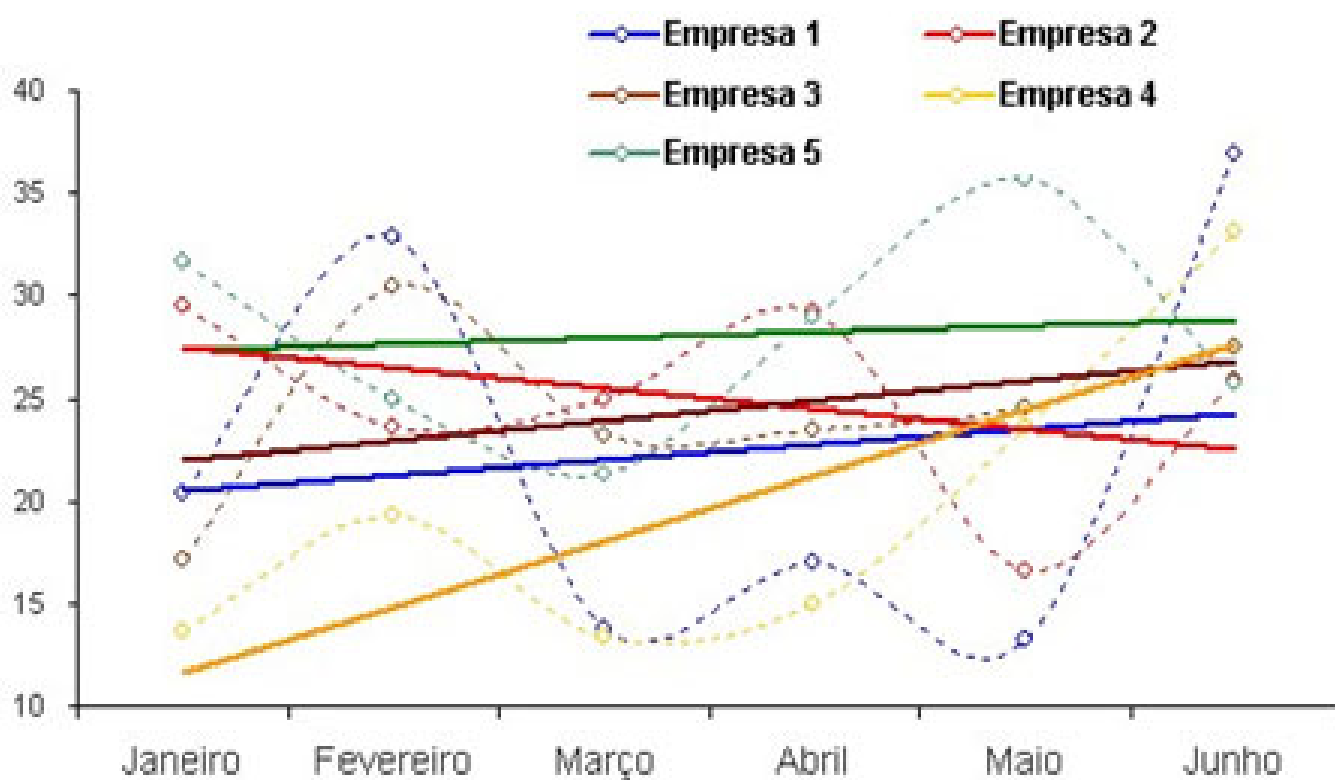




PUC - SP

ANÁLISE DISCRIMINANTE E INDICADORES DE GOVERANÇA DO BANCO MUNDIAL



GABRIELA AYER

SÃO PAULO
2010

1. APRESENTAÇÃO DOS DADOS

Este trabalho faz Análise Discriminante, com apuração apoiada pela aplicação do software estatístico MINITAB, dando continuidade aos estudos da pesquisa do WGI, que inicialmente eram 211 países analisados por 6 variáveis, que após tratamentos como Amostragem, Componentes Principais e Conglomerados obteve-se dois grupos avaliados por 3 principais das variáveis e que servirão de base neste trabalho.

1.1. AMOSTRA 60

60 países aleatórios aglomerados em 5 grupos não supervisionados, conforme tabela:

COUNTRY_60	CODE_60	VA_60	GE_60	CC_60	CLUSTER_60
AFGHANISTAN	AFG	-1,26415	-1,30946	-1,63753	1
ANGOLA	AGO	-1,06796	-0,98195	-1,22148	2
ARGENTINA	ARG	0,32106	-0,18052	-0,44245	2
ARUBA	ABW	1,00127	1,29395	1,3209	4
AZERBAIJAN	AZE	-1,23208	-0,63762	-1,00453	2
BANGLADESH	BGD	-0,60569	-0,76867	-1,10113	2
BELGIUM	BEL	1,37402	1,3642	1,34873	5
BOTSWANA	BWA	0,54504	0,66823	0,99788	4
BULGARIA	BGR	0,59626	0,0979	-0,17345	3
BURKINA FASO	BFA	-0,33061	-0,66702	-0,35933	2
CAMBODIA	KHM	-0,93649	-0,80506	-1,14097	2
CAMEROON	CMR	-1,01921	-0,79916	-0,90013	2
CANADA	CAN	1,42532	1,9331	2,02763	5
CAPE VERDE	CPV	0,94999	0,05215	0,75162	4
CHAD	TCD	-1,45179	-1,47766	-1,4536	1
CONGO	COG	-1,16366	-1,33778	-1,16126	2
Congo, Dem. Rep.	ZAR	-1,47529	-1,89313	-1,30702	1
COTE D'IVOIRE	CIV	-1,23618	-1,3879	-1,16981	1
CROATIA	HRV	0,48388	0,51569	0,12123	3
CUBA	CUB	-1,85204	-0,50713	-0,06435	2
CZECH REPUBLIC	CZE	1,02454	1,06811	0,36503	4
DOMINICA	DMA	1,0906	0,72337	0,66786	4
DOMINICAN REPUBLIC	DOM	0,1404	-0,39532	-0,61813	2
ECUADOR	ECU	-0,22151	-0,97383	-0,79279	2
FRANCE	FRA	1,24326	1,53673	1,43259	5
GAMBIA	GMB	-0,96985	-0,76695	-0,77775	2
GHANA	GHA	0,48366	-0,07547	-0,06265	3
GUATEMALA	GTM	-0,25562	-0,49183	-0,71876	2
GUYANA	GUY	0,16822	-0,17259	-0,47349	2
IRELAND	IRL	1,40438	1,61224	1,75506	5
ISRAEL	ISR	0,69447	1,29713	0,8741	4
KOSOVO	LWI	-0,64781	-0,81161	-0,66199	2
KUWAIT	KWT	-0,52908	0,11414	0,4975	3
LIBYA	LBY	-1,8962	-0,84004	-0,80576	2
LUXEMBOURG	LUX	1,50117	1,64596	2,02377	5
MACEDONIA	MKD	0,16193	-0,13556	-0,11039	3
MALAWI	MWI	-0,18463	-0,64837	-0,58593	2
MALDIVES	MDV	-0,3861	-0,35478	-0,60331	2

MARSHALL ISLANDS	MHL	1,19673	-1,21779	-0,54359	3
MAURITIUS	MUS	0,88321	0,60382	0,53177	4
MICRONESIA	FSM	1,02817	-0,62674	-0,43328	3
MOLDOVA	MDA	-0,26993	-0,75939	-0,64294	2
MYANMAR	MMR	-2,24333	-1,67576	-1,68505	1
NEW CALEDONIA	NCL	0,00516	0,00516	-0,92721	2
NIGER	NER	-0,41344	-0,79266	-0,8243	2
NORWAY	NOR	1,53057	1,9518	1,87667	5
PAKISTAN	PAK	-1,01064	-0,72784	-0,77343	2
PARAGUAY	PRY	-0,32678	-0,77664	-0,92534	2
PUERTO RICO	PRI	1,30547	0,65686	0,49889	4
RWANDA	RWA	-1,2354	-0,19888	0,02575	3
SAN MARINO	SMR	1,11752	-0,4179	1,18287	4
SINGAPORE	SGP	-0,40657	2,53125	2,34453	5
SLOVENIA	SVN	1,01978	1,09208	0,95179	4
SOLOMON ISLANDS	SLB	0,18748	-0,78937	-0,40873	2
ST. KITTS AND NEVIS	KNA	1,12111	0,65803	0,99956	4
SUDAN	SDN	-1,76582	-1,40892	-1,49021	1
SYRIA	SYR	-1,74677	-0,66828	-1,07246	2
TRINIDAD AND TOBAGO	TTO	0,53392	0,29837	-0,17033	3
TURKMENISTAN	TKM	-2,06436	-1,15859	-1,34476	2
YEMEN	YEM	-1,17617	-0,99346	-0,72608	2

1.2. AMOSTRA 30

30 países aleatórios aglomerados em 3 grupos, por classificação não supervisionada. Segue-se então, com a seguinte tabela:

COUNTRY_30	CODE_30	VA_30	GE_30	CC_30	CLUSTER_30
ANDORRA	ADO	1,33649	1,55613	1,3209	1
BANGLADESH	BGD	-0,60569	-0,76867	-1,10113	3
BRUNEI	BRN	-0,99979	0,89017	0,50657	2
BURKINA FASO	BFA	-0,33061	-0,66702	-0,35933	3
CAMEROON	CMR	-1,01921	-0,79916	-0,90013	3
CENTRAL AFRICAN REPUBLIC	CAF	-0,99764	-1,45008	-0,90223	3
COMOROS	COM	-0,43443	-1,87863	-0,74717	3
CROATIA	HRV	0,48388	0,51569	0,12123	2
GHANA	GHA	0,48366	-0,07547	-0,06265	2
IRAQ	IRQ	-1,25573	-1,40532	-1,48228	3
KIRIBATI	KIR	0,7142	-0,58022	0,00658	2
KOSOVO	LWI	-0,64781	-0,81161	-0,66199	3
LIBYA	LBY	-1,8962	-0,84004	-0,80576	3
LUXEMBOURG	LUX	1,50117	1,64596	2,02377	1
MARSHALL ISLANDS	MHL	1,19673	-1,21779	-0,54359	2
NAURU	NRU	1,03528	-0,41579	-0,2976	2
NIGER	NER	-0,41344	-0,79266	-0,8243	3
NIGERIA	NGA	-0,59651	-0,98447	-0,91918	3
NORWAY	NOR	1,53057	1,9518	1,87667	1

POLAND	POL	0,86073	0,47901	0,37577	2
QATAR	QAT	-0,77249	0,68093	1,2411	2
SAUDI ARABIA	SAU	-1,7356	0,00916	0,10827	2
SOLOMON ISLANDS	SLB	0,18748	-0,78937	-0,40873	2
SURINAME	SUR	0,56573	0,00343	-0,08659	2
SWITZERLAND	CHE	1,45323	2,06312	2,15464	1
SYRIA	SYR	-1,74677	-0,66828	-1,07246	3
TAIWAN	TWN	0,70244	0,87975	0,55043	2
TAJKISTAN	TJK	-1,31741	-0,88336	-0,989	3
THAILAND	THA	-0,55749	0,10977	-0,38436	2
TOGO	TGO	-1,1286	-1,4346	-0,98229	3

2. ANÁLISE DISCRIMINANTE

A Análise de Funções Discriminantes é utilizada para determinar quais variáveis discriminam entre dois ou mais grupos que ocorrem naturalmente para:

- Medir o poder de discriminação de cada variável ou grupo de variáveis;
- Descrever graficamente ou algebricamente diferentes grupos em termos de variáveis discriminadoras;
- Desenvolver regras para classificar novos elementos.

A ideia por trás da Análise de Discriminantes é a determinação do quanto os grupos diferem em relação as médias das variáveis e então usar estas variáveis para prever novos casos. Existem duas possibilidades para fazer a análise: Linear ou Quadráticas, apresentadas a seguir. Para ambas as análises se utiliza 3 das variáveis iniciais: Voice and Accountability, Government Effectiveness e Controlo f Corruption.

1.2. AMOSTRA 60

Discriminant Analysis: MODELO LINEAR					
Linear Method for Response: RECLUSTER_60					
Predictors: VA_60; GE_60; CC_60					
Group	1	2	3	4	5
Count	6	27	9	11	7
Summary of classification					
	True Group				
Put into Group	1	2	3	4	5
1	6	2	0	0	0
2	0	22	0	0	0
3	0	3	9	0	0
4	0	0	0	10	0
5	0	0	0	1	7
Total N	6	27	9	11	7
N correct	6	22	9	10	7
Proportion	1,000	0,815	1,000	0,909	1,000
N = 60	N Correct = 54		Proportion Correct = 0,900		

Squared Distance Between Groups					
	1	2	3	4	5
1	0,000	8,850	33,564	84,064	160,979
2	8,850	0,000	8,262	39,031	95,715
3	33,564	8,262	0,000	11,819	50,403
4	84,064	39,031	11,819	0,000	13,954
5	160,979	95,715	50,403	13,954	0,000

Linear Discriminant Function for Groups					
	1	2	3	4	5
Constant	-17,413	-4,402	-0,206	-5,415	-23,360
VA_60	-4,049	-1,808	0,816	2,541	2,912
GE_60	-5,782	-2,143	-0,719	1,923	6,392
CC_60	-13,477	-7,739	-0,703	8,423	17,422

Summary of Misclassified Observations					
Observation	True Group	Pred Group	Group	Squared Distance	Probability
3**	2	3	1	25,957	0,000
			2	4,886	0,159
			3	1,549	0,841
			4	19,889	0,000
			5	65,117	0,000
4**	4	5	1	118,605	0,000
			2	63,519	0,000
			3	27,598	0,000
			4	3,617	0,476
			5	3,427	0,524
16**	2	1	1	1,440	0,755
			2	3,695	0,245
			3	21,562	0,000
			4	64,258	0,000
			5	133,866	0,000
29**	2	3	1	24,124	0,000
			2	4,037	0,257
			3	1,916	0,743
			4	21,308	0,000
			5	67,137	0,000
54**	2	3	1	20,831	0,000
			2	4,403	0,334
			3	3,026	0,666
			4	24,429	0,000
			5	74,590	0,000
59**	2	1	1	1,576	0,972
			2	8,668	0,028
			3	33,335	0,000
			4	81,541	0,000
			5	153,517	0,000

A tabela nos apresenta que nos grupos 2 e 4 houveram reclassificação de cluster para alguns dados. Dos 27 indivíduos do grupo 2, vinte e dois permaneceram, dois foram remanejados para o grupo 1 (as linhas 16 e 59, que são os países Congo e Turcomenistan) e três foram para o grupo 3 (linhas 3, 29 e 54 que correspondem aos países Argentina, Guyana e Solomon Islands respectivamente). No grupo 4 ocorreu a reclassificação da Aruba (linha 4), que foi para o grupo 5.

Devido a todos estes reajustes, sendo que 18,5% do grupo 2 e 9,1% do grupo 4 sofreu reclassificação, o acerto da classificação inicial em relação a Análise Discriminante foi de 90%.

Discriminant Analysis: MODELO QUADRÁTICO

Quadratic Method for Response: RECLUSTER_60

Predictors: VA_60; GE_60; CC_60

Group	1	2	3	4	5
Count	6	27	9	11	7

Summary of classification

Put into Group	True Group				
	1	2	3	4	5
1	6	1	0	0	0
2	0	25	0	0	0
3	0	1	9	0	0
4	0	0	0	11	0
5	0	0	0	0	7
Total N	6	27	9	11	7
N correct	6	25	9	11	7
Proportion	1,000	0,926	1,000	1,000	1,000

N = 60

N Correct = 58

Proportion Correct = 0,967

From Generalized Squared Distance to Group

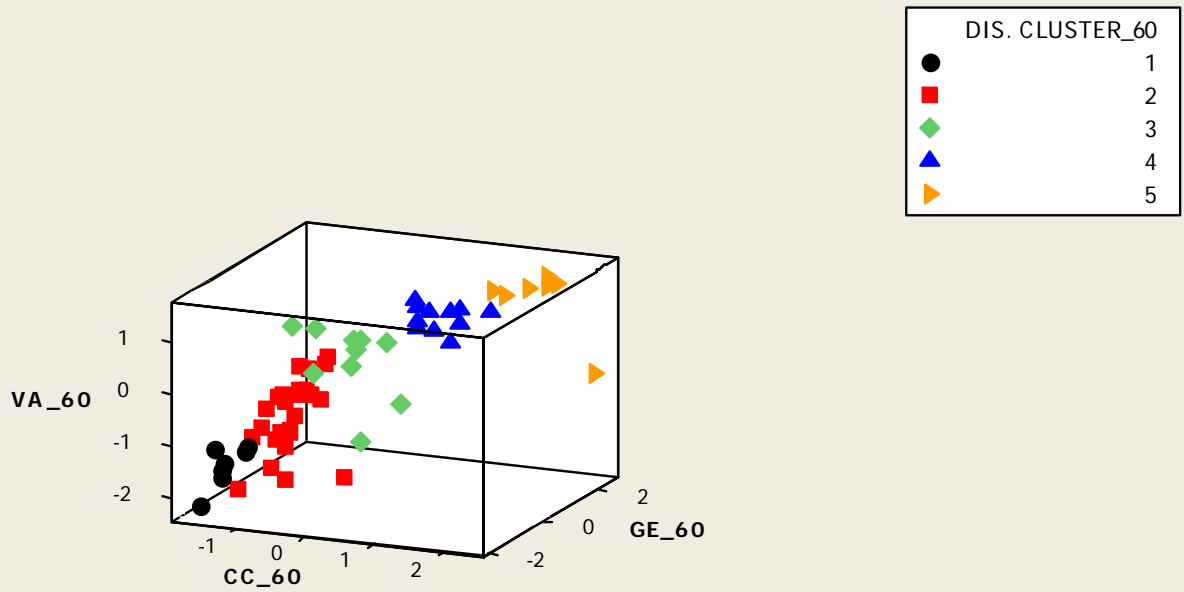
Group	1	2	3	4	5
1	-9,049	2,754	88,587	316,430	384,241
2	26,596	-6,305	20,597	137,106	196,818
3	103,897	-0,131	-5,709	26,024	90,367
4	325,732	29,225	26,610	-6,964	18,239
5	810,934	102,707	109,867	14,589	-7,313

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
16**	2	1	1	-5,513	0,896
			2	-1,203	0,104
			3	50,634	0,000
			4	230,684	0,000
			5	325,107	0,000
54**	2	3	1	35,949	0,000
			2	-1,363	0,321
			3	-2,863	0,679
			4	51,865	0,000
			5	169,066	0,000

Utilizando o modelo quadrático, o acerto da classificação é de 96,7%, apenas Congo e Turkmenistan são remanejados da mesma forma que ocorreu no modelo linear. Assim, defini-se que tais países devem passar a pertencer aos grupos indicados, o que gera um novo gráfico de posicionamento.

3D Scatterplot of VA_60 vs GE_60 vs CC_60



Visualmente, alguns países parecem estar bem próximos dos grupos subsequentes, mas através da análise de discriminantes tem-se a segurança de que a classificação está precisa.

2.2. AMOSTRA 30

Discriminant Analysis: MÉTODO LINEAR
 Linear Method for Response: RECLUSTER

Predictors: VA_30; GE_30; CC_30

Group	1	2	3
Count	4	13	13

Summary of classification

Put into Group	True Group		
	1	2	3
1	4	0	0
2	0	12	0
3	0	1	13
Total N	4	13	13
N correct	4	12	13
Proportion	1,000	0,923	1,000

N = 30 N Correct = 29 **Proportion Correct = 0,967**

Squared Distance Between Groups

	1	2	3
1	0,0000	29,5172	79,0926
2	29,5172	0,0000	12,1724
3	79,0926	12,1724	0,0000

Linear Discriminant Function for Groups

	1	2	3
Constant	-16,551	-0,065	-5,019
VA_30	4,949	0,410	-3,161
GE_30	3,907	-0,044	-3,244
CC_30	10,223	0,726	-4,079

Summary of Misclassified Observations

Observation	True Group	Pred Group	Group	Squared Distance	Probability
22**	2	3	1	55,326	0,000
			2	8,729	0,333
			3	7,342	0,667

Observa-se que a análise Linear identificou apenas um elemento (linha 22: Saudi Arabia) que, originalmente no segundo grupo, deveria ser remanejado para o terceiro grupo, permanecendo o grupo 2 com 12 elementos. Sendo assim, para os grupos 1 e 3 o acerto foi de 100%, já para o grupo 2 a proporção foi de 0,923 acertos. Isso implica num acerto geral bastante alto (de 96,7%).

Discriminant Analysis: MODELO QUADRÁTICO

Quadratic Method for Response: RECLUSTER

Predictors: VA_30; GE_30; CC_30

Group	1	2	3
Count	4	13	13

Summary of classification

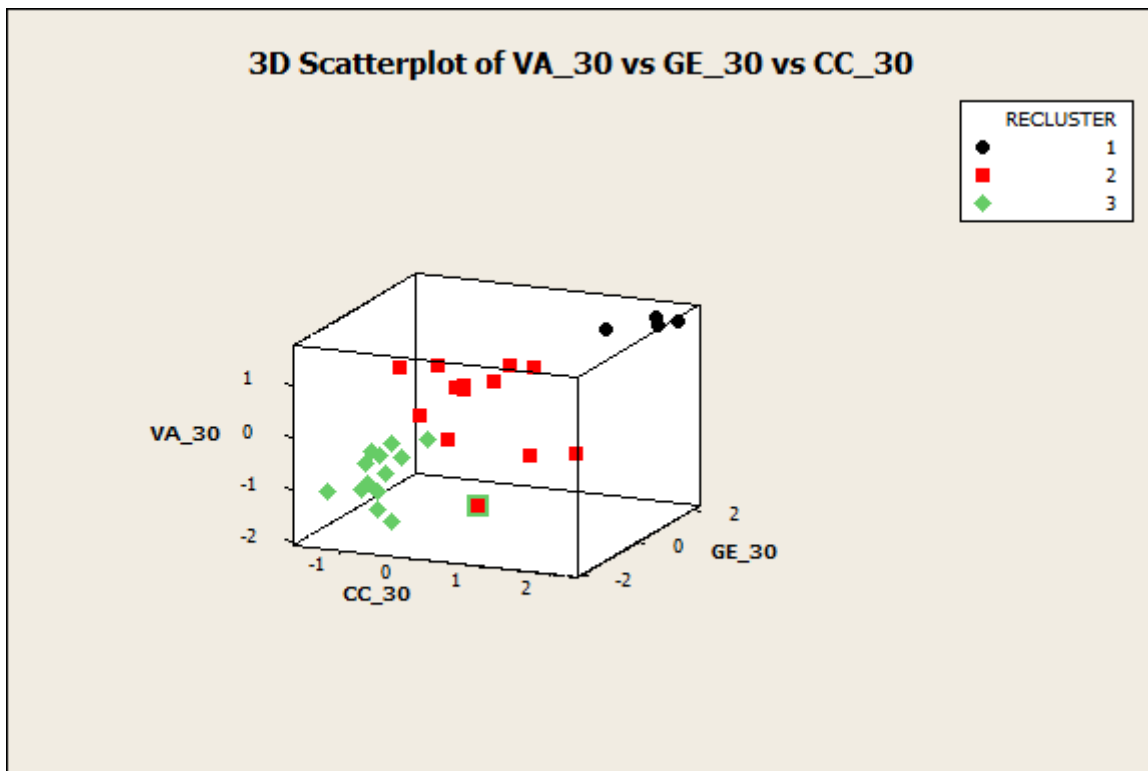
Put into Group	True Group		
	1	2	3
1	4	0	0
2	0	13	0
3	0	0	13
Total N	4	13	13
N correct	4	13	13
Proportion	1,000	1,000	1,000

N = 30 N Correct = 30 Proportion Correct = 1,000

From Generalized Squared Distance to Group

Group	1	2	3
1	-11,34	18,26	132,73
2	364,22	-3,66	13,09
3	1370,27	5,28	-6,34

Já ao fazer a análise pelo modelo Quadrático, todos os elementos são classificados no mesmo grupo em que estavam originalmente, tendo portanto um acerto de 100%. Desta forma, é possível visualizar graficamente a composição dos grupos em relação às três variáveis utilizadas para gerar a classificação.



O gráfico demonstra que o grupo 1, em preto, possui elevada avaliação nas três variáveis, localizando-se ao topo do gráfico. Já o grupo 2, vermelho, possui valores intermediários nas variáveis e ocupa a centralidade do gráfico. O grupo 3, em verde, fica na porção inferior do gráfico, por ter baixos valores nas variáveis. Saudi Arabia, que estaria reconsiderada no grupo 3, pelo modelo linear, aparece em destaque, com um contorno da cor deste grupo.

3. COMENTÁRIOS FINAIS

A Análise Discriminante serviu para confrontar a classificação estruturada com a classificação não supervisionada, e por ter apresentado grande número de acertos, válida esta, com poucos ajuste na amostra 60.

De acordo com todas as análises acima, percebemos que não há uma diferença muito grande entre os países dos diferentes grupos encontrados quanto aos assuntos responsabilidade social, efetividade de governo e controle de corrupção, conjuntamente.