



## **ÍNDICE DE DESENVOLVIMENTO HUMANO**

**Uma análise comparativa sócio-econômico entre os cinco continentes**

**PAULA AUGUSTA RODRIGUES COELHO**

**RODERICK CABRAL CASTELLO BRANCO**

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO**

**FEA - Faculdade de Economia e Administração**

**Programa de Estudos Pós-Graduados em Administração**

## 1. INTRODUÇÃO

O trabalho seguinte tratará de números que desenham o desenvolvimento nos países do mundo. Desenvolvimento em um conceito mais amplo, imprescindível para o mundo contemporâneo, fugindo do reducionismo do crescimento econômico como ator principal na redução da pobreza e desigualdade mundiais.

Nunca houve tanta riqueza no mundo e, ao mesmo tempo, tanta pobreza. Há quarenta anos, mesmo com o inquestionável crescimento econômico mundial, as desigualdades explodiram: o abismo entre os 20% mais pobres e os 20% mais ricos era de 1 para 30 em 1960, e é hoje de 1 para 80 (HARRIBEY, 2007).

Dito isso, entende-se impossível a construção de um espelho da situação mundial baseada somente no crescimento do Produto Interno Bruto (PIB). Assim, o Programa das Nações Unidas para o Desenvolvimento (PNUD), lançou o Índice de Desenvolvimento Humano (IDH), visando evitar o uso exclusivo da opulência econômica como critério de aferição, criando uma clara distinção entre desenvolvimento e crescimento. A partir de então, passou-se considerar as influências sociais, culturais e políticas como determinantes na qualidade da vida humana.

Até a década de 1960 não se enxergava a necessidade de outro medidor fora o PIB, pois eram poucas as nações desenvolvidas que haviam se tornado ricas pela industrialização (VEIGA, 2005). Porém, com a intensa industrialização, na década de 50, diversos países semi-industrializados (dentre eles o Brasil) passaram por um período de intenso crescimento econômico que não se traduziram em maior acesso das populações pobres a bens materiais ou culturais, a exemplo dos países desenvolvidos, provando a incapacidade de se pensar o futuro fora do paradigma do crescimento econômico permanente.

O estudo tem sua importância no sentido de ajudar na melhor compreensão do fenômeno do desenvolvimento sócio-econômico mundial, principalmente quando da recente colocação do Brasil no grupo de países com altos índices de desenvolvimento humano, apesar das gritantes desigualdades sociais ainda existentes. Para isso, este estudo tem como intuito realizar análise do comportamento de importantes variáveis econômico-sociais fornecidas pelo PNUD (dentre elas, o IDH), englobando todos os 177 países do globo.

## 2. ENTENDENDO OS DADOS

### 2.1 Os Indivíduos

Os indivíduos desta análise são os 177 países listados pelo PNUD, Programa das Nações Unidas para o Desenvolvimento, no Relatório de Desenvolvimento Humano 2007/2008.

### 2.2 As Variáveis

Foram consideradas sete variáveis para análise.

**Tabela 1.** As Variáveis

Variável	Significado	Tipo	Unidade de Medida
País	Nome do país	Variável Categórica	N/A
IDH	Índice de Desenvolvimento Humano	Variável Quantitativa	Índice
Saúde	Gastos com saúde per capita	Variável Quantitativa	US\$ 1,00
Educação	Percentual sobre o PIB gasto com educação	Variável Quantitativa	Percentual
Usuários de Internet	Percentual de usuários de internet no país	Variável Quantitativa	Unidade
Desenvolvimento humano	Grupo de desenvolvimento humano ao qual o país está inserido. Pode ser de alto desenvolvimento humano, médio e baixo.	Variável Categórica	N/A
Continente	É o continente em que o país se encontra.	Variável Categórica	N/A

### 2.3 A Tabela de Dados

Países	IDH	Gastos c/ saúde per capita-US\$	Gastos com educação (% do PIB)	Usuários de Internet (%)	Desenvolvimento humano	Continente
Iceland	0,968	3294	8,1	86,9	Alto	Europe
Norway	0,968	4080	7,7	73,5	Alto	Europe
Australia	0,962	3123	4,7	69,8	Alto	Oceania
Canada	0,961	3173	5,2	52	Alto	North and Central America
Ireland	0,959	2618	4,8	27,6	Alto	Europe
Sweden	0,956	2828	7,4	76,4	Alto	Europe
Switzerland	0,955	4011	6	49,8	Alto	Europe
Japan	0,953	2296	3,6	66,8	Alto	Asia
Netherlands	0,953	3092	5,4	73,9	Alto	Europe
France	0,952	3040	5,9	43	Alto	Europe
Finland	0,952	2203	6,5	53,4	Alto	Europe
United States	0,951	6096	5,9	63	Alto	North and Central

						America
Spain	0,949	2099	4,3	34,8	Alto	Europe
Denmark	0,949	2780	8,5	52,7	Alto	Europe
Austria	0,948	3418	5,5	49,6	Alto	Europe
United Kingdom	0,946	2560	5,4	47,3	Alto	Europe
Belgium	0,946	3133	6,1	45,8	Alto	Europe
Luxembourg	0,944	5178	3,6	69	Alto	Europe
New Zealand	0,943	2081	6,5	67,2	Alto	Oceania
Italy	0,941	2414	4,7	47,8	Alto	Europe
Hong Kong, China (SAR)	0,937	*	4,2	50,8	Alto	Asia
Germany	0,935	3171	4,6	45,5	Alto	Europe
Israel	0,932	1972	6,9	47	Alto	Asia
Greece	0,926	2179	4,3	18	Alto	Europe
Singapore	0,922	1118	3,7	57,1	Alto	Asia
Korea (Republic of)	0,921	1135	4,6	68,4	Alto	Asia
Slovenia	0,917	1815	6	54,5	Alto	Europe
Cyprus	0,903	1128	6,3	43	Alto	Europe
Portugal	0,897	1897	5,7	27,9	Alto	Europe
Brunei Darussalam	0,894	621	*	27,7	Alto	Asia
Barbados	0,892	1151	5,9	59,4	Alto	North and Central America
Czech Republic	0,891	1412	4,4	26,9	Alto	Europe
Kuwait	0,891	538	5,1	27,6	Alto	Asia
Malta	0,878	1733	4,5	31,5	Alto	Europe
Qatar	0,875	688	1,6	26,9	Alto	Asia
Hungary	0,874	1308	5,5	29,7	Alto	Europe
Poland	0,87	814	5,4	26,2	Alto	Europe
Argentina	0,869	1274	3,8	17,7	Alto	South America
United Arab Emirates	0,868	503	1,3	30,8	Alto	Asia
Chile	0,867	720	3,5	17,2	Alto	South America
Bahrain	0,866	871	*	21,3	Alto	Asia
Slovakia	0,863	1061	4,3	46,4	Alto	Europe
Lithuania	0,862	843	5,2	35,8	Alto	Europe
Estonia	0,861	752	5,3	51,3	Alto	Europe
Latvia	0,855	852	5,3	44,8	Alto	Europe
Uruguay	0,852	784	2,6	19,3	Alto	South America
Croatia	0,85	917	4,7	32,7	Alto	Europe
Costa Rica	0,846	592	4,9	25,4	Alto	North and Central America
Bahamas	0,845	1349	3,6	31,9	Alto	North and Central America

Seychelles	0,843	634	5,4	24,9	Alto	Africa
Cuba	0,838	229	9,8	1,7	Alto	North and Central America
Mexico	0,829	255	5,4	18,1	Alto	North and Central America
Bulgaria	0,824	671	4,2	20,6	Alto	Europe
Saint Kitts and Nevis	0,821	710	9,3	*	Alto	North and Central America
Tonga	0,819	316	4,8	2,9	Alto	Oceania
Libyan Arab Jamahiriya	0,818	328	2,7	3,6	Alto	Africa
Antigua and Barbuda	0,815	516	3,8	35	Alto	North and Central America
Oman	0,814	419	3,6	11,1	Alto	Asia
Trinidad and Tobago	0,814	523	4,2	12,3	Alto	North and Central America
Romania	0,813	433	3,4	20,8	Alto	Europe
Saudi Arabia	0,812	601	6,8	7	Alto	Asia
Panama	0,812	632	3,8	6,4	Alto	North and Central America
Malaysia	0,811	402	6,2	43,5	Alto	Asia
Belarus	0,804	427	6	34,7	Alto	Europe
Mauritius	0,804	516	4,5	14,6	Alto	Africa
Bosnia and Herzegovina	0,802	603	*	20,6	Alto	Europe
Russian Federation	0,802	583	3,6	15,2	Alto	Europe
Albania	0,801	339	2,9	6	Alto	Europe
Macedonia (TFYR)	0,801	471	3,5	7,9	Alto	Europe
Brazil	0,8	1520	4,4	19,5	Alto	South America
Dominica	0,798	309	5	36,1	Médio	North and Central America
Saint Lucia	0,795	302	5,8	33,9	Médio	North and Central America
Kazakhstan	0,794	264	2,3	2,7	Médio	Asia
Venezuela (Bolivarian Republic of)	0,792	285	*	12,5	Médio	South America
Colombia	0,791	570	4,8	10,4	Médio	South America
Ukraine	0,788	427	6,4	9,7	Médio	Europe
Samoa	0,785	218	4,5	3,2	Médio	Oceania
Thailand	0,781	293	4,2	11	Médio	Asia
Dominican Republic	0,779	377	1,8	16,9	Médio	North and Central America

Belize	0,778	339	5,4	13	Médio	North and Central America
China	0,777	277	1,9	8,5	Médio	Asia
Grenada	0,777	480	5,2	18,2	Médio	North and Central America
Armenia	0,775	226	3,2	5,3	Médio	Asia
Turkey	0,775	557	3,7	22,2	Médio	Asia
Suriname	0,774	376	*	7,1	Médio	South America
Jordan	0,773	502	4,9	11,8	Médio	Asia
Peru	0,773	235	2,4	16,4	Médio	South America
Lebanon	0,772	817	2,6	19,6	Médio	Asia
Ecuador	0,772	261	1	4,7	Médio	South America
Philippines	0,771	203	2,7	5,4	Médio	Asia
Tunisia	0,766	502	7,3	9,5	Médio	Africa
Fiji	0,762	284	6,4	7,7	Médio	Oceania
Saint Vincent and the Grenadines	0,761	418	8,2	8,4	Médio	North and Central America
Iran (Islamic Republic of)	0,759	604	4,7	10,3	Médio	Asia
Paraguay	0,755	327	4,3	3,4	Médio	South America
Georgia	0,754	171	2,9	3,9	Médio	Asia
Guyana	0,75	329	8,5	21,3	Médio	South America
Azerbaijan	0,746	138	2,5	8,1	Médio	Asia
Sri Lanka	0,743	163	*	1,4	Médio	Asia
Maldives	0,741	494	7,1	5,9	Médio	Asia
Jamaica	0,736	223	5,3	40,4	Médio	North and Central America
Cape Verde	0,736	225	6,6	4,9	Médio	Africa
El Salvador	0,735	375	2,8	9,3	Médio	North and Central America
Algeria	0,733	167	*	5,8	Médio	Africa
Vietnam	0,733	184	*	12,9	Médio	Asia
Occupied Palestinian Territories	0,731	*	*	6,7	Médio	Asia
Indonesia	0,728	118	0,9	7,3	Médio	Asia
Syrian Arab Republic	0,724	109	*	5,8	Médio	Asia
Turkmenistan	0,713	245	*	0,8	Médio	Asia
Nicaragua	0,71	231	3,1	2,7	Médio	North and Central America
Moldova	0,708	138	4,3	9,6	Médio	Europe
Egypt	0,708	258	*	6,8	Médio	Africa
Uzbekistan	0,702	160	*	3,4	Médio	Asia

Mongolia	0,7	141	5,3	10,5	Médio	Asia
Honduras	0,7	197	*	3,6	Médio	North and Central America
Kyrgyzstan	0,696	102	4,4	5,4	Médio	Asia
Bolivia	0,695	186	6,4	5,2	Médio	South America
Guatemala	0,689	256	*	7,9	Médio	North and Central America
Gabon	0,677	264	3,9	4,8	Médio	Africa
Vanuatu	0,674	123	9,6	3,8	Médio	Oceania
South Africa	0,674	748	5,4	10,9	Médio	Africa
Tajikistan	0,673	54	3,5	0,1	Médio	Asia
Sao Tome and Principe	0,654	141	*	13,1	Médio	Africa
Botswana	0,654	504	10,7	3,4	Médio	Africa
Namibia	0,65	407	6,9	3,7	Médio	Africa
Morocco	0,646	234	6,7	15,2	Médio	Africa
Equatorial Guinea	0,642	223	0,6	1,4	Médio	Africa
India	0,619	91	3,8	5,5	Médio	Asia
Solomon Islands	0,602	114	3,3	0,8	Médio	Oceania
Lao People's Democratic Republic	0,601	74	2,3	0,4	Médio	Asia
Cambodia	0,598	140	1,9	0,3	Médio	Asia
Myanmar	0,583	38	1,3	0,2	Médio	Asia
Bhutan	0,579	93	5,6	3,9	Médio	Asia
Comoros	0,561	25	3,9	3,3	Médio	Africa
Ghana	0,553	95	5,4	1,8	Médio	Africa
Pakistan	0,551	48	2,3	6,7	Médio	Asia
Mauritania	0,55	43	2,3	0,7	Médio	Africa
Lesotho	0,549	139	13,4	2,4	Médio	Africa
Congo	0,548	30	2,2	1,3	Médio	Africa
Bangladesh	0,547	64	2,5	0,3	Médio	Asia
Swaziland	0,547	367	6,2	3,2	Médio	Africa
Nepal	0,534	71	3,4	0,4	Médio	Asia
Madagascar	0,533	29	3,2	0,5	Médio	Africa
Cameroon	0,532	83	1,8	1,5	Médio	Africa
Papua New Guinea	0,53	147	*	2,3	Médio	Oceania
Haiti	0,529	82	*	7	Médio	North and Central America
Sudan	0,526	54	*	7,7	Médio	Africa
Kenya	0,521	86	6,7	3,2	Médio	Africa
Djibouti	0,516	87	7,9	1,3	Médio	Africa
Timor-Leste	0,514	143	*	*	Médio	Asia
Zimbabwe	0,513	139	4,6	7,7	Médio	Africa
Togo	0,512	63	2,6	4,9	Médio	Africa
Yemen	0,508	82	9,6	0,9	Médio	Asia
Uganda	0,505	135	5,2	1,7	Médio	Africa

Gambia	0,502	88	2	3,3	Médio	Africa
Senegal	0,499	72	5,4	4,6	Baixo	Africa
Eritrea	0,483	27	5,4	1,6	Baixo	Africa
Nigeria	0,47	53	*	3,8	Baixo	Africa
Tanzania (United Republic of)	0,467	29	2,2	0,9	Baixo	Africa
Guinea	0,456	96	2	0,5	Baixo	Africa
Rwanda	0,452	126	3,8	0,6	Baixo	Africa
Angola	0,446	38	2,6	1,1	Baixo	Africa
Benin	0,437	40	3,5	5	Baixo	Africa
Malawi	0,437	58	5,8	0,4	Baixo	Africa
Zambia	0,434	63	2	2	Baixo	Africa
Côte d'Ivoire	0,432	64	4,6	1,1	Baixo	Africa
Burundi	0,413	16	5,1	0,5	Baixo	Africa
Congo (Democratic Republic of the)	0,411	15	*	0,2	Baixo	Africa
Ethiopia	0,406	21	6,1	0,2	Baixo	Africa
Chad	0,388	42	2,1	0,4	Baixo	Africa
Central African Republic	0,384	54	*	0,3	Baixo	Africa
Mozambique	0,384	42	3,7	0,7	Baixo	Africa
Mali	0,38	54	4,3	0,4	Baixo	Africa
Niger	0,374	26	2,3	0,2	Baixo	Africa
Guinea- Bissau	0,374	28	5,2	2	Baixo	Africa
Burkina Faso	0,37	77	4,7	0,5	Baixo	Africa
Sierra Leone	0,336	34	4,6	0,2	Baixo	Africa

Observação: os dados não disponíveis foram substituídos por \*

### 3. ANÁLISE DAS VARIÁVEIS

A análise individual das variáveis faz-se indispensável para melhor visualização e entendimento do objeto de estudo.

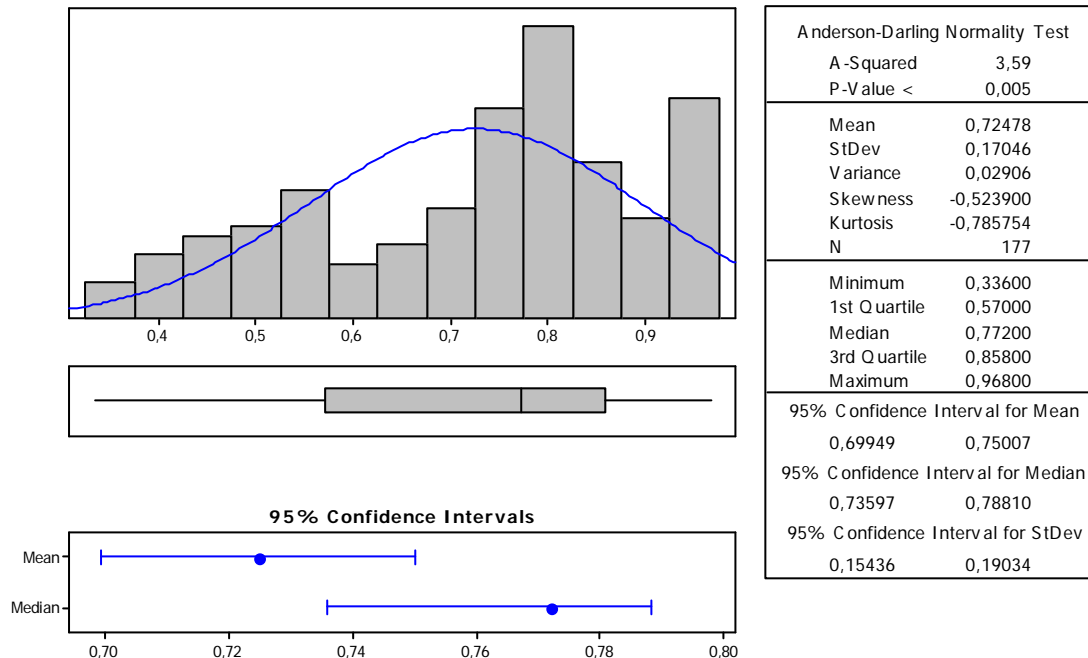
#### 3.1 Variável IDH: Índice de Desenvolvimento Humano

O IDH foi criado para medir o nível de desenvolvimento humano dos países a partir de indicadores de educação (alfabetização e taxa de matrícula), longevidade (expectativa de vida ao nascer) e renda (PIB per capita). Seus valores variam de 0 (nenhum desenvolvimento humano) a 1 (desenvolvimento humano total). Países com IDH até 0,499 são considerados de desenvolvimento humano baixo; com índices entre 0,500 e 0,799 são considerados de desenvolvimento humano médio; e com índices maiores que 0,800 são considerados de desenvolvimento humano alto.

Segue abaixo quadro contendo Histograma, Curva de Densidade, Box-Plot, Intervalo de confiança da média e mediana, além das medidas numéricas como média, desvio-padrão,

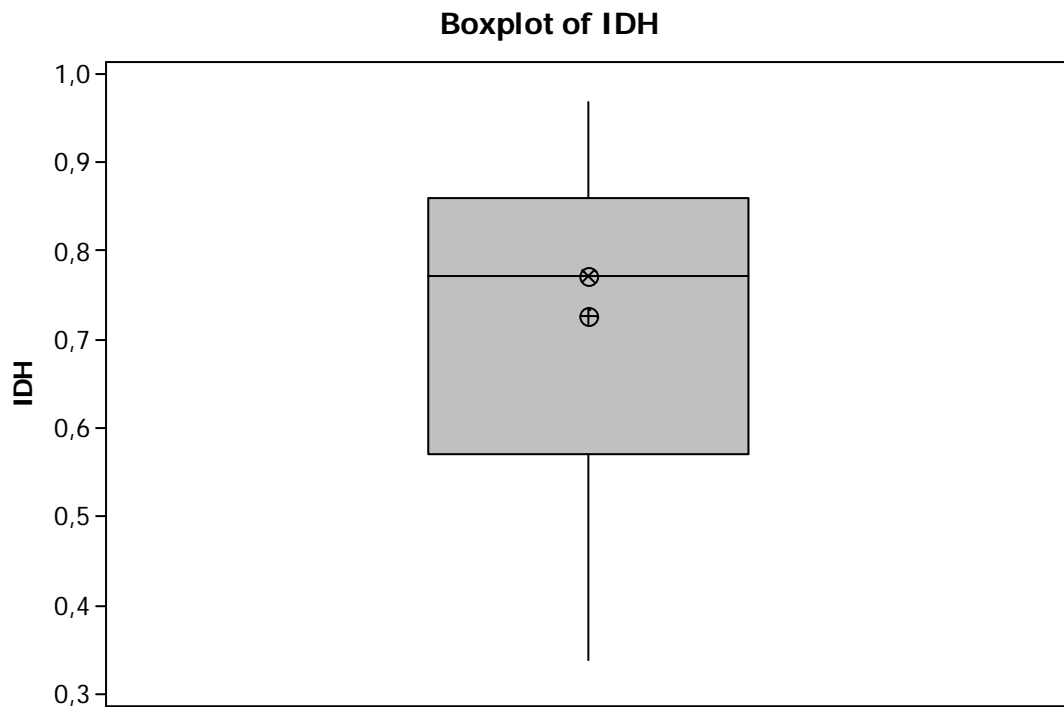
variância, quantidade de observações, valores mínimos, máximos, informações dos quartis e o teste de normalidade de Anderson-Darling (A-Squared e P-Value), para a variável RLV.

### Summary for IDH

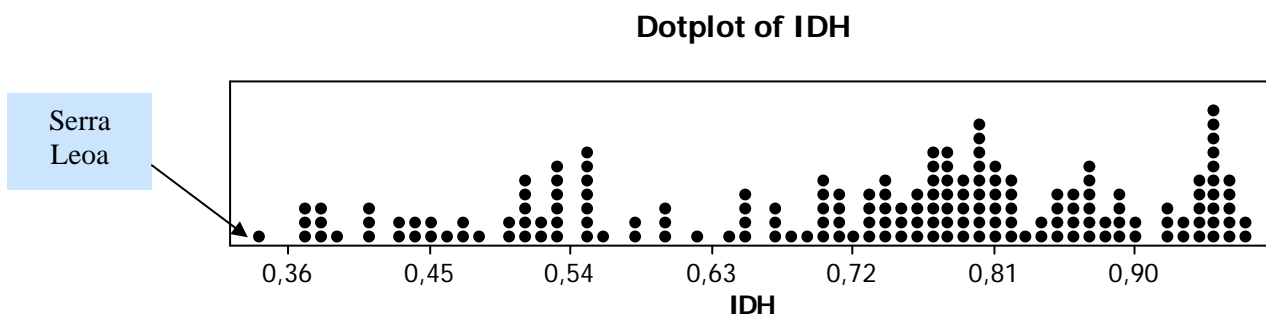


As principais observações que podemos fazer são:

- **Forma:** Visualmente, observam-se picos em diversas áreas do histograma. Nota-se o maior deles em IDH de 0,8, afirmando que há uma grande quantidade de países em situação próxima à do Brasil (IDH de exatamente 0,8), valor limítrofe entre os países de alto desenvolvimento humano e os de médio desenvolvimento humano. Outra observação é a grande quantidade de países com IDH entre o nível médio e alto, aproximadamente 80% dos países.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos países possui IDH abaixo de 0,77 e outra metade acima deste valor. Já o IDH médio é de 0,72, levando ao entendimento de que, tanto média quanto mediana representam a situação de países de médio desenvolvimento humano. O IDH mínimo é de 0,33 (Serra Leoa) e o máximo de 0,96 (Islândia), apresentando desvio padrão de 0,17. O boxplot abaixo demonstra com maior clareza a posição da média, representada pelo símbolo de uma cruz inscrita em um círculo, e a mediana, representada pelo X inscrito em um círculo.



Abaixo o dotplot facilita a visualização da situação do IDH no mundo.

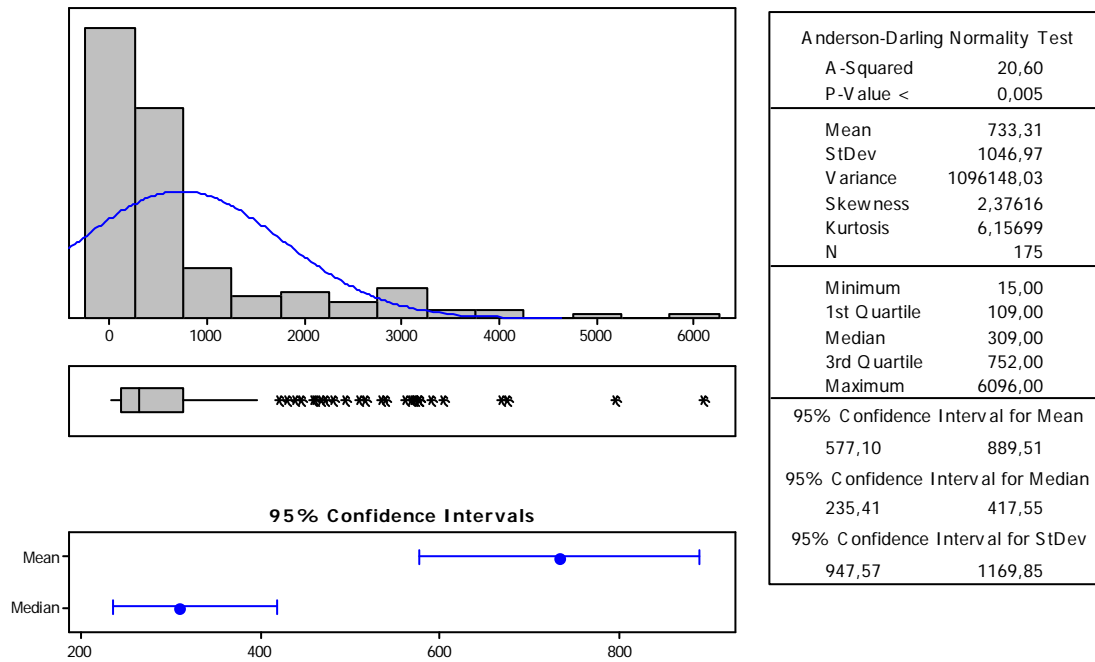


O pico observado no histograma acima está disperso entre os valores de IDH 0,72 e 0,81, mostrando a grande divergência de valores entre esses países instalados no limiar entre os países de médio e alto desenvolvimento humano. Porém, o último pico à direita apresenta grande similaridade e repetição de valores, significando uma quantidade significativa de países de com índices de alto desenvolvimento humano em situação muito parecida.

### 3.2 Variável Gastos com Saúde per capita

A variável diz respeito ao valor gasto em determinado país com saúde por pessoa no ano de 2004. Segue abaixo quadro contendo Histograma, Curva de Densidade, Box-Plot, Intervalo de confiança da média e mediana, além das medidas numéricas como média, desvio-padrão, variância, quantidade de observações, valores mínimos, máximos, informações dos quartis e A-Squared e P-Value para a variável Saúde per Capita.

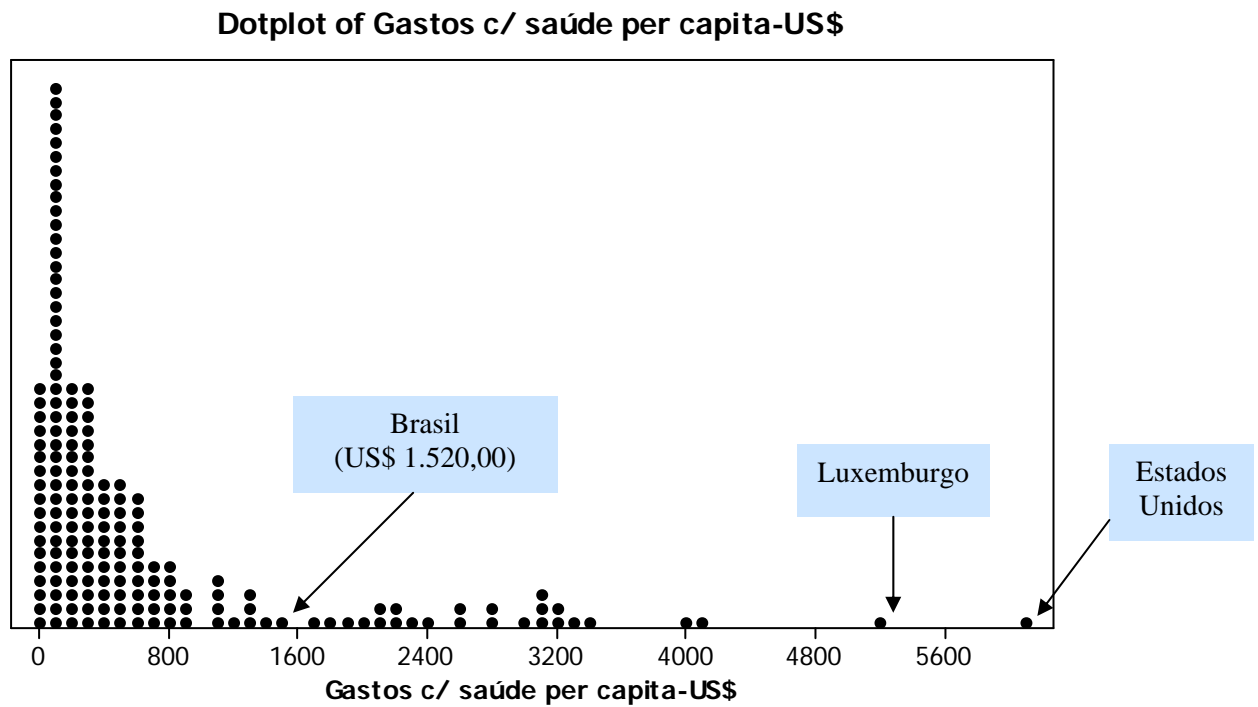
### Summary for Gastos c/ saúde per capita-US\$



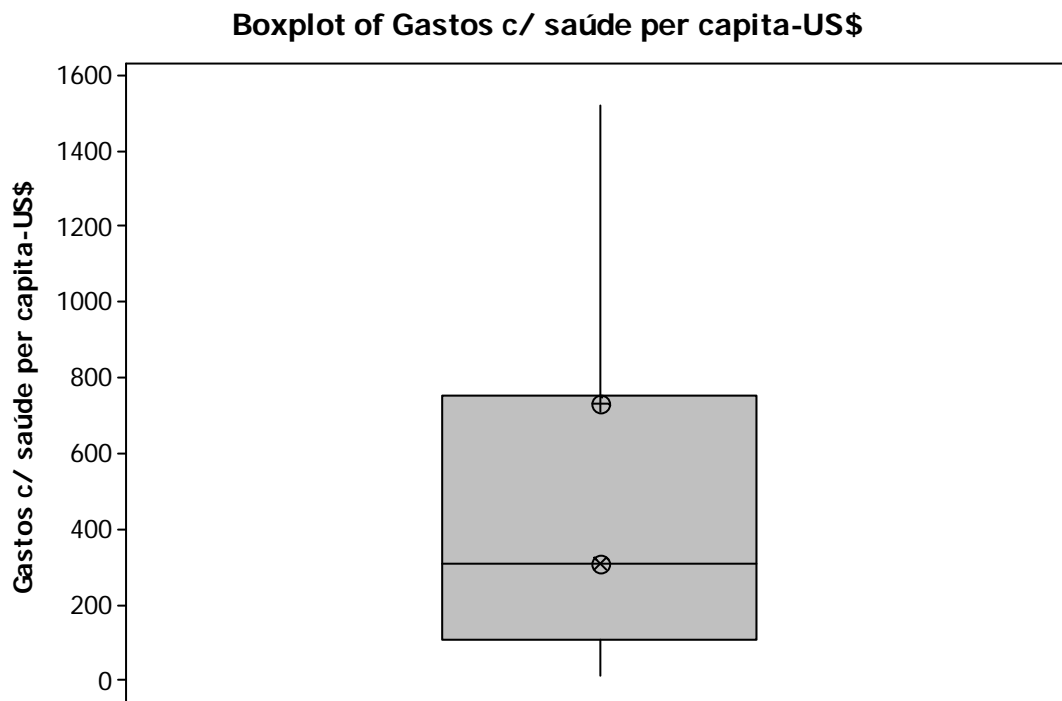
Com base nas informações acima, fazemos os seguintes comentários:

- **Forma:** É grande a concentração à esquerda do gráfico, acusando que a imensa quantidade de países com gastos com saúde por habitante próximo a 0 e menores que mil dólares per capita naquele ano. Do outro lado, uma pequena quantidade de países se dispersa apresentando gastos entre US\$ 2 mil e até US\$ 6 mil per capita/ano com saúde para sua população.
- **Valores Atípicos:** O boxplot apresenta vários valores atípicos em função da extrema desigualdade entre os países analisados.
- **Centro e Dispersão:** A mediana acusa que metade dos países gasta, com cada um de seus habitantes, menos de US\$ 309,00 ao ano, enquanto a outra metade dos países gasta acima deste valor. Porém, quando considerada a média gasta no mundo com saúde, este número sobre para US\$ 733,31, em função dos altos valores gasto em outros países. O desvio padrão é de US\$ 1.046,97, sendo que o valor mínimo gasto é de US\$ 15,00 (República Democrática do Congo) e o máximo é de US\$ 6.096,00 (Estados Unidos), despesa 400 vezes maior.

O gráfico dotplot abaixo confirma com maior nível de detalhes o afirmado no histograma acima: existe uma grande concentração de países com gastos inferiores à US\$ 800 per capita ao ano, enquanto poucos investindo em saúde para sua população.



No caso da exclusão dos outliers, ou seja, dos países com muitos recursos destinados à saúde em relação ao total de sua população, o resultado seria o observado no boxplot abaixo.

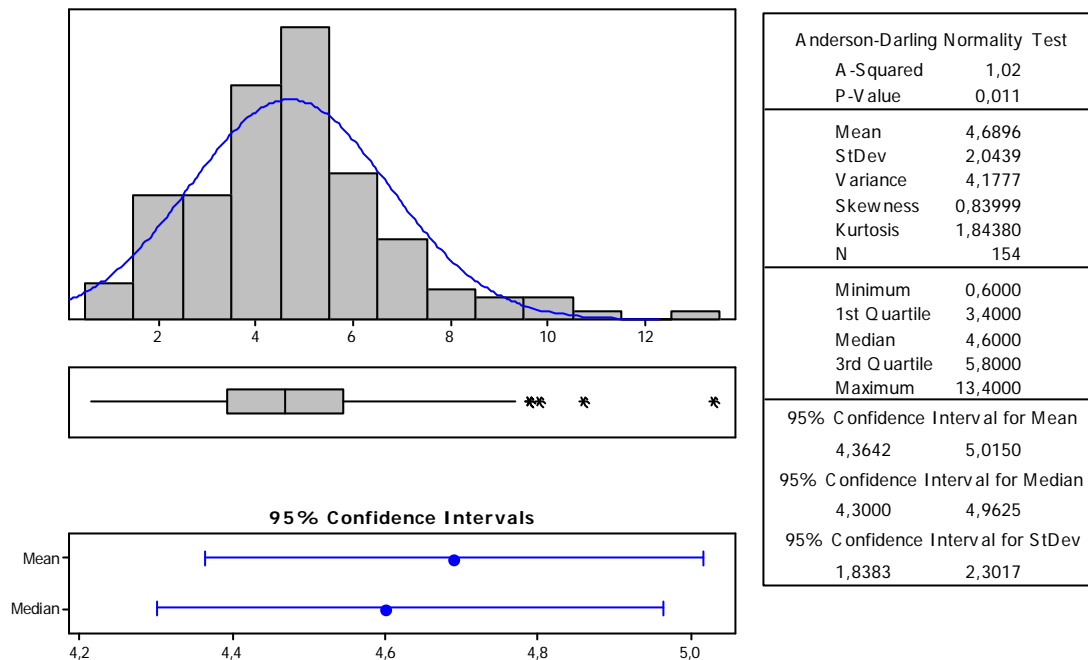


### 3.3 Gastos com Educação: % do Produto Interno Bruto

A variável analisada nos gráficos abaixo é referente ao total gasto por cada país em educação em relação ao seu produto interno bruto. Considerando a importância da educação e do

conhecimento como forças transformadoras da realidade, optamos por essa variável fornecida pelo Relatório Anual de Desenvolvimento para análise.

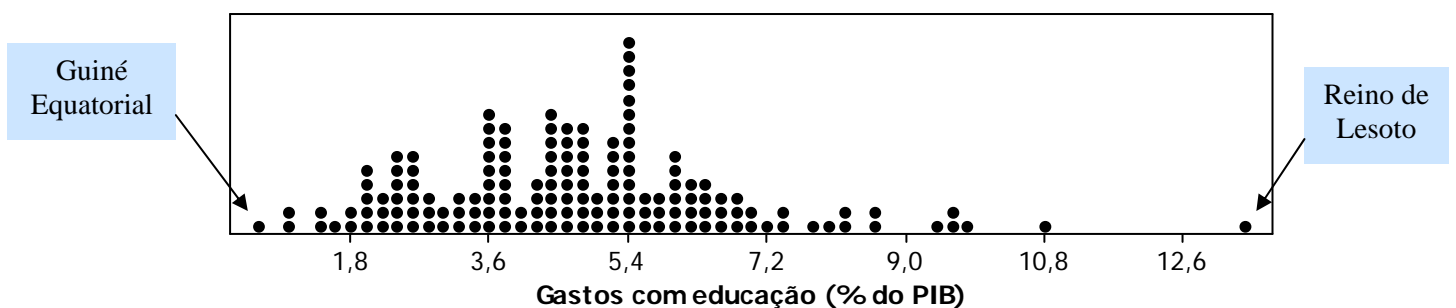
### Summary for Gastos com educação (% do PIB)



Analisando as informações acima, concluímos:

- **Forma:** O histograma apresenta uma concentração no centro do gráfico, levando à dedução da existência de muitos países realizando investimentos nessa área em torno de 5% do seu PIB. Os percentuais mais altos, à direita do gráfico, são representados por uma quantidade bem menor de países.
- **Valores Atípicos:** Foram observados valores atípicos, como confirmado pelo Dot-plot abaixo.

### Dotplot of Gastos com educação (% do PIB)



O país cujo investimento em educação representa a maior fatia do PIB é o Reino de Lesoto, um pequeno país da África Austral, a parte sul do continente, possuidor de IDH igual a 0,54, sendo considerado um país de médio desenvolvimento humano, porém dentre os mais baixos do grupo. Os investimentos em educação chegaram a 13,4% do PIB entre

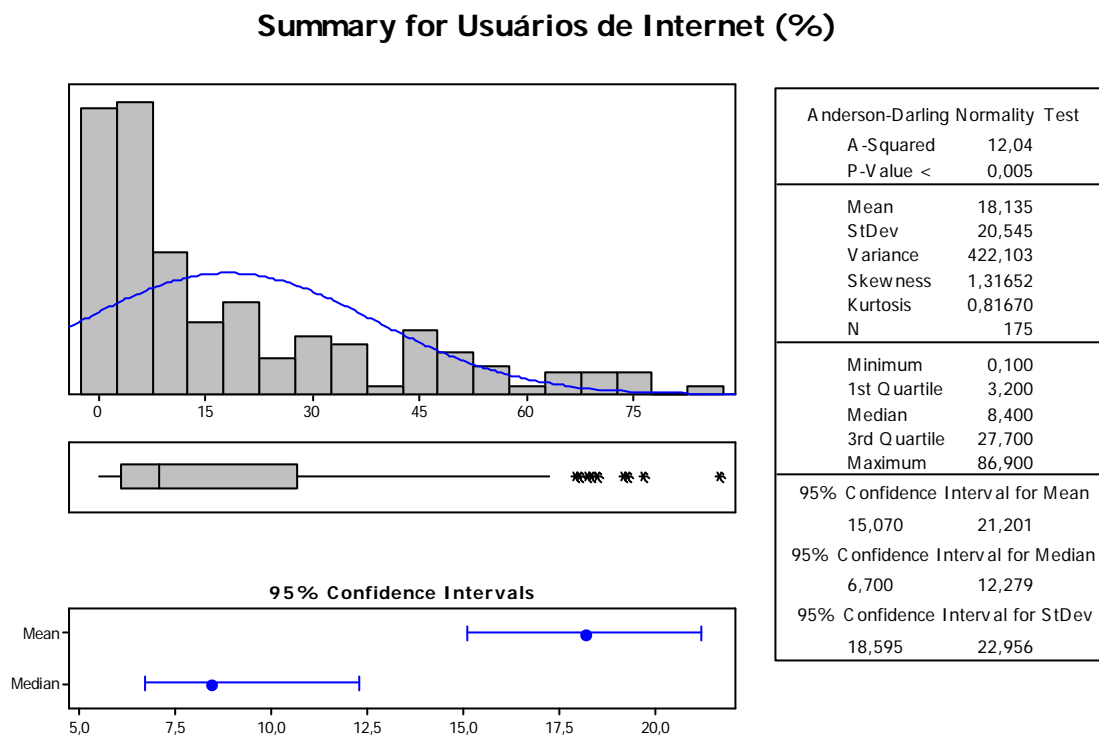
2002 e 2005. O segundo maior investimento também é de um país da África, Botswana, vizinho do Reino de Lesoto: 10,7% do PIB.

O Brasil investe em torno de 4,4% do PIB em educação, enquanto Islândia investe 8,1%, Estados Unidos, 5,9% e Alemanha, 4,6%. O país com menor índice de investimento em educação em relação ao seu PIB é a Guiné Equatorial, com 0,6%.

- **Centro e Dispersão:** A mediana nos indica que metade dos países investe em educação abaixo de 4,6% do seu PIB, enquanto outra metade investe acima deste valor. A média mundial não se diferencia muito da mediana, pois seu valor é de 4,68%. O menor valor gasto é de 0,6% e o maior de 13,4% - ambos de países africanos.

### 3.4 Usuários de Internet: % da população

A quantidade de usuários de internet nos países também foi considerada importante por medir o grau de conectividade do país e a intimidade de sua população com a informação, comunicação com o mundo e tecnologias. Os valores apresentados são referentes ao percentual da população que utiliza a internet, sendo analisados 175 países.

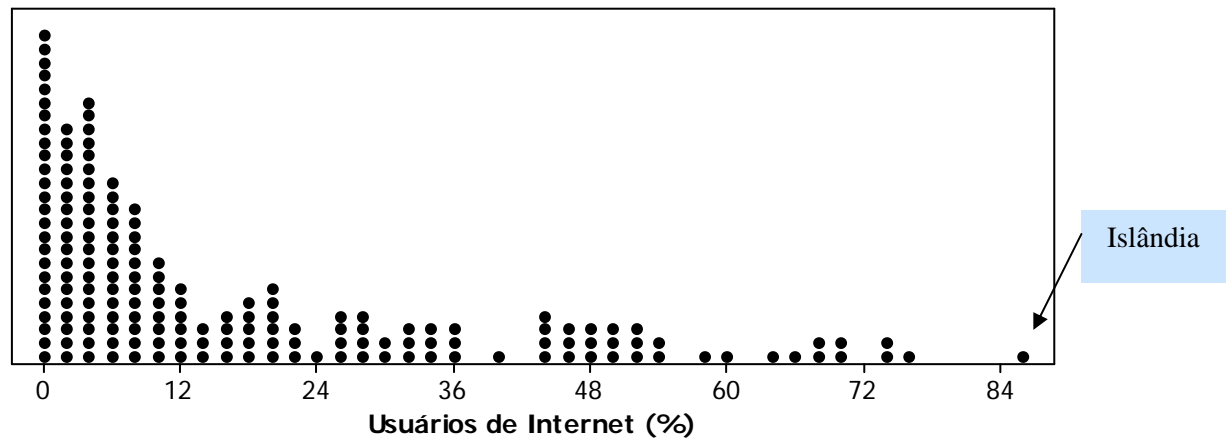


Com base nas informações acima, fazemos os seguintes comentários:

- **Forma:** A exemplo do ocorrido no histograma de gastos com saúde, a concentração de países está à esquerda, povoando os mais baixos índices de usuários de internet (entre 0 e 30% do total de habitantes). À direita os valores ficam dispersos, podendo chegar até 86% de conectividade.

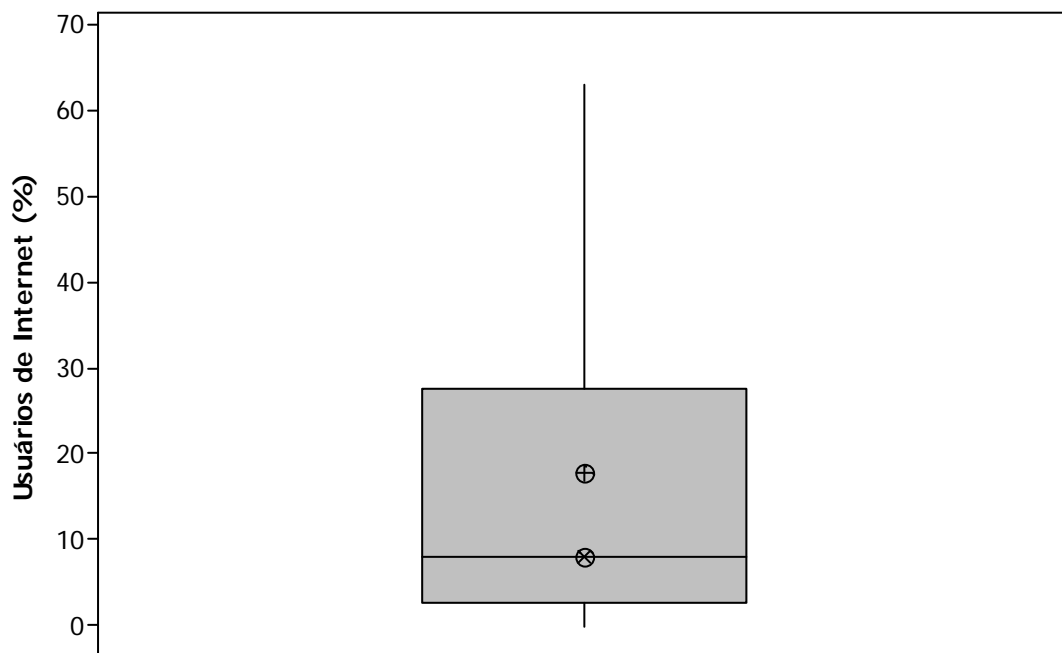
O dotplot abaixo facilita a visualização da distribuição desses índices entre os países.

Dotplot of Usuários de Internet (%)



- Valores Atípicos: No caso de exclusão dos valores atípicos, teríamos a seguinte visualização:

Boxplot of Usuários de Internet (%)

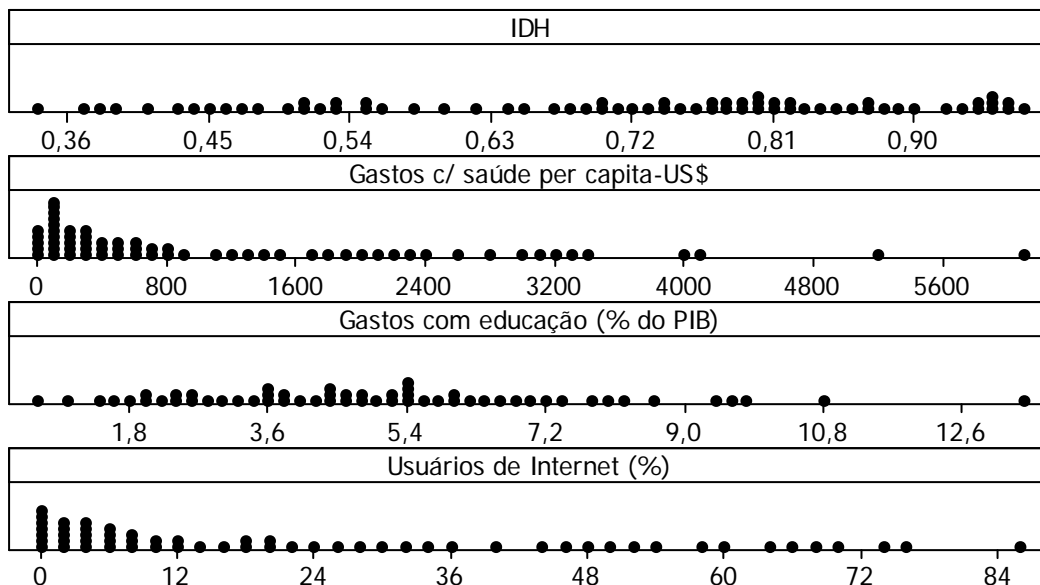


Confirmando a afirmação acima, ao se excluírem os outliers, os índices de conectividade não chegam a 30%.

- Centro e Dispersão: A mediana indica que metade dos países possui mais de 8,1% de sua população conectada à internet e a outra metade possui menos desse valor. Quando considerada a média, em função dos altos níveis de conectividade dos países desenvolvidos, o percentual sobe para 17,9%.

A seguir, comparações entre as variáveis para melhor visualização dos diferentes tipos de comportamento adotados por elas individualmente.

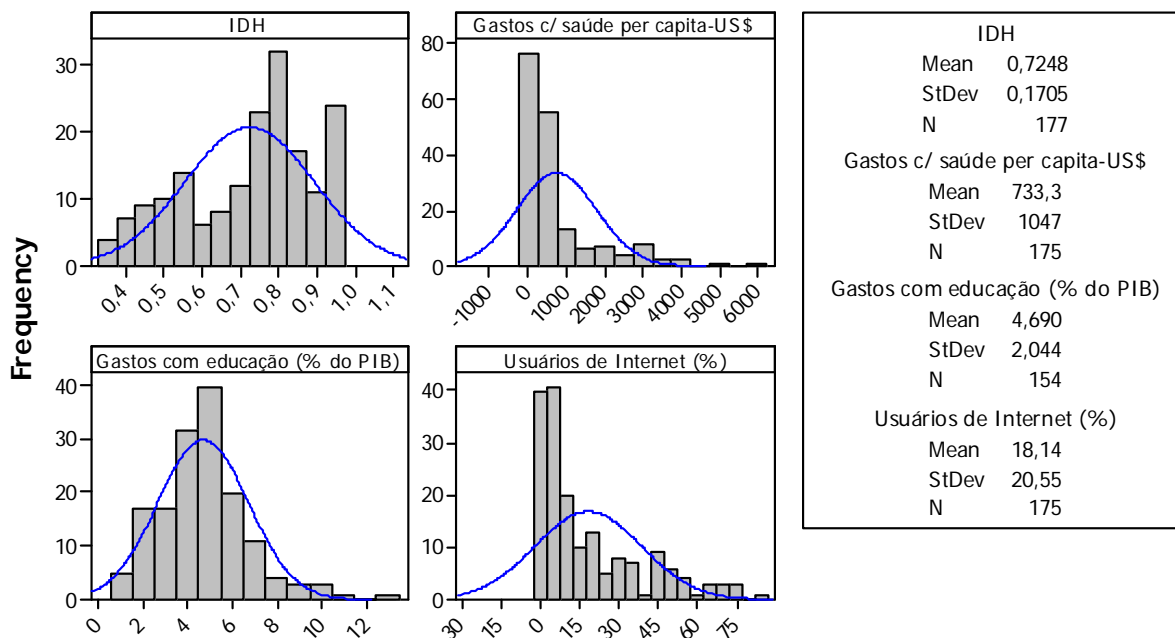
### Dotplot of IDH; Gastos c/ saúde ; Gastos com educa; Usuários de Inte



Each symbol represents up to 4 observations.

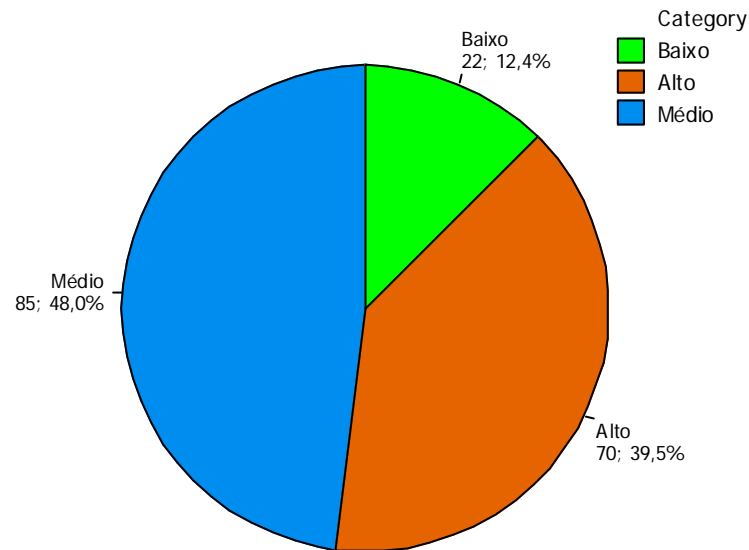
### Histogram of IDH; Gastos c/ saúde ; Gastos com educa; Usuários de Inte

Normal

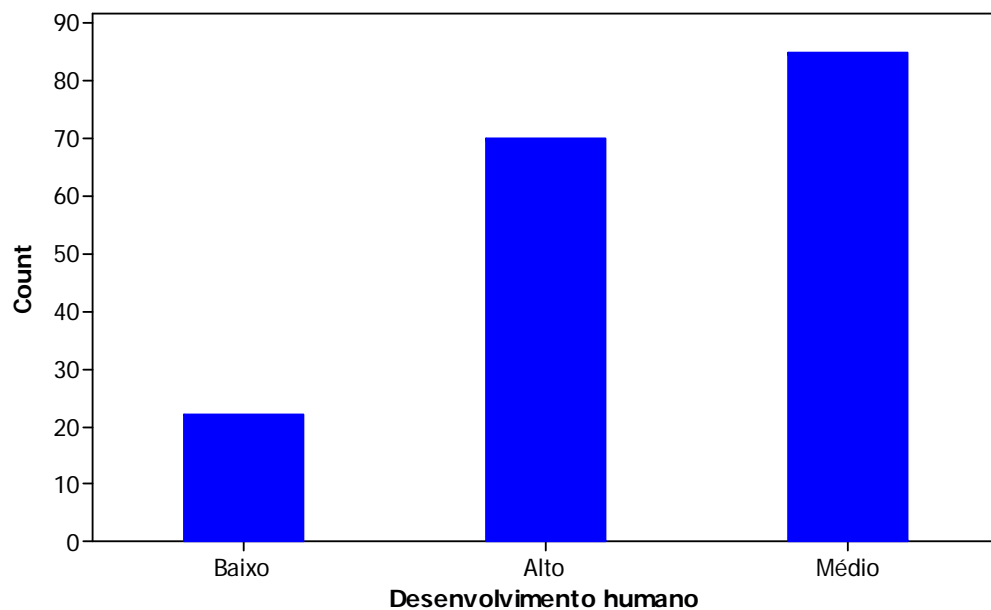


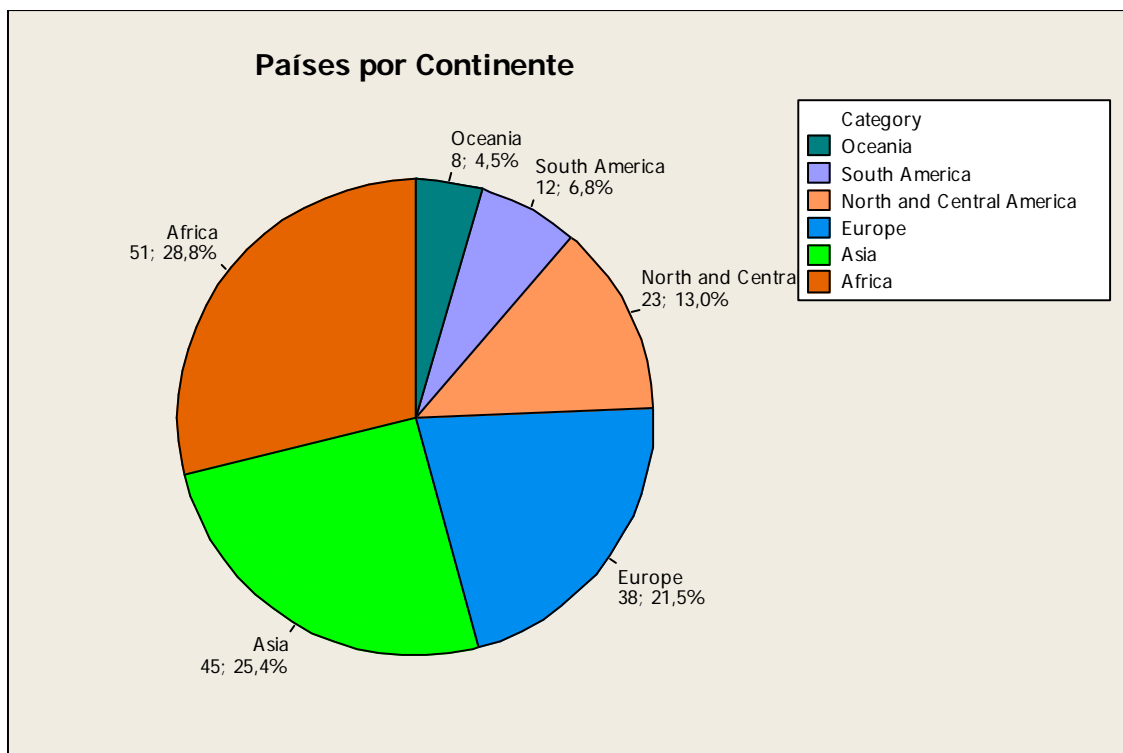
### 3.4 Variáveis Desenvolvimento Humano e Continente

Desenvolvimento Humano e Continente são as únicas variáveis categóricas a serem analisadas neste estudo. A primeira trata-se da classificação dos países em três níveis de desenvolvimento humano: alto, médio e baixo desenvolvimento. A segunda trata do continente que cada país se encontra. A seguir, alguns gráficos explicativos.

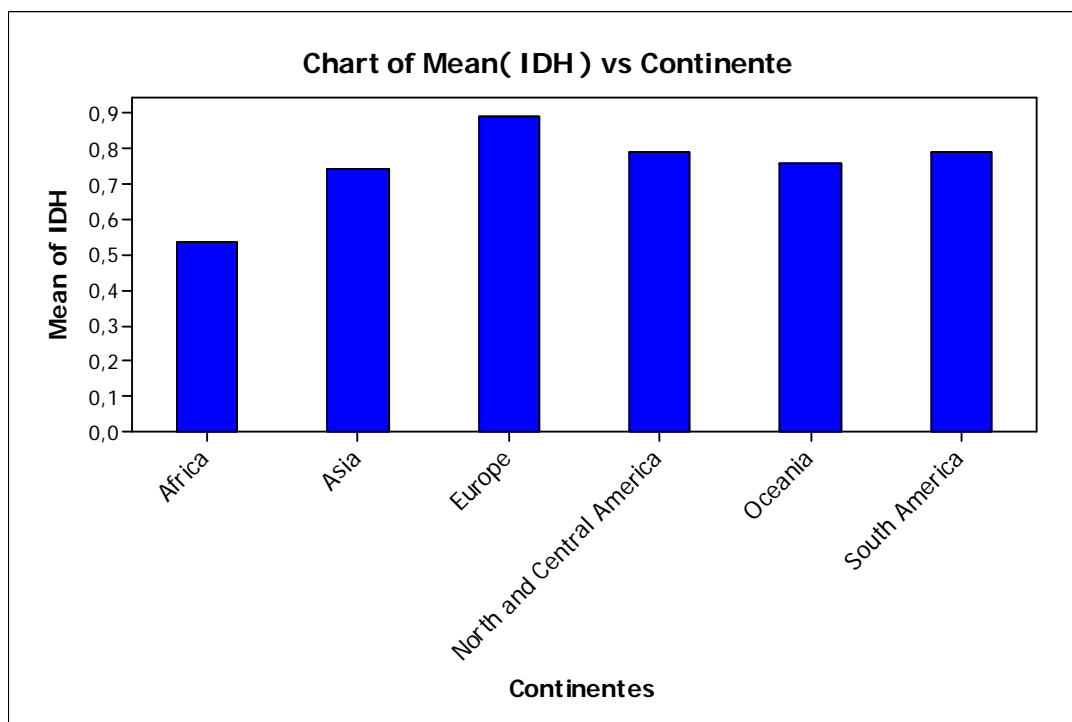
**Pie Chart of Desenvolvimento humano**

Acima observa-se a maior participação de países tidos como possuidores de médio desenvolvimento humano, com 85 ocorrências, o que significa 48% do total de países. Em segundo lugar, com 39,5% das ocorrências (70, no total), estão os de alto desenvolvimento humano. Em terceiro lugar, 22 países encontram-se atualmente com níveis de desenvolvimento humano baixo.

**Chart of Desenvolvimento humano**



Observa-se no gráfico acima que a maior concentração dos países está no continente africano com um total de 51 países, 28,8% do total. Seguido pelo continente asiático com 45 países, 25,4%; continente europeu com 38 países, 21,5%; continente norte e centro americano com 23 países, 13%; sul americano com 12 países, 6,8% e por último Oceania com 8 países, que representam apenas 4,5% do total de países analisados.

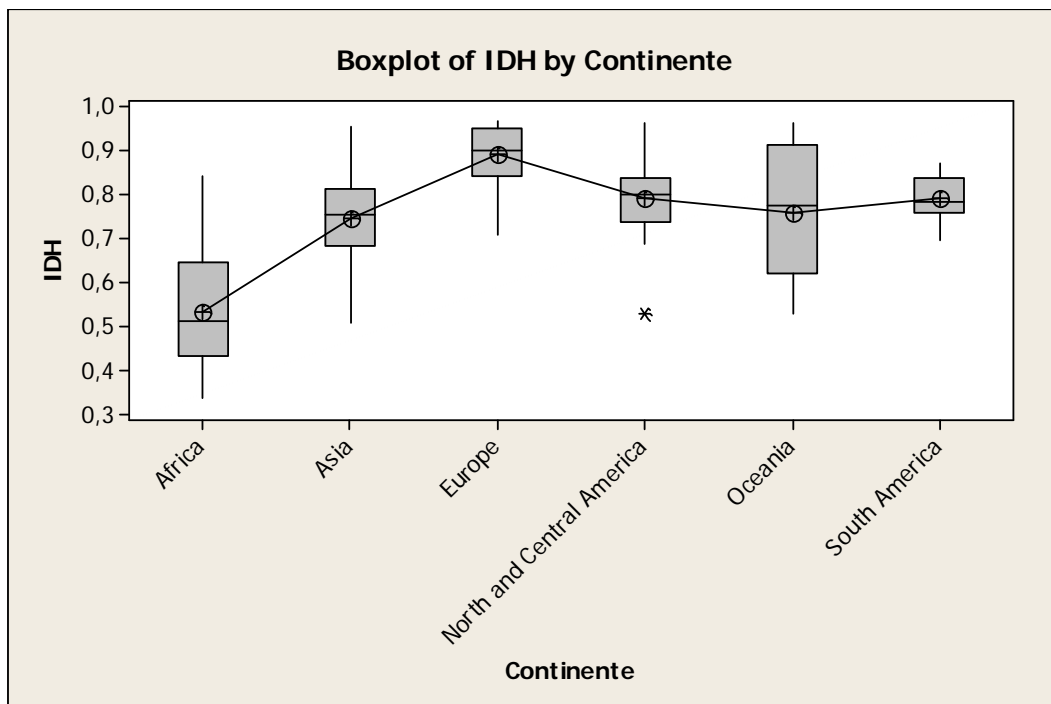
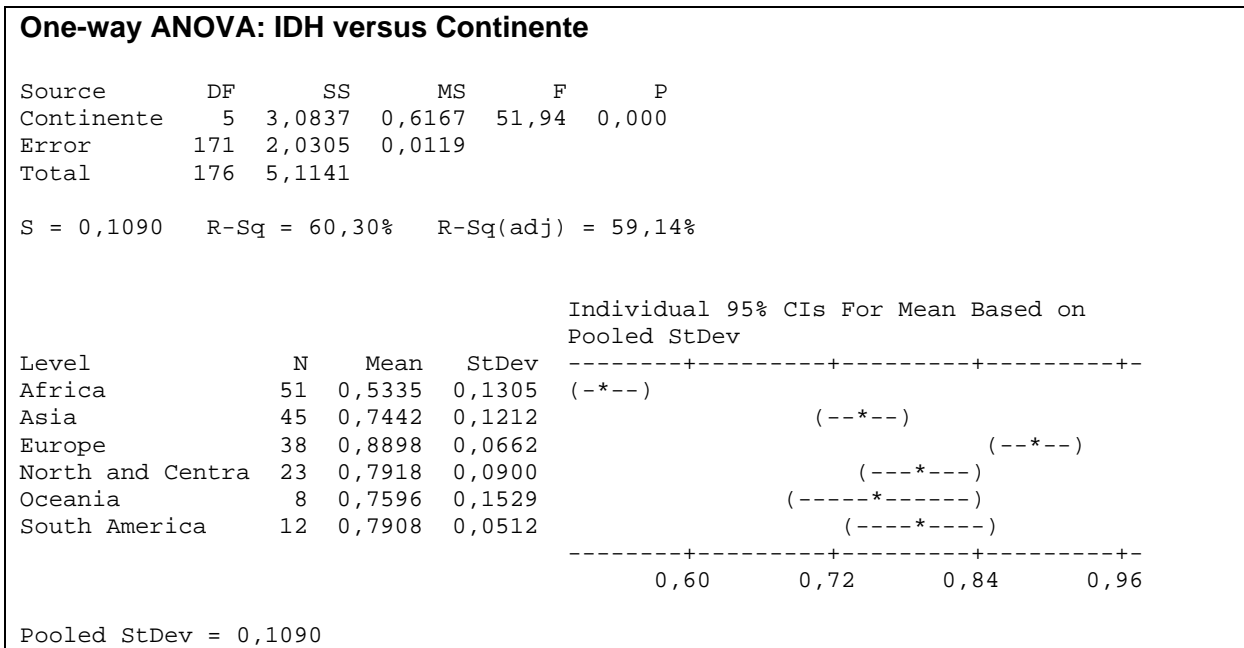


No gráfico acima percebe-se que dentre os seis continentes o que apresenta maior IDH médio é o continente europeu com média de aproximadamente 0,89. Seguido pelo continente

americano norte, sul e central com média de aproximadamente 0,79. Oceania e Ásia com respectivamente 0,76 e 0,74 e por último África com IDH médio de 0,53.

#### 4. ANÁLISE COMPARATIVA

Nesta etapa da análise utilizou-se o teste ANOVA para comparar cada variável quantitativa com a variável “Continente” dos 177 países e para melhor visualizar as diferenças entre cada continente o gráfico box-plot foi confeccionado. Seguem abaixo os quadros:



Como já vimos anteriormente, utilizando o teste ANOVA, tem-se que o maior IDH médio é o do continente Europeu. Observa-se no gráfico Box-plot acima que o IDH dos países europeus não estão muito distantes, o desvio padrão é de 0,0662.

O continente norte e centro americano e o sul americano tem praticamente o mesmo IDH médio, porém o primeiro tem uma dispersão maior que o segundo, o desvio padrão é respectivamente 0,0900 e 0,0512. Ou seja, os países do continente sul americano são mais homogêneos que os do continente norte e centro americano. O continente norte e centro americano apresenta um outlier que é o Haiti, seu IDH é de 0,529.

A Oceania é o continente que apresenta maior desvio padrão, 0,1529. Os países deste continente são os que apresentam IDH mais heterogêneos. Assim como a Oceania, o continente africano tem uma grande dispersão, porém sua média é a mais baixa entre os continentes.

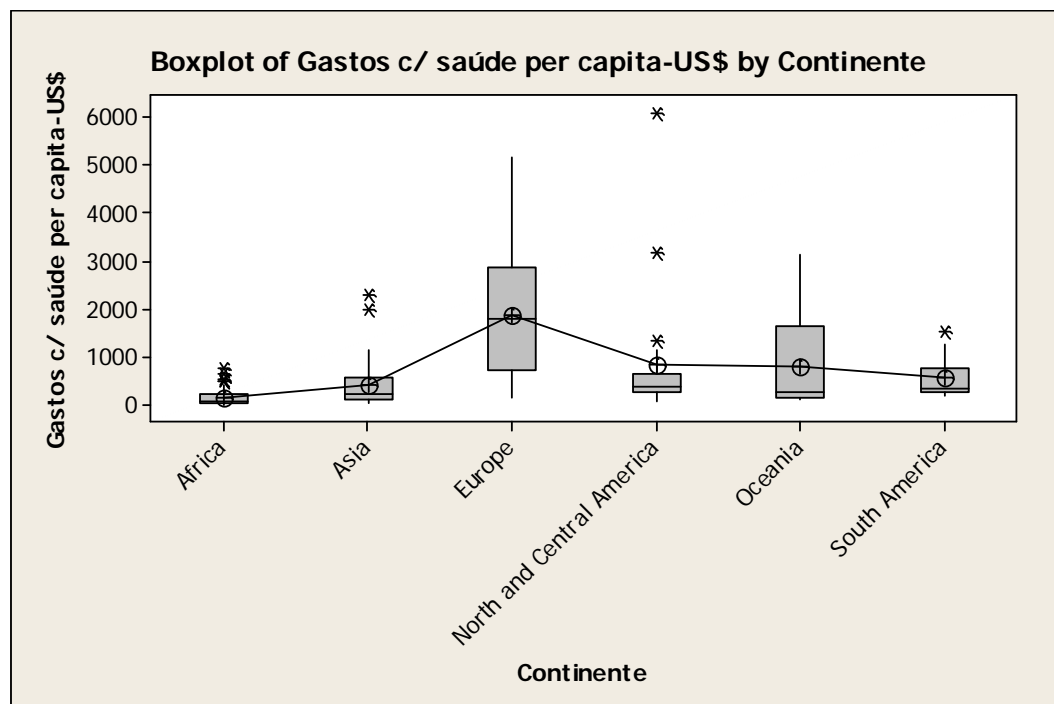
### One-way ANOVA: Gastos c/ saúde per capita-US\$ versus Continente

Source	DF	SS	MS	F	P
Continente	5	70519524	14103905	19,83	0,000
Error	169	120210233	711303		
Total	174	190729757			

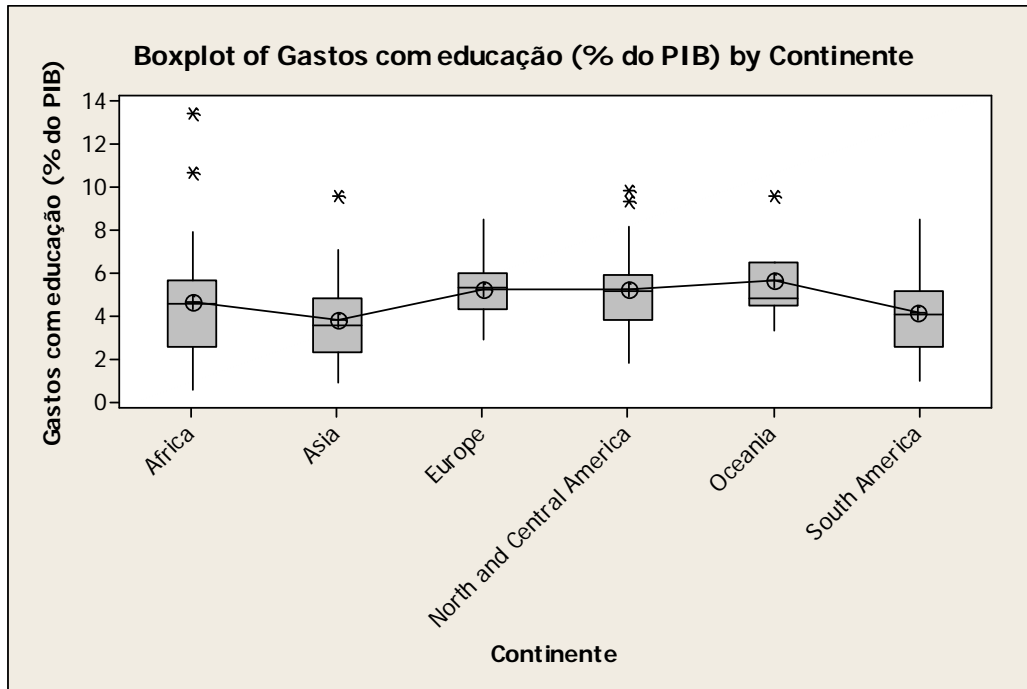
S = 843,4 R-Sq = 36,97% R-Sq(adj) = 35,11%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
Africa	51	150,8	173,0	(---*--)
Asia	43	414,7	479,6	(---*---)
Europe	38	1861,1	1269,9	(-----*-----)
North and Centra	23	818,0	1315,9	(-----*-----)
Oceania	8	800,8	1148,4	(-----*-----)
South America	12	572,3	432,5	(-----*-----)

Pooled StDev = 843,4







Esta é uma variável muito relativa, o montante gasto com educação depende do PIB do país, ou seja, se o país tem um PIB alto seus gastos com educação serão mais altos. Outro fator determinante que se deve considerar é a idade média da população do país. Em geral, quanto mais envelhecida é a população, menor deveriam ser os gastos, pois menor seria o número de habitantes em idade de formação.

A Oceania é o continente que apresenta maior porcentagem média em gastos com educação, 5,7% de seu PIB. Em seguida está o continente europeu e norte centro americano que investem em média, respectivamente, 5,3% e 5,2% de seus PIBs.

O continente africano investe em média 4,6% de seu PIB, acima da América do Sul com 4,2%, e é o que apresenta maior dispersão absoluta. A África possui o país que teve o menor percentual de investimento de seu PIB que foi a Guiné Equatorial (0,6%) e o país que teve o maior percentual de investimento do PIB entre os países analisados, o Lesotho (13,4%).

Em último lugar aparece a Ásia com, em média, 3,8% do seu PIB investido em educação no ano e com maior desvio relativo. Embora tenha países que investiram cerca de 10% de seu PIB em educação tais como Iêmen com 9,6%, têm, em contrapartida, a China com 1,9% e a Indonésia com 0,9% do PIB investidos em educação.

#### One-way ANOVA: Usuários de Internet (%) versus Continente

Source	DF	SS	MS	F	P
Continente	5	28568	5714	21,52	0,000
Error	169	44878	266		
Total	174	73446			

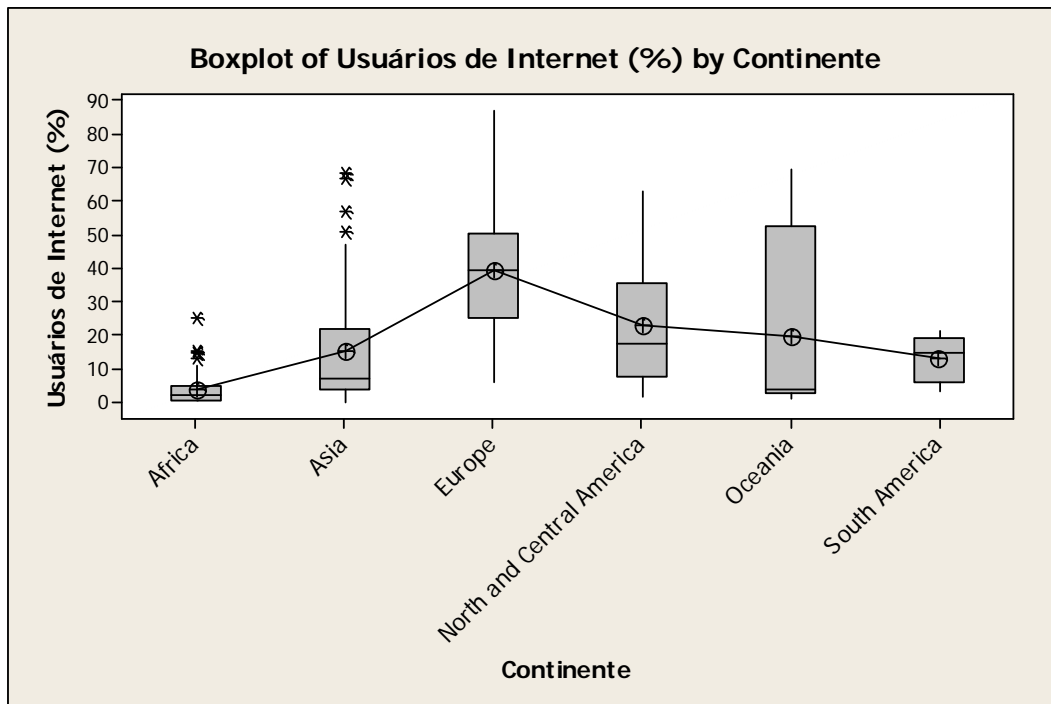
S = 16,30    R-Sq = 38,90%    R-Sq(adj) = 37,09%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	
Africa	51	3,81	4,80	(---*---)
Asia	44	15,31	18,45	(---*---)
Europe	38	39,23	20,11	(-----*-----)
North and Centra	22	22,85	18,63	(-----*-----)
Oceania	8	19,71	30,18	(-----*-----)
South America	12	12,89	6,51	(-----*-----)

0                      12                      24                      36

Pooled StDev = 16,30



O continente africano é o que apresenta menor percentual de usuários de internet, 3,8% da população, contudo há quatro outliers neste continente, Seychelles (24,9%), Marrocos (15,2%), Mauritius (14,6%) e São Tomé e Príncipe (13,1%). O segundo continente com menor percentual é o sul americano com, em média, 12,9% da população e é o que tem menor desvio relativo.

A Ásia tem um percentual médio de usuários de internet de 15,31% da população. Há quatro outliers: República da Coreia (68,4%), Japão (66,8%), Singapura (57,1%) e Hong Kong (50,8%). Seu desvio absoluto e relativo são altos.

O continente europeu tem em média 39,23% da população como usuários de internet, é a média mais alta entre os continentes. Seu desvio padrão absoluto é alto, 20,11%, porém o relativo não está entre os mais altos.

A Oceania e América Norte e Centro tem média de usuários de internet de, respectivamente, 19,71% e 22,85%. Todavia o desvio absoluto e relativo do primeiro é mais alto que o do segundo. O desvio absoluto da Oceania é de 30,18%, é o maior entre todos os continentes e seu intervalo de confiança também é o maior.

## 5. RELAÇÃO ENTRE AS VARIÁVEIS

Esta seção do trabalho está voltada para análise das relações entre as variáveis, incluindo variáveis categóricas e quantitativas.

### 5.1 Relação entre as variáveis categóricas

Para analisar a relação entre as variáveis categóricas utilizamos o teste do qui-quadrado. Este teste indica a probabilidade de haver relação entre as variáveis e indica também a intensidade desta relação. Para que o teste seja possível é necessário a criação prévia de uma tabela de dupla entrada, que é uma tabela que cruza os valores de duas variáveis categóricas, em nosso caso: Continente e Desenvolvimento Humano. Os valores dentro da tabela são as quantidades de países em cada grupo.

<b>Tabulated statistics: Desenvolvimento humano; Continente</b>							
Rows: Desenvolvimento humano				Columns: Continente			
				North and Central America	Oceania	South America	All
	Africa	Asia	Europe				
Alto	3	13	36	11	3	4	70
Baixo	22	0	0	0	0	0	22
Médio	26	32	2	12	5	8	85
All	51	45	38	23	8	12	177
Cell Contents:		Count					

A tabela acima indica que os países do continente africano estão distribuídos, na sua quase totalidade, entre os níveis de IDH médio e baixo. Enquanto que um terço dos países do continente asiático, da Oceania e do continente sul americano estão no nível alto de IDH e o restante no nível médio. Quase todos os países do continente europeu estão no nível alto de IDH. Já os países da América Central e Norte estão divididos igualmente entre os níveis alto e médio de IDH.

### 5.2 Relação entre as variáveis quantitativas

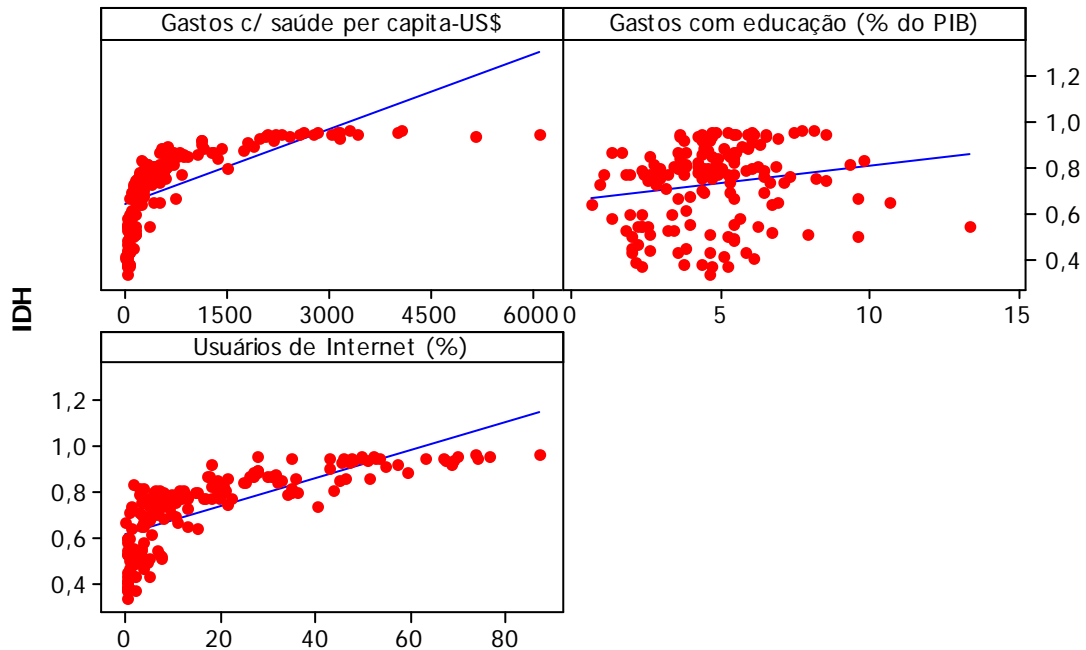
Verificam-se as relações entre as variáveis quantitativas utilizando a análise de gráficos de dispersão, matrizes de correlação e dendogramas.

#### 5.2.1 Análise gráfica

Abaixo, os gráficos de dispersão entre todas as variáveis quantitativas de nossa pesquisa, facilitando uma análise visual.

Inicialmente, gráfico de correlação da variável IDH com as demais.

### Scatterplot of IDH vs Gastos c/ sa; Gastos com e; Usuários de



A correlação é grandeza utilizada para analisar o grau de relacionamento entre duas variáveis ( $X$  e  $Y$ ). A correlação entre essas variáveis é escrita como  $r(X, Y)$  se os dados provêm de uma amostra. Acima, em cada um dos gráficos é analisada a relação entre duas variáveis quantitativas, como *IDH* e *Gastos com Saúde per capita*, por exemplo.

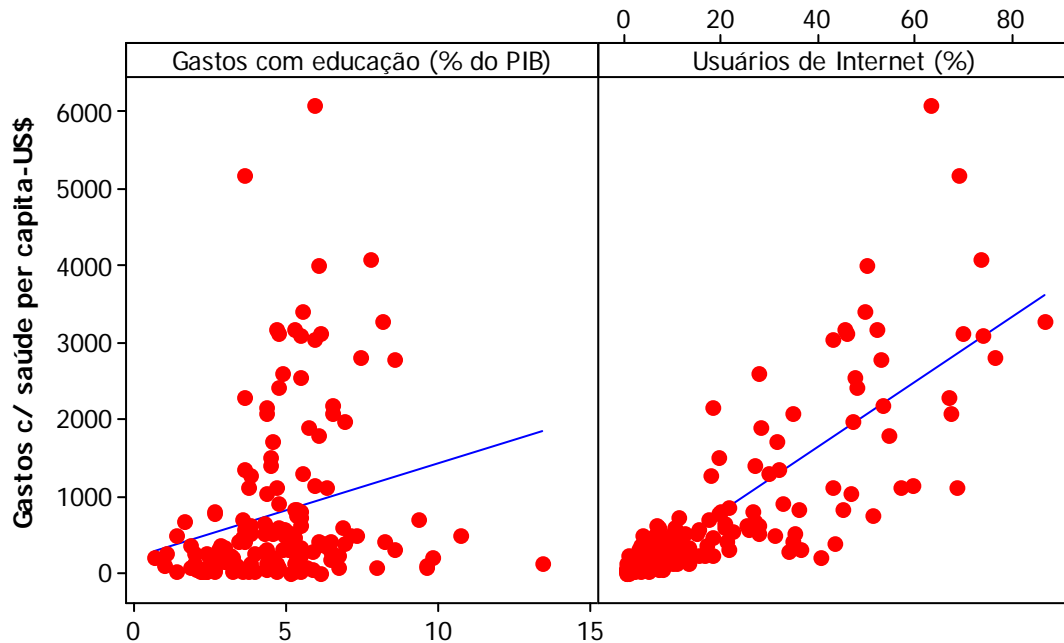
Da análise visual dessa primeira relação, nota-se que um grande grupo de países, independentemente de seu IDH, possui gastos com saúde por habitante muito parecidos. Já um segundo grupo países com índice de desenvolvimento humano mais elevado não apresentam relação clara dos gastos com saúde per capita com seu IDH.

A segunda correlação é feita entre a mesma variável IDH e os Gastos com Educação (como percentual do PIB). Os pontos dispersos no gráfico acusam a total falta de correlação com a forma como se investe em educação nos diversos países e o índice de desenvolvimento humano. Ou seja, não é possível afirmar que países com IDH mais altos investem percentual mais elevado do seu PIB em educação que os demais.

A última correlação é referente às variáveis IDH e percentual de usuários de internet naquele país. A correlação se dá da mesma forma que nas primeiras duas variáveis. Porém, é possível visualizar uma relação um pouco mais positiva, levando à seguinte possível conclusão: quanto mais elevado é o IDH de determinado país, maior será o percentual de pessoas com acesso à rede mundial de computadores.

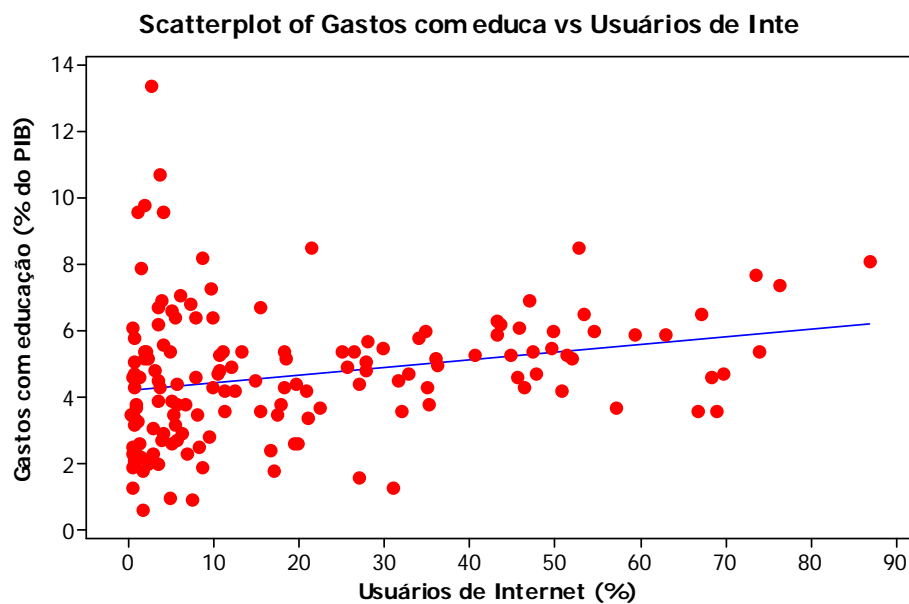
A seguir, será analisada a correlação entre a variável Gastos com Saúde per capita com as demais variáveis.

### Scatterplot of Gastos c/ saúde vs Gastos com educa; Usuários de Inte



No tocante às duas primeiras variáveis, Gastos com Saúde per capita (US\$) e Gastos com Educação (% do PIB), pouca correlação foi achada, em função da dispersão dos pontos, fugindo à reta. Quanto as próxima correlação, nota-se que é positiva, apesar da grande concentração de países se comportando da mesma maneira.

Abaixo, correlação entre a variável Gastos com Educação com Usuários de Internet:



Pode-se dizer que a correlação entre as duas últimas variáveis, Gastos com Educação (% do PIB) e Usuários de Internet (%), é positiva, levando à conclusão de que quanto maior o valor percentual em relação ao PIB é investido por um país em educação, maior será o percentual da população com acesso à rede de computadores mundial.

### 5.2.2 Coeficientes de Correlação

A propriedade mais importante do coeficiente de correlação é que seu valor está sempre entre -1 e 1. Se X e Y são independentes, então obviamente sua relação é zero. Se o coeficiente de relação for positivo, entende-se então que quando X tende a crescer, o mesmo ocorre com Y, caracterizando as variáveis como correlacionadas positivamente. Quanto mais próximo de 1, mais correlacionadas estarão as variáveis. Caso ocorra o inverso, ou seja, se Y tende a tornar-se pequeno enquanto X cresce, diz-se que as variáveis são correlacionadas negativamente.

Abaixo segue a matriz de correlação, utilizando-se do índice de Pearson.

<b>Correlations: IDH; Gastos c/ saúde ; Gastos com educa; Usuários de Inte</b>			
	IDH	Gastos c/ sa	Gastos com e
Gastos c/ sa	0,665 0,000		
Gastos com e	0,175 0,030	0,227 0,005	
Usuários de	0,746 0,000	0,817 0,000	0,245 0,002

Cell Contents: Pearson correlation  
P-Value

Confirma-se através da correlação de Pearson o demonstrado na análise visual das correlações: todas as correlações são positivas, umas mais, outras menos. A correlação mais significativa deu-se surpreendentemente entre as variáveis Gastos com Saúde e Usuários de Internet (0,817). Por ser positiva, induz à conclusão de que quanto mais se gasta com saúde per capita, maior será o índice de conectividade.

Em segundo lugar, a relação positiva entre IDH e Usuários de Internet também (0,746) sugere que quanto maior for o IDH de um país, maior será a quantidade de usuários conectados à internet.

Outra correlação relevante ocorre entre as variáveis IDH e Gastos com Saúde (0,665). Assim como na última correlação mencionada, pode haver a seguinte hipótese: quanto maior IDH de um país, maior será o gasto per capita com saúde. Porém, nenhuma das correlações acima apresenta força suficiente para uma afirmação conclusiva.

Quanto às outras variáveis, suas correlações não são significativas o suficiente para qualquer afirmação, pois não passam de 0,245.

## 6. MODELOS PARA EXPLICAR O IDH UTILIZANDO INDICADORES SÓCIO-ECONÔMICOS

### 6.1 Regressão Linear

Regressão significa “voltar para trás”. A utilização de modelos de regressões lineares requerem, como premissa elementar, a relação de causa-efeito entre as variáveis. Uma tem que necessariamente ser a variável dependente e a outra a variável independente ou explicativa. Os resultados podem se alterar apenas com a alternância entre qual é a variável explicativa e qual é a dependente.

A regressão linear assume sempre a forma de uma equação linear:

$$Y = a + bx, \text{ sendo:}$$

Y= Variável dependente;

a = uma constante, o intercepto;

b = a inclinação na reta;

x = variável independente ou explicativa.

O “b”, ou seja, a declividade é dado pela multiplicação do índice de correlação pela divisão dos desvios-padrão entre as variáveis x e y. E “a” é dado pela média de “Y” menos a multiplicação de “b” pela média de “x”. Assim, percebe-se muito claramente que a regressão depende da correlação entre as variáveis, além de medidas do centro de cada uma das variáveis.

Iniciamos com a regressão dos mínimos quadrados com as séries de dados. Segue abaixo o resultado desta regressão.

### Regression Analysis: IDH versus Gastos c/ sa; Gastos com e; Usuários de

The regression equation is

$$\begin{aligned} \text{IDH} = & 0,621 + 0,000029 \text{ Gastos c/ saúde per capita-US\$} \\ & - 0,00181 \text{ Gastos com educação (\% do PIB)} \\ & + 0,00495 \text{ Usuários de Internet (\%)} \end{aligned}$$

152 cases used, 25 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	0,62089	0,02348	26,44	0,000
Gastos c/ saúde per capita-US\$	0,00002898	0,00001441	2,01	0,046
Gastos com educação (% do PIB)	-0,001806	0,004729	-0,38	0,703
Usuários de Internet (%)	0,0049460	0,0007476	6,62	0,000

S = 0,113673    R-Sq = 57,9%    R-Sq(adj) = 57,1%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	2,63078	0,87693	67,87	0,000
Residual Error	148	1,91238	0,01292		
Total	151	4,54316			

Source	DF	Seq SS
Gastos c/ saúde per capita-US\$	1	2,06384
Gastos com educação (% do PIB)	1	0,00138
Usuários de Internet (%)	1	0,56557

## Unusual Observations

Obs	Gastos c/ saúde per capita-US\$	IDH	Fit	SE Fit	Residual	St Resid
1	3294	0,96800	1,13154	0,03235	-0,16354	-1,50 X
7	4011	0,95500	0,97262	0,03212	-0,01762	-0,16 X
12	6096	0,95100	1,09852	0,05408	-0,14752	-1,48 X
18	5178	0,94400	1,10574	0,04225	-0,16174	-1,53 X
26	1135	0,92100	0,98379	0,03427	-0,06279	-0,58 X
124	504	0,65400	0,63299	0,03263	0,02101	0,19 X
138	139	0,54900	0,61259	0,04477	-0,06359	-0,61 X
170	42	0,38800	0,62029	0,01592	-0,23229	-2,06R
172	42	0,38400	0,61889	0,01252	-0,23489	-2,08R
173	54	0,38000	0,61667	0,01237	-0,23667	-2,09R
174	26	0,37400	0,61848	0,01537	-0,24448	-2,17R
175	28	0,37400	0,62220	0,01275	-0,24820	-2,20R
176	77	0,37000	0,61711	0,01256	-0,24711	-2,19R
177	34	0,33600	0,61456	0,01255	-0,27856	-2,47R

R denotes an observation with a large standardized residual.

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

A base de cálculo da equação de regressão linear foi de 152 países, sendo que outros 25 não foram analisados por não apresentarem dados.

Tomando a variável “IDH” como variável resposta e as variáveis “Gastos c/ saúde per capita-US\$”, “Gastos com educação (% do PIB)” e “Usuários de Internet (%)” como variáveis explicativas, temos a seguinte equação de regressão:

$$\text{IDH} = 0,621 + 0,000029 \text{ Gastos c/ saúde per capita-US\$} - 0,00181 \text{ Gastos com educação (\% do PIB)} + 0,00495 \text{ Usuários de Internet (\%)}$$

Esta equação explica 57,9% da variação dos valores da variável resposta. Porém percebe-se que a variável “Gastos com educação (% do PIB)” não está ajudando na equação, ou seja, não é significativa para a regressão devido ao valor de seu p-value ser maior que 5%. Recalculando a regressão linear utilizando o método stepwise temos:

### Stepwise Regression: IDH versus Gastos c/ sa; Gastos com e; Usuários de

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is IDH on 3 predictors, with N = 152  
N(cases with missing observations) = 25 N(all cases) = 177

Step	1	2
Constant	0,6131	0,6133
Usuários de Internet (%)	0,00612	0,00492
T-Value	14,02	6,63
P-Value	0,000	0,000
Gastos c/ saúde per capita-US\$		0,00003
T-Value		2,00
P-Value		0,047
S	0,114	0,113
R-Sq	56,73	57,86
R-Sq(adj)	56,45	57,30
Mallows C-p	4,1	2,1

Como foi observado nos cálculos anteriores a variável “Gastos com educação (% do PIB)” não ajuda a explicar a variável “IDH”, portanto não entrou na equação de regressão stepwise, pois não agregaria nenhuma informação.

A amostra é extensa o suficiente para corroborar nossos resultados, portanto pode-se ter uma noção próxima da realidade a partir desta análise.

O valor do R-quadrado é de aproximadamente 57% utilizando apenas a variável “Usuários de Internet (%)”. Acrescentando a variável “Gastos c/ saúde per capita-US\$” o R-quadrado aumenta apenas 1%, ou seja, a equação passa a explicar aproximadamente 58% da variável resposta. Este resultado é satisfatório, pois o P-value é menor que 5%.

## 6.2 Regressão Logística Ordinal

### Ordinal Logistic Regression: Code versus Gastos c/ sa; Gastos com e; ...

Link Function: Logit

#### Response Information

Variable	Value	Count
Code	1	19
	2	68
	3	65
	Total	152

\* NOTE \* 152 cases were used  
 \* NOTE \* 25 cases contained missing values

#### Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio
Const(1)	0,0333111	0,532787	0,06	0,950	
Const(2)	5,43390	0,997656	5,45	0,000	
Gastos c/ saúde per capita-US\$	-0,0092174	0,0018580	-4,96	0,000	0,99
Gastos com educação (% do PIB)	0,0383317	0,101252	0,38	0,705	1,04
Usuários de Internet (%)	-0,0706109	0,0287015	-2,46	0,014	0,93

Predictor	95% CI	
	Lower	Upper
Const(1)		
Const(2)		
Gastos c/ saúde per capita-US\$	0,99	0,99
Gastos com educação (% do PIB)	0,85	1,27
Usuários de Internet (%)	0,88	0,99

Log-Likelihood = -67,074

Test that all slopes are zero: G = 164,699, DF = 3, P-Value = 0,000

#### Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	181,302	299	1,000
Deviance	134,149	299	1,000

#### Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
-------	--------	---------	------------------

Concordant	6641	95,6	Somers' D	0,91
Discordant	300	4,3	Goodman-Kruskal Gamma	0,91
Ties	6	0,1	Kendall's Tau-a	0,55
Total	6947	100,0		

Apesar do poder de concordância (95,6%) ter sido elevado, os resultados não podem ser levados em consideração, pois algumas variáveis apresentaram *p-value* superior a 10%. Assim, procedemos com a exclusão de cada variável que apresenta *p-value* maior do que 10%, uma a uma, sempre seguindo a regra de excluir primeiro a variável de maior *p-value*. Assim, chegamos a um modelo bastante simples, com duas variáveis explicativas, Gastos c/ saúde per capita-US\$ e Usuários de Internet (%), que por sinal é a mesma que foi mantida no modelo de regressão dos mínimos quadrados. Vejamos abaixo:

### Ordinal Logistic Regression: Code versus Gastos c/ sa; Usuários de

Link Function: Logit

#### Response Information

Variable	Value	Count
Code	1	22
	2	83
	3	68
	Total	173

\* NOTE \* 173 cases were used  
 \* NOTE \* 4 cases contained missing values

#### Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio
Const(1)	0,297018	0,327898	0,91	0,365	
Const(2)	6,27822	0,955762	6,57	0,000	
Gastos c/ saúde per capita-US\$	-0,0103144	0,0019179	-5,38	0,000	0,99
Usuários de Internet (%)	-0,0802897	0,0292300	-2,75	0,006	0,92

Predictor	95% CI	
	Lower	Upper
Const(1)		
Const(2)		
Gastos c/ saúde per capita-US\$	0,99	0,99
Usuários de Internet (%)	0,87	0,98

Log-Likelihood = -72,895

Test that all slopes are zero: G = 193,863, DF = 2, P-Value = 0,000

#### Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	221,855	342	1,000
Deviance	145,789	342	1,000

#### Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	8623	96,2	Somers' D 0,92
Discordant	335	3,7	Goodman-Kruskal Gamma 0,93
Ties	8	0,1	Kendall's Tau-a 0,56

Total	8966	100,0
-------	------	-------

O nível de concordância do modelo se manteve alto ao retirarmos a variável explicativa “Gastos com educação (% do PIB)”, ficando em um patamar bastante elevado: 96,2%. Isso significa que, dado os valores das duas variáveis restantes, o modelo é capaz de prever o nível de IDH, com capacidade de acerto de 96,2%. A tabela de frequências observadas e esperadas nos dão uma noção da qualidade do ajuste do modelo. Quanto mais próximos estão estes valores, melhor o ajuste e menores são os erros do modelo.

### 6.3 Análise Discriminante

Como o próprio nome já denota, este tipo de análise discrimina os indivíduos em grupos de acordo com as variáveis escolhidas. Da mesma forma que na regressão logística ordinal, os grupos são os três níveis de IDH (Alto, Médio e Baixo).

São dois os tipos de análise discriminante: linear e quadrática. Veremos qual das duas discrimina melhor os dados entre os grupos.

#### Discriminant Analysis: Desenvolvimento versus Gastos c/ sa; Gastos com e; ...

Linear Method for Response: Desenvolvimento humano

Predictors: Gastos c/ saúde per capita-US\$; Gastos com educação (% do PIB);  
Usuários de Internet (%)

Group	Alto	Baixo	Médio
Count	65	19	68

152 cases used, 25 cases contain missing values

Summary of classification

Put into Group	True Group		
	Alto	Baixo	Médio
Alto	48	0	3
Baixo	5	14	30
Médio	12	5	35
Total N	65	19	68
N correct	48	14	35
Proportion	0,738	0,737	0,515

N = 152

N Correct = 97

Proportion Correct = 0,638

Squared Distance Between Groups

	Alto	Baixo	Médio
Alto	0,00000	5,82493	4,05573
Baixo	5,82493	0,00000	0,23870
Médio	4,05573	0,23870	0,00000

Linear Discriminant Function for Groups

	Alto	Baixo	Médio
Constant	-5,0940	-2,0169	-2,5923
Gastos c/ saúde per capita-US\$	0,0003	-0,0002	-0,0003

Gastos com educação (% do PIB)	1,0160	1,0238	1,1310
Usuários de Internet (%)	0,1237	-0,0162	0,0155

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
38**	Alto	Médio	Alto	2,144	0,321
			Baixo	2,247	0,305
			Médio	1,844	0,373
40**	Alto	Médio	Alto	2,0066	0,244
			Baixo	1,3812	0,334
			Médio	0,9086	0,422
46**	Alto	Médio	Alto	2,353	0,316
			Baixo	2,398	0,309
			Médio	2,004	0,376
51**	Alto	Médio	Alto	14,363	0,023
			Baixo	8,778	0,370
			Médio	7,788	0,607
52**	Alto	Médio	Alto	2,9191	0,190
			Baixo	2,0178	0,299
			Médio	0,9420	0,511
53**	Alto	Médio	Alto	1,4739	0,311
			Baixo	1,7363	0,272
			Médio	0,8839	0,417
55**	Alto	Baixo	Alto	5,3356	0,038
			Baixo	0,2567	0,485
			Médio	0,2893	0,477
56**	Alto	Baixo	Alto	5,4391	0,047
			Baixo	0,6015	0,529
			Médio	1,0436	0,424
58**	Alto	Médio	Alto	3,2026	0,112
			Baixo	0,5466	0,422
			Médio	0,3486	0,466
59**	Alto	Médio	Alto	2,7786	0,130
			Baixo	0,5605	0,395
			Médio	0,1890	0,475
60**	Alto	Médio	Alto	2,239	0,296
			Baixo	2,315	0,285
			Médio	1,550	0,418
61**	Alto	Médio	Alto	5,843	0,064
			Baixo	2,185	0,402
			Médio	1,615	0,534
62**	Alto	Baixo	Alto	4,2694	0,073
			Baixo	0,5237	0,477
			Médio	0,6438	0,449
65**	Alto	Médio	Alto	2,4011	0,162
			Baixo	0,8142	0,358
			Médio	0,2308	0,480
67**	Alto	Médio	Alto	2,3612	0,187
			Baixo	1,0279	0,365
			Médio	0,6195	0,448
68**	Alto	Baixo	Alto	4,6787	0,062
			Baixo	0,5212	0,498
			Médio	0,7719	0,439
69**	Alto	Baixo	Alto	3,8872	0,081
			Baixo	0,3933	0,464
			Médio	0,4347	0,455
71**	Médio	Alto	Alto	4,002	0,678
			Baixo	8,201	0,083
			Médio	6,088	0,239
72**	Médio	Alto	Alto	3,895	0,602
			Baixo	7,458	0,101
			Médio	5,311	0,297
73**	Médio	Baixo	Alto	6,0112	0,042
			Baixo	0,8594	0,547
			Médio	1,4251	0,412

77**	Médio	Baixo	Alto	5,18855	0,038
			Baixo	0,09324	0,487
			Médio	0,14194	0,475
81**	Médio	Baixo	Alto	5,086	0,085
			Baixo	1,577	0,491
			Médio	1,866	0,425
83**	Médio	Baixo	Alto	4,7602	0,053
			Baixo	0,2811	0,497
			Médio	0,4780	0,450
89**	Médio	Baixo	Alto	7,117	0,056
			Baixo	2,542	0,555
			Médio	3,259	0,388
90**	Médio	Baixo	Alto	5,0460	0,054
			Baixo	0,5780	0,507
			Médio	0,8689	0,439
95**	Médio	Baixo	Alto	5,0470	0,042
			Baixo	0,1275	0,492
			Médio	0,2390	0,466
96**	Médio	Baixo	Alto	5,3012	0,044
			Baixo	0,3762	0,518
			Médio	0,7096	0,438
98**	Médio	Baixo	Alto	4,7542	0,073
			Baixo	0,9750	0,480
			Médio	1,1186	0,447
101**	Médio	Alto	Alto	5,799	0,761
			Baixo	11,104	0,054
			Médio	8,629	0,185
103**	Médio	Baixo	Alto	3,9979	0,093
			Baixo	0,8037	0,462
			Médio	0,8779	0,445
107**	Médio	Baixo	Alto	6,908	0,071
			Baixo	2,908	0,526
			Médio	3,439	0,403
110**	Médio	Baixo	Alto	5,4747	0,039
			Baixo	0,2751	0,526
			Médio	0,6595	0,434
119**	Médio	Baixo	Alto	4,67076	0,049
			Baixo	0,08165	0,486
			Médio	0,17151	0,465
122**	Médio	Baixo	Alto	6,18513	0,025
			Baixo	0,06238	0,537
			Médio	0,47258	0,438
127**	Médio	Baixo	Alto	8,619	0,038
			Baixo	3,086	0,599
			Médio	4,087	0,363
128**	Médio	Baixo	Alto	4,7138	0,048
			Baixo	0,1367	0,478
			Médio	0,1520	0,474
129**	Médio	Baixo	Alto	5,9933	0,029
			Baixo	0,1337	0,537
			Médio	0,5604	0,434
130**	Médio	Baixo	Alto	6,7229	0,028
			Baixo	0,7242	0,565
			Médio	1,3786	0,407
131**	Médio	Baixo	Alto	7,103	0,029
			Baixo	1,153	0,577
			Médio	1,920	0,393
132**	Médio	Baixo	Alto	7,905	0,028
			Baixo	1,827	0,590
			Médio	2,698	0,382
134**	Médio	Baixo	Alto	5,31393	0,036
			Baixo	0,04911	0,496
			Médio	0,16244	0,469
136**	Médio	Baixo	Alto	5,332	0,059
			Baixo	1,068	0,500
			Médio	1,316	0,441
137**	Médio	Baixo	Alto	6,6638	0,029
			Baixo	0,7157	0,561

			Médio	1,3419	0,410
139**	Médio	Baixo	Alto	6,6061	0,031
			Baixo	0,8135	0,557
			Médio	1,4193	0,412
140**	Médio	Baixo	Alto	6,5989	0,027
			Baixo	0,5583	0,561
			Médio	1,1732	0,412
142**	Médio	Baixo	Alto	6,10878	0,027
			Baixo	0,08972	0,537
			Médio	0,50746	0,436
143**	Médio	Baixo	Alto	6,1816	0,026
			Baixo	0,1486	0,540
			Médio	0,5903	0,433
144**	Médio	Baixo	Alto	6,913	0,033
			Baixo	1,239	0,567
			Médio	1,934	0,400
152**	Médio	Baixo	Alto	5,4016	0,047
			Baixo	0,6453	0,512
			Médio	0,9475	0,440
155**	Médio	Baixo	Alto	6,253	0,041
			Baixo	1,085	0,544
			Médio	1,624	0,415
156**	Baixo	Médio	Alto	5,4326	0,039
			Baixo	0,5585	0,447
			Médio	0,2821	0,514
157**	Baixo	Médio	Alto	6,2964	0,027
			Baixo	0,5349	0,475
			Médio	0,4348	0,499
164**	Baixo	Médio	Alto	6,9642	0,023
			Baixo	0,8940	0,476
			Médio	0,7933	0,501
169**	Baixo	Médio	Alto	7,387	0,021
			Baixo	1,217	0,469
			Médio	1,053	0,509
175**	Baixo	Médio	Alto	6,0419	0,028
			Baixo	0,3965	0,476
			Médio	0,3142	0,496

### Discriminant Analysis: Desenvolvime versus Gastos c/ sa; Gastos com e; ...

Quadratic Method for Response: Desenvolvimento humano

Predictors: Gastos c/ saúde per capita-US\$; Gastos com educação (% do PIB);  
Usuários de Internet (%)

Group	Alto	Baixo	Médio
Count	65	19	68

152 cases used, 25 cases contain missing values

Summary of classification

Put into Group	True Group		
	Alto	Baixo	Médio
Alto	52	0	5
Baixo	0	19	17
Médio	13	0	46
Total N	65	19	68
N correct	52	19	46
Proportion	0,800	1,000	0,676

N = 152

N Correct = 117

Proportion Correct = 0,770

## From Generalized Squared Distance to Group

Group	Alto	Baixo	Médio
Alto	20,45	3741,29	76,17
Baixo	23,30	7,97	17,43
Médio	22,46	75,54	16,19

## Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
39**	Alto	Médio	Alto	27,24	0,435
			Baixo	703,25	0,000
			Médio	26,71	0,565
51**	Alto	Médio	Alto	40,19	0,000
			Baixo	75,35	0,000
			Médio	21,50	1,000
52**	Alto	Médio	Alto	22,13	0,121
			Baixo	205,52	0,000
			Médio	18,16	0,879
55**	Alto	Médio	Alto	23,34	0,042
			Baixo	102,29	0,000
			Médio	17,07	0,958
56**	Alto	Médio	Alto	23,90	0,045
			Baixo	108,00	0,000
			Médio	17,77	0,955
58**	Alto	Médio	Alto	22,16	0,089
			Baixo	226,82	0,000
			Médio	17,51	0,911
59**	Alto	Médio	Alto	21,81	0,186
			Baixo	351,42	0,000
			Médio	18,86	0,814
60**	Alto	Médio	Alto	21,90	0,217
			Baixo	381,61	0,000
			Médio	19,33	0,783
61**	Alto	Médio	Alto	25,99	0,099
			Baixo	427,75	0,000
			Médio	21,57	0,901
65**	Alto	Médio	Alto	21,60	0,185
			Baixo	370,94	0,000
			Médio	18,64	0,815
67**	Alto	Médio	Alto	21,81	0,329
			Baixo	461,04	0,000
			Médio	20,39	0,671
68**	Alto	Médio	Alto	23,35	0,047
			Baixo	123,74	0,000
			Médio	17,35	0,953
69**	Alto	Médio	Alto	22,57	0,123
			Baixo	253,80	0,000
			Médio	18,65	0,877
71**	Médio	Alto	Alto	22,35	0,974
			Baixo	712,56	0,000
			Médio	29,57	0,026
72**	Médio	Alto	Alto	22,67	0,928
			Baixo	632,46	0,000
			Médio	27,78	0,072
88**	Médio	Alto	Alto	22,94	0,942
			Baixo	911,09	0,000
			Médio	28,52	0,058
101**	Médio	Alto	Alto	23,33	0,998
			Baixo	835,65	0,000
			Médio	36,15	0,002
121**	Médio	Alto	Alto	22,65	0,767
			Baixo	682,14	0,000
			Médio	25,04	0,233
122**	Médio	Baixo	Alto	23,583	0,001

			Baixo	8,700	0,988
			Médio	17,560	0,012
129**	Médio	Baixo	Alto	23,60	0,005
			Baixo	13,28	0,869
			Médio	17,14	0,126
130**	Médio	Baixo	Alto	24,84	0,001
			Baixo	10,11	0,979
			Médio	17,84	0,021
132**	Médio	Baixo	Alto	27,04	0,001
			Baixo	12,41	0,961
			Médio	18,85	0,038
133**	Médio	Baixo	Alto	24,12	0,009
			Baixo	15,47	0,701
			Médio	17,24	0,290
134**	Médio	Baixo	Alto	23,01	0,003
			Baixo	11,25	0,958
			Médio	17,63	0,040
135**	Médio	Baixo	Alto	24,20	0,002
			Baixo	12,22	0,924
			Médio	17,27	0,074
137**	Médio	Baixo	Alto	24,793	0,001
			Baixo	9,641	0,985
			Médio	18,021	0,015
139**	Médio	Baixo	Alto	24,90	0,001
			Baixo	10,44	0,978
			Médio	18,14	0,021
140**	Médio	Baixo	Alto	24,529	0,001
			Baixo	9,529	0,984
			Médio	17,802	0,016
142**	Médio	Baixo	Alto	23,590	0,001
			Baixo	8,952	0,985
			Médio	17,409	0,014
143**	Médio	Baixo	Alto	23,709	0,001
			Baixo	9,173	0,986
			Médio	17,807	0,013
144**	Médio	Baixo	Alto	25,69	0,001
			Baixo	11,33	0,965
			Médio	18,00	0,034
148**	Médio	Baixo	Alto	26,61	0,003
			Baixo	15,78	0,759
			Médio	18,10	0,238
149**	Médio	Baixo	Alto	31,04	0,001
			Baixo	18,93	0,575
			Médio	19,53	0,424
152**	Médio	Baixo	Alto	23,87	0,015
			Baixo	16,51	0,615
			Médio	17,53	0,370
155**	Médio	Baixo	Alto	25,08	0,003
			Baixo	13,72	0,876
			Médio	17,68	0,121

O primeiro quadro refere-se ao modelo linear e o segundo ao modelo quadrático. A proporção de acerto do modelo quadrático (70,0%) é maior do que a do modelo linear (63,8%). Desta maneira, o modelo quadrático é o que será utilizado para as análises.

A maior proporção de acerto foi a do grupo de países com baixo nível de IDH, acertando em 19 dos 19 casos (100%). A segundo maior proporção de acerto foi para o grupo de países com alto nível de IDH, discriminando corretamente 52 dos 65 países (80%). E, a pior proporção de acerto foi para o grupo de países com o IDH no nível médio, acertando na classificação de apenas 46 dos 68 países (67,6% dos casos). No geral, o modelo classificou corretamente 117 dos 152 países ou 77,0%.

## 7. ANÁLISE DE CLUSTERS

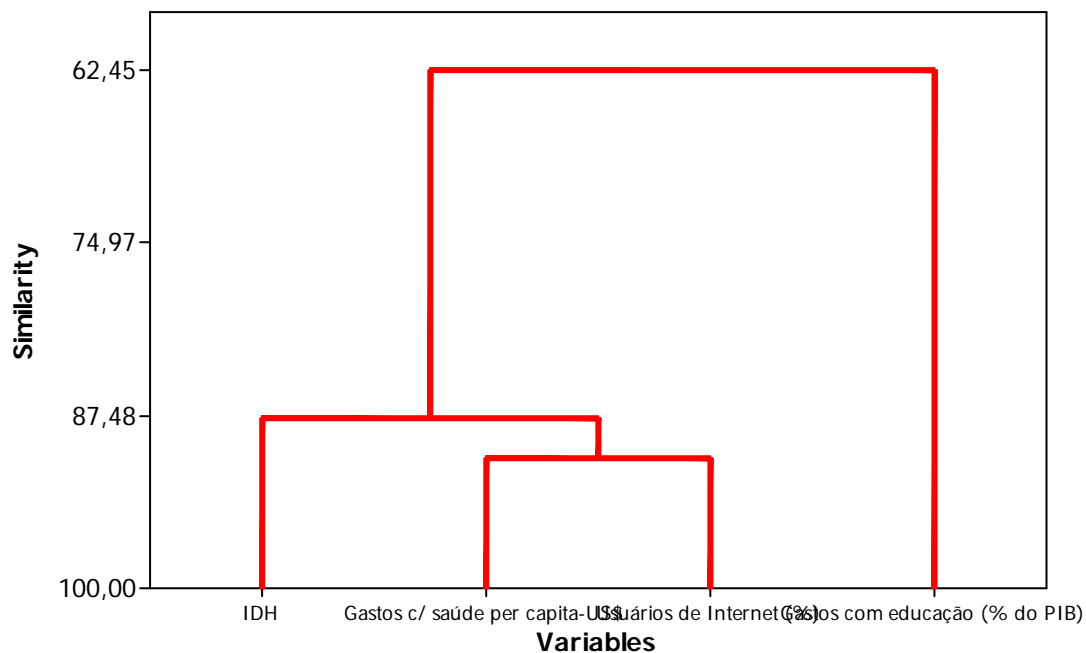
Esta seção dedica-se a análise de *clusters* ou agrupamentos. Estes agrupamentos podem se dar em relação às variáveis explicativas ou às observações.

O principal objetivo de agrupar as variáveis explicativas é possibilitar a criação de índices para simplificar as análises e modelos. Em nosso caso, o objetivo é tentar agrupar as quatro variáveis explicativas em um ou mais índices que bem representem o comportamento dos países em relação às variáveis sócio-econômicas. Já o objetivo de agrupar observações é tentar unir em grupos países com características semelhantes em relação a todas as variáveis, dependentes e explicativas.

### 7.1 Cluster de variáveis

A técnica classificatória multivariada pode ser utilizada quando se deseja explorar as similaridades entre variáveis definindo-as em grupos, considerando, simultaneamente, todos os indivíduos nos quais foram feitas as mensurações. Para a tentativa de agrupamento das variáveis explicativas em índices, utilizamos a análise dos componentes principais. Neste tipo de ferramenta, a correlação entre as variáveis é elemento chave, pois, mais eficiente é o agrupamento de variáveis quanto mais fortes são as correlações entre elas. No dendograma estão dispostas linhas ligadas segundo os níveis de similaridade que agruparam pares de variáveis. O objetivo deste algoritmo é o de unificar objetos em classes ou grupos sucessivamente maiores através da utilização de alguma medida de similaridade ou de distância. Um resultado típico deste enfoque é uma árvore hierárquica, como no dendograma abaixo:

**Dendrogram with Single Linkage and Correlation Coefficient Distance**



Assim, o dendograma acima reforça a afirmação supracitada de uma forte relação entre as variáveis “Gastos com Saúde per capita (US\$)” e “Usuários de Internet (%)”, seguido pela relação do IDH com esse par.

A seguir veremos como se dá a análise dos componentes principais:

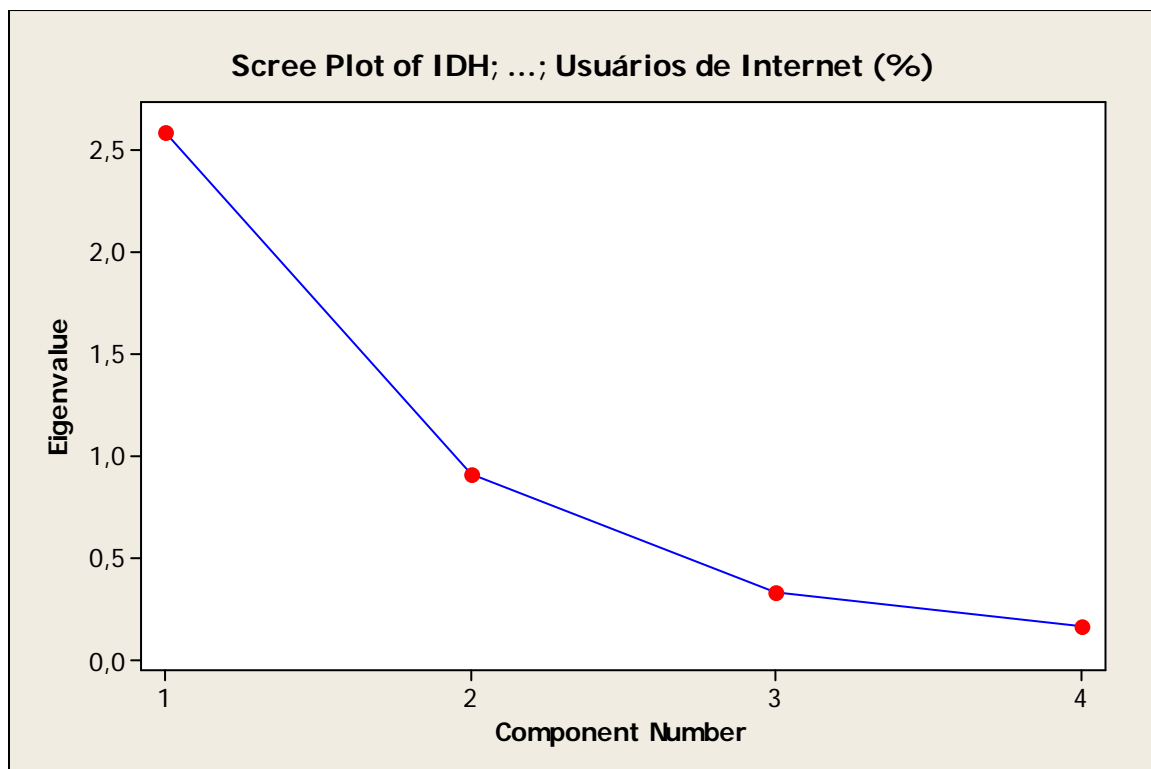
### Principal Component Analysis: IDH; Gastos c/ sa; Gastos com e; Usuários de

Eigenanalysis of the Correlation Matrix  
152 cases used, 25 cases contain missing values

Eigenvalue	2,5853	0,9129	0,3301	0,1717
Proportion	0,646	0,228	0,083	0,043
Cumulative	0,646	0,875	0,957	1,000

Variable	PC1	PC2	PC3	PC4
IDH	-0,541	-0,193	0,776	-0,261
Gastos c/ saúde per capita-US\$	-0,562	-0,107	-0,605	-0,554
Gastos com educação (% do PIB)	-0,233	0,970	0,070	-0,027
Usuários de Internet (%)	-0,581	-0,105	-0,165	0,790

De acordo com os resultados acima, o índice PC1 sozinho é capaz de explicar 64,6% do comportamento das quatro variáveis sócio-econômicas. Já os índices PC1 e PC2 somados, são capazes de explicar 87,5%. Ao agregarmos o PC3, este valor passa a 95,7%, e assim por diante. Desta forma, notamos claramente que apenas o indicador PC1 e PC2 já são satisfatoriamente suficientes para representar todas as variáveis. O primeiro índice poderia ser batizado como “Desenvolvimento Humano”, pois tem proporções iguais para as variáveis para o desenvolvimento da sociedade; o segundo como “Investimento em Educação”, pois tem uma grande proporção para a variável que se refere aos gastos com educação. O gráfico abaixo demonstra de forma visual que a contribuição do PC1 e PC2 é grande e que a contribuição marginal vai diminuindo à medida que se agregam novos índices.



## 8. CONSIDERAÇÕES FINAIS

O objetivo deste artigo foi realizar análises comparativas dos dados do IDH dos países incluídos no Relatório de Desenvolvimento Humano (RDH) de 2007/2008 e verificar suas relações com os quatro indicadores sócio-econômico divulgados pelo mesmo. As principais evidências encontradas estão resumidas abaixo:

- O continente europeu é o que apresentou maior IDH médio e foi um dos mais homogêneos. Os IDHs médio dos continentes sul, centro e norte americano foram bem próximos, contudo o sul americano foi mais homogêneo que o centro e norte, talvez pelo fato do último apresentar um outlier, o Haiti. A Oceania foi o continente que apresentou IDH mais heterogêneo.
- Assim como se destacou em relação ao IDH, a Europa também teve a maior média com gastos com saúde per capita em relação aos outros continentes. Seu desvio relativo foi alto, mas o relativo foi o mais baixo entre os continentes. O continente centro norte americano e a Oceania apresentaram médias e desvios semelhantes, e ambos apresentaram desvios relativo alto. No continente sul americano o Brasil se destacou positivamente, seus gastos com saúde per capita no ano foram bem mais altos que os dos demais países sul americanos. O gasto médio do continente africano é o mais baixo entre os continentes, seguido pelo asiático. Todavia os dois apresentaram países que se destacaram positivamente, no continente africano - Tunísia, Mauritius, Seychelles e África do sul e na Ásia - Israel e Japão.
- A variável “Gasto com educação (% do PIB)” é uma variável relativa, o montante gasto com educação depende do PIB do país, ou seja, se o país tem um PIB alto seus gastos com educação serão mais altos. Outro fator determinante que se deve considerar é a idade média da população do país. Em geral, quanto mais envelhecida é a população, menor deveriam ser os gastos, pois menor seria o número de habitantes em idade de formação. Talvez por esse motivo a Europa não teve uma média superior com gastos em educação comparado aos outros continentes. Um fato interessante foi a constatação de que em alguns países africanos e asiáticos, como Lesotho e Iêmen, houve uma vontade política maior de se investir em educação.
- O continente africano foi o que apresentou menor percentual médio de usuários de internet, enquanto a Europa, novamente, apresentou a maior média. A Ásia teve um percentual médio de usuários de internet de 15,31% da população, porém apresentou quatro outliers: República da Coreia (68,4%), Japão (66,8%), Singapura (57,1%) e Hong Kong (50,8%). Seu desvio absoluto e relativo foram altos.

Devemos ressaltar que o trabalho baseou-se em técnicas estatísticas, às quais possuem limitações e demandam mais rigor quanto a sua interpretação e extrapolação das conclusões. Outro fator limitante foi a questão do tempo, uma vez que o trabalho foi pontual (dados referentes a 2004/2005). Seria interessante uma análise temporal para verificar o comportamento dos países no decorrer dos anos. Desta forma, as evidências encontradas não podem ser interpretadas como respostas definitivas às questões levantadas neste trabalho, mas, sem dúvida, contribuíram para uma análise exploratória do tema e para o desenvolvimento de pesquisas futuras.

