PONTÍFICIA UNIVERSIDADE CATÓLICA DE SÃO PAULO

Faculdade de Economia, Administração, Contabilidade e Atuariais

UM ESTUDO DAS RELAÇÕES ENTRE INDICADORES DE DESENVOVLIMENTO HUMANO-HDI E INDICADORES DE INOVAÇÃO-GII

Aluno: Victor Werner Degenhardt

Prof. Arnoldo José de Hoyos Guevara

1° Semestre 2012

1. Introdução

O presente trabalho tem por objetivo comparar três técnicas de análise multivariada de países selecionados com dados relativos a indicadores de desenvolvimento humano direcionados à inovação.

2. Coleta de dados

Os dados foram colhidos nos sites abaixo:

International Human Development Indicators:
http://hdrstats.undp.org/en/tables

The Global Innovation Index da INSEAD:
http://www.globalinnovationindex.org/gii/

3. Variáveis

Usou-se as variáveis abaixo:

Variável	Significado	Variável calculada pelo	Tipo
GII	**Global Innovation Index**	Global Innovation Index da INSEAD, 2011.Score que mede a inovação resultante no país.	**Quantitativa**
IEI	**Innovation Efficiency Index**	Global Innovation Index da INSEAD, 2011.Posição no ranking, que mede a eficiência no uso das entradas.	**Quantitativa**
HDI	**Human Development Index**	HDRO- Human Development Report, 2011, United Nations	**Quantitativa**
EDI	**Education index**	HDRO, 2011, UN	**Quantitativa**
GDP	**GDP per capita (2005 PPP $)**	World Bank, 2009	**Quantitativa**
YSC	**Mean years of schooling (of adults)**	HDRO/ UNESCO, 2011	**Quantitativa**
% EX	**Public expenditure on education (% of GDP)**	World Bank, 2006- 2009	**Quantitativa**

Tomou-se dados de:
1-Argentina,
2-Brasil,
3-Coréia,
4-Canadá,
5-Chile,
6- China,
7-Colômbia,
8-Costa Rica,
9- Dinamarca,
10-Equador,
11-El Salvador,
12-Finlândia,

13-Alemanha,
14-Guatemala,
15-Honduras,
16-Hungria,
17-Índia,
18- Indonésia,
19-Irlanda,
20-Itália,
21- México,
22-Nicarágua,
23-Paquistão,
24-Paraguai,
25-Polônia,
26-Portugal,
27-Bélgica,
28-África do Sul,
29-Turquia,
30-Uruguai,
31- Rússia.

4. Análise das variáveis

4.1 Análise de discriminante

Iniciamos o exercício 9 com 5 clusters e reduzimos para 3 clusters, colocando África do Sul no cluster dos países pobres e Dinamarca nos ricos tendo-se uma proporção de acertos de 96,8%.

Discriminant Analysis: 3 Cluster versus 5 Cluster

```
Linear Method for Response: 3 Cluster


Predictors: 5 Cluster


Group          1         2         4
Count         20         8         3


Summary of classification

                      True Group
Put into Group        1         2         4
1                    19         0         0
2                     0         8         0
4                     1         0         3
Total N              20         8         3
N correct            19         8         3
Proportion       0,950     1,000     1,000

N = 31           N Correct = 30           Proportion Correct = 0,968


Squared Distance Between Groups
```

```
          1       2        4
1   0,0000  1,4904  13,6560
2   1,4904  0,0000   6,1236
4  13,6560  6,1236   0,0000


Linear Discriminant Function for Groups

                1       2        4
Constant   -1,254  -3,933  -13,935
5 Cluster   2,090   3,701    6,967


Summary of Misclassified Observations

                True   Pred          Squared
Observation    Group  Group  Group  Distance  Probability
         28**      1      4      1    25,152        0,000
                                  2    14,397        0,002
                                  4     1,742        0,998
```

Terminamos o Exercício 9 encaixando a África do Sul no cluster 4 da Índia, Paquistão e Guatemala resultando:

Discriminant Analysis: New 3 cluster versus 5 Cluster

```
Linear Method for Response: New 3 cluster


Predictors: 5 Cluster


Group         1         2         4
Count        19         8         4


Summary of classification

                     True Group
Put into Group       1      2      4
1                   19      0      0
2                    0      8      0
4                    0      0      4
Total N             19      8      4
N correct           19      8      4
Proportion       1,000  1,000  1,000

N = 31            N Correct = 31           Proportion Correct = 1,000


Squared Distance Between Groups

          1       2        4
1    0,000  21,808  182,000
2   21,808   0,000   77,808
4  182,000  77,808    0,000


Linear Discriminant Function for Groups

               1       2        4
Constant   -8,62  -38,90  -155,62
5 Cluster  17,23   36,62    73,23
```

Neste caso a proporção correta subiu para 100% mostrando que o encaixe foi perfeito.

Agora faremos a regressão logística para comparação.

4.2 Regressão logística

Juntando a África do Sul nos países do cluster 4 da Índia, Paquistão e Guatemala que tinha resultado em acerto pela análise de discriminante de 100% a regressão logística é:

Ordinal Logistic Regression: New 3 cluster versus EDI; GDP; ...

```
* WARNING * Algorithm has not converged after 20 iterations.
* WARNING * Convergence has not been reached for the parameter estimates
            criterion.
* WARNING * The results may not be reliable.
* WARNING * Try increasing the maximum number of iterations.


Link Function: Logit


Response Information

Variable       Value  Count
New 3 cluster  1         19
               2          8
               4          4
               Total     31


Logistic Regression Table

                                                                  95% CI
Predictor       Coef  SE Coef      Z      P   Odds Ratio  Lower         Upper
Const(1)    -7386,87   108494  -0,07  0,946
Const(2)    -6886,52   101528  -0,07  0,946
EDI         -5068,29   117122  -0,04  0,965         0,00   0,00             *
GDP        -0,103625  1,53548  -0,07  0,946         0,90   0,04         18,28
GII         -2,96610  338,333  -0,01  0,993         0,05   0,00  5,03162E+286
YSC          71,3720  3938,65   0,02  0,986  9,91877E+30   0,00             *
IEI          10,2607  149,582   0,07  0,945     28586,54   0,00  6,02828E+131
New%EX       47,5431  1020,22   0,05  0,963  4,44338E+20   0,00             *
HDI          15348,4   243183   0,06  0,950            *   0,00             *


Log-Likelihood = -0,000
Test that all slopes are zero: G = 56,657, DF = 7, P-Value = 0,000


Goodness-of-Fit Tests

Method    Chi-Square  DF      P
Pearson    0,0000097  53  1,000
Deviance   0,0000194  53  1,000


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

Pairs       Number  Percent  Summary Measures
```

Concordant	260	100,0	Somers' D	1,00
Discordant	0	0,0	Goodman-Kruskal Gamma	1,00
Ties	0	0,0	Kendall's Tau-a	0,56
Total	260	100,0		

Pela regressão logística o acerto também é de 100%, porém há advertências que o algoritmo empregado não convergiu e o resultado pode não ser crível. Isso também pode ser visto pelos altos p, sinal de problemas.

Method	Chi-Square	DF	P
Pearson	11,8936	54	1,000
Deviance	9,5163	54	1,000

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

Pairs	Number	Percent	Summary Measures	
Concordant	197	99,0	Somers' D	0,98
Discordant	2	1,0	Goodman-Kruskal Gamma	0,98
Ties	0	0,0	Kendall's Tau-a	0,42
Total	199	100,0		

A concordância foi de 99,0%.

4.3 Árvore de decisão

Lembrando que usaremos como clusters:

Cluster 1: 19 países: pobres
Cluster 2: 8 países: ricos
Cluster 4: 4 países: Índia, Paquistão, Guatemala, África do Sul

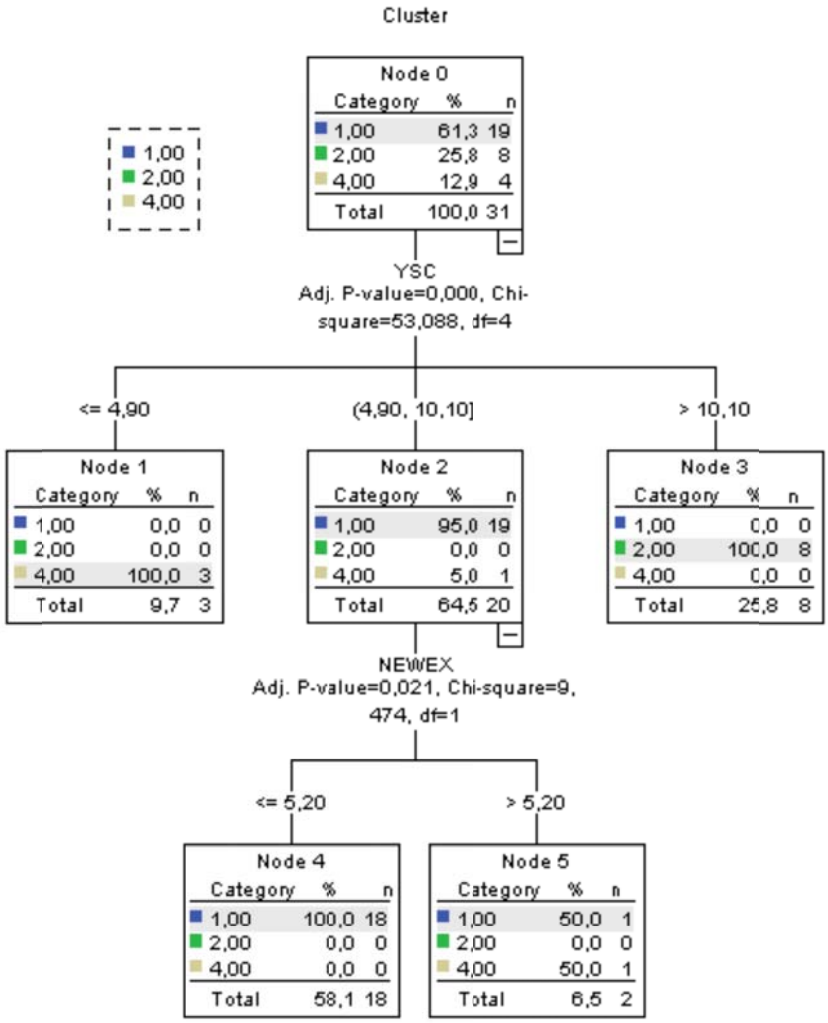
```
* Decision Tree.

TREE Dcluster [n] BY EDI [o] GDP [o] GII [o] YSC [o] IEE [o] NEWEX [o] HDI
[o]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES
SCALE=AUTO
  /DEPCATEGORIES USEVALUES=[1.00 2.00 4.00]
  /PRINT MODELSUMMARY CLASSIFICATION RISK
  /METHOD TYPE=CHAID
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=4 MINCHILDSIZE=2
  /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
  /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO CHISQUARE=PEARSON
CONVERGE=0.001 MAXITERATIONS=100 ADJUST=BONFERRONI
  /COSTS EQUAL.
```

Classification Tree

Model Summary

Specifications	Growing Method	CHAID
	Dependent Variable	Cluster
	Independent Variables	EDI, GDP, GII, YSC, IEE, NEWEX, HDI
	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	4
	Minimum Cases in Child Node	2
Results	Independent Variables Included	YSC, NEWEX
	Number of Nodes	6
	Number of Terminal Nodes	4
	Depth	2



Model Summary

Specifications	Growing Method	CHAID
	Dependent Variable	Cluster
	Independent Variables	EDI, GDP, GII, YSC, IEE, NEWEX, HDI
	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	4
	Minimum Cases in Child Node	2
Results	Independent Variables Included	YSC, NEWEX
	Number of Nodes	6

	Number of Terminal Nodes	4
	Depth	2

Risk

Estimate	Std. Error
,032	,032

Growing Method: CHAID

Dependent Variable: Cluster

Classification

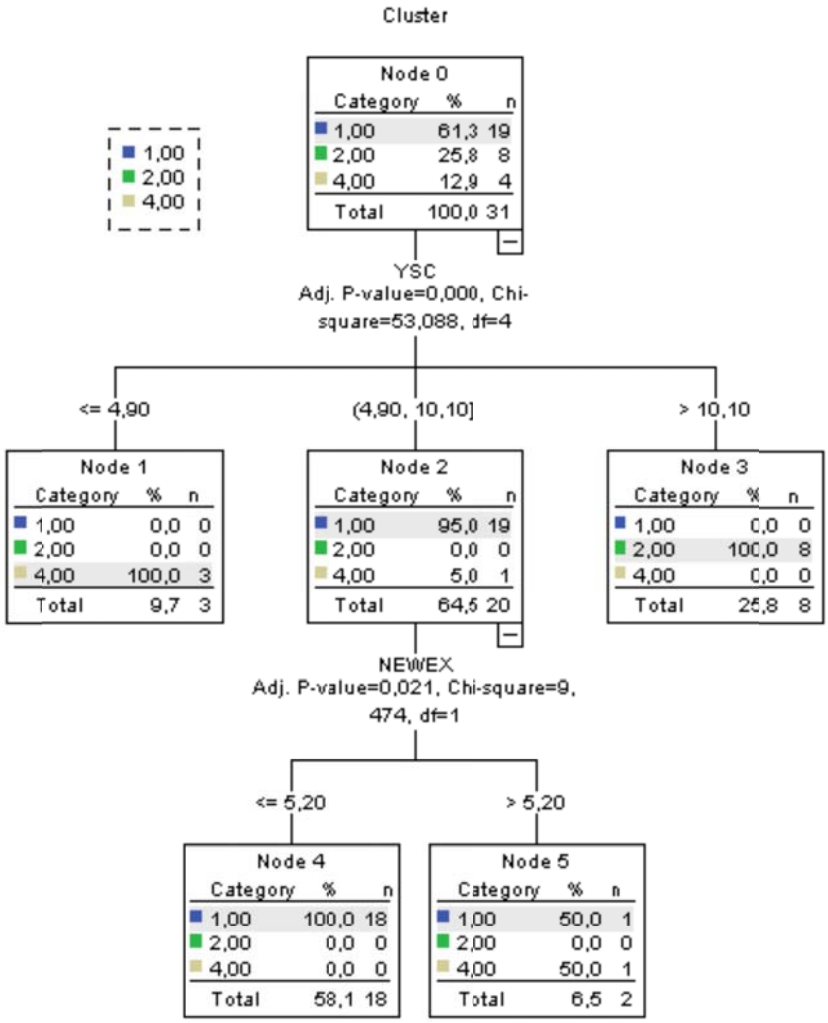
Observed	Predicted			
	1,00	2,00	4,00	Percent Correct
1,00	19	0	0	100,0%
2,00	0	8	0	100,0%
4,00	1	0	3	75,0%
Overall Percentage	64,5%	25,8%	9,7%	96,8%

Growing Method: CHAID

Dependent Variable: Cluster

O acerto foi de 96,8% havendo um país classificado como 4 que deveria ser classificado como 1.Refaremos classificando a África do Sul como país pobre.

```
* Decision Tree.
TREE Dcluster [n] BY EDI [o] GDP [o] GII [o] YSC [o] IEE [o] NEWEX [o] HDI
[o]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES
SCALE=AUTO
  /DEPCATEGORIES USEVALUES=[1.00 2.00 4.00]
  /PRINT MODELSUMMARY CLASSIFICATION RISK
  /METHOD TYPE=CHAID
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=4 MINCHILDSIZE=2
  /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
  /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO CHISQUARE=PEARSON
CONVERGE=0.001 MAXITERATIONS=100 ADJUST=BONFERRONI
  /COSTS EQUAL.
```



Classification Tree

Model Summary

Specifications	Growing Method	CHAID
	Dependent Variable	Cluster
	Independent Variables	EDI, GDP, GII, YSC, IEE, NEWEX, HDI
	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	4
	Minimum Cases in Child Node	2

Results	Independent Variables Included	YSC, NEWEX
	Number of Nodes	6
	Number of Terminal Nodes	4
	Depth	2

Risk

Estimate	Std. Error
,032	,032

Growing Method: CHAID

Dependent Variable: Cluster

Classification

Observed	Predicted			
	1,00	2,00	4,00	Percent Correct
1,00	19	0	0	100,0%
2,00	0	8	0	100,0%
4,00	1	0	3	75,0%
Overall Percentage	64,5%	25,8%	9,7%	96,8%

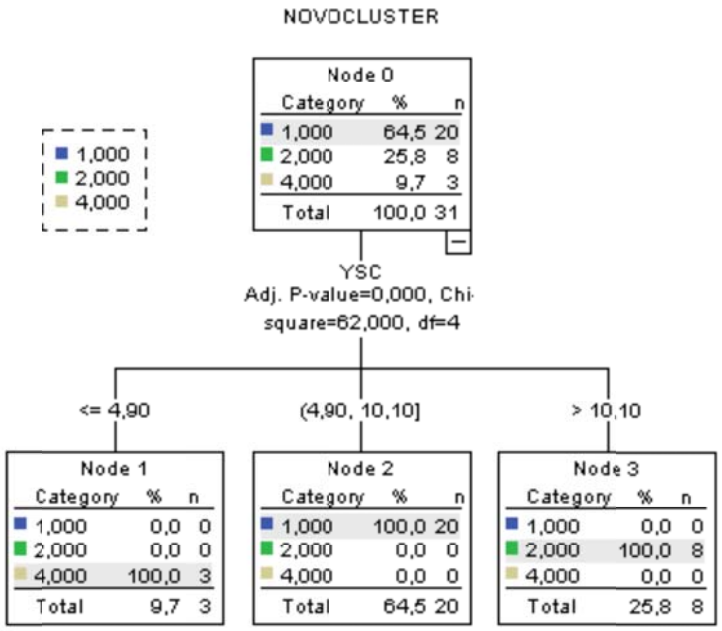
Growing Method: CHAID

Dependent Variable: Cluster

```
* Define Variable Properties.
*Novocluster.
VARIABLE LEVEL  Novocluster(NOMINAL).
VARIABLE LABELS  Novocluster 'NOVOCLUSTER'.
EXECUTE.
SAVE OUTFILE='C:\Users\Gateway\Documents\PUC\2012 aulas\Estatistica
Hoyos\Exercício 12 '+
    'SPSS\Arvore.sav'
 /COMPRESSED.
* Decision Tree.
TREE Novocluster [n] BY EDI [o] GDP [o] GII [o] YSC [o] IEE [o] NEWEX [o]
HDI [o]
  /TREE DISPLAY=TOPDOWN NODES=STATISTICS BRANCHSTATISTICS=YES NODEDEFS=YES
SCALE=AUTO
  /DEPCATEGORIES USEVALUES=[VALID]
  /PRINT MODELSUMMARY CLASSIFICATION RISK TREETABLE
  /METHOD TYPE=CHAID
  /GROWTHLIMIT MAXDEPTH=AUTO MINPARENTSIZE=4 MINCHILDSIZE=2
  /VALIDATION TYPE=NONE OUTPUT=BOTHSAMPLES
  /CHAID ALPHASPLIT=0.05 ALPHAMERGE=0.05 SPLITMERGED=NO CHISQUARE=PEARSON
CONVERGE=0.001 MAXITERATIONS=100 ADJUST=BONFERRONI.
```

Model Summary

Specifications	Growing Method	CHAID
	Dependent Variable	NOVOCLUSTER
	Independent Variables	EDI, GDP, GII, YSC, IEE, NEWEX, HDI
	Validation	None
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	4
	Minimum Cases in Child Node	2
Results	Independent Variables Included	YSC
	Number of Nodes	4
	Number of Terminal Nodes	3
	Depth	1



Classification

Observed	Predicted

	1,00	2,00	4,00	Percent Correct
1,00	20	0	0	100,0%
2,00	0	8	0	100,0%
4,00	0	0	3	100,0%
Overall Percentage	64,5%	25,8%	9,7%	100,0%

Growing Method: CHAID

Dependent Variable: NOVOCLUSTER

Agora temos 100% de acerto, mostrando haver diferenças entre as técnicas de análise.

5. Referências bibliográficas

LAS CASAS, A. L.; GUEVARA, A.J.H. Pesquisa de Marketing. São Paulo: Atlas, 2010.

MOORE, D.S.; McCABE, G.P. Introduction to the practice of Statistics. 2 ed, New York: Freeman, 1993.

BARTHOLOMEW, D. et al. The analysis and interpretation of multivariate data for social scientists. Florida- USA: Chapman & Hall, 2002.