 1, EXCETO Rússia que está no cluster 4..

Observando as distâncias, fizemos uma reclassificação conforme distância dos clusters, deixando o cluster 1 e agrupando o 2, 3, 4 e 5.

Primeiramente realizaremos uma análise exploratória da amostra 2, com a base tratada:

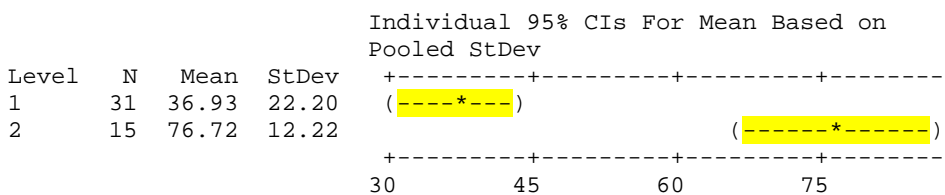


Baseado na análise do One-Way ANOVA, é possível identificar as variáveis com medianas mais distantes, que melhor servirão como base na classificação da amostra 2.

One-way ANOVA: DALY2 versus C10

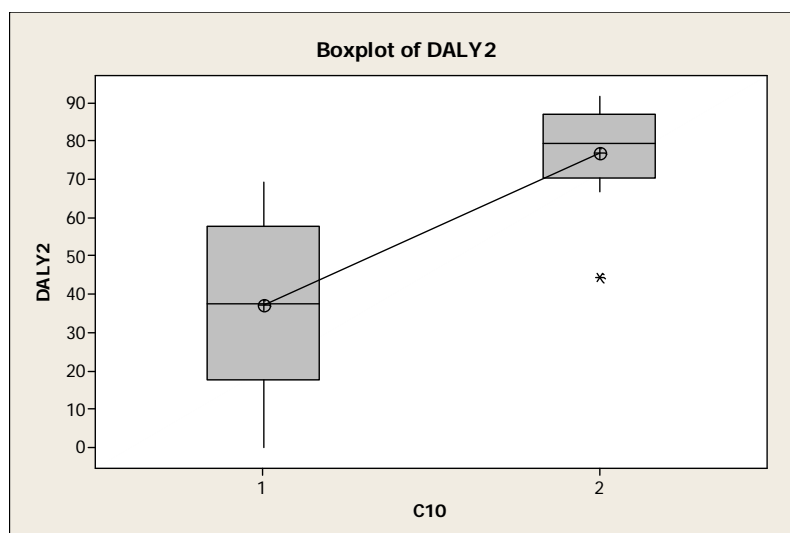
Source	DF	SS	MS	F	P
C10	1	16011	16011	41.75	0.000
Error	44	16875	384		
Total	45	32886			

S = 19.58 R-Sq = 48.69% R-Sq(adj) = 47.52%



Pooled StDev = 19.58

Boxplot of DALY2



One-way ANOVA: AIR_E2 versus C10

Source	DF	SS	MS	F	P
C10	1	308	308	3.03	0.089
Error	44	4473	102		
Total	45	4780			

S = 10.08 R-Sq = 6.44% R-Sq(adj) = 4.31%

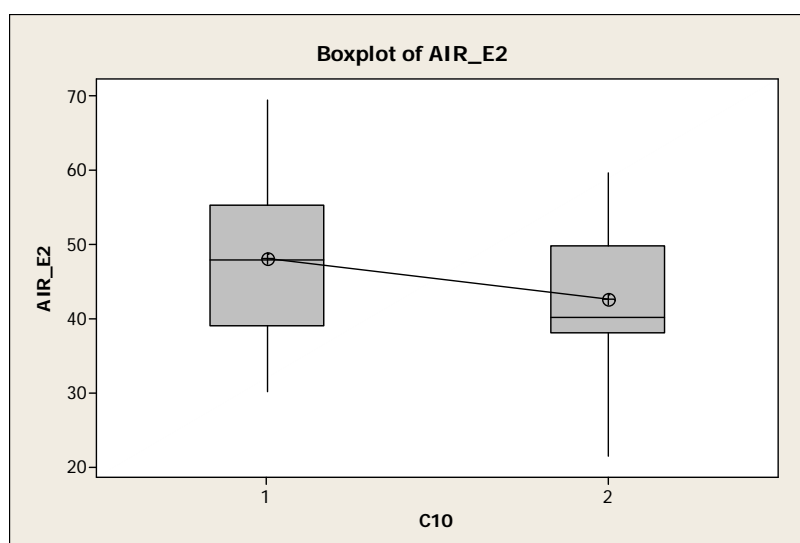
Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI Lower	CI Upper
1	31	48.08	10.44	37.64	58.52
2	15	42.56	9.28	33.28	51.84

40.0 44.0 48.0 52.0

Pooled StDev = 10.08

Boxplot of AIR_E2



One-way ANOVA: WATER_E2 versus C10

Source	DF	SS	MS	F	P
C10	1	422	422	1.49	0.228
Error	44	12448	283		
Total	45	12871			

Error 44 1391170339 31617508
 Total 45 4802276471

S = 5623 R-Sq = 71.03% R-Sq(adj) = 70.37%

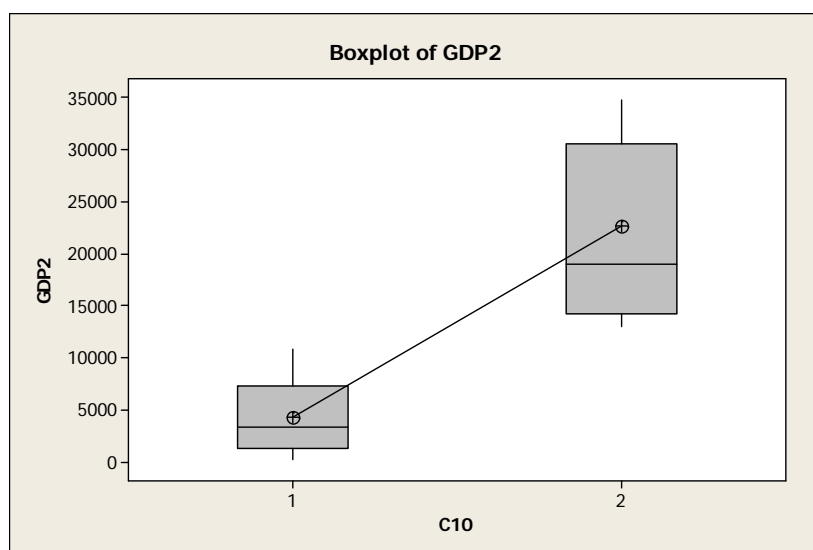
Individual 95% CIs For Mean Based on
 Pooled StDev

Level	N	Mean	StDev	CI
1	31	4242	3294	(---*---)
2	15	22612	8725	(-----*-----)

6000 12000 18000 24000

Pooled StDev = 5623

Boxplot of GDP2



One-way ANOVA: HDI2 versus C10

Source	DF	SS	MS	F	P
C10	1	0.8853	0.8853	49.94	0.000
Error	44	0.7800	0.0177		
Total	45	1.6653			

S = 0.1331 R-Sq = 53.16% R-Sq(adj) = 52.10%

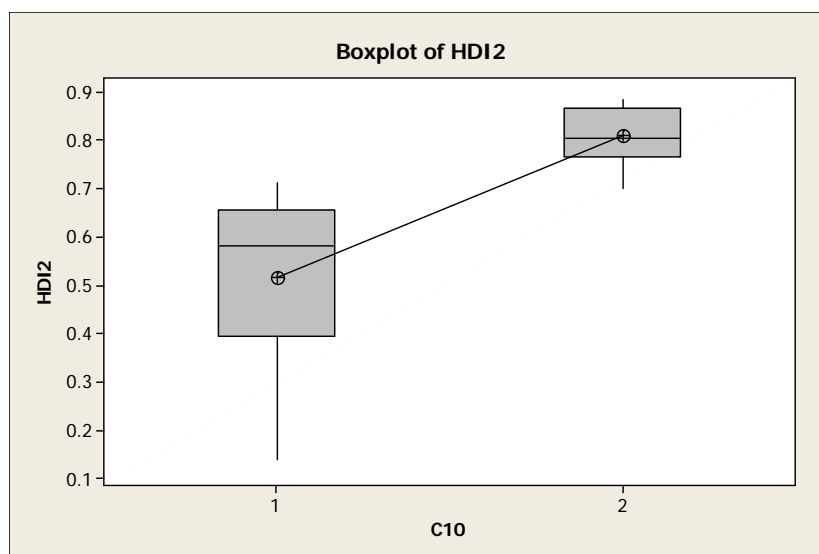
Individual 95% CIs For Mean Based on
 Pooled StDev

Level	N	Mean	StDev	CI
1	31	0.5151	0.1563	(---*---)
2	15	0.8110	0.0580	(-----*-----)

0.48 0.60 0.72 0.84

Pooled StDev = 0.1331

Boxplot of HDI2



Com P-value maior que 0,05, com intervalo de confiança de 95%, conclui-se que as médias de BIODIVERSIDADE, AR e ÁGUA podem ser as mesmas em relação aos clusters. Abaixo, veremos a análise do 2-Sample T:

Two-Sample T-Test and CI: DALY2, GDP2

Two-sample T for DALY2 vs GDP2

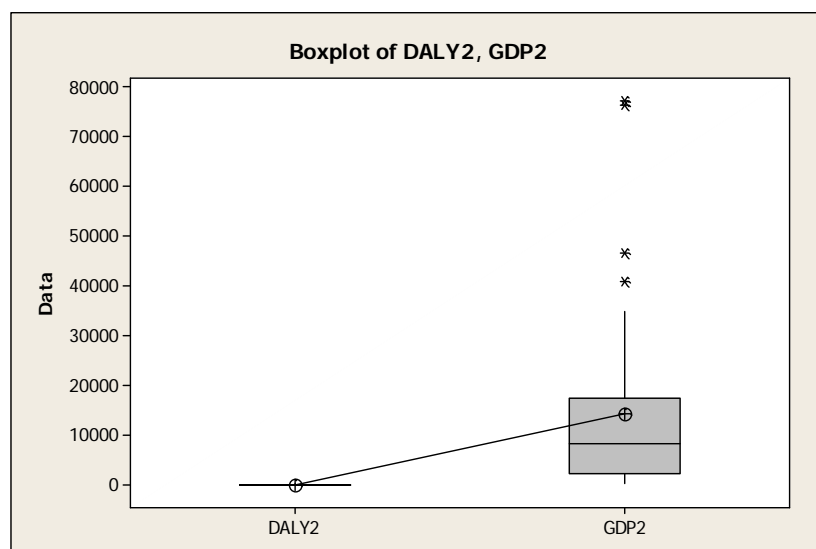
	N	Mean	StDev	SE Mean
DALY2	50	52.4	27.4	3.9
GDP2	50	14239	17577	2486

Difference = mu (DALY2) - mu (GDP2)

Estimate for difference: -14187

95% CI for difference: (-19182, -9192)

T-Test of difference = 0 (vs not =): T-Value = -5.71 P-Value = 0.000 DF = 49



Como o P-value foi igual a zero, a chance de serem iguais é igual a zero.

Os melhores indicadores pela análise de ANOVA ONE WAY foram Daly, GDP e HDI, sobre os quais realizaremos A Análise de Classificação através dos modelos: Análise Discriminante, Regressão Logística e Árvores de Classificação.

9b. ANÁLISE DISCRIMINANTE – AMOSTRA 2

Considerando primeiramente todas as variáveis:

Discriminant Analysis: C10 versus DALY2, AIR_E2, ...

Linear Method for Response: C10

Predictors: DALY2, AIR_E2, WATER_E2, BIODIVERSITY2, CLIMATE2, GDP2, HDI2

Group	1	2
Count	31	15

Summary of classification

Put into Group	True Group	
	1	2
1	31	2
2	0	13
Total N	31	15
N correct	31	13
Proportion	1.000	0.867

N = 46 N Correct = 44

Proportion Correct = 0.957

Squared Distance Between Groups

	1	2
1	0.0000	13.5713
2	13.5713	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-69.439	-89.957
DALY2	0.033	0.028
AIR_E2	0.790	0.984
WATER_E2	0.234	0.254
BIODIVERSITY2	0.164	0.149
CLIMATE2	0.534	0.525
GDP2	-0.000	0.001
HDI2	82.048	85.156

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
7**	2	1	1	4.593	0.579 México
			2	5.230	0.421
12**	2	1	1	6.985	0.581 Mauricius
			2	7.642	0.419

Com todas as variáveis conseguimos um acerto de 95,7%. Considerando o critério de parcimônia, verificaremos com quais variáveis também obtemos um modelo perfeito.

Discriminant Analysis: C10 versus DALY2, GDP2, HDI2

Linear Method for Response: C10

Predictors: DALY2, GDP2, HDI2

Group	1	2
Count	31	15

Summary of classification

Put into Group	True Group	
	1	2
1	31	1
2	0	14
Total N	31	15
N correct	31	14
Proportion	1.000	0.933

N = 46 N Correct = 45

Proportion Correct = 0.978

Squared Distance Between Groups

	1	2
1	0.0000	10.8663
2	10.8663	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-12.408	-22.467
DALY2	-0.255	-0.281
GDP2	-0.000	0.000
HDI2	69.014	75.326

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
12**	2	1	1	3.487	0.569
			2	4.046	0.431

Considerando DALY, GDP e HDI, também obtivemos 97,8% de acerto. Vejamos com menos variáveis:

Discriminant Analysis: C10 versus DALY2, GDP2

Linear Method for Response: C10

Predictors: DALY2, GDP2

Group	1	2
Count	31	15

Summary of classification

Put into Group	True Group	
	1	2
1	31	2
2	0	13
Total N	31	15
N correct	31	13
Proportion	1.000	0.867

N = 46 N Correct = 44

Proportion Correct = 0.957

Squared Distance Between Groups

	1	2
1	0.0000	10.6900
2	10.6900	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-1.8753	-9.9204
DALY2	0.1127	0.1211
GDP2	-0.0001	0.0005

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability	
12**	2	1	1	3.405	0.518	Mauricius
			2	3.548	0.482	
45**	2	1	1	2.820	0.598	Rússia
			2	3.614	0.402	

Com as variáveis DALY e GDP, também foi possível obter 95,7% de acerto. Vejamos o que acontece considerando apenas uma variável, otimizando o critério parcimonioso.

Discriminant Analysis: C10 versus GDP2

Linear Method for Response: C10

Predictors: GDP2

Group	1	2
Count	31	15

Summary of classification

Put into Group	True Group	
	1	2
1	31	2
2	0	13
Total N	31	15
N correct	31	13
Proportion	1.000	0.867

N = 46 N Correct = 44

Proportion Correct = 0.957

Squared Distance Between Groups

	1	2
1	0.0000	10.6727
2	10.6727	0.0000

Linear Discriminant Function for Groups

	1	2
Constant	-0.2846	-8.0858
GDP2	0.0001	0.0007

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability	
12**	2	1	1	2.482	0.547	Mauricius
			2	2.861	0.453	
45**	2	1	1	2.376	0.575	Russia
			2	2.977	0.425	

Com apenas o GDP, obtivemos 95,7% de acerto, através da equação linear.

Discriminant Analysis: C10 versus GDP2

Quadratic Method for Response: C10

Predictors: GDP2

Group	1	2
Count	31	15

Summary of classification

Put into Group	True Group	
	1	2
1	30	0
2	1	15
Total N	31	15
N correct	30	15
Proportion	0.968	1.000

N = 46 N Correct = 45

Proportion Correct = 0.978**From Generalized Squared** Distance to Group

Group	1	2
1	16.20	22.58
2	47.30	18.15

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
42**	1	2	1	20.22	0.468 Brasil
			2	19.97	0.532

Considerando apenas o GDP, obtivemos 97,8% de acerto, através da equação quadrática, considerando o modelo perfeito, atendendo ao critério parcimonioso.

10b. REGRESSÃO LOGÍSTICA

Não foi possível realizar o método de Regressão Logística neste modelo.

Binary Logistic Regression: C10 versus DALY2, GDP2

* 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	
C10	2	15	(Event)
	1	31	
	Total	46	

Logistic Regression Table

Odds

95% CI

Predictor	Coef	SE Coef	Z	P	Ratio	Lower	Upper
Constant	-160.183	8754.01	-0.02	0.985			
DALY2	-0.146075	101.993	-0.00	0.999	0.86	0.00	5.66299E+86
GDP2	0.0141824	0.750470	0.02	0.985	1.01	0.23	4.42

Log-Likelihood = -0.000
Test that all slopes are zero: G = 58.086, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0.0000006	43	1.000
Deviance	0.0000013	43	1.000
Hosmer-Lemeshow	0.0000001	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	0	0	0	0	1	4	5	5		15
Exp	0.0	0.0	0.0	0.0	0.0	0.0	1.0	4.0	5.0	5.0		
1												
Obs	4	5	4	5	5	4	4	0	0	0		31
Exp	4.0	5.0	4.0	5.0	5.0	4.0	4.0	0.0	0.0	0.0		
Total	4	5	4	5	5	4	5	4	5	5		46

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures		
Concordant	465	100.0	Somers' D		1.00
Discordant	0	0.0	Goodman-Kruskal Gamma		1.00
Ties	0	0.0	Kendall's Tau-a		0.45
Total	465	100.0			

11b. ÁRVORES DE CLASSIFICAÇÃO – AMOSTRA 2

Estatísticas descritivas:							
Variável	Categorias	Frequências		%			
nova	1		31				67.391
	2		15				32.609
Variável	Observações	Obs. com dados faltantes	Obs. sem dados faltantes	Mínimo	Máximo	Média	Desvio padrão
DALY2	46	0	46	0.000	91.499	49.904	27.033
AIR_E2	46	0	46	21.443	69.434	46.279	10.307
WATER_E	46	0	46	1.804	96.734	64.234	16.912
BIODIVER	46	0	46	0.000	100.000	54.848	29.117
CLIMATE2	46	0	46	32.341	90.312	58.764	14.868
GDP2	46	0	46	187.000	34734.000	10232.543	10330.404
HDI2	46	0	46	0.140	0.884	0.612	0.192

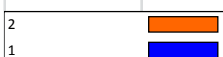
Matriz de correlação:

Variáveis	DALY2	AIR_E2	WATER_E2	BIODIVERSITY2	CLIMATE2	GDP2	HDI2
DALY2	1.000	-0.477	0.132	-0.346	-0.552	0.815	0.930
AIR_E2	-0.477	1.000	-0.270	-0.062	0.414	-0.506	-0.438
WATER_E2	0.132	-0.270	1.000	0.101	-0.100	0.220	0.208
BIODIVERSITY2	-0.346	-0.062	0.101	1.000	0.119	-0.143	-0.300
CLIMATE2	-0.552	0.414	-0.100	0.119	1.000	-0.461	-0.600
GDP2	0.815	-0.506	0.220	-0.143	-0.461	1.000	0.832
HDI2	0.930	-0.438	0.208	-0.300	-0.600	0.832	1.000

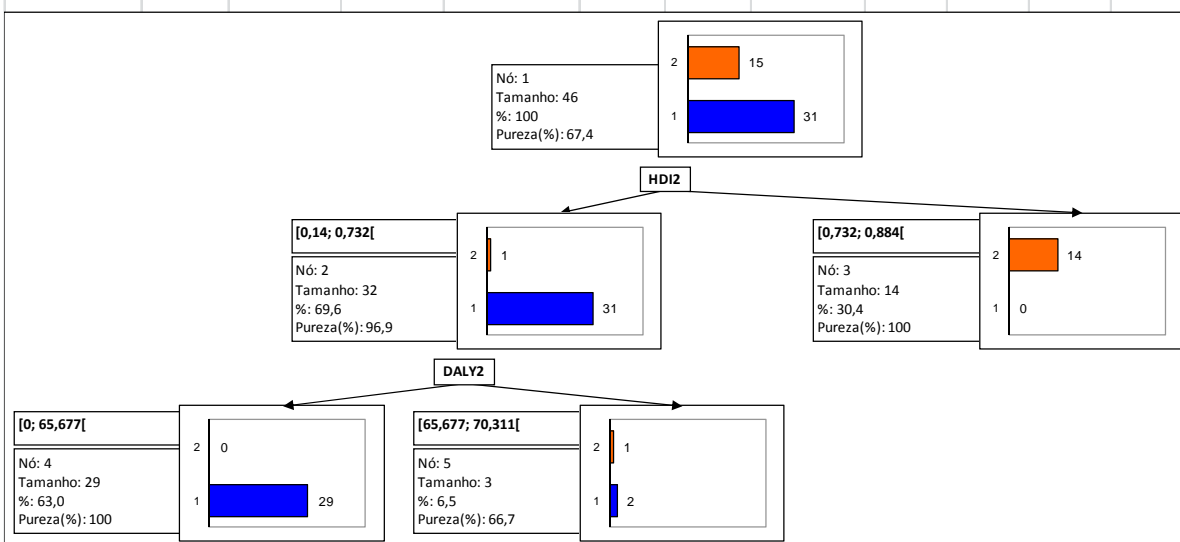
Estrutura da árvore:

Nó	p-valor	Objetos	%	Nó pai	Filhos	vel de sepa	Valores	Pureza
1	0.951	46	100.00%		2; 3			67.39%
2	0.558	32	69.57%	1	4; 5	HDI2	[0,14; 0,732[96.88%
3	0.000	14	30.43%	1		HDI2	[0,732; 0,884[100.00%
4	0.000	29	63.04%	2		DALY2	[0; 65,677[100.00%
5	1.000	3	6.52%	2		DALY2	[65,677; 70,311[66.67%

Legenda:



Árvore de classificação:



Réguas:

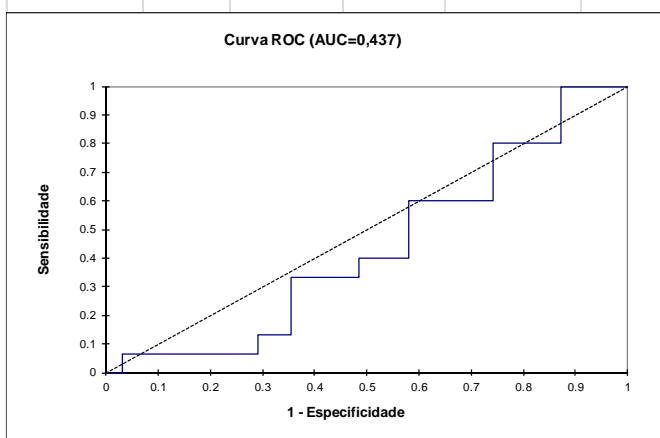
Nó	Pred(nova)	Frequência	Pureza	Réguas
Nó1	1.000	31	67.39%	
Nó2	1.000	31	96.88%	Se HDI2 em [0,14; 0,732[então nova = 1 em 96,9% dos casos
Nó3	2.000	14	100.00%	Se HDI2 em [0,732; 0,884[então nova = 2 em 100% dos casos
Nó4	1.000	29	100.00%	Se DALY2 em [0; 65,677[e HDI2 em [0,14; 0,732[então nova = 1 em 100% dos casos
Nó5	1.000	2	66.67%	Se DALY2 em [65,677; 70,311[e HDI2 em [0,14; 0,732[então nova = 1 em 66,7% dos casos

Resultados por objeto:				
Observação	A priori	A posteriori	Pr(1)	Pr(2)
Obs1	1	1	1.000	0.000
Obs2	1	1	1.000	0.000
Obs3	1	1	1.000	0.000
Obs4	1	1	1.000	0.000
Obs5	2	2	0.000	1.000
Obs6	2	2	0.000	1.000
Obs7	2	2	0.000	1.000
Obs8	1	1	1.000	0.000
Obs9	1	1	1.000	0.000
Obs10	1	1	1.000	0.000
Obs11	1	1	1.000	0.000
Obs12	2	1	0.667	0.333
Obs13	2	2	0.000	1.000
Obs14	2	2	0.000	1.000
Obs15	1	1	1.000	0.000
Obs16	1	1	1.000	0.000
Obs17	1	1	1.000	0.000
Obs18	1	1	0.667	0.333
Obs19	1	1	1.000	0.000
Obs20	2	2	0.000	1.000
Obs21	2	2	0.000	1.000
Obs22	2	2	0.000	1.000
Obs23	1	1	1.000	0.000
Obs24	1	1	1.000	0.000
Obs25	1	1	1.000	0.000
Obs26	2	2	0.000	1.000
Obs27	1	1	0.667	0.333
Obs28	1	1	1.000	0.000
Obs29	1	1	1.000	0.000
Obs30	1	1	1.000	0.000
Obs31	2	2	0.000	1.000
Obs32	2	2	0.000	1.000
Obs33	2	2	0.000	1.000
Obs34	1	1	1.000	0.000
Obs35	1	1	1.000	0.000
Obs36	2	2	0.000	1.000
Obs37	1	1	1.000	0.000
Obs38	1	1	1.000	0.000
Obs39	1	1	1.000	0.000
Obs40	1	1	1.000	0.000
Obs41	1	1	1.000	0.000
Obs42	1	1	1.000	0.000
Obs43	1	1	1.000	0.000
Obs44	1	1	1.000	0.000
Obs45	2	2	0.000	1.000
Obs46	1	1	1.000	0.000

Matriz de confusão para a amostra de estimação:

de \ a	1	2	Total	% correto
1	31	0	31	100.00%
2	1	14	15	93.33%
Total	32	14	46	97.83%

Tabela de classificação para a amostra de validação:



Área sob a curva: 0.437

Realizada a árvore de classificação, foi possível observar que nesta amostra 2, também tanto pelo aplicativo Minitab (pela Análise Discriminante, equação linear, pois a Regressão Logística não deu certo) quanto pelo aplicativo XLSTAT (Árvore de classificação e Regressão), a variável que apresenta maior importância na separação dos grupos foi o GDP.

AMOSTRA 3

Neste primeiro momento, realizaremos uma análise de conglomerados, para ver como a amostra 3 se divide, a fim de possibilitar o tratamento da amostra, para futura classificação.

Cluster Analysis of Observations: DALY3, AIR_E3, WATER_E3, BIODIVERSITY, ...

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	99.9800	43	4 31	4	2
2	48	99.9783	47	2 22	2	2
3	47	99.9770	49	27 50	27	2
4	46	99.9679	69	10 18	10	2
5	45	99.9672	70	10 17	10	3
6	44	99.9616	82	10 28	10	4
7	43	99.9577	91	41 43	41	2
8	42	99.9565	93	16 35	16	2
9	41	99.9534	100	2 41	2	4
10	40	99.9512	105	15 20	15	2
11	39	99.9330	144	1 11	1	2
12	38	99.9121	189	4 49	4	3
13	37	99.9038	206	10 16	10	6
14	36	99.8920	232	2 7	2	5
15	35	99.8888	239	2 8	2	6
16	34	99.8835	250	38 48	38	2
17	33	99.8835	250	5 42	5	2
18	32	99.8797	258	13 38	13	3
19	31	99.8424	338	6 29	6	2
20	30	99.8409	341	15 47	15	3
21	29	99.8383	347	9 15	9	4
22	28	99.8292	367	3 5	3	3
23	27	99.8286	368	4 36	4	4
24	26	99.8159	395	32 44	32	2
25	25	99.8153	396	26 30	26	2
26	24	99.8126	402	1 39	1	3
27	23	99.8047	419	2 10	2	12
28	22	99.7794	474	2 13	2	15
29	21	99.7588	518	1 9	1	7
30	20	99.7339	571	4 32	4	6
31	19	99.7293	581	12 45	12	2
32	18	99.6703	708	27 46	27	3
33	17	99.6584	733	1 21	1	8
34	16	99.6391	775	6 23	6	3
35	15	99.6304	793	2 3	2	18
36	14	99.4759	1125	1 27	1	11
37	13	99.3444	1407	25 26	25	3
38	12	99.3320	1434	19 33	19	2
39	11	99.3180	1464	19 25	19	5
40	10	99.3126	1475	1 2	1	29
41	9	99.1300	1867	6 40	6	4
42	8	99.0387	2063	1 4	1	35
43	7	98.1983	3867	6 19	6	9
44	6	98.1382	3996	1 12	1	37
45	5	96.2750	7995	14 37	14	2
46	4	96.0238	8534	1 6	1	46
47	3	94.7586	11249	1 14	1	48
48	2	91.5379	18162	1 34	1	49
49	1	35.0613	139376	1 24	1	50

Final Partition
Number of clusters: 3

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	48	9422316511	10972.0	44454.0
Cluster2	1	0	0.0	0.0
Cluster3	1	0	0.0	0.0

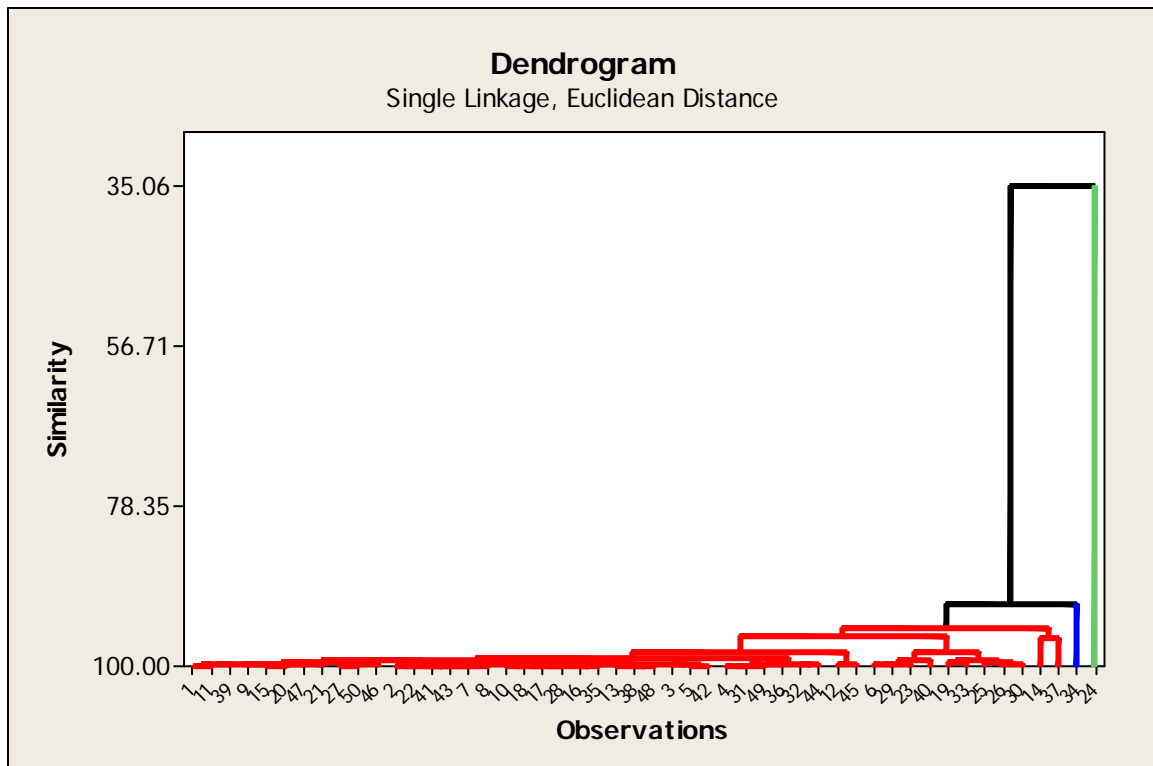
Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Grand centroid
DALY3	55.5	87	69.0	56.4
AIR_E3	47.7	22	69.8	47.6
WATER_E3	68.9	62	79.8	69.0
BIODIVERSITY3	53.6	75	40.2	53.7
CLIMATE3	55.3	59	36.5	55.0
GDP3	13824.0	215816	76440.0	19116.2
HDI3	0.7	1	0.9	0.7

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3
Cluster1	0	201992	62616
Cluster2	201992	0	139376
Cluster3	62616	139376	0

Dendrogram



Países outliers: 14 – Kwait 24-Malta 34 – Macedônia 37- Noruega

NOVA ANÁLISE DE CLUSTERS:

Cluster Analysis of Observations: DALY3, AIR_E3, WATER_E3, BIODIVERSITY, ...

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	45	99.8866	42.90	4	29	4	2
2	44	99.8770	46.57	2	21	2	2
3	43	99.8697	49.30	25	46	25	2
4	42	99.8182	68.81	10	17	10	2
5	41	99.8140	70.40	10	16	10	3
6	40	99.7823	82.38	10	26	10	4
7	39	99.7603	90.70	37	39	37	2
8	38	99.7534	93.34	15	32	15	2
9	37	99.7359	99.95	2	37	2	4
10	36	99.7233	104.73	14	19	14	2
11	35	99.6202	143.73	1	11	1	2
12	34	99.5015	188.65	4	45	4	3
13	33	99.4545	206.46	10	15	10	6
14	32	99.3877	231.75	2	7	2	5
15	31	99.3694	238.67	2	8	2	6
16	30	99.3396	249.93	34	44	34	2
17	29	99.3396	249.94	5	38	5	2
18	28	99.3175	258.29	13	34	13	3
19	27	99.1063	338.23	6	27	6	2
20	26	99.0978	341.45	14	43	14	3
21	25	99.0827	347.15	9	14	9	4
22	24	99.0315	366.53	3	5	3	3
23	23	99.0279	367.91	4	33	4	4
24	22	98.9561	395.09	30	40	30	2
25	21	98.9527	396.36	24	28	24	2
26	20	98.9372	402.22	1	35	1	3
27	19	98.8923	419.23	2	10	2	12
28	18	98.7487	473.56	2	13	2	15
29	17	98.6323	517.63	1	9	1	7
30	16	98.4910	571.09	4	30	4	6
31	15	98.4646	581.09	12	41	12	2
32	14	98.1303	707.62	25	42	25	3
33	13	98.0626	733.22	1	20	1	8
34	12	97.9533	774.60	6	22	6	3
35	11	97.9039	793.30	2	3	2	18
36	10	97.0279	1124.81	1	25	1	11
37	9	96.2821	1407.07	23	24	23	3
38	8	96.2119	1433.65	18	31	18	2
39	7	96.1323	1463.78	18	23	18	5
40	6	96.1016	1475.40	1	2	1	29
41	5	95.0663	1867.20	6	36	6	4
42	4	94.5485	2063.19	1	4	1	35
43	3	89.7822	3867.04	6	18	6	9
44	2	89.4414	3996.01	1	12	1	37
45	1	77.4506	8534.08	1	6	1	46

Final Partition

Number of clusters: 5

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	29	288130563	2827.81	6008.98
Cluster2	6	1780848	482.58	909.89
Cluster3	4	5270127	937.78	1864.86
Cluster4	2	168833	290.55	290.55
Cluster5	5	15468794	1559.29	2664.66

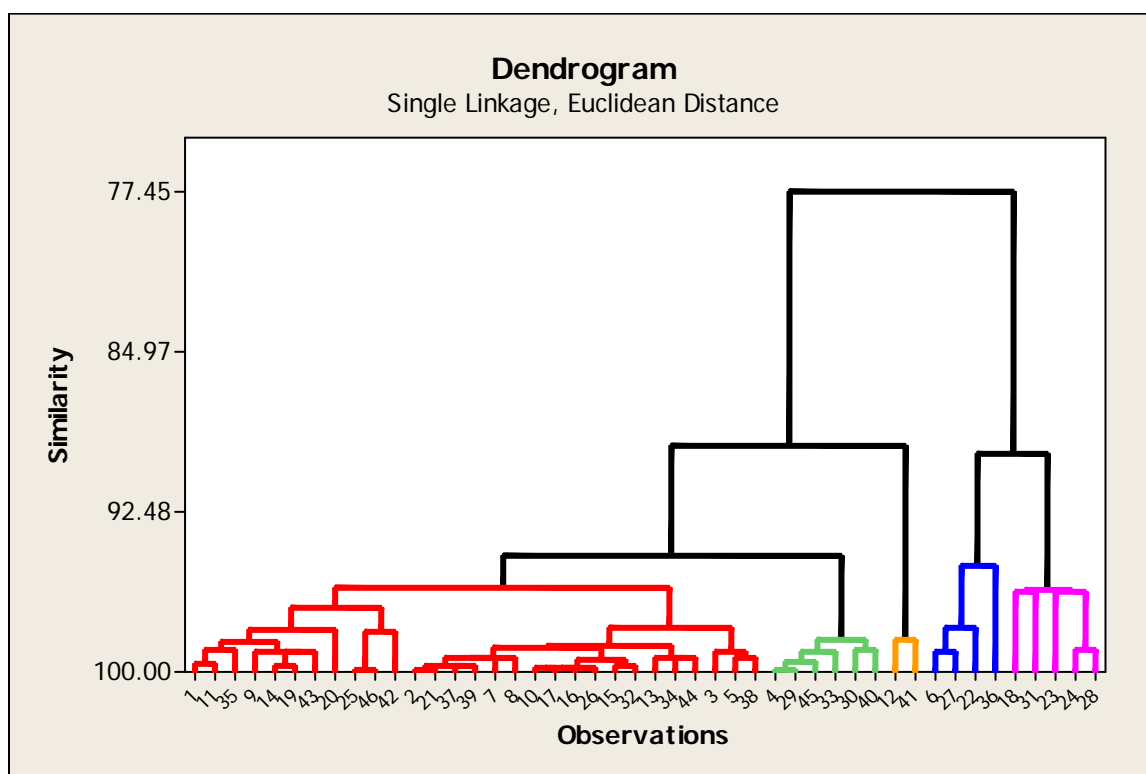
Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Grand centroid
DALY3	43.44	55.9	83.8	68.5	84.5	54.1
AIR_E3	50.47	51.7	37.1	45.6	36.6	47.7
WATER_E3	67.85	71.9	60.8	73.6	83.7	69.7
BIODIVERSITY3	52.40	61.9	43.1	65.2	62.7	54.5
CLIMATE3	60.98	49.9	35.7	48.6	49.2	55.5
GDP3	4838.10	13500.3	28610.3	18696.0	36370.4	12065.0
HDI3	0.55	0.7	0.9	0.8	0.9	0.6

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	0.0	8662.3	23772.2	13857.9	31532.3
Cluster2	8662.3	0.0	15110.0	5195.7	22870.1
Cluster3	23772.2	15110.0	0.0	9914.3	7760.2
Cluster4	13857.9	5195.7	9914.3	0.0	17674.4
Cluster5	31532.3	22870.1	7760.2	17674.4	0.0

Dendrogram

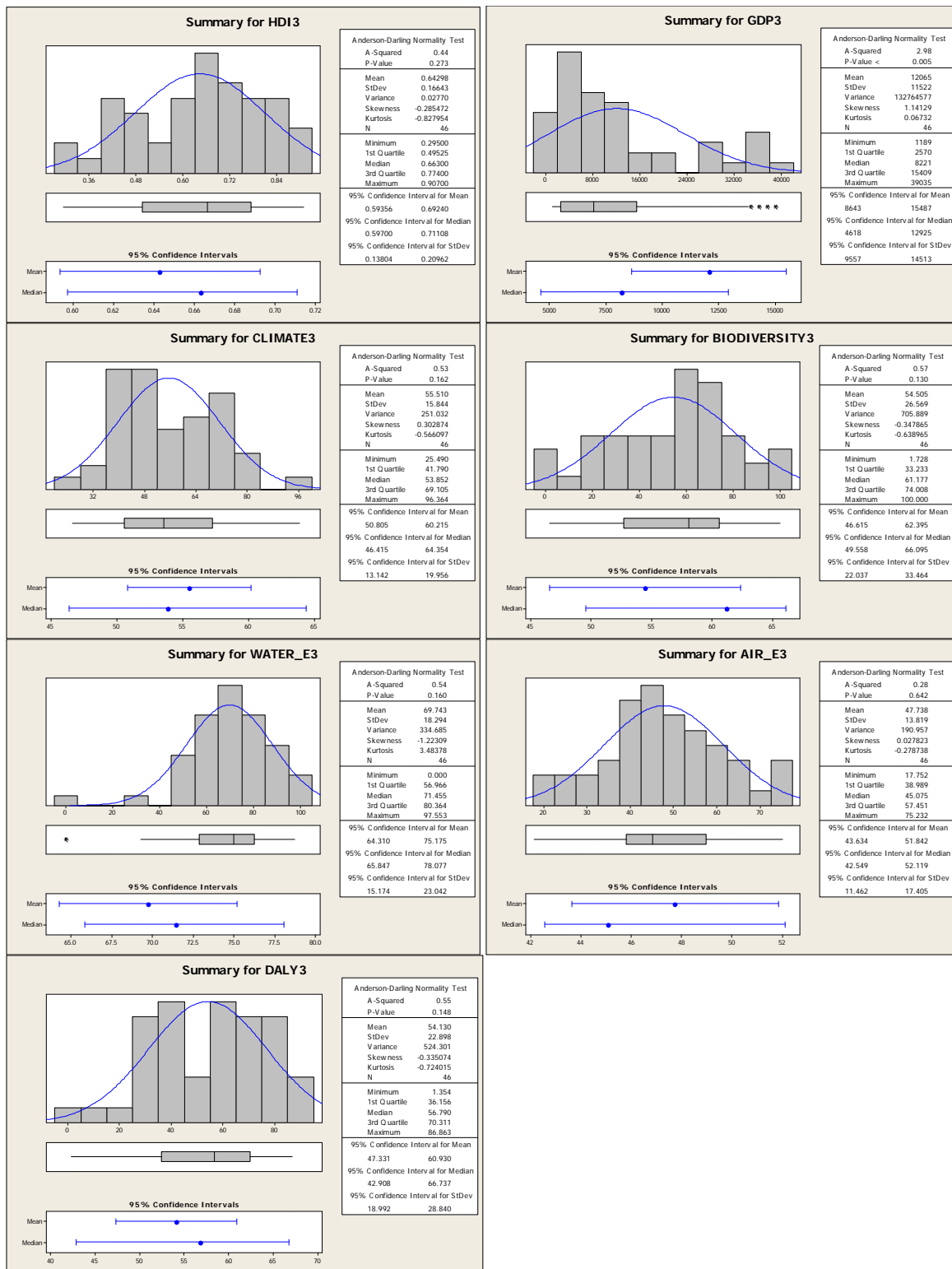


Na análise cluster x cluster observamos que a maior distância entre os centroides se dá entre o cluster 1 e 5, e a menor, entre o cluster 3 e 5.

O grande centroe do indicador DALY foi 54,1, AR 47,7, ÁGUA 69,7, CLIMA 55,5, BIODIVERSIDADE 54,5, HDI 0,6 e GDP 12065. No cluster 1 obtivemos 29 observações. OS Bric's estão todos localizados no cluster 1.

Para as análises a seguir, realizamos reclassificação de clusters de acordo com a proximidade entre os centroides. Agrupamos Cluster 2 com o 4 e 3 com o 5.

Primeiramente realizaremos uma análise exploratória da amostra 1, com a base tratada:



Baseado na análise do One-Way ANOVA, é possível identificar as variáveis com medianas mais distantes, que melhor servirão como base na classificação da amostra 3.

One-way ANOVA: DALY3 versus C10

Source	DF	SS	MS	F	P
C10	2	11639	5819	20.93	0.000
Error	43	11955	278		
Total	45	23594			

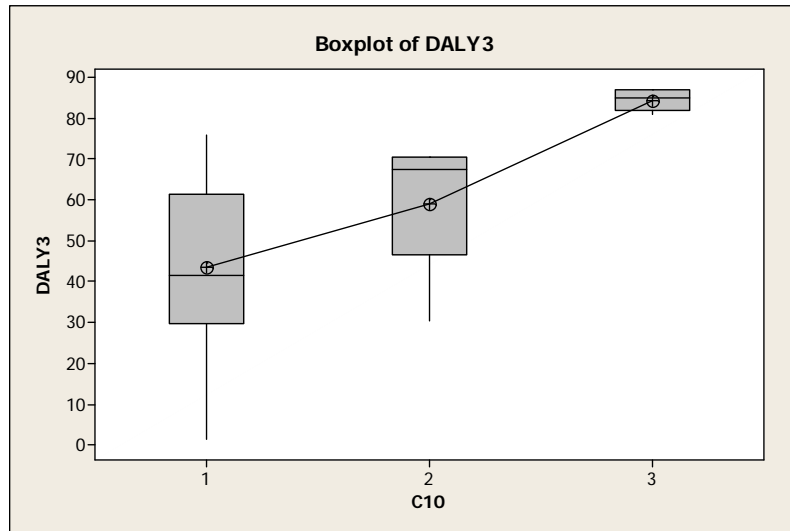
S = 16.67 R-Sq = 49.33% R-Sq(adj) = 46.97%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	
1	29	43.44	19.18	(---*---)
2	8	59.07	15.14	(---*---)
3	9	84.18	2.42	(---*---)

-----+-----+-----+-----+-----
 45 60 75 90
 -----+-----+-----+-----+-----

Pooled StDev = 16.67



One-way ANOVA: AIR_E3 versus C10

Source	DF	SS	MS	F	P
C10	2	1340	670	3.97	0.026
Error	43	7254	169		
Total	45	8593			

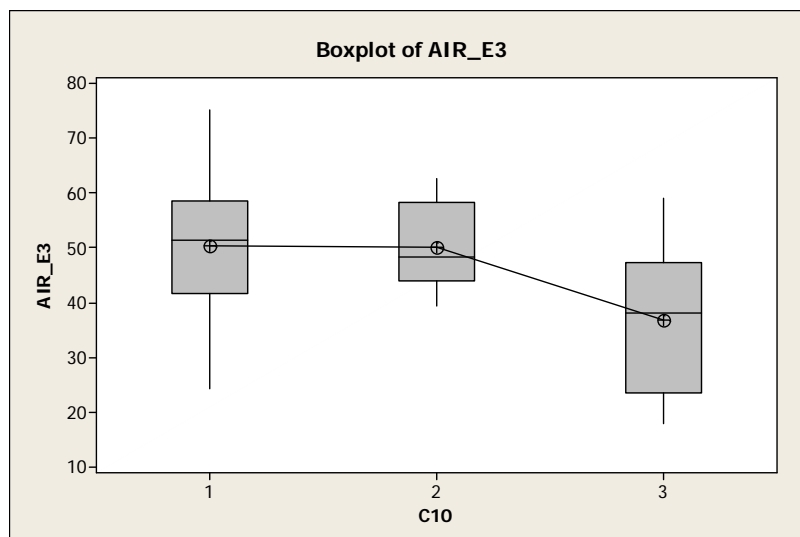
S = 12.99 R-Sq = 15.59% R-Sq(adj) = 11.66%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	
1	29	50.47	13.64	(---*---)
2	8	50.14	8.31	(---*---)
3	9	36.80	13.98	(---*---)

-----+-----+-----+-----+-----
 32.0 40.0 48.0 56.0
 -----+-----+-----+-----+-----

Pooled StDev = 12.99



One-way ANOVA: WATER_E3 versus C10

Source	DF	SS	MS	F	P
C10	2	288	144	0.42	0.661
Error	43	14773	344		
Total	45	15061			

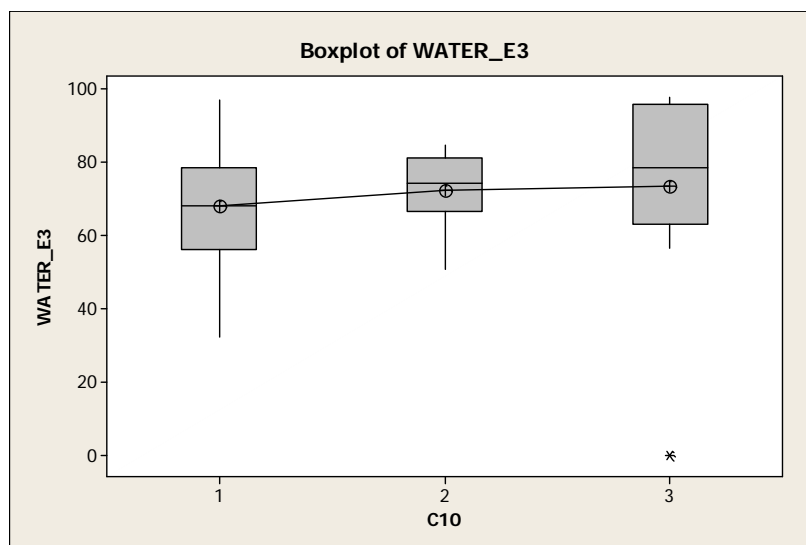
S = 18.54 R-Sq = 1.91% R-Sq(adj) = 0.00%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI
1	29	67.85	15.06	(-----*-----)
2	8	72.35	10.81	(-----*-----)
3	9	73.53	30.82	(-----*-----)

63.0 70.0 77.0 84.0

Pooled StDev = 18.54

**One-way ANOVA: BIODIVERSITY3 versus C10**

Source	DF	SS	MS	F	P
C10	2	669	335	0.46	0.633
Error	43	31096	723		
Total	45	31765			

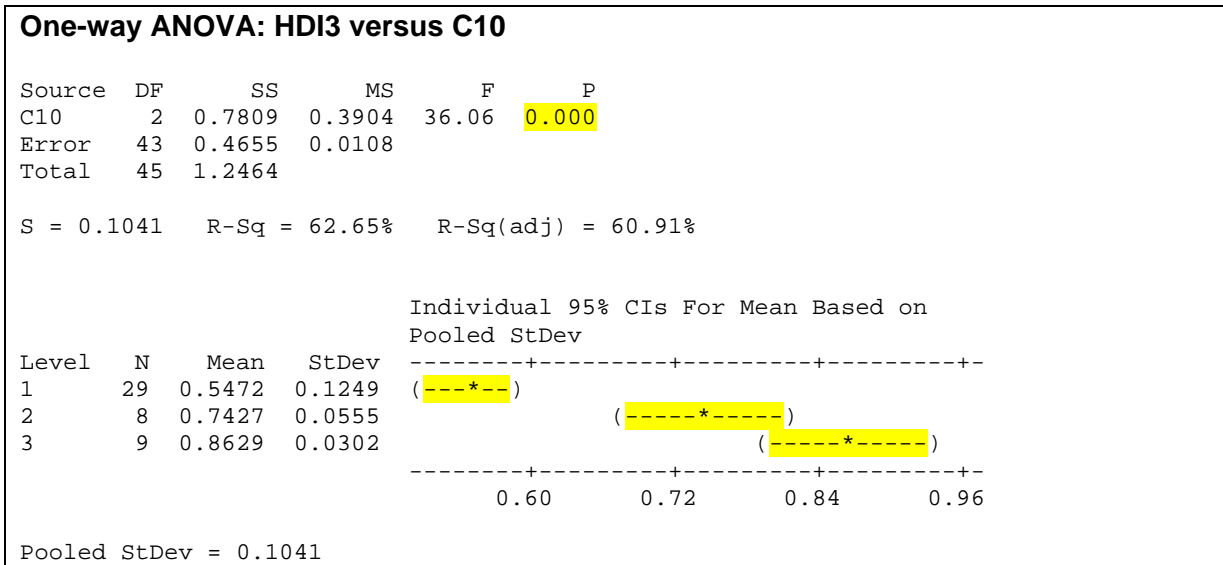
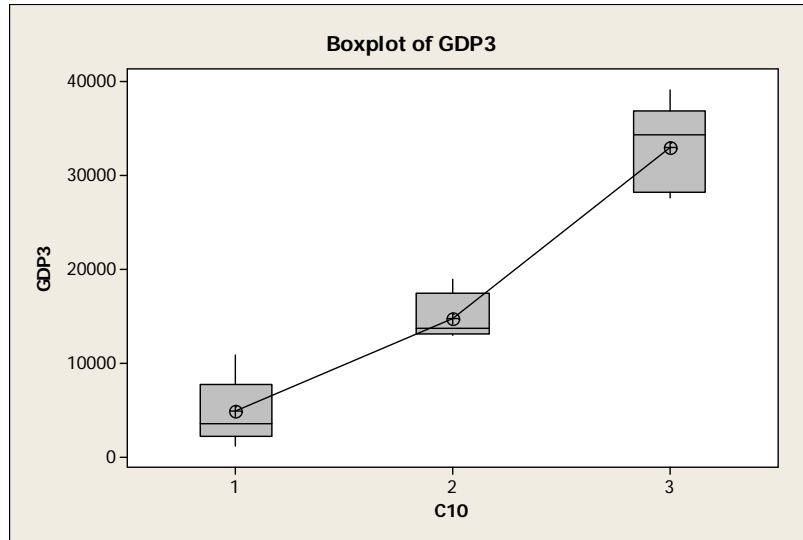
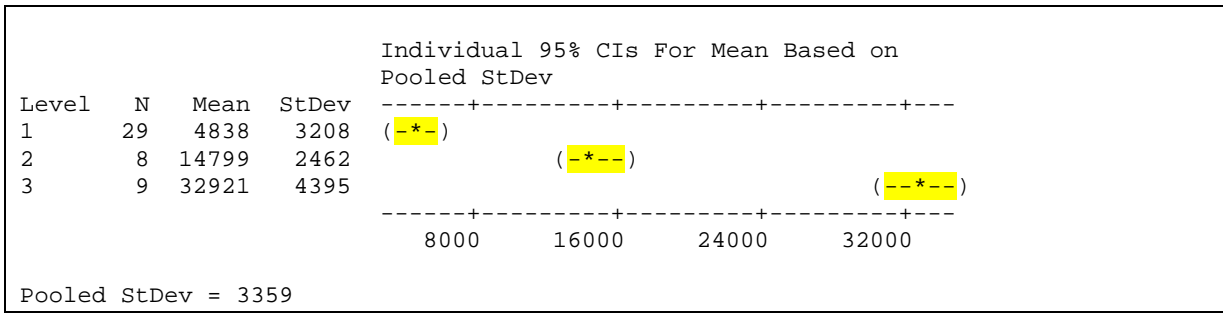
S = 26.89 R-Sq = 2.11% R-Sq(adj) = 0.00%

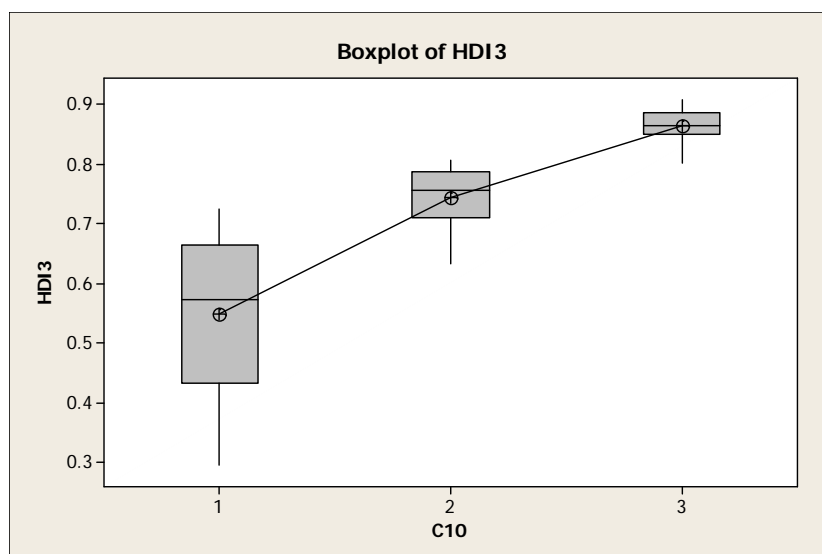
Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI
1	29	52.40	26.02	(-----*-----)
2	8	62.71	30.19	(-----*-----)
3	9	53.99	26.83	(-----*-----)

36 48 60 72

Pooled StDev = 26.89





Com P-value maior que 0,05, com intervalo de confiança de 95%, conclui-se que as médias de BIODIVERSIDADE e ÁGUA podem ser as mesmas em relação aos clusters. Abaixo veremos a análise do 2-Sample T:

Two-Sample T-Test and CI: DALY3, GDP3

Two-sample T for DALY3 vs GDP3

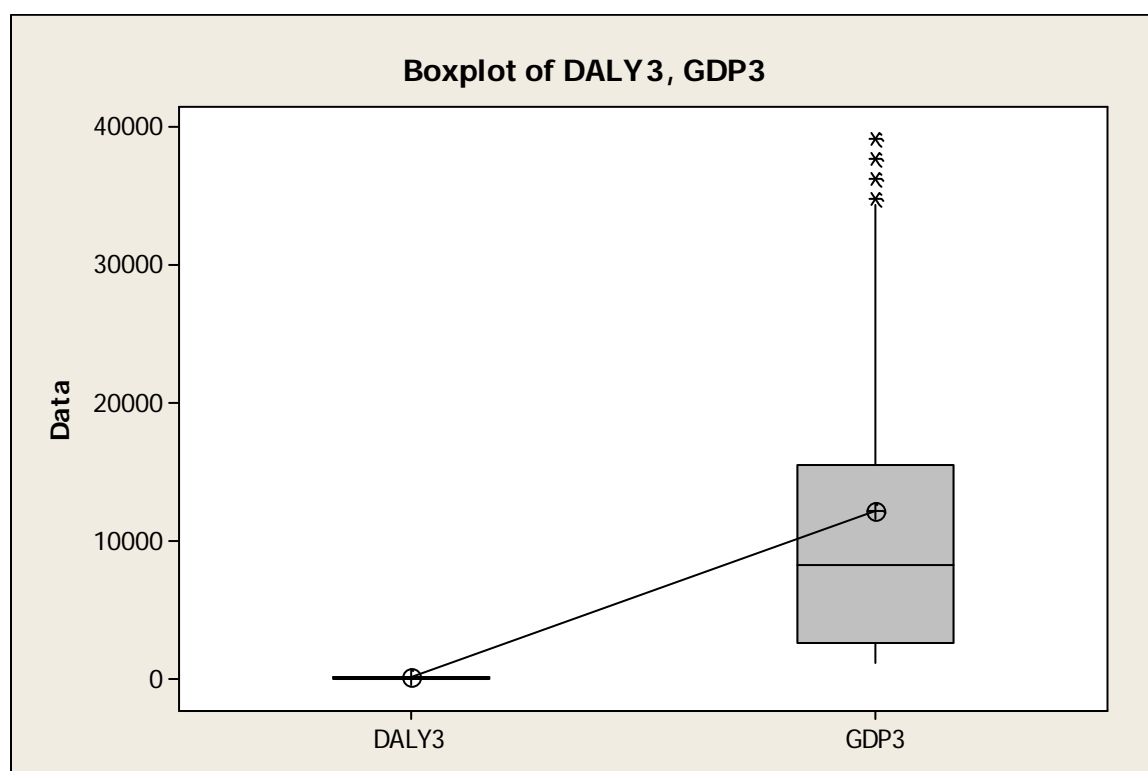
	N	Mean	StDev	SE Mean
DALY3	46	54.1	22.9	3.4
GDP3	46	12065	11522	1699

Difference = μ (DALY3) - μ (GDP3)

Estimate for difference: -12011

95% CI for difference: (-15433, -8589)

T-Test of difference = 0 (vs not =): T-Value = -7.07 **P-Value = 0.000** DF = 45



Com P-Value igual a zero, a chance de serem iguais é igual a zero.

Os melhores indicadores pela análise de ANOVA ONE WAY da amostra 3 foram Daly, GDP e HDI sobre os quais realizaremos A Análise de Classificação através dos modelos: Análise Discriminante, Regressão Logística e Árvores de Classificação.

9c. ANÁLISE DISCRIMINANTE – AMOSTRA 3

Primeiramente realizaremos a análise discriminante considerando todas as variáveis.

Discriminant Analysis: C10 versus DALY3, AIR_E3, ...						
Linear Method for Response: C10						
Predictors: DALY3, AIR_E3, WATER_E3, BIODIVERSITY3, CLIMATE3, GDP3, HDI3						
Group	1	2	3			
Count	29	8	9			
Summary of classification						
	True Group					
Put into Group	1	2	3			
1	28	0	0			
2	1	8	0			
3	0	0	9			
Total N	29	8	9			
N correct	28	8	9			
Proportion	0.966	1.000	1.000			
N = 46	N Correct = 45		Proportion Correct = 0.978			
Squared Distance Between Groups						
	1	2	3			
1	0.0000	11.7476	91.0684			
2	11.7476	0.0000	43.7261			
3	91.0684	43.7261	0.0000			
Linear Discriminant Function for Groups						
	1	2	3			
Constant	-53.95	-70.19	-101.47			
DALY3	-0.22	-0.32	-0.26			
AIR_E3	0.26	0.36	0.45			
WATER_E3	0.03	-0.01	-0.06			
BIODIVERSITY3	-0.03	-0.03	-0.05			
CLIMATE3	0.52	0.52	0.55			
GDP3	-0.00	0.00	0.00			
HDI3	139.11	151.19	103.57			
Summary of Misclassified Observations						
Observation	True Group	Pred Group	Squared Distance	Probability		
14**	1	2	17.66	0.086	Turcomenistão	
			12.93	0.914		
			82.75	0.000		

Foi possível obter 97,8% de acerto com todas as variáveis. Considerando o critério parcimonioso, verificaremos um modelo com menos variáveis.

Discriminant Analysis: C10 versus DALY3, GDP3

Linear Method for Response: C10

Predictors: DALY3, GDP3

Group	1	2	3
Count	29	8	9

Summary of classification

	True Group		
Put into Group	1	2	3
1	27	0	0
2	2	8	0
3	0	0	9
Total N	29	8	9
N correct	27	8	9
Proportion	0.931	1.000	1.000

N = 46 N Correct = 44

Proportion Correct = 0.957

Squared Distance Between Groups

	1	2	3
1	0.0000	9.7636	79.0326
2	9.7636	0.0000	33.2586
3	79.0326	33.2586	0.0000

Linear Discriminant Function for Groups

	1	2	3
Constant	-3.399	-10.415	-48.398
DALY3	0.161	0.088	-0.063
GDP3	-0.000	0.001	0.003

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability	
42**	1	2	1	3.232	0.355	Brasil
			2	2.035	0.645	
			3	51.275	0.000	
46**	1	2	1	5.002	0.204	África do Sul
			2	2.285	0.796	
			3	48.068	0.000	

Considerando DALY e GDP, o percentual continuou o mesmo, 95,7% de acerto, através da equação linear. Sem mudança para a equação quadrática

Discriminant Analysis: C10 versus GDP3

Linear Method for Response: C10

Predictors: GDP3

Group	1	2	3
Count	29	8	9

Summary of classification

	True Group		
Put into Group	1	2	3
1	26	0	0

2	3	8	0		
3	0	0	9		
Total N	29	8	9		
N correct	26	8	9		
Proportion	0.897	1.000	1.000		
N = 46 N Correct = 43 Proportion Correct = 0.935					
Squared Distance Between Groups					
	1	2	3		
1	0.0000	8.7961	69.9147		
2	8.7961	0.0000	29.1134		
3	69.9147	29.1134	0.0000		
Linear Discriminant Function for Groups					
	1	2	3		
Constant	-1.038	-9.708	-48.040		
GDP3	0.000	0.001	0.003		
Summary of Misclassified Observations					
Observation	True Group	Pred Group	Squared Distance	Probability	
25**	1	2	1 2.492	0.430	Coréia do Sul
			2 1.924	0.570	
			3 46.008	0.000	
42**	1	2	1 3.201	0.287	Brasil
			2 1.385	0.713	
			3 43.197	0.000	
46**	1	2	1 2.492	0.430	África do Sul
			2 1.924	0.570	
			3 46.008	0.000	

Considerando apenas o GDP, através da equação linear, foi possível obter 93,5% de acerto, que se mantém o mesmo na função quadrática.

Discriminant Analysis: C10 versus GDP3

Quadratic Method for Response: C10

Predictors: GDP3

Group	1	2	3
Count	29	8	9

Summary of classification

Put into Group	True Group		
	1	2	3
1	28	0	0
2	1	8	0
3	0	0	9
Total N	29	8	9
N correct	28	8	9
Proportion	0.966	1.000	1.000

N = 46 N Correct = 45 **Proportion Correct = 0.978**

From Generalized Squared Distance to Group

Group	1	2	3
1	16.15	31.99	57.60
2	25.79	15.62	33.78
3	92.80	69.80	16.78

Summary of Misclassified Observations

Observation	True Group	Pred Group	Squared Distance	Probability	
42**	1	2	1	19.66	0.325
			2	18.19	0.675
			3	42.00	0.000

Logo, o melhor modelo ocorreu considerando o GDP apenas, através da função quadrática e observando o critério parcimonioso, com 97,8% de acerto no modelo.

10c. REGRESSÃO LOGÍSTICA

Não foi possível realizar a Regressão Logística Ordinal

Ordinal Logistic Regression: C10 versus DALY3, GDP3

* 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
C10	1	29
	2	8
	3	9
Total		46

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Const(1)	155.071	6886.39	0.02	0.982			
Const(2)	268.337	10846.7	0.02	0.980			
DALY3	0.134215	74.2456	0.00	0.999	1.14	0.00	1.80411E+63
GDP3	-0.0136957	0.550657	-0.02	0.980	0.99	0.34	2.90

Log-Likelihood = -0.000

Test that all slopes are zero: G = 84.111, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	0.0000011	88	1.000
Deviance	0.0000022	88	1.000

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	565	100.0	Somers' D 1.00
Discordant	0	0.0	Goodman-Kruskal Gamma 1.00
Ties	0	0.0	Kendall's Tau-a 0.55
Total	565	100.0	

11c. ÁRVORES DE CLASSIFICAÇÃO – AMOSTRA 3

Estatísticas descritivas:								
Variável	Categorias	Frequências						
Nova	1	29	63.043					
	2	8	17.391					
	3	9	19.565					
Variável	Observações	Obs. com dados faltantes	Obs. sem dados faltantes	Mínimo	Máximo	Média	Desvio padrão	
DALY3	46	0	46	1.354	86.863	54.130	22.898	
AIR_E3	46	0	46	17.752	75.232	47.738	13.819	
WATER_E3	46	0	46	0.000	97.553	69.743	18.294	
BIODIVERSITY3	46	0	46	1.728	100.000	54.505	26.569	
CLIMATE3	46	0	46	25.490	96.364	55.510	15.844	
GDP3	46	0	46	1189.000	39035.000	12065.043	11522.351	
HDI3	46	0	46	0.295	0.907	0.643	0.166	
Matriz de correlação:								
Variáveis	DALY3	AIR_E3	WATER_E3	BIODIVERSITY3	CLIMATE3	GDP3	HDI3	
DALY3	1.000	-0.536	0.231	-0.088	-0.568	0.792	0.897	
AIR_E3	-0.536	1.000	-0.009	-0.058	0.499	-0.467	-0.466	
WATER_E3	0.231	-0.009	1.000	0.292	0.064	0.188	0.254	
BIODIVERSITY3	-0.088	-0.058	0.292	1.000	0.159	0.066	0.089	
CLIMATE3	-0.568	0.499	0.064	0.159	1.000	-0.522	-0.624	
GDP3	0.792	-0.467	0.188	0.066	-0.522	1.000	0.855	
HDI3	0.897	-0.466	0.254	0.089	-0.624	0.855	1.000	
Estrutura da árvore:								
Nó	p-valor	Objetos	%	Nó pai	Filhos	Variável de separação	Valores	Pureza
1	0.841	46	100.00%		2; 3			63.04%
2	0.753	37	80.43%	1	4; 5	GDP3	[1189; 23253[78.38%
3	0.000	9	19.57%	1		GDP3	[23253; 39035[100.00%
4	0.395	32	69.57%	2	6; 7	HDI3	[0,295; 0,738[90.63%
5	0.000	5	10.87%	2		HDI3	[0,738; 0,805[100.00%
6	0.693	26	56.52%	4		HDI3	[0,295; 0,683[96.15%
7	0.707	6	13.04%	4	8; 9	HDI3	[0,683; 0,738[66.67%
8	0.000	3	6.52%	7		CLIMATE3	[37,837; 49,006[100.00%
9	0.500	3	6.52%	7		CLIMATE3	[49,006; 72,916[66.67%

Réguas:				
Nó	Pred(Nova)	Frequência	Pureza	Réguas
Nó1	1.000	29	63.04%	
Nó2	1.000	29	78.38%	Se GDP3 em [1189; 23253[então Nova = 1 em 78,4% dos casos
Nó3	3.000	9	100.00%	Se GDP3 em [23253; 39035[então Nova = 3 em 100% dos casos
Nó4	1.000	29	90.63%	Se HDI3 em [0,295; 0,738[e GDP3 em [1189; 23253[então Nova = 1 em 90,6% dos casos
Nó5	2.000	5	100.00%	Se HDI3 em [0,738; 0,805[e GDP3 em [1189; 23253[então Nova = 2 em 100% dos casos
Nó6	1.000	25	96.15%	Se HDI3 em [0,295; 0,683[e GDP3 em [1189; 23253[então Nova = 1 em 96,2% dos casos
Nó7	1.000	4	66.67%	Se HDI3 em [0,683; 0,738[e GDP3 em [1189; 23253[então Nova = 1 em 66,7% dos casos
Nó8	1.000	3	100.00%	Se CLIMATE3 em [37,837; 49,006[e HDI3 em [0,683; 0,738[e GDP3 em [1189; 23253[então Nova = 1 em 100% dos casos
Nó9	2.000	2	66.67%	Se CLIMATE3 em [49,006; 72,916[e HDI3 em [0,683; 0,738[e GDP3 em [1189; 23253[então Nova = 2 em 66,7% dos casos

Resultados por objeto:					
Observação	A priori	A posteriori	Pr(1)	Pr(2)	Pr(3)
Obs1	1	1	0.962	0.038	0.000
Obs2	1	1	0.962	0.038	0.000
Obs3	1	1	0.962	0.038	0.000
Obs4	2	2	0.333	0.667	0.000
Obs5	1	1	0.962	0.038	0.000
Obs6	3	3	0.000	0.000	1.000
Obs7	1	1	0.962	0.038	0.000
Obs8	1	1	0.962	0.038	0.000
Obs9	1	1	1.000	0.000	0.000
Obs10	1	1	0.962	0.038	0.000
Obs11	1	1	0.962	0.038	0.000
Obs12	2	2	0.000	1.000	0.000
Obs13	1	1	0.962	0.038	0.000
Obs14	1	1	0.962	0.038	0.000
Obs15	1	1	0.962	0.038	0.000
Obs16	1	1	0.962	0.038	0.000
Obs17	1	1	0.962	0.038	0.000
Obs18	3	3	0.000	0.000	1.000
Obs19	1	1	1.000	0.000	0.000
Obs20	1	1	0.962	0.038	0.000
Obs21	1	1	0.962	0.038	0.000
Obs22	3	3	0.000	0.000	1.000
Obs23	3	3	0.000	0.000	1.000
Obs24	3	3	0.000	0.000	1.000
Obs25	1	1	0.962	0.038	0.000
Obs26	1	1	0.962	0.038	0.000
Obs27	3	3	0.000	0.000	1.000
Obs28	3	3	0.000	0.000	1.000
Obs29	2	2	0.333	0.667	0.000
Obs30	2	2	0.000	1.000	0.000
Obs31	3	3	0.000	0.000	1.000
Obs32	1	1	0.962	0.038	0.000
Obs33	2	1	0.962	0.038	0.000
Obs34	1	1	0.962	0.038	0.000
Obs35	1	2	0.333	0.667	0.000
Obs36	3	3	0.000	0.000	1.000
Obs37	1	1	0.962	0.038	0.000
Obs38	1	1	0.962	0.038	0.000
Obs39	1	1	0.962	0.038	0.000
Obs40	2	2	0.000	1.000	0.000
Obs41	2	2	0.000	1.000	0.000
Obs42	1	1	1.000	0.000	0.000
Obs43	1	1	0.962	0.038	0.000
Obs44	1	1	0.962	0.038	0.000
Obs45	2	2	0.000	1.000	0.000
Obs46	1	1	0.962	0.038	0.000

Matriz de confusão para a amostra de estimação:

de \ a	1	2	3	Total	% correto
1	28	1	0	29	96.55%
2	1	7	0	8	87.50%
3	0	0	9	9	100.00%
Total	29	8	9	46	95.65%

Realizada a árvore de classificação também na amostra 3, foi possível observar que pelo aplicativo Minitab (pela Análise Discriminante, equação quadrática), houve um acerto de 97,8% no modelo considerando apenas GDP. Já pelo aplicativo XLSTAT (Árvore de classificação e Regressão), a variável que apresenta maior importância na separação dos grupos foi o GDP, e a seguir pela variável HDI e depois CLIMA.

Como conclusão, através das análises quantitativas realizadas neste trabalho, pudemos observar como o índice DALY, que é a soma do número de anos de vida perdidos devido à mortalidade prematura causada pela doença influenciada pelo ambiente e os anos de vida saudável perdidos por incapacidade causada por essa doença, está relacionado ao GDP per capita e ao índice de desenvolvimento humano.