



**PONTÍFICA UNIVERSIDADE CATÓLICA DE  
SÃO PAULO**

**Faculdade de Economia, Administração,  
Contabilidade e Atuariais**

**UMA PESQUISA DAS RELAÇÕES ENTRE INDICADORES  
SOCIAIS, ECONOMICOS E TECNOLOGICOS DO BANCO  
MUNDIAL**

**Aluna: Maria Teresa Stefani**

**Prof. Arnaldo José de Hoyos Guevara**

**1º Semestre 2012**

## Sumário

1.	INTRODUÇÃO.....	4
2.	ENTENDENDO OS DADOS.....	4
2.1	Os Indivíduos.....	4
2.2	As Variáveis.....	4
2.3	A Tabela de Dados.....	5
2.4	A Fonte dos Dados.....	9
3.	ANÁLISE EXPLORATÓRIA.....	9
3.1	Variável: “Crescimento Populacional”.....	9
3.2	Variável: “Usuários de Internet”.....	10
3.3	Variável: “Assinaturas de Celular”.....	11
3.4	Variável: “Razão de dependência da idade”.....	12
3.5	Variável: “Taxa de Fertilidade”.....	13
3.6	Variável: “PIB Per Capita”.....	14
3.7	Variável: “% Crescimento PIB Per Capita”.....	15
3.8	Variável: “Expectativa de vida ao nascer”.....	16
3.9	Variável: “População Urbana”.....	17
3.10	Análise das médias das variáveis.....	18
3.11	Gráfico RADAR para as médias.....	19
4.	CORRELAÇÕES.....	19
4.1	Coefficientes de correlação entre variáveis quantitativas.....	19
4.2	Dendograma.....	22
5.	REGRESSÕES.....	22
5.1	Regressão linear simples.....	22
5.1.1	Regressão da variável Crescimento Populacional x Usuários de Internet.....	23
5.1.2	Regressão da variável Crescimento Populacional x Assinaturas de celular.....	24
5.1.3	Regressão da variável Crescimento Populacional x Razão de dependência da idade.....	25
5.1.4	Regressão da variável Crescimento Populacional x Taxa de Fertilidade.....	26
5.1.5	Regressão da variável Crescimento Populacional x PIB per capita.....	27
5.1.6	Regressão da variável Crescimento Populacional x % Crescimento PIB per capita.....	28
5.1.7	Regressão da variável Crescimento Populacional x Expectativa de vida ao nascer.....	30
5.1.8	Regressão da variável Crescimento Populacional x População Urbana.....	31
5.1.9	Regressão da variável Usuários de Internet x % Assinaturas de celular.....	32
5.1.10	Regressão da variável Usuários de Internet x Razão de dependência da idade.....	33
5.1.11	Regressão da variável Usuários de Internet x Taxa de Fertilidade.....	34
5.1.12	Regressão da variável Usuários de Internet x PIB Per Capita.....	35
5.1.13	Regressão da variável Usuários de Internet x % Crescimento do PIB Per Capita.....	36
5.1.14	Regressão da variável Usuários de Internet x Expectativa de vida ao nascer.....	37
5.1.15	Regressão da variável Usuários de Internet x População Urbana.....	38
5.1.16	Regressão da variável Razão de dependência da idade x Taxa de Fertilidade.....	39
5.1.17	Regressão da variável Razão de dependência da idade x PIB Per Capita.....	41
5.1.18	Regressão da variável Razão de dependência da idade x % Crescimento do PIB Per Capita.....	42

5.1.19	Regressão da variável Razão de dependência da idade x Expectativa de Vida ao nascer .....	43
5.1.20	Regressão da variável Razão de dependência da idade x População Urbana .....	44
5.1.21	Regressão da variável Taxa de Fertilidade x PIB per Capita .....	45
5.1.22	Regressão da variável Taxa de Fertilidade x % Crescimento do PIB per Capita .....	46
5.1.23	Regressão da variável Taxa de Fertilidade x Expectativa de Vida ao nascer .....	47
5.1.24	Regressão da variável Taxa de Fertilidade x População Urbana .....	48
5.1.25	Regressão da variável PIB per Capita x % Crescimento do PIB per Capita .....	49
5.1.26	Regressão da variável PIB per Capita x Expectativa de Vida ao nascer .....	50
5.1.27	Regressão da variável PIB per Capita x População Urbana .....	51
5.1.28	Regressão da variável % Crescimento do PIB per Capita x Expectativa de Vida ao nascer .....	52
5.1.29	Regressão da variável % Crescimento do PIB per Capita x População Urbana .....	53
5.1.30	Regressão da variável Expectativa de Vida ao nascer x População Urbana .....	54
5.2	Regressão Linear Múltipla .....	55
5.2.1	Regressão linear da variável Crescimento Populacional frente às demais .....	55
5.2.2	Regressão linear da variável Usuários de Internet frente às demais .....	57
5.2.3	Regressão linear da variável Assinaturas de celular frente às demais .....	58
5.2.4	Regressão linear da variável Razão de dependência da idade frente às demais .....	59
5.2.5	Regressão linear da variável Taxa de Fertilidade frente às demais .....	59
5.2.6	Regressão linear da variável PIB per Capita frente às demais .....	60
5.2.7	Regressão linear da variável % Crescimento do PIB per Capita frente às demais .....	61
5.2.8	Regressão linear da variável Expectativa de vida ao nascer frente às demais .....	61
5.2.9	Regressão linear da variável População Urbana frente às demais .....	62
6.	ANÁLISE DE VARIÂNCIA .....	62
7.	AMOSTRAS .....	64
7.1	Análise Exploratória das Amostras .....	64
7.1.1	ANOVA para a variável Crescimento Populacional .....	64
7.1.2	ANOVA para a variável Usuários de Internet .....	65
7.1.3	ANOVA para a variável Assinaturas de celular .....	66
7.1.4	ANOVA para a variável Razão de dependência da idade .....	67
7.1.5	ANOVA para a variável Taxa de Fertilidade .....	68
7.1.6	ANOVA para a variável PIB per Capita .....	69
7.1.7	ANOVA para a variável % Crescimento do PIB per Capita .....	70
7.1.8	ANOVA para a variável Expectativa de Vida ao nascer .....	71
7.1.9	ANOVA para a variável População Urbana .....	72
8.	ANÁLISE DE CLUSTERS e DENDOGRAMA POR PAÍS .....	73
8.1	Análise de clusters para todos os países .....	73
8.2	Análise de clusters para 30 países e 3 clusters .....	77
8.3	Análise de clusters para 30 países e 2 clusters .....	79
9.	ANÁLISE DE COMPONENTES .....	81
9.1	Análise de Componentes Principais .....	81
9.2	Análise Fatorial .....	84
9.2.1	Rotações .....	86

10.	ANÁLISE DISCRIMINANTE .....	90
11.	REGRESSÃO LOGÍSTICA.....	95
12.	ANÁLISE DE CORRESPONDÊNCIA .....	96
12.1	Análise de correspondência para 10 países.....	96
13.	CONSIDERAÇÕES FINAIS .....	98

## 1. **INTRODUÇÃO**

O presente trabalho tem por objetivo efetuar diversas análises de algumas variáveis fornecidas pelo Banco Mundial, verificando se há correlação entre indicadores sociais, econômicos e tecnológicos.

O trabalho foi estruturado em 11 etapas, brevemente descrito a seguir:

Etapa 1 – Introdução. Uma breve descrição do trabalho e seus objetivos.

Etapa 2 – Os dados. São descritas as variáveis, as fontes, os indivíduos, de modo a compreender os dados que serão analisados.

Etapa 3 – Análise Exploratória. São analisadas as variáveis de maneira isolada de modo a compreender seu comportamento.

Etapa 4 – Regressão. São feitas análises de regressão para investigar e modelar o relacionamento entre uma variável resposta e uma ou mais preditoras.

Etapa 5 – Correlação entre variáveis. É verificado se há correlação entre as variáveis, ou seja, se variáveis influenciam no comportamento de outras.

Etapa 6 – Amostras. São feitos exercícios com diversos tamanhos de amostras, de modo a identificar o grau de confiabilidade ao se definir universos menores para análise.

Etapa 7 – Análise de Componentes. São feitas análises para ajudar a selecionar um menor número de variáveis que não tenham correlação, de modo a não distorcer as análises.

Etapa 8 – Análises Discriminantes. São utilizadas para classificar observações em dois ou mais grupos, bem como para investigar como variáveis contribuem para separar os grupos.

Etapa 9 – Análise de Correspondência. São feitas análises para identificar o relacionamento existente entre variáveis, sendo uma delas categórica.

Etapa 10 – Regressões Logísticas. São feitas análises para investigar o relacionamento existente entre uma variável resposta e um ou mais preditoras.

Etapa 11 – Conclusões. São demonstradas as considerações finais do trabalho.

Os dados são compilados pelo Banco Mundial. O software estatístico utilizado foi o MINITAB, tendo sido utilizado também o Excel como auxiliar em alguns gráficos, como o Radar.

## 2. **ENTENDENDO OS DADOS**

### 2.1 **Os Indivíduos**

Os indivíduos desta análise são países que enviam seus dados para algum órgão que compartilha seus dados com o Banco Mundial. Os dados analisados de cada país são as variáveis descritas a seguir. Os dados foram coletados no banco de dados do Banco Mundial.

### 2.2 **As Variáveis**

São 9 as variáveis desta pesquisa, incluindo o nome dos países e seus respectivos clusters. As mesmas são explicadas na Tabela 1.

**Tabela 1. As Variáveis**

Variável	Significado	Tipo	Unidade de Medida
Crescimento Populacional	Porcentagem de crescimento da população no período	Variável Quantitativa	Porcentagem
Usuários de Internet	Quantidade de pessoas (por cada 100 habitantes) que tem acesso à Internet	Variável Quantitativa	Qtde
Assinaturas de celular	Corresponde à quantidade de assinantes de telefone celular incluindo pré e pós-pagos.	Variável Quantitativa	Qtde
Razão de dependência da idade	Taxa de dependência dos dependentes – menores de 15 anos e maiores de 64 anos – da população ativa – de 15 a 64 anos. Os dados são demonstrados como a proporção de dependentes por 100 habitantes	Variável Quantitativa	Porcentagem
Taxa de fertilidade	Taxa de fertilidade feminina	Variável Quantitativa	Porcentagem
PIB Per Capita	PIB Per Capita em Dólares	Variável Quantitativa	Moeda
% Crescimento do PIB Per Capita	Porcentagem anual de crescimento do PIB Per Capita	Variável Quantitativa	Porcentagem
Expectativa de Vida ao nascer	Expectativa de vida ao nascer	Variável Quantitativa	Número, com 2 casas decimais
População Urbana	Porcentagem do total da população que vive em cidades	Variável Quantitativa	Porcentagem
Cluster	Grupo de países onde determinado país está inserido. Foi montado a partir do valor da variável PIB per Capita: 1 – PIB <= 1000 – Países muito pobres 2 – 1000 < PIB < 3000 – Países pobres 3 – 3000 < PIB < 15000 – Países emergentes e de renda média 4 – 15000 < PIB < 30000 – Países ricos 5 – PIB > 30000 – Países muito ricos	Variável Quantitativa	Número natural, 1 dígito

### 2.3 A Tabela de Dados

**Tabela 2. Os dados**

**Todos os dados são do ano de 2010**

Cluster	Country Name	Population growth (annual %)	Internet users (per 100 people)	Mobile cellular subscriptions	Age dependency ratio (% of work)	Fertility rate, total (births p)	GDP per capita (current US\$)	GDP per capita growth (annual %)	Life expectancy at birth, total	Urban population (% of total)
5	Afghanistan	2,83130	3,654122	37,80718	94,85494	6,288	501,4719	5,22109	48,2822	24,8
3	Albania	0,36145	44,98995	141,8973	47,73985	1,536	3677,41	3,103535	76,90095	48
3	Algeria	1,47135	12,50007	92,4218	46,28614	2,264	4566,917	1,791822	72,85254	66,5
3	Angola	2,79954	9,999954	46,68903	96,28639	5,443	4451,159	2,959536	50,65366	58,5
3	Argentina	0,86961	36,00033	141,7896	54,92045	2,211	9124,42	8,216766	75,63215	92,4
3	Armenia	0,22966	44,00102	125,0114	45,55294	1,736	3030,785	1,864536	73,78356	63,7
1	Australia	1,57331	75,89565	100,9014	47,99575	1,92	50747,7	0,664317	81,69512	89,1
1	Austria	0,29240	72,73158	145,8999	47,73807	1,44	45181,08	2,013155	80,38293	67,6
3	Azerbaijan	1,18979	46,6797	100,5093	37,83866	2,3	5718,381	3,76193	70,50651	52,2
2	Bahamas, The	1,32673	42,98458	124,8913	41,39344	1,896	22453,53	-0,416373	75,22212	84,1
5	Bangladesh	1,12403	3,700003	46,16926	55,97713	2,245	674,9322	4,883857	68,6348	28,1
3	Belarus	-0,17898	32,05214	108,882	39,95871	1,44	5765,345	7,803138	70,40488	74,3
1	Belgium	0,91547	73,73393	111,5459	52,19274	1,84	43077,66	1,331616	79,93659	97,4

3	Belize	3,39316	12,64573	56,29014	63,50711	2,79	4060,87	-0,621526	75,83995	52,7
5	Benin	2,84371	3,129962	79,94253	87,73865	5,287	749,4978	0,111007	55,58559	42
4	Bhutan	1,70537	13,59888	54,3135	51,88285	2,399	2088,262	5,617933	66,90885	36,8
4	Bolivia	1,58767	20,00171	72,3063	68,75106	3,348	1979,024	2,495356	66,26856	66,5
3	Bosnia and Herzegovina	-0,20016	52,00206	82,71896	41,01313	1,148	4409,013	1,00597	75,40044	48,6
3	Botswana	1,27212	5,999836	117,7584	57,7498	2,75	7403,425	5,865514	53,10951	61,1
3	Brazil	0,87579	40,6501	104,1027	48,04903	1,83	10710,09	6,55263	73,09954	86,5
3	Bulgaria	-0,67254	45,98659	135,3855	45,45279	1,49	6333,221	0,880027	73,5122	71,7
5	Burkina Faso	2,98443	1,400061	34,66025	90,66914	5,85	535,6031	6,034697	54,9242	20,4
5	Burundi	2,56145	2,100213	13,72584	68,71981	4,338	192,1433	1,282704	49,87722	11
5	Cambodia	1,14065	1,259934	57,64739	55,51155	2,581	795,1246	4,755756	62,53622	22,8
4	Cameroon	2,18641	3,999977	44,0668	78,87195	4,487	1147,015	0,967544	51,06276	58,4
1	Canada	1,16864	81,33839	70,43712	43,94348	1,677	46212,27	2,016299	80,7978	80,6
3	Cape Verde	0,88658	29,99994	74,97399	60,91205	2,405	3322,768	4,479296	73,77405	61,1
5	Central African Republic	1,90214	2,300027	22,25022	79,71429	4,631	457,3994	1,354834	47,61846	38,9
5	Chad	2,61804	1,700031	23,82943	93,51948	5,981	675,8416	1,606697	49,19483	27,6
3	Chile	0,92724	45	116,0021	45,71312	1,862	12431,03	4,227209	78,88573	89
3	China	0,51838	34,37779	64,18613	38,20616	1,598	4428,463	9,82915	73,2731	44,9
3	Colombia	1,39383	36,49987	96,07442	52,34303	2,1	6240,103	2,561285	73,42968	75,1
5	Comoros	2,61658	5,098265	22,4868	82,58706	4,919	736,1871	-0,570714	60,62627	28,2
5	Congo, Dem. Rep.	2,70661	0,720009	17,91912	95,86698	5,775	199,2685	4,338685	48,06959	35,2
4	Congo, Rep.	2,54123	4,999875	93,95545	79,4405	4,544	2970,042	6,019898	56,9602	62,1
3	Costa Rica	1,47245	36,49911	65,14288	45,80726	1,848	7690,806	2,701412	79,19261	64,3
4	Cote d'Ivoire	1,98418	2,599974	76,13137	80,8926	4,432	1154,133	0,985215	54,74156	50,1
3	Croatia	-0,25043	60,11971	144,0042	47,53934	1,46	13773,62	-0,943514	76,47561	57,8
2	Cyprus	1,20086	53,02474	93,75077	41,35723	1,476	28779,19	0,569304	79,38039	70,3
2	Czech Republic	0,31051	68,63813	136,8152	40,55823	1,49	18254	2,027177	77,42439	73,5
1	Denmark	0,44420	88,77025	124,7966	52,58384	1,87	56244,59	0,859076	79,1	87,2
3	Dominican Republic	1,32294	39,53127	89,58178	59,4955	2,58	5214,705	6,338548	73,20002	70,5
3	Ecuador	1,41457	28,99948	102,1827	57,68644	2,479	4008,166	2,124843	75,46229	66,9
4	Egypt, Arab Rep.	1,74700	26,74003	87,10569	57,64313	2,733	2698,368	3,325763	72,97527	42,8
3	El Salvador	0,52730	15,89998	124,3394	63,80952	2,25	3425,593	0,892678	71,73237	61,3
5	Eritrea	3,00801	5,399667	3,526361	78,7951	4,453	402,9326	-0,831374	60,9942	21,6
3	Estonia	-0,00821	74,16304	123,344	48,39424	1,63	14340,72	3,125463	75,42927	69,5
5	Ethiopia	2,14681	0,749996	8,262809	81,16632	4,193	358,2521	7,797117	58,7151	17,6
3	Fiji	0,96910	14,83074	81,15349	51,05634	2,668	3708,484	-0,576946	69,22583	53,4
1	Finland	0,45749	86,89884	156,4131	51,01351	1,87	44377,53	3,245568	79,87073	63,9
1	France	0,54568	77,49862	97,38809	54,22677	2	39448,37	0,928943	81,36829	77,8
3	Gabon	1,87395	7,232224	106,9767	66,37168	3,25	8729,026	3,810739	62,28668	86
5	Gambia, The	2,73672	9,196775	85,50301	85,51502	4,896	466,4684	2,141879	58,16002	58,1
4	Georgia	0,94770	26,29725	89,39802	44,82759	1,555	2620,705	5,383325	73,32734	52,9
1	Germany	-0,15320	82,5269	127,8599	51,18975	1,39	40115,56	3,849245	79,9878	73,8

4	Ghana	2,35376	9,549931	71,48634	73,6033	4,17	1324,592	5,215551	63,83727	51,5
2	Greece	0,28983	44,57007	108,6313	49,52431	1,44	26606,86	-3,800131	80,3878	61,4
3	Grenada	0,37395	33,61668	117,2558	52,94118	2,235	7435,297	-0,736448	75,66044	31
4	Guatemala	2,50030	10,49995	125,5679	84,55816	3,983	2862,353	0,244511	70,82541	49,5
5	Guinea	2,23253	0,999959	40,07213	85,96982	5,246	451,8724	-0,321905	53,63859	35,4
5	Guinea-Bissau	2,07413	2,450362	39,21452	80,47619	5,063	579,8796	1,36346	47,70066	30
4	Guyana	0,19635	29,87992	73,56252	60,50955	2,262	2947,835	3,327347	69,54915	28,5
5	Haiti	1,29934	8,370207	40,02802	67,63966	3,34	671,4435	-6,275131	61,763	49,6
4	Honduras	2,00135	11,09076	125,0667	69,68073	3,139	2026,358	0,743794	72,82593	48,8
3	Hungary	-0,22601	65,16325	120,1182	45,29343	1,25	12863,16	1,487503	74,20732	68,3
2	Iceland	-0,14390	95,63811	107,2569	48,84793	2,2	39541,84	-3,849242	81,45122	92,3
3	India	1,38748	7,499996	61,42257	55,06363	2,625	1410,33	7,307105	65,13134	30,1
3	Indonesia	1,02935	9,900039	91,71635	48,34415	2,117	2945,588	5,018685	68,88966	53,7
4	Iraq	3,02360	2,471948	74,92741	86,66667	4,702	2564,713	-2,115881	68,48605	66,4
1	Ireland	0,34509	69,77039	105,0609	48,93333	2,07	46170,26	-0,78665	80,29195	61,9
2	Israel	1,82675	65,38779	129,5252	60,38292	3,03	28506,39	2,756714	81,50488	91,7
2	Italy	0,48176	53,74022	149,8024	52,46294	1,4	34075,11	1,050097	81,73659	68,4
3	Jamaica	0,24824	26,47722	117,7644	58,41642	2,329	5274,622	-0,83544	72,84712	53,7
1	Japan	-0,08431	77,63854	94,70986	56,36364	1,39	42830,87	4,087278	82,93268	66,8
3	Jordan	2,20708	38,8812	109,4758	70,77097	3,8	4559,87	0,856232	73,28966	78,5
3	Kazakhstan	1,41785	33,38213	118,8666	45,48128	2,59	9131,833	5,791246	68,29537	58,5
5	Kenya	2,62721	25,8998	61,6318	82,17915	4,718	794,761	2,813812	56,49707	22,2
2	Korea, Rep.	0,26224	82,51593	103,8716	38,08273	1,22	20756,69	5,883813	80,76195	81,9
5	Kyrgyz Republic	1,19472	19,58232	96,83328	52,61978	2,898	847,3137	-2,531075	69,36846	36,6
4	Lao PDR	1,44160	6,99988	64,56047	62,31353	2,747	1176,643	6,899196	67,064	33,2
2	Latvia	-0,70434	71,51472	102,9969	46,38436	1,17	10723,39	0,361258	73,48293	68,2
3	Lebanon	0,72661	31,00438	68,0096	47,2493	1,8	9227,874	6,240343	72,40876	87,2
4	Lesotho	1,02382	3,860565	45,48356	71,56398	3,199	1003,847	4,548291	47,36507	26,9
5	Liberia	4,04121	7,000214	39,34171	86,20047	5,238	246,9208	1,338738	56,14759	61,5
3	Lithuania	-1,58874	62,8119	148,7977	44,76232	1,55	11045,45	2,947283	73,26829	67,2
1	Luxembourg	1,82541	90,07953	143,3925	46,10951	1,63	105194,6	0,811434	80,0878	82,2
3	Macedonia, FYR	0,18429	51,91418	104,5352	41,64948	1,422	4460,9	1,640332	74,61885	67,9
5	Madagascar	2,88805	1,699985	37,22951	86,00808	4,651	420,9976	-1,326489	66,46707	30,2
5	Malawi	3,12569	2,259976	20,38433	95,7567	5,99	339,182	3,803051	53,46263	19,8
3	Malaysia	1,60294	56,30003	119,2176	54,10156	2,635	8372,836	5,489429	74,02456	72,2
3	Maldives	1,32119	28,2897	156,4402	46,2963	1,752	6039,433	8,379626	76,55141	40,5
5	Mali	3,03855	2,699966	48,40848	97,48169	6,294	601,9121	1,371207	50,95483	33,3
2	Malta	0,48290	63,07799	109,4745	41,01695	1,38	19845,26	2,646879	80,94878	94,7
4	Mauritania	2,40287	2,999803	79,33462	74,21954	4,533	1044,479	2,681813	58,21695	41,4
3	Mauritius	0,46104	28,73188	92,96643	40,32967	1,47	7590,833	3,676257	72,96732	42,6
3	Mexico	1,23278	31,05001	80,55046	54,91771	2,32	9132,812	4,226393	76,68378	77,8
4	Micronesia, Fed. Sts.	0,34996	20,01153	24,79099	71,875	3,451	2679,743	2,824022	68,76483	22,7
4	Moldova	-0,09936	40,12223	88,85604	38,51823	1,475	1630,768	7,054635	68,90371	41,2
4	Mongolia	1,62201	12,9	91,09107	46,46465	2,504	2249,767	4,653864	68,19498	57,5

3	Montenegro	0,16720	51,95804	185,1266	46,97674	1,656	6504,852	2,246183	74,3111	59,5
4	Morocco	0,99673	49,00063	100,0979	50,37421	2,279	2795,542	2,61175	71,86463	56,7
5	Mozambique	2,30136	4,170136	30,88575	90,09346	4,912	409,8412	4,76461	49,69693	38,4
3	Namibia	1,82138	6,500823	67,21542	67,00805	3,217	5330,851	2,921496	62,0701	38
5	Nepal	1,77342	7,930096	30,69382	67,67405	2,727	524,7781	2,716724	68,39483	18,2
1	Netherlands	0,51292	90,70356	115,4249	49,28109	1,79	46903,97	1,165628	80,70244	82,9
4	Nicaragua	1,35557	9,998554	65,13232	64,12365	2,622	1131,66	6,105947	73,72922	57,3
5	Niger	3,54119	0,829997	24,53321	104,8053	7,063	357,7111	5,019754	54,26563	16,7
4	Nigeria	2,51529	28,43003	55,10424	85,91311	5,525	1278,368	6,000097	51,41002	49,8
1	Norway	1,24567	93,27751	115,5384	50,10766	1,95	85388,56	-0,56467	80,99756	77,6
4	Pakistan	1,80134	16,78004	57,13701	65,76208	3,423	1018,875	2,285268	65,19885	37
3	Panama	1,57393	42,74781	184,7073	55,11464	2,484	7588,506	3,188736	75,97424	74,8
4	Papua New Guinea	2,28456	1,28005	27,83724	71,92279	3,951	1382,334	5,56474	62,44061	12,5
4	Paraguay	1,76079	19,80168	91,73936	63,02096	2,954	2840,291	13,04747	72,277	61,5
3	Peru	1,07657	34,3006	100,1346	56,38378	2,501	5401,465	7,622003	73,76498	71,6
4	Philippines	1,68438	24,99995	85,66887	64,14052	3,142	2140,117	5,834284	68,48432	66,4
3	Poland	0,08405	62,47123	122,9628	39,67737	1,38	12294,16	3,855857	76,24634	61,2
2	Portugal	0,04574	51,28048	142,836	49,39624	1,32	21486,29	1,33058	79,02683	60,7
3	Romania	-0,19758	40,02009	114,9361	43,09551	1,38	7539,12	1,14755	73,45854	54,6
3	Russian Federation	-0,07052	43,3665	167,682	38,57968	1,54	10439,64	4,107671	68,80488	72,8
5	Rwanda	2,98781	13,00001	33,40325	82,81142	5,371	529,7127	4,335661	55,05712	18,9
3	Samoa	0,37211	6,965038	90,97826	75	3,86	3116,899	0,827161	72,30639	23,4
4	Sao Tome and Principe	1,75783	18,79511	62,11939	79,34783	3,657	1192,767	2,926167	64,34929	62,2
2	Saudi Arabia	2,35549	41,00013	187,862	49,93992	2,811	15835,99	1,346771	73,85041	83,6
4	Senegal	2,66401	15,99965	67,10405	85,55224	4,819	1033,883	1,39397	58,95407	42,9
3	Serbia	-0,40201	43,05507	135,9943	46,98649	1,4	5269,955	1,367024	73,93659	52,4
3	Seychelles	-0,88942	40,77285	135,1575	39,0625	2,5	10765,62	6,563766	73,03415	55,3
1	Singapore	1,77066	71,13171	145,452	35,90909	1,15	41119,76	12,45557	81,64146	100
2	Slovak Republic	0,21217	79,88978	109,1162	37,31041	1,4	16071,47	4,024508	75,1122	56,8
2	Slovenia	0,43608	69,33997	103,5603	43,15642	1,57	22893,28	0,917864	79,42195	48
4	Solomon Islands	2,64034	5,001375	5,576208	75	4,229	1261,386	4,240474	67,4652	18,6
3	South Africa	1,35163	12,33489	100,7621	53,29163	2,458	7279,519	1,468979	52,08149	61,7
2	Spain	0,35307	65,80855	112,0033	46,91197	1,39	30548,62	-0,495756	81,62683	77,4
4	Sri Lanka	0,91891	11,99997	83,21818	49,33076	2,313	2375,443	7,02322	74,72261	15,1
3	St. Lucia	0,94120	40,06138	114,3678	48,30508	1,98	6883,96	2,133515	74,4399	28
3	Swaziland	1,14312	9,007736	69,38447	71,26623	3,364	3501,522	0,793483	48,3428	25,5
1	Sweden	0,85252	90,01619	116,0739	53,30173	1,98	48896,53	4,714809	81,45122	84,7
1	Switzerland	1,05745	82,16666	121,7352	46,92077	1,5	67457,19	1,6356	82,24634	73,6
4	Syrian Arab Republic	2,03860	20,66316	57,70632	69,02538	2,934	2892,7	1,136261	75,70256	54,9
5	Tajikistan	1,39435	11,54939	86,362	67,91992	3,297	819,9464	2,357302	67,2599	26,5
5	Tanzania	2,97985	11,00006	46,79613	91,75112	5,544	523,8029	3,976307	57,38749	26,4
3	Thailand	0,60381	21,20007	103,6199	41,68819	1,579	4608,117	7,161862	73,92766	34
5	Timor-Leste	2,20021	0,210066	53,43416	96,33508	5,578	623,6655	5,117271	61,99985	28,1



5	Togo	2,11144	5,37982	40,68402	75,68513	4,072	523,1256	1,205216	56,58871	43,4
3	Tonga	0,51933	12,00669	52,21154	79,31034	3,911	3349,331	-0,939496	72,15066	25,3
2	Trinidad and Tobago	0,38210	48,51682	141,2558	37,92395	1,639	15364,6	-0,247178	69,755	13,9
3	Tunisia	1,04343	36,56255	105,3579	43,74659	2,04	4198,582	2,624563	74,6	67,3
3	Turkey	1,25330	39,82018	84,90438	47,80082	2,088	10094,08	7,648428	73,69666	69,6
3	Turkmenistan	1,24378	2,199998	63,41975	50,07447	2,399	3966,819	7,850096	64,86351	49,5
5	Uganda	3,21272	12,50026	38,3804	103,8605	6,149	508,9387	1,858113	53,61463	13,3
3	Ukraine	-0,39729	44,58535	117,546	42,14881	1,445	3006,896	4,61411	70,27561	68,1
1	United Arab Emirates	7,93295	77,99678	145,4475	21,16129	1,749	39623,07	-6,307404	76,57361	78
1	United Kingdom	0,67769	84,7317	130,3437	51,40869	1,94	36343,25	1,401555	80,40244	90,1
1	United States	0,83691	74,24757	90,15707	49,5586	2,1	47153,01	2,141812	78,24146	82,3
3	Uruguay	0,34756	47,86747	132,1763	56,94315	1,9856	11632,84	8,087598	76,23683	92,5
4	Uzbekistan	1,64484	19,44502	74,22417	50,85507	2,499	1380,955	6,728054	68,001	36,9
4	Vanuatu	2,47604	7,988367	118,875	72,46377	3,869	2911,472	0,381774	70,81949	25,6
3	Venezuela, RB	1,57296	35,85044	96,69114	53,99776	2,49	13589,77	-3,026214	74,12732	94
4	Vietnam	1,04388	27,85182	177,1581	42,05641	1,822	1224,31	5,674168	74,82824	28,8
4	Yemen, Rep.	3,05760	12,34975	46,08573	87,90719	5,2	1300,048	4,761248	65,03046	31,8
4	Zambia	1,58024	10,12499	42,13654	97,75092	6,258	1252,638	5,918196	48,45549	35,7
5	Zimbabwe	0,77829	11,50042	61,25209	75,67077	3,29	594,7018	8,168401	49,86088	38,3

## 2.4 A Fonte dos Dados

Todos os dados desta pesquisa foram obtidos no site do próprio Banco Mundial, na opção Data Catalog:

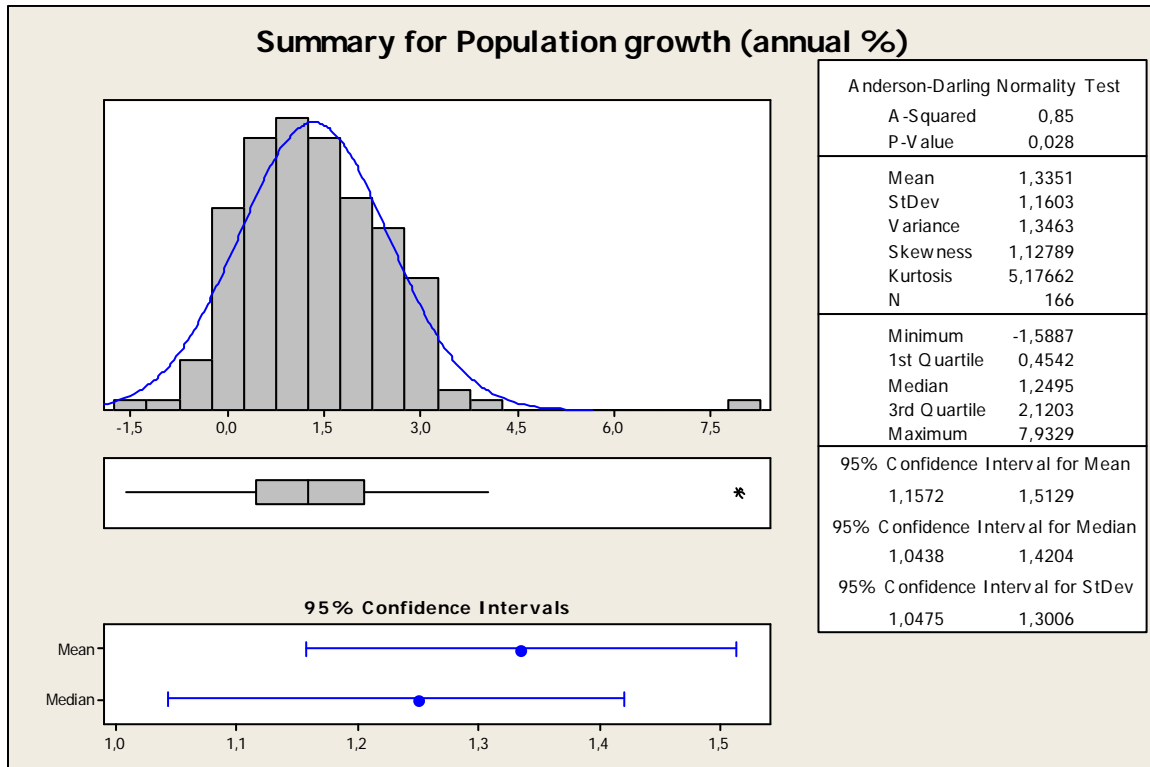
<http://data.worldbank.org/data-catalog>

## 3. ANÁLISE EXPLORATÓRIA

Este tópico exibe análise baseada em variáveis quantitativas utilizando Histograma, Box-Plot, Intervalo de Confiança, Média, Mediana, Mínimos e Máximos e Desvio-Padrão. Foram analisadas as variáveis separadamente quanto à sua forma de distribuição, os valores atípicos, medidas de centro e dispersão, sendo gerados diversos gráficos que permitissem realizar análises sobre os dados. Foram disponibilizadas também informações de medidas numéricas (média, mediana, quartis, desvio-padrão, variância, intervalo de confiança e teste de normalidade de Anderson-Darling).

### 3.1 Variável: “Crescimento Populacional”

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 o Crescimento Populacional.



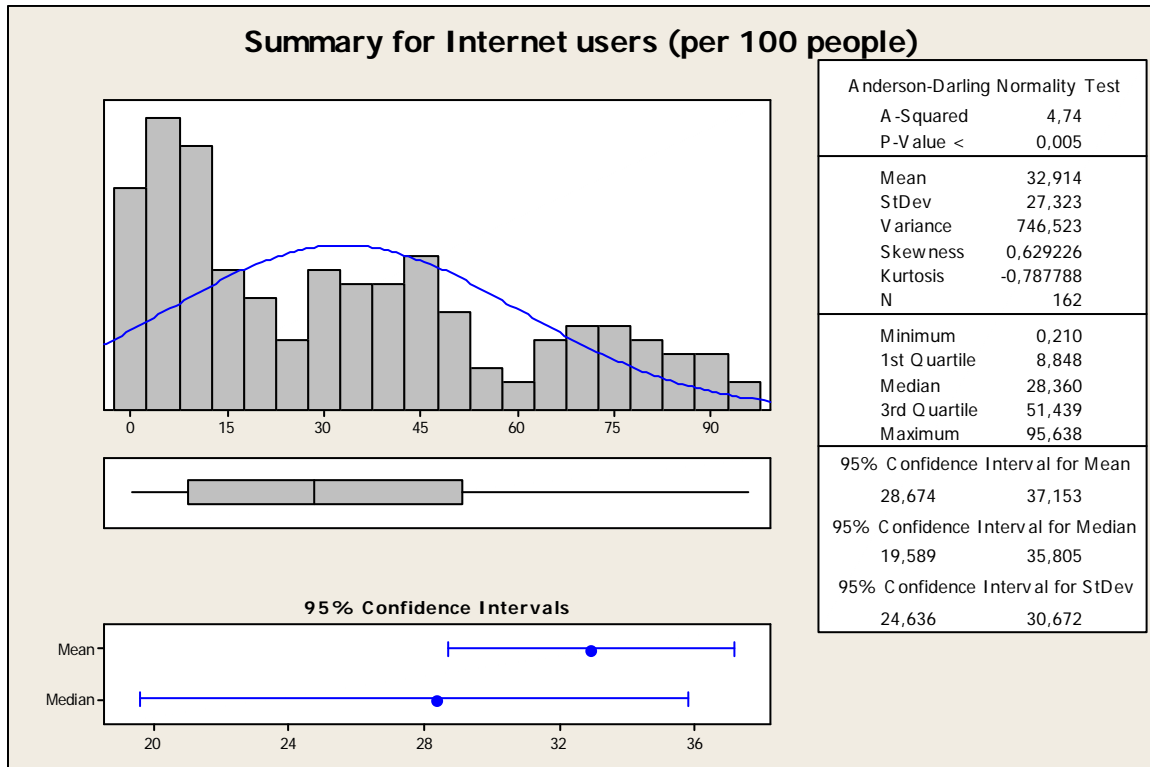
As principais observações da variável Crescimento Populacional são:

- **Forma:** O Histograma permite verificar que se trata de uma distribuição visivelmente assimétrica, revelando até valores negativos (decréscimo da população), valores positivos não muito altos, até um valor imenso, de 7,6%. A posição da linha da mediana à esquerda do centro do quadrado principal do Box-Plot nos indica também a assimetria. Considerando a variável curtose é possível afirmar que seu valor positivo indica que a curva é alongada, demonstrando valores altos (leptocúrtica).
- **Valores Atípicos:** esta variável apresentou um valor atípico, muito acima dos demais.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos países tem menos de 2 filhos (1,24) e a média está bastante próxima (1,33) da mediana.

Dado o tamanho da população mundial, acima de 7 bilhões, pode-se dizer que, quanto menor for o valor desta variável, melhor para o planeta. Até um valor negativo, por certo espaço de tempo, pode ser positivo para a sociedade.

### 3.2 Variável: “Usuários de Internet”

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 Usuários de Internet.



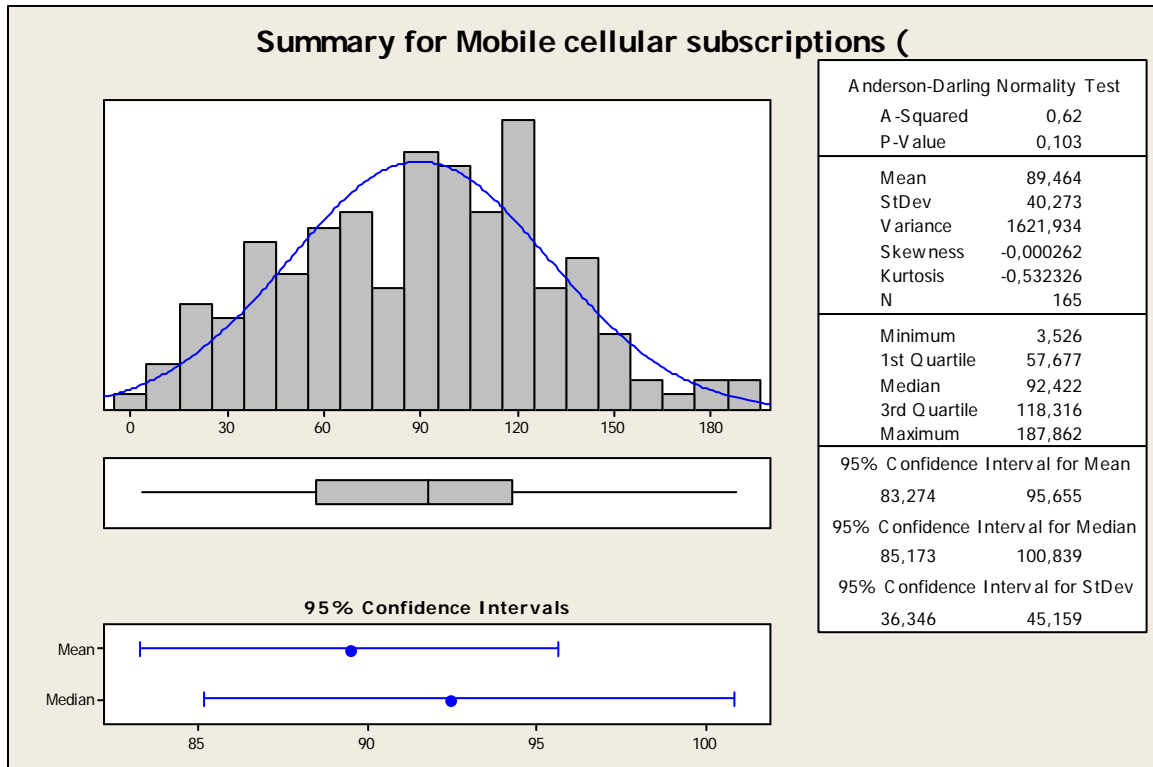
As principais observações da variável Usuários de Internet são:

- **Forma:** O Histograma permite verificar que se trata de uma distribuição visivelmente assimétrica, revelando que o acesso à Internet não é uniforme pelos países. A posição da linha da mediana à esquerda do centro do quadrado principal do Box-Plot nos indica também a assimetria. Considerando a variável curtose é possível afirmar que seu valor negativo indica que a curva é achatada, demonstrando valores próximos de zero (platicúrtica).
- **Valores Atípicos:** esta variável não apresentou valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que metade dos países tem menos de 30 pessoas com acesso à Internet, para cada 100 habitantes, o que reforça o fato da desigualdade no acesso à rede. Na distribuição desta variável média e mediana estão bastante próximas, muito baixas.

Dado que estamos na era da informação, quanto maior o número de pessoas com acesso à Internet, maior a inclusão digital dos países, o que não se demonstra no momento atual.

### 3.3 Variável: “Assinaturas de Celular”

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 Assinaturas de Celular.



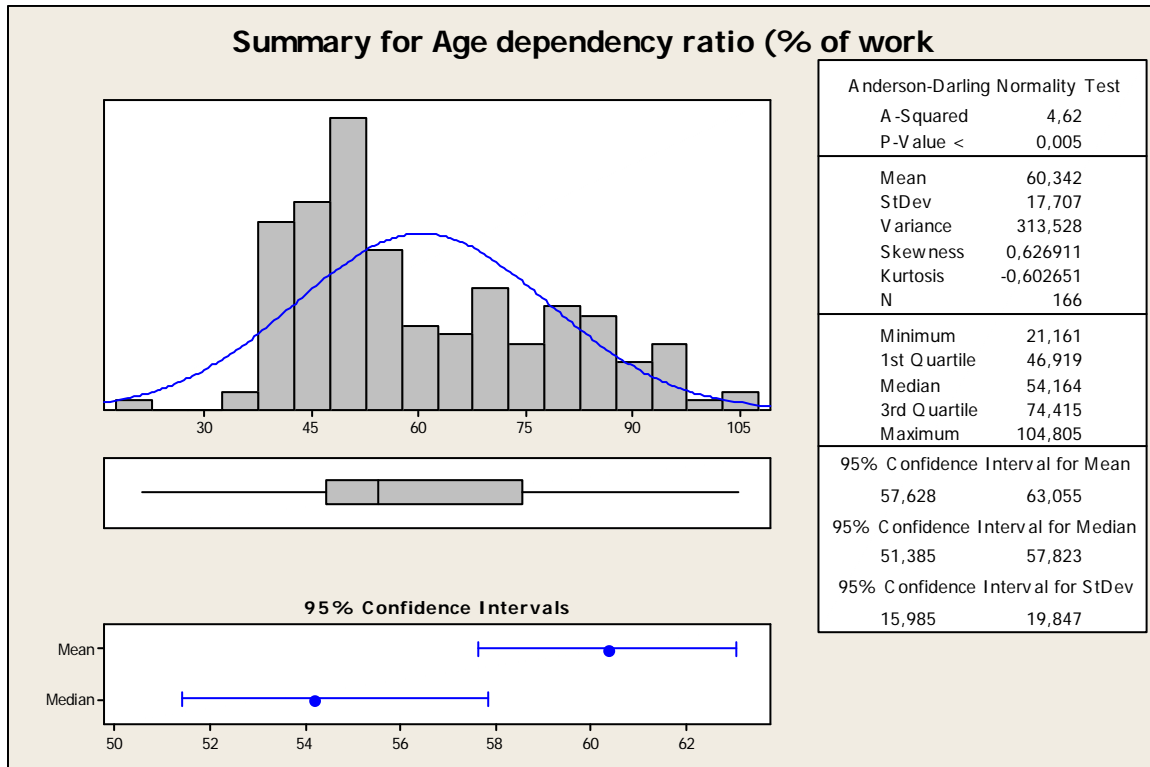
As principais observações da variável Assinaturas de Celular são:

- **Forma:** O Histograma permite verificar que se trata de uma distribuição quase simétrica, revelando também que há pouca uniformidade entre os países no aspecto ao acesso às telecomunicações. A posição da linha da mediana próxima do centro do quadrado principal do Box-Plot nos indica também a quase simetria. Considerando a variável curtose é possível afirmar que seu valor negativo, porém próximo de zero indica que a curva é bem pouco achatada, quase normal (mesocúrtica).
- **Valores Atípicos:** esta variável não apresentou valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos países tem cidadãos com acesso a Assinaturas de Celular. Na distribuição desta variável estão bastante próximas à média e a mediana.

Dado que estamos na era da informação, quanto maior o número de pessoas com acesso a serviços de telefonia, maior a inclusão digital dos países, o que não se demonstra no momento atual.

### 3.4 Variável: “Razão de dependência da idade”

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 Razão de dependência da idade.



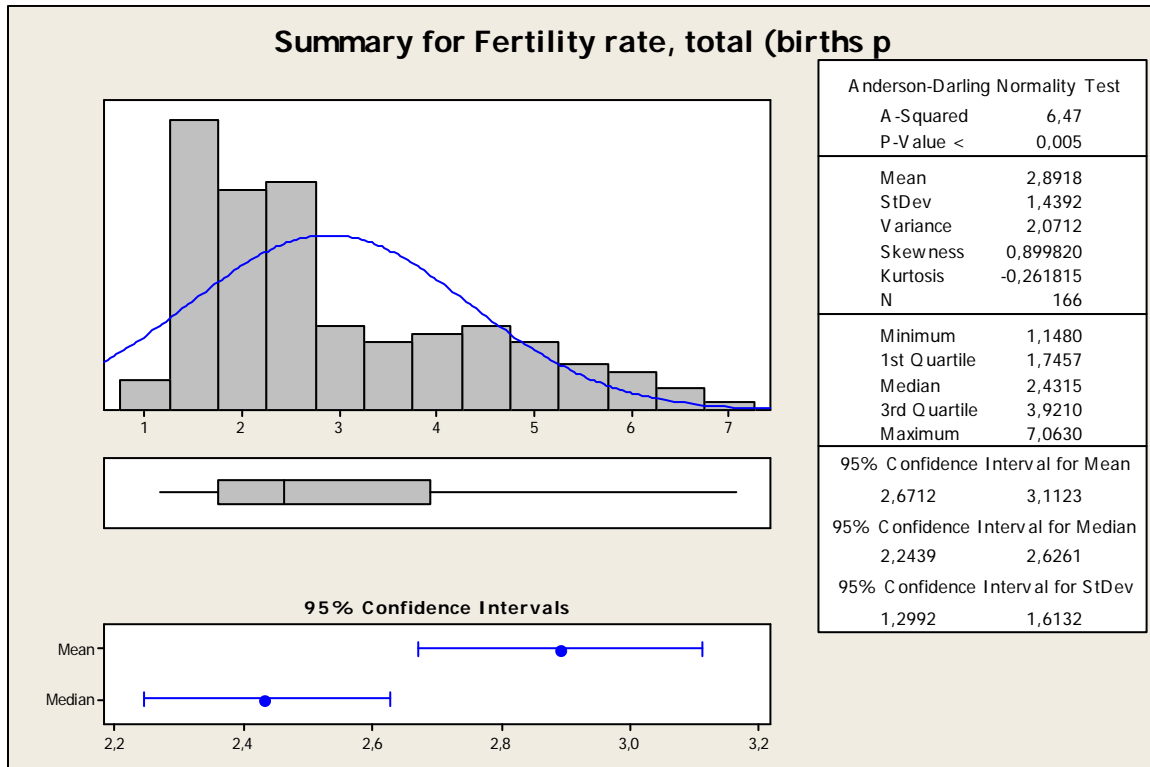
As principais observações da variável Razão de dependência da idade são:

- **Forma:** O Histograma permite verificar que se trata de uma distribuição visivelmente assimétrica, revelando que a razão de dependência da idade é alta. A posição da linha da mediana à direita do centro do quadrado principal do Box-Plot nos indica também a assimetria. Considerando a variável curtose é possível afirmar que seu valor negativo indica que a curva é achatada, mas demonstrando valores próximos de zero (mesocúrtica).
- **Valores Atípicos:** esta variável não apresentou valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos países tem razão de dependência da idade maior que 50%. Na distribuição desta variável, média e mediana estão relativamente próximas.

Essa variável indica que, quanto mais alto seu valor, maior o número de dependentes da população em idade ativa para trabalhar (15 a 64 anos), portanto, menos gente trabalhando para mais gente dependente.

### 3.5 Variável: “Taxa de Fertilidade”

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 Taxa de Fertilidade.



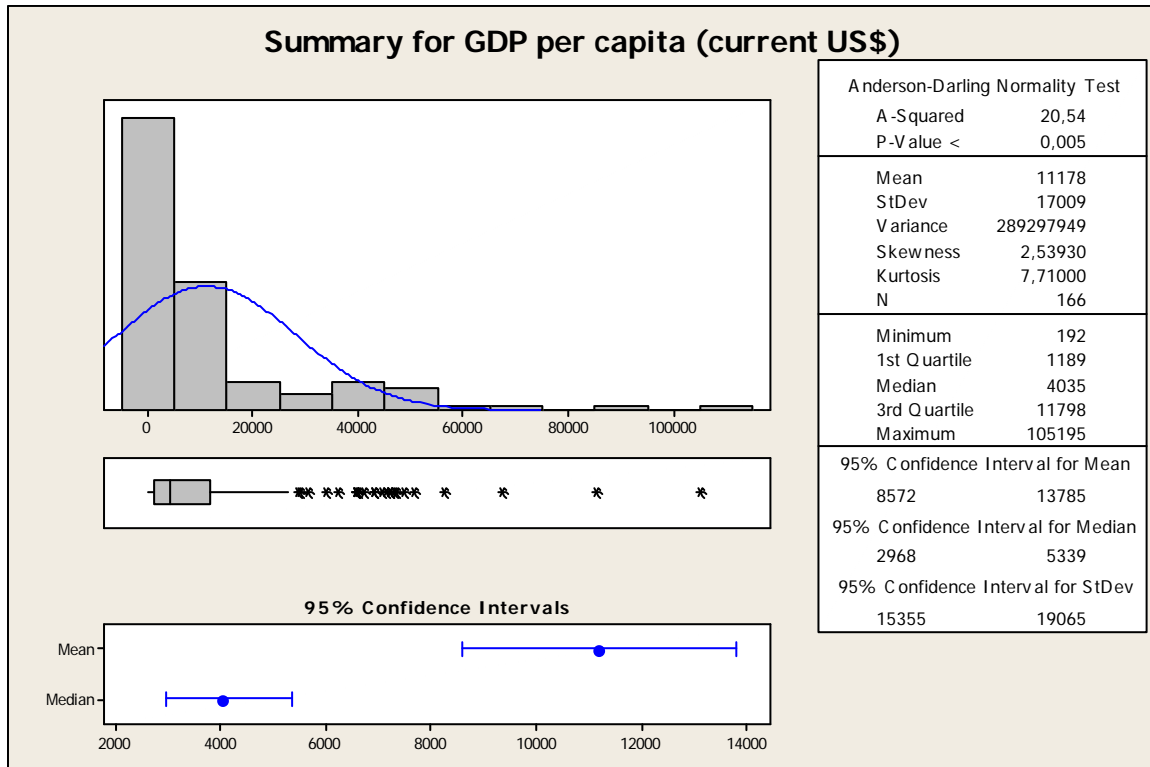
As principais observações da variável Taxa de Fertilidade são:

- **Forma:** O Histograma permite verificar que se trata de uma distribuição visivelmente assimétrica, revelando taxa de fertilidade menor que 3 para a maioria dos países. A posição da linha da mediana à esquerda do centro do quadrado principal do Box-Plot nos indica também a assimetria. Considerando a variável curtose é possível afirmar que seu valor negativo indica que a curva é achatada, demonstrando valores próximos de zero (mesocúrtica).
- **Valores Atípicos:** esta variável não apresentou valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos países tem menos de 2,5 filhos, embora haja países com taxa de fertilidade de até 7 filhos. Na distribuição desta variável média e mediana não estão muito próximas.

Dado o tamanho da população mundial, acima de 7 bilhões, pode-se dizer que, quanto menor for o valor desta variável, melhor para o planeta. Até um valor menor que um, por certo espaço de tempo, pode ser positivo para a sociedade.

### 3.6 Variável: “PIB Per Capita”

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 PIB Per Capita.



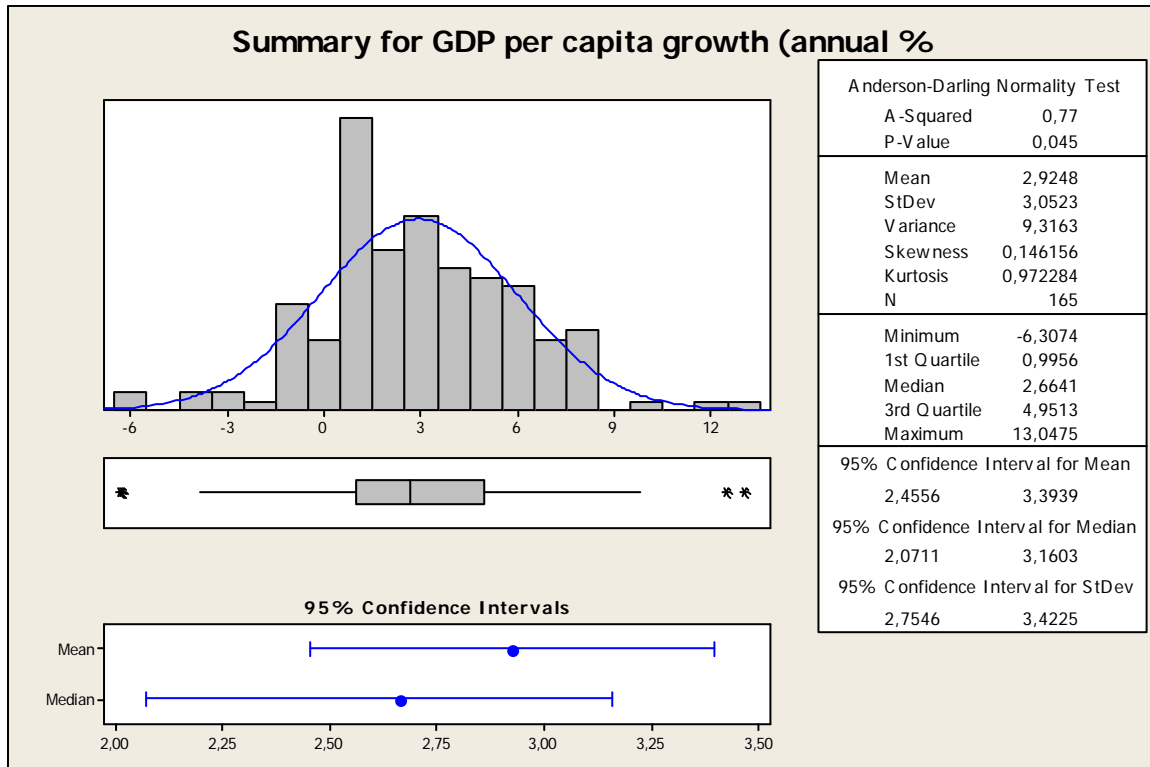
As principais observações da variável PIB Per Capita são:

- **Forma:** O Histograma permite verificar que se trata de uma distribuição visivelmente assimétrica, revelando a péssima distribuição de renda do planeta. A posição da linha da mediana acentuadamente à esquerda do centro do quadrado principal do Box-Plot nos indica também a assimetria. Considerando a variável curtose é possível afirmar que seu valor muito positivo indica que a curva é alongada, demonstrando valores próximos de zero (leptocúrtica).
- **Valores Atípicos:** esta variável apresentou muitos valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos países tem renda média menor que 4 mil dólares anuais, bastante baixa. Na distribuição desta variável média e mediana estão bastante distantes, demonstrando a imensa diferença na distribuição da riqueza econômica do planeta.

Quanto maior for o valor desta variável, melhor para os países, porém, sua distribuição muito irregular demonstra que há países muito ricos e países vivendo em condições miseráveis.

### 3.7 Variável: “% Crescimento PIB Per Capita”

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 % Crescimento PIB Per Capita.



As principais observações da variável % Crescimento PIB Per Capita são:

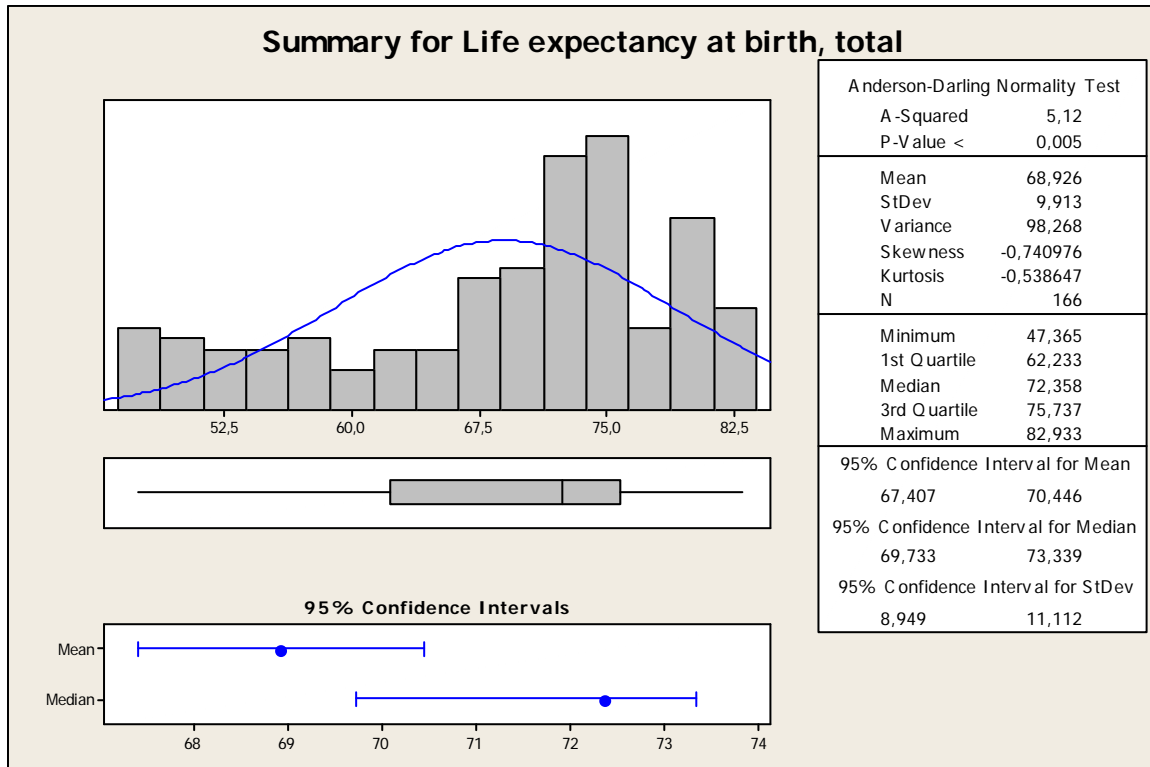
- **Forma:** O Histograma permite verificar que se trata de uma distribuição relativamente assimétrica, revelando que há países que tem sua economia encolhendo e outros com aumento significativo, acima de 10% ao ano. A posição da linha da mediana à esquerda do centro do quadrado principal do Box-Plot nos indica também a assimetria pouco acentuada. Considerando a variável curtose é possível afirmar que seu valor positivo indica que a curva é um pouco alongada (leptocúrtica).
- **Valores Atípicos:** esta variável apresentou dois valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos países teve crescimento do PIB per capita no ano de 2010. Na distribuição desta variável média e mediana estão bastante próximas.

Quanto maior for o valor desta variável, maior será a renda mundial. Porém, o aumento não significa que haverá distribuição igualitária da riqueza.

### 3.8 Variável: “Expectativa de vida ao nascer”

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 Expectativa de vida ao nascer.





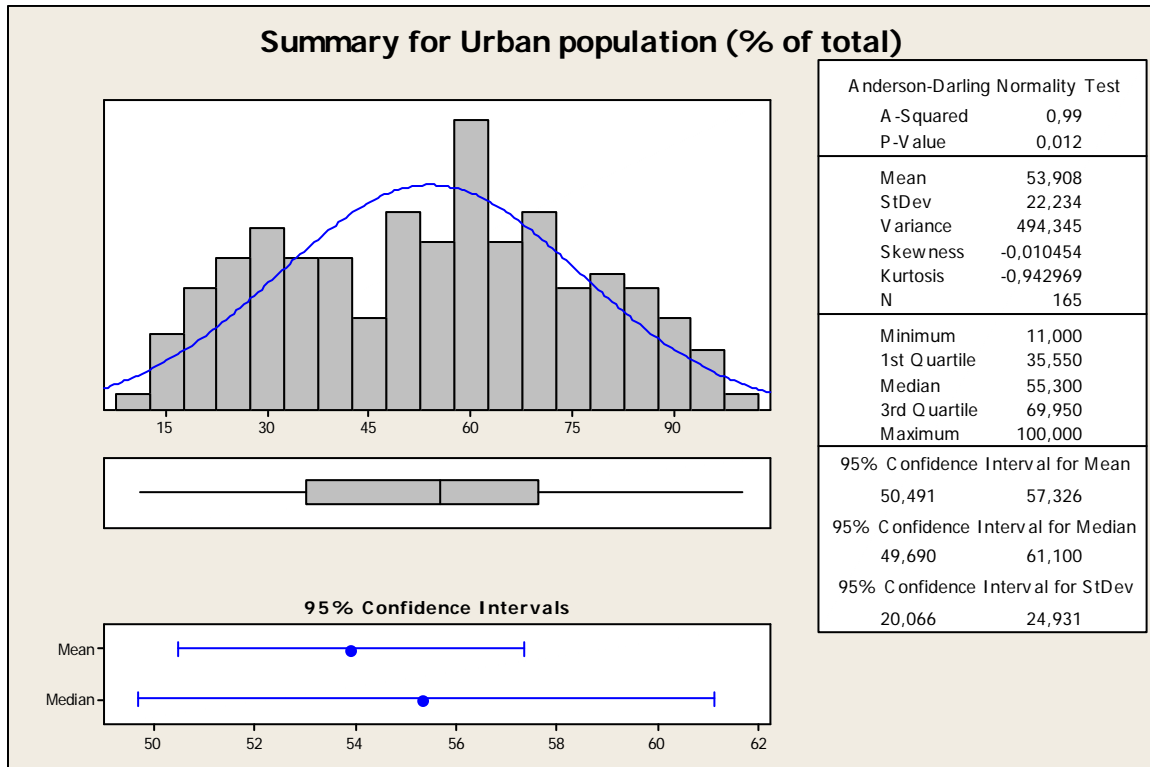
As principais observações da variável Expectativa de vida ao nascer são:

- **Forma:** O Histograma permite verificar que se trata de uma distribuição visivelmente assimétrica, revelando que a maioria dos países tem expectativa de vida acima dos 70 anos. A posição da linha da mediana à direita do centro do quadrado principal do Box-Plot nos indica também a assimetria. Considerando a variável curtose é possível afirmar que seu valor negativo indica que a curva é achatada (platicúrtica).
- **Valores Atípicos:** esta variável não apresentou valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que aproximadamente metade dos países tem expectativa de vida de sua população acima de 72 anos. Na distribuição desta variável média e mediana estão bastante distantes, demonstrando que há países com expectativa de vida ainda muito baixa, enquanto que outros bastante alta.

Quanto maior for o valor desta variável, mais tempo as pessoas irão viver. Se isso é bom ou ruim, depende das condições de vida oferecidas pelos países a seus idosos.

### 3.9 Variável: “População Urbana”

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 População Urbana.



As principais observações da variável População Urbana são:

- **Forma:** O Histograma permite verificar que se trata de uma distribuição relativamente assimétrica, revelando que há mais gente vivendo em cidades do que vivendo na zona rural. A posição da linha da mediana à esquerda do centro do quadrado principal do Box-Plot nos indica também a assimetria. Considerando a variável curtose é possível afirmar que seu valor negativo indica que a curva é achatada, demonstrando valores próximos de zero (platicúrtica).
- **Valores Atípicos:** esta variável não apresentou valores atípicos.
- **Centro e Dispersão:** A mediana nos indica que mais da metade da população mundial já vive em cidades. Na distribuição desta variável média e mediana estão bastante próximas.

### 3.10 Análise das médias das variáveis

O software Minitab oferece estatísticas básicas, que permitem verificar em uma só tabela, as médias, o desvio padrão, o quartil inferior, a mediana, o terceiro quartil e a máxima. Seguem abaixo as estatísticas básicas para as 9 variáveis numéricas selecionadas, padronizadas:

**Descriptive Statistics: Population g; Internet use; Mobile cellu; ...**

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
pops	161	0	3,7703	0,0867	1,0996	1	2,9358	3,6863	4,4798	10
intus	161	0	4,063	0,203	2,57	1	1,8	3,648	5,709	10
mobs	161	0	5,206	0,155	1,961	1	3,732	5,34	6,577	10
ages	161	0	5,217	0,15	1,909	1	3,771	4,558	6,676	10
ferts	161	0	3,645	0,173	2,193	1	1,905	2,993	5,172	10
gdppps	161	0	1,957	0,116	1,475	1	1,085	1,332	2,009	10
gdpgrs	161	0	5,307	0,113	1,429	1	4,396	5,18	6,267	10
lifes	161	0	6,483	0,197	2,5	1	4,795	7,443	8,188	10
urbpops	161	0	5,366	0,179	2,269	1	3,457	5,631	7,007	10

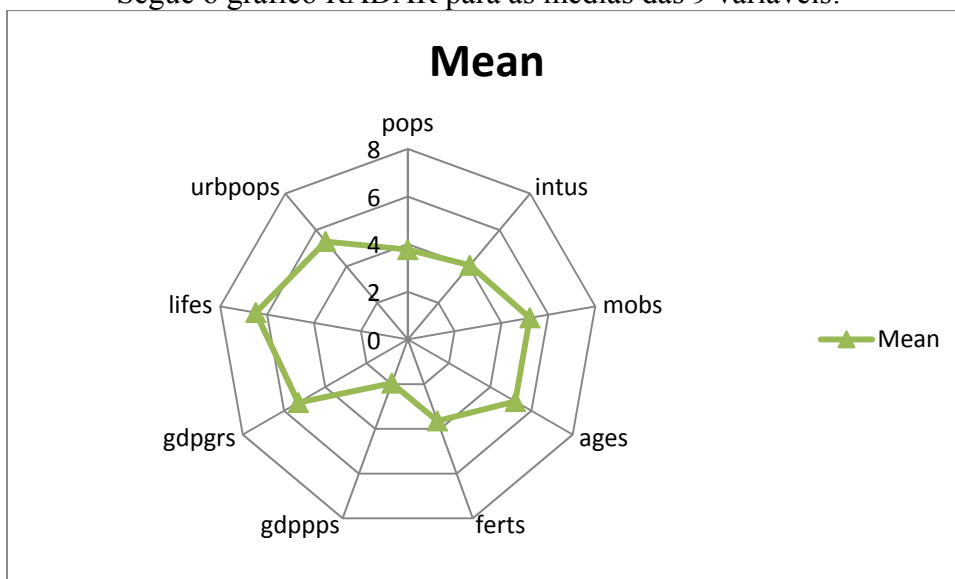
A média deve ser analisada para diferentes grupos de variáveis, já que algumas, quanto mais alta, melhor o resultado, ao passo que para outro grupo é o oposto, quanto mais baixa a média, melhor a situação encontrada. A variável com menor dispersão é a variável Usuários de Internet.

As variáveis Usuários de Internet, Assinaturas de Celular, PIB per Capita, % Crescimento do PIB per Capita e Expectativa de vida são aquelas que, quanto maior seu valor, melhor. Neste caso, a Expectativa de vida é a que tem a maior média.

Já as variáveis % Crescimento populacional, Razão de dependência da idade e Taxa de Fertilidade são aquelas que, quanto menor seu valor, melhor. Neste caso, a variável que apresenta a menor média é Taxa de fertilidade.

### 3.11 Gráfico RADAR para as médias

Segue o gráfico RADAR para as médias das 9 variáveis:



Pode-se concluir que, visualmente, o comportamento das variáveis não é muito uniforme.

## 4. CORRELAÇÕES

A proposta deste tópico é identificar quais variáveis influenciam no comportamento de outras.

### 4.1 Coeficientes de correlação entre variáveis quantitativas

**Correlations: Population g; Internet use; Mobile cellu; Age dependen; ...**

Internet users (	Population growt	Internet users (	Mobile cellular
	-0,491		
	0,000		
Mobile cellular	-0,487	0,655	
	0,000	0,000	
Age dependency r	0,601	-0,676	-0,695
	0,000	0,000	0,000
Fertility rate,	0,706	-0,700	-0,703
	0,000	0,000	0,000
GDP per capita (	-0,209	0,805	0,441
	0,008	0,000	0,000
GDP per capita g	-0,102	-0,152	-0,058
	0,197	0,054	0,463
Life expectancy	-0,513	0,772	0,681
	0,000	0,000	0,000
Urban population	-0,302	0,689	0,609
	0,000	0,000	0,000
Fertility rate,	Age dependency r	Fertility rate,	GDP per capita (
	0,953		
	0,000		
GDP per capita (	-0,419	-0,453	
	0,000	0,000	
GDP per capita g	-0,031	-0,016	-0,217
	0,693	0,840	0,006
Life expectancy	-0,766	-0,822	0,583
	0,000	0,000	0,000
Urban population	-0,533	-0,562	0,569
	0,000	0,000	0,000
Life expectancy	GDP per capita g	Life expectancy	
	-0,097		
	0,222		
Urban population	-0,059	0,614	
	0,459	0,000	
Cell Contents: Pearson correlation			
P-Value			

A matriz de correlação mostra diversos coeficientes de correlação acima de 0,5, ou seja, explicando um grau razoável do comportamento da outra variável, embora alguns sejam negativos, demonstrando que ao aumentar uma variável, a outra diminui e vice-versa. Há, porém, algumas variáveis que não tem correlação, seja por ter Pearson correlation menor que 0,5 ou P-Value maior que zero.

Alguns pontos de destaque:

- A maior correlação está entre as variáveis Taxa de Fertilidade e Razão de dependência da idade, com coeficiente de 0,953;
- A seguir vem Expectativa de vida com Taxa de Fertilidade, numa correlação inversa: -0,822;

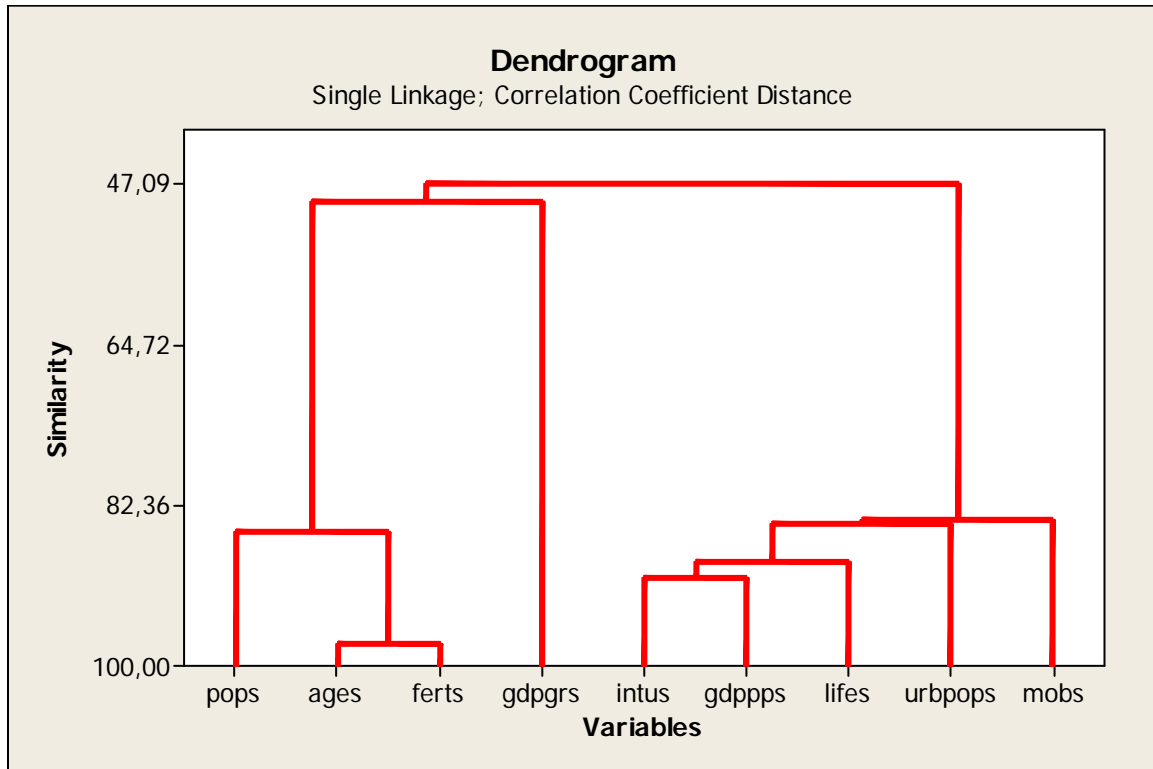
- Em terceiro lugar vem PIB per Capita com Usuários de Internet, com coeficiente 0,805.
- Em seguida, com grau de correlação acima de 0,6 estão as variáveis População Urbana e Expectativa de Vida, Expectativa de Vida com Razão de dependência da idade, População urbana com Usuários de Internet e Assinaturas de Celular.

As demais variáveis apresentam baixo grau de correlação.

## 4.2 Dendrograma

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.

Segue o Dendrograma para as 9 variáveis selecionadas:



O dendrograma das 9 variáveis demonstra de maneira gráfica que:

1. As variáveis Razão de dependência da idade e Taxa de fertilidade tem grande similaridade;
2. As duas variáveis anteriores tem similaridade alta com a variável Crescimento Populacional;
3. As variáveis Usuários de Internet e PIB per capita tem similaridade alta entre si;
4. As duas variáveis anteriores tem similaridade alta com a variável Expectativa de vida;
5. A variável % de Crescimento do PIB per capita tem baixa similaridade com as demais variáveis.

## 5. REGRESSÕES

Regressões são métodos de análise utilizados para identificar se o comportamento de uma variável numérica (Response) é explicado pelo comportamento de outra (Predictor). Esse método permite que sejam reduzidas as quantidades de variáveis utilizadas nos modelos estatísticos.

Neste tópico são analisadas as regressões entre as 9 variáveis numéricas selecionadas, mediante a análise da equação de regressão e do gráfico de dispersão.

### 5.1 Regressão linear simples

A seguir, as variáveis são analisadas comparativamente a cada uma das demais.

### 5.1.1 Regressão da variável Crescimento Populacional x Usuários de Internet

Segue o resultado da regressão simples das variáveis Crescimento Populacional e Usuários de Internet.

#### Regression Analysis: Population growth versus Internet users

The regression equation is

Population growth (annual %) = 2,03 - 0,0210 Internet users (per 100 people)

Predictor	Coef	SE Coef	T	P
Constant	2,0279	0,1253	16,18	0,000
Internet users (per 100 people)	-0,020982	0,002949	-7,12	0,000

S = 1,01629    R-Sq = 24,2%    R-Sq(adj) = 23,7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	52,301	52,301	50,64	0,000
Residual Error	159	164,222	1,033		
Total	160	216,523			

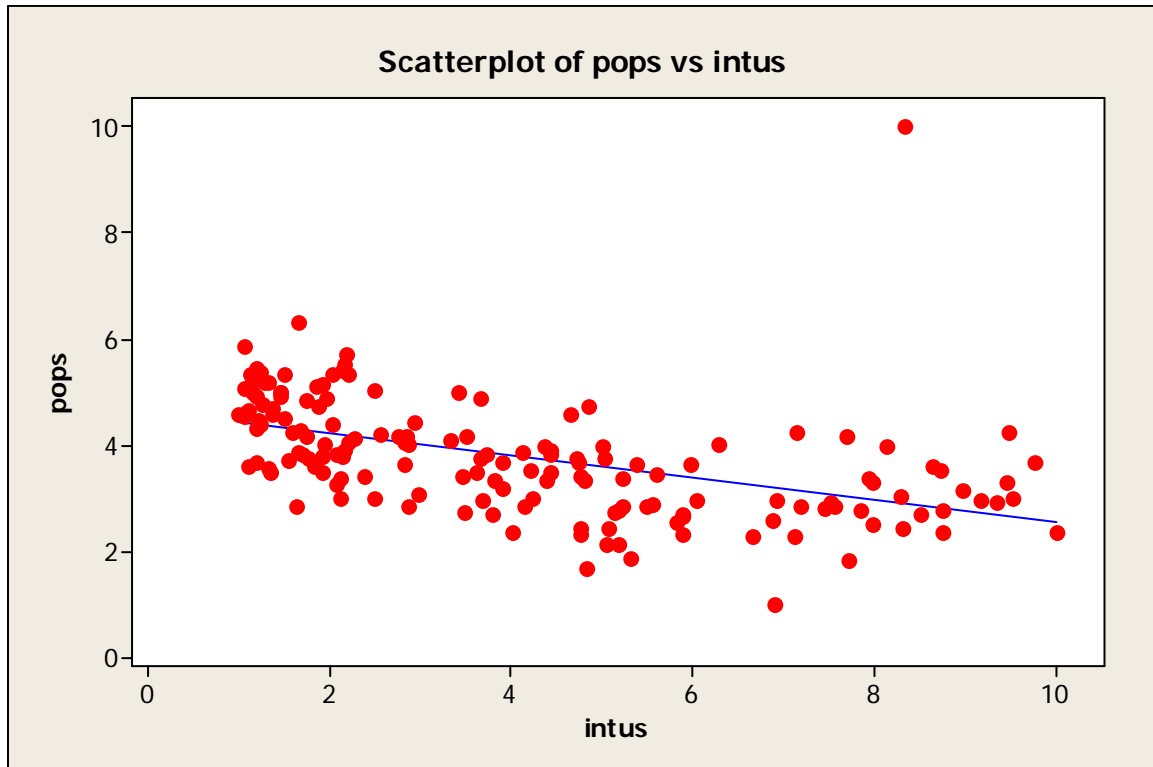
Unusual Observations

Obs	Internet users (per 100 people)	Population growth (annual %)	Fit	SE Fit	Residual	St Resid
66	95,6	-0,1439	0,0213	0,2022	-0,1652	-0,17 X
84	7,0	4,0412	1,8810	0,1102	2,1602	2,14R
85	62,8	-1,5887	0,7100	0,1196	-2,2988	-2,28R
126	40,8	-0,8894	1,1724	0,0836	-2,0619	-2,04R
151	78,0	7,9329	0,3914	0,1558	7,5415	7,51R

R denotes an observation with a large standardized residual.

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

A equação acima revela um grau muito baixo de adaptação da reta de regressão à amostra: a variável Usuários de Internet prevê apenas 24,2% do comportamento da variável Crescimento Populacional. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.2 Regressão da variável Crescimento Populacional x Assinaturas de celular

Segue o resultado da regressão simples das variáveis Crescimento Populacional e Assinaturas de Celular.

#### Regression Analysis: Population growth versus Mobile cellular

The regression equation is

Population growth (annual %) = 2,61 - 0,0141 Mobile cellular subscriptions (

Predictor	Coef	SE Coef	T	P
Constant	2,6071	0,1970	13,23	0,000
Mobile cellular subscriptions (	-0,014105	0,002006	-7,03	0,000

S = 1,01923    R-Sq = 23,7%    R-Sq(adj) = 23,2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	51,348	51,348	49,43	0,000
Residual Error	159	165,175	1,039		
Total	160	216,523			

Unusual Observations

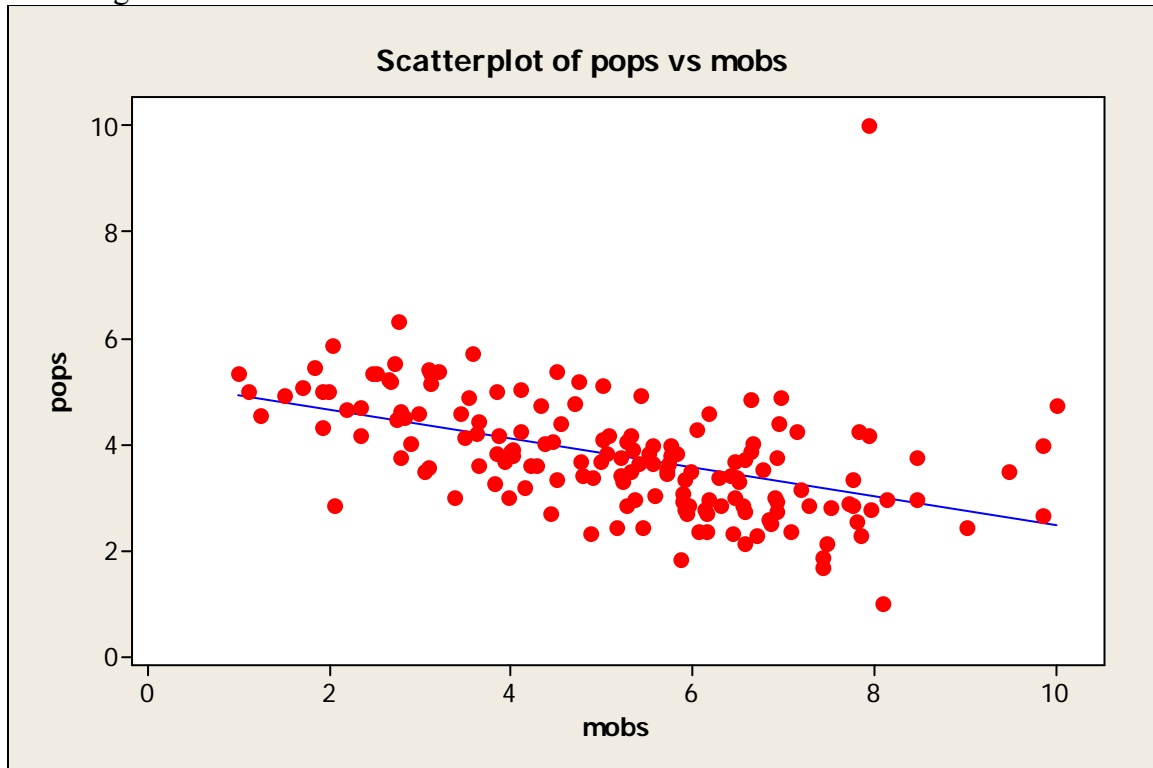
Obs	Mobile cellular subscriptions (	Population growth (annual %)	Fit	SE Fit	Residual	St Resid
85	149	-1,5887	0,5083	0,1433	-2,0970	-2,08R
100	185	0,1672	-0,0042	0,2077	0,1714	0,17 X
111	185	1,5739	0,0018	0,2069	1,5722	1,58 X
123	188	2,3555	-0,0427	0,2127	2,3982	2,41RX
151	145	7,9329	0,5555	0,1377	7,3774	7,31R

R denotes an observation with a large standardized residual.

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



A equação acima revela um grau baixo de adaptação da reta de regressão à amostra: a variável Assinaturas de Celular prevê apenas 23,7% do comportamento da variável Crescimento Populacional. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.3 Regressão da variável Crescimento Populacional x Razão de dependência da idade

Segue o resultado da regressão simples das variáveis Crescimento Populacional e Razão de dependência da idade.

#### Regression Analysis: Population growth versus Age dependency rate

The regression equation is

Population growth (annual %) = - 1,04 + 0,0394 Age dependency ratio (% of work

Predictor	Coef	SE Coef	T	P
Constant	-1,0373	0,2612	-3,97	0,000
Age dependency ratio (% of work	0,039426	0,004154	9,49	0,000

S = 0,932330 R-Sq = 36,2% R-Sq(adj) = 35,8%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	78,314	78,314	90,09	0,000
Residual Error	159	138,209	0,869		
Total	160	216,523			

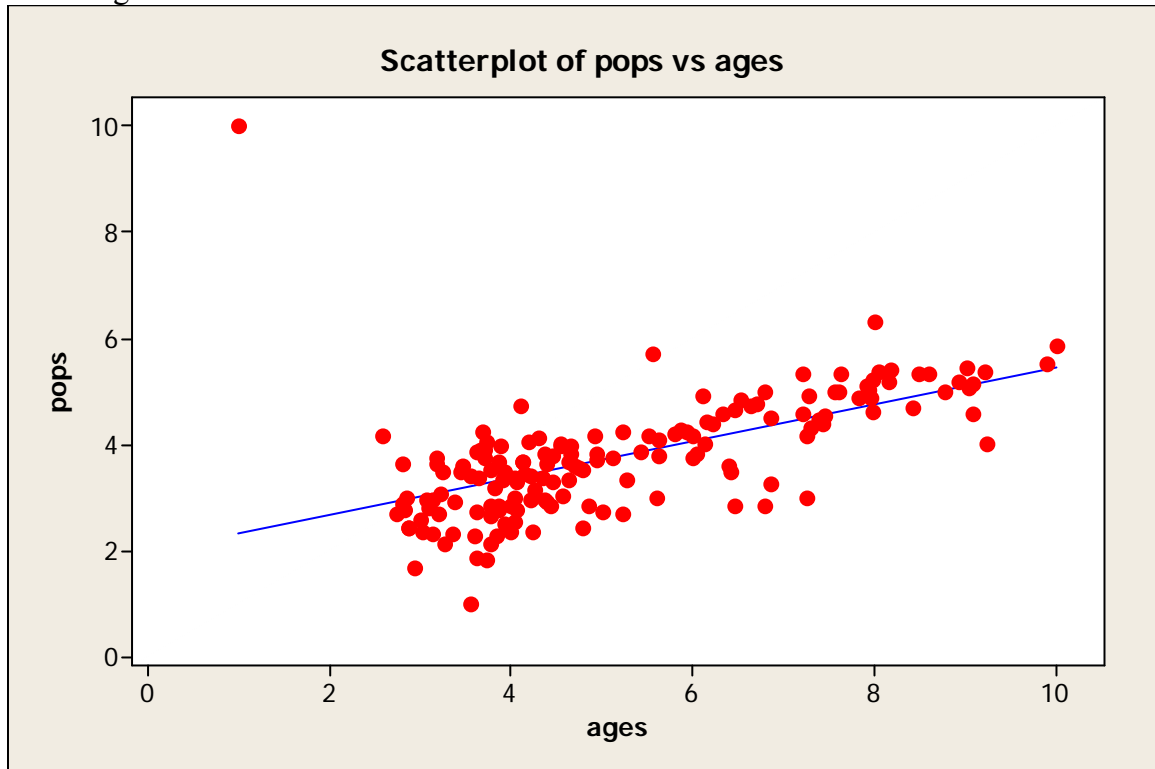
#### Unusual Observations

Obs	Age dependency ratio (% of work	Population growth (annual %)	Fit	SE Fit	Residual	St Resid
14	64	3,3932	1,4665	0,0746	1,9266	2,07R
85	45	-1,5887	0,7275	0,0979	-2,3162	-2,50R
107	105	3,5412	3,0948	0,1987	0,4464	0,49 X

149	104	3,2127	3,0575	0,1951	0,1552	0,17 X
151	21	7,9329	-0,2030	0,1786	8,1359	8,89R

R denotes an observation with a large standardized residual.  
 X denotes an observation whose X value gives it large leverage.

A equação acima revela um grau baixo de adaptação da reta de regressão à amostra: a variável Razão de dependência da idade prevê apenas 36,2% do comportamento da variável Crescimento Populacional. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.4 Regressão da variável Crescimento Populacional x Taxa de Fertilidade

Segue o resultado da regressão simples das variáveis Crescimento Populacional e Taxa de Fertilidade.

**Regression Analysis: Population growt versus Fertility rate,**

The regression equation is  
 Population growth (annual %) = - 0,302 + 0,570 Fertility rate, total (births p

Predictor	Coef	SE Coef	T	P
Constant	-0,3021	0,1463	-2,07	0,040
Fertility rate, total (births p	0,56963	0,04536	12,56	0,000

S = 0,826875    **R-Sq = 49,8%**    R-Sq(adj) = 49,5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	107,81	107,81	157,68	0,000
Residual Error	159	108,71	0,68		
Total	160	216,52			

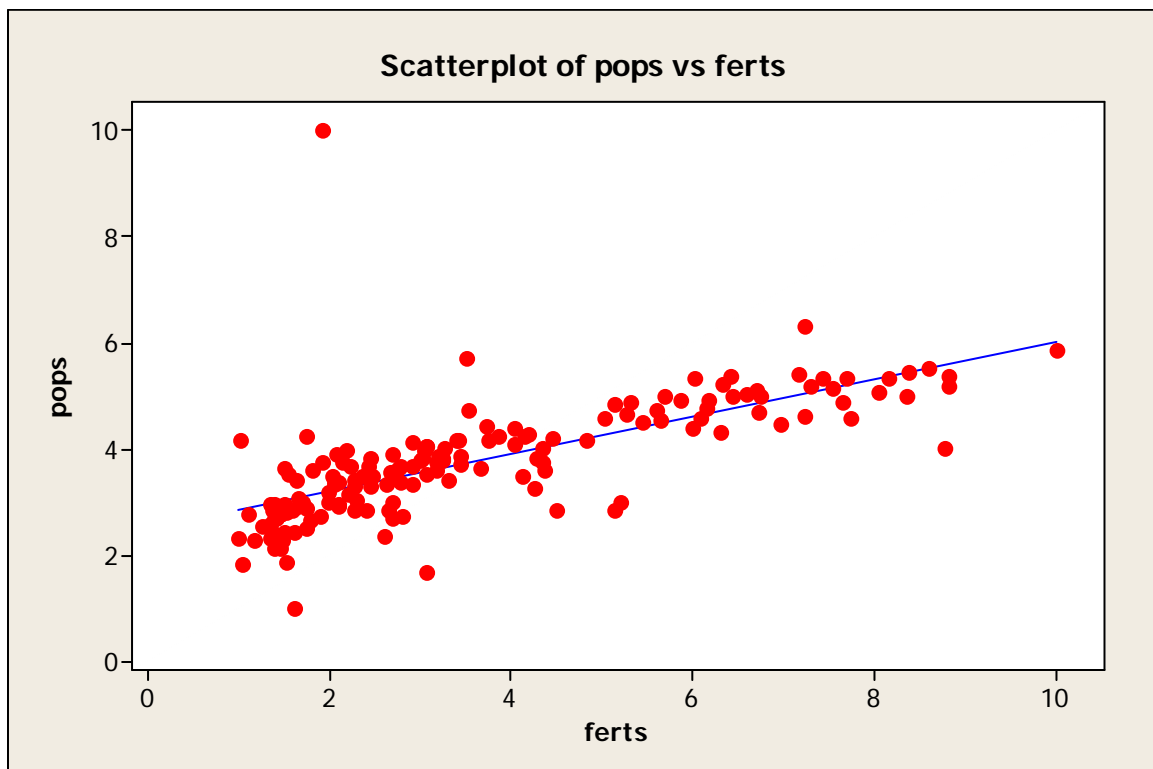
Unusual Observations

Fertility	Population
-----------	------------

Obs	rate, total (births p	growth (annual %)	Fit	SE Fit	Residual	St Resid
1	6,29	2,8313	3,2797	0,1675	-0,4484	-0,55 X
14	2,79	3,3932	1,2871	0,0653	2,1060	2,55R
85	1,55	-1,5887	0,5808	0,0890	-2,1695	-2,64R
92	6,29	3,0386	3,2831	0,1677	-0,2446	-0,30 X
107	7,06	3,5412	3,7212	0,2003	-0,1800	-0,22 X
126	2,50	-0,8894	1,1219	0,0675	-2,0114	-2,44R
149	6,15	3,2127	3,2005	0,1617	0,0122	0,02 X
151	1,75	7,9329	0,6941	0,0831	7,2388	8,80R
160	6,26	1,5802	3,2626	0,1662	-1,6824	-2,08RX

R denotes an observation with a large standardized residual.  
X denotes an observation whose X value gives it large leverage.

A equação acima revela um grau de adaptação baixo da reta de regressão à amostra: a variável Taxa de Fertilidade prevê 49,8% do comportamento da variável Crescimento Populacional. Surpreende esta correlação não ser ainda maior. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.5 Regressão da variável Crescimento Populacional x PIB per capita

Segue o resultado da regressão simples das variáveis Crescimento Populacional e PIB per Capita.

Regression Analysis: Population growth versus GDP per capita					
The regression equation is					
Population growth (annual %) = 1,50 - 0,000014 GDP per capita (current US\$)					
Predictor	Coef	SE Coef	T	P	
Constant	1,5023	0,1079	13,93	0,000	
GDP per capita (current US\$)	-0,00001409	0,00000524	-2,69	0,008	
S = 1,14130    R-Sq = 4,3%    R-Sq(adj) = 3,7%					

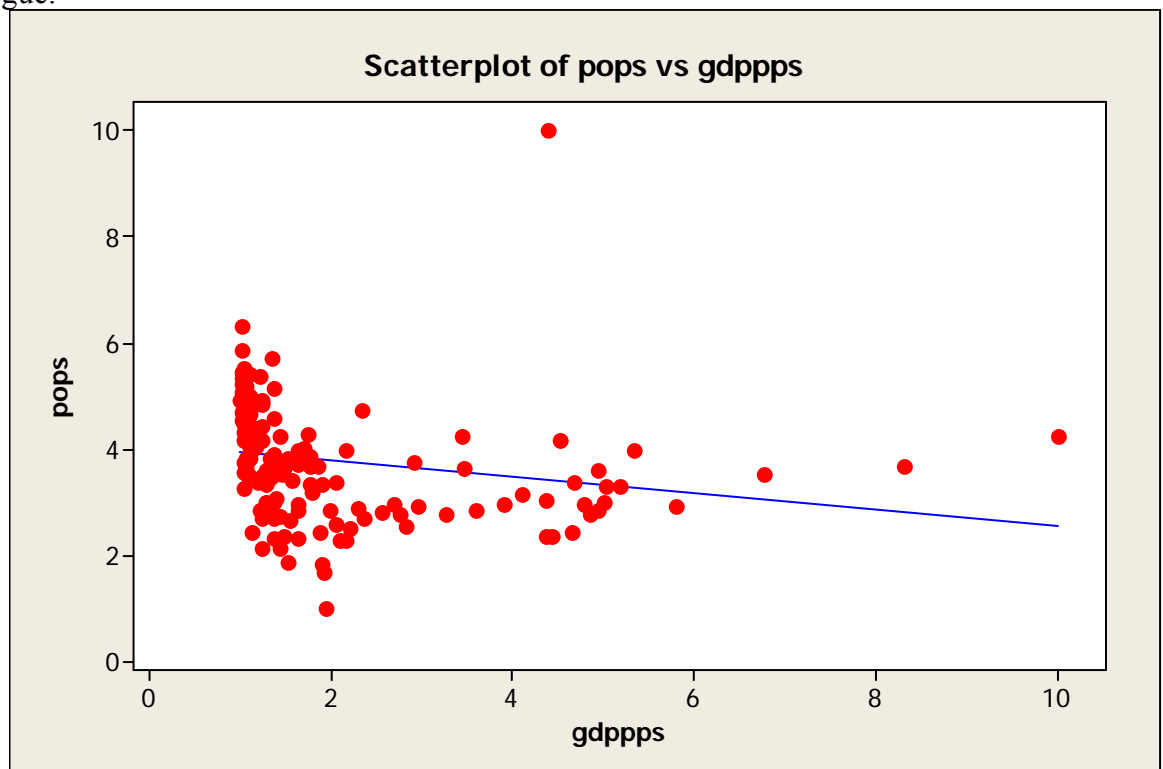
Analysis of Variance						
Source	DF	SS	MS	F	P	
Regression	1	9,413	9,413	7,23	0,008	
Residual Error	159	207,110	1,303			
Total	160	216,523				

Unusual Observations						
Obs	GDP per capita (current US\$)	Population growth (annual %)	Fit	SE Fit	Residual	St Resid
7	50748	1,5733	0,7872	0,2252	0,7861	0,70 X
41	56245	0,4442	0,7098	0,2519	-0,2656	-0,24 X
84	247	4,0412	1,4988	0,1072	2,5424	2,24R
85	11045	-1,5887	1,3466	0,0900	-2,9354	-2,58R
86	105195	1,8254	0,0201	0,5000	1,8053	1,76 X
109	85389	1,2457	0,2991	0,3983	0,9465	0,88 X
137	67457	1,0575	0,5518	0,3075	0,5057	0,46 X
151	39623	7,9329	0,9440	0,1733	6,9890	6,20R

R denotes an observation with a large standardized residual.  
X denotes an observation whose X value gives it large leverage.

A equação acima revela um grau de adaptação quase nulo da reta de regressão à amostra: a variável PIB per capita prevê apenas 4,3% do comportamento da variável Crescimento Populacional. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.6 Regressão da variável Crescimento Populacional x % Crescimento PIB per capita

Segue o resultado da regressão simples das variáveis Crescimento Populacional e 5 Crescimento PIB per capita.

**Regression Analysis: Population growth versus GDP per capita g**

The regression equation is  
 Population growth (annual %) = 1,46 - 0,0387 GDP per capita growth (annual %

Predictor	Coef	SE Coef	T	P
Constant	1,4565	0,1271	11,46	0,000
GDP per capita growth (annual %	-0,03869	0,02987	-1,30	0,197

S = 1,16084    R-Sq = 1,0%    R-Sq(adj) = 0,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2,261	2,261	1,68	0,197
Residual Error	159	214,262	1,348		
Total	160	216,523			

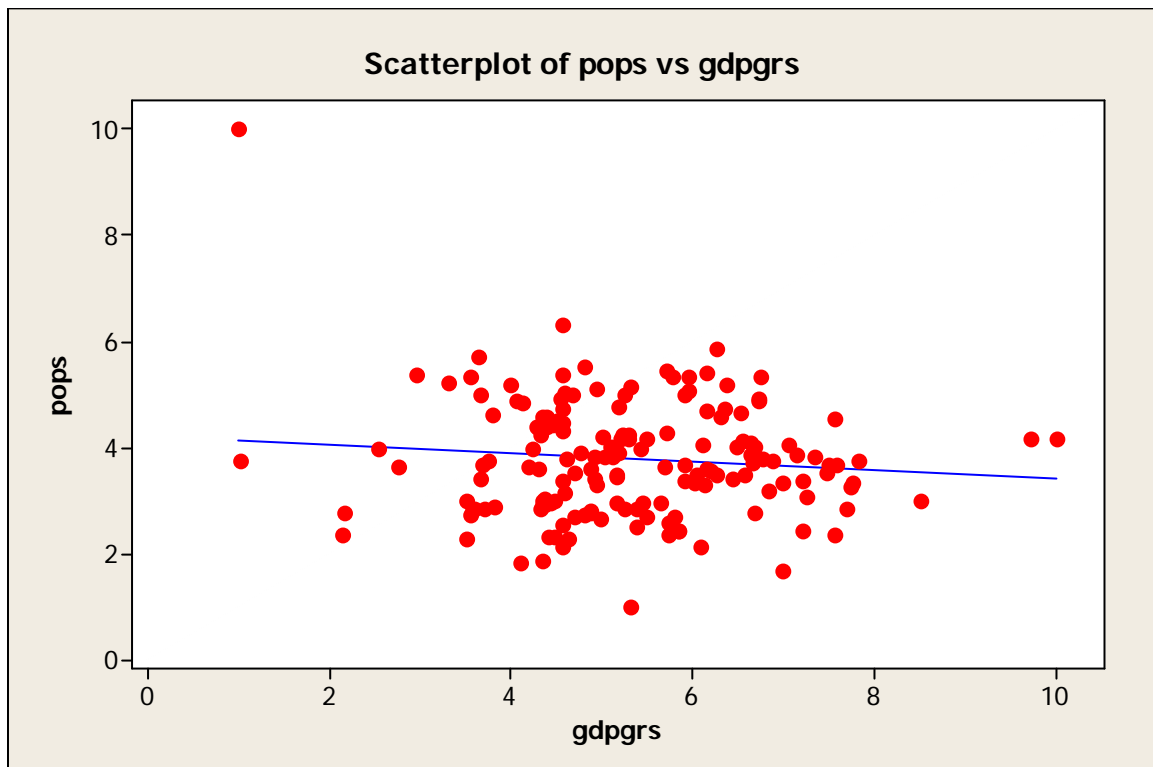
Unusual Observations

Obs	GDP per capita growth (annual %)	Population growth (annual %)	Fit	SE Fit	Residual	St Resid
31	9,8	0,5184	1,0762	0,2248	-0,5578	-0,49 X
63	-6,3	1,2993	1,6993	0,2905	-0,3999	-0,36 X
84	1,3	4,0412	1,4047	0,1034	2,6365	2,28R
85	2,9	-1,5887	1,3424	0,0915	-2,9312	-2,53R
113	13,0	1,7608	0,9517	0,3150	0,8091	0,72 X
127	12,5	1,7707	0,9746	0,2982	0,7961	0,71 X
151	-6,3	7,9329	1,7005	0,2914	6,2324	5,55RX

R denotes an observation with a large standardized residual.

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

A equação acima revela um valor inválido, porque o P-value deu acima de 0,05. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.7 Regressão da variável Crescimento Populacional x Expectativa de vida ao nascer

Segue o resultado da regressão simples das variáveis Crescimento Populacional e Expectativa de vida ao nascer.

#### Regression Analysis: Population growth versus Life expectancy

The regression equation is

Population growth (annual %) = 5,51 - 0,0604 Life expectancy at birth, total

Predictor	Coef	SE Coef	T	P
Constant	5,5149	0,5588	9,87	0,000
Life expectancy at birth, total	-0,060443	0,008013	-7,54	0,000

S = 1,00145 R-Sq = 26,4% R-Sq(adj) = 25,9%

Analysis of Variance

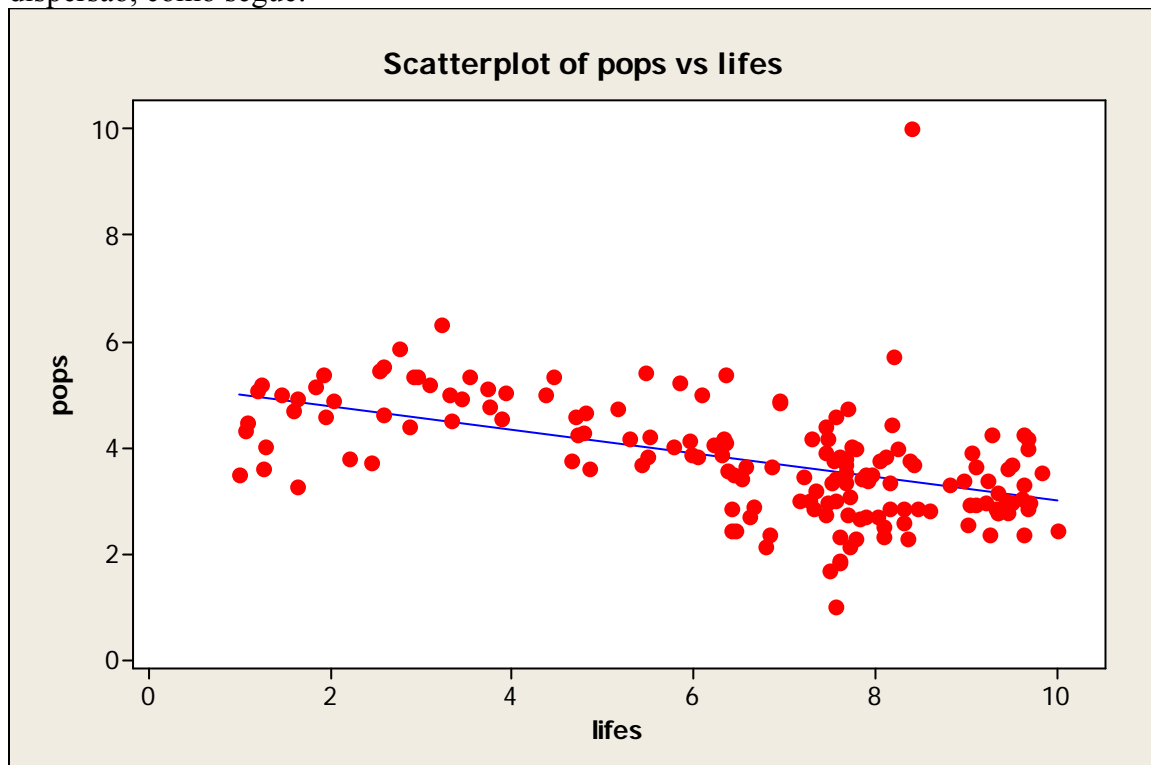
Source	DF	SS	MS	F	P
Regression	1	57,060	57,060	56,89	0,000
Residual Error	159	159,463	1,003		
Total	160	216,523			

Unusual Observations

Obs	Life expectancy at birth, total	Population growth (annual %)	Fit	SE Fit	Residual	St Resid
14	75,8	3,3932	0,9309	0,0959	2,4623	2,47R
85	73,3	-1,5887	1,0863	0,0859	-2,6751	-2,68R
151	76,6	7,9329	0,8865	0,0994	7,0464	7,07R

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação quase nulo da reta de regressão à amostra: a variável % Crescimento PIB per capita prevê apenas 26,4% do comportamento da variável Crescimento Populacional. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.8 Regressão da variável Crescimento Populacional x População Urbana

Segue o resultado da regressão simples das variáveis Crescimento Populacional e População Urbana.

#### Regression Analysis: Population growth versus Urban population

The regression equation is  
Population growth (annual %) = 2,19 - 0,0157 Urban population (% of total)

Predictor	Coef	SE Coef	T	P
Constant	2,1901	0,2298	9,53	0,000
Urban population (% of total)	-0,015651	0,003920	-3,99	0,000

S = 1,11253 R-Sq = 9,1% R-Sq(adj) = 8,5%

#### Analysis of Variance

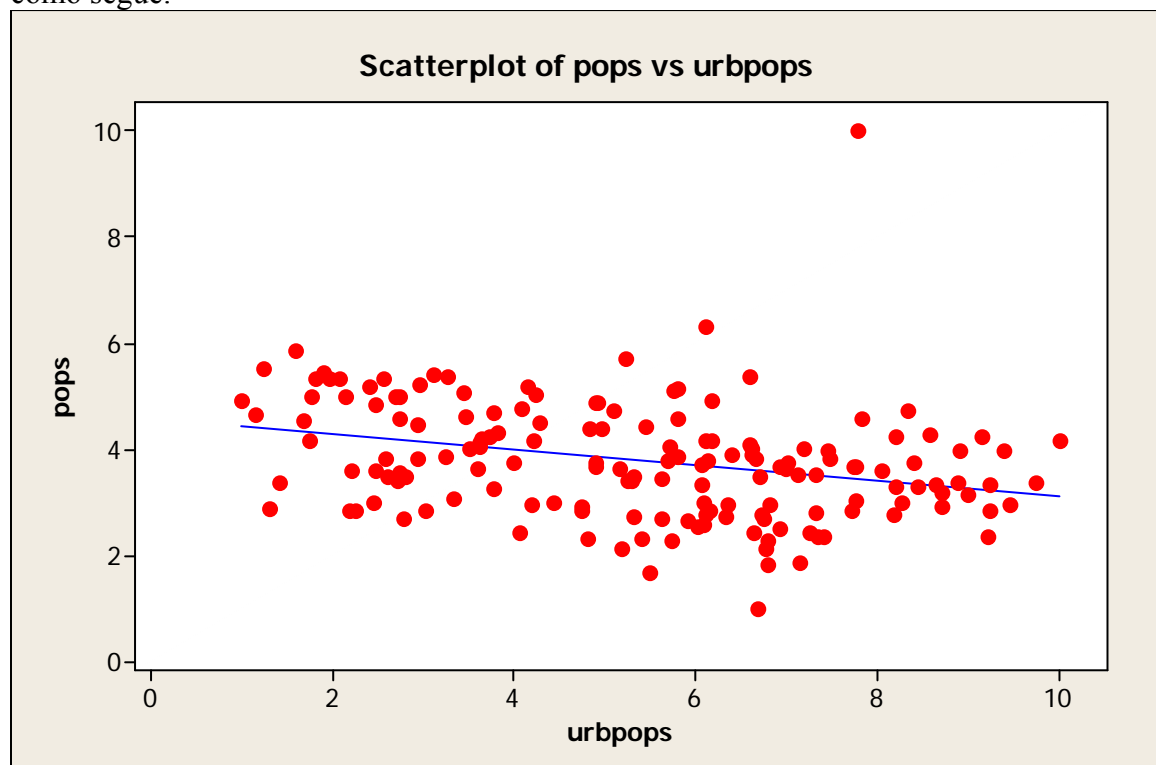
Source	DF	SS	MS	F	P
Regression	1	19,726	19,726	15,94	0,000
Residual Error	159	196,796	1,238		
Total	160	216,523			

#### Unusual Observations

Obs	Urban population (% of total)	Population growth (annual %)	Fit	SE Fit	Residual	St Resid
84	62	4,0412	1,2276	0,0923	2,8136	2,54R
85	67	-1,5887	1,1384	0,1015	-2,7271	-2,46R
151	78	7,9329	0,9693	0,1281	6,9636	6,30R

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação quase nulo da reta de regressão à amostra: a variável População urbana prevê apenas 9,1% do comportamento da variável Crescimento Populacional. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.9 Regressão da variável Usuários de Internet x % Assinaturas de celular

Segue o resultado da regressão simples das variáveis Usuários de Internet e Assinaturas de Celular.

#### Regression Analysis: Internet users versus Mobile cellular

The regression equation is

Internet users (per 100 people) = - 7,16 + 0,444 Mobile cellular subscriptions (

Predictor	Coef	SE Coef	T	P
Constant	-7,163	3,994	-1,79	0,075
Mobile cellular subscriptions (	0,44432	0,04066	10,93	0,000

S = 20,6581    R-Sq = 42,9%    R-Sq(adj) = 42,5%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	50950	50950	119,39	0,000
Residual Error	159	67855	427		
Total	160	118804			

#### Unusual Observations

Obs	Mobile cellular subscriptions (	Internet users (per 100 people)		SE Fit	Residual	St Resid
		Fit	SE Fit			
26	70	81,34	24,13	1,81	57,20	2,78R
51	97	77,50	36,11	1,66	41,39	2,01R
66	107	95,64	40,49	1,78	55,14	2,68R
74	95	77,64	34,92	1,64	42,72	2,07R
78	104	82,52	38,99	1,73	43,53	2,11R
100	185	51,96	75,09	4,21	-23,13	-1,14 X
105	115	90,70	44,12	1,94	46,58	2,26R
109	116	93,28	44,17	1,94	49,10	2,39R
111	185	42,75	74,91	4,19	-32,16	-1,59 X
123	188	41,00	76,31	4,31	-35,31	-1,75 X
136	116	90,02	44,41	1,95	45,61	2,22R
153	90	74,25	32,90	1,63	41,35	2,01R
158	177	27,85	71,55	3,91	-43,70	-2,15R

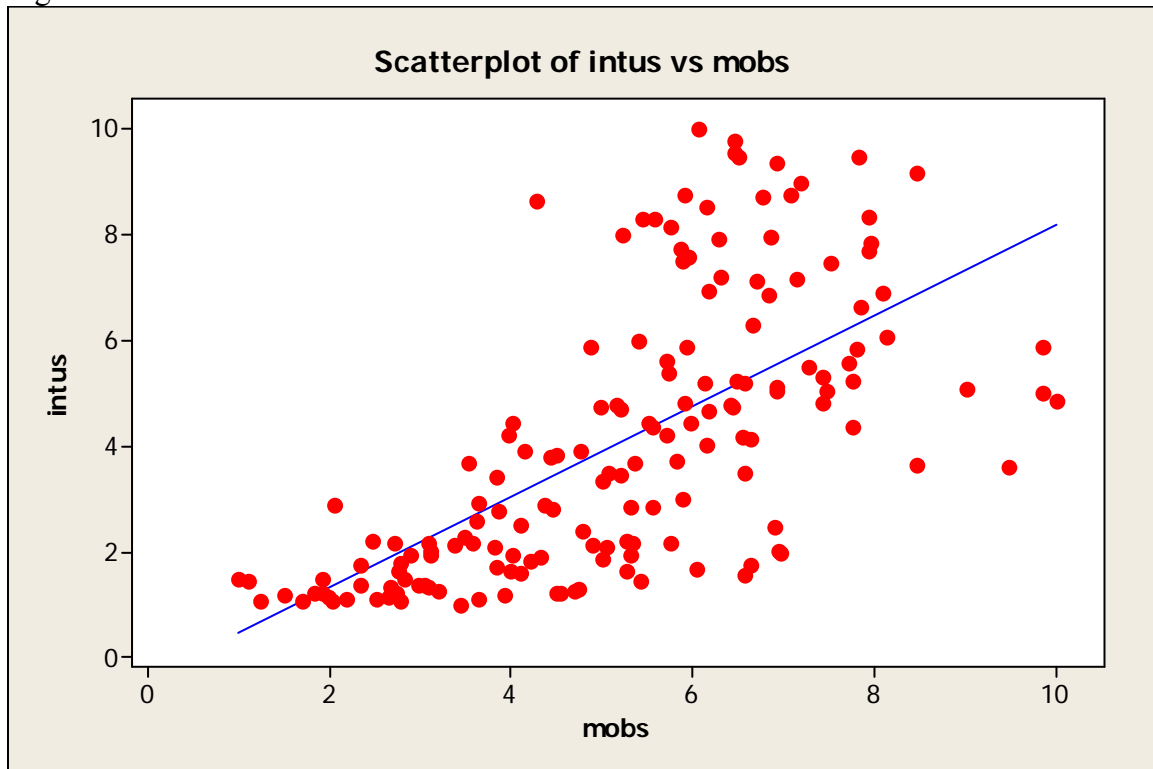
R denotes an observation with a large standardized residual.

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

A equação acima revela um grau de adaptação um pouco baixo da reta de regressão à amostra: a variável Assinaturas de Celular prevê apenas 42,9% do comportamento da variável Usuários de Internet. Essa conclusão é reforçada pela análise do gráfico de dispersão, como



segue:



### 5.1.10 Regressão da variável Usuários de Internet x Razão de dependência da idade

Segue o resultado da regressão simples das variáveis Usuários de Internet e Razão de dependência da idade.

#### Regression Analysis: Internet users versus Age dependency rate

The regression equation is

Internet users (per 100 people) = 95,3 - 1,04 Age dependency ratio (% of work

Predictor	Coef	SE Coef	T	P
Constant	95,304	5,646	16,88	0,000
Age dependency ratio (% of work	-1,03758	0,08978	-11,56	0,000

S = 20,1512 R-Sq = 45,7% R-Sq(adj) = 45,3%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	54239	54239	133,57	0,000
Residual Error	159	64565	406		
Total	160	118804			

#### Unusual Observations

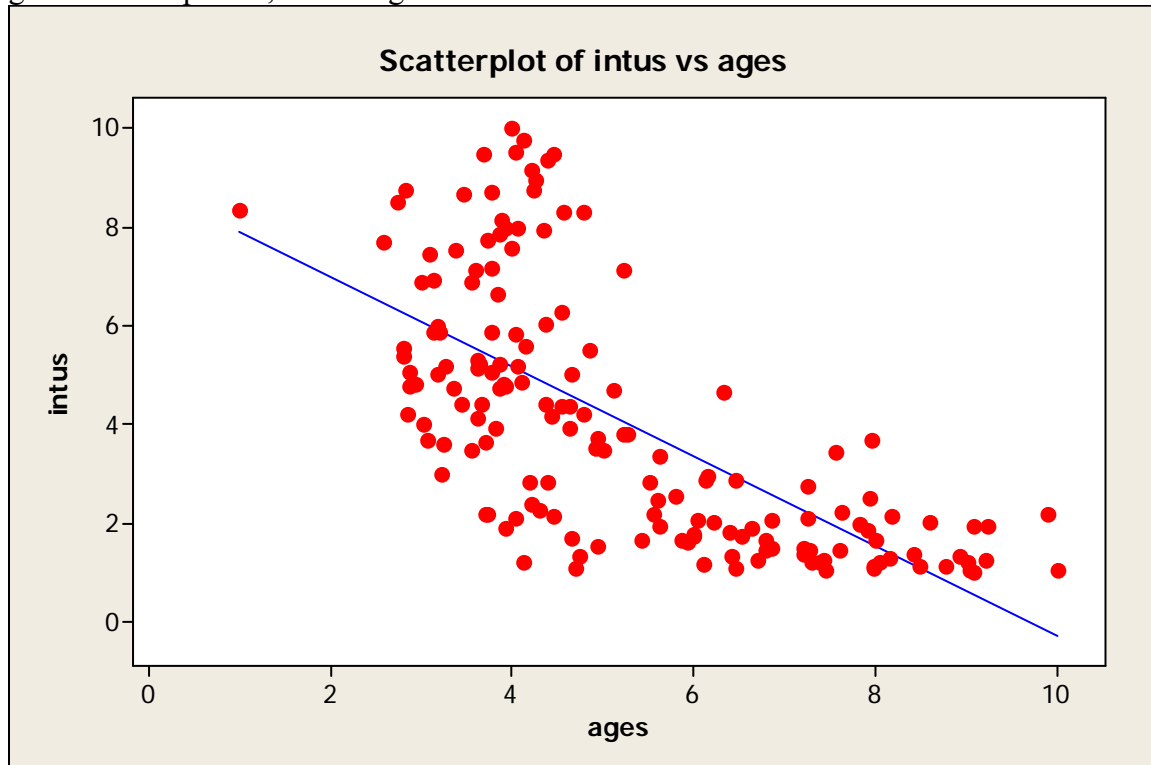
Obs	Age dependency ratio (% of work	Internet users (per 100 people)	Fit	SE Fit	Residual	St Resid
41	53	88,77	40,74	1,73	48,03	2,39R
50	51	86,90	42,37	1,80	44,53	2,22R
55	51	82,53	42,19	1,79	40,34	2,01R
66	49	95,64	44,62	1,89	51,02	2,54R
74	56	77,64	36,82	1,63	40,82	2,03R
86	46	90,08	47,46	2,04	42,62	2,13R
105	49	90,70	44,17	1,87	46,53	2,32R
107	105	0,83	-13,44	4,30	14,27	0,72 X
109	50	93,28	43,31	1,84	49,96	2,49R

136	53	90,02	40,00	1,71	50,02	2,49R
148	50	2,20	43,35	1,84	-41,15	-2,05R
149	104	12,50	-12,46	4,22	24,96	1,27 X
152	51	84,73	41,96	1,78	42,77	2,13R

R denotes an observation with a large standardized residual.

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

A equação acima revela um grau de adaptação um relativamente baixo da reta de regressão à amostra: a variável Razão de dependência da idade prevê 45,7% do comportamento da variável Usuários de Internet. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.11 Regressão da variável Usuários de Internet x Taxa de Fertilidade

Segue o resultado da regressão simples das variáveis Usuários de Internet e Taxa de Fertilidade.

#### Regression Analysis: Internet users versus Fertility rate

The regression equation is

Internet users (per 100 people) = 70,9 - 13,2 Fertility rate, total (births p

Predictor	Coef	SE Coef	T	P
Constant	70,882	3,454	20,52	0,000
Fertility rate, total (births p	-13,233	1,071	-12,35	0,000

S = 19,5264 R-Sq = 49,0% R-Sq(adj) = 48,7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	58181	58181	152,59	0,000
Residual Error	159	60623	381		
Total	160	118804			

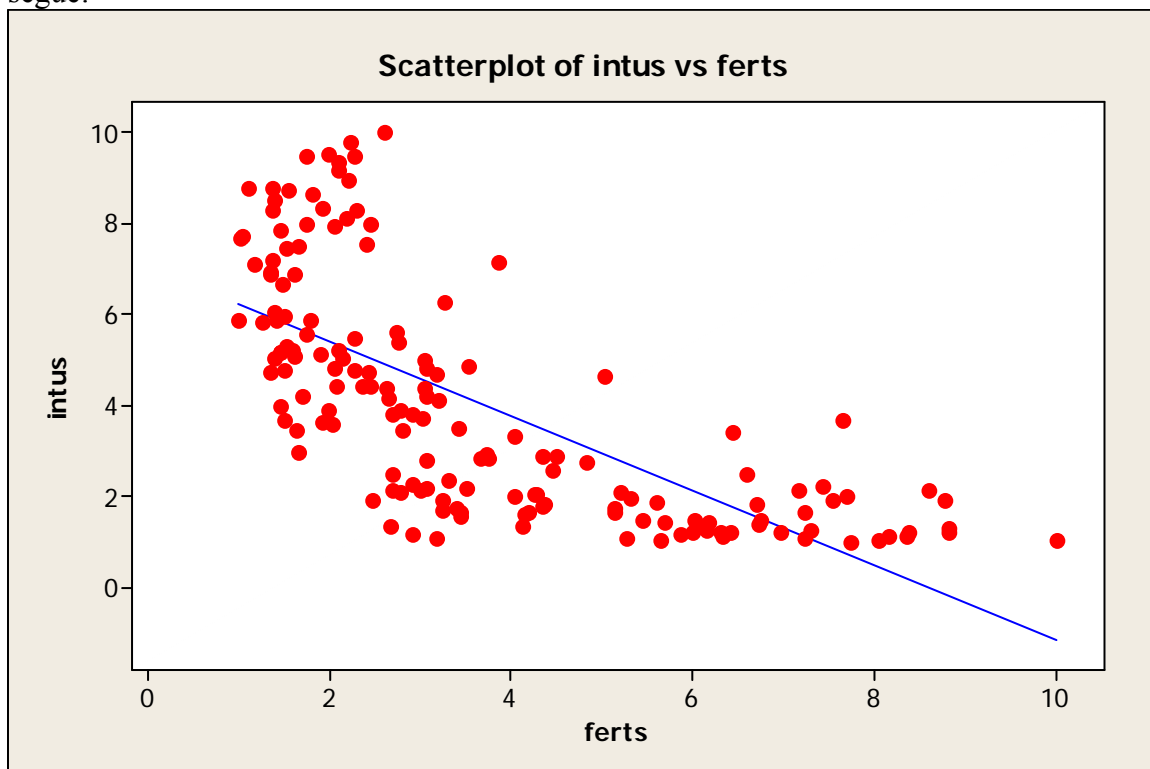
Unusual Observations

Fertility	Internet
-----------	----------

Obs	rate, total (births p	users (per 100 people)	Fit	SE Fit	Residual	St Resid
1	6,29	3,65	-12,33	3,96	15,98	0,84 X
41	1,87	88,77	46,14	1,89	42,63	2,19R
50	1,87	86,90	46,14	1,89	40,76	2,10R
66	2,20	95,64	41,77	1,71	53,87	2,77R
86	1,63	90,08	49,31	2,04	40,77	2,10R
92	6,29	2,70	-12,41	3,96	15,11	0,79 X
105	1,79	90,70	47,19	1,94	43,51	2,24R
107	7,06	0,83	-22,58	4,73	23,41	1,24 X
109	1,95	93,28	45,08	1,84	48,20	2,48R
136	1,98	90,02	44,68	1,82	45,34	2,33R
149	6,15	12,50	-10,49	3,82	22,99	1,20 X
152	1,94	84,73	45,21	1,84	39,52	2,03R
160	6,26	10,12	-11,93	3,93	22,05	1,15 X

R denotes an observation with a large standardized residual.  
X denotes an observation whose X value gives it large leverage.

A equação acima revela um grau de adaptação relativamente baixo da reta de regressão à amostra: a variável Taxa de fertilidade prevê 49% do comportamento da variável Usuários de Internet. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.12 Regressão da variável Usuários de Internet x PIB Per Capita

Segue o resultado da regressão simples das variáveis Usuários de Internet e PIB per Capita.

#### Regression Analysis: Internet users versus GDP per capita

The regression equation is  
Internet users (per 100 people) = 18,2 + 0,00127 GDP per capita (current US\$)

Predictor	Coef	SE Coef	T	P
Constant	18,198	1,532	11,88	0,000
GDP per capita (current US\$)	0,00127480	0,00007442	17,13	0,000

S = 16,2047 R-Sq = 64,9% R-Sq(adj) = 64,6%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	77052	77052	293,43	0,000
Residual Error	159	41752	263		
Total	160	118804			

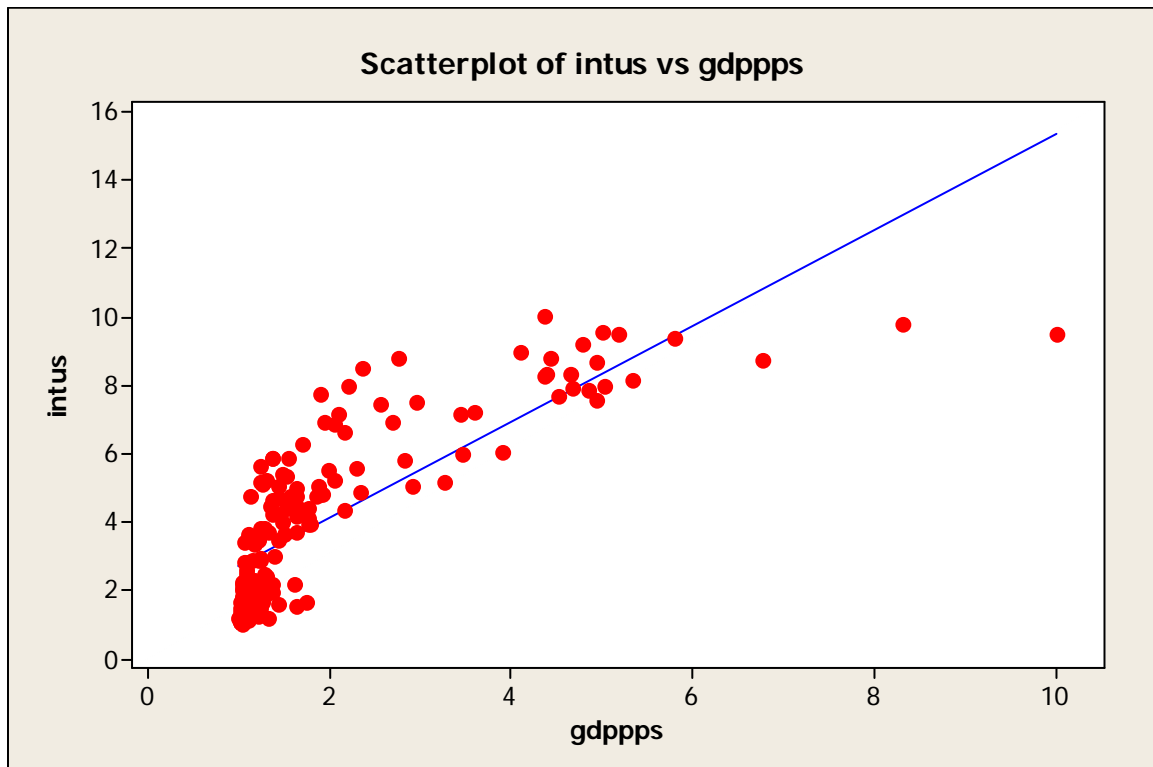
#### Unusual Observations

Obs	GDP per capita (current US\$)	Internet users (per 100 people)	Fit	SE Fit	Residual	St Resid
7	50748	75,90	82,89	3,20	-7,00	-0,44 X
41	56245	88,77	89,90	3,58	-1,13	-0,07 X
47	14341	74,16	36,48	1,30	37,68	2,33R
78	20757	82,52	44,66	1,46	37,86	2,35R
81	10723	71,51	31,87	1,28	39,65	2,45R
86	105195	90,08	152,30	7,10	-62,22	-4,27RX
109	85389	93,28	127,05	5,66	-33,77	-2,22RX
128	16071	79,89	38,69	1,32	41,20	2,55R
137	67457	82,17	104,19	4,37	-22,03	-1,41 X

R denotes an observation with a large standardized residual.

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

A equação acima revela um grau de adaptação bastante alto da reta de regressão à amostra: a variável PIB per capita prevê 64,9% do comportamento da variável Usuários de Internet. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.13 Regressão da variável Usuários de Internet x % Crescimento do PIB Per Capita

Segue o resultado da regressão simples das variáveis Usuários de Internet e % Crescimento PIB per Capita.

### Regression Analysis: Internet users versus GDP per capita growth

The regression equation is

Internet users (per 100 people) = 36,7 - 1,35 GDP per capita growth (annual %

Predictor	Coef	SE Coef	T	P
Constant	36,669	2,958	12,40	0,000
GDP per capita growth (annual %	-1,3492	0,6951	-1,94	0,054

S = 27,0167 R-Sq = 2,3% R-Sq(adj) = 1,7%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2749,5	2749,5	3,77	0,054
Residual Error	159	116054,6	729,9		
Total	160	118804,2			

#### Unusual Observations

Obs	GDP per capita growth (annual %	Internet users (per 100 people)	Fit	SE Fit	Residual	St Resid
31	9,8	34,38	23,41	5,23	10,97	0,41 X
50	3,2	86,90	32,29	2,14	54,61	2,03R
63	-6,3	8,37	45,14	6,76	-36,77	-1,41 X
66	-3,8	95,64	41,86	5,19	53,78	2,03R
78	5,9	82,52	28,73	2,95	53,78	2,00R
86	0,8	90,08	35,57	2,60	54,51	2,03R
105	1,2	90,70	35,10	2,47	55,61	2,07R
109	-0,6	93,28	37,43	3,24	55,85	2,08R
113	13,0	19,80	19,07	7,33	0,74	0,03 X
127	12,5	71,13	19,86	6,94	51,27	1,96 X
136	4,7	90,02	30,31	2,46	59,71	2,22R
151	-6,3	78,00	45,18	6,78	32,82	1,25 X

R denotes an observation with a large standardized residual.

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

A equação acima é inválida porque o P-value deu maior que 0,05.

### 5.1.14 Regressão da variável Usuários de Internet x Expectativa de vida ao nascer

Segue o resultado da regressão simples das variáveis Usuários de Internet e Expectativa de vida ao nascer.

### Regression Analysis: Internet users versus Life expectancy

The regression equation is

Internet users (per 100 people) = - 114 + 2,13 Life expectancy at birth, total

Predictor	Coef	SE Coef	T	P
Constant	-114,248	9,701	-11,78	0,000
Life expectancy at birth, total	2,1283	0,1391	15,30	0,000

S = 17,3850 R-Sq = 59,6% R-Sq(adj) = 59,3%

#### Analysis of Variance

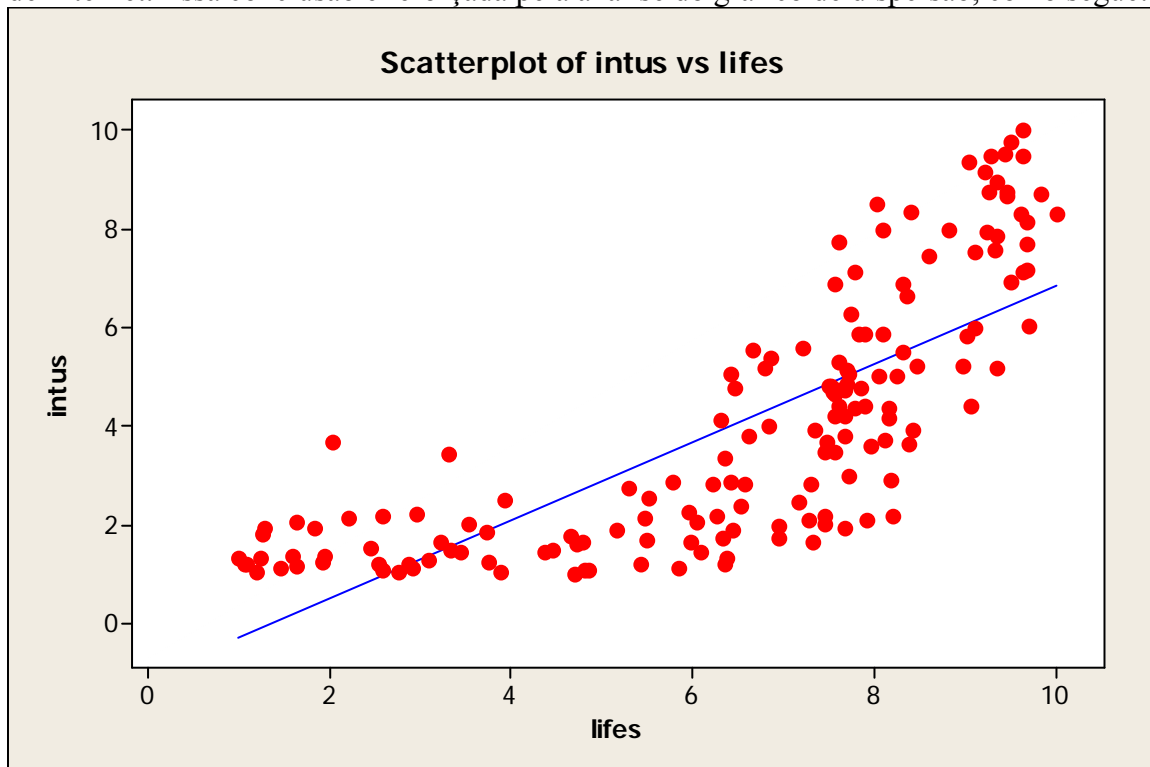
Source	DF	SS	MS	F	P
Regression	1	70748	70748	234,08	0,000
Residual Error	159	48056	302		
Total	160	118804			

## Unusual Observations

Obs	Life expectancy at birth, total	Internet users (per 100 people)	Fit	SE Fit	Residual	St Resid
41	79,1	88,77	54,10	1,96	34,67	2,01R
66	81,5	95,64	59,11	2,20	36,53	2,12R
109	81,0	93,28	58,14	2,16	35,14	2,04R

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação alto da reta de regressão à amostra: a variável Expectativa de vida ao nascer prevê 59,6% do comportamento da variável Usuários de Internet. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.15 Regressão da variável Usuários de Internet x População Urbana

Segue o resultado da regressão simples das variáveis Usuários de Internet e População urbana.

#### Regression Analysis: Internet users versus Urban population

The regression equation is  
 Internet users (per 100 people) = - 12,7 + 0,837 Urban population (% of total)

Predictor	Coef	SE Coef	T	P
Constant	-12,673	4,090	-3,10	0,002
Urban population (% of total)	0,83716	0,06979	12,00	0,000

S = 19,8044 R-Sq = 47,5% R-Sq(adj) = 47,2%

#### Analysis of Variance

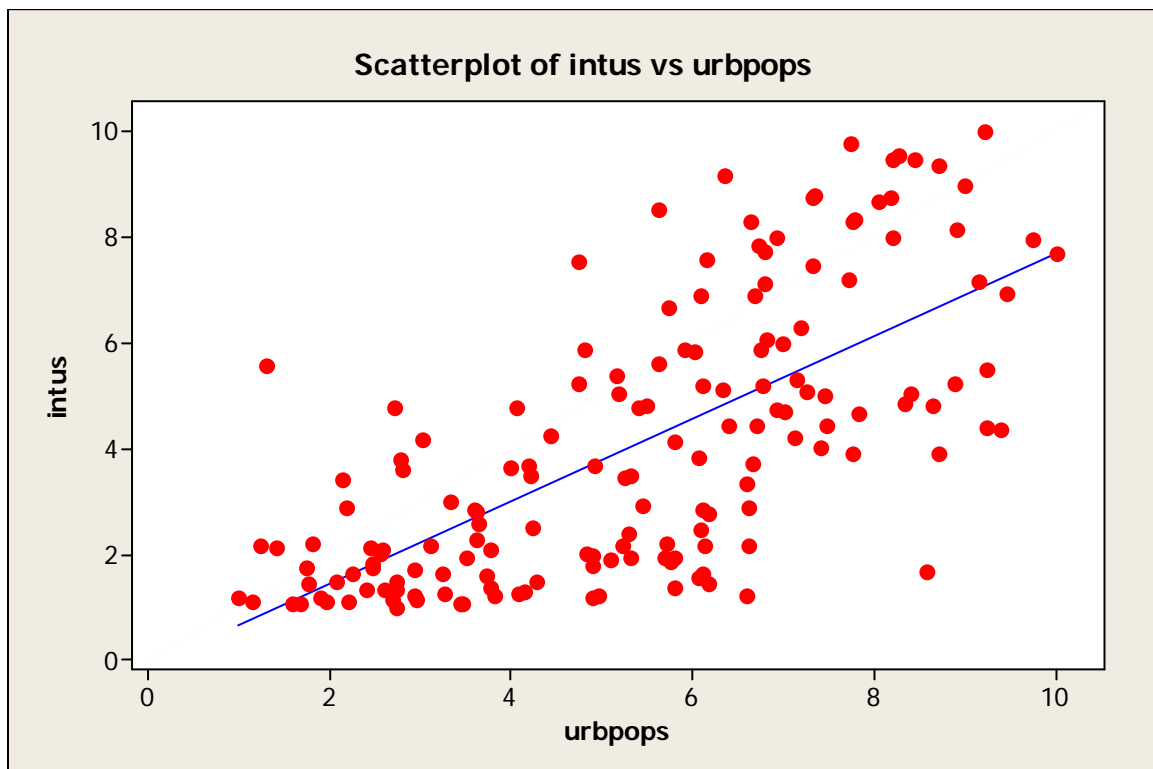
Source	DF	SS	MS	F	P
Regression	1	56442	56442	143,91	0,000
Residual Error	159	62362	392		
Total	160	118804			

## Unusual Observations

Obs	Urban population (% of total)	Internet users (per 100 people)	Fit	SE Fit	Residual	St Resid
50	64	86,90	40,82	1,70	46,08	2,34R
52	86	7,23	59,32	2,71	-52,09	-2,66R
69	66	2,47	42,91	1,78	-40,44	-2,05R
109	78	93,28	52,29	2,26	40,99	2,08R
128	57	79,89	34,88	1,57	45,01	2,28R
129	48	69,34	27,51	1,62	41,83	2,12R
145	14	48,52	-1,04	3,22	49,55	2,54R

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação baixo da reta de regressão à amostra: a variável população prevê apenas 47,5% do comportamento da variável Usuários de Internet. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.16 Regressão da variável Razão de dependência da idade x Taxa de Fertilidade

Segue o resultado da regressão simples das variáveis Razão de dependência da idade e Taxa de fertilidade.

#### Regression Analysis: Age dependency versus Fertility rate

The regression equation is  
Age dependency ratio (% of work = 26,5 + 11,7 Fertility rate, total (births p

Predictor	Coef	SE Coef	T	P
Constant	26,4666	0,9510	27,83	0,000
Fertility rate, total (births p	11,7390	0,2949	39,80	0,000

S = 5,37607    R-Sq = 90,9%    R-Sq(adj) = 90,8%

## Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	45786	45786	1584,16	0,000
Residual Error	159	4595	29		
Total	160	50381			

## Unusual Observations

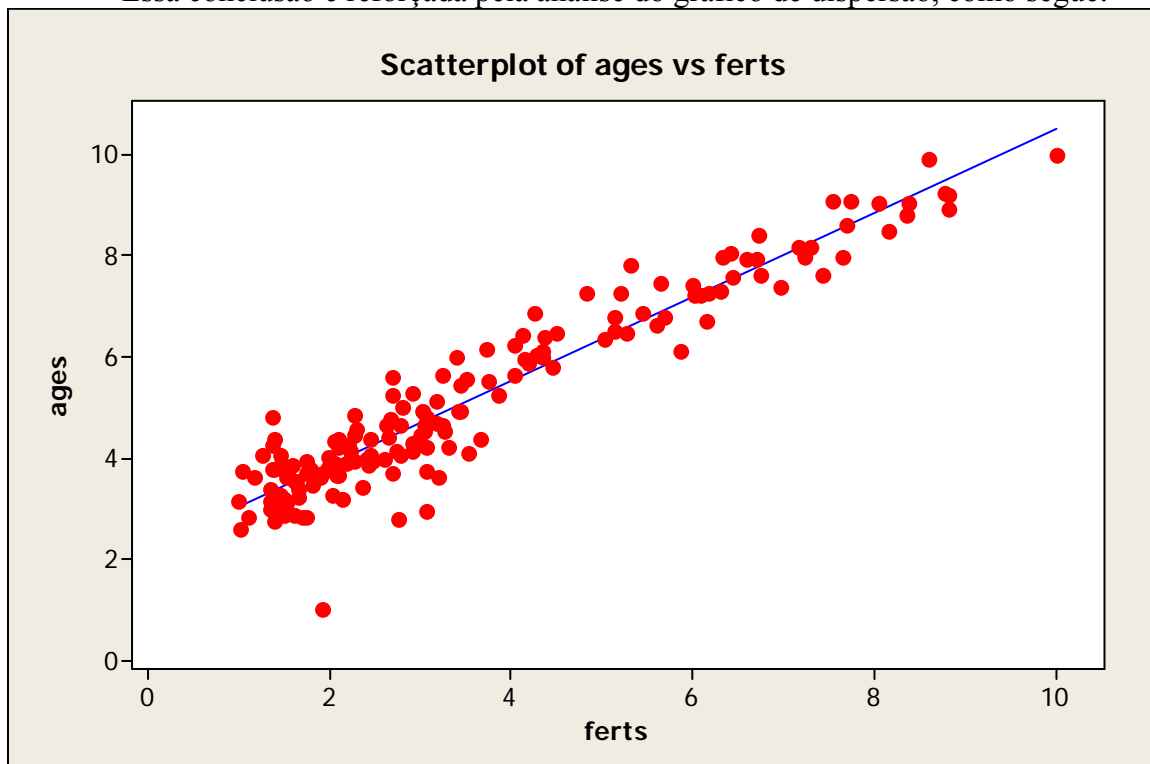
Obs	Age		Fit	SE Fit	Residual	St Resid
	Fertility rate, total (births p	dependency ratio (%) of work				
1	6,29	94,855	100,281	1,089	-5,426	-1,03 X
9	2,30	37,839	53,466	0,458	-15,628	-2,92R
45	2,25	63,810	52,879	0,463	10,930	2,04R
59	3,98	84,558	73,223	0,533	11,335	2,12R
74	1,39	56,364	42,784	0,612	13,580	2,54R
76	2,59	45,481	56,871	0,433	-11,389	-2,13R
92	6,29	97,482	100,352	1,091	-2,870	-0,55 X
107	7,06	104,805	109,379	1,303	-4,574	-0,88 X
126	2,50	39,063	55,814	0,439	-16,752	-3,13R
149	6,15	103,860	98,650	1,051	5,211	0,99 X
151	1,75	21,161	46,998	0,540	-25,837	-4,83R
160	6,26	97,751	99,929	1,081	-2,178	-0,41 X

R denotes an observation with a large standardized residual.

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

A equação acima revela um grau de adaptação altíssimo da reta de regressão à amostra: a variável Taxa de Fertilidade prevê 90,9% do comportamento da variável Razão de dependência da idade. Isso significa que a Razão de dependência da idade está sendo muito mais influenciada pelas crianças, que são dependentes dos adultos, hoje em número muito maior que os idosos. No futuro, talvez a Razão de dependência da idade venha a ser mais influenciada pela Expectativa de vida, mas só o tempo irá dizer.

Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:





### 5.1.17 Regressão da variável Razão de dependência da idade x PIB Per Capita

Segue o resultado da regressão simples das variáveis Razão de dependência da idade e PIB per capita.

#### Regression Analysis: Age dependency rate versus GDP per capita

The regression equation is

Age dependency ratio (% of work = 65,3 - 0,000432 GDP per capita (current US\$)

Predictor	Coef	SE Coef	T	P
Constant	65,266	1,527	42,73	0,000
GDP per capita (current US\$)	-0,00043240	0,00007421	-5,83	0,000

S = 16,1588 R-Sq = 17,6% R-Sq(adj) = 17,1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	8865,0	8865,0	33,95	0,000
Residual Error	159	41516,1	261,1		
Total	160	50381,1			

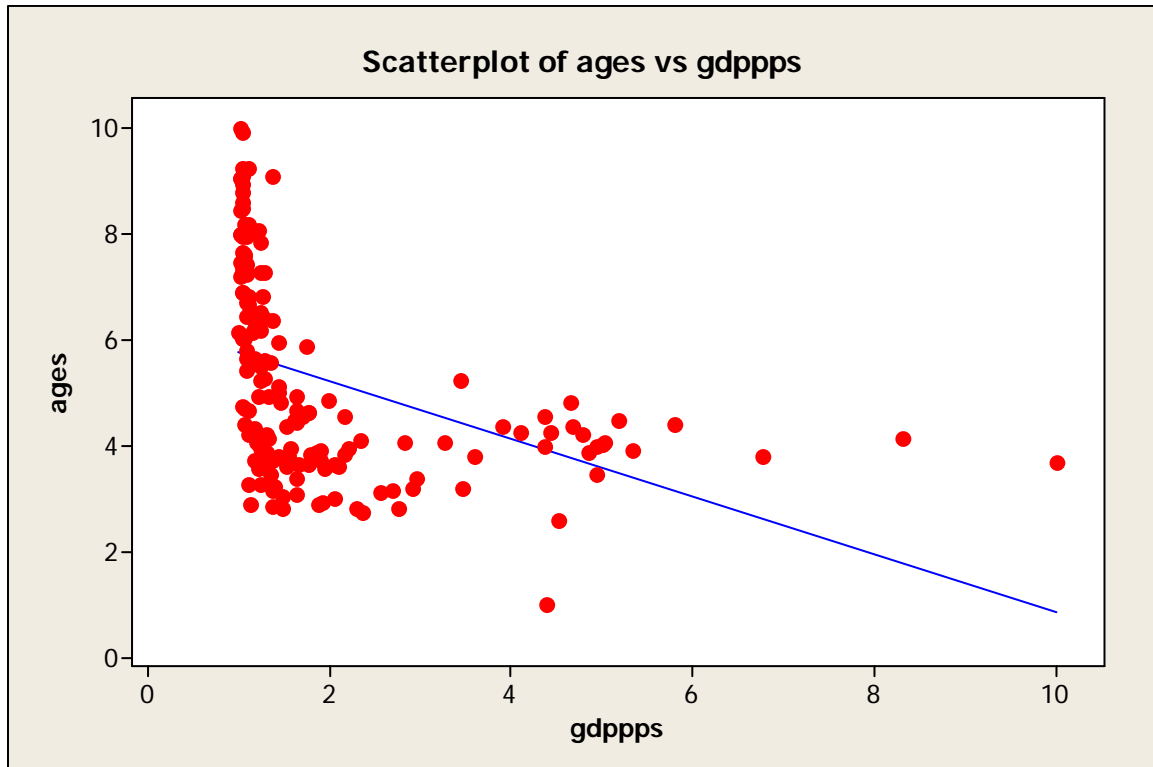
Unusual Observations

Obs	GDP per capita (current US\$)	Age dependency ratio (% of work)	Fit	SE Fit	Residual	St Resid
4	4451	96,29	63,34	1,37	32,95	2,05R
7	50748	48,00	43,32	3,19	4,67	0,30 X
41	56245	52,58	40,95	3,57	11,64	0,74 X
86	105195	46,11	19,78	7,08	26,33	1,81 X
92	602	97,48	65,01	1,50	32,48	2,02R
107	358	104,81	65,11	1,51	39,69	2,47R
109	85389	50,11	28,34	5,64	21,76	1,44 X
137	67457	46,92	36,10	4,35	10,82	0,70 X
149	509	103,86	65,05	1,51	38,81	2,41R
160	1253	97,75	64,72	1,48	33,03	2,05R

R denotes an observation with a large standardized residual.

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

A equação acima revela um grau de adaptação muito baixo da reta de regressão à amostra: a variável PIB per capita prevê apenas 17,6% do comportamento da variável Razão de dependência da idade. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.18 Regressão da variável Razão de dependência da idade x % Crescimento do PIB Per Capita

Segue o resultado da regressão simples das variáveis Razão de dependência da idade e % Crescimento PIB per capita.

#### Regression Analysis: Age dependency rate versus GDP per capita growth

The regression equation is  
Age dependency ratio (% of work = 60,9 - 0,181 GDP per capita growth (annual %

Predictor	Coef	SE Coef	T	P
Constant	60,887	1,948	31,25	0,000
GDP per capita growth (annual %	-0,1809	0,4578	-0,40	0,693

S = 17,7919 R-Sq = 0,1% R-Sq(adj) = 0,0%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	49,4	49,4	0,16	0,693
Residual Error	159	50331,7	316,6		
Total	160	50381,1			

#### Unusual Observations

Obs	GDP per capita growth (annual %	Age dependency ratio (% of work	Fit	SE Fit	Residual	St Resid
4	3,0	96,29	60,35	1,40	35,93	2,03R
31	9,8	38,21	59,11	3,45	-20,90	-1,20 X
34	4,3	95,87	60,10	1,54	35,76	2,02R
63	-6,3	67,64	62,02	4,45	5,62	0,33 X
89	3,8	95,76	60,20	1,46	35,56	2,01R
92	1,4	97,48	60,64	1,58	36,84	2,08R
107	5,0	104,81	59,98	1,69	44,83	2,53R
113	13,0	63,02	58,53	4,83	4,49	0,26 X

127	12,5	35,91	58,63	4,57	-22,72	-1,32 X
142	5,1	96,34	59,96	1,72	36,37	2,05R
149	1,9	103,86	60,55	1,49	43,31	2,44R
151	-6,3	21,16	62,03	4,47	-40,87	-2,37RX
160	5,9	97,75	59,82	1,95	37,93	2,15R

R denotes an observation with a large standardized residual.  
X denotes an observation whose X value gives it large leverage.

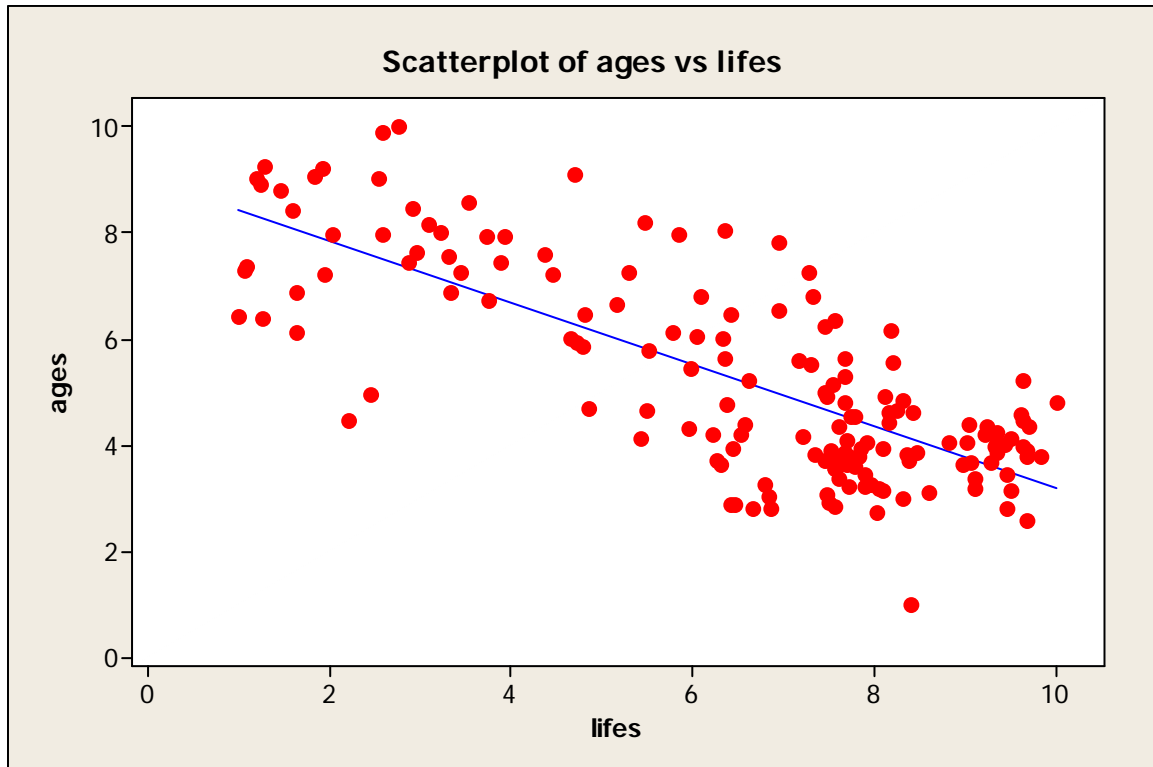
A equação acima é inválida porque o P Value é maior que 0,05.

### 5.1.19 Regressão da variável Razão de dependência da idade x Expectativa de Vida ao nascer

Segue o resultado da regressão simples das variáveis Razão de dependência da idade e Expectativa de vida ao nascer.

Regression Analysis: Age dependency rate versus Life expectancy						
The regression equation is						
Age dependency ratio (% of work = 155 - 1,38 Life expectancy at birth, total						
Predictor		Coef	SE Coef	T		P
Constant		155,287	6,389	24,30		0,000
Life expectancy at birth, total		-1,37516	0,09162	-15,01		0,000
S = 11,4501 R-Sq = 58,6% R-Sq(adj) = 58,4%						
Analysis of Variance						
Source	DF	SS	MS	F		P
Regression	1	29535	29535	225,28		0,000
Residual Error	159	20846	131			
Total	160	50381				
Unusual Observations						
Obs	Life expectancy at birth, total	Age dependency ratio (% of work)	Fit	SE Fit	Residual	St Resid
19	53,1	57,750	82,253	1,716	-24,504	-2,16R
59	70,8	84,558	57,891	0,917	26,667	2,34R
69	68,5	86,667	61,108	0,904	25,559	2,24R
107	54,3	104,805	80,663	1,627	24,142	2,13R
131	52,1	53,292	83,667	1,796	-30,375	-2,69R
142	62,0	96,335	70,028	1,109	26,307	2,31R
144	72,2	79,310	56,069	0,946	23,242	2,04R
151	76,6	21,161	49,986	1,136	-28,825	-2,53R
R denotes an observation with a large standardized residual.						

A equação acima revela um grau de adaptação alto da reta de regressão à amostra: a variável Expectativa de vida ao nascer prevê 58,6% do comportamento da variável Razão de dependência da idade. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.20 Regressão da variável Razão de dependência da idade x População Urbana

Segue o resultado da regressão simples das variáveis Razão de dependência da idade e População urbana.

#### Regression Analysis: Age dependency rate versus Urban population

The regression equation is  
Age dependency ratio (% of work = 83,2 - 0,421 Urban population (% of total)

Predictor	Coef	SE Coef	T	P
Constant	83,185	3,111	26,74	0,000
Urban population (% of total)	-0,42143	0,05308	-7,94	0,000

S = 15,0634 R-Sq = 28,4% R-Sq(adj) = 27,9%

#### Analysis of Variance

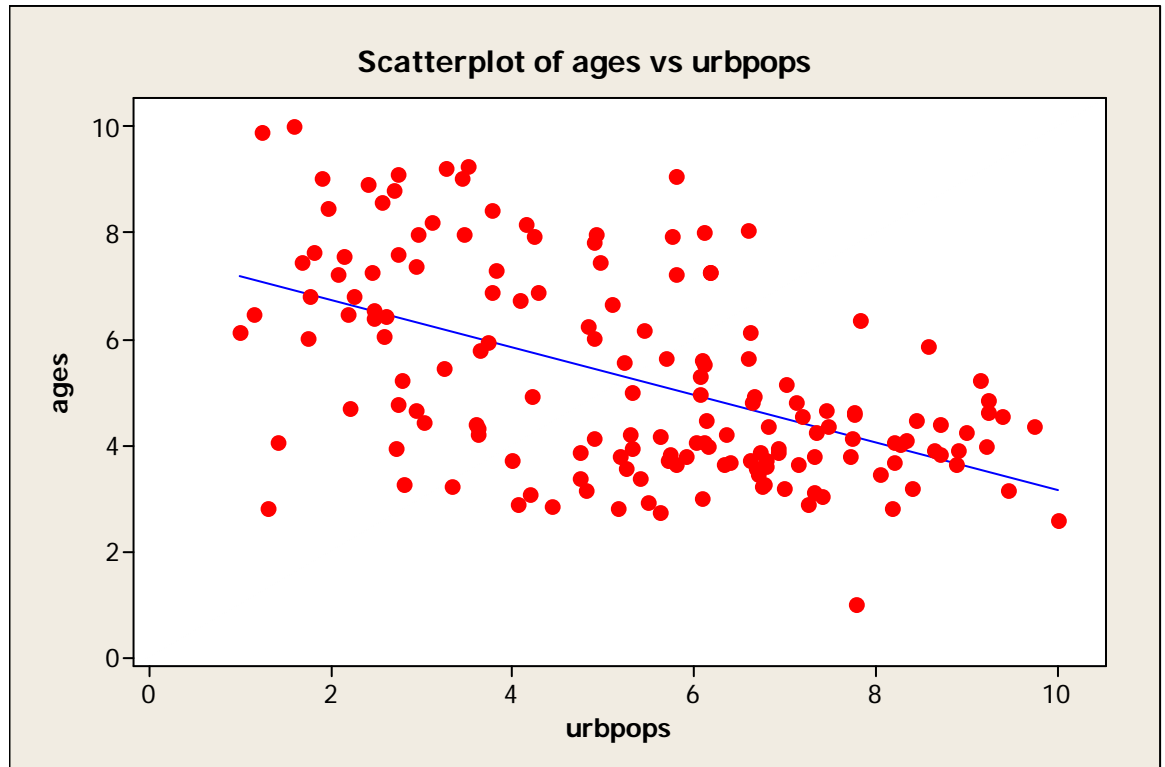
Source	DF	SS	MS	F	P
Regression	1	14303	14303	63,04	0,000
Residual Error	159	36078	227		
Total	160	50381			

#### Unusual Observations

Obs	Urban population (% of total)	Age dependency ratio (% of work)	Fit	SE Fit	Residual	St Resid
4	59	96,29	58,53	1,21	37,76	2,51R
69	66	86,67	55,20	1,35	31,46	2,10R
145	14	37,92	77,33	2,45	-39,40	-2,65R

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação relativamente baixo da reta de regressão à amostra: a variável População urbana capita prevê apenas 28,4% do comportamento da variável Razão de dependência da idade. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.21 Regressão da variável Taxa de Fertilidade x PIB per Capita

Segue o resultado da regressão simples das variáveis Taxa de fertilidade e PIB per capita.

#### Regression Analysis: Fertility rate versus GDP per capita

The regression equation is  
Fertility rate, total (births p = 3,32 - 0,000038 GDP per capita (current US\$)

Predictor	Coef	SE Coef	T	P
Constant	3,3175	0,1218	27,24	0,000
GDP per capita (current US\$)	-0,00003792	0,00000592	-6,41	0,000

S = 1,28872 R-Sq = 20,5% R-Sq(adj) = 20,0%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	68,186	68,186	41,06	0,000
Residual Error	159	264,067	1,661		
Total	160	332,253			

#### Unusual Observations

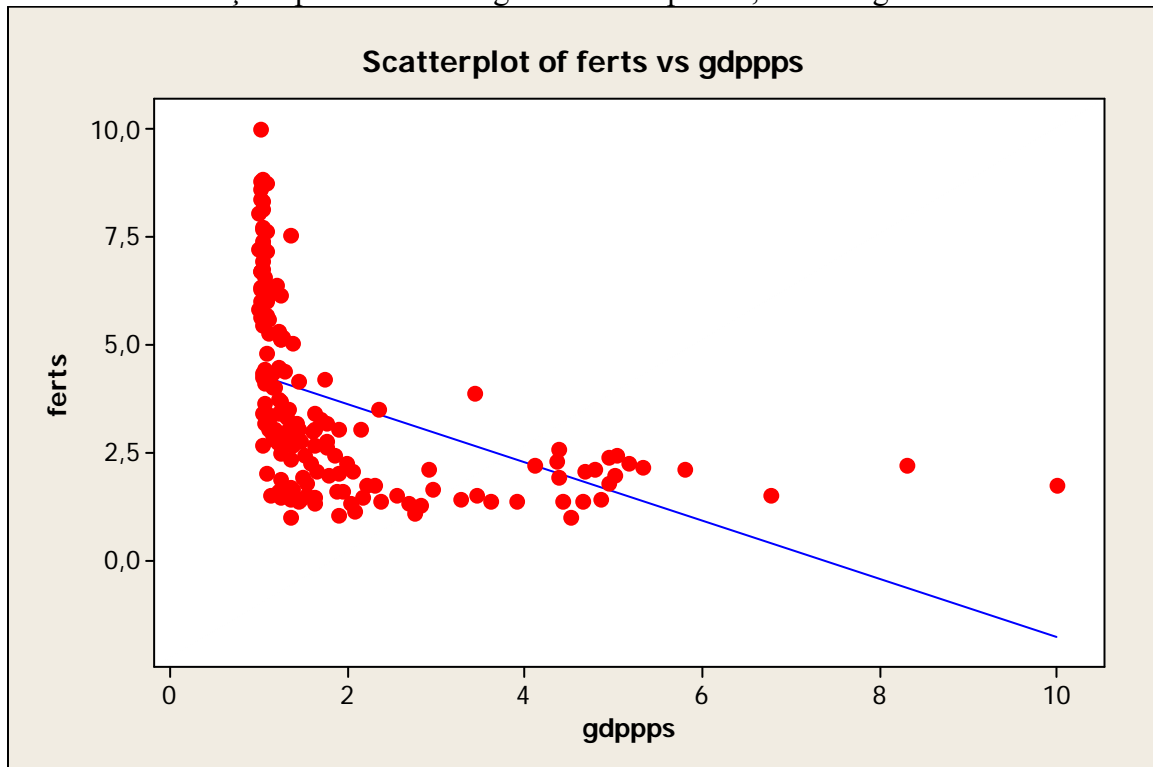
Obs	GDP per capita (current US\$)	Fertility rate, total (births p)	Fit	SE Fit	Residual	St Resid
1	501	6,288	3,298	0,120	2,990	2,33R
7	50748	1,920	1,393	0,254	0,527	0,42 X
29	676	5,981	3,292	0,120	2,689	2,10R
41	56245	1,870	1,185	0,284	0,685	0,55 X
86	105195	1,630	-0,672	0,565	2,302	1,99 X
89	339	5,990	3,305	0,121	2,685	2,09R
92	602	6,294	3,295	0,120	2,999	2,34R
107	358	7,063	3,304	0,121	3,759	2,93R
109	85389	1,950	0,079	0,450	1,871	1,55 X

137	67457	1,500	0,759	0,347	0,741	0,60 X
149	509	6,149	3,298	0,120	2,851	2,22R
160	1253	6,258	3,270	0,118	2,988	2,33R

R denotes an observation with a large standardized residual.

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

A equação acima revela um grau de adaptação baixo da reta de regressão à amostra: a variável PIB per capita prevê 20,5% do comportamento da variável Taxa de fertilidade. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



Visualmente parece claro que a interdependência entre as variáveis tem caráter exponencial e não linear, explicando a baixa aderência da reta de regressão. Esta observação está de acordo com o senso comum, pois, em geral, quanto mais rico o país, mais a população é informada de métodos contraceptivos e menor é a taxa de fertilidade.

### 5.1.22 Regressão da variável Taxa de Fertilidade x % Crescimento do PIB per Capita

Segue o resultado da regressão simples das variáveis Taxa de fertilidade e % Crescimento PIB per capita.

#### Regression Analysis: Fertility rate versus GDP per capita growth

The regression equation is

Fertility rate, total (births p = 2,91 - 0,0075 GDP per capita growth (annual %

Predictor	Coef	SE Coef	T	P
Constant	2,9088	0,1583	18,38	0,000
GDP per capita growth (annual %	-0,00750	0,03719	-0,20	0,840

S = 1,44537 R-Sq = 0,0% R-Sq(adj) = 0,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0,085	0,085	0,04	0,840

Residual Error	159	332,168	2,089			
Total	160	332,253				
Unusual Observations						
	GDP per capita growth (annual %)	Fertility rate, total (births p)	Fit	SE Fit	Residual	St Resid
Obs						
1	5,2	6,288	2,870	0,142	3,418	2,38R
22	6,0	5,850	2,864	0,162	2,986	2,08R
29	1,6	5,981	2,897	0,124	3,084	2,14R
31	9,8	1,598	2,835	0,280	-1,237	-0,87 X
34	4,3	5,775	2,876	0,125	2,899	2,01R
63	-6,3	3,340	2,956	0,362	0,384	0,27 X
89	3,8	5,990	2,880	0,118	3,110	2,16R
92	1,4	6,294	2,898	0,128	3,396	2,36R
107	5,0	7,063	2,871	0,137	4,192	2,91R
113	13,0	2,954	2,811	0,392	0,143	0,10 X
127	12,5	1,150	2,815	0,371	-1,665	-1,19 X
149	1,9	6,149	2,895	0,121	3,254	2,26R
151	-6,3	1,749	2,956	0,363	-1,207	-0,86 X
160	5,9	6,258	2,864	0,159	3,394	2,36R
R denotes an observation with a large standardized residual.						
X denotes an observation whose X value gives it large leverage.						

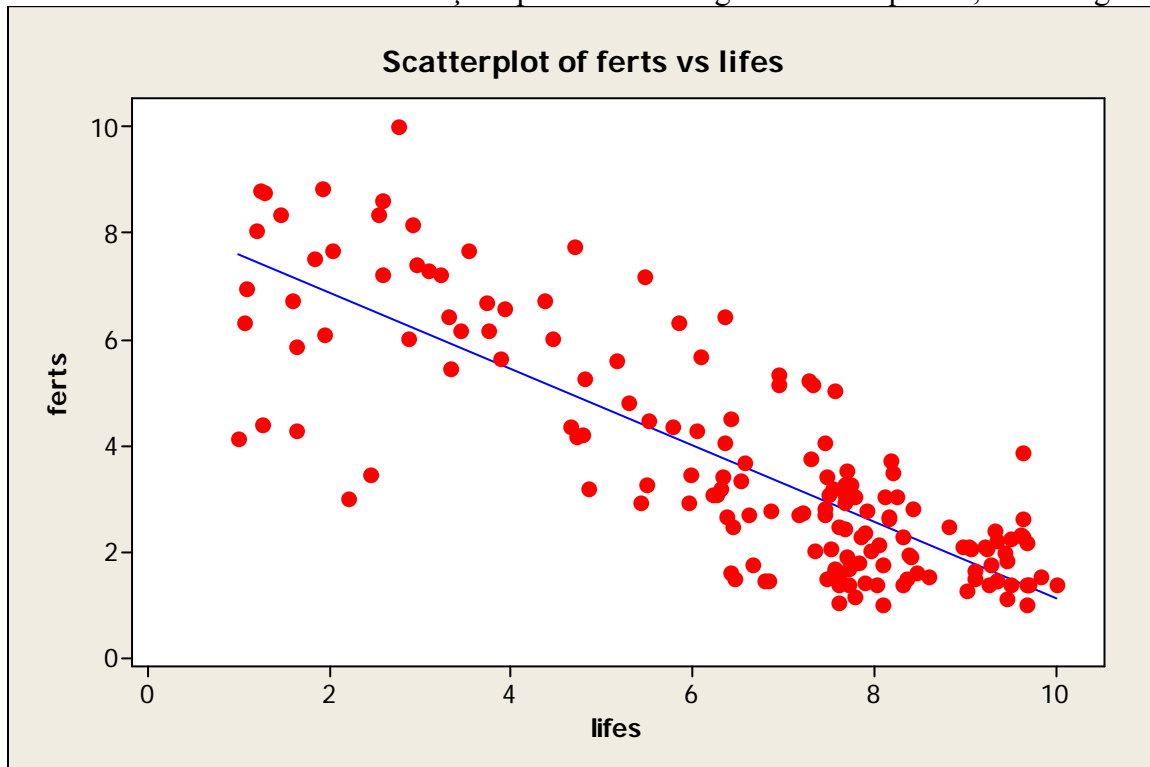
A equação acima resultou inválida, porque o P-value é maior que 0,05.

### 5.1.23 Regressão da variável Taxa de Fertilidade x Expectativa de Vida ao nascer

Segue o resultado da regressão simples das variáveis Taxa de fertilidade e Expectativa de vida ao nascer.

Regression Analysis: Fertility rate versus Life expectancy						
The regression equation is						
Fertility rate, total (births p = 11,2 - 0,120 Life expectancy at birth, total						
Predictor		Coef	SE Coef	T	P	
Constant		11,1604	0,4597	24,28	0,000	
Life expectancy at birth, total		-0,119848	0,006592	-18,18	0,000	
S = 0,823845 R-Sq = 67,5% R-Sq(adj) = 67,3%						
Analysis of Variance						
Source	DF	SS	MS	F	P	
Regression	1	224,34	224,34	330,53	0,000	
Residual Error	159	107,92	0,68			
Total	160	332,25				
Unusual Observations						
	Life expectancy at birth, total	Fertility rate, total (births p)	Fit	SE Fit	Residual	St Resid
Obs						
19	53,1	2,7500	4,7953	0,1234	-2,0453	-2,51R
69	68,5	4,7020	2,9525	0,0650	1,7495	2,13R
71	81,5	3,0300	1,3922	0,1047	1,6378	2,00R
83	47,4	3,1990	5,4838	0,1569	-2,2848	-2,83R
107	54,3	7,0630	4,6568	0,1170	2,4062	2,95R
131	52,1	2,4580	4,9185	0,1293	-2,4605	-3,02R
135	48,3	3,3640	5,3666	0,1511	-2,0026	-2,47R
142	62,0	5,5780	3,7298	0,0798	1,8482	2,25R
159	65,0	5,2000	3,3666	0,0701	1,8334	2,23R
161	49,9	3,2900	5,1847	0,1421	-1,8947	-2,33R
R denotes an observation with a large standardized residual.						

A equação acima revela um grau de adaptação alto da reta de regressão à amostra: a variável Expectativa de vida ao nascer prevê 67,5% do comportamento da variável Taxa de Fertilidade. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



#### 5.1.24 Regressão da variável Taxa de Fertilidade x População Urbana

Segue o resultado da regressão simples das variáveis Taxa de fertilidade e População urbana.

##### Regression Analysis: Fertility rate versus Urban population

The regression equation is  
Fertility rate, total (births p = 4,84 - 0,0361 Urban population (% of total)

Predictor	Coef	SE Coef	T	P
Constant	4,8421	0,2470	19,61	0,000
Urban population (% of total)	-0,036094	0,004213	-8,57	0,000

S = 1,19574 R-Sq = 31,6% R-Sq(adj) = 31,1%

##### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	104,92	104,92	73,38	0,000
Residual Error	159	227,34	1,43		
Total	160	332,25			

##### Unusual Observations

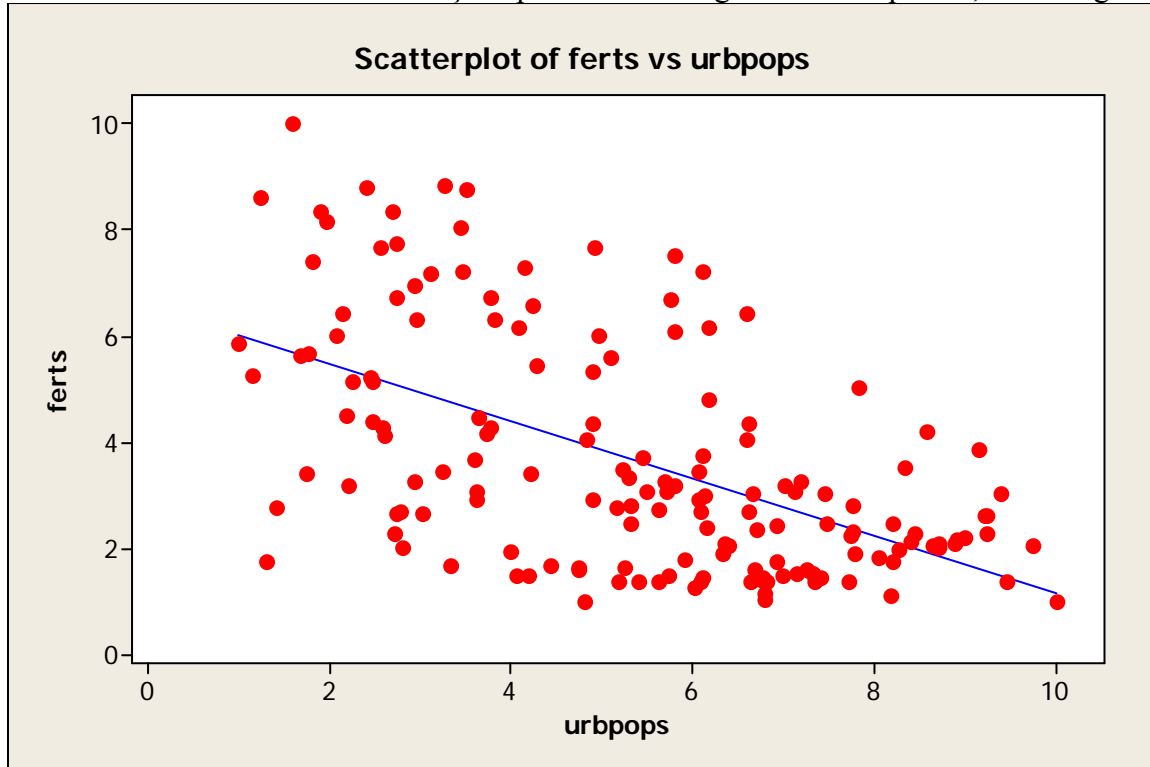
Obs	Urban population (% of total)	Fertility rate, total (births p	Fit	SE Fit	Residual	St Resid
4	59	5,4430	2,7306	0,0960	2,7124	2,28R
84	62	5,2380	2,6224	0,0992	2,6156	2,20R
92	33	6,2940	3,6402	0,1289	2,6538	2,23R
107	17	7,0630	4,2393	0,1839	2,8237	2,39R
108	50	5,5250	3,0446	0,0960	2,4804	2,08R
145	14	1,6390	4,3404	0,1941	-2,7014	-2,29R



160                      36                      6,2580    3,5536    0,1222    2,7044                      2,27R

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação baixo da reta de regressão à amostra: a variável População urbana prevê apenas 31,6 do comportamento da variável Taxa de Fertilidade. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.25 Regressão da variável PIB per Capita x % Crescimento do PIB per Capita

Segue o resultado da regressão simples das variáveis PIB per capita e % Crescimento PIB per capita.

#### Regression Analysis: GDP per capita versus GDP per capita growth

The regression equation is

GDP per capita (current US\$) = 14949 - 1214 GDP per capita growth (annual %

Predictor	Coef	SE Coef	T	P
Constant	14949	1846	8,10	0,000
GDP per capita growth (annual %	-1213,9	433,8	-2,80	0,006

S = 16858,1    R-Sq = 4,7%    R-Sq(adj) = 4,1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2225791686	2225791686	7,83	0,006
Residual Error	159	45187260137	284196605		
Total	160	47413051823			

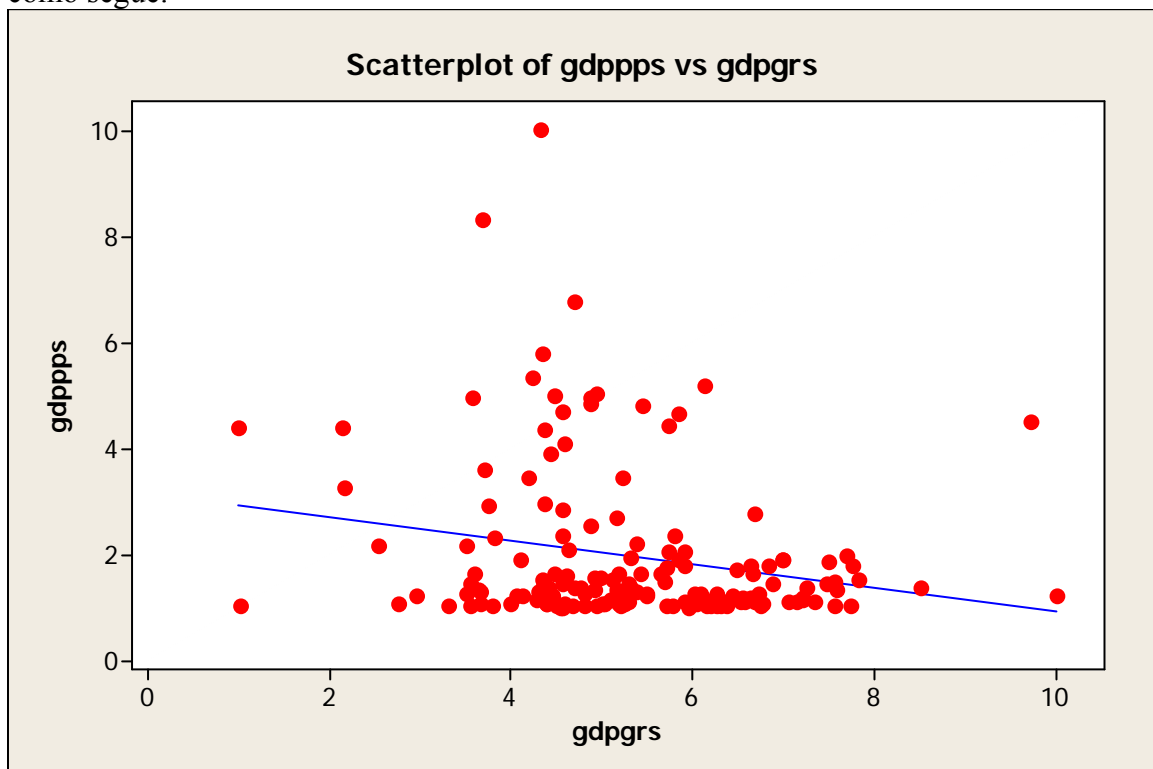
Unusual Observations

Obs	GDP per capita growth (annual %	GDP per capita (current US\$)	Fit	SE Fit	Residual	St Resid
7	0,7	50748	14142	1659	36605	2,18R

26	2,0	46212	12501	1390	33711	2,01R
31	9,8	4428	3017	3265	1411	0,09 X
41	0,9	56245	13906	1610	42339	2,52R
63	-6,3	671	22566	4218	-21894	-1,34 X
86	0,8	105195	13964	1621	91231	5,44R
109	-0,6	85389	15634	2024	69754	4,17R
113	13,0	2840	-889	4575	3730	0,23 X
127	12,5	41120	-171	4330	41291	2,53RX
136	4,7	48897	9225	1532	39671	2,36R
137	1,6	67457	12963	1447	54494	3,24R
151	-6,3	39623	22605	4231	17018	1,04 X
153	2,1	47153	12349	1375	34804	2,07R

R denotes an observation with a large standardized residual.  
X denotes an observation whose X value gives it large leverage.

A equação acima revela um grau de adaptação baixíssimo da reta de regressão à amostra: a variável % Crescimento PIB per capita prevê somente 4,7% do comportamento da variável PIB per Capita. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.26 Regressão da variável PIB per Capita x Expectativa de Vida ao nascer

Segue o resultado da regressão simples das variáveis PIB per capita e Expectativa de vida ao nascer.

#### Regression Analysis: GDP per capita versus Life expectancy

The regression equation is  
GDP per capita (current US\$) = - 58800 + 1016 Life expectancy at birth, total

Predictor	Coef	SE Coef	T	P
Constant	-58800	7827	-7,51	0,000
Life expectancy at birth, total	1016,3	112,2	9,06	0,000

S = 14026,2 R-Sq = 34,0% R-Sq(adj) = 33,6%

Analysis of Variance

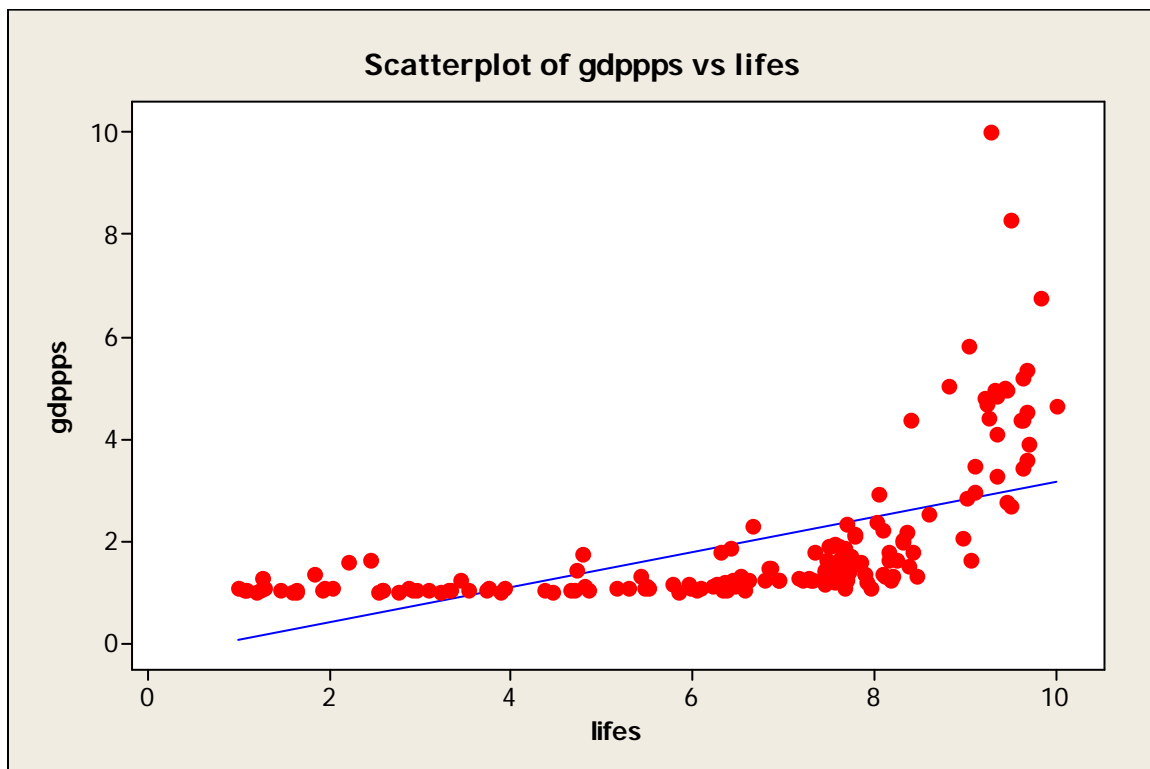
Source	DF	SS	MS	F	P
Regression	1	16132309354	16132309354	82,00	0,000
Residual Error	159	31280742469	196734229		
Total	160	47413051823			

Unusual Observations							
Obs	Life expectancy at birth, total	GDP per capita (current US\$)	Fit	SE Fit	Residual	St Resid	
41	79,1	56245	21591	1580	34653	2,49R	
86	80,1	105195	22595	1662	82599	5,93R	
109	81,0	85389	23520	1739	61869	4,45R	
137	82,2	67457	24789	1849	42668	3,07R	

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação baixo da reta de regressão à amostra: a variável Expectativa de vida ao nascer prevê 34% do comportamento da variável PIB per capita. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



Neste caso, mais uma vez a aderência não parece ser linear. É senso comum que, quanto maior a renda de um país, mais acesso à saúde, saneamento básico e condições de vida, portanto, maior a expectativa de vida esperada.

### 5.1.27 Regressão da variável PIB per Capita x População Urbana

Segue o resultado da regressão simples das variáveis PIB per capita e População urbana.

#### Regression Analysis: GDP per capita versus Urban population

The regression equation is  
 GDP per capita (current US\$) = - 12275 + 436 Urban population (% of total)

Predictor	Coef	SE Coef	T	P
Constant	-12275	2934	-4,18	0,000
Urban population (% of total)	436,29	50,05	8,72	0,000

S = 14204,9   R-Sq = 32,3%   R-Sq(adj) = 31,9%

Analysis of Variance

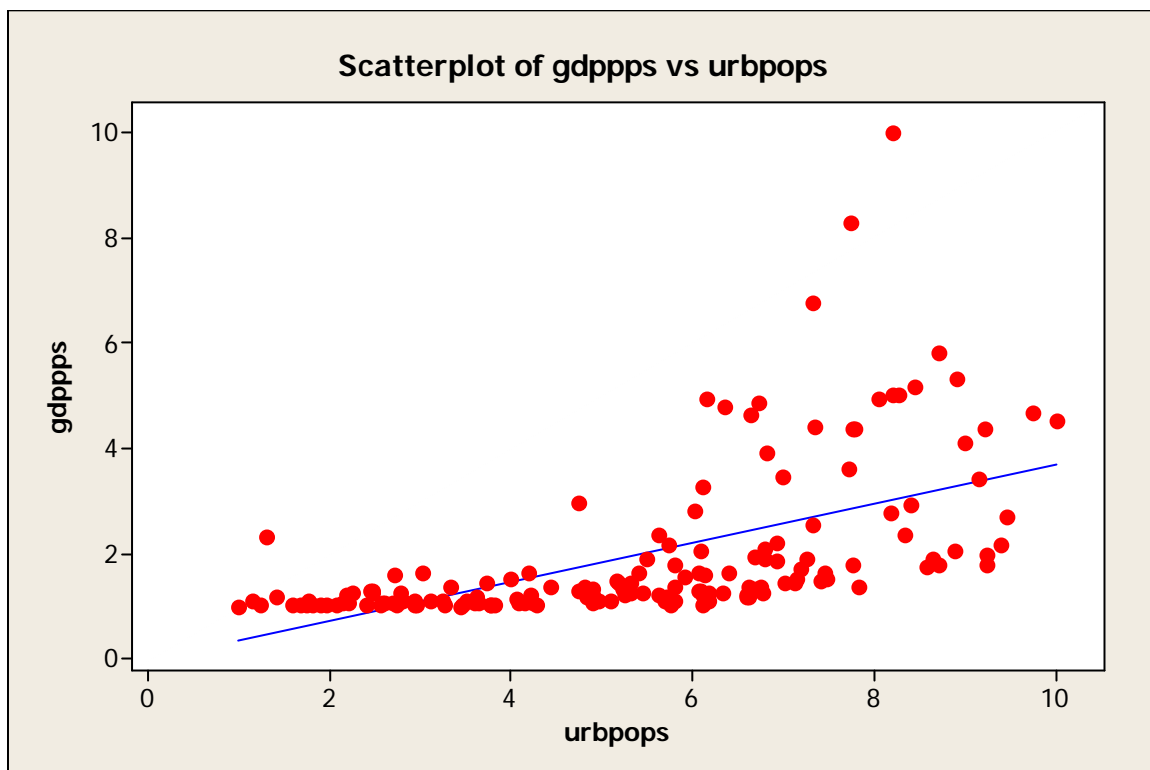
Source	DF	SS	MS	F	P
Regression	1	15330145940	15330145940	75,97	0,000
Residual Error	159	32082905883	201779282		
Total	160	47413051823			

Unusual Observations

Obs	Urban population (% of total)	GDP per capita (current US\$)	Fit	SE Fit	Residual	St Resid
41	87	56245	25770	1996	30475	2,17R
50	64	44378	15604	1221	28774	2,03R
70	62	46170	14731	1184	31439	2,22R
86	82	105195	23588	1795	81606	5,79R
109	78	85389	21581	1621	63807	4,52R
137	74	67457	19836	1483	47621	3,37R

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação baixo da reta de regressão à amostra: a variável População urbana prevê apenas 32,3% do comportamento da variável PIB per capita. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.28 Regressão da variável % Crescimento do PIB per Capita x Expectativa de Vida ao nascer

Segue o resultado da regressão simples das variáveis PIB per capita e Expectativa de vida ao nascer.

### Regression Analysis: GDP per capita growth versus Life expectancy

The regression equation is  
 GDP per capita growth (annual % = 5,03 - 0,0301 Life expectancy at birth, total

Predictor	Coef	SE Coef	T	P
Constant	5,034	1,712	2,94	0,004
Life expectancy at birth, total	-0,03012	0,02455	-1,23	0,222

S = 3,06777 R-Sq = 0,9% R-Sq(adj) = 0,3%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	14,166	14,166	1,51	0,222
Residual Error	159	1496,386	9,411		
Total	160	1510,551			

#### Unusual Observations

Obs	Life expectancy at birth, total	GDP per capita growth (annual %	Fit	SE Fit	Residual	St Resid
31	73,3	9,829	2,827	0,263	7,002	2,29R
57	80,4	-3,800	2,613	0,369	-6,413	-2,11R
63	61,8	-6,275	3,173	0,301	-9,449	-3,09R
66	81,5	-3,849	2,581	0,389	-6,430	-2,11R
113	72,3	13,047	2,857	0,255	10,191	3,33R
127	81,6	12,456	2,575	0,393	9,881	3,25R
151	76,6	-6,307	2,727	0,304	-9,035	-2,96R

R denotes an observation with a large standardized residual.

A equação acima resultou inválida, porque o P-value é maior que 0,05.

### 5.1.29 Regressão da variável % Crescimento do PIB per Capita x População Urbana

Segue o resultado da regressão simples das variáveis % Crescimento PIB per capita e População urbana.

#### Regression Analysis: Life expectancy versus Urban population

The regression equation is  
 Life expectancy at birth, total = 54,4 + 0,270 Urban population (% of total)

Predictor	Coef	SE Coef	T	P
Constant	54,382	1,615	33,66	0,000
Urban population (% of total)	0,27047	0,02756	9,81	0,000

S = 7,82145 R-Sq = 37,7% R-Sq(adj) = 37,3%

#### Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	5891,5	5891,5	96,31	0,000
Residual Error	159	9726,8	61,2		
Total	160	15618,3			

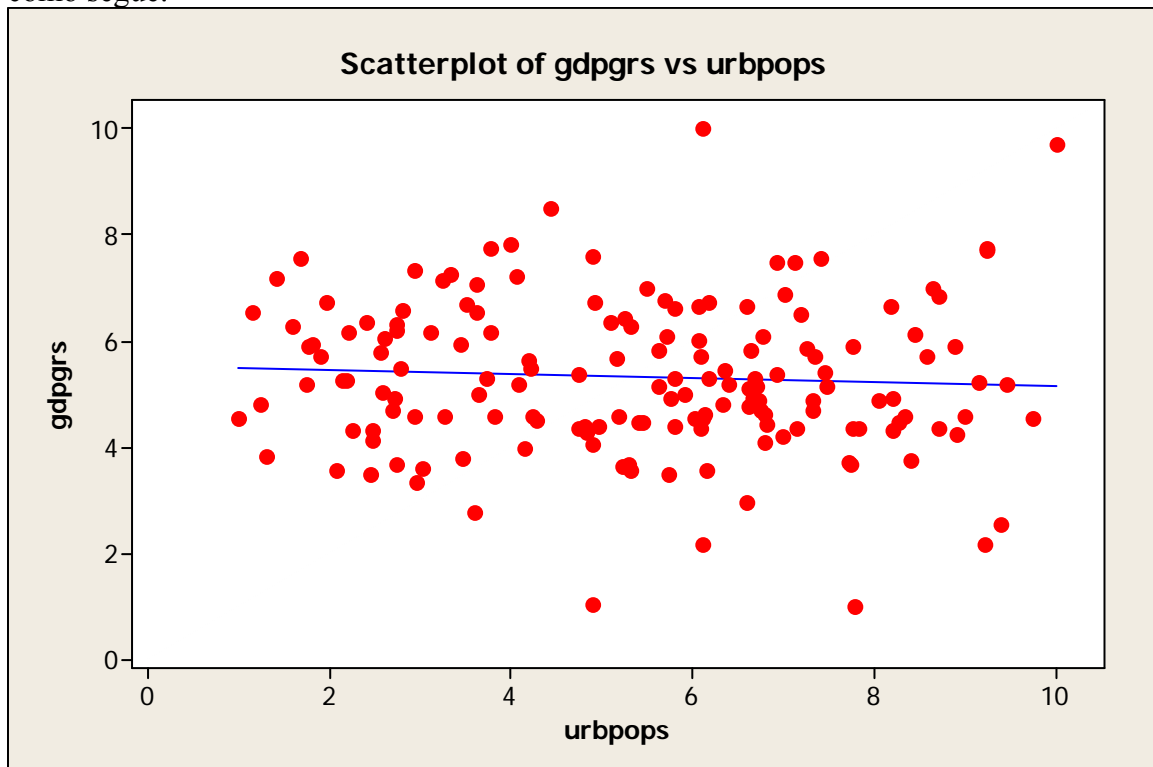
#### Unusual Observations

Obs	Urban population (% of total)	Life expectancy at birth, total	Fit	SE Fit	Residual	St Resid
4	59	50,654	70,204	0,628	-19,551	-2,51R
19	61	53,110	70,908	0,645	-17,798	-2,28R
25	58	51,063	70,177	0,627	-19,115	-2,45R

28	39	47,618	64,903	0,747	-17,285	-2,22R
34	35	48,070	63,902	0,808	-15,833	-2,04R
108	50	51,410	67,851	0,628	-16,441	-2,11R
131	62	52,081	71,070	0,650	-18,988	-2,44R
133	15	74,723	58,466	1,241	16,257	2,11R
160	36	48,455	64,038	0,800	-15,582	-2,00R

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação baixo da reta de regressão à amostra: a variável População urbana prevê apenas 37,7% do comportamento da variável % Crescimento PIB per capita. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



### 5.1.30 Regressão da variável Expectativa de Vida ao nascer x População Urbana

Segue o resultado da regressão simples das variáveis Expectativa de vida ao nascer e População urbana.

#### Regression Analysis: Life expectancy versus Urban population

The regression equation is  
Life expectancy at birth, total = 54,4 + 0,270 Urban population (% of total)

Predictor	Coef	SE Coef	T	P
Constant	54,382	1,615	33,66	0,000
Urban population (% of total)	0,27047	0,02756	9,81	0,000

S = 7,82145 R-Sq = 37,7% R-Sq(adj) = 37,3%

#### Analysis of Variance

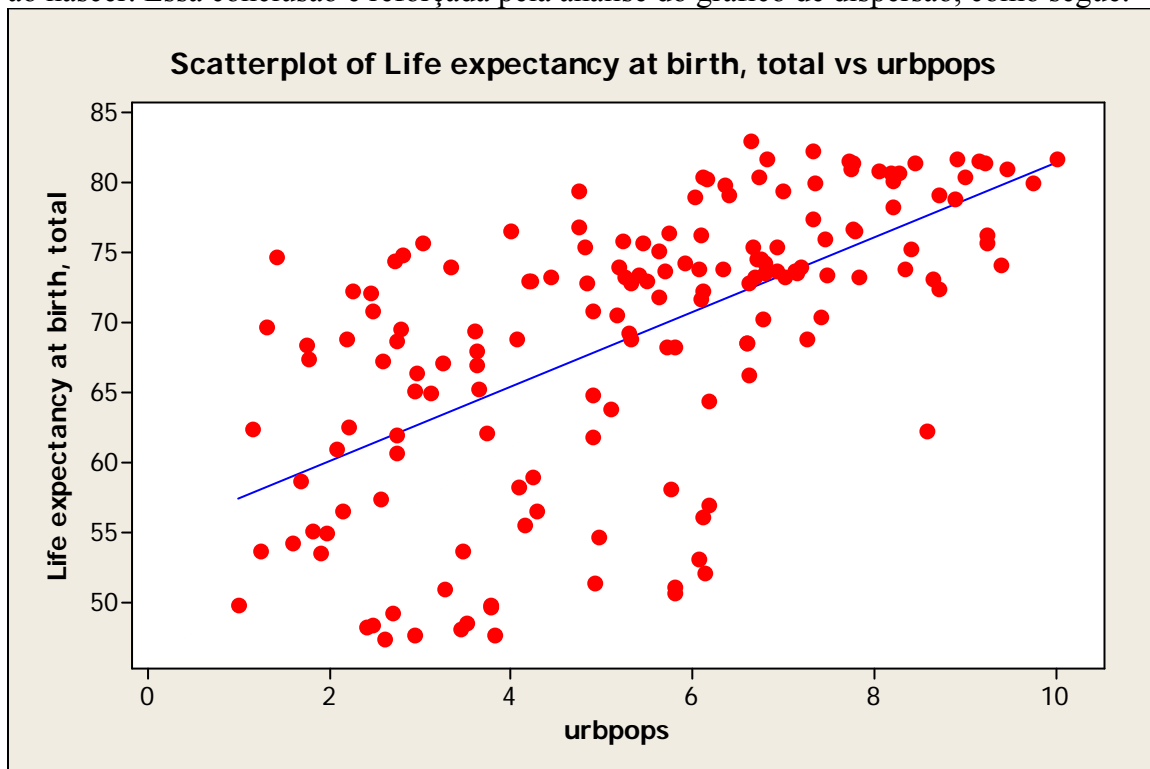
Source	DF	SS	MS	F	P
Regression	1	5891,5	5891,5	96,31	0,000
Residual Error	159	9726,8	61,2		
Total	160	15618,3			

Unusual Observations

Obs	Urban population (% of total)	Life expectancy at birth, total	Fit	SE Fit	Residual	St Resid
4	59	50,654	70,204	0,628	-19,551	-2,51R
19	61	53,110	70,908	0,645	-17,798	-2,28R
25	58	51,063	70,177	0,627	-19,115	-2,45R
28	39	47,618	64,903	0,747	-17,285	-2,22R
34	35	48,070	63,902	0,808	-15,833	-2,04R
108	50	51,410	67,851	0,628	-16,441	-2,11R
131	62	52,081	71,070	0,650	-18,988	-2,44R
133	15	74,723	58,466	1,241	16,257	2,11R
160	36	48,455	64,038	0,800	-15,582	-2,00R

R denotes an observation with a large standardized residual.

A equação acima revela um grau de adaptação baixo da reta de regressão à amostra: a variável População urbana prevê 37,7% do comportamento da variável Expectativa de vida ao nascer. Essa conclusão é reforçada pela análise do gráfico de dispersão, como segue:



A equação acima revela um grau de adaptação bastante alto da reta de regressão à amostra: a variável PIB per capita prevê 64,9% do comportamento da variável Usuários de Internet.

## 5.2 Regressão Linear Múltipla

A regressão linear múltipla tem o objetivo de demonstrar o quanto o comportamento de uma determinada variável pode ser previsto pelo comportamento de outras.

### 5.2.1 Regressão linear da variável Crescimento Populacional frente às demais

A seguir, a regressão Stepwise para a variável Crescimento Populacional frente às outras 8:

#### Stepwise Regression: pops versus intus; mobs; ...

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

Response is pops on 8 predictors, with N = 161						
Step	1	2	3	4	5	6
Constant	2,480	3,466	2,440	2,523	2,935	2,716
ferts	0,354	0,727	0,854	0,854	0,843	0,853
T-Value	12,56	8,24	8,69	8,75	9,09	9,31
P-Value	0,000	0,000	0,000	0,000	0,000	0,000
ages		-0,450	-0,478	-0,495	-0,539	-0,543
T-Value		-4,44	-4,78	-4,96	-5,66	-5,76
P-Value		0,000	0,000	0,000	0,000	0,000
lifes			0,110	0,148	0,136	0,126
T-Value			2,72	3,20	3,09	2,89
P-Value			0,007	0,002	0,002	0,004
intus				-0,060	-0,207	-0,236
T-Value				-1,67	-4,28	-4,78
P-Value				0,097	0,000	0,000
gdppps					0,276	0,266
T-Value					4,27	4,17
P-Value					0,000	0,000
urbpops						0,074
T-Value						2,25
P-Value						0,026
S	0,782	0,739	0,725	0,721	0,684	0,675
R-Sq	49,79	55,36	57,37	58,12	62,52	63,71
R-Sq(adj)	49,48	54,80	56,55	57,04	61,31	62,29
Mallows Cp	58,4	36,5	29,9	28,7	11,8	8,7
Step	7					
Constant	3,163					
ferts	0,846					
T-Value	9,28					
P-Value	0,000					
ages	-0,549					
T-Value	-5,85					
P-Value	0,000					
lifes	0,120					
T-Value	2,76					
P-Value	0,007					
intus	-0,240					
T-Value	-4,88					
P-Value	0,000					
gdppps	0,255					
T-Value	3,98					
P-Value	0,000					
urbpops	0,078					
T-Value	2,35					
P-Value	0,020					
gdpgrs	-0,062					
T-Value	-1,60					
P-Value	0,111					
S	0,672					
R-Sq	64,31					



R-Sq(adj)	62,68
Mallows Cp	8,1

A variável Taxa de fertilidade sozinha explica 49,79% da variável Crescimento populacional. As variáveis seguintes explicam muito pouco do comportamento da variável Crescimento populacional.

## 5.2.2 Regressão linear da variável Usuários de Internet frente às demais

A seguir, a regressão Stepwise para a variável Usuários de Internet frente às outras:

Stepwise Regression: intus versus pops; mobs; ...						
Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15						
Response is intus on 8 predictors, with N = 161						
Step	1	2	3	4	5	6
Constant	1,317	3,769	2,597	3,296	1,538	1,200
gdppps	1,403	1,070	0,944	0,972	0,888	0,886
T-Value	17,13	15,02	12,52	13,17	11,84	11,83
P-Value	0,000	0,000	0,000	0,000	0,000	
ferts		-0,494	-0,411	-0,264	-0,075	
T-Value		-10,30	-8,16	-3,93	-0,89	
P-Value		0,000	0,000	0,000	0,372	
urbpops			0,208	0,224	0,194	0,202
T-Value			3,95	4,35	3,84	4,06
P-Value			0,000	0,000	0,000	0,000
pops				-0,365	-0,399	-0,457
T-Value				-3,20	-3,61	-5,08
P-Value				0,002	0,000	0,000
lifes					0,235	0,273
T-Value					3,55	5,33
P-Value					0,001	0,000
S	1,53	1,19	1,13	1,10	1,06	1,06
R-Sq	64,86	78,97	80,87	82,05	83,40	83,31
R-Sq(adj)	64,64	78,71	80,50	81,59	82,86	82,88
Mallows Cp	198,5	57,7	40,5	30,6	19,0	17,8
Step	7	8				
Constant	2,408	2,740				
gdppps	0,895	0,900				
T-Value	12,13	12,44				
P-Value	0,000	0,000				
ferts		0,46				
T-Value		2,70				
P-Value		0,008				
urbpops	0,183	0,195				
T-Value	3,71	4,00				
P-Value	0,000	0,000				
pops	-0,369	-0,546				
T-Value	-3,86	-4,78				
P-Value	0,000	0,000				
lifes	0,193	0,266				
T-Value	3,24	4,13				
P-Value	0,001	0,000				

ages	-0,181	-0,541
T-Value	-2,47	-3,57
P-Value	0,015	0,000
S	1,05	1,03
R-Sq	83,94	84,67
R-Sq(adj)	83,43	84,07
Mallows Cp	13,4	8,1

A variável PIB per capita sozinha explica 64,86% da variável Usuários de Internet. A variável taxa de fertilidade explica quase 15% da mesma. As variáveis seguintes explicam muito pouco do comportamento da variável Usuários de Internet.

### 5.2.3 Regressão linear da variável Assinaturas de celular frente às demais

A seguir, a regressão Stepwise para a variável Assinaturas de celular frente às outras:

Stepwise Regression: mobs versus pops; intus; ...						
Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15						
Response is mobs on 8 predictors, with N = 161						
Step	1	2	3	4	5	6
Constant	7,498	5,477	5,001	5,606	5,804	4,826
ferts	-0,629	-0,472	-0,396	-0,174		
T-Value	-12,46	-8,28	-6,01	-1,09		
P-Value	0,000	0,000	0,000	0,278		
urbpops		0,270	0,201	0,205	0,211	0,196
T-Value		4,90	3,20	3,27	3,38	3,14
P-Value		0,000	0,002	0,001	0,001	0,002
intus			0,141	0,135	0,147	0,094
T-Value			2,19	2,11	2,33	1,35
P-Value			0,030	0,036	0,021	0,179
ages				-0,271	-0,446	-0,369
T-Value				-1,54	-6,13	-4,39
P-Value				0,127	0,000	0,000
lifes						0,135
T-Value						1,79
P-Value						0,075
S	1,40	1,31	1,29	1,29	1,29	1,28
R-Sq	49,42	56,10	57,40	58,03	57,71	58,57
R-Sq(adj)	49,10	55,54	56,58	56,96	56,90	57,50
Mallows Cp	32,6	9,6	6,7	6,3	5,5	4,3
Step	7	8				
Constant	4,839	5,279				
ferts						
T-Value						
P-Value						
urbpops	0,231	0,238				
T-Value	4,06	4,17				
P-Value	0,000	0,000				
intus						
T-Value						
P-Value						

ages	-0,389	-0,337
T-Value	-4,68	-3,76
P-Value	0,000	0,000
lifes	0,178	0,165
T-Value	2,62	2,42
P-Value	0,010	0,017
pops		-0,18
T-Value		-1,52
P-Value		0,131
S	1,28	1,28
R-Sq	58,08	58,69
R-Sq(adj)	57,28	57,63
Mallows Cp	4,1	3,8

A variável Taxa de fertilidade sozinha explica 49,42% da variável Assinaturas de celular. As variáveis seguintes explicam muito pouco do comportamento da variável Assinaturas de celular.

#### 5.2.4 Regressão linear da variável Razão de dependência da idade frente às demais

A seguir, a regressão Stepwise para a variável Razão de dependência da idade frente às outras:

<b>Stepwise Regression: ages versus pops; intus; ...</b>				
Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15				
Response is ages on 8 predictors, with N = 161				
Step	1	2	3	4
Constant	2,191	2,802	2,203	2,543
ferts	0,830	0,917	0,986	0,961
T-Value	39,80	32,95	23,46	22,24
P-Value	0,000	0,000	0,000	0,000
pops		-0,247	-0,266	-0,269
T-Value		-4,44	-4,78	-4,89
P-Value		0,000	0,000	0,000
lifes			0,066	0,083
T-Value			2,15	2,65
P-Value			0,033	0,009
mobs				-0,067
T-Value				-2,13
P-Value				0,035
S	0,578	0,547	0,541	0,535
R-Sq	90,88	91,89	92,12	92,35
R-Sq(adj)	90,82	91,79	91,97	92,15
Mallows Cp	40,3	20,4	17,3	14,5

A variável Taxa de fertilidade sozinha explica 90,88% da variável Razão de dependência da idade. As variáveis seguintes explicam muito pouco do comportamento da variável Razão de dependência da idade.

#### 5.2.5 Regressão linear da variável Taxa de Fertilidade frente às demais

A seguir, a regressão Stepwise para a variável Taxa de Fertilidade frente às outras:

<b>Stepwise Regression: ferts versus pops; intus; ...</b>
---

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15			
Response is fertst on 8 predictors, with N = 161			
Step	1	2	3
Constant	-2,0658	-2,8774	-0,7848
ages	1,095	0,952	0,789
T-Value	39,80	32,95	23,46
P-Value	0,000	0,000	0,000
pops		0,413	0,381
T-Value		8,24	8,69
P-Value		0,000	0,000
lifes			-0,173
T-Value			-7,24
P-Value			0,000
S	0,664	0,557	0,484
R-Sq	90,88	93,62	95,22
R-Sq(adj)	90,82	93,54	95,13
Mallows Cp	150,5	60,1	8,2

A variável Expectativa de vida ao nascer sozinha explica 90,88% da variável Taxa de fertilidade. As variáveis seguintes explicam muito pouco do comportamento da variável Taxa de fertilidade.

### 5.2.6 Regressão linear da variável PIB per Capita frente às demais

A seguir, a regressão Stepwise para a variável PIB per Capita frente às outras:

Stepwise Regression: gdppps versus pops; intus; ...				
Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15				
Response is gdppps on 8 predictors, with N = 161				
Step	1	2	3	4
Constant	0,07902	-1,45346	-1,89554	-2,51686
intus	0,462	0,532	0,569	0,553
T-Value	17,13	18,36	16,62	16,00
P-Value	0,000	0,000	0,000	0,000
pops		0,331	0,269	0,376
T-Value		4,89	3,65	4,36
P-Value		0,000	0,000	0,000
ages			0,100	0,344
T-Value			2,00	2,95
P-Value			0,048	0,004
ferts				-0,27
T-Value				-2,31
P-Value				0,022
S	0,877	0,820	0,813	0,802
R-Sq	64,86	69,48	70,24	71,22
R-Sq(adj)	64,64	69,10	69,67	70,48
Mallows Cp	30,6	7,9	5,9	2,7

A variável Usuários de internet sozinha explica 64,86% da variável PIB per capita. As variáveis seguintes explicam muito pouco do comportamento da variável PIB per capita.

### 5.2.7 Regressão linear da variável % Crescimento do PIB per Capita frente às demais

A seguir, a regressão Stepwise para a variável % Crescimento do PIB per Capita frente às outras:

Stepwise Regression: gdpgrs versus pops; intus; ...		
Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15		
Response is gdpgrs on 8 predictors, with N = 161		
Step	1	2
Constant	5,717	6,533
gdppps	-0,210	-0,241
T-Value	-2,80	-3,17
P-Value	0,006	0,002
pops		-0,20
T-Value		-1,96
P-Value		0,051
S	1,40	1,39
R-Sq	4,69	6,96
R-Sq(adj)	4,10	5,79
Mallows Cp	3,4	1,5

A regressão Stepwise resultou inválida, não sendo, portanto afirmar nada para o comportamento da variável % Crescimento do PIB per capita.

### 5.2.8 Regressão linear da variável Expectativa de vida ao nascer frente às demais

A seguir, a regressão Stepwise para a variável Expectativa de vida ao nascer frente às outras:

Stepwise Regression: lifes versus pops; intus; ...					
Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15					
Response is lifes on 8 predictors, with N = 161					
Step	1	2	3	4	5
Constant	9,899	7,255	6,522	5,109	4,231
ferts	-0,937	-0,629	-0,735	-1,179	-1,152
T-Value	-18,18	-9,93	-9,57	-6,67	-6,55
P-Value	0,000	0,000	0,000	0,000	0,000
intus		0,375	0,374	0,380	0,348
T-Value		6,93	7,02	7,28	6,36
P-Value		0,000	0,000	0,000	0,000
pops			0,30	0,42	0,42
T-Value			2,36	3,20	3,26
P-Value			0,019	0,002	0,001
ages				0,49	0,53
T-Value				2,78	2,99
P-Value				0,006	0,003
mobs					0,136

T-Value					1,87
P-Value					0,063
S	1,43	1,26	1,24	1,21	1,20
R-Sq	67,52	75,10	75,96	77,09	77,59
R-Sq(adj)	67,32	74,78	75,50	76,50	76,87
Mallows Cp	63,9	14,4	10,6	4,9	3,4

A variável Taxa de fertilidade sozinha explica 67,52% da variável Expectativa de vida. A seguir, a variável Usuários de Internet contribui com 6,93%. As variáveis seguintes explicam muito pouco do comportamento da variável Taxa de fertilidade.

### 5.2.9 Regressão linear da variável População Urbana frente às demais

A seguir, a regressão Stepwise para a variável População Urbana frente às outras:

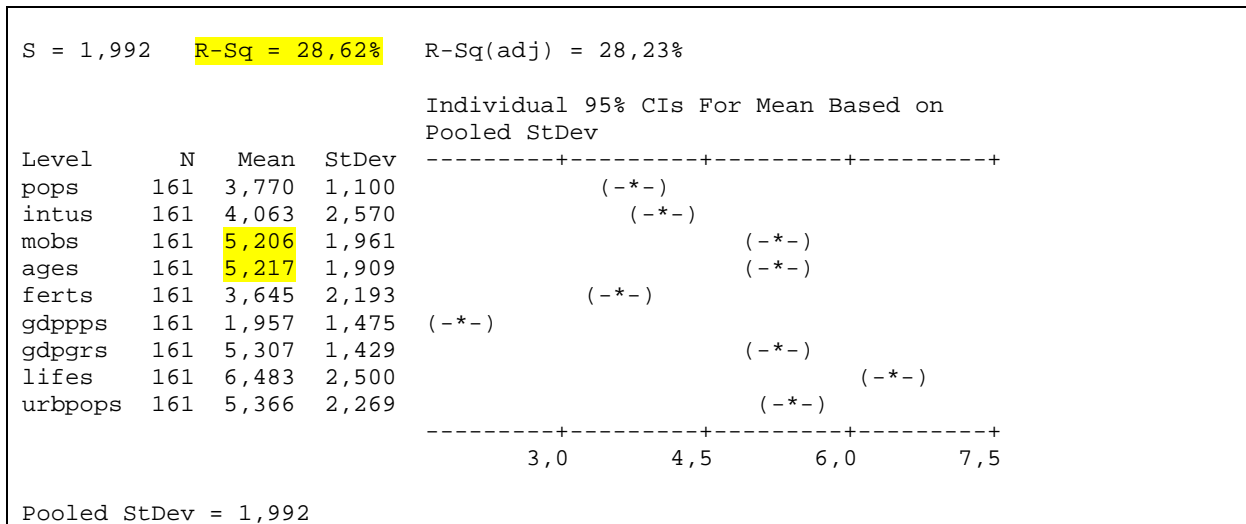
Stepwise Regression: urbpops versus pops; intus; ...					
Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15					
Response is urbpops on 8 predictors, with N = 161					
Step	1	2	3	4	5
Constant	2,89432	1,88136	0,65140	1,21477	-0,02435
intus	0,608	0,449	0,480	0,438	0,439
T-Value	12,00	6,96	7,24	6,16	6,20
P-Value	0,000	0,000	0,000	0,000	0,000
mobs		0,319	0,358	0,300	0,318
T-Value		3,77	4,13	3,20	3,39
P-Value		0,000	0,000	0,002	0,001
pops			0,24	0,37	0,47
T-Value			1,78	2,36	2,81
P-Value			0,076	0,020	0,006
ferts				-0,17	-0,52
T-Value				-1,57	-2,26
P-Value				0,118	0,025
ages					0,39
T-Value					1,73
P-Value					0,086
S	1,65	1,58	1,57	1,57	1,56
R-Sq	47,51	51,85	52,80	53,54	54,42
R-Sq(adj)	47,18	51,24	51,90	52,35	52,95
Mallows Cp	19,7	7,1	5,9	5,4	4,5

A variável Usuários de Internet sozinha explica 47,51% da variável População urbana. As variáveis seguintes explicam muito pouco do comportamento da variável População urbana.

## 6. ANÁLISE DE VARIÂNCIA

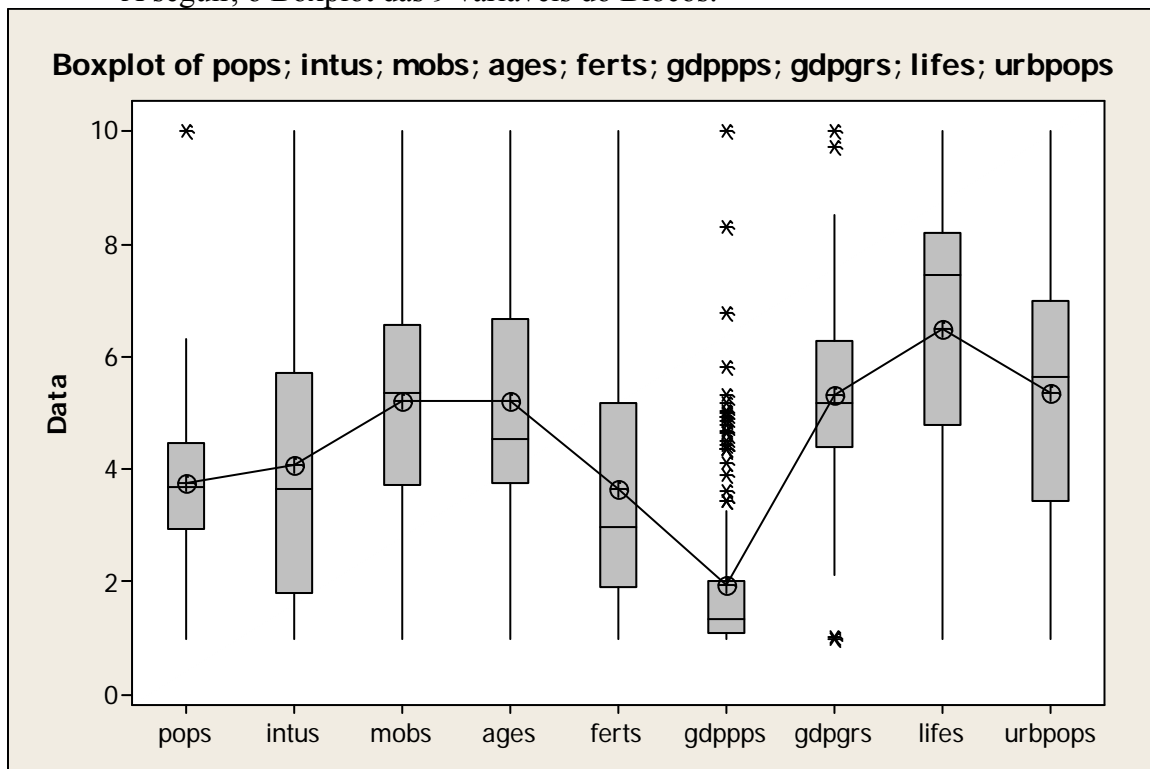
A seguir será realizada a análise de variância das 9 variáveis, através da função ANOVA:

One-way ANOVA: pops; intus; mobs; ages; ferts; gdppps; gdpgrs; lifes; urbpops					
Source	DF	SS	MS	F	P
Factor	8	2292,26	286,53	72,18	0,000
Error	1440	5716,39	3,97		
Total	1448	8008,64			



O R-Sq = 28,62% demonstra que a chance de as médias das variáveis selecionadas serem iguais é de 28,62%, afirmação que é comprovada pelo P=0,000.

A seguir, o Boxplot das 9 variáveis do Blocos:



Como o próprio gráfico aponta, a menor média é da variável % crescimento do PIB per Capita, o que reflete o momento de crise que o mundo vive desde 2008. A mediana também oscila muito. Há muitos outliers na variável PIB per capita e três na variável % Crescimento do PIB per capita.

O P-Value=0 indica que as chances das médias serem iguais é Zero.

Para uma nova análise, foram selecionadas as duas variáveis mais próximas, que são Assinaturas de Celular e Razão de dependência da idade, onde foi feita uma comparação entre elas, com teste “Maior Que”. O resultado apurado é exibido a seguir:

**Paired T-Test and CI: mobs; pops**

Paired T for mobs - pops

	N	Mean	StDev	SE Mean
mobs	161	5,206	1,961	0,155
pops	161	3,770	1,100	0,087
Difference	161	1,436	2,675	0,211

95% lower bound for mean difference: 1,087

T-Test of mean difference = 0 (vs > 0): T-Value = 6,81 P-Value = 0,000

O P-Value encontrado indica que, com a significância de 5%, as variáveis não tem relação entre si (P-Value = 0,000). Ou seja, não basta ter médias próximas, as variáveis não tem comportamentos parecidos.

## 7. AMOSTRAS

Amostragem é um instrumento muito utilizado por estatísticos porque permite inferir o comportamento de uma população sem que seja feita pesquisa com todos os elementos da população.

Porém, para definir o número de elementos de uma amostra não é simplesmente pegar um número aleatório de uma população, há que se escolher com cuidado para que ela seja representativa da população que está sendo estudada. Caso a amostra não seja bem feita, os resultados da pesquisa em questão serão distorcidos e as conclusões serão inválidas.

Quando uma amostra é bem feita, pode-se esperar que 2 situações:

1. Que a média de cada amostra se aproxime da média da população conforme aumenta o número de elementos de cada amostra;
2. Que a dispersão se reduza com o aumento do número de elementos de cada amostra.

Outro aspecto bastante delicado é definir qual o número ideal da amostra, já que quanto maior a amostra, melhores serão os resultados, porem maiores os custos de realizar a pesquisa.

A precisão da amostra aumenta de forma linear, enquanto que o número de elementos necessários para obter esse aumento de precisão aumenta de forma quadrática. Assim, para dobrar a precisão, é necessário multiplicar por 4 o número de elementos da amostra.

No exercício, foi utilizada a média das amostras para analisar sua convergência com a média da população e o desvio padrão como medida de dispersão de cada amostra. Para tanto, foram realizadas análises com 2 amostras, selecionadas de forma randômica, com 10 e 40 países. E o total da amostra, que se aproxima de 160 países, com 161.

Uma premissa adotada foi que a população se comporta como a amostra maior, de 161 elementos.

A proposta deste tópico é realizar análises entre as variáveis para identificar quais delas apresentam a maior diferenciação dos clusters.

### 7.1 **Análise Exploratória das Amostras**

Para a realização desta análise foram feitas amostras, de 10 e 40 países. E foi realizada a análise ANOVA com as 2 amostras, mais a população inteira de respondentes, 161 países.

#### 7.1.1 **ANOVA para a variável Crescimento Populacional**

Segue a análise ANOVA realizada para a variável Crescimento Populacional com as 2 amostras (10 e 40) e a população total (161):

**One-way ANOVA: Pop10; Pop40; pops**



Source	DF	SS	MS	F	P
Factor	2	2,96	1,48	1,25	0,288
Error	208	245,85	1,18		
Total	210	248,81			

S = 1,087    R-Sq = 1,19%    R-Sq(adj) = 0,24%

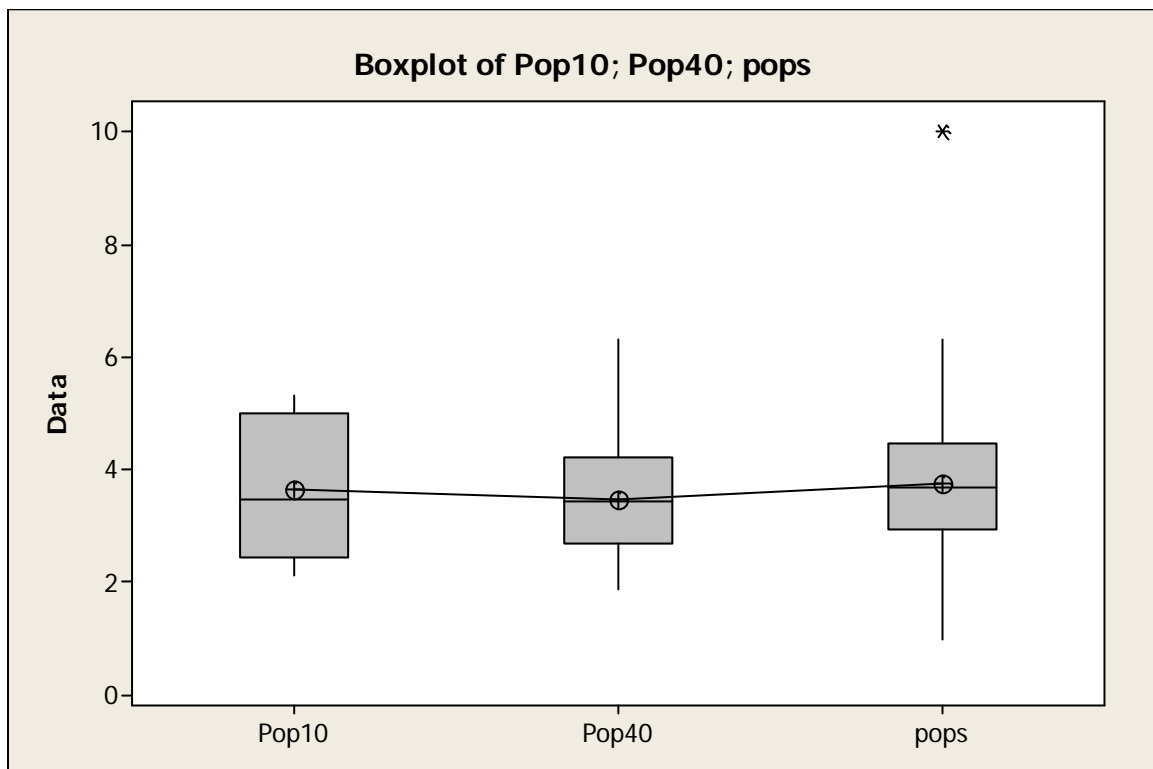
Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
Pop10	10	3,642	1,168
Pop40	40	3,468	1,014
pops	161	3,770	1,100

3,15      3,50      3,85      4,20

Pooled StDev = 1,087

Logo abaixo, segue o Boxplot resultante da ANOVA:



Conforme esperado, a dispersão da média de todos os países é menor do que a dispersão verificada nas amostras. Além disso, a dispersão é menor conforme aumenta o número de elementos de cada amostra.

As médias das amostras se tornam cada vez mais próximas da média do total da população a cada número maior de países na amostra.

### 7.1.2 ANOVA para a variável Usuários de Internet

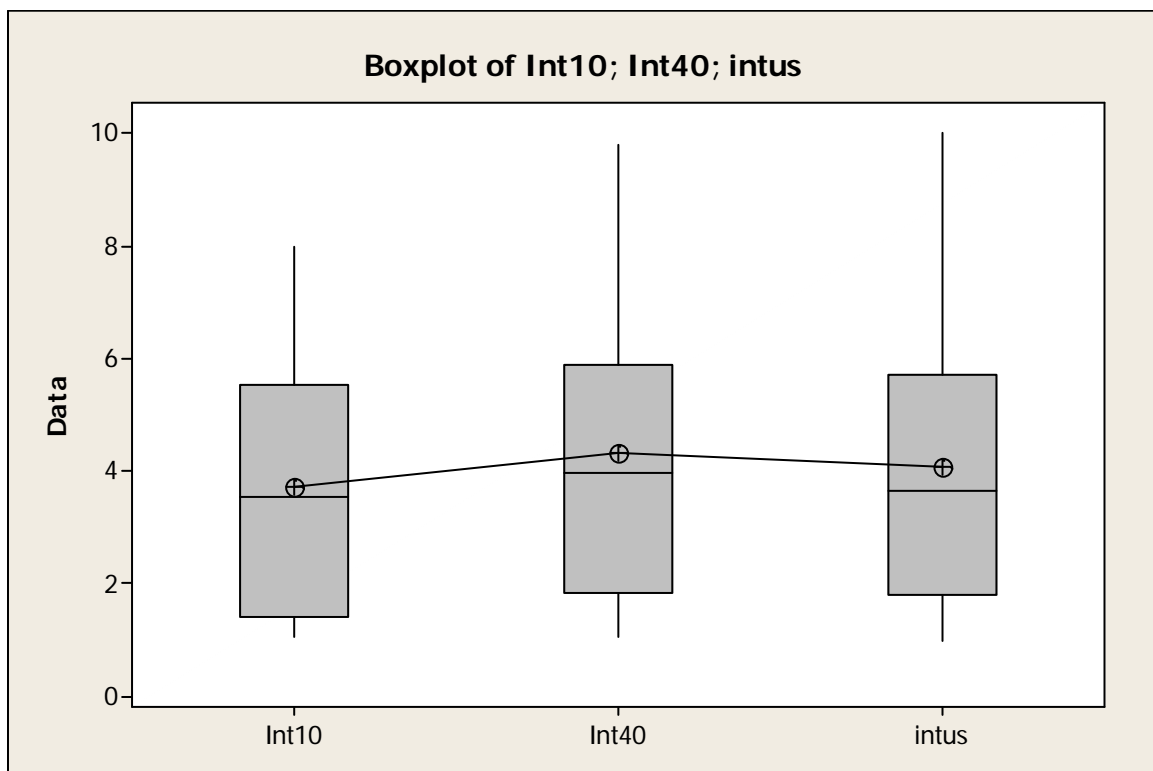
Segue a análise ANOVA realizada para a variável Crescimento populacional com as 2 amostras (10 e 40) e a população total (161):

#### One-way ANOVA: Int10; Int40; intus

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

Factor	2	3,54	1,77	0,27	0,765
Error	208	1372,34	6,60		
Total	210	1375,88			
S = 2,569    R-Sq = 0,26%    R-Sq(adj) = 0,00%					
Individual 95% CIs For Mean Based on Pooled StDev					
Level	N	Mean	StDev	-----+-----+-----+-----+-----+	
Int10	10	3,736	2,347	(-----*-----)	
Int40	40	4,323	2,612	(-----*-----)	
intus	161	4,063	2,570	(---*---)	
				-----+-----+-----+-----+-----+	
				3,0      4,0      5,0      6,0	
Pooled StDev = 2,569					

Logo abaixo, segue o Boxplot resultante da ANOVA:



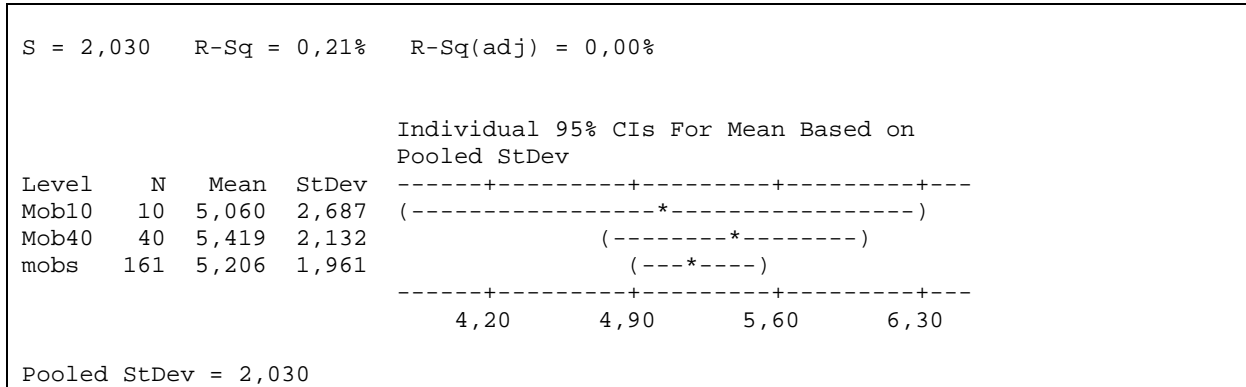
Neste caso, mais uma vez, a dispersão da média de todos os países é menor do que a dispersão verificada nas amostras. Além disso, a dispersão é menor conforme aumenta o número de elementos de cada amostra.

As médias das amostras se tornam cada vez mais próximas da média do total da população a cada número maior de países na amostra.

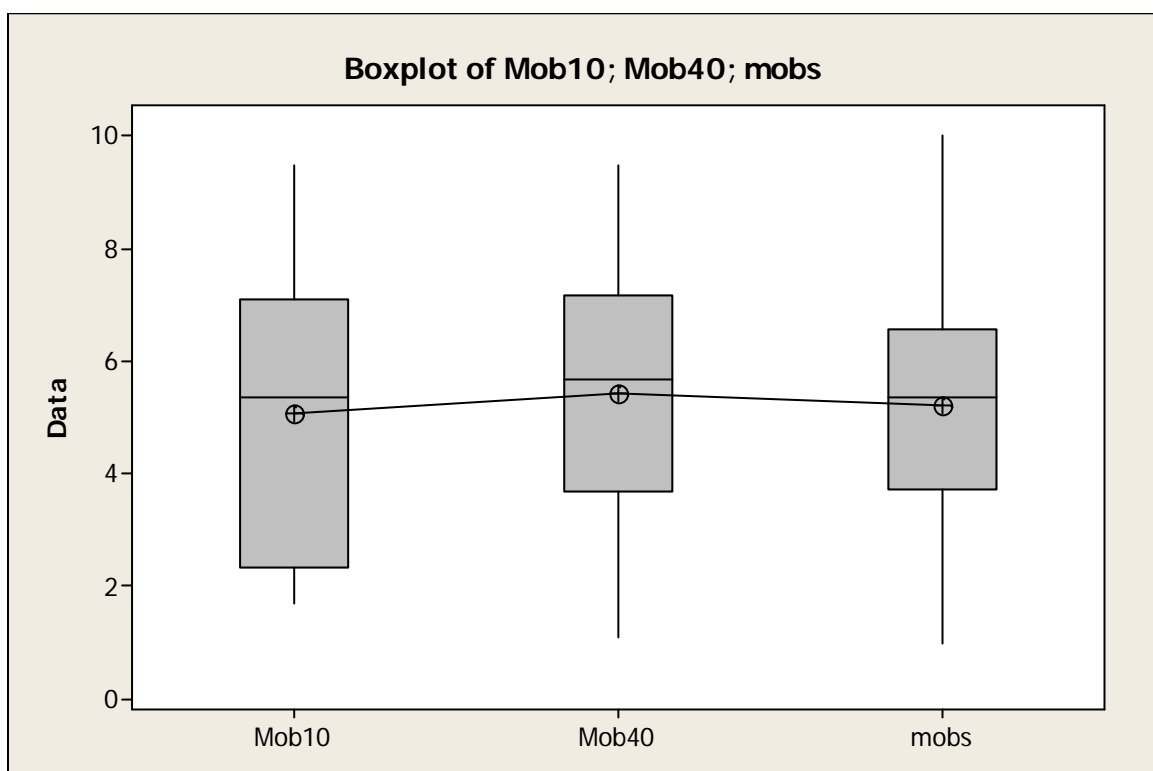
### 7.1.3 ANOVA para a variável Assinaturas de celular

Segue a análise ANOVA realizada para a variável Assinaturas de celular com as 2 amostras (10 e 40) e a população total (161):

One-way ANOVA: Mob10; Mob40; mobs					
Source	DF	SS	MS	F	P
Factor	2	1,78	0,89	0,22	0,806
Error	208	857,48	4,12		
Total	210	859,27			



Logo abaixo, segue o Boxplot resultante da ANOVA:



Conforme esperado, a dispersão da média de todos os países é menor do que a dispersão verificada nas amostras. Além disso, a dispersão é menor conforme aumenta o número de elementos de cada amostra.

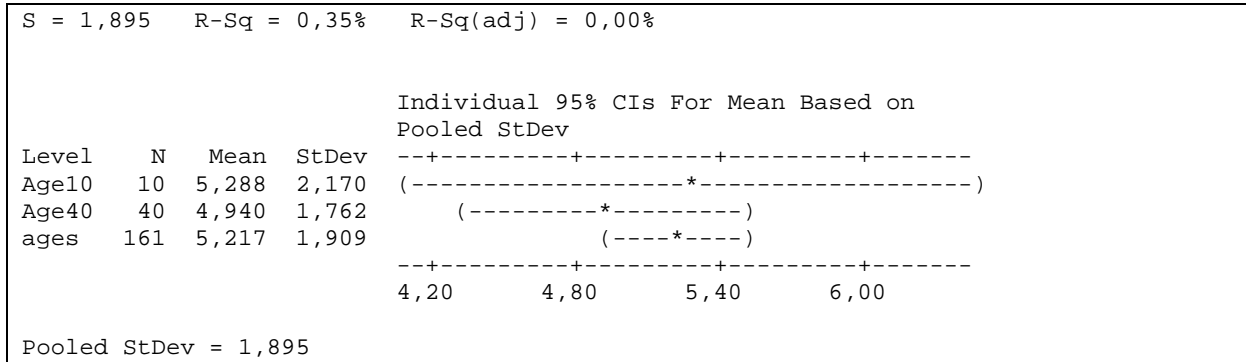
As médias das amostras se tornam cada vez mais próximas da média do total da população a cada número maior de países na amostra.

#### 7.1.4 ANOVA para a variável Razão de dependência da idade

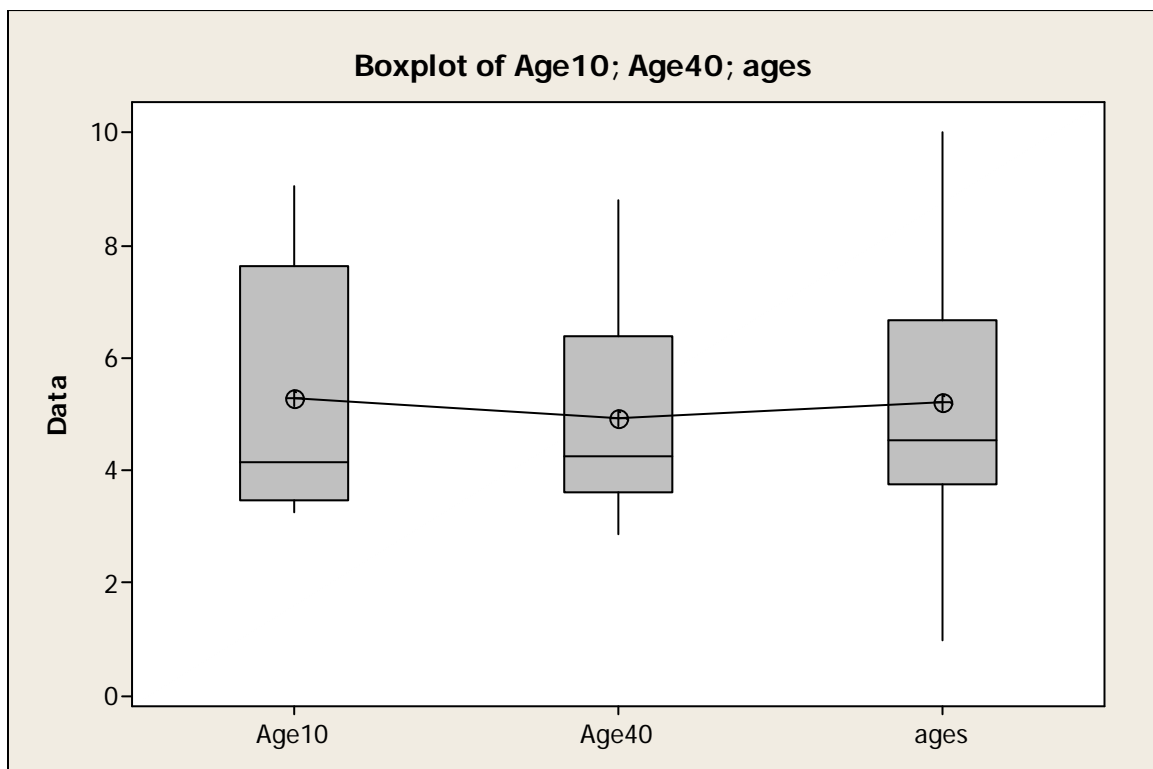
Segue a análise ANOVA realizada para a variável Razão de dependência da idade com as 2 amostras (10 e 40) e a população total (161):

##### One-way ANOVA: Age10; Age40; ages

Source	DF	SS	MS	F	P
Factor	2	2,61	1,30	0,36	0,696
Error	208	746,74	3,59		
Total	210	749,35			



Logo abaixo, segue o Boxplot resultante da ANOVA:

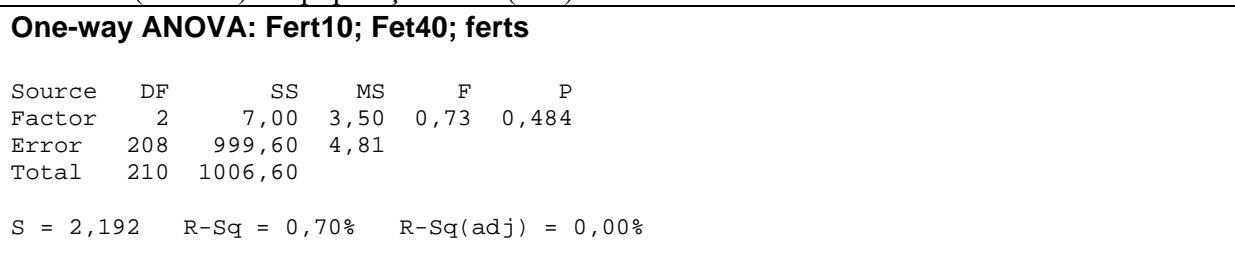


Conforme esperado, a dispersão da média de todos os países é menor do que a dispersão verificada nas amostras. Além disso, a dispersão é menor conforme aumenta o número de elementos de cada amostra.

As médias das amostras se tornam cada vez mais próximas da média do total da população a cada número maior de países na amostra.

### 7.1.5 ANOVA para a variável Taxa de Fertilidade

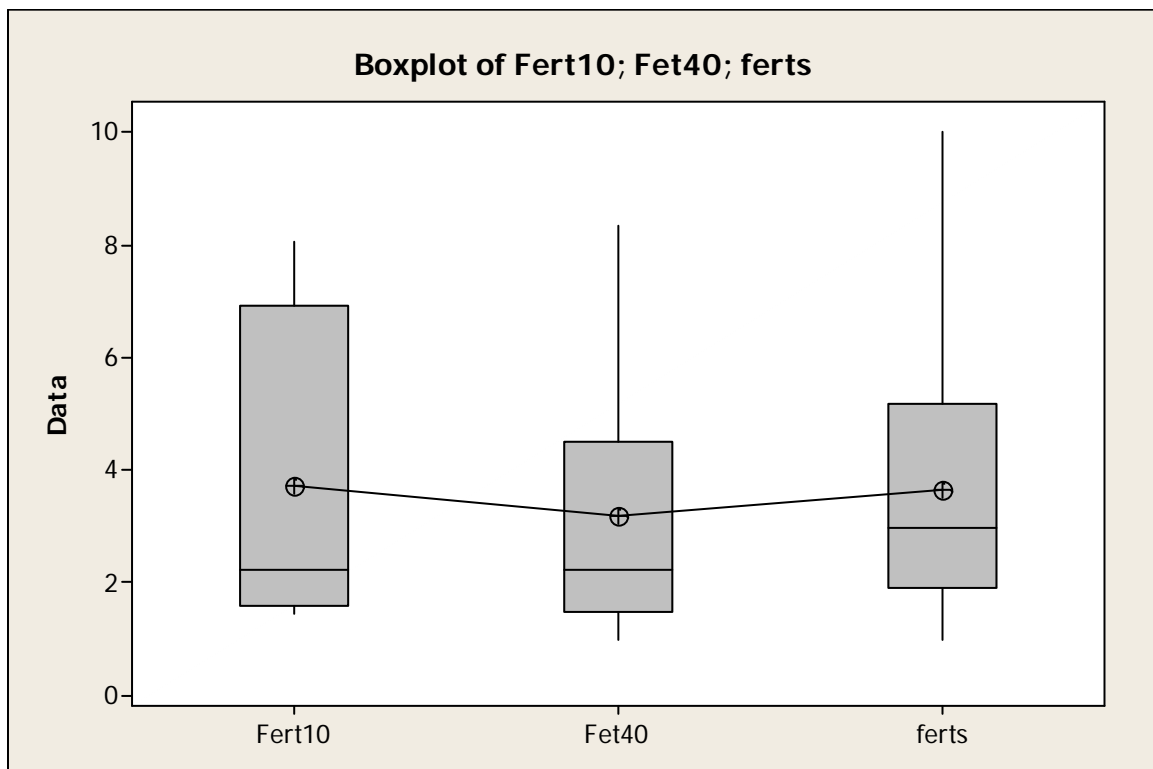
Segue a análise ANOVA realizada para a variável Taxa de Fertilidade com as 2 amostras (10 e 40) e a população total (161):



Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
Fert10	10	3,708	2,681	(-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+)
Fet40	40	3,186	2,061	(-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+)
ferts	161	3,645	2,193	(-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+)

Pooled StDev = 2,192

Logo abaixo, segue o Boxplot resultante da ANOVA:



Conforme esperado, a dispersão da média de todos os países é menor do que a dispersão verificada nas amostras. Além disso, a dispersão é menor conforme aumenta o número de elementos de cada amostra.

As médias das amostras se tornam cada vez mais próximas da média do total da população a cada número maior de países na amostra.

### 7.1.6 ANOVA para a variável PIB per Capita

Segue a análise ANOVA realizada para a variável PIB per Capita com as 2 amostras (10 e 40) e a população total (161):

#### One-way ANOVA: GDP10; GDP40; gdppps

Source	DF	SS	MS	F	P
Factor	2	5,84	2,92	1,22	0,298
Error	208	498,16	2,39		
Total	210	504,00			

S = 1,548    R-Sq = 1,16%    R-Sq(adj) = 0,21%

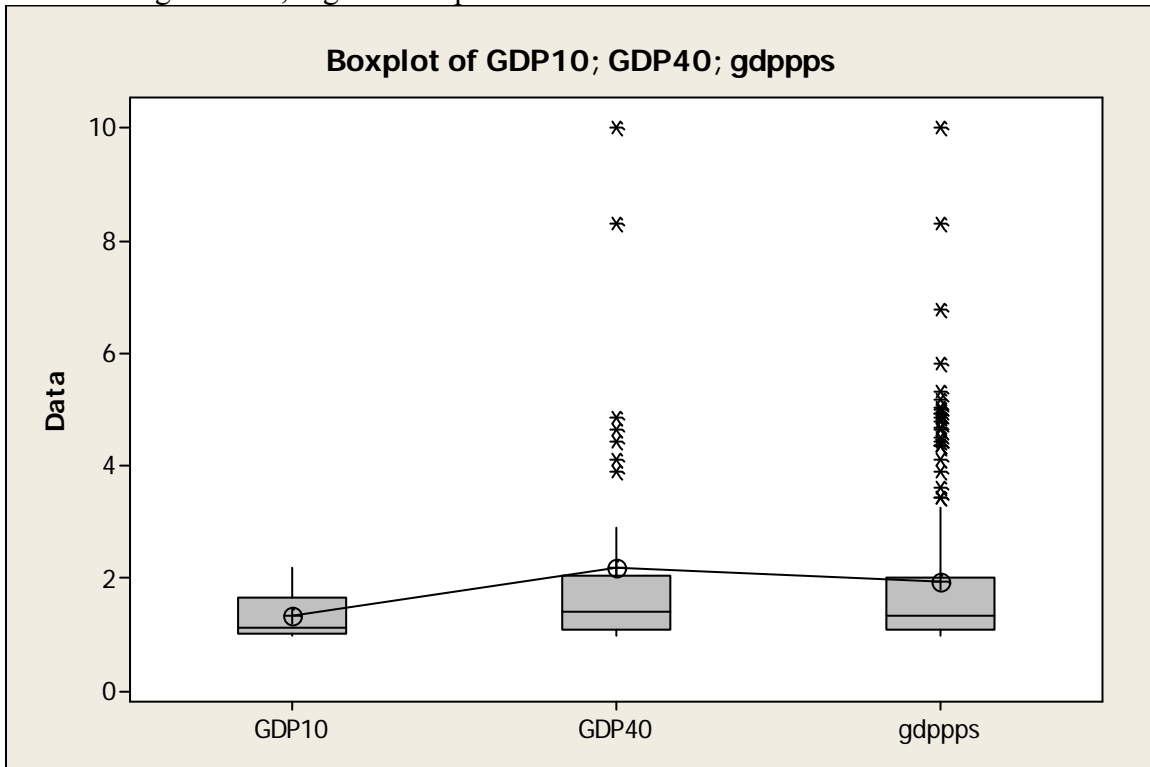
Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	
GDP10	10	1,358	0,463	(-----*-----)
GDP40	40	2,198	1,947	(-----*-----)
gdppps	161	1,957	1,475	(---*---)

0,60      1,20      1,80      2,40

Pooled StDev = 1,548

Logo abaixo, segue o Boxplot resultante da ANOVA:



Conforme esperado, a dispersão da média de todos os países é menor do que a dispersão verificada nas amostras. Além disso, a dispersão é menor conforme aumenta o número de elementos de cada amostra.

As médias das amostras se tornam cada vez mais próximas da média do total da população a cada número maior de países na amostra.

### 7.1.7 ANOVA para a variável % Crescimento do PIB per Capita

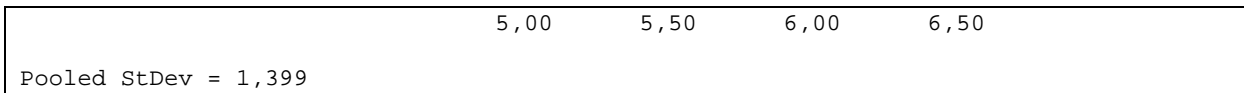
Segue a análise ANOVA realizada para a variável % Crescimento do PIB per Capita com as 2 amostras (10 e 40) e a população total (161):

One-way ANOVA: GDPT10; GDPT40; gdpgrs					
Source	DF	SS	MS	F	P
Factor	2	0,99	0,50	0,25	0,776
Error	208	406,92	1,96		
Total	210	407,91			

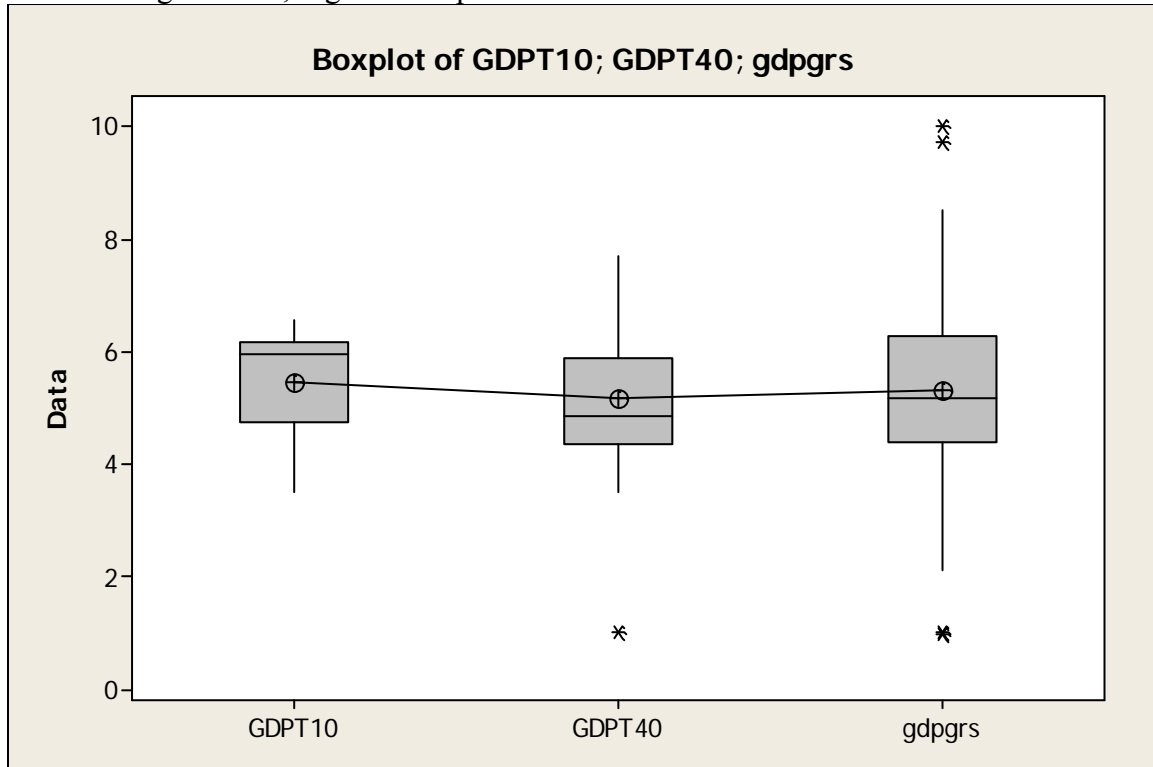
S = 1,399    R-Sq = 0,24%    R-Sq(adj) = 0,00%

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev	
GDPT10	10	5,471	1,085	(-----*-----)
GDPT40	40	5,165	1,337	(-----*-----)
gdpgrs	161	5,307	1,429	(---*---)



Logo abaixo, segue o Boxplot resultante da ANOVA:



Conforme esperado, a dispersão da média de todos os países é menor do que a dispersão verificada nas amostras. Além disso, a dispersão é menor conforme aumenta o número de elementos de cada amostra.

As médias das amostras se tornam cada vez mais próximas da média do total da população a cada número maior de países na amostra.

### 7.1.8 ANOVA para a variável Expectativa de Vida ao nascer

Segue a análise ANOVA realizada para a variável Expectativa de Vida ao nascer com as 2 amostras (10 e 40) e a população total (161):

#### One-way ANOVA: Lif10; Liv40; lifes

Source	DF	SS	MS	F	P
Factor	2	11,64	5,82	0,94	0,392
Error	208	1286,89	6,19		
Total	210	1298,53			

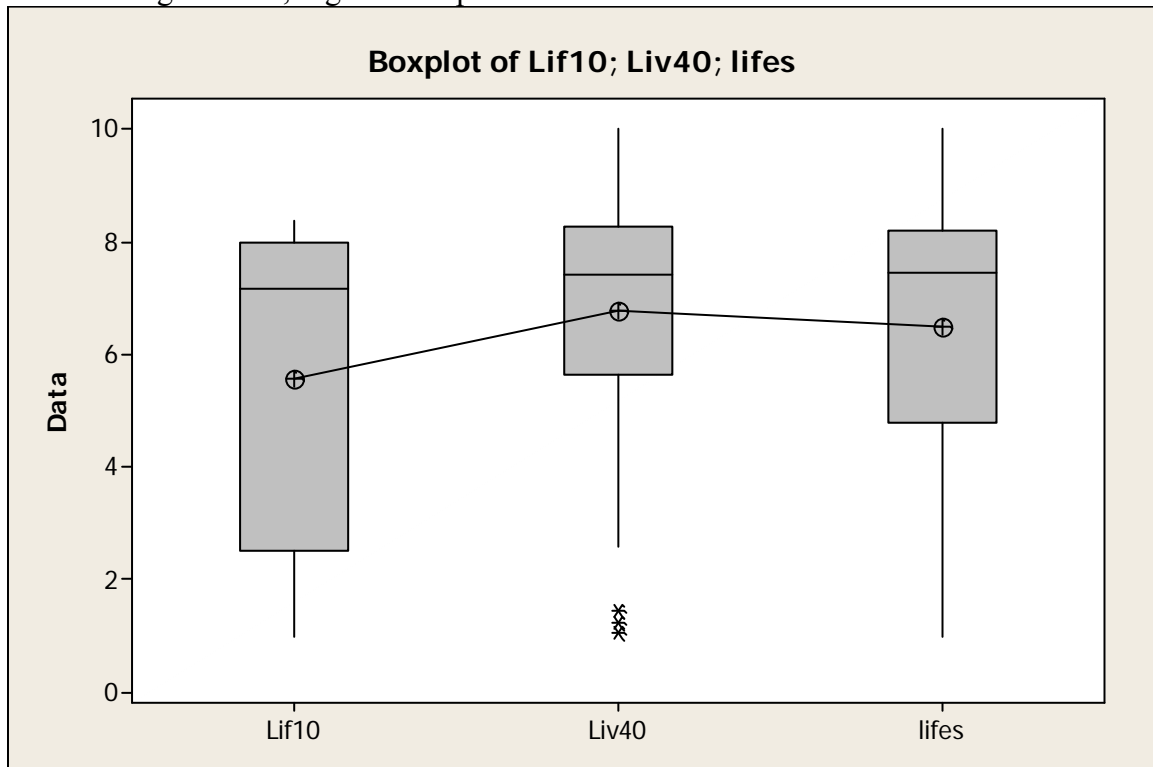
S = 2,487    R-Sq = 0,90%    R-Sq(adj) = 0,00%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
Lif10	10	5,586	2,944	+-----+ (-----*-----)
Liv40	40	6,785	2,314	(-----*-----)
lifes	161	6,483	2,500	(---*---)

4,0      5,0      6,0      7,0

Pooled StDev = 2,487

Logo abaixo, segue o Boxplot resultante da ANOVA:



Conforme esperado, a dispersão da média de todos os países é menor do que a dispersão verificada nas amostras. Além disso, a dispersão é menor conforme aumenta o número de elementos de cada amostra.

As médias das amostras se tornam cada vez mais próximas da média do total da população a cada número maior de países na amostra.

### 7.1.9 ANOVA para a variável População Urbana

Segue a análise ANOVA realizada para a variável População Urbana com as 2 amostras (10 e 40) e a população total (161):

#### One-way ANOVA: Urbp10; Urbp40; urbpops

Source	DF	SS	MS	F	P
Factor	2	8,92	4,46	0,88	0,417
Error	208	1057,61	5,08		
Total	210	1066,53			

S = 2,255    R-Sq = 0,84%    R-Sq(adj) = 0,00%

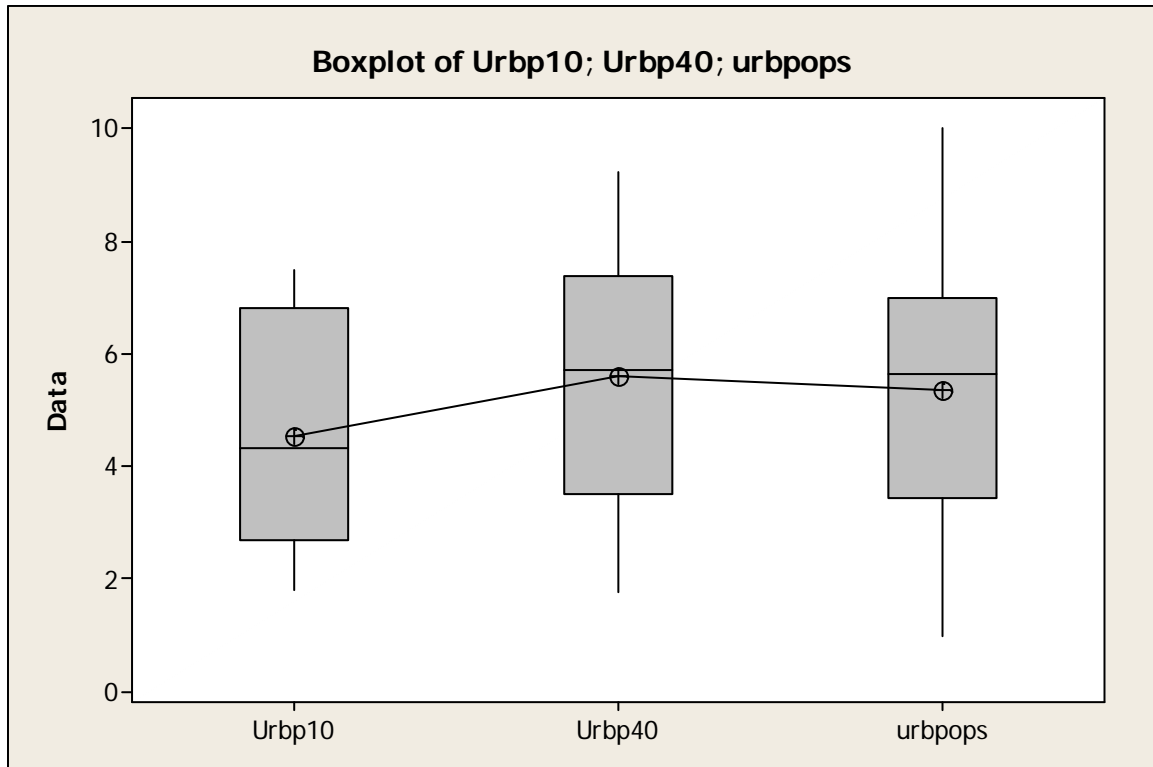
Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
Urbp10	10	4,553	2,106	(-----*-----)
Urbp40	40	5,607	2,231	(-----*-----)
urbpops	161	5,366	2,269	(---*---)

3,20    4,00    4,80    5,60

Pooled StDev = 2,255

Logo abaixo, segue o Boxplot resultante da ANOVA:





Conforme esperado, a dispersão da média de todos os países é menor do que a dispersão verificada nas amostras. Além disso, a dispersão é menor conforme aumenta o número de elementos de cada amostra.

As médias das amostras se tornam cada vez mais próximas da média do total da população a cada número maior de países na amostra.

## 8. ANÁLISE DE CLUSTERS e DENDOGRAMA POR PAÍS

Foi solicitado ao Minitab realizar a análise de clusters por observação.

### 8.1 Análise de clusters para todos os países

Segue o resultado para os 161 países que formam a base de dados:

<b>Cluster Analysis of Observations: Population g; Internet use; Mobile cellu; ...</b>							
Standardized Variables, Euclidean Distance, Single Linkage							
Amalgamation Steps							
Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	160	95,0090	0,50780	67	80	67	2
2	159	94,0638	0,60396	42	114	42	2
3	158	93,7766	0,63318	118	125	118	2
4	157	93,7480	0,63610	68	99	68	2
5	156	93,4891	0,66244	103	110	103	2
6	155	93,4277	0,66869	16	155	16	2
7	154	93,4103	0,67046	28	61	28	2
8	153	93,2753	0,68419	41	105	41	2
9	152	93,2478	0,68699	32	43	32	2
10	151	93,0774	0,70433	6	87	6	2
11	150	92,9633	0,71594	5	154	5	2
12	149	92,8448	0,72800	32	146	32	3
13	148	92,6425	0,74857	47	65	47	2
14	147	92,6061	0,75228	120	140	120	2
15	146	92,5977	0,75313	47	81	47	3

16	145	92,5021	0,76286	33	46	33	2
17	144	92,4917	0,76392	11	24	11	2
18	143	92,4729	0,76584	17	122	17	2
19	142	92,4632	0,76682	6	118	6	4
20	141	92,3463	0,77871	42	147	42	3
21	140	92,2712	0,78635	16	67	16	4
22	139	92,0529	0,80856	26	153	26	2
23	138	92,0128	0,81265	116	128	116	2
24	137	91,9248	0,82160	8	50	8	2
25	136	91,9176	0,82233	47	116	47	5
26	135	91,8958	0,82454	26	51	26	3
27	134	91,8645	0,82773	40	47	40	6
28	133	91,8229	0,83197	13	152	13	2
29	132	91,8074	0,83354	33	88	33	3
30	131	91,7978	0,83452	1	34	1	2
31	130	91,7896	0,83535	32	96	32	4
32	129	91,7729	0,83705	32	101	32	5
33	128	91,7429	0,84010	29	92	29	2
34	127	91,6748	0,84703	13	41	13	4
35	126	91,5898	0,85568	11	16	11	6
36	125	91,5516	0,85957	1	22	1	3
37	124	91,4190	0,87306	15	124	15	2
38	123	91,4190	0,87306	9	32	9	6
39	122	91,3587	0,87919	1	102	1	4
40	121	91,3241	0,88271	54	68	54	3
41	120	91,2975	0,88543	1	89	1	5
42	119	91,2877	0,88642	45	73	45	2
43	118	91,2864	0,88655	13	26	13	7
44	117	91,2134	0,89398	1	120	1	7
45	116	91,1828	0,89709	15	53	15	3
46	115	91,1512	0,90030	6	40	6	10
47	114	91,0198	0,91367	1	159	1	8
48	113	90,9487	0,92091	6	21	6	11
49	112	90,9179	0,92405	28	60	28	3
50	111	90,8774	0,92816	37	94	37	2
51	110	90,8708	0,92884	54	95	54	4
52	109	90,8346	0,93252	7	13	7	8
53	108	90,6958	0,94664	25	143	25	2
54	107	90,5496	0,96152	2	6	2	12
55	106	90,5161	0,96492	42	115	42	4
56	105	90,4405	0,97261	3	9	3	7
57	104	90,4312	0,97356	72	117	72	2
58	103	90,3871	0,97804	2	3	2	19
59	102	90,3687	0,97992	25	37	25	4
60	101	90,3652	0,98028	11	148	11	7
61	100	90,3018	0,98672	27	42	27	5
62	99	90,2978	0,98714	103	139	103	3
63	98	90,2927	0,98766	20	82	20	2
64	97	90,2416	0,99285	1	29	1	10
65	96	90,2284	0,99419	2	27	2	24
66	95	90,1532	1,00185	15	25	15	7
67	94	90,1465	1,00253	112	130	112	2
68	93	90,0780	1,00950	55	74	55	2
69	92	90,0677	1,01054	28	33	28	6
70	91	90,0644	1,01088	2	18	2	25
71	90	90,0581	1,01152	38	72	38	3
72	89	89,9844	1,01902	8	55	8	4
73	88	89,9129	1,02630	1	149	1	11
74	87	89,8704	1,03061	1	142	1	12
75	86	89,8496	1,03273	2	20	2	27
76	85	89,8274	1,03500	2	30	2	28
77	84	89,5619	1,06201	8	38	8	7
78	83	89,5153	1,06675	1	107	1	13
79	82	89,4968	1,06863	2	106	2	29
80	81	89,3401	1,08457	7	132	7	9
81	80	89,3212	1,08650	2	8	2	36
82	79	89,2434	1,09441	2	90	2	37
83	78	89,1600	1,10289	7	70	7	10

84	77	89,0256	1,11657	49	79	49	2
85	76	88,9437	1,12490	15	28	15	13
86	75	88,8942	1,12994	48	112	48	3
87	74	88,8784	1,13154	2	36	2	38
88	73	88,7881	1,14074	59	64	59	2
89	72	88,7869	1,14085	45	58	45	3
90	71	88,7718	1,14239	45	134	45	4
91	70	88,7537	1,14423	1	77	1	14
92	69	88,7299	1,14666	2	54	2	42
93	68	88,7250	1,14715	2	76	2	43
94	67	88,7100	1,14868	1	15	1	27
95	66	88,6744	1,15231	35	56	35	2
96	65	88,6178	1,15806	2	141	2	44
97	64	88,6145	1,15839	7	39	7	11
98	63	88,5693	1,16299	2	93	2	45
99	62	88,5028	1,16977	2	150	2	46
100	61	88,5003	1,17002	1	35	1	29
101	60	88,4521	1,17492	121	144	121	2
102	59	88,4240	1,17778	111	123	111	2
103	58	88,3186	1,18850	2	98	2	47
104	57	88,1695	1,20367	1	17	1	31
105	56	88,1540	1,20525	2	5	2	49
106	55	88,0946	1,21129	7	137	7	12
107	54	88,0815	1,21263	1	2	1	80
108	53	88,0107	1,21983	1	12	1	81
109	52	87,9792	1,22303	1	49	1	83
110	51	87,9036	1,23072	7	136	7	13
111	50	87,8086	1,24039	91	158	91	2
112	49	87,7975	1,24152	1	160	1	84
113	48	87,7541	1,24594	1	44	1	85
114	47	87,6878	1,25268	1	11	1	92
115	46	87,5687	1,26480	1	103	1	95
116	45	87,5228	1,26947	59	156	59	3
117	44	87,5218	1,26957	1	104	1	96
118	43	87,4695	1,27489	1	133	1	97
119	42	87,4630	1,27556	7	10	7	14
120	41	87,4582	1,27604	1	45	1	101
121	40	87,4072	1,28123	100	119	100	2
122	39	87,3816	1,28383	1	7	1	115
123	38	87,2935	1,29280	1	121	1	117
124	37	87,2426	1,29798	1	85	1	118
125	36	87,1965	1,30267	1	100	1	120
126	35	87,0227	1,32035	1	138	1	121
127	34	86,9020	1,33263	1	31	1	122
128	33	86,6244	1,36087	1	129	1	123
129	32	86,6177	1,36156	1	4	1	124
130	31	86,5271	1,37077	1	108	1	125
131	30	86,4805	1,37552	1	14	1	126
132	29	86,3038	1,39350	83	135	83	2
133	28	86,2178	1,40225	1	126	1	127
134	27	86,1294	1,41124	1	109	1	128
135	26	85,9707	1,42738	1	23	1	129
136	25	85,9535	1,42914	1	78	1	130
137	24	85,9428	1,43023	83	161	83	3
138	23	85,8165	1,44307	1	97	1	131
139	22	85,4444	1,48093	1	84	1	132
140	21	85,0847	1,51753	1	52	1	133
141	20	85,0302	1,52307	1	59	1	136
142	19	85,0106	1,52507	1	48	1	139
143	18	84,8609	1,54030	1	62	1	140
144	17	84,8573	1,54067	1	57	1	141
145	16	84,7703	1,54952	19	131	19	2
146	15	84,7162	1,55502	1	86	1	142
147	14	84,5392	1,57303	1	157	1	143
148	13	84,4600	1,58109	1	75	1	144
149	12	84,2189	1,60562	1	91	1	146
150	11	84,0406	1,62376	1	145	1	147
151	10	83,8242	1,64578	1	111	1	149

152	9	83,7686	1,65144	1	71	1	150
153	8	83,2459	1,70461	1	83	1	153
154	7	82,2922	1,80165	1	19	1	155
155	6	81,9944	1,83195	1	69	1	156
156	5	81,6057	1,87150	1	66	1	157
157	4	79,6738	2,06805	1	113	1	158
158	3	77,1566	2,32416	1	63	1	159
159	2	69,7986	3,07279	1	127	1	160
160	1	39,9142	6,11332	1	151	1	161

Final Partition  
Number of clusters: 5

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	157	1333,76	2,75646	6,41862
Cluster2	1	0,00	0,00000	0,00000
Cluster3	1	0,00	0,00000	0,00000
Cluster4	1	0,00	0,00000	0,00000
Cluster5	1	0,00	0,00000	0,00000

Cluster Centroids

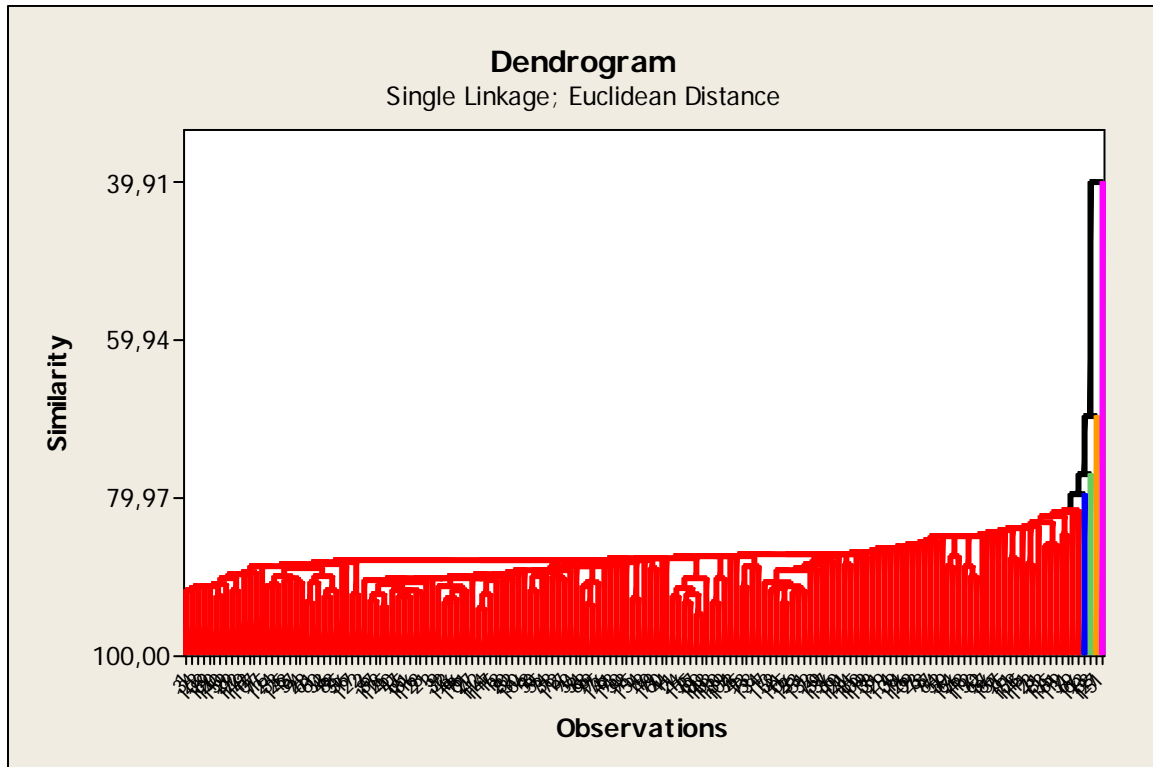
Variable	Cluster1	Cluster2	Cluster3	Cluster4
Population growth (annual %)	-0,0404902	-0,03682	0,35985	0,36834
Internet users (per 100 people)	-0,0108850	-0,89224	-0,47273	1,41099
Mobile cellular subscriptions (	-0,0101422	-1,23628	0,05130	1,38871
Age dependency ratio (% of work	0,0192678	0,41066	0,15038	-1,37749
Fertility rate, total (births p	0,0104024	0,31462	0,04676	-1,20512
GDP per capita (current US\$)	-0,0143582	-0,62105	-0,49506	1,72864
GDP per capita growth (annual %	-0,0022860	-3,00382	3,28484	3,09220
Life expectancy at birth, total	-0,0103878	-0,73608	0,32809	1,27591
Urban population (% of total)	-0,0205507	-0,20406	0,32635	2,04239

Variable	Cluster5	Grand centroid
Population growth (annual %)	5,66559	-0,0000000
Internet users (per 100 people)	1,66292	-0,0000000
Mobile cellular subscriptions (	1,38860	0,0000000
Age dependency ratio (% of work	-2,20859	-0,0000000
Fertility rate, total (births p	-0,78945	-0,0000000
GDP per capita (current US\$)	1,64170	-0,0000000
GDP per capita growth (annual %	-3,01432	-0,0000000
Life expectancy at birth, total	0,76297	-0,0000000
Urban population (% of total)	1,06179	0,0000000

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Cluster1	0,00000	3,53037	3,41571	5,10865	7,51698
Cluster2	3,53037	0,00000	6,56563	8,34529	7,93634
Cluster3	3,41571	6,56563	0,00000	4,25016	9,25955
Cluster4	5,10865	8,34529	4,25016	0,00000	8,21642
Cluster5	7,51698	7,93634	9,25955	8,21642	0,00000

E seu respectivo Dendograma:



Fica claro que o resultado é ruim, o Minitab pôs quase todos os países em um só cluster (157) e um país em cada um dos 4 clusters seguintes.

## 8.2 Análise de clusters para 30 países e 3 clusters

A seguir, a base de dados foi reduzida para 30 países, de modo a tornar possível a análise gráfica dos resultados, através do dendrograma. Segue o resultado apurado, para 3 clusters:

### Cluster Analysis of Observations: Population g; Internet use; Mobile cellu; ...

Standardized Variables, Euclidean Distance, Single Linkage  
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	29	92,8839	0,71596	8 20	8	2
2	28	91,4291	0,86233	8 28	8	3
3	27	90,1613	0,98989	12 14	12	2
4	26	89,6784	1,03847	1 13	1	2
5	25	88,1987	1,18734	8 29	8	4
6	24	86,8009	1,32798	8 25	8	5
7	23	86,7969	1,32839	8 26	8	6
8	22	85,8618	1,42247	10 12	10	3
9	21	85,7298	1,43574	10 11	10	4
10	20	85,4366	1,46524	8 23	8	7
11	19	85,3224	1,47673	1 8	1	9
12	18	85,2256	1,48647	9 21	9	2
13	17	85,1471	1,49437	6 30	6	2
14	16	84,3331	1,57627	1 7	1	10
15	15	84,1244	1,59726	9 10	9	6
16	14	83,8449	1,62538	3 18	3	2
17	13	83,5716	1,65289	1 15	1	11
18	12	82,3348	1,77732	9 17	9	7
19	11	82,2877	1,78206	6 9	6	9
20	10	81,6966	1,84153	2 24	2	2

21	9	81,1021	1,90134	3	19	3	3
22	8	81,0946	1,90210	1	16	1	12
23	7	80,5181	1,96010	2	6	2	11
24	6	79,2257	2,09013	2	3	2	14
25	5	78,3689	2,17633	2	4	2	15
26	4	76,9046	2,32366	1	2	1	27
27	3	75,2689	2,48823	1	22	1	28
28	2	72,2856	2,78839	1	5	1	29
29	1	37,8291	6,25510	1	27	1	30

Final Partition  
Number of clusters: 3

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	28	199,786	2,54161	5,64640
Cluster2	1	0,000	0,00000	0,00000
Cluster3	1	0,000	0,00000	0,00000

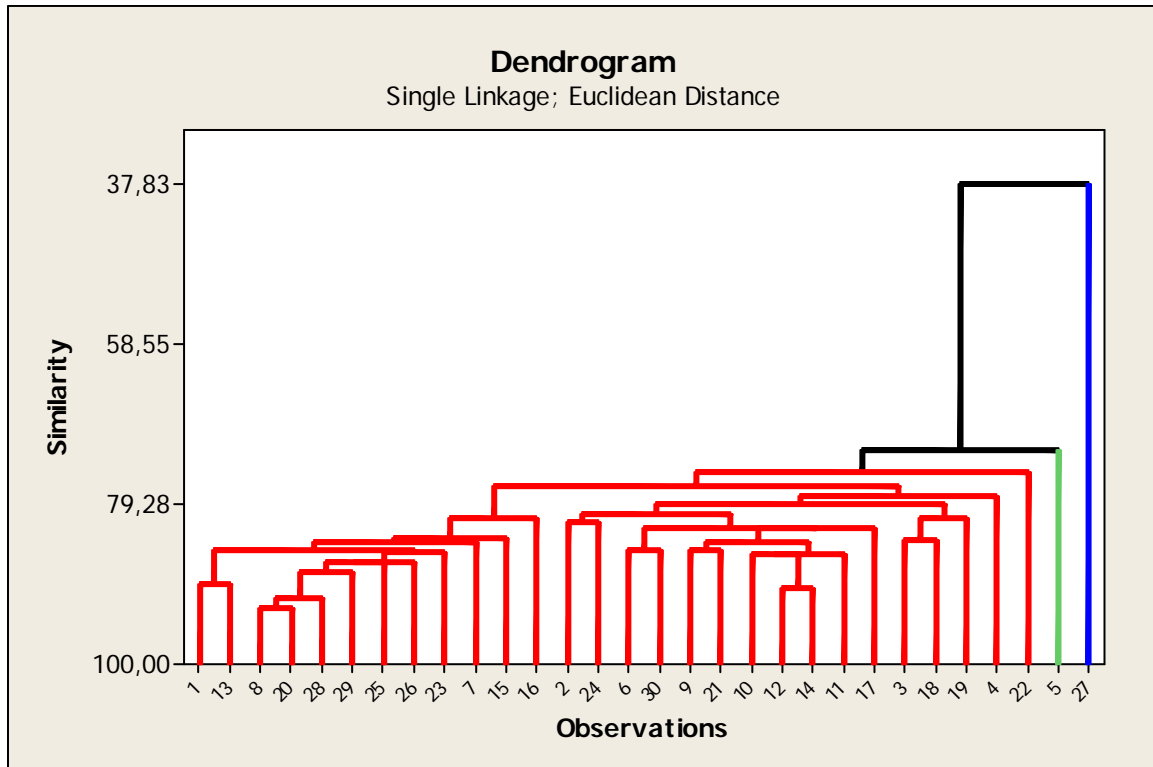
Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Grand centroid
Population growth (annual %)	-0,172709	0,09696	4,73890	0,000000
Internet users (per 100 people)	0,015290	-1,43283	1,00471	0,000000
Mobile cellular subscriptions (	-0,062630	0,44870	1,30494	-0,000000
Age dependency ratio (% of work	0,087851	0,43763	-2,89746	0,000000
Fertility rate, total (births p	-0,009458	0,83733	-0,57251	0,000000
GDP per capita (current US\$)	-0,010771	-0,63561	0,93722	0,000000
GDP per capita growth (annual %	0,066729	1,14585	-3,01428	0,000000
Life expectancy at birth, total	0,104477	-3,20991	0,28455	0,000000
Urban population (% of total)	-0,019812	-0,12953	0,68426	0,000000

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3
Cluster1	0,00000	3,97778	6,86452
Cluster2	3,97778	0,00000	8,60156
Cluster3	6,86452	8,60156	0,00000

E seu respectivo Dendograma:



Mais uma vez o Minitab não separou muito bem os clusters, colocando 28 países no primeiro e um país em cada um dos 2 clusters restantes.

### 8.3 Análise de clusters para 30 países e 2 clusters

A seguir, a base de dados foi reduzida para 30 países, de modo a tornar possível a análise gráfica dos resultados, através do dendrograma. Segue o resultado apurado para 2 clusters:

#### Cluster Analysis of Observations: Population g; Internet use; Mobile cellu; ...

Standardized Variables, Euclidean Distance, Single Linkage  
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined	New cluster	Number of obs. in new cluster
1	29	92,8839	0,71596	8 20	8	2
2	28	91,4291	0,86233	8 28	8	3
3	27	90,1613	0,98989	12 14	12	2
4	26	89,6784	1,03847	1 13	1	2
5	25	88,1987	1,18734	8 29	8	4
6	24	86,8009	1,32798	8 25	8	5
7	23	86,7969	1,32839	8 26	8	6
8	22	85,8618	1,42247	10 12	10	3
9	21	85,7298	1,43574	10 11	10	4
10	20	85,4366	1,46524	8 23	8	7
11	19	85,3224	1,47673	1 8	1	9
12	18	85,2256	1,48647	9 21	9	2
13	17	85,1471	1,49437	6 30	6	2
14	16	84,3331	1,57627	1 7	1	10
15	15	84,1244	1,59726	9 10	9	6
16	14	83,8449	1,62538	3 18	3	2
17	13	83,5716	1,65289	1 15	1	11
18	12	82,3348	1,77732	9 17	9	7
19	11	82,2877	1,78206	6 9	6	9

20	10	81,6966	1,84153	2	24	2	2
21	9	81,1021	1,90134	3	19	3	3
22	8	81,0946	1,90210	1	16	1	12
23	7	80,5181	1,96010	2	6	2	11
24	6	79,2257	2,09013	2	3	2	14
25	5	78,3689	2,17633	2	4	2	15
26	4	76,9046	2,32366	1	2	1	27
27	3	75,2689	2,48823	1	22	1	28
28	2	72,2856	2,78839	1	5	1	29
29	1	37,8291	6,25510	1	27	1	30

Final Partition  
Number of clusters: 2

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	29	215,063	2,59250	5,59041
Cluster2	1	0,000	0,00000	0,00000

Cluster Centroids

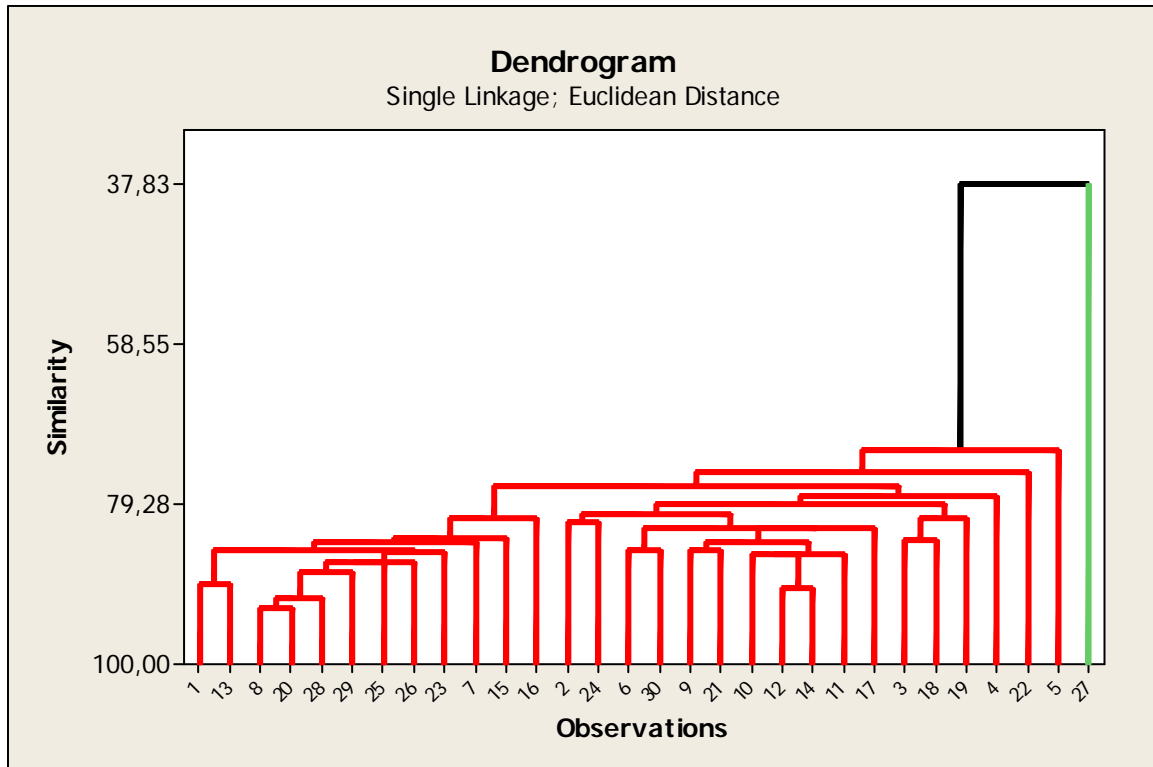
Variable	Cluster1	Cluster2	Grand centroid
Population growth (annual %)	-0,163410	4,73890	0,000000
Internet users (per 100 people)	-0,034645	1,00471	0,000000
Mobile cellular subscriptions (	-0,044998	1,30494	-0,000000
Age dependency ratio (% of work	0,099913	-2,89746	0,000000
Fertility rate, total (births p	0,019742	-0,57251	0,000000
GDP per capita (current US\$)	-0,032318	0,93722	0,000000
GDP per capita growth (annual %	0,103941	-3,01428	0,000000
Life expectancy at birth, total	-0,009812	0,28455	0,000000
Urban population (% of total)	-0,023595	0,68426	0,000000

Distances Between Cluster Centroids

	Cluster1	Cluster2
Cluster1	0,00000	6,89357
Cluster2	6,89357	0,00000

E seu respectivo Dendograma:





Mais uma vez o Minitab não separou muito bem os clusters, colocando 29 países no primeiro e um país no outro cluster.

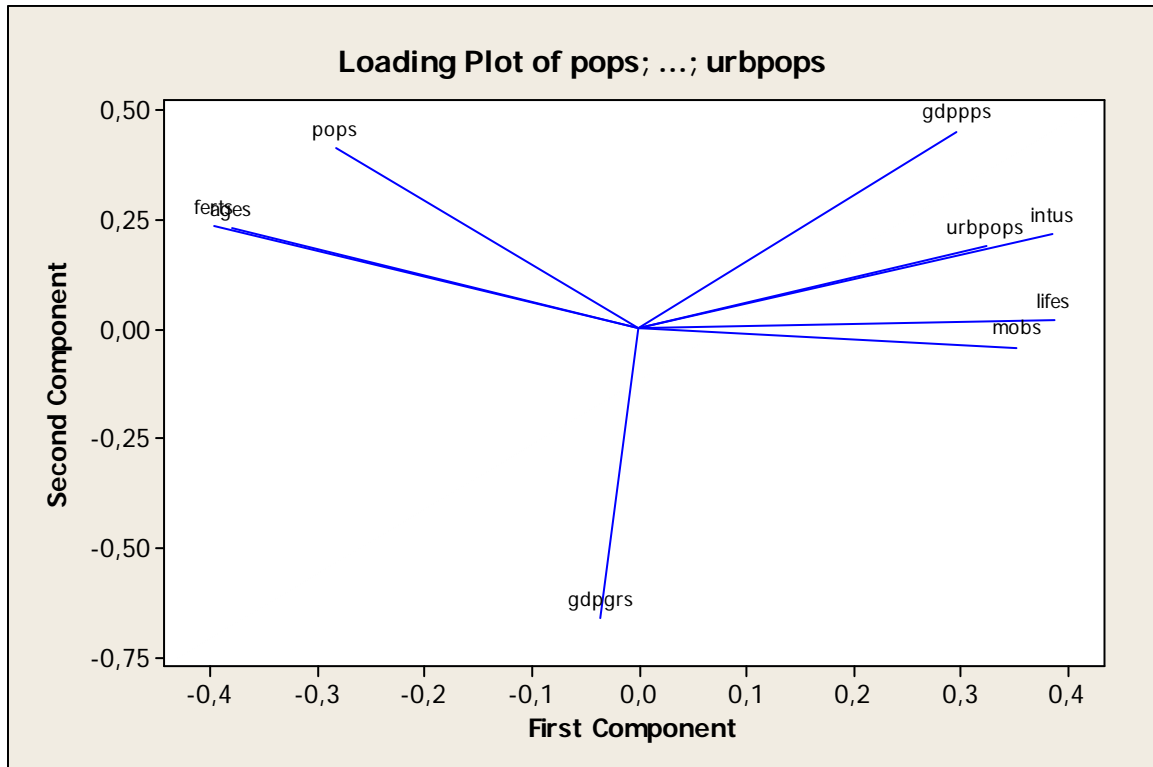
Portanto, será utilizado o cluster definido manualmente e não o apurado calculado pelo Minitab, para as análises dos tópicos seguintes.

## 9. ANÁLISE DE COMPONENTES

A proposta deste tópico é realizar análises entre as variáveis para identificar quais delas apresentam a maior diferenciação dos clusters.

### 9.1 **Análise de Componentes Principais**

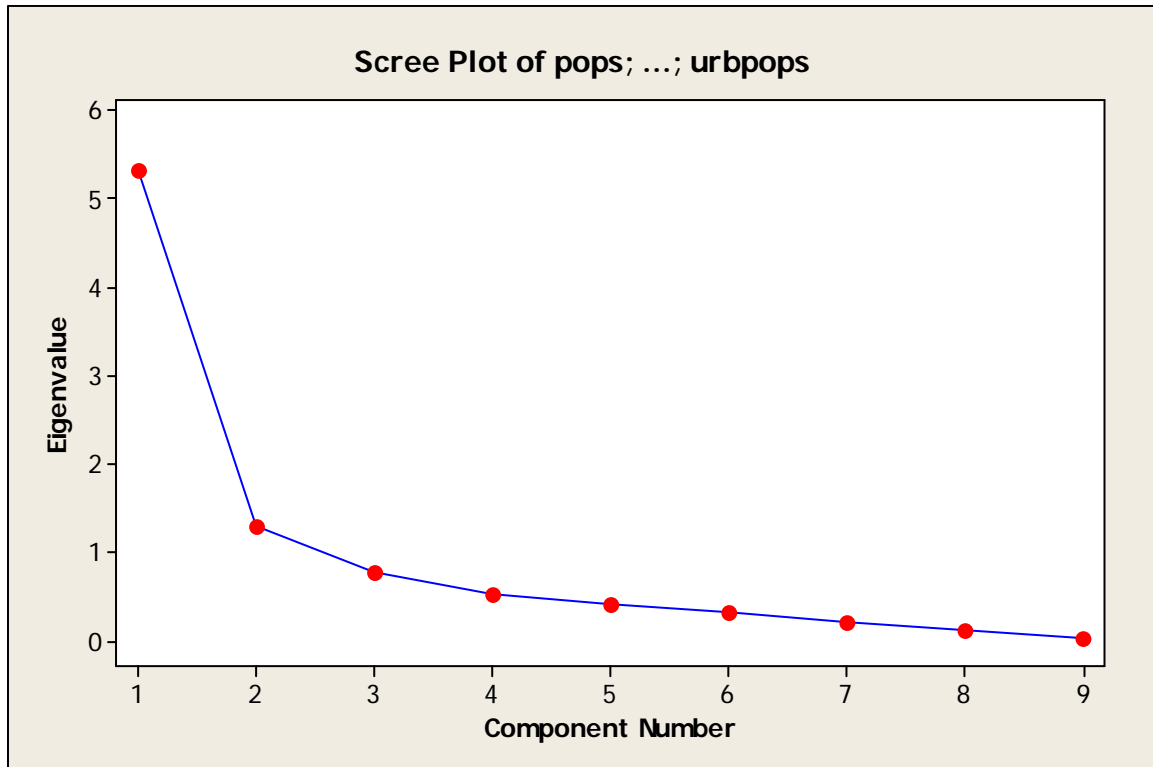
O objetivo deste tópico é, através da análise dos componentes principais, tentar reduzir o número de variáveis, ou seja, perceber as relações entre as variáveis e a possibilidade de agrupá-las. Certamente a análise de correlações e dendogramas anterior já dá uma ideia de que a possibilidade de agrupamento é grande pelos índices de correlação entre todas as variáveis:



Analisando o Loading Plot é possível identificar 6 grupos de variáveis que podem ser unidas, para simplificar a análise. O resultado será 6 grupos ao invés de 9 variáveis independentes. São eles:

1. GRUPO1 – PIB per Capita. Por ser composto de somente uma variável, o Grupo 1 será batizado de **Renda**;
2. GRUPO2 – Expectativa de vida ao nascer e Assinaturas de celular. A variável Grupo 2 será batizada de **Idade e celular**;
3. GRUPO3 – Usuários de Internet e População Urbana, que tem comportamento aproximado. Esta variável será batizada de **Cidadãos conectados**;
4. GRUPO4 – % Crescimento do PIB per Capita. Por ser composto de somente uma variável, o grupo 4 será batizado de **Aumento da renda**;
5. GRUPO 5 – Crescimento populacional. Por ser composto de somente uma variável, o Grupo 5 será batizado de **População**; e
6. GRUPO 6 – Razão de dependência da idade e Taxa de Fertilidade, o grupo 6 será batizado de **Crescimento Populacional**.

O Scree plot permite identificar a influência das variáveis agrupadas sobre o resultado final, como segue:



Como o Scree Plot aponta, o componente agrupado número 1, PC1, tem uma influência muito maior sobre o resultado geral que os demais componentes. O mesmo ocorre para os componentes agrupados PC2 e PC3. Os demais componentes continuam influenciando o resultado, mas de maneira mais discreta. Na abordagem mais usual, utilizam-se os componentes do Scree plot até que a curva se torne praticamente horizontal. No caso, a curva não chega a ficar praticamente horizontal, porém vamos utilizar até componente número PC3.

Através da análise de Componentes principais é possível validar a proposta de agrupamento, como segue:

#### Principal Component Analysis: pops; intus; mobs; ages; ferts; gdppps; gdpgrs; l

Eigenanalysis of the Correlation Matrix

Eigenvalue	5,3270	1,2937	0,7847	0,5152	0,4018	0,3179	0,2145	0,1180
Proportion	0,592	0,144	0,087	0,057	0,045	0,035	0,024	0,013
Cumulative	0,592	0,736	0,823	0,880	0,925	0,960	0,984	0,997

Eigenvalue	0,0272
Proportion	0,003
Cumulative	1,000

Variable	PC1	PC2	PC3	PC4
pops	-0,283	0,411	-0,348	0,446
intus	0,386	0,217	-0,156	-0,285
mobs	0,353	-0,043	0,034	0,520
ages	-0,381	0,227	-0,156	-0,118
ferts	-0,396	0,232	-0,188	-0,009
gdppps	0,296	0,451	-0,335	-0,490
gdpgrs	-0,035	-0,661	-0,727	-0,104
lifes	0,387	0,020	0,040	0,017
urbpops	0,324	0,189	-0,389	0,430

Olhando a matriz Principal Components Analysis, é possível afirmar que a variável agrupada PC1 explica 59,2% do total, enquanto que ela somada à variável agrupada PC2 explica 73,6%, e, ambas somadas à variável PC3 somam 82,3%. Aqui também, a abordagem mais aceita é a de selecionar os componentes com Eigenvalue superior a 1. No caso, somente os dois primeiros componentes possuem Eigenvalue superior a 1, tendo sido selecionado ainda o componente PC3, com Eigenvalue de 0,7847.

## 9.2 Análise Fatorial

A análise fatorial é uma ferramenta matemática de grande utilidade para os estatísticos, através da qual, é possível modelar dados cujas interdependências são desconhecidas ou emaranhadas.

A análise fatorial pode ser aplicada para explorar uma área de conteúdos, estruturar um domínio, classificar ou reduzir dados, esclarecer nexos causais, definir relações entre variáveis, formular teorias, etc.

Neste trabalho, a Análise Fatorial será utilizada para identificar relações causais entre as 9 variáveis selecionadas. Por exemplo, será possível identificar se Crescimento populacional tem relação com Usuários de Internet, e assim por diante. Uma relação que não é intuitiva poderá ser identificada, viabilizando a junção de blocos que não demonstram claramente correlação. Essa correlação entre variáveis, aparentemente não diretamente ou não intuitivamente ligadas, é “descoberta” pela Análise fatorial através dos passos que serão realizados a seguir.

O método busca agrupar as variáveis analisadas em fatores que sejam independentes entre si. A independência é necessária para demonstrar quanto um fator explica do total. Caso houvesse dependência, não seria possível explicar o comportamento do total por fatores diferentes. Cada fator é uma espécie de influência diferente sobre o total, como se fossem categorias que explicassem de maneira independente o comportamento do total.

### Factor Analysis: pops; intus; mobs; ages; ferts; gdppps; gdpgrs; lifes; urbpops

Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Factor4	Communality
pops	-0,654	0,467	-0,308	0,320	0,843
intus	0,891	0,247	-0,138	-0,204	0,916
mobs	0,815	-0,049	0,030	0,373	0,807
ages	-0,878	0,258	-0,138	-0,084	0,865
ferts	-0,915	0,264	-0,166	-0,007	0,934
gdppps	0,684	0,513	-0,296	-0,351	0,943
gdpgrs	-0,082	-0,752	-0,644	-0,075	0,993
lifes	0,894	0,023	0,035	0,012	0,801
urbpops	0,747	0,215	-0,345	0,309	0,819
Variance	5,3270	1,2937	0,7847	0,5152	7,9205
% Var	0,592	0,144	0,087	0,057	0,880

Factor Score Coefficients

Variable	Factor1	Factor2	Factor3	Factor4
pops	-0,123	0,361	-0,392	0,621
intus	0,167	0,191	-0,176	-0,397
mobs	0,153	-0,038	0,038	0,725
ages	-0,165	0,200	-0,176	-0,164
ferts	-0,172	0,204	-0,212	-0,013
gdppps	0,128	0,396	-0,378	-0,682
gdpgrs	-0,015	-0,581	-0,821	-0,145
lifes	0,168	0,018	0,045	0,023
urbpops	0,140	0,166	-0,439	0,599

Os fatores estão classificados de Factor1 ao Factor4 segundo a magnitude de variação nos dados descrita por cada padrão. Assim, o primeiro fator explica a maior quantidade de variação e o último fator a menor quantidade de variação.

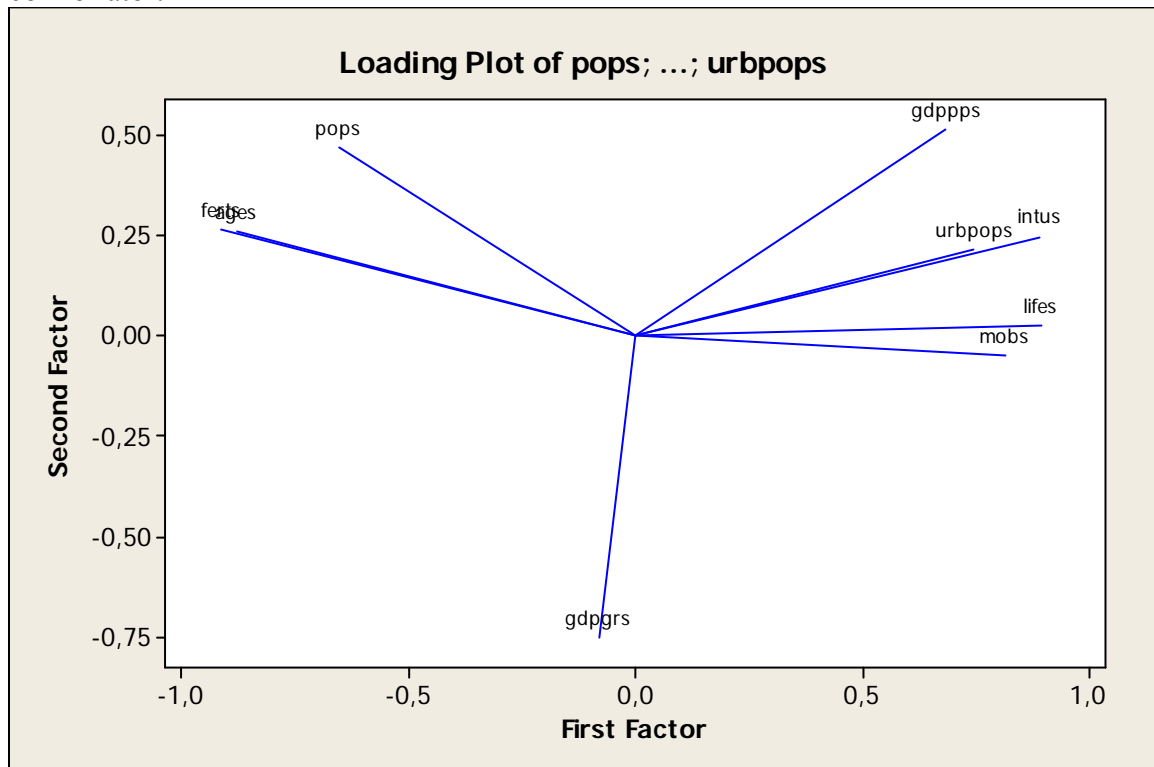
A coluna Commuality demonstra a “comunalidade” de cada variável, isto é, a proporção da variação total da variável que é explicada pelos fatores. E a comunalidade é sempre ser maior que qualquer fator.

#### Loading Plot (none)

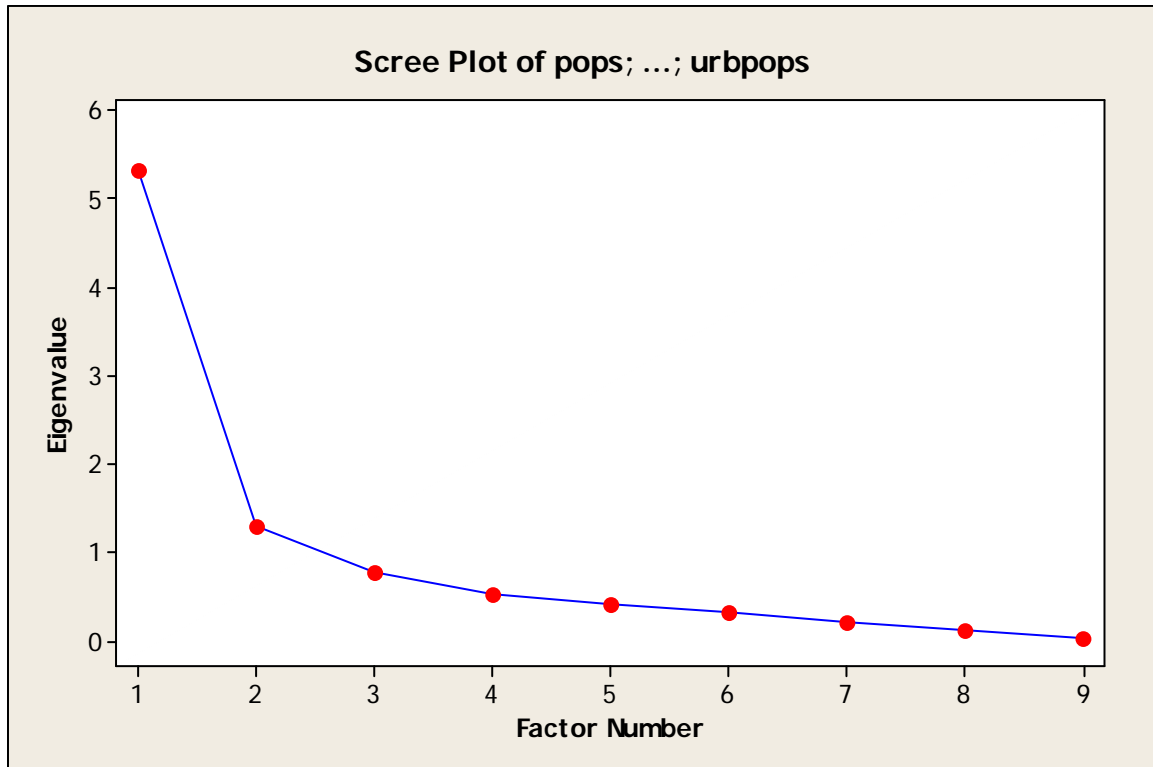
Para analisar o Loading Plot é necessário entender os vetores que refletem o comportamento de cada variável. A configuração das linhas do Loading Plot vai refletir a interrelação entre as Variáveis. As variáveis interrelacionadas estarão próximas umas das outras. Variáveis não relacionadas estarão a um ângulo reto uma da outra.

O número de dimensões seria igual ao número de países selecionados, ou seja, 161. Como seria impossível analisar este número de dimensões, a análise de Fatores identifica 2 fatores, através dos quais são plotados os comportamentos das variáveis analisadas. Esses 2 fatores são resultado da junção de variáveis que tem maior correlação entre si.

Os loadings podem ser interpretados como coeficientes de correlação das variáveis com o fator.



#### Scree plot (none)



Os fatores são selecionados pela combinação da análise do Scree plot e da variável Eigenvalue. No exemplo acima, somente os 2 primeiros fatores possuem Scree plot maior do que 1.

### 9.2.1 Rotações

Há várias estratégias de rotação que tem sido propostas pelos estudiosos. O objetivo de todas essas estratégias é obter um padrão mais simples dos loadings, ou seja, fatores que, de alguma maneira, são marcados por loadings altos para algumas variáveis e loadings baixos para outras. Esse modelo é também, algumas vezes, referido como rotação estrutural. As estratégias de rotação típica são VARIMAX, QUARTIMAX e EQUAMAX.

Todas as rotações efetuadas pelo Minitab são ortogonais, ou seja, resultam em escores de fatores não correlacionados e loadings de fatores ortogonais.

Essas rotações movem toda a estrutura de fatores em uma estrutura rígida em torno da origem. Portanto, as “communalities”, ficam invariáveis entre a situação não rotacionada e as rotacionadas.

#### **ROTAÇÃO VARIMAX**

A estratégia de Rotação VARIMAX procura maximizar a variância com relação aos novos eixos, ou seja, procura obter um padrão para os loadings de cada fator, que seja tão diverso quanto possível, permitindo uma interpretação mais fácil. Isso se evidencia através dos números pintados em amarelo na matriz abaixo:

#### **Factor Analysis: pops; intus; mobs; ages; ferts; gdppps; gdpgrs; lifes; urb pops**

Principal Component Factor Analysis of the Correlation Matrix

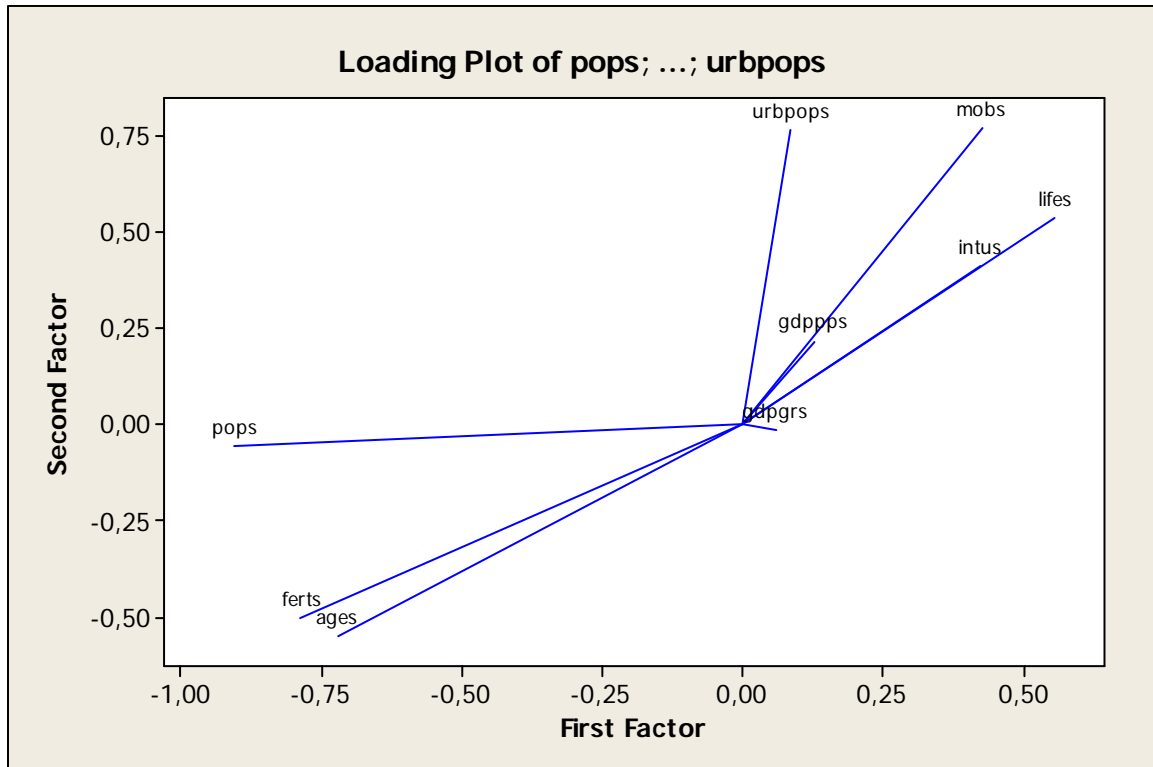
Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Factor4	Communality
pops	-0,654	0,467	-0,308	0,320	0,843
intus	0,891	0,247	-0,138	-0,204	0,916
mobs	0,815	-0,049	0,030	0,373	0,807

ages	-0,878	0,258	-0,138	-0,084	0,865
ferts	-0,915	0,264	-0,166	-0,007	0,934
gdppps	0,684	0,513	-0,296	-0,351	0,943
gdpgrs	-0,082	-0,752	-0,644	-0,075	0,993
lifes	0,894	0,023	0,035	0,012	0,801
urbpops	0,747	0,215	-0,345	0,309	0,819
Variance	5,3270	1,2937	0,7847	0,5152	7,9205
% Var	0,592	0,144	0,087	0,057	0,880
Rotated Factor Loadings and Communalities					
Varimax Rotation					
Variable	Factor1	Factor2	Factor3	Factor4	Communality
pops	-0,904	-0,059	-0,120	0,086	0,843
intus	0,424	0,410	0,749	0,086	0,916
mobs	0,426	0,768	0,177	0,069	0,807
ages	-0,721	-0,550	-0,205	0,018	0,865
ferts	-0,785	-0,506	-0,249	0,007	0,934
gdppps	0,127	0,215	0,930	0,122	0,943
gdpgrs	0,060	-0,017	-0,116	-0,988	0,993
lifes	0,557	0,536	0,443	0,087	0,801
urbpops	0,085	0,762	0,479	-0,040	0,819
Variance	2,6520	2,2328	2,0162	1,0196	7,9205
% Var	0,295	0,248	0,224	0,113	0,880
Factor Score Coefficients					
Variable	Factor1	Factor2	Factor3	Factor4	
pops	-0,657	0,503	-0,029	0,011	
intus	0,048	-0,166	0,471	-0,035	
mobs	-0,102	0,646	-0,341	0,086	
ages	-0,239	-0,182	0,184	-0,022	
ferts	-0,312	-0,062	0,120	-0,030	
gdppps	-0,089	-0,363	0,797	-0,078	
gdpgrs	-0,051	0,009	0,121	-1,008	
lifes	0,109	0,110	0,063	0,056	
urbpops	-0,406	0,642	0,034	-0,143	

No caso da rotação VARIMAX é possível ver que as variáveis Crescimento Populacional e Taxa de Fertilidade estão muito relacionadas ao fator 1, enquanto que as variáveis Assinaturas de Celular e População urbana estão mais relacionadas ao fator 2 e assim por diante, permitindo verificar o agrupamento de variáveis (clusters).

O Loading plot VARIMAX demonstra a explicação anterior de maneira gráfica:



### **ROTACÃO QUARTIMAX**

A estratégia de Rotação QUARTIMAX procura maximizar os loadings de cada variável, ou seja, olhando-se as linhas da matriz para cada variável, os loadings maiores são aumentados e os menores diminuídos. Em suma, cada variável terá load em menos fatores. Isso se evidencia através dos números pintados em amarelo na matriz abaixo:

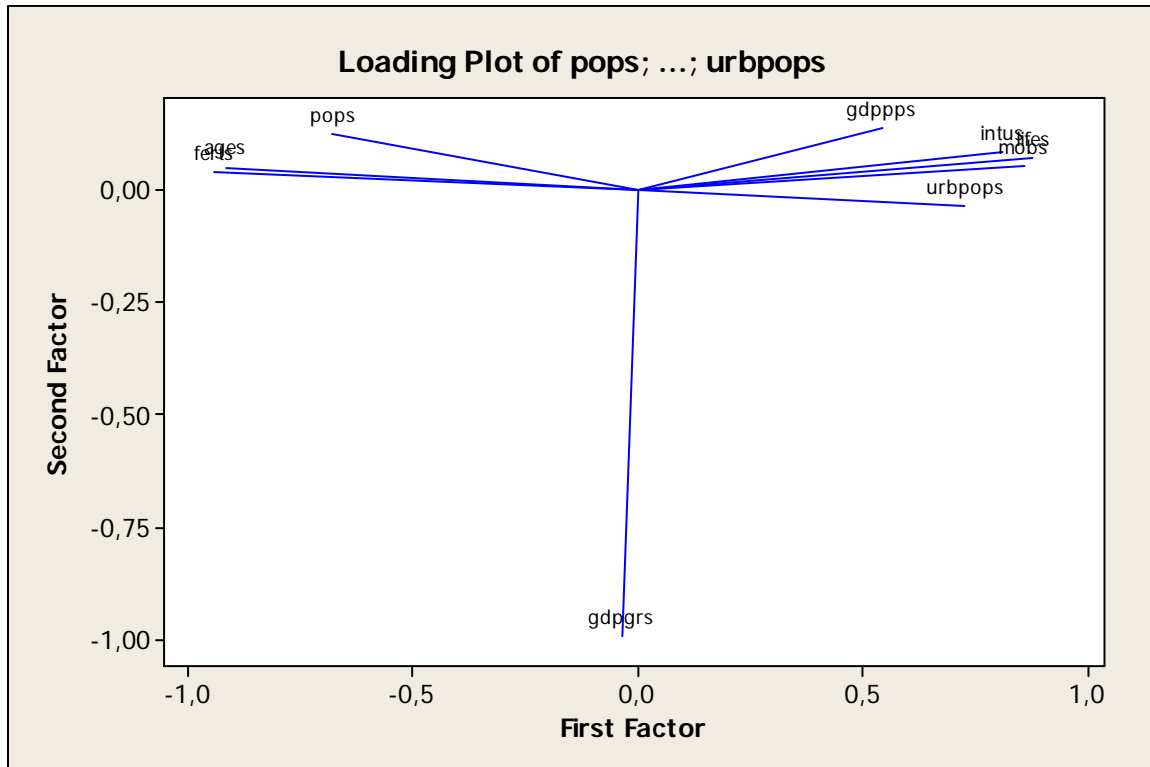
Rotated Factor Loadings and Communalities					
Quartimax Rotation					
Variable	Factor1	Factor2	Factor3	Factor4	Communality
pops	-0,677	0,124	-0,080	0,602	0,843
intus	<b>0,808</b>	0,082	-0,506	0,011	0,916
mobs	0,857	0,050	0,129	0,231	0,807
ages	<b>-0,914</b>	0,048	-0,100	0,129	0,865
ferts	<b>-0,942</b>	0,039	-0,060	0,204	0,934
gdppps	0,543	0,135	-0,788	0,097	0,943
gdppgs	-0,036	<b>-0,992</b>	0,086	-0,031	0,993
lifes	<b>0,878</b>	0,069	-0,158	-0,012	0,801
urbpops	0,725	-0,038	-0,228	0,489	0,819
Variance	5,1627	1,0361	0,9972	0,7245	7,9205
% Var	0,574	0,115	0,111	0,080	0,880

#### Factor Score Coefficients

Variable	Factor1	Factor2	Factor3	Factor4
pops	-0,108	0,037	0,034	0,819
intus	0,077	-0,026	-0,480	-0,126
mobs	0,253	0,080	0,479	0,501
ages	-0,219	-0,006	-0,271	0,055
ferts	-0,209	-0,013	-0,192	0,188
gdppps	-0,037	-0,055	-0,869	-0,147
gdppgs	-0,013	-1,002	-0,153	0,079
lifes	0,168	0,052	-0,006	-0,001
urbpops	0,168	-0,127	0,058	0,743

O Loading plot QUARTIMAX



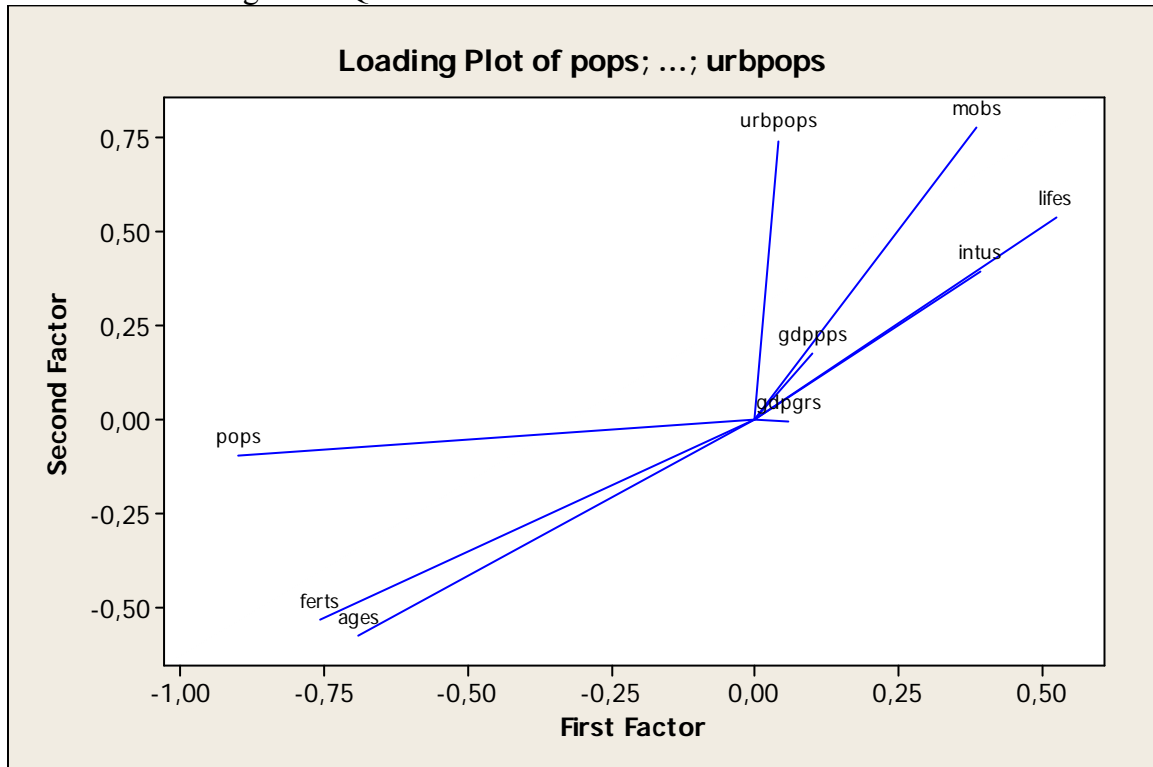


### ROTAÇÃO EQUAMAX

A seguir, é exibida a rotação EQUAMAX. A estratégia de rotação EQUAMAX também maximiza a variância, mas de forma ponderada entre os critérios da QUARTIMAX e VARIMAX, refletindo a preocupação de obter uma estrutura simplificada, tanto para variáveis quanto para os fatores. Segue a matriz resultante da rotação EQUAMAX, com os números evidenciados em amarelo demonstrando as variáveis mais ligadas aos Fatores:

Rotated Factor Loadings and Communalities					
Equamax Rotation					
Variable	Factor1	Factor2	Factor3	Factor4	Communality
pops	-0,898	-0,096	-0,142	0,085	0,843
intus	0,392	0,394	0,772	0,108	0,916
mobs	0,386	0,777	0,218	0,078	0,807
ages	-0,691	-0,573	-0,243	0,011	0,865
ferts	-0,756	-0,530	-0,286	-0,001	0,934
gdppps	0,102	0,177	0,938	0,147	0,943
gdppgrs	0,059	-0,004	-0,089	-0,991	0,993
lifes	0,524	0,540	0,475	0,101	0,801
urbpops	0,041	0,742	0,516	-0,023	0,819
Variance	2,4480	2,2512	2,1827	1,0386	7,9205
% Var	0,272	0,250	0,243	0,115	0,880
Factor Score Coefficients					
Variable	Factor1	Factor2	Factor3	Factor4	
pops	-0,679	0,473	-0,018	0,015	
intus	0,048	-0,185	0,464	-0,024	
mobs	-0,127	0,655	-0,315	0,082	
ages	-0,233	-0,202	0,172	-0,017	
ferts	-0,311	-0,082	0,112	-0,026	
gdppps	-0,085	-0,403	0,780	-0,059	
gdppgrs	-0,057	0,007	0,147	-1,004	
lifes	0,103	0,112	0,068	0,058	
urbpops	-0,437	0,621	0,060	-0,136	

O Loading Plot EQUAMAX



## 10. ANÁLISE DISCRIMINANTE

A variável dependente dessa análise será o Cluster, e, para tentar explicar em qual conglomerado um determinado país cai, foram utilizadas as 9 variáveis em análise neste trabalho.

Segue o resultado da Análise Discriminante, com opção LINEAR:

### Discriminant Analysis: Cluster versus pops; intus; ...

Linear Method for Response: Cluster

Predictors: pops; intus; mobs; ages; ferts; gdppps; gdpgrs; lifes; urb pops

Group	1	2	3	4	5
Count	19	16	59	35	32

Summary of classification

Put into Group	True Group				
	1	2	3	4	5
1	19	0	0	0	0
2	0	15	4	0	0
3	0	1	47	5	1
4	0	0	6	21	6
5	0	0	2	9	25
Total N	19	16	59	35	32
N correct	19	15	47	21	25
Proportion	1,000	0,938	0,797	0,600	0,781

**N = 161      N Correct = 127      Proportion Correct = 0,789**

Squared Distance Between Groups

	1	2	3	4	5
--	---	---	---	---	---

1	0,0000	20,7816	52,6158	66,8169	78,3822
2	20,7816	0,0000	9,4746	20,2510	31,2092
3	52,6158	9,4746	0,0000	4,1097	13,2095
4	66,8169	20,2510	4,1097	0,0000	3,6819
5	78,3822	31,2092	13,2095	3,6819	0,0000

## Linear Discriminant Function for Groups

	1	2	3	4	5
Constant	-136,52	-91,80	-76,77	-79,97	-75,99
pops	11,70	9,62	8,94	9,83	9,90
intus	6,45	5,27	3,67	3,47	3,79
mobs	2,78	3,22	3,42	3,13	2,66
ages	13,90	12,30	12,15	13,15	13,09
ferts	-8,28	-6,62	-6,42	-6,73	-6,15
gdppps	12,97	5,77	2,34	1,60	1,53
gdpgrs	5,10	4,37	4,84	5,15	4,79
lifes	3,24	3,49	3,40	3,44	2,89
urbpops	-0,73	-0,41	-0,26	-0,85	-1,26

## Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
4**	3	5	1	74,026	0,000
			2	34,501	0,000
			3	19,134	0,002
			4	10,761	0,128
			5	6,934	0,870
11**	5	4	1	80,039	0,000
			2	31,302	0,000
			3	11,130	0,065
			4	5,960	0,864
			5	10,963	0,071
14**	3	4	1	68,90	0,000
			2	28,40	0,000
			3	17,50	0,085
			4	12,84	0,872
			5	18,85	0,043
24**	5	4	1	82,583	0,000
			2	32,175	0,000
			3	10,882	0,062
			4	5,891	0,757
			5	8,763	0,180
25**	4	5	1	82,978	0,000
			2	33,192	0,000
			3	14,223	0,020
			4	9,578	0,207
			5	6,938	0,773
37**	4	5	1	81,721	0,000
			2	30,515	0,000
			3	10,767	0,034
			4	5,766	0,409
			5	5,150	0,557
38**	3	2	1	41,068	0,000
			2	5,623	0,825
			3	8,732	0,174
			4	18,595	0,001
			5	28,894	0,000
47**	3	2	1	33,926	0,000
			2	5,733	0,955
			3	11,874	0,044
			4	21,535	0,000
			5	31,798	0,000
53**	5	4	1	77,806	0,000
			2	28,854	0,000
			3	10,754	0,030
			4	4,895	0,552

			5	5,452	0,418
54**	4	3	1	62,905	0,000
			2	16,837	0,001
			3	2,864	0,720
			4	4,771	0,278
			5	15,033	0,002
65**	3	2	1	38,296	0,000
			2	5,526	0,828
			3	8,683	0,171
			4	18,766	0,001
			5	29,078	0,000
67**	3	4	1	75,911	0,000
			2	29,024	0,000
			3	8,962	0,077
			4	4,177	0,845
			5	8,934	0,078
79**	5	3	1	81,26	0,000
			2	24,19	0,001
			3	10,54	0,618
			4	11,83	0,325
			5	15,34	0,056
83**	4	5	1	92,06	0,000
			2	46,41	0,000
			3	25,58	0,006
			4	18,55	0,206
			5	15,87	0,788
98**	4	3	1	66,375	0,000
			2	18,765	0,002
			3	6,340	0,851
			4	9,889	0,144
			5	17,811	0,003
99**	4	3	1	75,474	0,000
			2	24,135	0,000
			3	6,177	0,695
			4	7,859	0,300
			5	15,958	0,005
101**	4	3	1	55,027	0,000
			2	10,248	0,034
			3	3,870	0,836
			4	7,659	0,126
			5	14,772	0,004
103**	3	4	1	65,928	0,000
			2	23,093	0,000
			3	7,047	0,072
			4	2,430	0,721
			5	4,922	0,207
104**	5	4	1	79,72	0,000
			2	35,75	0,000
			3	19,21	0,007
			4	10,05	0,689
			5	11,68	0,304
108**	4	5	1	76,629	0,000
			2	34,385	0,000
			3	20,788	0,003
			4	13,798	0,102
			5	9,461	0,895
112**	4	5	1	80,958	0,000
			2	37,015	0,000
			3	17,780	0,001
			4	6,387	0,326
			5	4,941	0,672
116**	3	2	1	37,841	0,000
			2	4,108	0,663
			3	5,476	0,334
			4	14,699	0,003
			5	25,733	0,000
121**	3	4	1	90,57	0,000
			2	36,10	0,000

			3	18,27	0,094
			4	14,05	0,776
			5	17,63	0,130
123**	2	3	1	52,02	0,000
			2	17,40	0,312
			3	15,83	0,682
			4	25,31	0,006
			5	38,28	0,000
124**	4	5	1	71,906	0,000
			2	26,478	0,000
			3	11,890	0,006
			4	4,274	0,284
			5	2,439	0,710
130**	4	5	1	81,755	0,000
			2	38,579	0,000
			3	21,813	0,001
			4	9,915	0,298
			5	8,206	0,701
135**	3	5	1	84,37	0,000
			2	38,86	0,000
			3	22,35	0,014
			4	17,45	0,158
			5	14,14	0,828
139**	5	4	1	74,901	0,000
			2	24,737	0,000
			3	7,461	0,059
			4	2,293	0,784
			5	5,510	0,157
144**	3	4	1	83,05	0,000
			2	32,17	0,000
			3	17,73	0,038
			4	12,25	0,597
			5	13,23	0,365
148**	3	4	1	78,712	0,000
			2	31,709	0,000
			3	10,461	0,332
			4	9,112	0,651
			5	16,396	0,017
158**	4	3	1	83,21	0,000
			2	31,62	0,000
			3	16,19	0,730
			4	18,19	0,269
			5	30,28	0,001
159**	4	5	1	76,700	0,000
			2	34,398	0,000
			3	18,754	0,001
			4	7,190	0,338
			5	5,850	0,661
160**	4	5	1	94,63	0,000
			2	46,74	0,000
			3	26,94	0,000
			4	17,02	0,031
			5	10,13	0,969
161**	5	4	1	89,03	0,000
			2	45,28	0,000
			3	24,17	0,021
			4	17,68	0,533
			5	18,04	0,446

O resultado pode ser considerado aceitável, já que o modelo realizou 127 opções corretas, em 161 opções possíveis:

N = 161	N Correct = 127	Proportion Correct = 0,789
---------	-----------------	----------------------------

A opção Quadrática deu um resultado um pouco diferente, conforme segue logo abaixo:

**Discriminant Analysis: Cluster versus pops; intus; ...**

Quadratic Method for Response: Cluster

Predictors: pops; intus; mobs; ages; fertis; gdppps; gdpgrs; lifes; urbpops

Group	1	2	3	4	5
Count	19	16	59	35	32

## Summary of classification

Put into Group	True Group				
	1	2	3	4	5
1	19	0	0	0	0
2	0	16	4	0	0
3	0	0	53	1	0
4	0	0	2	34	1
5	0	0	0	0	31
Total N	19	16	59	35	32
N correct	19	16	53	34	31
Proportion	1,000	1,000	0,898	0,971	0,969

N = 161	N Correct = 153	Proportion Correct = 0,950
---------	-----------------	----------------------------

## From Generalized Squared Distance to Group

Group	1	2	3	4	5
1	-7	65	285	6966	253695
2	7	-8	34	1242	49258
3	162	68	-2	56	3491
4	657	249	2	-6	133
5	1543	581	14	-2	-13

## Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
15**	5	4	1	1654,28	0,000
			2	490,28	0,000
			3	18,54	0,000
			4	0,30	0,586
			5	1,00	0,414
27**	3	4	1	206,416	0,000
			2	118,434	0,000
			3	2,689	0,272
			4	0,719	0,728
			5	677,127	0,000
38**	3	2	1	30,8	0,000
			2	4,2	0,833
			3	7,4	0,167
			4	360,7	0,000
			5	15567,9	0,000
47**	3	2	1	29,0	0,000
			2	-1,2	0,992
			3	8,3	0,008
			4	408,3	0,000
			5	17403,1	0,000
65**	3	2	1	60,4	0,000
			2	-1,0	0,964
			3	5,6	0,036
			4	314,8	0,000
			5	13760,3	0,000
67**	3	4	1	636,912	0,000
			2	332,203	0,000
			3	4,681	0,031
			4	-2,234	0,969
			5	59,854	0,000

99**	4	3	1	316,775	0,000
			2	154,815	0,000
			3	4,451	0,501
			4	4,456	0,499
			5	260,273	0,000
116**	3	2	1	33,0	0,000
			2	2,2	0,582
			3	2,9	0,418
			4	281,5	0,000
			5	12476,6	0,000

O resultado pode ser considerado excelente, já que o modelo realizou 153 opções corretas, em 161 opções possíveis, aproximando o resultado dos 100% de acerto:

N = 161	N Correct = 153	Proportion Correct = 0,950
---------	-----------------	----------------------------

## 11. REGRESSÃO LOGÍSTICA

A Regressão logística também tem a função de definir se a clusterização realizada anteriormente resultará em dados confiáveis.

Segue o resultado da Regressão Logística Ordinal:

Ordinal Logistic Regression: Cluster versus pops; intus; ...						
Link Function: Logit						
Response Information						
Variable	Value	Count				
Cluster	1	19				
	2	16				
	3	59				
	4	35				
	5	32				
	Total	161				
Logistic Regression Table						
Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI Lower
Const(1)	-115,154	21,8028	-5,28	0,000		
Const(2)	-60,8844	11,5275	-5,28	0,000		
Const(3)	-33,9970	6,96088	-4,88	0,000		
Const(4)	-28,2855	6,18550	-4,57	0,000		
pops	-0,476493	0,440922	-1,08	0,280	0,62	0,26
intus	-0,183154	0,329579	-0,56	0,578	0,83	0,44
mobs	0,132609	0,248027	0,53	0,593	1,14	0,70
ages	-0,0572898	0,540939	-0,11	0,916	0,94	0,33
ferts	-0,131975	0,548940	-0,24	0,810	0,88	0,30
gdppps	27,6390	5,23289	5,28	0,000	1,00806E+12	35419729,37
gdpgrs	0,218800	0,195103	1,12	0,262	1,24	0,85
lifes	-0,0258494	0,194858	-0,13	0,894	0,97	0,67
urbpops	0,214043	0,194957	1,10	0,272	1,24	0,85
Predictor	Upper					
Const(1)						
Const(2)						
Const(3)						
Const(4)						
pops	1,47					
intus	1,59					
mobs	1,86					
ages	2,73					
ferts	2,57					
gdppps	2,86898E+16					
gdpgrs	1,82					

lifes	1,43		
urbpops	1,82		
Log-Likelihood = -47,041			
Test that all slopes are zero: G = 389,688, DF = 9, P-Value = 0,000			
Goodness-of-Fit Tests			
Method	Chi-Square	DF	P
Pearson	3260,60	631	0,000
Deviance	94,08	631	1,000
Measures of Association:			
(Between the Response Variable and Predicted Probabilities)			
Pairs	Number	Percent	Summary Measures
Concordant	9616	98,3	Somers' D 0,98
Discordant	51	0,5	Goodman-Kruskal Gamma 0,99
Ties	120	1,2	Kendall's Tau-a 0,74
Total	9787	100,0	

Como se pode perceber pelos resultados, a regressão logística deu um percentual superior ao da Análise discriminante, de 98,3%, sendo, portanto a opção escolhida.

## 12. ANÁLISE DE CORRESPONDÊNCIA

A análise de correspondência é um método de análise fatorial para variáveis categóricas. Este método, basicamente, converte uma tabela de dados não negativos de duas ou múltiplas entradas em um tipo de representação gráfica em que as linhas e as colunas são simultaneamente representadas em dimensão reduzida, isto é, por pontos no gráfico. Este método permite mostrar como as variáveis dispostas em linhas e colunas estão relacionadas e não somente se a relação existe. A seguir, são apresentados os resultados da análise de correspondência para países e variáveis.

### 12.1 Análise de correspondência para 10 países

Uma tentativa foi realizada com 10 países escolhidos aleatoriamente, e as mesmas 9 variáveis. Seu resultado é exibido a seguir:

Simple Correspondence Analysis: pops; intus; mobs; ages; ferts; gdppps; gdpgrs;							
Analysis of Contingency Table							
Axis	Inertia	Proportion	Cumulative	Histogram			
1	0,1345	0,7888	0,7888	*****			
2	0,0144	0,0846	0,8734	***			
3	0,0104	0,0608	0,9343	**			
4	0,0052	0,0303	0,9646	*			
5	0,0043	0,0253	0,9899				
6	0,0013	0,0078	0,9977				
7	0,0004	0,0022	0,9998				
8	0,0000	0,0002	1,0000				
Total	0,1706						
Row Contributions							
ID	Name	Qual	Mass	Inert	Coord	Corr	Contr
1	Afghanistan	0,993	0,094	0,407	0,831	0,935	0,482
2	Austria	0,842	0,123	0,088	-0,319	0,831	0,093
3	Bhutan	0,585	0,085	0,039	0,188	0,451	0,022
4	Bolivia	0,620	0,100	0,041	0,207	0,619	0,032
5	Bosnia and Herzegovina	0,694	0,089	0,053	-0,260	0,671	0,045
6	Botswana	0,691	0,092	0,083	0,254	0,415	0,044
7	Cambodia	0,782	0,075	0,071	0,331	0,678	0,061
8	Canada	0,927	0,122	0,101	-0,319	0,717	0,092



9	Cyprus	0,908	0,108	0,051	-0,269	0,898	0,058			
10	Czech Republic	0,870	0,113	0,066	-0,292	0,848	0,071			
Component 2										
ID	Name	Coord	Corr	Contr						
1	Afghanistan	0,207	0,058	0,278						
2	Austria	0,036	0,011	0,011						
3	Bhutan	-0,103	0,134	0,062						
4	Bolivia	-0,008	0,001	0,000						
5	Bosnia and Herzegovina	-0,048	0,023	0,014						
6	Botswana	-0,207	0,276	0,273						
7	Cambodia	-0,130	0,104	0,088						
8	Canada	0,172	0,210	0,250						
9	Cyprus	0,029	0,010	0,006						
10	Czech Republic	-0,047	0,022	0,017						
Column Contributions										
ID	Name	Qual	Mass	Inert	Coord	Corr	Contr	Coord	Corr	Contr
1	pops	0,866	0,088	0,045	0,275	0,865	0,050	-0,008	0,001	0,000
2	intus	0,955	0,111	0,166	-0,467	0,853	0,180	0,161	0,101	0,199
3	mobs	0,631	0,126	0,052	-0,115	0,187	0,012	-0,178	0,444	0,275
4	ages	0,981	0,113	0,133	0,442	0,971	0,164	0,046	0,010	0,016
5	ferts	0,992	0,074	0,280	0,779	0,941	0,335	0,181	0,051	0,168
6	gdppps	0,714	0,057	0,080	-0,362	0,548	0,056	0,199	0,165	0,157
7	gdpgrs	0,847	0,134	0,065	0,227	0,623	0,051	-0,136	0,223	0,171
8	lifes	0,767	0,161	0,122	-0,315	0,766	0,118	-0,013	0,001	0,002
9	urbpops	0,499	0,136	0,055	-0,182	0,484	0,034	-0,033	0,016	0,010

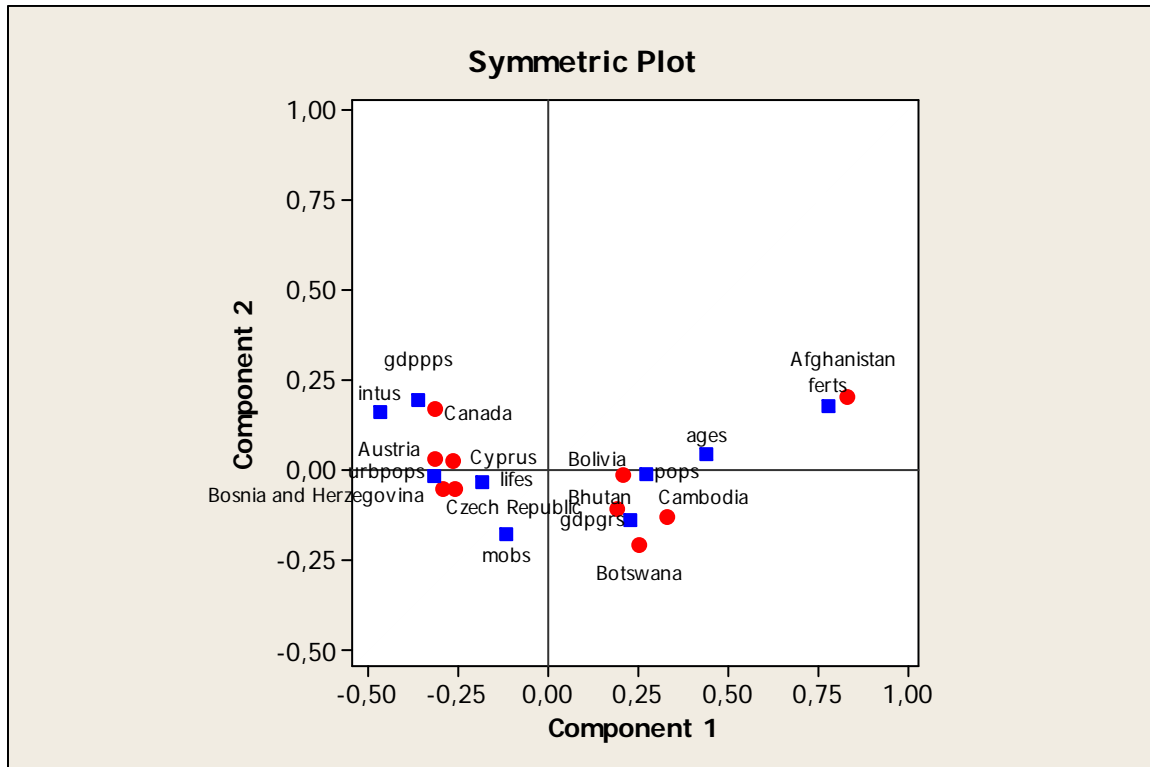
A análise da tabela de contingência mostra uma decomposição da inércia ( $\chi^2/n$ ). Do total da inércia da matriz de dados, 78,88% é contabilizada no primeiro componente, 8,46% é contabilizada no segundo componente e assim por diante.

Considerando a contribuição dos países (rows) para cada componente, identifica-se que o primeiro componente tem contribuição alta do Afeganistão, enquanto que Afeganistão, Botswana e Canadá contribuem mais para o componente 2.

Considerando a contribuição das variáveis (columns) para cada componente, identifica-se que o primeiro componente tem contribuição alta de Usuários de Internet, razão de dependência da idade e Taxa de Fertilidade, enquanto que Assinaturas de celular, PIB per capita e % de Crescimento do PIB per capita contribuem mais para o componente 2. As variáveis Usuários de Internet e Taxa de fertilidade aparecem nos dois componentes, mas de forma mais acentuada no componente 2 e 1 respectivamente.

Essas contribuições estão demonstradas nas tabelas anteriores com seus números pintados de amarelo.

O gráfico Symmetric plot gerado é exibido a seguir:



No Symmetric Plot observa-se:

1. Afeganistão tem a taxa de fertilidade mais alta do planeta, podendo ser considerado um outlier e é o país que mais contribui para a inércia do componente 1;
2. Botswana também pode ser considerado um outlier, contribuindo negativamente para a inércia do componente 2, junto com Afeganistão e Canadá, que tem contribuição é positiva;
3. Canadá está muito próximo das variáveis Usuários de Internet e PIB per capita;
4. A variável Assinaturas de Internet está sozinha, longe dos países ou outras variáveis;
5. Áustria está muito próxima da variável População urbana;
6. Não há países e variáveis muito próximos do zero;
7. A maioria dos países e indicadores estão muito próximos da linha do componente 1, estando distribuídos ao longo da mesma;
8. O componente 2 não tem variáveis nem países próximos de sua linha.

### **13. CONSIDERAÇÕES FINAIS**

Este trabalho teve como objetivo verificar se os indicadores que medem o acesso à tecnologia, como Usuários de Internet e Assinaturas de celular influenciam o comportamento de indicadores sociais e econômicos.

O resultado não foi muito animador, já que os países com maiores índices de renda são também os que tem maior acesso à Internet e à telefonia celular. São também aqueles que tem maior Expectativa de vida ao nascer.

Os países analisados correspondem à quase totalidade dos países do mundo, portanto, dão uma visão geral de como a situação é grave, já que não há uniformidade entre os mesmos.

Porem, algumas conclusões importantes puderam ser alcançadas:

- A grande importância do controle de natalidade para o desenvolvimento econômico das nações;
- A alta correlação entre tecnologia (Assinaturas de Celular e Usuários de Internet) e PIB não é capaz, por si só, de estabelecer uma relação causal, isto

é, é impossível, a priori, dizer se a difusão da tecnologia contribui para o crescimento econômico ou se, países mais ricos utilizam mais tecnologia por terem mais renda disponível;

- A análise das relações entre o crescimento econômico e as demais variáveis ficou prejudicada pela conjuntura pós-crise de 2008, em que, os países ricos vem apresentando baixo grau de crescimento e os países emergentes e mesmo os pobres produtores de commodities apresentam crescimento econômico elevado;
- A razão de dependência da idade ainda é muito maior nos países com alta taxa de fertilidade, indicando que os problemas da infância ainda são muito mais relevantes que os da velhice. Este quadro pode se inverter com o envelhecimento da população mundial;
- Já é esperado que países ricos cresçam, percentualmente, menos que países emergentes e pobres, ou seja, alta renda per capita não significa alto crescimento econômico futuro. Mas, com a crise de 2008, muitos países ricos sofreram queda de sua renda per capita, afetando suas políticas públicas (vide caso da Grécia) e, portanto, prejudicando a desempenho de algumas variáveis sociais aqui analisadas.

Análises de períodos mais longos podem trazer uma visão melhor do comportamento dos países ao longo do tempo.