

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
Programa de Pós Graduação em Economia e Administração FEA – PUC-SP



GPS

TEMA 1. RECURSOS BASICOS: AGUA, ALIMENTO, ENERGIA

(BASIC FEATURES: WATER, FOOD, ENERGY)

RESUMO

Este trabalho é resultado de um conjunto de doze trabalhos desenvolvidos ao longo do primeiro semestre de 2015 sobre 132 países divididos em 8 regiões sobre o tema: **recursos básicos: água, alimento, energia (basic features: water, food, energy)** elaborado pelo Programa Países Sustentáveis - Guia GPS – Gestão Pública Sustentável uma publicação do Núcleo de Estudos Futuros da PUC-SP –Pontifícia Universidade Católica de São Paulo.

Este tema é composto por 8 variáveis sintéticas – que são indicadores construídos considerando um conjunto de variáveis para representar uma situação e 8 variáveis analíticas que são indicadores que refletem uma determinada situação de uma realidade.

Aqui a análise dos dados foi condensada em três regiões a partir das oito originais: AIBER, AVECO e OTHERS para facilitar o estudo comparativo. Os dados apresentados estão sintetizados na apresentação com o auxílio de gráficos (*pie chart*, barras, histogramas, gráficos de ramos, box-plot, dot-plot e curvas de densidade) e de medidas numéricas (média, mediana, quartis, desvio-padrão, variância, intervalo de confiança e teste de normalidade de Anderson-Darling).

Houve a inclusão de Cuba na região AIBER que anteriormente fazia parte da Região CARLA e os dados ausentes na tabela original haviam sido obtidos do portal Knoema que é um Atlas de Dados que utiliza dados de fontes internacionais como ONU, Banco Mundial e outras fontes de renome internacional

(<http://pt.knoema.com/atlas/Cuba/Pessoas-com-desnutri%C3%A7%C3%A3o>).

O item 4 – Análise Exploratória dos dados é reservado para a apresentação das variáveis sem a separação das regiões do Estudo Final. A análise específica da região AIBER – Ibero Americana que tem dentre outros países o Brasil, Argentina, Espanha e Portugal, da região AVECO e a região denominada OTHERS serão tratadas e analisadas a partir do item 5 – Relação entre as Variáveis e as Regiões AIBER, AVECO e OTHERS. Os demais itens farão apresentação dos dados e análises com base nessa divisão em três regiões.

Todo o estudo emprega na sua maior parte o MINITAB. Apenas na construção da Árvore de Decisão que é empregado o SPSS.

SUMÁRIO

	P.
Introdução	5
1.1 – Estrutura do Trabalho Final	5
1.2 – Software utilizado	6
2. Entendendo os dados	7
2.1 O tema de pesquisa: recursos básicos: agua, alimento, energia (basic features: water, food, energy)	7
2.2 Os indivíduos	7
2.3 As Variáveis do tema	7
2.3.1 A Tabela de Dados	10
3. Análise das variáveis	11
3.1 – variáveis quantitativas	13
4 – Análise Exploratória dos Dados	15
4.1 – Análise Exploratória das variáveis da pesquisa	15
5 – Relação entre as variáveis e as regiões AIBER, AVECO e OTHERS	33
5.1 – região AIBER	33
5.2 – região AVECO	38
5.3 – Região OTHERS	39
5.4 – Relação entre as variáveis por região:	41
5.4.1. Região AIBER	41
5.4.2. Região AVECO	47
5.4.3. Região OTHERS	53
6 - Comparando as três regiões	59
7 - Análise de regressão das variáveis com similaridade	61
7.1 – Região AIBER	61
7.2 – Região AVECO	63
7.3 – Região OTHERS	65
7.4 – Comparação entre as três regiões: AIBER, AVECO E OUTROS	67
8 - Análise do componente principal	68
8.1 - Região AIBER	69
8.2 - Região AVECO	72
8.3 - Região OTHERS	75
8.4 – Comparação entre as três regiões: AIBER, AVECO e OUTROS	77

9. Análise de conglomerados	82
9.1 – Região AIBER	82
9.2 – Região AVECO	85
9.3 – Região OTHERS	87
10. Comparação de média, análise de variância e intervalo de Confiança.	88
10.1 – Para este item vamos apresentar a Região AIBER	89
10.2 – Região AVECO	91
10.3 – Região OTHERS	93
11 – One-way anova com as três regiões	98
12 – Regressão logística	103
12.1 – Variáveis Sintéticas e as Três Regiões: AIBER, AVECO e OTHERS	104
13 – Árvore de decisão para as três regiões – AIBER, AVECO e OTHERS	106
Conclusão	110
Referencias	

1. INTRODUÇÃO

O presente trabalho tem por objetivo aplicar os conceitos desenvolvidos durante o semestre na disciplina Métodos Quantitativos e Qualitativos da Pesquisa Empírica ministrada pelo prof. Dr. Arnoldo José de Hoyos Guevara.

O tema de estudo é **recursos básicos: água, alimento, energia (basic features: water, food, energy)** elaborado pelo Programa Países Sustentáveis – Guia GPS – Gestão Pública Sustentável uma publicação do Núcleo de Estudos Futuros da PUC-SP –Pontifícia Universidade Católica de São Paulo.

Quadro 1: Regiões de estudo ao longo do semestre: 8 regiões e 132 países

Região	Alguns países dessa região
AIBER – região Ibero-americana	Brasil, Paraguai e Peru
AVECO – região da Europa Ocidental	Suíça e Suécia
CWEAS – Europa oriental	Moldávia
CARLA – região do Caribe	Cuba e Jamaica
EMDEU – norte da Europa	Lituânia
ESEAS – sudeste da Ásia	Camboja, Filipinas, Malásia, Singapura
NSWUA – sul da Ásia	Paquistão, Arábia Saudita e Irã
SUSAF – África Subsaariana	Camarões e Republica Centro Africana

Fonte: autor

Os países que fazem parte de cada uma das regiões serão apresentados no item 5 juntamente com os respectivos mapas geográficos para facilitar a visualização e localização no globo terrestre.

No estudo atual houve a inclusão de Cuba e os dados ausentes na tabela original foram obtidos do portal Knoema que é um Atlas de Dados que utiliza dados de fontes internacionais como ONU, Banco Mundial e Outras fontes de renome internacional

(<http://pt.knoema.com/atlas/Cuba/Pessoas-com-desnutri%C3%A7%C3%A3o>).

1.2 – Estrutura do Trabalho Final

Como o estudo envolve o conjunto de 12 trabalhos que foram desenvolvidos ao longo do semestre, podemos resumir cada um deles no Quadro 2 com o título e o respectivo conteúdo:

Quadro 2: Título e conteúdo dos trabalhos

Título	Conteúdo
Análise Exploratória dos Dados	Estatística Descritiva das Variáveis Quantitativas e Média, Mínimo e Máximo das variáveis
Relação entre as variáveis	Stem-and-Leaf Display, Análise de Regressão, correlação, regressão e teste qui-quadrado
Regressão Linear Múltipla	Análise de Variância, <i>Cluster Variable</i> e Dendograma
Comparações	Teste de Hipóteses e Intervalo de Confiança, BoxPlot, Análise pela One-way ANOVA, Análise pelo Two-Sample T-Test and CI

Componentes principais	Loading Plot, Principal Component Analysis, Scree plot, One-way ANOVA
Análise de Conglomerados	Loading Plot, Scree plot,
Regressão Logística	Média, Análise de Variância, Intervalo de Confiança, Análise Discriminante, Fitted Line Plot, Loading Plot, Scree plot
Árvore de Decisão	Software SPSS – árvore de decisão (Tree) com Chad Extendido

1.3 – Software utilizado

No Trabalho 11 foi empregado o Software estatístico SPSS porque ele gera a Árvore de Decisão. De acordo com a IBM SPSS: Decision Tree – Árvore de decisão é uma ferramenta que nos ajuda a identificar grupos, descobrir relacionamentos entre eles e prever eventos futuros. O modelo gráfico gerado se assemelha a um organograma, muito embora seja chamado de árvore de decisão.

Os trabalhos de 1 ao 12, exceto o 11, foi empregado o software MINITAB.

Para o trabalho final esse software será utilizado em mais de 95% e apenas 5% para o SPSS.

2. ENTENDENDO OS DADOS

Aqui é importante destacar o que é o tema, quais são as principais as variáveis de estudo que serão denominadas indivíduos, a distribuição das regiões e os países que a compõem.

2.1 O tema de pesquisa: recursos básicos: água, alimento, energia (basic features: water, food, energy)

A análise dos conceitos que compõem o tema foi extraída do material de aula do Programa Países Sustentáveis - Guia GPS – Gestão Pública Sustentável uma publicação do Núcleo de Estudos Futuros da PUC-SP –Pontifícia Universidade Católica. Nesse estudo são apresentadas as informações teóricas de outros indicadores, cuja análise e interpretação estão sob responsabilidade de outros colegas desse curso.

2.2 Os indivíduos

Os indivíduos desta análise são os 132 países divididos em 3 regiões: AIBER, AVECO e Outros. Serão analisados por uma série de indicadores de referência internacional. Os dados analisados de cada região são as variáveis descritas a seguir.

2.3. As Variáveis do tema: recursos básicos: água, alimento, energia (basic features: water, food, energy)

Este tema de pesquisa apresenta vinte variáveis, sendo três categóricas e dezessete quantitativas. Elas ainda podem ser divididas em: índices sintéticos e variáveis-componentes ou indicadores, a saber:

- **Índices sintéticos:** São sete:
 - 1 – **SPI** - Índice de Progresso Social, Índice de Desenvolvimento Humano
 - 2 – **IDH**, publicado pelo PNUD – ONU,
 - 3 – Índice de **Governança**,
 - 4 – **EPI** (*Environmental Protection Index*) - Índice de Proteção Ambiental
 - 5 – **HPI** - *Happy Planet Index*
 - 6 – **OHI** (*Ocean Health Index*) - o Índice de Saúde dos Oceanos
 - 7 – Índice de **GINI**
 - 8 – **EV – water resources** (Nível de tratamento de águas residuais ponderadas pela ligação a taxa de tratamento de águas residuais)
- **Indicadores ou variáveis componentes:** Inclui oito variáveis como: Undernourishment (% of pop.) (DESNUTRICAÇÃO), Depth of food deficit (calories/undernourished person)(déficit alimentar ou calorias consumidas), Access to piped water (% of pop.)(água encanada), Rural vs. urban access to improved water source (absolute difference between % of pop.)(água tratada zona rural e urbana), Access to improved sanitation facilities (% of pop.)(água tratada zona rural e urbana), Access to electricity (% of pop.)(eletricidade), Quality of electricity supply (1=low; 7=high)(qualidade fornecimento eletricidade), Renewable internal freshwater

resources per capita (cubic meters) 2013(renovação água doce), conforme apresentado no Quadro 3 que classifica o tipo de variável e a unidade de medida.

Quadro 3. As 16 Variáveis do Eixo temático: Recursos Básicos: água, alimento e energia

Variável	Significado	Tipo	Unidade de Medida
Access to piped water (% of pop.)(água encanada)	Porcentagem da população que tem uma ou mais torneiras de água encanada tratada.	Variável Quantitativa	Percentual
Access to improved sanitation facilities (% of pop.)(Acesso a instalações sanitárias melhoradas)	Inclui ainda a população com melhores condições de saneamento, canalização de esgoto, fossas sépticas, latrinas com laje ou melhoradas para as condições sanitárias adequadas	Variável Quantitativa	Percentual
Rural vs. urban access to improved water source (absolute difference between % of pop.)(água tratada zona rural e urbana)	Acesso da água rural e urbana que é aproveitada para ser potável.	Variável Quantitativa	Percentual
Renewable internal freshwater resources per capita (cubic meters) 2013(renovação água doce)	Recursos de água doce internos renováveis em rios e águas subterrâneas interno da chuva.	Variável Quantitativa	Metros cúbicos
Depth of food deficit (calories/under nourished person)(Profundidade de déficit alimentar (calorias / pessoa subnutrida)	Ceteris paribus, considera o total de subnutridos multiplicado pelo número médio de consumo da dieta mínima. Isto revela o déficit de alimentos do país.	Variável Quantitativa	Percentual
Acesso a Eletricidade	Percentual de pessoas com acesso a energia elétrica.	Variável Quantitativa	Percentual

Qualidade no fornecimento de energia elétrica	Usando Escala Likert 1 não confiável e 7 extremamente confiável para mostrar a qualidade do fornecimento de energia elétrica.	Variável Quantitativa	Indicador nominal que varia de 1 a 7
Undernourishment (% of pop.) (desnutricao)	Percentual da população com ingestão de alimentos suficiente. Os dados com 5% ou menor indica desnutrição	Variável Quantitativa	Percentual
IDH – índice Desenvolvimento humano	Mede o progresso de um país por indicadores de qualidade de vida: renda, saúde e educação são os principais.	Variável Quantitativa	Indicador nominal que varia de 0 a 1
Governança	Indica como os governos são indicados. Como a autoridade do país é exercida e como as políticas são implementadas e qual a capacidade de formulá-las.	Variável Quantitativa	Percentual
EPI- Índice de desempenho ambiental	Classifica o desempenho dos países em questões ambientais principais: proteção à saúde humana e proteção do ecossistema ambiental	Variável Quantitativa	Percentual
HPI – Bem estar sustentável do país	Medida de bem estar sustentável país oferece uma vida feliz sustentável a longo prazo para as pessoas que vivem nelas.	Variável Quantitativa	Percentual
OHI – Índice de saúde do Oceano	Pontos de referência para a realização de dez objetivos sócio ecológicos e como os países colocam em prática para gerar a saúde dos oceanos.	Variável Quantitativa	Percentual
GINI index	Mede a distribuição de renda e de despesas das famílias	Variável Quantitativa	Percentual
Social Progress Index	Considera o progresso social de um país incluindo nutrição, água encanada, eletricidade, dentre outros.	Variável quantitativa	percentual
EV - Water Resources(tratamento águas residuais)	Mostra quanto da água utilizada é tratada e devolvida para a natureza	Variável quantitativa	percentual

Fonte: autor a partir dos dados da planilha estatística e do GPS p.16

2.3.1 A Tabela de Dados

Em função da extensão da tabela, vamos representar aqui apenas alguns países e **algumas** variáveis selecionadas para corroborar o Quadro 3.

Tabela 1: Países e variáveis selecionadas

país	Social Progress Index (índice de progresso social)	Undernourishment (% of pop.) (desnutrição)	Depth of food deficit (calories/undernourished person) (déficit alimentar ou calorias consumidas)	Access to piped water (% of pop.) (água encanada)	Rural vs. urban access to improved water source (absolute difference between % of pop.) (água tratada zona rural e urbana)	Access to improved sanitation facilities (% of pop.) (instalações sanitárias)
Albania	69,13	<5.0	52	81,8	1,8	93,9
Algeria	59,13	<5.0	22	73,7	6,0	95,1
Angola	39,93	27,4	176	20,0	31,6	58,7
Brazil	69,97	6,9	55	91,7	15,1	80,8
Bulgaria	70,24	<5.0	52	96,8	0,7	100,0
Burkina Faso	47,33	25,9	185	6,9	22,3	18,0
Canada	86,95	<5.0	5	87,7	1,0	99,8
El Salvador	64,70	12,3	80	71,8	12,8	70,0
Estonia	81,28	<5.0	23	90,3	2,3	97,9
Finland	86,91	<5.0	8	99,4	0,0	100,0
France	81,11	<5.0	2	100,0	0,0	100,0
Hungary	73,87	<5.0	9	94,3	0,0	100,0
Iceland	88,07	<5.0	6	100,0	0,0	100,0
India	50,24	17,5	125	25,2	6,8	35,1
Kuwait	70,66	<5.0	11		0,0	100,0
Paraguay	62,65	25,5	182	65,6	33,4	70,8
Peru	66,29	11,2	71	75,3	24,7	71,6
Philippines	65,86	17,0	106	42,7	0,7	74,2
Poland	77,44	<5.0	5	98,0		89,3
Portugal	80,49	<5.0	2	99,7	0,0	100,0
Russia	60,79	<5.0	12	82,0	6,6	70,4
Rwanda	49,46	28,9	196	3,6	13,1	61,3
Saudi Arabia	64,38	<5.0	18	66,7	0,0	100,0
South Africa	62,96	<5.0	16	68,8	19,7	74,0
Spain	80,77	<5.0	8	99,3	0,1	100,0
Sri Lanka	59,71	24,0	211	29,5	7,3	91,1
Venezuela	63,78	<5.0	16	86,8	19,1	90,9
Yemen	40,23	32,4	215	40,2	25,4	53,0
Zambia	49,88	47,4	345	15,0	35,9	42,1

Fonte: autor com base na Base de Dados da aula Métodos Quantitativos

Vale destacar que os dados da tabela da coluna Desnutrição foram corrigidos para o valor 5,0 em lugar do representado por <5.0.

3. ANÁLISE DAS VARIÁVEIS

Antes da apresentação e análise das variáveis, são apresentados os gráficos elaborados pelo recurso Pie Chart do MINITAB com as regiões AIBER, AVECO e OUTROS e seus respectivos países.

Gráfico 1: Participação das 8 regiões

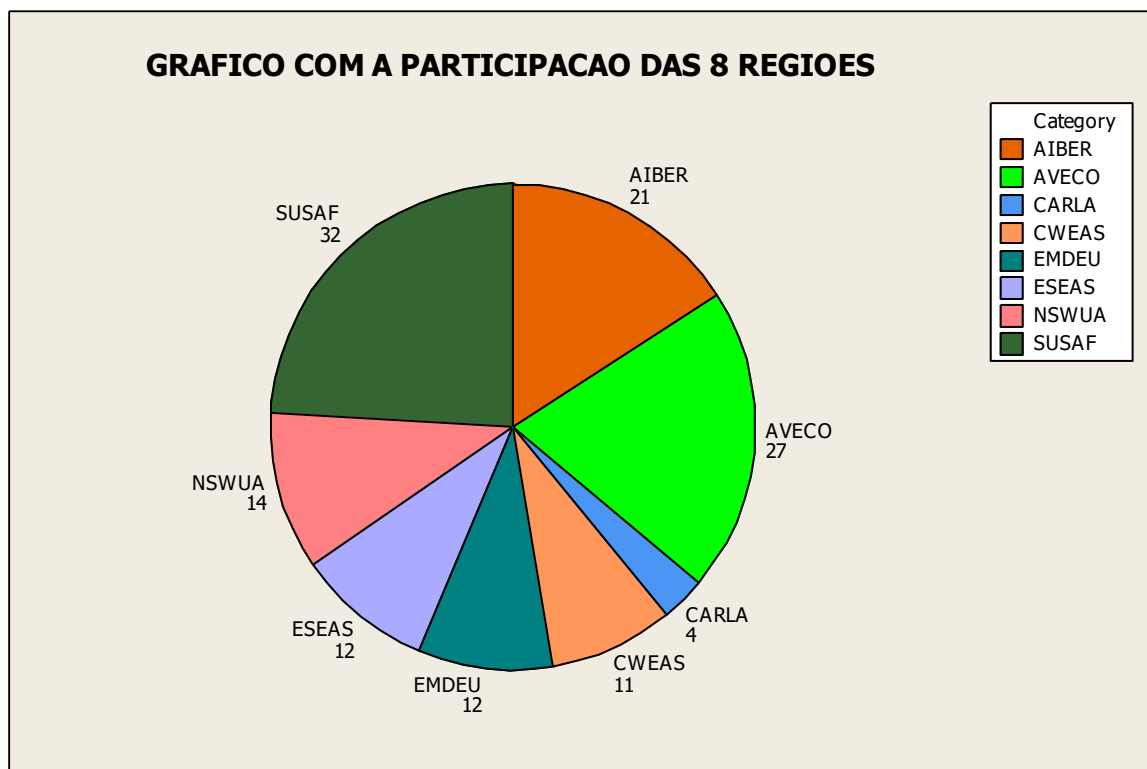


GRAFICO 2 – REGIAO AIBER COM SEUS PAISES

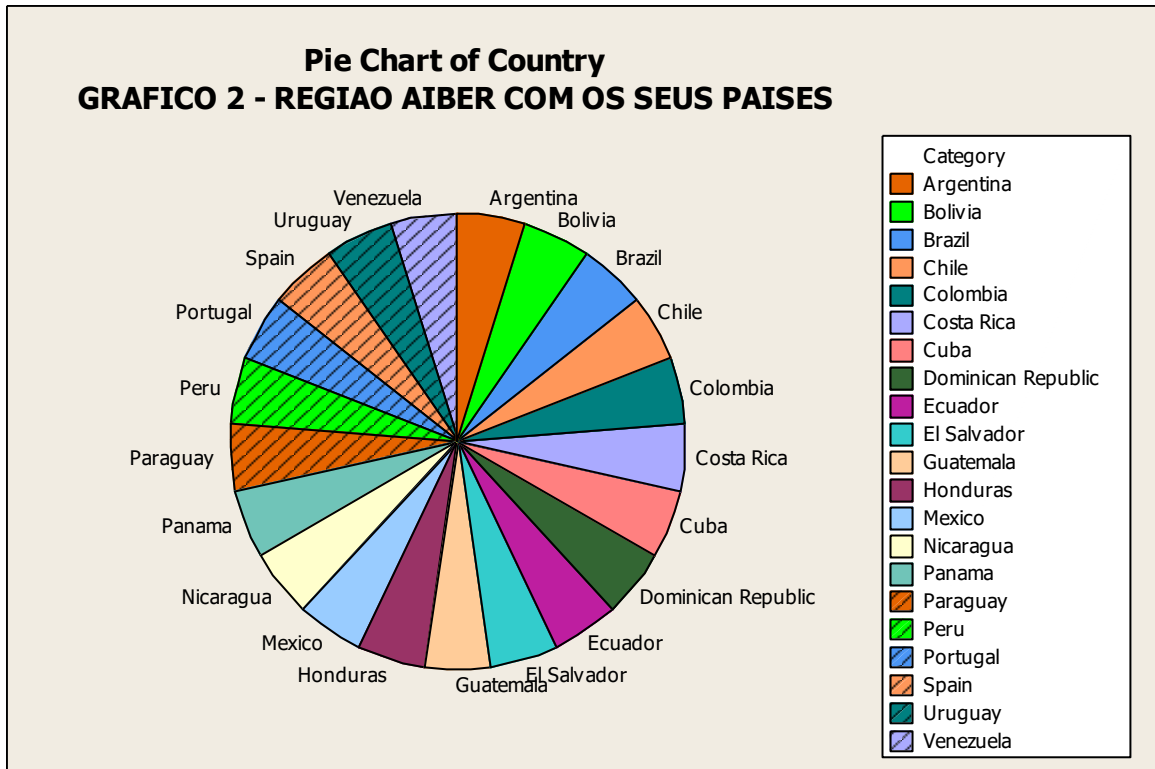
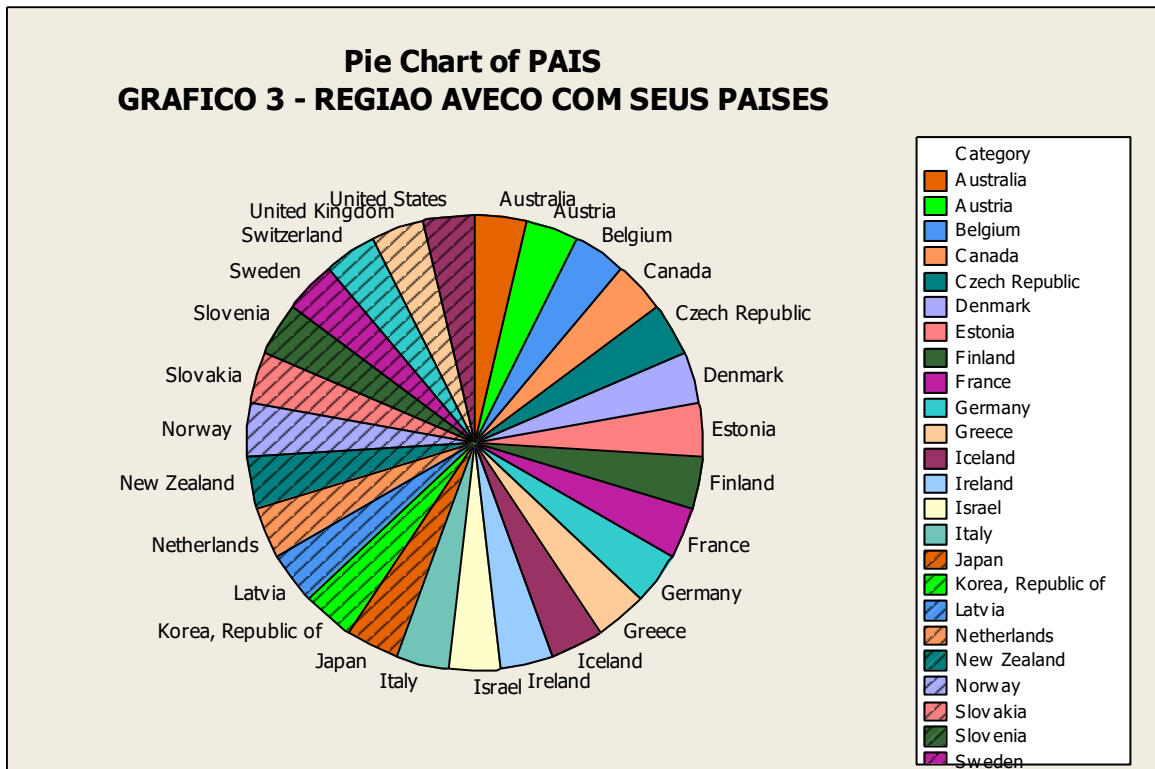
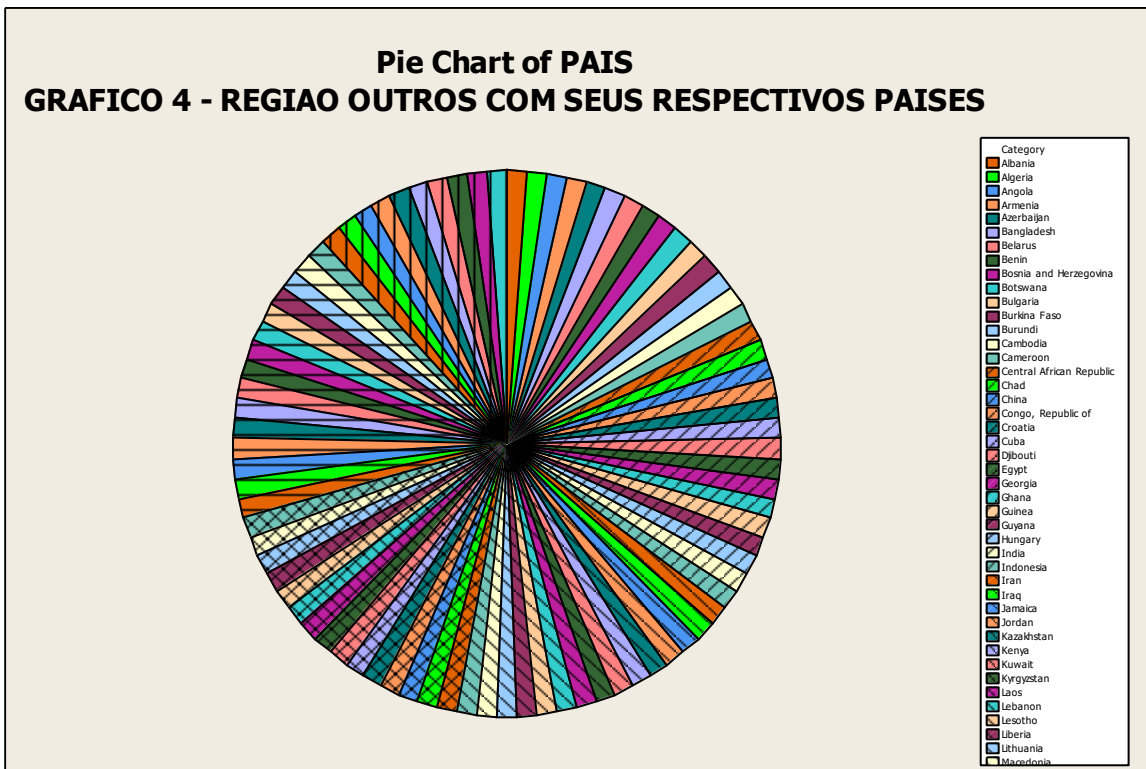


GRAFICO 3 – REGIAO AVECO COM SEUS PAISES





3.1 – Variáveis Quantitativas

Como apresentado terceira coluna e corroborado pelo indicador de medida na quarta coluna do Quadro 3, as variáveis são denominadas quantitativas. Essa informação aparece na Tabela 1 com uma seleção de países e indicadores com os números em percentual.

Com isso, a variável é denominada quantitativa pois mostra as que ela pode ser medida em uma escala quantitativa, isto é, com números (SANTOS; PARRA FILHO, 2011).

Elas são Variáveis Contínuas porque assumem valores fracionados já que estão em percentual (**IDH e GINI**).

Inclui oito variáveis componentes como: Undernourishment (% of pop.) (desnutricao), Depth of food deficit (calories/undernourished person)(déficit alimentar ou calorias consumidas), Access to piped water (% of pop.)(água encanada), Rural vs. urban access to improved water source (absolute difference between % of pop.)(agua tratada zona rural e urbana), Access to improved sanitation facilities (% of pop.)(agua tratada zona rural e urbana), Access to electricity (% of pop.)(eletricidade), Quality of electricity supply (1=low; 7=high)(qualidade fornecimento eletricidade), Renewable internal freshwater resources per capita (cubic meters) 2013(renovação água doce), conforme apresentado no Quadro 3 que classifica o tipo de variável e a unidade de medida. Estas são as variáveis ou indicadores analíticos.

- As variáveis sintéticas ou **Índices sintéticos**: São oito:
 1. – **SPI** - Índice de Progresso Social, Índice de Desenvolvimento Humano
 2. – **IDH**, publicado pelo PNUD – ONU,
 3. – Índice de **Governança**,
 4. – **EPI** (*Environmental Protection Index*) - Índice de Proteção Ambiental
 5. – **HPI** - *Happy Planet Index*
 6. – **OHI** (*Ocean Health Index*) - o Índice de Saúde dos Oceanos
 7. – Índice de **GINI**

8. – EV – water resources (Nível de tratamento de águas residuais ponderadas pela ligação a taxa de tratamento de águas residuais)

As variáveis categóricas são: País, Região e Código da Região.

Cada país está agrupado em uma determinada região que neste trabalho final está concentrado em três regiões: AIBER, AVECO e OUTROS, conforme identificadas nos gráficos 2,3,4.

4 – ANÁLISE EXPLORATÓRIA DOS DADOS

Como início do processo de análise e interpretação dos dados, foram realizados o tratamento das informações com o objetivo de deixá-los prontos para serem utilizados e sem margem de erro na pesquisa. Logo, avaliação de possíveis dados ausentes e a substituição deste pela média da amostra já foi realizada bem como correções de títulos e de nomes que pudessem levar a uma interpretação duvidosa do contexto. Para deixar as informações refinadas para esta pesquisa é apresentado o quadro Descriptive Statistics a seguir:

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1
Social Progress Index	132	0	55,84	2,22	25,52	0,00	34,74
Undernourishment (% of p	132	0	11,74	1,51	17,38	0,00	0,00
Depth of food deficit (c	132	0	12,42	1,32	15,22	0,00	1,21
Access to piped water (%)	132	0	63,05	3,03	34,85	0,00	24,39
Rural vs. urban access t	132	0	19,91	1,96	22,57	0,00	0,94
Access to improved sanit	132	0	70,97	2,78	31,93	0,00	48,02
Access to electricity (%)	132	0	77,43	2,92	33,57	0,00	56,25
Quality of electricity s	132	0	59,15	2,38	27,36	0,00	37,08
Renewable internal fresh	132	0	2,948	0,887	10,189	0,000	0,205
EV water resources	132	0	27,77	2,85	32,71	0,00	0,54
OHI	132	0	64,975	0,653	7,505	45,050	60,833
IDH - 2013	132	0	58,05	2,25	25,81	0,00	38,39
Governança	132	0	45,02	2,22	25,46	0,00	25,83
EPI Score	132	0	49,18	2,08	23,90	0,00	30,14
Happy Planet Index	132	0	48,45	1,88	21,56	0,00	33,77
GINI Index	132	0	37,07	1,84	21,13	0,00	21,52

ANÁLISE

Este quadro permite que sejam apresentadas todas as variáveis da pesquisa com seu respectivo nome, total de elementos da amostra, média, valor mínimo e máximo.

Como apresentado no item 3.1 são 16 variáveis ao todo distribuídas em 8 sintéticas como IDH, GINI e 8 variáveis componentes como Eletricidade e Desnutrição. As variáveis categóricas como País, Região e Código estão nos gráficos 2,3,4 e representam 132 países alocados em 3 regiões. Então, os dados da amostra estão perfeitos em termos de quantidade de informação.

A descrição das informações do quadro é:

- Valor de N é o total de elementos da amostra que nota caso ela é total: 132 países
- N* indica o total de dados ausentes na tabela em cada variável. Note que não existem dados ausentes informado pelo N* .
- Mean (média) mostra o valor médio da variável para os 132 países.
- Minimum mostra o valor mínimo dos dados dos 132 países
- Maximum mostra o valor máximo dos dados dos 132 países

4.1 – Análise Exploratória das variáveis da pesquisa

Inicialmente são apresentadas as variáveis sintéticas e depois as variáveis componentes.

Aqui são apresentados os gráficos gerados pelo Graphical Summary de acordo com a rotina:

Primeiro: clique em Stat em seguida passe o mouse sobre Basic Statistics e aparece uma janela onde o Graphical Summary é o terceiro de cima para baixo. Clique nele.

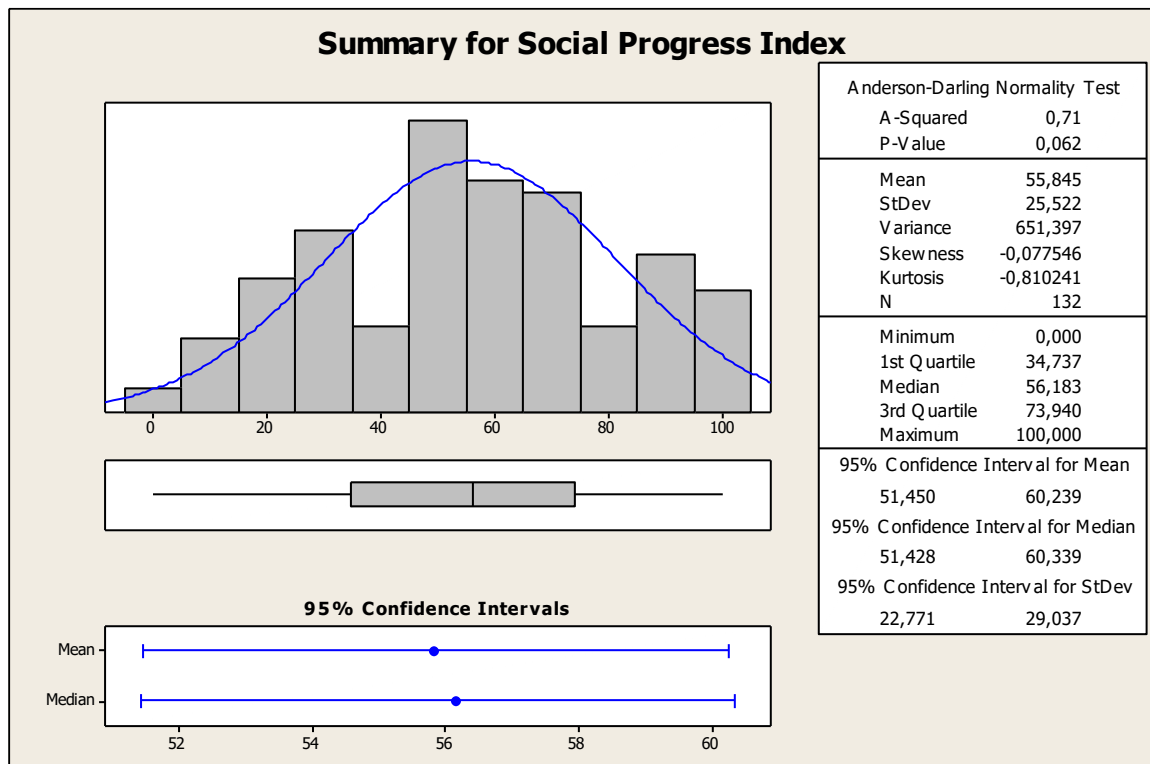
Segundo: Após clicar, abriu uma janela e você deve escolher a coluna que quer o gráfico na janela da Variables logo a primeira. Está escrito o grau de confiança de 95,0. Clique OK e o gráfico será gerado em uma janela.

a) SPI – Social Progress Index (variável sintética)

Variável: Social Progress Index – índice de Progresso Social

Origem e definição dado pelo portal: <http://pt.knoema.com/jklbfre/the-social-progress-index>

O Índice de Progresso Social é o resultado de um processo de dois anos de pesquisas envolvendo uma equipe renomada de especialistas que inclui, entre outros, os economistas Hernando de Soto e Michael Porter e do presidente da Fundação Rockefeller Dr. Judith Rodin. O índice sintetiza um total extenso de dados pesquisados para identificar as dimensões do desempenho das sociedades e medir o progresso social de forma abrangente e rigorosa. O Índice foi estruturado em torno de 12 componentes e 54 indicadores distintos consolidadas em três dimensões do Progresso Social: necessidades humanas básicas, Fundações de Bem-estar e oportunidade. A primeira dimensão, necessidades humanas básicas, avalia o quão bem um país prevê necessidades essenciais de seu povo através da medição se as pessoas têm comida suficiente e estão recebendo cuidados médicos básicos, se tiverem acesso a água potável, se tiverem acesso a uma habitação condigna com serviços básicos, e se eles são seguros e protegidos.



ANÁLISE

A princípio a distribuição aproxima-se da normal. O gráfico registra a pontuação cardinal no conjunto de observações para os 132 países analisados.

Percebe-se um número menor de países do que o esperado classificados entre 35 e 45 com apenas 7 observações, e o mesmo fenômeno se repete para a faixa de pontuação 75 à 85, que também apresenta apenas 7 observações. O maior número de observações encontra-se na faixa de 45 a 55, com 24 observações. Esta faixa corresponde a última faixa imediatamente

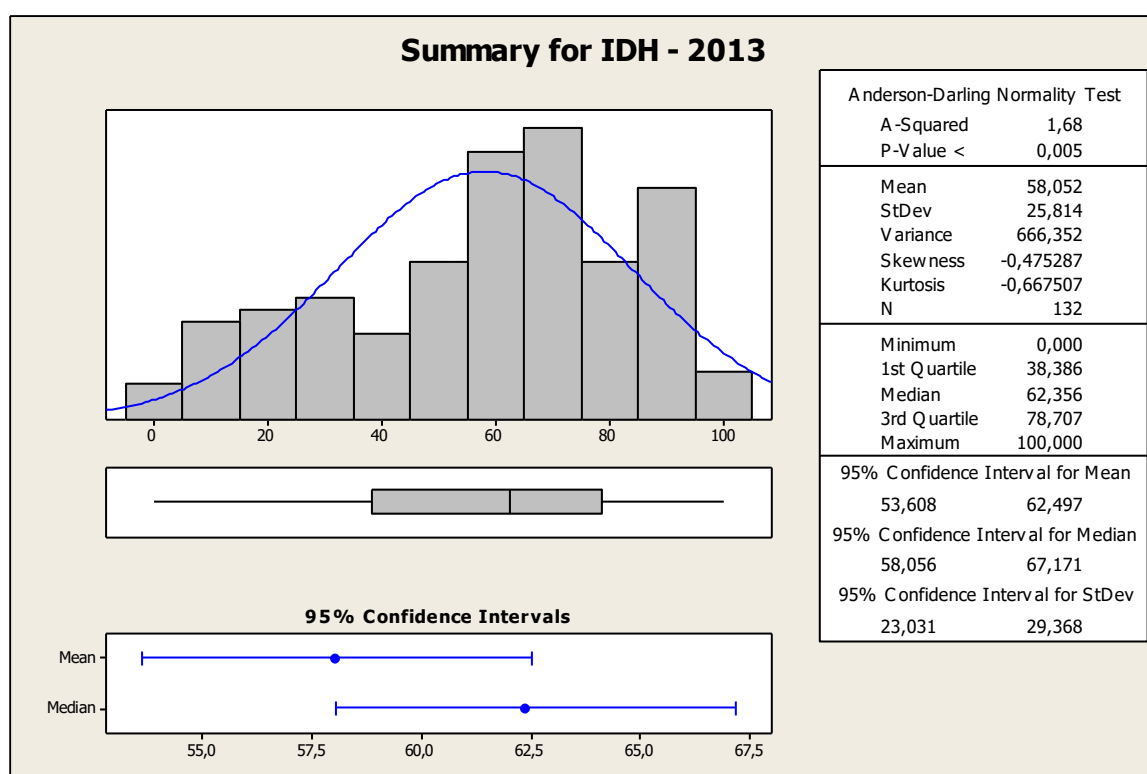
inferior à mediana e corresponde a um nível de desenvolvimento considerado “médio” pelo índice. A mediana de 56,183 e o terceiro quartil de 73,94 indicam o poder discricionário do índice, que consegue separar o grupo de países em 11 estratos bem definidos, o que pode ser bastante útil para classificações futura de dados.

b) IDH (2013) - Índice de Desenvolvimento Humano (IDH/PNUD)

O Índice de Desenvolvimento Humano (IDH) mede o progresso de uma nação a partir de três dimensões: renda, saúde e educação. Fonte: PNUD, 2013.

O trabalho publicado preliminar publicado por PNUD - Programa de las Naciones Unidas para el Desarrollo sobre o desenvolvimento humano em 2014 aponta que o IDH é progresso real da população em relação a expectativa de vida, educação, saúde, habitação, segurança e condições para um desenvolvimento sustentável.

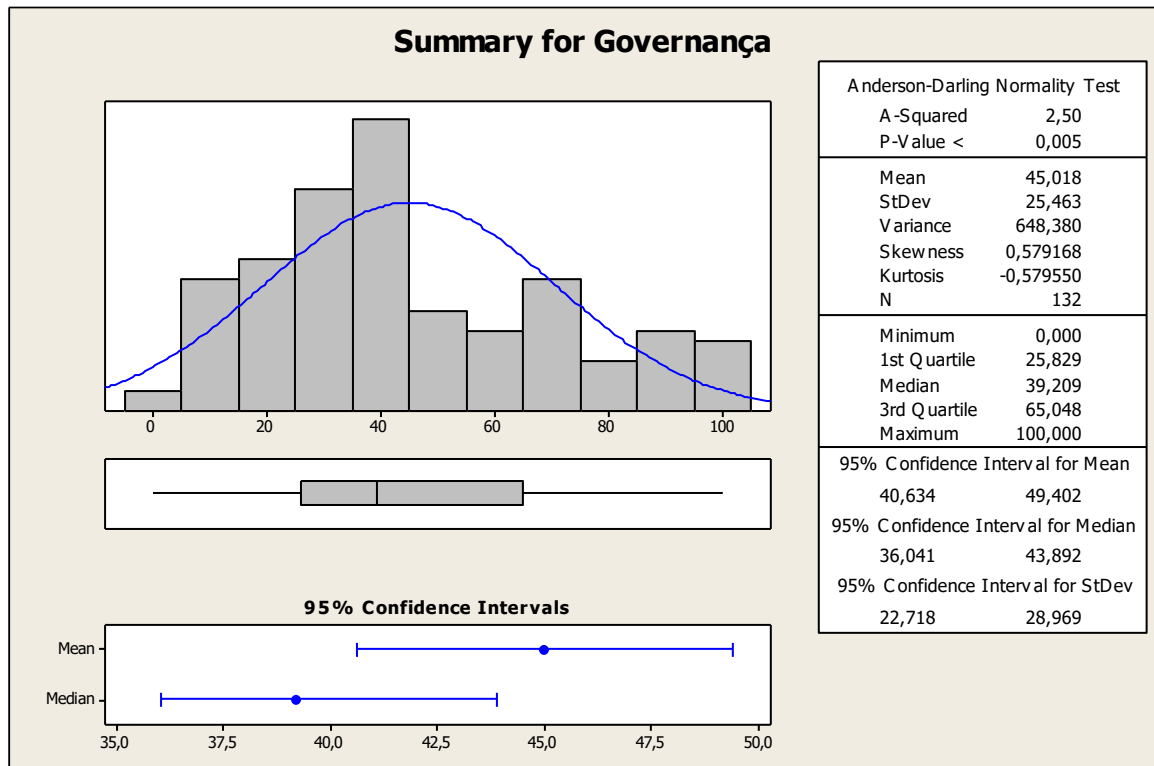
<http://hdr.undp.org/sites/default/files/hdr14-summary-es.pdf>



ANÁLISE

O gráfico para o IDH aproxima-se de uma distribuição normal, com uma concentração mais definida dos estratos deslocados para a direita, com uma concentração maior de países dos 65 aos 75 pontos, totalizando 24 países. Isso pode ser constatado também pela mediana, em 62,356. No estrato do 55 a 65 pontos, encontra-se o Paraguai, com IDH de 55,85 pontos. Do lado extremo esquerdo do gráfico estão 3 países no estrato de -5 e 5 pontos. Estes são considerados países de baixo desenvolvimento ou subdesenvolvidos.

- c) **Governança: World Economic Fórum** ressalta que as estratégias de atuação: Simplificar os processos administrativos e fazer uso de alianças regionais e de tecnologia para coibir a corrupção e os abusos; aprimorar os instrumentos e mecanismos regionais que possibilitem uma ação coletiva capaz de prevenir violações da democracia e garantir a preservação da ordem democrática ([http://www3.weforum.org/docs/LA11/WEF LA11 Report PT.pdf](http://www3.weforum.org/docs/LA11/WEF_LA11_Report_PT.pdf)).

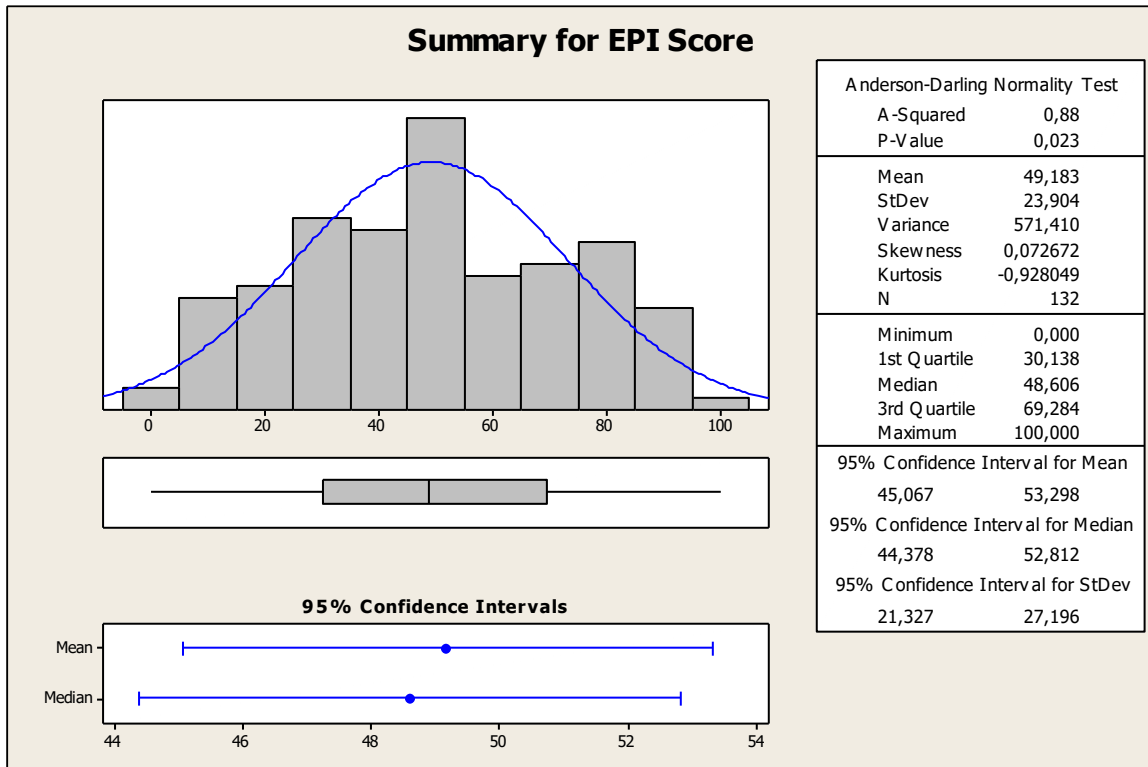


ANÁLISE

O gráfico para a variável GOV aproxima-se de uma distribuição normal deslocada para a esquerda, o primeiro quartil de 25,829 e a mediana de 39,209 confirmam esse deslocamento, indicando uma concentração crescente de países entre as faixas de 5 a 15 (13 países), de 15 a 25 (15 países), de 25 a 35 (22 países) até chegar ao pico – a faixa 35 a 45 (29 países). Essas faixas agrupam o equivalente a 60% dos países analisados.

Em relação ao pico, os países mais bem colocados são a Macedônia (44,74), a Turquia (44,69) e o Kuwait (44,24) e na faixa inferior ficam Argentina (35,30), Benin (35,75) e Indonésia (35,96). Aparentemente trata-se de grupo bastante heterogêneo, não sendo possível estabelecer relações imediatas.

d) EPI - Pontuação no Índice de Proteção Ambiental



ANÁLISE

A análise dos dados do Índice de Proteção Ambiental apresenta uma distribuição normal, com média de 49,183 e mediana de 48,606, ou seja, uma aproximação para o centro do gráfico. Há apenas um país no estrato superior (nota normalizada acima de 95), que é a Suíça, com EPI de 87,67, seguida por 9 países no segundo estrato (de 85 a 95), sendo todos países de alto IDH e IPS (Austrália, República Tcheca, Alemanha, etc) sendo o último país do estrato a Dinamarca com EPI de 76,92.

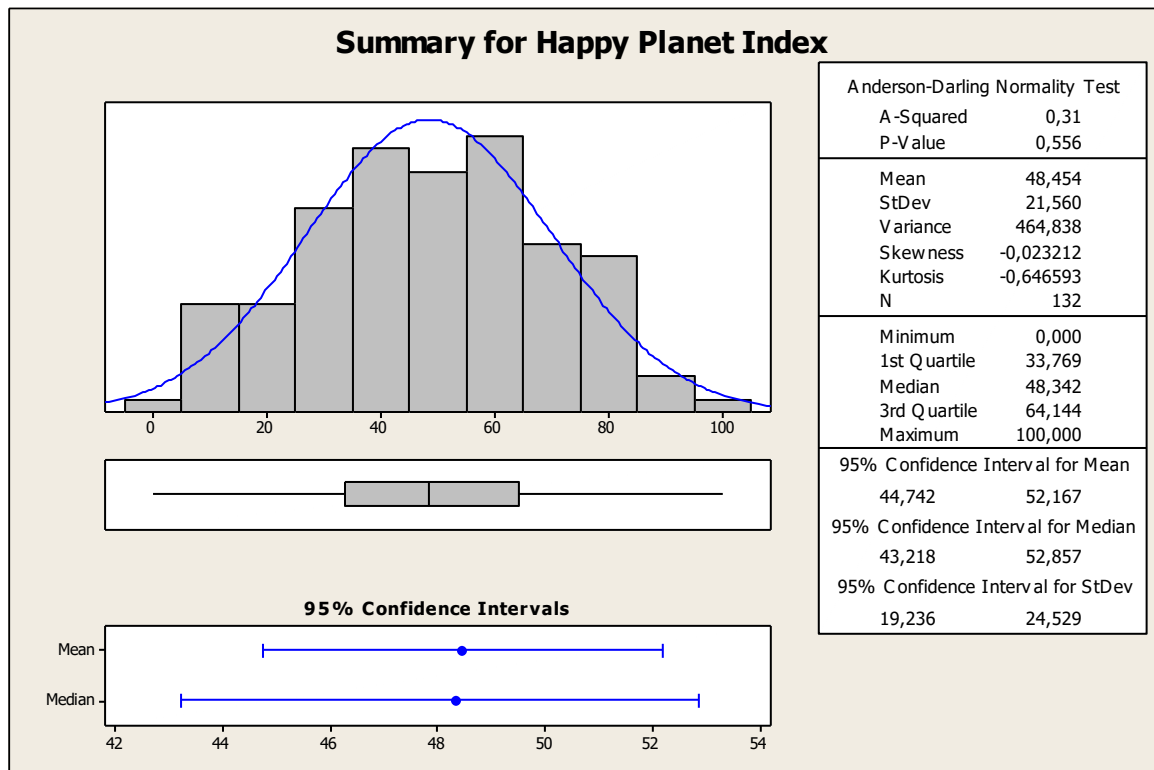
Já nos dois estratos inferiores temos a seguinte composição: no primeiro estrato (nota normalizada até 5) há dois países – Mali (EPI = 18,43) e Lesoto (EPI = 20,81), seguidos por um grupo de 10 países no segundo estrato que vai de 5 a 15. Destes 10 países, 9 encontram-se na África, sendo Bangladesh a única exceção do grupo. O país melhor colocado é Angola, com EPI de 39,93.

Já a faixa de pico, que vai de 45 a 55, é composta da seguinte maneira: na parte inferior, Argélia (EPI = 50,08) e Líbano (EPI = 50,15) seguindo por um grupo heterogêneo de 26 países nos quais Jordânia (EPI = 55,78) e Montenegro (EPI = 55,52).

e) HPI - Pontuação no Happy Planet Index

O Índice de Felicidade Bruta mede aspectos subjetivos e objetivos da vida humana, evidenciando pesquisas que apontam que, a partir de certo nível de renda, o nível de felicidade médio reportado não aumenta na mesma proporção do aumento da renda, até decaindo com o tempo. Outro tópico relaciona é a questão da resiliência construída por pessoas em situação de privação em países de baixo IDH e IPS, que acabam adaptando-se às situações adversas.

Uma análise mais aprofundada poderá revelar o baixo nível de correlação deste indicados com o PIB e inclusive com o IDH e o IPS, pois sua metodologia é bastante distinta.



ANÁLISE

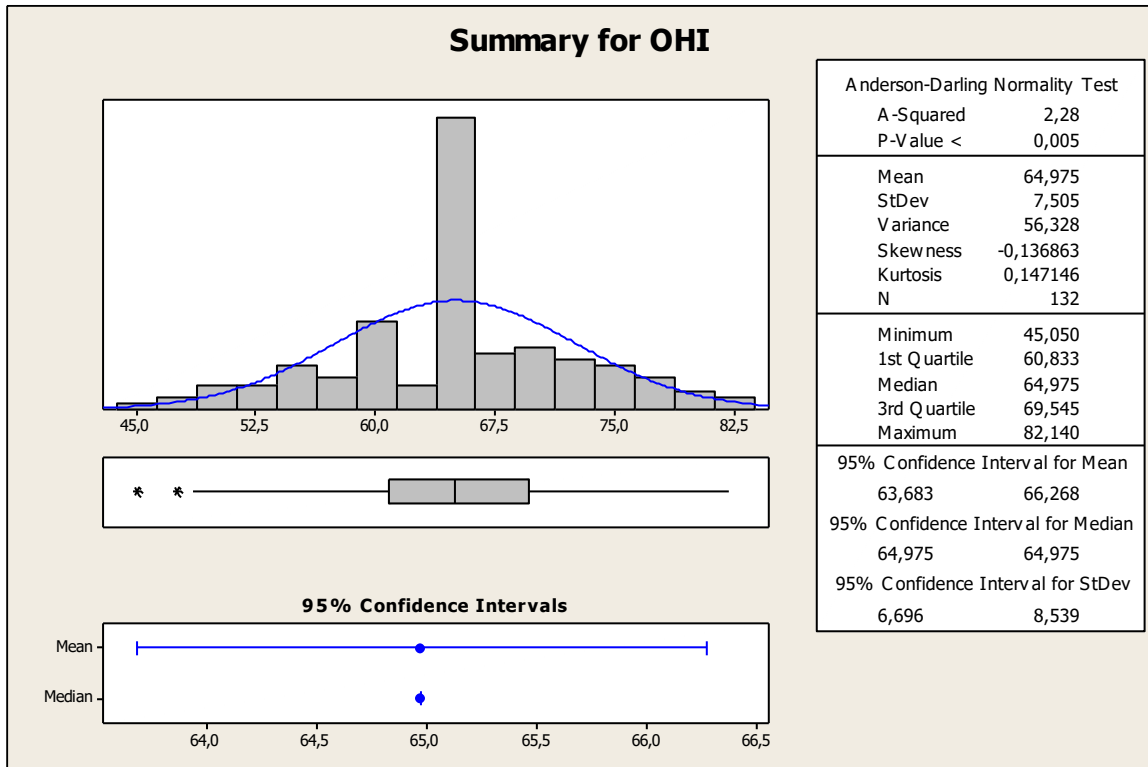
A distribuição para o Índice de Felicidade Bruta (FIB) aproxima-se de uma normal, com maior concentração de países entre a marca do 35 até 65, que formam quase um plateau, com os três maiores picos da distribuição. Nas extremidades encontramos um pequeno número de países : Costa Rica ficou um primeiro lugar, com um FIB de 64,03, seguida por um estrato de notas normalizadas de 85 a 95, que é composto por três países : Colômbia (FIB=59,75), El Salvador (58,88) e Jamaica (58,53).

No estrato inferior, que vai até 5 pontos, figura também apenas um país :Botswana, com FIB = 22,59, seguida pelo segundo estrato, que vai de 5 a 15 pontos e compreende 9 países. A menor pontuação é do Chade (FIB=24,68) e a maior é da Macedônia (FIB=28,27). Chama a atenção neste grupo de países a presença da África do Sul, que figura com um FIB de 28,19, o oitavo pior colocado na classificação geral do FIB.

A maior concentração em número de países está alocada entre 35 e 65 pontos, sendo Djibouti (FIB=37,23), Estados Unidos (FIB=37,34) e Hungria (FIB=37,40) os piores colocados do grupo, e Sri Lanka (FIB=49,38), Iraque (FIB=49,19) e Laos (FIB=49,14) os melhores colocados do grupo.

f) OHI - Pontuação no Índice de Saúde dos Oceanos

Este indicador mostra como os países estão cuidando de seus oceanos que são fontes de alimento, de riqueza e de poder geopolítico.



ANÁLISE

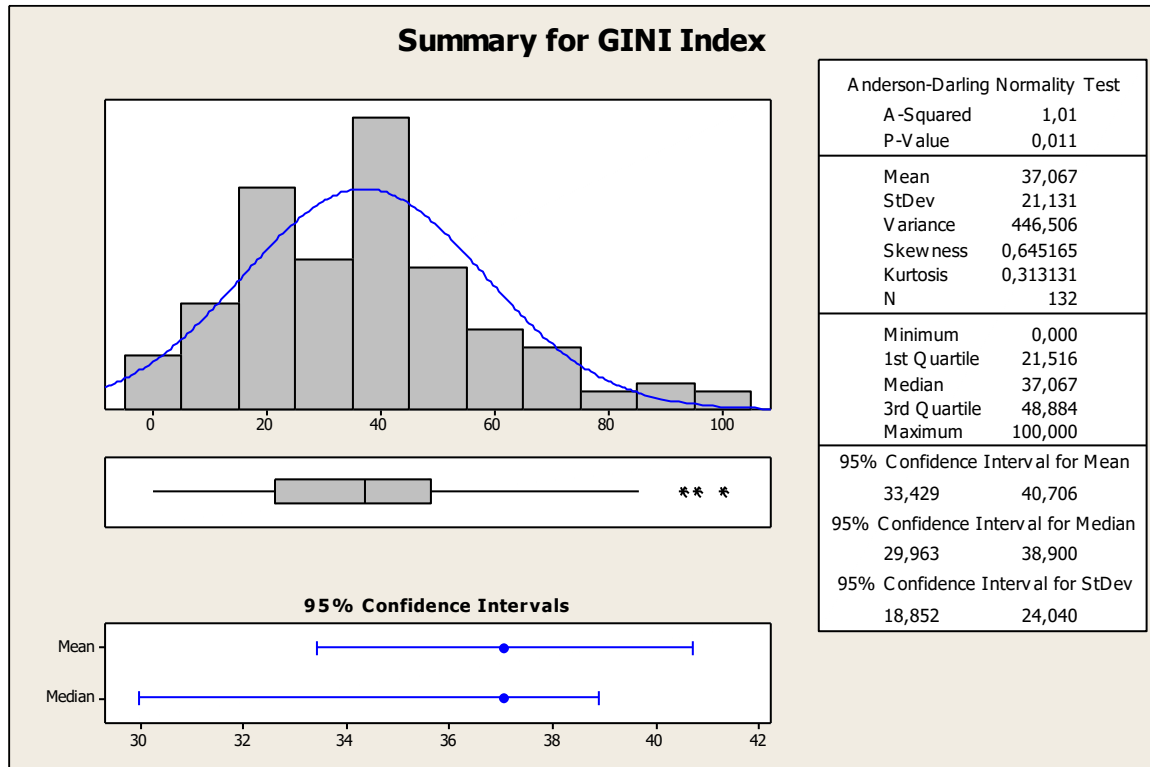
O gráfico para o Índice de Saúde dos Oceanos (OHI) chama atenção pela grande concentração de países entre o marco do 45 e 55 na escala normalizada (46 países), o que equivaleria as pontuações de Argélia (62,01), Namíbia (62,96) e Quênia (64,12) no estrato inferior e Togo (65,19) e Japão (65,09) no estrato superior.

É preciso registrar, contudo, que faltavam dados para 34 países do OHI, e que conforme relatado acima, aos valores faltantes, foi atribuída a média das pontuações do OHI dos países disponível no momento de elaboração deste estudo. A média foi de 64,975, que foi o valor atribuído a estes 34 países. Isso explica a grande concentração de países na faixa que vai de 45 a 55 do OHI, e por isso o pico não tem significância estatística.

Os melhores colocados foram a Dinamarca (82,14), Finlândia (81,4) e Noruega (80,11) e os piores colocados foram a Nicarágua (45,05), a Libéria (47,54) e Angola (42,66).

g) Índice de GINI

O Índice de Gini mede a desigualdade de renda de um país em uma variação de 0 a um. Quanto mais perto de um melhor é a distribuição de renda e mais perto de zero pior é a distribuição de renda. Em outras palavras, quanto mais perto de um menor o número de população pobre e quanto mais perto de zero maior a quantidade de pessoas pobres.



ANÁLISE

Para o Índice de Gini, que mede a desigualdade de renda, faltaram dados relativos a 13 países, que foram substituídos pela média da distribuição original, que foi de 39,02, que normalizado de 0 a 100 tornou-se 62,93, o que explica o pico registrado na coluna entre 55 e 65 – dos 33 países, na realidade apenas 20 tem sua pontuação original registrada neste intervalo.

Assim percebemos uma distribuição basicamente uniforme, com tendência para concentração de países mais a esquerda, como já demonstra o primeiro quartil, em 21,516.

São 6 países no estrato superior, acima de 95 pontos na escala normalizada, conforme a Tabela 2, abaixo, que como podemos ver é formada pelos antigos países socialistas

Tabela 2 – Países com menor índice de Gini

PAÍS	CÓD	REGIÃO	GINI
Ukraine	UKR	Eastern Europe	24,820
Slovenia	SVN	Southern Europe	24,870
Sweden	SWE	Northern Europe	26,080
Czech Republic	CZE	Eastern Europe	26,390
Belarus	BLR	Eastern Europe	26,460
Slovakia	SVK	Eastern Europe	26,580

Os países com maior desigualdade de renda estão divididos nos três primeiros estratos : de 5 a 15 (2 países – África do Sul e Namíbia), de 15 a 25 (3 países – Botswana, Zâmbia e Honduras) e de 25 a 35 (2 países – República Central Africana e Lesoto).

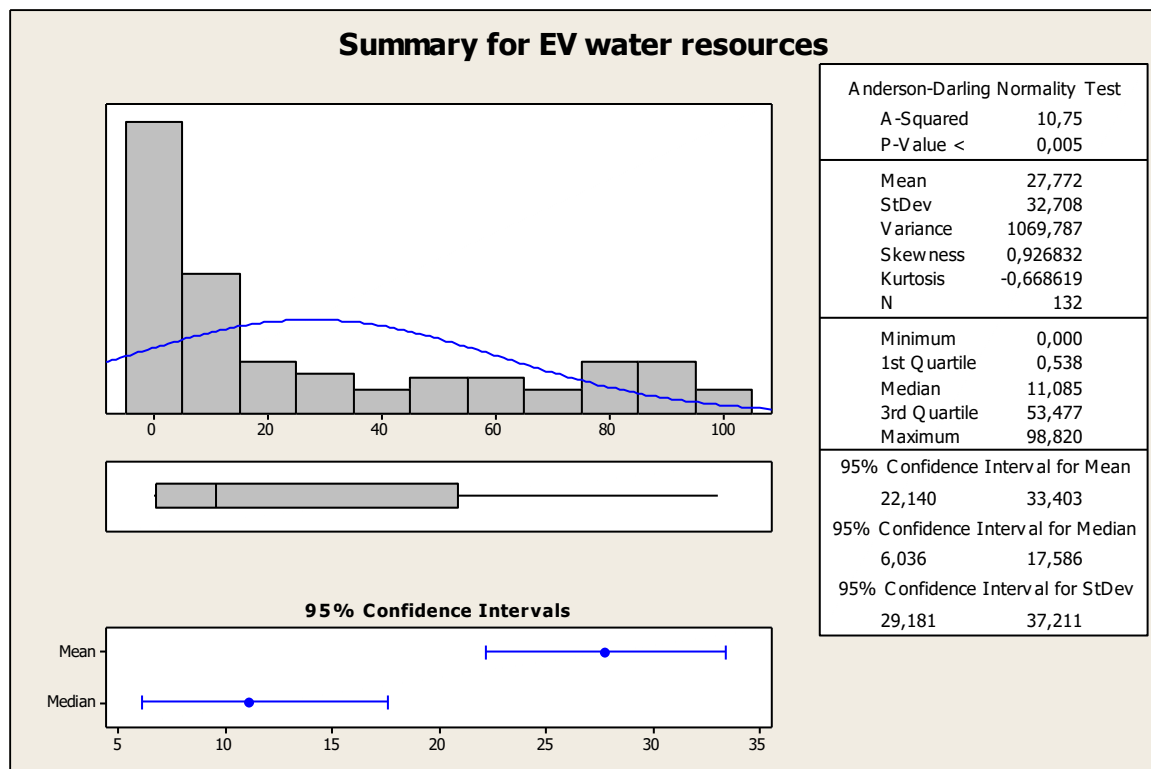
Tabela 3 – Países de maior índice de Gini (mais desiguais)

PAÍS	CÓD	REGIÃO	GINI
Lesotho	LSO	Southern Africa	54,170
Central African Republic	CAF	Eastern Africa	56,300
Honduras	HND	IBE	57,400
Zambia	ZMB	Middle Africa	57,490
Botswana	BWA	Southern Africa	60,460
Namibia	NAM	Southern Africa	61,320
South Africa	ZAF	Southern Africa	63,140

Botswana, Namíbia e África do Sul aparecem como *outliners* no gráfico.

h) EV – water resources

Este item trata do nível de tratamento de águas residuais ponderadas pela ligação a taxa de tratamento de águas residuais.



ANÁLISE

O gráfico mostra uma concentração de países no estrato Zero com um total de 50. Isto mostra que os países não estão se preocupando com o tratamento de águas residuais preferindo descartar de qualquer maneira e que a natureza tome conta do processo. A média de 27,772 é

baixa para o indicador deixando poucos países próximo do índice 100 que na análise corrobora o perfil de países desenvolvidos como Holanda (97,68) e Reino Unido (98,19).

Agora as variáveis componentes que vai do item i até p.

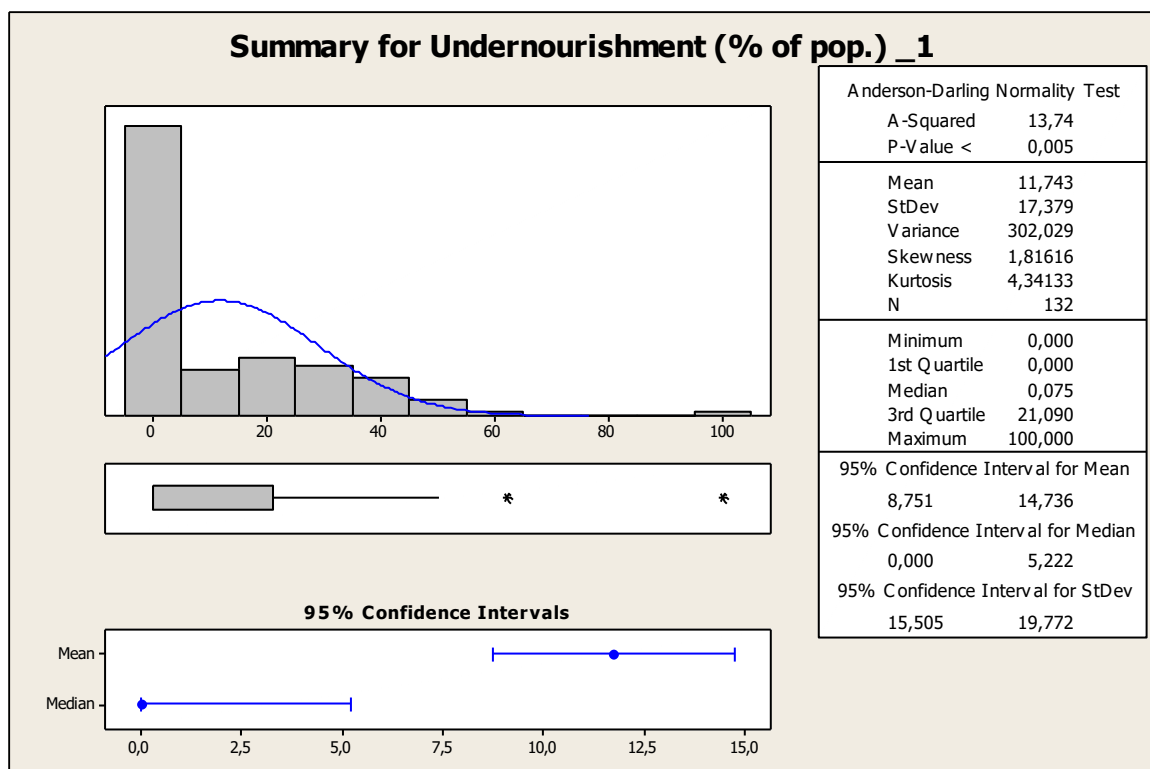
i) Undernourishment (% of pop.) (DESNUTRICA0)

Definição dado por:

<http://pt.knoema.com/search?query=undernourishment%20o%20que%20%C3%A9>

Proporção da população em estado de subnutrição. A desnutrição refere-se à condição de pessoas cujo consumo de energia alimentar é continuamente abaixo de um requisito de energia da ração mínima para a manutenção de uma vida saudável e da realização de uma atividade física leve, com um corpo de peso mínimo aceitável para atingido altura.

População abaixo do nível mínimo de consumo de energia da dieta (também referida como prevalência de desnutrição) mostra a percentagem da população com a ingestão de alimentos é insuficiente para atender às necessidades de energia na dieta de forma contínua. Mostrando os dados como 5 significa uma prevalência de desnutrição abaixo de 5%.



ANÁLISE

A análise na nossa pesquisa indica que a distribuição com a curtose (curva do gráfico) para a esquerda o que indica que os dados estão entre -5 e 5,0 com 77 observações. Nessa situação indicada pelo MINITAB está a Malásia com valor zero.

Na análise apresentada pelo Knoema em 2009 mostrava que a Malásia juntamente com Egito, Republica Dominicana, Jordânia e Kwait apresentavam fatores igual a 5,0 %.

Percebe-se um número menor de países do que o esperado classificados entre 45 e 55 com apenas 4 observações, e 10 observações para a faixa de pontuação 35 à 45. A mediana de

0,15 e o terceiro quartil de 21,09 indicam o poder discricionário do índice, que consegue separar o grupo de países em 6 estratos bem definidos, o que pode ser bastante útil para classificações futura de dados.

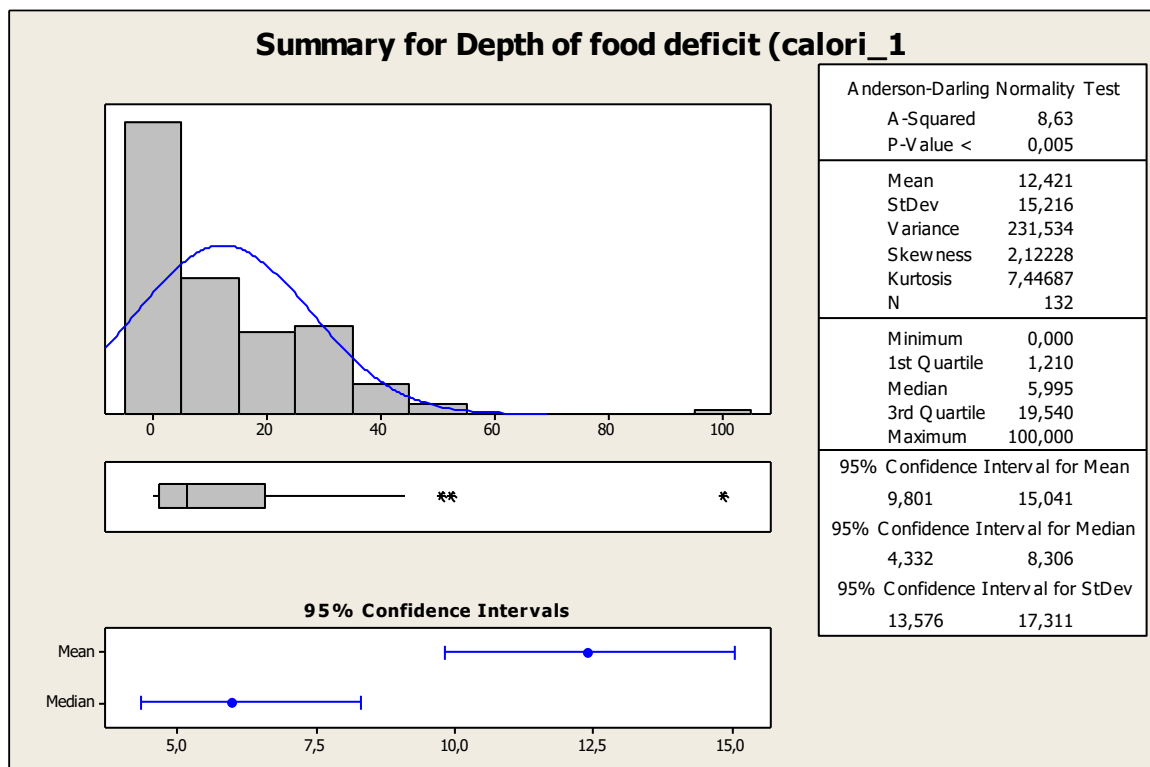
j) Depth of food deficit (calories/undernourished person)(déficit alimentar ou calorias consumidas)

Mostra a quantidade de calorias consumidas por cada habitante dentro de cada país. Pode dar uma ideia da oferta de alimentos, do poder de compra de cada pessoa para comprar alimentos, qual a capacidade do país de manter a população com saúde e mais agregado o Índice de Progresso Social.

De acordo com <http://www.fao.org/docrep/018/i3434e/i3434e.pdf> de 1990 para 1992 houve uma diminuição de 17% no total de pessoas subnutridas. No período 2011-2013 de cada oito pessoas no mundo pelo menos uma estava sofrendo de fome crônica, ou seja, com consumo de alimento muito inferior a qualquer dado considerado ideal. Este volume representava cerca de 842 milhões de pessoas sem comida suficiente para ter uma vida ativa.

Na nossa análise encontramos uma concentração de 60 países com intervalo de -5 e 5% no consumo de calorias diárias. Apenas 6 entre 35 e 45 % das calorias.

O estudo da FAO – FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS aponta ainda que subnutrição e desnutrição podem coexistir. No entanto, em alguns países, as taxas de desnutrição, indicado pela proporção de crianças raquíticas, são consideravelmente mais elevada do que a prevalência de subalimentação, tal como indicado pela incapacidade de abastecimento de energia da dieta. Nesses países, intervenções de melhoria da nutrição são cruciais para melhorar os aspectos nutricionais de segurança alimentar. Melhorias exigem uma gama de segurança alimentar e intervenções nutricionais-reforço na agricultura, saúde, higiene, abastecimento de água e educação, especialmente dirigidas às mulheres (2013).



ANÁLISE

Este gráfico apresenta semelhança estética com o gráfico da Desnutrição do indivíduo com a distribuição com a curtose (curva do gráfico) para a esquerda o que indica que os dados estão entre -5 e 5,0 com 61 observações. Dentre os países desta pesquisa estão: África do Sul (2,28), Gana (2,73), Tunisia (0,76) e Kwait (1,52). Nessa situação indicada pelo MINITAB está Cuba com 0,30.

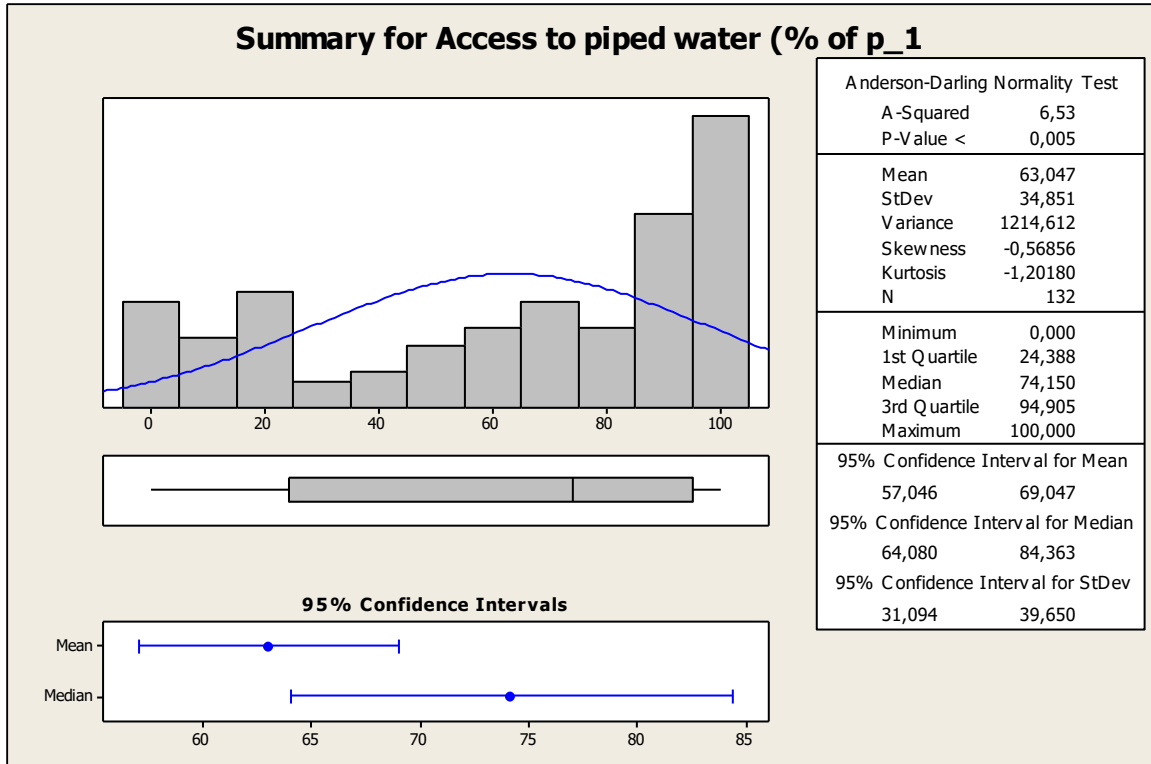
k) Access to piped water (% of pop.)(água encanada)

Access to piped water (% of pop.)(água encanada) mostra quantas residências possuem água encanada ou outro tipo de processo que leva a água tratada para o consumo.

De acordo com os dados da UNICEF (2010) a evolução da utilização de diferentes tipos de fontes de água a partir de 1990-2010, por regiões do MDG - Millenium Development Goals (objetivos de desenvolvimento do milênio) apresentou dois grupos distintos de evolução. O primeiro é um conjunto de regiões em que a utilização de água canalizada para uma habitação, terreno ou quintal é baixa (30 por cento ou menos). Ele inclui a África sub-saariana, Oceania, Sul da Ásia e do Sudeste Asiático. Embora os ganhos na utilização de água canalizada nas instalações foram feitas nessas regiões, o progresso foi principalmente na qualidade da origem da água. Destacamos que 65 por cento da população do sul da Ásia estão usando outras fontes naturais ao invés de água encanada no local.

O segundo conjunto de regiões é composto por Ásia oriental, África do Norte, Ásia Ocidental e na América Latina e no Caribe, onde pelo menos 70 por cento da população está usando água encanada. Ásia Oriental (destaque para a China) teve um aumento significativo no abastecimento de água canalizada desde 1990, ganhando 35 pontos percentuais na cobertura nesta categoria em 20 anos. Isto representa 562 milhões de novos usuários que foram adicionados durante um período em que o mundo como um todo aumentou apenas 9 %. Ásia Oriental é também a região com o aumento mais expressivo no uso de água potável de diversas fontes com qualidade, a partir de 68 por cento em 1990 e passou para 91 por cento de cobertura em 2010. Isso representa um aumento de 23 ponto percentual, muito maior que qualquer outra região.

<http://www.unicef.org/media/files/JMPreport2012.pdf>



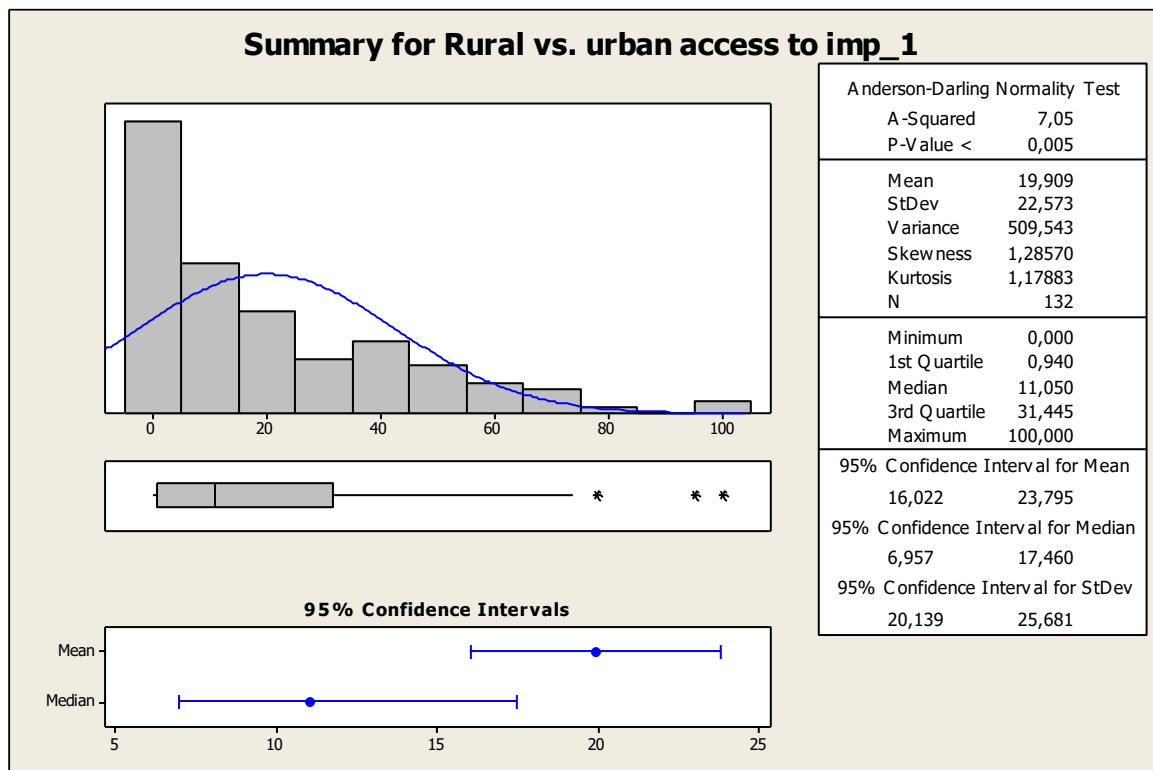
ANÁLISE

A análise da pesquisa mostrada no gráfico com uma tendência quase normal porque existem dois extremos quase distintos. Do lado direito do gráfico estão 3 estratos distintos, sendo o estrato 95-105 com 33 observações com destaque para Áustria da Região AVECO com 100%. Do outro extremo, está o estrato próximo da origem com 12 países dentro de um intervalo de -5 e 5%, do lado deste estrato está o estrato com 8 países entre 5 e 15% e colado está o estrato com 13 países entre 15 e 25.

A mediana é 74,15 com o terceiro quartil com o valor de 94,905 com 11 estratos distintos e visíveis.

1) Rural vs. urban access to improved water source (absolute difference between % of pop.)(agua tratada zona rural e urbana)

Mostra o percentual da população que possui acesso a água tratada na zona rural em relação a zona urbana.

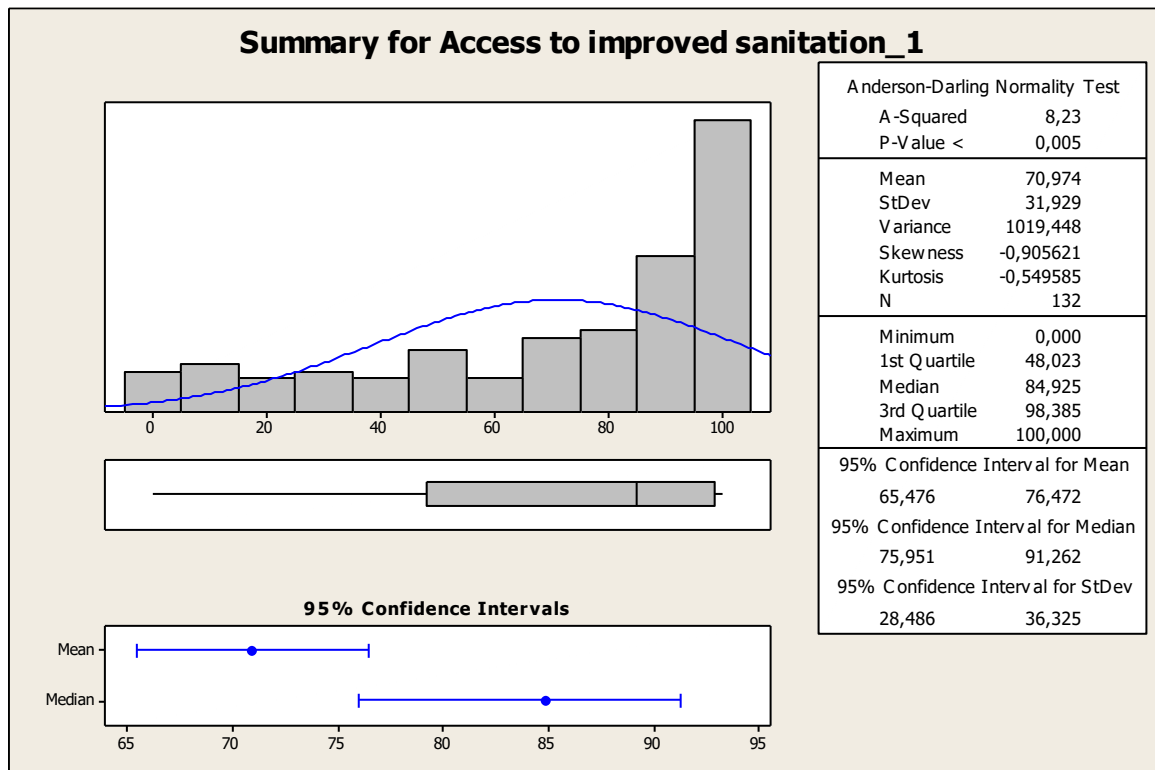


ANÁLISE

O gráfico apresenta uma curtose para a esquerda com a média de 19,109 e a mediana 11,05. Indicando que 50 países estão entre o indicado -5 e 5 positivo. Outro estrato bem acentuado está do lado direito com 25 países entre 5 e 15.

O gráfico identifica 3 outlier: um na linha 119 que é o Togo, outro na linha 91 que a República do Niger e por fim, o outro país é República do Congo. Eles estão com valores que expressam a falta de acesso a água com qualidade.

m) Access to improved sanitation facilities (% of pop)(acesso a instalações sanitárias melhoradas)



ANÁLISE

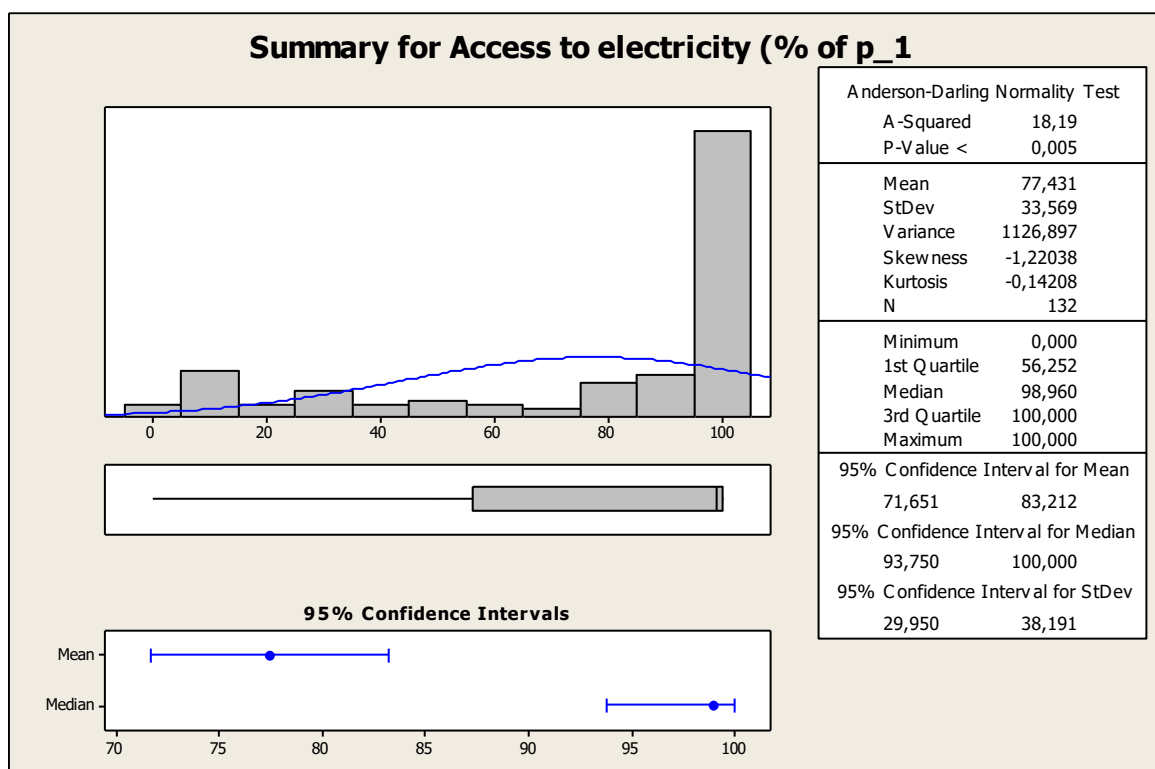
Os dados apresentados estão com a concentração do lado direito com uma média de 70,974 e mediana de 84,925, sendo que o terceiro quartil está com 98,385 indicando que esse estrato possui 43 países com melhores condições sanitárias.

Do lado oposto estão 6 países com indicadores -5 e 5 positivos. Do lado a este estrato estão 6 países com valores entre 5 e 15. Os países estão na tabela a seguir:

Tabela 4 – Países com valores entre -5 e 15 no item Acesso a instalações sanitárias melhoradas

País	Índice
Benin	5,08
Burkina Faso	9,33
República do Chade	2,38
República do Congo	9,06
Ghana	4,29
Republica da Guine	9,84
Libéria	9,53
Madagascar	4,51
Mali	13,31
Moçambique	10,48
Republica do Niger	0,00
Tanazania	2,58
Republica do Togo	2,01

n) Access to electricity (% of pop.) (eletricidade)



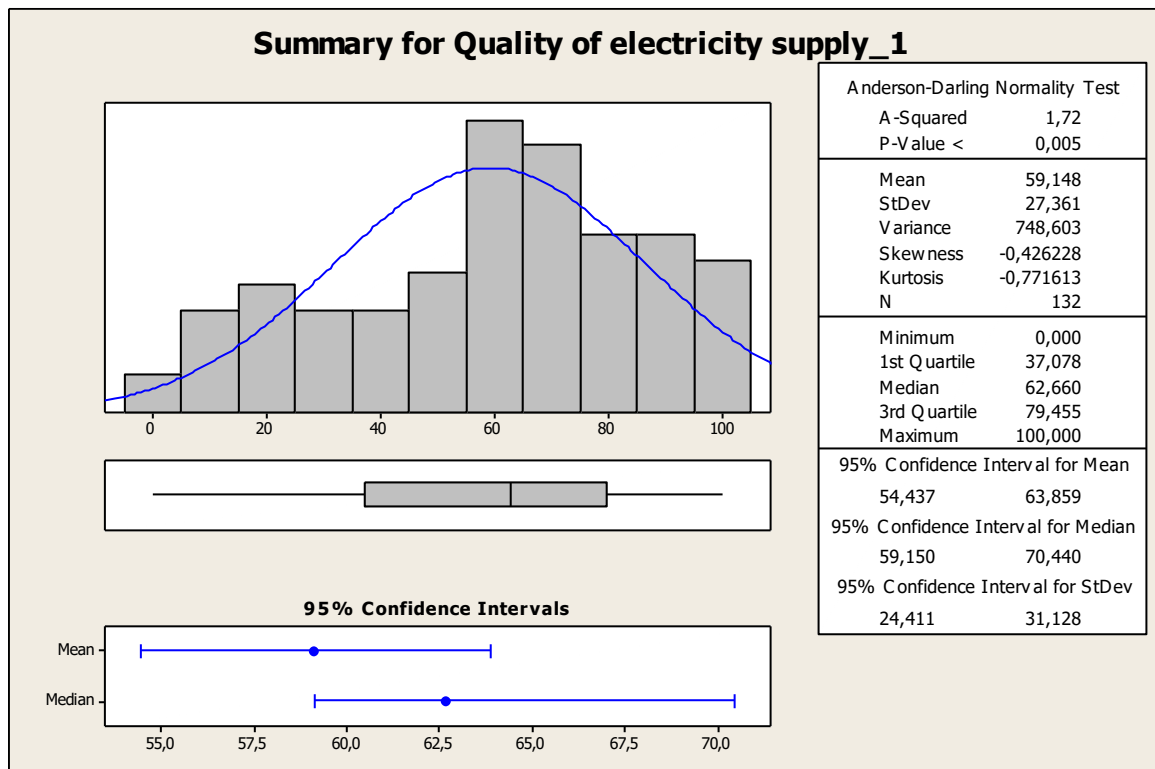
ANÁLISE

Access to electricity (% of pop.)(eletricidade)

Este indicador mostra o percentual da população de cada país que tem acesso a energia elétrica. De uma maneira geral existe um grande número de pessoas com acesso a energia elétrica com o estrato entre 95 e 105 para 76 países, com média de 77,431 e mediana de 98,96.

Isto quer dizer que metade da população mais significativa tem energia elétrica e outra metade bem menor não tem. Isto aparece no estrato entre 5 e 15 com 12 países.

o) Quality of electricity supply (1=low; 7=high)(qualidade fornecimento eletricidade)



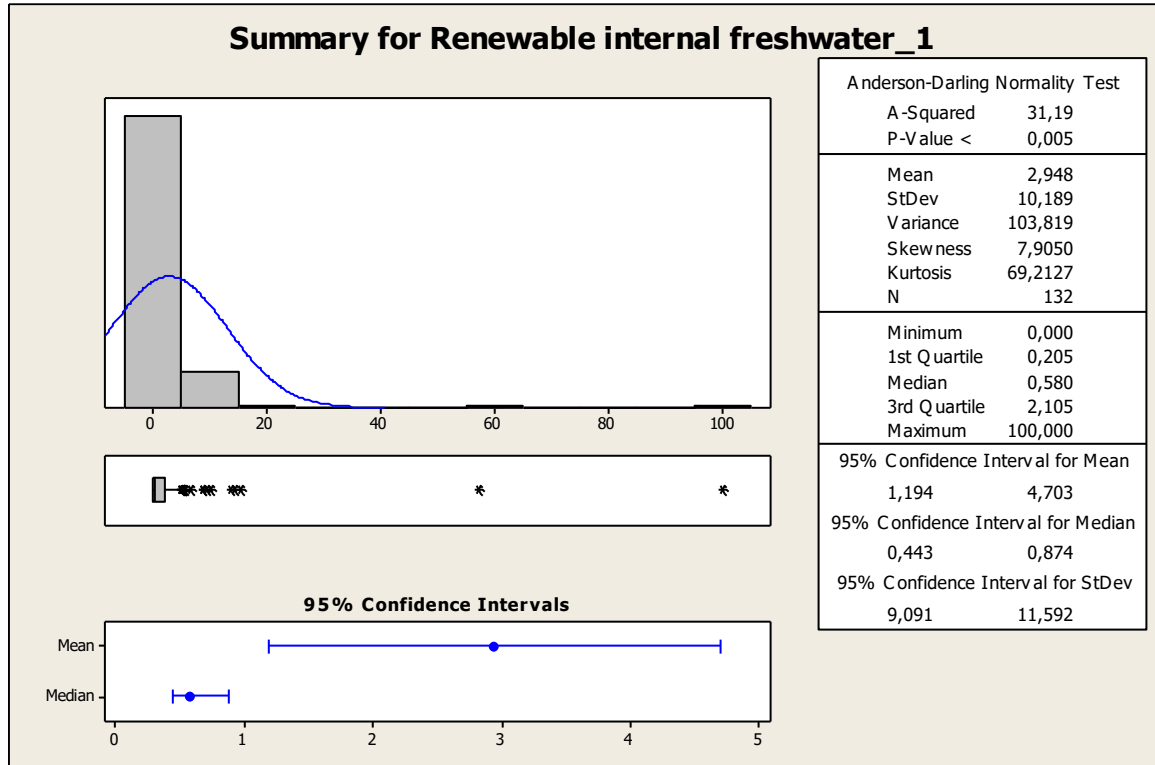
ANÁLISE

Quality of electricity supply (1=low; 7=high)(qualidade fornecimento eletricidade)

Ter acesso a energia não significa que ela esteja disponível 24 horas por dia ou sempre que o consumidor necessitar.

Esse item procura identificar a qualidade do fornecimento da energia e o gráfico mostra uma distribuição normal entre os dados. A média está com o valor de 59,148 e mediana 62,66, sendo que o terceiro quartil com 79,455 representado pelo estrato com 21 países.

p) Renewable internal freshwater resources per capita (cubic meters) 2013(renovação água doce)



ANÁLISE

A concentração dos valores estão voltados para o lado esquerdo do gráfico com estrato do terceiro quartil de 2,08 com 116 países.

Isto mostra que a renovação de água doce presente nesses países está sendo realizada em média de 2,948 com mediana de 0,58. Ou seja, é um valor muito pequeno considerando a necessidade de renovar por ser um recurso escasso.

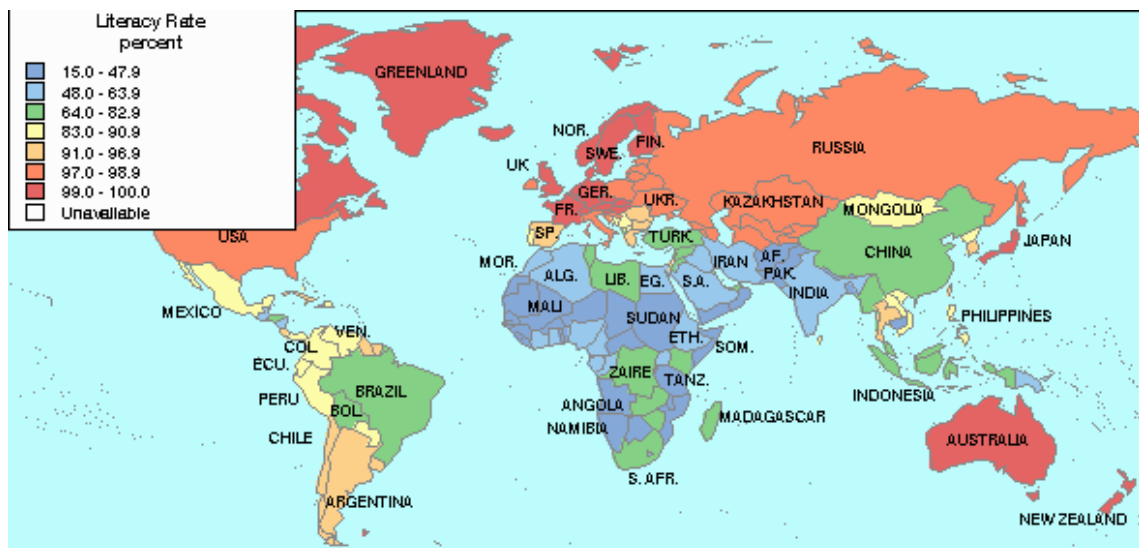
O gráfico mostra diversos outliers, sendo o mais extremo é a Islândia que tem renovação maior que os demais.

5 - RELAÇÃO ENTRE AS VARIÁVEIS E AS REGIÕES AIBER, AVECO e OTHERS

O item 4 mostrou a análise exploratória dos dados considerando as 16 variáveis (8 sintéticas e 8 componentes) com o objetivo de: identificar a média, mediana, desvio padrão, variância e a interpretação dos dados a partir da elaboração do gráfico histograma para melhor entender o significado e ao mesmo tempo conhecer cada variável.

O objetivo do item 5 é fazer um estudo exploratório mais clássico a partir da relação entre as variáveis procurando identificar e entender a correlação entre elas, fazer uma análise de regressão e o teste do Qui-Quadrado.

Até o item 4 foram apresentados os dados dos 132 países e seguir o mapa com alguns destes países:



Fonte: http://hdr.undp.org/sites/default/files/hdr2014_pt_web.pdf

5.1 – REGIAO AIBER

A partir deste item esta pesquisa trata exclusivamente dos dados das Regiões estabelecidas como objeto de pesquisa: AIBER, AVECO e OUTROS. Os países que compõem estas regiões estão nos quadros a seguir:

Países da região AIBER total de 21

Argentina	Costa Rica	Honduras	Peru
Bolivia	Dominican Republic	Mexico	Portugal
Brazil	Ecuador	Nicaragua	Spain
Chile	El Salvador	Panama	Uruguay
Colombia	Guatemala	Paraguay	Venezuela
			Cuba

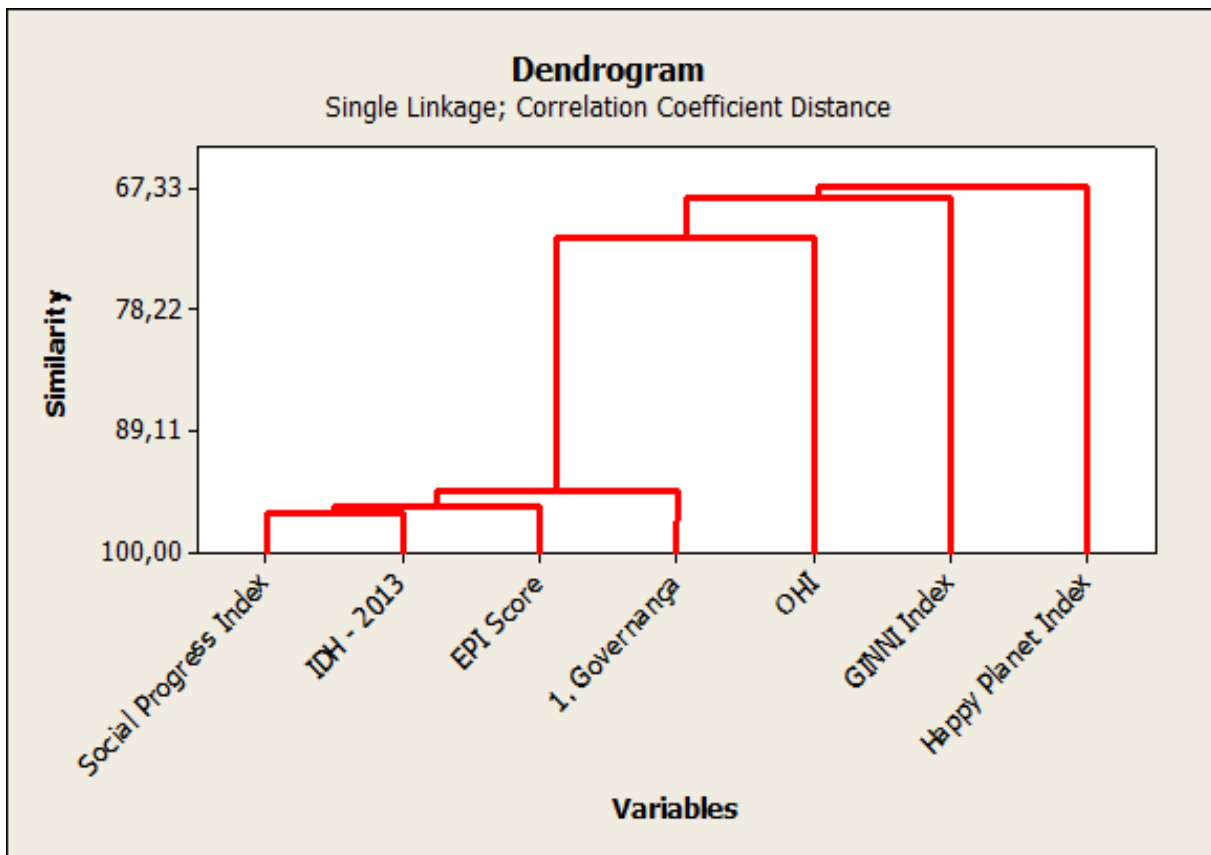
MAPA DA REGIAO AIBER



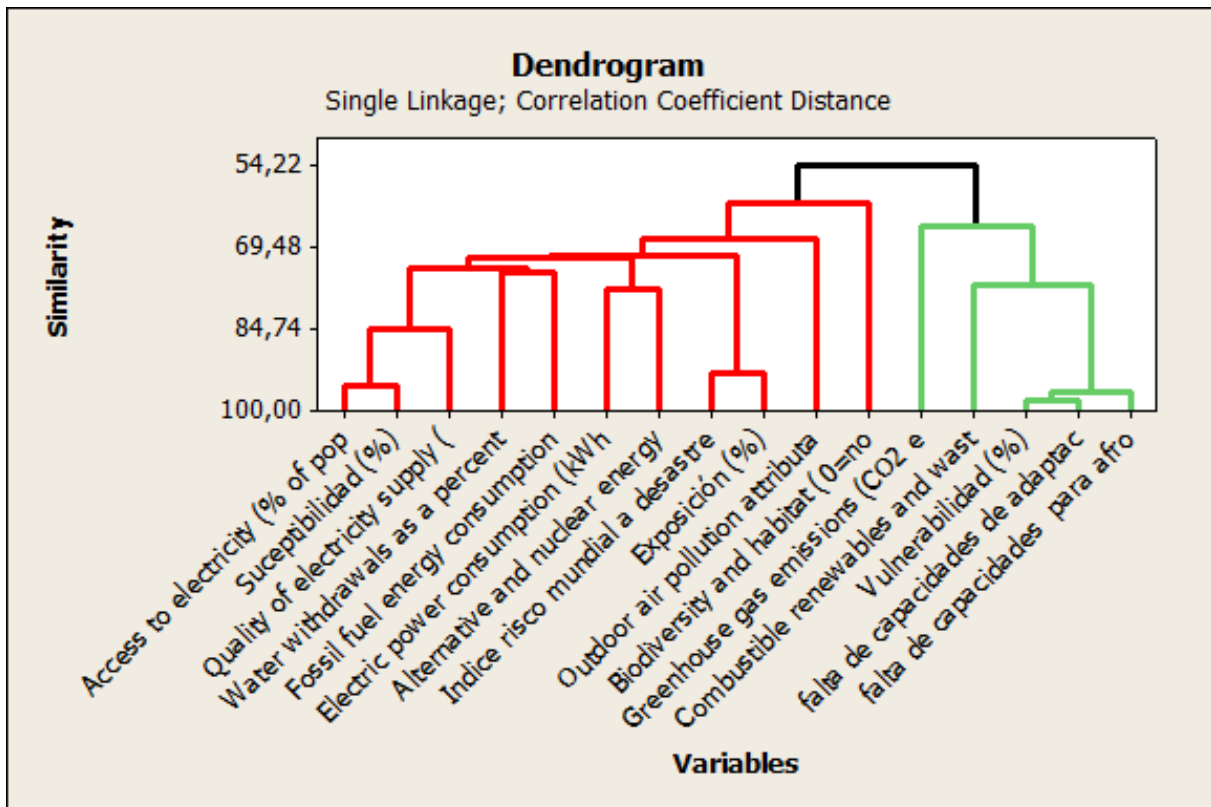
O Mapa foi pintado de acordo com as indicações de agrupamento feito pelo Dendograma indicado na tabela a seguir:

Grupo	País	Cor no mapa
1	Argentina	azul
1	Bolivia	azul
2	Brazil	amarela
2	Colombia	amarela
1	Domin Repub	azul
1	Ecuador	azul
1	El Salvador	azul
2	Guatemala	amarela
2	Honduras	amarela
1	Mexico	azul
1	Nicaragua	azul
2	Panama	amarela
1	Paraguay	azul
1	Peru	azul
1	Venezuela	azul
2	Chile	amarela
3	Costa Rica	verde
3	Portugal	verde
3	Spain	verde
3	Uruguay	verde
1	Cuba	azul

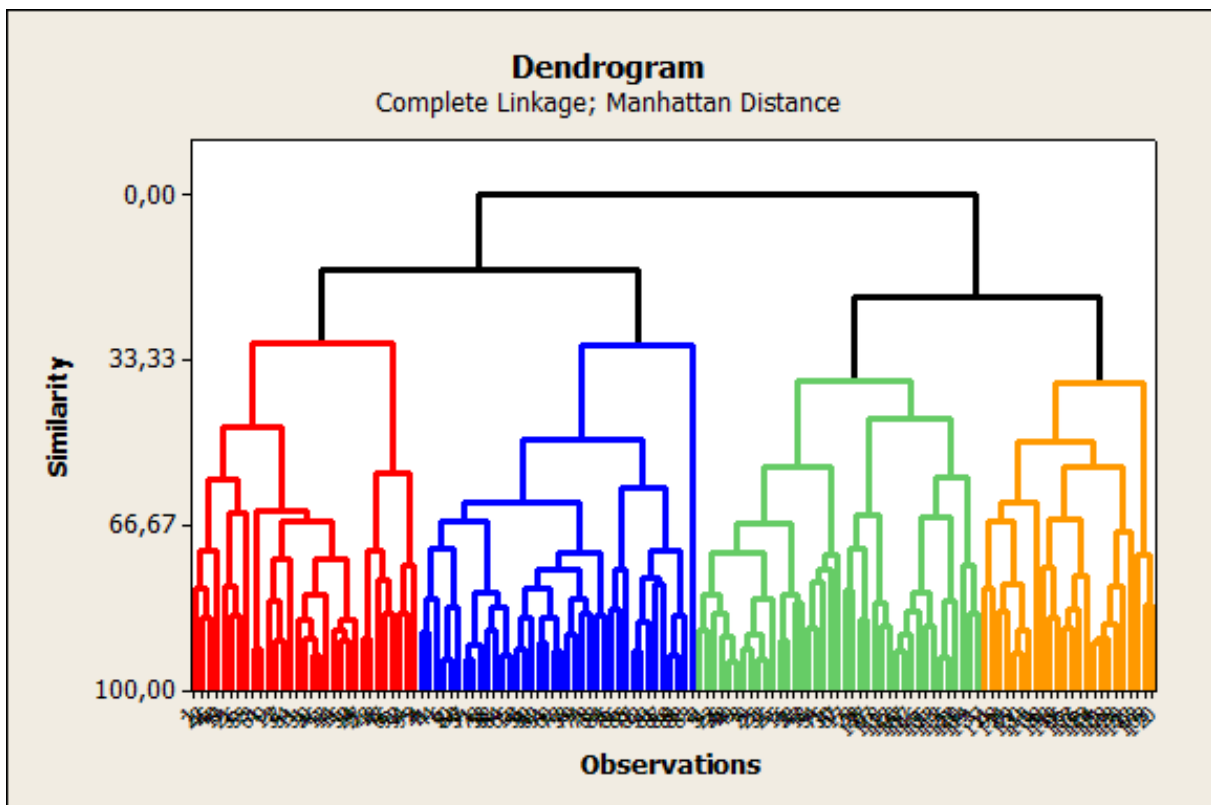
Com base nas principais variáveis sintéticas como: Social Progress Index, Happy Planet Index, GINI e OHI foi gerado o seguinte dendograma:



Em seguida foi realizado o agrupamento com base nas principais variáveis analíticas de todos os temas tratados ao longo do semestre e o MINITAB gerou a seguinte configuração para depois gerar o agrupamento em 4 países que reduzimos para três regiões:



A redução do agrupamento para 4 regiões:



Discriminant Analysis: 3REGIONS versus Social Progr; Happy Planet; ...

Quadratic Method for Response: 3REGIONS

Predictors: Social Progress Index; Happy Planet Index; GINNI Index; OHI;
Vulnerabilidad (%); Electric power consumption (kWh; Exposición (%))

Group	AIBER	AVECO	OTHERS
Count	21	27	84

Summary of classification

Put into Group	True Group		
	AIBER	AVECO	OTHERS
AIBER	20	0	5
AVECO	1	26	1
OTHERS	0	1	78
Total N	21	27	84
N correct	20	26	78
Proportion	0,952	0,963	0,929

N = 132

N Correct = 124

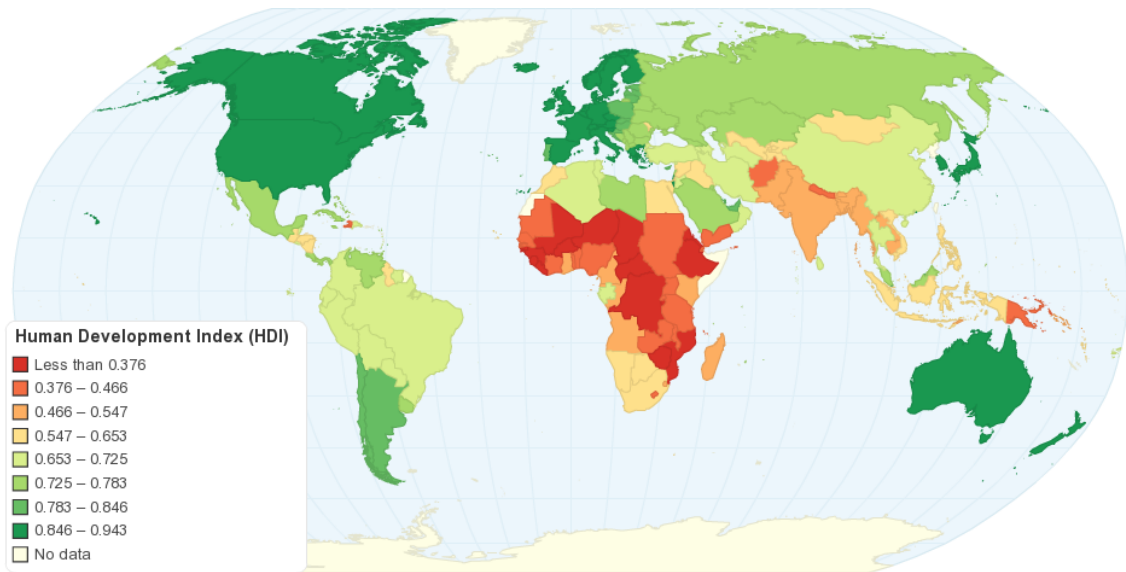
Proportion Correct = 0,939

5.2 – Região AVECO

27 Países da Região AVECO

Australia	Germany	Netherlands
Austria	Greece	New Zealand
Belgium	Iceland	Norway
Canada	Ireland	Slovakia
Czech Republic	Israel	Slovenia
Denmark	Italy	Sweden
Estonia	Japan	Switzerland
Finland	Korea, Republic of	United Kingdom
France	Latvia	United States

O mapa que melhor retrata essa região é o do IDH em especial a cor verde escura que está na faixa de 0,845 a 0,943.

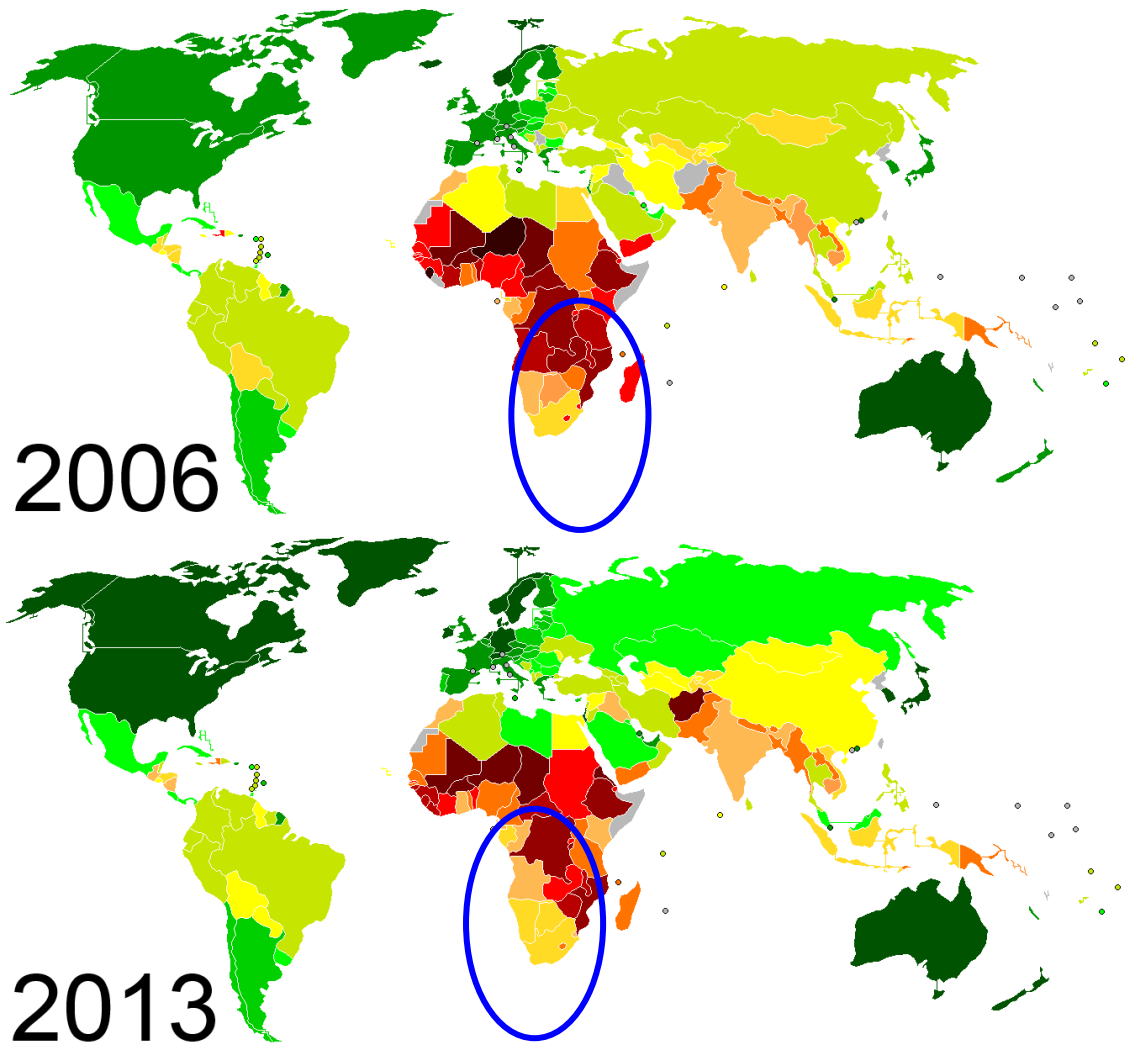


5.3 – Região OTHERS

84 países da Região OUTROS

CARLA	Guyana	EMDEU	Albania	ESEAS	Bangladesh	NSWUA	Algeria	SUSAF	Angola	SUSAF	Malawi
CARLA	Jamaica	EMDEU	Bosnia and	ESEAS	Cambodia	NSWUA	Egypt	SUSAF	Benin	SUSAF	Mali
CARLA	Trinidad and T	EMDEU	Bulgaria	ESEAS	China	NSWUA	Iran	SUSAF	Botswana	SUSAF	Mauritius
CWEAS	Armenia	EMDEU	Croatia	ESEAS	India	NSWUA	Iraq	SUSAF	Burkina Fas	SUSAF	Mozambique
CWEAS	Azerbaijan	EMDEU	Hungary	ESEAS	Indonesia	NSWUA	Jordan	SUSAF	Burundi	SUSAF	Namibia
CWEAS	Belarus	EMDEU	Lithuania	ESEAS	Laos	NSWUA	Kuwait	SUSAF	Cameroon	SUSAF	Niger
CWEAS	Georgia	EMDEU	Macedonia	ESEAS	Malaysia	NSWUA	Lebanon	SUSAF	Central Afri	SUSAF	Nigeria
CWEAS	Kazakhstan	EMDEU	Montenegro	ESEAS	Mongolia	NSWUA	Mauritania	SUSAF	Chad	SUSAF	Rwanda
CWEAS	Kyrgyzstan	EMDEU	Poland	ESEAS	Nepal	NSWUA	Morocco	SUSAF	Congo, Rep	SUSAF	Senegal
CWEAS	Moldova	EMDEU	Romania	ESEAS	Philippines	NSWUA	Pakistan	SUSAF	Djibouti	SUSAF	South Africa
CWEAS	Russia	EMDEU	Serbia	ESEAS	Sri Lanka	NSWUA	Saudi Arab	SUSAF	Ghana	SUSAF	Sudan
CWEAS	Tajikistan	EMDEU	Turkey	ESEAS	Thailand	NSWUA	Tunisia	SUSAF	Guinea	SUSAF	Swaziland
CWEAS	Ukraine					NSWUA	United Arab	SUSAF	Kenya	SUSAF	Tanzania
CWEAS	Uzbekistan					NSWUA	Yemen	SUSAF	Lesotho	SUSAF	Togo
								SUSAF	Liberia	SUSAF	Uganda
								SUSAF	Madagascar	SUSAF	Zambia

O mapa da Região Outros está citado com uma comparação entre os anos de 2006 e 2013 com algumas alterações sutis como no sul da África que altera o IDH.



Fonte: https://en.wikipedia.org/wiki/List_of_Latin_American_countries_by_Human_Development_Index

5.4 – Relação entre as variáveis por região:

As regiões deste trabalho final são:

Quadro 4: Regiões deste estudo: AIBER, AVECO e OUTROS

Região	Alguns países dessa região
AIBER – região Ibero-americana (21)	Brasil, Paraguai e Peru
AVECO – região da Europa Ocidental	Suíça e Suécia
OUTROS	
CWEAS – Europa oriental	Moldávia
CARLA – região do Caribe	Cuba e Jamaica
EMDEU – norte da Europa	Lituânia
ESEAS – sudeste da Ásia	Camboja, Filipinas, Malásia, Singapura
NSWUA – sul da Ásia	Paquistão, Arábia Saudita e Irã
SUSAF – África Subsaariana	Camarões e Republica Centro Africana

5.4.1. REGIAO AIBER

Antes de iniciarmos o trabalho é importante registrar o significado dos conceitos para ajudar no entendimento do que representam no projeto como um todo.

- Relação entre as variáveis: identificar as variáveis que estão correlacionadas a partir da identificação do Pvalor que deve estar com o limite 0,5. Se este valor for acima não é possível confira na correlação porque elas podem não ter nenhuma relação entre elas. Deve ficar atento para correlação que der resultado 0,00 (Zero) porque ela é nula e também não terá significado para a pesquisa.
- **PValor:** Mostra a probabilidade de que a relação entre as variáveis aleatórias tenha valor extremo em relação ao valor observado quando a hipótese H_0 é verdadeira.
- H_0 representa a primeira hipótese que o pesquisador estabelece como sendo verdadeira. O pesquisador irá criar outras hipóteses e identificar como H_1 , H_2 , por exemplo. Em geral ela obedece ao que está na observação. Exemplo: o valor observado da média pode ser o H_0 para iniciar a análise e verificar se existem mais valores da amostra com a mesma característica. Importante que o **grau de significância** esteja em 0,05 para que no gráfico da **distribuição normal** se tenha a região central com 0,95 de aceitação da hipótese e sobram os extremos com 0,025 lado esquerdo do gráfico e 0,024 lado direito do gráfico que é o 0,05 de rejeição.
- **Grau de significância ou nível de significância:** a estatística mostra que quanto maior o nível de significância maior o poder do teste e quando aumentamos o nível de significância menor fica a região de aceitação e isto leva a maior chance de rejeitar a hipótese nula (H_0) principalmente quando ela é falsa. Então, restam as outras hipóteses (H_1 , H_2) para verificação e posterior aceitação ou rejeição.
- Por exemplo, o pesquisador pode ter uma H_0 dada por uma valor igual a média e outra hipótese H_1 com um valor observado diferente da média.

- Distribuição normal ou curva de Gauss (Johann Carl Friedrich Gauss -1777-1825) ou forma de sino:** Define-se como curva normal, quando as medidas de tendência central, média, mediana e moda, têm valores muito próximos. Quando a média, a moda e a mediana são coincidentes, a curva é *simétrica*. Graficamente, a curva normal apresenta-se na *forma de um sino*. **Exemplo:** O ponto mais alto da curva é o valor que ocorre com maior probabilidade; esse valor é a moda que coincide com a média. É evidente que esta distribuição pode não ser normal, ou seja, não ter a forma de sino, ou ter as medidas de tendência central diferentes; aparecem então, as curvas assimétricas (SANTOS; PARRA FILHO,2011).

CORRELAÇÃO ENTRE OS VALORES DA REGIAO AIBER

Correlations: Social Progr; Undernourish; Depth of foo; Access to pi; ...			
	Social Progress	Undernourishment	Depth of food de
Undernourishment	-0,574 0,007		
Depth of food de	-0,542 0,011	0,984 0,000	
Access to piped	0,765 0,000	-0,586 0,005	-0,581 0,006
Rural vs. urban	-0,575 0,006	0,544 0,011	0,575 0,006
Access to improv	0,614 0,003	-0,656 0,001	-0,667 0,001
Access to electr	0,485 0,026	-0,558 0,009	-0,538 0,012
Quality of elect	0,683 0,001	-0,228 0,321	-0,248 0,279
Renewable intern	0,106 0,646	0,015 0,947	0,108 0,640
EV water resourc	0,634 0,002	-0,442 0,045	-0,508 0,019
OHI	0,385 0,084	0,052 0,824	0,019 0,936
IDH - 2013	0,744 0,000	-0,766 0,000	-0,775 0,000
Governança	0,892 0,000	-0,471 0,031	-0,439 0,047
EPI Score	0,750 0,000	-0,519 0,016	-0,573 0,007
Happy Planet Ind	-0,348 0,122	0,004 0,985	0,042 0,855
GINI Index	-0,321 0,155	0,288 0,206	0,361 0,108
Rural vs. urban	Access to piped -0,598	Rural vs. urban	Access to improv

IDH - 2013	0,630 0,002	0,422 0,057	
Governança	0,643 0,002	0,391 0,080	0,607 0,003
EPI Score	0,858 0,000	0,373 0,096	0,713 0,000
Happy Planet Ind	-0,431 0,051	-0,490 0,024	-0,296 0,193
GINI Index	-0,341 0,130	-0,162 0,484	-0,523 0,015
	Governança	EPI Score	Happy Planet Ind
EPI Score	0,664 0,001		
Happy Planet Ind	-0,347 0,124	-0,322 0,155	
GINI Index	-0,239 0,297	-0,384 0,086	0,492 0,024
Cell Contents: Pearson correlation P-Value			

ANÁLISE

Os dados pintados de lilás indicam que o Pvalor é muito acima do aceitável e para as variáveis terem uma correlação é preciso obter valores próximos de zero e menor que 0,05. No caso da Região AIBER as variáveis SPI e Renewable internal freshwater_1 apresenta um Pvalor de 0,830 indicando que não existe uma correlação entre elas.

Isto pode ser visto pelo Dendograma

Cluster Analysis of Observations: Social Progr; Undernourish; Depth of foo; ...

Manhattan Distance, Complete Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	20	88,2323	67,622	17	18	2
2	19	84,2301	90,620	3	14	3
3	18	79,6654	116,851	9	16	9
4	17	78,4949	123,576	2	13	2
5	16	77,9944	126,453	3	5	3
6	15	77,7558	127,824	6	19	6
7	14	77,2872	130,517	12	21	12
8	13	75,8335	138,870	8	12	8
9	12	73,2644	153,633	10	11	10
10	11	71,6557	162,878	1	8	1
11	10	71,3673	164,535	3	9	3
12	9	69,9618	172,612	7	15	7
13	8	67,2704	188,077	1	20	1
14	7	67,1765	188,617	4	17	4
15	6	64,4804	204,110	2	10	2
16	5	59,7564	231,256	2	7	2

17	4	58,7821	236,854	1	3	1	10
18	3	50,3280	285,435	1	6	1	12
19	2	30,8613	397,298	1	2	1	18
20	1	0,0000	574,639	1	4	1	21
Final Partition							
Number of clusters: 1							
			Within	Average	Maximum		
	Number of	cluster sum	distance	from	distance		
	observations	of squares	centroid	centroid			
Cluster1	21	70485,7	53,6843	109,664			

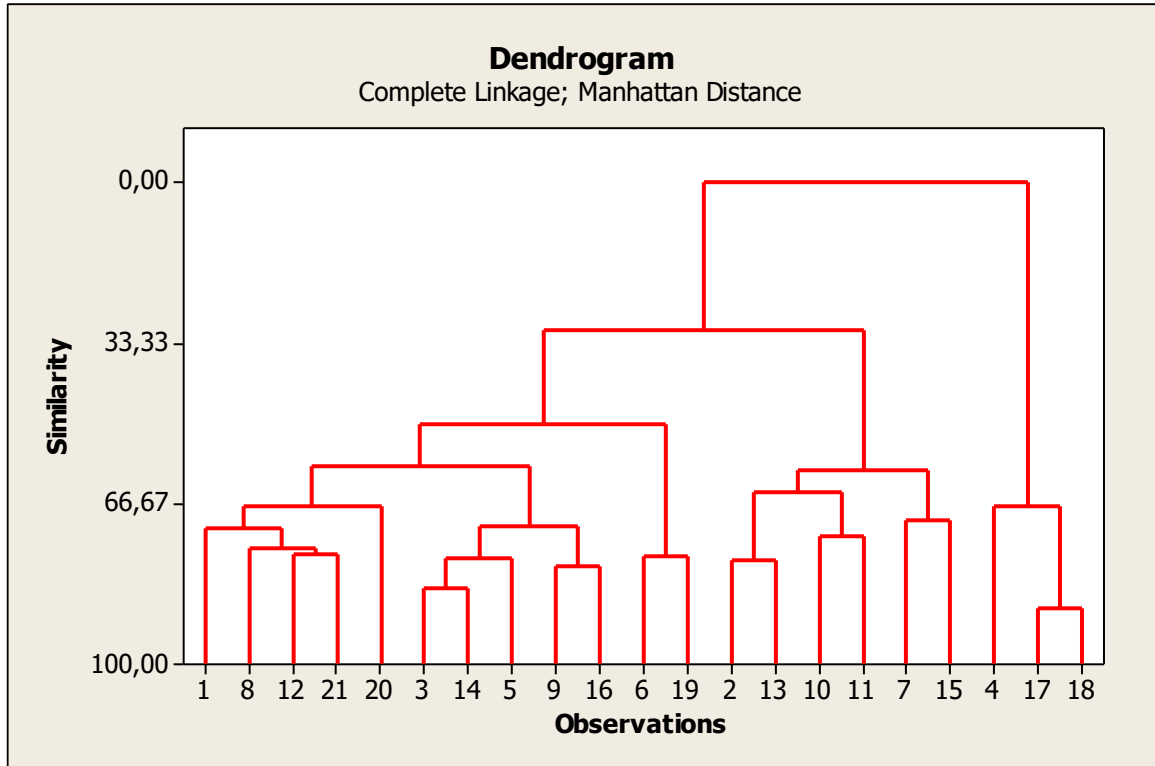
ANÁLISE

O grau de similaridade entre o cluster 1 e 2 é de 88,2323 e 84,2301.

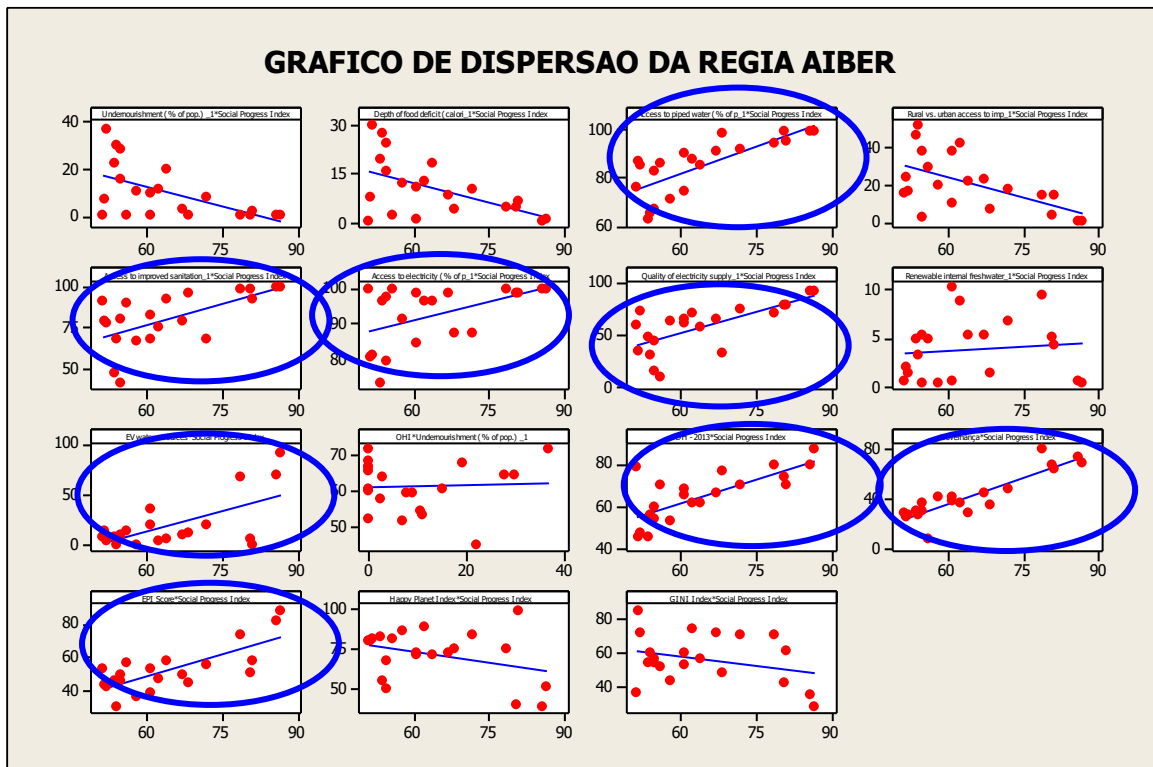
Um **Dendrograma** (dendr(o) = árvore) é um tipo específico de diagrama ou representação icônica que organiza determinados fatores e variáveis. É um diagrama de similaridade.

A interpretação de um dendrograma de similaridade entre amostras fundamenta-se na intuição: duas amostras próximas devem ter também valores semelhantes para as variáveis medidas. Ou seja, elas devem ser próximas matematicamente no espaço multidimensional. Portanto, quanto maior a proximidade entre as medidas relativas às amostras, maior a similaridade entre elas. O dendrograma hierarquiza esta similaridade de modo que podemos ter uma visão bidimensional da similaridade ou dissimilaridade de todo o conjunto de amostras utilizado no estudo.

Os países que apresentam similaridade são o indicado pelo dendrograma no eixo X 1, 8,12,21 e outros. Os países que estão nesta linha são: 1 – Argentina, 8- Equador, 12- México e 21 – Cuba porque apresentam os indicadores próximos das variáveis citadas exceto aquelas que apresentam Pvalor acima dos 0,05 indicados de lilás no quadro.



Gráficos de Dispersão da Região AIBER



ANÁLISE

Gráficos de dispersão devem ser inicialmente analisados quanto a seu padrão geral e seus desvios relativos ao padrão. A descrição do padrão geral pode ser feita pela verificação de sua forma, direção e intensidade.

Direção: Da análise das correlações entre as variáveis dos países da região AIBER indicamos com círculo azul aquelas que apresentam associações positivas, ou seja, o crescimento de uma variável é acompanhado do crescimento da outra.

Quanto as associações negativas ficaram para as demais exceto Renewable internal freshwater e OHI com o SPI que apresentam uma linha quase paralela com o eixo da abscissa (eixo X).

Intensidade: Alguns dos gráficos acima parecem indicar a existência de relações lineares, como Depth of food déficit, Rural vs. urban access to imp_1 Access to improved sanitation_1 Access to electricity (% of p_1, EV water resources, IDH - 2013 Governança EPI Score.

Forma: A concentração maior está com as variáveis: Depth of food déficit, EV water resources, IDH - 2013 Governança EPI Score.

As demais apresentam dados pouco pulverizados em torno da média, sendo que as variáveis Renewable internal freshwater e OHI estão com mais dispersão visual.

Valores Atípicos: Todos os gráficos indicam a existência de valores atípicos, ou seja, indivíduos ou empresas, que estão localizados longe dos demais.

5.4.2. REGIAO AVECO

Correlations: Social Progr; Undernourish; Depth of foo; Access to pi; ...			
	Social Progress	Undernourishment	Depth of food de
Undernourishment	*	*	
Depth of food de	-0,099 0,623	*	*
Access to piped	0,068 0,734	*	-0,077 0,701
Rural vs. urban	-0,333 0,089	*	0,025 0,900
Access to improv	0,222 0,265	*	-0,170 0,397
Access to electr	0,234 0,240	*	0,061 0,761
Quality of elect	0,755 0,000	*	-0,349 0,075
Renewable intern	0,306 0,121	*	-0,051 0,799

EV water resourc	0,191	*	-0,435
	0,340	*	0,023
OHI	0,275	*	0,149
	0,165	*	0,458
IDH - 2013	0,421	*	-0,297
	0,029	*	0,133
Governança	0,931	*	-0,140
	0,000	*	0,486
EPI Score	0,630	*	-0,150
	0,000	*	0,455
Happy Planet Ind	0,133	*	-0,203
	0,510	*	0,310
GINI Index	-0,309	*	-0,138
	0,116	*	0,492
	Access to piped	Rural vs. urban	Access to improv
Rural vs. urban	-0,394		
	0,042		
Access to improv	0,021	-0,026	
	0,916	0,898	
Access to electr	0,034	-0,850	-0,075
	0,865	0,000	0,712
Quality of elect	0,249	-0,413	0,389
	0,210	0,032	0,045
Renewable intern	0,079	-0,096	0,020
	0,696	0,633	0,923
EV water resourc	0,205	-0,170	0,119
	0,304	0,398	0,553
OHI	-0,106	-0,065	-0,153
	0,599	0,749	0,445
IDH - 2013	0,023	-0,556	0,184
	0,909	0,003	0,358
Governança	0,075	-0,310	0,220
	0,710	0,116	0,271
EPI Score	0,168	-0,571	0,177
	0,401	0,002	0,377
Happy Planet Ind	0,400	-0,257	-0,065
	0,039	0,195	0,748
GINI Index	-0,315	0,180	-0,271
	0,110	0,370	0,171
	Access to electr	Quality of elect	Renewable intern
Quality of elect	0,233		
	0,242		
Renewable intern	0,061	0,211	
	0,764	0,291	
EV water resourc	-0,059	0,193	-0,374

	0,770	0,336	0,054
OHI	0,123	0,054	-0,036
	0,540	0,789	0,859
IDH - 2013	0,516	0,382	0,153
	0,006	0,050	0,447
Governança	0,223	0,710	0,183
	0,263	0,000	0,360
EPI Score	0,390	0,515	0,086
	0,044	0,006	0,670
Happy Planet Ind	-0,008	0,225	-0,058
	0,967	0,260	0,775
GINI Index	0,039	-0,400	0,260
	0,847	0,039	0,191
	EV water resourc	OHI	IDH - 2013
OHI	0,035		
	0,864		
IDH - 2013	0,138	-0,082	
	0,493	0,685	
Governança	0,286	0,416	0,396
	0,148	0,031	0,041
EPI Score	0,424	-0,064	0,435
	0,028	0,752	0,023
Happy Planet Ind	0,487	-0,252	0,330
	0,010	0,205	0,093
GINI Index	-0,034	-0,155	0,051
	0,866	0,442	0,799
	Governança	EPI Score	Happy Planet Ind
EPI Score	0,494		
	0,009		
Happy Planet Ind	0,225	0,230	
	0,258	0,248	
GINI Index	-0,243	-0,416	0,105
	0,222	0,031	0,601
Cell Contents: Pearson correlation			
P-Value			
* NOTE * All values in column are identical.			

ANÁLISE

Os dados pintados de lilás indicam que é o Pvalor é muito acima do aceitável e para as variáveis terem uma correlação é preciso obter valores próximos de zero e menor que 0,05. No caso da Região AVECO diversas variáveis apresentaram valores acima dos aceitáveis para o Pvalor, como exemplo o SPI e GINI apresenta um Pvalor de 0,866 indicando que não existe uma correlação entre elas.

Correlação entre os valores da região AVECO

Cluster Analysis of Observations: Social Progr; Undernourish; Depth of foo; ...

Manhattan Distance, Complete Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	26	88,4031	36,160	8	24	2
2	25	86,1126	43,301	10	26	2
3	24	85,1950	46,163	9	13	2
4	23	85,0719	46,547	5	23	2
5	22	84,3574	48,775	6	8	3
6	21	82,9284	53,230	2	16	2
7	20	79,0503	65,322	9	10	4
8	19	78,6918	66,440	5	22	3
9	18	75,5999	76,081	6	19	4
10	17	75,2844	77,065	11	15	2
11	16	74,8252	78,496	20	21	2
12	15	73,4234	82,867	1	4	2
13	14	73,0776	83,945	2	9	6
14	13	69,5354	94,990	3	27	2
15	12	69,2267	95,953	6	25	5
16	11	65,0938	108,839	11	14	3
17	10	60,5423	123,031	3	7	3
18	9	58,8703	128,245	1	6	7
19	8	57,8700	131,364	2	20	8
20	7	52,7557	147,310	11	17	4
21	6	51,2191	152,101	3	5	6
22	5	48,9995	159,022	1	2	15
23	4	37,2043	195,800	3	11	10
24	3	29,6203	219,447	3	18	11
25	2	25,3735	232,689	1	12	16
26	1	0,0000	311,805	1	3	27

Final Partition

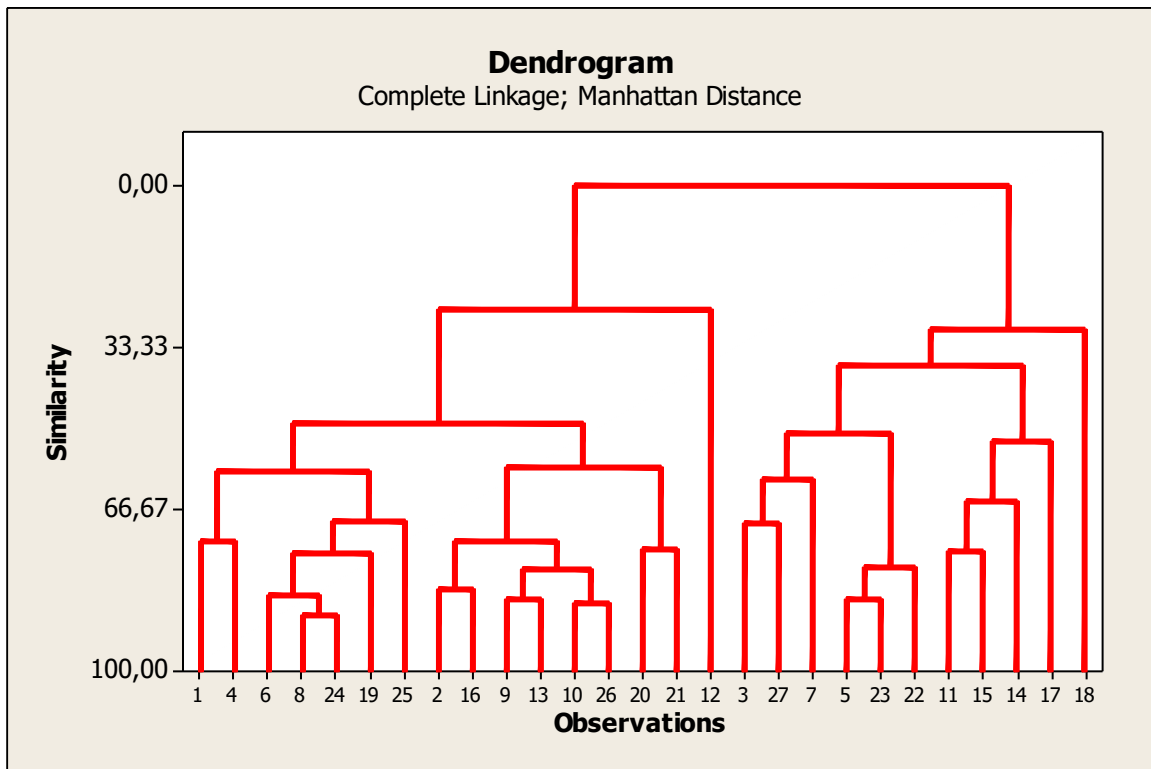
Number of clusters: 1

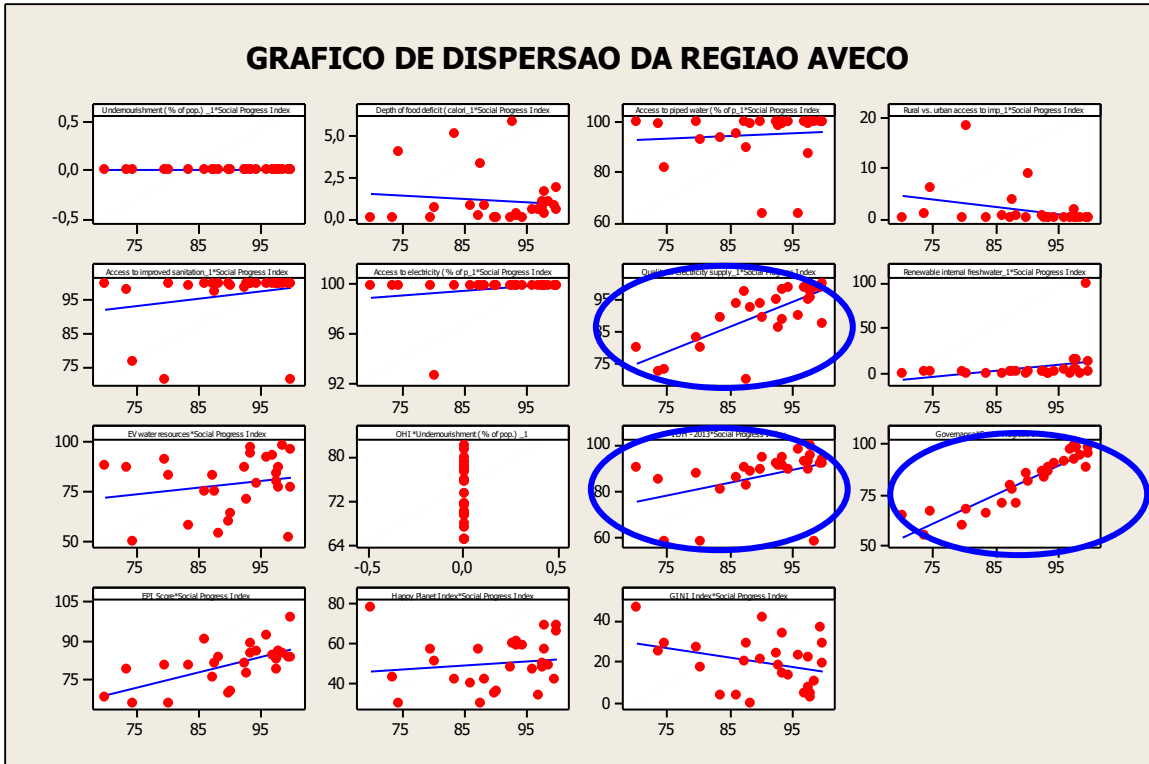
	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	27	43361,2	36,3909	100,705

ANÁLISE

O grau de similaridade entre o cluster 1 e 2 é de 88,4031 e 86,1126.

Os países que apresentam similaridade são o indicado pelo dendograma no eixo X 1, 4,6, 8 e outros. Os países que estão nesta linha são: 1 – Austrália, 4 - Canadá, 6 - Dinamarca e 8 – Finlândia porque apresentam os indicadores próximos das variáveis citadas exceto aquelas que apresentam Pvalor acima dos 0,05 indicados de lilás no quadro.





ANÁLISE

Gráficos de dispersão devem ser inicialmente analisados quanto a seu padrão geral e seus desvios relativos ao padrão. A descrição do padrão geral pode ser feita pela verificação de sua forma, direção e intensidade.

Direção: Da análise das correlações entre as variáveis dos países da região AVECO indicamos com círculo azul aquelas que apresentam associações positivas, ou seja, o crescimento de uma variável é acompanhado do crescimento da outra. Os gráficos das variáveis: Access to electricity, OHI e IDH apresentam correlação forte positiva.

Quanto as associações negativas mais acentuadas são: Access to piped water e GINI.

Intensidade: Alguns dos gráficos acima parecem indicar a existência de relações lineares, Access to electricity, OHI e IDH., Access to improved sanitation, Access to electricity.

Forma: A concentração maior está com as variáveis: Undernourishment, Access to electricity, OHI e IDH, Depth of food deficit, EV water resources, IDH - 2013 Governança EPI Score. As demais apresentam dados pouco pulverizados em torno da média, sendo que as variáveis Renewable internal freshwater e OHI estão com mais dispersão visual.

Valores Atípicos: Todos os gráficos indicam a existência de valores atípicos, ou seja, indivíduos ou empresas, que estão localizados longe dos demais.

5.4.3. Região OTHERS

CORRELAÇÃO ENTRE OS VALORES DA REGIAO OTHERS

Correlations: Social Progr; Undernourish; Depth of foo; Access to pi; ...

	Social Progress	Undernourishment	Depth of food de
Undernourishment	-0,650 0,000		
Depth of food de	-0,594 0,000	0,965 0,000	
Access to piped	0,815 0,000	-0,634 0,000	-0,602 0,000
Rural vs. urban	-0,631 0,000	0,475 0,000	0,400 0,000
Access to improv	0,789 0,000	-0,568 0,000	-0,508 0,000
Access to electr	0,803 0,000	-0,700 0,000	-0,634 0,000
Quality of elect	0,655 0,000	-0,424 0,000	-0,394 0,000
Renewable intern	0,016 0,886	-0,059 0,593	-0,049 0,659
EV water resourc	0,521 0,000	-0,427 0,000	-0,412 0,000
OHI	0,403 0,000	-0,185 0,092	-0,149 0,176
IDH - 2013	0,903 0,000	-0,653 0,000	-0,606 0,000
Governança	0,780 0,000	-0,378 0,000	-0,342 0,001
EPI Score	0,851 0,000	-0,594 0,000	-0,550 0,000
Happy Planet Ind	0,256 0,019	-0,227 0,038	-0,191 0,082
GINI Index	-0,211 0,054	0,325 0,003	0,284 0,009
	Access to piped	Rural vs. urban	Access to improv
Rural vs. urban	-0,624 0,000		
Access to improv	0,850 0,000	-0,659 0,000	
Access to electr	0,834 0,000	-0,590 0,000	0,870 0,000
Quality of elect	0,603 0,000	-0,287 0,008	0,596 0,000
Renewable intern	0,026 0,811	-0,039 0,723	0,039 0,724

EV water resourc	0,559 0,000	-0,329 0,002	0,490 0,000
OHI	0,285 0,009	-0,245 0,024	0,317 0,003
IDH - 2013	0,843 0,000	-0,678 0,000	0,863 0,000
Governança	0,514 0,000	-0,411 0,000	0,388 0,000
EPI Score	0,786 0,000	-0,546 0,000	0,773 0,000
Happy Planet Ind	0,241 0,027	-0,356 0,001	0,380 0,000
GINI Index	-0,293 0,007	0,172 0,118	-0,351 0,001
	Access to electr	Quality of elect	Renewable intern
Quality of elect	0,579 0,000		
Renewable intern	0,009 0,939	-0,104 0,345	
EV water resourc	0,481 0,000	0,438 0,000	-0,120 0,278
OHI	0,263 0,016	0,328 0,002	-0,182 0,097
IDH - 2013	0,890 0,000	0,631 0,000	-0,009 0,933
Governança	0,380 0,000	0,465 0,000	-0,017 0,880
EPI Score	0,738 0,000	0,640 0,000	-0,084 0,447
Happy Planet Ind	0,472 0,000	0,175 0,112	0,134 0,225
GINI Index	-0,438 0,000	-0,167 0,129	0,078 0,479
	EV water resourc	OHI	IDH - 2013
OHI	0,289 0,008		
IDH - 2013	0,559 0,000	0,331 0,002	
Governança	0,378 0,000	0,405 0,000	0,586 0,000
EPI Score	0,672 0,000	0,366 0,001	0,827 0,000
Happy Planet Ind	0,056 0,614	-0,110 0,319	0,335 0,002
GINI Index	-0,248 0,023	-0,021 0,847	-0,299 0,006

	Governança	EPI Score	Happy Planet Ind
EPI Score	0,606 0,000		
Happy Planet Ind	0,002 0,987	0,130 0,240	
GINI Index	0,145 0,189	-0,239 0,028	-0,241 0,027

Cell Contents: Pearson correlation
P-Value

ANÁLISE

Das três regiões esta denominada OUTROS foi a que apresentou o menor número de variáveis com Pvalor fora do padrão. Os dados pintados de lilás indicam que é o Pvalor é muito acima do aceitável e para as variáveis terem uma correlação é preciso obter valores próximos de zero e menor que 0,05.

No caso da Região OUTROS as variáveis SPI e Renewable internal freshwater_1 apresenta um Pvalor de 0,886 indicando que não existe uma correlação entre elas. A outra relação com Pvalor fora do normal é Acesso a Eletricidade e Renewable internal freshwater com valor de 0,939, outra relação é Governança e HPI com 0,987.

Isto pode ser visto pelo Dendograma

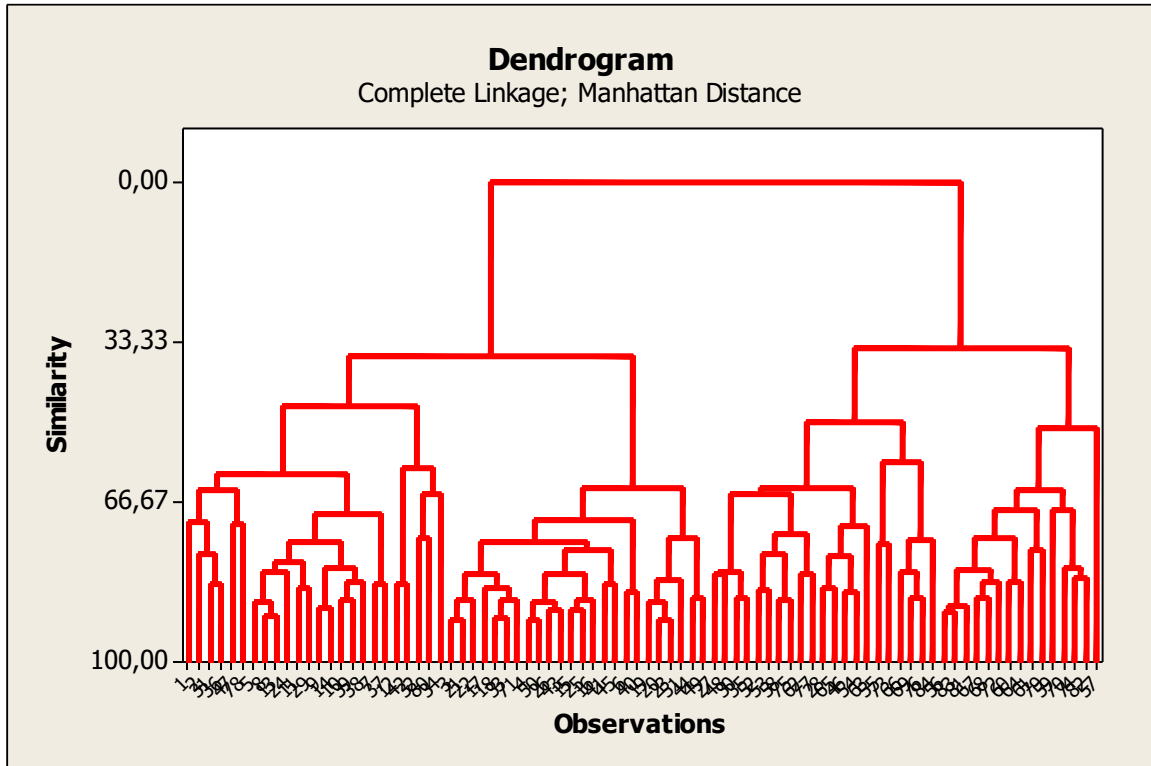
Cluster Analysis of Observations: Social Progr; Undernourish; Depth of foo; ...							
Manhattan Distance, Complete Linkage Amalgamation Steps							
Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster	
1	83	91,2529	76,184	3	21	3	2
2	82	91,1892	76,738	20	23	20	2
3	81	91,1428	77,143	4	50	4	2
4	80	90,8594	79,611	18	33	18	2
5	79	90,3122	84,377	8	13	8	2
6	78	89,5642	90,891	56	83	56	2
7	77	89,0141	95,683	15	25	15	2
8	76	88,9862	95,926	26	43	26	2
9	75	88,7427	98,047	9	14	9	2
10	74	88,1679	103,053	56	81	56	3
11	73	87,6394	107,656	4	26	4	4
12	72	87,5718	108,245	19	20	19	3
13	71	87,4861	108,991	5	8	5	3
14	70	87,2077	111,416	58	75	58	2
15	69	87,0640	112,667	10	39	10	2
16	68	87,0187	113,062	15	16	15	3
17	67	87,0038	113,192	18	71	18	3
18	66	86,9606	113,568	3	22	3	3
19	65	86,7882	115,070	69	76	69	2
20	64	86,7826	115,118	30	35	30	2

21	63	86,6577	116,206	67	68	67	2
22	62	86,6324	116,427	44	49	44	2
23	61	85,5605	125,762	6	40	6	2
24	60	85,5217	126,100	46	54	46	2
25	59	84,8320	132,107	52	53	52	2
26	58	84,7661	132,681	11	29	11	2
27	57	84,5971	134,153	28	65	28	2
28	56	84,5916	134,201	17	18	17	4
29	55	83,8273	140,858	41	45	41	2
30	54	83,7576	141,465	12	42	12	2
31	53	83,6418	142,474	31	36	31	2
32	52	83,5764	143,043	7	37	7	2
33	51	83,5399	143,361	60	64	60	2
34	50	83,5255	143,486	67	72	67	3
35	49	83,2458	145,922	10	38	10	3
36	48	83,1017	147,177	19	51	19	4
37	47	82,5678	151,827	74	82	74	2
38	46	81,8265	158,284	27	48	27	2
39	45	81,7284	159,138	3	17	3	7
40	44	81,6771	159,585	62	77	62	2
41	43	81,6723	159,627	4	15	4	7
42	42	81,3797	162,176	27	30	27	4
43	41	81,2680	163,148	5	24	5	4
44	40	81,2449	163,350	66	69	66	3
45	39	80,7794	167,404	56	67	56	6
46	38	80,5724	169,206	9	10	9	5
47	37	80,4147	170,580	70	74	70	3
48	36	79,1227	181,833	5	11	5	6
49	35	78,0246	191,397	28	46	28	4
50	34	77,5757	195,307	2	31	2	3
51	33	77,4077	196,770	52	58	52	4
52	32	76,5077	204,608	4	41	4	9
53	31	76,4905	204,758	61	79	61	2
54	30	75,5625	212,841	55	73	55	2
55	29	75,0135	217,622	3	4	3	16
56	28	74,9284	218,364	5	9	5	11
57	27	74,5608	221,565	66	84	66	4
58	26	74,4026	222,943	56	60	56	8
59	25	74,1201	225,404	32	80	32	2
60	24	74,0789	225,762	19	44	19	6
61	23	73,2147	233,289	52	62	52	6
62	22	71,5485	247,801	28	63	28	5
63	21	71,1515	251,259	47	78	47	2
64	20	70,7898	254,409	1	2	1	4
65	19	70,6020	256,045	3	6	3	18
66	18	69,1236	268,921	5	7	5	13
67	17	68,5496	273,920	59	70	59	4
68	16	68,3940	275,276	56	61	56	10
69	15	65,0091	304,757	32	34	32	3
70	14	64,8656	306,007	27	52	27	10
71	13	64,1820	311,960	56	59	56	14
72	12	64,0958	312,711	1	47	1	6
73	11	63,6835	316,302	27	28	27	15
74	10	63,6652	316,462	3	19	3	24
75	9	60,8722	340,787	1	5	1	19
76	8	59,7652	350,429	12	32	12	5
77	7	58,2115	363,961	55	66	55	6
78	6	51,3968	423,315	56	57	56	15
79	5	50,1119	434,505	27	55	27	21
80	4	46,8058	463,301	1	12	1	24
81	3	36,3829	554,079	1	3	1	48
82	2	34,8221	567,674	27	56	27	36
83	1	0,0000	870,960	1	27	1	84

Final Partition

Number of clusters: 1

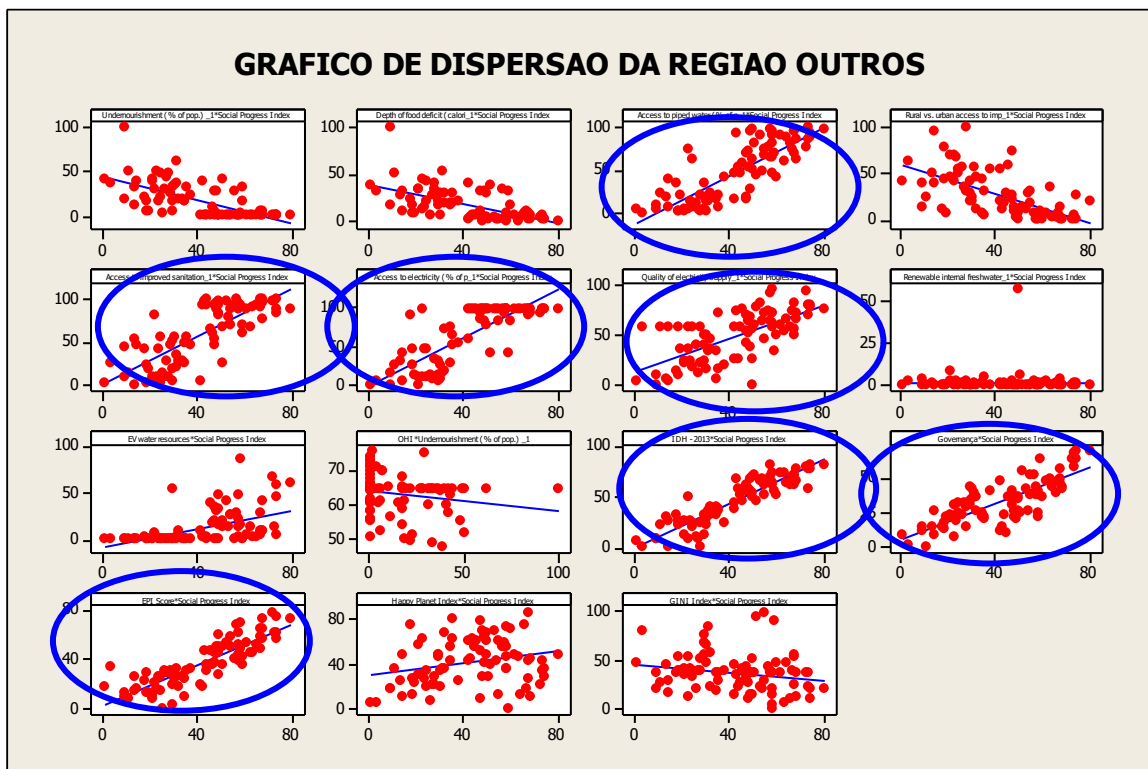
	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	84	701209	88,6772	163,542



ANÁLISE

Como são 84 observações o Dendrograma não consegue destacar os países, mas no relatório eles estão indicados. Pode ser visto que a relação entre Togo e Uganda (cluster 82 e 83) estão grau de similaridade de 91,2529 e 91,1892.

Gráficos de Dispersão da Região OUTROS



ANÁLISE

Gráficos de dispersão devem ser inicialmente analisados quanto a seu padrão geral e seus desvios relativos ao padrão. A descrição do padrão geral pode ser feita pela verificação de sua forma, direção e intensidade.

Direção: Da análise das correlações entre as variáveis dos países da região OUTRO indicamos com círculo azul aquelas que apresentam associações positivas, ou seja, o crescimento de uma variável é acompanhado do crescimento da outra. Nesse caso temos um total de 7 variáveis. Os gráficos das variáveis: Access to piped water, Access to improved sanitation, Access to electricity, Quality of electricity supply, IDH – 2013, Governança, EPI Score apresentam correlação forte positiva.

Quanto as associações negativas mais acentuadas são: Undernourishment, Depth of food deficit, Rural vs. urban access to imp.

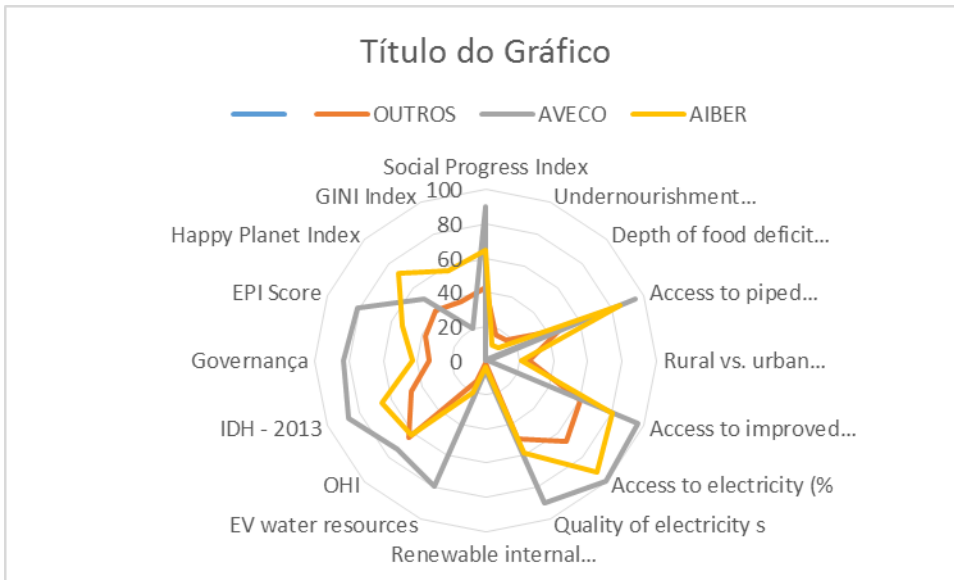
Intensidade: Alguns dos gráficos acima parecem indicar a existência de relações lineares, Access to electricity, OHI e IDH., Access to improved sanitation, Access to electricity.

Forma: A concentração maior está com as variáveis: Undernourishment, Access to electricity, OHI e IDH, Depth of food deficit, EV water resources, IDH - 2013 Governança EPI Score. As demais apresentam dados pouco pulverizados em torno da média, sendo que as variáveis Renewable internal freshwater e OHI estão com mais dispersão visual.

Valores Atípicos: Poucos gráficos indicam a existência de valores atípicos, ou seja, indivíduos ou empresas, que estão localizados longe dos demais. Apenas o Renewable internal freshwater tem um outlier.

6 - COMPARANDO AS TRES REGIOES

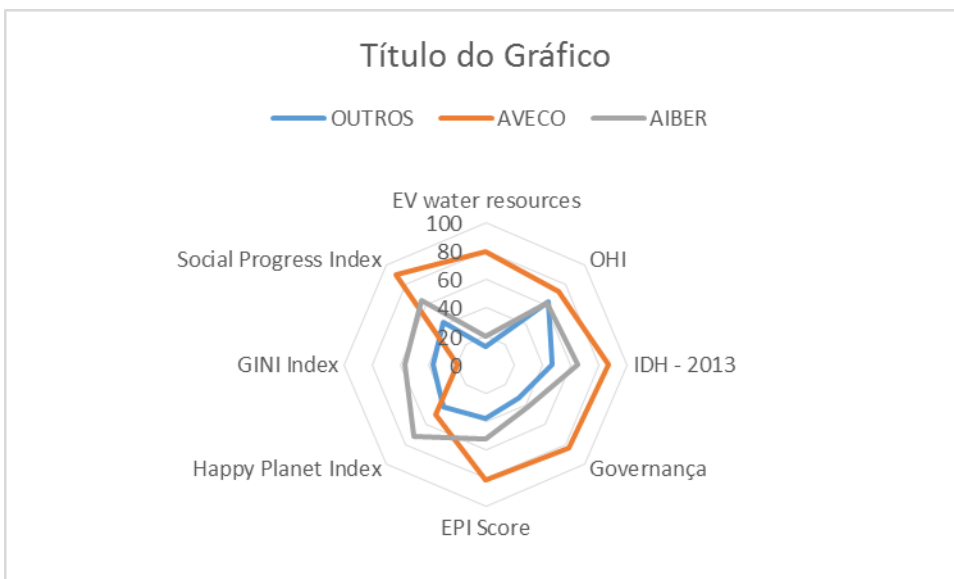
A comparação entre as três regiões: AIBER, AVECO e OUTROS será feita pela média com a apresentação do gráfico RADAR elaborado pelo EXCEL. Considerando todas as variáveis:



ANÁLISE

O gráfico mostra uma similaridade entre as três regiões em relação a variável Acesso a água encanada, Instalações sanitárias e Eletricidade.

Considerando as variáveis sintéticas:



ANÁLISE

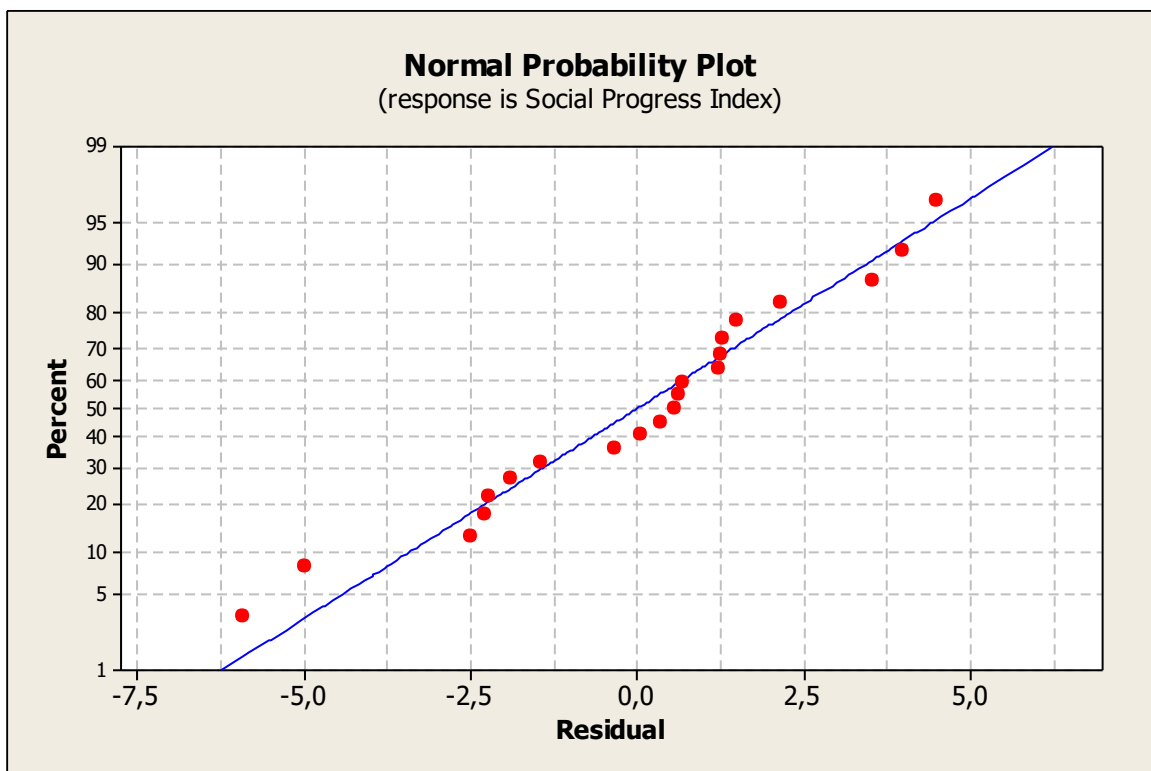
A Região AVECO aparece com mais destaque em todas as variáveis porque os países desta região estão em um padrão de desenvolvimento acima da média dos demais como pode ser visto nas análises de outros indicadores.

A Região AIBER e OUTROS que contra países emergentes e em desenvolvimento se aproximam pela média em variáveis como OHI e IDH.

7 - ANÁLISE DE REGRESSÃO DAS VARIÁVEIS COM SIMILARIDADE

A correlação mede a direção e a intensidade da relação linear (linha reta) entre duas variáveis quantitativas. Se um diagrama de dispersão mostra uma relação linear, é interessante resumirmos esse padrão geral traçando uma reta no diagrama de dispersão. Uma reta de regressão resume a relação entre duas variáveis, mas somente em um contexto específico: quando uma das variáveis ajuda a explicarmos ou predizermos a outra, ou seja, a regressão descreve uma relação entre uma variável explanatória e uma variável resposta.

7.1 – REGIAO AIBER



ANÁLISE

NA Região AIBER a variável SPI apresenta valores com pouca dispersão. Eles apresentam o menor desvio em torno da média conforme os pontos se apresentam próximos da linha de regressão e a constante P é de 86% (0,859).

Regression Analysis: Social Progr versus Undernourish; Depth of foo; ...

The regression equation is

$$\begin{aligned} \text{Social Progress Index} = & - 3,7 - 1,20 \text{ Undernourishment (\% of pop.) }_1 \\ & + 1,81 \text{ Depth of food deficit (calori}_1 \\ & + 0,322 \text{ Access to piped water (\% of p}_1 \\ & + 0,163 \text{ Rural vs. urban access to imp}_1 \\ & + 0,079 \text{ Access to improved sanitation}_1 \\ & - 0,137 \text{ Access to electricity (\% of p}_1 \\ & + 0,0423 \text{ Quality of electricity supply}_1 \\ & - 0,741 \text{ Renewable internal freshwater}_1 \\ & - 0,032 \text{ EV water resources} - 0,236 \text{ OHI} \\ & + 0,387 \text{ IDH} - 2013 + 0,318 \text{ Governança} + 0,242 \text{ EPI Score} \end{aligned}$$

Predictor	Coef	SE Coef	T	P
Constant	-3,69	19,95	-0,18	0,859
Undernourishment (% of pop.) _1	-1,195	1,150	-1,04	0,333
Depth of food deficit (calori_1	1,809	1,412	1,28	0,241
Access to piped water (% of p_1	0,3224	0,1870	1,72	0,128
Rural vs. urban access to imp_1	0,1633	0,1829	0,89	0,402
Access to improved sanitation_1	0,0785	0,1361	0,58	0,582
Access to electricity (% of p_1	-0,1370	0,2478	-0,55	0,598
Quality of electricity supply_1	0,04234	0,09354	0,45	0,664
Renewable internal freshwater_1	-0,7410	0,6168	-1,20	0,269
EV water resources	-0,0316	0,1189	-0,27	0,798
OHI	-0,2363	0,3525	-0,67	0,524
IDH - 2013	0,3865	0,2283	1,69	0,134
Governança	0,3184	0,1684	1,89	0,101
EPI Score	0,2415	0,2267	1,07	0,322

S = 4,52997 R-Sq = 95,0% R-Sq(adj) = 85,6%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	13	2713,15	208,70	10,17	0,002
Residual Error	7	143,64	20,52		
Total	20	2856,79			

Source	DF	Seq SS
Undernourishment (% of pop.) _1	1	939,67
Depth of food deficit (calori_1	1	43,50
Access to piped water (% of p_1	1	814,07
Rural vs. urban access to imp_1	1	64,04
Access to improved sanitation_1	1	6,71
Access to electricity (% of p_1	1	46,49
Quality of electricity supply_1	1	365,66
Renewable internal freshwater_1	1	2,44
EV water resources	1	206,30
OHI	1	17,09
IDH - 2013	1	101,18
Governança	1	82,71
EPI Score	1	23,29

Unusual Observations

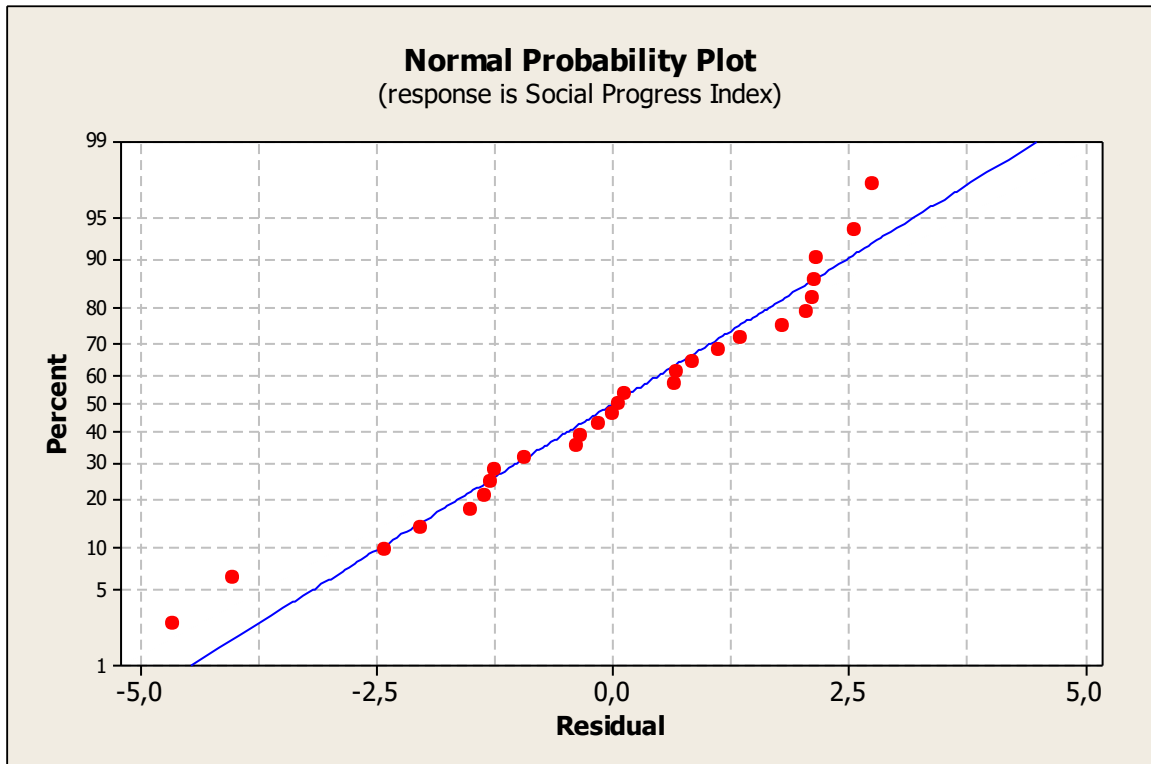
Obs	Undernourishment (% of pop.) _1	Social Progress Index	Fit	SE Fit	Residual	St Resid
21	0,0	51,168	57,085	3,802	-5,917	-2,40R

R denotes an observation with a large standardized residual.

ANÁLISE

O quadro mostra que a regressão nada mais é do que a expressão numérica da reta de tendência que vimos nos itens acima. Esta equação tem um poder explicativo de 95,0%, que é o R-Quadrado. Inclusive o R-quadrado ajustado tem percentual significativo de 85,6%.

7.2 – REGIAO AVECO



ANÁLISE

A Região AVECO que tem mais países estáveis do ponto de vista de desenvolvimento e com indicadores mostrados até aqui com relevância, apresenta neste gráfico quatro pontos fora da reta de regressão. A linha 18 que representa o país Látvia, Israel e na parte superior a Holanda e Canadá.

Regression Analysis: Social Progr versus Undernourish; Depth of foo; ...

* Undernourishment (% of pop.) _1 has all values = 0
* Undernourishment (% of pop.) _1 has been removed from the equation.

The regression equation is

$$\text{Social Progress Index} = 12 + 0,480 \text{ Depth of food deficit (calori_1)} \\ - 0,0075 \text{ Access to piped water (\% of p_1)} \\ + 0,329 \text{ Rural vs. urban access to imp_1} \\ - 0,0436 \text{ Access to improved sanitation_1} \\ + 0,03 \text{ Access to electricity (\% of p_1)} \\ + 0,184 \text{ Quality of electricity supply_1} \\ + 0,0441 \text{ Renewable internal freshwater_1} \\ - 0,0404 \text{ EV water resources} - 0,023 \text{ OHI} \\ + 0,0399 \text{ IDH - 2013} + 0,463 \text{ Governança} + 0,305 \text{ EPI Score}$$

Predictor

Coef SE Coef T P

Constant	12,2	117,4	0,10	0,919
Depth of food deficit (calori_1	0,4802	0,4631	1,04	0,317
Access to piped water (% of p_1	-0,00753	0,07624	-0,10	0,923
Rural vs. urban access to imp_1	0,3287	0,4630	0,71	0,489
Access to improved sanitation_1	-0,04360	0,06762	-0,64	0,529
Access to electricity (% of p_1	0,030	1,054	0,03	0,978
Quality of electricity supply_1	0,1838	0,1179	1,56	0,141
Renewable internal freshwater_1	0,04413	0,03407	1,30	0,216
EV water resources	-0,04041	0,06148	-0,66	0,522
OHI	-0,0227	0,1227	-0,19	0,856
IDH - 2013	0,03986	0,06901	0,58	0,573
Governança	0,46346	0,08274	5,60	0,000
EPI Score	0,3055	0,1009	3,03	0,009

S = 2,62359 R-Sq = 95,1% R-Sq(adj) = 90,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	1869,52	155,79	22,63	0,000
Residual Error	14	96,37	6,88		
Total	26	1965,89			

Source	DF	Seq SS
Depth of food deficit (calori_1	1	19,31
Access to piped water (% of p_1	1	7,30
Rural vs. urban access to imp_1	1	219,20
Access to improved sanitation_1	1	79,62
Access to electricity (% of p_1	1	30,37
Quality of electricity supply_1	1	872,95
Renewable internal freshwater_1	1	41,60
EV water resources	1	134,96
OHI	1	44,61
IDH - 2013	1	80,73
Governança	1	275,84
EPI Score	1	63,04

Unusual Observations

Obs	Depth of food deficit (calori_1	Social Progress Index	Fit	SE Fit	Residual	St Resid
14	0,00	69,734	74,395	1,829	-4,661	-2,48R
17	0,61	80,122	80,122	2,624	-0,000	* X
18	4,10	74,245	78,278	2,108	-4,033	-2,58R
19	1,06	98,436	95,685	2,259	2,751	2,06R

R denotes an observation with a large standardized residual.

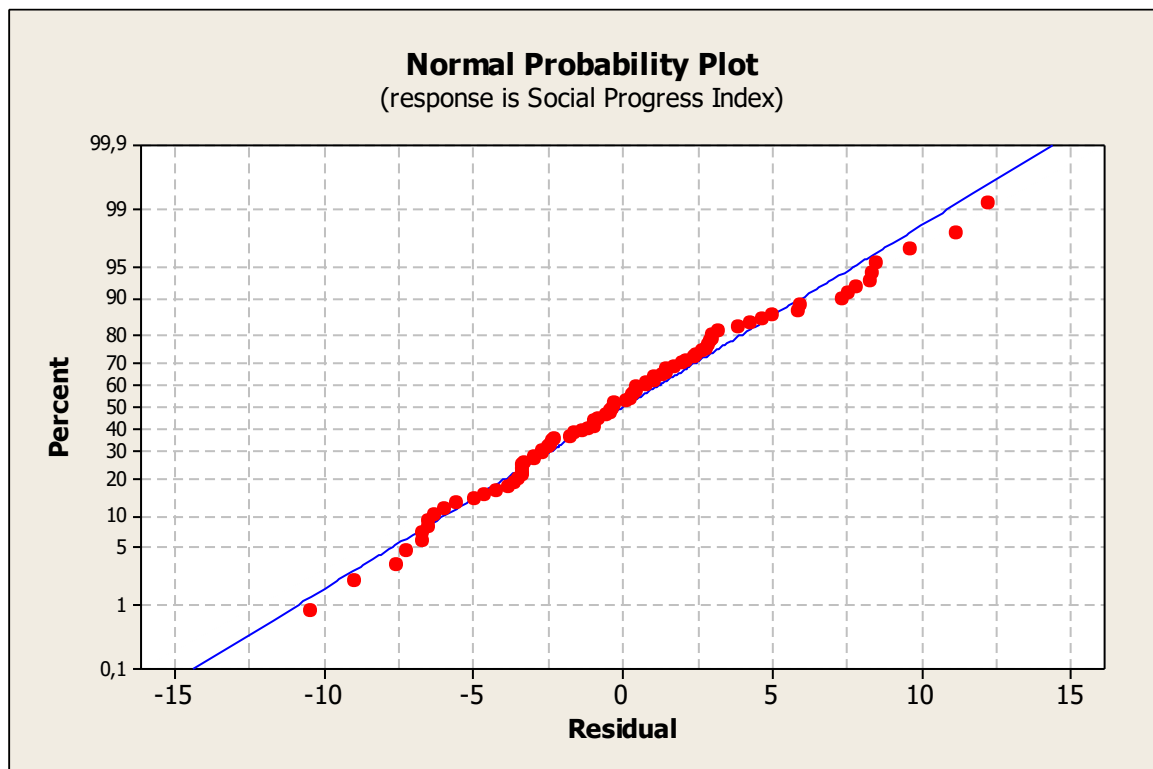
X denotes an observation whose X value gives it large leverage.

ANÁLISE

A regressão encontrada indica que os valores estão ajustados conforme o R-quadrado indica 95,1 e o R-quadrado ajustado também é valor significativo com 90,9. O MINITAB gerou a seguinte observação: R designa uma observação com um grande resíduo padronizado.

X denota uma observação cujo valor X lhe dá grande alavancagem

7.3 – REGIAO OUTROS



ANÁLISE

A reta de regressão apresenta uma tendência crescente para as variáveis da Região OUTROS embora tenhamos países com diferenças de desenvolvimento tão discrepantes como os extremos estatísticos entre Zâmbia e Emirados Árabes Unidos.

Pela disposição dos pontos existe uma concentração muito forte em torno da média e isto significa pouca dispersão entre os dados da variável e a reta de regressão.

Regression Analysis: Social Progr versus Undernourish; Depth of foo; ...

The regression equation is

```
Social Progress Index = - 2,25 - 0,067 Undernourishment (% of pop.) _1
                        + 0,015 Depth of food deficit (calori_1
                        - 0,0005 Access to piped water (% of p_1
                        - 0,0169 Rural vs. urban access to imp_1
                        + 0,0457 Access to improved sanitation_1
                        + 0,0989 Access to electricity (% of p_1
                        + 0,0369 Quality of electricity supply_1
                        + 0,0980 Renewable internal freshwater_1
                        - 0,0651 EV water resources + 0,054 OHI
                        + 0,199 IDH - 2013 + 0,479 Governança + 0,185 EPI Score
```

Predictor	Coef	SE Coef	T	P
Constant	-2,247	6,592	-0,34	0,734
Undernourishment (% of pop.) _1	-0,0673	0,1297	-0,52	0,605
Depth of food deficit (calori_1	0,0147	0,1380	0,11	0,916
Access to piped water (% of p_1	-0,00047	0,03904	-0,01	0,991
Rural vs. urban access to imp_1	-0,01692	0,03537	-0,48	0,634
Access to improved sanitation_1	0,04571	0,04460	1,02	0,309
Access to electricity (% of p_1	0,09888	0,04497	2,20	0,031
Quality of electricity supply_1	0,03687	0,03244	1,14	0,260
Renewable internal freshwater_1	0,09797	0,09346	1,05	0,298

EV water resources	-0,06507	0,04076	-1,60	0,115
OHI	0,0541	0,1037	0,52	0,604
IDH - 2013	0,19899	0,07781	2,56	0,013
Governança	0,47900	0,05381	8,90	0,000
EPI Score	0,18517	0,06580	2,81	0,006

S = 5,07593 R-Sq = 94,4% R-Sq(adj) = 93,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	13	30445,3	2341,9	90,90	0,000
Residual Error	70	1803,6	25,8		
Total	83	32248,8			

Source	DF	Seq SS
Undernourishment (% of pop.) _1	1	13639,1
Depth of food deficit (calori_1	1	520,6
Access to piped water (% of p_1	1	8551,5
Rural vs. urban access to imp_1	1	407,4
Access to improved sanitation_1	1	512,5
Access to electricity (% of p_1	1	240,5
Quality of electricity supply_1	1	1089,3
Renewable internal freshwater_1	1	7,8
EV water resources	1	56,8
OHI	1	455,7
IDH - 2013	1	2170,8
Governança	1	2589,2
EPI Score	1	204,1

Unusual Observations

Obs	Undernourishment (% of pop.) _1	Social Progress Index	Fit	SE Fit	Residual	St Resid
1	0	49,353	50,318	4,958	-0,965	-0,89 X
10	0	49,461	52,936	3,646	-3,475	-0,98 X
12	39	42,146	32,501	2,443	9,645	2,17R
36	18	59,777	48,574	1,499	11,203	2,31R
42	31	21,999	30,940	3,008	-8,941	-2,19R
48	22	17,613	28,067	2,488	-10,454	-2,36R
61	47	27,660	15,403	2,773	12,257	2,88R

R denotes an observation with a large standardized residual.

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

ANÁLISE

Pelos dados apresentados no gráfico com a reta de regressão com uma concentração dos pontos em torno da reta e pouquíssima variação era de se esperar que o R-quadrado ficasse perto do 100%. Observado o resultado final gerado pelo MINITAB e o R-quadrado ficou com 94,4% e o R-quadrado ajustado 93,4%.

7.4 – Comparação entre as três regiões: AIBER, AVECO E OUTROS

A comparação vista pelo gráfico com a reta de regressão não deixa dúvida que a região AVECO tem dados mais consistentes e países com dados de desenvolvimento mais significativos.

A Região AIBER tem diferenças políticas e econômicas entre os países e com a entrada de CUBA na estatística os dados ficarão mais consistentes quando este país puder apresentar informações socioeconômicas próximas da metodologia dos demais países.

Mesmo assim, Brasil tem IDH ainda pequeno, mas significativo em relação aos demais enquanto Chile ainda faz a diferença para mais próximo de dados dos países da região AVECO.

Quando se olha os dados do R-quadrado das regiões percebemos que todos se ajustam a consistência das variáveis e os dados obtidos. A Região AIBER tem 95,0%, que é o R-Quadrado e o R-quadrado ajustado tem percentual significativo de 85,6%. A Região AVECO com R-quadrado indica 95,1 e o R-quadrado ajustado também é valor significativo com 90,9. A Região OUTROS tem R-quadrado de 94,4% e o R-quadrado ajustado 93,4%.

Quando se olha para a composição de cada região percebe-se que a região OUTROS é a que mais apresenta diferença entre os países como já citado entre Zâmbia e Emirados Árabes Unidos. O primeiro sabe-se das deficiências na distribuição de renda e o outro do potencial de riqueza por conta do petróleo e agora também do turismo e investimentos estrangeiro em outros países no setor do futebol.

8 - ANÁLISE DO COMPONENTE PRINCIPAL

Para fazer a Análise dos Componentes Principais é preciso que as variáveis tenham uma certa homogeneidade entre elas. Até o presente momento este trabalho final tem comprovado que existe uma homogeneidade entre as variáveis tanto em relação a elas e as regiões juntas (como foi feito na análise exploratória dos dados com valores indicados pelo desvio padrão por exemplo) como em relação a elas e as regiões separadas (com apresentado no Pvalor e nos R-quadrados).

Neste item será apresentado a Análise de Componente principais seguindo a metodologia anterior de mostrar e analisar região por região e depois em conjunto.

A análise multivariada de dados representa um conjunto de métodos estatísticos que torna possível a análise simultânea de medidas múltiplas para cada indivíduo, objeto ou fenômeno observado, logo os métodos que permitem análise simultânea de mais de duas variáveis estão na categoria de análise multivariada de dados (CORRAR; PAULO; DIAS FILHO org. 2012). O trabalho é composto de uma análise com a aplicação do **Scatterplot**, depois usamos o Multivariate e nele o **Principal Component Analysis** e por fim o **Scree Plot**.

- **Scatterplot:** é um gráfico de dispersão usado para explorar a relação potencial entre um par de variáveis contínuas indicadas para os eixos X e Y mas o Scatterplot 3D permite que seja inserido uma terceira variável denominada Z.
- **ScatterPlot 3D:** é um gráfico que permite explorar a relação potencial entre três variáveis. Um gráfico de dispersão 3D gráficos os valores das três variáveis contínuas uns contra os outros dados reais sobre o X, Y, e Z-eixos. Normalmente, o gráfico 3D traça a linha das variáveis de previsão no eixo X e do eixo Y e a variável de resposta sobre o eixo Z.
- **Principal Component Analysis:** quando se tem um grande conjunto de dados usamos este recurso para identificar um pequeno número de variáveis não correlacionadas as quais são chamadas de componentes principais. O objetivo da análise de componentes principais é explicar a quantidade máxima de variância com o menor número de componentes principais. Análise de componentes principais é comumente usado nas ciências sociais, pesquisa de mercado, e outras indústrias que utilizam grandes conjuntos de dados. Análise de componentes principais é comumente usado como um passo em uma série de análises (<http://support.minitab.com/>)
- **Scree Plot:** é um gráfico que relaciona duas variáveis da amostra e cria valores próprios associados a ela para identificar quantos destes valores explicam a relação. Como o gráfico mostra uma linha decrescente desta relação a parte que ficar no eixo da ordenada (Y) mostra a maior parte da variabilidade e quanto mais próximo do eixo da abscissa (X) os fatores explicam uma variabilidade muito pequena e muitas vezes não tem importância para a análise do fato ou fenômeno

(<http://support.minitab.com/>)

8.1 – Região AIBER

Principal Component Analysis: Social Progr; Undernourish; Depth of foo; Access

Eigenanalysis of the Correlation Matrix

Eigenvalue	7,8817	1,8923	1,8038	1,0830	0,8777	0,6571	0,5250	0,3719
Proportion	0,493	0,118	0,113	0,068	0,055	0,041	0,033	0,023
Cumulative	0,493	0,611	0,724	0,791	0,846	0,887	0,920	0,943

Eigenvalue	0,2775	0,2523	0,1561	0,1245	0,0386	0,0348	0,0219	0,0017
Proportion	0,017	0,016	0,010	0,008	0,002	0,002	0,001	0,000
Cumulative	0,961	0,976	0,986	0,994	0,996	0,999	1,000	1,000

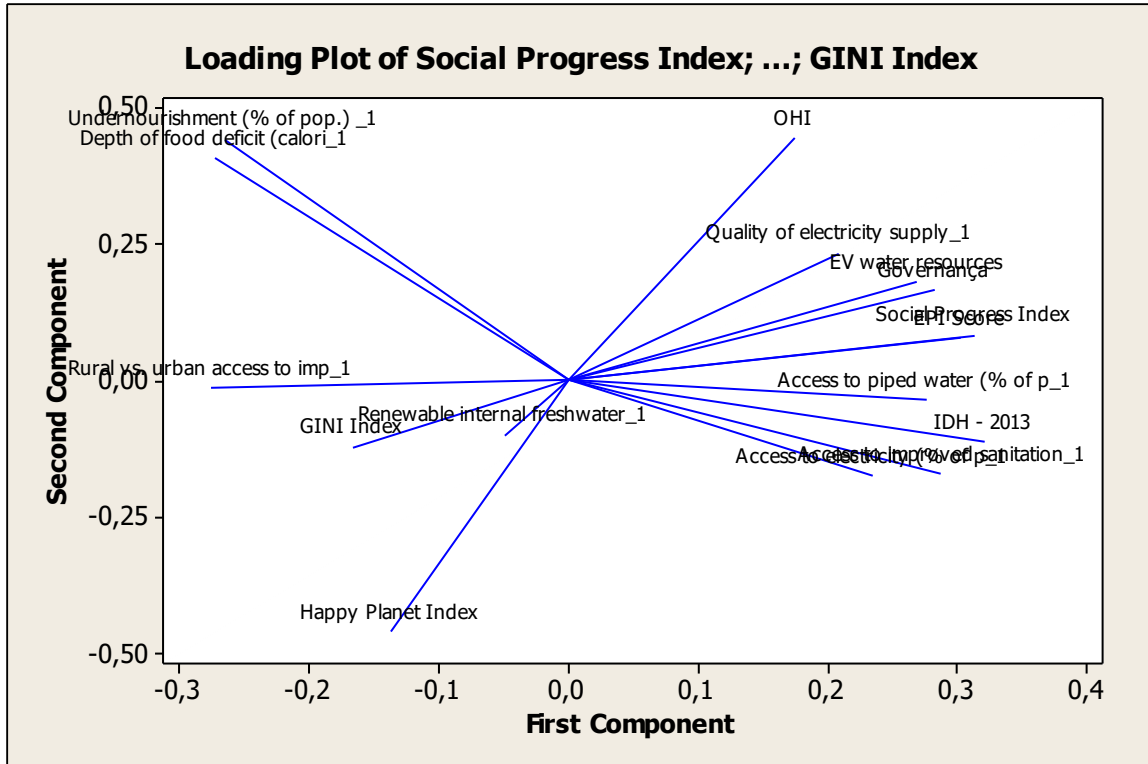
Variable	PC1	PC2
Social Progress Index	0,313	0,084
Undernourishment (% of pop.)_1	-0,264	0,443
Depth of food deficit (calori_1	-0,271	0,409
Access to piped water (% of p_1	0,276	-0,034
Rural vs. urban access to imp_1	-0,275	-0,012
Access to improved sanitation_1	0,288	-0,172
Access to electricity (% of p_1	0,234	-0,174
Quality of electricity supply_1	0,209	0,234
Renewable internal freshwater_1	-0,049	-0,099
EV water resources	0,269	0,182
OHI	0,175	0,444
IDH - 2013	0,321	-0,113
Governança	0,283	0,167
EPI Score	0,302	0,080
Happy Planet Index	-0,137	-0,461
GINI Index	-0,166	-0,121

ANÁLISE

Os componentes principais gerados pelo MINITAB de acordo com as variáveis indicadas para a Região AIBER apresentou uma correlação entre as componentes de 49,3% para a PC1 e 11,8% para a PC2 porque o *Eigenvalue* (que é um valor próprio ou auto-valor gerado pelo MINITAB para as PC1 e PC2) é de 7,8817 do primeiro fator que explica a variância global de 49,3%, ou seja, a divisão de 7,8817 por 16 variáveis apresentadas neste estudo.

Isto significa que ao invés de 16 variáveis, o estudo poderia ser feito apenas com uma delas que a SPI que tem influencia total de 49,26 %.

Quando se analisa a componente 2 (PC2) Desnutrição com 44,3% e OHI com 44,4% nota-se que é mais significativa que o SPI para a PC1.

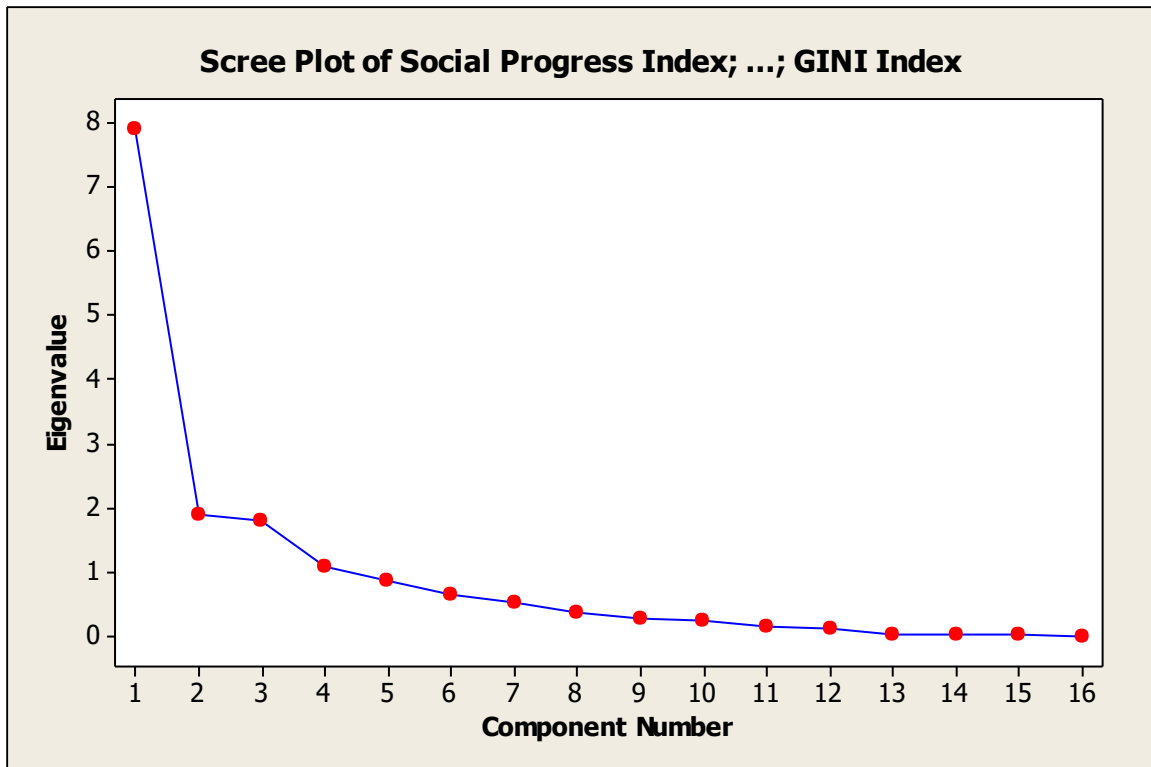


ANÁLISE

Este gráfico apresenta os dados das variáveis com uma correlação entre elas indicadas pela direção das linhas para a direita com valores positivos e para a esquerda com valores negativos. Também ela pode ter direções totalmente diferentes destas, pode ir para cima e para baixo ou se afastarem muito das demais.

Neste caso, as variáveis que estão na direção da direita do gráfico e concentradas umas próximas das outras como exemplo EV, Governança, SPI que quase estão uma reta em cima da outra, mostra que estão fortemente correlacionadas.

Aquelas retas que apontam para o lado esquerdo, embora estejam próximas como é o caso de Desnutrição e Déficit alimentar tem uma leitura que representa que uma delas pode ficar de fora da análise porque possuem o mesmo sentido. As outras quatro estão apontando para o lado esquerdo, mas com tendência negativa representam que os países da AIBER não tem grande preocupação com a distribuição de renda (GINI) ou o tratamento de águas residuais.



ANÁLISE

O Scree Plot é um gráfico que relaciona duas variáveis da amostra e cria valores próprios associados a ela para identificar quantos destes valores explicam a relação. Como o gráfico mostra uma linha decrescente desta relação a parte que ficar no eixo da ordenada (Y) mostra a maior parte da variabilidade e quanto mais próximo do eixo da abscissa (X) os fatores explicam uma variabilidade muito pequena e muitas vezes não tem importância para a análise do fato ou fenômeno (<http://support.minitab.com/>).

Como foi lido anteriormente com os valores do o *Eigenvalue* – que é um valor próprio ou auto-valor gerado pelo MINITAB para as PC1 e PC2 é de 7,8817 do primeiro fator que explica a variância global de 49,3%, ou seja, a divisão de 7,8817 por 16 variáveis apresentadas neste estudo), Apenas duas variáveis explicam as demais.

8.2 – Região AVECO

Importante registrar que a variável Undernourishment (% of pop.) foi deixada fora desta análise porque o sistema MINITAB não gerava informação porque a região apresenta este dado muito positivo que é igual a Zero.

Principal Component Analysis: Social Progr; Depth of foo; Access to pi; Rural v

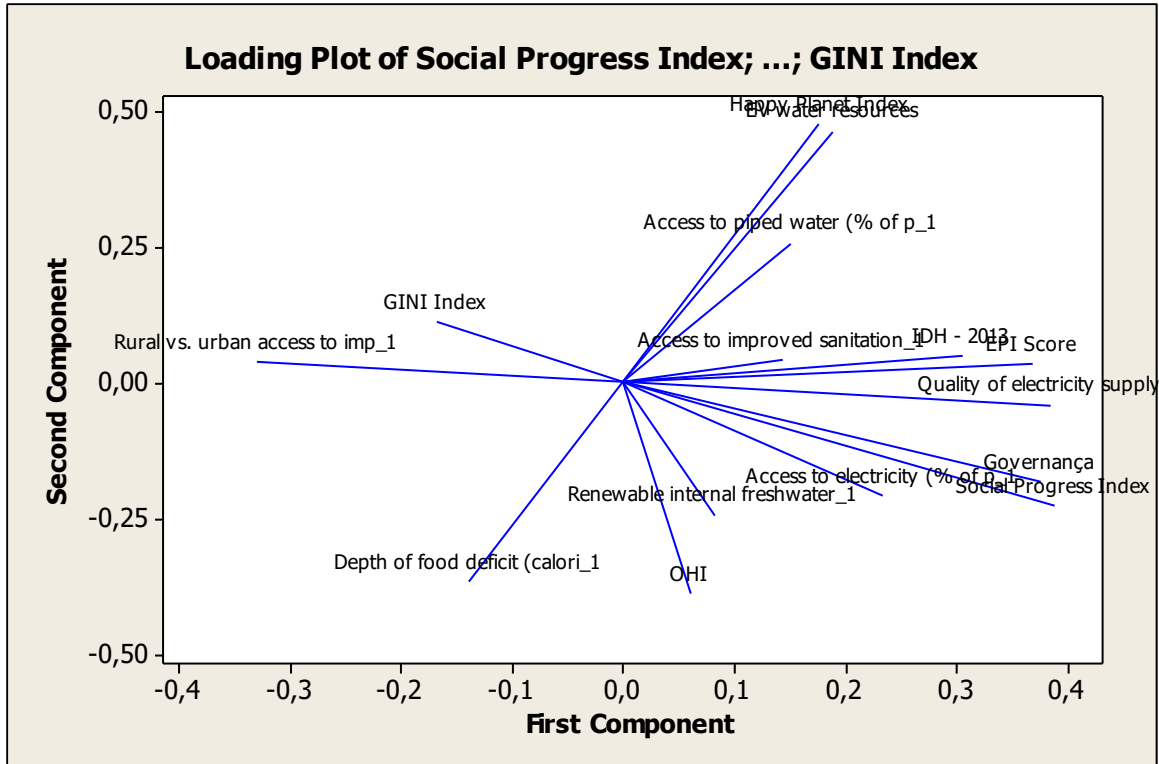
Eigenanalysis of the Correlation Matrix

Eigenvalue	4,5394	1,9663	1,8115	1,5109	1,2953	1,1338	0,7564	0,6372
Proportion	0,303	0,131	0,121	0,101	0,086	0,076	0,050	0,042
Cumulative	0,303	0,434	0,554	0,655	0,742	0,817	0,868	0,910
Eigenvalue	0,5044	0,3754	0,2264	0,1153	0,0698	0,0364	0,0216	
Proportion	0,034	0,025	0,015	0,008	0,005	0,002	0,001	
Cumulative	0,944	0,969	0,984	0,991	0,996	0,999	1,000	

Variable	PC1	PC2
Social Progress Index	0,388	-0,227
Depth of food deficit (calori_1	-0,139	-0,367
Access to piped water (% of p_1	0,151	0,257
Rural vs. urban access to imp_1	-0,330	0,037
Access to improved sanitation_1	0,143	0,042
Access to electricity (% of p_1	0,233	-0,209
Quality of electricity supply_1	0,385	-0,041
Renewable internal freshwater_1	0,081	-0,244
EV water resources	0,189	0,463
OHI	0,060	-0,388
IDH - 2013	0,305	0,049
Governança	0,375	-0,182
EPI Score	0,369	0,035
Happy Planet Index	0,176	0,476
GINI Index	-0,168	0,111

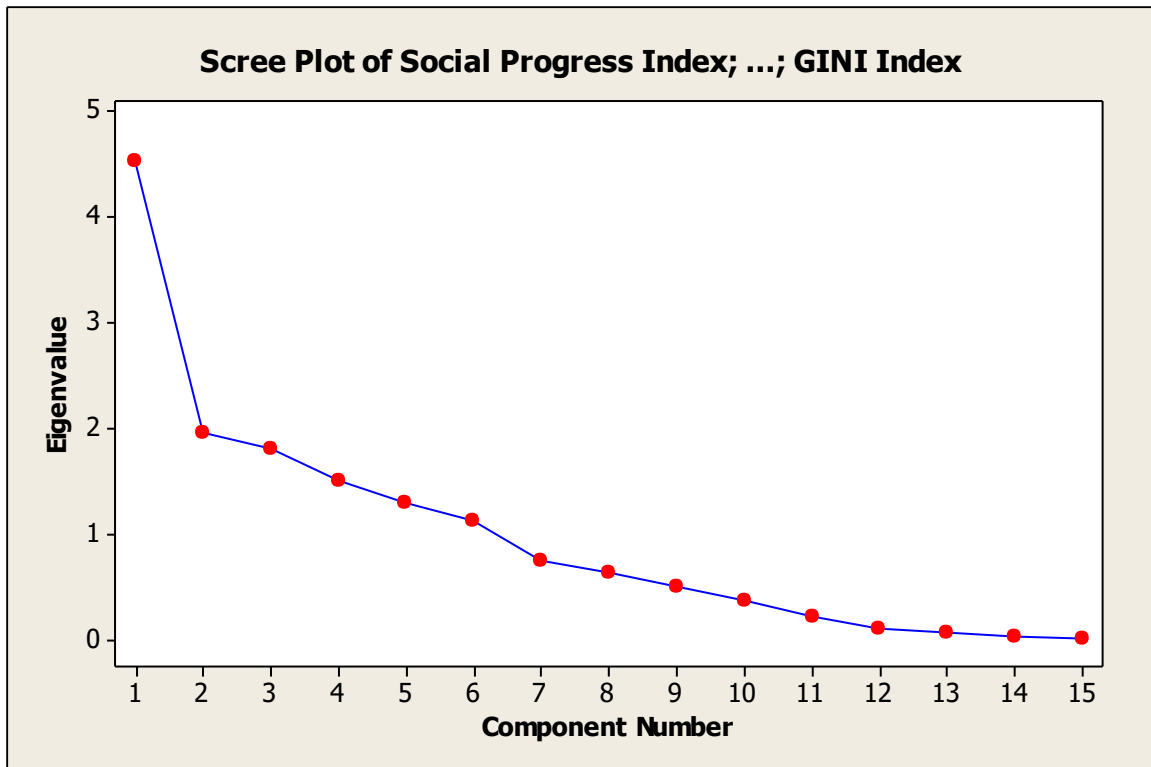
ANÁLISE

O Eigenvalue que é um valor criado pelo sistema para justificar a participação das variáveis PC1 e PC2 na explicação da variável selecionada para o SPI indicou que ela é explicada por 30,30% pela PC1, cujo resultado é a divisão do Engenvalue 4,5394 por 15 variáveis porque aqui tiramos a Variável Desnutrição que apresentava valor zero e o sistema não gerava os dados para análise.



ANÁLISE

Aqui os dados apresentam uma direção mais definida que o anterior. Note que as variáveis da região AVECO HPI e EV estão bem próximas e em uma mesma direção que é a diagonal superior do lado positivo. As outras variáveis concentradas para a direita quase paralela ao eixo da PC1 (first componente) IDH, EPI e Qualidade do fornecimento da Energia Elétrica apontam que existe uma forte correlação entre elas e as três podem ser explicadas pela PC1. Com o mesmo sentido, porém em direção decrescente para o lado direito estão as variáveis Governança e SPI muito paralelas e pouco recuada está o Acesso a Eletricidade. A variável Déficit de alimentos e calorías poderia ser deixada de fora da análise porque ela não é explicada pelas componentes PC1 nem a PC2.



ANÁLISE

Como o dado do gráfico loading plot mostrou com as componentes PC1 e PC2 e mais o engenvale de 4,5394 para 15 variáveis que dava um percentual de explicação da variável SPI pela PC1 de 30,3%, aqui no Scree Plot a informação é nítida com o ponto iniciando no próximo do 5 no eixo vertical e com o ponto no 2 na descendência em direção ao lado esquerdo quase paralelo ao eixo.

8.3 – Região OUTROS

Para análise dos dados da Região OUTROS foi inserida a variável Undernourishment (% of pop.) porque nesta região temos países com diferenças de desenvolvimento como Zâmbia e Emirados Árabes Unidos.

Principal Component Analysis: Social Progr; Depth of foo; Access to pi; Rural v

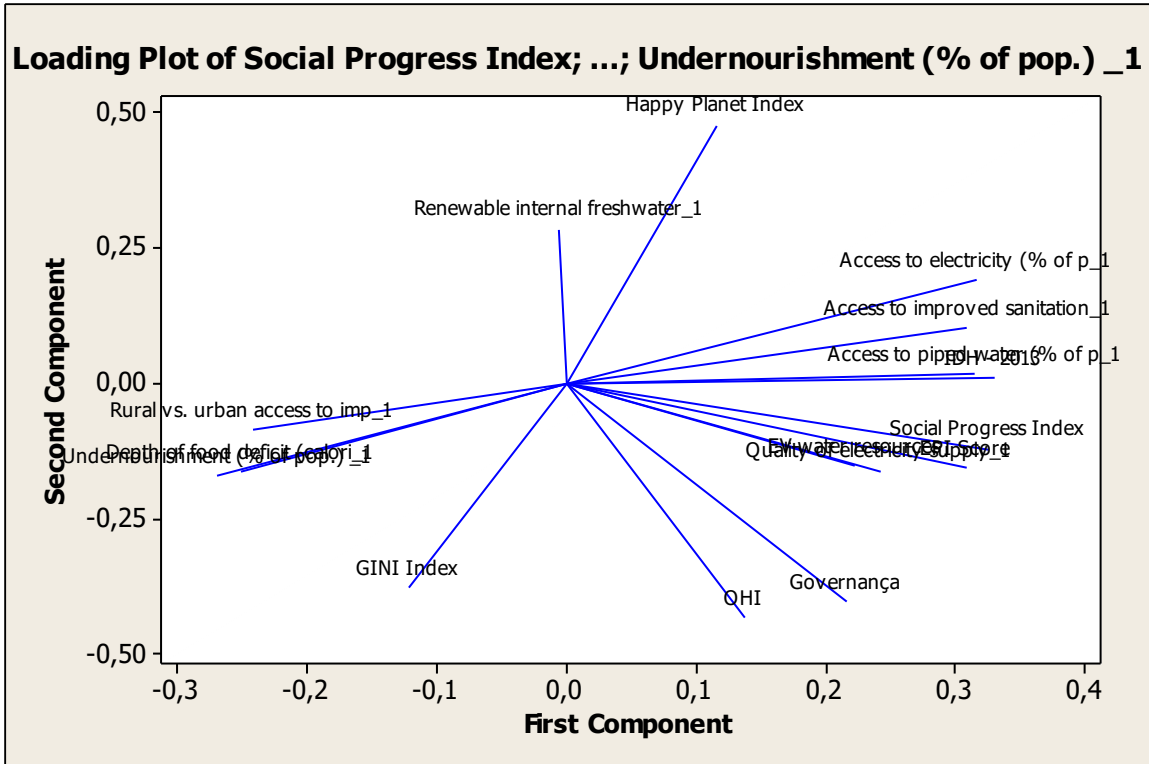
Eigenanalysis of the Correlation Matrix

Eigenvalue	8,2065	1,6817	1,1940	1,0511	0,7378	0,6937	0,6302	0,5124
Proportion	0,513	0,105	0,075	0,066	0,046	0,043	0,039	0,032
Cumulative	0,513	0,618	0,693	0,758	0,804	0,848	0,887	0,919
Eigenvalue	0,4472	0,3056	0,1806	0,1379	0,0938	0,0636	0,0407	0,0233
Proportion	0,028	0,019	0,011	0,009	0,006	0,004	0,003	0,001
Cumulative	0,947	0,966	0,978	0,986	0,992	0,996	0,999	1,000

Variable	PC1	PC2
Social Progress Index	0,325	-0,121
Depth of food deficit (calori_1	-0,250	-0,165
Access to piped water (% of p_1	0,315	0,017
Rural vs. urban access to imp_1	-0,241	-0,085
Access to improved sanitation_1	0,309	0,101
Access to electricity (% of p_1	0,316	0,191
Quality of electricity supply_1	0,242	-0,162
Renewable internal freshwater_1	-0,005	0,284
EV water resources	0,223	-0,154
OHI	0,138	-0,433
IDH - 2013	0,330	0,008
Governança	0,216	-0,403
EPI Score	0,309	-0,155
Happy Planet Index	0,116	0,476
GINI Index	-0,122	-0,378
Undernourishment (% of pop.) _1	-0,268	-0,171

ANÁLISE

Para a região OUTROS o eigenvalue ficou com 8,2065 e a variável SPI é explicada por 51,3 % da componente PC1. São consideradas todas as variáveis do estudo que somam 16.



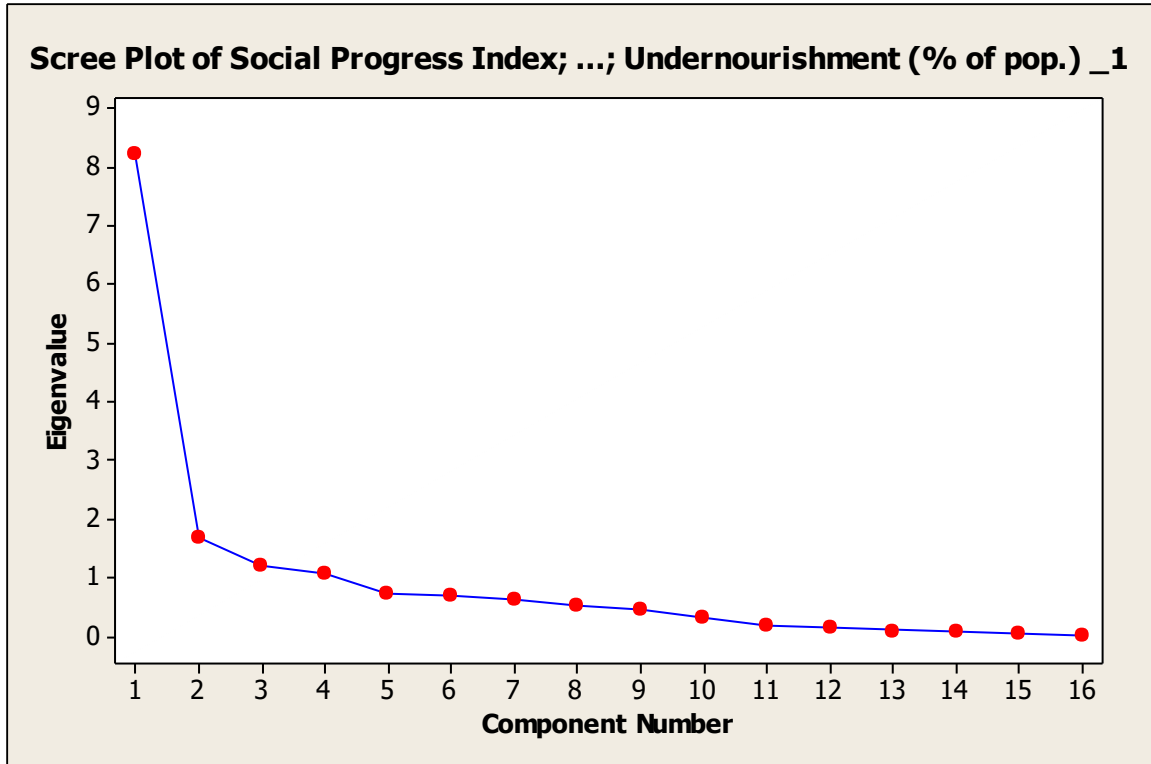
ANÁLISE

O gráfico loading plot tem muita semelhança com da região AVECO apenas a qualidade das variáveis são diferentes e para o negativo.

Aqui as linhas de Acesso a eletricidade, instalações sanitárias, água encanada e EPI estão quase que sobrepostas na direção da direita.

Para o lado oposto e com sentido descendente do lado negativo estão a Desnutrição e Déficit Alimentar estão sobrepostas indicando que elas poderiam ser unidas em uma ou eliminar uma e ficar com a outra que não faz diferença.

O fato de ter países como Zâmbia, Togo e Tanzânia nesta região explica esta variável negativa aparecer com destaque dada a falta de alimentos da população.



ANALISE

Quando se olha para o Scree Plot nota-se que o engenvalue estar pouco acima do 8 e com a direção descendente apontada para duas componentes mas com uma explicando a SPI, mostra que poderia ser utilizada apenas uma para analisar o contexto em função da qualidade baixa das demais variáveis.

8.4 – Comparação entre as três regiões: AIBER, AVECO e OUTROS

A análise feita pelos Componentes Principais separando cada uma das três regiões mostrou a diferença entre os dados de cada uma e deixa uma evidencia quanto a diferença entre elas tanto pela qualidade das variáveis para a região AVECO, por exemplo, e pela baixa qualidade das variáveis para as regiões AIBER e OTHERS mas isto não é suficiente.

Enquanto na avaliação da AVECO apareceu variável como SPI nas outras duas aparecia a variável Desnutrição com relevância, ou sejam a qualidade referida aqui está evidenciada nesta citação. Então, neste item é apresentado uma comparação entre elas em função da variável sintética SPI.

Componentes Principais das Regiões AIBER, AVECO e OUTROS

Principal Component Analysis: Social Progr; Depth of foo; Access to pi; Rural v

Eigenanalysis of the Correlation Matrix

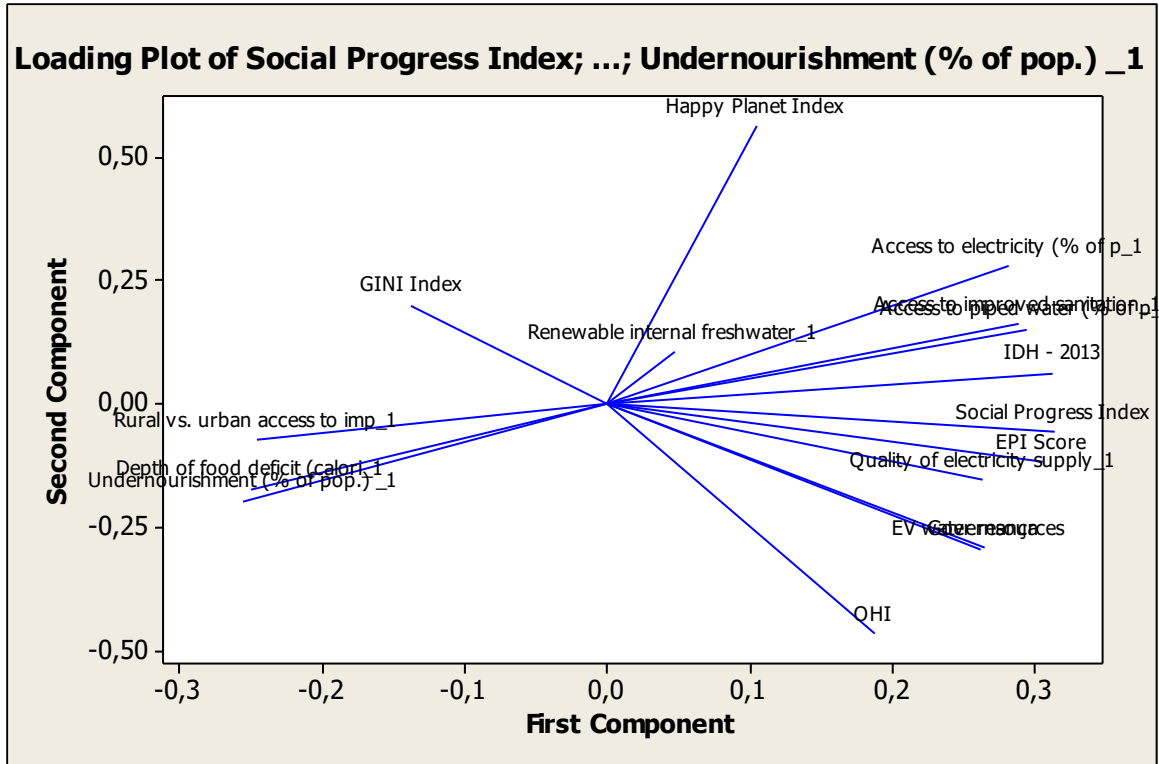
Eigenvalue	9,2544	1,6093	1,1286	0,9195	0,7316	0,5811	0,4517	0,4118
Proportion	0,578	0,101	0,071	0,057	0,046	0,036	0,028	0,026
Cumulative	0,578	0,679	0,750	0,807	0,853	0,889	0,917	0,943

Eigenvalue	0,2806	0,2042	0,1341	0,0978	0,0829	0,0648	0,0245	0,0231
Proportion	0,018	0,013	0,008	0,006	0,005	0,004	0,002	0,001
Cumulative	0,961	0,973	0,982	0,988	0,993	0,997	0,999	1,000

Variable	PC1	PC2
Social Progress Index	0,313	-0,057
Depth of food deficit (calori_1	-0,250	-0,173
Access to piped water (% of p_1	0,294	0,151
Rural vs. urban access to imp_1	-0,246	-0,075
Access to improved sanitation_1	0,288	0,162
Access to electricity (% of p_1	0,282	0,277
Quality of electricity supply_1	0,263	-0,156
Renewable internal freshwater_1	0,047	0,104
EV water resources	0,261	-0,294
OHI	0,187	-0,465
IDH - 2013	0,312	0,061
Governança	0,265	-0,291
EPI Score	0,306	-0,120
Happy Planet Index	0,104	0,562
GINI Index	-0,137	0,199
Undernourishment (% of pop.) _1	-0,255	-0,197

ANÁLISE

Com a junção das três regiões os dados ficaram um pouco mais robusto com tendência a proximidade dos dados apresentados para a região OUTROS. Note que o Eigenvalue está com 9,2544 e a PC1 explica 57,8 % da variável SPI, ou seja, como são 16 variáveis (8 sintética e 8 componentes) a divisão de 9,2544 por 16 variáveis. Como a amostra é maior os dados ficam mais representativos, logo, o fato de ter mais de 50% de explicação da variável SPI pelo PC1 tem valor no contexto.



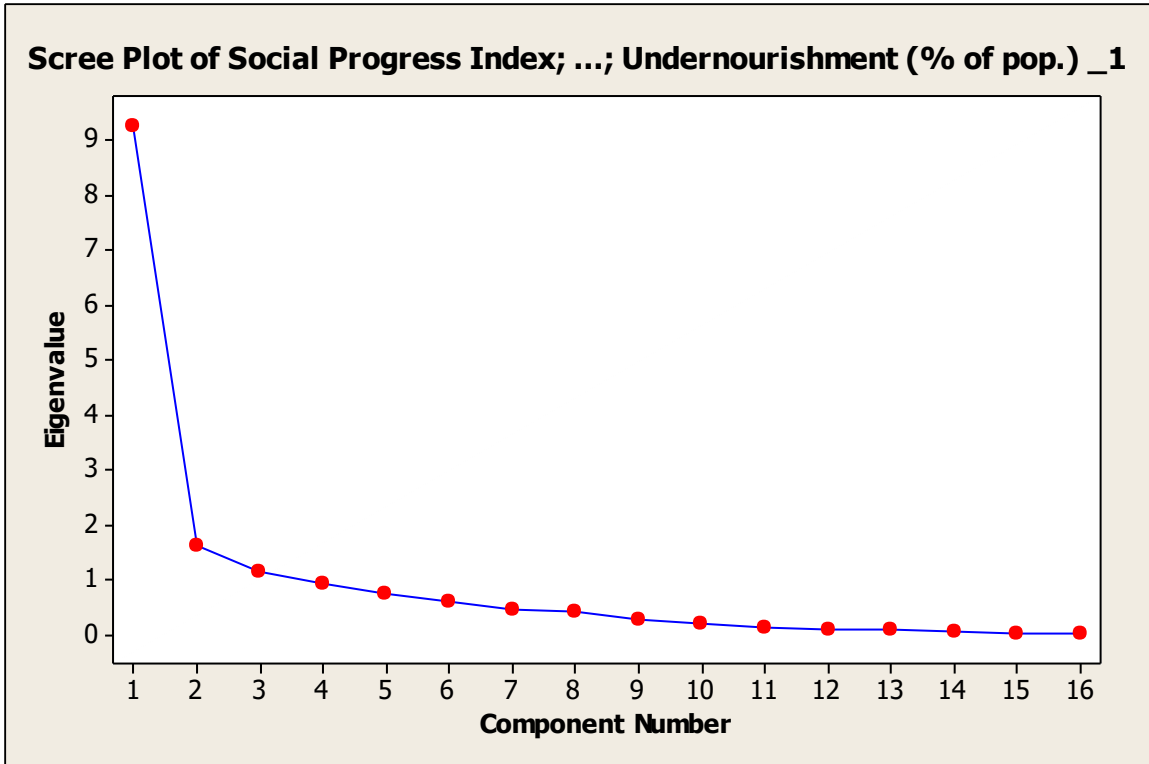
ANÁLISE

Aqui é possível notar que o gráfico apresenta dois conjuntos distintos de variáveis convergindo para a direita, o maior deles, e o menor para a esquerda.

Do lado direito as variáveis Acesso a eletricidade, SPI EPI qualidade no fornecimento de energia elétrica, por exemplo, estão mostrando que para os países destas regiões elas são significativas.

O Acesso a recurso de água da zona rural e urbana, Desnutrição e Déficit de calorias estão concentradas no sentido negativo e com tendência para mais negativo.

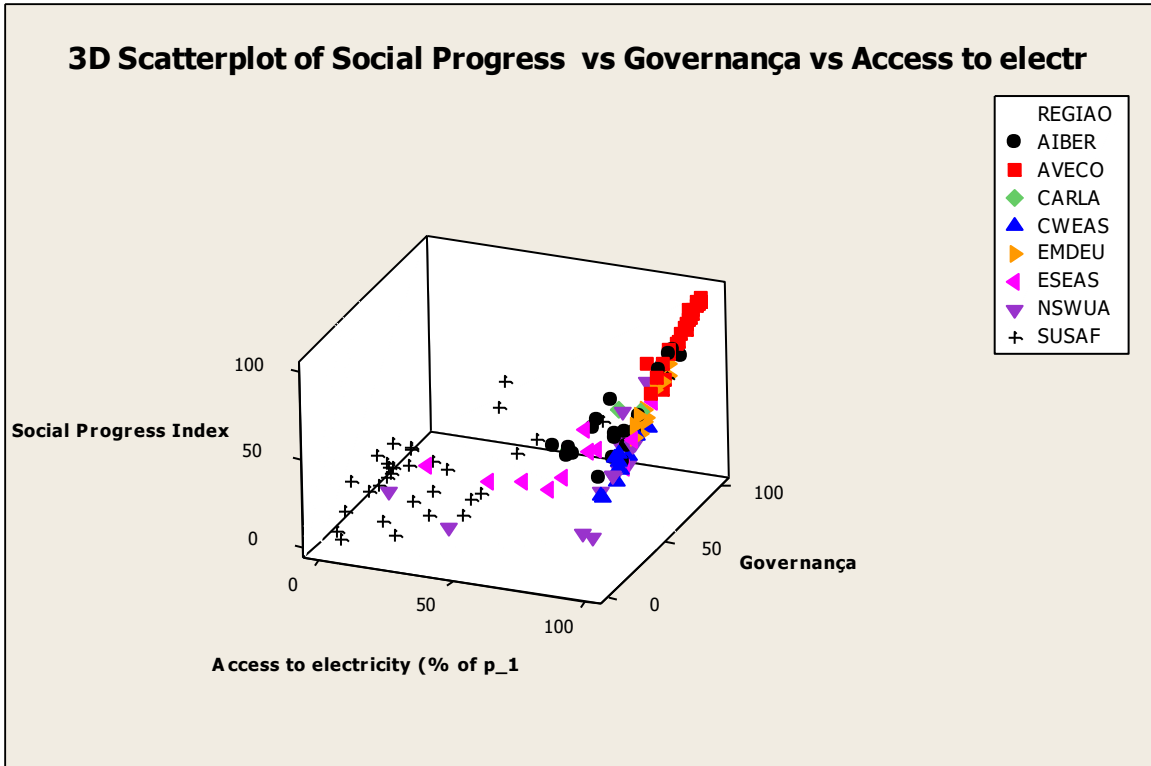
De um lado as regiões AVECO e parte dos OUTROS puxaram os indicadores mais qualitativos para o lado direito e de outro sentido países da região OUTROS de baixa renda e mais a região AIBER levaram os resultados para o baixo.



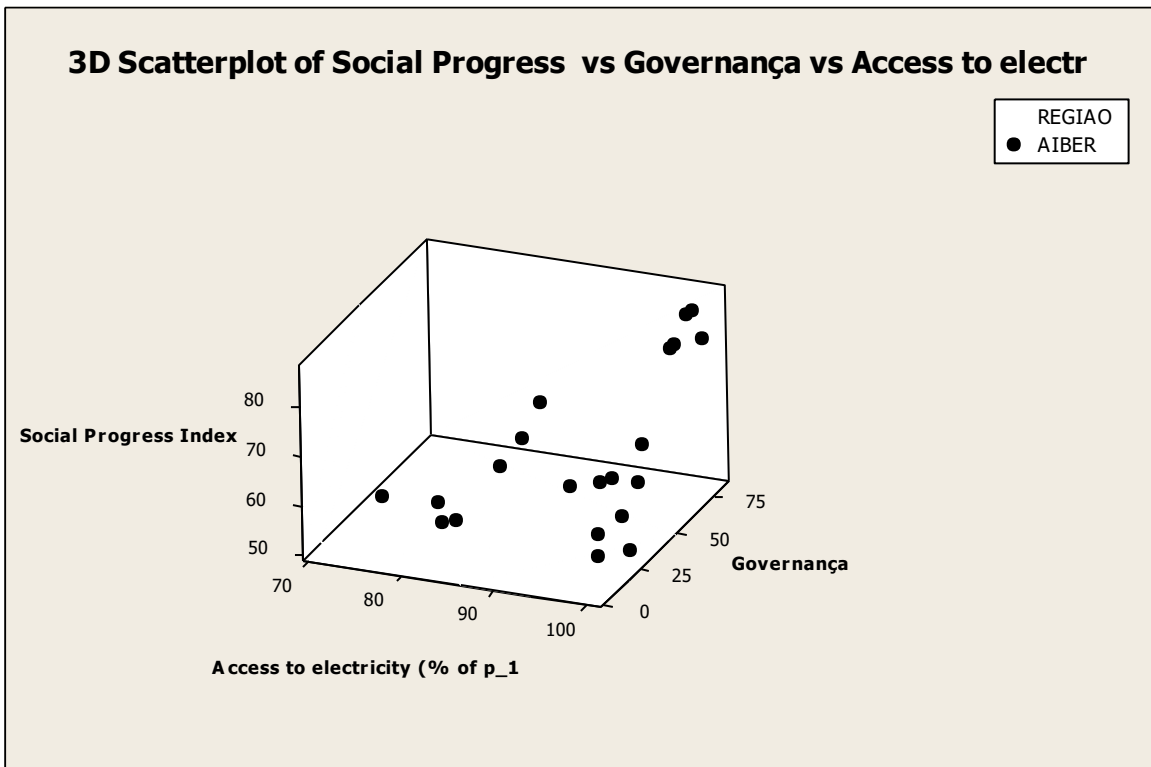
ANÁLISE

O gráfico do Scree Plot que mostra quantos componente explicam a variável, fica mais evidente que mesmo com um Eigenvalue maior isto não deixou que a componente PC1 fosse mostrar a importância para a variável SPI. Ainda continuam com no máximo duas componentes para explicar a variável.

Fazendo a representação com o Scatterplot 3D com as três regiões e as variáveis sintéticas SPI e Governança mais a variável analítica Acesso a eletricidade nota-se que há uma concentração da Região AVECO para SPI e Governança.



Com o gráfico apenas da Região AIBER os valores são menores mas a concentração do lado superior direito com o Chile para Governança e SPI.

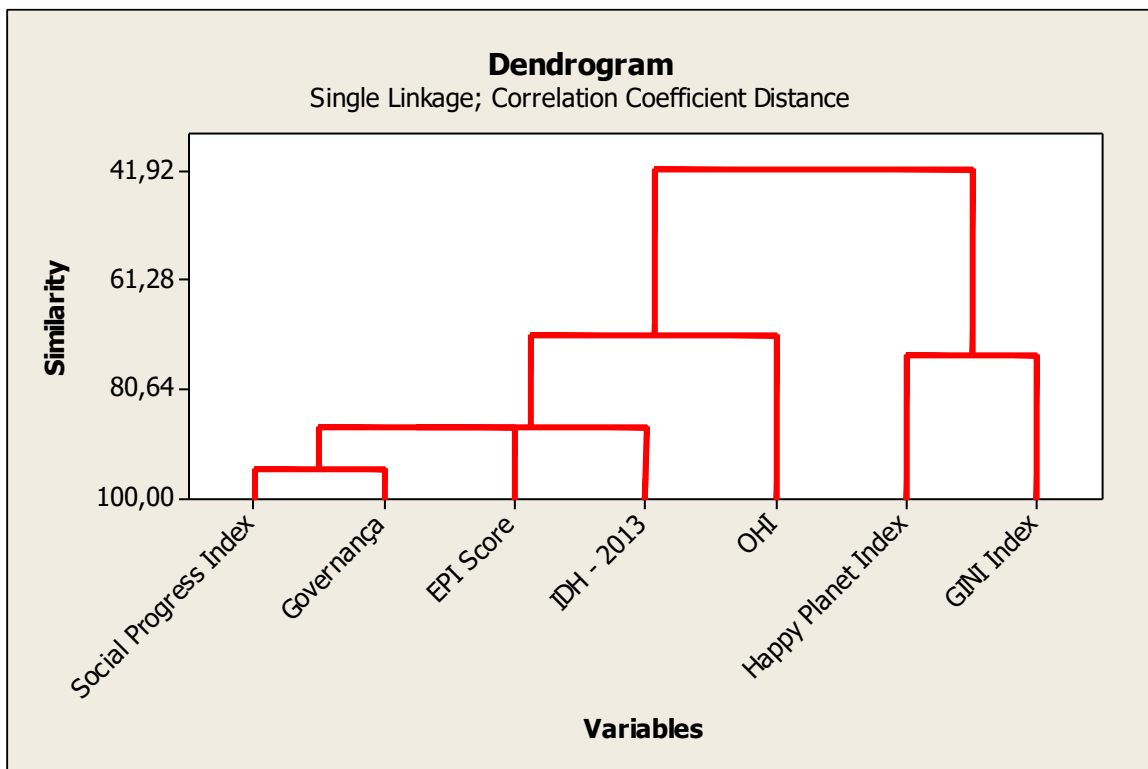


9. ANÁLISE DE CONGLOMERADOS

A análise de conglomerados ou análise de agrupamento ou o agrupamento é a tarefa de agrupar um conjunto de variáveis de um mesmo grupo de tal forma que as variáveis sejam semelhantes entre si e em relação a outras de outro grupo. Muitas vezes o processo deve ser repetido várias vezes combinando as variáveis para encontrar o maior número de semelhanças.

9.1 – Região AIBER

Para a exposição deste item, a primeira região é AIBER com a apresentação do Dendrograma pois permite uma análise do grau de similaridade entre os países com relação as variáveis estudadas. Iniciando pelas variáveis sintéticas: SPI, GOVERNANÇA, EV, HPI, EPI, IDH, GINI e OHI:



ANÁLISE

O SPI e Governança formam um cluster com grau de similaridade de 94,6 e o EPI mais IDH formam outro cluster e somam com SPI e Governança e seu grau de similaridade é 87,18.

Outro cluster é HPI e GINI e mais OHI e este por sua vez com os anteriores, porém seu grau de similaridade não é representativo para a análise da região AIBER.

Cluster Analysis of Variables: Social Progr; OHI; IDH - 2013; Governança; ...

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	7	94,6		2	Social Progress Index, Governança	2
2	6	87,18		2	EPI Score, IDH - 2013	2
3	5	80,64		2	Happy Planet Index, GINI Index	2
4	4	61,28		3	(Societal Cluster), (EPI/IDH Cluster), OHI	5
5	1	41,92		7	All Variables	7

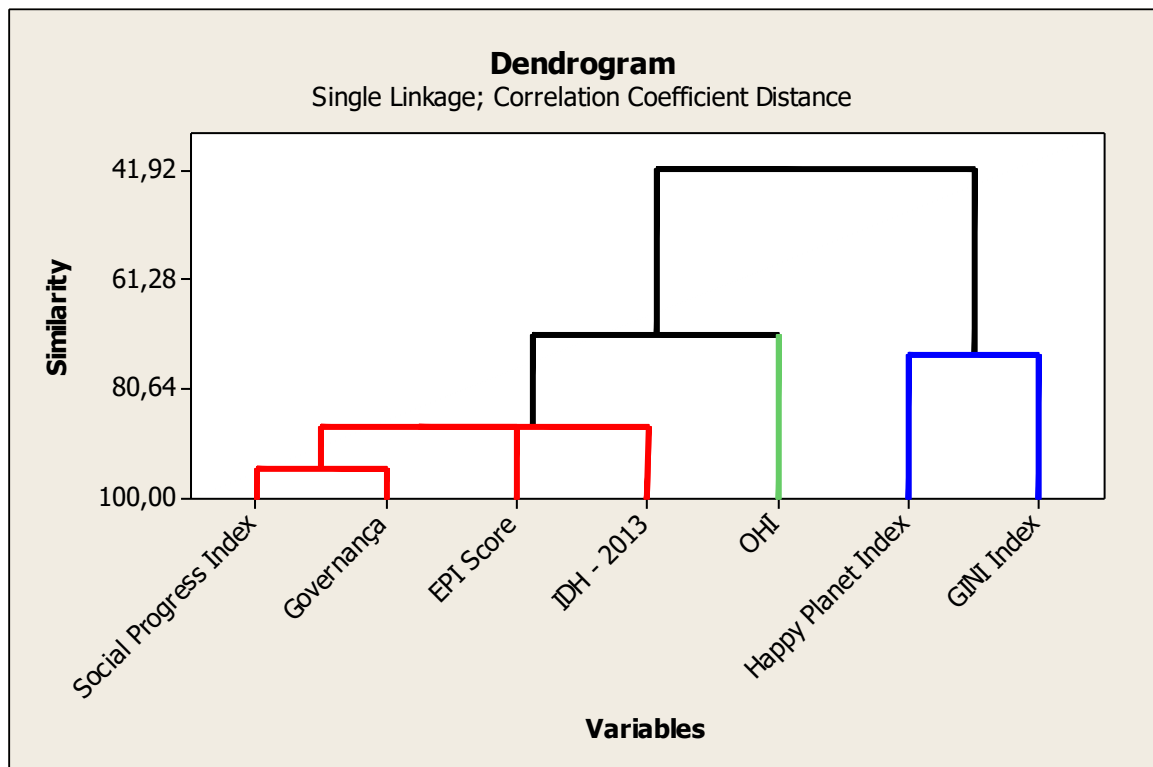
1	6	94,6018	0,10796	1	4	1	2
2	5	87,4755	0,25049	1	5	1	3
3	4	87,1804	0,25639	1	3	1	4
4	3	74,5863	0,50827	6	7	6	2
5	2	71,0819	0,57836	1	2	1	5
6	1	41,9172	1,16166	1	6	1	7

Final Partition

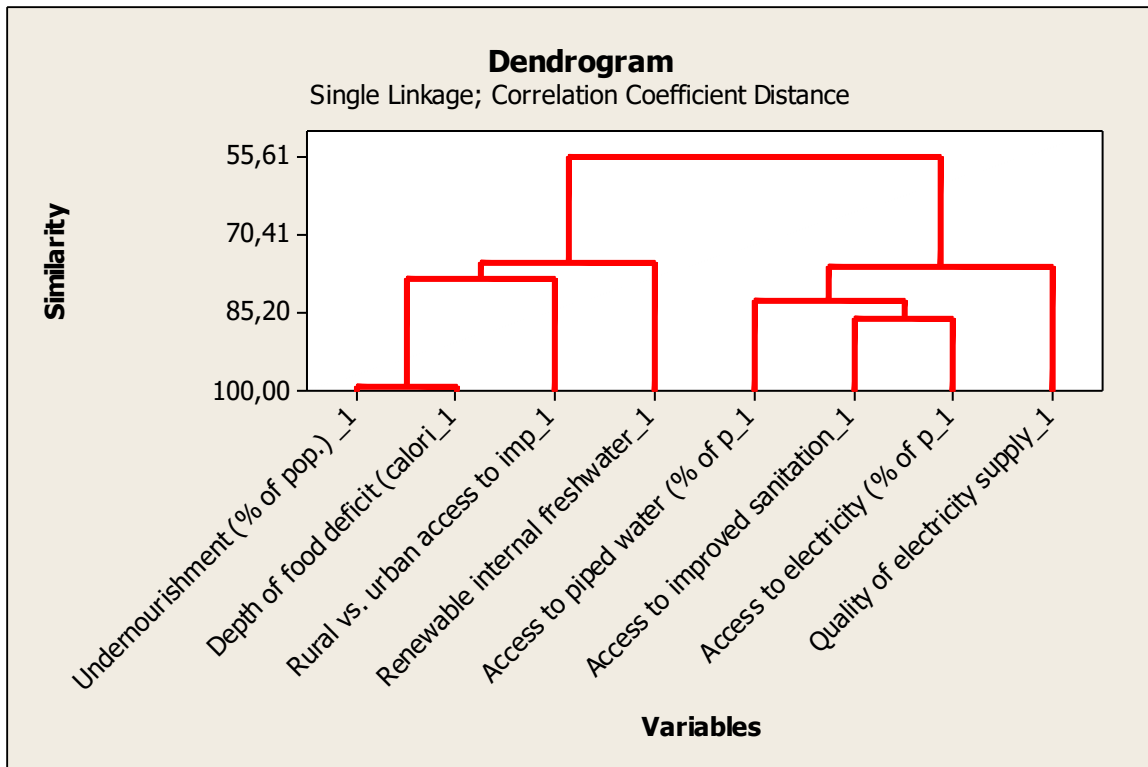
Cluster 1
Social Progress Index OHI IDH - 2013 Governança EPI Score

Cluster 2
Happy Planet Index GINI Index

Pedindo novo agrupamento com 2 e 3 clusters não houve alteração conforme mostra o de 3.



Agora para as variáveis analíticas da região AIBER



ANÁLISE

Conforme observado anteriormente em outras análises as variáveis Desnutrição e Déficit de Alimentos poderiam ser unidas, ou seja, para efeito de análise do tema poderia ficar uma ou outra por causa da sua similaridade próxima de 100%, com 99,21%.

Outro cluster informado é o Acesso as instalações sanitárias e Acesso a Eletricidade com grau de similaridade de 86,32%.

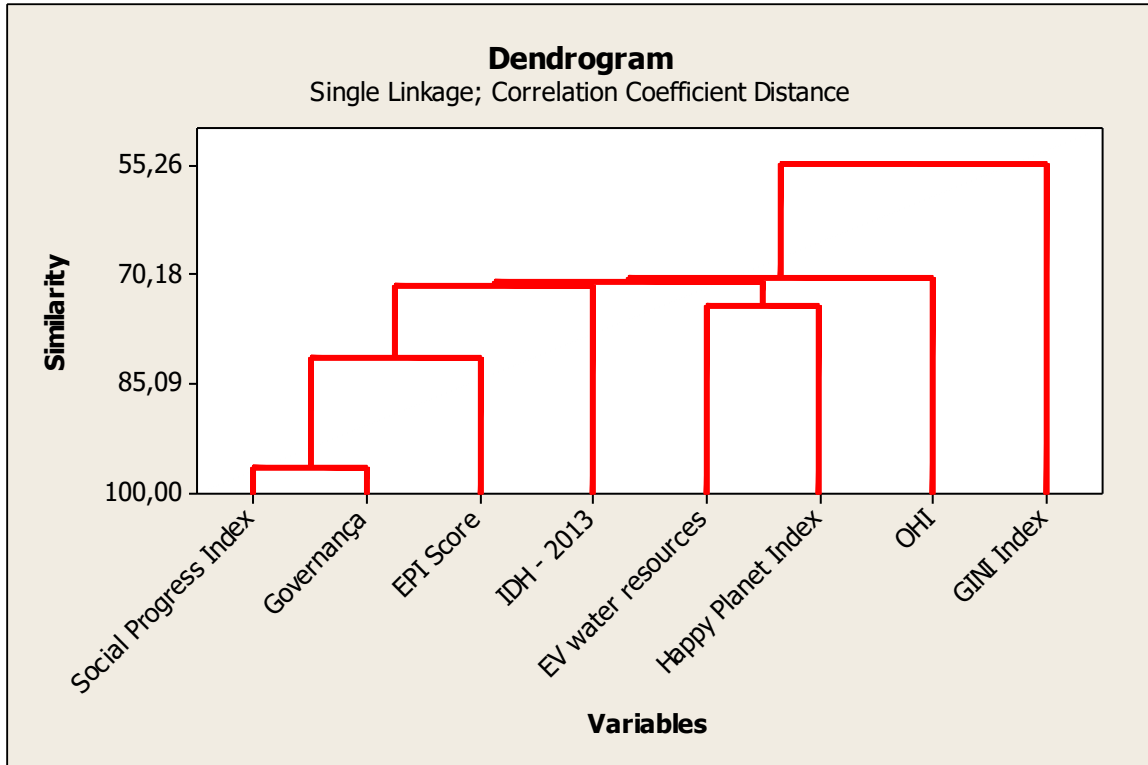
Cluster Analysis of Variables: Undernourish; Depth of foo; Access to pi; ...

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	7	99,2023	0,015954	1 2	1	2
2	6	86,3194	0,273612	5 6	5	2
3	5	82,9040	0,341920	3 5	3	3
4	4	78,7601	0,424799	1 4	1	3
5	3	76,2970	0,474061	3 7	3	4
6	2	75,6186	0,487627	1 8	1	4
7	1	55,6087	0,887826	1 3	1	8

9.2 – Região AVECO

Primeiro Variáveis Sintéticas:



ANÁLISE

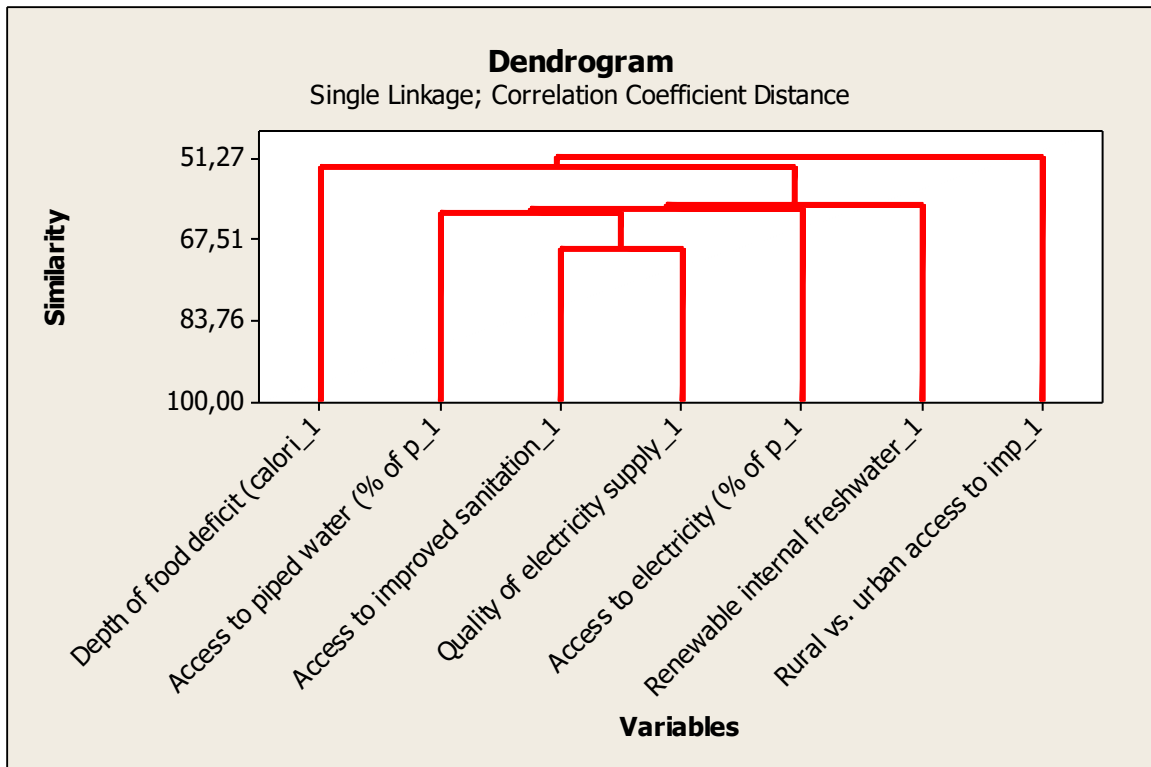
A similaridade foi mostrada entre SPI e Governança e este cluster com o EPI. A variável SPI e mais Governança com 96,55% de similaridade.

Cluster Analysis of Variables: Social Progr; EV water res; OHI; IDH - 2013; ...

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	7	96,5494	0,069012	1 5	1	2
2	6	81,4950	0,370099	1 6	1	3
3	5	74,3521	0,512957	2 7	2	2
4	4	71,7688	0,564625	1 4	1	4
5	3	71,1852	0,576295	1 2	1	6
6	2	70,7894	0,584212	1 3	1	7
7	1	55,2640	0,894720	1 8	1	8

As variáveis analíticas:



ANÁLISE

Foi preciso tirar a variável da Desnutrição porque para esta região seus valores é Zero para todos os países.

O resultado final ficou com grupos distintos, mas fora do habitual com similaridade próxima de 100% e depois os mais afastados com menor similaridade.

Aqui o resultado foi de 69,4641 até a menor 51,2659. Isto fica melhor evidenciado no quadro a seguir:

Cluster Analysis of Variables: Depth of foo; Access to pi; Rural vs. ur; ...

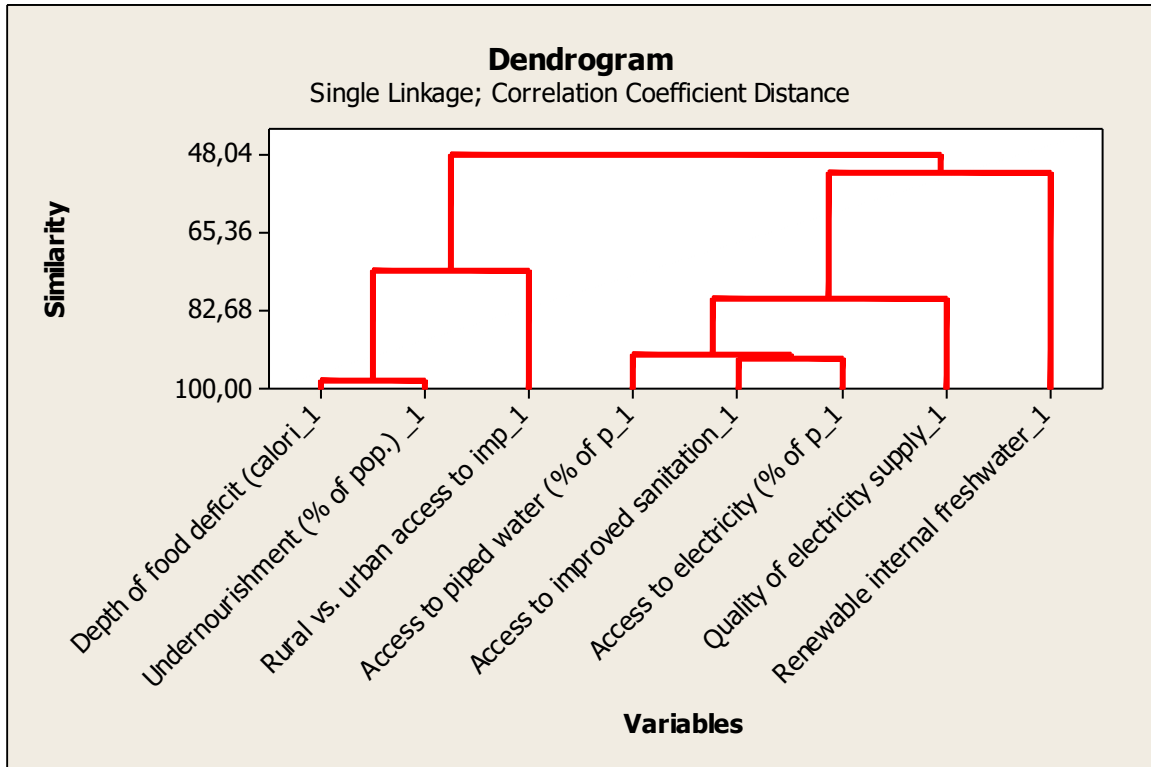
Correlation Coefficient Distance, Single Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	6	69,4641	0,610718	4	6	2
2	5	62,4645	0,750711	2	4	3
3	4	61,6423	0,767154	2	5	4
4	3	60,5377	0,789246	2	7	5
5	2	53,0714	0,938571	1	2	6
6	1	51,2659	0,974681	1	3	7

9.3 – Região OUTROS

Dendrograma das variáveis sintéticas:

Variáveis analíticas:



ANÁLISE

Embora esta região tenha um total de 84 países o que representa 4 vezes a região AIBER e 3,11 vezes a região da AVECO, os dados são bem próximos da região AIBER.

O cluster do déficit de alimentos e desnutrição poderia ser uma variável. O cluster Acesso a água encanada, Acesso as instalações sanitárias e acesso a eletricidade com similaridade de 92,50%.

Cluster Analysis of Variables: Depth of foo; Access to pi; Rural vs. ur; ...

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	7	98,2277	0,03545	1	8	2
2	6	93,5121	0,12976	4	5	2
3	5	92,4972	0,15006	2	4	3
4	4	80,1456	0,39709	2	6	4
5	3	73,7533	0,52493	1	3	3
6	2	51,9584	0,96083	2	7	5
7	1	48,0406	1,03919	1	2	8

10. COMPARAÇÃO DE MÉDIA, ANÁLISE DE VARIÂNCIA E INTERVALO DE CONFIANÇA.

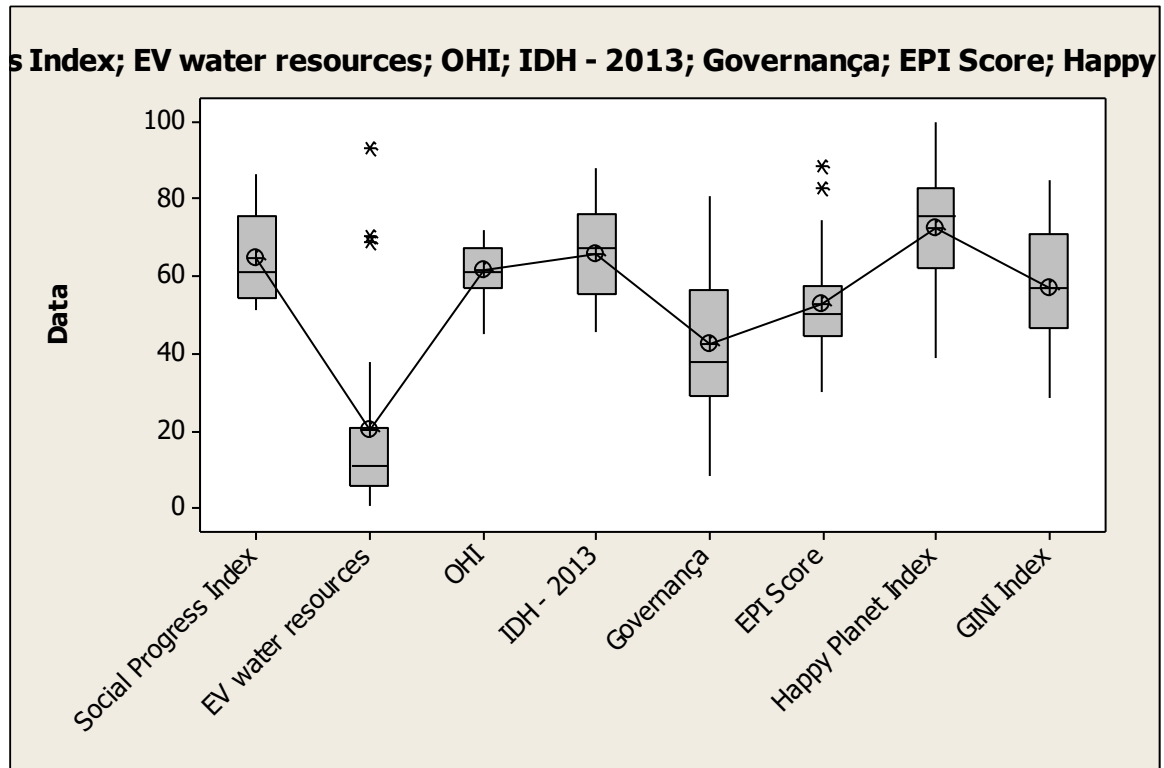
Análise de Variância: permite a verificação e visualização das médias e desvios padrões da variável a ser analisada

Média: é uma medida de posição que serve de referência para os demais valores de uma amostra

Intervalo de confiança: mostra a relação entre as variáveis no contexto de força para um intervalo de 95% de confiança.

10.1 – Para este item vamos apresentar a Região AIBER

Variáveis sintéticas



ANÁLISE

Apenas o EPI está com outlier: Portugal com 82,8567 e a Espanha com 88,6193 média é 52,6987.

A Espanha aparece novamente em EV com 92,83 e Portugal com 70,33 e a média é 20,1584.

One-way ANOVA: Social Progr; EV water res; OHI; IDH - 2013; Governança; ...

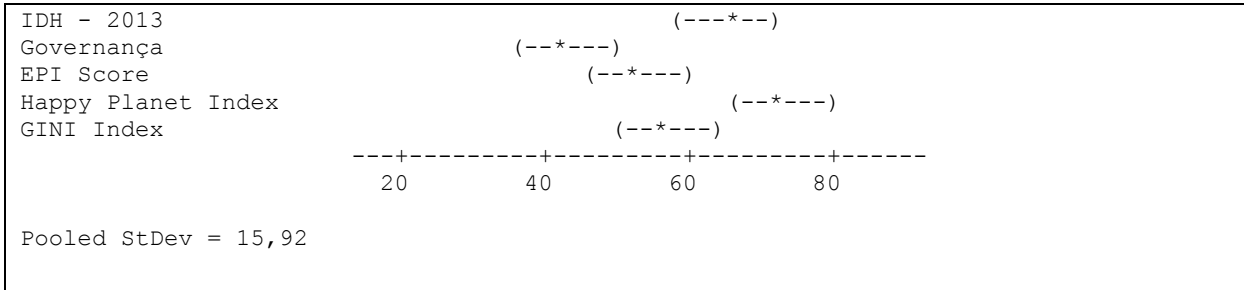
Source	DF	SS	MS	F	P
Factor	7	40399	5771	22,77	0,000
Error	160	40557	253		
Total	167	80957			

S = 15,92 R-Sq = 49,90% R-Sq(adj) = 47,71%

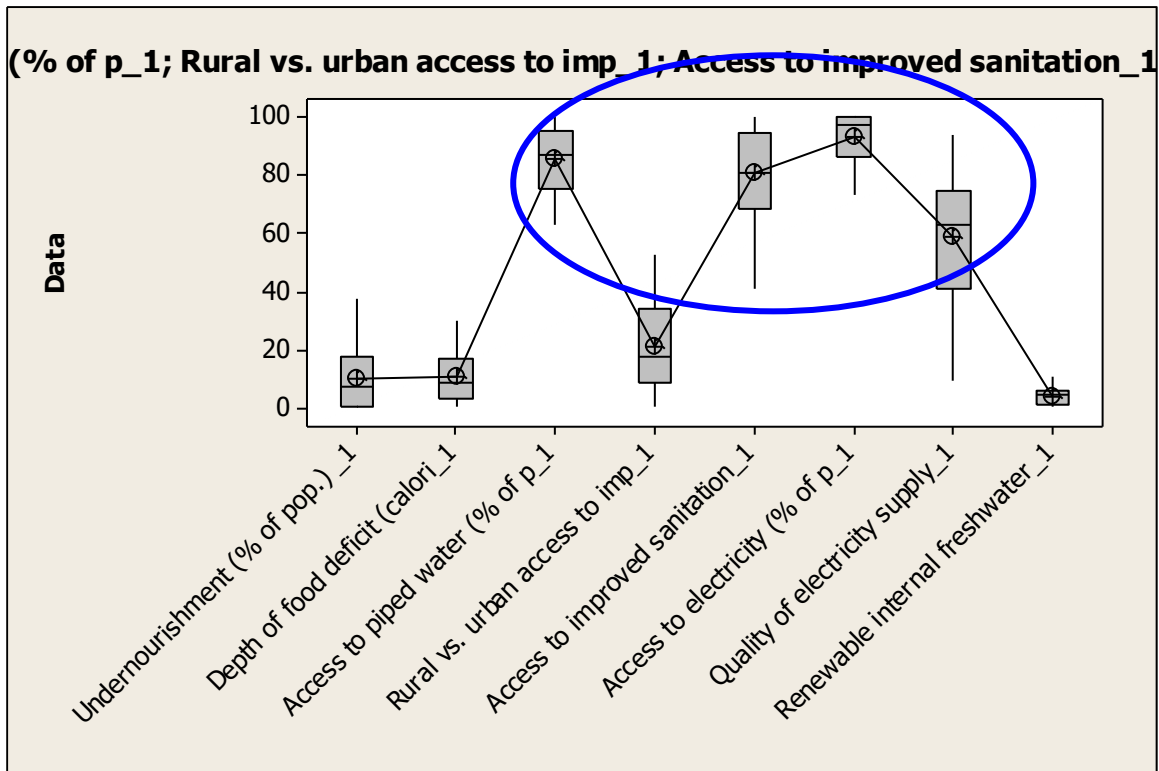
Level	N	Mean	StDev
Social Progress Index	21	64,41	11,95
EV water resources	21	20,15	25,69
OHI	21	61,57	7,06
IDH - 2013	21	65,62	12,03
Governança	21	42,48	18,82
EPI Score	21	52,70	14,31
Happy Planet Index	21	72,36	16,25
GINI Index	21	56,92	14,40

Individual 95% CIs For Mean Based on Pooled StDev

Level	CI
Social Progress Index	(---*---)
EV water resources	(---*---)
OHI	(---*---)



Variáveis Analíticas



ANÁLISE

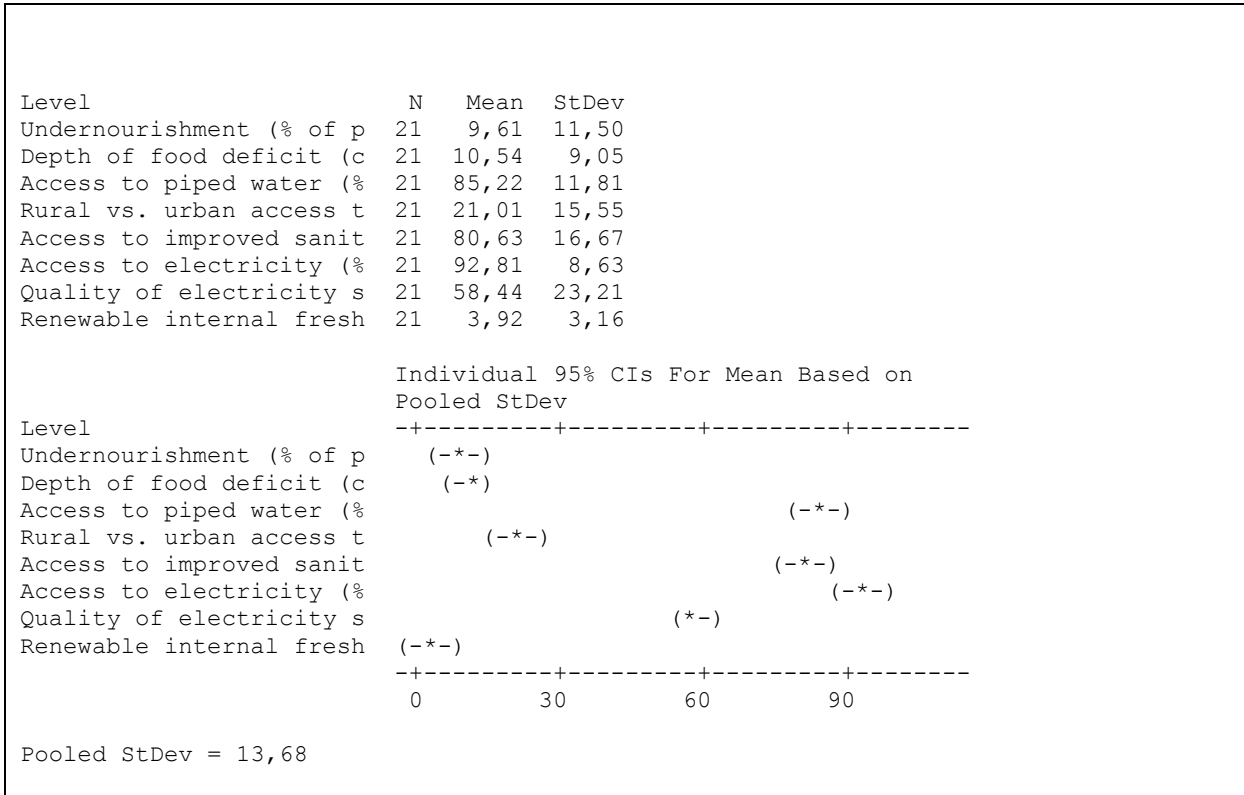
Olhando para este gráfico percebemos que a situação dos países da Região AIBER das variáveis analíticas não está muito boa. Verificamos o Déficit de calorias com média de 10,5352.

As variáveis indicadas no círculo estão boa média é o caso de Acesso a água encanada com média 85,2186.

One-way ANOVA: Undernourish; Depth of foo; Access to pi; Rural vs. ur; ...

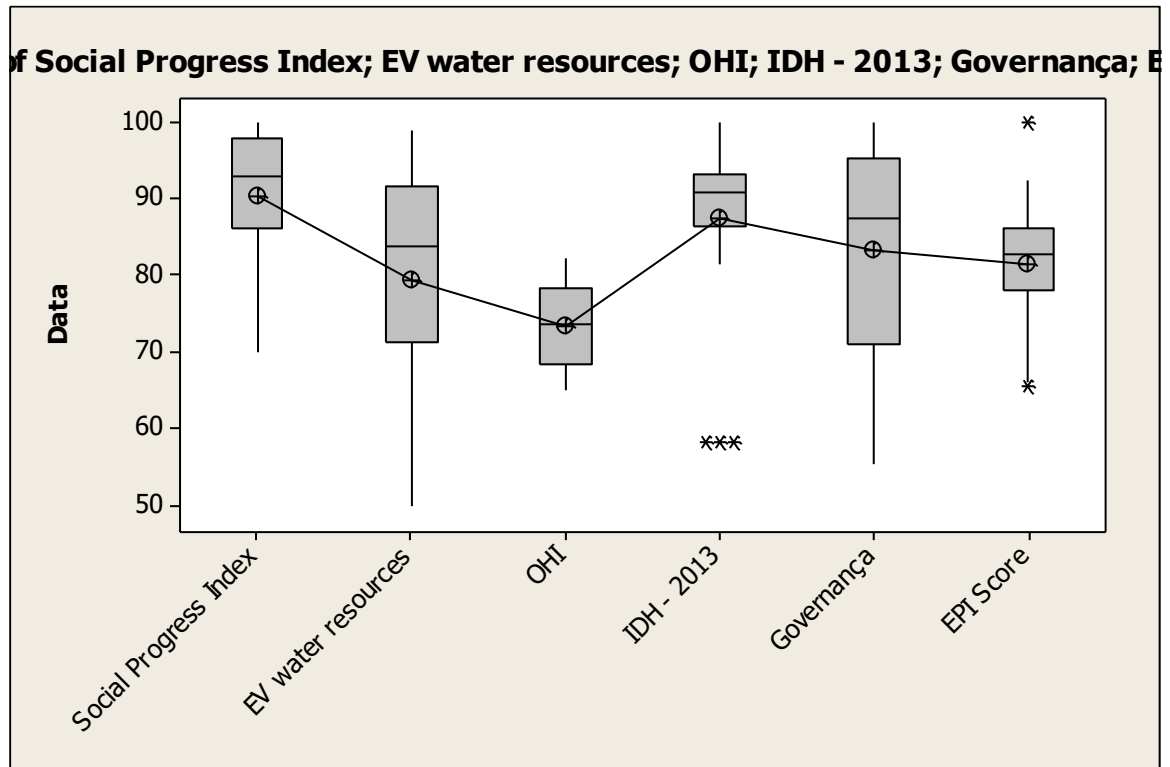
Source	DF	SS	MS	F	P
Factor	7	211177	30168	161,28	0,000
Error	160	29929	187		
Total	167	241106			

S = 13,68 R-Sq = 87,59% R-Sq(adj) = 87,04%



10.2 – Região AVECO

Para as variáveis sintéticas:



ANÁLISE

Para o EPI da região AVECO apareceram dois outliers: Suécia com 100 e República da Coreia com 65,511,.

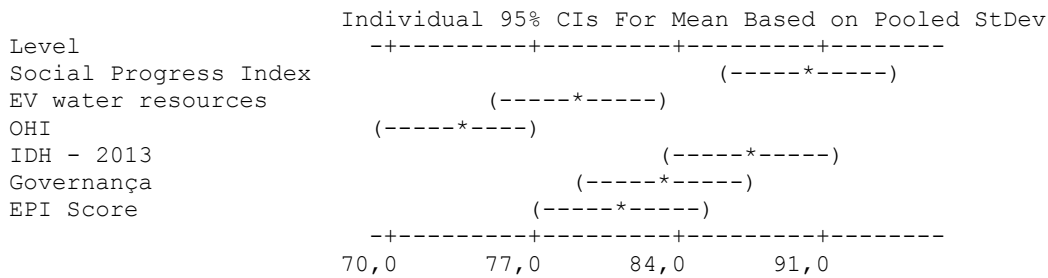
O IDH temos a Holanda, República da Coreia e Látvia.

One-way ANOVA: Social Progr; EV water res; OHI; IDH - 2013; Governança; ...

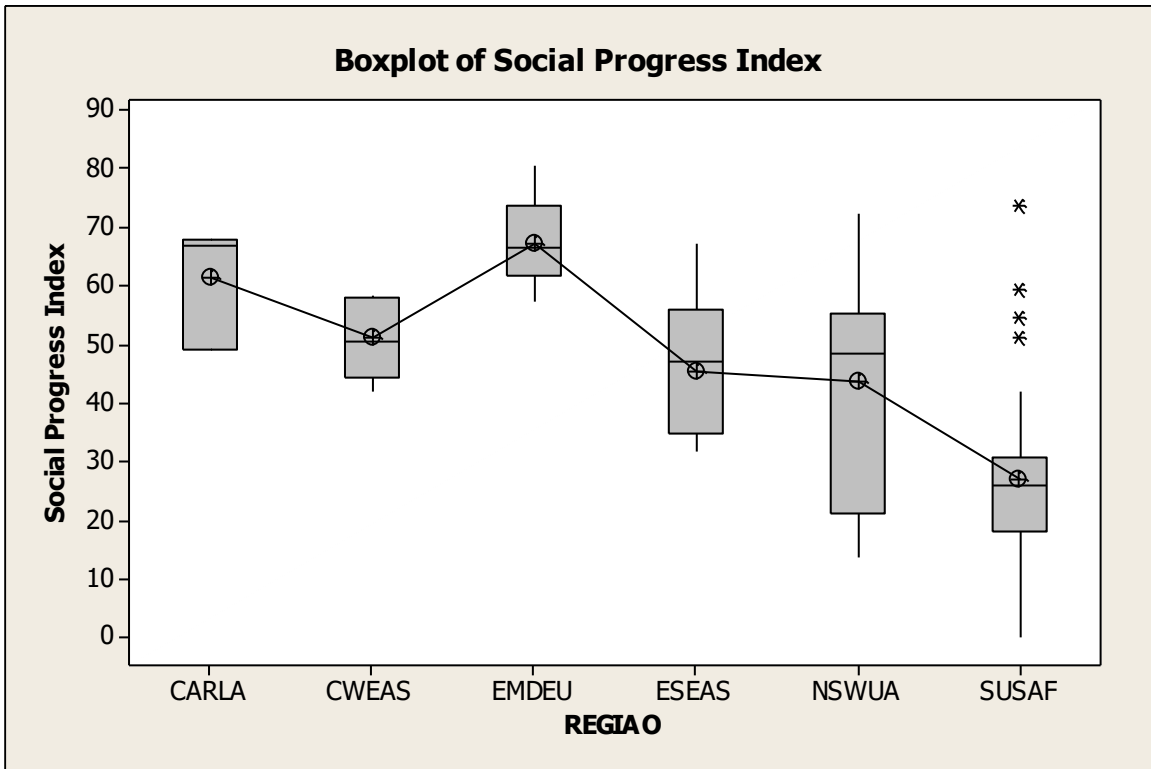
Source	DF	SS	MS	F	P
Factor	5	4956	991,8,51		0,000
Error	156	18168	116		
Total	161	23124			

S = 10,79 R-Sq = 21,43% R-Sq(adj) = 18,92%

Level	N	Mean	StDev
Social Progress Index	27	90,30	8,70
EV water resources	27	79,35	14,70
OHI	27	73,15	5,74
IDH - 2013	27	87,36	11,35
Governança	27	83,25	13,38
EPI Score	27	81,39	8,14



Pooled StDev = 10,79



ANÁLISE

Para a região OUTROS foi apresentado alguns outliers: Ilhas Mauricius, África do Sul, Namíbia e Botsuana todos na região SUSAF.

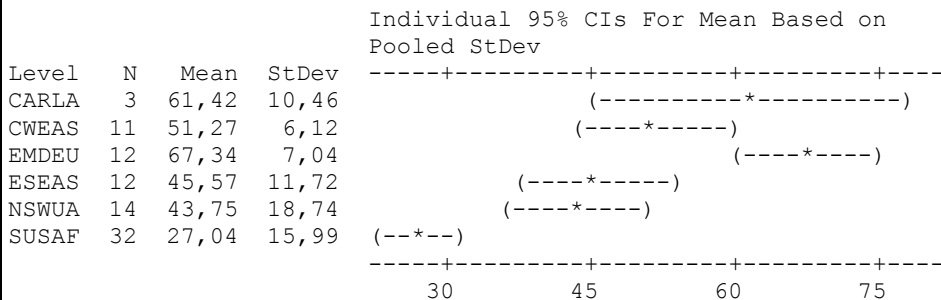
Os pontos circulares no retângulo representa a média. No exemplo a SUSAF tem uma média de 27,03 para o SPI isto significa que outras variáveis que condensam esta sintética estão com valores muito baixo como desnutrição, água encanada e outros. Isto demonstra a qualidade de vida das pessoas desta região que é ruim.

O valor de F é muito baixo entre variável e os países da região com 17,63 e o R-quadrado de 53,06.

One-way ANOVA: Social Progress Index versus REGIAO

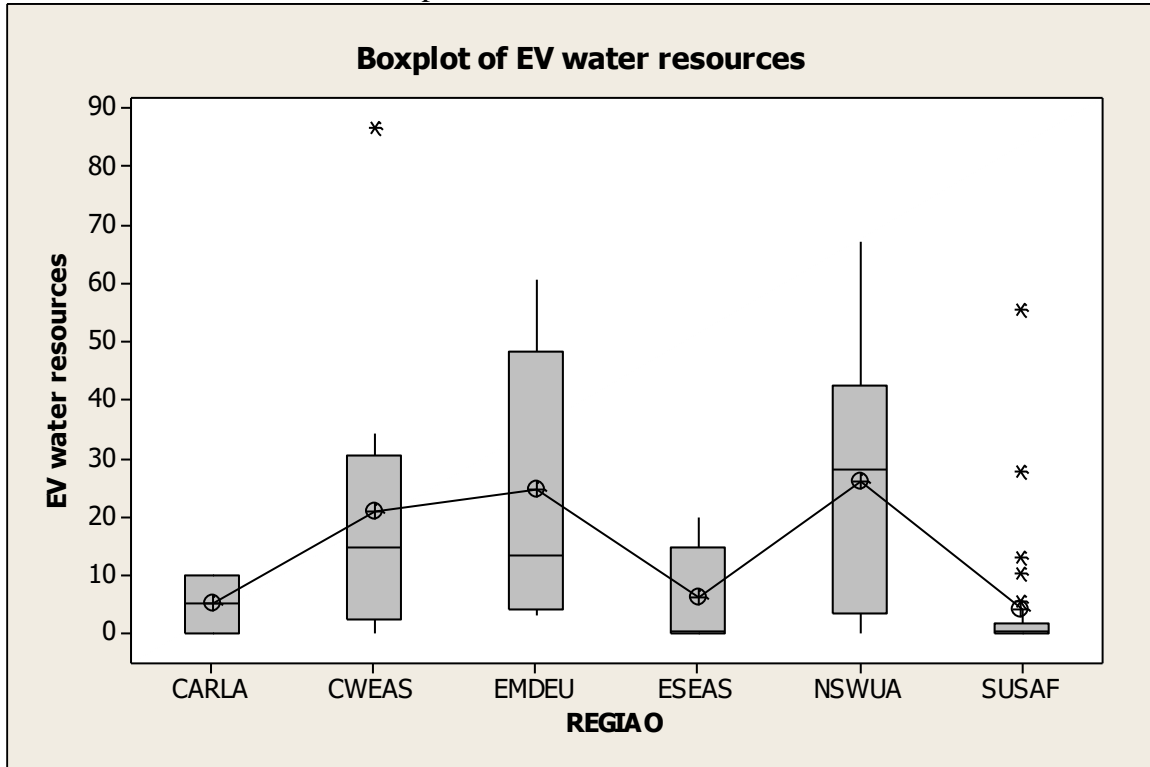
Source	DF	SS	MS	F	P
REGIAO	5	17111	3422	17,63	0,000
Error	78	15138	194		
Total	83	32249			

S = 13,93 R-Sq = 53,06% R-Sq(adj) = 50,05%



Pooled StDev = 13,93

Outra variável sintética EV e os piores resultados estão na SUSAF com diversos outliers



One-way ANOVA: EV water resources versus REGIAO

Source	DF	SS	MS	F	P
REGIAO	5	7955	1591	5,58	0,000
Error	78	22245	285		
Total	83	30199			

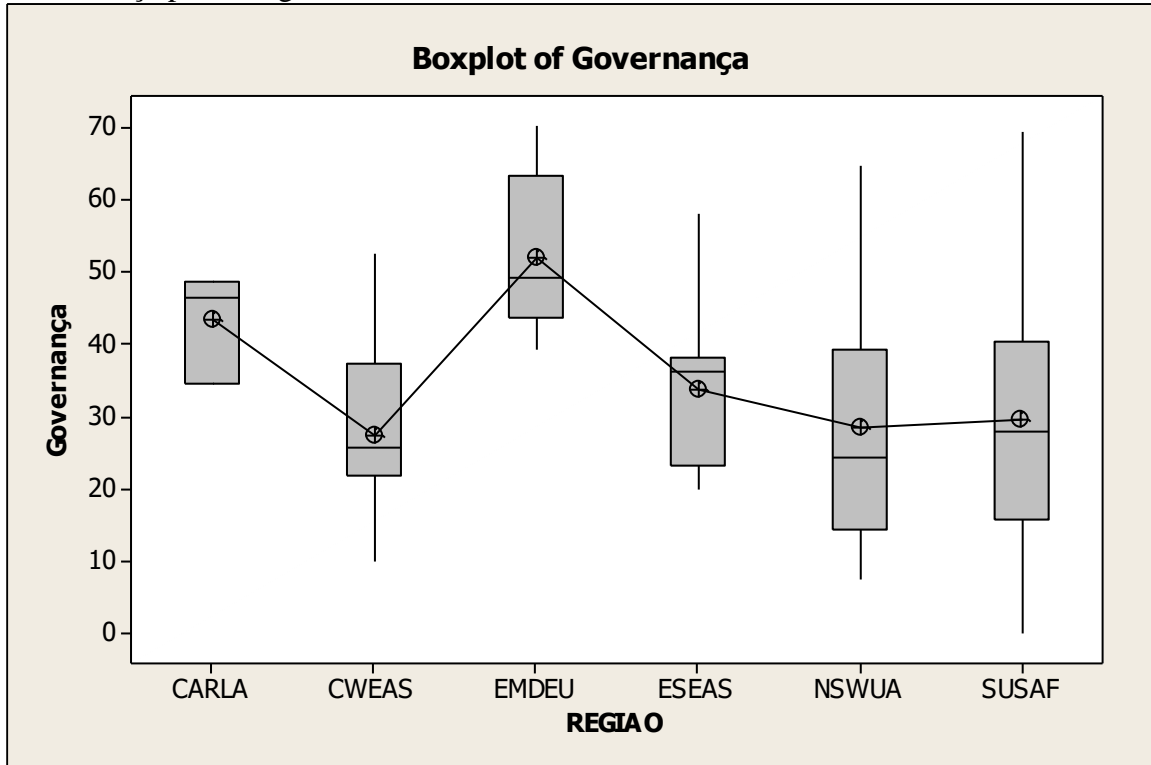
S = 16,89 R-Sq = 26,34% R-Sq(adj) = 21,62%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	CI
CARLA	3	5,08	5,00	(-10,00 ; 20,00)
CWEAS	11	20,86	24,83	(-20,00 ; 60,00)
EMDEU	12	24,45	22,88	(-10,00 ; 60,00)
ESEAS	12	6,14	8,04	(-10,00 ; 20,00)
NSWUA	14	25,88	21,23	(-10,00 ; 60,00)
SUSAF	32	3,94	10,92	(-10,00 ; 20,00)

Pooled StDev = 16,89

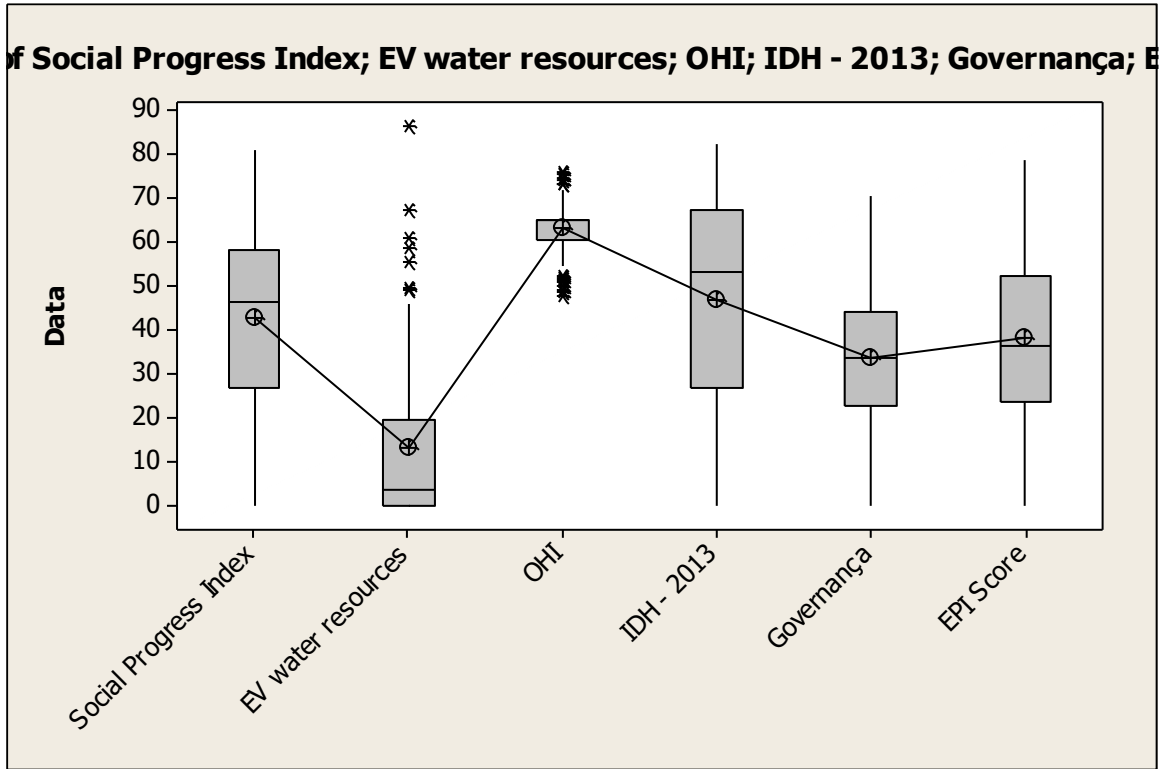
Governança para a região OUTROS

**ANÁLISE**

A região SUSAF apresenta uma média de 29,5827 enquanto que a região EMDEU está com 51,9274.

Por fim, aplicando o One-Way Unstacked para as variáveis sintéticas percebemos que o OHI é o mais descuidado de todos os países.

O valor de F é de 69,46 e o R-quadrado de 41,09.



One-way ANOVA: Social Progr; EV water res; OHI; IDH - 2013; Governança; ...

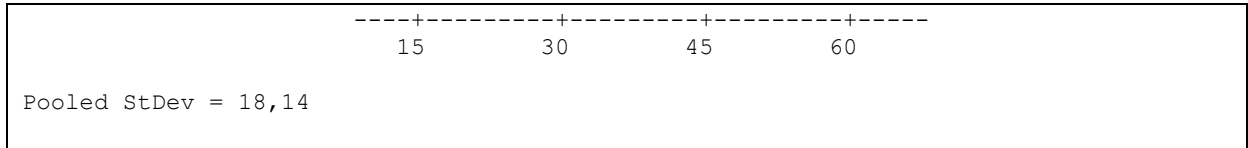
Source	DF	SS	MS	F	P
Factor	5	114320	22864	69,46	0,000
Error	498	163915	329		
Total	503	278234			

S = 18,14 R-Sq = 41,09% R-Sq(adj) = 40,50%

Level	N	Mean	StDev
Social Progress Index	84	42,63	19,71
EV water resources	84	13,10	19,07
OHI	84	63,20	6,22
IDH - 2013	84	46,74	23,52
Governança	84	33,36	16,35
EPI Score	84	37,95	19,06

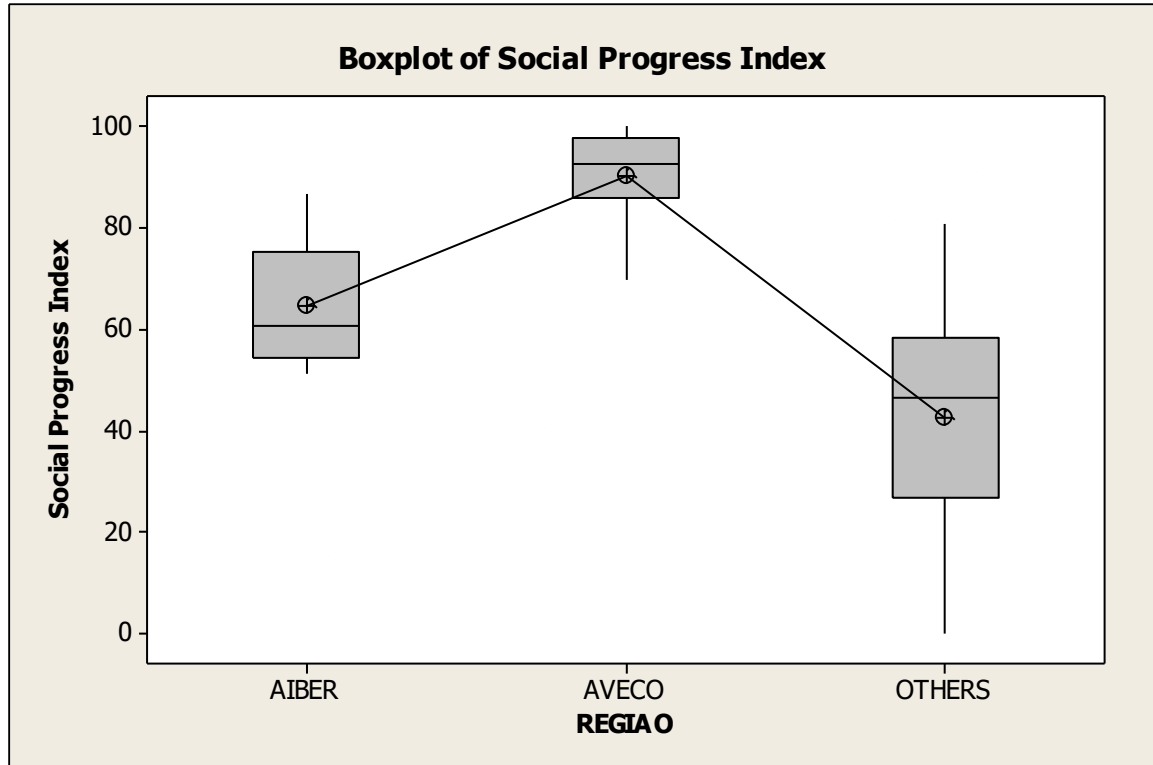
Individual 95% CIs For Mean Based on Pooled StDev

Level	95% CI (Approx.)
Social Progress Index	(---) (-*--)
EV water resources	(---)
OHI	(---) (-*--)
IDH - 2013	(---) (-*--)
Governança	(---) (-*--)
EPI Score	(---) (-*--)



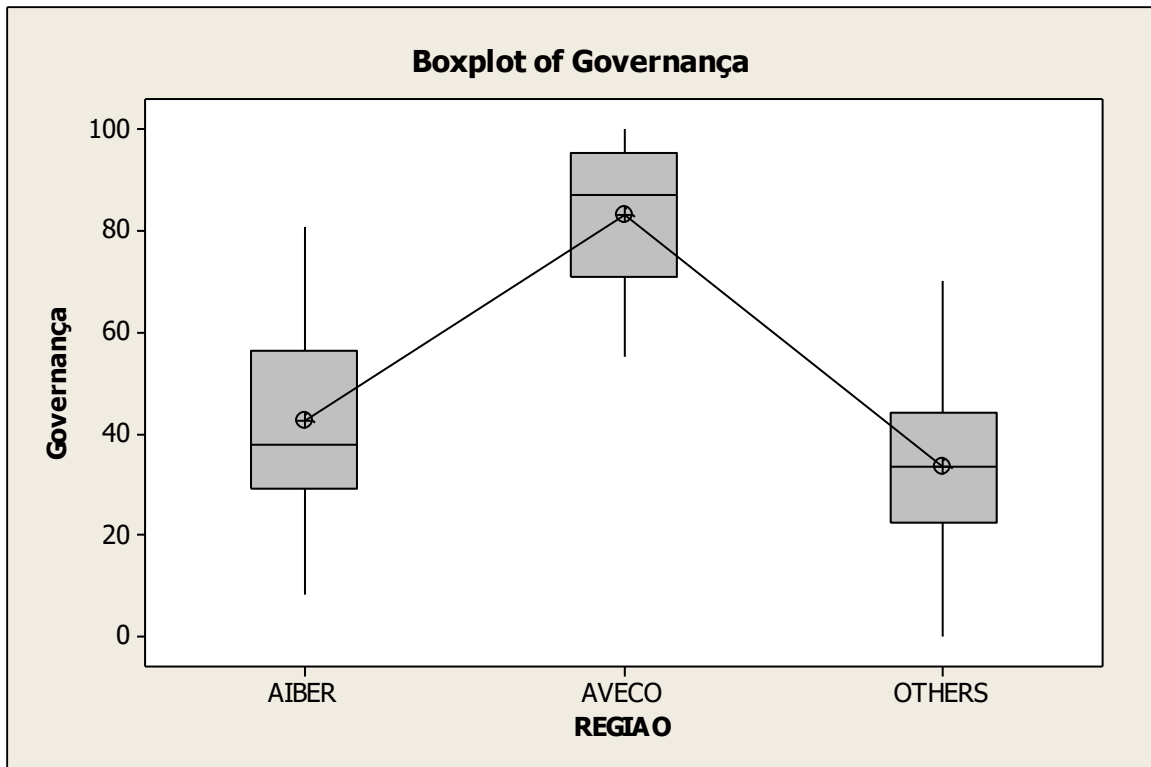
11 – ONE-WAY ANOVA COM AS TRES REGIOES

Primeiro variável Sintética quatro principais: SPI, GOVERNANÇA, IDH e OHI



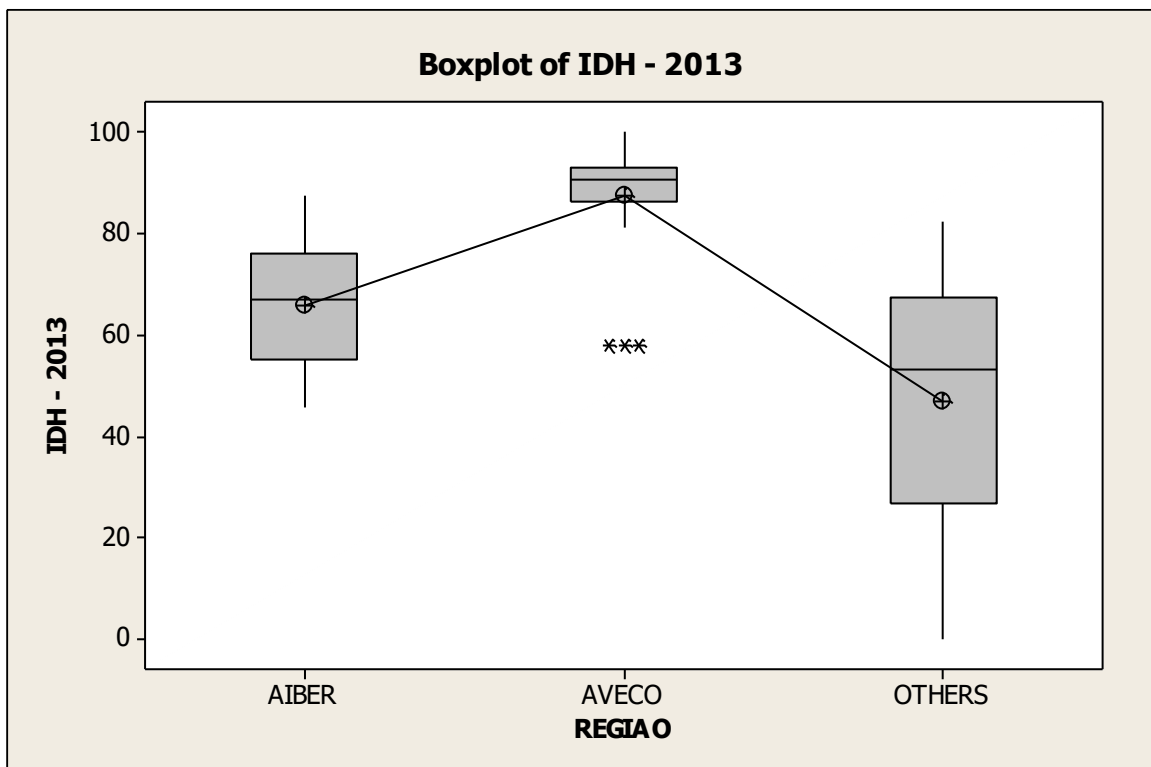
ANÁLISE

A região AIBER apresenta média de 64,4072, AVECO com 90,2994 e OTHERS com 42,6297 que é o pior resultado mas tem dentro os países Zâmbia e Togo com os piores índices.



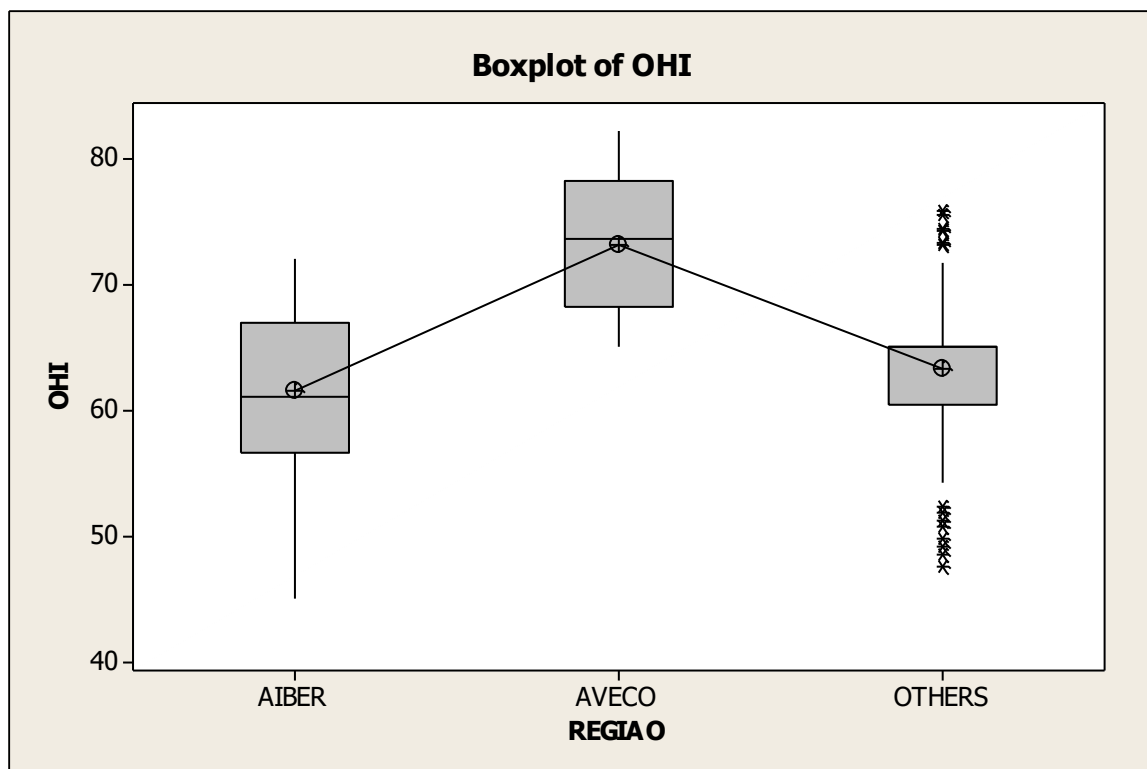
ANÁLISE

A média da AIBER é 42,4840, da AVECO com 83,2528 e OTHERS próximo da AIBER com 33,3617.



ANÁLISE

Apresentou os outliers: Holanda, LAtívia e República da Coreia com 58,056 e a média da região é 87,3586.

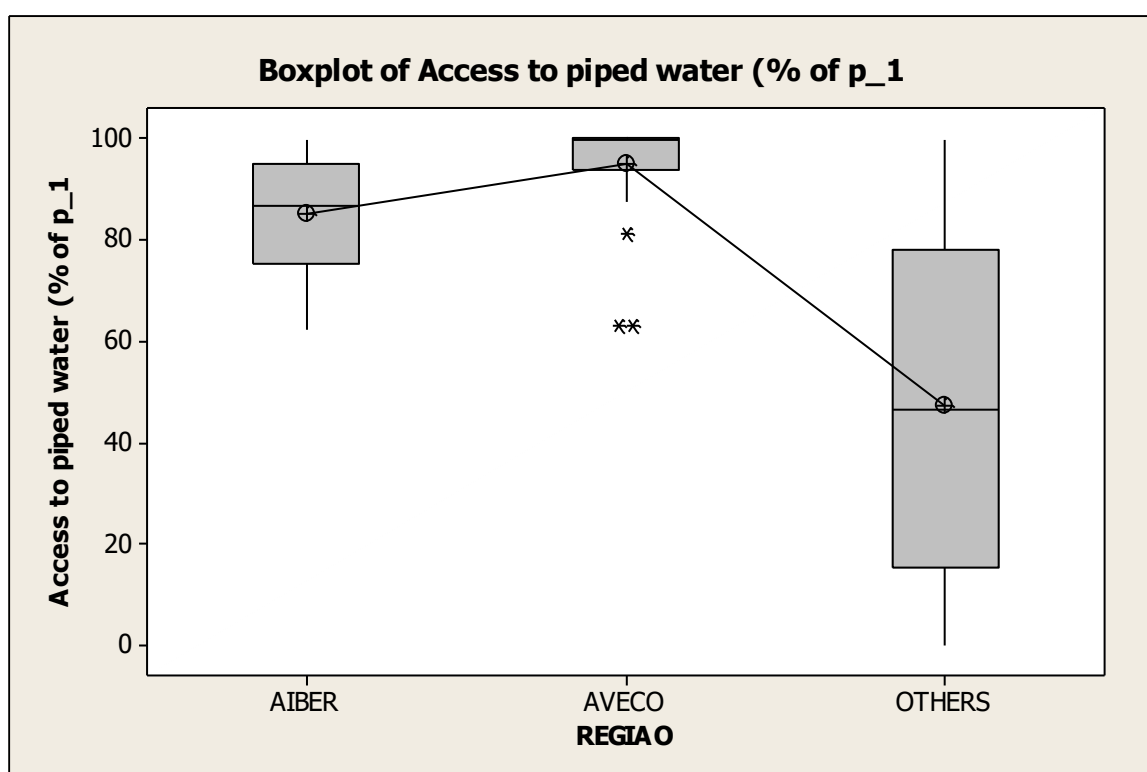


ANÁLISE

AVECO apresenta resultados melhores em relação a saúde dos oceanos com média de 73,1511.

A região OTHERS com outliers por exemplo a Libéria com 47,54 e a média dos outros países é 63,1982.

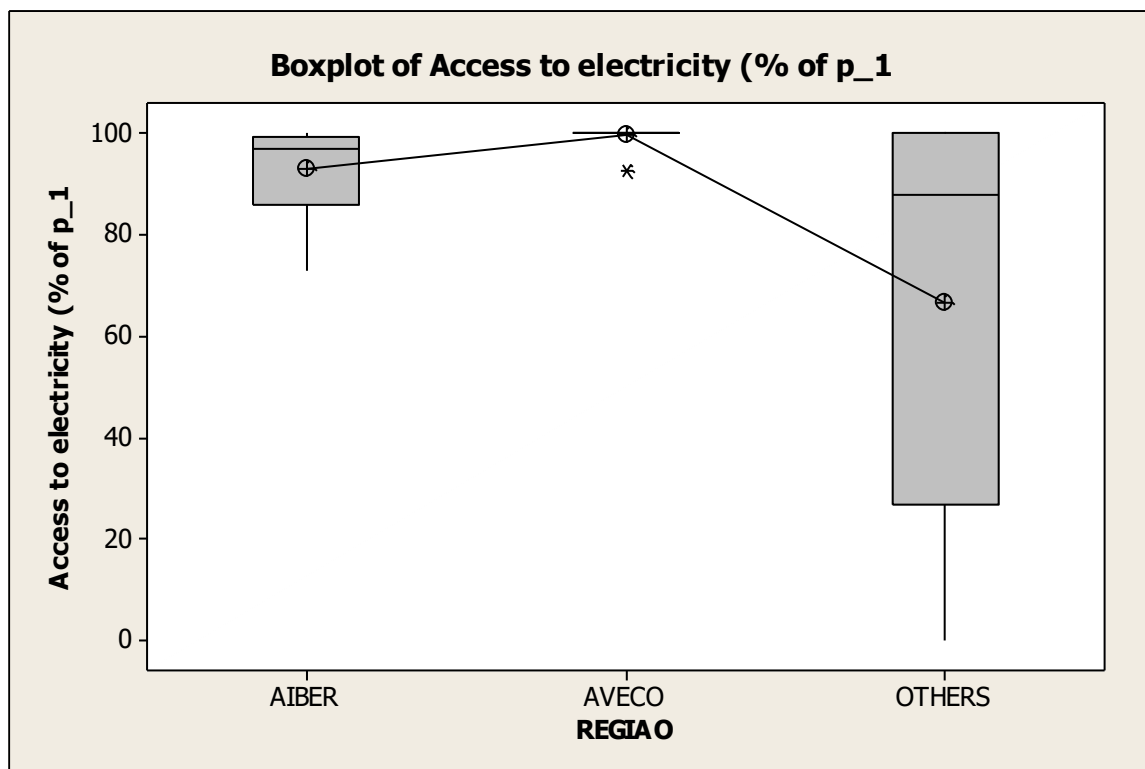
Variáveis Analíticas as principais são: Eletricidade, Acesso a água encanada e



ANÁLISE

AVECO com outliers: EUA e Austrália com 63,04 e Látvia com 81,26 sendo a média 94,8944.

AIBER está com média de 85,2186 não muito distante da AVECO.



ANÁLISE

AVECO com outlier LATívia com 92,871 enquanto que a média do setor é 100.

AIBER está em uma posição também melhor que OTHERS com média de 92,8095.

One-way ANOVA: Access to piped water (% of p_1 versus REGIAO)

Source	DF	SS	MS	F	P
REGIAO	2	58625	29313	37,63	0,000
Error	129	100489	779		
Total	131	159114			

S = 27,91 R-Sq = 36,84% R-Sq(adj) = 35,87%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
AIBER	21	85,22	11,81
AVECO	27	94,89	10,29
OTHERS	84	47,27	33,82

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

60 80 100 120

Pooled StDev = 27,91

One-way ANOVA: Access to electricity (% of p_1 versus REGIAO)

Source	DF	SS	MS	F	P
--------	----	----	----	---	---

```

REGIAO      2   28577  14289  15,48  0,000
Error      129  119046   923
Total      131  147624

```

S = 30,38 R-Sq = 19,36% R-Sq(adj) = 18,11%

```

                                Individual 95% CIs For Mean Based on
                                Pooled StDev
Level      N   Mean  StDev  +-----+-----+-----+-----+
AIBER     21  92,81   8,63  (-----*-----)
AVECO     27  99,73   1,40  (-----*-----)
OTHERS    84  66,42  37,63  (---*---)
                                +-----+-----+-----+
                                60       75       90       105

```

Pooled StDev = 30,38

12 – REGRESSAO LOGISTICA

A regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias^{1 2}. A regressão logística é amplamente usada em ciências médicas e sociais, e tem outras denominações, como modelo logístico e classificador de máxima entropia.

12.1 – Variáveis Sintéticas e as Três Regiões: AIBER, AVECO e OTHERS

Ordinal Logistic Regression: REGIAO versus Social Progr; EV water res; ...

Link Function: Logit

Response Information

Variable	Value	Count
REGIAO	AIBER	21
	AVECO	27
	OTHERS	84
	Total	132

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower	95% CI Upper
Const(1)	-23,8541	5,25080	-4,54	0,000			
Const(2)	-20,9957	5,06716	-4,14	0,000			
Social Progress Index	0,356047	0,0756404	4,71	0,000	1,43	1,23	1,66
EV water resources	0,0198791	0,0163778	1,21	0,225	1,02	0,99	1,05
OHI	0,0326605	0,0504688	0,65	0,518	1,03	0,94	1,14
IDH - 2013	-0,0262217	0,0310118	-0,85	0,398	0,97	0,92	1,04
Governança	-0,175699	0,0446424	-3,94	0,000	0,84	0,77	0,92
EPI Score	-0,0226671	0,0412405	-0,55	0,583	0,98	0,90	1,06
Happy Planet Index	0,0633045	0,0167102	3,79	0,000	1,07	1,03	1,10
GINI Index	0,0895862	0,0189783	4,72	0,000	1,09	1,05	1,14

Log-Likelihood = -54,973

Test that all slopes are zero: G = 128,892, DF = 8, P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	199,275	254	0,995
Deviance	109,946	254	1,000

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	4364	94,9	Somers' D 0,90
Discordant	234	5,1	Goodman-Kruskal Gamma 0,90
Ties	1	0,0	Kendall's Tau-a 0,48
Total	4599	100,0	

ANÁLISE

O P valor de EV, IDH, OHI e EPI estão bem acima dos 0,05 aceitáveis, portanto para análises futuras devem ficar de fora.

12.2 – Variáveis Analíticas

Ordinal Logistic Regression: REGIAO versus Undernourish; Depth of foo; ...

Link Function: Logit

Response Information

Variable	Value	Count
REGIAO	AIBER	21
	AVECO	27
	OTHERS	84
	Total	132

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio
Const(1)	-12,6955	2,98207	-4,26	0,000	
Const(2)	-11,0631	2,92720	-3,78	0,000	
Undernourishment (% of pop.) _1	0,0659933	0,0693570	0,95	0,341	1,07
Depth of food deficit (calori_1	-0,0189020	0,0785167	-0,24	0,810	0,98
Access to piped water (% of p_1	0,107863	0,0243623	4,43	0,000	1,11
Rural vs. urban access to imp_1	0,0496989	0,0216813	2,29	0,022	1,05
Access to improved sanitation_1	-0,0164232	0,0236201	-0,70	0,487	0,98
Access to electricity (% of p_1	0,0279713	0,0263650	1,06	0,289	1,03
Quality of electricity supply_1	-0,0010533	0,0110376	-0,10	0,924	1,00
Renewable internal freshwater_1	0,0101318	0,0184837	0,55	0,584	1,01

Predictor	95% CI	
	Lower	Upper
Const(1)		
Const(2)		
Undernourishment (% of pop.) _1	0,93	1,22
Depth of food deficit (calori_1	0,84	1,14
Access to piped water (% of p_1	1,06	1,17
Rural vs. urban access to imp_1	1,01	1,10
Access to improved sanitation_1	0,94	1,03
Access to electricity (% of p_1	0,98	1,08
Quality of electricity supply_1	0,98	1,02
Renewable internal freshwater_1	0,97	1,05

Log-Likelihood = -83,768

Test that all slopes are zero: G = 71,301, DF = 8, P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	223,622	254	0,916
Deviance	167,537	254	1,000

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	3962	86,1	Somers' D 0,72
Discordant	629	13,7	Goodman-Kruskal Gamma 0,73
Ties	8	0,2	Kendall's Tau-a 0,39

Total	4599	100,0
-------	------	-------

ANÁLISE

Aqui metade das variáveis apresentaram Pvalor acima dos 0,05 aceitáveis.

13 – ARVORE DE DECISAO PARA AS TRES REGIOES – AIBER, AVECO e OTHERS

De acordo com a IBM SPSS Decision Tree a árvore de decisão é uma ferramenta que nos ajuda a identificar grupos, descobrir relacionamentos entre eles e prever eventos futuros. O modelo gráfico gerado se assemelha a um organograma, muito embora seja chamado de árvore de decisão.

A rotina para a geração da Árvore de decisão é:

- 1 – tenha instalado o software SPSS;
- 2 – Acesse seus dados diretamente de uma planilha de Excel e cole na planilha do SPSS;
- 3 – Verifique se as variáveis do seu trabalho estão classificadas como valores numéricos (aparece o desenho de uma régua do lado direito do nome da variável) ou se elas estão classificadas como texto (aparece um símbolo com três balões colados na cor azul, vermelho e verde)
- 4 – Clique em Analyze e desça o cursor até o link Classify e aparecerá do lado outras opções e a quarta está escrito Tree.
- 5 – Clicando em Tree abrirá uma janela com indicando que vai trabalhar com as variáveis clique Ok para ir para a janela com o título Decision Tree onde você deverá escolher a Variável Dependente, em seguida as variáveis independentes e uma terceira que pode ser indicada que a Variável de influência;

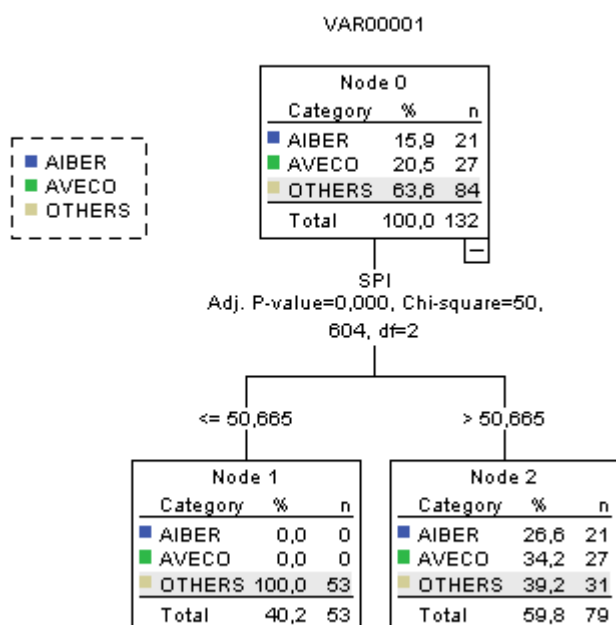
Você pode escolher o processo de crescimento da árvore. São três ou quatro, dependendo da versão do SPSS:

- 1 – Growing Method você pode escolher o CHAID que gera as possibilidades de combinação de forma rápida, explora os dados e constrói os segmentos e perfis com relação ao resultado desejado;
- 2 – Nesse item pode-se escolher o EXHAUSTIVE CHAID ou CHAID EXAUSTIVO que é uma modificação da primeira opção (CHAID) que examina todas as opções possíveis para cada preditor – são as variáveis que antecipam ou dizem algo antes de qualquer outra simulação que seja feita.
- 3 – Nessa caixa existe ainda o CRT - Classification and regression tree – é um algoritmo de árvore binária completo que particiona (divide os dados) os dados e produz subgrupos homogêneos precisos.

Model Summary

Specifications	Growing Method	EXHAUSTIVE CHAID
	Dependent Variable	VAR00001
	Independent Variables	SPI, DEFICTI ALIMENTO, RURAL E URBANA AGUA, INSTALACOES SANITARIAS, ELETRICIDADE, OHI, IDH, GOVERNANÇA, EPI, HPI, GINI, EV, RENOVACAO AGUA, QUALIDADE ELETRICIDADE, AGUA ENCANADA
	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent	100
	Node	

Results	Minimum Cases in Child Node	50
	Independent Variables Included	SPI
	Number of Nodes	3
	Number of Terminal Nodes	2
	Depth	1



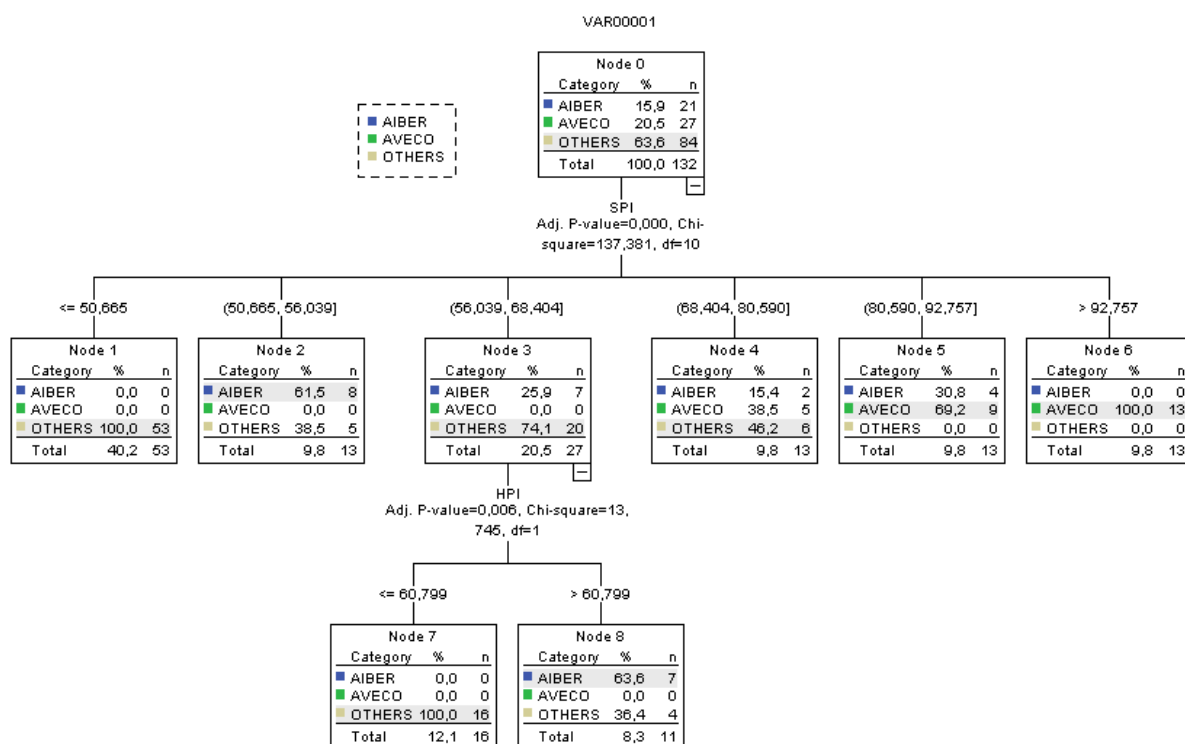
ANÁLISE

Como são poucos os dados para análise o recurso apresentou apenas uma interação com a variável SPI, mesmo tendo sido incluída todas as variáveis tanto sintéticas como as analíticas.

Para gerar diversas interações é preciso aplicar o recurso CRITERIA com valor de 10,5 e 5. O resultado é:

Model Summary		
Specifications	Growing Method	EXHAUSTIVE CHAID
	Dependent Variable	VAR00001

Independent Variables	OHI, IDH, GOVERNANÇA, EPI, HPI, GINI, EV, RENOVACAO AGUA, QUALIDADE ELETRICIDADE, AGUA ENCANADA, DESNUTRICAO, DEFICTI ALIMENTO, RURAL E URBANA AGUA, INSTALACOES SANITARIAS, SPI, ELETRICIDADE
Validation	None
Maximum Tree Depth	5
Minimum Cases in Parent Node	10
Minimum Cases in Child Node	5
Results Independent Variables Included	SPI, HPI
Number of Nodes	9
Number of Terminal Nodes	7
Depth	2



ANÁLISE

O recurso CRITERIA permitiu varias interações entre as variáveis das Três REGIOES. A variável SPI é a principal com Qui-quadrado de 137,381 e com mais seis interações.

Classification

Observed	Predicted			
	AIBER	AVECO	OTHERS	Percent Correct
AIBER	15	4	2	71,4%
AVECO	0	22	5	81,5%
OTHERS	9	0	75	89,3%
Overall Percentage	18,2%	19,7%	62,1%	84,8%

Growing Method: EXHAUSTIVE CHAID

Dependent Variable: VAR00001

ANÁLISE

A participação da região AIBER foi de 71,4%, AVECO 81,5% e OTHERS com 84,8% e o acerto de 84,8%.

Considerações Finais

Este trabalho final analisou dos dados de 3 regiões AIBER, AVECO e OTHERS identificadas a partir de 8 regiões e 132 países que fizeram parte de estudos durante o semestre.

O objetivo era analisar a região AIBER em comparação com AVECO e com os dados da região OTHERS que condensava outras cinco regiões com potencial de desenvolvimento diferente entre si e entre os países que a compõem.

Como considerações finais, os índices apresentados refletem o grau de evolução no quesito humano, principalmente o tema deste trabalho que é: **RECURSOS BASICOS: AGUA, ALIMENTO, ENERGIA (BASIC FEATURES: WATER, FOOD, ENERGY)** para a região AIBER não com indicadores satisfatórios, mas com evolução que é evidente nas informações como SPI – Índice de Progresso Social que é uma variável sintética que incorpora dados de IDH – índice de desenvolvimento humano, saúde, distribuição de renda.

O mesmo foi percebido para a região AVECO, porém esta região possui um diferencial que está nas suas estruturas econômicas e políticas consolidadas em diversos países sendo que os mais representativos são: Holanda, Suécia e Suíça.

Depois de ter feito a apresentação dos dados pela estatística descritiva como média, desvio padrão e variância; por exemplo, é possível deduzir que os países de uma forma geral corroboram o que as análises macroeconômicas de diversos organismos tais como: UNICEF, Banco Mundial, WEF – World Economic Fórum, FAO e outros órgãos preocupados em identificar a evolução dos povos do globo.

Indicadores como o IDH – índice de Desenvolvimento econômico apresentado nesse trabalho com a normalização não muda a situação real da economia analisada. Por exemplo: quando são considerados os dados de países da África Subsaariana, República centro-africana, Lesoto, apresentam resultados que mostram que as variáveis que fazem parte deste indicador como saúde, educação, uso de água potável, instalações sanitárias que podem refletir a qualidade de vida estão baixos, evidenciando as diferenças principalmente quando comparados com outros países desenvolvidos da região da AVECO.

Especificamente a Zâmbia difere dos demais principalmente quando analisamos seus dados de IDH de 1998 para 2013. Esse indicador salta de 0,41 em 1998 para 0,56 em 2013, ou seja, o indicador nos diz que quanto mais próximo de um melhor a qualidade de vida da população.

<http://pt.knoema.com/atlas/Z%C3%A2mbia/%C3%8Dndice-de-Desenvolvimento-Humano>

Para melhor entendimento dos dados estatísticos sugerimos que o pesquisador examine a definição do conceito na estatística para aplicar ao fato analisado, como é o caso da mediana que divide a metade inferior da metade superior. Como temos 132 países, a mediana é calculada pela média dos dois centrais depois de coladas em ordem crescente os valores. Por exemplo: se pegarmos a variável IDH (poderia ser qualquer uma) e colocarmos em ordem crescente os valores dos 132 países, a mediana estará entre os números da posição 66 e 67 na linha do excel. Logo, se somarmos os valores e dividirmos por 2 temos a mediana. Isto que o MINITAB fez e encontrou valores como 56,185 para o IPS, para o Déficit de Calorias 5,995 dentre outros.

Analisando a mediana do SPI 56,185 mostra que existe metade com a média de 55,84 e a outra metade abaixo da média. O mesmo vale para o Déficit de Alimentos com 5,995 mostrando que existe metade acima da média de 12,42 e a outra metade abaixo da média.

Importante ressaltar nessas considerações finais que a análise dos dados da Região AVECO para o indicador analítico Desnutrição teve que ser suprimida porque esta região apresentava dado observado igual a Zero, ou seja, é uma região que está muito em vantagem em termos de nutrição da população. De qualquer maneira, como o valor da média é baixo a parte que fica acima da média também é baixo porque a mediana também indica valor proporcionalmente baixo.

A análise de componentes principais permite que sejam feitas as comparações entre as variáveis mais correlacionadas geradas pelo MINITAB e entre elas e as novas variáveis. Com a junção das três regiões os dados ficaram um pouco mais robusto com tendência a proximidade dos dados apresentados para a região OUTROS. Note que o Eigenvalue está com 9,2544 e a PC1 explica 57,8 % da variável SPI, ou seja, como são 16 variáveis (8 sintética e 8 componentes) a divisão de 9,2544 por 16 variáveis. Como a amostra é maior os dados ficam mais representativos, logo, o fato de ter mais de 50% de explicação da variável SPI pelo PC1 tem valor no contexto.

Por fim, a Árvore de Decisão é uma ferramenta que nos ajuda a identificar grupos, descobrir relacionamentos entre eles e predizer eventos futuros. O modelo gráfico gerado se assemelha a um organograma, muito embora seja chamado de árvore de decisão.

Quando foi solicitado ao SPSS gerar uma as relações entre a região e as variáveis sintéticas e analíticas, o resultado foi muito pequeno em termo de nós. Foram gerados apenas dois nós. Isto aconteceu pela escolha do processo denominado CHAID que gera as possibilidades de combinação de forma rápida, explora os dados e constrói os segmentos e perfis com relação ao resultado desejado.

Com a solicitação do recurso CRITERIA depois da indicação de EXHAUSTIVE CHAID ou CHAID EXAUSTIVO que é uma modificação da primeira opção (CHAID) que examina todas as opções possíveis para cada preditor – são as variáveis que antecipam ou dizem algo antes de qualquer outra simulação que seja feita, a Árvore gerada ficou mais robusta com o SPI encabeçando a primeira linha e mais seis nós logo abaixo. Em seguida na terceira linha a variável sintética encabeçava mais dois nós com as relações das três regiões do estudo: AIBER, AVECO e OTHERS.

Em termos gerais a evolução dos países da região AVECO tem sido constante embora os saltos de qualidade sejam ainda pequenos em função da velocidade política com que acontecem, do crescimento populacional, da escolaridade, dos investimentos e da escolaridade que interfere em todo o processo.

REFERENCIAS

<http://pt.knoema.com/atlas/>

Programa Países Sustentáveis – Guia GPS – Gestão Pública Sustentável uma publicação do Núcleo de Estudos Futuros da PUC-SP –Pontifícia Universidade Católica de São Paulo.

SANTOS, Joao Almeida; PARRA FILHO, Domingos. Metodologia Científica. São Paulo: Cengage Learning, 2011. 3 edição.

