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 trabalho, educação, saúde e muito particularmente HABITAÇÃO

MÉTODOS QUANTITATIVOS NA PESQUISA EMPÍRICA

Professor: Dr. Arnoldo Jose de Hoyos

HANNAH DE CARVALHO

CAP I ANÁLISE DE CONGLOMERADOS

1. INTRODUÇÃO

O presente trabalho tem por objetivo efetuar uma análise exploratória dos dados de Habitação incluindo as variáveis de água encanada, esgotamento sanitário, coleta de lixo, energia elétrica, domicilio próprio e densidade por dormitório, que compõe o ISDM- Indice de desenvolvimento dos municípios brasileiros.
Para tal, iniciamos com o entendimento dos dados, incluindo a definição dos indivíduos e das variáveis, suas classificações em variáveis categóricas ou quantitativas, os significados e unidades de medida, além da apresentação da tabela de dados. Na seqüência, analisamos cada uma das variáveis separadamente quanto a sua forma de distribuição, os valores atípicos, medidas de centro e dispersão. Para tal contamos com o auxílio de gráficos (*pie chart*, barras, histogramas, gráficos de ramos, box-plot, dot-plot e curvas de densidade) e de medidas numéricas (média, mediana, quartis, desvio-padrão, variância, intervalo de confiança e teste de normalidade de Anderson-Darling). No final, buscamos comparar as análises efetuadas para cada variável. O software estatístico utilizado é o **MINITAB 14.**

2. ENTENDENDO OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são os municípios brasileiros, dados referentes ao ano de 2010. Trata-se de um total de 5565 munícipios distribuídos em 27 unidades federativas, sendo 26 estados e um distrito federal. Os dados analisados de cada munícipio são as variáveis descritas abaixo.

2.2 As Variáveis

São 13 as variáveis desta pesquisa, incluindo o os três principais índices sintéticos; ISDM, IFGF e IFDM são melhor explicadas na Tabela 1. Ressaltamos que todos os dados desta pesquisa são referentes ao ano de 2010.

Tabela 1. As Variáveis

Variável	Significado	Tipo	Unidade de Medida
UF	Abreviação de Unidade Federativa (ou Unidade da Federação) do Brasil. As UF do Brasil são entidades autônomas, com governo e constituição próprias, que em seu conjunto constituem a República Federativa do Brasil. (IBGE, 2013)	Variável Categórica	N/A
Município	O município é a divisão administrativa autônoma da UF. São as unidades de menor hierarquia dentro da organização político administrativa do Brasil, criadas através de leis ordinárias das Assembléias Legislativas de cada Unidade da Federação e sancionadas pelo Governador. (IBGE, 2013)	Variável Categórica	N/A

UF2	Apresenta a sigla que representa as Unidades Federativas (ou Unidades da Federação) do Brasil.	Variável Categórica	N/A
H- Habitação	Indicador do ISDM composto por H1, H2, H3, H4, H5, H6.	Variável Quantitátiva	Percentual
H1- Água Encanada	Proporção de pessoas que vivem em domicilio com acesso à água canalizada em pelo menos um cômodo.	Variável Quantitátiva	Percentual
H2- Esgotamento Sanitário	Proporção de pessoas que vivem em domicilio com esgotamento sanitário do tipo rede geral ou esgoto pluvial.	Variável Quantitátiva	Percentual
H3- Coleta de Lixo	Proporção de pessoas que vivem em domicilio atendido por coleta de lixo (realizada por serviço de limpeza, ou cujo lixo é colocado em caçamba de serviço de limpeza).	Variável Quantitátiva	Percentual
H4- Energia Elétrica	Proporção de pessoas que vivem em domicilio que tem acesso à energia elétrica provida por companhia distribuidora.	Variável Quantitátiva	Percentual
H5- Domicilio Próprio	Proporção de pessoas que vivem em domicilio próprio de algum morador (Já pago ou ainda pagando).	Variável Quantitátiva	Percentual
H6- Densidade por Dormitório	Percentual de pessoas que vivem em domicilio que tem densidade de moradores por dormitório inferior à dois.	Variável Quantitátiva	Percentual
ISDM	Indicador Social de Desenvolvimento dos Municipios, calculado pelo Centro de Economia Aplicada da Fundação Getulio Vargas (C-Micro-FGV)- pretende contribuir para o debate de políticas publicas brasileira fornecendo uma medida sintética de bem-estar dos municípios que considere algumas de suas caracteristicas importantes relacionadas à dimensão de Renda, Habitação, Educação, Trabalho, Saude e Segurança.	Variável Quantitátiva	Percentual
IFDM	Indice Firjan de Desenvolvimento Municipal é um estudo anual que acompanha o desenvolvimento dos 5565 municipios do Brasil em três áreas: Emprego e Renda, Educação e Saúde, variando de 0à 1, sendo que quanto mais próximo de 1, maior é o desenvolvimento da localidade.	Variável Quantitátiva	0-1 Proporção
IFGF	Indice Firjan de Gestão Fiscal, para estimular a cultura de responsabilidade administrativa para aperfeiçoamento das decisões quanto à alocação de recursos públicos afim de contribuir com uma gestão eficiente e democrática e maior controle social da gestão fiscal dos municípios. Indicadores: Receita própria, pessoal, investimentos, liquidez e custo da divida.	Variável Quantitátiva	0-1 Proporção

2.3 A Tabela de Dados

Tabela 2. Tabela de Dados

Tabela de Dados- 2010												
UF	**Município**	**UF2**	**ISDM**	**H**	**H1**	**H2**	**H3**	**H4**	**H5**	**H6**	**IFGF**	**IFDM**
Acre	Acrelândia	AC	3.37	3.15	47.53	94	68.34	0	78.73	39.74	0.57	0.6108
Acre	Assis	AC	2.91	2.93	58.82	81.38	52.07	2.21	85.22	34.09	0.44	0.5459
Acre	Brasiléia	AC	3.5	3.58	65.47	89.59	64.98	19.71	79.96	37.62	0.55	0.5772
Acre	Bujari	AC	3.37	2.72	46.77	91.87	56.26	0.22	72.55	31.87	0.46	0.5402
Acre	Capixaba	AC	3.05	2.49	43.39	92	36.8	0.96	78.19	32.78	0.28	0.5291
Acre	Cruzeiro	AC	3.71	3.3	65.05	91.72	48.95	4.26	91.07	37.56	0.66	0.58
Acre	Epitaciolândia	AC	3.8	3.35	66.59	92.18	62.04	4.03	78.64	38.1	0.66	0.5472
Acre	Feijó	AC	2.09	2.03	46.06	62.4	31.31	3.23	86.95	23.42	0.43	0.4929
Acre	Jordão	AC	1.42	1.17	34.72	32.39	15.54	0.14	94.24	18.8	0.54	0.3941
Acre	Mâncio	AC	2.82	2.71	49.52	83.11	39.13	0.28	95.81	31.17	0.64	0.5393
Acre	Manoel	AC	2.51	2.39	49.29	73.26	38.72	6.72	81.01	28.39	0.59	0.5075
Acre	Marechal	AC	1.44	1.28	30.78	33.34	18.65	1.14	97.19	22.24	0.31	0.471
Acre	Plácido	AC	3.61	3.18	54.4	98.37	58.84	4.01	77.97	36.62	0.15	0.5682
Acre	Porto	AC	3.24	2.85	48.56	95.85	53.11	0.13	79.87	30.73	0.27	0.5418
Acre	Porto	AC	1.53	1.27	18.55	47.28	21.74	0	95.07	19.96	0.59	0.4641
Acre	Rio	AC	4.92	4.59	92.59	98.23	68.25	45.47	81.19	39.11	0.72	0.7691
Acre	Rodrigues	AC	1.94	2.16	36.77	87.02	17.13	0	95	26.5	0.48	0.5365
Acre	Santa	AC	0.99	1.38	36.87	34.22	31.98	0.06	93.35	13.74	0.34	0.4585
Acre	Sena	AC	3.15	2.84	61.36	79.28	44.08	6.35	86.74	30.03	0.48	0.5632
Acre	Senador	AC	4.35	3.25	63.52	98.81	63.42	0.22	76.1	33.13	0.35	0.5828
Acre	Tarauacá	AC	2.26	2.11	46.06	66.71	33	1.2	90.44	21.76	0.49	0.4633
Acre	Xapuri	AC	3.49	3.38	59.93	80.18	64.59	17.66	84.64	36.71	0.52	0.5277
Alagoas	Água	AL	2.62	3.26	33.46	95.95	47.79	31.31	87.17	40.01	0.33	0.5073
Alagoas	Anadia	AL	3.51	4.17	79.73	99.41	82.58	8.89	78.82	47.28	0.26	0.5433
Alagoas	Arapiraca	AL	4.23	4.34	90.33	99.21	85.24	10.87	75.14	47.63	0.58	0.6749
Alagoas	Atalaia	AL	3.42	3.75	75.42	97.51	73.11	16.87	67.86	37.78	0.28	0.7449
Alagoas	Barra	AL	3.52	4.07	87.14	96.61	92.21	16.09	66.72	30.99	0.47	0.5438
Alagoas	Barra	AL	4.3	4.17	94.41	99.02	95.31	17.25	57.07	33.82	0.83	0.5975
Alagoas	Batalha	AL	3.53	3.64	62.06	98.46	63.31	17.9	83.67	35.72	0.26	0.5037
Alagoas	Belém	AL	3.3	3.49	55.52	97.93	56.27	6.53	87.15	48.1	0.43	0.5155
Alagoas	Belo	AL	2.09	2.71	27.99	97.54	41.52	6.76	87.83	37.38	0.35	0.5485
Alagoas	Boca	AL	4.12	4.86	85.68	98.96	82.58	62.1	68.6	44.94	0.49	0.6326
Alagoas	Branquinha	AL	3.56	3.97	60.9	98.08	72.16	36.63	76.51	39.15	0.45	0.5502
Alagoas	Cacimbinhas	AL	2.59	2.74	53.21	97.47	24.32	0	80.42	45.82	0.49	0.4708
Alagoas	Cajueiro	AL	3.54	3.78	76.57	98.9	82.12	5.07	72.95	34.93	0.42	0.5396
Alagoas	Campestre	AL	3.82	4.48	73.54	98.94	78.29	61.8	63.67	40.93	0.15	0.5737

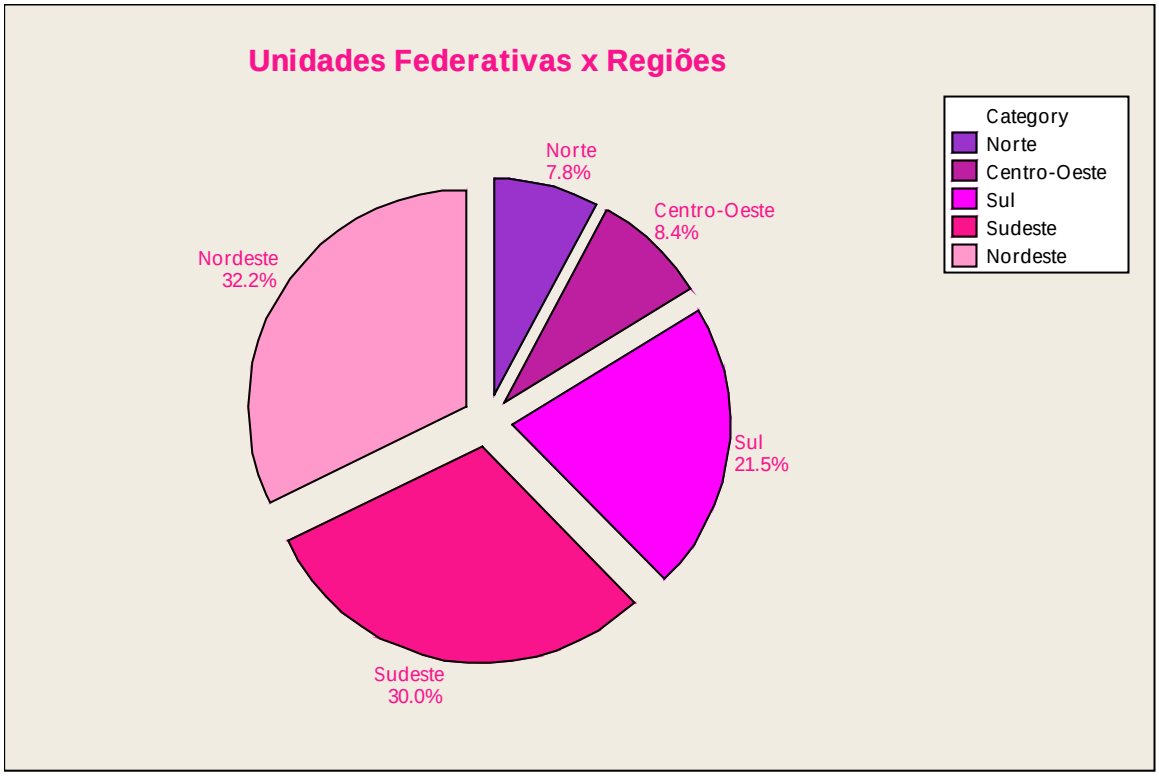
3. ANÁLISE DAS VARIÁVEIS

3.1 Variáveis Categóricas ou qualitativas.

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: “UF” e “UF2”

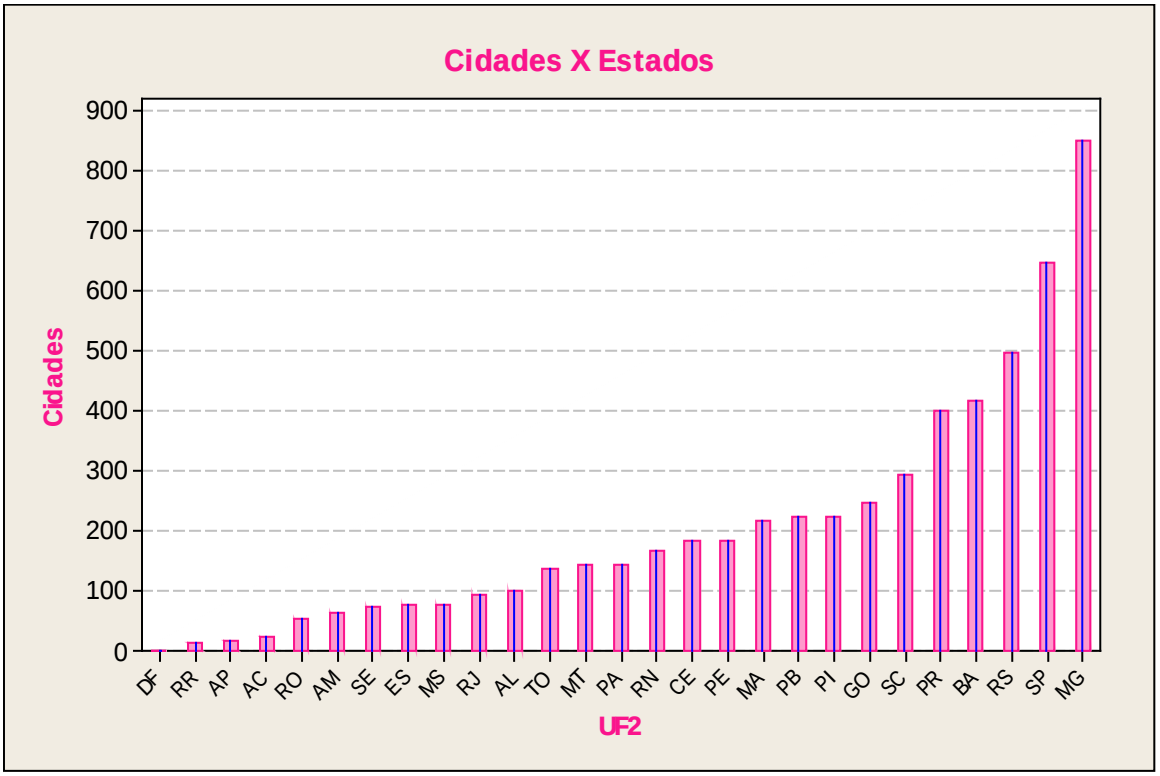
Nossa amostra totaliza 26 unidades federativas e 1 distrito federal. As unidades federativas estão distribuídas em 5 regiões.

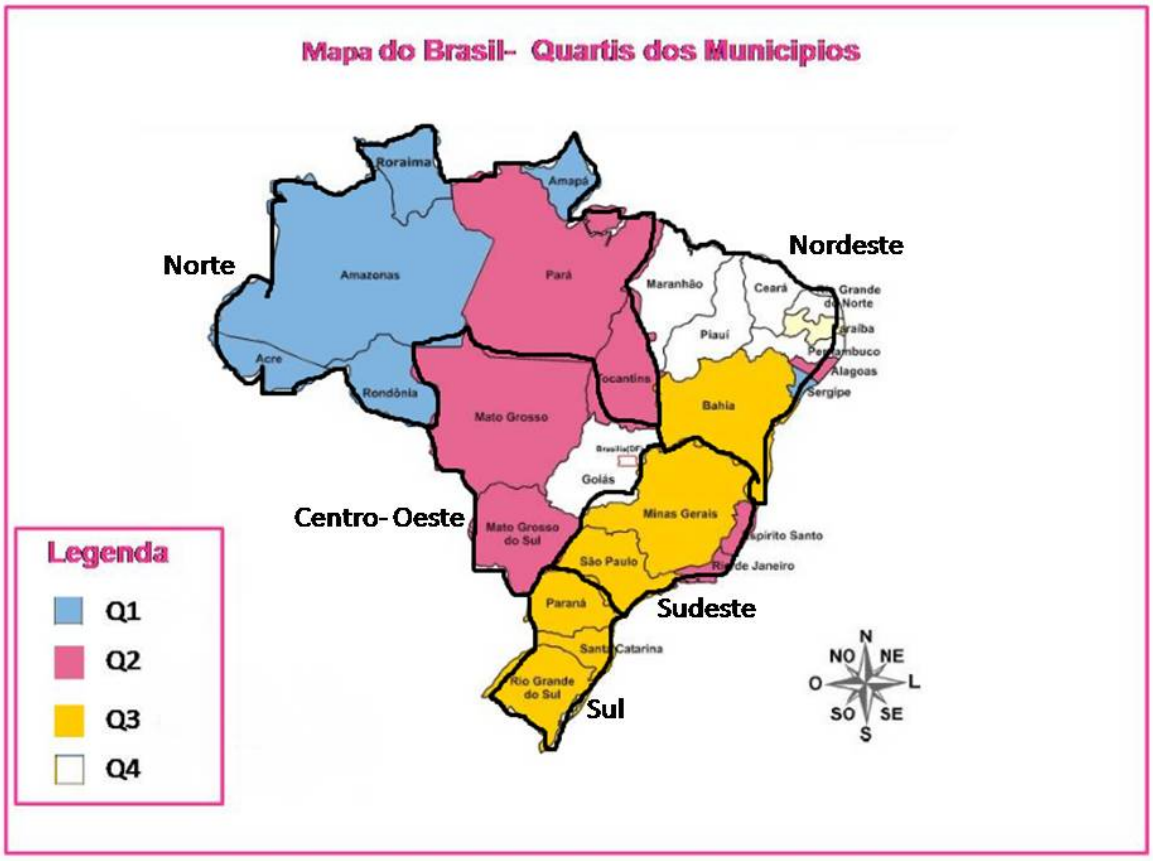


No que diz respeito a relação regiões e cidades pode-se observar no gráfico acima que as regiões Nordeste (32,2%), Sudeste (30,0%) e Sul (21,5%) concentram 83, 7% dos municípios do território nacional, enquanto as demais regiões, Norte (7.8%) e Centro-Oeste (8,4%) somam apenas 16, 2% dos munícipios. Além da concentração dos municípios brasileiros, as três regiões tem em comum o fato de serem as três regiões banhadas significativamente pelo oceano Atlântico. Fato este, que nos ajuda a entender a concentração nestas regiões.

3.1.2 Variável: “Munícipios”

Os gráficos abaixo nos ajudam a entender melhor o comportamento desta variável





Análise:

- O comportamento dos municipios por Unidades Federativas (UF2) não consiste em igualdade conforme demonstra os gráficos acima, pois enquanto o estado de Minas Gerais que contém a maior quantidade de municípios brasileiros tem 851 cidades que correspondem à 15,3 % , Roraima tem apenas 15 municipios que é correspondente à 0,3%.
Portanto Minas Gerais tem 57 vezes mais municípios que Roraima.
A distância aumenta ao considerarmos o Distrito Federal que tem somente uma cidade.

- O Primeiro e o segundo quartil concentram-se nas regiões Norte e Centro-Oeste, de maneira que tem somente dois estados no Sudeste: Rio de Janeiro e Espirito Santo e no Nordeste apenas: Alagoas e Sergipe, exclui-se deste contexto Goiás que corresponde ao quarto quartil Portanto podemos afirmar que nestas regiões concentram-se os estados com menor quantidade de municípios que totalizam 1.015, ou seja, as Regiões Norte e Centro-Oeste somadas aos quatro estados descritos acima correspondem 18% do total de municípios brasileiros.

- No terceiro Quartil os estados possuem a quantidade de municípios entre 167 e 223 concentrados na Região Sul e Sudeste, incluindo a Bahia que pertence à região Nordeste , exclui-se deste contexto Rio de Janeiro e Espirito Santo.
Este quartil é composto por 1.198 municipios que correspondem à 22% do total de municípios brasileiros.

-No ultimo Quartil visualizamos os estados que possuem as maiores quantidades de municípios, com forte concentração na região Nordeste, excluindo-se destes os estados da Bahia, Alagoas e Sergipe e incluimos Goias correspondente à região centro-oeste.
Deste total temos 3.352 municipios que correspondem à 60% do total de municípios brasileiros., portanto a Região Nordeste é composta pelos estados que mais contém municípios.

.

3.2 VARIÁVEIS QUANTITATIVAS

A análise deste tipo de variável permite a utilização de uma maior gama de ferramentas de análise como histogramas, curvas de densidade, gráfico de ramos, box-plot e dot-plot, além de informações numéricas como média, desvio-padrão, mediana, quartis, 5 números, intervalo de confiança e teste de normalidade de Anderson-Darling.

3.2.1. DENDOGRAMA DE ISDM POR ESTADO (-DF)

O Dendograma permite uma análise do grau de similaridade dos dados para geramos o Dendrograma de ISDM Estado.

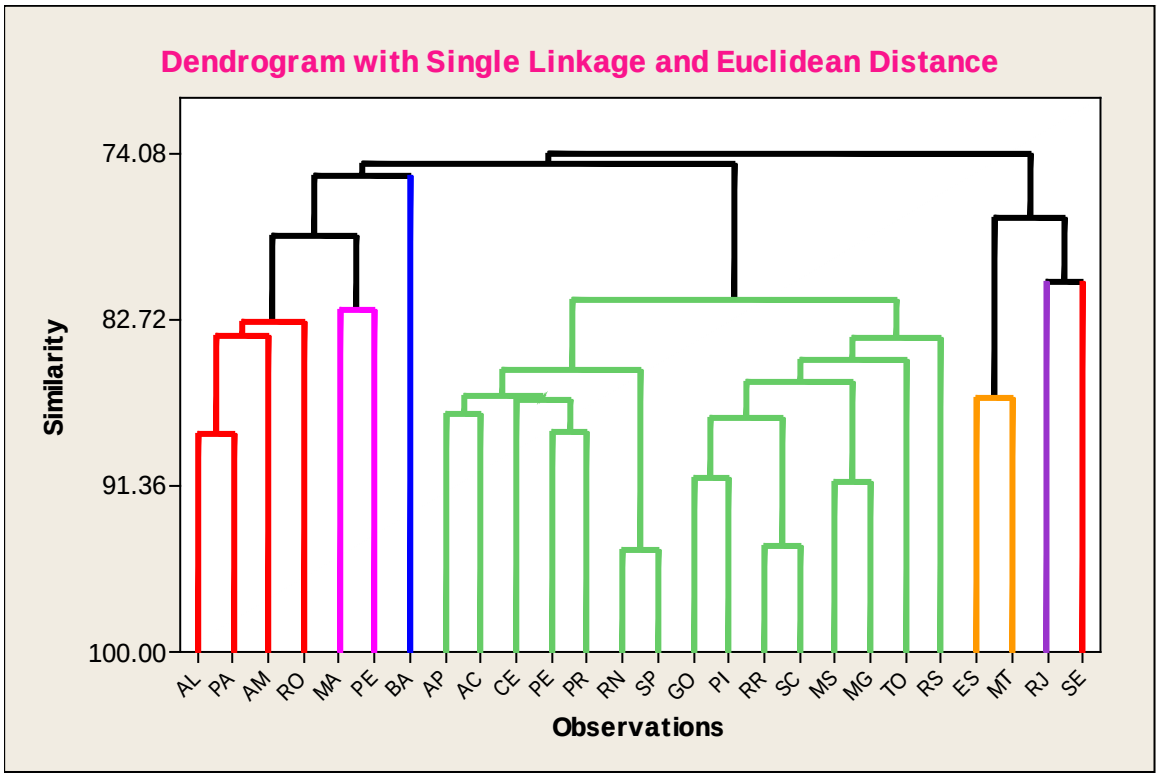


Figura 2. Dendograma da variável ISDM por estados do Brasil (classificação não supervisionada)

Na figura acima podemos verificar três grandes grupos de variáveis, agrupadas pela similaridade dos dados. O nível de similaridade dos dados destes estados está entre 74,08 a 82.72%, conforme indicado na escala apresentada no eixo Y do gráfico.

Os 2 grandes agrupamento de dados, compostos pelos grupos de 7 a 13 estados do Brasil, além de sete estados que ficaram isolados por não terem seus dados em similaridade com os outros estados. Estes estados isolados são: SE, RJ, MT, ES, BA. PE e MA , onde PE e MA apresentam certa similaridade, assim como ES e MT.

Na classificação não supervisionada não se tem informações prévias sobre estes grupos. Não se tem informações sobre os por quês ou os critérios de agrupamento utilizados neste agrupamento.

Podemos observar estados com alto nível de similaridade o que significa que a desigualdade é baixa. O menor nível de desigualdade se encontra nos estados mais próximos do eixo X, por exemplo, RR e SC, que tem um nível de similaridade próximo de 94,47%.

Figura 3. Mapa do Brasil representando os três grupos de estados por similaridades de ISDM (classificação não supervisionada)

Quando o nível de desigualdade é baixo poderíamos erroneamente dizer que a situação é boa. **Isso não é verdade.** Baixa desigualdade não significa que as coisas vão bem, e sim que existe um padrão nos municípios do estado em termos de ISDM, uma maior similaridade entre estes municípios, e não é possível responder se esta similaridade é boa ou não.

Na figura acima (figura 3) podemos observar o agrupamento por similaridade de ISDM do dendrograma no mapa político do Brasil.

3.2.2. ANÁLISE DAS VARIANCIAS DE HABITAÇÃO E ISDM POR UF (– DF)

A análise das variâncias permite a verificação e visualização das médias e desvios padrões da variável a ser analisada. O gráfico BOXPLOT ilustra os agrupamentos, o seu tamanho (largura) varia de acordo com a quantidade de dados de cada grupo e amplitude dos dados (comprimento), e também é possível visualizar as ocorrências de *outliers* (marcas fora das caixas) dentro de um grupo de dados.

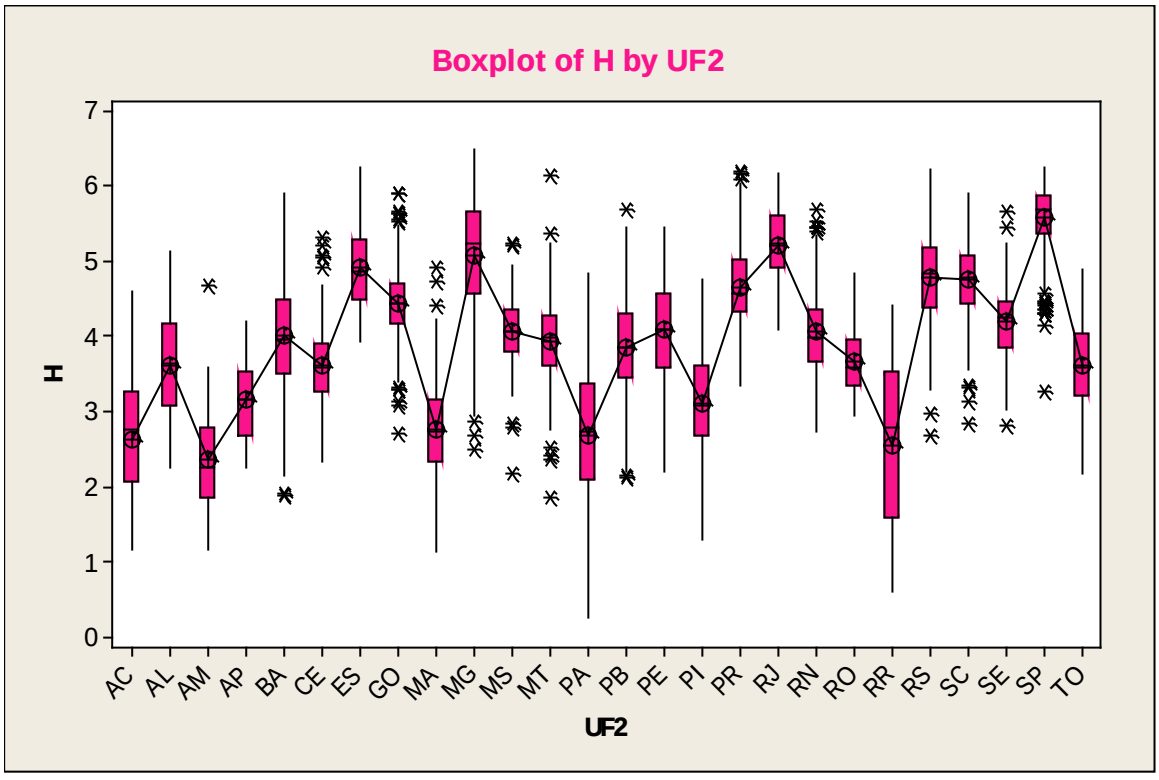
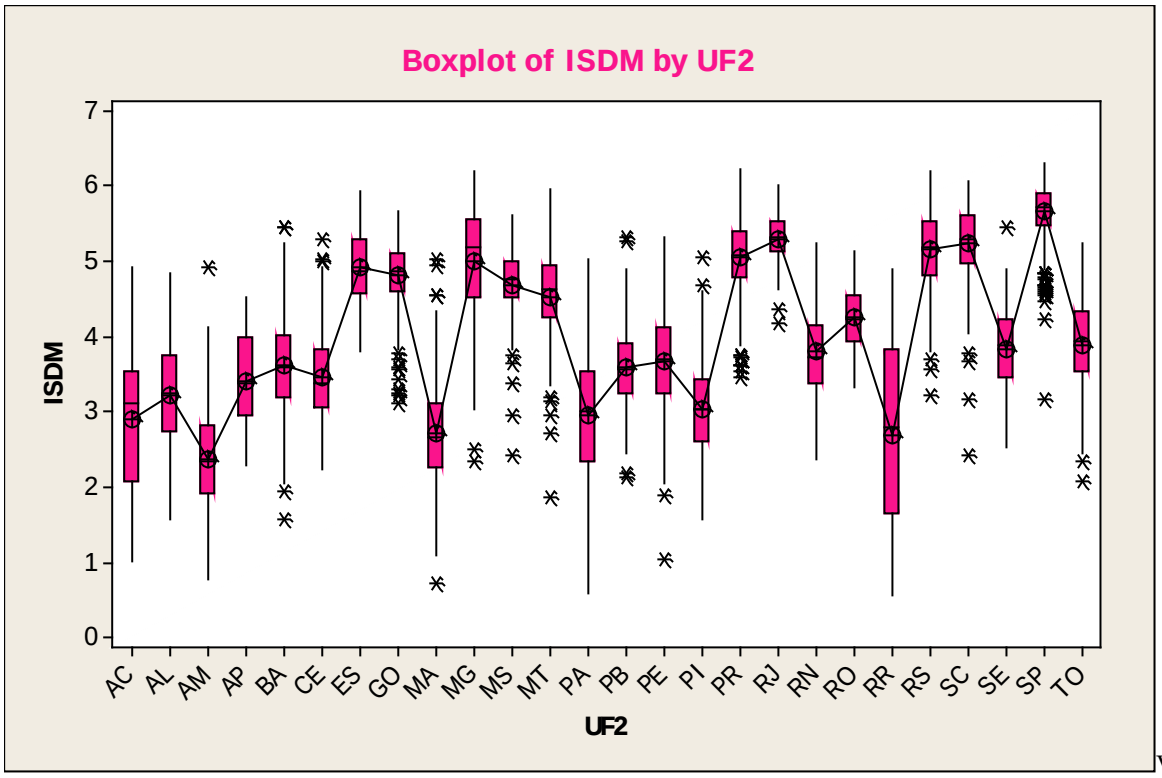


Figura 4. Gráfico BOXPLOT de HABITAÇÃO por Unidade Federativa

Podemos visualizar no gráfico da figura 4, uma grande variabilidade sobre as médias de HABITAÇÃO por unidades federativas. A UF que apresenta maior variabilidade dos dados é PA. E SP apresenta uma baixa variabilidade dos dados de HABITAÇÃO, embora tenha muitos *outliers* (o maior de todos) que são os dados muito distantes das médias.

O resultado deste comando não fica armazenado na base de dados, é necessário copiar da área *session* para a área *worksheet*, para cada variável gerada. Com isso temos os dados dos 5565 municípios do Brasil, resumidos pela média e pelo desvio padrão. A partir destes dados resumidos, fica mais fácil trabalhar os dados, uma vez que estando resumido se torna mais simples a sua manipulação e análise.



V

Figura 5. Gráfico BOXPLOT de ISDM por Unidade Federativa

Podemos verificar na figura 5 que existe uma variação grande entre as médias das UFs do Brasil, no que diz respeito à ISDM. O tamanho das caixas de cada estado representa a variância dos dados de ISDM de cada UF, e os sinais * representam o *outliers* ou pontos fora da curva, que são dados ou muito acima ou abaixo da média dos dados do estado. O estado que apresenta a maior média de ISDM é também o SP (acima de 5,6459), e o estado que apresenta a menor média é RR, pouco acima de 2,6680.

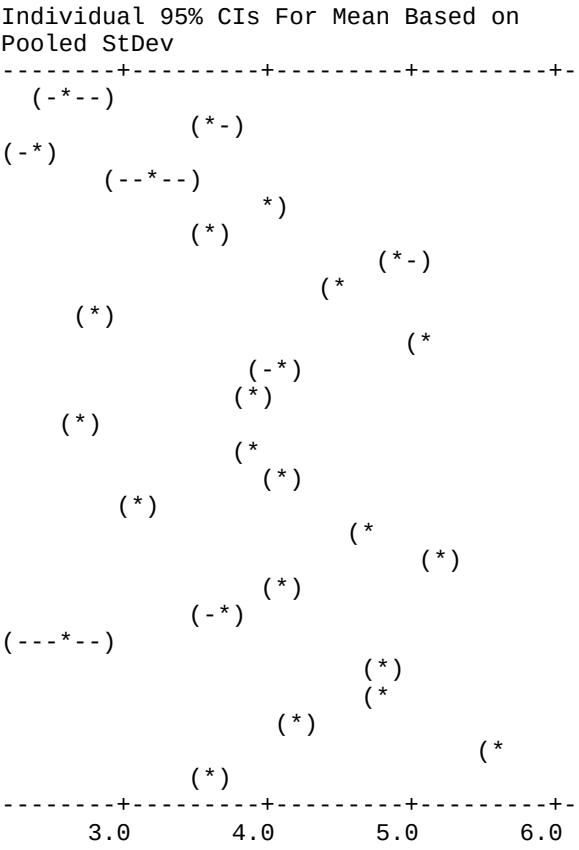
Abaixo podemos visualizar os dados descritivos gerados pelo comando, para a variável H e, na sequência, pelo ISDM.

One-way ANOVA: H versus UF2

```
Source    DF        SS       MS       F      P
UF2       25  3768.934  150.757  407.17  0.000
Error   5538  2050.482    0.370
Total   5563  5819.416

S = 0.6085   R-Sq = 64.76%   R-Sq(adj) = 64.61%
```

Level	N	Mean	StDev
AC	22	2.6414	0.8652
AL	101	3.6325	0.6895
AM	62	2.3610	0.6600
AP	16	3.1813	0.5750
BA	416	4.0163	0.7142
CE	184	3.6176	0.5231
ES	78	4.9196	0.5159
GO	247	4.4550	0.5007
MA	217	2.7811	0.6068
MG	851	5.0825	0.7365
MS	78	4.0787	0.5171
MT	142	3.9377	0.5799
PA	143	2.6953	0.8319
PB	222	3.8620	0.6545
PE	185	4.0929	0.6733
PI	223	3.1024	0.6796
PR	399	4.6604	0.5563
RJ	92	5.2208	0.4462
RN	167	4.0578	0.6093
RO	52	3.6842	0.4266
RR	15	2.5507	1.1191
RS	497	4.7991	0.5501
SC	295	4.7588	0.4913
SE	75	4.1992	0.5543
SP	647	5.5774	0.4137
TO	138	3.6131	0.5414

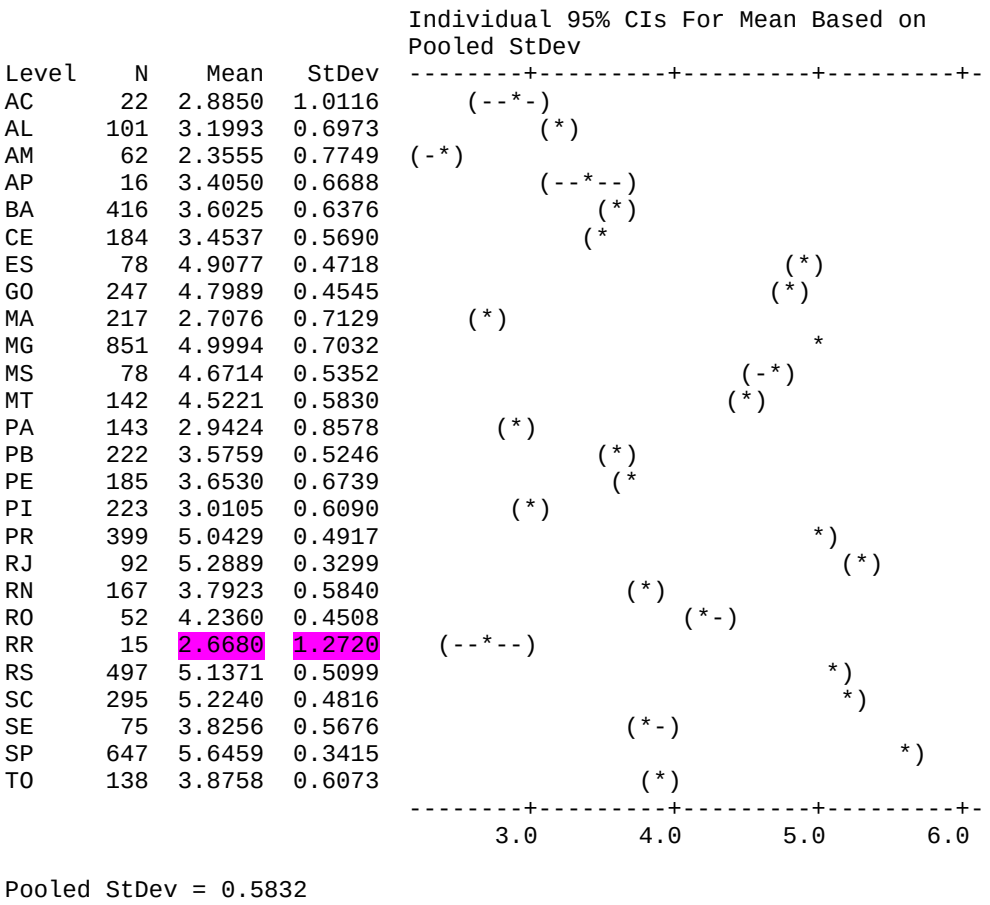



```
Pooled StDev = 0.6085
```

One-way ANOVA: ISDM versus UF2

```
Source    DF        SS       MS       F      P
UF2       25  4760.670  190.427  559.97  0.000
Error   5538  1883.279    0.340
Total   5563  6643.950

S = 0.5832   R-Sq = 71.65%   R-Sq(adj) = 71.53%
```




```
Pooled StDev = 0.5832
```

Assim como podemos observar nos gráficos de Boxplot, nas tabelas de dados, as unidades da federação que apresentam maior amplitude dos dados, ou seja, possuem alta variabilidade dos dados em relação à média, e o que destaca-se é RR, enquanto as que apresentaram menor variabildade doISDM e H é SP.

Observando ainda os dados descritivos, notamos que que alguns estados possuem valores médios maiores do ISDM e H, como SP, e outros com menor média RR.

3.2.5. CONSIDERAÇÕES FINAIS

As análises comparativas dos dados nos permitem um resumo dos dados através de cálculos específicos como médias e desvios padrões, tornando a análise dos dados mais fácil e simples. Os gráficos de Boxplot e Dendograma são excelentes figuras visuais para podermos analisar e interpretar os diferentes comportamentos dos dados. No dendograma podemos analisar as similaridades dos dados e no Boxplot podemos ver as relações entre as médias e as variâncias dos agrupamentos analisados. Trata-se de ferramentas úteis para análise de grandes volumes de dados.

CAP II ANALISE DISCRIMINANTE

1. INTRODUÇÃO

O presente trabalho tem por objetivo efetuar uma análise exploratória dos dados de Habitação incluindo as variáveis de água encanada, esgotamento sanitário, coleta de lixo, energia elétrica, domicilio próprio e densidade por dormitório, que compõe o ISDM- Indice de desenvolvimento dos municípios brasileiros.
Para tal, iniciamos com o entendimento dos dados, incluindo a definição dos indivíduos e das variáveis, suas classificações em variáveis categóricas ou quantitativas, os significados e unidades de medida, além da apresentação da tabela de dados. Na seqüência, analisamos cada uma das variáveis separadamente quanto a sua forma de distribuição, os valores atípicos, medidas de centro e dispersão. Para tal contamos com o auxílio de gráficos (*pie chart*, barras, histogramas, gráficos de ramos, box-plot, dot-plot e curvas de densidade) e de medidas numéricas (média, mediana, quartis, desvio-padrão, variância, intervalo de confiança e teste de normalidade de Anderson-Darling). No final, buscamos comparar as análises efetuadas para cada variável. O software estatístico utilizado é o **MINITAB 14.**

2. ENTENDENDO OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são os municípios brasileiros, dados referentes ao ano de 2010. Trata-se de um total de 5565 munícipios distribuídos em 27 unidades federativas, sendo 26 estados e um distrito federal. Os dados analisados de cada munícipio são as variáveis descritas abaixo.

2.2 As Variáveis

São 13 as variáveis desta pesquisa, incluindo o os três principais índices sintéticos; ISDM, IFGF e IFDM são melhor explicadas na Tabela 1. Ressaltamos que todos os dados desta pesquisa são referentes ao ano de 2010.

Tabela 1. As Variáveis

Variável	Significado	Tipo	Unidade de Medida
UF	Abreviação de Unidade Federativa (ou Unidade da Federação) do Brasil. As UF do Brasil são entidades autônomas, com governo e constituição próprias, que em seu conjunto constituem a República Federativa do Brasil. (IBGE, 2013)	Variável Categórica	N/A
Município	O município é a divisão administrativa autônoma da UF. São as unidades de menor hierarquia dentro da organização político administrativa do Brasil, criadas através de leis ordinárias das Assembléias Legislativas de cada Unidade da Federação e sancionadas pelo Governador. (IBGE, 2013)	Variável Categórica	N/A
UF2	Apresenta a sigla que representa as Unidades Federativas (ou Unidades da Federação) do Brasil.	Variável Categórica	N/A

H- Habitação	Indicador do ISDM composto por H1, H2, H3, H4, H5, H6.	Variável Quantitátiva	Percentual
H1- Água Encanada	Proporção de pessoas que vivem em domicilio com acesso à água canalizada em pelo menos um cômodo.	Variável Quantitátiva	Percentual
H2- Esgotamento Sanitário	Proporção de pessoas que vivem em domicilio com esgotamento sanitário do tipo rede geral ou esgoto pluvial.	Variável Quantitátiva	Percentual
H3- Coleta de Lixo	Proporção de pessoas que vivem em domicilio atendido por coleta de lixo (realizada por serviço de limpeza, ou cujo lixo é colocado em caçamba de serviço de limpeza).	Variável Quantitátiva	Percentual
H4- Energia Elétrica	Proporção de pessoas que vivem em domicilio que tem acesso à energia elétrica provida por companhia distribuidora.	Variável Quantitátiva	Percentual
H5- Domicilio Próprio	Proporção de pessoas que vivem em domicilio próprio de algum morador (Já pago ou ainda pagando).	Variável Quantitátiva	Percentual
H6- Densidade por Dormitório	Percentual de pessoas que vivem em domicilio que tem densidade de moradores por dormitório inferior à dois.	Variável Quantitátiva	Percentual
ISDM	Indicador Social de Desenvolvimento dos Municipios, calculado pelo Centro de Economia Aplicada da Fundação Getulio Vargas (C-Micro-FGV)- pretende contribuir para o debate de políticas publicas brasileira fornecendo uma medida sintética de bem-estar dos municípios que considere algumas de suas caracteristicas importantes relacionadas à dimensão de Renda, Habitação, Educação, Trabalho, Saude e Segurança.	Variável Quantitátiva	Percentual
IFDM	Indice Firjan de Desenvolvimento Municipal é um estudo anual que acompanha o desenvolvimento dos 5565 municipios do Brasil em três áreas: Emprego e Renda, Educação e Saúde, variando de 0à 1, sendo que quanto mais próximo de 1, maior é o desenvolvimento da localidade.	Variável Quantitátiva	0-1 Proporção
IFGF	Indice Firjan de Gestão Fiscal, para estimular a cultura de responsabilidade administrativa para aperfeiçoamento das decisões quanto à alocação de recursos públicos afim de contribuir com uma gestão eficiente e democrática e maior controle social da gestão fiscal dos municípios. Indicadores: Receita própria, pessoal, investimentos, liquidez e custo da divida.	Variável Quantitátiva	0-1 Proporção

2.3 A Tabela de Dados

Tabela 2. Tabela de Dados

Tabela de Dados- 2010												
UF	**Município**	**UF2**	**ISDM**	**H**	**H1**	**H2**	**H3**	**H4**	**H5**	**H6**	**IFGF**	**IFDM**
Acre	Acrelândia	AC	3.37	3.15	47.53	94	68.34	0	78.73	39.74	0.57	0.6108
Acre	Assis	AC	2.91	2.93	58.82	81.38	52.07	2.21	85.22	34.09	0.44	0.5459
Acre	Brasiléia	AC	3.5	3.58	65.47	89.59	64.98	19.71	79.96	37.62	0.55	0.5772
Acre	Bujari	AC	3.37	2.72	46.77	91.87	56.26	0.22	72.55	31.87	0.46	0.5402
Acre	Capixaba	AC	3.05	2.49	43.39	92	36.8	0.96	78.19	32.78	0.28	0.5291
Acre	Cruzeiro	AC	3.71	3.3	65.05	91.72	48.95	4.26	91.07	37.56	0.66	0.58
Acre	Epitaciolândia	AC	3.8	3.35	66.59	92.18	62.04	4.03	78.64	38.1	0.66	0.5472
Acre	Feijó	AC	2.09	2.03	46.06	62.4	31.31	3.23	86.95	23.42	0.43	0.4929
Acre	Jordão	AC	1.42	1.17	34.72	32.39	15.54	0.14	94.24	18.8	0.54	0.3941
Acre	Mâncio	AC	2.82	2.71	49.52	83.11	39.13	0.28	95.81	31.17	0.64	0.5393
Acre	Manoel	AC	2.51	2.39	49.29	73.26	38.72	6.72	81.01	28.39	0.59	0.5075
Acre	Marechal	AC	1.44	1.28	30.78	33.34	18.65	1.14	97.19	22.24	0.31	0.471
Acre	Plácido	AC	3.61	3.18	54.4	98.37	58.84	4.01	77.97	36.62	0.15	0.5682
Acre	Porto	AC	3.24	2.85	48.56	95.85	53.11	0.13	79.87	30.73	0.27	0.5418
Acre	Porto	AC	1.53	1.27	18.55	47.28	21.74	0	95.07	19.96	0.59	0.4641
Acre	Rio	AC	4.92	4.59	92.59	98.23	68.25	45.47	81.19	39.11	0.72	0.7691
Acre	Rodrigues	AC	1.94	2.16	36.77	87.02	17.13	0	95	26.5	0.48	0.5365
Acre	Santa	AC	0.99	1.38	36.87	34.22	31.98	0.06	93.35	13.74	0.34	0.4585
Acre	Sena	AC	3.15	2.84	61.36	79.28	44.08	6.35	86.74	30.03	0.48	0.5632
Acre	Senador	AC	4.35	3.25	63.52	98.81	63.42	0.22	76.1	33.13	0.35	0.5828
Acre	Tarauacá	AC	2.26	2.11	46.06	66.71	33	1.2	90.44	21.76	0.49	0.4633
Acre	Xapuri	AC	3.49	3.38	59.93	80.18	64.59	17.66	84.64	36.71	0.52	0.5277
Alagoas	Água	AL	2.62	3.26	33.46	95.95	47.79	31.31	87.17	40.01	0.33	0.5073
Alagoas	Anadia	AL	3.51	4.17	79.73	99.41	82.58	8.89	78.82	47.28	0.26	0.5433
Alagoas	Arapiraca	AL	4.23	4.34	90.33	99.21	85.24	10.87	75.14	47.63	0.58	0.6749
Alagoas	Atalaia	AL	3.42	3.75	75.42	97.51	73.11	16.87	67.86	37.78	0.28	0.7449
Alagoas	Barra	AL	3.52	4.07	87.14	96.61	92.21	16.09	66.72	30.99	0.47	0.5438
Alagoas	Barra	AL	4.3	4.17	94.41	99.02	95.31	17.25	57.07	33.82	0.83	0.5975
Alagoas	Batalha	AL	3.53	3.64	62.06	98.46	63.31	17.9	83.67	35.72	0.26	0.5037
Alagoas	Belém	AL	3.3	3.49	55.52	97.93	56.27	6.53	87.15	48.1	0.43	0.5155
Alagoas	Belo	AL	2.09	2.71	27.99	97.54	41.52	6.76	87.83	37.38	0.35	0.5485
Alagoas	Boca	AL	4.12	4.86	85.68	98.96	82.58	62.1	68.6	44.94	0.49	0.6326
Alagoas	Branquinha	AL	3.56	3.97	60.9	98.08	72.16	36.63	76.51	39.15	0.45	0.5502
Alagoas	Cacimbinhas	AL	2.59	2.74	53.21	97.47	24.32	0	80.42	45.82	0.49	0.4708
Alagoas	Cajueiro	AL	3.54	3.78	76.57	98.9	82.12	5.07	72.95	34.93	0.42	0.5396
Alagoas	Campestre	AL	3.82	4.48	73.54	98.94	78.29	61.8	63.67	40.93	0.15	0.5737

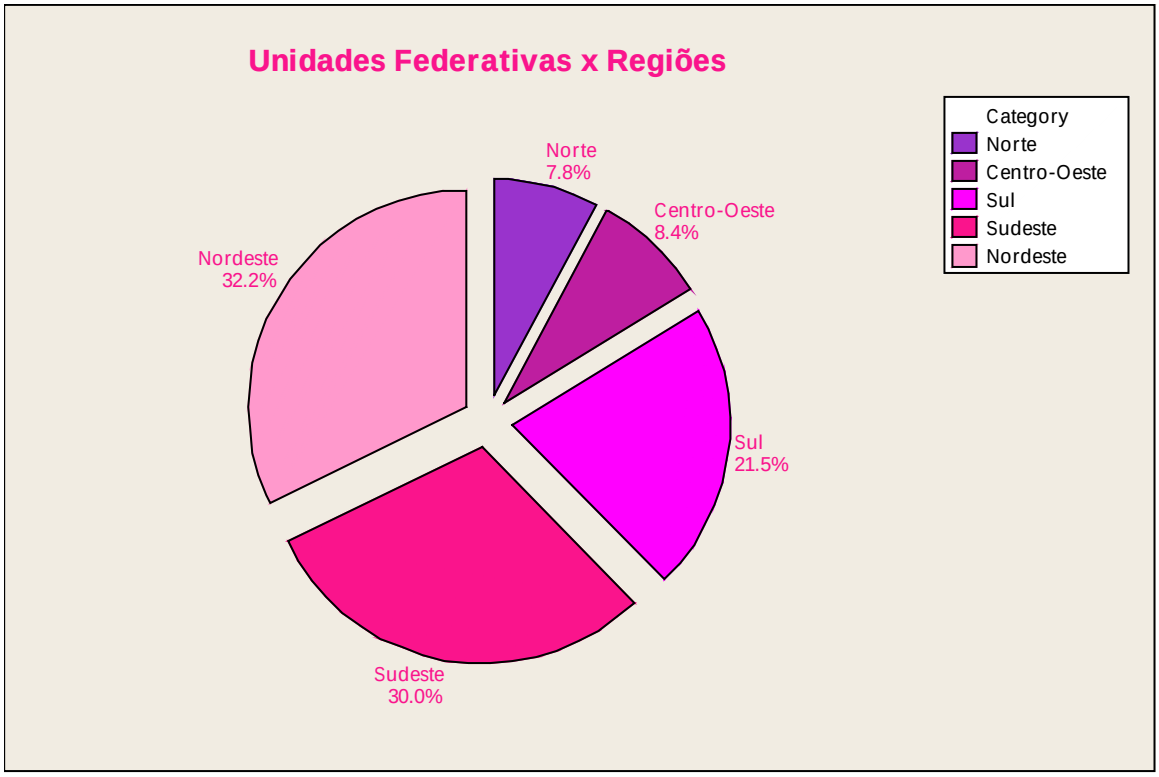
3. ANÁLISE DAS VARIÁVEIS

3.1 Variáveis Categóricas ou qualitativas.

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: “UF” e “UF2”

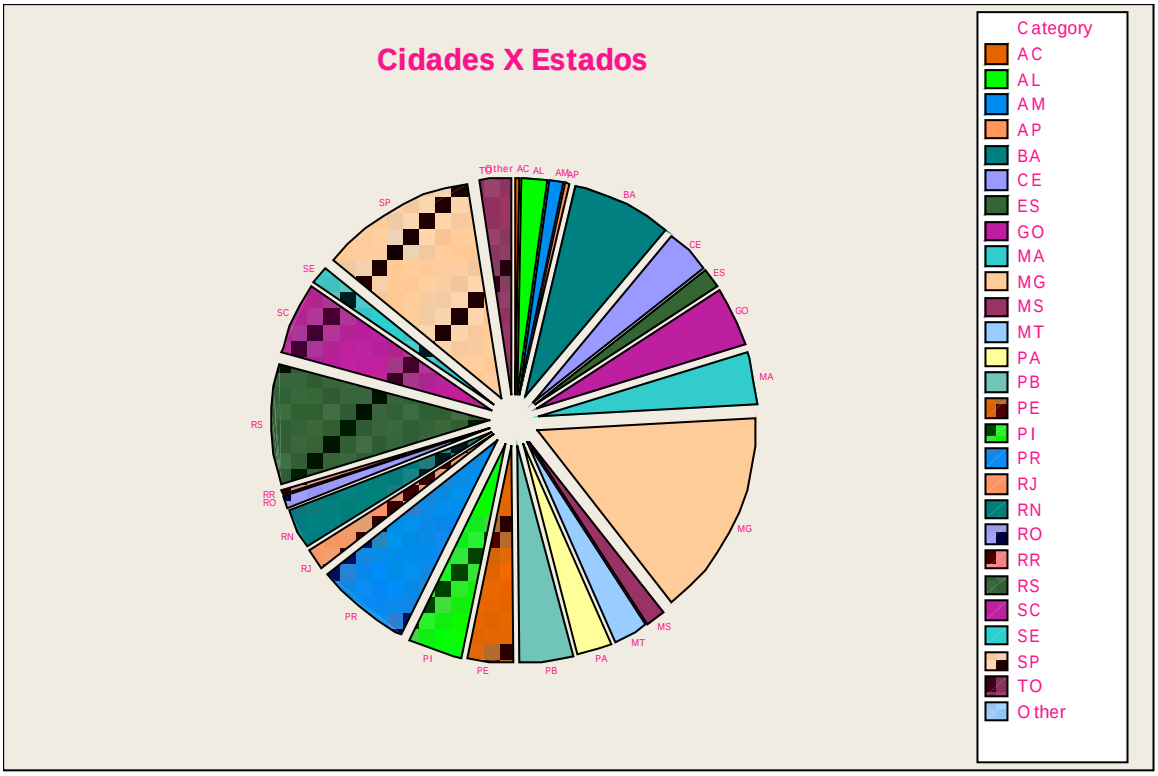
Nossa amostra totaliza 26 unidades federativas e 1 distrito federal. As unidades federativas estão distribuídas em 5 regiões.

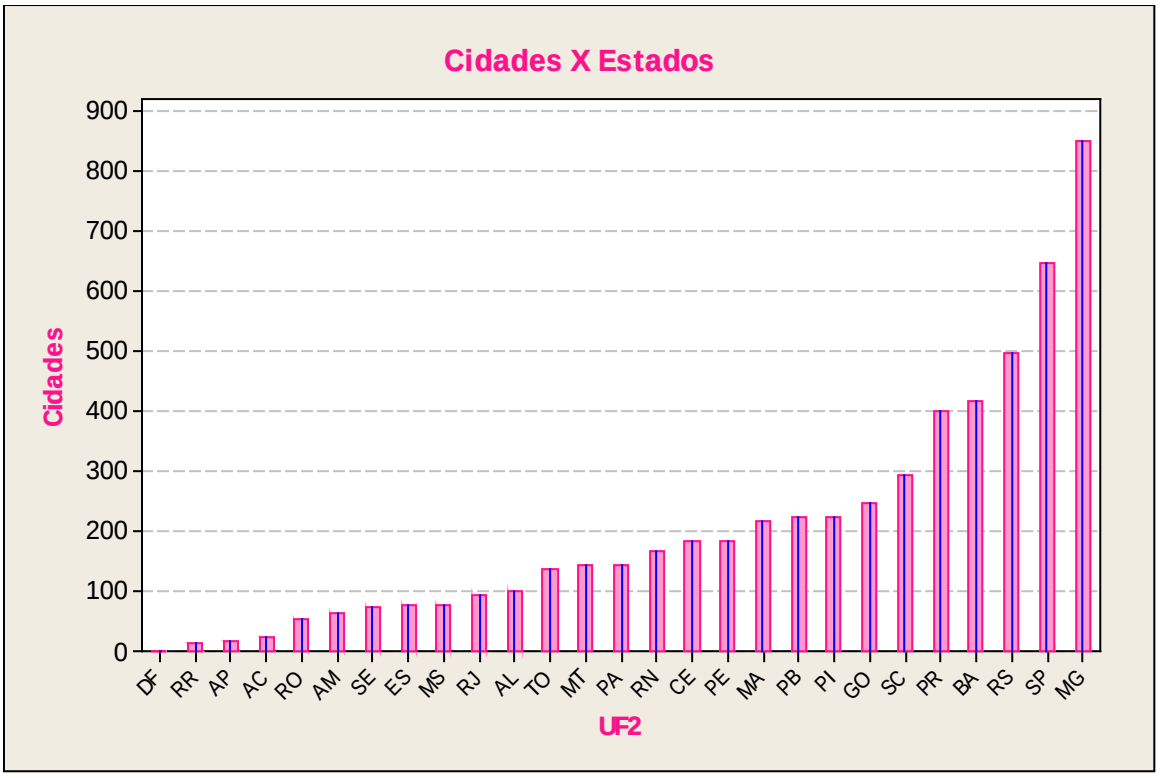


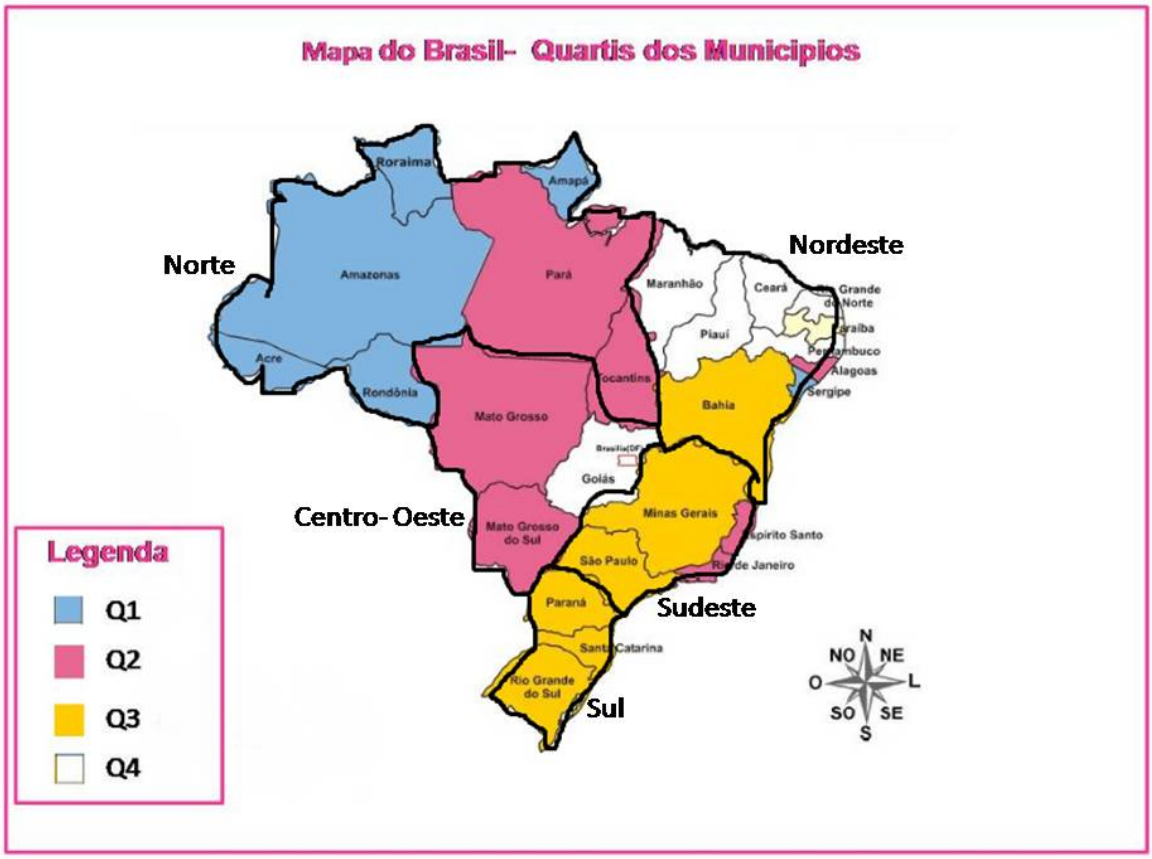
No que diz respeito a relação regiões e cidades pode-se observar no gráfico acima que as regiões Nordeste (32,2%), Sudeste (30,0%) e Sul (21,5%) concentram 83, 7% dos municípios do território nacional, enquanto as demais regiões, Norte (7.8%) e Centro-Oeste (8,4%) somam apenas 16, 2% dos munícipios. Além da concentração dos municípios brasileiros, as três regiões tem em comum o fato de serem as três regiões banhadas significativamente pelo oceano Atlântico. Fato este, que nos ajuda a entender a concentração nestas regiões.

3.1.2 Variável: “Munícipios”

Os gráficos abaixo nos ajudam a entender melhor o comportamento desta variável







Análise:

- O comportamento dos municipios por Unidades Federativas (UF2) não consiste em igualdade conforme demonstra os gráficos acima, pois enquanto o estado de Minas Gerais que contém a maior quantidade de municípios brasileiros tem 851 cidades que correspondem à 15,3 % , Roraima tem apenas 15 municipios que é correspondente à 0,3%.
Portanto Minas Gerais tem 57 vezes mais municípios que Roraima.
A distância aumenta ao considerarmos o Distrito Federal que tem somente uma cidade.

- O Primeiro e o segundo quartil concentram-se nas regiões Norte e Centro-Oeste, de maneira que tem somente dois estados no Sudeste: Rio de Janeiro e Espirito Santo e no Nordeste apenas: Alagoas e Sergipe, exclui-se deste contexto Goiás que corresponde ao quarto quartil Portanto podemos afirmar que nestas regiões concentram-se os estados com menor quantidade de municípios que totalizam 1.015, ou seja, as Regiões Norte e Centro-Oeste somadas aos quatro estados descritos acima correspondem 18% do total de municípios brasileiros.

- No terceiro Quartil os estados possuem a quantidade de municípios entre 167 e 223 concentrados na Região Sul e Sudeste, incluindo a Bahia que pertence à região Nordeste , exclui-se deste contexto Rio de Janeiro e Espirito Santo.
Este quartil é composto por 1.198 municipios que correspondem à 22% do total de municípios brasileiros.

-No ultimo Quartil visualizamos os estados que possuem as maiores quantidades de municípios, com forte concentração na região Nordeste, excluindo-se destes os estados da Bahia, Alagoas e Sergipe e incluimos Goias correspondente à região centro-oeste.
Deste total temos 3.352 municipios que correspondem à 60% do total de municípios brasileiros., portanto a Região Nordeste é composta pelos estados que mais contém municípios.

3.2 VARIÁVEIS QUANTITATIVAS

A análise deste tipo de variável permite a utilização de uma maior gama de ferramentas de análise como histogramas, curvas de densidade, gráfico de ramos, box-plot e dot-plot, além de informações numéricas como média, desvio-padrão, mediana, quartis, 5 números, intervalo de confiança e teste de normalidade de Anderson-Darling. Também podemos fazer classificações supervisionadas das variáveis quantitativas, através da análise discriminante.

3.2.1. ANÁLISE DISCRIMINANTE LINEAR DOS MUNIPIOS POR REGIÃO

A análise discriminante é uma técnica da estatística multivariada utilizada para discriminar e classificar objetos, e estuda a separação de objetos de uma população em duas ou mais classes. Neste caso queremos discriminar os valores de IFGF dos municípios do Brasil, e utilizaremos inicialmente a variável categórica Região.

Discriminant Analysis: Região versus ISDM, H, H1, H2, H3, H5, H6

```
After subtracting group means,
H4 is highly correlated with other predictors.


Linear Method for Response: Região

Predictors: ISDM, H, H1, H2, H3, H5, H6


Group  Centro-Oeste  Nordeste  Norte  Sudeste   Sul
Count           470      1793    433     1672  1196


Summary of classification

                                 True Group
Put into Group  Centro-Oeste  Nordeste  Norte  Sudeste    Sul
Centro-Oeste             102       288     75      245    206
Nordeste                  25       126     27      101    101
Norte                    130       451    136      444    278
Sudeste                   97       371     90      395    233
Sul                      116       557    105      487    378
Total N                  470      1793    433     1672   1196
N correct                102       126    136      395    378
Proportion             0.217     0.070  0.314    0.236  0.316

N = 5564          N Correct = 1137          Proportion Correct = 0.204
```

A região que errou mais é Nordeste (0,07 ou 7%) e a que acertou mais é o Sul (0,316 ou 32%). O gráfico exibe o cruzamento de dados entre as regiões. Por exemplo, a região Sudeste possui 1672 municípios e apenas 395 correspondem a região, sendo que 378 são semelhantes aos dados da região Sul. O nome desta matriz é *confusion matrix* ou matriz de confusão. Podemos concluir que o agrupamento por região não é uma boa escolha segundo esta avaliação, pois o percentual de acerto é muito baixo.

3.2.2. ANÁLISE DISCRIMINANTE LINEAR POR "2 BRASIS"

Esta segunda análise está interessada em verificar os possíveis agrupamento de dados utilizando a variável 2 Brasis, calculada no exercício anterior, e demonstra os agrupamentos do Brasil segundo sua proximidade de dados de habitação e ISDM.

Discriminant Analysis: 2 Brasis versus ISDM, H, H1, H2, H3, H5, H6

```
Linear Method for Response: 2 Brasis


Predictors: ISDM, H, H1, H2, H3, H5, H6


Group  Centro-Oeste  SSNN
Count           470  5094


Summary of classification

                     True Group
Put into Group  Centro-Oeste   SSNN
Centro-Oeste            219   2098
SSNN                    251   2996
Total N                 470   5094
N correct               219   2996
Proportion            0.466  0.588

N = 5564          N Correct = 3215          Proportion Correct = 0.578
```

Como pode se observar a % de acertos continua muito baixa devido a alta variabilidade dos dados entre os municípios em relação as variáveis sendo utilizadas.

4. CONSIDERAÇÕES FINAIS

A tarefa da análise discriminante é encontrar a melhor função discriminante linear de um conjunto de variáveis que reproduza, tanto quanto possível, um agrupamento a priori de casos considerados.

Um procedimento em passos é utilizado nesse programa, e em cada passo a variável mais poderosa é introduzida na função discriminante. A função critério para selecionar a próxima variável depende do número de grupos especificados (o número de grupos varia de 2 a 20).

Quando o número de variáveis é maior do que dois, então o critério de seleção de variáveis é o traço do produto da matriz de covariância para as variáveis envolvidas e a matriz de covariância interclasse em um passo particular.

Os cálculos podem ser realizados em toda a população ou em amostra de dados ou mesmo em dados previamente agrupados.

Nos nossos exemplos com as variáveis de ISDM e Habitação, utilizamos a análise discriminante linear e conseguimos um resultado de 57,8 % de proporção correta para 2 Brasis. Muito baixo seguramente devido a alta variabilidade dos dados.

CAP III REGRESSÃO LOGÍSTICA

1. INTRODUÇÃO

A regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, frequentemente binária, a partir de uma série de variáveis explicativas contínuas e/ou binárias1 2. A regressão logística é amplamente usada em ciências médicas e sociais, e tem outras denominações, como modelo logístico e classificador de máxima entropia.

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.

Enquanto método de predição para variáveis categóricas, a regressão logística é comparável às técnicas supervisionadas propostas em aprendizagem automática (árvores de decisão, redes neuronais, etc.), ou ainda a análise discriminante preditiva em estatística exploratória. É possível de colocá-las em concorrência para escolha do modelo mais adaptado para certo problema preditivo a resolver.

Trata-se de um modelo de regressão para variáveis dependentes ou de resposta binomialmente distribuídas. É útil para modelar a probabilidade de um evento ocorrer como função de outros fatores. Os dados são originários da pesquisa de ISDM sobre o desenvolvimento dos municípios do Brasil. Neste trabalho abordaremos as variáveis referentes à Habitação dos municípios. O software estatístico utilizado é o **MINITAB14.**

2. ENTENDENDO OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são os municípios brasileiros, dados referentes ao ano de 2010. Trata-se de um total de 5565 munícipios distribuídos em 27 unidades federativas, sendo 26 estados e um distrito federal. Os dados analisados de cada munícipio são as variáveis descritas abaixo.

2.2 As Variáveis

São 13 as variáveis desta pesquisa, incluindo o os três principais índices sintéticos; ISDM, IFGF e IFDM são melhor explicadas na Tabela 1. Ressaltamos que todos os dados desta pesquisa são referentes ao ano de 2010.

Tabela 1. As Variáveis

Variável	Significado	Tipo	Unidade de Medida
UF	Abreviação de Unidade Federativa (ou Unidade da Federação) do Brasil. As UF do Brasil são entidades autônomas, com governo e constituição próprias, que em seu conjunto constituem a República Federativa do Brasil. (IBGE, 2013)	Variável Categórica	N/A
Município	O município é a divisão administrativa autônoma da UF. São as unidades de menor hierarquia dentro da organização político administrativa do Brasil, criadas através de leis ordinárias das Assembléias Legislativas de cada Unidade da Federação e sancionadas pelo Governador. (IBGE, 2013)	Variável Categórica	N/A
UF2	Apresenta a sigla que representa as Unidades Federativas (ou Unidades da Federação) do Brasil.	Variável Categórica	N/A
H- Habitação	Indicador do ISDM composto por H1, H2, H3, H4, H5, H6.	Variável Quantitátiva	Percentual
H1- Água Encanada	Proporção de pessoas que vivem em domicilio com acesso à água canalizada em pelo menos um cômodo.	Variável Quantitátiva	Percentual
H2- Esgotamento Sanitário	Proporção de pessoas que vivem em domicilio com esgotamento sanitário do tipo rede geral ou esgoto pluvial.	Variável Quantitátiva	Percentual
H3- Coleta de Lixo	Proporção de pessoas que vivem em domicilio atendido por coleta de lixo (realizada por serviço de limpeza, ou cujo lixo é colocado em caçamba de serviço de limpeza).	Variável Quantitátiva	Percentual
H4- Energia Elétrica	Proporção de pessoas que vivem em domicilio que tem acesso à energia elétrica provida por companhia distribuidora.	Variável Quantitátiva	Percentual

H5- Domicilio Próprio	Proporção de pessoas que vivem em domicilio próprio de algum morador (Já pago ou ainda pagando).	Variável Quantitátiva	Percentual
H6- Densidade por Dormitório	Percentual de pessoas que vivem em domicilio que tem densidade de moradores por dormitório inferior à dois.	Variável Quantitátiva	Percentual
ISDM	Indicador Social de Desenvolvimento dos Municipios, calculado pelo Centro de Economia Aplicada da Fundação Getulio Vargas (C-Micro-FGV)- pretende contribuir para o debate de políticas publicas brasileira fornecendo uma medida sintética de bem-estar dos municípios que considere algumas de suas caracteristicas importantes relacionadas à dimensão de Renda, Habitação, Educação, Trabalho, Saude e Segurança.	Variável Quantitátiva	Percentual
IFDM	Indice Firjan de Desenvolvimento Municipal é um estudo anual que acompanha o desenvolvimento dos 5565 municipios do Brasil em três áreas: Emprego e Renda, Educação e Saúde, variando de 0à 1, sendo que quanto mais próximo de 1, maior é o desenvolvimento da localidade.	Variável Quantitátiva	0-1 Proporção
IFGF	Indice Firjan de Gestão Fiscal, para estimular a cultura de responsabilidade administrativa para aperfeiçoamento das decisões quanto à alocação de recursos públicos afim de contribuir com uma gestão eficiente e democrática e maior controle social da gestão fiscal dos municípios. Indicadores: Receita própria, pessoal, investimentos, liquidez e custo da divida.	Variável Quantitátiva	0-1 Proporção

2.3 A Tabela de Dados

Tabela 2. Tabela de Dados

Tabela de Dados- 2010

UF	Município	UF2	ISDM	H	H1	H2	H3	H4	H5	H6	IFGF	IFDM
Acre	Acrelândia	AC	3.37	3.15	47.53	94	68.34	0	78.73	39.74	0.57	0.6108
Acre	Assis	AC	2.91	2.93	58.82	81.38	52.07	2.21	85.22	34.09	0.44	0.5459
Acre	Brasiléia	AC	3.5	3.58	65.47	89.59	64.98	19.71	79.96	37.62	0.55	0.5772
Acre	Bujari	AC	3.37	2.72	46.77	91.87	56.26	0.22	72.55	31.87	0.46	0.5402
Acre	Capixaba	AC	3.05	2.49	43.39	92	36.8	0.96	78.19	32.78	0.28	0.5291
Acre	Cruzeiro	AC	3.71	3.3	65.05	91.72	48.95	4.26	91.07	37.56	0.66	0.58
Acre	Epitaciolândia	AC	3.8	3.35	66.59	92.18	62.04	4.03	78.64	38.1	0.66	0.5472
Acre	Feijó	AC	2.09	2.03	46.06	62.4	31.31	3.23	86.95	23.42	0.43	0.4929
Acre	Jordão	AC	1.42	1.17	34.72	32.39	15.54	0.14	94.24	18.8	0.54	0.3941
Acre	Mâncio	AC	2.82	2.71	49.52	83.11	39.13	0.28	95.81	31.17	0.64	0.5393

Acre	Manoel	AC	2.51	2.39	49.29	73.26	38.72	6.72	81.01	28.39	0.59	0.5075
Acre	Marechal	AC	1.44	1.28	30.78	33.34	18.65	1.14	97.19	22.24	0.31	0.471
Acre	Plácido	AC	3.61	3.18	54.4	98.37	58.84	4.01	77.97	36.62	0.15	0.5682
Acre	Porto	AC	3.24	2.85	48.56	95.85	53.11	0.13	79.87	30.73	0.27	0.5418
Acre	Porto	AC	1.53	1.27	18.55	47.28	21.74	0	95.07	19.96	0.59	0.4641
Acre	Rio	AC	4.92	4.59	92.59	98.23	68.25	45.47	81.19	39.11	0.72	0.7691
Acre	Rodrigues	AC	1.94	2.16	36.77	87.02	17.13	0	95	26.5	0.48	0.5365
Acre	Santa	AC	0.99	1.38	36.87	34.22	31.98	0.06	93.35	13.74	0.34	0.4585
Acre	Sena	AC	3.15	2.84	61.36	79.28	44.08	6.35	86.74	30.03	0.48	0.5632
Acre	Senador	AC	4.35	3.25	63.52	98.81	63.42	0.22	76.1	33.13	0.35	0.5828
Acre	Tarauacá	AC	2.26	2.11	46.06	66.71	33	1.2	90.44	21.76	0.49	0.4633
Acre	Xapuri	AC	3.49	3.38	59.93	80.18	64.59	17.66	84.64	36.71	0.52	0.5277
Alagoas	Água	AL	2.62	3.26	33.46	95.95	47.79	31.31	87.17	40.01	0.33	0.5073
Alagoas	Anadia	AL	3.51	4.17	79.73	99.41	82.58	8.89	78.82	47.28	0.26	0.5433
Alagoas	Arapiraca	AL	4.23	4.34	90.33	99.21	85.24	10.87	75.14	47.63	0.58	0.6749
Alagoas	Atalaia	AL	3.42	3.75	75.42	97.51	73.11	16.87	67.86	37.78	0.28	0.7449
Alagoas	Barra	AL	3.52	4.07	87.14	96.61	92.21	16.09	66.72	30.99	0.47	0.5438
Alagoas	Barra	AL	4.3	4.17	94.41	99.02	95.31	17.25	57.07	33.82	0.83	0.5975
Alagoas	Batalha	AL	3.53	3.64	62.06	98.46	63.31	17.9	83.67	35.72	0.26	0.5037
Alagoas	Belém	AL	3.3	3.49	55.52	97.93	56.27	6.53	87.15	48.1	0.43	0.5155
Alagoas	Belo	AL	2.09	2.71	27.99	97.54	41.52	6.76	87.83	37.38	0.35	0.5485
Alagoas	Boca	AL	4.12	4.86	85.68	98.96	82.58	62.1	68.6	44.94	0.49	0.6326
Alagoas	Branquinha	AL	3.56	3.97	60.9	98.08	72.16	36.63	76.51	39.15	0.45	0.5502
Alagoas	Cacimbinhas	AL	2.59	2.74	53.21	97.47	24.32	0	80.42	45.82	0.49	0.4708
Alagoas	Cajueiro	AL	3.54	3.78	76.57	98.9	82.12	5.07	72.95	34.93	0.42	0.5396
Alagoas	Campestre	AL	3.82	4.48	73.54	98.94	78.29	61.8	63.67	40.93	0.15	0.5737

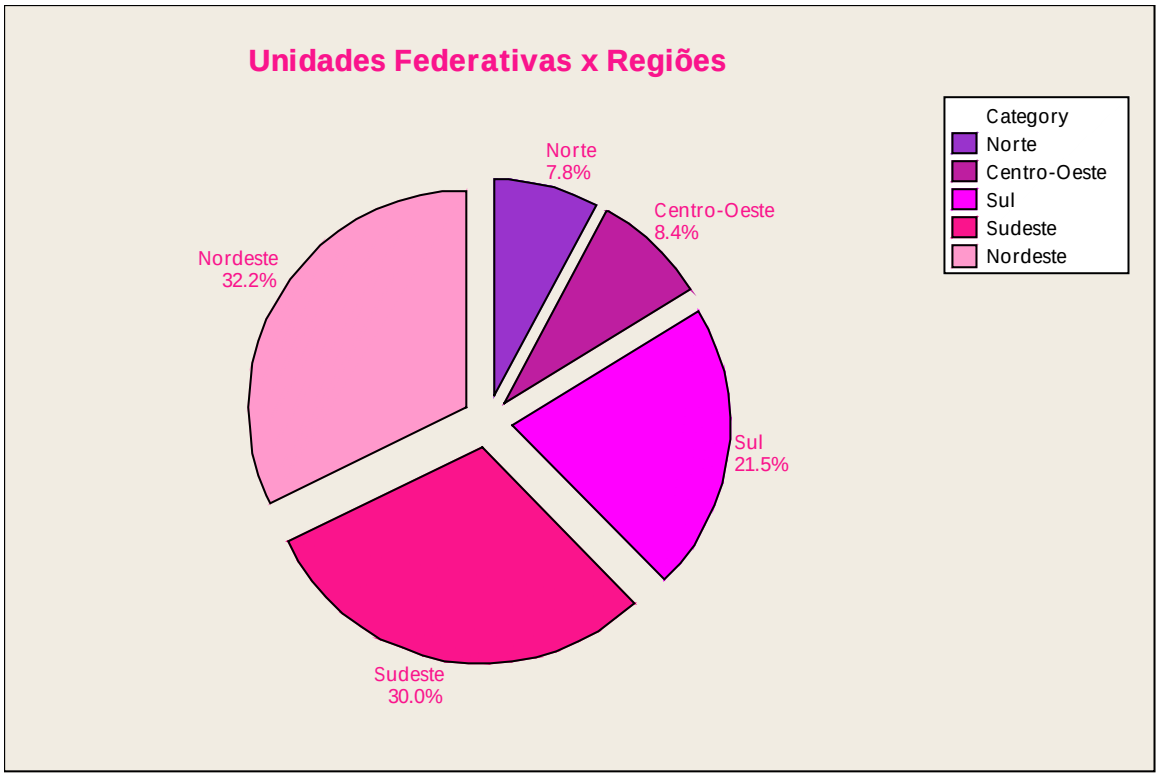
3. ANÁLISE DAS VARIÁVEIS

3.1 Variáveis Categóricas ou qualitativas.

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: “UF” e “UF2”

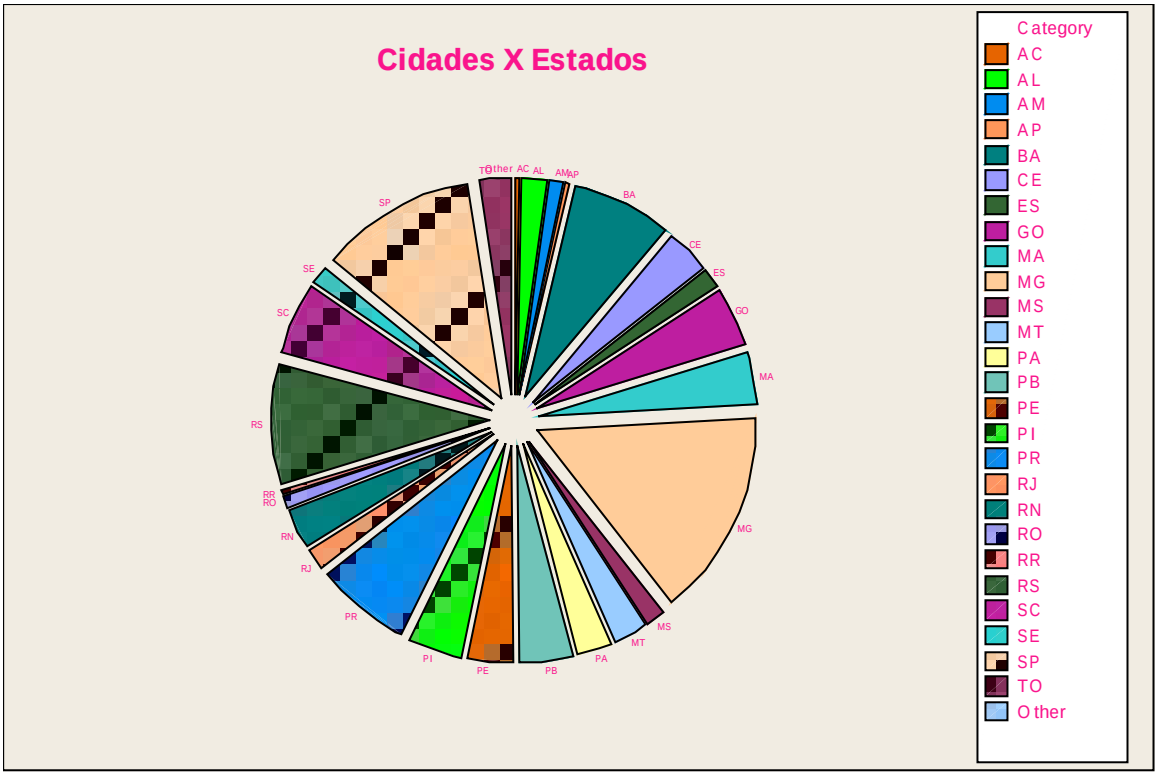
Nossa amostra totaliza 26 unidades federativas e 1 distrito federal. As unidades federativas estão distribuídas em 5 regiões.

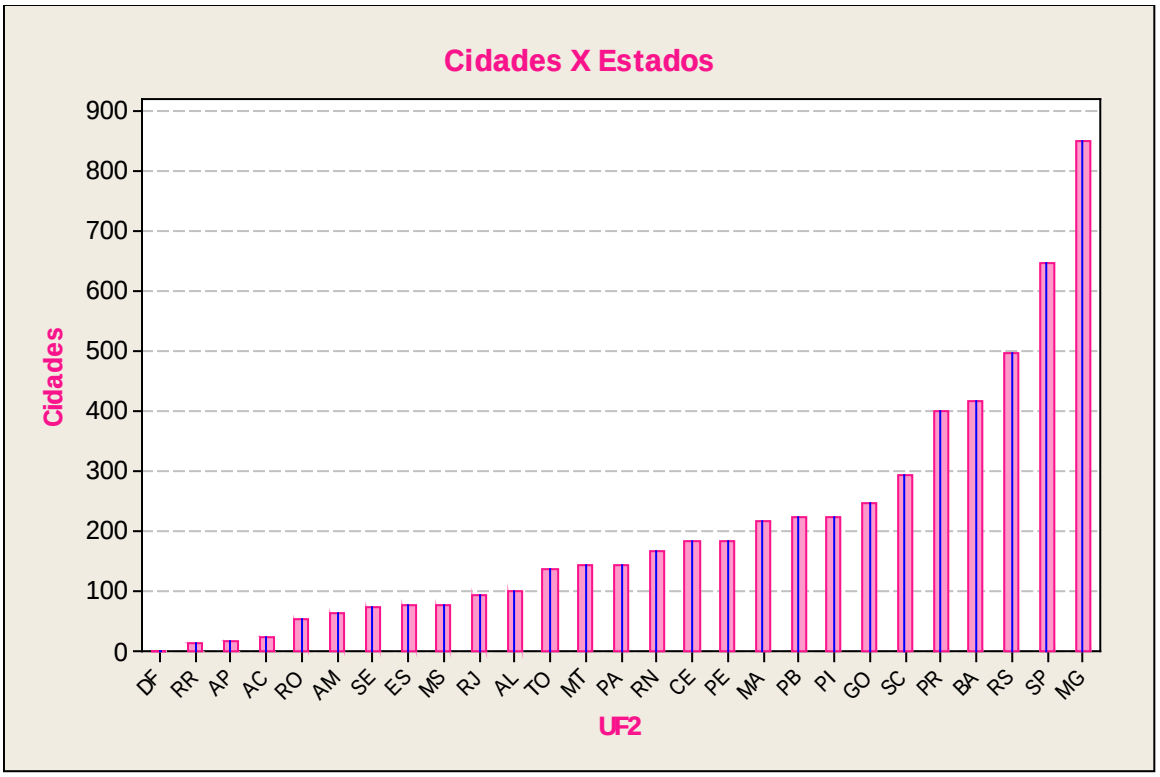


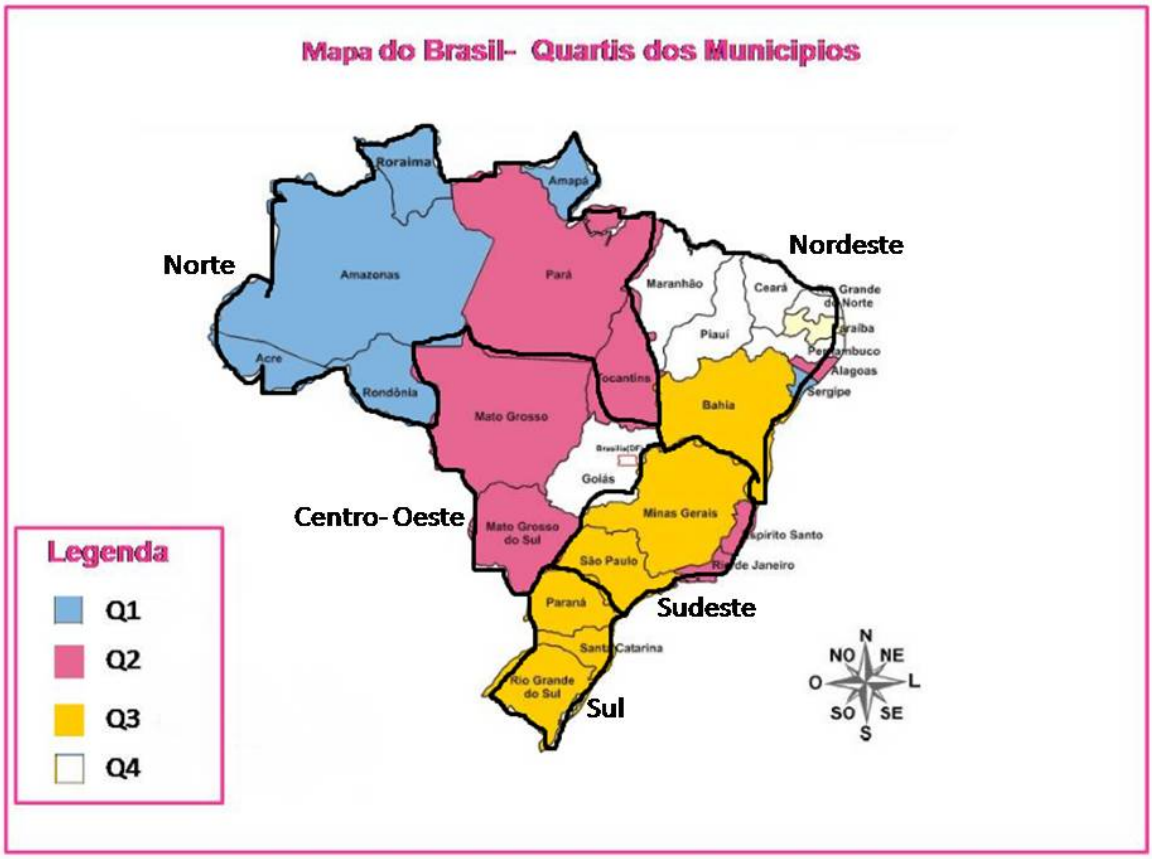
No que diz respeito a relação regiões e cidades pode-se observar no gráfico acima que as regiões Nordeste (32,2%), Sudeste (30,0%) e Sul (21,5%) concentram 83, 7% dos municípios do território nacional, enquanto as demais regiões, Norte (7.8%) e Centro-Oeste (8,4%) somam apenas 16, 2% dos munícipios. Além da concentração dos municípios brasileiros, as três regiões tem em comum o fato de serem as três regiões banhadas significativamente pelo oceano Atlântico. Fato este, que nos ajuda a entender a concentração nestas regiões.

3.1.2 Variável: “Munícipios”

Os gráficos abaixo nos ajudam a entender melhor o comportamento desta variável


Cidades X Estados
Category
AC
AL
AM
AP
BA
CE
ES
GO
MA
MG
MS
MT
PA
PB
PE
PI
PR
RJ
RN
RO
RR
RS
SC
SE
SP
TO
Other


Cidades X Estados
Cidades
900
800
700
600
500
400
300
200
100
0
DF
RR
AP
AC
RO
AM
SE
ES
MS
RJ
AL
TO
MT
PA
RN
CE
PE
MA
PB
PI
GO
SC
PR
BA
RS
SP
MG
UF2



Análise:

- O comportamento dos municipios por Unidades Federativas (UF2) não consiste em igualdade conforme demonstra os gráficos acima, pois enquanto o estado de Minas Gerais que contém a maior quantidade de municípios brasileiros tem 851 cidades que correspondem à 15,3 % , Roraima tem apenas 15 municipios que é correspondente à 0,3%.
Portanto Minas Gerais tem 57 vezes mais municípios que Roraima.
A distância aumenta ao considerarmos o Distrito Federal que tem somente uma cidade.

- O Primeiro e o segundo quartil concentram-se nas regiões Norte e Centro-Oeste, de maneira que tem somente dois estados no Sudeste: Rio de Janeiro e Espirito Santo e no Nordeste apenas: Alagoas e Sergipe, exclui-se deste contexto Goiás que corresponde ao quarto quartil Portanto podemos afirmar que nestas regiões concentram-se os estados com menor quantidade de municípios que totalizam 1.015, ou seja, as Regiões Norte e Centro-Oeste somadas aos quatro estados descritos acima correspondem 18% do total de municípios brasileiros.

- No terceiro Quartil os estados possuem a quantidade de municípios entre 167 e 223 concentrados na Região Sul e Sudeste, incluindo a Bahia que pertence à região Nordeste , exclui-se deste contexto Rio de Janeiro e Espirito Santo.
Este quartil é composto por 1.198 municipios que correspondem à 22% do total de municípios brasileiros.

-No ultimo Quartil visualizamos os estados que possuem as maiores quantidades de municípios, com forte concentração na região Nordeste, excluindo-se destes os estados da Bahia, Alagoas e Sergipe e incluimos Goias correspondente à região centro-oeste.
Deste total temos 3.352 municipios que correspondem à 60% do total de municípios brasileiros., portanto a Região Nordeste é composta pelos estados que mais contém municípios.

3.2 VARIÁVEIS QUANTITATIVAS

A análise deste tipo de variável permite a utilização de uma maior gama de ferramentas de análise como histogramas, curvas de densidade, gráfico de ramos, box-plot e dot-plot, além de informações numéricas como média, desvio-padrão, mediana, quartis, 5 números, intervalo de confiança e teste de normalidade de Anderson-Darling. Também podemos fazer classificações supervisionadas das variáveis quantitativas, através da análise discriminante.

3.2.1. REGRESSÃO LOGÍSTICA

Nominal Logistic Regression: Região versus ISDM, H, ...

Response Information

Variable	Value	Count	
Região	Sul	1196	(Reference Event)
	Sudeste	1672	
	Norte	433	
	Nordeste	1793	
	Centro-Oeste	470	
	Total	5564	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio
Logit 1: (Sudeste/Sul)					
Constant	16.2437	23.3116	0.70	0.486	
ISDM	0.0079606	0.111008	0.07	0.943	1.01
H	9.81770	13.2275	0.74	0.458	18355.78
H1	-0.150871	0.197790	-0.76	0.446	0.86
H2	-0.136307	0.197892	-0.69	0.491	0.87
H3	-0.140336	0.197885	-0.71	0.478	0.87
H4	-0.149195	0.197787	-0.75	0.451	0.86
H5	-0.141813	0.197824	-0.72	0.473	0.87
H6	-0.150888	0.197725	-0.76	0.445	0.86
Logit 2: (Norte/Sul)					
Constant	19.2960	34.5711	0.56	0.577	
ISDM	-0.254029	0.159157	-1.60	0.110	0.78
H	12.0095	19.6176	0.61	0.540	164312.38
H1	-0.183109	0.293347	-0.62	0.532	0.83
H2	-0.174926	0.293459	-0.60	0.551	0.84
H3	-0.166635	0.293480	-0.57	0.570	0.85
H4	-0.181014	0.293339	-0.62	0.537	0.83
H5	-0.168671	0.293398	-0.57	0.565	0.84
H6	-0.181838	0.293239	-0.62	0.535	0.83
Logit 3: (Nordeste/Sul)					
Constant	21.3855	22.9702	0.93	0.352	
ISDM	0.0255699	0.109428	0.23	0.815	1.03

H	12.5141	13.0339	0.96	0.337	272136.47
H1	-0.189101	0.194896	-0.97	0.332	0.83
H2	-0.176310	0.194986	-0.90	0.366	0.84
H3	-0.188471	0.194989	-0.97	0.334	0.83
H4	-0.188548	0.194893	-0.97	0.333	0.83
H5	-0.183925	0.194934	-0.94	0.345	0.83
H6	-0.187589	0.194830	-0.96	0.336	0.83
Logit 4: (Centro-Oeste/Sul)					
Constant	8.68596	33.5344	0.26	0.796	
ISDM	-0.216972	0.155551	-1.39	0.163	0.80
H	5.23201	19.0292	0.27	0.783	187.17
H1	-0.0799218	0.284543	-0.28	0.779	0.92
H2	-0.0758805	0.284637	-0.27	0.790	0.93
H3	-0.0729769	0.284680	-0.26	0.798	0.93
H4	-0.0786308	0.284539	-0.28	0.782	0.92
H5	-0.0777643	0.284608	-0.27	0.785	0.93
H6	-0.0785788	0.284440	-0.28	0.782	0.92

	95% CI	
Predictor	Lower	Upper
Logit 1: (Sudeste/Sul)		
Constant		
ISDM	0.81	1.25
H	0.00	3.33597E+15
H1	0.58	1.27
H2	0.59	1.29
H3	0.59	1.28
H4	0.58	1.27
H5	0.59	1.28
H6	0.58	1.27
Logit 2: (Norte/Sul)		
Constant		
ISDM	0.57	1.06
H	0.00	8.21293E+21
H1	0.47	1.48
H2	0.47	1.49
H3	0.48	1.50
H4	0.47	1.48
H5	0.48	1.50
H6	0.47	1.48
Logit 3: (Nordeste/Sul)		
Constant		
ISDM	0.83	1.27
H	0.00	3.38425E+16
H1	0.56	1.21
H2	0.57	1.23
H3	0.57	1.21
H4	0.57	1.21
H5	0.57	1.22
H6	0.57	1.21
Logit 4: (Centro-Oeste/Sul)		
Constant		
ISDM	0.59	1.09
H	0.00	2.95240E+18
H1	0.53	1.61
H2	0.53	1.62
H3	0.53	1.62
H4	0.53	1.61
H5	0.53	1.62
H6	0.53	1.61

Log-Likelihood = -8121.352
Test that all slopes are zero: G = 50.212, DF = 32, P-Value = 0.021

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	22257.9	22220	0.427
Deviance	16242.7	22220	1.000

4. CONSIDERAÇÕES FINAIS

Enquanto método de predição para variáveis categóricas, a regressão logística é comparável às técnicas supervisionadas propostas em aprendizagem automática (árvores de decisão, redes neuronais, etc.), ou ainda a análise discriminante preditiva em estatística exploratória. É possível de colocá-la em concorrência para escolha do modelo mais adaptado para certo problema preditivo a resolver.

Como tinha sido observado no caso de Analise Discriminante a grande variabilidade em termos das variáveis ISDM e Habitação não permitiu uma % de acertos razoável (20%) , seguramente devido a isto a Regressão Logística para 5 Brasis não convergiu, e mesmo para 2 Brasis a % não tinha sido alta e por tanto era pouco provável que melhorasse no caso da Regressão Logistica. Um outra possibilidade teria sido considerar estados no lugar de municípios.

CAP IV ÁRVORES DE CLASSIFICAÇÃO

1. INTRODUÇÃO

.As árvores de classificação são modelos estatísticos que utilizam técnicas supervisionadas para a classificação e previsão de dados. Ou seja, em sua construção é utilizado um conjunto de dados formados por entradas (*predictors*) e saídas (classes).

Os dados utilizados são originários da pesquisa da ISDM- FGV sobre o desenvolvimento dos municípios do Brasil. Neste trabalho abordaremos as variáveis referentes à habitação dos municípios. O software estatístico utilizado é o **SPSS21.**

2. ENTENDENDO OS DADOS

2.1 Os Indivíduos

Os indivíduos desta análise são os municípios brasileiros, dados referentes ao ano de 2010. Trata-se de um total de 5565 munícipios distribuídos em 27 unidades federativas, sendo 26 estados e um distrito federal. Os dados analisados de cada munícipio são as variáveis descritas abaixo.

2.2 As Variáveis

São 13 as variáveis desta pesquisa, incluindo o os três principais índices sintéticos; ISDM, IFGF e IFDM são melhor explicadas na Tabela 1. Ressaltamos que todos os dados desta pesquisa são referentes ao ano de 2010.

Tabela 1. As Variáveis

Variável	Significado	Tipo	Unidade de Medida
UF	Abreviação de Unidade Federativa (ou Unidade da Federação) do Brasil. As UF do Brasil são entidades autônomas, com governo e constituição próprias, que em seu conjunto constituem a República Federativa do Brasil. (IBGE, 2013)	Variável Categórica	N/A
Município	O município é a divisão administrativa autônoma da UF. São as unidades de menor hierarquia dentro da organização político administrativa do Brasil, criadas através de leis ordinárias das Assembléias Legislativas de cada Unidade da Federação e sancionadas pelo Governador. (IBGE, 2013)	Variável Categórica	N/A
UF2	Apresenta a sigla que representa as Unidades Federativas (ou Unidades da Federação) do Brasil.	Variável Categórica	N/A

H- Habitação	Indicador do ISDM composto por H1, H2, H3, H4, H5, H6.	Variável Quantitátiva	Percentual
H1- Água Encanada	Proporção de pessoas que vivem em domicilio com acesso à água canalizada em pelo menos um cômodo.	Variável Quantitátiva	Percentual
H2- Esgotamento Sanitário	Proporção de pessoas que vivem em domicilio com esgotamento sanitário do tipo rede geral ou esgoto pluvial.	Variável Quantitátiva	Percentual
H3- Coleta de Lixo	Proporção de pessoas que vivem em domicilio atendido por coleta de lixo (realizada por serviço de limpeza, ou cujo lixo é colocado em caçamba de serviço de limpeza).	Variável Quantitátiva	Percentual
H4- Energia Elétrica	Proporção de pessoas que vivem em domicilio que tem acesso à energia elétrica provida por companhia distribuidora.	Variável Quantitátiva	Percentual
H5- Domicilio Próprio	Proporção de pessoas que vivem em domicilio próprio de algum morador (Já pago ou ainda pagando).	Variável Quantitátiva	Percentual
H6- Densidade por Dormitório	Percentual de pessoas que vivem em domicilio que tem densidade de moradores por dormitório inferior à dois.	Variável Quantitátiva	Percentual
ISDM	Indicador Social de Desenvolvimento dos Municipios, calculado pelo Centro de Economia Aplicada da Fundação Getulio Vargas (C-Micro-FGV)- pretende contribuir para o debate de políticas publicas brasileira fornecendo uma medida sintética de bem-estar dos municípios que considere algumas de suas caracteristicas importantes relacionadas à dimensão de Renda, Habitação, Educação, Trabalho, Saude e Segurança.	Variável Quantitátiva	Percentual
IFDM	Indice Firjan de Desenvolvimento Municipal é um estudo anual que acompanha o desenvolvimento dos 5565 municipios do Brasil em três áreas: Emprego e Renda, Educação e Saúde, variando de 0à 1, sendo que quanto mais próximo de 1, maior é o desenvolvimento da localidade.	Variável Quantitátiva	0-1 Proporção
IFGF	Indice Firjan de Gestão Fiscal, para estimular a cultura de responsabilidade administrativa para aperfeiçoamento das decisões quanto à alocação de recursos públicos afim de contribuir com uma gestão eficiente e democrática e maior controle social da gestão fiscal dos municípios. Indicadores: Receita própria, pessoal, investimentos, liquidez e custo da divida.	Variável Quantitátiva	0-1 Proporção

2.3 A Tabela de Dados

Tabela 2. Tabela de Dados

Tabela de Dados- 2010												
UF	**Município**	**UF2**	**ISDM**	**H**	**H1**	**H2**	**H3**	**H4**	**H5**	**H6**	**IFGF**	**IFDM**
Acre	Acrelândia	AC	3.37	3.15	47.53	94	68.34	0	78.73	39.74	0.57	0.6108
Acre	Assis	AC	2.91	2.93	58.82	81.38	52.07	2.21	85.22	34.09	0.44	0.5459
Acre	Brasiléia	AC	3.5	3.58	65.47	89.59	64.98	19.71	79.96	37.62	0.55	0.5772
Acre	Bujari	AC	3.37	2.72	46.77	91.87	56.26	0.22	72.55	31.87	0.46	0.5402
Acre	Capixaba	AC	3.05	2.49	43.39	92	36.8	0.96	78.19	32.78	0.28	0.5291
Acre	Cruzeiro	AC	3.71	3.3	65.05	91.72	48.95	4.26	91.07	37.56	0.66	0.58
Acre	Epitaciolândia	AC	3.8	3.35	66.59	92.18	62.04	4.03	78.64	38.1	0.66	0.5472
Acre	Feijó	AC	2.09	2.03	46.06	62.4	31.31	3.23	86.95	23.42	0.43	0.4929
Acre	Jordão	AC	1.42	1.17	34.72	32.39	15.54	0.14	94.24	18.8	0.54	0.3941
Acre	Mâncio	AC	2.82	2.71	49.52	83.11	39.13	0.28	95.81	31.17	0.64	0.5393
Acre	Manoel	AC	2.51	2.39	49.29	73.26	38.72	6.72	81.01	28.39	0.59	0.5075
Acre	Marechal	AC	1.44	1.28	30.78	33.34	18.65	1.14	97.19	22.24	0.31	0.471
Acre	Plácido	AC	3.61	3.18	54.4	98.37	58.84	4.01	77.97	36.62	0.15	0.5682
Acre	Porto	AC	3.24	2.85	48.56	95.85	53.11	0.13	79.87	30.73	0.27	0.5418
Acre	Porto	AC	1.53	1.27	18.55	47.28	21.74	0	95.07	19.96	0.59	0.4641
Acre	Rio	AC	4.92	4.59	92.59	98.23	68.25	45.47	81.19	39.11	0.72	0.7691
Acre	Rodrigues	AC	1.94	2.16	36.77	87.02	17.13	0	95	26.5	0.48	0.5365
Acre	Santa	AC	0.99	1.38	36.87	34.22	31.98	0.06	93.35	13.74	0.34	0.4585
Acre	Sena	AC	3.15	2.84	61.36	79.28	44.08	6.35	86.74	30.03	0.48	0.5632
Acre	Senador	AC	4.35	3.25	63.52	98.81	63.42	0.22	76.1	33.13	0.35	0.5828
Acre	Tarauacá	AC	2.26	2.11	46.06	66.71	33	1.2	90.44	21.76	0.49	0.4633
Acre	Xapuri	AC	3.49	3.38	59.93	80.18	64.59	17.66	84.64	36.71	0.52	0.5277
Alagoas	Água	AL	2.62	3.26	33.46	95.95	47.79	31.31	87.17	40.01	0.33	0.5073
Alagoas	Anadia	AL	3.51	4.17	79.73	99.41	82.58	8.89	78.82	47.28	0.26	0.5433
Alagoas	Arapiraca	AL	4.23	4.34	90.33	99.21	85.24	10.87	75.14	47.63	0.58	0.6749
Alagoas	Atalaia	AL	3.42	3.75	75.42	97.51	73.11	16.87	67.86	37.78	0.28	0.7449
Alagoas	Barra	AL	3.52	4.07	87.14	96.61	92.21	16.09	66.72	30.99	0.47	0.5438
Alagoas	Barra	AL	4.3	4.17	94.41	99.02	95.31	17.25	57.07	33.82	0.83	0.5975
Alagoas	Batalha	AL	3.53	3.64	62.06	98.46	63.31	17.9	83.67	35.72	0.26	0.5037
Alagoas	Belém	AL	3.3	3.49	55.52	97.93	56.27	6.53	87.15	48.1	0.43	0.5155
Alagoas	Belo	AL	2.09	2.71	27.99	97.54	41.52	6.76	87.83	37.38	0.35	0.5485
Alagoas	Boca	AL	4.12	4.86	85.68	98.96	82.58	62.1	68.6	44.94	0.49	0.6326
Alagoas	Branquinha	AL	3.56	3.97	60.9	98.08	72.16	36.63	76.51	39.15	0.45	0.5502
Alagoas	Cacimbinhas	AL	2.59	2.74	53.21	97.47	24.32	0	80.42	45.82	0.49	0.4708
Alagoas	Cajueiro	AL	3.54	3.78	76.57	98.9	82.12	5.07	72.95	34.93	0.42	0.5396
Alagoas	Campestre	AL	3.82	4.48	73.54	98.94	78.29	61.8	63.67	40.93	0.15	0.5737

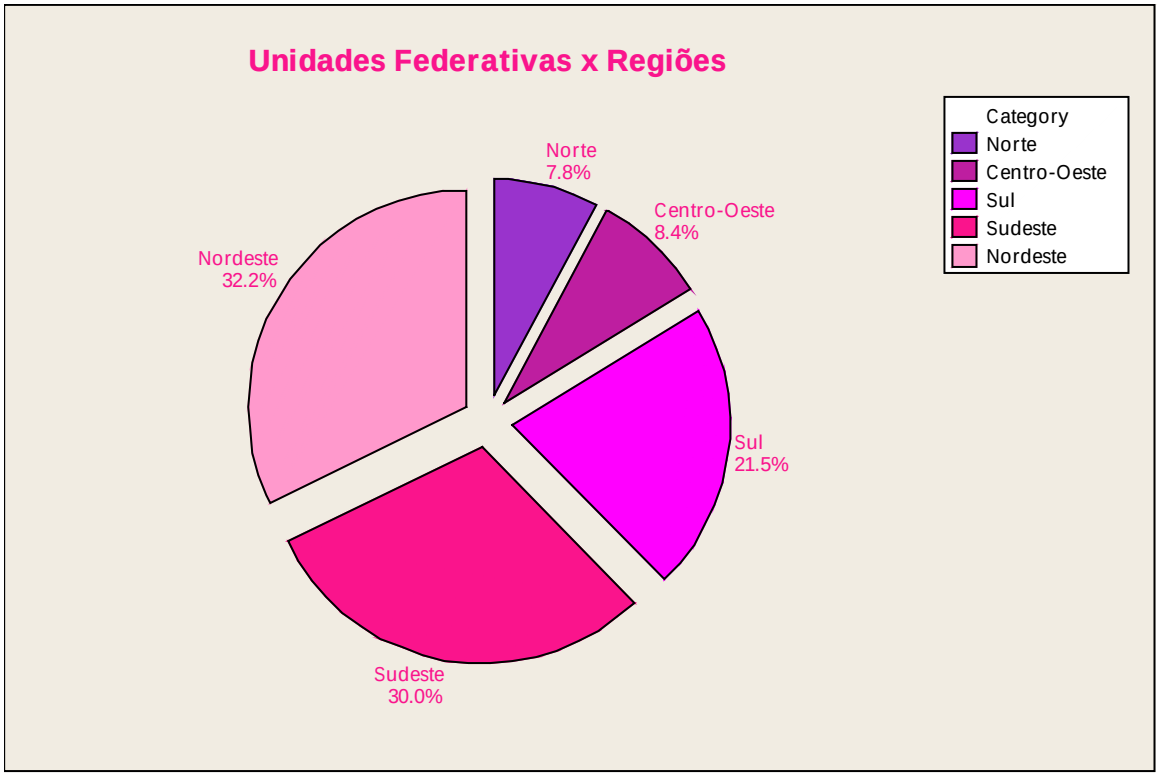
3. ANÁLISE DAS VARIÁVEIS

3.1 Variáveis Categóricas ou qualitativas.

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: "UF" e "UF2"

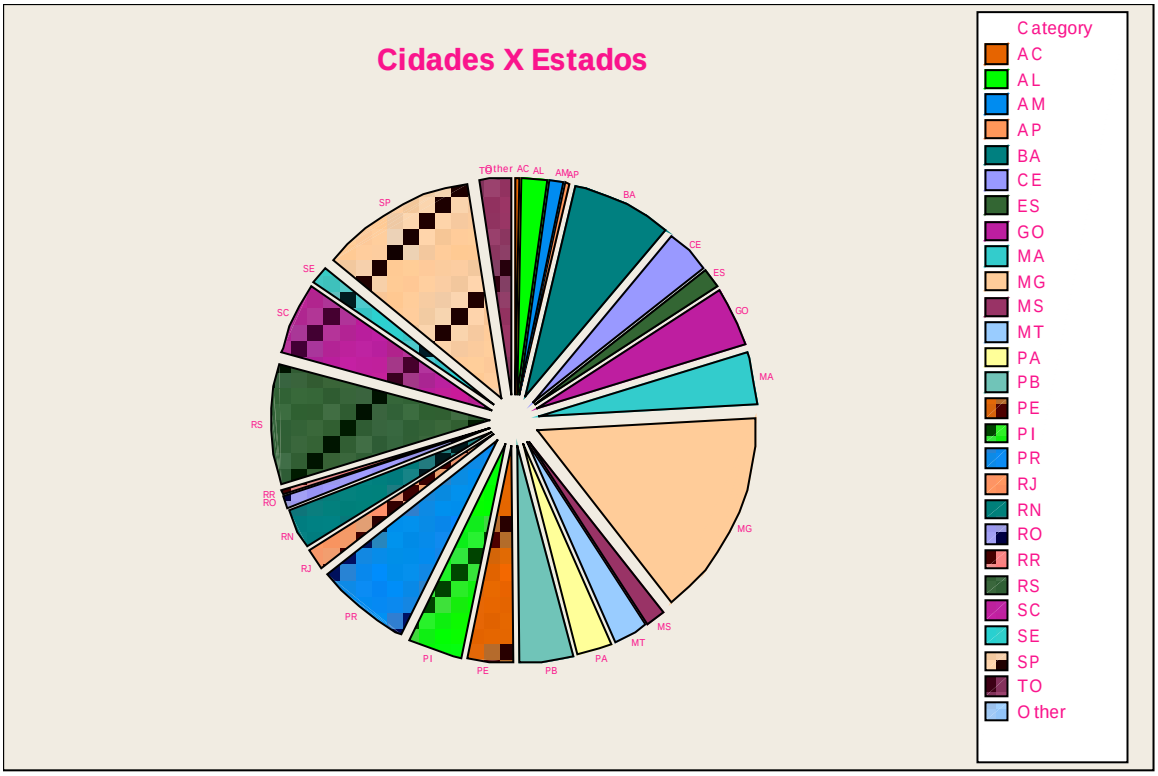
Nossa amostra totaliza 26 unidades federativas e 1 distrito federal. As unidades federativas estão distribuídas em 5 regiões.

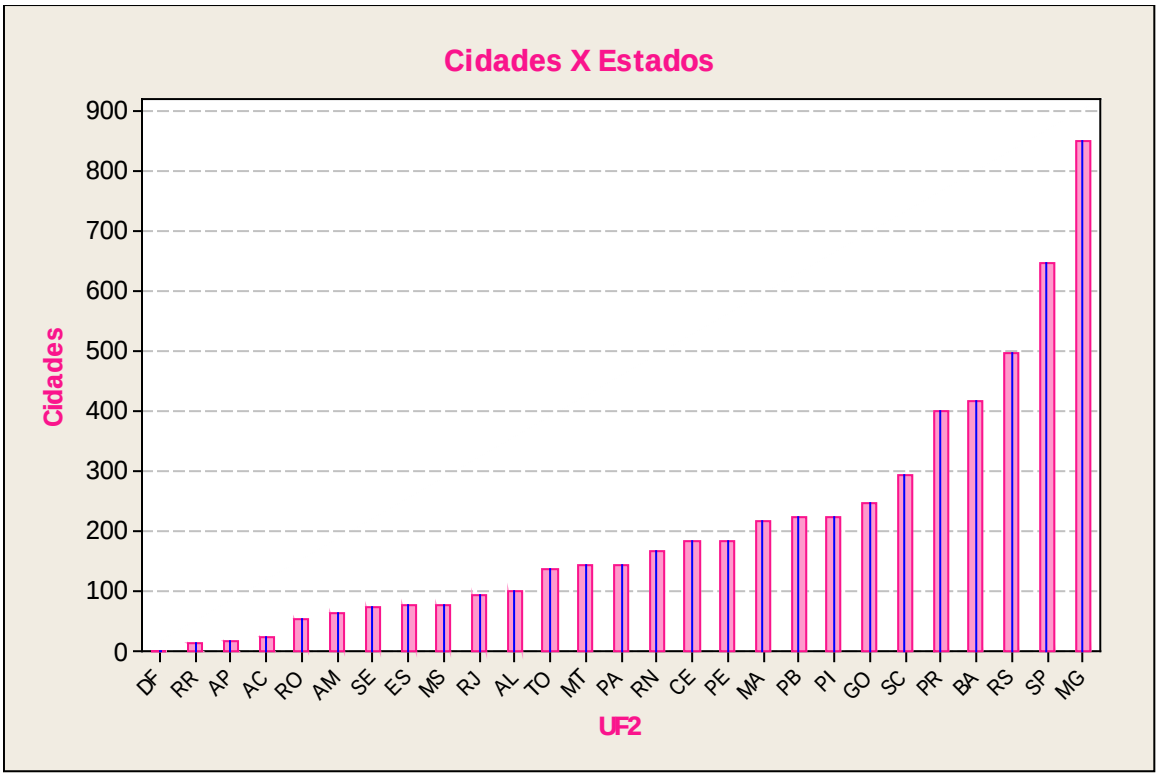


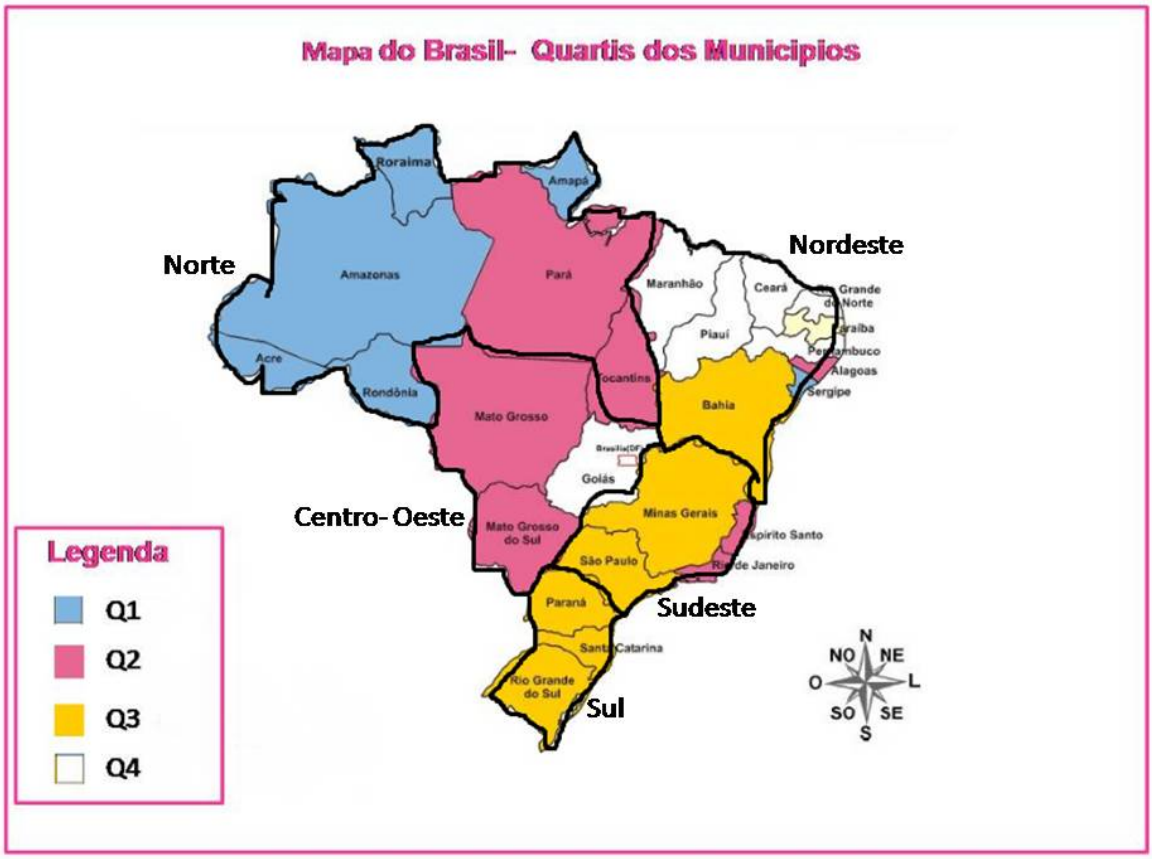
No que diz respeito a relação regiões e cidades pode-se observar no gráfico acima que as regiões Nordeste (32,2%), Sudeste (30,0%) e Sul (21,5%) concentram 83, 7% dos municípios do território nacional, enquanto as demais regiões, Norte (7.8%) e Centro-Oeste (8,4%) somam apenas 16, 2% dos munícipios. Além da concentração dos municípios brasileiros, as três regiões tem em comum o fato de serem as três regiões banhadas significativamente pelo oceano Atlântico. Fato este, que nos ajuda a entender a concentração nestas regiões.

3.1.2 Variável: “Munícipios”

Os gráficos abaixo nos ajudam a entender melhor o comportamento desta variável







Análise:

- O comportamento dos municipios por Unidades Federativas (UF2) não consiste em igualdade conforme demonstra os gráficos acima, pois enquanto o estado de Minas Gerais que contém a maior quantidade de municípios brasileiros tem 851 cidades que correspondem à 15,3 % , Roraima tem apenas 15 municipios que é correspondente à 0,3%.
Portanto Minas Gerais tem 57 vezes mais municípios que Roraima.
A distância aumenta ao considerarmos o Distrito Federal que tem somente uma cidade.

- O Primeiro e o segundo quartil concentram-se nas regiões Norte e Centro-Oeste, de maneira que tem somente dois estados no Sudeste: Rio de Janeiro e Espirito Santo e no Nordeste apenas: Alagoas e Sergipe, exclui-se deste contexto Goiás que corresponde ao quarto quartil Portanto podemos afirmar que nestas regiões concentram-se os estados com menor quantidade de municípios que totalizam 1.015, ou seja, as Regiões Norte e Centro-Oeste somadas aos quatro estados descritos acima correspondem 18% do total de municípios brasileiros.

- No terceiro Quartil os estados possuem a quantidade de municípios entre 167 e 223 concentrados na Região Sul e Sudeste, incluindo a Bahia que pertence à região Nordeste , exclui-se deste contexto Rio de Janeiro e Espirito Santo.
Este quartil é composto por 1.198 municipios que correspondem à 22% do total de municípios brasileiros.

-No ultimo Quartil visualizamos os estados que possuem as maiores quantidades de municípios, com forte concentração na região Nordeste, excluindo-se destes os estados da Bahia, Alagoas e Sergipe e incluimos Goias correspondente à região centro-oeste.
Deste total temos 3.352 municipios que correspondem à 60% do total de municípios brasileiros., portanto a Região Nordeste é composta pelos estados que mais contém municípios.

3.2 VARIÁVEIS QUANTITATIVAS

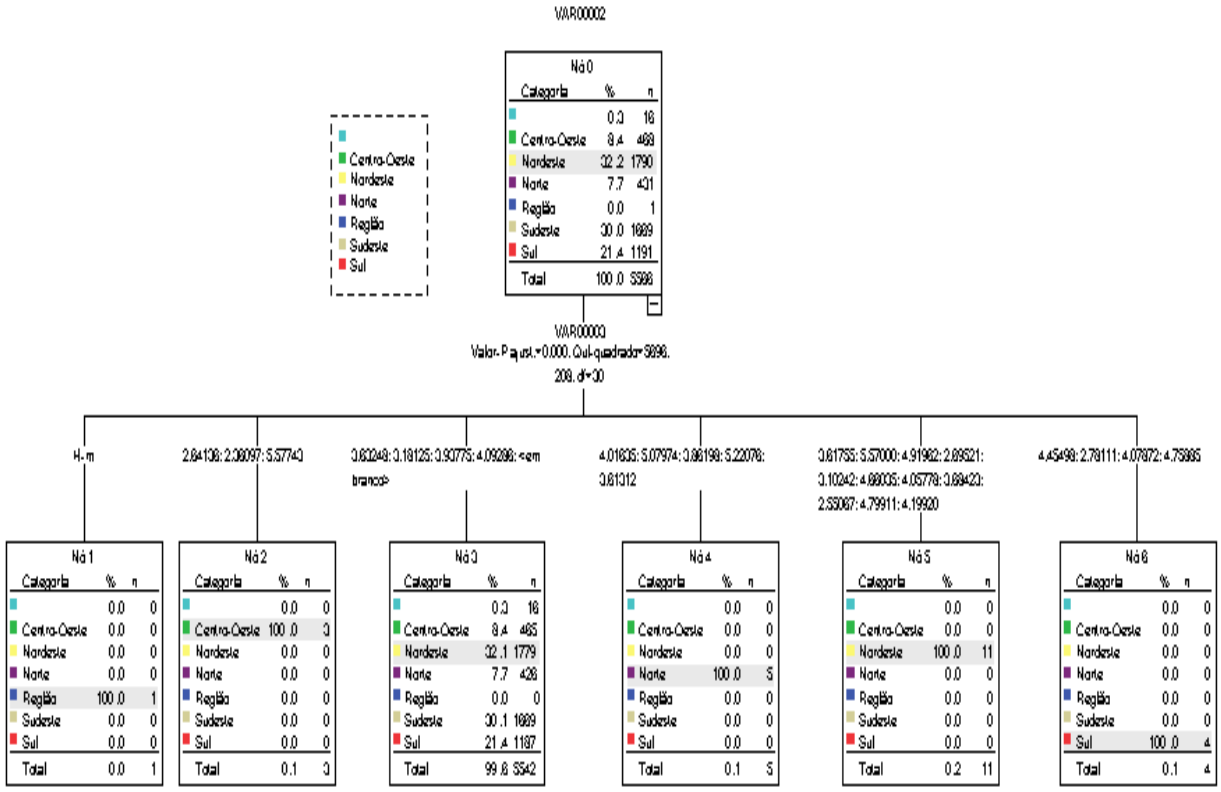
A análise deste tipo de variável permite a utilização de uma maior gama de ferramentas de análise como histogramas, curvas de densidade, gráfico de ramos, box-plot e dot-plot, além de informações numéricas como média, desvio-padrão, mediana, quartis, 5 números, intervalo de confiança e teste de normalidade de Anderson-Darling. Também podemos fazer classificações supervisionadas das variáveis quantitativas, através da análise discriminante.

3.2.1. ÁRVORES DE CLASSIFICAÇÃO DAS VARIÁVEIS HABITAÇÃO

Este resultado se refere à variável dependente REGIÃO e as variáveis independente: ISDM, H, H1, H2, H3, H4, H5, H6,

Resumo do modelo

Especificações	Método de crescimento	CHAID
	Variável dependente	VAR00002
	Variáveis independentes	VAR00001, VAR00003, VAR00004, VAR00005, VAR00006, VAR00007, VAR00008, VAR00009
	Validação	Nenhum
	Profundi00dade de árvore máxima	3
	Casos mínimos em nó pai	2
	Casos mínimos em nó filho	1
	Variáveis independentes incluídas	VAR00003
Resultados	Número de nós	7
	Número de nós de terminal	6
	Profundidade	1



Posto

Observado	Previsto							
		Centro-Oeste	Nordeste	Norte	Região	Sudeste	Sul	Porcentagem Correta
	0	0	16	0	0	0	0	0.0%
Centro-Oeste	0	3	465	0	0	0	0	0.6%
Nordeste	0	0	1790	0	0	0	0	100.0%
Norte	0	0	426	5	0	0	0	1.2%
Região	0	0	0	0	1	0	0	100.0%
Sudeste	0	0	1669	0	0	0	0	0.0%
Sul	0	0	1187	0	0	0	4	0.3%
Porcentagem global	0.0%	0.1%	99.8%	0.1%	0.0%	0.0%	0.1%	32.4%

Método de crescimento: CHAID

Variável dependente: VAR00002

Risco

Estimativas	Modelo padrão
.676	.006

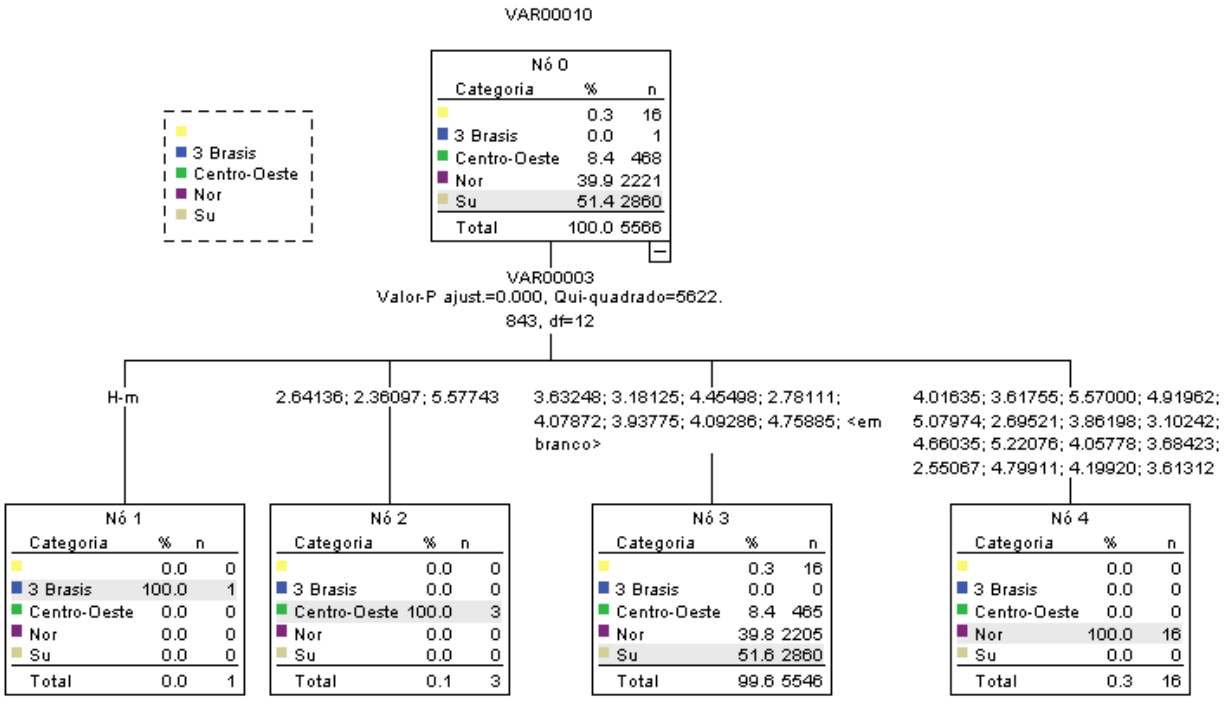
Método de crescimento: CHAID

Variável dependente: VAR00002

Esta baixa % de acertos de 32,4 % se refere à 5 Brasis e as variáveis independente: ISDM, H, H1, H2, H3, H4, H5, H6,

Resumo do modelo

Especificações	Método de crescimento	CHAID
	Variável dependente	VAR00010
	Variáveis independentes	VAR00001, VAR00003, VAR00004, VAR00005, VAR00006, VAR00007, VAR00008, VAR00009
	Validação	Nenhum
	Profundidade de árvore máxima	3
	Casos mínimos em nó pai	2
	Casos mínimos em nó filho	1
	Variáveis independentes incluídas	VAR00003
Resultados	Número de nós	5
	Número de nós de terminal	4
	Profundidade	1



Risco

Estimativas	Modelo padrão
.483	.007

Método de crescimento: CHAID

Variável dependente: VAR00010

Posto

Observado	Previsto					
		3 Brasis	Centro-Oeste	Nor	Su	Porcentagem Correta
	0	0	0	0	16	0.0%
3 Brasis	0	1	0	0	0	100.0%
Centro-Oeste	0	0	3	0	465	0.6%
Nor	0	0	0	16	2205	0.7%
Su	0	0	0	0	2860	100.0%
Porcentagem global	0.0%	0.0%	0.1%	0.3%	99.6%	51.7%

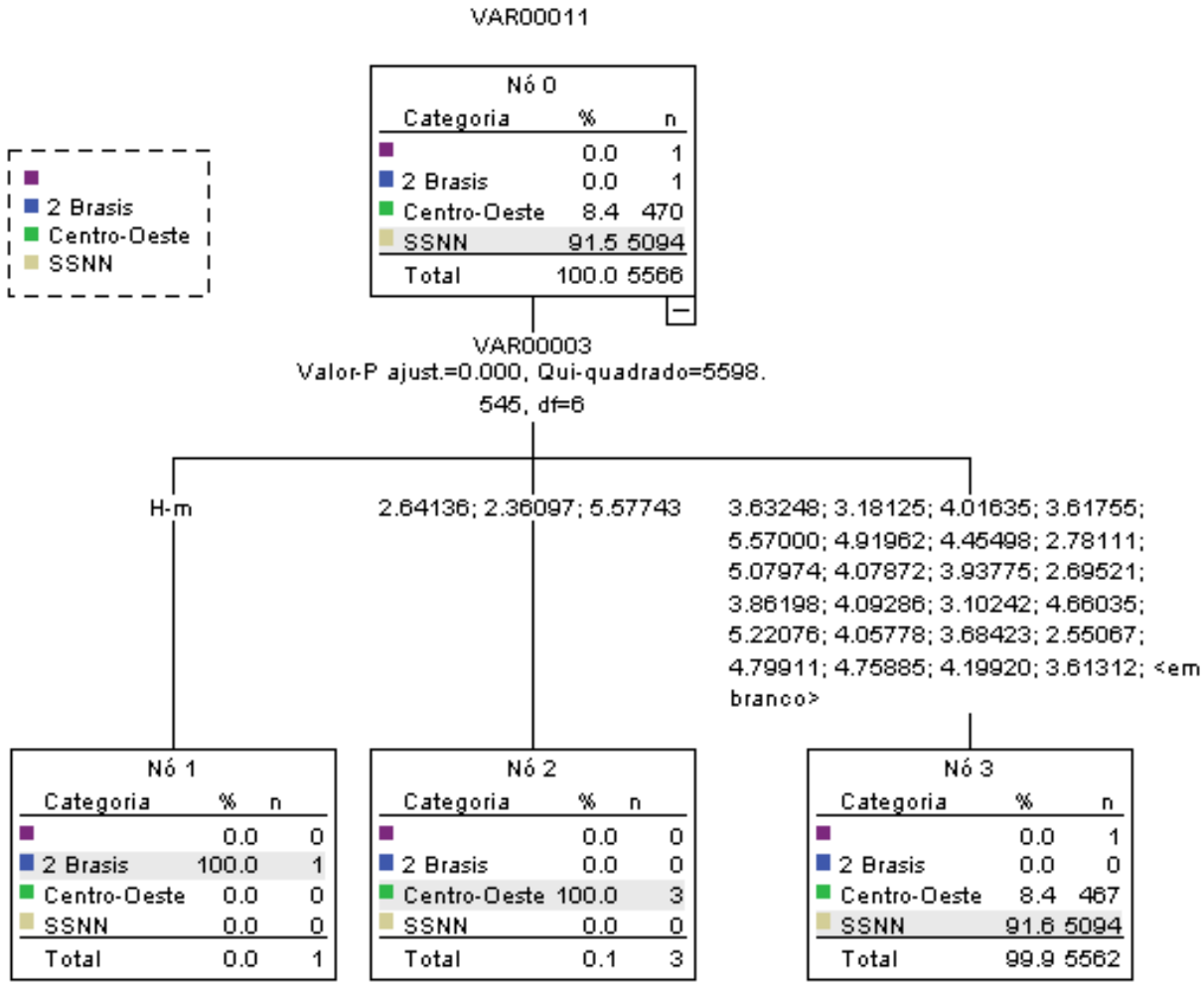
Método de crescimento: CHAID

Variável dependente: VAR00010

Esta % de acertos de 51,7% se refere à variável dependente 3 Brasis e as variáveis: ISDM, H, H1, H2, H3, H4, H5, H6,

Resumo do modelo

Especificações	Método de crescimento	CHAID
	Variável dependente	VAR00011
	Variáveis independentes	VAR00001, VAR00003, VAR00004, VAR00005, VAR00006, VAR00007, VAR00008, VAR00009
	Validação	Nenhum
	Profundidade de árvore máxima	3
	Casos mínimos em nó pai	2
	Casos mínimos em nó filho	1
	Variáveis independentes incluídas	VAR00003
Resultados	Número de nós	4
	Número de nós de terminal	3
	Profundidade	1



Posto

Observado	Previsto				
		2 Brasis	Centro-Oeste	SSNN	Porcentagem Correta
	0	0	0	1	0.0%
2 Brasis	0	1	0	0	100.0%
Centro-Oeste	0	0	3	467	0.6%
SSNN	0	0	0	5094	100.0%
Porcentagem global	0.0%	0.0%	0.1%	99.9%	91.6%

Método de crescimento: CHAID

Variável dependente: VAR00011

Risco

Estimativas	Modelo padrão
.084	.004

Método de crescimento: CHAID

Variável dependente: VAR00011

Esta % de acertos de 91,6 % se refere à 2 Brasis(Centro-Oeste e o resto) e as variáveis: ISDM, H, H1, H2, H3, H4, H5, H6 .

O resultado superou todas as expectativas !

Conclusão:

Observou-se um índice de previsibilidade para dos 2 Brasis (Variável 11), que alcançou 91,6 % de acerto, contra 2 Brasis (Variável 10), que alcançou 51,7 % e 32,4 % para 5 Brasis(Variável 2).