

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
FEA - FACULDADE DE ECONOMIA E ADMINISTRAÇÃO
PROGRAMA DE ESTUDOS PÓS-GRADUADOS EM
ADMINISTRAÇÃO**

CLASSIFICAÇÃO DO BRASIL

**Focando principalmente indicadores relacionados a habitação,
educação, trabalho e muito particularmente EMPREGO e RENDA**

MÉTODOS QUANTITATIVOS DA PESQUISA EMPÍRICA.

Professor: Dr. Arnaldo José de Hoyos

José Felipe Ferreira de Souza

CAP I ANÁLISE DE CONGLOMERADOS

1 – INTRODUÇÃO

O presente trabalho propõe efetuar uma análise de conglomerados dos indicadores ISDM, Educação e Emprego e Renda com base nos dados Pesquisa Firjan/FGV sobre o Desenvolvimento dos Municípios nos períodos de 2000 e 2010.

O software estatístico a ser utilizado é o MINITAB16.

2 – OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são os 5565 Municípios brasileiros cujas prefeituras disponibilizaram os dados referentes ao ISDM, IFDM, IFGF, Emprego e Renda, Educação e Saúde.

Fonte: Tabela ISDM, IFDM e IFGF por municípios brasileiros; e <http://www.firjan.org.br>

2.2 As Variáveis

Para o objetivo do presente trabalho, foram trabalhados os dados de 7 variáveis: IFDM, ISDM, IFGF, Emprego e Renda, Educação e Saúde.

Variável	Significado	Tipo	Unidade de Medida
ISDM	Índice Social de Desenvolvimento Municipal. Consiste na média ponderada dos indicadores das dimensões: Habitação, Renda, Trabalho, Saúde e Segurança e Educação (H, R, T, S e E) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Renda (R)	Média ponderada dos indicadores da dimensão Renda (R1 e R2) padronizada pela média do	Variável Quantitativa	Numérico

	Brasil.		
Emprego/Trabalho (T)	Média ponderada dos indicadores da dimensão Trabalho (T1_1, T1_2 e T2_1) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Educação (E)	Média ponderada dos indicadores da dimensão Educação (E1_1, E1_2, E2_1, E2_2, E2_3, E2_4, E2_5, E2_6, E3_1, E3_2 e E3_3) padronizada pela média do Brasil.	Variável Quantitativa	Numérico

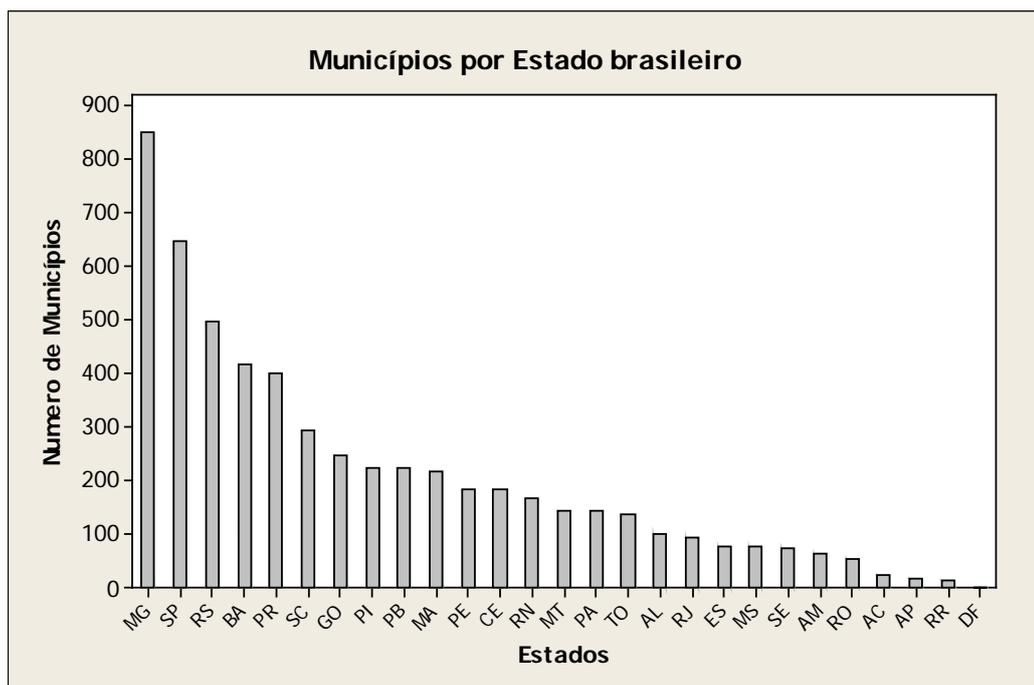
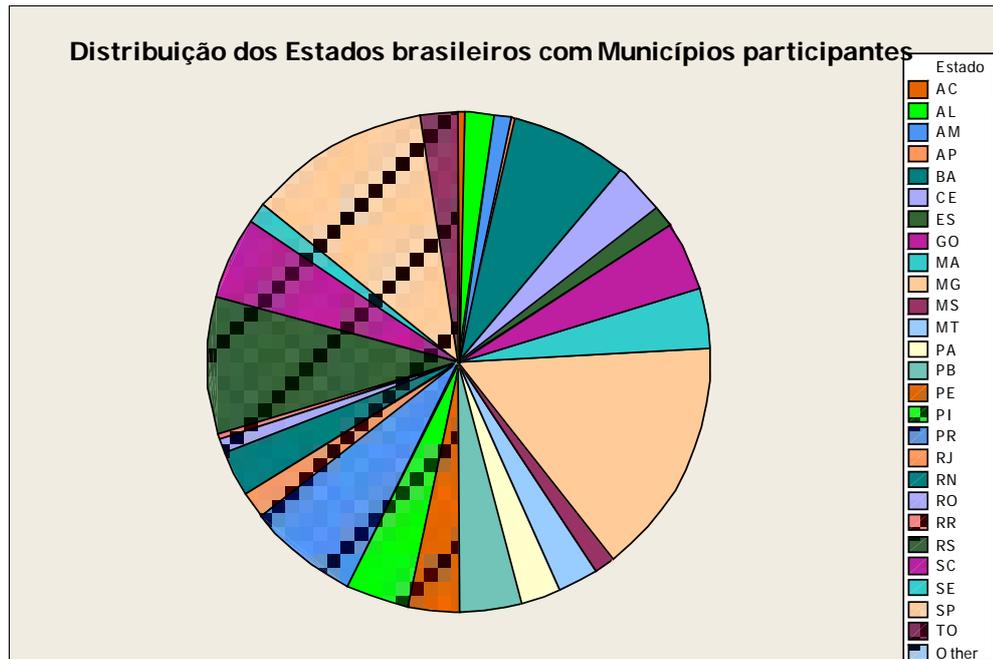
3. ANÁLISE DAS VARIÁVEIS

3.1 Variáveis Categóricas

Este tipo de variável indica que o foco de concentração deve ser a análise de gráficos do tipo *pie chart* e barras.

3.1.1 Variável: “Estados”

Fazem parte desta pesquisa os 27 Estados brasileiros e seus municípios. Nos gráficos abaixo é possível visualizar a distribuição de municípios por Estado brasileiro.



Os dados dos gráficos nos permite visualizar que, os Estados com maior participação de municípios são: Minas Gerais (852 cidades), São Paulo (647 cidades), Rio Grande do Sul (497 cidades), Bahia (416 cidades) e Paraná (399 cidades).

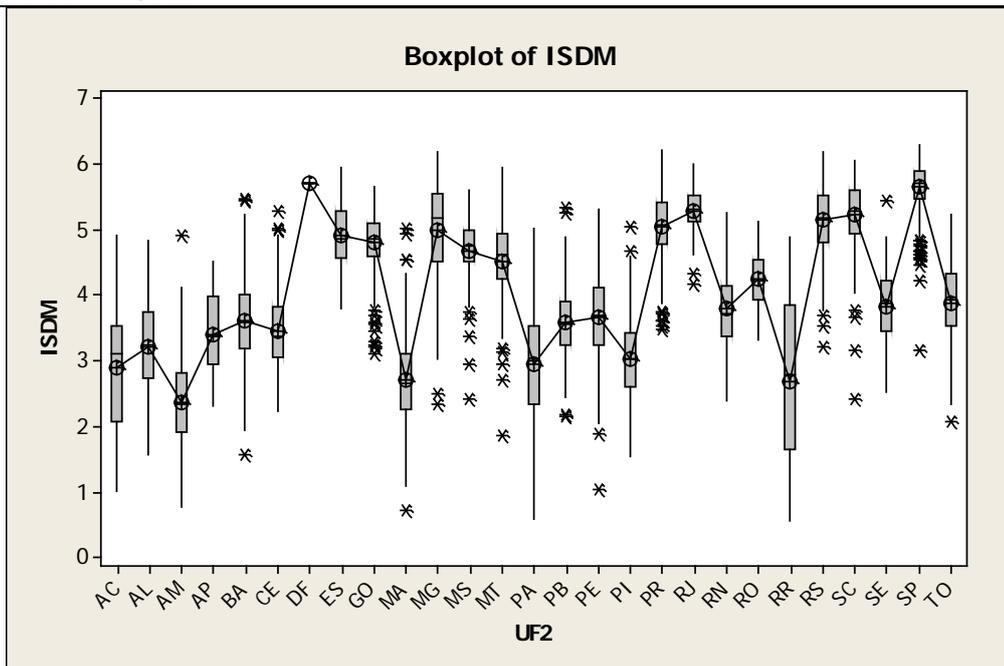
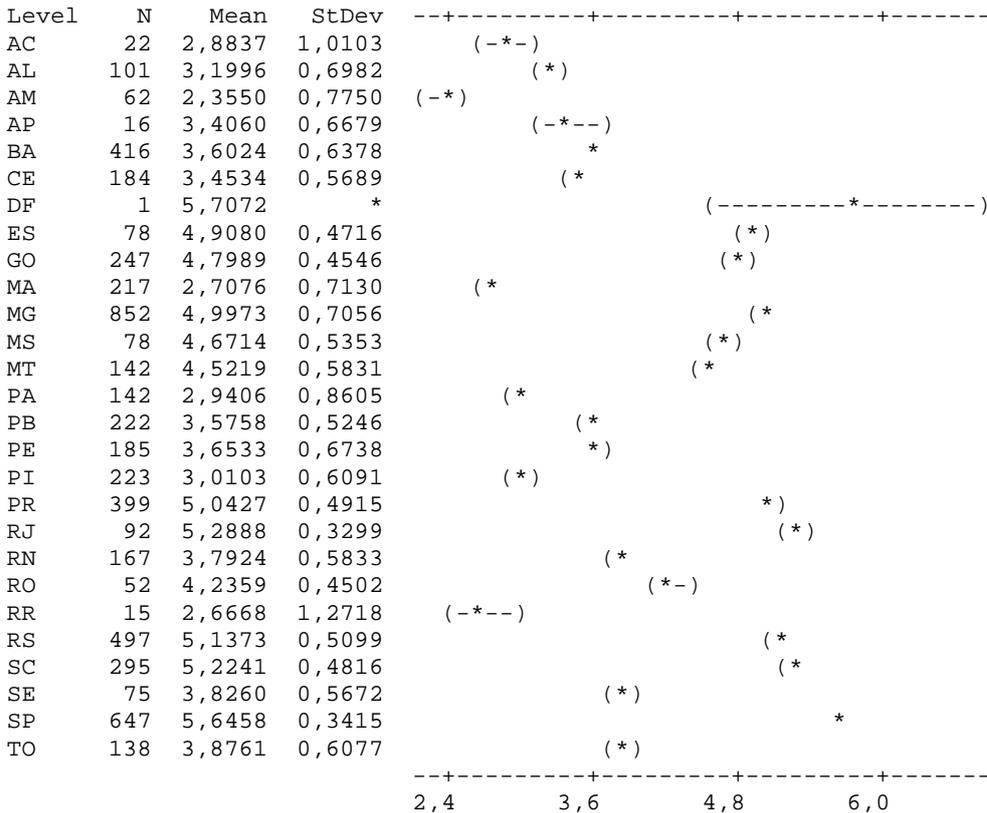
Já entre os Estados com menor número de cidades participantes, temos: Distrito Federal (1 cidade), Roraima (15 cidades), Amapá (16 cidades) e Acre (22 cidades).

One-way ANOVA: ISDM versus UF2

Source	DF	SS	MS	F	P
UF2	26	4759,395	183,054	537,35	0,000
Error	5538	1886,584	0,341		
Total	5564	6645,979			

S = 0,5837 R-Sq = 71,61% R-Sq(adj) = 71,48%

Individual 95% CIs For Mean Based on Pooled StDev



4.2. EDUCAÇÃO

One-way ANOVA: Educação versus Região

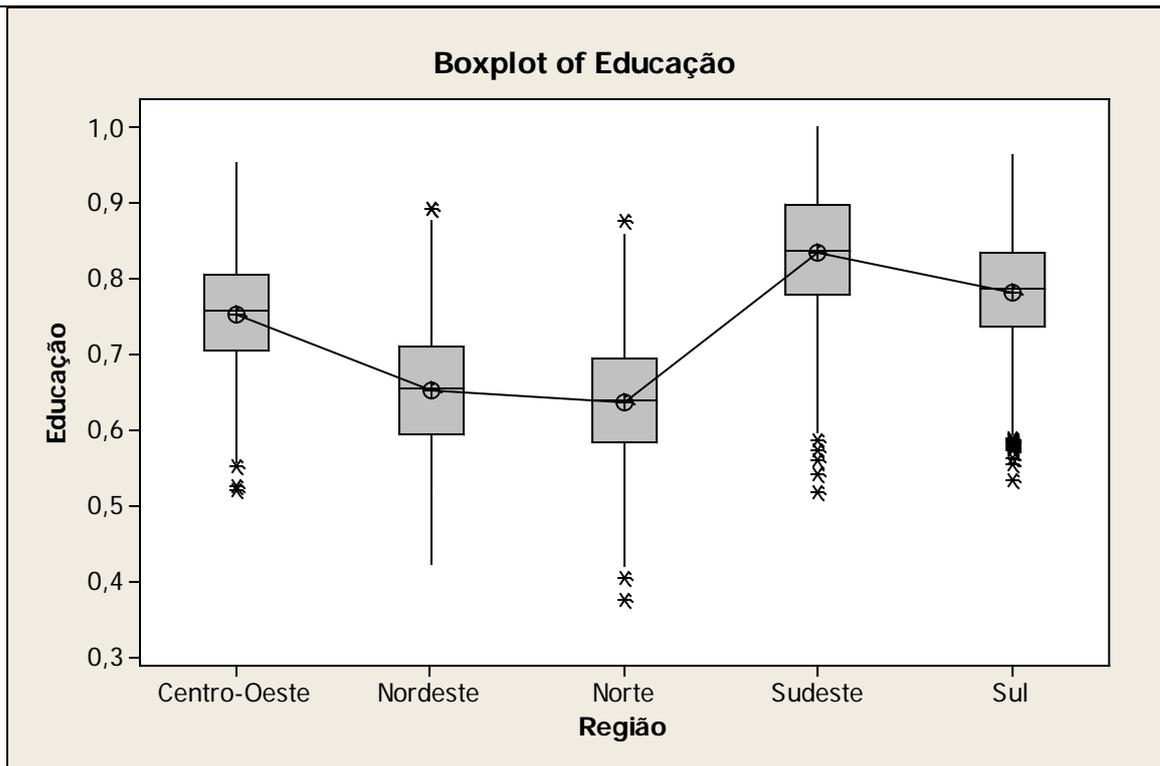
Source	DF	SS	MS	F	P
Região	4	35,82737	8,95684	1446,69	0,000
Error	5560	34,42333	0,00619		
Total	5564	70,25070			

S = 0,07868 R-Sq = 51,00% R-Sq(adj) = 50,96%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
Centro-Oeste	468	0,75299	0,07162	(*-)
Nordeste	1790	0,65090	0,08127	(*)
Norte	447	0,63739	0,08599	(*)
Sudeste	1669	0,83427	0,08138	(*)
Sul	1191	0,78106	0,07021	(*)

-----+-----+-----+-----+-----
0,660 0,720 0,780 0,840

Pooled StDev = 0,07868

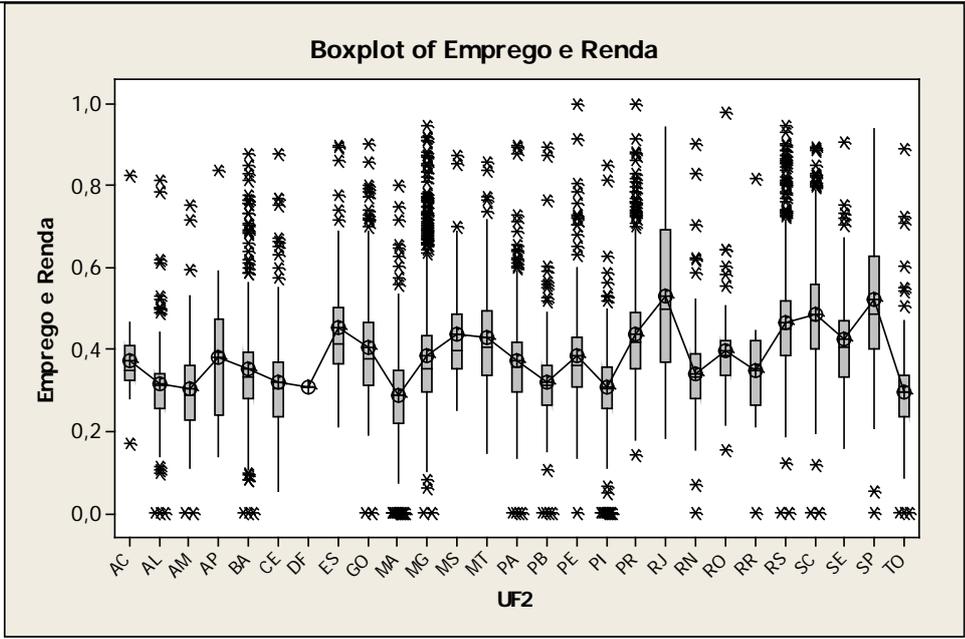
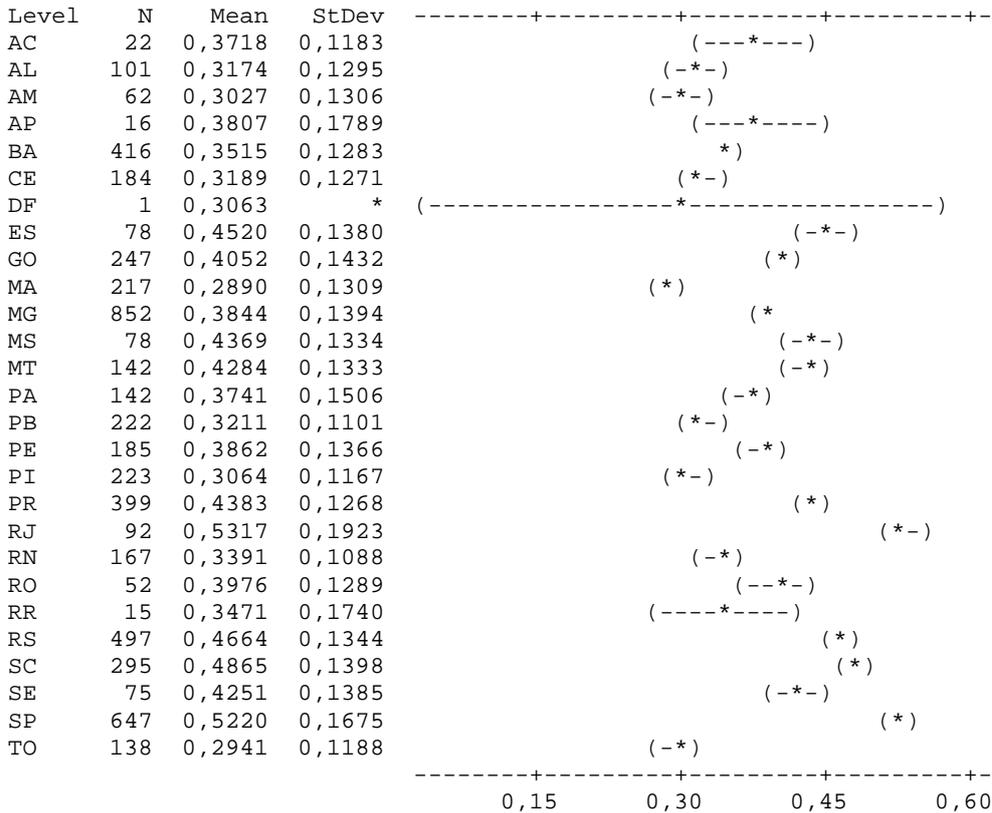


One-way ANOVA: Emprego e Renda versus UF2

Source	DF	SS	MS	F	P
UF2	26	28,6765	1,1029	57,94	0,000
Error	5538	105,4184	0,0190		
Total	5564	134,0949			

S = 0,1380 R-Sq = 21,39% R-Sq(adj) = 21,02%

Individual 95% CIs For Mean Based on Pooled StDev

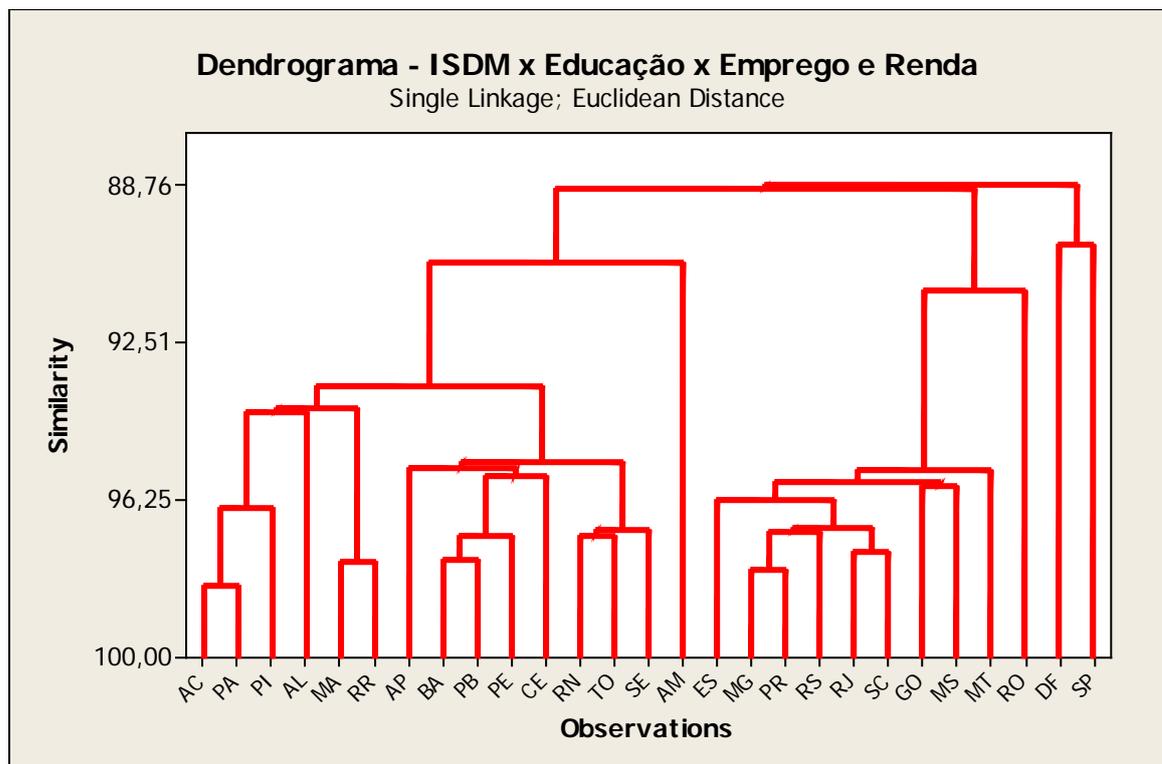


5. DENDOGRAMA

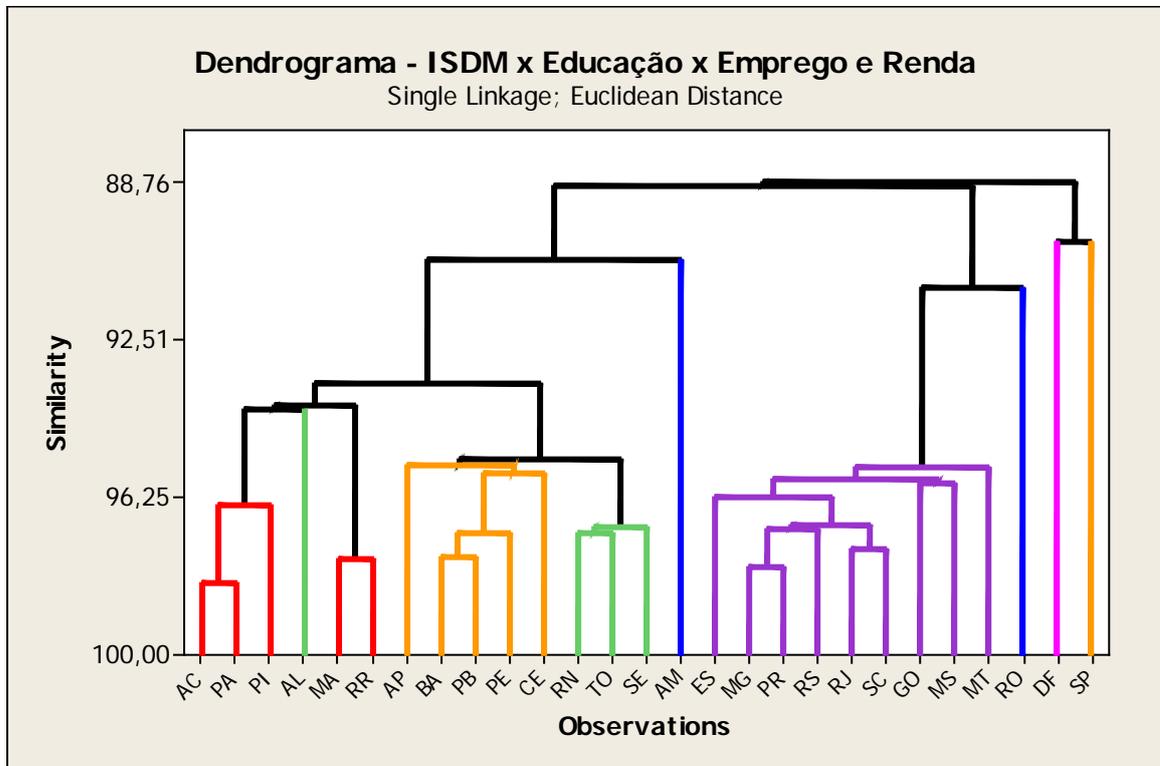
Um **Dendrograma** (dendr(o) = árvore) é um tipo específico de diagrama ou representação icônica que organiza determinados fatores e variáveis. É um diagrama de similaridade.

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

5.1. ISDM x EDUCAÇÃO x EMPREGO E RENDA



5.1. ISDM x EDUCAÇÃO x EMPREGO E RENDA COLORIDOS POR CONGLOMERADOS



O Dendrograma pareceria indicar basicamente 3 Brasis a um nível de similaridade maior que 93.

CAP II ANALISE DISCRIMINANTE

1 – INTRODUÇÃO

O presente trabalho propõe efetuar uma análise discriminantes dos dados da Pesquisa Firjan/FGV sobre o Desenvolvimento dos Municípios nos períodos de 2000 e 2010 e utilizará a classificação por regiões brasileiras: Norte, Nordeste, Centro-Oeste, Sudeste e Sul.

O software estatístico a ser utilizado é o MINITAB16.

2 – OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são os 5565 Municípios brasileiros classificados em suas regiões cujas prefeituras disponibilizaram os dados referentes ao ISDM, IFDM, IFGF, Emprego e Renda, Educação e Saúde.

Fonte: Tabela ISDM, IFDM e IFGF por municípios brasileiros; e <http://www.firjan.org.br>

2.2 As Variáveis

Para o objetivo do presente trabalho, foram trabalhados os dados de 8 variáveis: Emprego e Renda, Educação e Saúde e as regiões Norte, Nordeste, Centro-Oeste, Sudeste e Sul.

Variável	Significado	Tipo	Unidade de Medida
Renda (R)	Média ponderada dos indicadores da dimensão Renda (R1 e R2) padronizada pela média do Brasil.	Variável Quantitativa	Numérico

Emprego/Trabalho (T)	Média ponderada dos indicadores da dimensão Trabalho (T1_1, T1_2 e T2_1) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Educação (E)	Média ponderada dos indicadores da dimensão Educação (E1_1, E1_2, E2_1, E2_2, E2_3, E2_4, E2_5, E2_6, E3_1, E3_2 e E3_3) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Saúde	Média ponderada dos indicadores da dimensão Saúde e Segurança (S1_1, S1_2, S1_3, S2_1, S2_2 e S3_1) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Norte	A Região Norte é a região que possui a maior área (3 869 637,9 km ² , ou 45% do território nacional) e com população de 16,3 milhões de habitantes. É a região com a menor densidade demográfica (3,77 hab./km ² , segundo o censo IBGE 2010). A cidade mais populosa da região, Manaus, com 1,8 milhão de habitantes, é a sétima mais populosa do Brasil.	Variável Quantitativa	Numérico
Nordeste	A Região Nordeste possui um território de 1 556 001 km ² (18,2% do território nacional), dentro dos quais está localizado o Polígono das secas. Sua população é pouco superior a 50 milhões de habitantes. A região possui nove estados.	Variável Quantitativa	Numérico
Centro-Oeste	Ocupa 18,86% do território brasileiro, com uma área de 2.612.077,2 km ² . Sua população é de cerca de 12 milhões de habitantes. A região possui três estados mais um distrito federal.	Variável Quantitativa	Numérico
Sudeste	Possui um território de 927 286 km ² (10,6% do território nacional). Sua população é de cerca de 77 milhões de habitantes. Possui o maior PIB bem como as duas cidades mais populosas do Brasil: São Paulo, com pouco	Variável Quantitativa	Numérico

	mais de 11 milhões de habitantes e Rio de Janeiro com cerca de 6 milhões.		
Sul	A Região Sul é a que possui a menor área (575 316 km ² , ou 6,8% do território nacional) e sua população é de mais de 26 milhões de habitantes, é a segunda região mais rica do país, depois da Região Sudeste, e a que possui o maior IDH, a maior taxa de alfabetização e os melhores níveis de educação, saúde e bem estar social do país. A região possui três estados.	Variável Quantitativa	Numérico

3. ANÁLISE DAS VARIÁVEIS

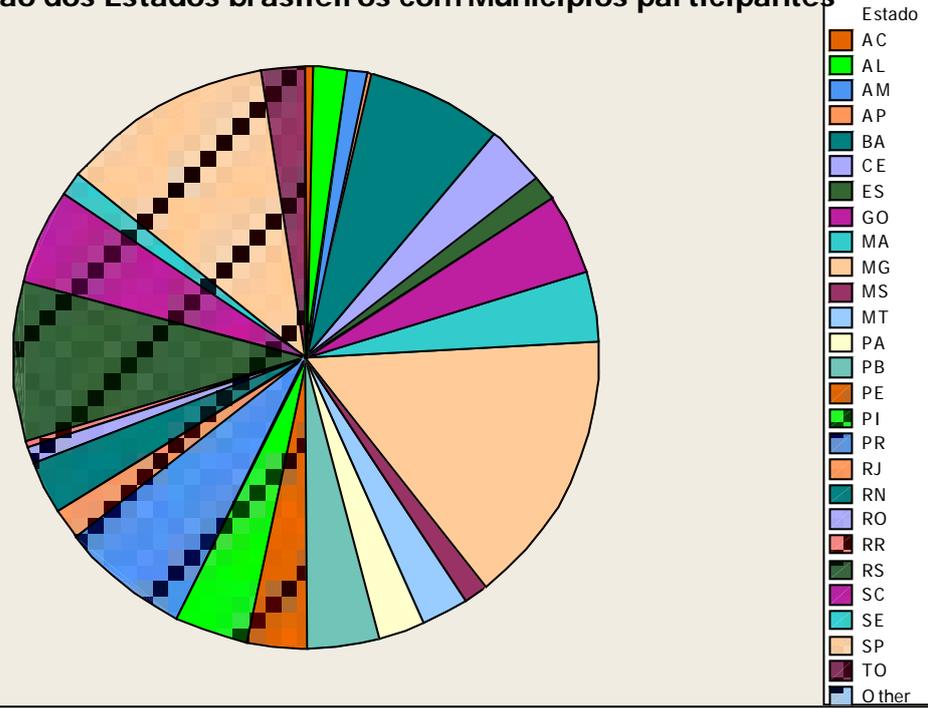
3.1 Variáveis Categóricas

Este tipo de variável indica que o foco de concentração deve ser a análise de gráficos do tipo *pie chart* e barras.

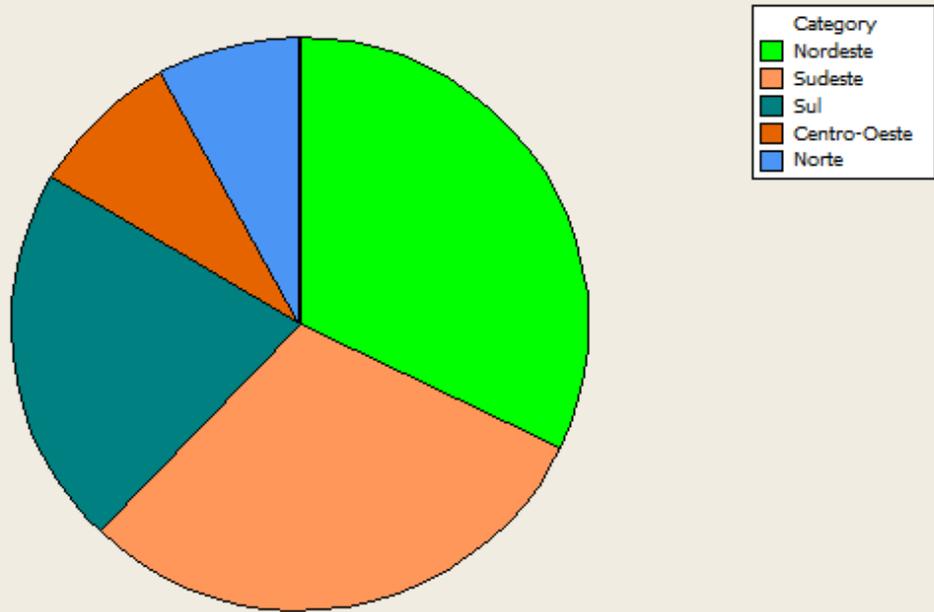
3.1.1 Variável: “Estados”

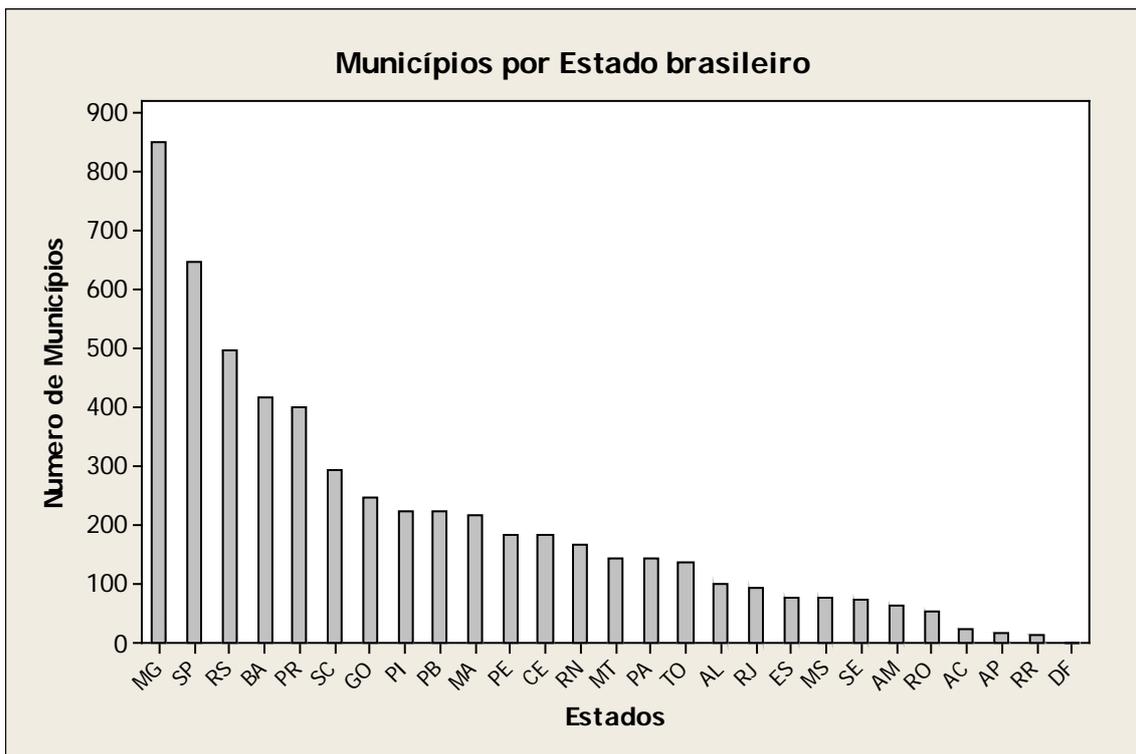
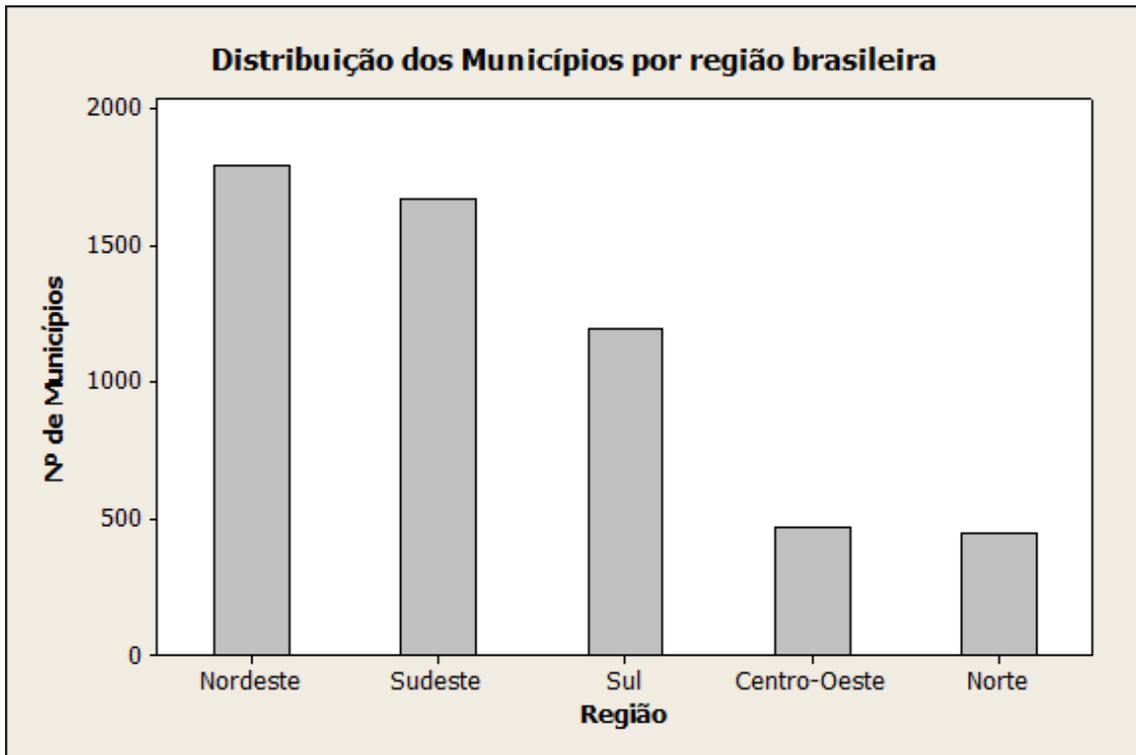
Fazem parte desta pesquisa os 27 Estados brasileiros e seus municípios classificados em regiões. Nos gráficos abaixo é possível visualizar a distribuição de municípios por Estado brasileiro e as cinco regiões brasileiras.

Distribuição dos Estados brasileiros com Municípios participantes



Distribuição das cinco regiões brasileiras

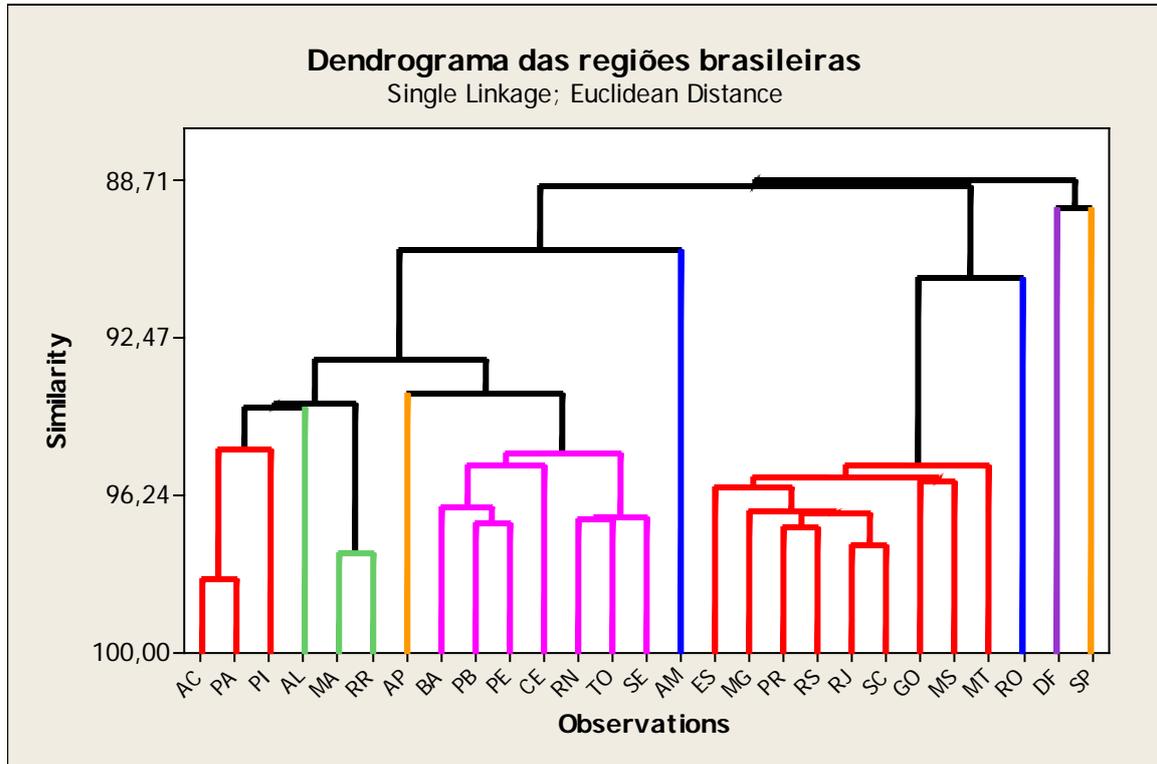




Os dados dos gráficos nos permite visualizar que, os Estados com maior participação de municípios são: Minas Gerais (852 cidades), São Paulo (647 cidades), Rio Grande do Sul (497 cidades), Bahia (416 cidades) e Paraná (399 cidades).

Já entre os Estados com menor número de cidades participantes, temos: Distrito Federal (1 cidade), Roraima (15 cidades), Amapá (16 cidades) e Acre (22 cidades).

4. DENDOGRAMA DAS REGIÕES BRASILEIRAS



5. ANÁLISE DISCRIMINANTE DAS CINCO REGIÕES *VERSUS* ISDM, EMPREGO E RENDA, EDUCAÇÃO E SAÚDE

Discriminant Analysis: Região versus ISDM; Emprego e Re; Educação; Saúde					
Linear Method for Response: Região					
Predictors: ISDM; Emprego e Renda; Educação; Saúde					
Group	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
Count	468	1790	447	1669	1191
Summary of classification					
		True Group			
Put into Group	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
Centro-Oeste	169	231	67	158	179
Nordeste	34	900	103	37	22
Norte	18	603	248	78	6
Sudeste	99	27	27	1137	261
Sul	148	29	2	259	723
Total N	468	1790	447	1669	1191
N correct	169	900	248	1137	723
Proportion	0,361	0,503	0,555	0,681	0,607
N = 5565		N Correct = 3177		Proportion Correct = 0,571	

Essa análise permite verificar que a região sudeste apresenta maior número de acertos, com 1137. Enquanto o Centro-Oeste é a que menos apresenta acertos. A proporção de acerto geral foi de 57%.

Com o objetivo de buscar uma proporção de acerto maior e como consequência da quantidade de municípios que aparecem dentro de duas regiões, serão unificadas as regiões Sudeste x Sul (SULD) e Nordeste x Norte (NOR).

Discriminant Analysis: 3 Brasis versus ISDM; Emprego e Renda; ...

Linear Method for Response: 3 Brasis

Predictors: ISDM; Emprego e Renda; Educação; Saúde

Group	Centro-Oeste	NOR	SULD
Count	468	2237	2860

Summary of classification

Put into Group	True Group		
	Centro-Oeste	NOR	SULD
Centro-Oeste	267	378	756
NOR	48	1817	137
SULD	153	42	1967
Total N	468	2237	2860
N correct	267	1817	1967
Proportion	0,571	0,812	0,688

N = 5565

N Correct = 4051

Proportion Correct = 0,728

É possível notar que, ao transformar as cinco regiões brasileiras em 3 regiões brasileiras a proporção de acertos melhoram consideravelmente em todos os sentidos.

A proporção de acertos geral agora é de 72,8% e o grupo NOR (Nordeste e Norte) apresenta os maiores números de acerto enquanto o Centro-Oeste continua com o menor numero de acertos, porém com melhora quando comparado com a análise anterior.

CAP III REGRESSÃO LOGÍSTICA

1 – INTRODUÇÃO

O presente trabalho propõe efetuar uma regressão logística dos dados da Pesquisa Firjan/FGV sobre o Desenvolvimento dos Municípios nos períodos de 2000 e 2010 e utilizará a classificação por regiões brasileiras: Norte, Nordeste, Centro-Oeste, Sudeste e Sul. As variáveis trabalhadas são ISDM, Educação, Emprego e Renda e Saúde.

O software estatístico a ser utilizado é o MINITAB16.

O objetivo da regressão logística é produzir um modelo que permita a predição de valores tomados por uma variável categórica, normalmente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias.

O êxito da regressão logística assenta, sobretudo, nas numerosas ferramentas que permitem interpretar de modo aprofundado os resultados obtidos. Em comparação com as técnicas conhecidas em regressão, em especial a regressão linear, a regressão logística distingue-se essencialmente pelo fato de a variável resposta ser categórica.

2 – OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são os 5565 Municípios brasileiros classificados em suas regiões cujas prefeituras disponibilizaram os dados referentes ao ISDM, IFDM, IFGF, Emprego e Renda, Educação e Saúde.

Fonte: Tabela ISDM, IFDM e IFGF por municípios brasileiros; e <http://www.firjan.org.br>

2.2 As Variáveis

Para o objetivo do presente trabalho, foram trabalhados os dados de 8 variáveis: Emprego e Renda, Educação e Saúde e as regiões Norte, Nordeste, Centro-Oeste, Sudeste e Sul.

Variável	Significado	Tipo	Unidade de Medida
Renda (R)	Média ponderada dos indicadores da dimensão Renda (R1 e R2) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Emprego/Trabalho (T)	Média ponderada dos indicadores da dimensão Trabalho (T1_1, T1_2 e T2_1) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Educação (E)	Média ponderada dos indicadores da dimensão Educação (E1_1, E1_2, E2_1, E2_2, E2_3, E2_4, E2_5, E2_6, E3_1, E3_2 e E3_3) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Saúde	Média ponderada dos indicadores da dimensão Saúde e Segurança (S1_1, S1_2, S1_3, S2_1, S2_2 e S3_1) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Norte	A Região Norte é a região que possui a maior área (3 869 637,9 km ² , ou 45% do território nacional) e com população de 16,3 milhões de habitantes. É a região com a menor densidade demográfica (3,77 hab./km ² , segundo o censo IBGE 2010). A cidade mais populosa da região, Manaus, com 1,8 milhão de habitantes, é a sétima mais populosa do Brasil.	Variável Quantitativa	Numérico
Nordeste	A Região Nordeste possui um território de 1 556 001 km ² (18,2% do território nacional), dentro dos quais está localizado o Polígono das secas. Sua população é pouco superior a 50 milhões de habitantes. A região possui nove estados.	Variável Quantitativa	Numérico

Centro-Oeste	Ocupa 18,86% do território brasileiro, com uma área de 2.612.077,2 km ² . Sua população é de cerca de 12 milhões de habitantes. A região possui três estados mais um distrito federal.	Variável Quantitativa	Numérico
Sudeste	Possui um território de 927 286 km ² (10,6% do território nacional). Sua população é de cerca de 77 milhões de habitantes. Possui o maior PIB bem como as duas cidades mais populosas do Brasil: São Paulo, com pouco mais de 11 milhões de habitantes e Rio de Janeiro com cerca de 6 milhões.	Variável Quantitativa	Numérico
Sul	A Região Sul é a que possui a menor área (575 316 km ² , ou 6,8% do território nacional) e sua população é de mais de 26 milhões de habitantes, é a segunda região mais rica do país, depois da Região Sudeste, e a que possui o maior IDH, a maior taxa de alfabetização e os melhores níveis de educação, saúde e bem estar social do país. A região possui três estados.	Variável Quantitativa	Numérico

3. ANÁLISE DAS VARIÁVEIS

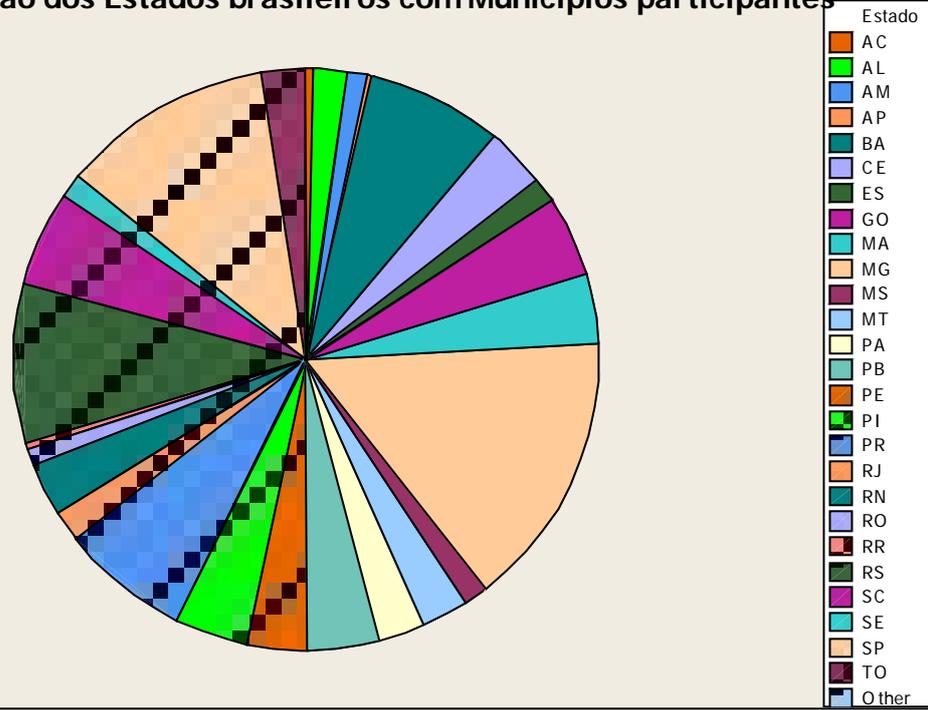
3.1 Variáveis Categóricas

Este tipo de variável indica que o foco de concentração deve ser a análise de gráficos do tipo *pie chart* e barras.

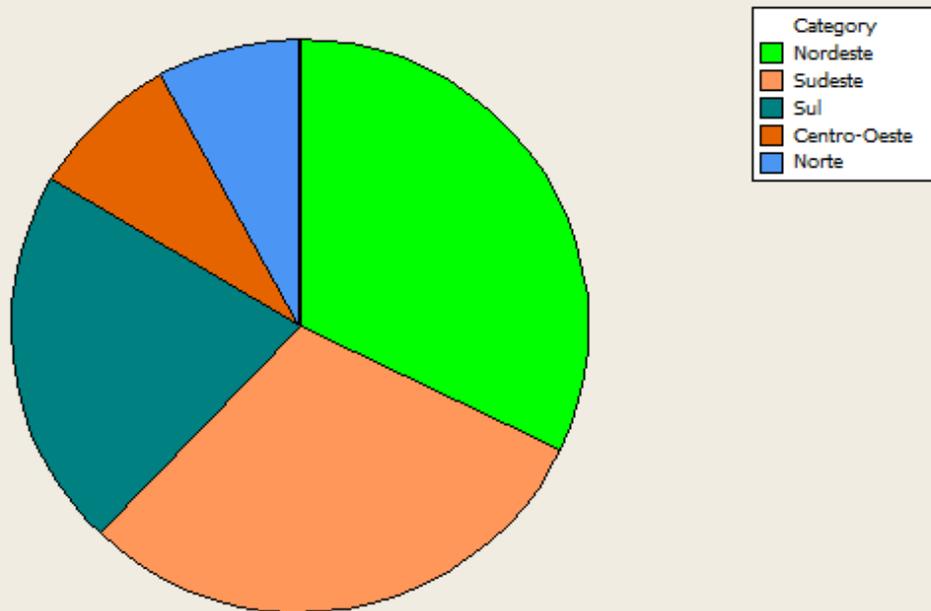
3.1.1 Variável: “Estados”

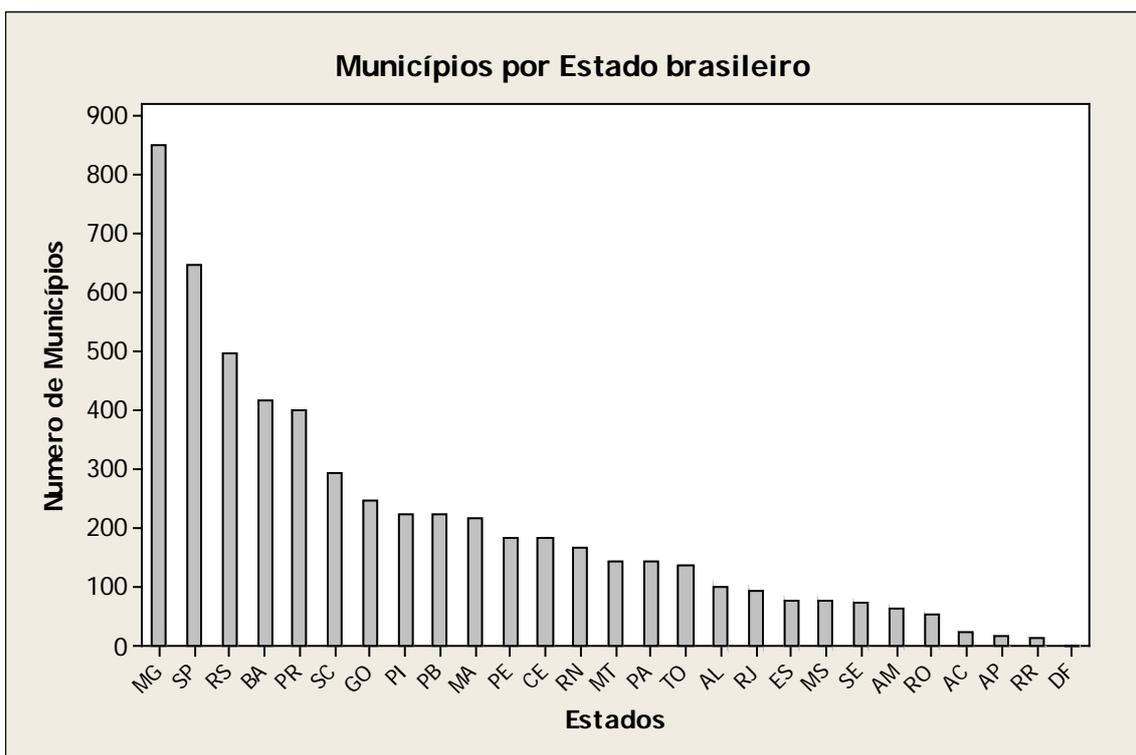
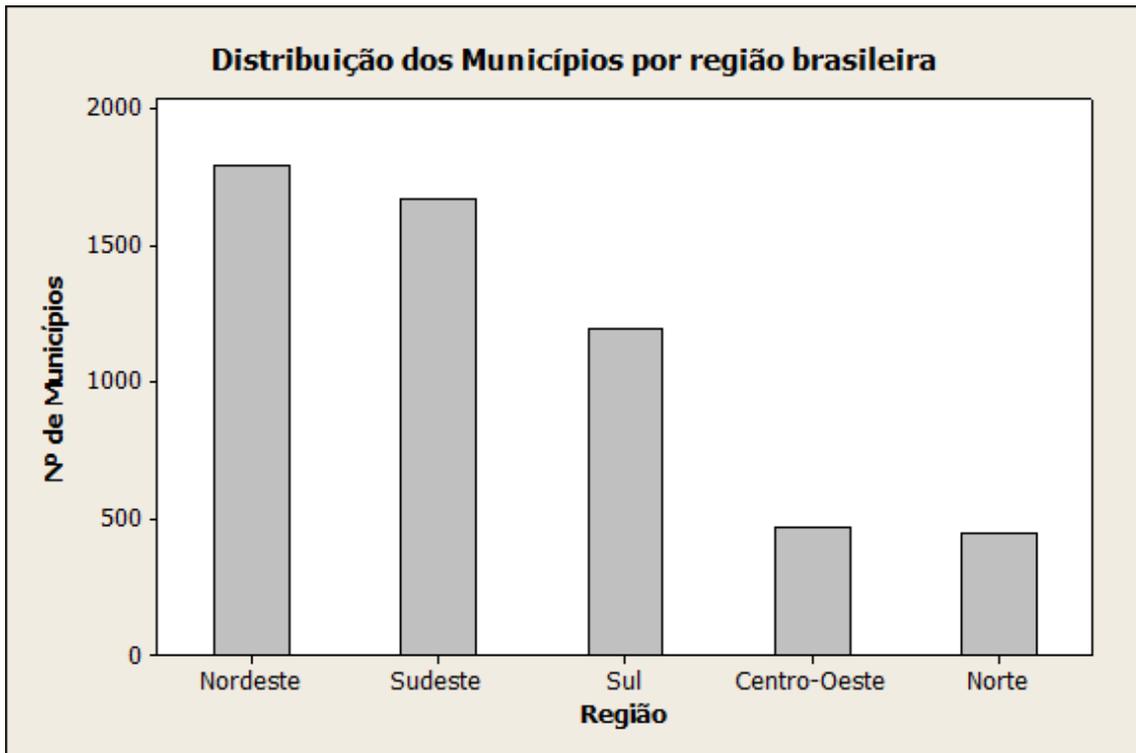
Fazem parte desta pesquisa os 27 Estados brasileiros e seus municípios classificados em regiões. Nos gráficos abaixo é possível visualizar a distribuição de municípios por Estado brasileiro e as cinco regiões brasileiras.

Distribuição dos Estados brasileiros com Municípios participantes



Distribuição das cinco regiões brasileiras





Os dados dos gráficos nos permite visualizar que, os Estados com maior participação de municípios são: Minas Gerais (852 cidades), São Paulo (647 cidades), Rio Grande do Sul (497 cidades), Bahia (416 cidades) e Paraná (399 cidades).

Já entre os Estados com menor número de cidades participantes, temos: Distrito Federal (1 cidade), Roraima (15 cidades), Amapá (16 cidades) e Acre (22 cidades).

4. REGRESSÃO LOGÍSTICA

Ordinal Logistic Regression: Região versus ISDM; Emprego e Renda; ...

Link Function: Logit

Response Information

Variable	Value	Count
Região	Centro-Oeste	468
	Nordeste	1790
	Norte	447
	Sudeste	1669
	Sul	1191
Total		5565

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Const(1)	2,95343	0,232409	12,71	0,000			
Const(2)	5,27909	0,237742	22,21	0,000			
Const(3)	5,78785	0,239866	24,13	0,000			
Const(4)	7,63908	0,248532	30,74	0,000			
ISDM	-0,848336	0,0441828	-19,20	0,000	0,43	0,39	0,47
Emprego e Renda	0,107100	0,188368	0,57	0,570	1,11	0,77	1,61
Educação	0,134947	0,364992	0,37	0,712	1,14	0,56	2,34
Saúde	-2,93169	0,356290	-8,23	0,000	0,05	0,03	0,11

Log-Likelihood = -7225,729

Test that all slopes are zero: G = 1873,197, DF = 4, P-Value = 0,000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	28572,9	22252	0,000
Deviance	14451,5	22252	1,000

Measures of Association:

(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures
Concordant	8520996	73,6	Somers' D 0,48
Discordant	2997672	25,9	Goodman-Kruskal Gamma 0,48
Ties	52457	0,5	Kendall's Tau-a 0,36
Total	11571125	100,0	

A % de acertos para 5 Brasis se elevou para 73% utilizando Regressão Logística, sendo que são os coeficientes de ISDM e Saúde tem a precisão necessária.

CAP IV ÁRVORES DE CLASSIFICAÇÃO

1 – INTRODUÇÃO

O presente trabalho propõe elaborar uma árvore de classificação dos dados da Pesquisa Firjan/FGV sobre o Desenvolvimento dos Municípios nos períodos de 2000 e 2010 e utilizará a classificação por regiões brasileiras: Norte, Nordeste, Centro-Oeste, Sudeste e Sul. As variáveis trabalhadas são ISDM, Educação, Emprego e Renda e Saúde. Para a elaboração da árvore de classificação foi utilizado os dados da divisão regional brasileira que resultou na divisão do Brasil em 3 regiões principais, classificadas neste trabalho como: Centro-Oeste, NORD (Nordeste e Norte), SUD (Sudeste e Sul).

O software estatístico a ser utilizado é o MINITAB16 e o SPSS.

A técnica árvore de classificação alcançou o segmento de negócios através da utilização em pesquisas de mercado. Tendo como pontos fortes a simplicidade de sua representação gráfica baseado em árvores e a facilidade de entender as regras e perfis derivados de cada segmento (nós), rapidamente foi adotada por outras áreas de marketing, sobretudo aquelas formadas por gestores com menor grau de sofisticação analítica.

Um dos vários algoritmos criados (CHAID) tornou-se popular em marketing direto, sobretudo para selecionar grupos de consumidores e prever a taxa de resposta de uma campanha em função do perfil determinado pelo algoritmo.

2 – OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são os 5565 Municípios brasileiros classificados em suas regiões cujas prefeituras disponibilizaram os dados referentes ao ISDM, IFDM, IFGF, Emprego e Renda, Educação e Saúde.

Fonte: Tabela ISDM, IFDM e IFGF por municípios brasileiros; e <http://www.firjan.org.br>

2.2 As Variáveis

Para o objetivo do presente trabalho, foram trabalhados os dados de 8 variáveis: Emprego e Renda, Educação e Saúde e as regiões Norte, Nordeste, Centro-Oeste, Sudeste e Sul.

Variável	Significado	Tipo	Unidade de Medida
Renda (R)	Média ponderada dos indicadores da dimensão Renda (R1 e R2) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Emprego/Trabalho (T)	Média ponderada dos indicadores da dimensão Trabalho (T1_1, T1_2 e T2_1) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Educação (E)	Média ponderada dos indicadores da dimensão Educação (E1_1, E1_2, E2_1, E2_2, E2_3, E2_4, E2_5, E2_6, E3_1, E3_2 e E3_3) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Saúde	Média ponderada dos indicadores da dimensão Saúde e Segurança (S1_1, S1_2, S1_3, S2_1, S2_2 e S3_1) padronizada pela média do Brasil.	Variável Quantitativa	Numérico
Norte	A Região Norte é a região que possui a maior área (3 869 637,9 km ² , ou 45% do território nacional) e com população de 16,3 milhões de habitantes. É a região com a menor densidade demográfica (3,77 hab./km ² , segundo o censo IBGE 2010). A cidade mais populosa da região, Manaus, com 1,8 milhão de habitantes, é a sétima mais populosa do Brasil.	Variável Quantitativa	Numérico
Nordeste	A Região Nordeste possui um território de 1 556 001 km ² (18,2% do território nacional), dentro dos quais está localizado o Polígono das secas. Sua população é pouco superior a 50 milhões de habitantes. A região possui nove	Variável Quantitativa	Numérico

	estados.		
Centro-Oeste	Ocupa 18,86% do território brasileiro, com uma área de 2.612.077,2 km ² . Sua população é de cerca de 12 milhões de habitantes. A região possui três estados mais um distrito federal.	Variável Quantitativa	Numérico
Sudeste	Possui um território de 927 286 km ² (10,6% do território nacional). Sua população é de cerca de 77 milhões de habitantes. Possui o maior PIB bem como as duas cidades mais populosas do Brasil: São Paulo, com pouco mais de 11 milhões de habitantes e Rio de Janeiro com cerca de 6 milhões.	Variável Quantitativa	Numérico
Sul	A Região Sul é a que possui a menor área (575 316 km ² , ou 6,8% do território nacional) e sua população é de mais de 26 milhões de habitantes, é a segunda região mais rica do país, depois da Região Sudeste, e a que possui o maior IDH, a maior taxa de alfabetização e os melhores níveis de educação, saúde e bem estar social do país. A região possui três estados.	Variável Quantitativa	Numérico

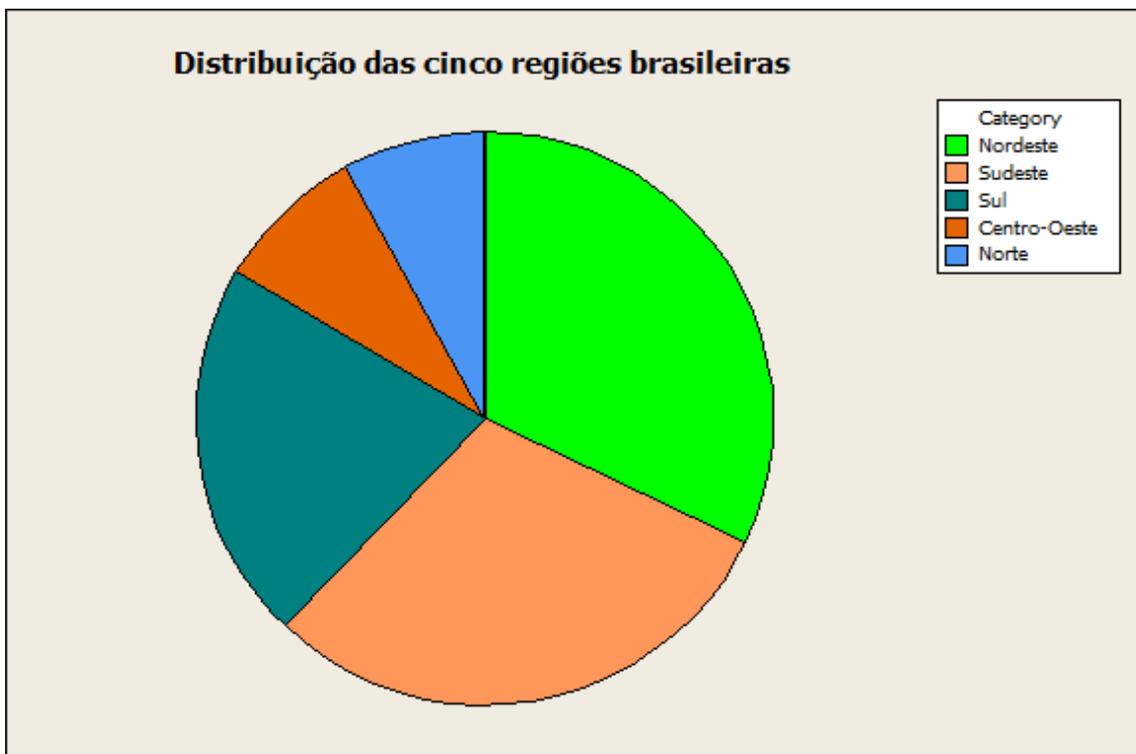
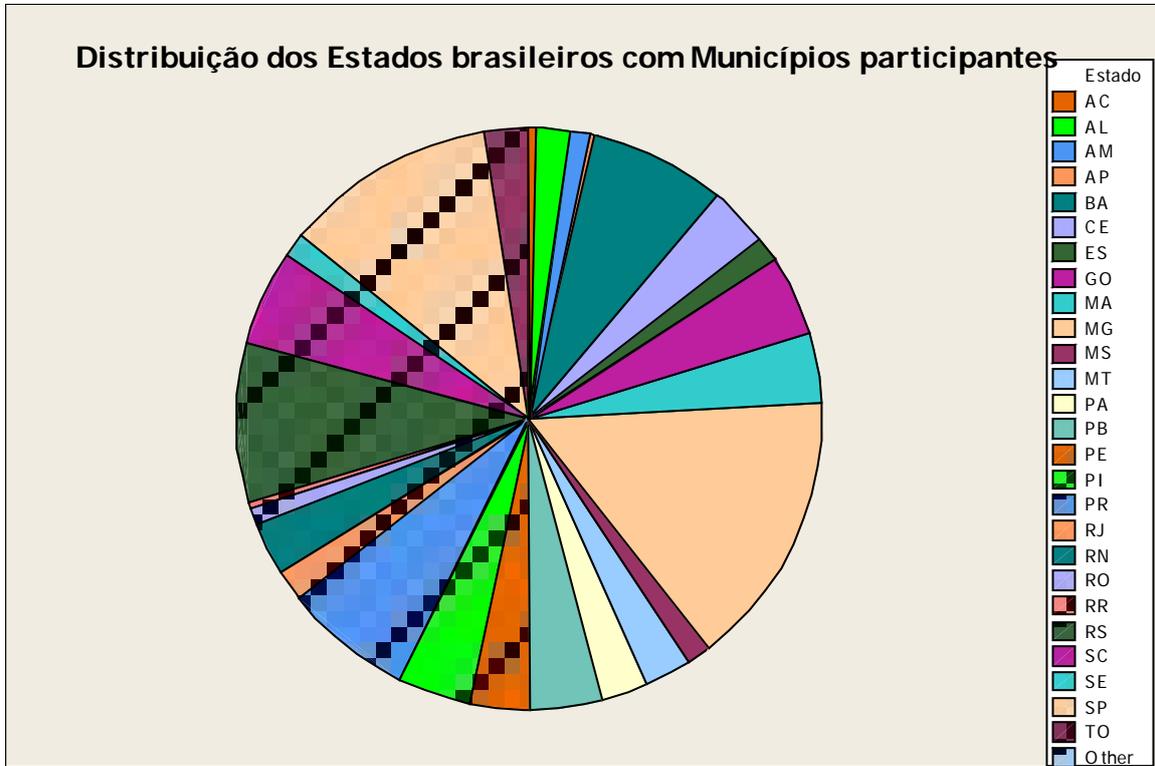
3. ANÁLISE DAS VARIÁVEIS

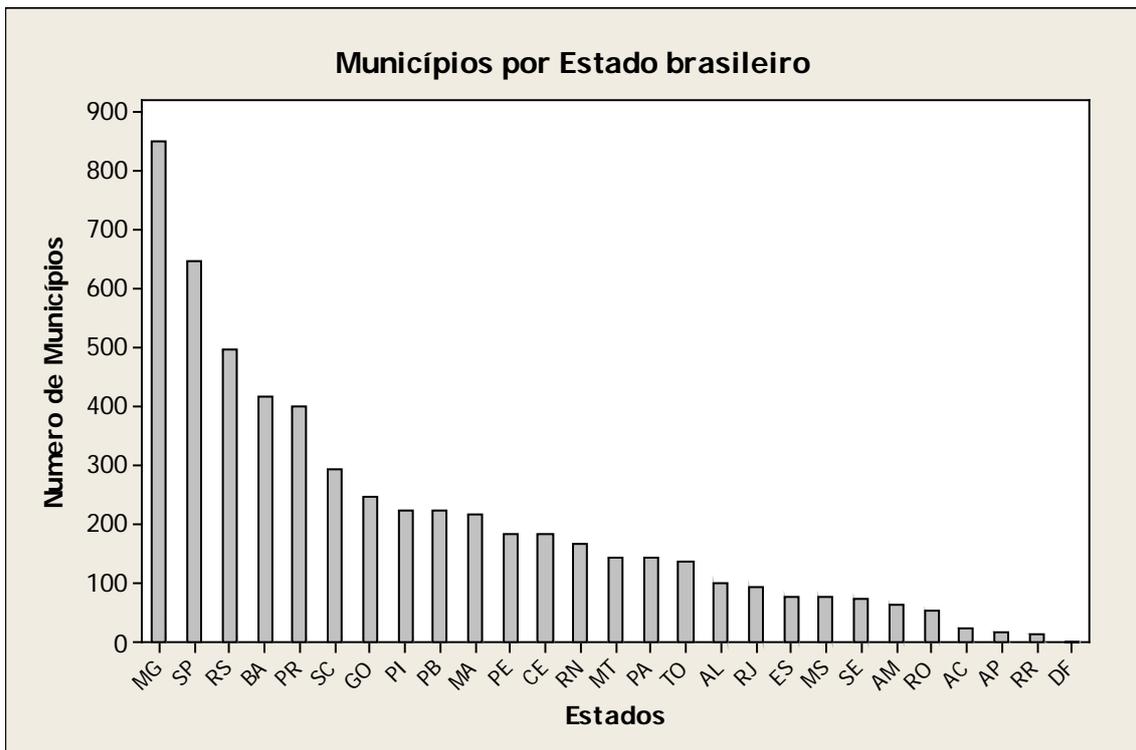
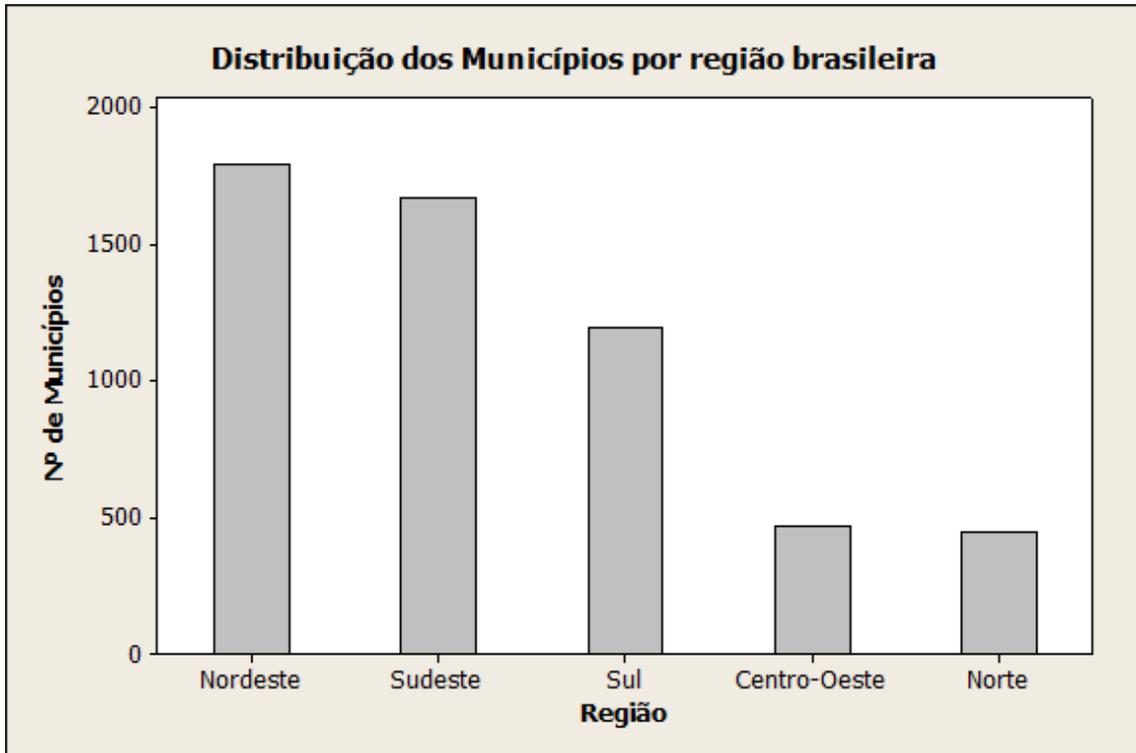
3.1 Variáveis Categóricas

Este tipo de variável indica que o foco de concentração deve ser a análise de gráficos do tipo *pie chart* e barras.

3.1.1 Variável: “Estados”

Fazem parte desta pesquisa os 27 Estados brasileiros e seus municípios classificados em regiões. Nos gráficos abaixo é possível visualizar a distribuição de municípios por Estado brasileiro e as cinco regiões brasileiras.





Os dados dos gráficos nos permite visualizar que, os Estados com maior participação de municípios são: Minas Gerais (852 cidades), São Paulo (647 cidades), Rio Grande do Sul (497 cidades), Bahia (416 cidades) e Paraná (399 cidades).

Já entre os Estados com menor número de cidades participantes, temos: Distrito Federal (1 cidade), Roraima (15 cidades), Amapá (16 cidades) e Acre (22 cidades).

4. ÁRVORE DE CLASSIFICAÇÃO DAS VARIÁVEIS: ISDM, EDUCAÇÃO, EMPREGO E RENDA E SAÚDE

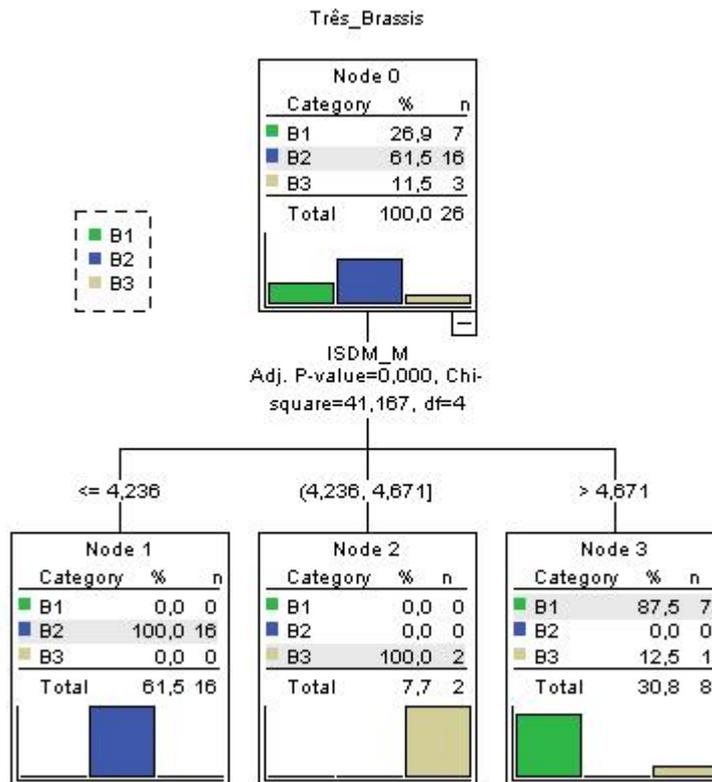
Foram transportados os seguintes dados do Minitab para o *software* SPSS:

Região1	Região2	N	ISDM_M	ISDM_SD	EmpregoRenda_M	EmpregoRenda_SD	Educação_M	Educação_SD	Saúde_M	Saúde_SD	Região3	3Brassis
Nordeste	AL	101	3,1996	0,6982	0,3174	0,1295	0,60578	0,06785	0,71957	0,05788	NORD	B2
Norte	AM	62	2,3550	0,7750	0,3027	0,1306	0,61479	0,06003	0,62033	0,07749	NORD	B2
Norte	AP	16	3,4060	0,6679	0,3807	0,1789	0,60567	0,03819	0,62593	0,05655	NORD	B2
Nordeste	BA	416	3,6024	0,6378	0,3515	0,1283	0,59061	0,06767	0,68630	0,07114	NORD	B2
Nordeste	CE	184	3,4534	0,5689	0,3189	0,1271	0,73541	0,04756	0,81269	0,05808	NORD	B2
Sudeste	ES	78	4,9080	0,4716	0,4520	0,1380	0,84199	0,05457	0,84231	0,05686	SUD	B1
Centro-Oeste	GO	247	4,7989	0,4546	0,4052	0,1432	0,76734	0,07622	0,83356	0,06867	Centro-Oeste	B3
Nordeste	MA	217	2,7076	0,7130	0,2890	0,1309	0,63245	0,07051	0,66166	0,06948	NORD	B2
Sudeste	MG	852	4,9973	0,7056	0,3844	0,1394	0,78568	0,06210	0,80049	0,08317	SUD	B1
Centro-Oeste	MS	78	4,6714	0,5353	0,4369	0,1334	0,72983	0,05297	0,82405	0,07894	Centro-Oeste	B3
Centro-Oeste	MT	142	4,5219	0,5831	0,4284	0,1333	0,74137	0,06685	0,82897	0,08224	Centro-Oeste	B3
Norte	PA	142	2,9406	0,8605	0,3741	0,1506	0,58226	0,07192	0,63492	0,06429	NORD	B2
Nordeste	PB	222	3,5758	0,5246	0,3211	0,1101	0,65751	0,06866	0,77246	0,06596	NORD	B2
Nordeste	PE	185	3,6533	0,6738	0,3862	0,1366	0,66662	0,06338	0,75181	0,05659	NORD	B2
Nordeste	PI	223	3,0103	0,6091	0,3064	0,1167	0,65258	0,07257	0,74471	0,05934	NORD	B2
Sul	PR	399	5,0427	0,4915	0,4383	0,1268	0,78194	0,06330	0,88912	0,05848	SUD	B1
Sudeste	RJ	92	5,2888	0,3299	0,5317	0,1923	0,78439	0,06950	0,82826	0,06287	SUD	B1
Nordeste	RN	167	3,7924	0,5833	0,3391	0,1088	0,71961	0,07062	0,78813	0,07119	NORD	B2
Norte	RO	52	4,2359	0,4502	0,3976	0,1289	0,69039	0,05534	0,75802	0,05249	NORD	B2
Norte	RR	15	2,6668	1,2718	0,3471	0,1740	0,60517	0,05907	0,68890	0,07725	NORD	B2
Sul	RS	497	5,1373	0,5099	0,4664	0,1344	0,76149	0,07711	0,89180	0,06212	SUD	B1
Sul	SC	295	5,2241	0,4816	0,4865	0,1398	0,81286	0,05324	0,84725	0,06902	SUD	B1
Nordeste	SE	75	3,8260	0,5672	0,4251	0,1385	0,67580	0,05106	0,75034	0,07763	NORD	B2
Sudeste	SP	647	5,6458	0,3415	0,5220	0,1675	0,90442	0,05082	0,86731	0,05239	SUD	B1
Norte	TO	138	3,8761	0,6077	0,2941	0,1188	0,69921	0,07383	0,74138	0,06498	NORD	B2

Nesta etapa faz-se importante observar que os dados trabalhados se referem a média e o desvio padrão das variáveis ISDM, Educação, Emprego e renda e Saúde, com suas classificações por regiões e também uma nova classificação baseada na divisão do Brasil e três regiões principais (“3 Brasis”).

A próxima etapa será elaborar a árvore de classificação no *software* SPSS com os dados copiados do MiniTab. Serão montadas árvores de classificação para as variáveis “3brassis” e “região3”, ambos referentes a nova classificação brasileira em 3 divisões: Centro-Oeste, NORD (Nordeste e Norte) e SUD (Sudeste e Sul).

4.1 Árvore de classificação referente as médias dos estados – Variáveis: ISDM, Educação, Emprego e Renda e Saúde x Três Brasis



Risk

Estimate	Std. Error
,038	,038

Growing Method: CHAID

Dependent Variable:

Três_Brassis

Classification

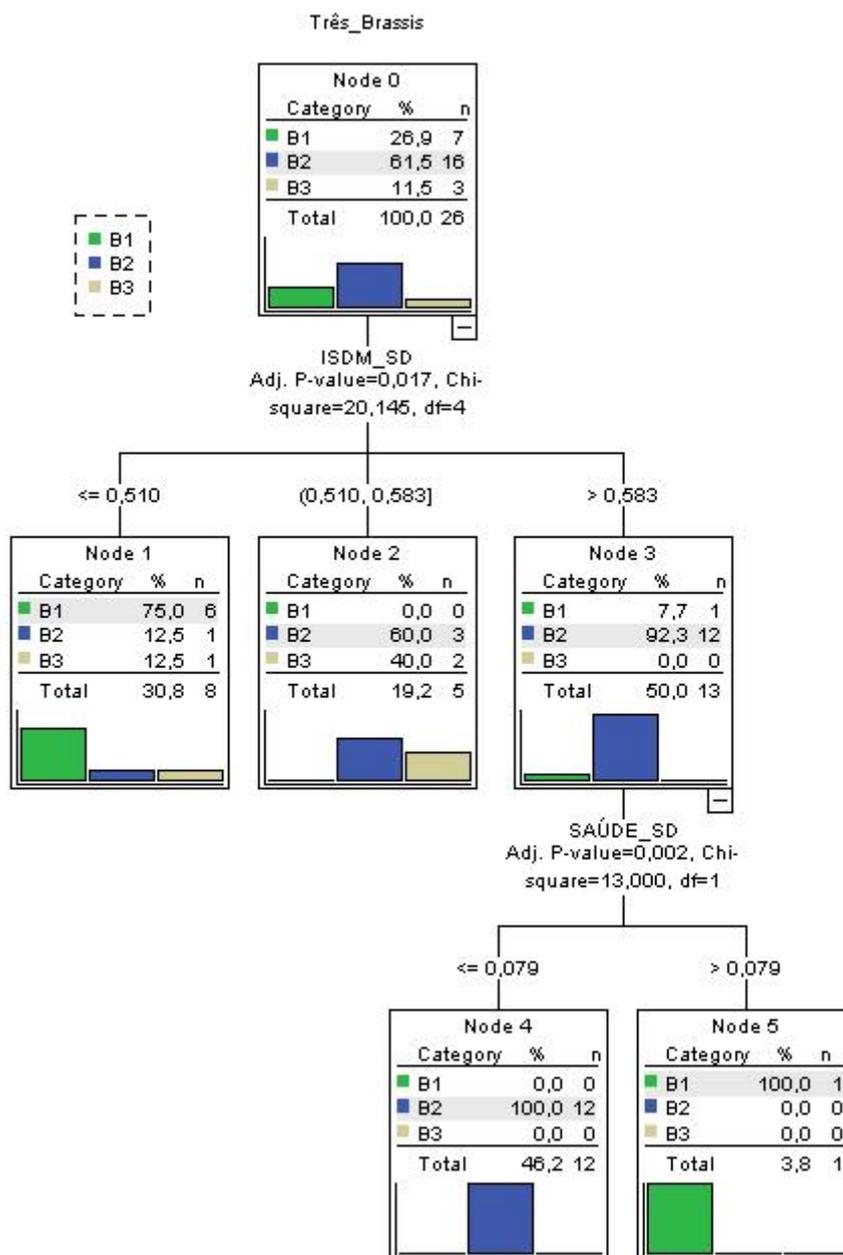
Observed	Predicted			
	B1	B2	B3	Percent Correct
B1	7	0	0	100,0%
B2	0	16	0	100,0%
B3	1	0	2	66,7%
Overall Percentage	30,8%	61,5%	7,7%	96,2%

Growing Method: CHAID

Dependent Variable: Três_Brassis

Se observou uma alta % de acertos (96 %) para 3 Brasis utilizando as medias dos estados referentes as variáveis : ISDM, Educação, Emprego e Renda e Saúde

4.2 Árvore de classificação dos desvios padrões (disparidades) – Variáveis : ISDM, Educação, Emprego e Renda e Saúde x Três Brasis



Risk

Estimate	Std. Error
,154	,071

Growing Method: CHAID

Dependent Variable:

Três_Brassis

Classification

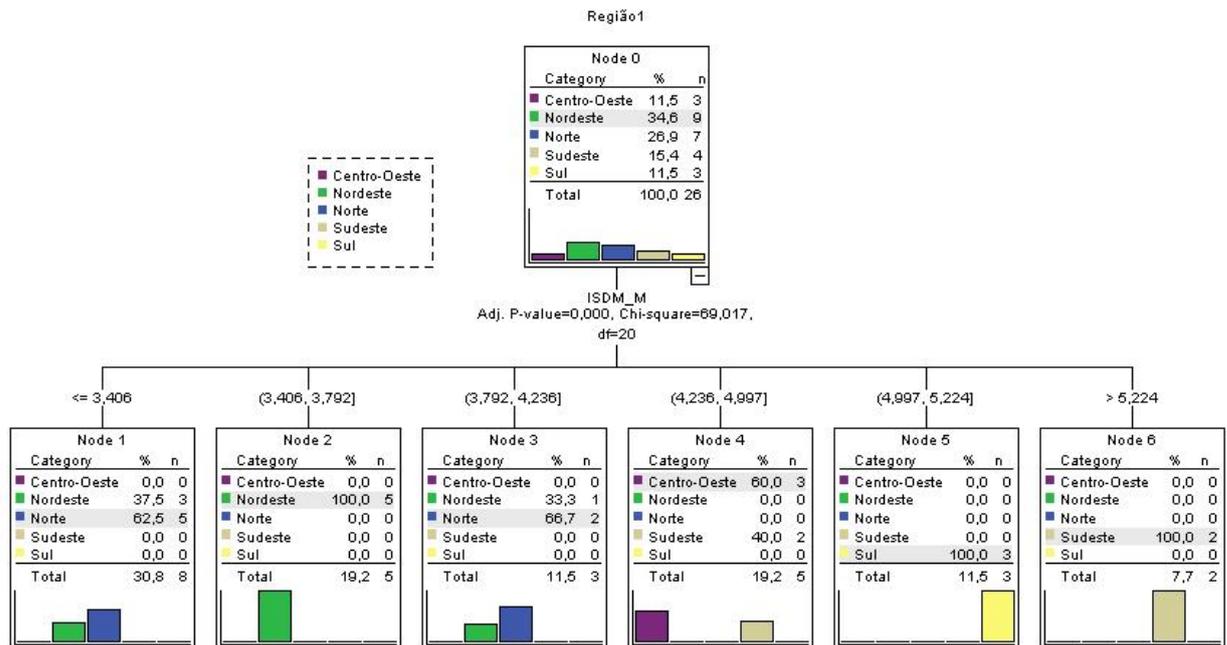
Observed	Predicted			
	B1	B2	B3	Percent Correct
B1	7	0	0	100,0%
B2	1	15	0	93,8%
B3	1	2	0	0,0%
Overall Percentage	34,6%	65,4%	0,0%	84,6%

Growing Method: CHAID

Dependent Variable: Três_Brassis

Similarmente ao caso das médias se observou uma alta % de acertos(84,6 %) para 3 Brasis utilizando as disparidades nos estados referentes as variáveis : ISDM, Educação, Emprego e Renda e Saúde

4.3 Árvore de classificação das médias dos estados para 5 Brasis – Variáveis : ISDM, Educação, Emprego e Renda e Saúde x Região



Risk

Estimate	Std. Error
,231	,083

Growing Method: CHAID

Dependent Variable:

Região1

Classification

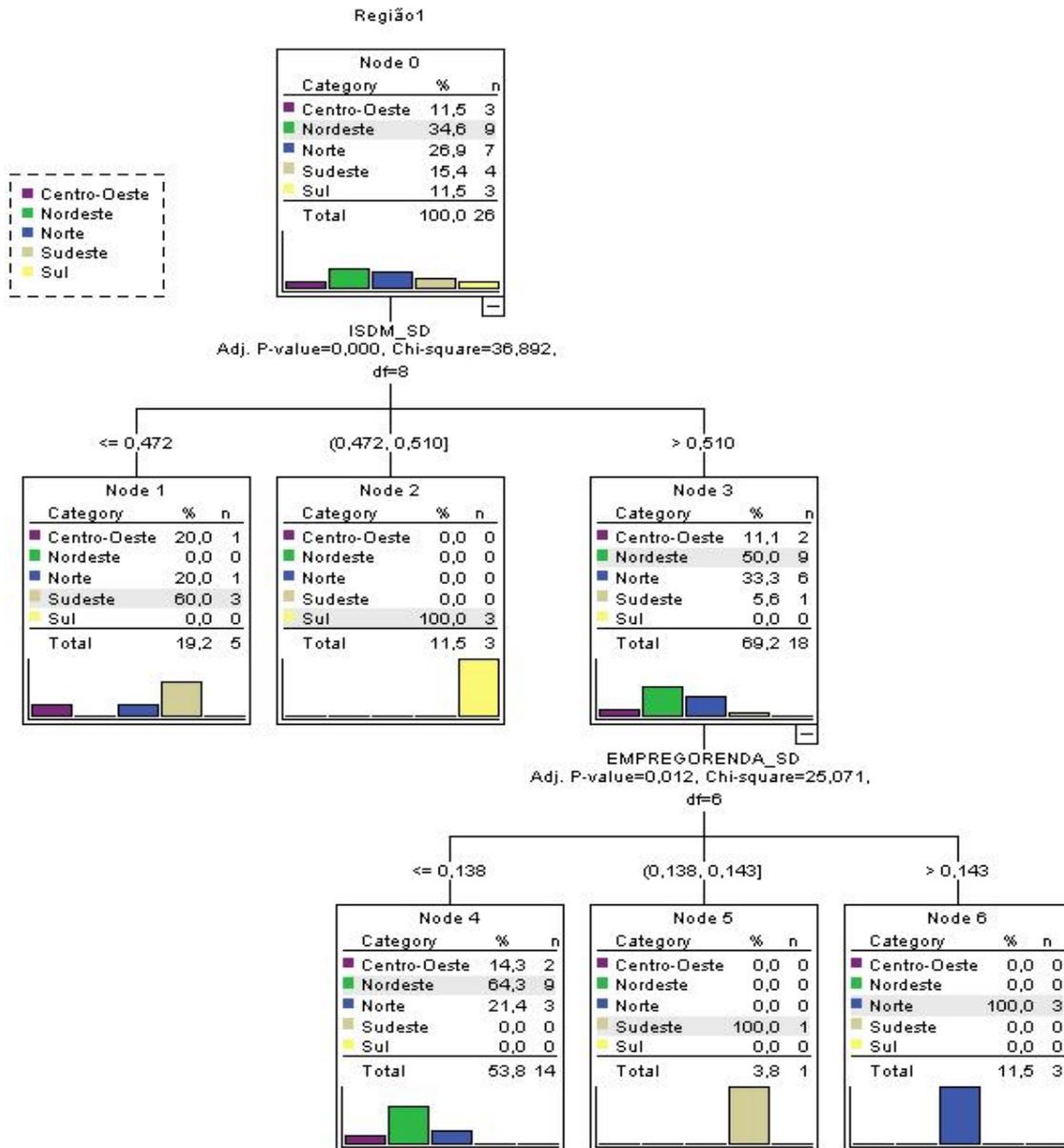
Observed	Predicted					Percent Correct
	Centro-Oeste	Nordeste	Norte	Sudeste	Sul	
Centro-Oeste	3	0	0	0	0	100,0%
Nordeste	0	5	4	0	0	55,6%
Norte	0	0	7	0	0	100,0%
Sudeste	2	0	0	2	0	50,0%
Sul	0	0	0	0	3	100,0%
Overall Percentage	19,2%	19,2%	42,3%	7,7%	11,5%	76,9%

Growing Method: CHAID

Dependent Variable: Região1

Para 5 Brasis o nível de precisão cai de 96% para 76,9 % no caso das medias.

4.4 Árvore de classificação dos desvios padrões por estado (disparidades) para 5 Brasis – Variáveis: ISDM, Educação, Emprego e Renda e Saúde x Região



Risk

Estimate	Std. Error
,269	,087

Growing Method: CHAID

Dependent Variable:

Região1

Classification						
Observed	Predicted					Percent Correct
	Centro-Oeste	Nordeste	Norte	Sudeste	Sul	
Centro-Oeste	0	2	0	1	0	0,0%
Nordeste	0	9	0	0	0	100,0%
Norte	0	3	3	1	0	42,9%
Sudeste	0	0	0	4	0	100,0%
Sul	0	0	0	0	3	100,0%
Overall Percentage	0,0%	53,8%	11,5%	23,1%	11,5%	73,1%

Growing Method: CHAID

Dependent Variable: Região1

A porcentagem no caso das disparidades cai de 84,6 %(medias) para 73,1 % para 5 Brasis.