PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO

Programas de Pós Graduação em Economia e Administração da PUC-SP

BOLETIM DE ANÁLISE ESTATÍSTICO BASTA 2017

Vol. 1

IDHM

ÍNDICE DE DESENVOLVIMENTO HUMANO MUNICIPAL ATLAS BRASIL

CLASSIFICAÇÃO:

Da Visão Não Supervisionada a Supervisionada

Utilizando 18 variáveis selecionadas do Atlas Brasil.

DISCIPLINA: MÉTODOS QUALITATIVOS E QUANTITATIVOS EM ADMINISTRAÇÃO
PROF. ARNOLDO JOSÉ DE HOYOS GUEVARA

Danilo Nunes

1º SEMESTRE DE 2017

São Paulo - SP

2017

SUMÁRIO

INTRODUÇÃO 3
CAPÍTULO I. ENTENDENDO OS DADOS.....3
1.1. Os Indivíduos..... 3
1.2. As Variáveis 3
1.3. Análise de dados..... 5
1.4. Base de dados 6
CAPÍTULO II. ANÁLISE DISCRIMINANTE..... 7
CONSIDERAÇÕES FINAIS 15
REFERÊNCIAS.....;.....21

INTRODUÇÃO

O presente trabalho tem por objetivo efetuar a análise discriminante de dados dimensionadores do IDHM, Plataforma Atlas Brasil, Base IBGE. Por fim, após as análises fazemos as considerações finais, com a distribuição demonstrada no mapa do Brasil, considerados os Estados correlacionados. O software estatístico utilizado é o **MINITAB**

CAPÍTULO I. ENTENDENDO OS DADOS

1.1 Os Indivíduos

O foco de estudo deste trabalho são todos os municípios brasileiros, com os 232 indicadores disponibilizados no portal Atlas Brasil, dados de 2010, abertos em 08 categorias, a saber:

Demografia
Educação
Renda
Trabalho
Habitação
Vulnerabilidade
População
IDHM (Longevidade, Renda e Educação

O IDHM demonstrado no Atlas de Desenvolvimento Humano é baseado exclusivamente nos dados dos Censos Geográficos e tem por objetivo disponibilizar informações aos gestores público e privado, bem como, professores, alunos e todas as pessoas interessadas em conhecer e poder atuar em ações que nos levem a melhorar nossos referenciais de cidadania.

1.2. As variáveis

Dos 232 indicadores que compõem a base de dados disponibilizada no Atlas Brasil, por metodologia previamente definida, foram escolhidas 02 variáveis de cada uma das 07 classificações que somadas aos 04 indicadores dos IDHM, perfazem o total de 18 (dezoito) variáveis a serem abordadas, exploradas e analisadas neste trabalho.

Na busca por esses indicadores procurou-se, mesmo que de forma empírica, uma provável associação que possa nos levar a conhecer melhor a composição dos dados e, consequentemente, ter mais subsídios em ações que possam ser adotadas.

Quadro 01 – As variáveis

Código	Indicador	Abrangência	Tipo	Medida
ESPVIDA	Esperança de vida ao nascer	Número médio de anos que as pessoas deverão viver a partir do nascimento, se permanecerem constantes ao longo da vida o nível e o padrão de mortalidade por idade prevalecente no ano do Censo.	Variável Quantitativa	Anos
T_ENV	Taxa de envelhecimento	Razão entre a população de 65 anos ou mais de idade e a população total multiplicado por 100.	Variável Quantitativa	Percentual
T_FUND15A17	Percentual da população de 15 a 17 anos com fundamental completo	Razão entre a população de 15 a 17 anos de idade que concluiu o ensino fundamental, em qualquer de suas modalidades (regular seriado, não seriado, EJA ou supletivo), e o total de pessoas nesta faixa etária, multiplicada por 100.	Variável Quantitativa	Percentual

T_MED18a20	Percentual da população de 18 a 20 anos de idade com o ensino médio completo	Razão entre a população de 18 a 20 anos de idade que já concluiu o ensino médio em qualquer de suas modalidades (regular seriado, não seriado, EJA ou supletivo), e o total de pessoas nesta faixa etária, multiplicada por 100. As pessoas de 18 a 20 anos frequentando a 4ª série do ensino médio foram consideradas como já tendo concluído esse nível de ensino.	Variável Quantitativa	Percentual
RIND	Renda domiciliar per capita média dos extremamente pobres	Média da renda domiciliar *per capita* das pessoas com renda domiciliar *per capita* igual ou inferior a R$ 70,00 mensais, a preços de agosto de 2010. O universo de indivíduos é limitado àqueles que vivem em domicílios particulares permanentes.	Variável Quantitativa	Absoluto
RDPC	Renda per capita média	Razão entre o somatório da renda de todos os indivíduos residentes em domicílios particulares permanentes e o número total desses indivíduos. Valores em reais de 01/agosto de 2010.	Variável Quantitativa	Absoluto
TRABCC	Percentual de ocupados de 18 anos ou mais que são empregados com carteira	Razão entre o número de empregados de 18 anos ou mais de idade com carteira de trabalho assinada e o número total de pessoas ocupadas nessa faixa etária, multiplicada por 100.	Variável Quantitativa	Percentual
TRABSC	Percentual de ocupados de 18 anos ou mais que são empregados sem carteira	Razão entre o número de empregados de 18 anos ou mais de idade sem carteira de trabalho assinada e o número total de pessoas ocupadas nessa faixa etária, multiplicada por 100.	Variável Quantitativa	Percentual
T_AGUA	Percentual da população que vive em domicílios com água encanada	Razão entre a população que vive em domicílios particulares permanentes com água canalizada para um ou mais cômodos e a população total residente em domicílios particulares permanentes, multiplicada por 100. A água pode ser proveniente de rede geral, de poço, de nascente ou de reservatório abastecido por água das chuvas ou carro-pipa.	Variável Quantitativa	Percentual
T_LUZ	Percentual da população que vive em domicílios com energia elétrica	Razão entre a população que vive em domicílios particulares permanentes com iluminação elétrica e a população total residente em domicílios particulares permanentes, multiplicada por 100. Considera-se iluminação proveniente ou não de uma rede geral, com ou sem medidor.	Variável Quantitativa	Percentual
AGUA_ESGOTO	Percentual de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados	Razão entre as pessoas que vivem em domicílios cujo abastecimento de água não provem de rede geral e cujo esgotamento sanitário não é realizado por rede coletora de esgoto ou fossa séptica e a população total residente em domicílios particulares permanentes, multiplicada por 100. São considerados apenas os domicílios particulares permanentes.	Variável Quantitativa	Percentual
T_SLUZ	% de pessoas em domicílios sem energia elétrica	Razão entre as pessoas que vivem em domicílios sem energia elétrica e população total residente em domicílios particulares permanentes multiplicado por 100.	Variável Quantitativa	Percentual
PESORUR	População rural	População residente na área rural	Variável Quantitativa	Absoluto
PESOURB	População urbana	População residente na área urbana	Variável Quantitativa	Absoluto
IDHM	Índice de Desenvolvimento Humano Municipal	Índice de Desenvolvimento Humano Municipal. Média geométrica dos índices das dimensões Renda, Educação e Longevidade, com pesos iguais.	Variável Quantitativa	Índice

IDHM_E	Índice de Desenvolvimento Humano Municipal - Dimensão Educação	Índice sintético da dimensão Educação, é um dos 3 índices que compõem o IDHM. É obtido através da média geométrica do subíndice de frequência escolar, com peso de 2/3, e do subíndice de escolaridade, com peso de 1/3.	Variável Quantitativa	Índice
IDHM_R	Índice de Desenvolvimento Humano Municipal - Dimensão Renda	Índice da dimensão Renda, é um dos 3 índices que compõem o IDHM. É obtido a partir do indicador *Renda per capita*, através da fórmula: [ln (valor observado do indicador) - ln (valor mínimo)] / [ln (valor máximo) - ln (valor mínimo)], onde os valores mínimo e máximo são R$ 8,00 e R$ 4.033,00 (a preços de agosto de 2010).	Variável Quantitativa	Índice
IDHM_L	Índice de Desenvolvimento Humano Municipal - Dimensão Longevidade	Índice da dimensão Longevidade, é um dos 3 índices que compõem o IDHM. É obtido a partir do indicador *Esperança de vida ao nascer*, através da fórmula: [(valor observado do indicador) - (valor mínimo)] / [(valor máximo) - (valor mínimo)], onde os valores mínimo e máximo são 25 e 85 anos, respectivamente.	Variável Quantitativa	Índice
Município	Nome do Município	Nome do Município	Categórica	n/a
UF	Unidade da Federação (Estado)	Unidade da Federação (Estado)	Categórica	n/a

Fonte: Atlas Brasil, 2010

1.3 Análise De Dados

Buscando um foco direto nos indicadores que interessam a este pesquisador, sobretudo àqueles que podem ter uma ligação direta com o desenvolvimento da tese, foram selecionadas as seguintes variáveis para o estudo dos componentes principais:

Código	Indicador	Abrangência	Tipo	Medida
ESPVIDA	Esperança de vida ao nascer	Número médio de anos que as pessoas deverão viver a partir do nascimento, se permanecerem constantes ao longo da vida o nível e o padrão de mortalidade por idade prevalecentes no ano do Censo.	Variável Quantitativa	Anos
T_ENV	Taxa de envelhecimento	Razão entre a população de 65 anos ou mais de idade e a população total multiplicado por 100.	Variável Quantitativa	Percentual
T_FUND15A17	Percentual da população de 15 a 17 anos com fundamental completo	Razão entre a população de 15 a 17 anos de idade que concluiu o ensino fundamental, em qualquer de suas modalidades (regular seriado, não seriado, EJA ou supletivo), e o total de pessoas nesta faixa etária, multiplicada por 100.	Variável Quantitativa	Percentual
T_MED18a20	Percentual da população de 18 a 20 anos de idade com o ensino médio completo	Razão entre a população de 18 a 20 anos de idade que já concluiu o ensino médio em qualquer de suas modalidades (regular seriado, não seriado, EJA ou supletivo), e o total de pessoas nesta faixa etária, multiplicada por 100. As pessoas de 18 a 20 anos frequentando a 4ª série do ensino médio foram consideradas como já tendo concluído esse nível de ensino.	Variável Quantitativa	Percentual

RIND	Renda domiciliar per capita média dos extremamente pobres	Média da renda domiciliar *per capita* das pessoas com renda domiciliar *per capita* igual ou inferior a R$ 70,00 mensais, a preços de agosto de 2010. O universo de indivíduos é limitado àqueles que vivem em domicílios particulares permanentes.	Variável Quantitativa	Absoluto
RDPC	Renda per capita média	Razão entre o somatório da renda de todos os indivíduos residentes em domicílios particulares permanentes e o número total desses indivíduos. Valores em reais de 01/agosto de 2010.	Variável Quantitativa	Absoluto
TRABCC	Percentual de ocupados de 18 anos ou mais que são empregados com carteira	Razão entre o número de empregados de 18 anos ou mais de idade com carteira de trabalho assinada e o número total de pessoas ocupadas nessa faixa etária, multiplicada por 100.	Variável Quantitativa	Percentual
TRABSC	Percentual de ocupados de 18 anos ou mais que são empregados sem carteira	Razão entre o número de empregados de 18 anos ou mais de idade sem carteira de trabalho assinada e o número total de pessoas ocupadas nessa faixa etária, multiplicada por 100.	Variável Quantitativa	Percentual
T_AGUA	Percentual da população que vive em domicílios com água encanada	Razão entre a população que vive em domicílios particulares permanentes com água canalizada para um ou mais cômodos e a população total residente em domicílios particulares permanentes, multiplicada por 100. A água pode ser proveniente de rede geral, de poço, de nascente ou de reservatório abastecido por água das chuvas ou carro-pipa.	Variável Quantitativa	Percentual
T_LUZ	Percentual da população que vive em domicílios com energia elétrica	Razão entre a população que vive em domicílios particulares permanentes com iluminação elétrica e a população total residente em domicílios particulares permanentes, multiplicada por 100. Considera-se iluminação proveniente ou não de uma rede geral, com ou sem medidor.	Variável Quantitativa	Percentual
AGUA_ESGOTO	Percentual de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados	Razão entre as pessoas que vivem em domicílios cujo abastecimento de água não provem de rede geral e cujo esgotamento sanitário não é realizado por rede coletora de esgoto ou fossa séptica e a população total residente em domicílios particulares permanentes, multiplicada por 100. São considerados apenas os domicílios particulares permanentes.	Variável Quantitativa	Percentual
T_SLUZ	% de pessoas em domicílios sem energia elétrica	Razão entre as pessoas que vivem em domicílios sem energia elétrica e população total residente em domicílios particulares permanentes multiplicado por 100.	Variável Quantitativa	Percentual
PESORUR	População rural	População residente na área rural	Variável Quantitativa	Absoluto
PESOURB	População urbana	População residente na área urbana	Variável Quantitativa	Absoluto

Fonte: Atlas Brasil, 2010

1.4 Base de dados

A base de dados normalizada para a análise deste trabalho teve como referência as 14 variáveis descritas na tabela acima, sendo que, 03 delas, foram positivadas, a saber:

TRABSC	Percentual de ocupados de 18 anos ou mais que	Razão entre o número de empregados de 18 anos ou mais de idade sem carteira de trabalho assinada	Variável Quantitativa	Percentual

	são empregados sem carteira	e o número total de pessoas ocupadas nessa faixa etária, multiplicada por 100.		
AGUA_ESGOTO	Percentual de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados	Razão entre as pessoas que vivem em domicílios cujo abastecimento de água não provem de rede geral e cujo esgotamento sanitário não é realizado por rede coletora de esgoto ou fossa séptica e a população total residente em domicílios particulares permanentes, multiplicada por 100. São considerados apenas os domicílios particulares permanentes.	Variável Quantitativa	Percentual
T_SLUZ	% de pessoas em domicílios sem energia elétrica	Razão entre as pessoas que vivem em domicílios sem energia elétrica e população total residente em domicílios particulares permanentes multiplicado por 100.	Variável Quantitativa	Percentual

Fórmulas utilizadas
Passos no Minitab: CALC / CALCULADORA

Para normalização
Construir equação selecionando na caixa de opções funções a opção MINIMO e MAXIMO.

A equação: (c9-MIN(c9))/(MAX(c9)-MIN(c9))

Para positivação
Para os dados que representam valores
A Equação: 1-((c8-MIN (c8)) /(MAX(c8)-MIN(c8)))

CAPÍTULO II. ANÁLISE DISCRIMINANTE

O primeiro passo para esta análise foi resgatar a base de dados utilizada no trabalho anterior, ou seja, análise de conglomerados, e as variáveis naquele nesse estudo consideradas e suas médias por Estado, a seguir, reforçando que já foram normalizadas e no caso da agua_esgoto, positivada.

T_ENV	Taxa de envelhecimento	Razão entre a população de 65 anos ou mais de idade e a população total multiplicado por 100.	Variável Quantitativa	Percentual
T_FUND15A17	Percentual da população de 15 a 17 anos com fundamental completo	Razão entre a população de 15 a 17 anos de idade que concluiu o ensino fundamental, em qualquer de suas modalidades (regular seriado, não seriado, EJA ou supletivo), e o total de pessoas nesta faixa etária, multiplicada por 100.	Variável Quantitativa	Percentual
T_MED18a20	Percentual da população de 18 a 20 anos de idade com o ensino médio completo	Razão entre a população de 18 a 20 anos de idade que já concluiu o ensino médio em qualquer de suas modalidades (regular seriado, não seriado, EJA ou supletivo), e o total de pessoas nesta faixa etária, multiplicada por 100. As pessoas de 18 a 20 anos frequentando a 4ª série do ensino médio foram consideradas como já tendo concluído esse nível de ensino.	Variável Quantitativa	Percentual
RDPC	Renda per capita média	Razão entre o somatório da renda de todos os indivíduos residentes em domicílios particulares permanentes e o número total desses indivíduos. Valores em reais de 01/agosto de 2010.	Variável Quantitativa	Absoluto
AGUA_ESGOTO	Percentual de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados	Razão entre as pessoas que vivem em domicílios cujo abastecimento de água não provem de rede geral e cujo esgotamento sanitário não é realizado por rede coletora de esgoto ou fossa séptica e a população total residente em domicílios particulares permanentes, multiplicada por 100.	Variável Quantitativa	Percentual

		São considerados apenas os domicílios particulares permanentes.		

O Dendograma base desse estudo apresentou as características a seguir, reforçando que foram excluídos desta análise, os dados referentes ao Distrito Federal, Brasília.

Cluster Analysis of Observations: MEVsBR; MFUNsBR; MMEDsBR; MRENsBR; MAESIsBR

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	25	97,0947	0,024389	7	8	7	2
2	24	96,7259	0,027485	14	16	14	2
3	23	93,7155	0,052757	4	13	4	2
4	22	93,6106	0,053637	5	14	5	3
5	21	93,4264	0,055184	5	20	5	4
6	20	92,7966	0,060471	5	24	5	5
7	19	92,5739	0,062340	21	26	21	2
8	18	92,1748	0,065691	18	25	18	2
9	17	91,7140	0,069559	10	12	10	2
10	16	91,6655	0,069966	7	15	7	3
11	15	90,2528	0,081826	5	17	5	6
12	14	90,2159	0,082135	7	10	7	5
13	13	90,1295	0,082860	18	23	18	3
14	12	88,6761	0,095062	7	11	7	6
15	11	88,2123	0,098955	7	19	7	7
16	10	87,7740	0,102634	1	3	1	2
17	9	87,0860	0,108411	1	4	1	4
18	8	86,5656	0,112779	7	18	7	10
19	7	86,5187	0,113173	2	5	2	7
20	6	83,1901	0,141115	2	6	2	8
21	5	83,0687	0,142135	1	9	1	5
22	4	82,3811	0,147907	1	22	1	6
23	3	82,2123	0,149324	1	2	1	14
24	2	80,2797	0,165547	1	21	1	16
25	1	79,5995	0,171258	1	7	1	26

Final Partition
Number of clusters: 7

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	4	0,025541	0,076275	0,111915
Cluster2	7	0,036586	0,065816	0,124489
Cluster3	1	0,000000	0,000000	0,000000
Cluster4	10	0,139993	0,107429	0,186490
Cluster5	1	0,000000	0,000000	0,000000
Cluster6	2	0,001943	0,031170	0,031170
Cluster7	1	0,000000	0,000000	0,000000

Cluster Centroids

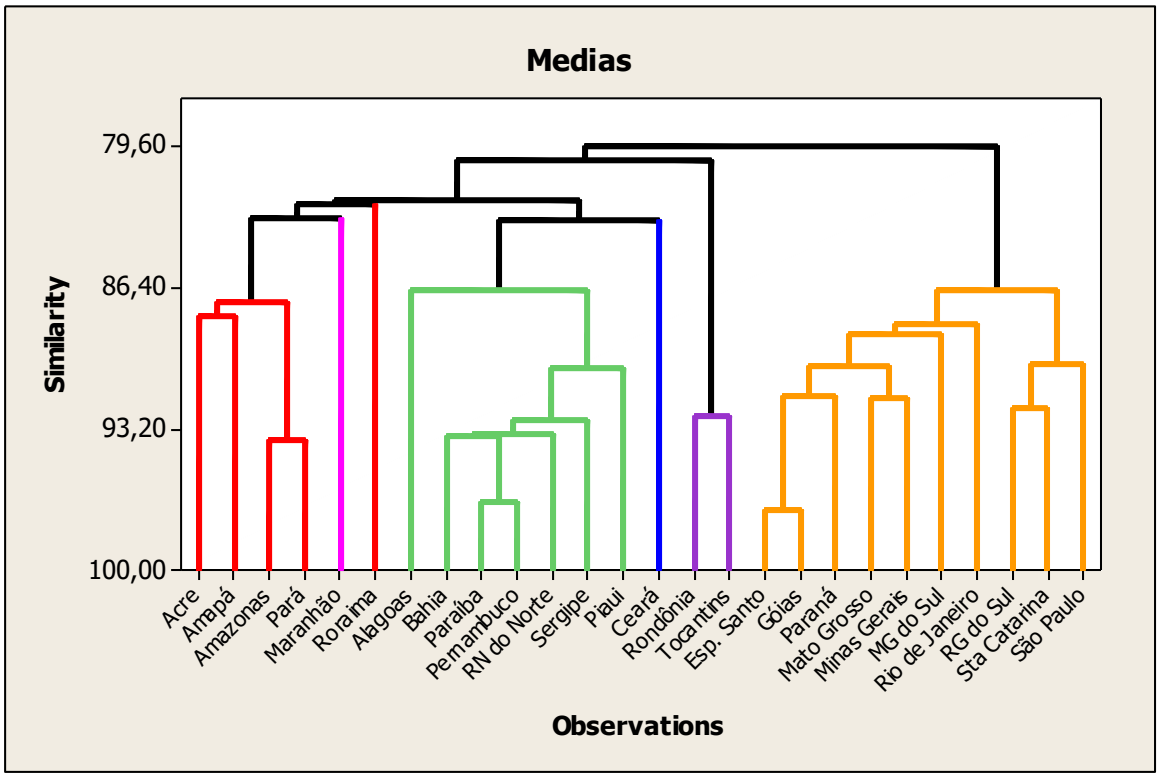
Variable	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
MEVsBR	0,453700	0,370371	0,4029	0,699930	0,30970	0,528700	0,56850
MFUNsBR	0,345750	0,357600	0,5340	0,604300	0,38880	0,494900	0,41220
MMEDsBR	0,207550	0,255100	0,3587	0,468790	0,23390	0,341600	0,29350
MRENsBR	0,114245	0,096187	0,0880	0,275395	0,06963	0,173745	0,11713
MAESIsBR	0,574200	0,805971	0,7948	0,966800	0,65250	0,871450	0,72230

Variable	Grand centroid
MEVsBR	0,528662
MFUNsBR	0,471308
MMEDsBR	0,341273
MRENsBR	0,173327
MAESIsBR	0,827658

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Cluster1	0,000000	0,251772	0,331991	0,613048	0,177215	0,371131	0,216619
Cluster2	0,251772	0,000000	0,207605	0,522604	0,171355	0,248403	0,226164
Cluster3	0,331991	0,207605	0,000000	0,412298	0,256770	0,175711	0,229378
Cluster4	0,613048	0,522604	0,412298	0,000000	0,628504	0,277290	0,411984
Cluster5	0,177215	0,171355	0,256770	0,628504	0,000000	0,359995	0,279652
Cluster6	0,371131	0,248403	0,175711	0,277290	0,359995	0,000000	0,190231
Cluster7	0,216619	0,226164	0,229378	0,411984	0,279652	0,190231	0,000000

Dendrogram



Observa-se que as similaridades entre os Estados na distribuição com os 07 clusters está por volta dos 90%. O Estado do Ceará teve o maior reflexo dessa nova análise (com a exclusão de Brasília)

Acre a Amapá
Amazonas e Pará
Paraíba e Pernambuco
Rondônia e Tocantins
Espírito Santo e Goiás
Mato Grosso e Minas Gerais
Rio Grande do Sul e Santa Catarina

Para análise discriminante, o objetivo proposto foi trabalhar com 03 clusters, cabendo ao autor a construção das novas similaridades.

O primeiro passo foi determinar com os 03 clusters principais que permaneceriam na análise, sendo que o critério estabelecido foi a escolha do que já continham o maior número de observações.

Vejamos um recorte da distribuição anterior:

Average Maximum

	Number of observations	Within cluster sum of squares	distance from centroid	distance from centroid
Cluster1	4	0,025541	0,076275	0,111915
Cluster2	7	0,036586	0,065816	0,124489
Cluster3	1	0,000000	0,000000	0,000000

Cluster4	10	0,139993	0,107429	0,186490
Cluster5	1	0,000000	0,000000	0,000000
Cluster6	2	0,001943	0,031170	0,031170
Cluster7	1	0,000000	0,000000	0,000000

Respeitado os critérios estabelecidos, os clusters escolhidos foram os clusters 01, 02 e 04.

O próximo passo é associar os clusters eliminados com os clusters escolhidos e o critério adotado foram as menores distâncias entre eles.

Clusters a ser associados

CLUSTER 3, COM 01 COMPONENTE DE OBSERVAÇÃO
CLUSTER 5, COM 01 COMPONENTE DE OBSERVAÇÃO
CLUSTER 6, COM 02 COMPONENTES DE OBSERVAÇÃO
CLUSTER 7, COM 01 COMPONENTE DE OBSERVAÇÃO

BUSCA DE PROXIMIDADE DE DISTÂNCIA

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Cluster1	0,000000	0,251772	0,331991	0,613048	0,177215	0,371131	0,216619
Cluster2	0,251772	0,000000	0,207605	0,522604	0,171355	0,248403	0,226164
Cluster3	0,331991	0,207605	0,000000	0,412298	0,256770	0,175711	0,229378
Cluster4	0,613048	0,522604	0,412298	0,000000	0,628504	0,277290	0,411984
Cluster5	0,177215	0,171355	0,256770	0,628504	0,000000	0,359995	0,279652
Cluster6	0,371131	0,248403	0,175711	0,277290	0,359995	0,000000	0,190231
Cluster7	0,216619	0,226164	0,229378	0,411984	0,279652	0,190231	0,000000

Cluster 03 tem maior proximidade, dentre os clusters considerados, com o cluster 2
Cluster 05 tem maior proximidade, dentre os clusters considerados, com o cluster 2
Cluster 06 tem maior proximidade, dentre os clusters considerados, com o cluster 2
Cluster 07 tem maior proximidade, dentre os clusters considerados, com o cluster 1

Discriminant Analysis: AnDisc versus MEVsBR; MFUNsBR; ...

Linear Method for Response: AnDisc

Predictors: MEVsBR; MFUNsBR; MMEDsBR; MRENsBR; MAESIsBR

Group	1	2	4
Count	5	11	10

Summary of classification

	True Group		
Put into Group	1	2	4
1	5	0	0
2	0	11	0
4	0	0	10
Total N	5	11	10
N correct	5	11	10
Proportion	1,000	1,000	1,000

N = 26 N Correct = 26 Proportion Correct = 1,000

Squared Distance Between Groups

	1	2	4
1	0,0000	54,0367	60,0787
2	54,0367	0,0000	26,2981
4	60,0787	26,2981	0,0000

Linear Discriminant Function for Groups

	1	2	4
Constant	-80,36	-144,66	-179,87
MEVsBR	42,73	-167,44	-82,32
MFUNsBR	158,22	219,70	204,26
MMEDsBR	-238,04	-245,92	-257,69
MRENsBR	-34,50	44,66	87,74
MAESIsBR	233,51	412,50	403,98

Percebe-se que, com a nova distribuição, o percentual de acerto (proportion= 1,000) entre os novos clusters definidos é de 100%

O próximo passo é elaborar o novo Dendograma para verificar como fica a distribuição por Estados.

CONSTRUÇÃO DO NOVO DENDOGRAMA COM 3 CLUSTERS

Cluster Analysis of Observations: MEVsBR; MFUNsBR; MMEDsBR; MRENsBR; MAESIsBR

Manhattan Distance, Complete Linkage
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	25	97,8348	0,03974	7	8	7	2
2	24	97,0307	0,05450	14	16	14	2
3	23	95,3357	0,08561	21	26	21	2
4	22	94,3512	0,10368	4	13	4	2
5	21	94,0216	0,10973	5	24	5	2
6	20	93,2446	0,12399	18	25	18	2
7	19	92,7249	0,13353	14	20	14	3
8	18	92,6235	0,13539	10	12	10	2
9	17	92,1375	0,14431	7	15	7	3
10	16	90,4295	0,17566	5	17	5	3
11	15	89,5845	0,19117	11	19	11	2
12	14	88,9628	0,20258	1	3	1	2
13	13	88,6217	0,20884	18	23	18	3
14	12	86,9148	0,24017	10	11	10	4
15	11	85,3092	0,26964	2	9	2	2
16	10	84,1514	0,29089	6	14	6	4
17	9	81,1510	0,34596	2	5	2	5
18	8	80,6410	0,35532	7	10	7	7
19	7	79,0300	0,38489	1	22	1	3
20	6	71,0798	0,53081	6	21	6	6
21	5	68,6455	0,57549	2	4	2	7
22	4	65,7121	0,62933	1	2	1	10
23	3	64,6524	0,64878	7	18	7	10
24	2	51,2343	0,89506	1	6	1	16

25	1	0,0000	1,83543	1	7	1	26

Final Partition
Number of clusters: 3

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	10	0,238199	0,152136	0,193578
Cluster2	6	0,074873	0,108755	0,139195
Cluster3	10	0,139993	0,107429	0,186490

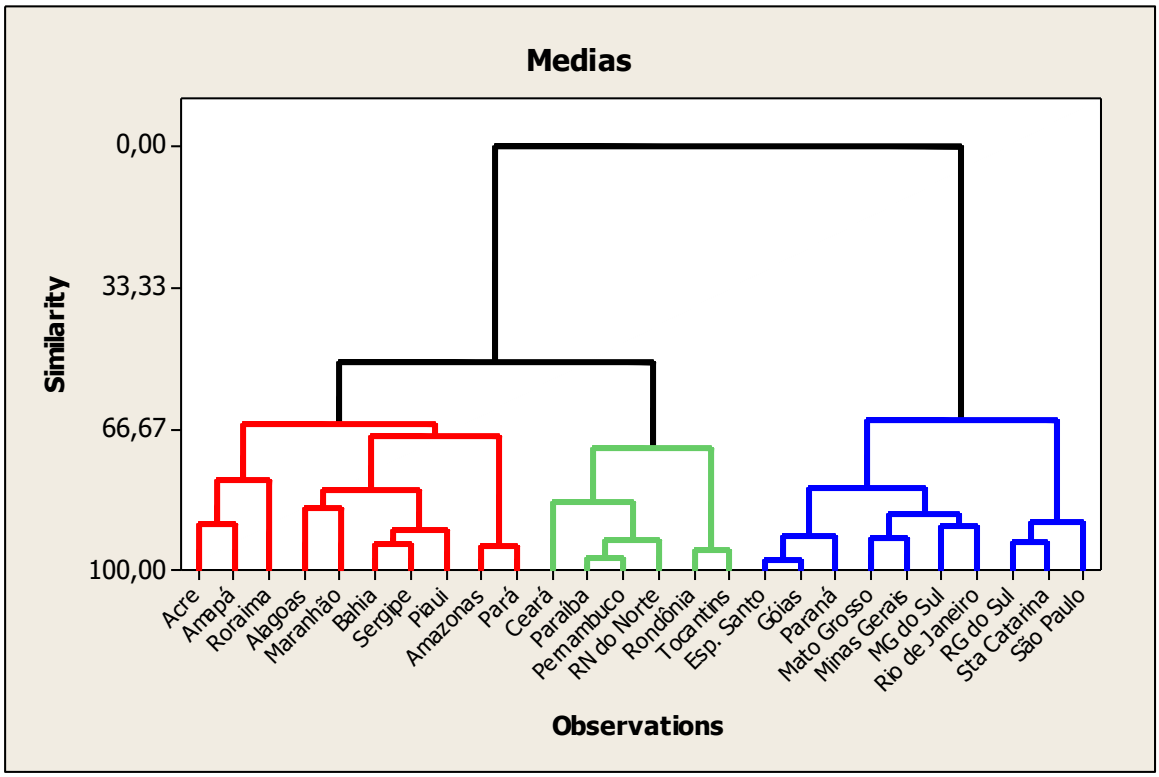
Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Grand centroid
MEVsBR	0,412120	0,437450	0,699930	0,528662
MFUNsBR	0,352590	0,447517	0,604300	0,471308
MMEDsBR	0,227490	0,318383	0,468790	0,341273
MRENsBR	0,101115	0,123565	0,275395	0,173327
MAESIsBR	0,695770	0,815567	0,966800	0,827658

Distances Between Cluster Centroids

	Cluster1	Cluster2	Cluster3
Cluster1	0,000000	0,181024	0,555202
Cluster2	0,181024	0,000000	0,402521
Cluster3	0,555202	0,402521	0,000000

Dendrogram



Os Estados, dentre as variáveis analisadas, como maior similaridade são:
Acre com Amapá; Alagoas com Maranhão; Bahia com Sergipe, Amazonas com Pará, Paraíba com Pernambuco, Rondônia com Tocantins, Espírito Santo com Goiás, Mato Grosso com Minas Gerais, Mato Grosso do Sul com Rio de Janeiro, Rio Grande do Sul com Santa Catarina.

UF ESTADOS	Cluster Observation	UF ESTADOS	Cluster Observation
Acre	1	Paraíba	2
Alagoas	1	Paraná	3
Amapá	1	Pernambuco	2
Amazonas	1	Piaui	1
Bahia	1	RG do Sul	3
Ceará	2	Rio de Janeiro	3
Esp. Santo	3	RN do Norte	2
Góias	3	Rondônia	2
Maranhão	1	Roraima	1
Mato Grosso	3	São Paulo	3
MG do Sul	3	Sergipe	1
Minas Gerais	3	Sta Catarina	3
Pará	1	Tocantins	2

CONSIDERAÇÕES FINAIS

Percebeu-se que com a nova distribuição dos clusters baseada na menor distância entre os clusters analisados, no início com 07, há mais estados em similaridades entre si, e uma boa parte delas próximas ao percentual de 80%.
Observa-se, nitidamente, que há um recorte de 03 "BRASIS" diferentes, ficando todo NORTE e NORDESTE distribuídos em 02 clusters e o terceiro ocupado, integralmente, pelas regiões SUL, SUDESTE e CENTRO-OESTE

Isto sinaliza, pelas variáveis estudadas, que as questões sociais e econômicas interferem diretamente nas condições de vida e pode demostrar uma ineficácia nas questões de políticas públicas, já que estamos falando de um único país.

Com o objetivo de subsidiar as próximas análises, a seguir a construção e demonstração pelos BOXPLOTS, das ANOVAS das variáveis estudadas, obedecidas esta nova distribuição, abertas por regiões.

Variável: Expectativa de vida (dados normalizados)

One-way ANOVA: MEVsBR versus EST REG

Source	DF	SS	MS	F	P
EST REG	4	0,51922	0,12981	40,02	0,000
Error	21	0,06811	0,00324		
Total	25	0,58734			

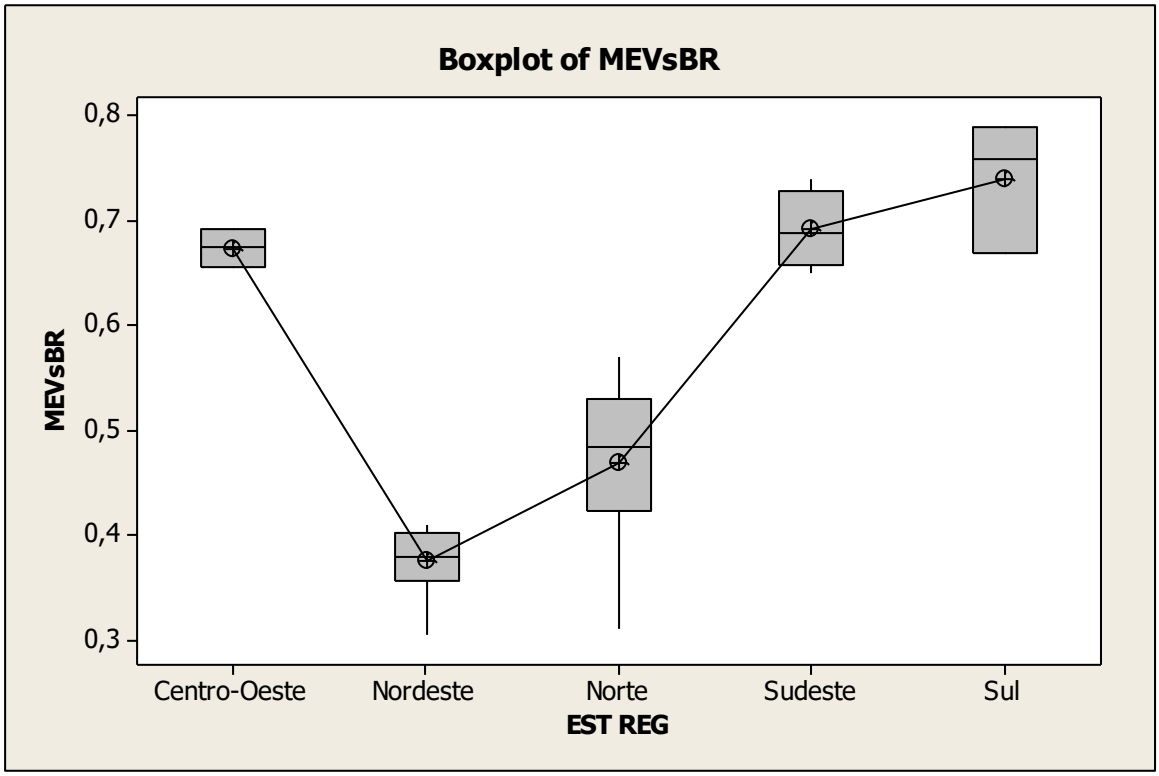
S = 0,05695 R-Sq = 88,40% R-Sq(adj) = 86,19%

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
Centro-Oeste	3	0,67313	0,01835	(-----*-----)
Nordeste	8	0,37444	0,03363	(--*---)
Norte	8	0,46880	0,08259	(--*---)
Sudeste	4	0,69105	0,03689	(----*----)
Sul	3	0,73857	0,06200	(-----*----)

```
                                  --+---------+---------+---------+-------
                                  0,36      0,48      0,60      0,72

Pooled StDev = 0,05695
```

Boxplot of MEVsBR



As relações não se alteraram em comparação com as análises anteriores, ou seja, a REGIÃO SUL com os melhores indicadores e a região NORDESTE com os piores indicadores. Não nota-se outliers nesta análise.

Variável: Taxa conclusão do ensino fundamental, de 15 a 17(dados normalizados)

One-way ANOVA: MFUNsBR versus EST REG

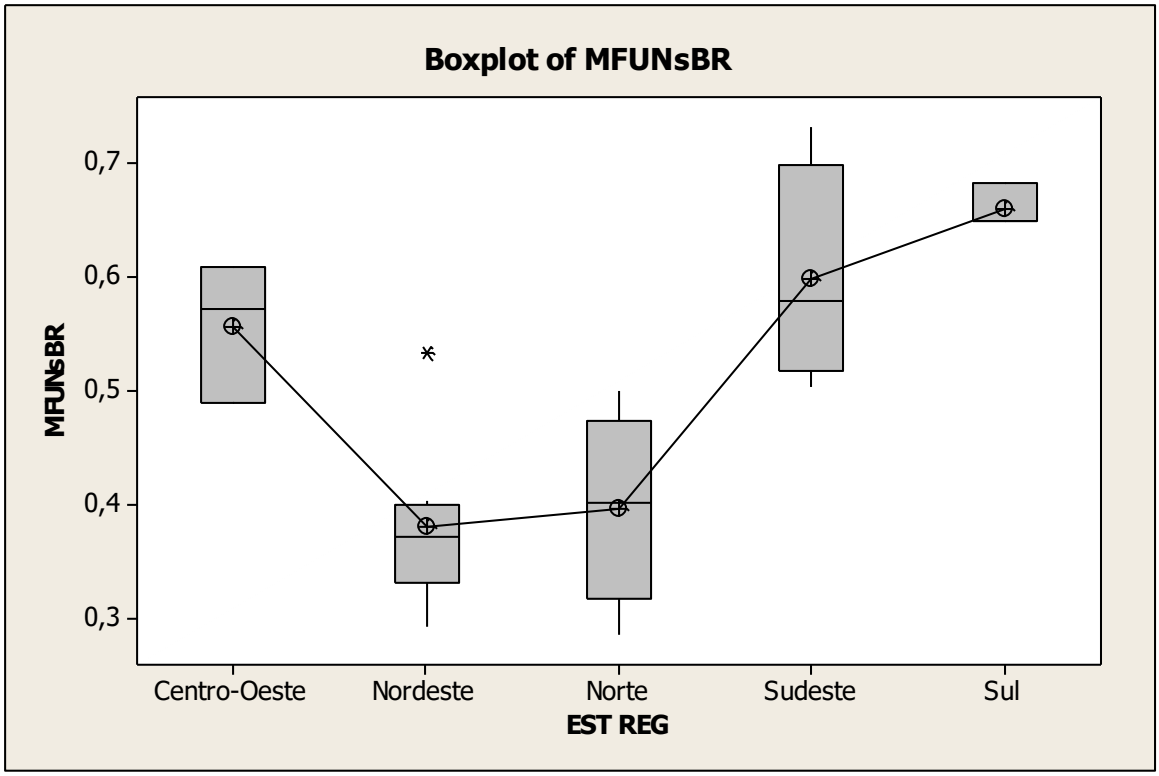
```
Source   DF       SS       MS      F      Pll
EST REG   4  0,30491  0,07623  13,89  0,000
Error    21  0,11525  0,00549
Total    25  0,42016

S = 0,07408   R-Sq = 72,57%   R-Sq(adj) = 67,35%

                                  Individual 95% CIs For Mean Based on
                                  Pooled StDev
Level         N     Mean    StDev ---+---------+---------+---------+------
Centro-Oeste  3  0,55667  0,06087             (------*-------)
Nordeste      8  0,37965  0,07172  (----*---)
Norte         8  0,39673  0,07817    (---*----)
Sudeste       4  0,59810  0,09718                  (------*-----)
Sul           3  0,66020  0,01905                       (------*------)
                                  ---+---------+---------+---------+------
                                   0,36      0,48      0,60      0,72

Pooled StDev = 0,07408
```

Boxplot of MFUNsBR



Segue a mesma distribuição da análise anterior, com a região SUL apresentando as melhores médias e a região nordeste as menores médias. Observa-se uma grande dispersão entre os municípios da região Norte, porém com a média e mediana próximas, mas no eixo inferior, ou seja, abaixo da média nacional. Essa mesma dispersão é encontrada na região SUL, mas no eixo superior, com pequeno distanciamento entre a média e a mediana.

Variável: Taxa conclusão do ensino médio, de 18 a 20(dados normalizados)

One-way ANOVA: MMEDsBR versus EST REG

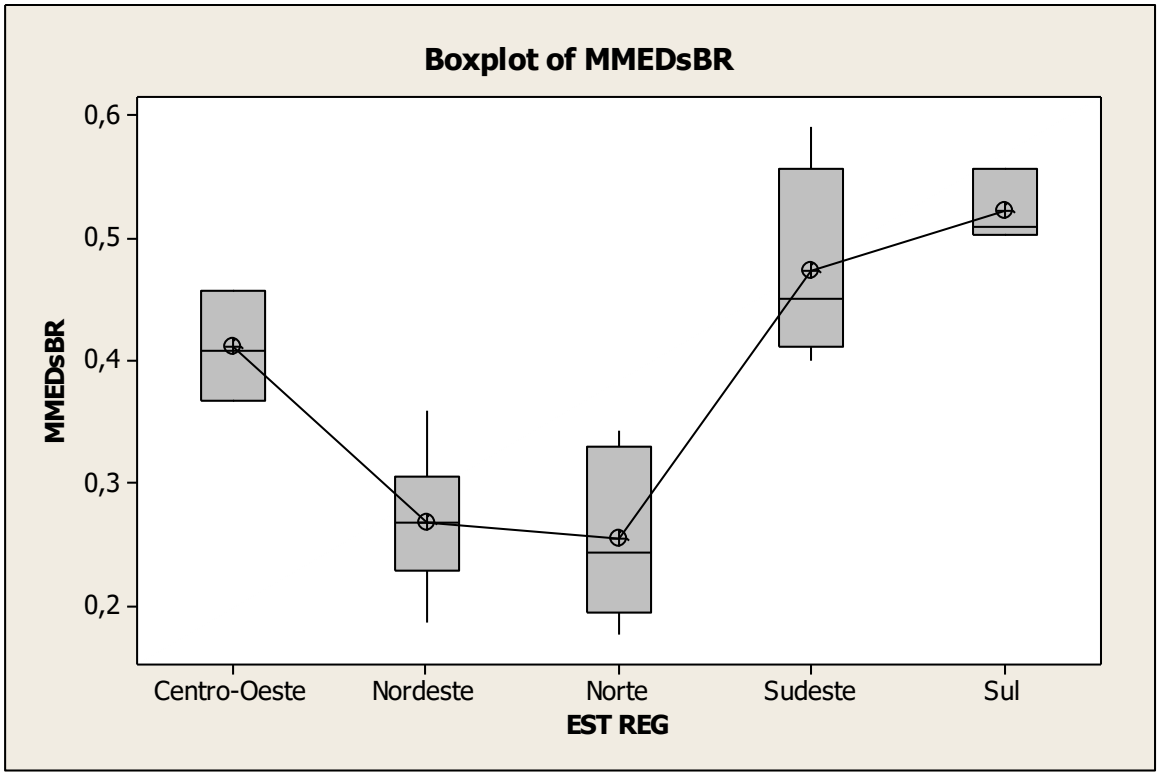
```
Source   DF       SS       MS      F      P
EST REG   4  0,28376  0,07094  19,89  0,000
Error    21  0,07491  0,00357
Total    25  0,35868

S = 0,05973   R-Sq = 79,11%   R-Sq(adj) = 75,14%

                                        Individual 95% CIs For Mean Based on
                                        Pooled StDev
Level          N     Mean    StDev  ---------+---------+---------+---------+
Centro-Oeste   3  0,41043  0,04451              (------*------)
Nordeste       8  0,26805  0,05322   (----*---)
Norte          8  0,25510  0,06468  (----*---)
Sudeste        4  0,47242  0,08197                     (-----*-----)
Sul            3  0,52230  0,02899                         (------*------)
                                    ---------+---------+---------+---------+
                                           0,30      0,40      0,50      0,60

Pooled StDev = 0,05973
```

Boxplot of MMEDsBR



Segue a mesma distribuição da análise anterior, com a região SUL apresentando as melhores médias e neste item, a região norte apresenta a menor média. Observa-se uma distribuição não uniforme na região Sul, com uma concentração muito grande dos municípios entre a mediana e o quartil superior. As outras regiões apresentam uma distribuição mais uniforme.

Variável: Renda per capita (dados normalizados)

One-way ANOVA: MRENsBR versus EST REG

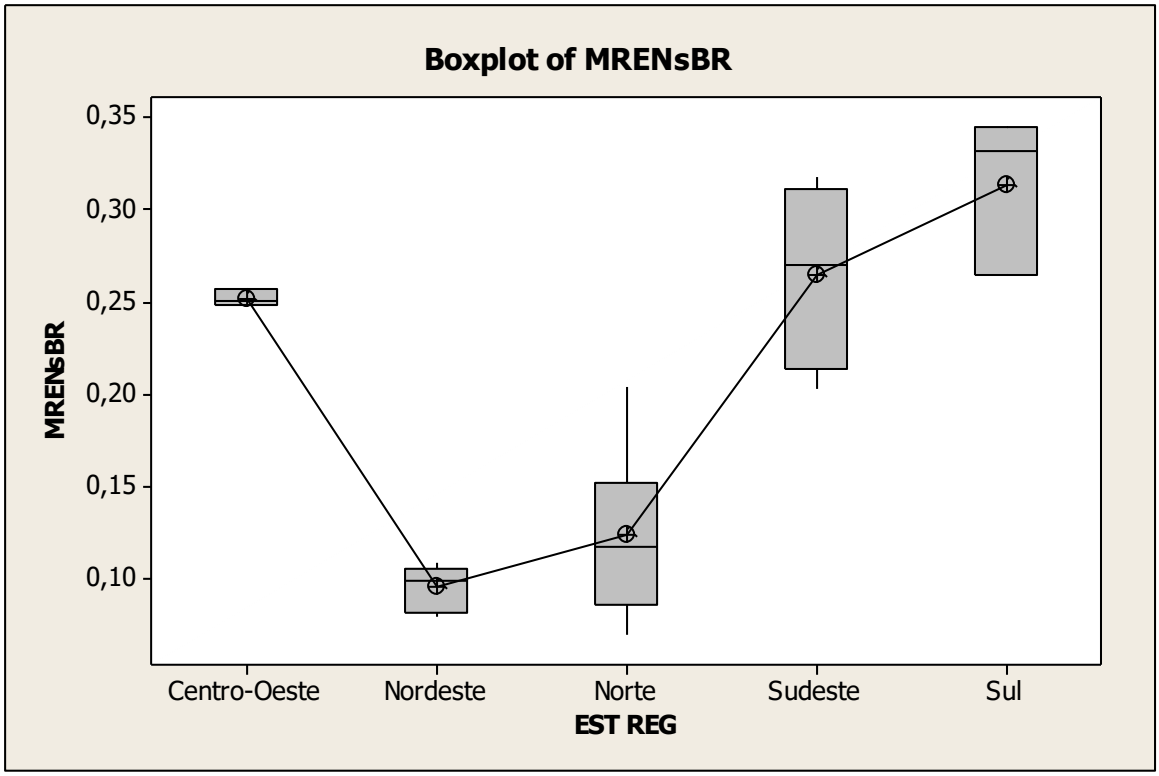
```
Source   DF       SS       MS      F      P
EST REG   4  0,17909  0,04477  36,72  0,000
Error    21  0,02560  0,00122
Total    25  0,20469

S = 0,03492   R-Sq = 87,49%   R-Sq(adj) = 85,11%

                                        Individual 95% CIs For Mean Based on
                                        Pooled StDev
Level         N     Mean    StDev  -+---------+---------+---------+--------
Centro-Oeste  3  0,25170  0,00464                   (----*-----)
Nordeste      8  0,09516  0,01203   (--*--)
Norte         8  0,12390  0,04318      (--*---)
Sudeste       4  0,26463  0,05079                      (---*----)
Sul           3  0,31344  0,04334                           (----*----)
                                     -+---------+---------+---------+--------
                                     0,080     0,160     0,240     0,320

Pooled StDev = 0,03492
```

Boxplot of MRENsBR



Nesta análise evidencia-se a diferença do poder aquisitivo, pois a variável estudada é renda per capita, entre os Estados do SUL e os Estados do Norte do Brasil. Enquanto a região SUL apresenta a melhor média nacional, mas mesmo assim percebemos uma grande concentração dos municípios dos estados dessa região situados entre a mediana e o quartil inferior, a região Nordeste apresenta a menor média nacional, com o agravante de ter a maior parte dos seus municípios situados entre a mediana e o quartil inferior.

Variável: Domicílios sem água e saneamento (dados normalizados e *positivados*)

One-way ANOVA: MAESIsBR versus EST REG

```
Source   DF       SS       MS      F      P
EST REG   4  0,38347  0,09587  15,04  0,000
Error    21  0,13382  0,00637
Total    25  0,51729

S = 0,07983   R-Sq = 74,13%   R-Sq(adj) = 69,20%

                                  Individual 95% CIs For Mean Based on
                                  Pooled StDev
Level         N    Mean   StDev  --------+---------+---------+---------+-
Centro-Oeste  3  0,9381  0,0223                    (-------*-------)
Nordeste      8  0,8046  0,0369           (----*----)
Norte         8  0,6768  0,1324  (---*----)
Sudeste       4  0,9769  0,0136                        (------*------)
Sul           3  0,9820  0,0012                        (-------*-------)
                                  --------+---------+---------+---------+-
                                        0,72      0,84      0,96      1,08
```

```
Pooled StDev = 0,0798
```

Boxplot of MAESIsBR

One-way ANOVA: MAESIsBR versus EST REG

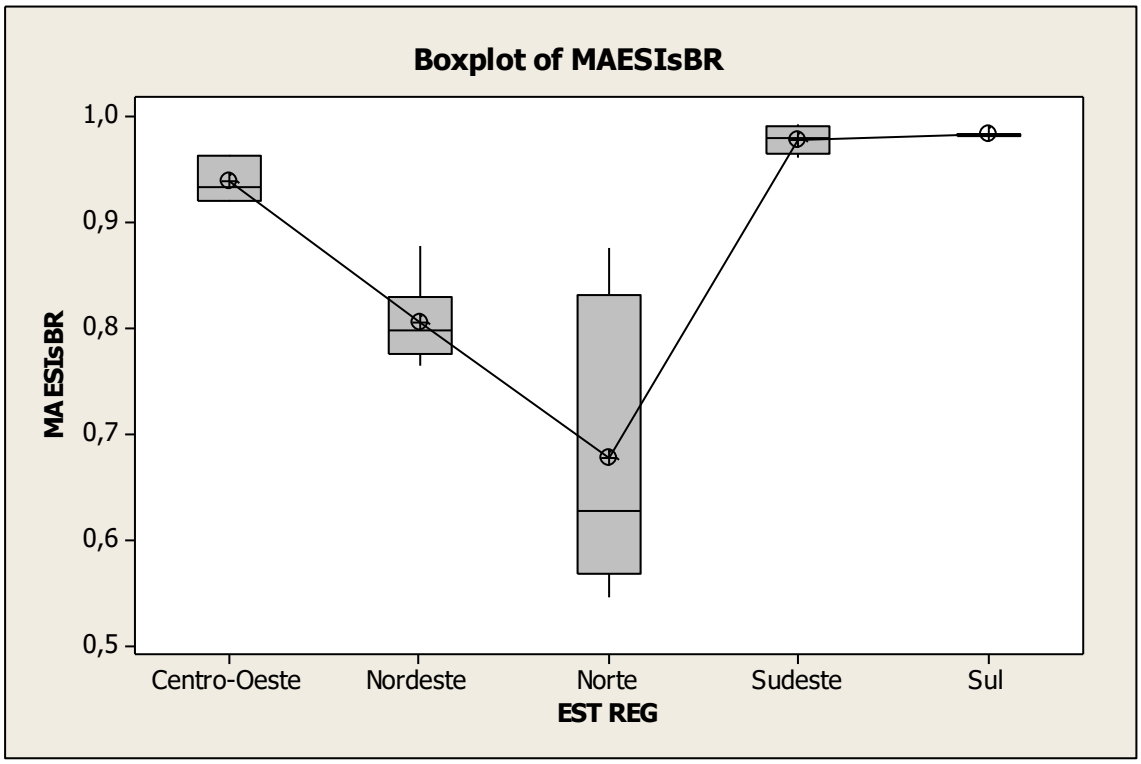
```
Source   DF       SS       MS      F      P
EST REG   4  0,38347  0,09587  15,04  0,000
Error    21  0,13382  0,00637
Total    25  0,51729

S = 0,07983   R-Sq = 74,13%   R-Sq(adj) = 69,20%

                                  Individual 95% CIs For Mean Based on
                                  Pooled StDev
Level          N    Mean   StDev  --------+---------+---------+---------+-
Centro-Oeste   3  0,9381  0,0223                    (-------*-------)
Nordeste       8  0,8046  0,0369           (----*----)
Norte          8  0,6768  0,1324  (---*----)
Sudeste        4  0,9769  0,0136                        (------*------)
Sul            3  0,9820  0,0012                        (-------*-------)
                                  --------+---------+---------+---------+-
                                        0,72      0,84      0,96      1,08

Pooled StDev = 0,0798
```

Boxplot of MAESIsBR



Praticamente essa condição inexiste nos municípios da região SUL, o que seria excelente se acontecesse em todo o Brasil. Mas não é isso que observamos na análise. Há um percentual baixo nos municípios que compõem as regiões Sudeste e Centro-Oeste, porém alto nas regiões Nordeste e, principalmente na região NORTE e esta

ainda apresenta uma distribuição não uniforme observadas a média, a mediana e a distribuição entre os quartis.

REFERÊNCIAS

ANDERSON, David R.; SWEENEY, Dennis J.; WILLIAMS, Thomas A. Estatística aplicada à administração e economia. 2. ed. São Paulo: Thomson Learning, 2007.

ATLAS DO DESENVOLVIMENTO HUMANO NO BRASIL. Disponível em: <http://www.atlasbrasil.org.br/2013/>. Acessado em: 17 mar. 2017.

IBGE, Instituto Brasileiro de Geografia e Estatística. Séries Históricas e Estatísticas. Disponível em: <http://seriesestatisticas.ibge.gov.br/apresentacao.aspx>. Acessado em: 30 mar. 2017.

LAS CASAS A., DE HOYOS A. Pesquisa de Marketing