

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

**MÉTODOS ESTATÍSTICOS APLICADOS AO ESTUDO DA DISTRIBUIÇÃO TERRITORIAL DAS
DESIGUALDADES SOCIAIS EM CABO VERDE**

Teresa Brito Oliveira Barros Moniz

Mestrado em Bioestatística e Biometria

2017

UNIVERSIDADE ABERTA



UNIVERSIDADE
AbERTA
www.uab.pt

**MÉTODOS ESTATÍSTICOS APLICADOS AO ESTUDO DA DISTRIBUIÇÃO TERRITORIAL DAS
DESIGUALDADES SOCIAIS EM CABO VERDE**

Teresa Brito Oliveira Barros Moniz

Mestrado em Bioestatística e Biometria

Dissertação orientada pela Prof. Doutora Maria do Rosário Olaia Duarte Ramos e co-orientada pela
Prof. Doutora Sandra Sofia Ferreira da Silva Caeiro

2017

RESUMO

As desigualdades sociais vêm sendo estudadas há vários anos, no entanto a sua mais-valia para a elaboração de estratégias políticas para a sua diminuição iniciou com o desenvolvimento de medidas que as quantificam. O aparecimento da análise e ferramentas de estatística espacial permitiu estudar as desigualdades, não apenas em termos globais, mas também por regiões, por forma a serem comparáveis, e considerar a relação entre o fenómeno e o espaço.

É objectivo deste trabalho explorar técnicas de análise e estatística espacial para caracterizar a distribuição espacial das desigualdades em Cabo Verde. Utilizam-se índices para quantificar as desigualdades sociais em diferentes vertentes, como a desigualdade de género, económica, ambiental e em saúde.

Para a realização do estudo foram utilizados dados oriundos dos inquéritos e recenseamento realizados pelo Instituto Nacional de Estatística de Cabo Verde (INE-CV), dados do Ministério de Saúde e Conselho Nacional das Eleições (CNE), registados dos anos 2000 e 2010 a 2015. A quantificação da desigualdade económica e em saúde, deu-se através do índice de Gini e índice de concentração, respectivamente, e as desigualdades de género e ambiental através de medidas sintéticas. Utilizaram-se medidas de estatística descritiva, a análise de autocorrelação para averiguar a associação espacial entre concelhos, a análise de *Hot Spot* para identificar concelhos com desigualdade elevada e a análise de *clusters*, para identificar concelhos semelhantes. Recorreu-se ao *software* de Sistemas de Informação Geográfica (ArcGIS), para a análise e estatística espacial, SPSS, para análises clássicas (não espacial) e Excel para organização dos dados e cálculos intermédios.

A desigualdade social abrange todo o país, sendo mais concentrada ao norte das ilhas de Santo Antão, Santiago e Fogo. A desigualdade de género apresenta um desvio menos significativo da desigualdade quando comparado a desigualdade ambiental. Os concelhos da ilha de Santiago são semelhantes, com excepção do concelho da Praia. Com o passar dos anos verificou-se semelhança entre os concelhos próximos e que partilham a mesma fronteira.

Palavras-chaves: Desigualdades sociais, índices de desigualdade, estatística espacial, análise de *clusters*, Sistemas de Informação Geográfica

ABSTRACT

The Social inequalities have been studied for several years, however its added value to the elaboration of political strategies for its decrease has begun with the development of measures that quantify them. The emergence of the analysis and tools of spatial statistics has allowed to study inequalities, not only in global terms, but also by regions, in order to be comparable, and to consider the relationship between the phenomenon and space.

It is the objective of this work to explore techniques of analysis and spatial statistics to characterize the spatial distribution of inequalities in Cape Verde. Indexes are used to quantify social inequalities in different strands such as gender inequality, economic, environmental and health.

For the realization of the study, data originating from the surveys and census conducted by the National Institute of Statistics of Cape Verde (INE-CV), data from the Ministry of Health and National Council of Elections (CNE), registered from the years 2000 and 2010 to 2015. The quantification of economic and health inequality is given by the index of Gini and concentration index, respectively, and gender and environmental inequalities through synthetic measures. Descriptive statistical measures were used the autocorrelation analysis to ascertain the spatial association between municipalities, the analysis of Hot Spot to identify municipalities with high inequality and the analysis of clusters, to identify similar municipalities. The software of Geographic Information Systems (ArcGIS), for analysis spatial and statistics, SPSS, for classical (non-spatial) analyses, and Excel for organizing data and intermediate calculations.

Social inequality covers the whole country, being more concentrated north of the islands of Santo Antão, Santiago and fire. Gender inequality shows a less significant deviation from inequality when compared to environmental inequality. The municipalities of the island of Santiago are similar, except for the municipality of the beach. Over the years, there was a similarity between the nearby municipalities and sharing the same frontier.

Keywords: Social inequalities, inequality indexes, spatial statistics, cluster analysis, geographic information systems

AGRADECIMENTOS

Gostaria, primeiramente de agradecer à Deus, por ter-me dado mais uma oportunidade de adquirir conhecimentos e experiências e capacidade para que, hoje, conseguisse ser hábil de chegar até aqui.

À minha mãe (e pai), pelo apoio, cafezinho, paciência, oportunidade de estudar e tempo que ficou tomando conta da minha filha na minha ausência. Obrigada!

Aos meus irmãos, namorado e amigos, que sempre me apoiaram ao longo da minha vida acadêmica. Obrigada!

Não podia de deixar de agradecer, também, as orientadoras desta dissertação, a Prof. Doutora. Maria do Rosário e Prof. Doutora Sandra Caeiro. Obrigada, pelo apoio, tempo e, principalmente, pela paciência ao longo da elaboração da dissertação. Pois, se não fosse esses apoios, não seria possível chegar até aqui. Muito obrigada!

Por fim, agradeço a minha filha, pois, mesmo sem ter a mínima ideia, contribuiu muito ao permanecer acordada, fazendo-me companhia ao longo da noite. Obrigada!

ÍNDICE

RESUMO	ii
ABSTRACT	iii
AGRADECIMENTOS	iv
SIGLAS E ABREVIATURAS	xi
INTRODUÇÃO	1
1. ENQUADRAMENTO TEÓRICO	3
1.1. Análise Espacial.....	4
1.1.1. Análise Espacial de áreas	5
1.2. Estatística Espacial	9
1.2.1. Distribuição direccional geográfica dos dados: Centro, dispersão e orientação	10
1.2.2. Índice de vizinho mais próximo	14
1.2.3. Análise Hot Spot	16
1.3. Estatística Multivariada Espacial.....	18
1.3.1. Análise de Clusters	18
2. ESTUDO DAS DESIGUALDADES SOCIAIS	25
2.1. Revisão da Literatura.....	26
2.1.1. Principais Indicadores – índices e medidas sintéticas	27
2.2. Caso de estudo: Desigualdades sociais em Cabo Verde.....	36
3. MATERIAL E MÉTODOS	39
3.1. Área geográfica de Estudo	40
3.2. Dados – variáveis e fontes	41
3.3. Softwares utilizados.....	41
3.4. Cálculo dos índices de desigualdade	43
3.4.1. Pré-processamento dos dados	43
3.4.1.1. Tratamento de zeros	43
3.4.2. Índice de Desigualdade Económica (IDE).....	44
3.4.3. Índice de desigualdade em Saúde (IDS).....	45
3.4.4. Índice de Desigualdade de Género (IDG)	46
3.4.5. Índice de Desigualdade Ambiental (IDA).....	48
3.5. Análise dos Dados	49
3.5.1. Análise descritiva da distribuição espacial	49
3.5.2. Índice de vizinho mais próximo	51
3.5.3. Autocorrelação espacial.....	51

3.5.4.	Análise Hot Spot	52
3.5.5.	Análise de clusters	53
4.	ANÁLISE DOS RESULTADOS	57
4.1.	Distribuição geográfica dos dados	58
4.1.1.	Distribuição espacial da população de Cabo Verde: Censo 2000 e 2010	58
4.1.2.	Distribuição espacial dos Índices de Desigualdade	59
4.1.3.	Características espaciais da distribuição do rendimento total e per capita, IDE, IDA e IDG	67
4.2.	Autocorrelação espacial	74
4.3.	Análise Hot Spot	80
4.4.	Análise de Clusters (ou de Agrupamento)	83
4.4.1.	Identificação de soluções de clusters: Método índice R^2	86
4.4.2.	Identificação de soluções de clusters: método Mojena (1977)	90
4.4.3.	Resultados: método Mojena (1977)	94
	CONSIDERAÇÕES FINAIS	101
	REFERÊNCIAS BIBLIOGRÁFICAS	105
	ANEXO I – MATERIAL E MÉTODOS DE ANÁLISES	108
	ANEXO II – ESTATÍSTICAS DESCRITIVAS	113
	ANEXO III – DISTRIBUIÇÃO ESPACIAL	120
	ANEXO IV – AUTOCORRELAÇÃO ESPACIAL	130
	ANEXO V – ANÁLISE HOT SPOT	140
	ANEXO VI – ANÁLISE DE CLUSTERS	148

ÍNDICE DE TABELAS

Tabela 1.1:	Exemplo de matriz de proximidade para seis sujeitos	20
Tabela 3.1:	Variáveis e índices de caracterização da desigualdade social e suas respectivas fontes	42
Tabela 3.2:	Classificação dos índices de desigualdade	43
Tabela 4.1:	Estatísticas descritivas dos Índices de Desigualdade, 2000 a 2015	60
Tabela 4.2:	Índice Global de Moran das desigualdades	75
Tabela 4.3:	Ponto referencial de corte do dendrograma, método Mojena (1977)	84
Tabela 4.4:	Corte referencial do dendrograma e número ideal de grupos para cada método, 2000 – 2015	85
Tabela 4.5:	Teste de Levene, Soma dos Quadrados entre e dentro dos <i>clusters</i> e teste F da ANOVA segundo os métodos Mojena e índice R^2	94
Tabela III.1:	Centro médio do rendimento total de Cabo Verde, 2002 e 2007	121
Tabela III.2:	Centro médio do rendimento <i>per capita</i> de Cabo Verde, 2002 e 2007	121
Tabela III.3:	Centro médio do IDE de Cabo Verde, 2002 e 2007	122

Tabela III.4: Distância padrão do IDE, rendimento total e <i>per capita</i> da ilha de Santo Antão, Santiago e Fogo, 2002 e 2007	122
Tabela III.5: Elipse de desvio padrão do rendimento total da ilha de Santo Antão, Santiago e Fogo, 2002 e 2007	122
Tabela III.6: Elipse de desvio padrão do rendimento <i>per capita</i> da ilha de Santo Antão, Santiago e Fogo, 2002 e 2007	123
Tabela III.7: Elipse de desvio padrão IDE da ilha de Santo Antão, Santiago e Fogo, 2002 e 2007	123
Tabela III.8: Centro médio da Desigualdade de Género, Cabo Verde, 2000 a 2015	124
Tabela III.9: Centro médio da Desigualdade de Género, Cabo Verde, 2000 a 2015, continuação.....	125
Tabela III.10: Centro médio da Desigualdade Ambiental, Cabo Verde, 2000 a 2015	126
Tabela III.11: Centro médio da Desigualdade Ambiental, Cabo Verde, 2000 a 2015, continuação.....	127
Tabela III.12: Distância padrão da Desigualdade Ambiental e de Género da ilha de Santo Antão, Santiago e Fogo, 2000 a 2015	128
Tabela III.13: Elipse de distância padrão da Desigualdade de Género da ilha de Santo Antão, Santiago e Fogo, 2002 e 2007	128
Tabela III.14: Elipse de distância padrão da Desigualdade Ambiental da ilha de Santo Antão, Santiago e Fogo, 2002 e 2007	129
Tabela IV.1: score-z índice Local de Morans para o índice de desigualdade de género	131
Tabela IV.2: Score-z índice Global e Local de Moran para o índice de desigualdade de ambiental	135
Tabela V.1: Parâmetro, e suas respetivas descrições, e tipo da ferramenta “ <i>Hot Spot</i> ”	147
Tabela VI.1: Matriz de distância entre concelhos, 2012	149
Tabela VI.2: Matriz de distância entre concelhos, 2013.....	151
Tabela VI.3: Matriz de distância entre concelhos, 2014.....	153
Tabela VI.4: Matriz de distância entre concelhos, 2015.....	155

ÍNDICE DE FIGURAS

Figura 1.1: Exemplo e um gráfico usando o método <i>K-Funtion</i>	16
Figura 1.2: Exemplo de um dendrograma para 10 classes.....	21
Figura 2.1: Exemplo da Curva de Lorenz.....	30
Figura 3.1: Área de estudo – Arquipélago de Cabo Verde.....	40
Figura 4.1: Distribuição espacial da População de Cabo Verde por concelho, pelo método <i>Natural Breaks</i> (Jenks), 2000.	58
Figura 4.2: Distribuição espacial da População de Cabo Verde por concelho, pelo método <i>Natural Breaks</i> (Jenks), 2010	59
Figura 4.3: Distribuição espacial do índice de Desigualdade Económico, 2002	61
Figura 4.4: Distribuição espacial do índice de Desigualdade Económico, 2007	61
Figura 4.5: Distribuição espacial do índice de Desigualdade de Género, 2000	62
Figura 4.6: Distribuição espacial do índice de Desigualdade de Género, 2012	63

Figura 4.7: Distribuição espacial do índice de Desigualdade Ambiental, 2000	64
Figura 4.8: Distribuição espacial do índice de Desigualdade Ambiental, 2010	64
Figura 4.9: Distribuição espacial do índice de Desigualdade Ambiental, 2012	65
Figura 4.10: Distribuição espacial do índice de Desigualdade Ambiental, 2013	65
Figura 4.11: Distribuição espacial do índice de Desigualdade Ambiental, 2014	66
Figura 4.12: Distribuição espacial do índice de Desigualdade Ambiental, 2015	66
Figura 4.13: Centro médio e distância padrão do rendimento total e <i>per capita</i> e IDE da ilha de Santo Antão, 2002 e 2007	69
Figura 4.14: Centro médio e distância padrão do rendimento total e <i>per capita</i> e IDE da ilha de Santiago, 2002 e 2007	69
Figura 4.15: Centro médio e distância padrão da desigualdade ambiental da ilha de Santo Antão, Santiago e Fogo, 2000 a 2015	70
Figura 4.16: Centro médio e distância padrão da desigualdade de Género da ilha de Santo Antão, Santiago e Fogo, 2000 a 2015	70
Figura 4.17: Centro médio e orientação da média do rendimento familiar <i>per capita</i> , total e IDE, 2002 e 2007	72
Figura 4.18: Centro médio e orientação do rendimento total, rendimento familiar <i>per capita</i> . total e IDE, 2002 e 2007	73
Figura 4.19: Centro médio e orientação desigualdade ambiental, 2000 a 2015	73
Figura 4.20: Centro médio e orientação desigualdade de género, 2000 a 2015.....	74
Figura 4.21: Mapa do índice Local Morans ponderado pelo índice de desigualdade económico, 2002	76
Figura 4.22: Mapa do índice Local Morans ponderado pelo índice de desigualdade de género, 2000	78
Figura 4.23: Mapa do índice Local Morans ponderado pelo índice de desigualdade de género, 2010 a 2015.....	78
Figura 4.24: Mapa da autocorrelação espacial para o índice de desigualdade ambiental: índice Local de Moran	79
Figura 4.25: Gráfico <i>K-Function</i> : Distribuição dos dados	80
Figura 4.26: Mapa do <i>Hot spot</i> do IDG, 2010 a 2015.....	81
Figura 4.27: Mapa do <i>Hot spot</i> do IDA, 2000.....	81
Figura 4.28: Mapa do <i>Hot spot</i> do IDA, 2000 a 2015.....	82
Figura 4.29: Dendrograma da análise de <i>clusters</i> dos concelhos, 2000.....	85
Figura 4.30: Dendrograma da análise de <i>clusters</i> dos concelhos, 2010.....	86
Figura 4.31: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2000.....	96
Figura 4.32: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2010.....	96
Figura 4.33: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2012.....	97
Figura 4.34: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2013.....	97
Figura 4.35: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2014.....	98
Figura 4.36: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2015.....	98
Figura I.1: Organização das Informações	109

Figura I.2: Modelos das medidas de distribuição espaciais	109
Figura I.3: Modelo do índice de vizinho mais próximos.....	110
Figura I.4: Modelo da construção do Mapa de autocorrelação espacial: <i>clusters</i> e <i>outliers</i>	110
Figura I.5: Modelo da Autocorrelação de Morans Global	111
Figura I.6: Modelo da Análise de <i>Hot Spot</i>	111
Figura I.7: Modelo de construção do dendrograma usando os índices de desigualdade	112
Figura I.8: Modelo de construção do agrupamento, análise de <i>clusters</i>	112
Figura II.1: Distribuição espacial do índice de Desigualdade de Género, 2010.....	117
Figura II.2: Distribuição espacial do índice de Desigualdade de Género, 2011	117
Figura II.3: Distribuição espacial do índice de Desigualdade de Género, 2013.....	118
Figura II.4: Distribuição espacial do índice de Desigualdade de Género, 2014.....	118
Figura II.5: Distribuição espacial do índice de Desigualdade de Género, 2015.....	119
Figura IV.1: Mapa do índice Local Morans ponderado pelo índice de desigualdade de género, 2010.....	132
Figura IV.2: Mapa do índice Local Morans ponderado pelo índice de desigualdade de género, 2011	132
Figura IV.3: Mapa do índice Local Morans ponderado pelo índice de desigualdade de género, 2012.....	133
Figura IV.4: Mapa do índice Local Morans ponderado pelo índice de desigualdade de género, 2013.....	133
Figura IV.5: Mapa do índice Local Morans ponderado pelo índice de desigualdade de género, 2014.....	134
Figura IV.6: Mapa do índice Local Morans ponderado pelo índice de desigualdade de género, 2015.....	134
Figura IV.7: Mapa do índice Local Morans ponderado pelo índice de desigualdade ambiental, 2000.....	136
Figura IV.8: Mapa do índice Local Morans ponderado pelo índice de desigualdade ambiental, 2010.....	136
Figura IV.9: Mapa do índice Local Morans ponderado pelo índice de desigualdade ambiental, 2012.....	137
Figura IV.10: Mapa do índice Local Morans ponderado pelo índice de desigualdade ambiental, 2013.....	137
Figura IV.11: Mapa do índice Local Morans ponderado pelo índice de desigualdade ambiental, 2014.....	138
Figura IV.12: Mapa do índice Local Morans ponderado pelo índice de desigualdade ambiental, 2015.....	138
Figura IV.13: Mapa do índice Local Morans ponderado pelo índice de desigualdade económico, 2007.....	139
Figura IV.14: Gráfico de resultado da análise do índice de vizinho mais próximo	139
Figura V.1: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade de género, 2000.....	141
Figura V.2: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade de género, 2010.....	141
Figura V.3: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade de género, 2011	142
Figura V.4: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade de género, 2012.....	142
Figura V.5: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade de género, 2013.....	143
Figura V.6: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade de género, 2014.....	143
Figura V.7: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade de género, 2015.....	144
Figura V.8: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade ambiental, 2010.....	144
Figura V.9: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade ambiental, 2012.....	145
Figura V.10: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade ambiental, 2013.....	145

Figura V.11: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade ambiental, 2014.....	146
Figura V.12: Mapa dos <i>hot</i> e <i>cold spot</i> ponderado pelo índice de desigualdade ambiental, 2015.....	146
Figura VI.1: Dendrograma da análise de clusters dos concelhos com índices de desigualdades, 2012	150
Figura VI.2: Dendrograma da análise de clusters dos concelhos com índices de desigualdades, 2013	152
Figura VI.3: Dendrograma da análise de <i>clusters</i> dos concelhos com índices de desigualdades, 2014	154
Figura VI.4: Dendrograma da análise de <i>clusters</i> dos concelhos com índices de desigualdades, 2015	156
Figura VI.5: Gráfico da estatística F- <i>pseudo</i> , 2012	156
Figura VI.6: Gráfico da estatística F- <i>pseudo</i> , 2013	157
Figura VI.7: Gráfico da estatística F- <i>pseudo</i> , 2014	157
Figura VI.8: Gráfico da estatística F- <i>pseudo</i> , 2015	158

ÍNDICE DE GRÁFICOS

Gráfico 4.1: <i>Box-plot</i> dos <i>clusters</i> , determinado pelo método índice R^2 , 2000 a 2015.....	90
Gráfico 4.2: <i>Box-plot</i> dos <i>clusters</i> , determinado pelo método índice R^2 , 2000 a 2015.....	93
Gráfico II.1: <i>Box plot</i> do índice de desigualdade em saúde, 2007	115
Gráfico II.2: <i>Box Plot</i> do índice de desigualdade económica, 2002 e 2007	115
Gráfico II.3: <i>Box Plot</i> do índice de desigualdade económica, 2000 a 2015	116
Gráfico II.4: <i>Box Plot</i> do índice de desigualdade ambiental, 2000 a 2015	116

SIGLAS E ABREVIATURAS

AF – Agregados Familiares

ANN – *Average Nearest Neighbor*

B.O. – Boletim Oficial

BIT – *Bureau International du Travail*

IAA – Índice de Abastecimento de Água

ICH – Índice de Condições de Habitação

IDA – Índice de Desigualdade Ambiental

IDE – Índice de Desigualdade Económico

IDG – Índice de Desigualdade de Género

IDRF – Inquérito às Despesas e Receitas Familiares

IDS – Índice de Desigualdade em Saúde

IE – Inquérito ao Emprego

IEA – Índice de Evacuações de Águas residuais

IG – Índice de Gini

IMC – Inquérito Multi-Objectivo Continuo

IQA – Índice de Qualidade Ambiental

IRS – Índice de Resíduos Sólidos

MUT – Módulo Uso do Tempo

ODM – Objectivo do Desenvolvimento do Milénio

PND – Plano Nacional de Desenvolvimento

PNUD – Programa das Nações Unidas para o Desenvolvimento

QUIBB – Questionário Unificado de Indicadores Básicos de Bem-estar

UNICEF – *United Nations Children's Fund*

INTRODUÇÃO

A desigualdade social é um fenómeno multidimensional que concerne a todos os países e remonta a um passado longínquo. De acordo com Costa (2012), as dinâmicas sociais atuais voltaram a dar grande visibilidade às desigualdades recolocando-as sob intensa atenção pública. Há uma multiplicação de produção e disponibilização de indicadores para medir a desigualdade social e um número crescente de estudos e pesquisas sobre diversos aspectos das desigualdades sociais contemporânea. Esta persistência das desigualdades sociais indica a necessidade de novas propostas de políticas públicas que minimizam este processo (Teague & Wilson, 1995). Em Cabo Verde a informação sobre a temática desigualdade é escassa. Em 2002, o Instituto Nacional de Estatística de Cabo Verde (INE-CV) calculou a desigualdade de rendimento através das despesas *per capita* dos agregados familiares, em que este valor situou em 0,57 (IDRF, 2002). Mas, a desigualdade social não se reduz somente à desigualdade de rendimento, mas também a muitos outros factores.

Existem vários tipos de desigualdades sociais, sendo frequentemente destacados os relacionados com o factor económico, género, posição social e saúde, no entanto, o estudo não está limitado a estes factores, podendo considerar-se o factor geográfico, nomeadamente as regiões de um país. Na produção e disseminação das desigualdades sociais, a existência de padrões de separação socio-espacial diferentes evidência o território como um factor de grande importância. Assim, caracterizar e compreender as desigualdades sociais a nível territorial foi a motivação central para o desenvolvimento desta dissertação.

Neste contexto, os Sistemas de Informação Geográfica (SIG), aliados a base de dados geográficos e às técnicas de análise espacial destes dados, são instrumentos relevantes para a definição de políticas públicas na medida em que permitem produzir, analisar e integrar dados diversos em uma mesma base territorial e visualizar dados georreferenciados.

O objectivo principal deste trabalho é, portanto, quantificar em forma de indicadores as desigualdades sociais (de género, ambiental, económico e em saúde) e explorar as técnicas de análise e visualização de dados e estatística espacial com o propósito de analisar as relações entre aspectos da distribuição espacial e identificar concelhos com semelhanças, em termos da desigualdade social nas vertentes consideradas. Os objectivos específicos remontam a:

- Analisar os indicadores socioeconómicos e demográficos capazes de reflectir nas desigualdades sociais, e produzir uma base de dados com os indicadores disponíveis e úteis para medi-las.
- Determinar as desigualdades sociais em diferentes aspectos: género, ambiente, saúde e económico e, conhecer o nível de desigualdade entre os concelhos;
- Aplicar técnicas da análise e estatística espacial e visualização para diagnosticar e caracterizar as desigualdades a nível territorial e comparar os métodos espaciais com algumas técnicas clássicas;
- Aplicar análise de *clusters* “espacial” a fim de construir grupos (concelhos) com semelhanças nas desigualdades.

A dissertação encontra-se estruturada, para além desta introdução, e das Considerações Finais, em quatro capítulos: no capítulo 1, Enquadramento teórico, são introduzidos os conceitos de Análise Espacial e Estatística Espacial bem como as metodologias mais usadas neste processo que são utilizadas na componente empírica. No capítulo 2, Estudo das Desigualdades Sociais, apresenta-se uma síntese da literatura neste tema e os principais indicadores que são utilizados nas respectivas pesquisas. Faz-se, ainda, uma introdução ao tema das desigualdades sociais em Cabo Verde. No capítulo 3, Material e Métodos é apresentada a área geográfica em estudo, os dados e variáveis e a análise de dados, na qual se inclui a determinação dos índices de desigualdade nas vertentes económica, saúde, género e ambiental. Apresenta-se também a análise de dados realizada com as técnicas de estatística espacial e de agrupamentos. Finalmente, o capítulo 4, Análise dos Resultados, apresenta as análises e descrições dos resultados com as aplicações das técnicas de análise espacial, estatísticas clássicas e análise multivariada espacial. A análise inicia-se com a distribuição geográfica dos dados, onde é apresentada a distribuição espacial dos índices e as características espaciais da distribuição. É apresentada os resultados da autocorrelação espacial e análise de *Hot Spot*, bem como os resultados da análise de *clusters*. A dissertação, também é constituído por uma secção de referências bibliográficas onde estão referenciadas os documentos e artigos consultados e, um anexo com as informações relevantes sobre alguns resultados e processo de construção.

1. ENQUADRAMENTO TEÓRICO

Neste capítulo do Enquadramento Teórico pretendeu-se realizar um resumo sobre a teoria e métodos relativos à análise e estatísticas espaciais, com o objectivo de justificar e fundamentar a proposta central deste trabalho. Portanto, apresenta os conceitos da análise e estatística espacial, destacando alguns princípios básicos, os efeitos e a classificação dos dados espaciais, com maior ênfase para os dados de áreas, bem como algumas ferramentas do Software de Sistemas de Informação Geográfica (SIG) que se utilizou neste trabalho, ArcGIS (em particular o *ArcMap*), para aplicação das técnicas da análise e estatística espacial.

1.1. Análise Espacial

A análise espacial consiste num modelo de informação que permite conhecer os fenómenos, considerando o espaço onde está envolvido. Isto é, compreende um conjunto de técnicas que permite explicar a relação entre um dado fenómeno e o espaço.

No estudo específico sobre as desigualdades sociais, o território Cabo-Verdiano precisa ser amplamente estudado, e neste processo o espaço não pode ser desconsiderado, pois, as desigualdades parecem seguir algum padrão no espaço e no tempo. Assim, a análise espacial mostra-se como uma ferramenta muito importante na medida em que permite dar resposta as questões como: existem grandes diferenças geográficas nas desigualdades? Em que concelho há maior desigualdade? Neste contexto, é de extrema importância recorrer ao uso de Softwares de SIG, como o ArcGIS, na medida em que permite produzir, analisar e integrar dados diversos em uma mesma base geográfica.

Existem diferentes tipos de dados espaciais, que possuem um índice que faz referência a uma área geográfica, geralmente representada em um mapa bidimensional, e que são classificados em quatro categorias (Assunção, 2001):

- Dados de processos pontuais
- Dados de superfícies aleatórios
- Dados de área
- Dados de interacção espacial

Os dados de processos pontuais apresentam os eventos distribuídos no espaço, por forma a ser possível estimar o valor esperado dos eventos por unidade de área, ou seja, estimar a intensidade. As estimativas podem ser calculadas através de interpolações, em função da distância, por modelos determinísticos ou probabilísticos, sendo os segundos mais precisos. Os modelos probabilísticos baseiam-se na autocorrelação ou dependência espacial dos dados geográficos, baseando-se esta dependência para a realização de estimativas. Podem ser por exemplo o estimador *Kriging*, modelos locais de regressão, estimador de Kernel, entre outros. Para os dados de superfície aleatória, a superfície é a variável aleatória (Assunção, 2001) do fenómeno que está sendo estudado. Em dados de área o espaço geográfico é dividido em diferentes regiões e cada região representa uma variável. Em dados de interacção espacial os espaços geográficos são interdependentes, ou seja, existe interdependência entre dois ou mais lugares.

A representação formal dos fenómenos geográficos corresponde a especificações que envolvem um processo computacional de organização de uma taxonomia dos objectos de um dado domínio (representação de uma cidade, de uma linha de água, de uma área protegida, entre outros), para uma forma que permita a sua tradução automática para o SIG (representação em pontos, linhas ou polígonos) (Painho, 1998). As perspectivas de objectos discretos ou superfícies contínuas correspondem a dois modelos conceptuais de representação do fenómeno geográfico. Esta forma de representação associa-se dois principais métodos de representação digital de dados geográficos, o modelo vectorial e o modelo matricial ou *raster* (Painho, 1998).

1.1.1. Análise Espacial de áreas

Os dados de processos observados em áreas são objectos espaciais que ocorrem em eventos agregados por concelho, ilha, distrito de recenseamento, cidades, zonas, dos quais se tem apenas o seu valor por área (como é o caso dos dados analisados neste trabalho). Como exemplos temos o número de furtos/roubos, número de população, número de óbitos, rendimento médio familiar, taxa de alfabetização, entre outros. Nesses tipos de estudo, assim como em qualquer estudo estatístico-demográficos, as técnicas de análise exploratória são de extrema importância, pois, auxiliam no desenvolvimento da modelagem estatística espacial. Essas técnicas ajudam na visualização e extracção de informações de maneira imediata, quando utilizado procedimentos comuns de classificação e visualização de dados espaciais.

A caracterização da dependência espacial é um dos aspectos fundamentais da análise exploratória dos dados. A forma como os dados encontram-se distribuídos no espaço é um indicador significativo de interacção espacial (Anselin, 1992), a qual é revelada através de medidas de associação ou autocorrelação espacial. Por outras palavras, pode-se dizer que a análise exploratória de dados pode fornecer pistas que permite revelar como os valores das áreas estão relacionados no espaço, e estimar a magnitude dessa dependência espacial.

Para cada uma das categorias de dados anteriormente referidas existem diferentes métodos estatísticos para descrever e analisar os dados. As estatísticas utilizadas para descrever os dados de áreas, assim como os de pontos, podem ser subdivididas em três categorias (Levine, 1996 *apud* Krempi, 2004):

- Medidas de distribuição espacial – descrevem o centro, a dispersão, direcção e forma de distribuição de uma variável;

- Medidas de autocorrelação espacial – descrevem a relação entre as diferentes localizações para uma variável simples, indicando o grau de concentração ou dispersão (por exemplo, análise de *clusters*);
- Medidas de associação espacial entre duas ou mais variáveis – descrevem a correlação ou associação entre variáveis distribuídos no espaço, por exemplo, a correlação entre a localização de lojas de bebidas alcoólicas com pontos onde ocorrem muitos acidentes de trânsito.

1.1.1.1. Matriz de Proximidade espacial

A matriz de proximidade espacial, também denominada de matriz de vizinhança, é uma ferramenta básica utilizada para estimar a variabilidade espacial de dados de área.

Dado um conjunto de n áreas $\{A_1, A_2, \dots, A_n\}$, construi-se a matriz $W_{n \times n}$, em que cada um dos elementos ($w_{i \times j}$) dessa matriz representa uma medida de proximidade entre às áreas A_i e A_j . A escolha e a determinação da medida de proximidade da matriz ocorrem com base em critérios adoptados, de entre os quais (Druck et al., 2004):

- Distância – um objecto está próximo do outro caso os centroides de suas áreas estejam a uma distância específica;
- Contiguidade – um objecto está próximo apenas daqueles com os quais compartilha uma fronteira e;
- Vizinhos mais próximos – um objecto está próximo apenas dos n vizinhos mais próximo (considerando seus centroides).

1.1.1.2. Autocorrelação espacial

Em estatística, a autocorrelação é uma medida que quantifica a correlação cruzada de uma variável aleatória, com ela própria, ou seja, mede a associação em relação aos seus vizinhos. No caso da correlação serial, mede-se a correlação de uma variável, em um dado momento, consigo mesma, em um momento posterior. Já a autocorrelação espacial leva em conta o espaço geográfico que está sendo estudado. Portanto, refere-se à correlação espacial entre atributos de uma mesma variável aleatória em diferentes localizações do espaço (Bailey & Gatrell, 1995). Por outras palavras, a autocorrelação espacial, permite estimar o quanto o valor observado de um atributo, em uma determinada localização, é dependente dos valores desse mesmo atributo nas localizações vizinhas.

Existem diferentes formas de medir a autocorrelação espacial, mas as mais difundidas são os índices de Morans I (Global e Local). Estes índices medem a autocorrelação espacial através do produto dos desvios em relação a média, indicando o grau de associação espacial presente no conjunto dos dados. O índice serve como um teste cuja hipótese nula é de independência espacial. O valor do índice varia entre -1 e 1. Quando mais próximo de 1 ou -1 estiver, o índice representa dependência espacial positiva e negativa, respectivamente.

O índice Global de Morans é uma medida descritiva de autocorrelação espacial para as localizações. Entretanto, nas divisões que compõem a área existe modificações de influência em que, haverá necessidade de estudá-las em cada divisão. Neste caso, o mais adequado é o índice Local de Morans. A diferença entre os dois índices é que o índice Local de Morans calcula para cada divisão e Morans Global para toda a região.

Segundo Druck et al. (2004), o objectivo dos dois índices na análise exploratória é verificar como ocorre a dependência espacial, a partir da comparação entre os valores de uma amostra e seus vizinhos. Portanto, um parâmetro utilizado nessas medidas consiste numa matriz de proximidade espacial (ou matriz de vizinhança).

1.1.1.3. Índice Global e Local de Morans

O índice Global e Local de Morans, como referido anteriormente, funcionam como um teste estatístico, que se baseia numa hipótese a testar:

H_0 : não existe dependência espacial

H_1 : existe dependência espacial.

O índice Global consiste num dos indicadores de autocorrelação espacial, que se caracteriza por fornecer medida global de associação espacial existente no conjunto dos dados e, o índice Local caracteriza-se por gerar um índice de associação espacial para cada área considerada, evidenciando àqueles que possuem maiores semelhanças e, portanto, que criam grupos (*clusters*).

O cálculo do indicador global considera a matriz de proximidade de 1ª ordem e, é dada pela expressão (ESRI, 2016):

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{X})(x_j - \bar{X})}{\sum_{i=1}^n (x_i - \bar{X})^2} \quad (1)$$

em que x_i é a *feature* i ,¹ \bar{X} é a média correspondente ao atributo, w_{ij} é o peso espacial entre as *features* i e j , n é o número total de *features*.

O índice Local de Morans é determinado na mesma lógica que o índice Global de Morans. Mas, este é determinado para cada área e, é dada pela expressão (ESRI, 2016):

$$I_i = \frac{x_i - \bar{X}}{S_i^2} \sum_{j=1, j \neq i}^n w_{ij} (x_j - \bar{X}) \quad (2)$$

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^n w_{ij}}{n - 1} - \bar{X}^2. \quad (3)$$

O *score-z* z_{I_i} é determinado da seguinte forma

$$z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{V[I_i]}} \quad (4)$$

onde

$$E[I_i] = -1/(n_i - 1) \quad (5)$$

$$V[I_i] = E[I_i^2] - E[I_i]^2 \quad (6)$$

A interpretação do valor do índice é semelhante à interpretação do valor de correlação entre duas variáveis aleatórias, como já referido. Os valores próximos de zero correspondem a autocorrelação não significativa, entre o valor da *feature* e o valor médio dos vizinhos. Um valor positivo de I implica em autocorrelação positiva, isto é, o valor, elevado ou baixo, de uma *feature* tende a ser semelhante ao valor do atributo de seus vizinhos (essa *feature* é parte de um *cluster*). Um valor negativo para I indica autocorrelação espacial negativa, isto é, o valor do atributo numa região não é dependente dos valores desse mesmo atributo em áreas diferentes. Ou seja, uma *feature* tem *features* vizinhas com valores diferentes; esta *feature* é um *outlier* (casos atípicos, pois, apresenta valores distintos em suas proximidades).

Obtido o índice é necessário validar a estatística, ou seja, testar a hipótese de que o índice é diferente de zero (independência espacial). Os testes comumente utilizados para validar a estatística são a *pseudo-significância* e *scores-z*.

¹ *Features* são atributos, elementos que definem a estrutura de uma classe, as denominadas de *features class*. Ou seja, são, basicamente, a estrutura de dados que vão representar a classe. As *features class* são representadas por um tipo de geometria, que podem ser ponto, linha ou polígono e um conjunto comum de colunas de atributos.

No entanto, muitos testes estatísticos paramétricos exigem que as *features* sejam independentes. Portanto, quando há dependência espacial os resultados podem não ser satisfatórios. Há pelo menos três abordagens para lidar com os problemas de dependência (ESRI, 2016). A primeira abordagem é ignorar o problema baseado no facto de que o teste individual realizado para cada *feature* no conjunto de dados deve ser considerado isoladamente. A segunda é aplicar um procedimento de teste múltiplo clássico, como as correcções de Bonferroni ou Sidak. A terceira abordagem é aplicar a correcção *False Discovery Rate* (FDR) que estima o número de falsos positivos para um determinado nível de confiança e ajusta o valor-p crítico em conformidade. Esta última abordagem, apesar de não ser perfeito, funciona melhor do que considerar que cada teste local é realizado isoladamente, ou aplicar os métodos do teste múltiplos clássicos, excessivamente conservadores. São retiradas da análise as associações mais fracas (ou seja, os valor-p significativos mais elevados – tendo como referência para comparação a estimativa dos falsos positivos).

1.2. Estatística Espacial

A Estatística espacial é um conjunto de métodos de análise de dados em que a informação da localização, ou seja, o espaço geográfico, é usada na análise. É a área da estatística que estuda os fenómenos no espaço geográfico. Segundo Assunção (2001), a estatística espacial

“é um ramo da estatística que estuda métodos científicos para a coleta, descrição, visualização e análise de dados que possam ser modelados como processos estocásticos, onde o espaço índice é um conjunto de dimensão maior que um”.

(Assunção, 2001)

O desenvolvimento computacional tem dando grandes avanços no desenvolvimento da estatística espacial. Portanto, compreender os fenómenos que ocorrem no espaço constitui cada vez mais um desafio para o esclarecimento de questões em diferentes áreas do conhecimento, tais como, saúde, segurança, ambiente, migrações, entre outras. A disponibilidade de sistema de informação geográfica (SIG) tornou estes estudos cada vez mais comum, devido ao baixo custo e a interface amigáveis.

Segundo Paelinck e Klaassen (1979 *apud* Krempi, 2004), existem cinco princípios básicos no campo da estatística espacial:

- Interdependência – qualquer modelo espacial tem que se caracterizar por sua interdependência, ou seja, devem ser incorporadas as relações recíprocas entre as observações e as variáveis. Um exemplo pode ser verificado em áreas onde

residem pessoas com elevado nível de rendimento, uma vez que seus gastos não se restringem à área em que vivem, mas ocorrem também nas localizações vizinhas, igualmente estimulando o crescimento económico desta última.

- Assimetria – as relações espaciais são em princípio assimétricas. Um exemplo ocorre no comércio, onde a probabilidade de um residente da periferia ir fazer compras no centro é maior do que a do centro (ou próximo dele) ir fazer compras na periferia.
- Alotropia – as causas de um fenómeno espacial podem não se manifestar directamente no lugar onde ele ocorre. Assim, por exemplo, os fenómenos migratórios se explicam não só pela comparação de vantagens e inconvenientes nos espaços de origem e destino, mas também devido as causas que ocorrem em outros lugares do espaço (nível dos salários, por exemplo).
- Não linearidade – a não linearidade de soluções espaciais óptimas obtidas à priori conduz a modelos que requerem especial atenção para sua especificação, uma vez que estes poderão ter comportamento não linear.
- Inclusão de variáveis topológicas – um modelo espacial deve incorporar variáveis topológicas: por exemplo, coordenadas, distâncias, superfícies e densidade.

Apesar da sua importância, segundo Krempf (2004) nem sempre será possível observar estes cinco princípios na construção dos modelos espaciais.

1.2.1. Distribuição direccional geográfica dos dados: Centro, dispersão e orientação

Um valor amostral (uma estatística) representa uma imagem das medidas da distribuição subjacente aos dados, tais como o valor central e a dispersão em torno desse valor, bem como a sua orientação, pode servir para a visualização dos dados, sendo um apoio útil na análise global. Estas medidas podem ser determinadas usando valores dos atributos como ponderadores nas distribuições, pois, num conjunto de dados em que seu centro médio apresenta uma determinada localização, pode não ser o mesmo quando as *features* são ponderadas. Assim, é possível encontrar realidades diferentes do conjunto de *features* em causa.

A medida que indica a localização central apresenta três tipos, a *feature* central, o centro mediano e o centro médio. A *feature* central permite identificar a *feature* que mostra menor distância para todas as outras. O centro mediano apresenta o local com menor distância para todas as *features* e, o centro médio é a média das coordenadas *x* e *y* de todas as *features* na área de estudo.

Assim, pode-se dizer que estas medidas se aproximam das medidas de tendências centrais, neste caso podem ser denominadas de medidas de tendências centrais espaciais, pois, estes envolvem o espaço geográfico do fenómeno que se quer representar.

O centro médio é dado pela seguinte expressão

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad \bar{Y} = \frac{\sum_{i=1}^n y_i}{n}, \quad (7)$$

em que x_i e y_i são as coordenadas das *features* i e n é o número total das *features*.

A expressão do centro médio ponderado é

$$\bar{X}_w = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i} \quad \bar{Y}_w = \frac{\sum_{i=1}^n y_i w_i}{\sum_{i=1}^n w_i}, \quad (8)$$

onde w_i é o peso da *feature* i .

A dispersão é representada graficamente por um círculo ao redor do centro médio anteriormente referido. E, sendo ela a dispersão das observações, neste caso *features*, em torno do centro médio (medida de tendência central), que no fundo é uma extensão do conceito de desvio-padrão para uma população multivariada. Definimos “distância padrão” pela fórmula:

$$SD = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n} + \frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n}} \quad (9)$$

onde x_i e y_i são coordenadas das *features* i , $\{\bar{X}, \bar{Y}\}$ representam o centro médio das *features* e n é o número total das *features*. A distância padrão ponderada se estende

$$SD_w = \sqrt{\frac{\sum_{i=1}^n w_i (x_i - \bar{X}_w)^2}{\sum_{i=1}^n w_i} + \frac{\sum_{i=1}^n w_i (y_i - \bar{Y}_w)^2}{\sum_{i=1}^n w_i}}, \quad (10)$$

onde w_i é o peso da *feature* i e $\{\bar{X}_w, \bar{Y}_w\}$ representam o centro médio ponderado.

O valor obtido pelas expressões é a distância que corresponde ao raio do círculo que é gerado no mapa. Quando maior for o resultado maior será a dispersão das *features* e quando menor o resultado menor será a dispersão.

Finalmente, a medida orientação das *features* no espaço, é onde os valores determinados são representados graficamente por uma elipse (Dias, 2013). A distância padrão é calculada separadamente nas direcções x , y e z . Estas medidas definem os eixos de uma elipse (ou elipsóide) que engloba a distribuição das *features*. A elipse é referida como o *standard deviational ellipse*, (do

inglês, elipse de desvio padrão) uma vez que o método calcula o desvio padrão das coordenadas x e y do centro médio para definir os eixos da elipse. A elipse ou elipsóide permite verificar se a distribuição das *features* é ampliada e, portanto, tem uma orientação particular.

A elipse é determinada por meio de três parâmetros: o ângulo de rotação, a dispersão do eixo maior e do eixo menor. Em termos matemáticos, a elipse de desvio padrão origina-se da distribuição bivariada que se caracteriza por dois tipos de desvio padrão, nas direcções X (longitudes) e Y (latitudes), ortogonais e que definem uma elipse, expressa da seguinte forma (Ebdon, 1985 *apud* Tartaruga, 2008):

$$\sqrt{\frac{\sigma_x^2 + \sigma_y^2}{2}} \quad (11)$$

em que σ_x e σ_y é o desvio padrão na direcção X e Y , respectivamente.

A elipse de desvio padrão é determinada em duas etapas. Na primeira etapa, a orientação dos eixos que definem a elipse é calculada de modo que a soma dos quadrados das distâncias entre os pontos da distribuição e os eixos da elipse seja mínimo. Portanto, obtém-se o ângulo (θ) desta orientação por meio da seguinte equação (Tartaruga, 2008):

$$\theta = \text{Arctang} \frac{\pi + [\pi^2 + 4 \times (\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))^2]^{\frac{1}{2}}}{2 \times \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})} \quad (12)$$

onde

$$\pi = \sum_{i=1}^n (x_i - \bar{x})^2 - \sum_{i=1}^n (y_i - \bar{y})^2, \quad (13)$$

em que θ é o ângulo de rotação da elipse de desvio padrão, \bar{x} e \bar{y} são os centros médios das coordenadas X e Y , respectivamente, x_i e y_i são as coordenadas X e Y , respectivamente, do centroide do concelho i e n é o número de centroides (concelhos).

Na segunda etapa são determinados dois desvios padrões (σ_x , σ_y), utilizando os eixos X e Y , por meio das duas equações seguintes (Tartaruga, 2008):

$$\sigma_x = \sqrt{\left\{ 2 \times \sum_{i=1}^n \frac{[(x_i - \bar{x}) \cos \theta - (y_i - \bar{y}) \sin \theta]^2}{n} - 2 \right\}} \quad (14)$$

$$\sigma_y = \sqrt{\left\{ 2 \times \sum_{i=1}^n \frac{[(x_i - \bar{x}) \sin \theta - (y_i - \bar{y}) \cos \theta]^2}{n} - 2 \right\}} \quad (15)$$

em que σ_x e σ_y são os desvios padrão na direção X e Y, respectivamente. Desta forma, para além do centro médio, a elipse de desvio padrão define-se pelos seguintes parâmetros:

- Ângulo de rotação = θ
- Comprimento do eixo X = σ_x
- Comprimento do eixo Y = σ_y .

O cálculo ponderado da elipse do desvio padrão da distribuição pontual no espaço é matematicamente, adicionando às equações anteriores o peso correspondente ao centroide do concelho i , expressas das seguintes formas (Tartaruga, 2008):

$$\theta = \text{Arctang} \frac{\pi_w + \left[\pi_w^2 + 4 \times (\sum_{i=1}^n w_i^2 (x_i - \bar{x})(y_i - \bar{y}))^2 \right]^{\frac{1}{2}}}{2 \times \sum_{i=1}^n w_i^2 (x_i - \bar{x})(y_i - \bar{y})} \quad (16)$$

onde

$$\pi_w = \sum_{i=1}^n w_i (x_i - \bar{x})^2 - \sum_{i=1}^n w_i (y_i - \bar{y})^2, \quad (17)$$

$$\sigma_{xw} = \sqrt{\left\{ 2 \times \frac{\sum_{i=1}^n [w_i (x_i - \bar{x}) \cos \theta - w_i (y_i - \bar{y}) \sin \theta]^2}{w_i} \right\}} \quad (18)$$

$$\sigma_{yw} = \sqrt{\left\{ 2 \times \frac{\sum_{i=1}^n [w_i (x_i - \bar{x}) \sin \theta - w_i (y_i - \bar{y}) \cos \theta]^2}{w_i} \right\}} \quad (19)$$

Esses métodos, de estatística espacial, apresentados aqui, em termos gerais, podem auxiliar na determinação de padrões tendenciais ou de relações da distribuição. Em termos específicos, visam responder as seguintes questões com respeito à distribuição espacial de

fenómenos (dados quantitativos): onde está o centro médio? Como os dados estão dispersos ao redor do centro? Além disso, o uso destes métodos é indicado para a realização de comparações entre distribuições de diferentes elementos, por exemplo, de homens e de mulheres, ou entre distribuições ao longo do tempo (Tartaruga, 2008).

1.2.2. Índice de vizinho mais próximo

O índice de vizinho mais próximo foi desenvolvido por dois botânicos, Philip Clark e Frances Evan, com o intuito de quantificar padrões em distribuições espaciais (Mitchell, 2005; Smith & Bruce 2010 *apud* Dias, 2013). Ele possibilita verificar se as distribuições espaciais são aleatórias, concentradas ou dispersas, usando como hipótese nula de que a distribuição dos pontos é aleatória. É expresso como a razão da distância observada dividida pela distância esperada (20). A distância esperada é a distância média entre vizinhos numa distribuição aleatória hipotética. Se o índice for menor que 1 o padrão exhibe agrupamento, se for maior, a tendência é para a dispersão.

A ferramenta *Average Nearest Neighbor* (ANN) do ArcGIS calcula este índice, retornando cinco valores como resultados: a distância média observada (\bar{d}_o) e esperada (\bar{d}_e), índice de vizinho mais próximo, score-z e o valor-p. Nesta, o score-z é baseado no cálculo da hipótese nula de randomização.

$$ANN = \frac{\bar{d}_o}{\bar{d}_e} \quad (20)$$

onde

$$\bar{d}_o = \frac{\sum_{i=1}^n d_i}{n} \quad (21)$$

$$\bar{d}_e = \frac{1}{2\sqrt{n/A}} \quad (22)$$

em que d é a distância entre a *feature* i e seus vizinhos mais próximos, A é a área total e, n é o número total de *feature*.

A partir do resultado se pode chegar a três possíveis conclusões (Dias, 2013):

- Quando o índice vizinho mais próximo for 1 a distribuição observada é aleatória;
- Para um índice inferior a 1 a distribuição é concentrada. Quando mais próximo de zero mais concentrada é a distribuição
- E, para um índice superior a 1 a distribuição é dispersa.

A credibilidade do índice é averiguada através do score-z (23) ou valor-p. Para ambos, deve-se ter em conta sempre o nível de significância fixado. O score-z é uma medida de significância estatística que nos diz se devemos ou não rejeitar a hipótese nula.

$$z = \frac{\bar{d}_o - \bar{d}_e}{SE} \quad (23)$$

$$SE = \frac{0,26136}{\sqrt{n^2/A}} \quad (24)$$

O índice de vizinho mais próximo inclui a distância para um único vizinho mais próximo de cada *feature*. Portanto, quando se pretende determinar o índice incluindo todas as distâncias este método não é apropriado. Assim sendo, ao invés do índice de vizinho mais próximo é utilizada a função K (*K-Function*) que inclui todos os vizinhos que estejam dentro de uma determinada distância.

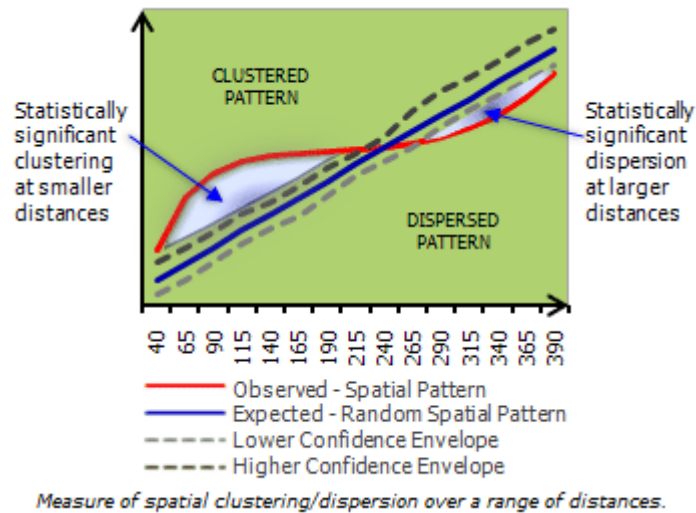
O *K-Function* encontra, primeiramente, a distância de cada *feature* para cada *feature*, e em seguida, para cada *feature*, conta o número de *feature* dentro da distância especificada. Se o número de *features* encontrado, dentro da distância, for maior do que para uma distribuição aleatória, a distribuição é concentrada, e o inverso (disperso) se for menor. Por último, os valores são automaticamente calculados as várias distâncias e apresentada num gráfico, permitindo observar a evolução do padrão de distribuição de *features* à medida que a distância especificada vai aumentando, ou seja, permite verificar a que distância a concentração é maior (exemplo do gráfico na Figura 1.1).

A expressão dada para o *K-Function* é a seguinte (ESRI, 2016):

$$L(d) = \sqrt{\frac{A \sum_{i=1}^n \sum_{j=1, j \neq i}^n k(i, j)}{\pi n(n-1)}}, \quad (25)$$

onde d é a distância entre cada *target feature* (*feature* alvo) i e todas as outras *features* j , A é a área, n é o número de *features* e $k(i, j)$ é o peso de valor 0 ou 1 (se não há correção de limite), sendo 1 quando o vizinho está dentro da distância do *target feature* e zero caso contrário. Quando a correção de limite é aplicada, o peso $k(i, j)$ é ligeiramente modificado.

Figura 1.1: Exemplo e um gráfico usando o método *K-Função*.



Fonte: *Multi-Distance Spatial Cluster Analysis (Ripley's K Function)*, Estatística Espacial, Tutorial ArcGIS 10.4

1.2.3. Análise *Hot Spot*

A análise de *Hot Spot* permite determinar se existe algum grupo com significância estatística no padrão espacial dos dados. Ela atribui um grau de certeza, a confiança de que os pontos pertencem a algum grupo, ou seja, está agrupada, identificada em algum grupo. Portanto, encontrar um grupo com significância estatística nos dados tem-se acesso a informação relevante e confiável.

Esta análise funciona num contexto de vizinhança e tem por base o cálculo da estatística *Getis-Ord G_i^** (26) que identifica quais os grupos são estatisticamente significativa (ESRI, 2016)

$$G_i^* = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j}, \quad \forall j \neq i, \quad (26)$$

em que x_i e x_j são, respectivamente, os valores das *features* i e j , w_{ij} é o peso espacial entre as *features* i e j e n é o número total de *features*.

Um ponto importante é a definição da distância de cálculo, uma vez que, este método apresenta resultados apenas às *features* vizinhas dentro da distância escolhida (Mitchell, 2005 *apud* Dias, 2013). O *ArcMap* tem disponíveis duas distâncias de cálculo que são retratadas mais a frente, as distâncias Euclidiana e Manhattan. Um *hot* ou *cold spot* é considerado estatisticamente significativo se o seu valor é alto (ou baixo) e é rodeado por vizinhos com valores também altos (ou baixos).

O score-z para a estatística é determinado (ESRI, 2016)

$$z_G = \frac{G - E[G]}{\sqrt{V[G]}}. \quad (27)$$

Os scores-z são números de desvios-padrão acima ou abaixo da média populacional. Se por exemplo, uma ferramenta retorna um score-z de 1,8, isto significa que o dado está acima da média em um desvio de 1,8. Assim como o valor-p, o score-z também é associado à distribuição normal padrão. Para score-z positivo, e quando maior for o valor, maior é a aglomeração de valores de *hot spot*, e quando o score-z for negativo e menor for o valor, a aglomeração de *cold spot* é maior.

No entanto, a escolha do método de conceptualização de relações espaciais entre as *features* é fundamental, na medida em que a sua escolha influencia, na maioria das vezes, os resultados finais, pois, as aplicabilidades e regras intrínsecas em cada método são diferentes. Para esta análise é possível escolher seis métodos (Dias, 2013)

- Distância inversa: para esta distância as *features* de vizinhanças mais próximas têm uma influência maior sobre os cálculos para um *target feature* do que as *features* mais distantes;
- Quadrado da distância inversa: esta difere da distância inversa pelo facto da sua influência diminuir mais rapidamente e apenas os vizinhos mais próximos de uma *target feature* exerce influência sobre os cálculos;
- “Banda” de distância fixa: neste método cada *feature* é analisada no contexto de *features* vizinhas. *Features* vizinhas dentro da distância especificada recebe um peso de valor 1 e os que estão fora recebem valor 0. Assim, as *features* dentro da distância exercem influência sobre os cálculos e os que estão fora não têm nenhuma influência.
- Zona de indiferença: neste método, as *features* dentro da distância especificada de uma *target feature* recebe um peso de valor 1 e influencia os cálculos. Uma vez que a distância é excedida, os pesos diminuem com a distância.
- Contiguidade de polígono (1ª ordem): neste método apenas as *features* polígonos vizinhas que compartilham uma fronteira vão influenciar os cálculos para a *target feature* polígono.
- Ficheiro de pesos espaciais: neste método as relações espaciais são definidas por um ficheiro de pesos espaciais com extensão .wmf.

Da mesma forma que acontece na análise estatística não espacial, as técnicas inferenciais no domínio espacial são usadas com base em hipóteses estatísticas. É fixado um nível de confiança

antes de se proceder à estimação e testes espaciais. Os níveis de confiança comum usados são 90%, 95% ou 99%.

Com a utilização desta ferramenta é criada uma nova *feature* com score-z, valor-p e um bin de nível de confiança (Gi_Bin). O Gi_Bin -3 ou 3 reflecte o nível de significância de 99%, o valor -2 ou 2 o nível de confiança de 95%, os valores -1 e 1 o nível de significância de 90%, e zero indica que o valor não é estatisticamente significativo.

1.3. Estatística Multivariada Espacial

A análise Multivariada, segundo Alencar (2009), tem um vasto significado na geografia, pois, possibilita exames de um conjunto de informações permitindo esclarecer os fenómenos geográficos, estudar as tendências e padrões espaciais, formular modelos e realizar previsões.

Portanto, diversas técnicas multivariadas podem ser utilizadas na organização, classificação e análise de dados geográficos. Neste trabalho, é abordada uma das técnicas, por meio dos resultados e mapas obtidos, utilizando a Análise de *Clusters*.

1.3.1. Análise de *Clusters*

A Análise de *Clusters*, ou Agrupamento, é uma técnica estatística utilizada para agrupar sujeitos ou variáveis com valores ou comportamentos semelhantes. O agrupamento de sujeitos, em *clusters* homogêneos a partir de medidas de dissemelhança, pode ser realizado através de duas técnicas: a Hierárquica e a Não-hierárquica. Através da técnica hierárquica é possível utilizar dois métodos: o aglomerante e o divisivo. No primeiro método, cada sujeito é considerado um *cluster* e depois vão sendo agrupados de acordo com as suas proximidades. No segundo acontece o oposto, ou seja, todos os sujeitos fazem parte do mesmo grupo (existe um único grupo) e depois vão sendo divididos para grupos distintos de acordo com as suas distâncias (Alencar, 2009; Marôco, 2014).

As técnicas hierárquicas permitem agrupar sujeitos ou variáveis, enquanto as não-hierárquicas permitem agrupamento apenas de sujeitos. Contudo, em ambas as técnicas, a escolha da medida da distância é fundamental, pois, definem os critérios que avaliam se dois sujeitos (ou variáveis) farão parte de um mesmo grupo ou não. Estas medidas denominam-se de medidas de semelhança e dissemelhança (Alencar, 2009; Marôco, 2014).

1.3.1.1. Medidas de Semelhança e de Dissemelhança

Para construir grupos a partir de um conjunto de sujeitos na Análise de *Clusters*, é necessário utilizar medidas de proximidade que possibilite a comparação entre os componentes desse conjunto. As medidas fornecem distância dimensional entre os sujeitos, possibilitando verificar o quanto os sujeitos são semelhantes pelas suas distâncias. Estas medidas são conhecidas como medidas de semelhança ou dissemelhança, sendo esta última abordagem a adoptada nesta dissertação. Porém, em alguns momentos, chamaremos a medida utilizada simplesmente de distância.

Quando é utilizado o valor mais elevado de uma medida para indicar que dois sujeitos são parecidos, diz-se que se usa o critério de semelhança, em contrapartida, quando se utiliza o menor valor está-se a utilizar uma medida de dissemelhança. A correlação é um exemplo de uma medida de semelhança e a distância euclidiana, uma das medidas mais utilizadas, é um exemplo da medida de dissemelhança. Não se pode dizer que uma determinada medida é melhor que a outra, o que se sabe é que uma determinada medida se adapta melhor a determinados tipos de variáveis ou situações. Em geral, as medidas de dissemelhança são mais adequadas para variáveis quantitativas, e as de semelhanças para as variáveis qualitativas.

As medidas de dissemelhança normalmente utilizadas são (Marôco, 2014):

- Distância Euclidiana: esta medida métrica mede o comprimento do segmento de recta que une duas observações num espaço p -dimensional. Define-se por

$$d_{ij} = \left(\sum_{k=1}^P (x_{ik} - x_{jk}) \right)^{1/2} \quad (28)$$

x_{ik} e x_{jk} correspondem aos valores observados na k -ésima variável nos sujeitos i e j , respectivamente, e P é o número de variáveis observadas;

- Distância Manhattan: esta medida mede a distância numa configuração rectilínea, numa configuração de cidade. Dados dois sujeitos, a distância Manhattan define-se por

$$d_{ij} = \sum_{k=1}^P |x_{ik} - x_{jk}|. \quad (29)$$

- Distância MinKowski: esta medida é considerada uma generalização da distância euclidiana e da distância Manhattan. Pois, a métrica euclidiana é substituída por

uma nova métrica, onde a distância entre dois sujeitos é a soma das diferenças absolutas de suas coordenadas. Ela é dada pela expressão

$$d_{ij} = \left[\sum_{k=1}^P |x_{ik} - x_{jk}|^r \right]^{1/r} \quad r \geq 1 \quad (30)$$

Quando r assume valor igual a 1 obtêm-se a distância de Manhattan, quando assume valor igual a 2 obtêm-se a distância euclidiana, e quando vai ao infinito é denominada de distância Chebyshev;

- Distância de Mahalanobis: esta medida mede a distância estatística entre dois sujeitos. Ela difere da distância euclidiana pelo facto de levar em conta as correlações do conjunto de dados e não depende da escala das medições.

É ainda de referir que algumas das distâncias, como por exemplo a distância euclidiana, é influenciada pela existência de *outliers* e, também, pelas diferenças das escalas das variáveis. Assim, uma análise descritiva antes da definição da distância é válida. Uma outra nota importante, é que a distância euclidiana não deverá ser usada quando as variáveis são medidas em unidades diferentes, quando são correlacionadas ou quando têm variâncias muito diferentes, porque as variáveis contribuirão com pesos diferentes no cálculo dos valores da dissemelhança. Esta dissemelhança é ainda sensível a mudanças de escala, ou seja, quando se muda de escala, os valores da distância são alterados, assim como as suas ordens e conseqüentemente os resultados da análise. Portanto, quando as variáveis utilizadas, nesta técnica, estão em escalas diferentes há necessidade de padronizá-las, ou seja, criar variáveis com média zero e desvio padrão um, evitando assim que a análise sofra qualquer influência de grandeza de escala.

A matriz de proximidade (ou semelhança) é uma das componentes da análise de agrupamento. Esta serve como uma técnica que ajuda na identificação da análise de agrupamentos. A Tabela 1.1 exhibe um exemplo da matriz de proximidade para seis sujeitos.

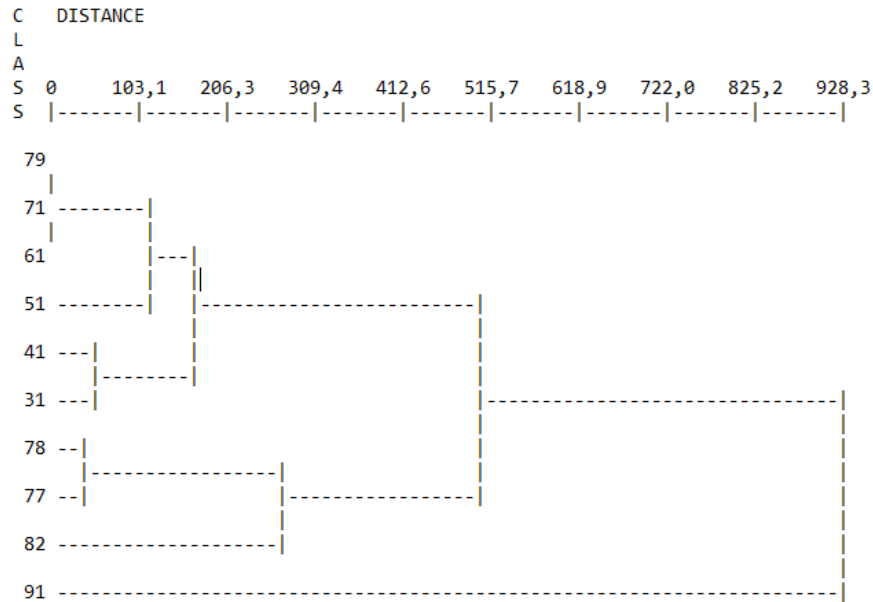
Tabela 1.1: Exemplo de matriz de proximidade para seis sujeitos

	A	B	C	D	E
A	0				
B	1	0			
C	0	1	0		
D	0	0	1	0	
E	1	0	1	1	0

Fonte: Elaborado pela autora

O dendrograma é uma técnica visual auxiliar para a identificação e validação de agrupamentos, que em forma de árvore é utilizada para representar os agrupamentos com base nos valores da matriz de proximidade (exemplo na Figura 1.2). São úteis para observar saltos que ocorrem na construção dos grupos, detectando a formação de grupos heterogéneos.

Figura 1.2: Exemplo de um dendrograma para 10 classes



Fonte: Elaborado pela autora

Os dendrogramas, de modo geral, apresentam estruturas de agrupamentos de sujeitos homogéneos. Existem muitos critérios para determinar o seu ponto de corte (número óptimo de *clusters*). Um dos métodos muito utilizado é o método de Mojena (1977). Este método baseia-se no tamanho relativo dos níveis das distâncias no dendrograma. A proposta é seleccionar o número de *clusters* na etapa j que satisfaça, primeiramente, a inequação:

$$\alpha_j > \theta_k, \quad (31)$$

em que α_j é o valor das distâncias dos níveis da combinação correspondentes a etapa j ($j=1, \dots, g-1$), θ_k é o valor do ponto de corte, dado por:

$$\theta_k = \bar{\alpha} + k\hat{\sigma}_\alpha \quad (32)$$

sendo $\bar{\alpha}$ (33) e $\hat{\sigma}_\alpha$ (34) a média e o desvio, respectivamente, dos valores de α , e k é uma constante. Mojena (1977) sugeriu valores de k em torno de 2,75 e 3,50. No entanto, Milligan e

Copper (1985) sugeriram o valor de $k = 1,25$ como regra de parada na definição do número de *clusters*.

$$\bar{\alpha} = \frac{\sum_{j=1}^{g-1} \alpha_j}{g-1} \quad (33)$$

$$\hat{\sigma}_\alpha = \sqrt{\frac{\sum_{j=1}^{g-1} \alpha_j^2 - \frac{1}{g-1} \left(\sum_{j=1}^{g-1} \alpha_j\right)^2}{g-2}} \quad (34)$$

em que g é o número de *clusters*. Para além do método de Mojena, existem outros métodos que determinam o número óptimo de *clusters* que iremos ver a seguir.

1.3.1.2. Índices de validação dos *clusters*

Os índices de validação dos *clusters* têm como principal objectivo permitir que os investigadores escolhessem um número óptimo de *clusters* no sujeito dos dados para as condições pré-definidas. Existem inúmeros destes índices, sendo estudados vários deles por Milligan e Cooper em 1985 (Wilkinson, Engelman, Corter, & Coward, 1998). Um deles é o *Root Mean Square Standard Deviation* (RMSSTD), cuja tradução é raiz-quadrada do desvio-padrão médio. Nesta subsecção, iremos apresentar três índices para o agrupamento hierárquico.

Índice *Root Mean Square Standard Deviation* (RMSSTD)

O índice RMSSTD (35) é a raiz-quadrada do desvio-padrão médio para todas as variáveis dentro de cada *cluster*, ou seja, calcula a soma dos quadrados de cada *cluster* e normaliza-a pelo produto do número de elementos em cada *cluster* e o número de variáveis (Sharma, 1996), isto é, calcula a homogeneidade dos agrupamentos. Portanto, quando menor for o valor do índice mais homogêneos são os *clusters* (agrupamentos).

$$RMSSTD = \sqrt{\frac{w_k}{v(N_k - 1)}}, \quad (35)$$

onde w_k é a soma dos quadrados dentro dos *clusters* k , N_k é o número de sujeito no *cluster* k e v é o número de variáveis (Wilkinson, Engelman, Corter, & Coward, 1998).

A decisão sobre o número óptimo de *cluster* é baseada no valor do índice e, pode ser verificado através do gráfico do RMSSTD em função do número de grupo. Este, é o número de

cluster onde há um salto íngreme do valor do índice de maior para menor número de *clusters*. Ou seja, é o ponto máximo da curva.

Índice R-Square (R²)

O R², ou coeficiente de determinação (37), é utilizado, também, para determinar o número óptimo de *cluster*. Este calcula a dissimilaridade entre os *clusters*, e quando maior é o seu valor maior é a dissimilaridade entre eles (Sharma, 1996).

Índice Pseudo-F

A estatística *pseudo-F* proposta pelo Calinski e Harabosz, em 1974, é o rácio da variância entre os *clusters* e a variância dentro dos *clusters*. Esta estatística avalia a eficácia do agrupamento e é expressa da seguinte forma (ESRI, 2016):

$$F = \frac{R^2(n - n_c)}{(n_c - 1)(1 - R^2)} \quad (36)$$

em que

$$R^2 = \frac{SST - SSE}{SST} \quad (37)$$

$$SST = \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} \sum_{k=1}^{n_v} (v_{ij}^k - \overline{v^k})^2 \quad (38)$$

$$SSE = \sum_{i=1}^{n_c} \sum_{j=1}^{n_i} \sum_{k=1}^{n_v} (v_{ij}^k - \overline{v_i^k})^2 \quad (39)$$

onde

n – é o número da *feature*

n_i – é o número da *feature* no *cluster* i

n_c – é o número de classes

n_v – é o número de variáveis usado no *cluster* das *features*

v_{ij}^k – é o valor da k -ésima variável do j -ésimo *feature* no i -ésimo grupo

$\overline{v^k}$ – é o valor médio da k -ésima variável

$\overline{v_i^k}$ – é o valor médio da k -ésima variável do *cluster* i

Quando mais elevada for a estatística *pseudo-F* maior é capacidade dos grupos distinguir as classes e as variáveis que foram especificadas para a análise. Portanto, o grupo com maior estatística é considerado o número óptimo de grupo.

2. ESTUDO DAS DESIGUALDADES SOCIAIS

Neste capítulo é realizado um enquadramento sobre os conceitos e estudos sobre as desigualdades sociais tendo em conta os objectivos deste trabalho. Ele está dividido em duas secções. Na primeira secção é realizada uma revisão dos estudos sobre as desigualdades sociais, apresentando os indicadores mais utilizados, para quantificar as desigualdades consideradas neste estudo. Na segunda secção, é dada uma especial atenção as desigualdades sociais em Cabo Verde.

2.1. Revisão da Literatura

A desigualdade social é um problema social presente em todos os países do mundo. Esta desigualdade ocorre pela má distribuição de rendimento e, pela falta de investimento na área social. Na literatura (Hoffmann, 1998; Castro, 2006; Coelho, 2009; Hellebrandt & Mauro, 2015; Banco Mundial, 2014), ela é também denominada de desigualdade económica, pois, os estudos realizados inicialmente sobre as desigualdades sociais eram direccionados as desigualdades de rendimentos. Em 2001, Sen contestou a desigualdade estudada a partir de rendimento e, propôs uma reavaliação sobre a temática. Para este autor, a desigualdade não deve ser estudada só a partir do rendimento, mas sim através da capacidade de realizações e funcionamento dos indivíduos. Pois, para ele o que determina a posição de uma pessoa na sociedade são as suas realizações e a liberdade de a realizar. Magalhães e Carvalho (2010) defenderam que também deve ser levada em consideração as importantes diversidades de cada contexto, mas sem deixar de questionar qual será o espaço de investigação da desigualdade, quais as variáveis disponíveis no espaço determinado e qual o tipo de desigualdade que interessa estudar, desigualdades de quê?

Tilly (2006) relacionou a desigualdade ao grupo (categoria) e ao território (fronteira). Isto é, as desigualdades dão-se entre grupos diferentes separadas por fronteiras, onde acontece interacção entre indivíduos. Para o mesmo autor o quotidiano é demarcado por fronteira que divide os grupos, em que a desigualdade pode ser gerada através destes, quando as transacções constantes através das fronteiras geram vantagens aos que estão em um dos lados (Magalhães & Carvalho, 2010). Assim, Tilly (2006) define a desigualdade como uma relação entre indivíduos, ou conjuntos de indivíduos, em que a interacção gera mais vantagem para um dos lados.

Para Coelho (2009), a desigualdade parece mudar de origem, abrindo espaços as novas desigualdades. Portanto, a desigualdade social conduz a vários outros tipos de desigualdades, que para além das mencionadas anteriores, estão a desigualdade de género, desigualdade entre raças, desigualdades regionais, desigualdade em saúde, entre outras. O que sucede é que as novas desigualdades originadas pode ser *Between* (entre) ou *Within* (dentro) dos grupos sociais. Tilly (2006) denominou de desigualdade categórica a desigualdade que se dá por categoria (grupos) e, entende-se por este como formas de benefício desigual em que um conjunto de indivíduos não recebe o mesmo tratamento.

Para Proença (2009), a ideia da desigualdade social pode ser reduzida na definição de política económica e social, assim, seria relevante não só quantificar as desigualdades, mas

também conhecer as suas causas e a sua natureza. Para quantificar a desigualdade são utilizados um conjunto de indicadores sociais, que segundo Sen (2001) estes têm como premissas que os seres humanos são iguais no que diz respeito a uma série de características pessoais relevantes que influenciam no processo de interação social.

De entre as desigualdades existentes, a mais falada e medida é a desigualdade de rendimento. Porém, existem outras desigualdades, no qual o estudo implica, como desigualdade de género, ambiental e em saúde.

Existe uma vasta sucessão de medidas sintéticas para mensurar a desigualdade na distribuição de rendimento de uma dada região, que partem de diferentes hipóteses e conceituações. Porém, a utilização de um ou de outro pode conduzir a diferentes resultados e conclusão em termos de evolução da desigualdade de rendimento. É difícil identificar a melhor forma de a medir, mas cada uma consegue alcançar a desigualdade sob uma determinada óptica e a comparação entre diferentes resultados pode ser elucidativa/informativa caso se tenha a definição clara de cada uma dessas medidas.

2.1.1. Principais Indicadores – índices e medidas sintéticas

Existem inúmeras medidas, na literatura, para mensurar o grau de desigualdade. Alguns são: Fórmula de Pareto, Amplitude, Desvio da média absoluta, Coeficiente de variação, Curva de Lorenz, Índice de Gini, Índices de Theil, índice de concentração, índice angular, entre outras. Entretanto, a curva de Lorenz, o índice de Gini e os índices de Theil são as mais comuns.

2.1.1.1. Fórmula de Pareto

A primeira representação analítica da curva de rendimento foi apresentada por Vilfredo Pareto.

Seja $n_x = f(x)$ a distribuição de frequência dos rendimentos x (x variável contínua e $f(x)$ uma função incógnita), então $n_x dx$ representa o número de indivíduos com rendimento entre x e $x + dx$. Se N_x é o número de possuidores de rendimento iguais ou superiores a x , teremos

$$N_x = \int_x^w f(x) dx, \quad (40)$$

onde w é o rendimento máximo possível, em geral considerar-se-á $w = \infty$.

Assim, Pareto concluiu que, a partir de um nível de rendimento adequado, N_x exprime-se mediante uma função hiperbólica

$$N_x = \frac{A}{x^\alpha}, \quad (41)$$

em que A e α são constantes positivas.

Esta fórmula aponta que existe uma estreita correlação funcional entre os níveis de rendimento de cada estrato da distribuição e o número de pessoas economicamente activas. Em todos os casos examinados, Pareto observou que o rendimento distribuía-se com regularidade estatística. De um lado, situava-se grande número de indivíduos com rendimentos inferiores à média do respectivo grupo e do outro, apenas pequeno grupo de pessoas com rendimento sensivelmente acima do valor médio.

2.1.1.2. Amplitude

A Amplitude é uma medida de dispersão que é utilizada também para medir a desigualdade de rendimento. Este indicador (42) é expresso de seguinte forma:

$$Amplitude = \frac{R_{max} - R_{min}}{\mu}, \quad (42)$$

onde R_{max} e R_{min} são os rendimentos máximo e mínimos, respectivamente, e μ é o rendimento médio (Castro, 2006).

2.1.1.3. Desvio da Média Absoluta

O Desvio da Média Absoluta é também uma medida de dispersão e uma medida absoluta utilizada para determinar a desigualdade de rendimento. Segundo Castro (2006), este indicador é o quociente entre a soma das distâncias dos rendimentos em relação ao rendimento médio e total, expressa da seguinte forma:

$$D = \frac{\sum_{i=1}^n n_j R_j}{\mu n} - \mu, \quad (43)$$

onde R_j é o rendimento do individuo j e n é o total da população.

2.1.1.4. Coeficiente de Variação

O coeficiente de variação (CV) é uma medida de dispersão relativa. Ele é o quociente entre o desvio-padrão e a média (44).

$$CV = \frac{\sqrt{\sum_{i=j}^n n_j (R_j - \mu)^2}}{\mu n} \quad (44)$$

Segundo Castro (2006) este indicador é uma forma de evitar a insensibilidade do indicador Desvio da Média Absoluta. Portanto, ao invés de utilizar o Desvio da Média Absoluta é mais apropriado utilizar o Coeficiente de Variação, pois, ele satisfaz os quatros primeiros requisitos para obter um “bom” indicador².

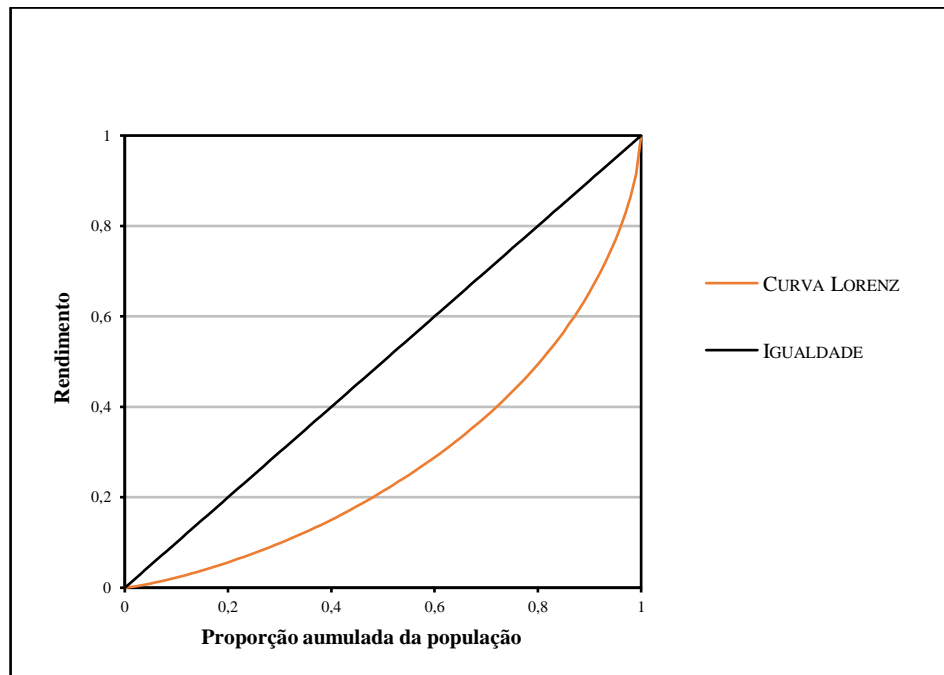
2.1.1.5. Índice de Gini e Curva de Lorenz

O índice de Gini, ou coeficiente de Gini, é um dos índices comumente utilizados para medir o grau da desigualdade de rendimento. Ele é uma medida de desigualdade desenvolvida pelo estatístico italiano Corrado Gini em 1912. O Gini varia entre 0 e 1, em que zero corresponde a perfeita igualdade de rendimento (onde todos detêm o mesmo rendimento *per capita*) e um corresponde a desigualdade perfeita (onde um individuo, ou uma pequena parte de uma população, detém todo o rendimento e os demais não têm). A sua construção é baseada na Curva de Lorenz.

A Curva de Lorenz, também denominada de curva de concentração de Lorenz, consiste num gráfico que procura ilustrar a desigualdade existente na distribuição do rendimento entre as famílias numa determinada economia ou sociedade. Este gráfico consiste num diagrama em que num dos eixos é colocada a variável rendimento e no outro a população, ambos representados por classes percentuais. Nesse diagrama, é então representada uma linha representativa da percentagem de rendimento que cabe a cada grupo da população. Quando mais afastada da diagonal estiver a linha, maior é a concentração do rendimento, ou seja, maior é a desigualdade na distribuição do rendimento entre os grupos. Esta curva mostra como a proporção acumulada do rendimento, varia em proporção acumulada da população, ordenados de ordens crescentes. Na Figura 2.1 podemos visualizar um exemplo da Curva de Lorenz.

² Requisitos para um bom indicador (Castro, 2006): (a) Anonimato – não interessa quem está ganhando os rendimentos; (b) Rendimento relativo – deve ser insensível as mudanças de escala ou transferências proporcionais; (c) População – não interessa o tamanho da população, o que é relevante é a proporções de população que ganham diferentes níveis de rendimento; (d) Dalton-Pigou – a transferência de rendimento de um rico para um pobre deverá acusar a queda na desigualdade, e vice-versa; (e) Sensibilidade – apresentar maior sensibilidade a transferência entre pessoas mais pobres do que aquelas das pessoas mais ricas; (f) Aditivamente decomponível – este requisito deve ser satisfeito a ponto do indicador expressar a soma de dois, ou mais, termos: desigualdade entre e dentro dos grupos.

Figura 2.1: Exemplo da Curva de Lorenz



Fonte: Elaborado pela autora

O índice de Gini (IG) complementa a Curva de Lorenz, pois, ele quantifica o grau de concentração dos rendimentos. Este indicador (45) é a relação entre a área da desigualdade indicada por área de concentração (A) e a área do triângulo.

$$IG = 2A \quad (45)$$

2.1.1.6. Índice de Theil – L de Theil

Em 1967, Theil apresentou dois índices para medir a desigualdade de rendimento, denominados de T e L de Theil. O L de Theil é o único indicador que satisfaz todos os requisitos de um 'bom' indicador de desigualdade, pois, é o único estritamente decomponível. É uma medida de desigualdade baseada na teoria de informação expressa como (Hoffmann, 1998)

$$L = \frac{1}{n} \sum_{i=1}^n \ln \frac{1}{nR_i} \quad (46)$$

Em que R_i é o rendimento do indivíduo i e n é o total de indivíduos. O L é igual a zero quando todos os rendimentos são iguais. Porém, não existe um limite superior para este indicador. Assim, por analogia é padronizada o valor do L (L_p) por forma a obter um valor entre 0 e 1:

$$L_p = 1 - e^{-L} \quad (47)$$

O L de Theil, ou Entropia Generalizada (Theil, 1967), é muito utilizado no estudo da desigualdade dado a sua capacidade estatística de decomposição. A medida de Entropia Generalizada (GE) é em função de um parâmetro α , e é calculada pela seguinte fórmula (IDRF, 2002):

$$GE(\alpha) = \frac{1}{\alpha^2 - \alpha} \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i}{\mu(y)} \right)^\alpha - 1 \right] \quad (48)$$

em que y_i é o rendimento do agregado ou individuo i , $\mu(y)$ é a média da distribuição do rendimento na população e n é o número total de agregados familiares ou indivíduos.

Os indicadores de entropia, geralmente utilizados são $GE(0)$, $GE(1)$ e $GE(2)$, e são determinados pelas seguintes expressões: (IDRF, 2002)

$$GE(0) = \frac{1}{n} \sum_{i=1}^n \log \frac{\mu(y)}{y_i} \quad (49)$$

$$GE(1) = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{\mu(y)} \log \frac{y_i}{\mu(y)} \quad (50)$$

$$GE(2) = \frac{1}{2n\mu(y)^2} \sum_{i=1}^n (y_i - \mu(y))^2 \quad (51)$$

sendo as notações com mesmo significado apresentada anterior, na fórmula da GE.

Se $\alpha = 0$ é dado um maior peso às distâncias entre rendimentos na parte inferior da distribuição, isto é, mede a desigualdade entre os mais pobres. No caso em que $\alpha = 1$ é atribuído peso igual a desigualdade de toda a distribuição. Se $\alpha = 2$ é atribuído maior peso a distância entre rendimentos dos mais ricos. Isto significa que se dá uma maior importância à desigualdade entre os agregados com rendimentos mais elevados.

2.1.1.7. Índice Angular de Desigualdade

O índice angular de desigualdade é obtido através da análise de regressão da variável de estudo sobre um indicador de posição relativa acumulada de cada grupo em relação a uma variável socioeconómica dos grupos como o tamanho da população. Ele é o coeficiente angular do modelo estimado. Para obter o modelo, os grupos são ordenados de forma crescente, da pior para melhor

situação socioeconómica e, cada grupo é caracterizado por um valor que corresponde a frequência média acumulada de cada grupo, ordenados em função da variável socioeconómica.

Em 1991, Wagstaff *et al.* analisam alguns indicadores de medida da desigualdade em saúde e, chegaram a conclusão que o índice angular de desigualdade e o índice de concentração são os mais adequados para mensurar a desigualdade em saúde, pois, estes indicadores reflectem a dimensão socioeconómica das desigualdades em saúde, utilizam informação de toda a população e são sensíveis a redistribuição da população entre os diferentes grupos sociais. O problema com as restantes medidas (índice de Gini, Curva de Lorenz, índice de dissimilaridade, entre outras) é que eles não fornecem uma medida de desigualdade social em saúde. Os resultados apenas indicam o grau de desigualdade, mas não relaciona com o *status* socioeconómico dos indivíduos envolvidos.

2.1.1.8. Curva e Índice de Concentração

A curva e índice de concentração são derivados da Curva de Lorenz e do índice de Gini, respectivamente, e fornece uma medida da extensão da desigualdade social em saúde. A diferença entre estes indicadores é que o primeiro considera uma variável socioeconómica em que os indivíduos são ordenados. Isto é, a curva de Lorenz representa exclusivamente a desigualdade relativa, ou seja, é indiferente ao nível da distribuição. Ela pode ser entendida como a curva de concentração em que a variável de ordenação é a mesma da distribuição. A outra diferença é que o índice de Gini varia entre 0 e 1, enquanto o índice de concentração varia entre -1 e 1.

O índice de concentração (IC) é definido como o dobro da área entre a curva de concentração, $L(p)$ e a linha de igualdade. Para dados agrupados é expressa como (ScotPHO, 2007):

$$IC = (P_1L_2 - P_2L_1) + (P_2L_3 - P_3L_2) + \dots + (P_{T-1}L_T - P_TL_{T-1}), \quad (52)$$

em que P é a percentagem acumulada da amostra classificada, $L(p)$ é a correspondente curva de concentração ordenada e T é o número de grupos.

Apesar do índice de Gini ser a mais comum, o índice de concentração mostra ser mais adequado quando se pretende medir a desigualdade considerando uma variável socioeconómica de ordenação, por exemplo, quando se pretende medir a desigualdade de rendimento levando em consideração o nível de educação dos indivíduos. Todavia, o índice de Theil é considerado o mais apropriado para medir a desigualdade de rendimento, pela sua característica de decompor. Já o

índice de concentração é utilizado quando se pretende medir a desigualdade em saúde pelo facto de levar em consideração a variável socioeconómica, que quase sempre é o rendimento familiar.

Em 1998, David Champernowne e Frank Cowell *apud* Medeiros (2012) mostraram que existem pelo menos duas abordagens para medir as desigualdades: a absoluta e a relativa. A abordagem absoluta, segundo os mesmos autores, está relacionada a diferença e a relativa a razão. Ou seja, a desigualdade absoluta pode ser medida pela diferença entre os rendimentos e, a relativa pela razão (divisão). Esta foi apenas um exemplo, apresentado pelos autores, de como estas duas abordagens medem a desigualdade de forma diferente e que cada um conduz ao seu próprio resultado, não comparáveis pelo facto das duas adoptar metodologias diferentes.

Em 2002 o Instituto Nacional de Estatística de Cabo Verde (INE-CV) mediu a desigualdade de rendimento usando como indicador o índice de Gini, rácio decil, *Share Ratio* e a entropia generalizada. Utilizou outras estatísticas adicionais como a *Skewness* (assimetria), *Kurtosis* (curtose) e teste de Jaque-Bera para tirar conclusões *à priori* sobre a desigualdade de rendimento. Mediu-se a desigualdade de rendimento considerando a desigualdade apenas como a dispersão da distribuição das despesas entre os agregados familiares. Porém, a desigualdade quantificada levou em conta apenas as despesas de cada agregado, deixando de lado um indicador importante como o nível de educação dos indivíduos.

A *Skewness* (coeficiente de assimetria) é uma medida de assimetria da distribuição em torno da média. Permite medir a desigualdade do rendimento *per capita* entre os agregados. O coeficiente de curtose, que tal como na assimetria, é vulgarmente o termo usado é em inglês, *kurtosis*, mede o tamanho das caudas da distribuição em relação a uma distribuição simétrica. Uma distribuição normal tem *kurtosis* igual a 3

O rácio decil é o rácio entre o percentil 90 e o percentil 10 da distribuição. Mede a relação entre o rendimento acima do qual se encontra os 10% dos agregados familiares de rendimento mais elevado, comparativamente ao rendimento abaixo do qual se encontra os 10% da população mais pobre. O *Share Ratio* (no estudo do INE-CV foi adoptada o rácio S80/S20) é o rácio entre a proporção do rendimento dos 20% da população mais rico e a proporção do rendimento dos 20% mais pobres. Apesar de facilitar a manipulação, este indicador pouco informa sobre os extremos da distribuição, levando as vezes a interpretações ambíguas.

Os exemplos emblemáticos nas análises relacionadas as desigualdade de género são o *Global Gender Gap Index* (Índice Global de Desigualdade de Género) elaborado pelo Fórum

Economico Mundial, o *Women's Economic Opportunity Index* (Índice de Oportunidade Económica das Mulheres) determinado pela Unidade de Inteligência Económica, o *Social Institutions and Gender Index* (Índice de Género e Instituições Sociais) da Organização para a Cooperação e Desenvolvimento Economico e o *Gender Inequality Index* (Índice de Desigualdade de Género) publicado pelo Programa das Nações Unidas para o Desenvolvimento (PNUD). Todos esses estudos adoptaram na sua metodologia o uso de índices sintéticos para quantificar a desigualdade de género.

Segundo, Freudenberg (2003) os índices sintéticos causam geralmente divisões. Há que se ter em conta que nem todos os aspectos da desigualdade de género podem ser quantificados e expressos em número (assim como qualquer outra desigualdade social), o que pode levar qualquer tentativa de quantificação ser alvo de críticas. Os riscos podem decorrer da omissão de indicadores relevantes, perda de informação decorrente da agregação, inconsistências metodológicas, entre outros. Por outro lado, o seu uso é sugerido quando se quer captar diferentes dimensões de um conceito multidimensional (Mainali et al., 2014).

Ainda na revisão da literatura, foram encontrados vários indicadores consideradas para medir a desigualdade em saúde, como a razão de taxa, diferença de taxa, risco atribuível populacional, índice de efeito, índice de dissimilaridade, índice relativo de desigualdade, índice angular de desigualdade, bem como os indicadores da área económica como o índice de Gini, com a sua correspondente curva de Lorenz e as variantes de ambos, como o índice de concentração e Curva de concentração. Segundo Schneider et al. (2002), a razão de taxa, diferença de taxa e o índice de efeito são medidas de efeito, pois, baseiam na categoria fixa das variáveis e os indicadores, risco atribuível populacional, índice de dissimilaridade, índice relativo de desigualdade, índice angular de desigualdade são medidas de impacto por utilizarem categorias definidas por um indicador socioeconómico quantificável em termos populacionais, de modo que, se a distribuição do indicador socioeconómico variar, também varia o nível de desigualdade. Segundo o mesmo autor, o uso do índice de efeito, implica necessariamente conhecimentos estatísticos para a escolha do melhor modelo e interpretação dos resultados. Também, deve-se considerar que os pressupostos da regressão nem sempre são garantidos, sendo restritivos, tornando-a inviável em muitos casos.

As razões de taxas e diferenças de taxas, para ScotPHO (2007), é a amplitude. Para este autor, este indicador é a mais comum para estudar a desigualdade em saúde e, envolve comparação de experiências dos grupos socioeconómicos superiores e inferiores. Às vezes, essa

comparação é apresentada na forma de intervalo propriamente dita (amplitude absoluta, ou diferença de taxas), mas mais frequentemente como a proporção de um valor extremo para outro (amplitude relativa, ou razão de taxas). Estas são medidas simples e directas. Uma das vantagens da amplitude relativa em comparação com a absoluta é que ela é neutra em escala, ou seja, a desigualdade pode ser comparada para as taxas de resultado medido em diferentes escalas (ScotPHO, 2007). Por outro lado, a amplitude tem vista para mudanças nos grupos intermediários e não leva em conta os tamanhos dos grupos que estão sendo comparados (Kunst & Mackenbach, 1994; ScotPHO, 2007).

O Risco Atribuível Populacional (RAP) é fácil de calcular e interpretar e, além de medir o indicador de saúde dos grupos com alto nível socioeconómico, leva em consideração o tamanho da população, pois um número elevado de grupos com indicador elevado, aumenta a redução potencial do indicador global (Kunst & Mackenbach, 1994). Ele é uma medida de frequência de doença amplamente utilizada na epidemiologia, que requer a escolha de um grupo não exposto ou de referência. A estatística é normalmente expressa em percentagem e, é dada pela seguinte forma (ScotPHO, 2007):

$$RAP = \frac{P_e \times (RR - 1)}{1 + P_e \times (RR - 1)} \quad (53)$$

Onde P_e é a proporção da população total no grupo exposto e RR é o risco relativo para o grupo exposto, comparado ao risco de 1 para o grupo não exposto.

O índice de dissimilaridade é recomendável em análises de situação de saúde porque pressupõe a redistribuição da carga de doença ou morte (Schneider, et al., 2002). O índice angular e relativo da desigualdade leva em consideração o tamanho da população e a posição socioeconómica relativa dos grupos, além de serem sensíveis à situação de saúde média (Kunst & Mackenbach, 1994; ScotPHO, 2007). Entretanto, eles são complexos relativamente ao cálculo e interpretação, podendo fornecer resultados poucos confiáveis quando aplicados em amostra pequena de dados agregados. Quando se trabalha com dados agrupados, a equação de regressão deve ser transformada para evitar heterocedasticidade.³ O índice de Gini e a curva de Lorenz utilizam informação de todos os indivíduos ou grupos da população, entretanto, como já foi referido,

³ Em estatísticas, um vetor de variáveis aleatórias é heterocidástico se as variáveis aleatórias possuírem diferentes variações. Ao utilizar um modelo Mínimos Quadrados Ponderados um dos pressupostos é que o erro seja constante.

não consideram a situação socioeconómica (Kunst & Mackenbach, 1994), como o índice de concentração.

Em 1987, Le Grand *apud* Noronha e Andrade (2001), utilizou três indicadores para medir a desigualdade em saúde entre 32 países desenvolvidos: o índice de Gini, a diferença da média absoluta e o índice de Atkinson. Entretanto, estes índices não fornecem uma medida da desigualdade social em saúde. Os resultados encontrados pelo autor apenas indicaram o grau de desigualdade entre os países, não informando a qual classe de rendimento essa desigualdade favorece.

Ainda não existem muitos estudos sobre a quantificação da desigualdade ambiental, como as restantes desigualdades sociais e económicas, até então apresentadas nesta dissertação. Dos que existem, adoptam a metodologia do PNUD para a normalização dos dados e optam sempre pelos índices sintéticos. Morato, Kawakubo, & Luchiarri (2005) optaram pela mesma metodologia, normalizando os indicadores básicos, como condições de abastecimento de água, destino da água servida e do lixo, ocorrência de domicílios improvisados e a presença de cobertura vegetal, através da criação dos seus índices, denominados de índices básicos, e determinou a desigualdade em saúde pela média aritmética desses índices.

2.2. Caso de estudo: Desigualdades sociais em Cabo Verde

A desigualdade de rendimento é o único tipo de desigualdade medida em Cabo Verde. Entretanto, não existe uma série, ao longo do tempo, sobre a desigualdade de rendimento. Esta foi medida apenas nos anos 1989, 2002 e, o último incidiu-se sobre os dados do Inquérito às Despesas e Receitas Familiares (IDRF) de 2015, onde existe até o momento resultados preliminares.

O Banco Mundial realizou em 1993 o primeiro estudo sobre a pobreza em Cabo Verde, com a utilização dos dados sobre as despesas das famílias oriundos do inquérito às Despesas e Receitas Familiares de 1988/89. Esse estudo revelou que 30% da população Cabo-Verdiana eram pobres e 14% eram muito pobres (pobreza extrema), ou seja, viviam com menos de 26 mil e 18 mil Escudos Cabo-Verdianos (ECV), respectivamente, por ano. Em 2002 o percentual de população pobre era de 36,7%, sendo 62% residentes no meio rural. A pobreza extrema passou de 14% para 19.7% da população total, onde 68% viviam em meio rural (IDRF, 2002).

Em 2004, O banco Mundial realizou o cálculo da evolução da pobreza absoluta de Cabo Verde, onde também calculou o índice de desigualdade de Gini. Este índice revelou que em

1988/89 a desigualdade era de 0,5 e em 2001/02 aumentou para 0,53. A nível nacional, segundo os dados do INE-CV, em 1988/89 o país apresentou um índice de desigualdade de 0,43 e 0,57 em 2001/02. A ilha com maior índice de desigualdade foi Santo Antão, em ambos os anos, passando de 0,49 para 0,65 e, a ilha com menor índice, em 1988/89, foi o Fogo com 0,39 e, em 2001/02 a ilha de Boavista com 0,42. É de realçar que no ano 1988/89 não estavam disponíveis as informações sobre as ilhas de Boavista, Maio e Sal (IDRF, 1988/89, 2001/02). O meio urbano supera o meio rural na matéria da desigualdade, apresentando um índice de 0,58 (superior ao índice global) por 0,47 do meio rural. A desigualdade é ainda mais elevada entre os representantes dos agregados familiares com idade compreendida entre 45 e 64 anos.

É importante referir que as metodologias utilizadas nos inquéritos de 1988/89 e 2001/02 não foram exactamente iguais. No primeiro inquérito foi estabelecido como limiar da pobreza, o consumo anual médio por habitante de 26 mil ECV e a pobreza extrema de 18 mil ECV. Já no segundo inquérito, o limiar aumentou para 43.250 ECV e 28.833 ECV, respectivamente, para os pobres e muito pobres. Assim, as comparações realizadas não são muito confiáveis, pois, o facto do limiar ter sido estabelecido em 67% da despesa anual média em 1988/89, contra 60% do consumo mediano em 2001/02, pode-se constatar que os dados subestimam o aumento da pobreza que poderia ser mais elevada. Note-se que o aumento do consumo médio anual do limiar da pobreza extrema aumentou cerca de 60%, um valor inferior ao aumento dos preços no consumo, estimado em 80% naquele período, o que indica que, o aumento da pobreza extrema seria mais significativo conservando a mesma definição.

Segundo os dados do IDRF 2001/02, 10% da população mais rica tinham 47% do rendimento total, enquanto 10% da população mais pobre tinham apenas 1%. Esses valores traduziram a excessiva concentração de riqueza no país. Incidindo sobre os mesmos dados, chegou-se a conclusão que os atributos mais importantes para explicar a desigualdade em Cabo Verde, entre o conjunto dos atributos considerados, são a qualificação na actividade, a dimensão do agregado familiar, o nível de instrução do chefe do agregado, o sector de actividade e o grau de urbanização da residência do agregado (IDRF, 2002).

Em 2007, Cabo Verde foi considerado, pelas Nações Unidas, um país de rendimento médio por ter apresentado um Índice de Desenvolvimento Humano (IDH) de 0,736. Desde então, Cabo Verde passou por uma transição com bastante sucesso de um nível muito baixo de desenvolvimento para um nível mais alto, porém, a pobreza é ainda um problema que esse país

enfrenta. O Questionário Unificado de Indicadores Básicos de Bem-estar (QUIBB) de 2007, um inquérito realizado pelo INE-CV, revelaram uma percentagem de agregados familiares pobres de 38,6%, sendo 6,9% consideradas em pobreza extrema. A população total de pobre incidu-se em 26,6% e, apesar do aumento da população entre os anos 2001/02 e 2007, houve uma diminuição do número de pobres de 172.572 em 2001/02 para 130.892 em 2007, cerca de 41.680 indivíduos (INE-QUIBB 2007).

A desigualdade na distribuição do rendimento em Cabo Verde é um facto que pode explicar o aumento da pobreza apresentada no ano 2001/02, face ao ano 1988/89. Se esta distribuição for muito elevada pode levar a um impacto negativo das políticas de redistribuição sobre a pobreza, por outras palavras, podem levar a colisão nos domínios da educação, saúde, ambiente, justiça e segurança, entre outras. A proporção elevada dos pobres no meio rural é um indício, ou uma consequência, da desigual distribuição do rendimento a nível do país, pois, a maioria dos pobres são mulheres, com nível de escolaridade baixa.

3. MATERIAL E MÉTODOS

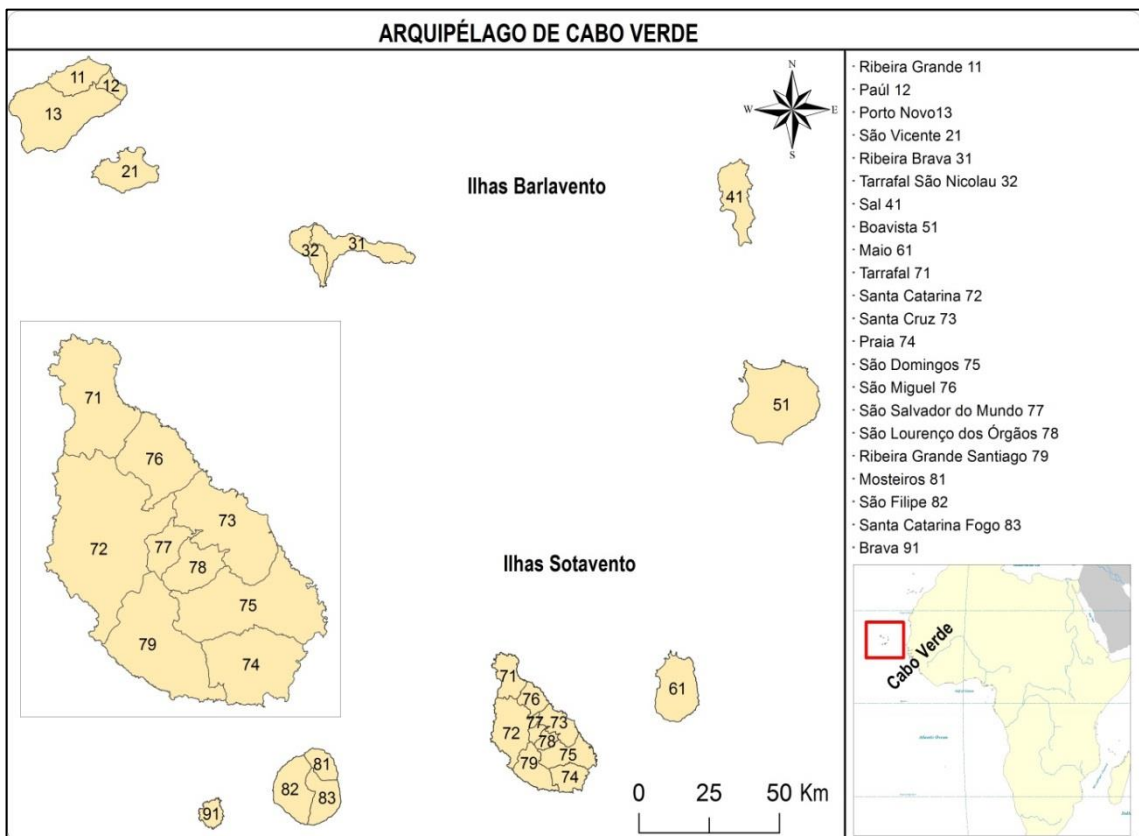
Esta secção apresenta os métodos que foram seguidos no sentido de alcançar os objectivos do presente trabalho. Primeiramente são apresentadas as áreas de estudo, de seguida as variáveis, e as suas respectivas fontes, utilizadas para os cálculos dos índices de desigualdades, bem como os *Softwares* utilizados. Dando seguimento, apresentam-se as metodologias adoptadas para determinar os índices de desigualdade e as técnicas de análise dos dados adoptadas.

3.1. Área geográfica de Estudo

Cabo Verde é um arquipélago composto por 10 ilhas, sendo uma desabitada, 13 ilhotas, 22 concelhos e 32 freguesias, situado a cerca de 450 km ao largo da costa da África Ocidental. O país dispõe de 4.033 Km² de área. As ilhas são divididas em Barlavento e Sotavento. O Barlavento é composto pelas ilhas de Santo Antão, São Vicente, Santa Luzia (desabitada), São Nicolau, Sal e Boavista e, o Sotavento é composto pelas ilhas do Maio, Santiago, Fogo e Brava. A ilha de Santiago é a maior ilha do país, e é onde se situa a capital do país, Cidade da Praia. Só esta ilha contém mais da metade da população de Cabo Verde e, apenas o concelho da Praia tem quase 1/3 da população (Projeções Demográficas de Cabo Verde, 2013).

As áreas de estudo seleccionadas para a aplicação da metodologia deste estudo são os 22 concelhos das nove ilhas habitadas do país (Figura 3.1).

Figura 3.1: Área de estudo – Arquipélago de Cabo Verde



Fonte: Elaborado pela autora

3.2. Dados – variáveis e fontes

Os dados utilizados para este estudo são dados provenientes de fontes administrativas (Ministério de Saúde e Boletins Oficiais) e de inquéritos por amostragem e recenseamento realizado pelo INE-CV desde 2000 até 2015. Os dados do Ministério de Saúde foram obtidos dos relatórios estatísticos anuais dos anos 2005 a 2014. Os Boletins oficiais (B.O.) utilizados para obter os eleitos municipais foram: B.O. 1996, 2001, 2004, 2008 e 2012. Os dados dos inquéritos por amostragem foram obtidos do Inquérito às Despesas e Receitas Familiares (IDRF, 2001/02), Inquérito Multi-Objectivo Contínuo (IMC, 2012 a 2015), Inquérito ao Emprego (IE, 2011) e Questionário Unificados dos Indicadores Básicos do Bem-Estar (QUIBB, 2007), e dados dos Censos (2000 e 2010). As variáveis e as respectivas fontes, utilizadas para o cálculo de cada indicador se resumem na Tabela 3.1.

As informações foram organizadas numa estrutura simples de base de dados com o nome “*inequality_SIG_TMoniz.mbd*”, onde as informações de cada ano são uma *feature class* independente (Ver Anexo I). Toda a informação utilizada inicialmente neste trabalho é do tipo polígono (dados por concelho). Com a análise houve necessidade da criação de uma nova *geodatabase* por motivo de espaço e melhor organização. Assim, a nova *geodatabase* nomeada de “*inequality_Analysis_TMoniz.mbd*” serviu para armazenar os resultados das análises.

3.3. Softwares utilizados

O principal *Software* utilizado para o desenvolvimento deste trabalho é o *ArcGIS* versão 10.4. Este *Software* é, hoje, considerado como uma poderosa ferramenta, possibilitando as análises espaciais, assim como a aplicação das estatísticas espaciais. Uma outra ferramenta, também utilizada neste trabalho é o *Statistical Package for the Social Sciences* (SPSS) 22.0 para fins exploratórios do qual não seria possível obtê-lo através do *ArcGIS*. O *Software* Excel foi utilizado para organização de informações em tabelas e gráficos e cálculos intermédios.

Tabela 3.1: Variáveis e índices de caracterização da desigualdade social e suas respectivas fontes

Índice	Dimensão	Variáveis/indicadores	Fonte
ÍNDICE DE DESIGUALDADE DE GÉNERO (IDG)	Saúde (S)	Taxa de Mortalidade Materna (TMM)	Ministério de Saúde
		Taxa de Fertilidade na adolescência (TFA)	INE - Censo 2000 e 2010
	Educação (Edu)	Eleitos Municipais	B.O.
		População adulta com pelo menos o ensino secundário	INE - Censos, IMC
		Taxa de alfabetização	INE - Censos, IMC
	Trabalho (T)	Taxa de participação no mercado de trabalho remunerado	INE - IDRF, Censos, IMC, IE
		Taxa de participação no mercado de trabalho não remunerado	INE - MUT, 2012
		Médias de horas habitualmente trabalhadas	INE - IMC, MUT
Esperança de Vida (E0)	Esperança de vida ao nascer	INE	
ÍNDICE DE DESIGUALDADE AMBIENTAL (IDA)	Qualidade ambiental	Percentagem de AF com ligação a rede pública de distribuição de água	INE - Censos, IMC
		Percentagem de AF com ligação a rede de esgoto ou fossa séptica	INE - Censos, IMC
		Modo de deposição dos resíduos sólidos (lixos)	INE - Censos, IMC
	Condições da habitação	Percentagem de Alojamento improvisados	INE - Censos, IMC
		Percentagens de AF com acesso a cozinha	INE - Censos, IMC
		Percentagens de AF com electricidade	INE - Censos, IMC
ÍNDICE DE DESIGUALDADE EM SAÚDE (IDS)	Saneamento de qualidade	Número de indivíduos vivendo com saneamento básico	INE - QUIBB 2007
	Rendimento	Rendimento familiar <i>per capita</i>	INE - QUIBB 2007
	Doença	Número de indivíduos doentes	INE - QUIBB 2007
ÍNDICE DE DESIGUALDADE ECONÓMICA (IDE)	Rendimento	Rendimento familiar <i>per capita</i>	INE - IDRF 2001/02, QUIBB 2007
	Educação	Nível de educação	INE - IDRF 2001/02, QUIBB 2007
		Rendimento total	INE - IDRF 2001/02, QUIBB 2007

Fonte: Elaborada pela autora

3.4. Cálculo dos índices de desigualdade

3.4.1. Pré-processamento dos dados

Trabalhar com indicadores muito diferentes, como é o caso, requer uma normalização com o objectivo de obter uma escala comum para agregação. Assim, foram utilizados os parâmetros máximos e mínimos para cada indicador. A normalização foi realizada através de uma equação pela qual o valor observado é convertido em uma proporção da distância entre valores mínimo e máximo escolhidos como parâmetros

$$I = \frac{v_o - v_{m\acute{a}x}}{v_{m\acute{i}n} - v_{m\acute{a}x}} \quad (54)$$

em que I é o indicador, v_o é o valor observado, $v_{m\acute{i}n}$ e $v_{m\acute{a}x}$ são os valores mínimo e máximo, respectivamente, fixados. Assim, quando mais próximo o valor observado estiver do valor máximo, menor será o numerador e menor o índice, indicando uma boa situação. Em contrapartida, quando mais próximo estiver do valor mínimo, maior será o índice e pior é a situação. No entanto, quando os indicadores forem obtidos em taxas ou percentagem, não há necessidade de normalização, pois, já se encontram no intervalo 0 e 1.

A classificação da desigualdade adaptada da classificação do IDH, utilizando os valores dos coeficientes, é conforme apresentada na Tabela 3.2.

Tabela 3.2: Classificação dos índices de desigualdade

Índices	Classificação
$0 \leq ID \leq 0,2$	Desigualdade muito baixa
$0,2 < ID \leq 0,4$	Desigualdade baixa
$0,4 < ID \leq 0,6$	Desigualdade média
$0,6 < ID \leq 0,8$	Desigualdade elevada
$0,8 < ID \leq 1$	Desigualdade muito elevada

Fonte: adaptada (Gao, 1999) da classificação do IDH, PNUD

3.4.1.1. Tratamento de zeros

Alguns dos índices medidos são médias geométricas, portanto, não podem ter indicadores com valores nulos. Para tal, é definida estratégias capazes de reflectir da melhor forma os indicadores.

A morte materna é nula em quase todos os concelhos do país. Portanto, a fim de garantir a confidencialidade das informações é utilizada a taxa de mortalidade materna a nível nacional. Ou seja, a taxa de mortalidade materna é igual para todos os concelhos.

A representatividade parlamentar é a taxa de assentos parlamentar a nível do país. Esses assentos envolvem deputados eleitos por circo eleitoral e não por concelho. Em Cabo Verde existe 10 círculos eleitorais, cada círculo corresponde a uma ilha, com excepção da ilha de Santiago que é dividida em dois círculos, o Norte e o Sul. Entretanto, para medir a desigualdade de género de cada concelho não podemos usar essas informações. Assim, considerou o uso dos eleitos municipais, ou seja, os deputados eleitos para assembleia e câmara municipal.

Através das eleições autárquicas são eleitos deputados para as câmaras municipais e para a assembleia municipal. Sendo assim, estes dados parecem mais adequados, pois, reflecte a realidade de cada município.

Para a medição dos índices foram tidas em consideração os indicadores apresentados no capítulo anterior, excepto para o índice de desigualdade de género e ambiental que se optou pelas médias geométricas e aritméticas.

3.4.2. Índice de Desigualdade Económica (IDE)

A desigualdade económica é o que Therborn (2006) designou de desigualdade de recursos, pois, esses não se resumem ao rendimento, referem-se também ao conhecimento e à educação. Assim, para determinar o índice de desigualdade económica foram utilizados como indicadores o rendimento *per capita* dos agregados e o nível de educação dos membros. As medidas utilizadas para medir a desigualdade económica foram a Curva de Lorenz e o índice de Gini.

Para obter as duas medidas, foi agregada, primeiramente, o rendimento *per capita* dos agregados segundo uma análise multivariada denominada de análise de *Clusters*. Esta técnica permitiu agrupar os rendimentos com base no nível de educação dos indivíduos. Após a obtenção dos grupos foi calculado o grau da desigualdade através do índice de Gini da seguinte forma

$$IG = 2 \times A_A, \quad (55)$$

em que

$$A_A = 0,5 - A_B \quad (56)$$

e

$$A_B = \sum_{i=1}^{N-1} \frac{(R_{i+1} + R_i)}{2 \times \frac{1}{N}}. \quad (57)$$

A_A é a área da curva, A_B é a área abaixo da curva, R_i é a percentagem acumulada do rendimento do grupo i , N é o número total de grupo.

As bases de dados utilizadas para medir esta desigualdade foram o IDRF 2001/02 e o QUIBB 2007. O IDRF 2001/02 tinha como principal objectivo elaborar o perfil de pobreza da população. O inquérito foi a nível nacional por amostragem do agregado familiar. Este era composto por quatro questionários. Assim, forneceu informações sobre as condições dos alojamentos dos agregados, as características dos indivíduos, a saúde, a educação, a migração, a actividade económica dos indivíduos, as despesas e as receitas individuais e dos agregados. O QUIBB 2007 foi implementado para recolher quantidade mínimas de informações necessárias à identificação e a classificação de grupos alvos, e para fornecer os indicadores de base de bem-estar que permite acompanhar os programas de luta contra a pobreza. O questionário foi desenvolvido de modo a recolher informações sobre as características das famílias, informações, essas, que permitem medir o acesso a utilização e o grau de satisfação em serviços sociais. O inquérito utilizou uma metodologia desenvolvida pelo Banco Mundial, *Bureau International du Travail* (BIT), UNICEF e o PNUD. O inquérito abrangeu todo o território nacional, por uma amostra de 8.175 agregados familiares, correspondendo a 355 distrito de recenseamento distribuído por concelho.

3.4.3. Índice de desigualdade em Saúde (IDS)

O índice de desigualdade em saúde considera como medida de saúde a doença. Este indicador reflecte todos os indivíduos que tiveram alguma doença nas quatro semanas antes do inquérito. O que se pretende saber é o grau de desigualdade da distribuição dos doentes causada pela diferença nos valores do rendimento familiar *per capita* tendo em conta o saneamento básico. Este índice, ao contrário dos outros, varia entre -1 e 1.

O Saneamento básico é um dos indicadores considerados essenciais para as condições de acesso a benefícios sociais na área da saúde. Assim, é utilizado um indicador, nomeado de saneamento de qualidade. Este indicador é classificado segundo as seguintes variáveis: água potável, sistema de descargas de águas residuais, e depósitos de resíduos sólidos. A água potável considera a população com ligação a rede pública de distribuição de água e que é a principal fonte usada no agregado; O sistema de descargas de águas residuais considera a população cujo agregado tem ligação a fossa séptica ou rede de esgoto e que é o principal modo de descargas; e a deposição de resíduos sólidos considera a população cujo principal modo de depósitos de lixo do agregado é contentor ou recolha por carro. Portanto, são considerados população com saneamento de qualidade as que possui ligação a rede pública de distribuição de água e que a água principal de

uso provém da mesma fonte, que possui um sistema de descargas de águas residuais e faz uso para tal, e que o principal modo de deposição de lixo é por carro ou contentor.

O indicador utilizado para mensurar a desigualdade em saúde foi o índice de concentração. Assim, a variável socioeconómica utilizada foi o rendimento familiar *per capita* dos agregados familiares. Então, o índice de concentração foi obtido através da relação entre a população acumulada com saneamento de qualidade, ordenada de forma crescente segundo o seu rendimento familiar *per capita* (X), em relação a população acumulada de indivíduos doentes (Y)

$$IC = 1 - \sum_{i=1}^{n-1} (Y_{i+1} + Y_i) \times (X_{i+1} - X_i). \quad (58)$$

3.4.4. Índice de Desigualdade de Género (IDG)

A estratégia metodológica empregada na construção do índice de desigualdade de género adopta o mesmo procedimento matemático do PNUD, sendo os pontos de diferença a inclusão de indicadores como média de horas habitualmente trabalhadas, taxa de participação no trabalho não remunerado, taxa de alfabetização e esperança de vida ao nascer.

Portanto, as dimensões utilizadas para o cálculo deste índice são a educação, a saúde reprodutiva, o trabalho e a esperança de vida. A educação considera o percentual de mulheres eleitas municipalmente, a percentagem de população adulta (maior ou igual a 25 anos) com pelo menos o ensino secundário e a taxa de alfabetização. A saúde reprodutiva leva em consideração a taxa de mortalidade materna e a taxa de fertilidade na adolescência. O trabalho considera a taxa de participação no mercado de trabalho remunerado e não remunerado e média de horas trabalhadas habitualmente. Por fim, a esperança de vida é a esperança de vida ao nascer.

Os efeitos da educação da mulher se manifestam e interferem de formas variadas e abrangentes na vida pessoal, familiar e da sociedade. Mulheres com mais acesso a educação promovem aumento do potencial de geração de rendimento e têm autonomia nas decisões pessoais e controlam a própria fertilidade. Têm também maior participação na vida pública, tendo assim mais capacidade de melhorar a própria qualidade de vida e a de suas famílias. Portanto, os indicadores relacionados a educação neste aspecto é de extrema importância. Para isso, foi também acrescentada um indicador nesta dimensão que é a taxa de alfabetização. A média de horas trabalhadas habitualmente reflecte o tempo gasto em trabalhos e a taxa de participação no trabalho

não remunerado é incluída porque, apesar dos homens superarem as mulheres no trabalho remunerado, as mulheres transpõem os homens (MUT, 2012).

O índice baseia-se na média geral das médias gerais de ordens diferentes – a primeira agregação foi pela média geométrica entre dimensões, estas médias, calculadas separadamente para mulheres e homens, foram depois agregadas usando uma média harmónica entre géneros.

Para mulheres a fórmula de agregação é

$$G_F = \sqrt[4]{\left(\frac{1}{TMM} \times \frac{1}{TFA}\right)^{1/2} \times (PM_F \times ES_F \times TA_F)^{1/3} \times \left(TPR_F \times TPNR_F \times \frac{1}{\mu_h(F)}\right)^{1/3} \times E_0(F)} \quad (59)$$

e para homens é

$$G_M = \sqrt[4]{(PM_M \times ES_M \times TA_M)^{1/3} \times \left(TPR_M \times TPNR_M \times \frac{1}{\mu_h(M)}\right)^{1/3} \times E_0(M)} \quad (60)$$

em que TMM é a taxa de mortalidade materna, TFA a taxa de fertilidade na adolescência, PM a percentagem de eleitos municipais, ES a percentagem da população adulta com pelo menos o ensino secundário, TA é a taxa de alfabetizados, TPR a taxa de participação no mercado de trabalho remunerado, TPNR a taxa de participação no mercado de trabalho não remunerado, μ_h é a média de horas habitualmente trabalhadas e E_0 é a esperança de vida ao nascer.

A harmonização dos índices femininos e masculinos agregados foi obtida da seguinte forma

$$HARM(G_F, G_M) = \left[\frac{(G_F)^{-1} + (G_M)^{-1}}{2} \right]^{-1}. \quad (61)$$

A utilização da média harmónica das médias geométricas dentro dos grupos capta a desigualdade entre mulheres e homens e faz o ajustamento para a associação entre dimensões (PNUD, 2011).

O próximo passo foi calcular a média geométrica das médias aritméticas para cada indicador. O padrão de referência para o cálculo da desigualdade foi obtido pela agregação dos índices feminino e masculino, usando ponderações iguais e agregando depois os índices entre dimensões. As médias das dimensões saúde (\bar{S}), educação (\bar{Edu}), trabalho (\bar{T}) e esperança de vida (\bar{E}_0), foram obtidas pelas seguintes expressões

$$\bar{S} = \left(\sqrt{\frac{1}{TMM} \times \frac{1}{TFA} + 1} \right) / 2, \quad (62)$$

$$\overline{Edu} = \frac{\sqrt[3]{PM_F \times ES_F \times TA_F} + \sqrt[3]{PM_M \times ES_M \times TA_M}}{2}, \quad (63)$$

$$\bar{T} = \frac{TPR_F + TPNR_F + \frac{1}{\mu_h(F)} + TPR_M + TPNR_M + \frac{1}{\mu_h(M)}}{6} \quad (64)$$

e,

$$\bar{E}_0 = \frac{E_0(F) + E_0(M)}{2}. \quad (65)$$

A agregação foi dada pela fórmula

$$G_{\bar{F}, \bar{M}} = \sqrt[4]{\bar{S} \times \overline{Edu} \times \overline{TPMT} \times \bar{E}_0} \quad (66)$$

O índice de desigualdade de género é expresso

$$IDG = 1 - \frac{HARM(G_F, G_M)}{G_{\bar{F}, \bar{M}}}. \quad (67)$$

A média aritmética da dimensão saúde não deve ser interpretada como uma média dos índices femininos e masculinos correspondentes, mas como um ponto a meio caminho das normas estabelecidas para os indicadores de saúde reprodutiva – menos mortes maternas e menos gravidez na adolescência (PNUD, 2011).

3.4.5. Índice de Desigualdade Ambiental (IDA)

A quantificação do Índice de Desigualdade Ambiental (IDA) foi baseada na metodologia adoptada pelo Schneider et al. (2002) e Morato et al. (2005). Porém, para este estudo, é levada em consideração as condições de habitação dos Agregados Familiares (AF).

O IDA foi mensurado através de duas dimensões: a qualidade ambiental e a condição de habitação. O índice da qualidade ambiental foi obtido através de três índices simples: o índice de abastecimento de água (IAA), índice de evacuações de águas residuais (IEA), e índice de resíduos sólidos (IRS). O primeiro índice é a percentagem de AF com ligação a rede pública de distribuição

de água, o segundo considera a percentagem de ligação a rede de esgoto ou fossa séptica e, o terceiro é o modo de depósito dos resíduos sólidos (lixos). Para este último, foi considerado a percentagem de AF que são recolhidos os lixos na sua residência (através de carro) e/ou são colocadas nos contentores. O índice da qualidade ambiental (IQA) é a média dos índices simples:

$$IQA = \frac{IEA + IAA + IRS}{3}. \quad (68)$$

A condição de habitação considera quatro índices simples: percentagem de alojamentos improvisados, percentagem de agregados familiares com instalações sanitárias, percentagens de agregados familiares com acesso a cozinha e percentagens de agregados familiares com luz eléctrica. O índice de condições de habitação (ICH) é a média aritmética dos índices simples.

O índice de desigualdade ambiental (69) é a raiz quadrada dos índices das duas dimensões, qualidade ambiental e condições de habitação.

$$IDA = \sqrt{IQA \times ICH}. \quad (69)$$

3.5. Análise dos Dados

3.5.1. Análise descritiva da distribuição espacial

A análise descritiva dos dados iniciou-se com a distribuição espacial da população de Cabo Verde, com informações dos dois últimos Censos (2000 e 2010). O método utilizado para representar a população, espacialmente, foi o método de classificação de dados *Natural Breaks* (que pode ser traduzido para “quebras naturais”), também denominado de método de Jenks. Este método utiliza um algoritmo iterativo que procura diminuir a variância dentro dos grupos e maximizar a variância entre os grupos. A maior vantagem do método é que ele considera explicitamente a distribuição dos dados. Por outro lado, a desvantagem do método é a dificuldade de entendimento da lógica para a maioria dos utilizadores do mapa, além disso, os valores da legenda da classificação (intervalos das classes), geralmente, parecem não ser intuitivos. De qualquer maneira, este método foi utilizado aqui em razão da capital do país (Cidade da Praia) possuir valores muito discrepantes (maiores) em comparação aos restantes concelhos (característica de um *outlier*), pois este método discrimina a capital dos restantes concelhos (ESRI, 2016). De seguida, foram aplicadas as estatísticas descritivas clássicas, como média, desvio padrão, amplitude, coeficiente de variação, bem como as análises de valores ausentes e *outliers*. Para além destas estatísticas, foram aplicadas as técnicas da estatística descritiva espacial, descritas teoricamente na secção 2.2.1.

Estas técnicas foram implementadas com o objectivo de resumir e apresentar as informações de forma a facilitar a sua interpretação. Com este intuito, de um lado, utilizou-se uma medida de tendência central espacial: *Mean Center* (Centro médio) espacial, ponderado pelo rendimento *per capita*, rendimento total, IDE, IDA e IDG. De outro lado, duas medidas de dispersão espacial: distância padrão e elipse de desvio Padrão. Estas técnicas de estatística espacial, juntamente com os SIG, foram aplicadas para descrever a distribuição espacial de dados económicos, bem como as desigualdades. Como é possível verificar na Figura 1.2 estas medidas são simples, pois, necessitam apenas da ligação com a *feature class* inicial. Todavia, a representação gráfica da informação da dispersão e da orientação não é apresentada em conjunto, sendo apenas a informação do centro médio apresentada em conjunto com uma das anteriores.

Para a análise espacial foram utilizadas as informações estatísticas por ilha – rendimento *per capita*, rendimento total e desigualdades (IDE, IDA e IDG). Estes indicadores, rendimento *per capita* e rendimento total, foram utilizados para esta análise em razão de serem mais representativas da riqueza do concelho. Analisando a realidade de cada ilha (Cabo Verde), comparou-se a dispersão do rendimento total do rendimento *per capita* e do IDE, IDA e IDG.

As ferramentas da distribuição espacial possuem os seguintes campos a serem preenchidos (ESRI, 2016):

- *Input* (entrada) da *feature class* – é o campo de entrada da *feature class* que se pretende analisar;
- *Output* (saída) da *feature class* – nesse campo é especificado o local onde se pretende guardar os resultados
- *Peso* (opcional) – nesse campo é especificado a variável que servirá como ponderador para a determinação da medida de distribuição. Este campo é opcional;
- *Caso* (opcional) – campo usado pelos grupos de *features* para separar a medida de distribuição calculada. O caso pode ser inteiro, data ou *string*.
- *Dimensão* (opcional) – este campo é disponível apenas para a ferramenta *mean center*. É um campo numérico que deverá conter valores de atributos a partir dos quais é determinado o valor médio.

A medida distância padrão (ou desvio padrão) contém um campo denominado de *circle size* (tamanho dos círculos). O tamanho dos círculos de *output* em distância padrão. O tamanho do

círculo padrão é 1; as escolhas válidas são 1, 2 ou 3 desvios padrão. E, a elipse de desvio padrão (ou elipse de distância padrão) contém um campo denominado de tamanho da elipse (*ellipse size*).

Para que a aplicação das medidas distâncias padrão e elipse de desvio padrão sejam aplicáveis é necessário que a *feature class* de entrada tenha pelo menos três *features*. Em razão, foi utilizada como campo de separação a ilha, porém, a maioria das ilhas não possuem o mínimo de três concelhos. Neste caso, pode-se dizer que as medidas seriam válidas apenas para as ilhas de Santo Antão, Santiago e Fogo. Entretanto, no ano 2002 não existia o concelho de Santa Catarina do Fogo, o que levou a ter duas *features* para a ilha do Fogo e, conseqüentemente, não foi possível obter as duas medidas, distância padrão e elipse de desvio padrão, neste ano para esta ilha.

Como a unidade espacial deste estudo é o concelho (representado por um polígono) procedeu-se, primeiramente, a transformação dos polígonos em pontos, mais especificamente nos centros geométricos. De seguida, foram determinadas medidas de distribuição, iniciando pelo cálculo do centro médio e terminando na elipse de desvio padrão. Após a obtenção dos resultados, procedeu-se a conversão destes, que se encontrava num formato *Shape file*, para graus minutos e segundos usando a ferramenta *Convert Coordinate Notation* (Converter notação de coordenadas). De seguida foram exportados os resultados para o Excel para as devidas organizações tabulares.

3.5.2. Índice de vizinho mais próximo

A primeira hipótese em que se baseia para fazer inferência para uma população a partir de sua amostra é de que esta é aleatória. Assim, com a finalidade de verificar se as distribuições em análise são aleatórias, foi determinado o índice de vizinho mais próximo. Este foi obtido pela ferramenta *Average Nearest Neighbor* que compõe as ferramentas da estatística espacial. Nos campos que fazem parte desta ferramenta especifica-se o método de distância que se pretende utilizar, sendo que das duas distâncias disponíveis, euclidiana e a distância Manhattan, foi a euclidiana. O campo área é opcional e é onde se pode especificar o valor do tamanho da área (valor numérico).

3.5.3. Autocorrelação espacial

Para o teste de autocorrelação espacial foram determinados os índices Global e Local de Morans utilizando como campo os índices de desigualdade dos anos em análise. O índice Global de Morans foi obtido pela ferramenta *Spatial Autocorrelation* (Morans I) e o índice Local de Morans pela ferramenta "*Clusters and Outliers Analysis*". Esta última, identifica valores extremos espaciais, bem

como os *hot* e *cold spots* estatisticamente significativos, utilizando a estatística de Anselin Local Morans I (2). Dado um conjunto de *features* (Input *feature class*) e um campo de análise (campo de entrada), esta ferramenta identifica agrupamentos espaciais de *features* com valores altos e baixos. Para fazer isso, a ferramenta calcula o valor de um Morans Local I, *score-z*, valor *pseudo-p* e um código que representa o tipo de *cluster* para cada *feature* estatisticamente significativa. Os valores *pseudo-p* estão associados ao *scores-z* e, representam a significância estatística dos valores dos índices calculados. Na representação espacial, o resultado é uma tipologia de padrões COType que distingue entre uns *clusters* estatisticamente significativos:

- HH: Alto-Alto – agrupamento de valores altos e próximos. Indica que tanto o valor do atributo, quanto o valor médio para os seus vizinhos, estão acima da média do conjunto;
- HL: Alto-Baixo – *outlier* de valores altos que não se agrupam, pois se encontram em meio de valores baixos. Ou seja, valor positivo e média dos vizinhos negativa;
- LL: Baixo-Baixo – agrupamento de valores baixos e próximos. Indica que o atributo e média dos vizinhos estão abaixo da média do conjunto;
- LH: Baixo-Alto – *outlier* de valores baixos que não se agrupam, pois se encontram em meio de valores altos. Ou seja, valor negativo e média dos vizinhos positivos;

Pode-se verificar a dependência espacial através dessa representação espacial. As tipologias HH e LL indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes, ou seja, a área para o valor do atributo considerado está cercada por áreas que tem comportamento semelhantes. E, as tipologias HL e LH indicam pontos de associação espaciais negativas (valores baixos estão cercados por valores altos). Essas duas últimas tipologias representam os valores que não são possíveis de serem agrupados, *outliers*, pois apresentam valores distintos em suas proximidades. Contudo, a significância estatística pode ser observada através do *score-z* ou valor-p do índice.

3.5.4. Análise Hot Spot

A análise *Hot Spot* foi aplicada com a finalidade de verificar concelhos com índices de desigualdades elevados (*Hot Spot*), onde deve-se dar especial atenção. Para que o cálculo se realize é fundamental a escolha do campo de atributos ao qual se pretende analisar, sendo restrita aos campos que apresentam valores numéricos (Dias, 2013). Assim, foi utilizada como valor de atributo os índices de desigualdade de cada ano em estudo. O modelo de conceptualização

seleccionado foi a “banda” de distância fixa (*fixed distance band*), onde cada *feature* é analisada num contexto de *features* vizinhas, dentro de uma distância específica. Por sua vez, a escolha da distância é também um dos parâmetros, sendo esta a distância euclidiana. Esta deve-se ao facto, de que com esta distância seja definida uma distância em que uma *feature* tenha pelo menos uma *feature* vizinha. Para um índice de Morans significativo foi aplicada a correcção FDR na análise *Hot Spot*. Este último procedimento não se fez necessário apenas para o IDG do ano 2000.

Os parâmetros e seus respectivos tipos, bem como as suas descrições utilizados para a análise *Hot Spot*, encontram-se descritos na Tabela V.1.

3.5.5. Análise de *clusters*

Para a realização da análise de *cluster* foram utilizadas como variáveis os índices de desigualdade ambiental e de género (IDA e IDG), pois, estes índices não apresentaram *outliers*, valores atípicos capazes de influenciar os resultados, em nenhum dos anos em análise, ao contrário do índice de desigualdade em saúde e índice de desigualdade económica (Ver Gráfico II.1 – II.4) e, apresentam dependência espacial, ou seja, são passíveis de serem analisadas espacialmente. Por falta de dados suficientes para o cálculo do IDA no ano 2011, não foi aplicada a análise de *cluster* para este ano. Assim, foi possível agrupar concelhos de acordo com os índices dos anos 2000, 2010 e 2012 a 2015.

A ferramenta utilizada para aplicar a análise de *cluster* é o *Grouping Analysis*.⁴ Esta ferramenta cria grupos com restrições espaciais e temporais ou sem os mesmos. As restrições podem ser seleccionadas no campo restrição espacial (*constraint spatial*) onde é especificada a limitação espacial desejada (ESRI, 2016):

- Agrupamento por Contiguidade dos limites/bordas apenas (*Contiguity Edges Only*) – indica que as classes só podem ser parte do mesmo grupo se eles partilham um limite;
- Ângulo de Contiguidade das fronteiras/bordas (*Contiguity Edges Corners*) – indica que as classes só podem fazer parte do mesmo grupo se eles partilham um limite ou uma fronteira;
- Triangulação Delaunay (*Delaunay Triangulation*) – indica que uma classe é incluída em um grupo se pelo menos uma outra classe é vizinho natural;

⁴ As análises, usando as ferramentas do *ArcToolBox*, foram realizadas com o apoio do *BuildModel*. Estes modelos podem ser visualizados no Anexo I.

- k vizinhos mais próximo (*k nearest neighbors*) – indica que uma classe faz parte de um grupo se pelo menos uma outra classe é um dos k vizinhos mais próximo. k indica o número de vizinhos que devem ser tidos em conta;
- Ficheiro como base para os pesos espaciais (*Get spatial weights from file*) – define as relações espaciais e temporais entre classes;
- Nenhuma restrição espacial (*No constraint spatial*) – não impor restrições contiguidade ou proximidade.

Se nenhuma restrição espacial for aplicada a ferramenta *grouping analysis* usa um algoritmo k-médias, do inglês *k-means*, que é uma das técnicas não hierárquicas.

a) Número óptimo de Grupos

O número óptimo de grupos a criar na análise de *clusters* é de extrema importância na medida em que se pretende criar grupos com elementos homogéneos dentro do mesmo grupo e heterogéneos entre os grupos. Há casos em que não se sabe ao certo o número de grupos que se pretende criar. Assim, uma mais-valia que a ferramenta *grouping analysis* trás é a possibilidade de avaliar o número de grupo “ideal”. Quando é seleccionado o campo “avaliar o número ideal de grupo” esta calcula a estatística *pseudo-F* para cada número de grupo e o R^2 para cada variável dentro de cada grupo. Para este último quando mais elevado for o valor melhor a variável diferencie os grupos.

Neste sentido, antes da obtenção dos grupos foi aplicada a ferramenta *grouping analysis* com a finalidade de obter o número óptimo de grupos.

Foram, também, utilizadas a matriz de proximidades e o dendrograma para validar o número de grupos. A matriz de proximidade foi criada com a ajuda do SPSS, pelo facto de se pretender obter a distância entre os concelhos, e não entre os anos (*features layer*), aplicando os métodos vizinhos mais próximo e distância euclidiana. Havia possibilidade de obter esta matriz através da ferramenta *Generate Near Table* ou *Near*. Mas, para isso era preciso que cada concelho seja uma *feature class*, pois, estas ferramentas calculam as distâncias de uma *feature* à *feature* mais próxima. O dendrograma foi obtido através da ferramenta *dendrogram*. O processo da construção do dendrograma pode ser visualizado na Figura I.7.

O ponto de corte do dendrograma foi determinado utilizando o método de Mojena (1977), apresentado no capítulo 2, utilizando o valor de *k* igual a 1,25, como definido por Milligan e Copper (1985), e o índice RMSSTD. Para determinar o ponto de corte pelo método de Mojena (1977) foi

utilizado o resultado das distâncias combinadas entre cada classe, dado pela determinação do dendrograma. O cálculo foi realizado com o apoio do Excel. Para a determinação do índice R^2 também foi levada em consideração a tabela de distância dada pelo resultado do dendrograma, sendo cada combinação um grupo. Assim, foi determinado o índice R^2 para cada grupo, e o maior valor foi considerado o número óptimo de grupo a ser criado. Aquando do *pseudo-F*, o grupo com maior estatística é o número óptimo de grupo a ser produzido.

Como o número óptimo de grupo determinado para cada método foram muitos distintos, foram, então, criados grupos de acordo com o número de grupos conseguido pelo índice R^2 e pelo método de Mojena (1977). Assim, foram comparados os resultados destes dois métodos, chegando assim a conclusão que o número de grupo dado pelo método de Mojena é plausível. Os outros métodos foram excluídos pelo facto de apresentarem número de grupos elevados, tendo em conta o número de casos que se pretende agrupar, apesar de demonstrarem ser os mais adequados.

b) Técnicas de agrupamento

O algoritmo de agrupamento escolhido foi o hierárquico, utilizando como restrições o k vizinho mais próximo, e a distância escolhida foi a euclidiana (medida de dissemelhança), pois, as variáveis são quantitativas. O valor do k definido foi 21, referente ao total dos concelhos que não sejam a *target feature*, pois, sendo todos pertencentes ao mesmo país, considerou-se sendo todos vizinhos.

Os agrupamentos, por sua vez, foram determinados através dos métodos de Mojena (1977) e índice R^2 , pela aplicação da ferramenta *Grouping Analysis*. Os resultados dos agrupamentos foram mapeados por forma a apresentá-las no relatório. De seguida, esses resultados foram importados para o SPSS 22.0 para verificar a homogeneidade dentro dos grupos e heterogeneidade entre os grupos formados. Para tal foi aplicada o teste de Levene e a estatística ANOVA.⁵

⁵ A Análise de Variância (ANOVA) só é capaz de ser executada se as suas premissas são válidas: (a) Homoscedasticidade (variâncias entre grupo iguais); (b) Normalidade dos resíduos; (c) Aleatoriedade e independência das observações.

4. ANÁLISE DOS RESULTADOS

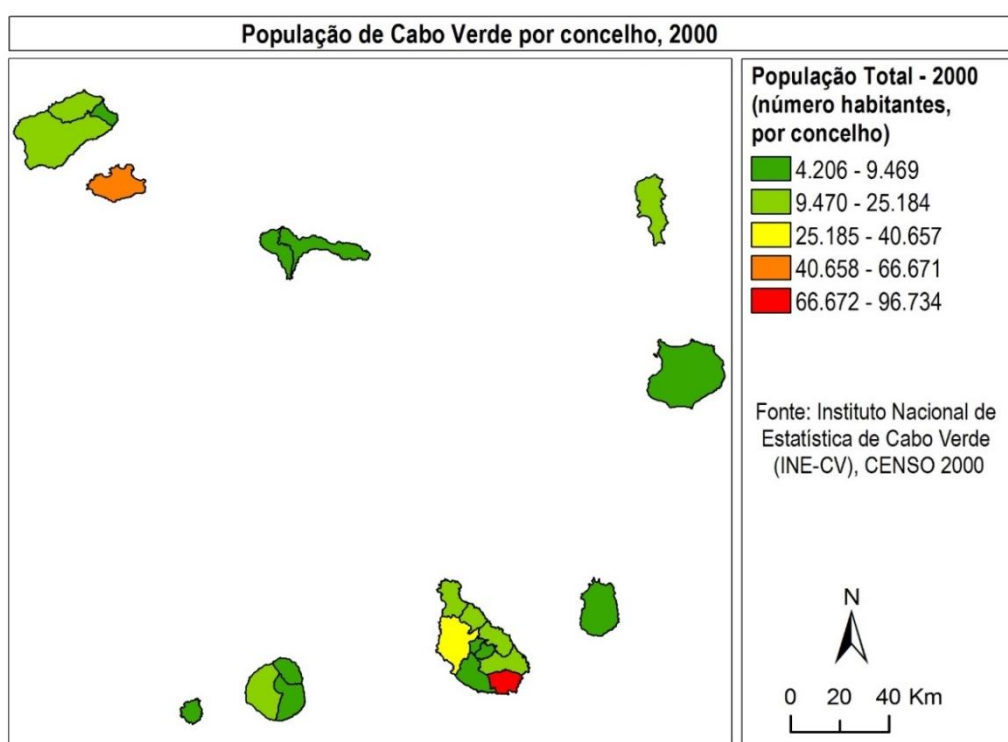
Neste capítulo são apresentados os resultados obtidos através do tratamento e análise dos dados e, esta dividida em quatro secções. A primeira secção é destinada a distribuição geográfica dos dados. Nela é apresentada a distribuição da população Cabo-Verdiana nos dois anos censitários, 2000 e 2010, a distribuição espacial das médias dos índices de desigualdades dos diferentes anos em análise e as medidas de distribuição espaciais (centro médio, distância padrão e elipse de distância padrão). Na segunda secção, é apresentada o resultado da análise de autocorrelação espacial, bem como o índice de vizinho mais próximo que foi aplicado com o intuito de verificar a aleatoriedade dos dados. Na terceira secção, encontra-se o resultado alcançado com a análise de *Hot Spot* e, finalmente, na quarta secção são apresentados os resultados da análise de *clusters*.

4.1. Distribuição geográfica dos dados

4.1.1. Distribuição espacial da população de Cabo Verde: Censo 2000 e 2010

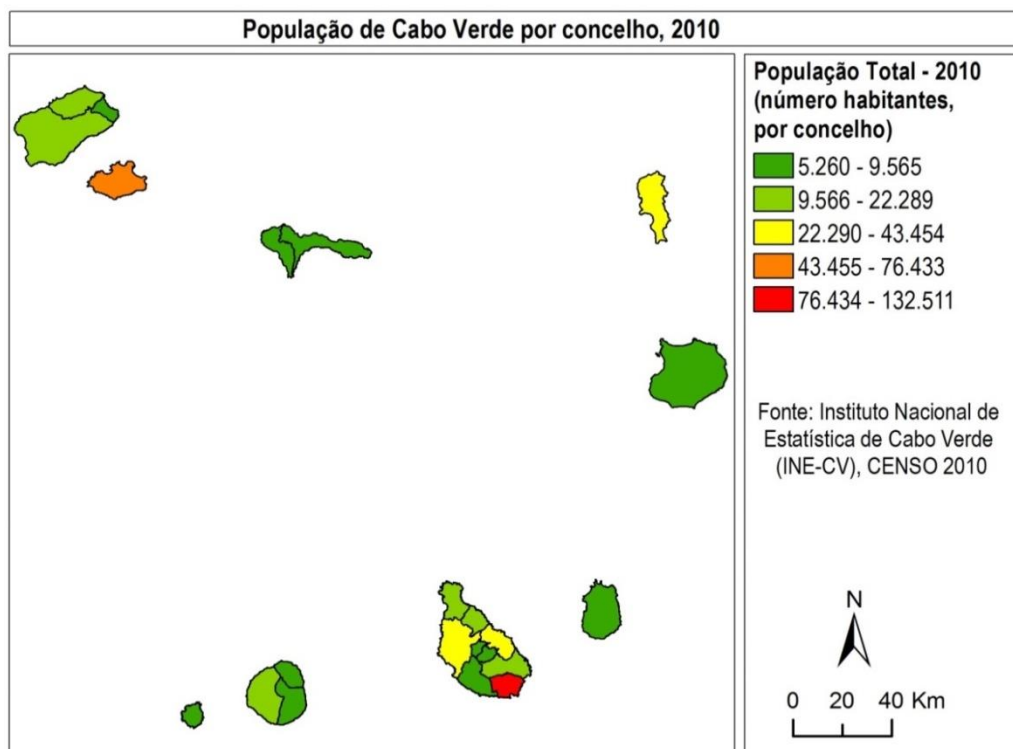
Nas Figuras 4.1 e 4.2, pode-se visualizar a distribuição espacial da população de Cabo Verde do ano 2000 e 2010, respectivamente, pelo método de *Natural Breaks* (Jenks). Nota-se que o concelho da Praia é, nos dois anos, a com maior número de população seguido pelo concelho de São Vicente. Percebe-se um aumento relativo do número de população dos concelhos Sal e Santa Cruz do ano 2000 para 2010.

Figura 4.1: Distribuição espacial da População de Cabo Verde por concelho, pelo método Natural Breaks (Jenks), 2000.



Fonte: Elaborado pela autora.

Figura 4.2: Distribuição espacial da População de Cabo Verde por concelho, pelo método *Natural Breaks* (Jenks), 2010



Fonte: Elaborado pela autora

4.1.2. Distribuição espacial dos Índices de Desigualdade

Na Tabela 4.1 encontra-se algumas estatísticas descritivas como média, desvio padrão, amplitude e coeficiente de variação dos índices de desigualdade. Nota-se uma ligeira diminuição nas médias dos índices de desigualdades com o passar dos anos, excepto para o índice de desigualdade ambiental. Este apresenta variação no valor das médias dos índices e, o índice do ano 2010 é duas vezes maior que do ano 2000. As Figuras 4.3 – 4.12 apresentam as distribuições espaciais dos índices de desigualdade conforme a classificação da desigualdade apresentada anteriormente

na Tabela 3.2.

Tabela 4.1: Estatísticas descritivas dos Índices de Desigualdade, 2000 a 2015

		Estatísticas			
		Média	Desvio Padrão	Amplitude	C.V.
IDS	2007	0,000 ^(a)	0,000 ^(b)	0,001	10,07
IDE	2002	0,322	0,057	0,214	0,176
	2007	0,238	0,053	0,192	0,225
IDG	2000	0,548	0,031	0,123	0,056
	2010	0,503	0,019	0,066	0,038
	2011	0,501	0,020	0,070	0,040
	2012	0,420	0,018	0,057	0,042
	2013	0,488	0,015	0,053	0,031
	2014	0,421	0,016	0,059	0,038
	2015	0,419	0,013	0,044	0,031
IDA	2000	0,216	0,115	0,398	0,533
	2010	0,495	0,110	0,370	0,223
	2012	0,612	0,105	0,350	0,172
	2013	0,618	0,108	0,365	0,175
	2014	0,508	0,090	0,289	0,177
	2015	0,519	0,085	0,281	0,164

C.V. - Coeficiente de Variação

(a) o valor é de aproximadamente 2,407E-05

(b) o valor é de aproximadamente 2,42E-04

Fonte: Elaborada pela autora

O índice de desigualdade económico variou de baixa a média no ano 2002 (Figura 4.3) e de muito baixa a baixa no ano 2007 (Figura 4.4), verificando assim a diminuição do índice. Em relação ao ano 2002, os dois concelhos com maiores índices foram os concelhos São Vicente e Maio (com desigualdade média), sendo os restantes concelhos com desigualdade baixa (Figura 4.3). Relativamente ao ano 2007, estes dois concelhos apresentam desigualdade baixa. Enquanto, o Paúl, Tarrafal S.N, São Miguel, São Salvador do Mundo, São Lourenço dos Órgãos e Santa Catarina do Fogo apresentam índices muito baixos, mantendo-se os restantes concelhos nos mesmos níveis de desigualdade (Figura 4.4).

Figura 4.3: Distribuição espacial do índice de Desigualdade Económico, 2002

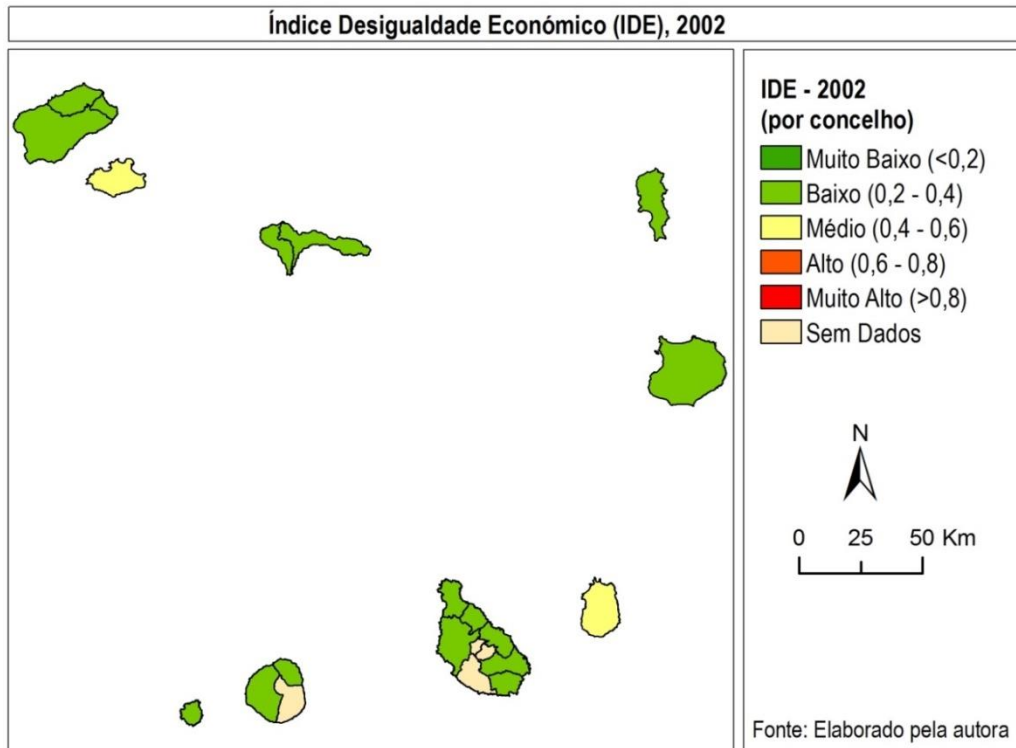
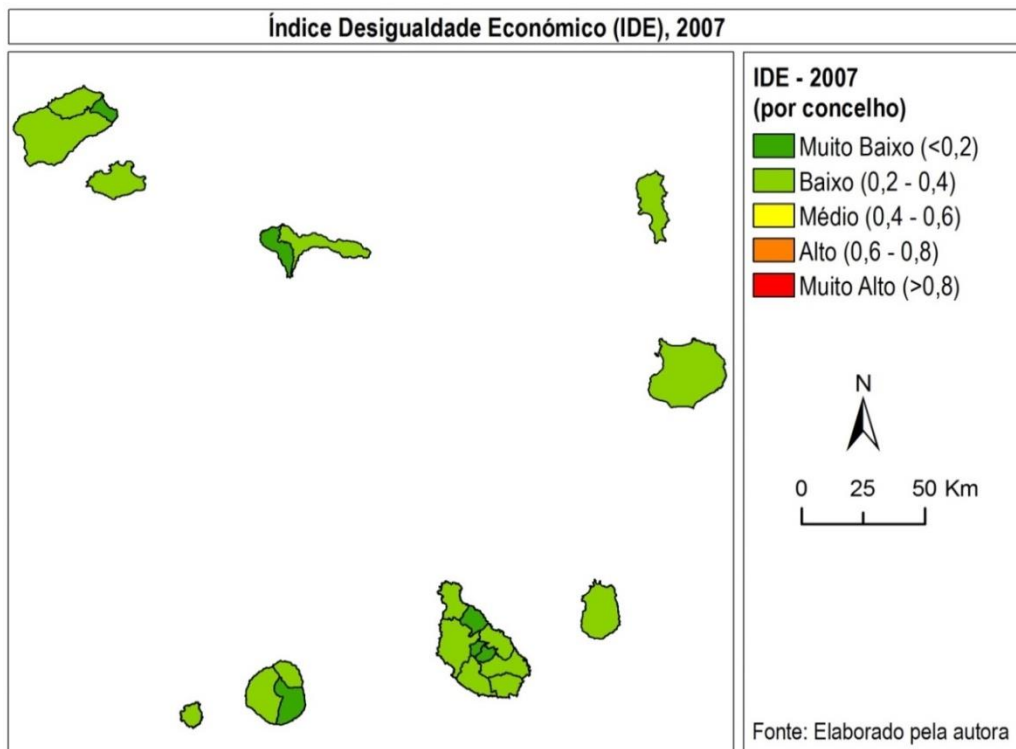


Figura 4.4: Distribuição espacial do índice de Desigualdade Económico, 2007



O índice de desigualdade de género do ano 2000 (Figura 4.5) é elevado para os concelhos Santa Cruz e Brava, sendo os restantes concelhos com índices médio. Nos anos 2010, 2011 e 2013 (Figuras II.1, II.2, II.3, respectivamente) todos os concelhos apresentam desigualdade de género média e, o mesmo acontece para os anos 2014 e 2015, com excepção dos concelhos de São Salvador do Mundo e São Lourenço dos Órgãos que apresentam índices baixos (Figura II.4, II.5). Estes dois últimos concelhos apresentam o mesmo comportamento no ano 2012, em que os concelhos Tarrafal e São Miguel também apresentam índices baixos (Figura 4.6).

Figura 4.5: Distribuição espacial do índice de Desigualdade de Género, 2000

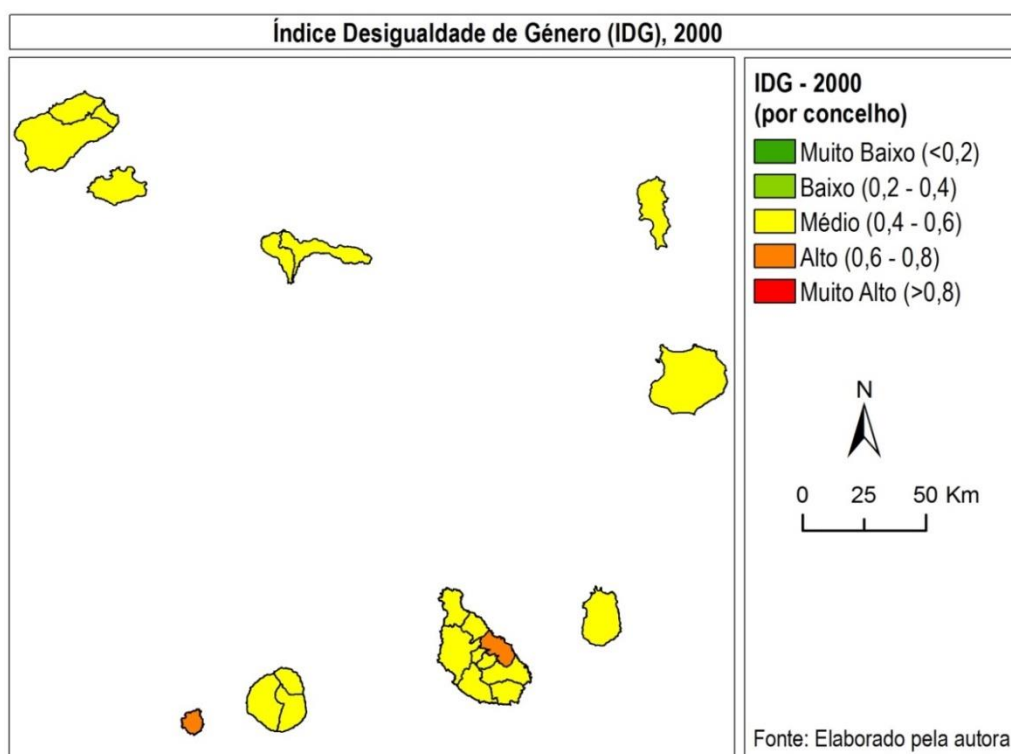
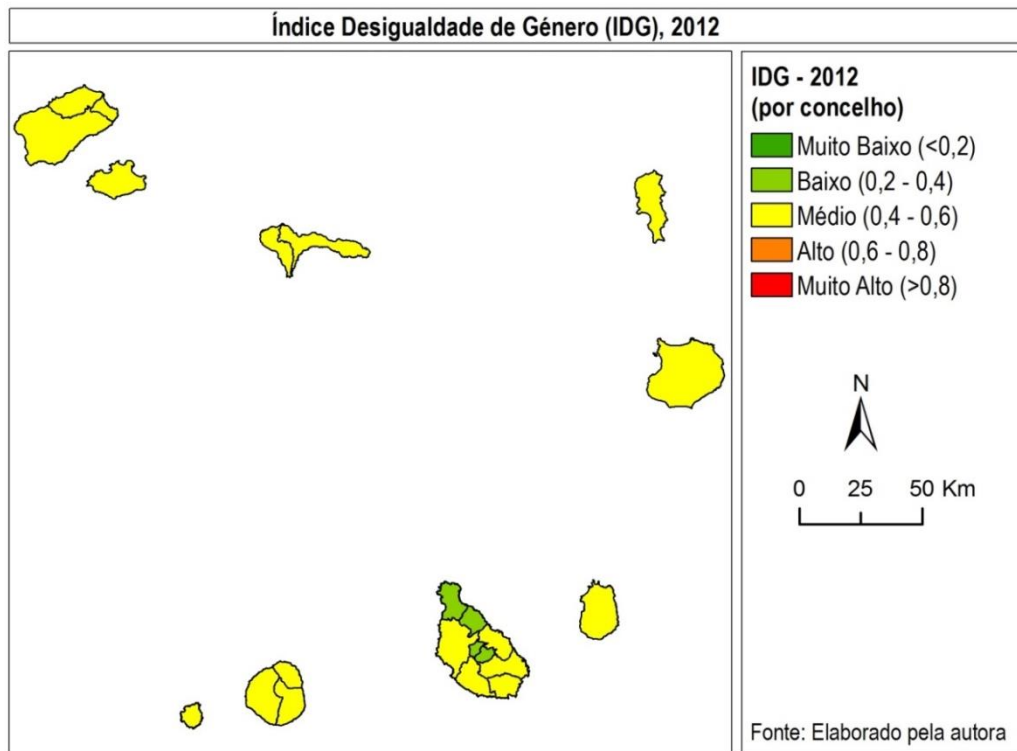


Figura 4.6: Distribuição espacial do índice de Desigualdade de Género, 2012



O IDA apresenta comportamento oscilante com o passar dos anos. Esse comportamento pode ser ainda visualizado com a distribuição espacial deste índice (Figura 4.7 – 4.12). Pode-se verificar através da distribuição espacial do IDA dos anos 2000 a 2015, que os índices mais baixos são do ano 2000, em que os concelhos São Vicente e Sal apresentam índices mais elevados, índices médios (Figura 4.7), comparativamente aos restantes anos. Já no ano 2010 (Figura 4.8) a desigualdade ambiental aumentou passando de muito baixa a média, no ano 2000, à baixa e elevada. De tal modo, houve um aumento significativo deste índice no ano 2012, com a maioria dos concelhos com índices elevados (Figura 4.9). Este índice diminuiu no ano 2014, em que a maioria dos concelhos apresentaram índices médios (Figura 4.11). É ainda de referir que a desigualdade ambiental manteve-se no mesmo nível para todos os concelhos, relativamente ao ano 2015 quando comparado com 2014, com excepção do concelho São Domingos, que passou de um nível de desigualdade baixa à média (Figura 4.12).

Figura 4.7: Distribuição espacial do índice de Desigualdade Ambiental, 2000

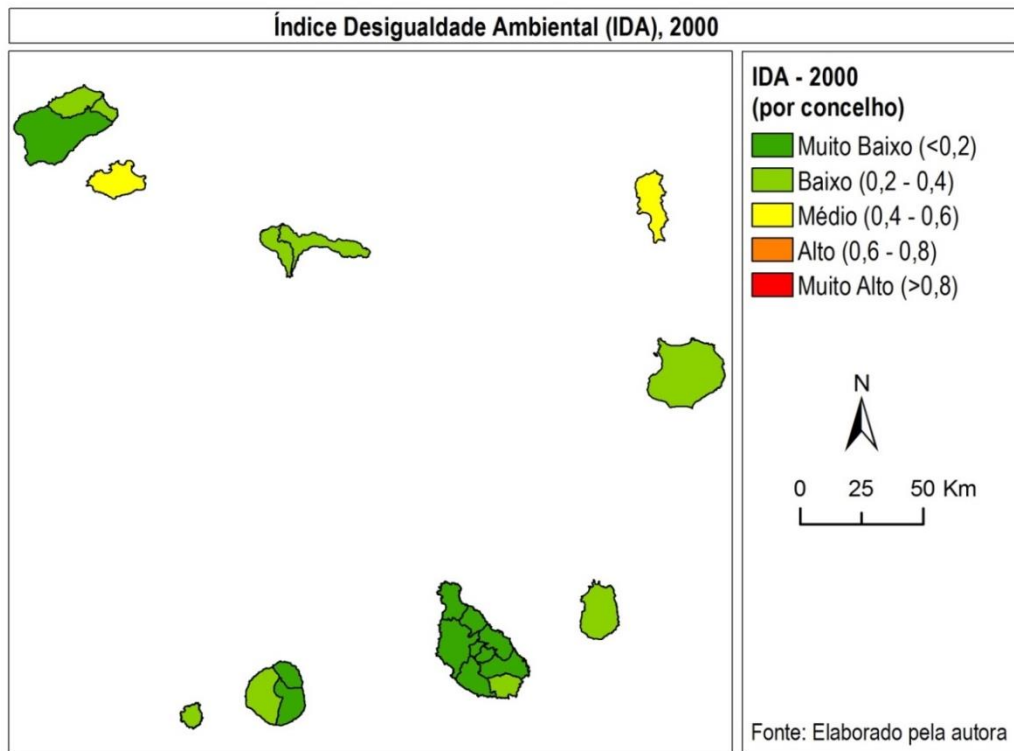


Figura 4.8: Distribuição espacial do índice de Desigualdade Ambiental, 2010

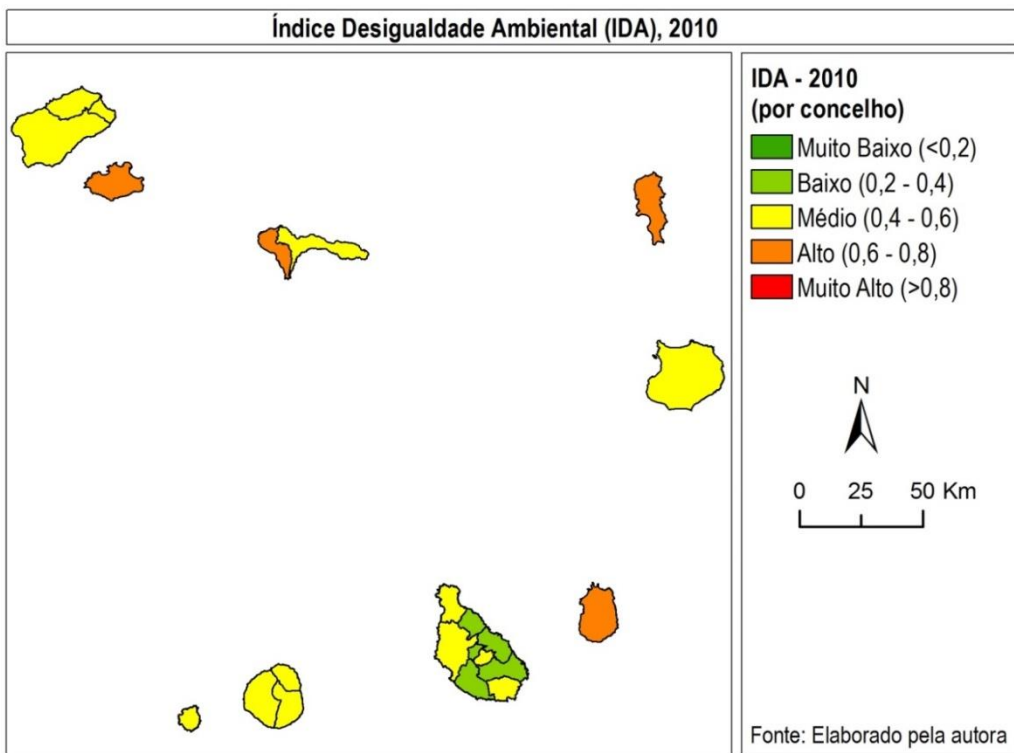


Figura 4.9: Distribuição espacial do índice de Desigualdade Ambiental, 2012

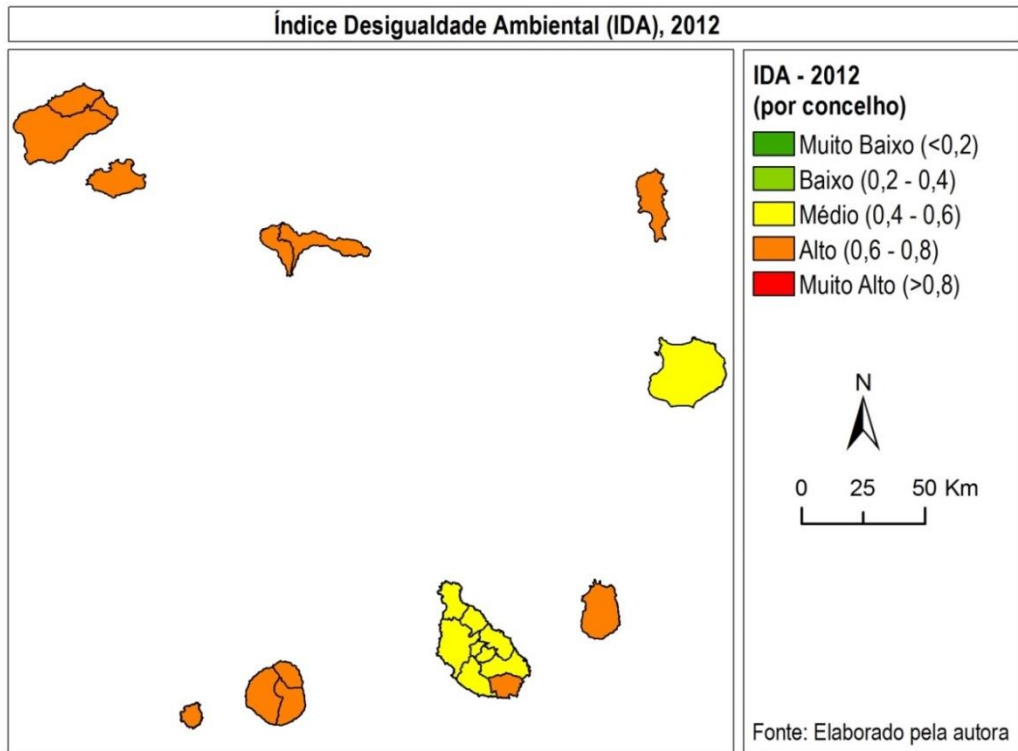


Figura 4.10: Distribuição espacial do índice de Desigualdade Ambiental, 2013

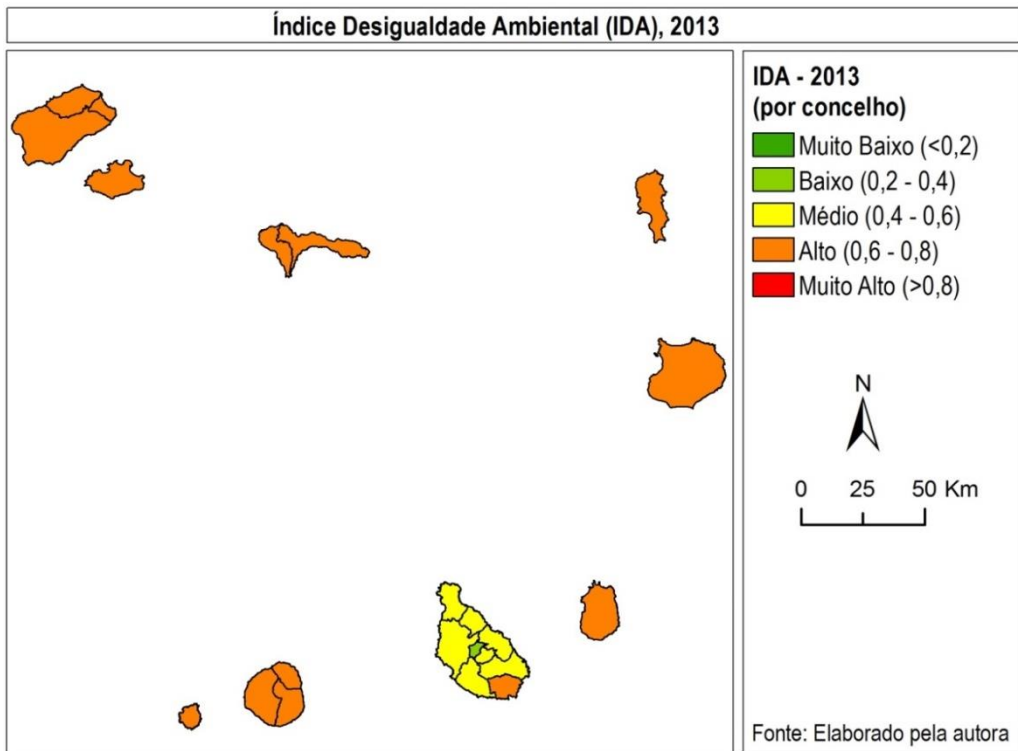


Figura 4.11: Distribuição espacial do índice de Desigualdade Ambiental, 2014

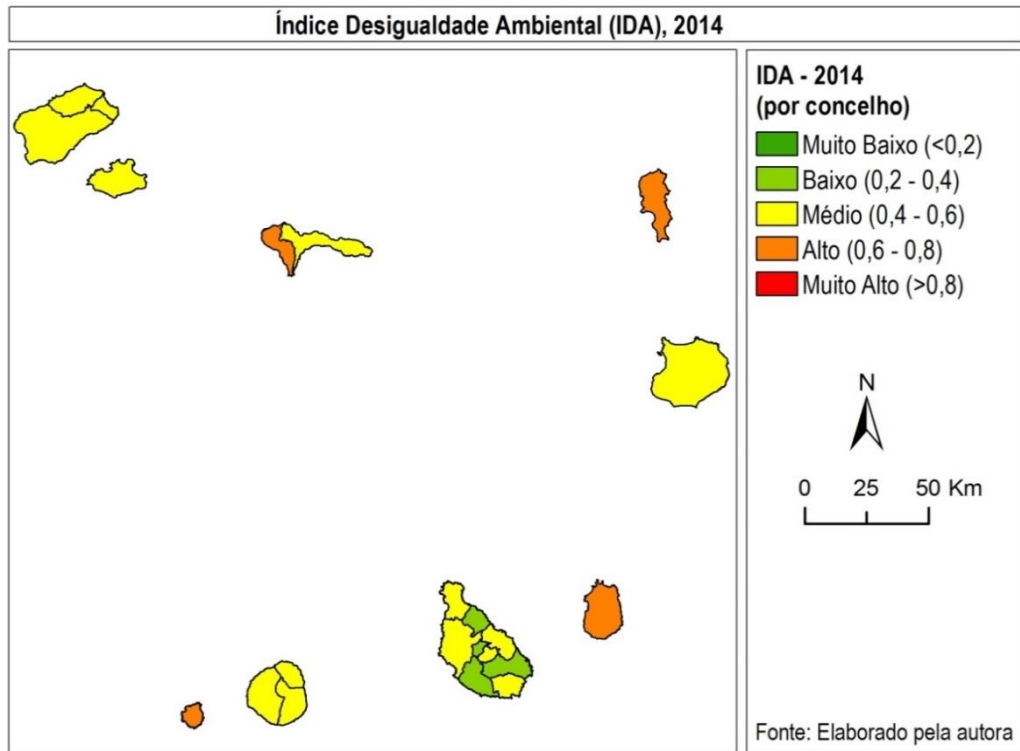
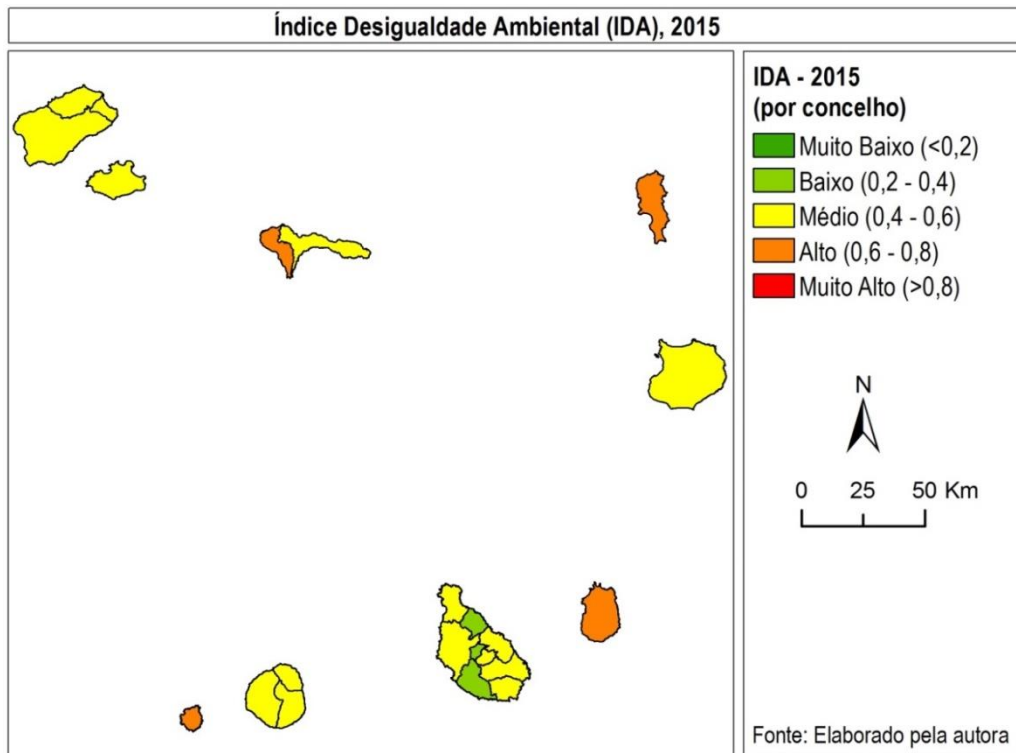


Figura 4.12: Distribuição espacial do índice de Desigualdade Ambiental, 2015



O comportamento, dos índices de desigualdade, pode ser ainda observado através dos gráficos *Box Plot* (Gráfico II.1 – II.4). Através deles pode-se verificar o aumento e redução da desigualdade com o passar dos anos, bem como os *outliers*. No ano 2007 o índice de desigualdade em saúde mostra que os concelhos Terrafal S.N. e Santa Cruz são *outliers* severos (apresentada por asteriscos), ou apresentam valores atípicos, ou seja, apresentam índices elevados em relação aos outros concelhos (Gráfico II.1). Neste ano, Terrafal S.N. apresentou a maior taxa de mortalidade (8,7%), em relação aos restantes concelhos (Relatório Estatístico 2007, Ministério de Saúde). Estes dois concelhos (Terrafal S. N. e Santa Cruz) tinham infra-estruturas de saúde “pobres”, quando comparado aos outros concelhos. No ano 2002, o IDE apresenta como *outliers* moderados os concelhos São Vicente e Maio (Gráfico II.2). Os índices de desigualdade ambiental e de género não apresentam *outliers* em nenhum dos anos em análise (Gráficos II.3 e II.4). Este facto podia ser argumentado, tendo em conta que, tanto a ilha de São Vicente como Maio, possuíam a maioria da população com níveis de rendimento muito baixo e com nível de educação, também, baixo.

4.1.3. Características espaciais da distribuição do rendimento total e *per capita*, IDE, IDA e IDG

A distribuição espacial dos rendimentos, total e *per capita*, e das desigualdades (IDE, IDA e IDG) não apresentariam uma interpretação satisfatória e fácil dos dados em função da complexidade das informações. Neste sentido, optou-se por utilizar métodos que generalizam espacialmente as informações e, portanto, facilitam sua interpretação em termos gerais. Primeiro, foram determinados os centros médios (ponderados) do rendimento total, rendimento *per capita* e dos índices de desigualdade, ou seja, os pontos imaginários que representam os centros de gravidade destas variáveis no período 2000 a 2015 (Anexo III).

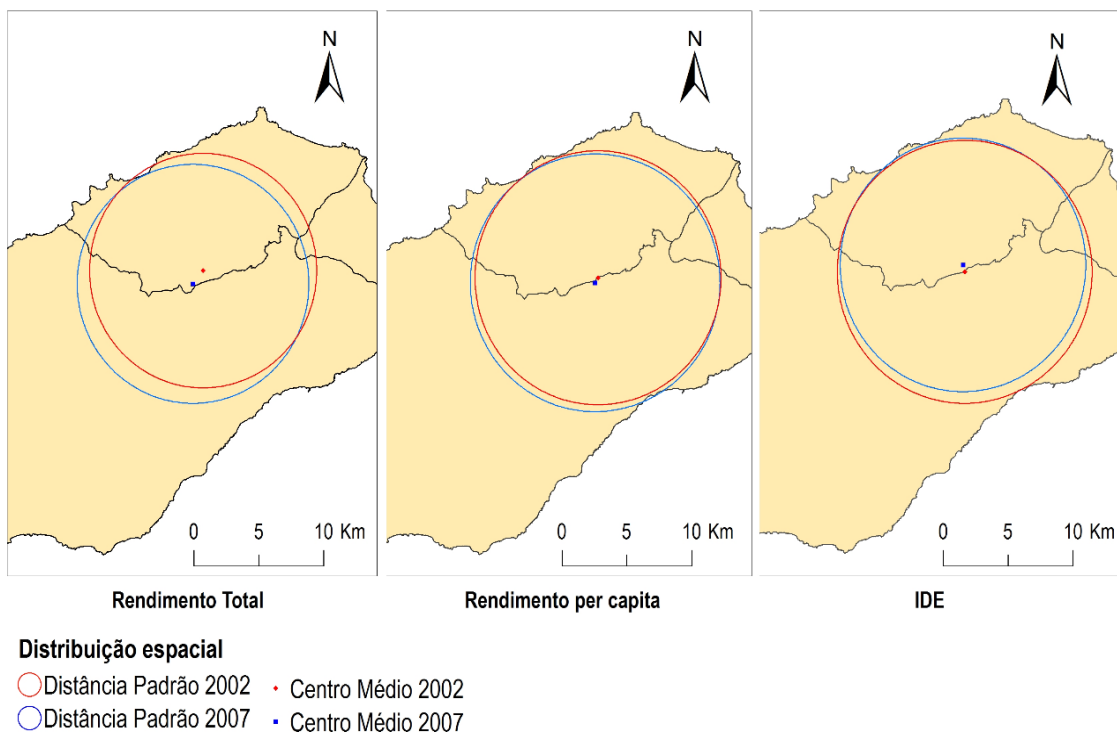
Em resumo, o centro médio do rendimento total da Ilha de Santo Antão está localizado no concelho Ribeira Grande, a da Ilha de Santiago no concelho São Domingos e, do Fogo no concelho São Filipe e proximidades (Tabela III.1). Ou seja, o rendimento total, destas ilhas, está concentrado nestes concelhos, e/ou concelhos próximos. Porém, o mais interessante verifica-se no deslocamento do rendimento total na ilha de São Nicolau, passando do concelho Terrafal S.N, no ano 2002, para Ribeira Brava, no ano 2007. Relativamente ao centro médio do rendimento *per capita* da ilha de Santo Antão, pode-se notar que este situou-se, tanto no ano 2002 como 2007, no concelho de Terrafal S.N. (Tabela III.2). É notável também, o deslocamento do rendimento *per capita* da ilha de Santiago e Fogo, passando de São Lourenço dos Órgãos para São Salvador do

Mundo e de São Filipe para Santa Catarina do Fogo, respectivamente. Já para o centro médio do IDE, pode-se dizer que não houve deslocamento significativo em nenhuma das ilhas (Santo Antão, São Nicolau, Santiago e Fogo). Ou seja, o centro médio situou-se nos mesmos concelhos quando comparado os dois anos, 2002 e 2007 (Tabela III.3). No que concerne ao centro médio das desigualdades de género e ambiental, praticamente não houve deslocamento das desigualdades nas ilhas (Tabela III.8 – III.11). O centro médio da desigualdade ambiental apresenta um ligeiro deslocamento, mais precisamente na ilha de Santiago, passando do concelho de São Salvador do Mundo, em 2000, para São Lourenço dos Órgãos, nos anos 2010 a 2014, e para o limite entre este concelho e o concelho de São Lourenço dos Órgãos, em 2015 (Tabela III.10, III.11).

A distância padrão (ponderada) do rendimento total, *per capita* e das desigualdades, indica um comportamento regular com respeito à medida de dispersão espacial do rendimento total, rendimento *per capita* e desigualdades da ilha de Santo Antão (Tabela III.4). Observa-se, em primeiro lugar, que não houve modificações tão significativas no período analisado. A distância padrão do IDE diminuiu tanto na Ilha de Santo Antão como Santiago, sendo que nesta última houve uma diminuição de aproximadamente 1,42 Km. Ao contrário, se observa com a distância padrão do rendimento total, onde houve um aumento não muito significativo. Relativamente a distância padrão do rendimento *per capita*, é notável a sua diminuição na ilha de Santiago, passando de aproximadamente 16 Km à 13,27 Km. Em relação ao IDG não houve modificações significativas da distância padrão, e para o IDA houve um ligeiro aumento, na ilha de Santiago, de aproximadamente 1,42 Km em 2010, quando comparado com o ano 2000 (Tabela III.12).

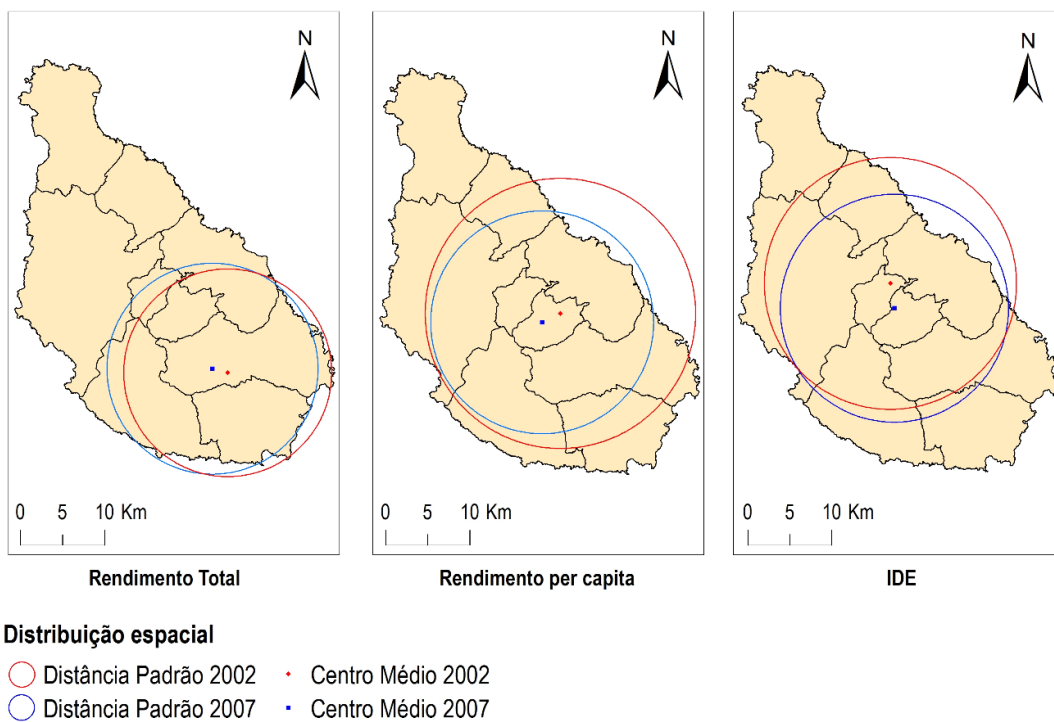
Pode-se visualizar as distâncias padrão e os centros médios das variáveis em análise nas Figuras 4.13 – 4.16. Nas figuras podemos observar as distâncias em relação ao centro médio dos diferentes anos em análise para a ilha de Santo Antão, Santiago e Fogo analisados anteriormente através dos resultados contidos nas tabelas em anexo.

Figura 4.13: Centro médio e distância padrão do rendimento total e *per capita* e IDE da ilha de Santo Antão, 2002 e 2007



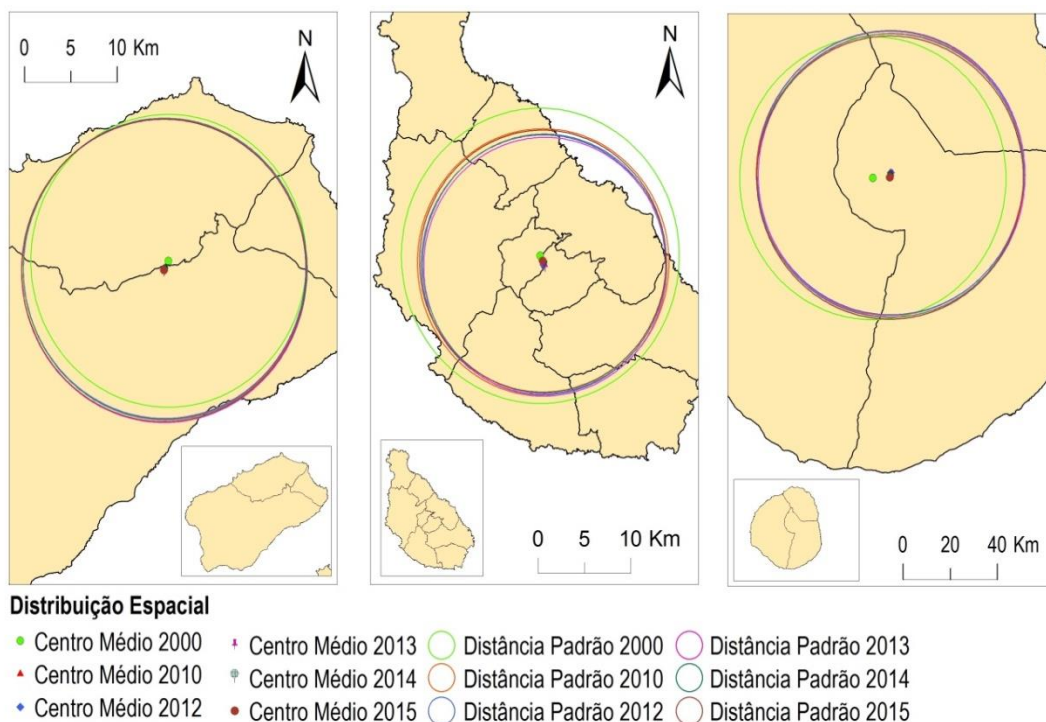
Fonte: Elaborado pela autora

Figura 4.14: Centro médio e distância padrão do rendimento total e *per capita* e IDE da ilha de Santiago, 2002 e 2007



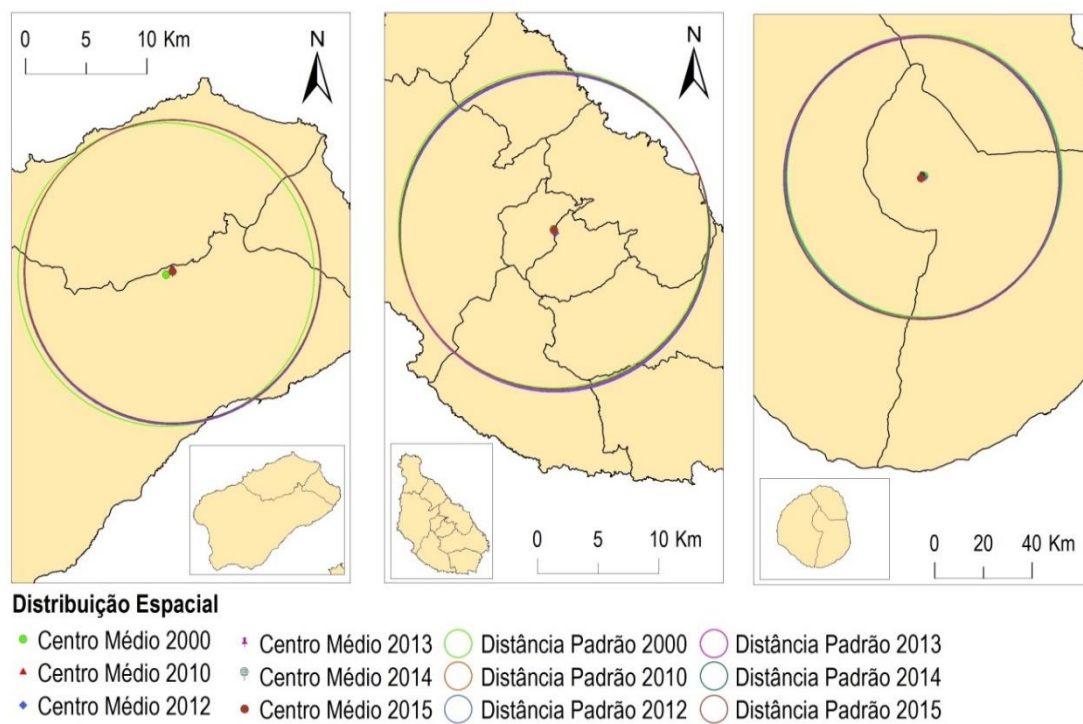
Fonte: Elaborado pela autora

Figura 4.15: Centro médio e distância padrão da desigualdade ambiental da ilha de Santo Antão, Santiago e Fogo, 2000 a 2015



Fonte: Elaborado pela autora

Figura 4.16: Centro médio e distância padrão da desigualdade de Género da ilha de Santo Antão, Santiago e Fogo, 2000 a 2015

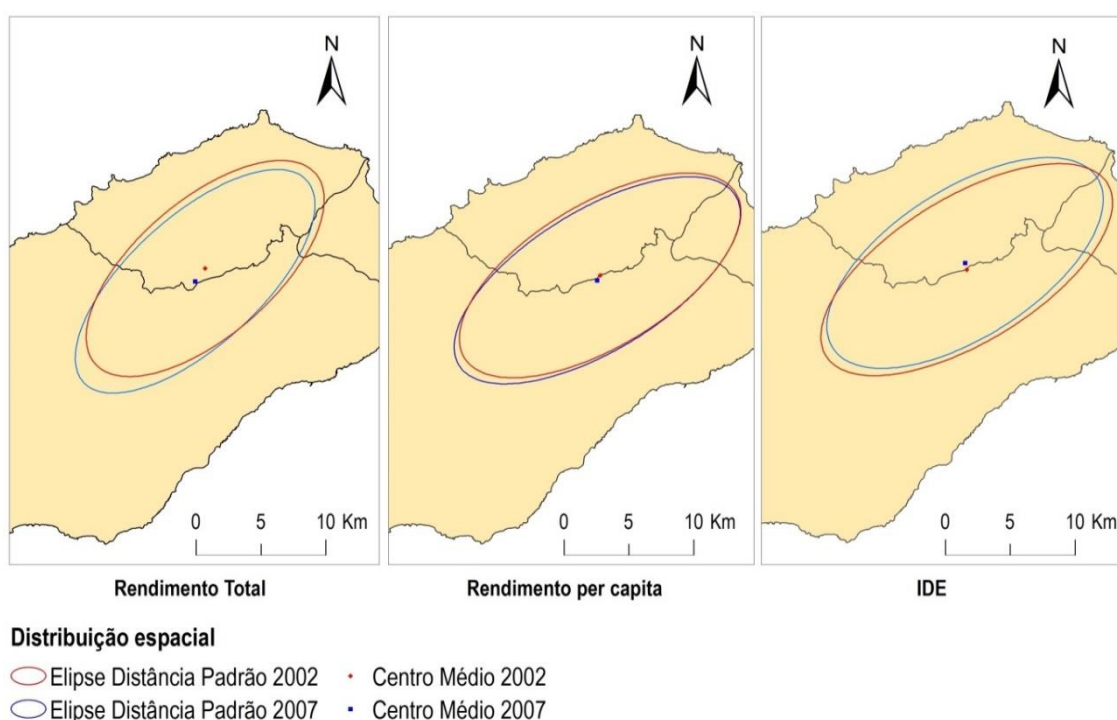


Fonte: Elaborado pela autora

Por fim, procede-se a construção das elipses de desvio padrão que, além de indicarem a localização da concentração do rendimento total e *per capita* e das desigualdades, indicam a orientação geral da distribuição espacial nos dois anos analisados. Ao analisarmos as elipses da ilha de Santo Anão, observa-se que o rendimento total se concentra no concelho Ribeira Grande e Porto Novo em ambos os anos, porém, no ano 2002 esta direcciona-se para o concelho Ribeira Grande (Figura 4.17). Já no ano 2007, esta aponta para o concelho Porto Novo. É notável também que o rendimento total da ilha não abrange o concelho Paúl (o concelho não se encontra dentro da elipse, uma pequena parte, praticamente o limite do concelho, encontra-se dentro da elipse). Ou seja, o rendimento total não é predominante neste concelho. Este resultado vem ao encontro do esperado, pois, este concelho, em comparação com os outros dois concelhos da ilha, não é, significativamente, desenvolvido, representado apenas 18% da população desta ilha. O concelho de Ribeira Grande era, e é, o principal concelho da ilha, apesar do Porto Novo ser o maior concelho em termos de área. Assim, faz todo o sentido a concentração de rendimento neste concelho. O rendimento *per capita*, por sua vez, abrange os três concelhos, assim como a desigualdade económica, retratando a deposição do rendimento *per capita* e da desigualdade para todos os concelhos da ilha. As elipses da ilha de Santiago (Figura 4.18) mostram que o rendimento total está, especialmente, concentrado no espaço demarcado pelo concelho São Domingos e Praia, a sul da ilha. Tanto no ano 2002 como no ano 2007, os concelhos Tarrafal e São Miguel não se encontram dentro da elipse e, ainda no ano 2007, o concelho Santa Cruz também apresenta o mesmo comportamento. O rendimento *per capita* está concentrado nos concelhos São Salvador Mundo, São Lourenço dos Órgãos e São Domingos, no ano 2002. No ano 2007, este direcciona-se para o concelho da Praia, porém, abrangendo todos os concelhos da ilha, com excepção do concelho Ribeira Grande de Santiago. Este resultado comprova a centralização da economia de Cabo Verde, no concelho da Praia, onde uma das estratégias do governo, agora, é a regionalização. A concentração do rendimento mais a sul da ilha de Santiago deve-se ao facto do concelho da Praia englobar mais de 70% do rendimento total no ano 2002. Ele é o maior concelho do país e, também, a maior cidade, tem aproximadamente 1/3 da população de Cabo Verde e com maior economia e infra-estruturas. Quando se analisa as elipses do IDE, verifica-se que a desigualdade abrange todos os concelhos da ilha, em ambos os anos. Porém, ela direcciona-se mais a Norte da Ilha no ano 2002 e mais a Sul no ano 2007. A desigualdade ambiental (Figura 4.19), assim como a desigualdade de género (Figura 4.20), também abrangem todos os concelhos da ilha de Santo Antão, Santiago e Fogo, sendo que para a desigualdade de género o seu deslocamento não é

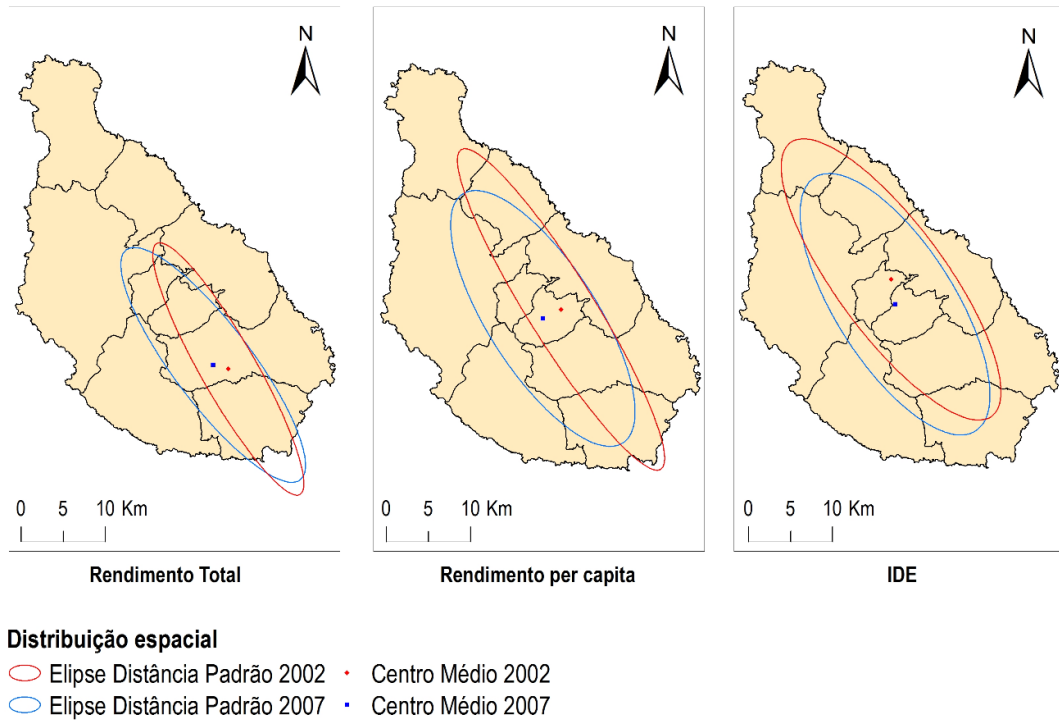
significativo. Para a desigualdade ambiental o deslocamento é significativo quando comparado a desigualdade do ano 2000 em relação ao ano 2010. A estabilidade da desigualdade de género pode dever-se as decisões políticas e medidas institucionais, tomadas com base nas desigualdades de género, que teve início em 1997 com a criação do Plano Nacional de Desenvolvimento (PND), e que veio apresentar resultados satisfatórios entre os anos 2002 e 2005. Um outro objectivo deste plano era assegurar a sustentabilidade ambiental, porém, este não teve impacto muito significativo, como a igualdade e equidade de género.

Figura 4.17: Centro médio e orientação da média do rendimento familiar *per capita*, total e IDE, 2002 e 2007



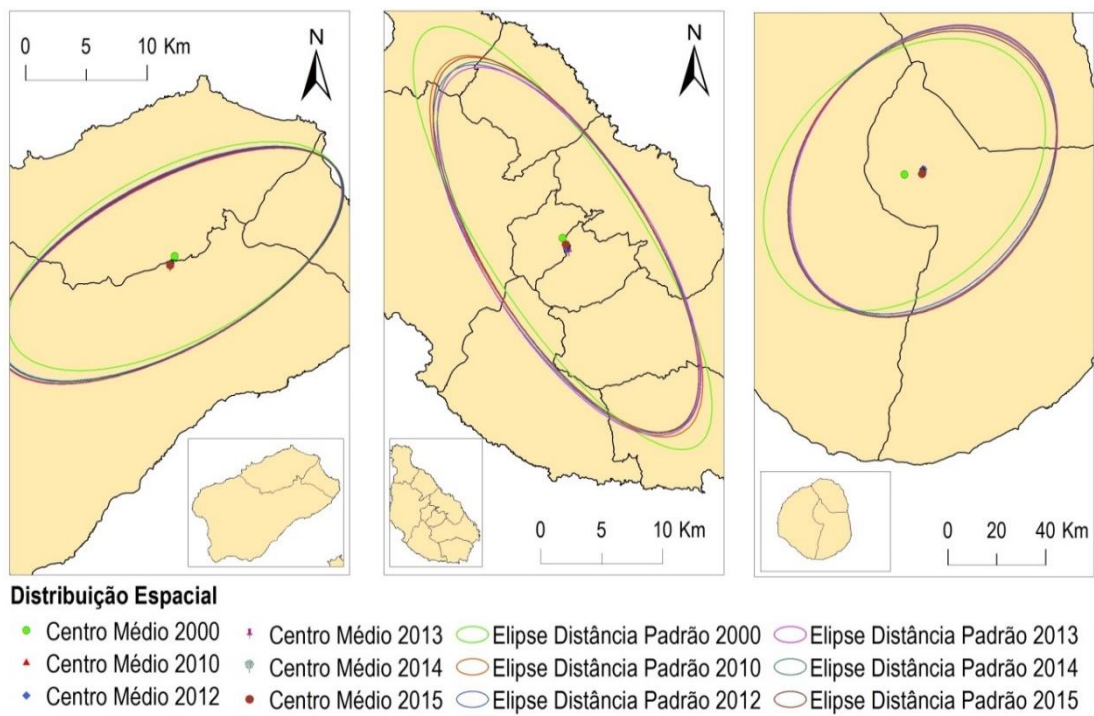
Fonte: Elaborado pela autora

Figura 4.18: Centro médio e orientação do rendimento total, rendimento familiar *per capita*. total e IDE, 2002 e 2007



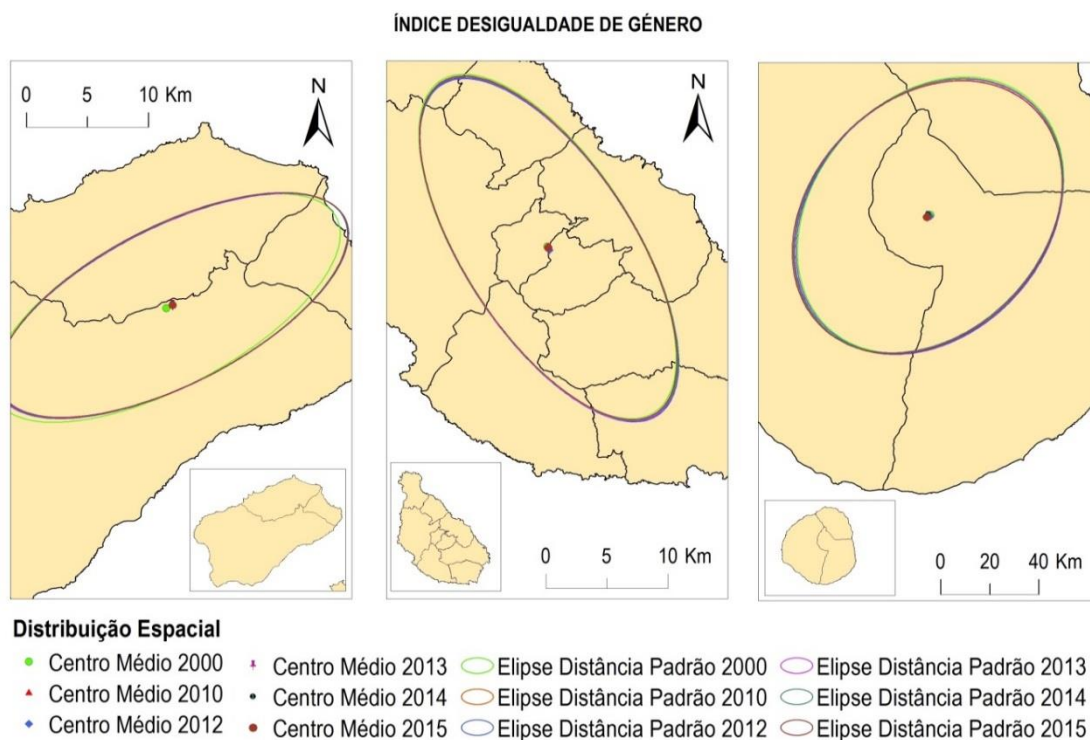
Fonte: Elaborado pela autora

Figura 4.19: Centro médio e orientação desigualdade ambiental, 2000 a 2015



Fonte: Elaborado pela autora

Figura 4.20: Centro médio e orientação desigualdade de género, 2000 a 2015



Fonte: Elaborado pela autora

4.2. Autocorrelação espacial

A autocorrelação espacial, como foi referido no capítulo 2, permite estimar o quanto o valor observado de um atributo é dependente dos valores desse mesmo atributo nas localizações vizinhas. Assim, da análise da Tabela 4.2, observa-se que o índice de desigualdade de género do ano 2000 é independente dos valores dos seus vizinhos, ou seja, o IDG de um concelho é independente do IDG dos concelhos vizinhos. Em contrapartida, a autocorrelação espacial do IDG dos outros anos em análise (2010 a 2015) é significativa a um nível de confiança de 99%, exibindo dependência espacial. O índice Global de Morans comprova um padrão de dependência espacial da distribuição da desigualdade de género em Cabo Verde, bem como da desigualdade ambiental, pois, é significativamente maior que zero. Este, apresenta uma autocorrelação espacial e, sendo positiva, indica que a desigualdade de género e ambiental de um concelho tende a ser mais semelhante ao concelho vizinho do que concelhos mais distantes.

Tabela 4.2: Índice Global de Moran das desigualdades

	Ano	Índice	Z-score	p-value
IDS	2007	-0,071	-0,281	0,779
IDE	2002	-0,207	-1,078	0,281
	2007	-0,117	-0,731	0,465
IDG	2000	-0,035	0,132	0,895
	2010	0,434	5,053	0,000
	2011	0,716	6,451	0,000
	2012	0,310	3,747	0,000 ^(a)
	2013	0,269	3,340	0,001
	2014	0,257	3,204	0,001
	2015	0,286	3,490	0,000 ^(b)
IDA	2000	0,240	3,039	0,002
	2010	0,379	4,479	0,000 ^(c)
	2012	0,326	3,920	0,000 ^(d)
	2013	0,389	4,612	0,000 ^(e)
	2014	0,345	4,107	0,000 ^(f)
	2015	0,342	4,090	0,000 ^(g)

(a) o valor é de aproximadamente 1,79E-04

(b) o valor é de aproximadamente 4,82E-04

(c) o valor é de aproximadamente 8,00E-06

(d) o valor é de aproximadamente 8,90E-05

(e) o valor é de aproximadamente 4,00E-06

(f) o valor é de aproximadamente 4,00E-05

(g) o valor é de aproximadamente 4,30E-05

Fonte: Elaborada pela autora

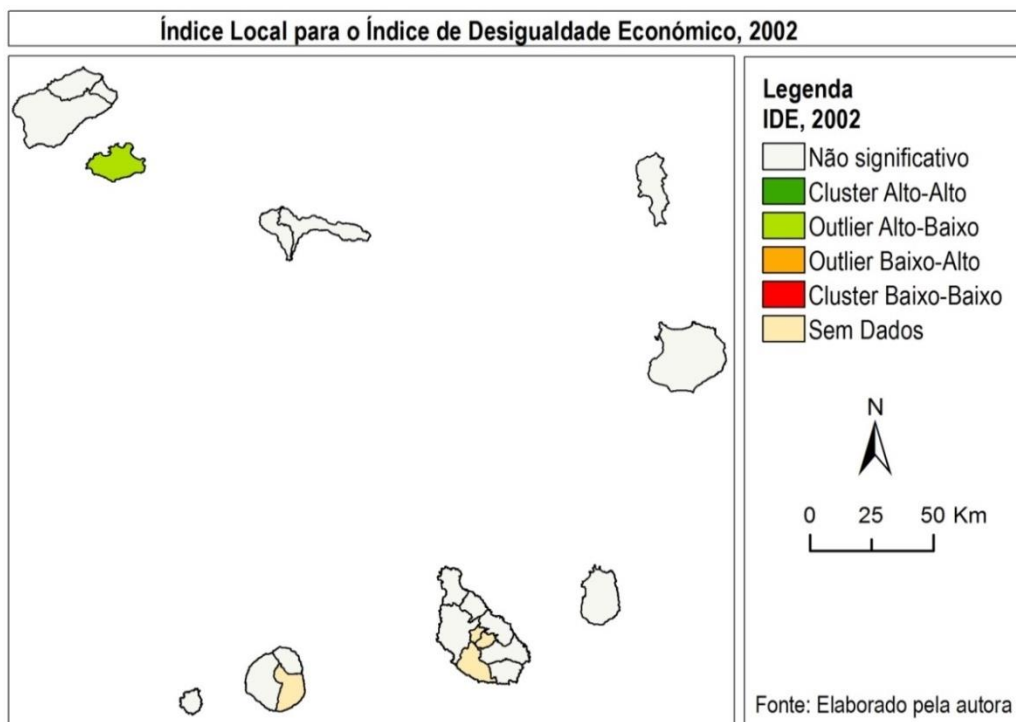
Nota-se ainda na Tabela IV.1, no índice Local de Morans considerando o índice de desigualdade de género, que os concelhos Ribeira Grande, Paúl, Tarrafal S.N., Sal, Boavista, Mosteiros, São Filipe, Santa Catarina do Fogo e Brava não apresentam índices significativos em nenhum dos anos. Em contrapartida, os concelhos Tarrafal, São Salvador do Mundo e São Lourenço dos Órgãos, apresentam índices estatisticamente significativos à 99% nos anos 2010 a 2015.

Para o índice de desigualdade ambiental o índice Local de Morans é significativo à 99%, em todos os anos em análise, para os concelhos São Salvador do Mundo e São Lourenço dos Órgãos (Tabela IV.2). O concelho Praia tem um índice significativo em todos os anos, porém, o nível de confiança nos anos 2000, 2010 e 2013 é de 90%. Para os concelhos Ribeira Grande, Paúl, Porto Novo, São Vicente, Tarrafal S.N., Ribeira Brava, Mosteiros, São Filipe, Santa Catarina do Fogo e Brava os índices não foram significativos em nenhum dos anos.

A partir do mapa de autocorrelação espacial que considerou os índices de desigualdade dos concelhos é possível observar a espacialização dos diferentes níveis de desigualdade do país. Nelas os concelhos classificados em LL têm índices baixos e são vizinhos de áreas com baixos índices, os classificados com HH têm índices altos assim como os seus vizinhos. Estas duas tipologias indicam que os concelhos são passíveis de serem agrupados. Já os concelhos com classificação HL possuem índices altos e seus vizinhos possuem índices diferentes, baixos, e os com LH apresenta o inverso (índices baixos e seus vizinhos com índices altos).

O índice local de Morans permite comparar o valor de cada concelho com seus vizinhos, indicando autocorrelação espacial local, ou seja, o nível de semelhança dos vizinhos. Uma aplicação prática de acção embasada no resultado obtido nesta análise é priorizar os *clusters* de concelhos de altos índices vizinhos de concelhos de altos índices como região prioritária de controlo da desigualdade. Na Figura 4.21 podemos visualizar o mapa do índice local de Morans do ano 2002 considerando o índice de desigualdade económico. Podemos observar que o concelho São Vicente é o único com significância estatística, porém, é um *outlier*, ou seja, não apresenta agrupamento. Este resultado vai ao encontro da análise de *outlier* clássica realizada anteriormente na secção 4.1.2, por outro lado, na análise espacial, ao contrário da análise clássica, o concelho do Maio não é estatisticamente significativa. Este facto pode dever-se-á ao nível elevado de desenvolvimento económico que a ilha de São Vicente apresenta, e que a ilha do Maio não possui.

Figura 4.21: Mapa do índice Local Morans ponderado pelo índice de desigualdade económico, 2002



Relativamente ao índice Local de Morans considerando o IDG, no ano 2000 o concelho São Vicente apresenta agrupamento estatisticamente significativo com índices altos, assim como os seus vizinhos e, o concelho de São Filipe apresenta *outlier* estatisticamente significativo, sendo os restantes concelhos não significativos (Figura 4.22). De acordo com os mapas, os *clusters* que apresentam valores baixos e estão próximos (LL) concentram-se a sul do país, precisamente na ilha de Santiago (Ver Anexo IV). Os concelhos mais a norte da ilha de Santiago, Tarrafal, São Salvador do Mundo e São Lourenço dos Órgãos apresentam esse comportamento em todos os anos em análise, isto é, o índice local de Morans para estes concelhos são positivos (Figura 4.23). O concelho mais populoso do país (Praia) apresenta o mesmo comportamento nos anos 2010 e 2015, não sendo significativo nos outros anos. O concelho Santa Catarina apresenta o mesmo resultado para o ano 2015, Ribeira Grande para o ano 2010 e, finalmente, o concelho São Miguel para os anos 2011, 2012 e 2013. No ano 2014, os concelhos Maio e Santa Cruz, apresentam exceções, indicadas pela tipologia HL que são valores do índice de desigualdade altas, porém, não se encontram agrupados, pois sua vizinhança apresenta valores baixos. O concelho Maio apresenta o mesmo comportamento nos anos 2013 e 2015.

É de referir que os restantes concelhos do país não foram apresentados neste mapa, e nem referidos, por não serem estatisticamente significativos, ou seja, não se enquadram nos agrupamentos, pois apresentam níveis de desigualdade variados assim como os valores dos concelhos vizinhos (Ver Anexo IV).

Figura 4.22: Mapa do índice Local Morans ponderado pelo índice de desigualdade de gênero, 2000

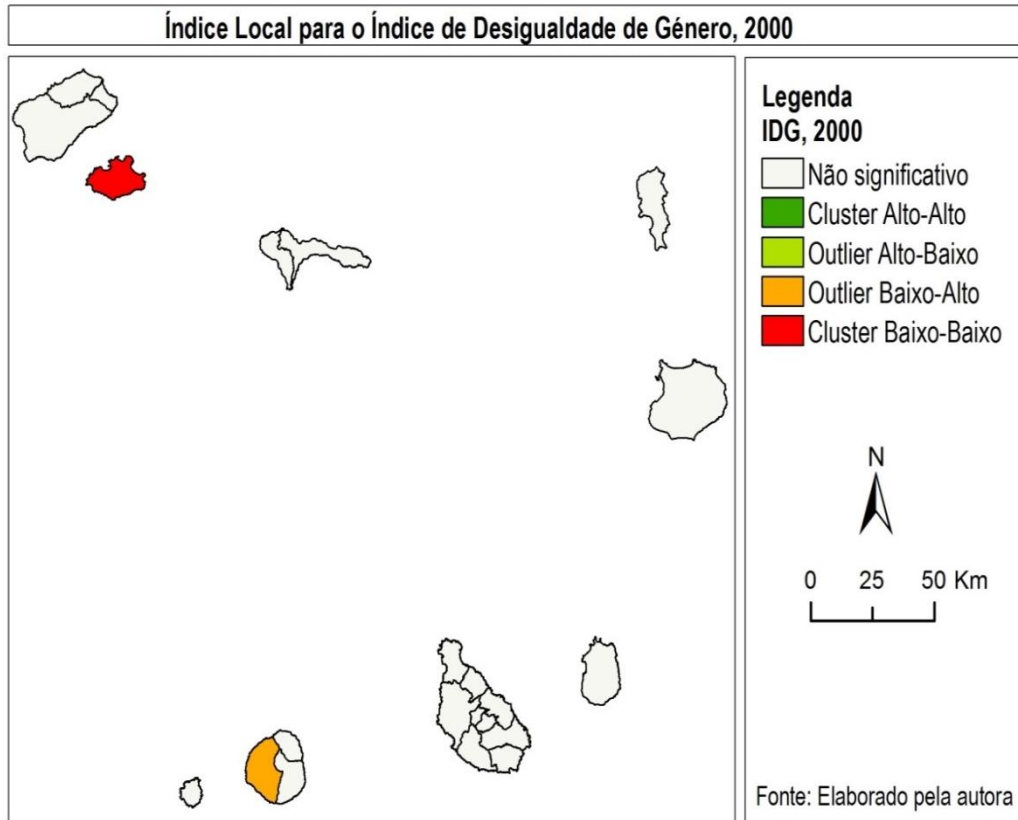
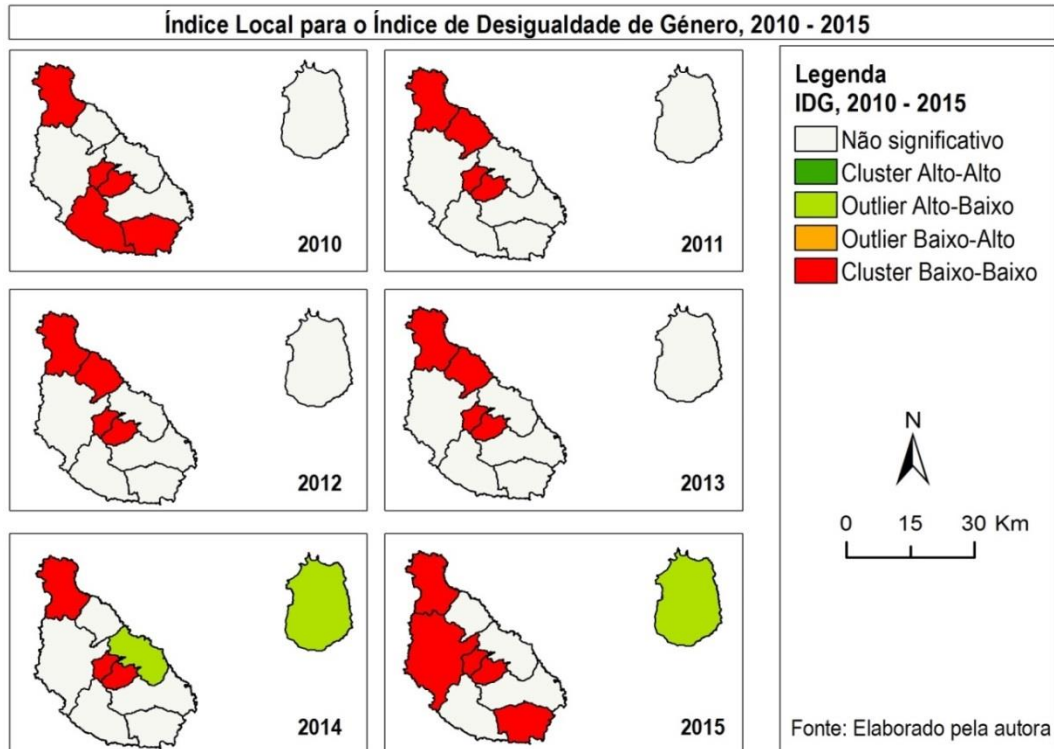
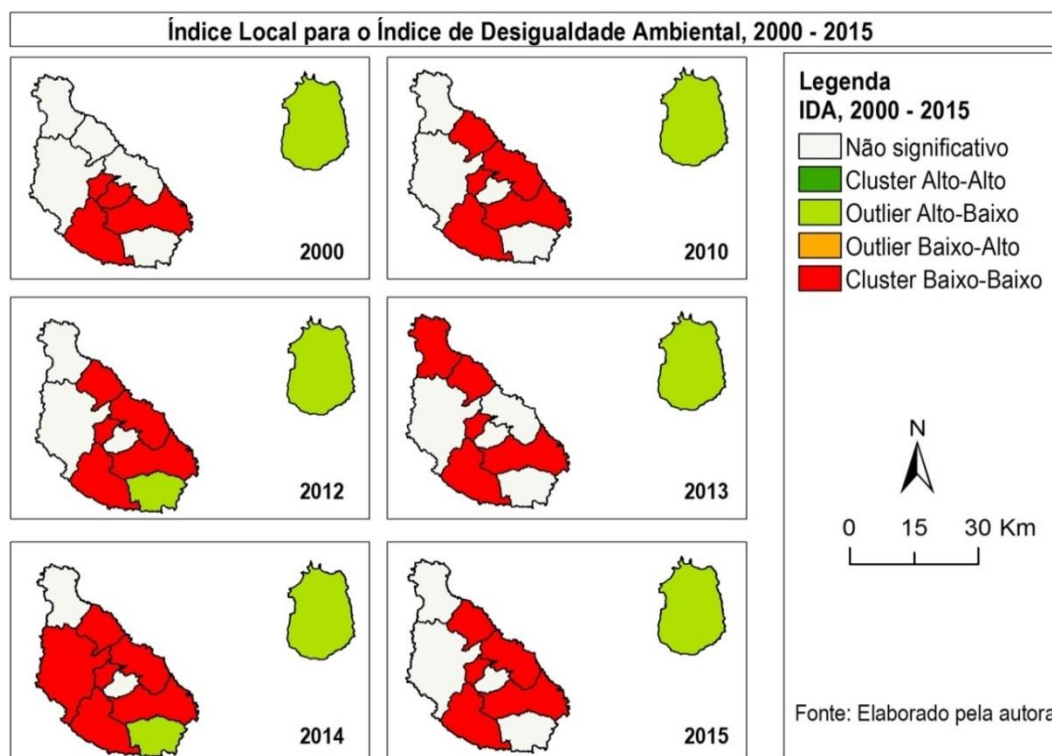


Figura 4.23: Mapa do índice Local Morans ponderado pelo índice de desigualdade de gênero, 2010 a 2015



A representação espacial mostrou que não foi possível obter agrupamentos de valores altos e próximos (HH) para a análise de autocorrelação espacial quando considerado o índice de desigualdade ambiental (Figura 4.24). Observa-se que no ano 2014, há maior número de concelhos com tipologia LL, isto é, maior número de concelhos apresentando agrupamento. Os concelhos Maio e Praia são os únicos *outliers* estatisticamente significativos, neste mesmo ano. Entretanto, o concelho Maio é um *outlier* para todos os anos em análise e o concelho Praia para o ano 2012 e 2014. É ainda de realçar que os concelhos São Domingos, São Salvador do Mundo e Ribeira Grande de Santiago apresentam valores dos índices baixos, assim como os seus vizinhos, em todos os anos. O que está de acordo com as características destes concelhos, pois, além de serem próximas e essencialmente rural, a actividade económica baseia-se essencialmente no tradicional sector da agricultura, pecuária e comércio de artesanato.

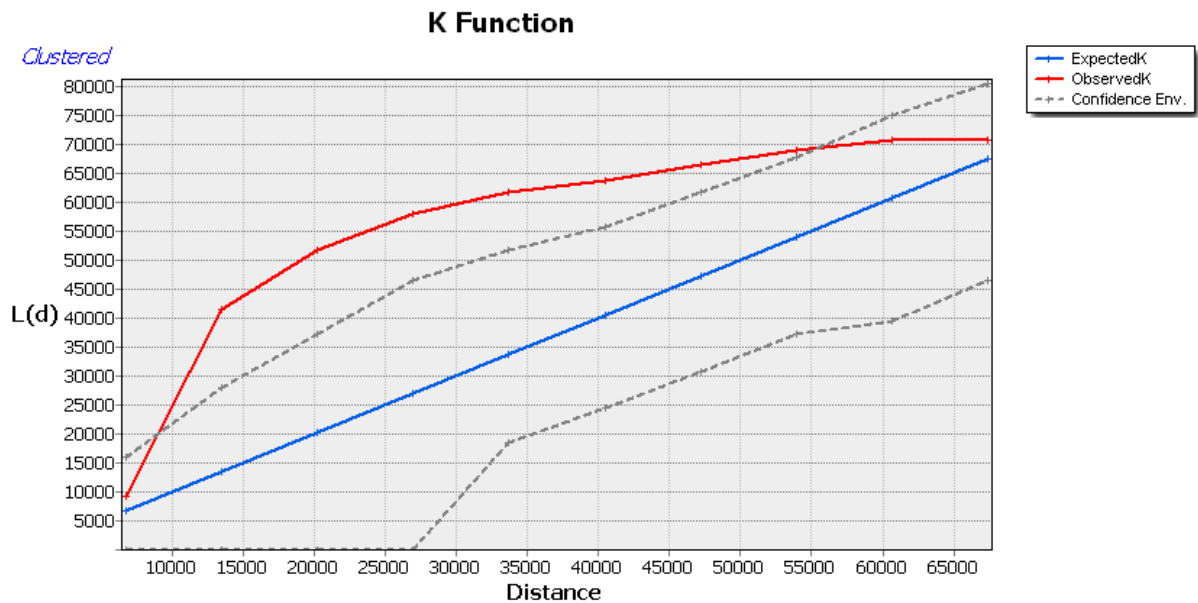
Figura 4.24: Mapa da autocorrelação espacial para o índice de desigualdade ambiental: índice Local de Morans



A presença de padrão de autocorrelação espacial na desigualdade ambiental e de género permite concluir sobre a não aleatoriedade dos dados. O índice vizinho mais próximo (*Nearest Neighbor Index*) calculado é de aproximadamente 0,73 (Figura IV.14). O valor-p do índice é estatisticamente significativo a um nível de significância de 5%, concluindo assim a não aleatoriedade dos dados. E, sendo o índice inferior a 1 e estatisticamente significativo conclui-se também que a distribuição observada é concentrada, ou seja, os dados apresentam agrupamento. Podemos

visualizar ainda o gráfico do método *K-Function* (Figura 4.25) permitindo verificar que os dados são concentrados, pois, a linha dos valores observados (linha vermelha) estão acima da linha dos valores esperados (linha azul).

Figura 4.25: Gráfico *K-Function*: Distribuição dos dados



Fonte: Elaborado pela autora

4.3. Análise *Hot Spot*

O primeiro aspecto a salientar nesta análise é a disposição de aglomerados dos *cold spot* (ponto frio, concelhos com desigualdades baixas assim como os seus vizinhos) somente nos concelhos da ilha de Santiago e do concelho Maio, ilhas a sul do país. Estes, com excepção do concelho de Santa Catarina, apresentam *cold spot* estatisticamente significativos a 95% de confiança em todos os anos em análise (Figura 4.26). O concelho de Santa Catarina apresenta *cold spot* estatisticamente significativo, a um nível de confiança de 90%, nos anos 2011, 2013 e 2014, 95% de confiança para os anos 2010 e 2015 e, não apresenta significância estatística para o ano 2012.

É de realçar que a análise de *Hot Spot* da desigualdade de género destes anos, não mostram existir *hot spot* (concelhos com índices elevados assim como os seus vizinhos) estatisticamente significativos. Já para a mesma análise ponderada pelo IDA, no ano 2000, os concelhos Terrafal S.N., Sal e Boavista são *hot spots* estatisticamente significativos a um nível de confiança de 90% (Figura 4.27). O concelho Terrafal S.N., também, é um *hot spot* estatisticamente significativo no ano 2010 (Figura V.8).

Figura 4.26: Mapa do *Hot spot* do IDG, 2010 a 2015

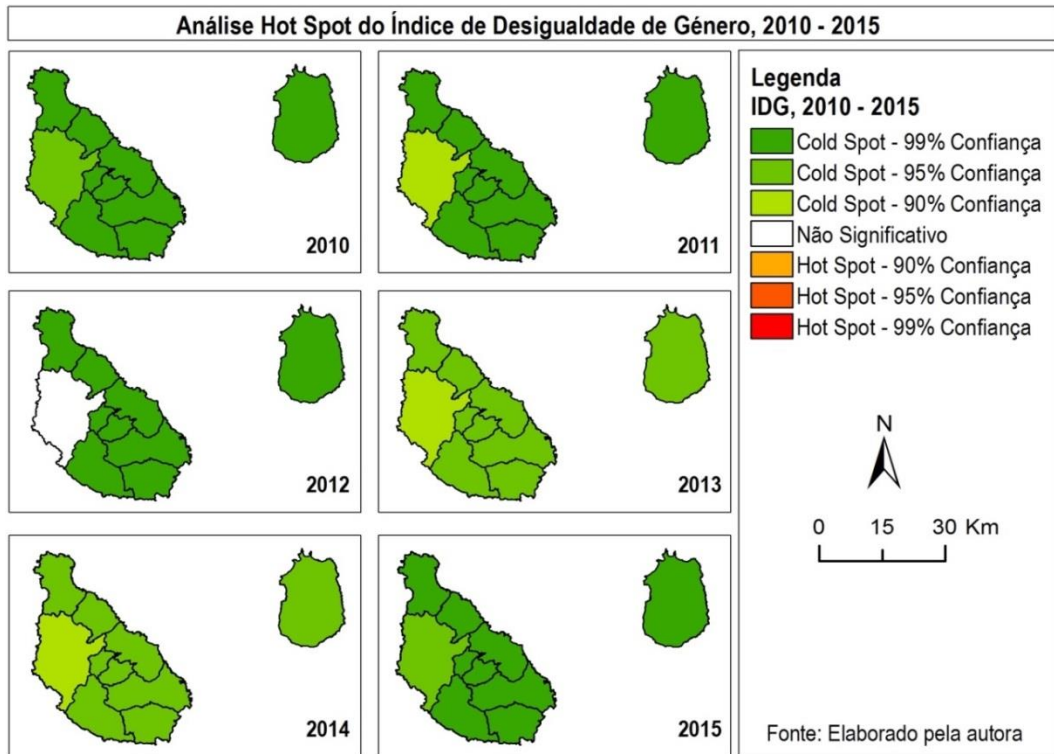
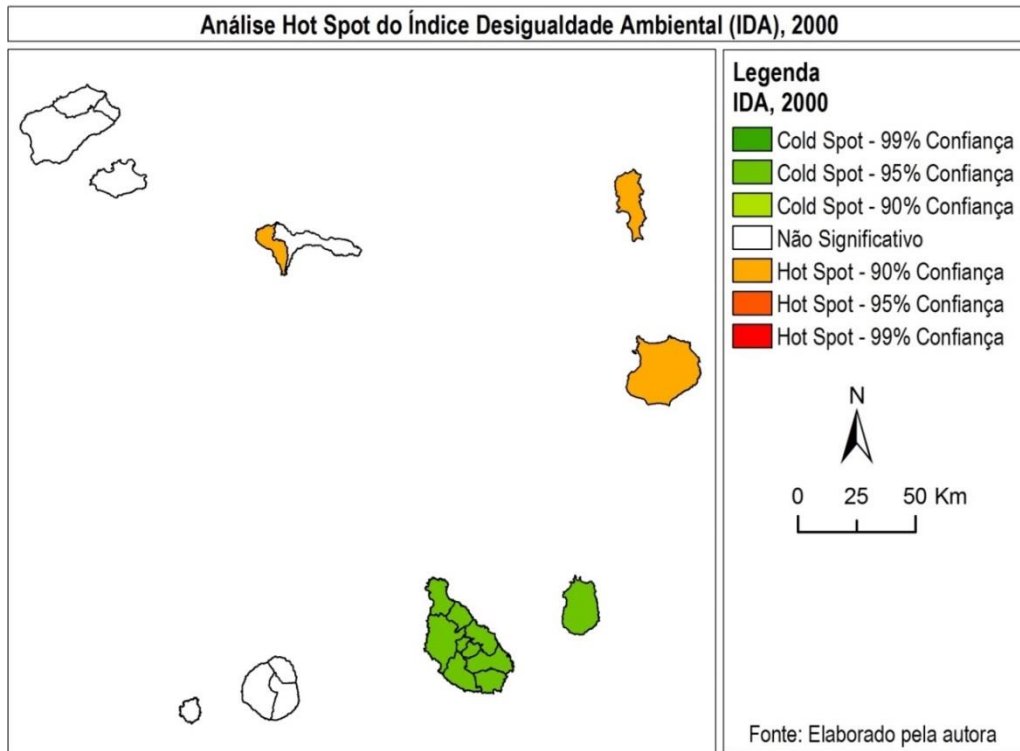


Figura 4.27: Mapa do *Hot spot* do IDA, 2000



Apesar dos dados do ano 2000 e 2010 apresentarem *hot spot* estatisticamente significativo, nos anos mais recentes, 2012 a 2015, não há concelhos com este comportamento. Este resultado pode ser justificado pela implementação de política na área ambiental, onde foi criado o Plano de Acção Nacional para o Ambiente II (2004-2014). Entretanto, até o ano 2011 a antiga ministra do ambiente, Sara Lopes, considerou os resíduos sólidos como o maior problema ambiental em Cabo Verde.

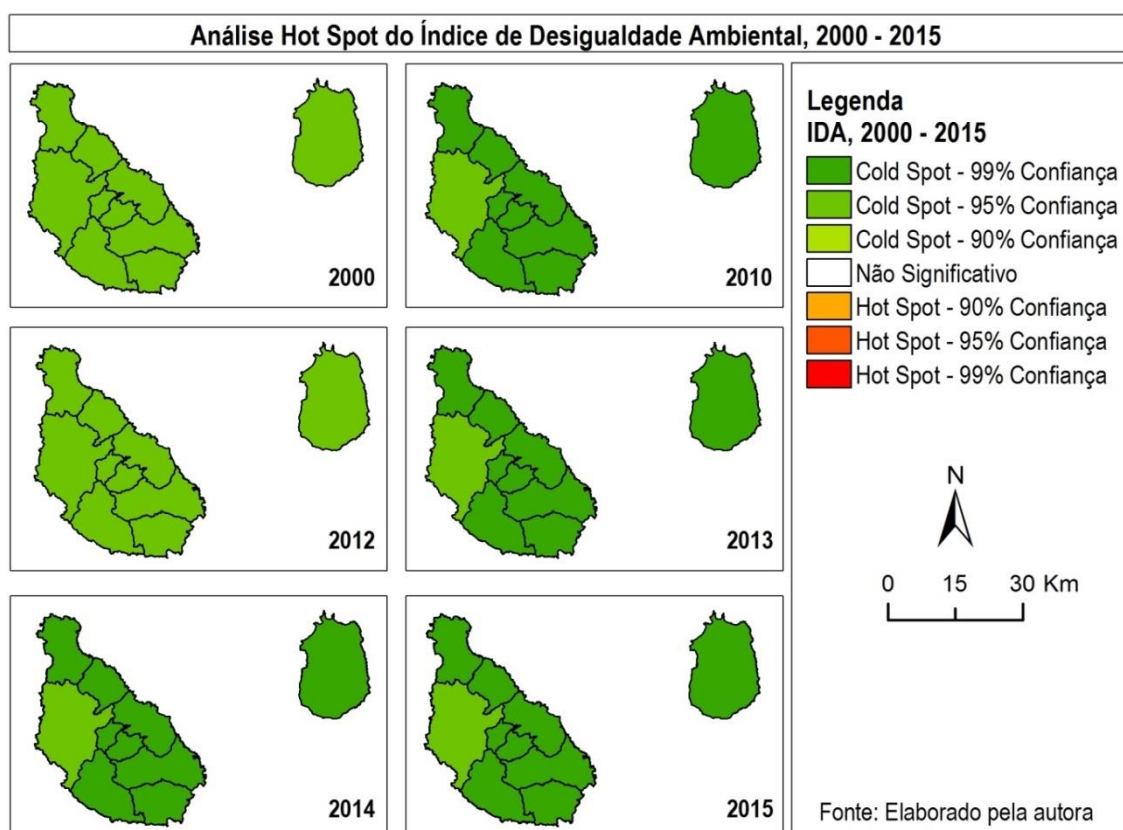
"Um dos piores problemas ambientais que temos em Cabo Verde é a forma como nos relacionamos com os resíduos sólidos e como facilmente abrimos a janela do carro e jogamos fora uma lata de cerveja. É uma atitude ambientalmente condenável",

disse, citada pela agência noticiosa cabo-verdiana Inforpress.

Isto, justifica o facto de alguns concelhos apresentarem *hot spot* estatisticamente significativo antes deste período.

Observa-se uma ligeira mudança nos concelhos *cold spot* estatisticamente significativos aquando o índice de desigualdade ambiental, passando de 90% de significância (ano 2012) para 99% (ano 2015), com excepção do concelho de Santa de Catarina que passa para 95% (Figura 4.28).

Figura 4.28: Mapa do *Hot spot* do IDA, 2000 a 2015



4.4. Análise de *Clusters* (ou de Agrupamento)

Para a realização da análise de *clusters* foram utilizados como variáveis os índices de desigualdade de género e o índice de desigualdade ambiental. Procurou-se criar grupos de concelhos com características homogéneas dentro de cada grupo e heterogéneas entre os grupos.

Analisando os dendrogramas verifica-se um lento agrupamento em todos os anos em análise (Figuras 4.29, 4.30, VI.1 – VI.4). O método de Mojena (1977) apresenta número de agrupamento menor quando comparado com os outros métodos determinados (*pseudo-F*, índice R^2 e RMSSTD). Por exemplo, no ano 2000, enquanto através do método de Mojena (1977) quatro grupos é o número óptimo a ser criado, para a estatística *pseudo-F* este é 13 grupos (Tabela 4.4).

O resultado do método de Mojena (1977) é apresentado na Tabela 4.3. Podemos verificar que com o passar dos anos os concelhos são mais homogéneos no que tange a desigualdade social, pois, o número de grupo ideal considerado pelo método Mojena (1977) diminui. Este facto pode ser explicado pelas políticas sociais e de investimentos adoptados em Cabo Verde nos últimos anos. A realização das grandes opções de desenvolvimento cujos objectivos se justapõem aos Objectivos do Desenvolvimento do Milénio (ODM) aprovados em 2000, precede o macro-histórico da adopção da cimeira mundial para o desenvolvimento. O percurso começa com a execução do plano nacional de desenvolvimento 1997-2000, conhece evoluções significativas com a execução do PND 2002-2005 e com os dois Documentos de Estratégia de Crescimento e de Redução da Pobreza (DECRP): DECRP I 2006-2007, DECRP II 2008-2011 e com o DECRP III 2012-2016.

Tabela 4.3: Ponto referencial de corte do dendrograma, método Mojena (1977)

Etapa (j)	2000	2010	2012	2013	2014	2015
2	0,0085	0,0055	0,0124	0,0045	0,0032	0,0061
3	0,0094	0,0056	0,0133	0,0057	0,0051*	0,0082*
4	0,0126*	0,0075*	0,0142	0,0067	0,0071*	0,0094*
5	0,0184*	0,0091*	0,0156	0,0101*	0,0080	0,0100
6	0,0220*	0,0166*	0,0169*	0,0130*	0,0087	0,0118*
7	0,0243*	0,0199*	0,0201*	0,0150*	0,0112*	0,0134*
8	0,0259	0,0226*	0,0220*	0,0162	0,0127*	0,0146*
9	0,0282*	0,0248*	0,0235*	0,0186*	0,0139*	0,0154*
10	0,0306*	0,0267*	0,0249*	0,0207*	0,0149*	0,0165*
11	0,0326*	0,0280	0,0259	0,0224*	0,0167*	0,0175*
12	0,0356*	0,0291	0,0271*	0,0238*	0,0186*	0,0182*
13	0,0401*	0,0300	0,0293*	0,0257*	0,0211*	0,0202*
14	0,0491*	0,0305	0,0312*	0,0278*	0,0254*	0,0223*
15	0,0558*	0,0319*	0,0327*	0,0299*	0,0289*	0,0239*
16	0,0615*	0,0354*	0,0345*	0,0332*	0,0319*	0,0266*
17	0,0661*	0,0415*	0,0372*	0,0387*	0,0361*	0,0300*
18	0,0701*	0,0466*	0,0403*	0,0445*	0,0420*	0,0343*
19	0,0784*	0,0537*	0,0491*	0,0569*	0,0481*	0,0411*
20	0,0966*	0,0651*	0,0674*	0,0692*	0,0577*	0,0506*
21	0,1176*	0,0893*	0,0843*	0,0893*	0,0779*	0,0714*

* $\alpha_j > \theta_k$

Fonte: Elaborado pela autora

A penúltima alinha da Tabela 4.3, apresenta o ponto referencial de corte do dendrograma para determinar o número óptimo de grupo. Verifica-se que o número óptimo de grupos dado pelo método de Mojena, com excepção do ano 2014, é semelhante em todos os anos em análise (Tabela 4.4). Por outro lado, pela estatística *pseudo-F*, o número óptimo de grupo considerado oscila entre 13 e 15 grupos. Podemos verificar ainda nesta tabela, que os quatro métodos apresentam número óptimo de grupos muito distintos.

Um dos objectivos da análise de *clusters* é criar grupos com características homogéneas dentro de cada *cluster*, mas também com características heterogéneas entre eles. Assim, apesar de quando maior o número de grupo mais homogéneas elas são, considerou-se identificar *clusters* a ser criado segundo o método de Mojena e o índice R^2 , pelo facto destes dois métodos apresentarem resultados mais próximos do que os outros métodos (índice RMSSTD e estatística *pseudo-F*) e, porque os *clusters* a serem produzidos, tendo em conta os outros dois métodos, é muito grande quando comparado ao número de concelhos a classificar.

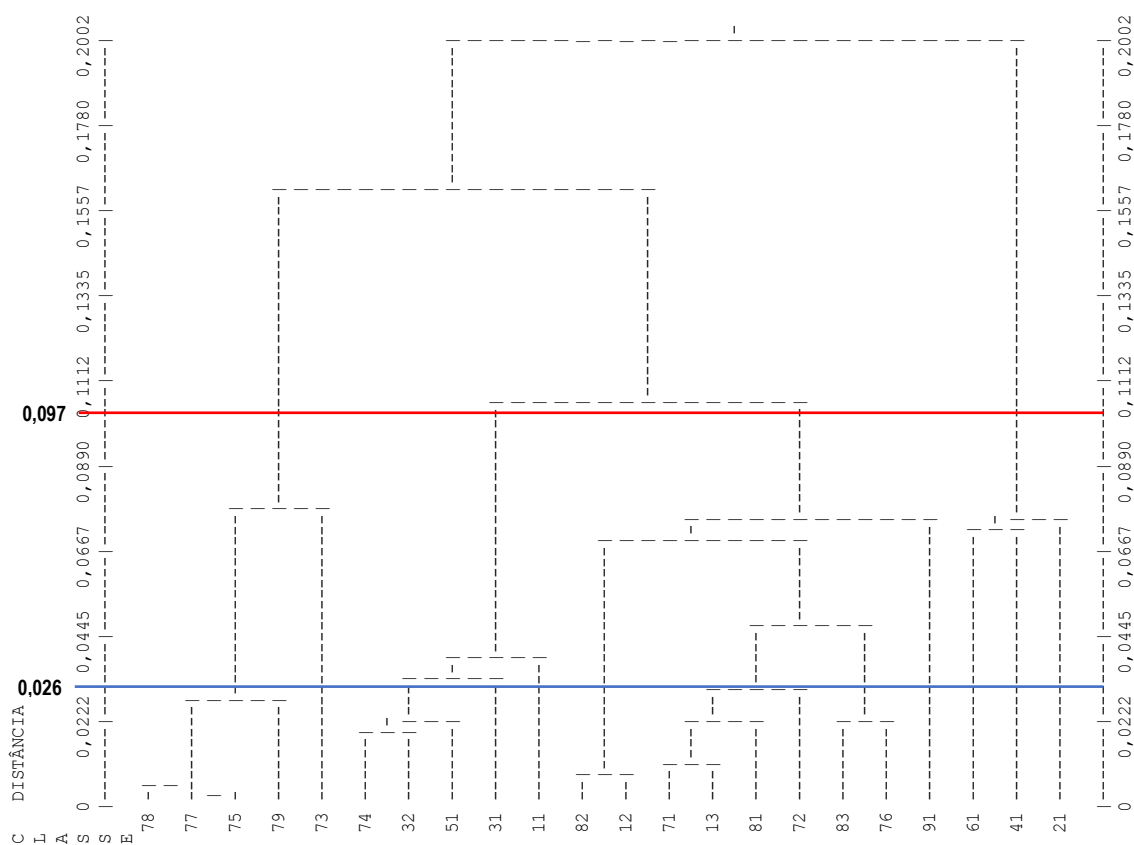
Tabela 4.4: Corte referencial do dendrograma e número ideal de grupos para cada método, 2000 – 2015

Métodos	2000	2010	2012	2013	2014	2015
R ²	0,996	0,995	0,998	0,999	0,999	0,999
Número de grupo	7	5	6	7	6	7
<i>Pseudo-F</i>	24,064	59,723	41,392(a)	78,450	102,326	114,737
Número de grupo	13	15	14	15	15	14
Métodos	Ponto de corte dendrograma					
Mojena (1977)	0,097	0,065	0,067	0,069	0,058	0,051
Número de grupo	4	4	4	4	5	4
RMSSTD	0,026	0,017	0,014	0,017	0,013	0,011
Número de grupo	13	18	21	17	16	21

Fonte: Elaborado pela autora

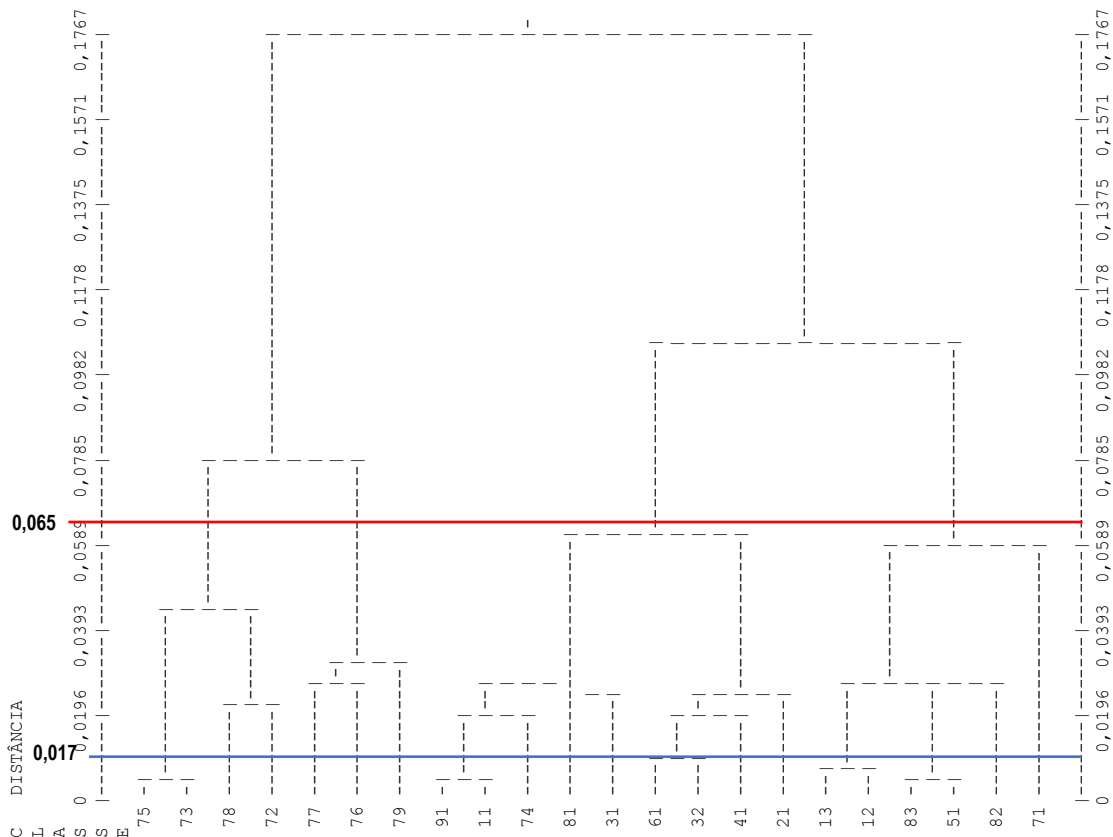
Através dos dendrogramas (Figuras 4.29 4.30 VI.1 – VI.4) podemos visualizar os agrupamentos dos concelhos tendo em conta o método de Mojena (linha cor vermelha) e o índice RMSSTD (linha cor azul).

Figura 4.29: Dendrograma da análise de *clusters* dos concelhos, 2000



Fonte: Elaborado pela autora

Figura 4.30: Dendrograma da análise de *clusters* dos concelhos, 2010



Fonte: Elaborado pela autora

4.4.1. Identificação de soluções de *clusters*: Método índice R^2

O método índice R^2 caracteriza-se como sendo um método de dissimilaridade, pois, quanto maior é o valor do índice R^2 , maior é a distância entre os grupos. Como foi visto anteriormente, na Tabela 4.4, o coeficiente do R^2 foi muito elevado, aproximando de um, ou seja, os grupos são muito distintos.

No ano 2000 os *clusters* formados possuem as seguintes características

- *Cluster* 1: formado por 4,5% da amostra (1 caso), com valores de desigualdades (IDA e IDG) médias;
- *Cluster* 2: formado por 9,1% da amostra (2 casos), valores de IDG elevados e IDA muito baixos;
- *Cluster* 3: formado por 13,6% da amostra (3 casos), valores IDG Médios e IDA muito baixos;

- *Cluster 4*: formado por 18,2% da amostra (4 casos), valores IDG médios e IDA baixos;
- *Cluster 5*: formado por 22,7% da amostra (5 casos), valores IDG médios e IDA muito baixos;
- *Cluster 6*: também formado por 22,7% da amostra (5 casos), valores IDG médios e IDA baixos;
- *Cluster 7*: formado por 9,1% (2 casos), valores IDG médios e IDA muito baixos.

No ano 2010 os *clusters* passam a ter as seguintes características

- *Cluster 1*: formado por 9,1% da amostra (2 casos), valores IDG Médios e IDA elevados;
- *Clusters 2*: formado por 13,6% da amostra (3 casos), valores IDG médios assim como o IDA;
- *Cluster 3*: formado por 27,3% da amostra (6 casos), valores IDG e IDA médios;
- *Cluster 4*: formado por 22,7% da amostra (5 casos), valores IDG médios e IDA baixos;
- *Cluster 5*: formado por 27,3% da amostra (6 casos), também, com valores IDG e IDA médios.

No ano 2012 os *clusters* têm as seguintes características

- *Cluster 1*: formado por 22,7% da amostra (5 casos), valores IDG médios e IDA elevados;
- *Clusters 2*: formado por 9,1% da amostra (2 casos), valores IDG baixos e IDA médios
- *Cluster 3*: formado por 13,6% da amostra (3 casos), valores IDG médios e IDA elevados;
- *Clusters 4*: formado por 13,6% da amostra (3 casos), valores de IDA e IDG médios;
- *Cluster 5*: formado por 18,2% da amostra (4 casos), valores de IDA e IDG médios;
- *Cluster 6*: formado por 22,7% da amostra (5 casos), valores IDG médios e IDA elevados.

No ano 2013 os *clusters* formados têm as seguintes características

- *Cluster 1*: formado por 13,6% da amostra (3 casos), valores IDA elevados e IDG médios;
- *Cluster 2*: formado por 18,2% da amostra (4 casos), valores IDA e IDG médios;
- *Cluster 3*: formado por 4,5% da amostra (1 caso), valor IDA e IDG médios;
- *Cluster 4*: formado por 27,3% da amostra (6 casos), valores IDA elevados e IDG médios;
- *Cluster 5*: formado por 9,1% da amostra (2 casos), valores IDA e IDG médios;
- *Cluster 6*: formado por 4,5% da amostra (1 caso), valor IDA e IDG médios;
- *Cluster 7*: formado por 22,7% da amostra (5 casos), valores IG médios e IDA elevados.

No ano 2014 as características dos *clusters* são as seguintes:

- *Cluster 1*: formado por 22,7% da amostra (5 casos), valores IDA e IDG médios;
- *Cluster 2*: formado por 18,2% da amostra (4 casos), valores IDA e IDG médios;
- *Cluster 3*: formado por 18,2% da amostra (4 casos), valores IDA e IDG médios;
- *Cluster 4*: formado por 27,3% da amostra (6 casos), valores IDA variando de médios a elevados e IDG médios;
- *Cluster 5*: formado por 4,5% da amostra (1 caso), valores IDA e IDG médios;
- *Cluster 6*: formado por 9,1% da amostra (2 casos), valores IDA e IDG médios.

E, finalmente, no ano 2015 os *clusters* têm as seguintes características

- *Cluster 1*: formado por 9,1% da amostra (2 casos), valores IDG e IDA médios;
- *Cluster 2*: formado por, também, por 9,1% da amostra (2 casos), com valores de IDA e IDG muito baixo;
- *Cluster 3*: formado por 13,6% da amostra (3 casos), com valores de IDA e IDG médios;
- *Cluster 4*: formado por, também, por 13,6% da amostra (3 casos), com valores de IDA e IDG médios;
- *Cluster 5*: formado por 22,7% da amostra (5 casos), com valores de IDA elevados e IDG médios;
- *Cluster 6*: formado por 13,6% da amostra (3 casos), com valores IDA muito baixos e IDG baixos;

- *Cluster 7*: formado por 18,2% da amostra com valores de IDA e IDG médios.

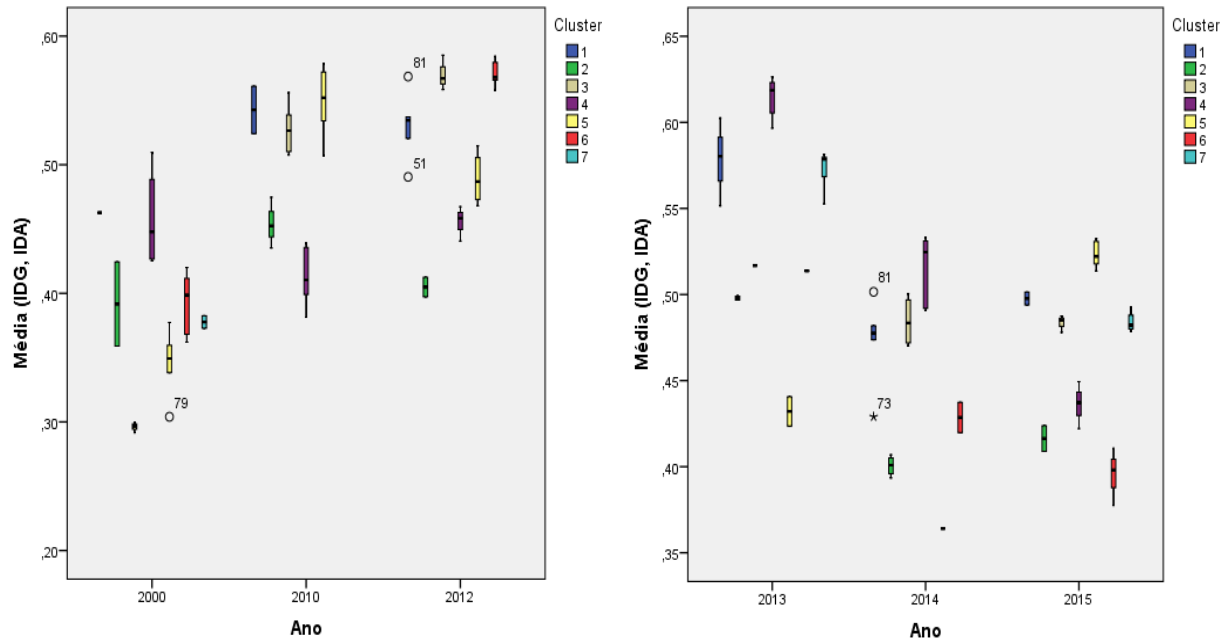
Os *clusters* formados no ano 2000, usando o número ótimo de grupo dado pelo método índice R^2 , mostram semelhanças em alguns *clusters*. Os *clusters* 3, 5 e 7 são constituídos por valores de IDA muito baixos e IDG médios, e *clusters* 4 e 6 constituídos por valores de IDG médios e IDA baixos. Portanto, estes *clusters* podem ser representados por dois *clusters*, um formado pelos *clusters* 3, 5 e 7 e outro pelos *clusters* 4 e 6. Neste sentido, o número de *clusters* para este ano podia ser 4. No ano 2010 os *clusters* 2, 3 e 5 apresentam características de IDA e IDG semelhantes, ou seja, valores de IDA e IDG médios. Sendo assim, a formação de 3 grupos também seria uma solução plausível. No ano 2012, os *clusters* 1, 3 e 6 apresentam valores de IDA e IDG semelhantes (IDG médios e IDA elevados), podendo formar um único grupo. Os *clusters* 4 e 5 também apresentam semelhanças no IDA e IDG, com valores médios para IDG assim como IDA, podendo formar também um grupo. Assim, no ano 2012 os grupos a formar pode ser 3. No ano 2013 o número de grupo a se formar também pode ser de 3, pois, os *clusters* 1, 3 e 6 apresentam as mesmas características e *clusters* 4 e 5, também. Já no ano 2014, o número de grupo a se formar pode ser 5, pois, dos 6 *clusters* formados existem dois *clusters* com características semelhantes. E, por último, o ano 2015, formado por 7 *clusters*, apresenta semelhanças nos valores do IDA e IDG entre os *clusters* 1, 2, 3, 4 e 7. Portanto, a solução dada para este ano passa de 7 para 3 *clusters*.

Então, verifica-se que a solução do número de *clusters*, dado para cada ano, aproxima do número de *clusters* dado pelo método de Mojena (1977). No Gráfico 4.1, podemos observar o *Box-plot* de cada grupo nos diferentes anos. Verifica-se que alguns dos grupos formados não são homogêneos, como é o caso do grupo 2 e 3 do ano 2000. Ainda, nota-se que alguns grupos apresentam valores extremos, como é o caso do grupo 5 no ano 2000, e grupo 1 no ano 2012 e 2014. Estes resultados vêm contrariar a solução do número de *clusters* dado anteriormente, e mostra que os *clusters* criados não são muito homogêneos, mesmo que os grupos entre si são heterogêneos.

Esta diferença pode ser justificada pelo critério adotado neste trabalho de criar um número pequeno de *cluster*, tendo em conta o número de concelhos que se pretende agrupar. Pois, isto levaria a ter concelhos com características muito distintas a fazerem parte do mesmo *cluster*. Por exemplo, temos os concelhos Praia e São Vicente agrupados com outros concelhos, que não apresentam características semelhantes. No entanto, levando em consideração que os indicadores utilizados para agrupar os concelhos foram o IDA e o IDG, os restantes concelhos podem

apresentar características semelhantes, pois, como foi vista na secção 4.1.3 estas desigualdades abrangem todos os concelhos da Ilha de Santo Antão, Santiago e Fogo, o que, inferencialmente, se pode dizer que acontece em todos os concelhos do país.

Gráfico 4.1: *Box-plot* dos *clusters*, determinado pelo método índice R^2 , 2000 a 2015



Fonte: Elaborado pela autora

4.4.2. Identificação de soluções de *clusters*: método Mojena (1977)

O método de Mojena (1977), de entre os métodos aqui apresentados, é o método que apresenta menor *cluster* a ser produzido. No ano 2000, 2010, 2012, 2013 e 2015 este método apresenta uma solução de 4 *clusters*, e no ano 2014 apresenta uma solução de 5 *cluster*.

No ano 2000 os 4 *clusters* apresentam as seguintes características:

- *Cluster* 1: constituído por 27,3% da amostra (6 casos), com valores de IDG baixos e de IDA médios;
- *Cluster* 2: constituído por 18,2% da amostra (4 casos), com valores elevados do IDG e muito baixos do IDA;
- *Cluster* 3: constituído por 36,4% da amostra (8 casos), com valores de IDG médios e IDA muito baixos;
- *Cluster* 4: constituído por 18,2% da amostra (5 casos), com valores elevados do IDG e IDA baixos.

No ano 2010 as características dos *clusters* são as seguintes:

- *Cluster 1*: constituído por 36,4% da amostra (8 casos), com valores de IDG médios e de IDA elevados;
- *Cluster 2*: constituído por 13,6% da amostra (3 casos), com valores do IDG e IDA médios;
- *Cluster 3*: constituído por 27,3% da amostra (6 casos), com valores do IDG e IDA médios;
- *Cluster 4*: constituído por 22,7% da amostra (5 casos), com valores do IDG médios e IDA baixos.

Em 2012 os *clusters* passam a ter as seguintes características:

- *Cluster 1*: constituído por 22,7 % da amostra (5 casos), com valores de IDG médios e de IDA elevados;
- *Cluster 2*: constituído por 9,1% da amostra (2 casos), com valores do IDG baixos e IDA médios;
- *Cluster 3*: constituído por 36,4% da amostra (8 casos), com valores do IDG médios e IDA elevados;
- *Cluster 4*: constituído por 31,8% da amostra (7 casos), com valores do IDG e IDA médios.

No ano 2013 as características dos *clusters* são:

- *Cluster 1*: constituído por 13,6% da amostra (3 casos), com valores de IDG médios e de IDA elevados;
- *Cluster 2*: constituído por 31,8% da amostra (7 casos), com valores do IDG e IDA médios;
- *Cluster 3*: constituído por 27,3% da amostra (6 casos), com valores do IDG médios e IDA elevados;
- *Cluster 4*: constituído por 27,3% da amostra (6 casos), também, com valores do IDG médios e IDA elevados.

No ano 2014 os *clusters* têm as seguintes características:

- *Cluster 1*: constituído por 22,7% da amostra (5 casos), com valores de IDG e IDA médios;

- *Cluster 2*: constituído por 27,3% da amostra (6 casos), com valores do IDG e IDA médios;
- *Cluster 3*: constituído por 18,2% da amostra (4 casos), com valores de IDG e IDA médios;
- *Cluster 4*: constituído por 27,3% da amostra (6 casos), também, com valores do IDG médios e IDA elevados;
- *Cluster 5*: constituído por 4,5% da amostra (1 caso), com valores de IDA e IDG baixos.

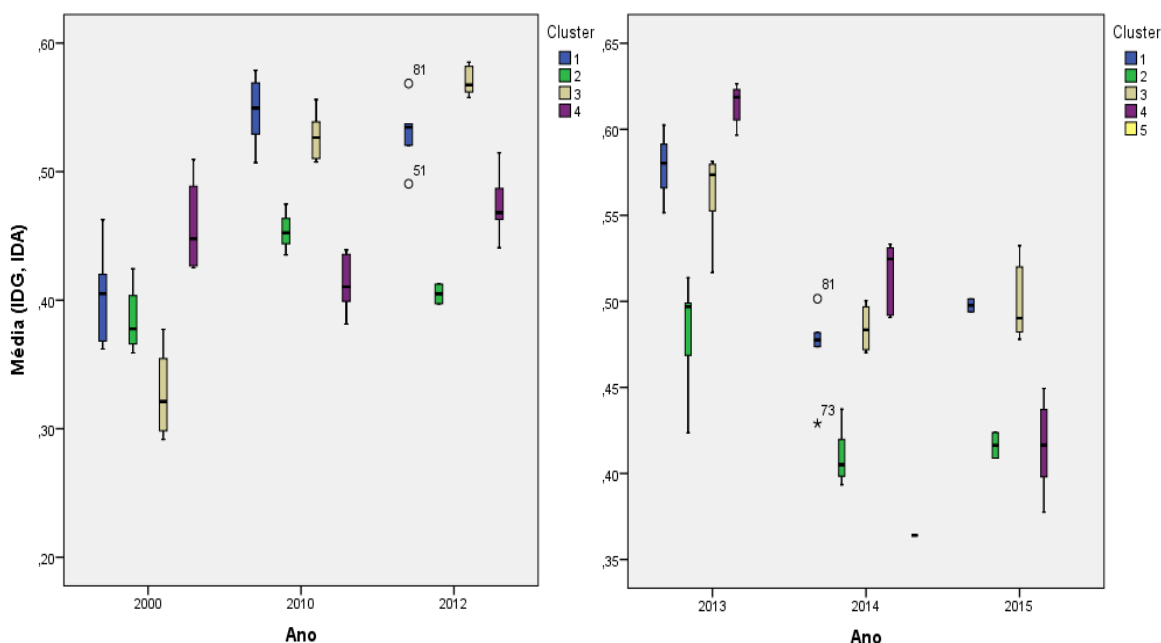
E, em 2015 os clusters apresentam as seguintes características:

- *Cluster 1*: constituído por 9,1% da amostra (2 casos), com valores de IDG médios e IDA elevados;
- *Cluster 2*: constituído por 9,1% da amostra (2 casos), com valores do IDG e IDA médios;
- *Cluster 3*: constituído por 54,5% da amostra (12 casos), com valores de IDG e IDA médios;
- *Cluster 4*: constituído por 27,3% da amostra (6 casos), também, com valores de IDG e IDA médios;

O mesmo que aconteceu com os *clusters* formados pelo método do índice R^2 acontece, também, com os *clusters* formados pelo método de Mojena (1977). Isto é, existem *clusters* com valores de IDA e IDG semelhantes, excepto no ano 2000. No ano 2010, os *clusters* 2 e 3 apresentam características de IDA e IDG semelhantes, no ano 2012 os *clusters* 1 e 3, no ano 2013 os *clusters* 1, 3 e 4, no ano 2014 os *clusters* 1, 2 e 3, e no ano 2015 os *clusters* 2, 3 e 4. Portanto, uma solução aceitável passa para 3 *clusters* no ano 2010, assim como nos anos 2012 e 2014, e 2 *clusters* para os anos 2013 e 2015.

Neste sentido, verifica-se que os agrupamentos podem ser ainda menores dos que apresentados nos métodos aqui adoptados. Portanto, dos métodos apresentados para este estudo, o método Mojena (1977) mostrou ser mais adequado. Entretanto, como foi verificada através da análise de *Box-plot* dos agrupamentos determinado pelo método índice R^2 , o agrupamento deveria ser maior para que os grupos sejam homogéneos. Assim, observando o Gráfico 4.2 pode-se chegar a mesma conclusão, pois, com a criação de grupos ainda menores (determinado pelo método Mojena, 1977) apresentam caixa maiores, ou seja, grupos menos homogéneos.

Gráfico 4.2: Box-plot dos *clusters*, determinado pelo método índice R^2 , 2000 a 2015



Fonte: Elaborado pela autora

A Tabela 4.5 mostra os resultados do teste F da ANOVA e teste de Levene, que foi aplicada com a finalidade de verificar a diferença entre e dentro dos grupos, utilizando os dois métodos, Mojena (1977) e índice R^2 . Pode-se verificar, que os grupos determinados pelo método índice R^2 são, ligeiramente, mais homogêneos quando comparado com o método Mojena (1977), pois, a soma dos quadrados (SQ) dentro dos grupos são menores, com exceção do ano 2010, em que a SQ dentro dos grupos são iguais para os dois métodos, sendo que o método de Mojena (1977) determina 4 *clusters* e o índice R^2 determina 5 *clusters*. Os testes F da ANOVA apresentam significância inferior a 1%, o que demonstra existência de diferença entre os grupos.

Tabela 4.5: Teste de Levene, Soma dos Quadrados entre e dentro dos *clusters* e teste F da ANOVA segundo os métodos Mojena e índice R²

Grupo	2000	2010	2012	2013	2014	2015
Mojena (1977)						
Teste Levene (sig.)	0,366 (0,778)	0,921 (0,451)	1,285 (0,310)	2,233 (0,119)	0,302 (0,824) ^a	2,687 (0,077)
SQ Entre	0,050	0,067	0,062	0,061	0,046	0,036
SQ Dentro	0,022	0,009	0,008	0,012	0,007	0,009
Teste F (sig.)	13,853 (0,000)	42,734 (0,000)	47,723 (0,000)	29,970 (0,000)	28,929 (0,000)	25,171 (0,000)
índice R²						
Teste Levene (sig.)	2,770 (0,058) ^a	0,750 (0,571)	0,906 (0,501)	2,621 (0,077) ^a	1,180 (0,357) ^a	1,284 (0,322)
SQ Entre	0,059	0,067	0,064	0,071	0,047	0,043
SQ Dentro	0,013	0,009	0,006	0,003	0,006	0,001
Teste F (sig.)	11,655 (0,000)	30,524 (0,000)	35,052 (0,000)	65,321 (0,000)	26,386 (0,000)	71,894 (0,000)

a. Os *clusters* com apenas com um caso são ignorados no cálculo do teste de homogeneidade

Fonte: Elaborado pela autora

Sendo os resultados dos dois métodos não muito diferentes, optou-se por apresentar os resultados dos *clusters*, especialmente, utilizando o método de Mojena (1977), pois, apresenta menos agrupamento quando comparado ao método índice R².

4.4.3. Resultados: método Mojena (1977)

No ano 2000, os concelhos Ribeira Grande, Paúl, São Vicente, Tarrafal S.N., Praia e São Filipe fazem parte do mesmo *cluster*, ou seja, apresentam características de IDA e IDG semelhantes (Figura 4.31). Os concelhos Porto Novo, Santa Cruz, Mosteiros e Brava se agrupam, apresentando semelhanças nas desigualdades. Neste mesmo ano ainda existem mais dois agrupamentos, um dado pelos concelhos Tarrafal, Santa Catarina, São Domingos, São Miguel, São Salvador do Mundo, São Lourenço dos Órgãos, Ribeira Grande de Santiago e Santa Catarina do Fogo, e outro dado pelos concelhos Ribeira Brava, Sal, Boavista e Maio. No ano 2010, também com 4 *clusters*, apresentam agrupamento os concelhos Ribeira Grande, São Vicente, Tarrafal S. N., Sal, Maio, Praia, São Filipe e Brava (Figura 4.32). Um outro agrupamento é composto pelos concelhos Tarrafal, Santa Catarina e São Lourenço dos Órgãos. Os concelhos Paúl, Porto Novo, Ribeira Brava, Boavista, Mosteiros e Santa Catarina do Fogo, também fazem parte do mesmo *cluster*, sendo os concelhos Santa Cruz, São Domingos, São Miguel, São Salvador do Mundo e Ribeira Grande de Santiago, um outro *cluster*. Em 2012, assim como nos anos 2000 e 2010, os concelhos Ribeira Grande, São Vicente, Tarrafal S. N. e Praia fazem parte do mesmo *cluster*, juntamente com os concelhos Ribeira Brava, Sal, Maio e Brava (Figura 4.33). Neste mesmo ano os concelhos Paúl, Porto Novo, Sal, Mosteiros e Santa Catarina do Fogo se agrupam. Os concelhos São Miguel e São

Salvador do Mundo constituem um *cluster* e, finalmente, os concelhos Tarrafal, Santa Catarina, Santa Cruz, São Domingos, São Lourenço dos Órgãos, Ribeira Grande de Santiago e São Filipe fazem parte do mesmo *cluster*. Em relação ao ano 2013, assim como nos anos anteriores descritos, foi plausíveis quatro agrupamentos (Figura 4.34). Sendo um pelos concelhos São Vicente, Praia e São Filipe, outro pelos concelhos Tarrafal, Santa Catarina, São Domingos, São Miguel, São Salvador do Mundo, São Lourenço dos Órgãos e Ribeira Grande de Santiago, outro pelos concelhos Ribeira Grande, Paúl, Porto Novo, Boavista, Santa Cruz e Santa Catarina do Fogo, e por último, um outro *cluster* constituído pelos concelhos Ribeira Brava, Tarrafal S. N., Sal, Maio, Mosteiros e Brava. Em 2014 o número de *clusters* pelo método de Mojena (1977) foi maior, quando comparado com os outros anos, sendo considerado cinco agrupamentos. Um agrupamento é composto pelos concelhos Paúl, Porto Novo, Boavista, Santa Cruz e Mosteiros, um outro agrupamento é constituído pelos concelhos São Vicente, Praia, São Filipe e Santa Catarina do Fogo (Figura 4.35). Um outro agrupamento é composto pelos concelhos da ilha de Santiago, Tarrafal, Santa Catarina, São Domingos, São Miguel, São Lourenço dos Órgãos e Ribeira Grande de Santiago. Um outro agrupamento é constituído pelos concelhos Ribeira Grande, Ribeira Brava, Tarrafal S. N., Sal, Maio e Brava e, finalmente um *cluster* constituído por um único concelho, o concelho São Salvador do Mundo. Em 2015, o primeiro *cluster* e o segundo *clusters* são constituído por apenas dois concelhos, sendo o primeiro composto pelos concelhos São Vicente e Praia e, o segundo pelos concelhos Santa Cruz e São Miguel (Figura 4.36). Estes dois últimos concelhos partilham as mesmas fronteiras. Um outro agrupamento é constituído pelos concelhos Ribeira Grande, Paúl, Porto Novo, Ribeira Brava, Tarrafal S. N., Sal, Boavista, Maio, Mosteiros São Filipe, Santa Catarina do Fogo e Brava e, um outro agrupamento é composto pelos concelhos Tarrafal, Santa Catarina, São Domingos, São Salvador do Mundo, São Lourenço dos Órgãos e Ribeira Grande de Santiago.

Figura 4.31: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2000

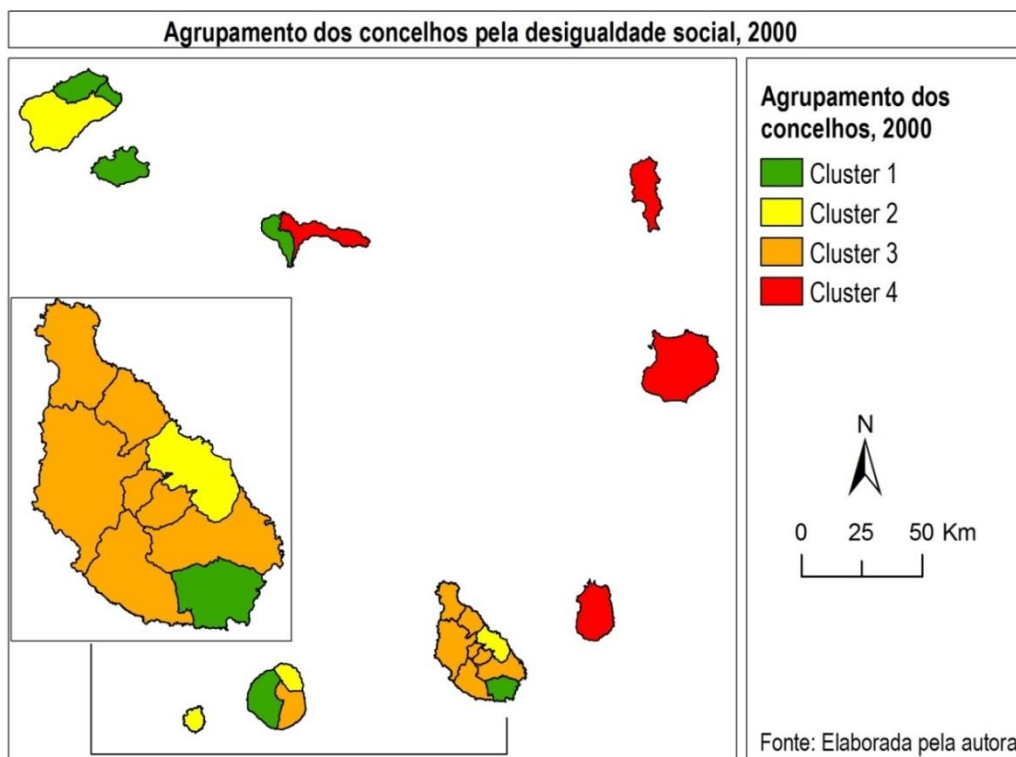


Figura 4.32: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2010

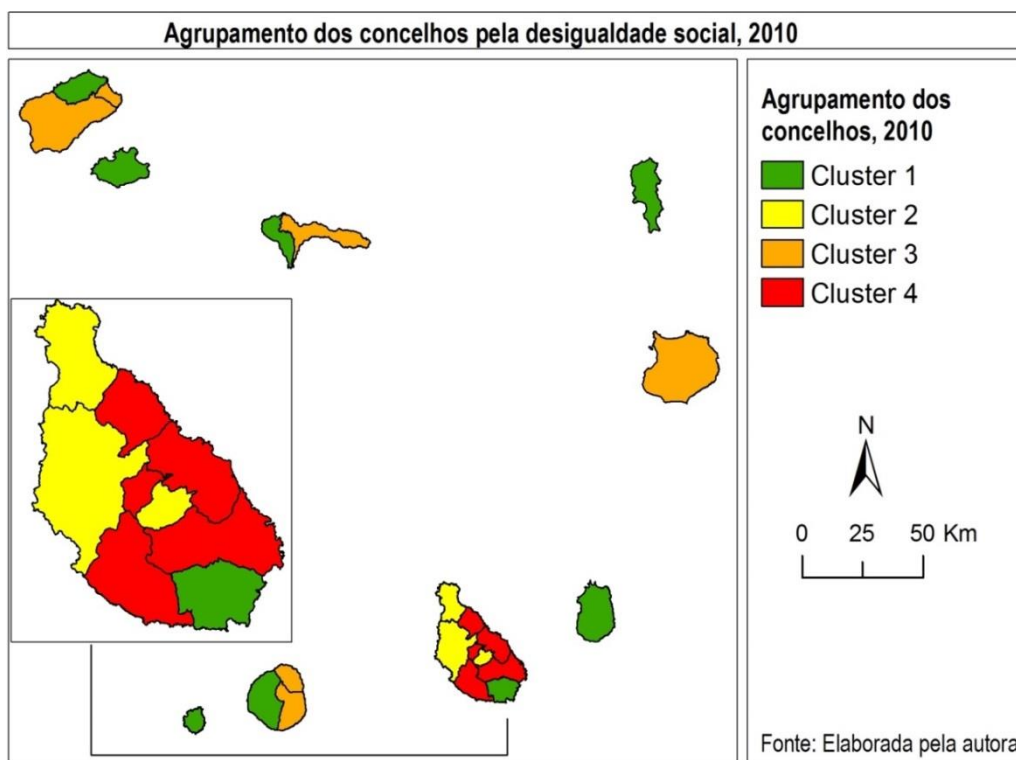


Figura 4.33: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2012

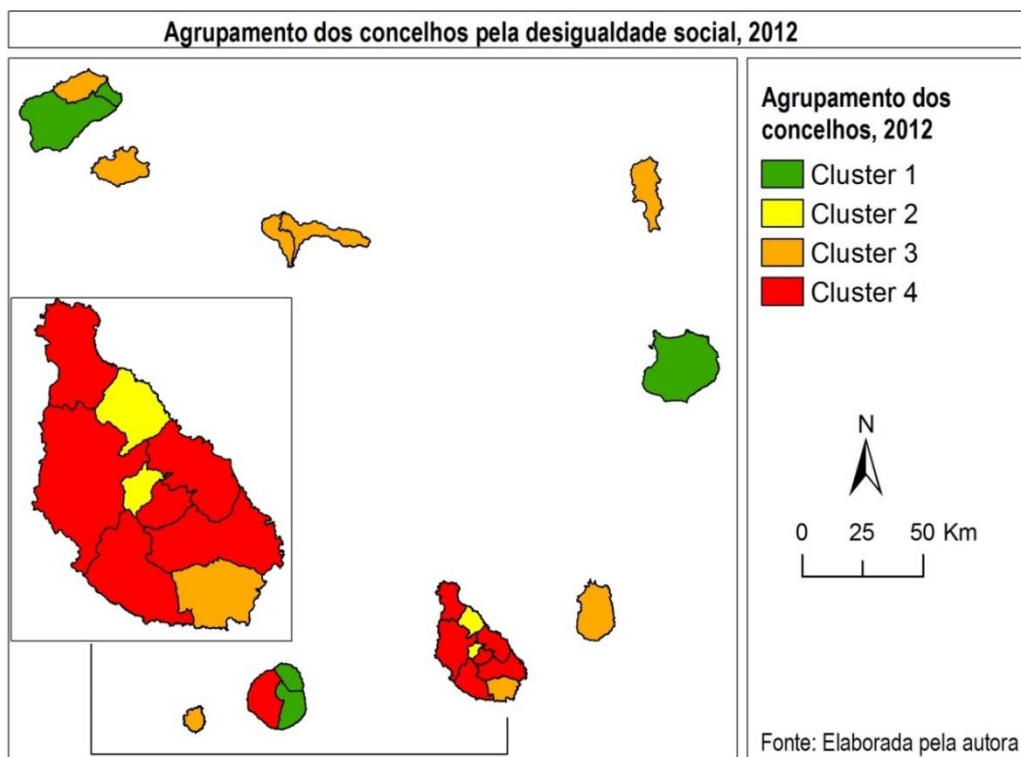


Figura 4.34: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2013

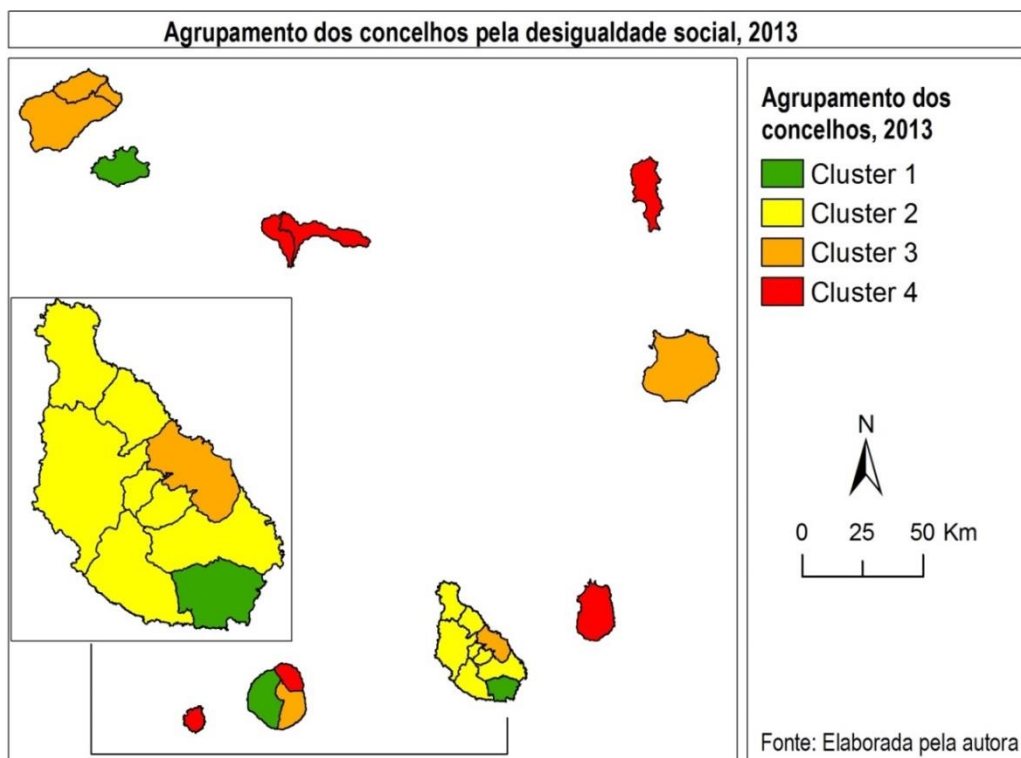


Figura 4.35: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2014

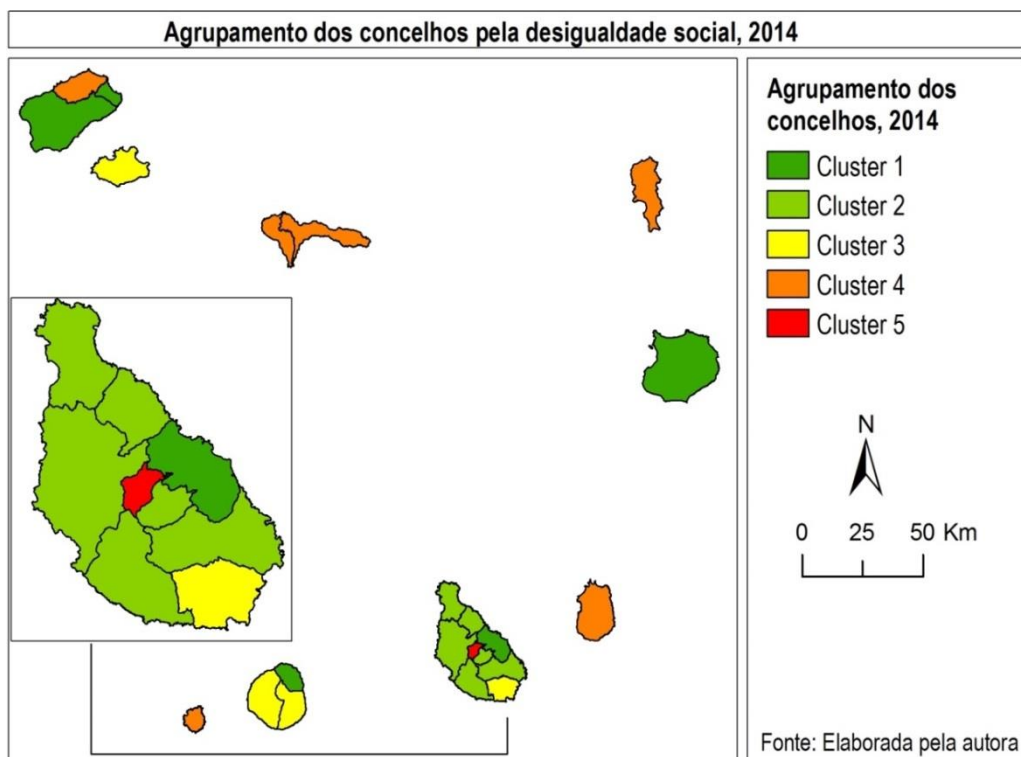
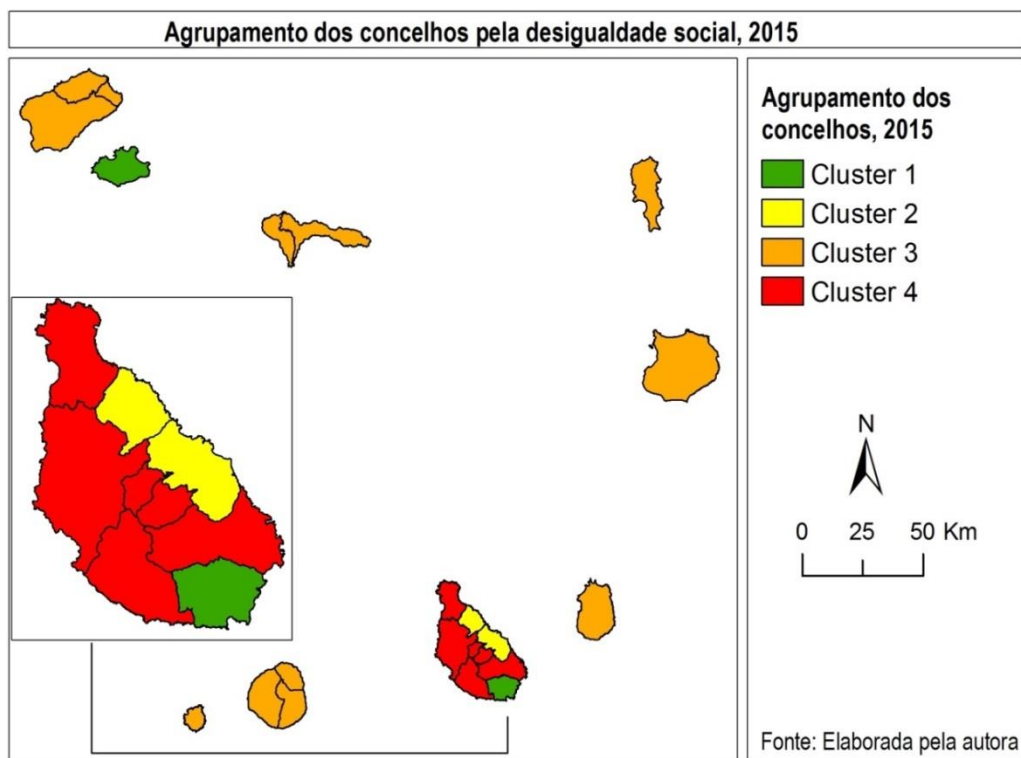


Figura 4.36: Agrupamento pelo Método Mojena (1977): IDA e IDG, 2015



No que respeita à semelhança dos concelhos em relação a desigualdade social, verifica-se, através da análise de *clusters* que os concelhos São Vicente e Praia, Sal e Maio, Tarrafal, Santa Catarina e São Lourenço dos Órgãos, e São Domingos e Ribeira Grande de Santiago são semelhantes. Pois, estes concelhos pertencem ao mesmo *cluster* em todos os anos em análise. É notável, também, que os concelhos da ilha de Santiago são bem semelhantes em relação a desigualdade social, com excepção do concelho da Praia. Este último, em nenhum dos anos se agrupa com os demais concelhos da ilha. Estes agrupamentos explicam-se face às características económicas destes Concelhos. Por exemplo o concelho da Praia tem um desenvolvimento económico, turístico (...) muito mais acentuado, sendo um polo de atracção para as populações dos outros concelhos da mesma ilha, como das outras ilhas, por estar concentrada maiores economias, e, portanto, maiores oportunidades de negócios, a custo mais baixos em termos gerais.

No ano 2015, os concelhos com características semelhantes do IDA e IDG pertencem a mesma fronteira, com excepção do concelho Praia que apresenta características semelhantes ao concelho São Vicente (Figura 4.36). Este facto comprova-se pelo elevado desenvolvimento económico destes dois concelhos, em relação aos restantes. Ou seja, Praia e São Vicente são concelhos com características bem semelhantes, para além de ser os dois grandes concelhos do país com desenvolvimento económico bem destacado.

CONSIDERAÇÕES FINAIS

A análise espacial é hoje uma ferramenta de grande porte com grandes potencialidades para a análise da distribuição territorial de dados com informação de interesse, particularmente para um país composto por ilhas, como é o caso de Cabo Verde, permitindo identificar padrões espaciais e algum tipo de interacção entre os lugares observados. Na mesma linha aparece a estatística espacial que possibilita a extracção dos resultados mais significativos, fornece mais visibilidade e confiança as análises.

Um dos importantes aspectos na análise espacial é a forma de organização dos dados antes do seu tratamento. A forma de organização dos dados na *geodatabase* deve ir de encontro a análise e estatística espacial que se pretende realizar. Neste estudo em concreto, a forma de organização dos dados foi uma das maiores dificuldades encontradas, pois, como foi anteriormente referido, foram realizados diferentes tipos de análise, sendo que num caso a estrutura em que se encontravam não era a mais apropriada. Assim, chegou-se à conclusão que a melhor forma de organizar os dados, para este estudo, é por ano, ou seja, cada ano é uma *feature layer*. Concluiu-se também que as informações deveriam ser organizadas por distrito de recenseamento, e não por concelho ou freguesia, permitindo assim obter o número recomendado de *features* (pelo menos 30 *features*) para se aplicar as análises pretendidas.

O foco deste trabalho foi medir as desigualdades sociais através de indicadores reconhecidos, compará-las e analisar os seus aspectos quanto à sua distribuição espacial mediante as técnicas de análise e estatística espacial em ambiente SIG. O estudo dos indicadores socioeconómicos e demográficos, que melhor reflectem as desigualdades sociais, foi realizado de acordo com a sua importância na quantificação da desigualdade e disponibilidade da informação. Os indicadores utilizados para determinar a dimensão educação foram avaliadas, não só pelo nível de importância destes indicadores reflectirem na desigualdade, mas levando em conta, também, a tradição antiga dos Cabo-Verdianos. Por exemplo, a taxa de alfabetização foi levada em consideração, na quantificação da desigualdade de género, pelo facto de muitas mulheres, hoje com mínimo de 50 anos, não terem frequentado escolas porque os pais não permitiram, por um único motivo de não apreenderem a escrever cartas para os namorados (podemos designar de “desigualdade de ignorância”?). O saneamento foi considerado para determinar a dimensão saúde, porquanto, regiões com população mais pobres tende a ter saneamento de má qualidade.

A quantificação dos índices de desigualdades deu-se através dos índices compostos. Estes índices sintetizam as diferentes dimensões utilizadas para o seu cálculo. Podem ser considerados indicadores resumo, pois, representam as diversas dimensões analisadas em um único valor. Eles são úteis no cálculo da desigualdade social, pois, permitem tornar compatível, conciliar, dados de diversas fontes e origens. Entretanto, a escolha das variáveis a ser utilizadas é determinante e afecta consideravelmente o índice, podendo gerar, também, resultados inesperados. O que se espera é que as regiões de alto padrão possuam melhores índices do que as regiões de baixo padrão. Neste sentido, a análise espacial apresenta extrema importância, pois, dá possibilidade de verificar espacialmente se o esperado é facto. Assim, com a utilização destes dois mecanismos, índices compostos e análise espacial, foi possível quantificar os índices de desigualdade desejados e analisá-las espacialmente ao nível do concelho.

Foram analisados dados de quatro índices de desigualdade: índice de desigualdade ambiental, de género, em saúde e económico. A desigualdade ambiental em Cabo Verde é a mais elevada quando comparado com os outros tipos de desigualdades aqui medidas. É de notar que em todos os anos a desigualdade ambiental foi mais elevada, e com tendência a aumentar. Constatou-se, também, que os concelhos de alto padrão, como Praia e São Vicente, não estão isentos de certos problemas ambientais.

No início foi avaliada a autocorrelação espacial através de determinação do índice global de Morans. O índice foi significativo para a desigualdade de género, com excepção do ano 2000, e desigualdade ambiental em todos os anos em análise. Analisando os valores dos índices Global de Morans verificou-se que a dependência espacial existente é positiva, mas os valores dos índices são baixos, com excepção do índice do ano 2011, ponderada pelo IDG. Assim, conclui-se que os valores dos índices de desigualdade para alguns concelhos não são semelhantes aos dos seus concelhos mais próximos. Através do índice Local de Morans pode-se verificar que concelhos apresentam esta característica, sendo que para a desigualdade de género todos os índices são muito baixos, inferior a 0,01. Foi possível perceber, que os concelhos Tarrafal S.N., Sal e Boavista apresentavam índices elevados, no ano 2000, tendo o primeiro apresentar esta mesma característica em 2010.

Para validar a análise de *clusters*, que não é robusta na presença de *outliers*, realizou-se inicialmente uma análise exploratória para identificar dados discrepantes. Entretanto, sendo um dos objectivos principais deste trabalho analisar a distribuição espacial das desigualdades, este deve ser

feito através da análise espacial, especificamente a análise de autocorrelação espacial, feita desta através do índice global e local de Morans, por forma a verificar a interdependência regional. O índice global de Morans mostrou que existe autocorrelação espacial e que alguns concelhos são *ouliers* estatisticamente significativos. Neste sentido, os resultados da análise de *clusters* podem parecer estar deturpados.

Os métodos de determinação do número de *clusters* apresentaram discrepância nos resultados, sendo que alguns apresentam número de agrupamentos menores, como foi o caso do método Mojena (1977), e outros bem maiores, caso do RMSSTD. Através do teste da ANOVA pode-se verificar, que apesar dos métodos RMSSTD e *pseudo-F* determinaram número de agrupamentos muito maiores, estes determinam agrupamento mais homogêneos dentro dos grupos e heterógenos entre os grupos. Pois, como foi visto, o método índice R^2 apresenta resultados mais satisfatórios do que o método Mojena (1977), com número menor de grupos. Entretanto, com a existência de poucos elementos para o agrupamento não se considerou pertinente criar muitos grupos.

Pode-se daqui concluir que os métodos que determinam grupos pelas suas proximidades tende a determinar maiores agrupamentos (RMSSTD e *pseudo-F*), quando comparado com os que determinam grupos pelas suas distâncias (Mojena, 1977 e índice R^2).

A semelhança, na desigualdade social, entre os concelhos da ilha de Santiago é notável. Estes, com excepção do concelho Praia, pertenciam, quase sempre, aos mesmos *clusters*. Em 2000 e 2013 os concelhos da ilha, com excepção dos concelhos Praia e Santa Cruz, formaram um *cluster* e, o mais impressionante, é que com o passar dos anos parecem ser mais semelhantes. Este pode ser explicado pelas características comuns apresentadas por esses concelhos (por exemplo, meios de subsistência das populações, dada praticamente pela agricultura). No ano 2015 os concelhos da ilha do Fogo, Santo Antão e São Nicolau pertencem ao mesmo *cluster*, ou seja, são semelhantes no que tange a desigualdade social. Estas três ilhas, também, caracterizam pela prática da agricultura. A ilha de São Nicolau é frequentemente comparada com à de Santo Antão, pois, ambas possuem montanhas espectaculares, que as torna excelentes pontos turísticos. Estes resultados vêm ao encontro do resultado da autocorrelação espacial. Isto é, concelhos com dependência espacial de índice negativo não se agruparam com os concelhos mais próximos, mas sim com os concelhos mais distantes, e concelhos com índice positivos com os concelhos mais próximos.

Uma das valências da análise espacial que encontramos nos estudos realizados, que poderia ser aproveitada para trabalhos futuros sobre a desigualdade social, é a utilização da análise de autocorrelação espacial para a escolha de variáveis espaciais que servirão para a determinação dos índices compostos. Este estudo deve abranger principalmente a desigualdade ambiental, pois, como foi concluído neste estudo, ela apresenta uma tendência crescente. Neste sentido, com a análise de autocorrelação espacial das variáveis que podem influenciar a desigualdade ambiental, vê-se claramente quais destas variáveis influencia no aumento desta desigualdade.

Tendo em conta que cada indicador conduz a resultados diferentes, conseguindo alcançar a desigualdade sob uma determinada óptica, propõe-se medir as desigualdades sociais aqui estudadas, através dos diferentes indicadores aqui apresentados no capítulo 1, e outros indicadores de dependência espacial como por exemplo, a regressão espacial.

Um outro estudo importante, seria quantificar a desigualdade social nos jovens e compará-las com as desigualdades aqui determinadas. A fim de poder tirar alguma conclusão sobre o impacto das desigualdades sociais dos jovens na desigualdade total. A percepção actual é que a população jovem é, em Cabo Verde, a mais prejudicada. Isto pode ser vista através da taxa de desemprego que atingiu 41% em 2016, nos jovens com idade compreendida entre 15 e 24 anos.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alencar, B. J. (2009). *A análise multivariada aplicada ao tratamento da informação espacial: uma abordagem matemático-computacional em análise de agrupamentos e análise de componentes principais*. Tese de doutoramento, Universidade Católica de Minas Gerais, Belo Horizonte.
- Alves, H. (2007). *Desigualdade ambiental no município de São Paulo: análise da exposição diferenciada de grupos sociais a situações de risco ambiental através do uso de metodologias de geoprocessamento*. Revista Brasileira Est. Pop., São Paulo, v. 24, n. 2, p. 301-316, jul./dez.
- Anselin, L. (1992). *Spatial data analysis with GIS: an introduction to application in the social sciences*. National Center for Geographic Information and Analysis University of California Santa Barbara, CA 93106.
- Assunção, R. M. (2001). *Estatística Espacial com aplicações em Epidemiologia, Economia e Sociologia*. Associação Brasileira de Estatística, São Carlos.
- Bailey, T. C., & Gatrell, A. C. (1995). *Interactive Spatial Data Analysis. Essex-England: Longman Scientific and Technical*.
- Bem-Estar, Q. U. (2007). Cabo Verde: Instituto Nacional de Estatística de Cabo Verde.
- Castro, R. S. (2006). *Efeitos da Desigualdade de Rendimento sobre o Crescimento Económico no Brasil: uma análise não linear*. UFRGS/FCE, Porto Alegre.
- Coelho, M. M. (26 de janeiro de 2009). As desigualdades e Rousseau.
- Costa, A. F. (2012). Desigualdades globais. *Sociologia Problemas e Práticas*(nº 68), pp. 9-32.
- Dias, P. V. (2013). *Análise exploratória de dados espaciais de actores de sectores estratégicos na cidade de Lisboa*. Relatório de Estágio de Mestrado em Gestão do Território, Universidade Nova de Lisboa, Faculdade de Ciências Sociais e Humanas (FCSH).
- Druck, S., Carvalho, S. M., Câmara, G., & Monteiro, A. M. (2004). *Análise espacial de dados geográficos*. Brasília: Embrapa.
- Druck, S., Carvalho, S. M., Câmara, G., & Monteiro, A. M. (2004). *Análise especial de dados geográficos*. Brasília: Embrapa.
- Ebdon, D. (1985). *Statistics ind Geography*. Oxford (UK): Blackwell.
- ESRI. (2016). *ArcGIS Desktop*, 10.4. (Redlands, CA: Environmental Systems Research Institute) Obtido em 25 de Outubro de 2016, de ArcGIS Pro: <http://pro.arcgis.com/fr/pro-app/tool-reference/spatial-statistics/how-grouping-analysis-works.htm#GUID-2F9A290B-DA01-4CA3-9AF0-A10E56C9A23F>
- Freudenberg, M. (2003). *Composite Indicators of Country Performance*. Paris: OECD.
- Gao, A. (1999). Equity and health. *Presented at the XI World Congress of Psychiatry*. Hamburg, Germany.
- Genovez, P. C. (2005). *Território e desigualdade: análise espacial intraurbana no estudo da dinâmica de exclusão/inclusão social no espaço urbano de São José dos Campos*. Dissertação de Mestrado do Curso de Pós-Graduação em Sensoriamento Remoto, INPE - 12177 - TDI/973.
- Hellebrandt, T., & Mauro, P. (Abril de 2015). The Future of Worldwide Income Distribution. (P. I. Economics, Ed.) *Working Paper Series 15-7*.
- Hoffmann, R. (1998). *Distribuição de Renda: medidas de desigualdade e pobreza*. São Paulo, Brasil: Editora da Universidade de São Paulo.

- IDRF. (2002). *O Perfil da Pobreza em Cabo Verde. Inquérito às Despesas e Receitas Familiares 2001/02*. INECV.
- Imprensa Nacional de Cabo Verde. (6 de Março de 2000). Boletim Oficial. *Comissão Nacional de Eleições. Edital nº 5/2000, I Série - Nº 6*, pp. 107-126.
- Imprensa Nacional de Cabo Verde. (5 de Abril de 2004). Boletim Oficial. *Comissão Nacional de Eleições. Edital nº 7/2004, I série - Nº 10*, pp. 2-37.
- Imprensa Nacional de Cabo Verde. (3 de Junho de 2008). Boletim oficial. *Comissão Nacional de Eleições. Edital nº 02/2008, I Série - Nº 20*, pp. 2-25.
- Imprensa Nacional de Cabo Verde. (1 de Agosto de 2012). Boletim Oficial. *Comissão Nacional das Eleições. Edital nº 3/2012, I Série - Nº 45*, pp. 932-955.
- Krempi, A. P. (2004). *Explorando recursos de estatística espacial para análise de acessibilidade na Cidade de Bauru*. Dissertação de mestrado, Escola de Engenharia, Universidade de São Paulo, Departamento de Transportes, São Carlos.
- Kunst, A., & Mackenbach, J. (1994). *Measuring socioeconomic inequalities in health*. document EUR/ICP/RPD 416, WHO Regional Office for Europe.
- Magalhães, A., & Carvalho, K. F. (2010). *Educação, Trabalho e Desigualdade: desafios para a promoção da mobilidade social*. UFJF. Ed. 9: 228-239.
- Mainali, B., Swaha, A., Varenyan, T., Dheemahi, B., & Prachodayat, D. (2014). Assessing rural energy sustainability in developing countries. *Energy for Sustainable Development*.
- Marôco, J. (2014). *Análise Estatística com o SPSS Statistics* (6.ª ed.). ReportNumber, Lda. Pero Pinheiro.
- Medeiros, M. (2012). *Medidas de Desigualdade e Pobreza*. Fundação Universidade de Brasília. Brasília: Universidade de Brasília.
- Mojena, R. (1997). *Hierarchical grouping methods and stopping rules: an evaluation*. Universidade de Rhode Island, Departamento de Management Science, Kingston, Rhode Island.
- Morato, R. G., Kawakubo, F. S., & Luchiari, A. (2005). *Geografia da desigualdade ambiental na subprefeitura de campo limpo município de São Paulo*. Universidade de São Paulo, Departamento de Geografia. Goiânia, Brasil: Simpósio Brasileiro de Sensoriamento Remoto .
- Mundial, B. (2014). *World Development Indicators database*. Washington, DC.
- MUT. (2012). *Relatório do Módulo Uso do Tempo e Trabalho Não Remunerado em Cabo Verde*. Instituto Nacional de Estatística, Cabo Verde.
- Noronha, K. V., & Andrade, M. V. (2001). *Desigualdade Sociais em Saúde: evidências empíricas sobre o caso brasileiro*. Fortaleza: Revista Económica do Nordeste.
- Painho, M. (1998). Os sistemas de informação geográfica como ferramenta de suporte de integração de planos. In: *Actas do Workshop sobre Instrumentos de Planeamento: Oportunidade e Condicionantes no Ambiente*. Associação Portuguesa de Engenharia do Ambiente: Lisboa, 17 e 18 de Junho.
- PNUD. (2011). *Sustentabilidade e Equidade: um futuro melhor para todos. Relatório do Desenvolvimento Humano de 2011*.
- Proença, C. S. (2009). *Exclusão Social em Cabo Verde. Uma Abordagem Preliminar*. CEsa. Lisboa.
- Projeções Demográficas de Cabo Verde. (2013). *Projeções Demográficas de Cabo Verde, 2010-2030*. Cabo Verde: Instituto Nacional de Estatística.
- Questionário Unificados dos Indicadores Básicos do Bem-Estar. (2007). Cabo Verde: Instituto Nacional de Estatística de Cabo Verde.

- Schneider, M. C., Salgado, C. C., Bacallao, J., Loyola, E., Mujica, O. J., Vidaurre, M., et al. (2002). *Métodos de mensuração das desigualdades em saúde*. Rev Panam Salud Publica.
- ScotPHO. (2007). *Measuring Socio-Economic Inequalities in Health: Practical Guide*. Public Health Information for Scotland.
- Sen, A. (2001). *Desigualdade reexaminada*. Rio de Janeiro: Record.
- Sharma, S. (1996). *Applied multivariate techniques*. Universidade de South Carolina, New York: John Wiley & Sons.
- Tartaruga, I. G. (2008). *Análise espacial da centralidade e da dispersão da população do Estado do Rio Grande do Sul de 1970 a 2000: notas preliminares*. Brasil: Fundação de Economia e Estatística (FEE).
- Teague, P., & Wilson, R. (1995). *Towards an inclusive society. Social Exclusion, Social inclusion, Democratic Dialog* (Vol. 2).
- Therborn, G. (2006). Meaning, mechanisms, patterns, and forces: an introduction, in Göran Therborn, (ed.). In *Inequalities of the world. New Theoretical Frameworks, Multiple Empirical Approaches* (pp. 1-58). London, Verso.
- Tilly, C. (2006). *O acesso desigual ao conhecimento científico* (2 ed., Vol. 18). Tempo Social.
- Wagstaff, A., Paci P, V., & Doorslaer, E. (1991). On the measurement of inequalities in health. *Soc Sci Med.*, pp. 545-547.
- Wilkinson, L., Engelman, L., Corter, j., & Coward, M. (1998). Cluster Analysis. In L. Wilkinson, *SYSTAT 8.0 Statistics* (pp. 65-123). Chicago.