

Introdução à Quantificação em Geografia

(Com uso do Excel)

Paulo Fernando Braga Carvalho

Este trabalho é uma adaptação do texto extraído da dissertação de mestrado do autor, em Tratamento da Informação Espacial, pela PUCMINAS, desenvolvida com orientação do Prof. Dr. Leônidas Conceição Barroso e co-orientação do Prof. Dr. João Francisco de Abreu

SUMÁRIO

1. Introdução à quantificação em Geografia.....	4
1.1 Introdução.....	4
1.2 População, amostra e amostragem.....	6
1.3 Dado e informação.....	7
1.4 Tipos de dados.....	8
1.4.1 Dados nominais.....	8
1.4.2 Dados ordinais.....	9
1.4.3 Dados intervalares.....	10
1.4.4 Dados resultantes de razões matemáticas.....	10
1.5 Outros tipos de classificação dos dados.....	11
1.5.1 Dados qualitativos e dados quantitativos.....	11
1.5.2 Dados discretos e dados contínuos.....	11
2. Apresentação de dados.....	13
2.1 Grupamento de dados.....	13
2.2 Gráficos.....	24
2.2.1 Histograma.....	25
2.2.2 Polígono de freqüência.....	26
2.2.3 Ogiva.....	27
2.2.4 Gráfico em linha.....	28
2.2.5 Gráfico em Barras (ou em Colunas).....	30
2.2.6 Gráfico em setores.....	31
3. Medidas Estatísticas.....	34
3.1 Medidas de Tendência Central.....	34
3.1.1 Média aritmética simples.....	34
3.1.2 Média aritmética ponderada.....	37
3.1.3 Média geométrica.....	39
3.1.4 Mediana.....	45
3.1.5 Moda.....	47
3.1.6 Percentil.....	48
3.2 Medidas de Dispersão.....	49

3.2.1 Amplitude	50
3.2.2 Desvio Médio Absoluto	50
3.2.3 Variância e Desvio Padrão.....	52
3.2.4 Coeficiente de Variação	56
3.3 Medidas Estatísticas para dados agrupados	57
3.3.1 Média	58
3.3.2 Mediana	59
3.3.3 Moda	60
3.3.4 Percentil	60
3.3.5 Variância e Desvio Padrão.....	61
3.3.6 Comparação entre Média, Mediana e Moda	62
4. Estatística Espacial	64
4.1 Introdução.....	64
4.2 Medidas de tendência central em padrões de pontos.....	70
4.2.1 Centro Modal.....	70
4.2.2 Centro Mediano.....	70
4.2.3 Centro Médio.....	71
4.2.4 Centro Médio Ponderado	74
4.2.5 Centro Médio para dados agrupados.....	77
4.2.6 Aplicação	78
4.3 Medidas de Variabilidade ou Dispersão em distribuições espaciais de pontos	94
4.3.1 Retângulo Quartílico	94
4.3.2 Índice de Dispersão	97
4.3.3 Distância Padrão.....	98
4.3.4 Distância Padrão Ponderada	102
4.3.5 Dispersão Relativa	105
5. Alguns exemplos resolvidos usando o EXCEL.....	109
6. Exercício proposto.....	124

1 – INTRODUÇÃO À QUANTIFICAÇÃO EM GEOGRAFIA

1.1 - Introdução

Você está iniciando o estudo de um dos instrumentos mais importantes na tomada de decisões e que quando aplicada à Geografia, em meados da década de 1950, causou a denominada “revolução quantitativa” .

Trata-se da Quantificação em Geografia. Segundo TAYLOR (1977, p.23), citado por GERARDI e SILVA (1981), “deriva basicamente de analogias entre os diagramas de frequência dos estatísticos e os mapas de frequência dos geógrafos resultando numa estatística descritiva bidimensional”.

Embora a palavra Estatística ainda não existisse, há indícios de que 3000 anos a.C já se faziam censos na Babilônia, China e Egito. Também encontramos referências a recenseamentos na Bíblia:

No primeiro dia do segundo mês do segundo ano da saída do Egito, Javé falou a Moisés na tenda da reunião, no deserto do Sinai: Façam um recenseamento completo da comunidade dos filhos de Israel: todos os homens, um a um, conforme clãs e famílias, registrando seus nomes. Você e Aarão registrarão, por esquadrões, todos os homens maiores de vinte anos e capacitados para a guerra.(Nm1,1-3)

Este recenseamento, citado no Antigo Testamento, tem por objetivo organizar o povo para a conquista da terra que Deus vai lhe dar.

Outra citação encontrada na Bíblia, no Novo Testamento, trata da necessidade de saber quantas pessoas deviam pagar o tributo cobrado pelo imperador:

Naqueles dias, o imperador Augusto publicou um decreto, ordenando o recenseamento em todo o império. Esse primeiro recenseamento foi feito quanto Quirino era governador da Síria.³ Todos iam registrar-se, cada um na sua cidade natal.(Lc 1,1-3)

Sob a palavra estatística acumularam-se descrições e dados relativos ao estado.

Os censos existem há milhares de anos e constituem um esforço imenso e caro feito pelos governos com o objetivo de conhecer seus habitantes, sua condição sócio-econômica, sua cultura, religião, etc. Portanto, associar-se estatística a censo é perfeitamente correto do ponto de vista histórico, sendo interessante salientar que as palavras estatística e estado têm a mesma origem latina: status. (SOARES et al., 1991, p.1)

Segundo COSTA (1992, p.6), a palavra estatística foi cunhada pelo acadêmico alemão Gottfried Achenwall por volta da metade do século XVIII e apareceu como verbete “Statistic” (Estatística) na Enciclopédia Britânica em 1797.

O grande volume de dados com que o geógrafo trabalha, precisa ser organizado e apresentado de uma maneira sintética, mais fácil de ser manipulada e de fácil compreensão.

Atualmente, várias ciências lançam mão dos instrumentos quantitativos oferecidos pela Estatística. Porém, a Geografia pede uma atenção especial, pois lida, também, com a grandeza espacial, o que leva a um tratamento bidimensional dos dados. Procura-se descobrir a relação entre os fenômenos estudados e sua distribuição no tempo e no espaço.

Buscando desenvolver um curso de caráter prático e introdutório, procurou-se apresentar cada tópico de modo que o aluno tenha acesso ao conteúdo da Estatística Espacial fundamental, explorando o aspecto intuitivo das definições e propriedades dos entes matemáticos, deixando em segundo plano as demonstrações das fórmulas e teoremas, que podem ficar para um segundo curso.

Para que a linguagem fique mais próxima do geógrafo, uma parte significativa dos exemplos selecionados são aplicados à ciência de interesse: a Geografia. Busca-se, assim, que o estudante consiga perceber as vantagens e desvantagens das técnicas estatísticas aplicadas, de tal maneira que possa fazer uma escolha criteriosa da técnica a ser empregada em seu trabalho.

O curso original foi desenvolvido com o suporte do software matemático MATLAB® (MATrix LABoratory) que é um sistema interativo e uma linguagem de programação para computação técnica e científica. É um programa de computação numérica, de análise e visualização de dados que permite ao usuário obter respostas e entender como obtê-las.

Entretanto, na prática, percebeu-se que, apesar de bastante difundido no meio acadêmico, devido ao custo financeiro, o MATLAB não tem sido disponibilizado com muita frequência para cursos fora da área exata, como é o caso da Geografia, que prioriza a aquisição de softwares específicos de geoprocessamento. Portanto, optou-se por criar esta versão do curso de Introdução à Quantificação em Geografia, adotando o software EXCEL® como ferramenta de suporte computacional.

A opção pelo software EXCEL justifica-se por possuir recursos relativamente simples para cálculos e construções de gráficos que facilitam muitas tarefas extremamente laboriosas, quando feitas apenas com o uso de calculadoras, lápis e papel. Além disso, é um software bastante difundido, tanto na área comercial quanto na acadêmica.

O estudante terá a oportunidade de ver a resolução de boa parte dos exemplos usados no texto com a utilização do EXCEL. A correta utilização desta poderosa ferramenta computacional pode trazer benefícios importantes como, por exemplo, a economia de tempo no desenvolvimento das questões, confiabilidade dos resultados numéricos obtidos e a interatividade com os resultados e conceitos aprendidos, já que agrega a capacidade de fazer cálculos com a visualização gráfica.

1.2 - População, Amostra e Amostragem

População (ou universo) é o conjunto de todas as informações sobre um fenômeno, que tenham, entre si, alguma característica em comum.

Por exemplo, o conjunto das alturas de todos os picos do mundo, constitui a população das alturas dos picos do mundo. O conjunto dos números de habitantes de todos os municípios do Brasil constitui a população dos números de habitantes dos municípios do Brasil.

Mas, na maioria das vezes, a população é um conjunto muito grande, o que dificulta ou impossibilita o estudo de elemento por elemento. Esta dificuldade pode ser representada pelo alto custo no levantamento de dados, pelo tempo gasto no levantamento das informações, pelo custo com mão-de-obra, etc.

Usa-se, portanto, uma amostra. Amostra é qualquer subconjunto representativo com as características essenciais de uma população. A amostragem, ou seja, o processo de obtenção da amostra, não será discutida neste texto.

1.3 - Dado e Informação

Uma quantidade muito grande de dados apresentados de modo casual ou mesmo organizados de modo inadequado pode confundir o leitor, indo contra o objetivo de buscar elucidacões sobre o assunto estudado. As características essenciais podem ficar ocultas em meio aos detalhes, alguns de pouca ou nenhuma importância, prejudicando a análise do fenômeno.

O processamento de dados reduz a quantidade de detalhes e facilita a constataçã de relações. Quando são processados e analisados, os dados são transformados em informações. Para isso, os dados são organizados e condensados em uma quantidade menor de números que devem transmitir sua essência. Com este processamento, seus aspectos mais importantes são enfatizados.

Para promover o processamento e a visualizaçã dos dados, usa-se tabelas, mapas e gráficos. Estes dois últimos, são especialmente atraentes pelo apelo visual das principais características dos dados. Neste caso em particular, leitores interessados na quantificaçã geográfica, a parceria dos mapas com as informações estatísticas mostra-se de fundamental relevância.

As tabelas, apesar de não possuírem o mesmo apelo visual apresentado pelos mapas e gráficos, têm sua importância pautada nas vantagens em termos de análise matemática. Os dados tabelados permitem manipulações matemáticas que podem retornar números que auxiliam na análise do contexto e verificação da relação entre as variáveis.

1.4 - Tipos de dados

Os métodos estatísticos envolvem a análise e a interpretação de dados que podem ter sido gerados, por exemplo, pela medição das temperaturas registradas nos postos de controle de um instituto de meteorologia, pela avaliação do índice de criminalidade em uma determinada região, pela contagem de consumidores que passam diariamente por determinado centro de compras, pela ocorrência ou não de determinado fenômeno em algumas cidades, etc.

O geógrafo lida constantemente com variáveis de tipos e origens diversos. Com a análise do dados, surgem informações que precisam ser interpretadas com o objetivo de dar significado e explicações para os fenômenos estudados. “A variedade de tipos de dados usados na Geografia reflete os interesses e as preocupações da pesquisa geográfica.” (O’Brien, 1992)

O geógrafo trabalha constantemente com dados que podem ser representados por áreas e pontos sobre a superfície terrestre e aspectos de relacionamento entre essas áreas e pontos. Na busca de informações encontram-se, basicamente, quatro tipos de dados que se distinguem, de acordo com a escala de mensuração:

1.4.1 - Dados Nominais

A escala nominal, ou classificatória, é encontrada quando os dados são agrupados em categorias diferentes, de acordo com seus atributos. Os dados consistem em nomes, rótulos, etc.

Este tipo de escala é muito comum para dados que acusam a presença ou ausência de determinado fenômeno. Por exemplo, determinadas regiões podem apresentar o atributo de possuir vulcões ativos ou não possuir vulcões ativos. Cada pessoa pode ser classificada de acordo com o sexo: masculino ou feminino.

Os dados nominais envolvem categorias tais como sim ou não, região (norte, sul, leste ou oeste), nível de desenvolvimento de determinada região (alto, médio ou baixo). Estas características não são naturalmente numéricas, mas, cada item do conjunto de dados pode ser atribuído a uma única categoria e a frequência da categoria pode ser quantificada.

1.4.2 - Dados Ordinais

Os dados ordinais buscam apresentar uma hierarquia no conjunto de dados. Os dados são organizados segundo sua importância, grandeza, preferência, etc, dentro do conjunto de dados.

Temos exemplos numerosos do uso desta escala em Geografia como a divisão da população em classes sociais, a hierarquia urbana, o padrão habitacional de uma população, a preferência locacional e a escala de dureza dos minerais que é um exemplo clássico de escala ordinal. (GERARDI & SILVA, 1981)

Esta escala de mensuração é aplicável quando se deseja trabalhar com valores qualitativos, encontrando pouca aplicação das operações matemáticas sobre esses dados. Em diversas situações podemos supor que determinado elemento é melhor que outro, mas nem sempre podemos quantificar o número de vezes que um é melhor que outro. Por exemplo, podemos apontar o nível de conforto encontrado em diversos cinemas, mas não podemos dizer que um cinema de nível “classe A” é duas ou três vezes melhor que outro de nível “classe B”.

Ao estudar as medidas estatísticas, veremos que a mediana é uma boa medida representativa para dados ordinais.

1.4.3 - Dados Intervalares

São os dados que podem assumir qualquer valor em um intervalo de números reais. Como exemplo, encontramos a distância entre dois pontos turísticos de uma cidade, a temperatura máxima registrada num posto meteorológico num intervalo de tempo, a quantidade de água consumida diariamente por uma determinada cidade, a velocidade do vento, etc.

Em GERARDI & SILVA (1981, p.26), encontra-se uma interessante discussão sobre este tipo de dado, que é apresentado com o nome de escala de intervalo:

A escala de intervalo é apropriada a operações de transformação linear, isto é, podemos converter °C em °F¹ através da equação $^{\circ}F = 32 + 1,8^{\circ}C$. Cada transformação linear desta forma preserva a informação no dado original. Transformações mais complexas alteram relações. . . . não tem sentido dizer que a temperatura de 30°C seria duas vezes mais quente do que 15°C. Caso fosse possível, isto significaria que 86°F (=30°C) seria duas vezes superior a 59°F (=15°C).

O que foi dito, exemplifica o fato de não existir o zero absoluto, o ponto de partida inerente para os dados deste tipo. A temperatura de 0°F (32°C) não indica ausência de calor, apesar de parecer um ponto de partida. Ou seja, para este tipo de dados o ponto de partida é arbitrário.

1.4.4 - Dados Resultantes de Razões Matemáticas

Este tipo de dado é semelhante aos dados de intervalos, diferenciando-se pela inclusão do zero como ponto de partida, indicando ausência de ocorrência do fenômeno. As razões matemáticas tomam sentido, ou seja, dizer que a massa de

¹ °C: Graus centígrados °F: Graus Fahrenheit
Paulo Fernando Braga Carvalho

uma determinada pedra é de 20kg, implica em dizer que a massa de outra pedra com 40kg é o dobro da primeira. A duração de uma peça de teatro ser igual 0 minutos indica, que não houve apresentação teatral.

Alguns exemplos para este tipo de dados são as escalas métricas, idades, distâncias, áreas, produção.

1.5 - Outros Tipos de Classificação dos Dados

1.5.1 - Dados Qualitativos e Dados Quantitativos

Os dois primeiros tipos de dados, nominais e ordinais, lidam com as características qualitativas dos dados, ou seja, os dados se diferenciam por alguma característica não-numérica, enquanto que os dois últimos, intervalares e de razão, lidam com características quantitativas, portanto, originados de contagens ou medições.

1.5.2 - Dados Discretos e Dados Contínuos

Os dados discretos são resultado de contagem do número de ocorrências de determinado fenômeno, portanto são representados por números inteiros não-negativos.

Os dados contínuos podem assumir qualquer valor em um intervalo de números reais, não existindo lacunas entre dois possíveis valores. Muito comumente encontrados em situações como tempo, peso, distância, área, etc.

Estes dados são denominados contínuos, mas, na prática, raramente depara-se com a possibilidade de lidar com resultados realmente contínuos, dadas as limitações físicas encontradas em nossos instrumentos de medidas, que afetam sua precisão.

O QUADRO 2, resume os tipos de dados e as medidas estatísticas mais apropriadas para o seu tratamento.

QUADRO 2 - Tipos de dados

Tipo de dado	Características	Medidas estatísticas apropriadas		Exemplo
		Medida de Tendência Central	Medida de dispersão	
Nominal	-Qualitativo; -Dispostos em categorias de acordo com os seus atributos.	Moda	% na classe modal	Presença ou ausência de produção pecuária em fazendas; Tipos de climas; Sexo.
Ordinal	-Qualitativo; -Os dados podem ser agrupados de acordo com alguma hierarquia; -Determinadas operações matemáticas não possuem sentido.	Mediana	Percentis	Cidade : "pequena" "média" "grande"; Nível escolar. Divisão de classes sociais: "A", "B", "C", etc.
Intervalo	-Quantitativo; -Os dados podem assumir qualquer valor de um intervalo de números reais; -Não existe o zero absoluto; -Aceita a operação de diferença; -As razões não têm sentido.	Média aritmética	Desvio padrão	Temperaturas
Razão	-Quantitativo; -Os dados podem assumir qualquer valor dentro de um intervalo de números reais; -As operações matemáticas de diferença e razão fazem sentido; -Existe o zero absoluto.	Média geométrica	Coefficiente de variação	Taxa de crescimento populacional Densidade demográfica Áreas. Tempo.

Observação: Quadro adaptado a partir de GERARDI & SILVA (1981, p.28)

2 - APRESENTAÇÃO DE DADOS

2.1 - Grupamento de dados

Neste capítulo serão discutidas técnicas que permitem determinar as principais características dos dados, perceber relacionamentos e, até mesmo, detectar possíveis erros ocorridos no processo de coleta de dados. Vamos organizar os dados em grupos, construir tabelas denominadas distribuições de frequências e apresentar estas informações em gráficos.

Coletar informações é uma coisa, fazer uso é outra. Como os conjuntos de dados geográficos são muito grandes, algumas formas de sínteses matemáticas são normalmente solicitadas na transmissão das informações para outros pesquisadores. (O'BRIEN, 1992)

Algumas etapas importantes a serem cumpridas:

- Determinação da amplitude do conjunto de dados
- Definir número de classes
- Calcular amplitude de cada classe
- Determinar os intervalos
- Efetuar a contagem dos dados
- Montar tabela distribuição de frequência

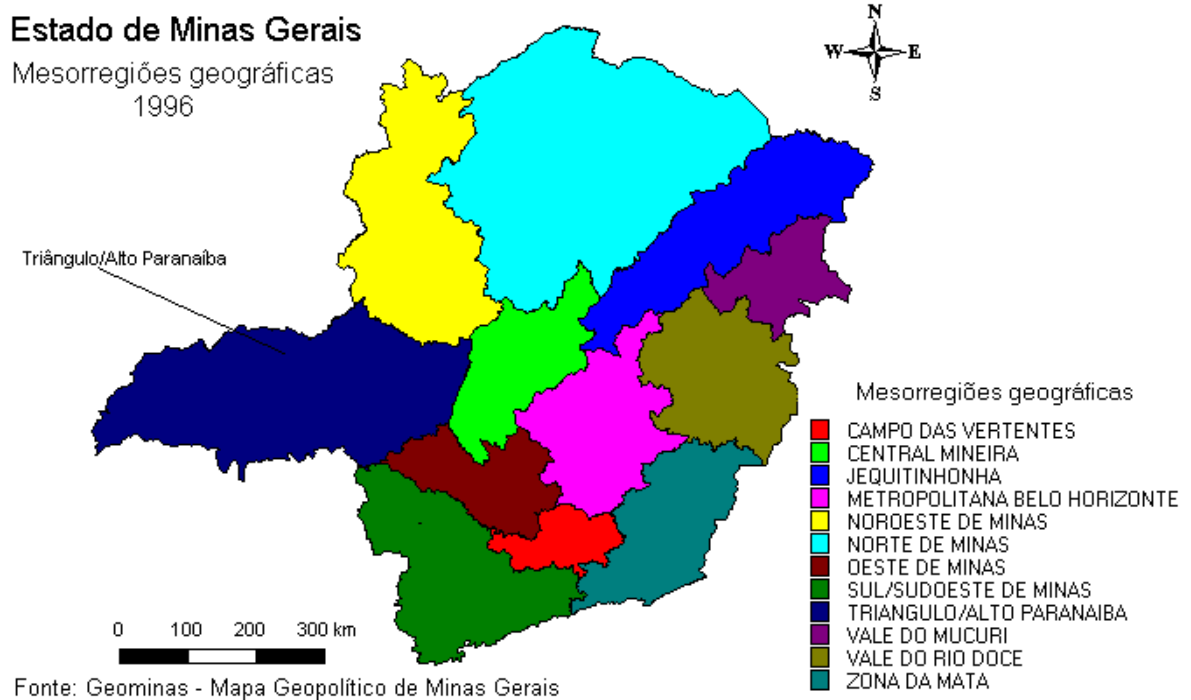
Para discutir as etapas acima, serão considerados os dados sobre o índice de criminalidade dos 64 municípios do estado de Minas Gerais, localizados na mesorregião do Triângulo Mineiro e Alto Paranaíba, no ano de 1995, expostos na TABELA1.

TABELA 1
Índices de criminalidade dos municípios do Triângulo Mineiro e Alto Paranaíba no ano de 1995

i	Municípios	Índice de criminalidade	i	Municípios	Índice de criminalidade
1	Abadia dos Dourados	17,28	33	Itapagipe	12,63
2	Água Comprida	31,72	34	Ituiutaba	24,31
3	Araguari	31,57	35	Iturama	25,11
4	Araporã	29,23	36	Lagoa Formosa	13,59
5	Arapuá	8,73	37	Limeira do Oeste	19,33
6	Araxá	25,52	38	Matutina	11,78
7	Cachoeira Dourada	43,86	39	Monte Alegre de Minas	17,04
8	Campina Verde	19,22	40	Monte Carmelo	38,59
9	Campo Florido	15,59	41	Nova Ponte	21,31
10	Campos Altos	20,89	42	Patos de Minas	28,61
11	Canápolis	10,39	43	Patrocínio	23,57
12	Capinópolis	18,14	44	Pedrinópolis	4,67
13	Carmo do Paranaíba	15,8	45	Perdizes	20,98
14	Carneirinho	14,71	46	Pirajuba	20,68
15	Cascalho Rico	13,24	47	Planura	42,06
16	Centralina	13,29	48	Prata	14,22
17	Comendador Gomes	7,13	49	Pratinha	7,86
18	Conceição das Alagoas	35,05	50	Rio Paranaíba	18,39
19	Conquista	6,66	51	Romaria	23,11
20	Coromandel	25,86	52	Sacramento	16,74
21	Cruzeiro da Fortaleza	8,17	53	Santa Juliana	9,29
22	Douradoquara	11,51	54	Santa Rosa da Serra	4,5
23	Estrela do Sul	10,48	55	Santa Vitória	18,13
24	Fronteira	32,45	56	São Francisco de Sales	30,01
25	Frutal	24,67	57	São Gotardo	9,87
26	Grupiara	39,01	58	Serra do Salitre	17,08
27	Guimarânia	8,64	59	Tapira	13,63
28	Gurinhatã	9,57	60	Tiros	9,59
29	Ibiá	22,85	61	Tupaciguara	30,55
30	Indianópolis	14,63	62	Uberaba	42,34
31	Ipiaçu	29,54	63	Uberlândia	35,96
32	Iraí de Minas	16,29	64	Veríssimo	30,52

Fonte: Site do Geominas, conforme levantamento da PMMG. Acesso em 05 jul 2001

O MAPA 1, apresenta as mesorregiões geográficas do estado de Minas Gerais, dando destaque à mesorregião do Triângulo Mineiro e Alto Paranaíba, estudadas no exemplo dos índices de criminalidade.



MAPA 1: Mesorregiões do estado de Minas Gerais



Observe que, analisando os dados , na forma como estão expostos na TABELA1, pouca ou nenhuma informação pode se obtida. Portanto, estes dados devem ser processados.

A primeira pergunta a ser feita pode ser: “Qual o maior e qual o menor índice de criminalidade encontrado?”

Analisando a TABELA1, constata-se que o maior índice 43,86 é o de Cachoeira Dourada e o menor índice, de Santa Rosa da Serra, 4,50.

Logo, nossos dados variam de 4,50 a 43,86. Ou seja a **amplitude, A**, (diferença entre o maior e menor valor) dos nossos dados é $A = 43,86 - 4,50 = 39,36$.



[5.2]

Procurando entender o significado da amplitude: Esta medida apresenta o “tamanho” do intervalo onde os dados se encontram. Ou seja, neste caso, os dados se encontram num intervalo de tamanho 39,36. Quanto maior a amplitude, maior o intervalo onde os dados estão distribuídos. Mas uma grande amplitude, não indica, necessariamente, que os dados variam muito. Pode-se encontrar um conjunto de dados que, por natureza, já são valores muito altos, mas bem concentrados. Analogamente, pode-se pensar nos casos das amplitudes pequenas.

Como exemplo, considere duas situações extremantes, que tomam uma mesma unidade de medida como referência : a primeira onde tenham sido feitas medidas das distâncias dos pontos turísticos de uma cidade à prefeitura desta mesma cidade. E outra em que se mede a distância entre as cidades de cunho turístico de um estado e sua capital. É evidente que os números encontrados na segunda situação serão bem maiores do que na primeira, mas isto não significa que a distribuição dos dados seja menos uniforme. Ou seja, termos como grande ou pequeno pedem um referencial.

Voltando ao processamento dos dados, pode-se, então, pensar em dividi-los em grupos, com o objetivo de concentrar as informações e destacar os detalhes mais importantes. Mas em quantos grupos, classes, estes dados devem ser divididos? Na bibliografia estatística, encontram-se sugestões para que este número varie de 5 a 25 classes, veja SOARES et al. (1991) e STEVENSON (1986). É sugerido um número não muito grande de classes, pois, corre-se o risco de ficar perdido com a quantidade de números apresentados, diminuindo, assim, a capacidade de

visualização das informações. Já um número muito pequeno de classes poderia reduzir excessivamente a informação causando grande perda de detalhes.

Para encontrar um número razoável de classes, uma opção é a utilização da fórmula de Sturges, que fornece estimativa para o **número de classes, k** , a ser utilizado. Esta fórmula deve ser vista como uma sugestão, não sendo inflexível.

Fórmula de Sturges para cálculo do número de classes

$$k = 1 + \log_2 N, \tag{2.1.1}$$

sendo N o número de observações.

Mas, a maioria das calculadoras científicas efetua apenas cálculos de logaritmos de base 10 ou neperianos. Aplicando a propriedade da mudança de base dos logaritmos, pode-se escrevê-la na forma $k = 1 + \frac{\log N}{\log 2}$ (2.1.2), aplicando, assim, apenas logaritmos de base 10, mais comum nas calculadoras.

Para o exemplo dos índices de criminalidade, tem-se:

Número de observações: $N = 64$

Número de classes: $k = 1 + \frac{\log 64}{\log 2} = 1 + 5,999... = 6,999... \text{ classes.}$

Arredondando este número, obtém-se a sugestão de $k = 7$ classes.



[5.3]

Portanto, para este exemplo, será feita a opção por dividir o conjunto de dados em 7 classes de mesmo tamanho. Mas qual o intervalo de cada classe ?

A **amplitude da classe** será indicada por **h**. Para obter seu valor, basta dividir a amplitude dos dados (A) pelo número de classes (k).

$$h = \frac{A}{k} \text{ (tamanho do intervalo)} \quad (2.1.3)$$

Logo, calculando o valor de h, encontra-se::

$$h = \frac{39,36}{7} \cong 5,6229.$$

A próxima etapa é determinar os limites de cada classe. Para tanto, é importante lembrar que estes limites devem se tocar mas não interceptar. Ou seja, cada um dos dados deve pertencer a uma única classe.

Classes
4,50 ↦ 10,12
10,12 ↦ 15,74
15,74 ↦ 21,36
21,36 ↦ 26,98
26,98 ↦ 32,60
32,60 ↦ 38,22
38,22 ↦ 43,84

O símbolo ↦ indica que o intervalo é fechado no valor inicial e aberto no valor final. Ou seja, um índice de criminalidade igual a 4,50, da cidade de Santa Rosa da Serra, pertence à classe 4,50 ↦ 10,12. Mas, caso existisse, entre os dados um índice de 10,12, este pertenceria à segunda classe 10,12 ↦ 15,74.

Observe que, por motivo de arredondamento no cálculo da amplitude da classe, o final da última classe é menor que o maior dos índices encontrados no conjunto dos dados. Ou seja, o índice igual a 43,86 está sendo excluído da análise. Para evitar

que isto ocorra, deve-se efetuar o arredondamento da amplitude da classe sempre para cima.

Seguindo este raciocínio, pode-se pensar em usar $h=5,63$ e o final da última classe deve ser calculado como apresentado abaixo:

Final da última classe = Início da primeira classe + (número de classes) × (amplitude da classe)

$$\text{Final da última classe} = 4,50 + 7 \times 5,63$$

$$\text{Final da última classe} = 43,91.$$



Por motivo deste arredondamento, o limite superior da última classe ficou muito grande, deixando uma parte significativa desta classe sem função, de acordo com o conjunto de dados. Mas, como é interessante que todas as classes tenham a mesma amplitude, sugere-se uma adequação, reduzindo o valor inicial da primeira classe.

Com isto, todas as classe serão alteradas. O resultado é que aquela parcela “inútil” que existia na última classe, será dividida com a primeira classe. E, finalmente, fica definida a divisão de classes:

Classes	
4,48	↔ 10,11
10,11	↔ 15,74
15,74	↔ 21,37
21,37	↔ 27,00
27,00	↔ 32,63
32,63	↔ 38,26
38,26	↔ 43,89

Após a determinação das classes, parte-se para a contagem dos dados pertencentes a cada uma delas. A tabela, portanto, será ampliada sendo incluída mais uma coluna: a coluna da frequência absoluta, indicada por n_i .

Confira a contagem e confirme o que está exposto na tabela abaixo.

i	Classes	n_i
1	4,48 \mapsto 10,11	12
2	10,11 \mapsto 15,74	13
3	15,74 \mapsto 21,37	15
4	21,37 \mapsto 27,00	8
5	27,00 \mapsto 32,63	9
6	32,63 \mapsto 38,26	2
7	38,26 \mapsto 43,89	5
	Soma	64

Portanto, existem 12 cidades com índice de criminalidade no intervalo 4,48 \mapsto 10,11, 13 cidades com índice de criminalidade no intervalo 10,11 \mapsto 15,74 e assim, sucessivamente. Observe que a soma dos n_i 's deve ser igual ao número de observações.



Como medir a representatividade de alguma classe em relação às observações? Para ajudar a responder a esta pergunta, pode-se calcular o valor da **frequência relativa da classe**, f_i . Ou seja, o tamanho da amostra é considerado como o inteiro e procura-se determinar que fração deste inteiro determinada classe representa. Para tanto, aplica-se uma regra de três simples.

Veja o exemplo para a primeira classe:

Número de dados	fração
64	1 (N ^o de observações)
12	x (Primeira classe)

$$\frac{64}{12} = \frac{1}{x}$$

$$64 \cdot x = 1 \cdot 12$$

$$x = \frac{12}{64} = 0,19.$$

Ou seja, o número de dados da primeira classe representa 0,19 (ou 19%) do número total de observações.

De modo geral, temos:

Número de dados	fração
N	1 (N ^o de observações)
n _i	x (Primeira classe)

$$\frac{N}{n_i} = \frac{1}{x}$$

$$N \cdot x = 1 \cdot n_i$$

$$x = \frac{n_i}{N}. \tag{2.1.4}$$

Fazendo o mesmo cálculo para as demais classes, cria-se mais uma coluna para a tabela. Observe que o somatório das freqüências relativas deve ser igual a 1, pois, ao efetuar a soma das partes, deve-se obter o inteiro.

i	Classes	n _i	f _i
1	4,48 ↔ 10,11	12	0,19
2	10,11 ↔ 15,74	13	0,20

3	15,74 ↦ 21,37	15	0,23
4	21,37 ↦ 27,00	8	0,13
5	27,00 ↦ 32,63	9	0,14
6	32,63 ↦ 38,26	2	0,03
7	38,26 ↦ 43,89	5	0,08
Somam		64	1



[5.6]

Outras informações que podem ajudar em uma análise dos dados são a **Freqüência Acumulada da classe, N_i** , e a **Freqüência Relativa Acumulada da classe, F_i** .

A freqüência acumulada da classe representa a quantidade de dados que pertencem a uma dada classe ou estão abaixo dela. Por exemplo, para a terceira classe (15,74 ↦ 21,37), temos: $N_3 = 12 + 13 + 15 = 40$. Ou seja, existem 40 dados na terceira classe ou abaixo dela (nas 1ª e 2ª classes).

De modo análogo ao que foi feito para a construção da coluna das Freqüências Relativas, pode-se determinar a Freqüência Relativa Acumulada, F_i da classe em relação ao conjunto de observações. Ou seja:

$$F_i = \frac{N_i}{N} . \quad (2.1.5)$$

Portanto, para a terceira classe:

$$F_3 = \frac{N_3}{N} = \frac{40}{64}$$

$$F_3 = 0,63.$$

Este resultado indica que 63% dos dados estão na terceira classe ou abaixo dela.

E, efetuando cálculos semelhantes, as colunas N_i e F_i serão adicionadas à tabela:

i	Classes	n_i	f_i	N_i	F_i
1	4,48 ↪ 10,11	12	0,19	12	0,19
2	10,11 ↪ 15,74	13	0,20	25	0,39
3	15,74 ↪ 21,37	15	0,23	40	0,62
4	21,37 ↪ 27,00	8	0,13	48	0,75
5	27,00 ↪ 32,63	9	0,14	57	0,89
6	32,63 ↪ 38,26	2	0,03	59	0,92
7	38,26 ↪ 43,89	5	0,08	64	1
Sommas		64	1		

Apenas para reforçar os conceitos, observe que $N_4 = 48$ e $F_4 = 0,75$. Isto significa que 48 municípios da amostra possuem índice inferior a 27,00, o que representa 75% da amostra. Ou ainda, $100\% - 75\% = 25\%$ dos municípios apresentam índice de criminalidade maior ou igual a 27,00. Experimente efetuar esta mesma análise para as outras classes.



[5.7]

E, finalmente, para concluir a tabela denominada Tabela de Distribuição de Freqüência, falta acrescentar o **ponto médio de cada classe, x_i** . Este valor poderá se tornar o representante da classe em vários cálculos de medidas estatísticas, como será visto mais adiante.

Para obter o ponto médio da classe basta somar o limite inferior da classe com seu o limite superior e dividir o resultado por dois (média aritmética simples):

$$x_i = \frac{\text{limite inferior da classe} + \text{limite superior da classe}}{2} \tag{2.1.6}$$

Para a primeira classe, tem-se: $x_1 = \frac{4,48 + 10,11}{2} = 7,295$.

Fazendo o cálculo para as demais classes, fica concluída a construção da Tabela de Distribuição de Freqüência.

TABELA 2
Distribuição de frequência dos índices de criminalidade no
ano de 1995 dos municípios da mesorregião do Triângulo
Mineiro e Alto Paranaíba

i	Classes	x_i	n_i	f_i	N_i	F_i
1	4,48 \mapsto 10,11	7,295	12	0,19	12	0,19
2	10,11 \mapsto 15,74	12,925	13	0,20	25	0,39
3	15,74 \mapsto 21,37	18,555	15	0,23	40	0,62
4	21,37 \mapsto 27,00	24,185	8	0,13	48	0,75
5	27,00 \mapsto 32,63	29,815	9	0,14	57	0,89
6	32,63 \mapsto 38,26	35,445	2	0,03	59	0,92
7	38,26 \mapsto 43,89	41,075	5	0,08	64	1,00
Somadas			64	1,00		

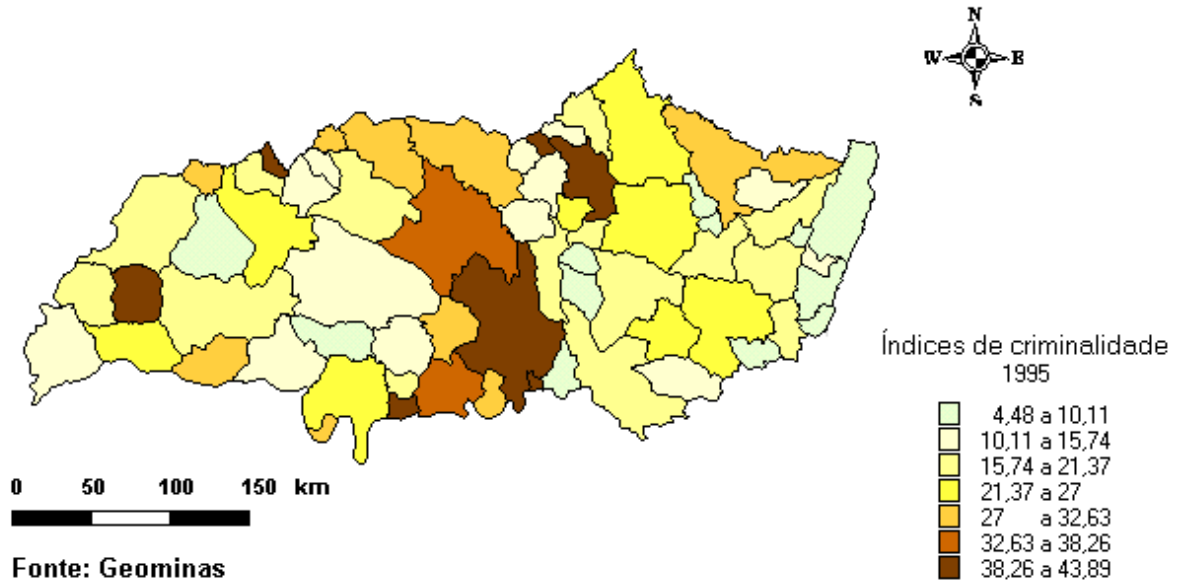
Fonte: Site do Geominas/PMMG. Acesso em 05 jul 2001



[5.8]

O MAPA 2 é exemplo do uso de outros recursos computacionais disponíveis, que podem complementar o trabalho do geógrafo na análise de dados. Sugere-se o estudo dos mapas temáticos, pois estes favorecem a visualização da distribuição dos fenômenos estudados.

Índices de criminalidade nas mesorregiões geográficas do Triângulo Mineiro e Alto Paranaíba - 1995



MAPA 2 : Índices de criminalidade nas mesorregiões do Triângulo Mineiro e Alto Paranaíba - 1995
 Mapa elaborado com uso do software Mapinfo.

2.2 - Gráficos

Outro importante recurso para apresentação de dados é o uso de gráficos. Um gráfico bem escolhido e bem trabalhado destaca as informações mais importantes e a forma de distribuição dos dados. Ou seja, a representação gráfica fornece uma visão de conjunto mais rápida que a observação do conjunto de dados no seu estado bruto ou na forma tabular.

Os gráficos mais usados na Estatística são:

- Histograma
- Polígono de frequência
- Ogiva
- Gráfico em linhas
- Gráfico em barras (ou em colunas)

- Gráfico em setores

2.2.1 - Histograma

O Histograma é a representação gráfica da tabela de distribuição de frequência. No eixo das abscissas (dos x's) são colocados os intervalos de classe, podendo também apresentar os pontos médios de cada classe. No eixo das ordenadas (dos y's) ficam dispostas as frequências absolutas ou relativas de cada classe.

Veja o Histograma obtido com os dados da TABELA 2:

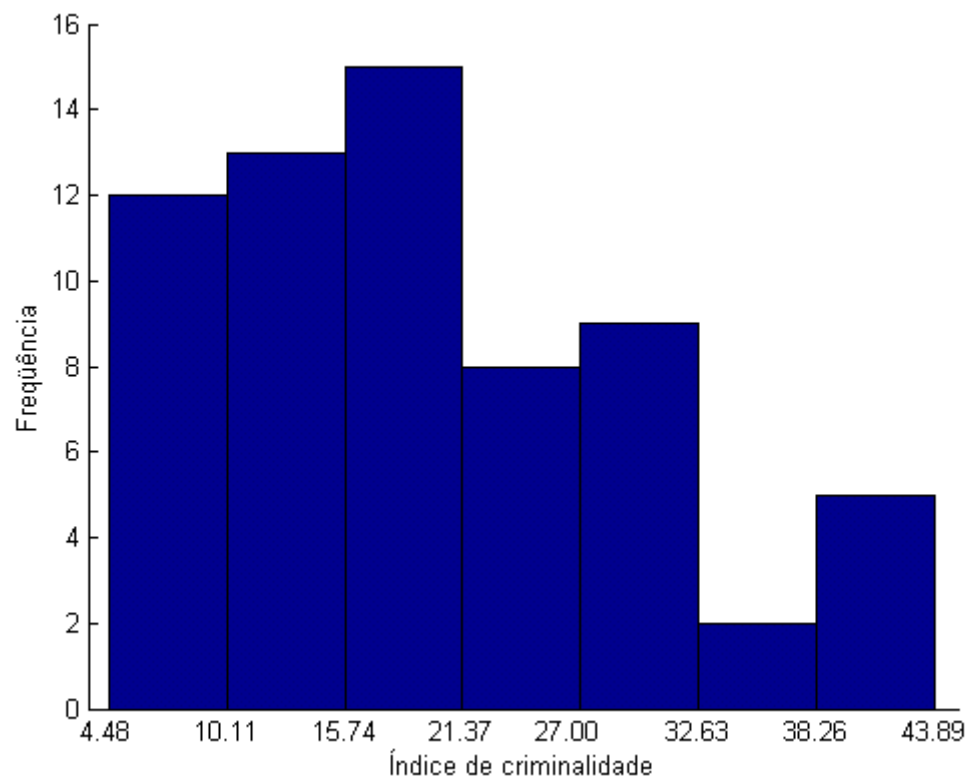


GRÁFICO 1: Histograma para a variável Índice de criminalidade



5.9 (Comandos básicos para construção dos gráficos)

2.2.2 - Polígono de frequência

O polígono de freqüência é obtido dispondo os pontos médios de cada classe no eixo das abscissas e as freqüências relativas ou absolutas no eixo das ordenadas. Os pontos obtidos são ligados por segmentos de reta. Antes da primeira classe e ao final da última classe são criados pontos médios de uma classe com freqüência nula, com o objetivo de fechar a linha poligonal.

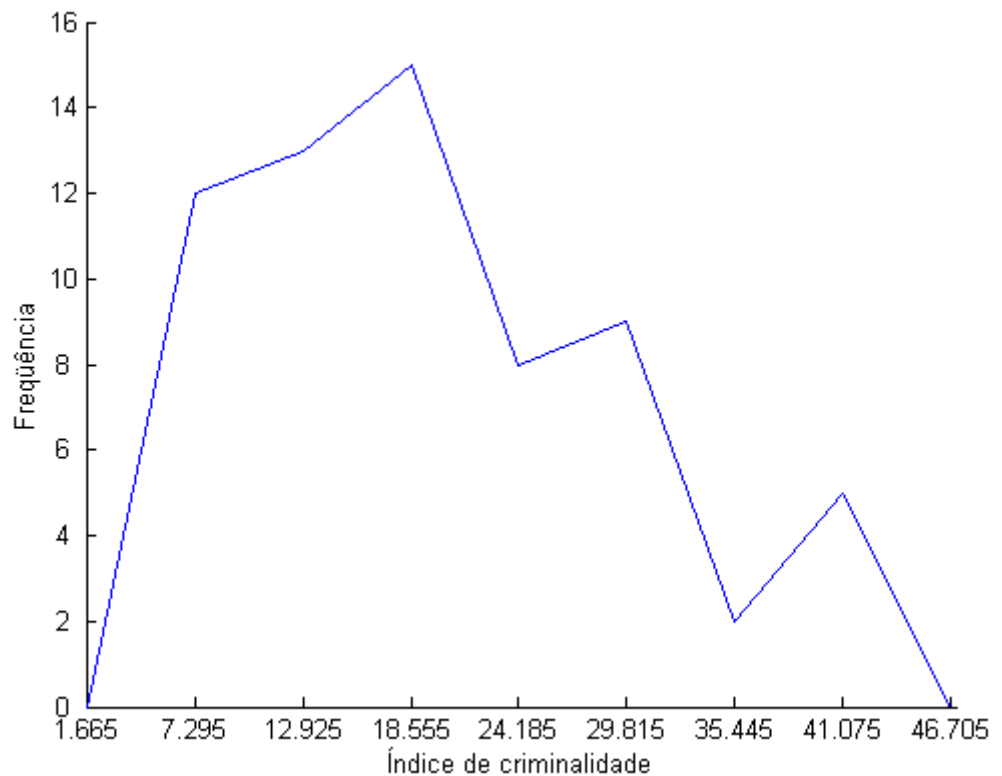


GRÁFICO 2: Polígono de freqüência para a variável Índice de criminalidade

É comum encontrar o Histograma construído junto com o Polígono de Frequência, conforme GRÁFICO 3, para enfatizar a forma de distribuição dos dados.

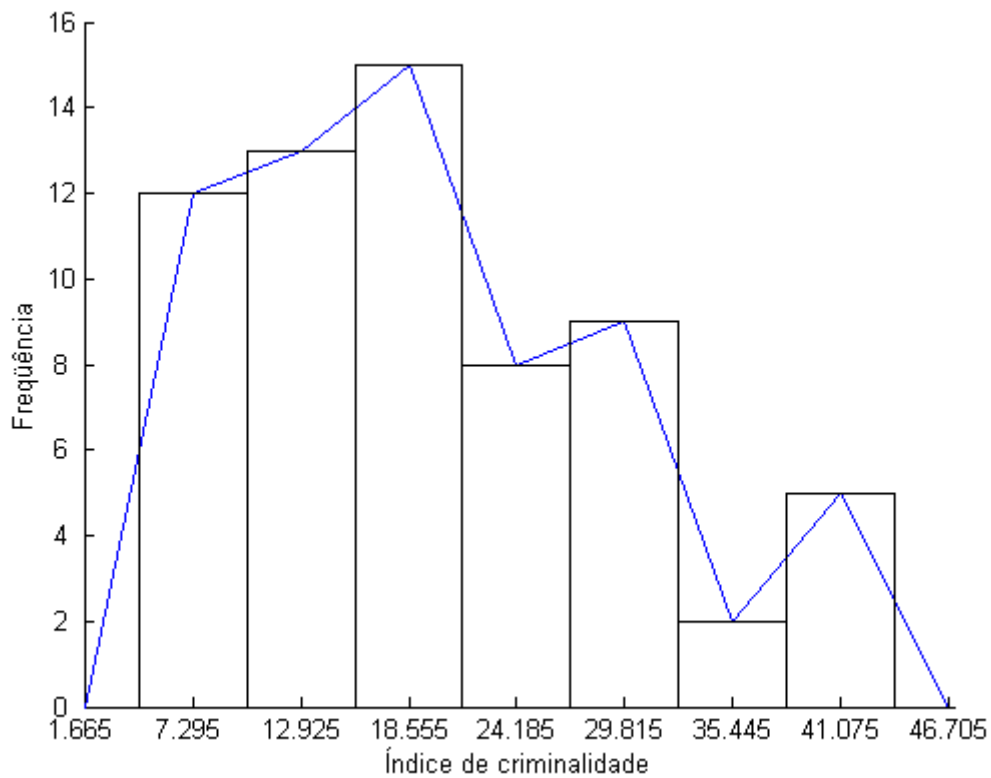


GRÁFICO 3: Histograma e Polígono de Frequência para a variável índice de criminalidade

2.2.3 - Ogiva

O gráfico que utiliza a distribuição de frequência acumulada recebe o nome de Ogiva. Observe que é uma poligonal ascendente. No eixo das abscissas ficam dispostos os limites das classes e no eixo das ordenadas as frequências acumuladas.

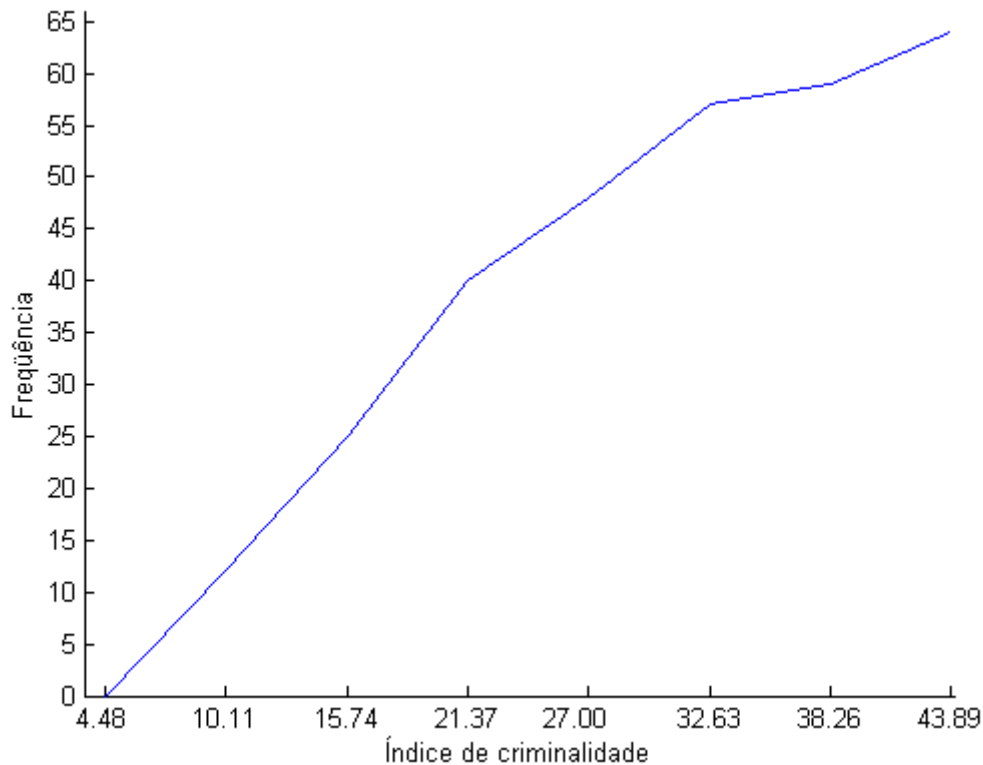


GRÁFICO 4: Ogiva para a variável Índice de criminalidade

2.2.4 - Gráfico em linha

O gráfico em linha é muito usado para dados temporais, representando observações feitas ao longo do tempo. Com este gráfico pode-se estudar a evolução das frequências dos valores de uma variável durante um certo período. A inclinação de cada segmento sinaliza a intensidade do crescimento ou do decréscimo.

Exemplo 2.1: Conforme dados do Dieese, a TABELA 3 mostra o quanto gastou um trabalhador com a parte de alimentação da cesta básica no primeiro semestre de 2001, em Belo Horizonte:

TABELA 3
Gasto do trabalhador com alimentação

Meses	Gasto (R\$)
Janeiro	116,46
Fevereiro	114,61
Março	118,83
Abril	127,38
Maio	123,72
Junho	115,91

Fonte: Dieese. Acesso em 04 jan 2002

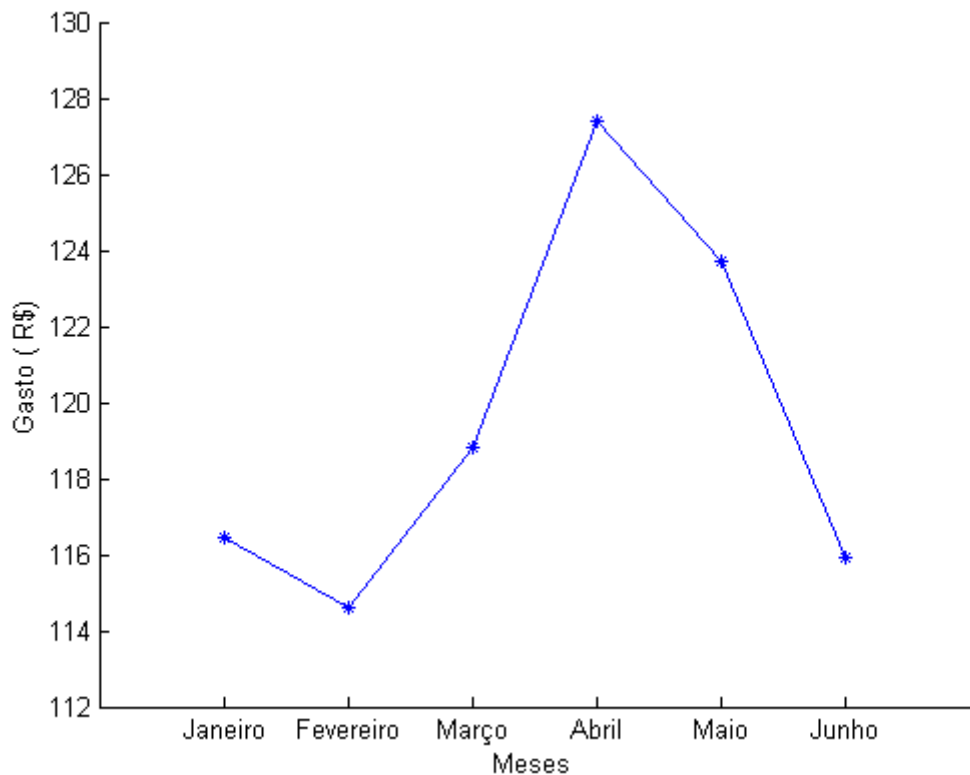


GRÁFICO 5: Gráfico em linha para o gasto do trabalhador com alimentação

No gráfico, os pontos são unidos por uma poligonal apenas para facilitar a visualização. Na realidade, não há observações intermediárias.

2.2.5 - Gráfico em Barras (ou em Colunas)

Quando os dados são provenientes de contagens e os valores distintos não são muito numerosos – como o número de filhos por famílias, o número de televisores por residência, número de questões corretas em uma prova, etc - os próprios valores são usados como “classes” . Assim, o gráfico em Barras é construído colocando cada um dos valores no eixo das abscissas e suas respectivas freqüências no eixo das ordenadas, sendo, a altura das barras, proporcional à freqüência de cada valor.

Exemplo 2.2: Construir o gráfico em barras para as freqüências do número de leitos ocupados por dia em uma determinada enfermaria hospitalar de 6 leitos, num dado mês.

TABELA 4
Número de leitos ocupados por dia em enfermaria

Nº de leitos ocupados	n_i
0	1
1	1
2	3
3	4
4	7
5	8
6	6

Fonte: Dados hipotéticos

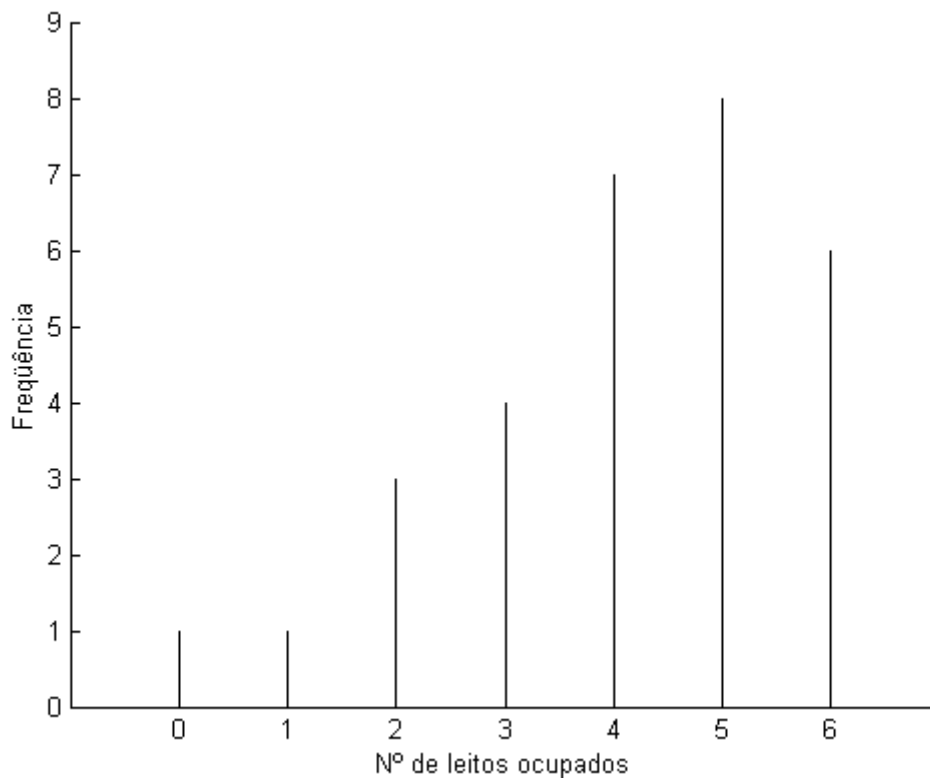
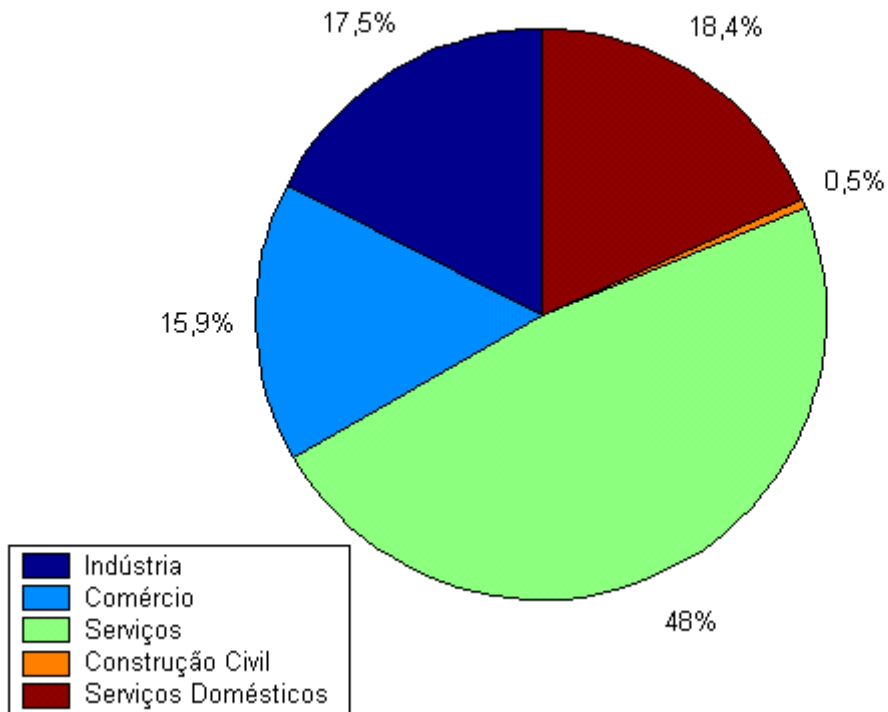


GRÁFICO 6: Gráfico em barras para a variável número de leitos ocupados

2.2.6 - Gráfico em setores

No gráfico em setores o círculo indica o total das observações e cada setor indica a representatividade de uma determinada classe. Na construção deste gráfico o ângulo correspondente a cada setor é obtido através de regra de três simples. Forma-se assim um círculo dividido em setores com áreas proporcionais às frequências das várias classes.

Exemplo 2.3: Observe a distribuição das ocupações das mulheres, segundo setor de atividade na região metropolitana de São Paulo - 1996 - (em %), conforme dados do Dieese.



*GRÁFICO 7: Gráfico em setores para a variável ocupações das mulheres, segundo setor de atividade na região metropolitana de São Paulo - 1996
Fonte: Site do Dieese. Acesso em 04 jan 2002*

Observe o cálculo do ângulo central obtido para a componente indústria:

Ângulo central	Porcentagem
360°	100
x	17,5

$$\frac{360}{x} = \frac{100}{17,5}$$

$$x = \frac{360 \times 17,5}{100}$$

$$x = 63^\circ$$

Como comentário final, vale dizer que, no estudo geográfico, podem ser observados muitos gráficos em que são usados dois eixos verticais. Um apresentando a

freqüência absoluta e outro apresentando a freqüência relativa. Veja como fica no caso do Polígono de Freqüência do exemplo dos índices de criminalidade:

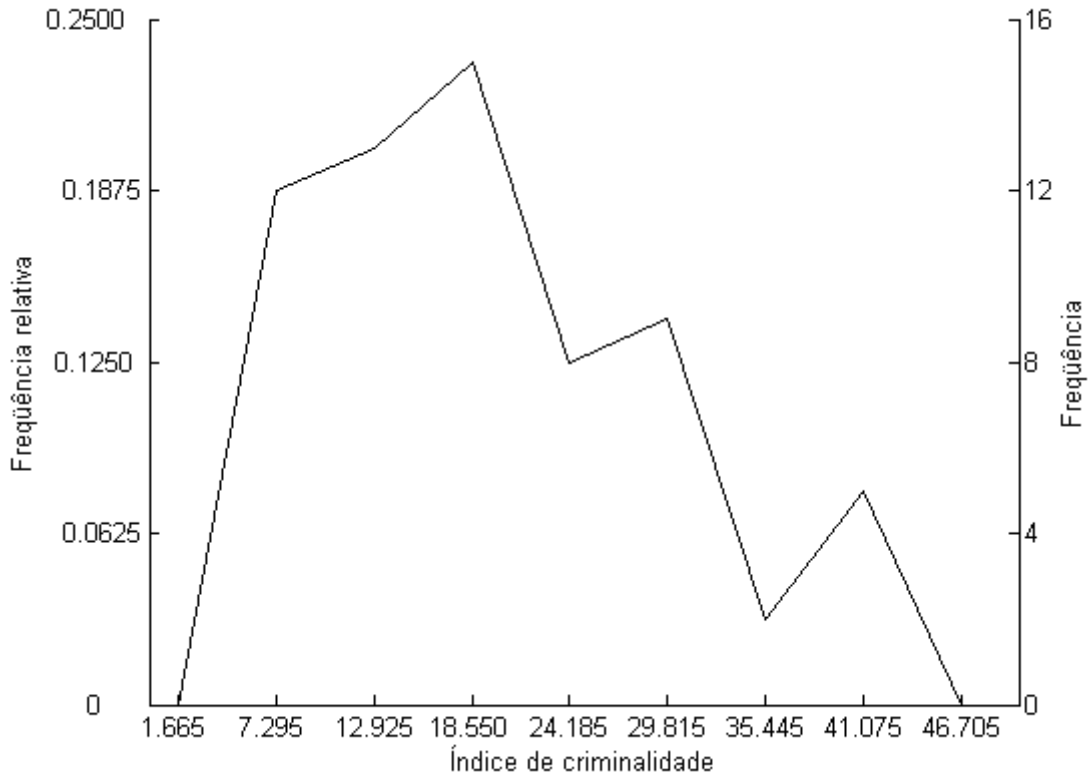


GRÁFICO 8: Gráfico com dois eixos verticais para a variável Índice de criminalidade

3 - MEDIDAS ESTATÍSTICAS

Nesta unidade desenvolve-se o estudo de medidas que buscam sintetizar as características de um conjunto de observações usando uma pequena quantidade de parâmetros.

Este estudo será dividido em duas partes: Num primeiro momento serão discutidas as medidas estatísticas para pequenos grupos de dados, buscando “sentir” e entender o significado de cada medida. Em seguida, será desenvolvido basicamente o mesmo trabalho, mas, desta vez, com os dados agrupados. Ou seja, será feito o estudo de como obter as mesmas medidas estudadas na primeira etapa, no caso de estarem à disposição apenas as tabelas de distribuição de frequência.

3.1 - Medidas de Tendência Central

As Medidas de Tendência Central são calculadas como resposta à busca de parâmetros que concentrem em poucos números as características do conjunto de dados em estudo. As medidas mais importantes são: Média, Mediana e Moda.

3.1.1 - Média Aritmética Simples

É a Medida de Tendência Central mais conhecida, sendo largamente difundida e utilizada pelos meios de comunicação mais diversos. Pode ser encontrada em situações esportivas, resultados de notas de estudantes, temperaturas, índices econômicos como a inflação, etc.

Por possuir certas propriedades matemáticas convenientes, pode-se citá-la como a mais importante das medidas conhecidas.

A Média Aritmética Simples, também chamada simplesmente de média, de um conjunto de n valores observados x_1, x_2, \dots, x_n é dada pelo resultado da divisão por n da soma destes valores.

Indica-se a média por \bar{x} . Logo,

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}. \quad (3.1.1.1)$$

Exemplo 3.1: As temperaturas mínimas, em graus centígrados, registradas nos postos de observação do Instituto Nacional de Meteorologia no estado de Santa Catarina em 18 de outubro de 2001 estão dispostos na TABELA 5, extraídas do MAPA 3, produzido pelo Instituto Nacional de Meteorologia:

TABELA 5
Temperaturas mínimas registradas nos postos de observação do Instituto Nacional de Meteorologia no estado de Santa Catarina em 18 de outubro de 2001

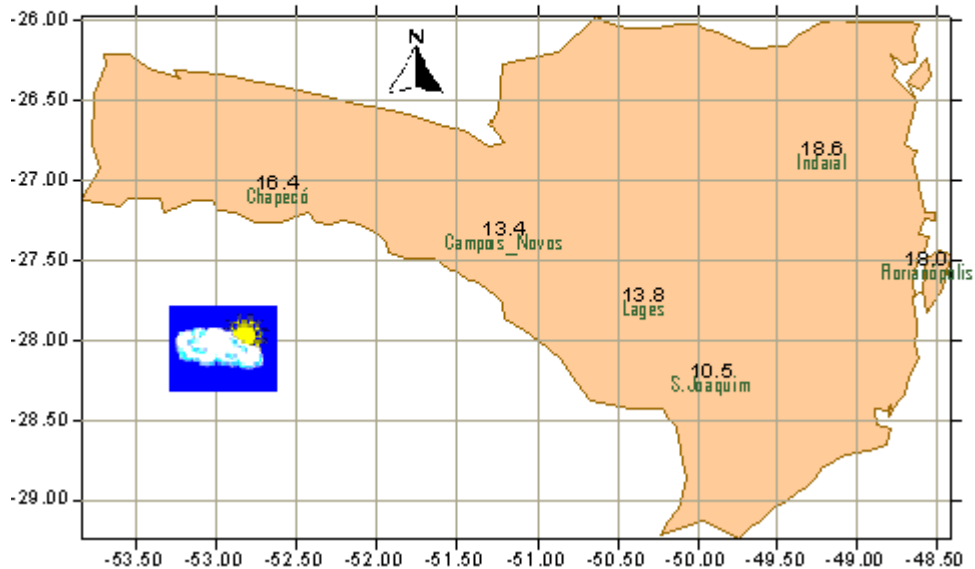
Posto de observação	Temperatura Mínima (°C)
Chapecó	16,4
Campos Novos	13,4
Lages	13,8
São Joaquim	10,5
Indaial	18,6
Florianópolis	18,0

Fonte: Instituto Nacional de Meteorologia. Acesso em 19 out 2001

Santa Catarina

Temperatura Mínima (°C)

Data: 18/10/2001



MAPA 3: Temperatura mínima nos pontos de controle de Santa Catarina
 Fonte: Instituto Nacional de Meteorologia. Acesso em 19 out 2001.

Portanto, a temperatura mínima média obtida na região, foi:

$$\bar{x} = \frac{16,4 + 13,4 + 13,8 + 10,5 + 18,6 + 18,0}{6} = 15,1167^{\circ}\text{C}$$



[5.10]

Cada valor exerce influência no cálculo da média aritmética, este fato, que a princípio é uma importante qualidade matemática, pode ser revertido em problema se existirem dados discrepantes. Veja o exemplo:

Exemplo 3.2: Considere a situação hipotética de uma empresa com dez funcionários onde os salários, em Reais, são:

2.500,00	250,00	300,00	2.000,00	200,00
250,00	200,00	320,00	250,00	200,00

A média salarial da empresa será:

$$\bar{x} = \frac{2.500 + 250 + 300 + 2.000 + 200 + 250 + 200 + 320 + 250 + 200}{10} = 647,00.$$

Observe que este número não reflete a realidade salarial da empresa já que 80% de seus funcionários possuem salários inferiores ou iguais a R\$ 320,00. Os valores discrepantes, os dois maiores salários, provocaram um aumento excessivo na média salarial da empresa.

Em situações como esta não é recomendável o uso da média como Medida de Tendência Central de um conjunto de dados. Mais adiante outras medidas mais adequadas serão estudadas.

3.1.2 - Média Aritmética Ponderada

A média aritmética simples toma como princípio o fato de todos os dados possuírem a mesma importância. Realmente, esta situação é muito comum. Porém, em determinadas situações alguns dados têm mais influência sobre o conjunto do que outros. Nestes casos, aplica-se a Média Aritmética Ponderada.

A média ponderada dos números x_1, x_2, \dots, x_n com pesos respectivamente iguais a p_1, p_2, \dots, p_n é dada por

$$\bar{x}_p = \frac{p_1 \cdot x_1 + p_2 \cdot x_2 + \dots + p_n \cdot x_n}{p_1 + p_2 + \dots + p_n} \quad (3.1.2.1)$$

Exemplo 3.3: Considere a situação de um concurso em que são realizadas provas de Matemática, Português e Digitação com pesos respectivamente iguais a 4, 3 e 2.

As notas de dois candidatos estão apresentadas na TABELA 6. Determine o candidato com melhor média final.

TABELA 6
Notas de dois candidatos nas provas de um concurso

	Pesos	Candidato A	Candidato B
Matemática	4	8	6
Português	3	7	8
Digitação	2	6	7

Fonte: Dados Hipotéticos

Candidato A

$$\bar{x}_p = \frac{4 \cdot 8 + 3 \cdot 7 + 2 \cdot 6}{4 + 3 + 2} = 7,222$$

Candidato B

$$\bar{x}_p = \frac{4 \cdot 6 + 3 \cdot 8 + 2 \cdot 7}{4 + 3 + 2} = 6,889$$

Sendo assim, o candidato A obteve o melhor resultado.

Neste exemplo, o peso transmite uma idéia semelhante a de o candidato ter feito 4 provas de Matemática, obtendo notas iguais em todas. O mesmo raciocínio serve para as nota de Português e Digitação. Por este motivo, a soma final é dividida pela soma dos pesos (“quantidade de provas feitas”).

Observe que, se neste concurso, a importância de cada prova fosse a mesma, os candidatos teriam obtido o mesmo resultado final, pois os conjuntos de notas são iguais a { 6, 7, 8 }.

A média aritmética simples pode ser vista como um caso particular da ponderada se usarmos todos os pesos iguais a 1.



Veja um outro exemplo:

Exemplo 3.4: A TABELA 7 apresenta a área e a densidade demográfica de três municípios. Deseja-se estimar a densidade demográfica média da região composta por estes municípios.

TABELA 7
Área e densidade demográfica de municípios

Município	Área (km ²)	Densidade demográfica (hab/km ²)
A	235	15,18
B	812	14,12
C	367	14,19

Fonte: Dados hipotéticos

$$\bar{x}_p = \frac{235 \cdot 15,18 + 812 \cdot 14,12 + 367 \cdot 14,19}{235 + 812 + 367} = 14,31 \text{ hab/km}^2$$

Neste caso, o peso é a área do município, o número de km² onde estão localizados seus habitantes. No caso do município A a cada km² existe em média 15,18 habitantes. Logo, 235×15,18 indica o número total de habitantes.

Portanto, multiplicando a área pela densidade demográfica de cada município e somando estes produtos, obtém-se a população total da região, que deve ser dividida pela área total para encontrar a densidade demográfica média .

3.1.3 - Média Geométrica

A Média Geométrica é uma medida pouco encontrada na bibliografia estatística, por ser usada em situações muito particulares. O uso da Média Geométrica pode ser observado na determinação da taxa média de crescimento populacional, nas situações de juros compostos e em outras situações onde as atualizações dos dados envolvam correção sobre correção.

A média geométrica simples dos n números positivos, x_1, x_2, \dots, x_n é definida como o valor positivo g tal que $x_1 \cdot x_2 \cdot \dots \cdot x_n = g \cdot g \cdot \dots \cdot g = g^n$. Portanto, a média geométrica simples dos n números positivos x_1, x_2, \dots, x_n é calculada da seguinte forma:

$$g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (3.1.3.1)$$

Veja a seguir dois exemplos onde deve-se aplicar o conceito de Média Geométrica.

Exemplo 3.5: Uma empresa aumentou sua produção durante os meses de novembro e dezembro de 2001. Em novembro, a taxa de aumento foi de 11% e, em dezembro, de 23%. Qual a taxa média de aumento mensal nesse bimestre?

Em geral, a primeira idéia é a de calcular a média aritmética das taxas mensais.

Assim,

$$\begin{aligned} \bar{x} &= \frac{11\% + 23\%}{2} \\ \bar{x} &= 17\% \end{aligned}$$

Esta resposta não está correta, pois um aumento mensal de 17% levaria a um aumento maior do que o verificado.

Considere uma produção inicial igual a 100 e veja o resultado ao final dos dois meses:

Aplicando a taxa mensal de 17%:

$$100 \mapsto 100 \cdot 1,17 \mapsto (100 \cdot 1,17) \cdot 1,17 \mapsto 136,89,$$

e assim, o aumento bimestral registrado teria sido de 36,89%.

Aplicando as taxas de novembro e dezembro, indicadas no problema:

$$100 \mapsto 100 \cdot 1,11 \mapsto (100 \cdot 1,11) \cdot 1,23 \mapsto 136,53,$$

resultado que indica um aumento bimestral de 36,53%.

Observe que os resultados encontrados são diferentes. Quanto maior o valor inicial, maior será a diferença absoluta entre os resultados (diferença entre o número de unidades efetivamente produzidas e o número estimado).

Sendo assim, deseja-se encontrar uma taxa média i , tal que, aplicando-a nos dois meses consecutivos, o aumento bimestral será o mesmo que o obtido com a aplicação das taxas de 11% e 23%.

Portanto, aplicando-se a taxa i por dois meses consecutivos

$$100 \mapsto 100 \cdot (1+i) \mapsto 100 \cdot (1+i)^2$$

deve-se obter resultado final igual ao resultado obtido com a aplicação das taxas de 11% e 23%, ou seja:

$$100(1+i)^2 = 100 \cdot 1,11 \cdot 1,23$$

$$(1+i)^2 = 1,11 \cdot 1,23$$

$$1+i = \sqrt{1,11 \cdot 1,23}$$

$$1+i \cong 1,1685,$$

$$\text{e finalmente, } i \cong 0,1685 = 16,85\%$$

Pode-se notar que a taxa média igual a 0,1685, aumentada de uma unidade, corresponde à média geométrica das taxas mensais aumentadas de uma unidade.

Exemplo 3.6 :

A TABELA 8 apresenta a população residente no Brasil no período 1940-1980, conforme dados colhidos no site do IBGE, na seção de dados históricos dos censos.

TABELA 8
População presente no Brasil no período de
1940 a 1980

Anos	População presente
1940	41.236.315,00
1950	51.944.397,00
1960	70.070.457,00
1970	93.139.037,00
1980	119.002.706,00

Fonte: IBGE

Aplicando regra de três simples, podemos calcular a taxa de crescimento em cada período de dez anos. Veja o caso particular do crescimento populacional no período de 1940-1950:

Aumento da população de 1940 a 1950: $51.944.397 - 41.236.315 = 10.708.082$

Habitantes	%
41.236.315	100
10.708.082	x

$$x = \frac{10.708.082 \times 100}{41.236.315} = 25,97\%$$

A TABELA 9 será construída com o objetivo de auxiliar este cálculo para os outros períodos.

TABELA 9
Construção auxiliar para cálculo da taxa de crescimento populacional

Anos	População presente	Varição populacional (em número de habitantes)	Taxa de crescimento da população por 10 anos (%)
1940	41.236.315	-	-
1950	51.944.397	10.708.082	25,97
1960	70.070.457	18.126.060	34,90
1970	93.139.037	23.068.580	32,92
1980	119.002.706	25.863.669	27,77

Fonte: IBGE

Indicando por \bar{c} a média das taxas de crescimento da população, tem-se:

$$\bar{c} = \frac{25,97 + 34,90 + 32,92 + 27,77}{4}$$

$$\bar{c} = 30,39 \%$$

Este resultado pode estar sugerindo que, aplicando sucessivamente a taxa de 30,39% a partir da população existente em 1940, chega-se à população existente em 1980. Faça as contas e verifique que esta observação é falsa.

Calculando a média geométrica das taxas de crescimento, obtém-se:

$$g = \sqrt[4]{1,2597 \cdot 1,3490 \cdot 1,3292 \cdot 1,2777}$$

$$g = 1,3034$$

Aplicando-se sucessivamente a taxa de 30,34%, a partir da população de 1940 chega-se, a não ser por erros de arredondamento, à população de 1980.

Portanto, a taxa média de crescimento populacional, no período foi de 30,34%.

Taxa de crescimento da população

A fórmula simplificada para o cálculo desta taxa de crescimento da população, obtida com uso da definição de média geométrica é dada por:

$$i = \left(\frac{P(t+n)}{P(t)} \right)^{\frac{1}{n}} - 1, \text{ onde} \quad (3.1.3.2)$$

$P(t)$ é a população no início do período

$P(t+n)$ é a população no ano $t+n$

n é o intervalo de tempo entre os dois períodos

i é a taxa de crescimento considerada.

Aplicando esta fórmula no problema anterior, encontra-se:

$$i = \left(\frac{119.002.706}{41.236.315} \right)^{\frac{1}{4}}$$

$$i = 1,3034$$

A taxa média de crescimento populacional, no período foi de 30,34%, resultado igual ao calculado anteriormente.

Esta taxa indica a intensidade de crescimento da população em determinado período, sendo usada para realizar estimativas e projeções populacionais e quantificar as variações do contingente populacional. O crescimento pode estar sendo influenciado por diversos fatores como natalidade, mortalidade e migrações.

Porém, o uso desta taxa possui limitações. Projeções da taxa para anos distantes do censo demográfico mais recente podem não refletir as transformações da dinâmica populacional atual.

3.1.4 - Mediana

A mediana é o elemento “do meio” de um conjunto de valores. Esta medida estatística é muito utilizada para dados ordinais.

Sua principal característica é dividir um conjunto ordenado de valores em dois grupos iguais. Sendo assim, a primeira metade deve conter valores menores ou iguais à mediana e a outra metade deve ser composta de valores maiores ou iguais à mediana.

Em geral, a mediana ocupa a posição $\frac{n+1}{2}$. Assim, em um conjunto com n valores, a mediana será:

- O valor que ocupar a posição central caso n seja ímpar;
- A média dos dois valores centrais se n for par.

Ou seja, se um conjunto de dados tiver 5 elementos, a mediana será o elemento que ocupa a terceira posição. Mas, se o conjunto tiver 6 elementos a mediana será a média obtida entre os valores que ocupam a terceira e quarta posições.

Exemplo 3.7: A umidade relativa do ar foi medida em algumas cidades tendo sido constatados os seguintes resultados, em %, registrados na TABELA 10:

TABELA 10

Umidade relativa do ar

Cidade	Umidade relativa do ar (%)
A	63
B	62
C	81
D	75
E	94
F	97
G	75

Fonte: Dados hipotéticos

Nosso primeiro passo será colocar os dados em ordem crescente:

62 63 75 75 81 94 97

Como o número de valores é $n = 7$ (ímpar), a posição ocupada pela mediana deve ser a quarta $\left(\frac{7+1}{2} = 4\right)$, ou seja, a mediana deste conjunto de valores é igual a 75%.



[5.12]

Exemplo 3.8: Como primeiro passo, para calcular a mediana para os dados do exemplo 3.2, deve-se colocar os salários em ordem crescente:

200,00 200,00 200,00 250,00 250,00 250,00
 300,00 320,00 2.000,00 2.500,00

Como são $n = 10$ valores (um número par), a mediana será obtida calculando a média entre os quinto e sexto termos $\left(\frac{10 + 1}{2} = 5,5 \right)$:

$$\text{Mediana} = \frac{250,00 + 250,00}{2} = \text{R\$ } 250,00.$$

Observe que, neste exemplo, a mediana está mais próxima da maior parte dos salários da empresa do que a média de R\$ 647,00, calculada anteriormente.

Para obter a mediana, os valores devem ser dispostos em ordem crescente (ou decrescente), o que pode provocar um trabalho considerável se n for muito grande. Este trabalho é relativamente simples de ser feito em computadores, mas não é realizado pelas calculadoras científicas mais comuns.

Outra importante observação é o fato de o valor do dado não influenciar tanto na obtenção da mediana, mas sim sua posição no conjunto ordenado. Esta é, sem dúvida, uma das principais características da mediana.

Esta é uma medida muito usada em situações salariais e de valores imobiliários, onde é mais comum encontrar uma pequena quantidade de valores nos extremos (geralmente, valores muito altos em relação ao conjunto de dados).

3.1.5 - Moda

A moda é o valor de maior frequência no conjunto de dados.

A moda é freqüentemente utilizada em situações em que a variedade de valores ou de classes é pequena. Por exemplo, situações que indicam presença ou ausência de determinado fenômeno, questões de verdadeiro ou falso, vitória, derrota ou empate ou para valores que quando agrupados apresentam pequeno número de classes.

A utilização da Moda como Medida de Tendência Central é pouco comum para dados não agrupados.

Um conjunto de dados pode não ter moda, ter apenas uma ou mais de uma moda.

Exemplo 3.9: O conjunto de dados 2, 3, 5, 2, 4, 3, 1, 2 é denominado unimodal pois tem apenas uma moda: o valor 2.



Exemplo 3.10: O conjunto 3, 2, 1, 5, 2, 3, 4, 4, 7 é trimodal, pois possui três modas: 2, 3, e 4.

Exemplo 3.11: O conjunto de dados do exemplo 3.1, da umidade relativa, não possui moda, ou seja, é classificado como amodal.

Exemplo 3.12: O exemplo 3.2, dos salários, apresenta duas modas R\$ 200,00 e R\$ 250,00 sendo, portanto, um conjunto bimodal.

Um Conjunto de dados com mais de 3 modas é denominado multimodal.

3.1.6 - Percentil

Outra medida estatística importante é o Percentil. Enquanto a Mediana é calculada com a divisão do conjunto de dados em duas partes iguais, os percentis usam a divisão do conjunto de dados ordenados em 100 partes iguais. Ou seja, o Percentil 80, P_{80} , separa os dados deixando 80% das informações nele, ou abaixo dele e, obviamente, os 20% restantes nele ou acima dele.

Exemplo 3.13: Calcular o P_{80} para o conjunto de notas em Estatística de 15 estudantes:

10 8 3 6 5 4 8 9 7 5 7 8 5 10 6

Ordenando as notas:

3 4 5 5 5 6 6 7 7 8 8 8 9 10 10

Ao calcular o Percentil 80, será determinada a nota que deixa 80% das demais nela ou abaixo dela. Neste caso, 80% de 15 é igual a 12. Logo, o P_{80} é a 12ª nota, ou seja, 8. Isto significa que 80% dos estudantes obtiveram nota menor ou igual a 8.

Exemplo 3.14: Calcular o Percentil 70 para os dados do exemplo 3.2.

Dispondo os dados em ordem crescente, encontramos:

200,00 200,00 200,00 250,00 250,00 250,00 300,00 320,00 2.000,00 2.500,00

Logo, como 70% de 10 é igual 7, o P_{70} será o 7º dado, ou seja, $P_{70} = R\$ 300,00$.

Portanto, 70% dos funcionários recebem no máximo R\$ 300,00. Realmente, podemos constatar mais uma vez que, a média salarial de R\$ 647,00, obtida anteriormente, não é a medida mais adequada a ser tomada como uma medida de tendência salarial para esta empresa.

Alguns percentis notáveis merecem atenção especial:

Os decis, que dividem os dados em 10 grupos: P_{10} , P_{20} , P_{30} , ... , P_{90} , indicados, respectivamente, por D_1 , D_2 , D_3 , . . . , D_9 . Ou seja, o Decil 4 corresponde ao Percentil 40.

Os quartis, que dividem os dados em 4 grupos: P_{25} , P_{50} , P_{75} , também denotados respectivamente, por Q_1 , Q_2 e Q_3 .

Observe que a Mediana também pode ser indicada por P_{50} , D_5 ou Q_2 .

3.2 - Medidas de dispersão

O estudo das Medidas de Dispersão justifica-se, pois raramente um único número será suficiente para transmitir a essência das características de um conjunto de dados. Mas, determinando a Medida de Tendência Central escolhida, que busca apresentar o “meio” do conjunto de dados e, também, conseguindo exprimir o grau de dispersão destes valores, o contexto estará melhor explicado.

Ou seja, inicia-se o estudo das medidas estatísticas que indicam se os valores de um determinado conjunto estão relativamente próximos uns dos outros, ou se estão dispersos.

Veja o que pode acontecer quando calculadas a média e a mediana dos conjuntos de valores $X = \{ 13, 15, 16, 20, 24, 25, 27 \}$ e $Y = \{ 7, 9, 12, 20, 26, 32, 34 \}$:

Os dois conjuntos apresentam média e mediana iguais a 20. Confira! No entanto, o conjunto Y apresenta um conjunto de valores que estão muito mais “dispersos” do que os do conjunto X. É por situações como essa que se torna necessário calcular medidas que apresentem o grau de dispersão de um conjunto de valores em relação ao valor central. A discussão destas medidas de dispersão, será feita em relação à média, pelas importantes propriedades matemáticas que esta medida apresenta.

As medidas mais importantes são: Amplitude (ou intervalo), Desvio Médio Absoluto, Variância, Desvio Padrão e Coeficiente de Variação.

3.2.1- Amplitude (ou intervalo)

Amplitude é a diferença entre o maior e o menor valor do conjunto de dados, representando, portanto, o “tamanho” do intervalo onde os dados se encontram.

Quanto maior a amplitude, maior o intervalo onde os dados estão distribuídos. Mas, lembre-se: uma grande amplitude, não indica, necessariamente, que os dados variam muito. É possível encontrar dados que, por natureza, já são valores muito altos, mas que estão bem concentrados.

A amplitude é a medida mais simples de calcular. Seu inconveniente é o fato de levar em conta apenas o maior e o menor valor do conjunto de observações. Ou seja, dois conjuntos de valores bastante diferentes, podem apresentar uma mesma amplitude.

Exemplo 3.15: Determinar amplitude das temperaturas mínimas registradas em seis localidades do estado de Santa Catarina no dia 18 de Outubro de 2001, apresentadas na TABELA 5 (ver exemplo 3.1).

Maior valor: 18,6°C

Menor valor: 10,5°C

Logo, a amplitude é igual a 18,6°C – 10,5°C, ou seja, 8,1°C. Este número indica que o conjunto das temperaturas “encaixa-se” num intervalo de tamanho 8,1°C.

3.2.2 - Desvio Médio Absoluto

Já foram discutidos a importância e o processo de determinação do tamanho do intervalo onde os valores “encaixam-se”. Mas fica a pergunta: Qual o grau de dispersão do conjunto de valores dentro deste intervalo?

Para tentar responder a esta pergunta, calcula-se a média dos desvios de cada valor em relação à média (seu ponto central).

Exemplo 3.16: Veja o que acontece com o conjunto de valores:

2 4 7 8 9

Este conjunto apresenta média igual a 6.

Logo, os desvios de cada valor em relação à média são:

$$2 - 6 = -4$$

$$4 - 6 = -2$$

$$7 - 6 = 1$$

$$8 - 6 = 2$$

$$9 - 6 = 3$$

Portanto, o Desvio Médio (DM), ou seja, a média dos desvios é:

$$DM = \frac{-4 - 2 + 1 + 2 + 3}{5} = 0.$$

O fato do Desvio Médio ser igual a zero indica que os valores são todos iguais, não existindo variação? É claro que não. Observe o conjunto de valores deste exemplo.

O problema é que o Desvio Médio, da forma como foi feito e pela definição de Média, será sempre nulo, pelas propriedades da média. Para contornar este problema faz-se a opção por trabalhar com a média dos módulo dos desvios: Desvio Médio Absoluto, denotado por DMA.

$$DMA = \frac{\sum |x_i - \bar{x}|}{n} \tag{3.2.2.1}$$

Calculando o Desvio Médio Absoluto para os dados do exemplo 3.16:

$$|2 - 6| = 4$$

$$|4 - 6| = 2$$

$$|7 - 6| = 1$$

$$|8 - 6| = 2$$

$$|9 - 6| = 3$$

Assim, o valor do Desvio Médio Absoluto (DMA) é:

$$DMA = \frac{4 + 2 + 1 + 2 + 3}{5} = 2,4$$

Logo, se conclui que boa parte dos nossos dados gira em torno da média num raio de tamanho 2,4. No exemplo atual, 60% das informações, encaixam-se neste perfil.



Uma discussão mais apurada sobre o nível de aglutinação dos dados em torno da Medida de Tendência Central escolhida é feita quando se estuda Distribuições de Probabilidade, mas este tópico foge ao escopo deste trabalho.

3.2.3 - Variância e Desvio Padrão

O problema encontrado no cálculo do Desvio Médio Absoluto (DMA) é a utilização do módulo. Esta operação respeita poucas propriedades matemáticas, trazendo certas dificuldades para seu cálculo. Mas, seguindo quase a mesma idéia, faz-se o cálculo de outra Medida de Dispersão importante: a Variância.

Ao calcular o Desvio Médio de qualquer conjunto de valores, o resultado obtido será sempre igual a zero, pelas propriedades da média. Tentando contornar este problema, apresenta-se outra solução que não seja a aplicação do módulo a cada desvio: elevar cada um destes desvios ao quadrado, com o objetivo de trabalhar apenas com valores positivos ou iguais a zero.

Portanto, a **Variância**, denotada por s^2 é definida como a média dos quadrados dos desvios:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} \quad (3.2.3.1)$$

Mas, para trabalhar com esta fórmula, é preciso calcular o desvio de cada um dos valores do conjunto de dados em relação à média, elevar ao quadrado e somar os resultados. Mas, esta fórmula pode ser modificada com o objetivo de simplificar o trabalho de cálculo.

Pelas regras de produtos notáveis temos que:

$$(x_i - \bar{x})^2 = x_i^2 - 2x_i\bar{x} + \bar{x}^2. \quad (3.2.3.2)$$

Logo,

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n} \quad (3.2.3.3)$$

Usando algumas propriedades do somatório, pode-se escrever:

$$s^2 = \frac{\sum x_i^2 - \sum 2x_i\bar{x} + \sum \bar{x}^2}{n} = \frac{\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2}{n} \quad (3.2.3.4)$$

Mas, sendo $\bar{x} = \frac{\sum x_i}{n}$, tem-se que $\sum x_i = n\bar{x}$. Substituindo este resultado em s^2 :

$$s^2 = \frac{\sum x_i^2 - 2\bar{x} \cdot n\bar{x} + n\bar{x}^2}{n} = \frac{\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2}{n} \quad (3.2.3.5)$$

$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n} \quad (3.2.3.6)$$

$$s^2 = \frac{\sum x_i^2}{n} - \frac{n\bar{x}^2}{n} \quad (3.2.3.7)$$

e, portanto:

$$\boxed{s^2 = \frac{\sum x_i^2}{n} - \bar{x}^2} \quad (3.2.3.8)$$

Observe, que esta fórmula, além de ser equivalente à primeira, é bem mais simples de ser trabalhada, já que não exige o cálculo do desvio de cada valor em relação à média.

O inconveniente da variância é o fato de trabalhar com uma medida cuja unidade está ao quadrado, o que dificulta sua interpretação. É possível entender o significado

de m² (metro quadrado), mas, o que significa graus centígrados ao quadrado ou pontos obtidos em uma prova ao quadrado?

Por este motivo, chega-se a outra Medida de Dispersão importante: o **Desvio Padrão**. Denotado por s, o Desvio Padrão é a raiz quadrada da variância, indicando a dispersão em torno da média nas mesmas unidades dos dados originais.

$$s = \sqrt{s^2} \quad \therefore \quad s = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2} \quad (3.2.3.9)$$

Exemplo 3.17: De acordo com os dados da TABELA 10 (veja exemplo 3.7), calcular:
 a) a variância (usando a fórmula original e a melhorada, para efeito de comparação);
 b) o desvio padrão.

Cálculo da média:

$$\bar{x} = \frac{63 + 62 + 81 + 75 + 94 + 97 + 75}{7} = 78,143$$

TABELA 11

Construção auxiliar para cálculo da variância e do desvio padrão da umidade relativa do ar

i	x _i	x _i ²	(x _i - \bar{x}) ²
1	63	3.969	229,31
2	62	3.844	260,596
3	81	6.561	8,162
4	75	5.625	9,878
5	94	8.836	251,444
6	97	9.409	355,586
7	75	5.625	9,878
Somos		43.869	1.124,854

a) Cálculo da variância:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$s^2 = \frac{1.124,854}{7} = 160,693$$

ou

$$s^2 = \frac{\sum x_i^2}{n} - \bar{x}^2$$

$$s^2 = \frac{43869}{7} - (78,143)^2 = 160,672$$

Observe que os resultados são praticamente idênticos. A pequena diferença apresentada deve-se aos arredondamentos feitos durante os cálculos.

b) Cálculo do Desvio Padrão:

$$s = \sqrt{160,693} = 12,676 \%$$

ou

$$s = \sqrt{160,672} = 12,676 \%$$



[5.15]

Observação importante:

Ao trabalhar com uma amostra, pode-se estimar a Variância e o Desvio Padrão da população substituindo o denominador por $(n - 1)$. Nos casos em que n é maior que 30 dados, o resultado é praticamente o mesmo.

Portanto, as fórmulas da variância e do desvio padrão, para amostras, são:

Variância:
$$s^2 = \frac{\sum x_i^2}{n-1} - \bar{x}^2 \tag{3.2.3.10}$$

Desvio Padrão:
$$s = \sqrt{\frac{\sum x_i^2}{n-1} - \bar{x}^2} \tag{3.2.3.11}$$

3.2.4 - Coeficiente de Variação

O Coeficiente de Variação, indicado por CV, é usado quando se faz necessário exprimir a variabilidade em termos relativos. Um desvio padrão igual a 1 pode ser considerado pequeno num conjunto de dados com média 1.000, mas pode ser bastante significativo para outro com média igual a 10.

O Coeficiente de Variação, pode ser obtido por uma regra de três simples, tomando a média como valor de referência (o inteiro):

Medida	Parte do inteiro
\bar{x}	1
s	CV

$$\frac{\bar{x}}{s} = \frac{1}{CV}$$

$$\frac{\bar{x}}{s} = \frac{1}{CV}$$

$$\boxed{CV = \frac{s}{\bar{x}}}$$

(3.2.4.1)

O Coeficiente de Variação é uma medida adimensional, já que o Desvio Padrão possui a mesma unidade da Média.

A principal característica do Coeficiente de Variação é permitir a comparação do grau de variabilidade de dois ou mais conjuntos de dados bem distintos.

Exemplo 3.18: Duas empresas foram avaliadas quanto à média salarial de seus funcionários obtendo, o seguinte resultado:

Empresa A:

$$\bar{x} = \text{R\$ } 350,00 \qquad s = \text{R\$ } 100,00$$

Empresa B:

$$\bar{x} = \text{R\$ } 420,00 \qquad s = \text{R\$ } 150,00$$

Calculando os coeficientes de variação, encontra-se:

Paulo Fernando Braga Carvalho

professor@paulofernando.mat.br

$$CV_A = \frac{100}{350} = 0,2857 \qquad CV_B = \frac{150}{420} = 0,3571$$

Portanto, a empresa B apresenta uma variabilidade salarial maior entre seus funcionários.

3.3 - Medidas Estatísticas para dados agrupados

Até o momento, as Medidas Estatísticas mais importantes foram estudadas trabalhando com pequenos conjuntos de dados. Mas, muitas vezes, no trabalho de pesquisa, os dados originais não estão disponíveis e sim tabelas como, por exemplo, a que foi construída com os índices de criminalidade. Em outras situações, estas tabelas de distribuição de frequência são construídas pela grande quantidade de dados.

Sendo assim inicia-se, o estudo do cálculo destas medidas para dados agrupados. Para tanto, serão utilizados os dados da TABELA 2, de distribuição de frequência dos índices de criminalidade, vista anteriormente

TABELA 2
Distribuição de frequência dos índices de criminalidade no
ano de 1995 dos municípios da mesorregião do Triângulo
Mineiro e Alto Paranaíba

i	Classes	x_i	n_i	f_i	N_i	F_i
1	4,48 ↪ 10,11	7,295	12	0,19	12	0,19
2	10,11 ↪ 15,74	12,925	13	0,20	25	0,39
3	15,74 ↪ 21,37	18,555	15	0,23	40	0,62
4	21,37 ↪ 27,00	24,185	8	0,13	48	0,75
5	27,00 ↪ 32,63	29,815	9	0,14	57	0,89
6	32,63 ↪ 38,26	35,445	2	0,03	59	0,92
7	38,26 ↪ 43,89	41,075	5	0,08	64	1,00
Somam			64	1		

Fonte: Site do Geominas/PMMG

Observe que existem 8 municípios com índice de criminalidade no intervalo 21,37 \mapsto 27,00, ou seja, na 4ª classe. Mas, destes, quantos possuem índice realmente igual a 22,00, por exemplo? Não é possível responder a esta pergunta, a partir do momento em que apenas a Tabela de Distribuição de Freqüência está disponível. Portanto, cada classe precisa ter seu elemento representativo: o **ponto médio da classe**, x_i . Sendo assim, considera-se, para efeito de cálculo de algumas medidas estatísticas, que todos os 8 elementos da 4ª classe são iguais a 24,185. O mesmo ocorre para as demais classes: o elemento representativo da primeira classe é o 7,295, o da segunda classe 12,925, etc. Como consequência, geralmente só é possível obter aproximações para as medidas estatísticas

3.3.1 - Média

O cálculo da Média para dados agrupados é feito usando o conceito de Média Ponderada. Os valores são representados pelos pontos médios das classes e o peso é a freqüência de cada classe. Ou seja,

$$\bar{x} = \frac{\sum(n_i \cdot x_i)}{\sum n_i} \tag{3.3.1.1}$$

Exemplo 1.1: Para os dados da TABELA 2, temos:

$$\bar{x} = \frac{12 \cdot 7,295 + 13 \cdot 12,925 + 15 \cdot 18,555 + 8 \cdot 24,185 + 9 \cdot 29,815 + 2 \cdot 35,445 + 5 \cdot 41,075}{12 + 13 + 15 + 8 + 9 + 2 + 5}$$

$$\bar{x} = 19,8745$$



[5.16]

3.3.2 - Mediana

Ao calcular a mediana, pretende-se determinar o valor que divide o conjunto de dados ordenado, em duas partes iguais. Para tanto, faz-se a suposição de que os dados estão em ordem crescente e igualmente espaçados dentro das classes.

No caso da TABELA 2, como ao todo são 64 dados, a mediana é o 32º elemento pertencendo, portanto, à terceira classe (15,74 \mapsto 21,37). Até o final da segunda classe já existem 25 dados, logo, a mediana é o 7º elemento da terceira. A amplitude desta classe é igual a 5,63 (21,37 – 15,74 = 5,63). Logo, num intervalo de tamanho 5,63 “cabem” 15 dados (confira na tabela). Portanto, se em um intervalo de tamanho 5,63 “cabem” 15 dados, num intervalo de que tamanho encaixam-se os 7 dados necessários para se alcançar a mediana? Para este cálculo aplica-se a regra de três simples:

Tamanho do intervalo	quantidade de dados
5,63	15
x	7

$$\frac{5,63}{x} = \frac{15}{7}$$

$$x = 2,63$$

Portanto, é necessário um intervalo de tamanho 2,63 para “encaixar” os 7 dados que faltam (a mediana é o 32º elemento). Como a terceira classe começa em 15,74, temos:

$$\text{Md} = 15,74 + 2,63$$

$$\text{Md} = 18,37.$$

3.3.3 - Moda

Para dados agrupados fala-se em classe modal. Portanto, a classe modal de um conjunto de dados é a classe com maior frequência.

No exemplo da TABELA 2, a classe modal é a terceira, 15,74 \mapsto 21,37, com frequência igual a 15.

Esta medida é mais usada para dados agrupados do que para dados em seu estado bruto.

3.3.4 - Percentil

Os percentis, para dados agrupados, são calculados com o auxílio da regra de três simples, da mesma forma como foi feito para calcular a mediana. Lembre-se que a mediana é o percentil 50.

Como exemplo, com a intenção de determinar qual índice de criminalidade deixa 70% dos índices nele ou abaixo dele, deve-se calcular o P_{70} .

Resolução:

70% de 64 = 44,8, então, P_{70} é o 45º dado e pertence à 4ª classe (21,37 \mapsto 27,00).

Amplitude da classe: $27 - 21,37 = 5,63$

Número de dados na classe: 8

P_{70} é o 5º termo da 4ª classe ($40 + 5 = 45$)

Tamanho do intervalo	quantidade de dados
5,63	8
x	5

$$\frac{5,63}{x} = \frac{8}{5}$$

$$x = 3,52$$

$$P_{70} = 21,37 + 3,52$$

$$P_{70} = 24,89.$$

3.3.5 - Variância e Desvio Padrão

Seguindo raciocínio análogo ao apresentado no cálculo da média, quando o ponto médio foi usado como elemento representativo da classe, as fórmulas da variância e do desvio padrão podem ser modificadas para o trabalho com dados agrupados:

Variância e Desvio Padrão

$$s^2 = \frac{\sum (n_i \cdot x_i^2)}{\sum n_i} - \bar{x}^2 \quad (3.3.5.1)$$

$$s = \sqrt{\frac{\sum (n_i \cdot x_i^2)}{\sum n_i} - \bar{x}^2} \quad (3.3.5.2)$$

Portanto, para auxílio nos cálculos, a TABELA 12 será criada selecionando as colunas necessárias da TABELA 2 e incluindo duas novas colunas, veja:

TABELA 12
Construção auxiliar para cálculo da variância e do desvio padrão dos índices de criminalidade dos municípios da mesorregião do Triângulo Mineiro e Alto Paranaíba em 1995

i	Classes	x_i	n_i	x_i^2	$n_i \cdot x_i^2$
1	4,48 ↪ 10,11	7,295	12	53,217	638,604
2	10,11 ↪ 15,74	12,925	13	167,056	2.171,728
3	15,74 ↪ 21,37	18,555	15	344,288	5.164,320
4	21,37 ↪ 27,00	24,185	8	584,914	4.679,312
5	27,00 ↪ 32,63	29,815	9	888,934	8.000,406
6	32,63 ↪ 38,26	35,445	2	1.256,348	2.512,696
7	38,26 ↪ 43,89	41,075	5	1.687,156	8.435,780
	Somas		64		31.602,846

Fonte: Site do Geominas/PMMG. Acesso em 05 jul 2001

$$s^2 = \frac{31.602,846}{64} - (19,875)^2$$

$$s^2 = 98,779 \text{ (Variância)}$$

e

$$s = \sqrt{s^2} = \sqrt{98,779}$$

$$s=9,939 \text{ (Desvio Padrão)}$$



[5.17]

3.3.6 - Comparação entre Média, Mediana e Moda

É difícil especificar regras que apontem situações onde é mais adequado usar cada uma das medidas de tendência central. Em situações onde os dados estão distribuídos de forma simétrica os valores obtidos para a Média, Mediana e Moda podem ser iguais, mas o mesmo não ocorre para conjuntos que apresentam assimetria.

A escolha da Medida de Tendência Central depende muito dos objetivos da pesquisa e do conhecimento das vantagens e desvantagens de cada medida.

A moda tem como principal característica apontar o que é mais comum, ou seja, o dado com maior frequência. Esta medida é mais comumente usada para dados agrupados, onde determina-se a classe com maior frequência.

Para comparar a média e a mediana, deve-se levar em conta que a média é sensível a todos os dados do conjunto, ou seja, cada valor tem uma parcela de influência em seu resultado. Já a mediana leva em conta apenas a posição e quantidade de dados do conjunto analisado. Uma grande desvantagem da mediana é a necessidade de ordenação dos dados para sua obtenção.

Para decidir qual medida de tendência central deve ser escolhida, uma boa idéia é construir o histograma ou o gráfico de colunas do conjunto de dados, e analisar o nível de simetria ou assimetria apresentado.

Quanto mais próximo da simetria o gráfico estiver, mais a média apresenta-se como uma boa referência de tendência central. Quando se constata uma forte assimetria, a mediana deve ser a medida escolhida, já que ela é menos sensível aos valores extremos. Esta situação é muito comum em situações monetárias, como nos casos onde os dados estão relacionados a salários e valores imobiliários.

Veja os gráficos a seguir:

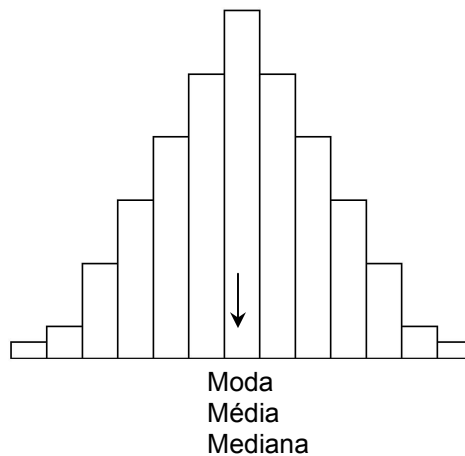


GRÁFICO 9: Simetria

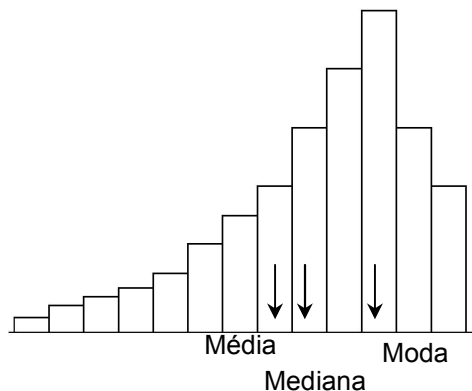


GRÁFICO 10: Assimetria à esquerda

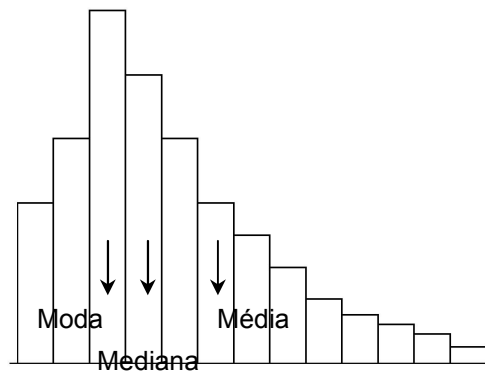


GRÁFICO 11: Assimetria à direita

4 - ESTATÍSTICA ESPACIAL

4.1 – Introdução

Ao realizar análises de dados geográficos, é importante ir além da estatística convencional e da simples visualização dos dados, dando atenção à componente espacial. Para tanto, é preciso adicionar, aos estudos desenvolvidos até este ponto, recursos da Estatística Espacial .

Estatística Espacial é um ramo da Estatística que estuda métodos científicos para a coleta, descrição, visualização e análise de dados que possuem coordenadas de localização. Sua marca distintiva é o uso explícito ou implícito destas coordenadas de localização na modelagem. (adaptado de ASSUNÇÃO, 2001, p.2).

Na realidade, a Estatística Espacial lida com dados que possuem coordenadas de localização, não sendo necessariamente dados geográficos, como por exemplo, os obtidos em estudos de neuroanatomia. O que distingue as técnicas de Estatística Espacial é o uso das coordenadas espaciais dos dados no processo de coleta, descrição ou análise de dados. Sem dúvidas, esta área da Estatística tem se mostrado de grande importância no trabalho com dados geográficos.

O enorme desenvolvimento computacional recente é o responsável pelo florescimento da área espacial em estatística. Em particular, o desenvolvimento de grandes sistemas geográficos de informação a um custo reduzido tem possibilitado e estimulado a criação de grandes bases de dados georeferenciados (ASSUNÇÃO, 2001, p.2)

O que existe de tão especial com os dados geográficos? O dado geográfico possui uma localização geográfica (expressa como coordenadas em um mapa) e atributos descritivos (que podem ser representados num banco de dados convencional).

Os mapas por si só não são úteis, se não for possível descrevê-los, compará-los e interpretá-los. É de importância relevante o uso da Estatística descritiva ou inferencial, para tornar explícito o que está apresentado nos mapas possibilitando análises.

Um interessante exemplo de aplicação da Análise Espacial, trata-se dos estudos do Dr. John Snow em 1854, que relacionou os casos de cólera em Londres com a distribuição espacial de poços de água. Naquela época ainda não se conhecia a forma de contaminação. Dr Snow colocou no mapa da cidade a localização dos doentes de cólera e dos poços de água, principal fonte de água dos habitantes da cidade naquela época. Com a espacialização dos dados, percebeu-se que a maioria dos casos estava concentrada em torno do poço da “Broad Street”. Como conseqüência, foi decretada a lacração do poço, contribuindo para debelar a epidemia. Esta constatação forneceu evidências empíricas para a hipótese, posteriormente comprovada, de que o cólera é transmitido por ingestão de água contaminada.



MAPA 4: Mapa de Londres com casos de cólera (pontos) e poços de água (cruzes)
 Fonte: CAMARA, Gilberto et al. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/>>

O trabalho desenvolvido pelo Dr. Snow é considerado como uma das primeiras aplicações de sucesso da análise espacial.

A partir deste ponto, considere a situação hipotética de três países com a forma do diagrama da FIGURA 1 e sua respectiva divisão em regiões:

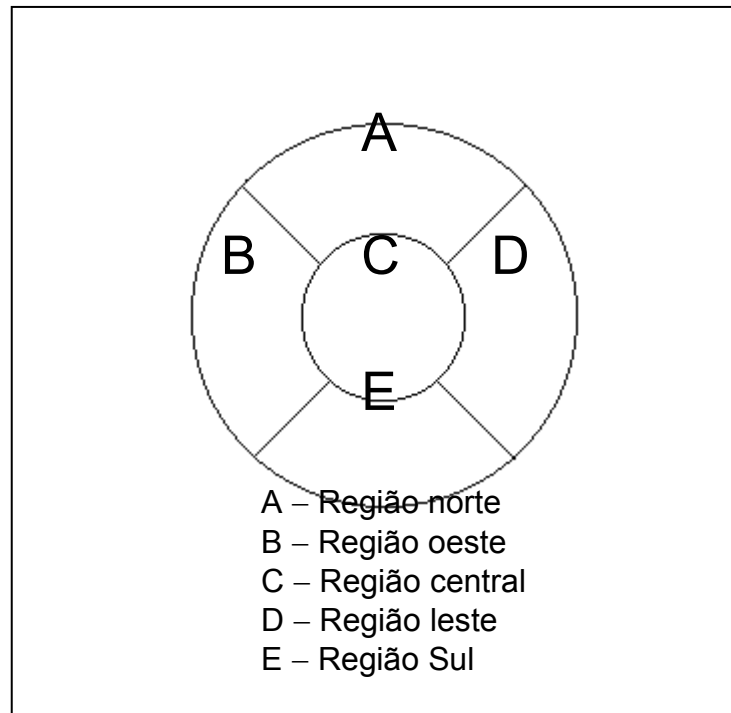


FIGURA 1: País hipotético e suas divisões regionais

Considere, ainda, que os números indicados nos diagramas da FIGURA 2, representem a densidade demográfica, em habitantes por km² de cada região, em cada um dos países:

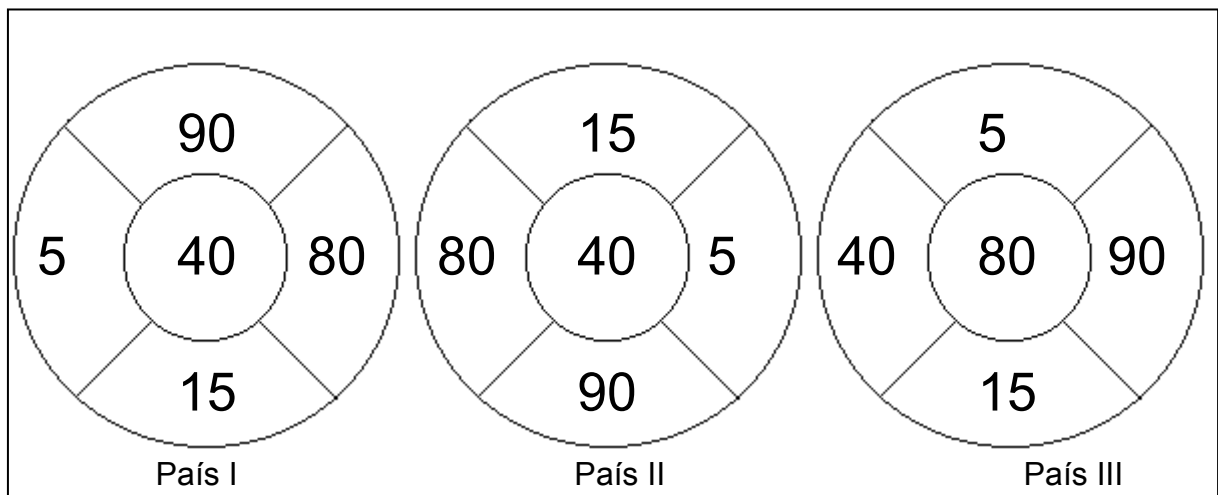


FIGURA 2: Distribuição da densidade demográfica em três países hipotéticos

Observe que para todas as configurações encontramos:

Média = 46 hab/km²

Desvio padrão = 33,97 hab/km²

Mediana = 40 hab/km²

Amplitude = 85 hab/km²

A situação hipotética acima apresenta um problema específico enfrentado pelos geógrafos: as medidas estatísticas, como estudadas até agora, não levam em conta a localização de cada fenômeno estudado. Até o momento, seria possível considerar que os três países apresentam distribuições populacionais idênticas. Mas, observe que em cada configuração a distribuição espacial da população é diferente, apesar de todos os países apresentarem resultados idênticos para média, desvio padrão, mediana e amplitude dos dados.

Na ótica de um trabalho geográfico, os países I, II e III representados pelos diagramas são bastante diferentes. No país I, encontra-se uma maior densidade populacional nas regiões Norte e Leste. No país II, a população concentra-se, principalmente, nas regiões Sul e Oeste, enquanto que no país III, percebe-se uma tendência de concentração da população nas regiões Central e Leste. Portanto, não é aceitável considerar que os 3 países são idênticos em relação à sua distribuição populacional no espaço.

Alguns problemas que envolvem a variável espacial são de grande importância quando se procura minimizar, por exemplo, o custo de redes telefônicas, rodovias, redes de distribuição de água e esgoto, transporte de mercadorias, etc. Ao tentar resolver problemas como estes, entende-se o porquê da Cartografia e da Geodésia serem campos do conhecimento de enorme importância.

Problemas semelhantes a estes foram estudados por Pierre de Fermat (1608-1647) e Jacob Steiner (1786-1863), professor de Matemática na Universidade de Berlim. O enunciado geral dos problemas estudados por estes cientistas é o de determinar a rede mínima de comunicação entre n pontos, ou seja: dados n pontos, determinar um sistema conexo de segmentos retilíneos que os una e que tenha comprimento mínimo.

O primeiro registro conhecido de solução para este problema, no caso de três pontos, é de cerca de 1640, solução apresentada por Evangelista Torricelli e Francesco Cavalieri, em trabalhos independentes.

Estes exemplos, foram citados com o objetivo de justificar a necessidade de incorporação da componente de localização espacial às medidas Estatística já estudadas. Sendo assim, inicia-se o estudo de uma área bastante específica da Estatística e de grande importância para o geógrafo: a Estatística Espacial.

As medidas de tendência central e as de dispersão serão estudadas sob o ponto de vista de interesse do geógrafo, além de outras formas de apresentação dos dados. Um importante recurso para a apresentação de dados georreferenciados é o mapa temático. Apesar de a construção de mapas temáticos não figurar como objetivo do curso, busca-se provocar o interesse no leitor de avançar seus estudos sobre o assunto.

Com o objetivo de trabalhar com a localização espacial, toma-se como referência o sistema de coordenadas cartesianas ortogonais, explorando a dicotomia existente na dimensão espacial, representada por valores como Norte/Sul, Leste/Oeste, Vertical/Horizontal. Assim, torna-se possível o estudo dos conceitos de Centro Médio, Distância Padrão e outros.

Os sistemas de coordenadas utilizados respeitam as seguintes regras:

- os eixos são ortogonais (formam ângulo de 90°);
- as unidades de medidas aplicadas aos eixos são iguais.

Como exemplo do sistema de coordenadas cartesianas ortogonais, tem-se o plano xOy , estudado nos cursos de Matemática do Ensino Médio, onde cada ponto é localizado por um par ordenado (x,y) , veja FIGURA 3.

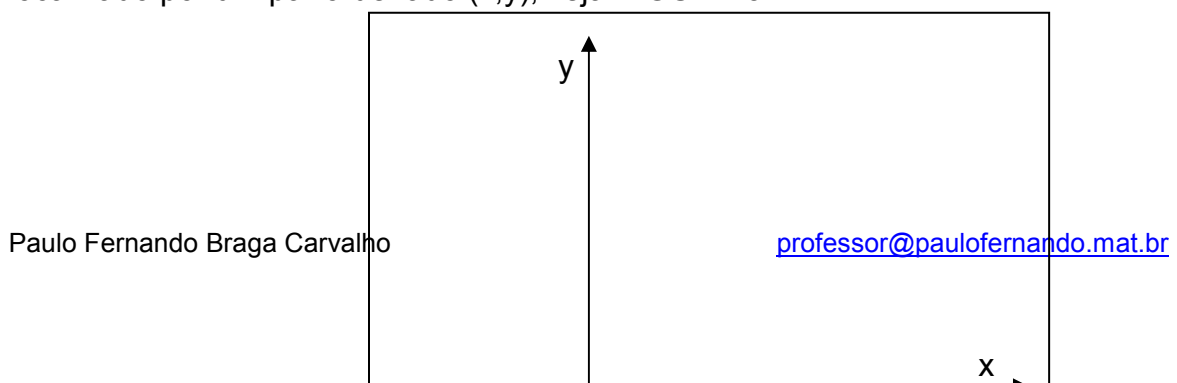


FIGURA 3: Plano cartesiano

4.2 - Medidas de tendência central em padrões de pontos

A localização do centro da distribuição espacial, dos pontos que representam os dados coletados a respeito de determinado fenômeno, será determinada pelas medidas de tendência central em padrões de pontos, explorando o conceito geométrico de ponto e admitindo que área e reta são algumas das representações de conjunto de pontos.

4.2.1 - Centro Modal

Centro Modal é a região ou ponto do espaço onde o fenômeno estudado ocorre com maior frequência.

Nos mapas temáticos o Centro Modal é apresentado com cores mais intensas ou colunas mais altas.

A opção pelo Centro Modal como medida de tendência central deve ser feita com cuidado, já que esta medida não é sensível ao que ocorre nas demais regiões, preocupando apenas com a região que se destaca pela maior frequência e ainda,

dependendo do problema, esta medida pode não se mostrar aplicável a estudos ao longo do tempo.

4.2.2 - Centro Mediano

O conceito de Centro Mediano está diretamente relacionado ao conceito de Mediana.

Para dados dispostos no espaço o Centro Mediano é determinado pela interseção de duas retas perpendiculares, que dividem o conjunto de pontos em duas partes iguais.

Considerando os pontos cardeais Norte, Sul, Leste, Oeste como referência, o Centro Mediano é o ponto de interseção da reta que deixa 50% dos pontos ao Sul e 50% dos pontos ao Norte com a reta que deixa 50% dos pontos a Leste e 50% dos pontos a Oeste.

O trabalho com o referencial do exemplo anterior não é obrigatório. Pode-se, por algum interesse, preferir o uso das retas Nordeste-Sudoeste e Noroeste-Sudeste, ou quaisquer outros eixos que sejam perpendiculares entre si e que apresentem escalas na mesma unidade de Medida. Mas, é importante salientar que, mudando o referencial, o Centro Mediano provavelmente terá mudanças nas coordenadas

Os casos mais complexos, nos quais se faz necessário trabalhar com a intensidade de cada ponto em relação aos demais além de sua localização, não serão discutidos neste curso.

4.2.3 - Centro Médio

O Centro Médio, também conhecido como centróide, é o ponto (\bar{x}, \bar{y}) obtido pela interseção das retas perpendiculares entre si, traçadas a partir da localização de \bar{x} , média das abscissas (coordenadas x) e \bar{y} , média das ordenadas (coordenadas y). Este ponto minimiza a soma das distâncias quadráticas a todos os outros pontos do plano.

Ou seja,

$$\bar{x} = \frac{\sum x_i}{n} \text{ e } \bar{y} = \frac{\sum y_i}{n} . \quad (4.2.3.1) \text{ e } (4.2.3.2)$$

Exemplo 4.1:

Considere algumas cidades, em uma região hipotética, e as respectivas distribuições de freqüências de ocorrência de determinado fenômeno: número de postos de saúde, postos bancários, supermercados, instituições educacionais, residências rurais, etc.

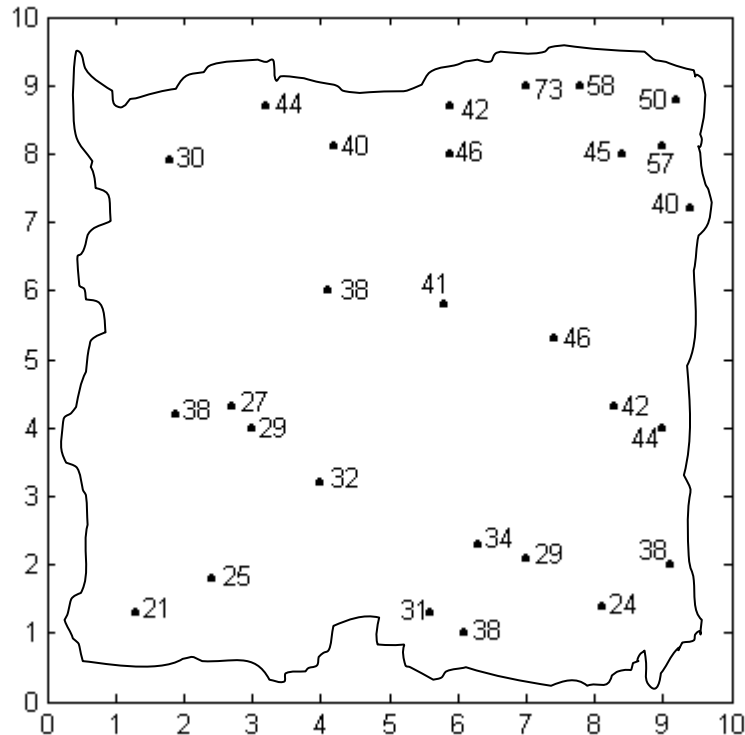
A TABELA 13 apresenta as coordenadas x (abscissa) e y (ordenada) de cada ponto (localidade) e a respectiva freqüência, p , do fenômeno estudado:

TABELA 13

Distribuição de determinado respectivo	<hr/>			um fenômeno e peso
	x	y	p	
	1,3	1,3	21	
	1,8	7,9	30	
	1,9	4,2	38	
	2,4	1,8	25	
	2,7	4,3	27	
	3,0	4,0	29	
	3,2	8,7	44	
	4,0	3,2	32	
	4,1	6,0	38	
	4,2	8,1	40	
	5,6	1,3	31	
	5,8	5,8	41	
	5,9	8,7	42	
	5,9	8,0	46	
	6,1	1,0	38	
	6,3	2,3	34	
	7,0	2,1	29	
	7,0	9,0	73	
	7,4	5,3	46	
	7,8	9,0	58	
	8,1	1,4	24	
	8,3	4,3	42	
	8,4	8,0	45	
	9,0	4,0	44	
	9,0	8,1	57	
	9,1	2,0	38	
	9,2	8,8	50	
	9,4	7,2	40	

Fonte: Dados hipotéticos

O MAPA 5 representa a sobreposição do mapa da região sobre o plano de eixos



ortogonais adotado como referência, tendo sua origem (0,0) no canto inferior esquerdo.

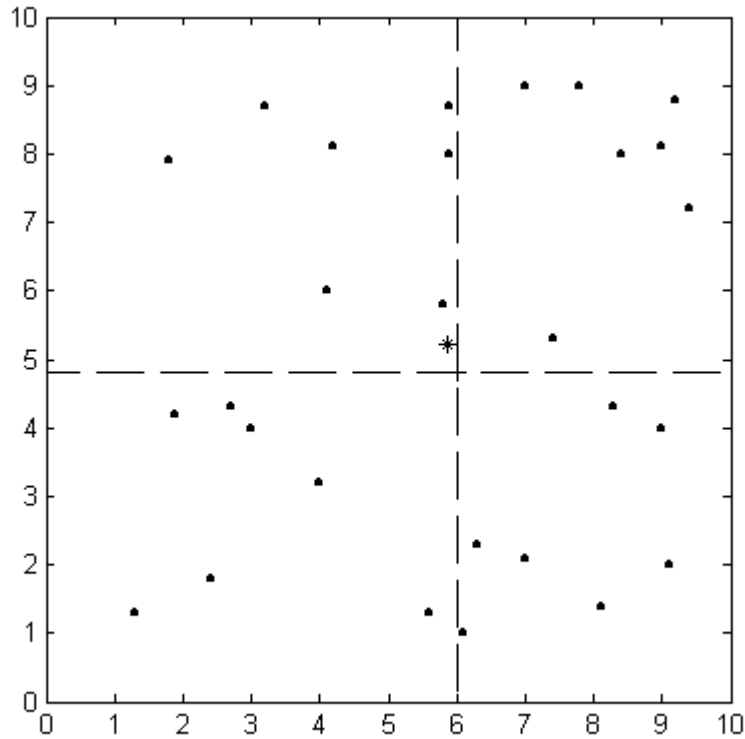
MAPA 5: Sobreposição do mapa com a região estudada sobre o plano de eixos coordenados

O MAPA 6, apresenta o mapa da região com os pontos relativos aos Centro Modal, Centro Médio e Centro Mediano.

O Centro Modal é o ponto de coordenadas (7,9), que representa a cidade com maior frequência, igual a 73.

Para encontrar o Centro Médio, calcula-se a média das abscissas (dos x's) e a média das ordenadas (dos y's), obtendo $\bar{x} \cong 5,8536$ e $\bar{y} \cong 5,2071$.

Calculando a Mediana das abscissas vamos encontrar 6, enquanto a Mediana da ordenadas é igual a 4,8. Logo, o Centro Mediano é o ponto de coordenadas (6; 4,8).



Confira no MAPA 6 onde existem 14 pontos abaixo da reta horizontal que passa pelo Centro Mediano e 14 pontos acima. O mesmo ocorrendo em relação à reta vertical.

MAPA 6: Centro Médio, Centro Modal e Centro Mediano

4.2.4 - Centro Médio ponderado

Quando a intensidade com que determinado fenômeno ocorre mostra-se de importância na análise de sua distribuição, determina-se o Centro Médio Ponderado, uma medida muito aplicada nos estudos geográficos. Para o cálculo do Centro Médio Ponderado (CMP), aplica-se o mesmo princípio envolvido na Média Aritmética Ponderada.

Assim, as coordenadas do CMP são obtidas calculando a média ponderada das abscissas e das ordenadas, ou seja:

$$\bar{x}_p = \frac{\sum p_i \cdot x_i}{\sum p_i} \quad \bar{y}_p = \frac{\sum p_i \cdot y_i}{\sum p_i} \quad (4.2.4.1) \text{ e } (4.2.4.2)$$

$$CMP = (\bar{x}_p, \bar{y}_p) \quad (4.2.4.3)$$

Para facilitar o cálculo do Centro Médio Ponderado do exemplo anterior, foi construída a TABELA 14, acrescentando mais duas colunas:

TABELA 14
Construção auxiliar para os cálculos do Centro
Médio Ponderado

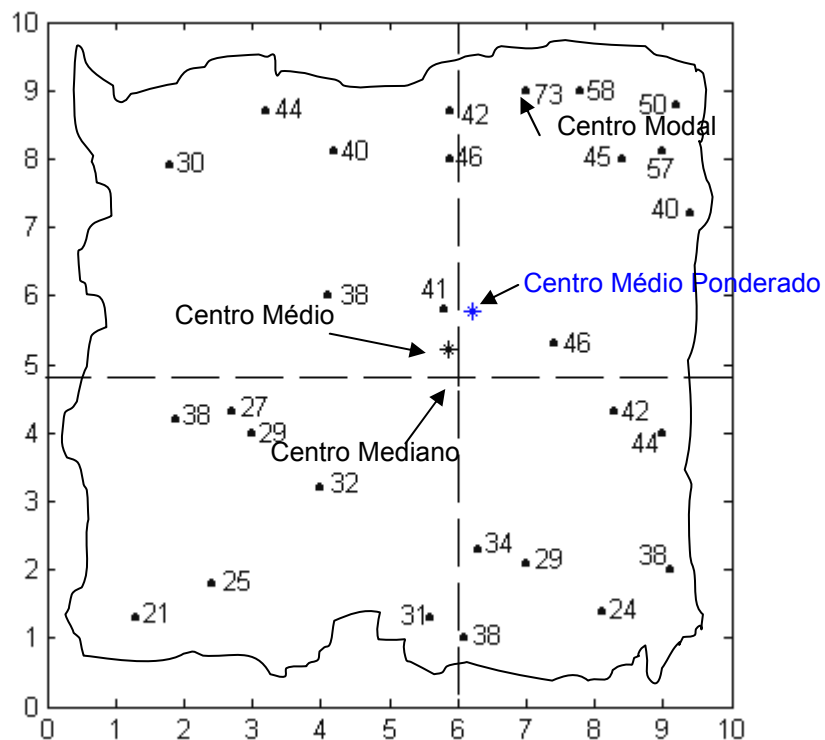
X	Y	p	$x_i \cdot p_i$	$y_i \cdot p_i$
1,3	1,3	21	27,3	27,3
1,8	7,9	30	54,0	237,0
1,9	4,2	38	72,2	159,6
2,4	1,8	25	60,0	45,0
2,7	4,3	27	72,9	116,1
3,0	4,0	29	87,0	116,0
3,2	8,7	44	140,8	382,8
4,0	3,2	32	128,0	102,4
4,1	6,0	38	155,8	228,0
4,2	8,1	40	168,0	324,0
5,6	1,3	31	173,6	40,3
5,8	5,8	41	237,8	237,8
5,9	8,7	42	247,8	365,4
5,9	8,0	46	271,4	368,0
6,1	1,0	38	231,8	38,0
6,3	2,3	34	214,2	78,2
7,0	2,1	29	203,0	60,9
7,0	9,0	73	511,0	657,0
7,4	5,3	46	340,4	243,8
7,8	9,0	58	452,4	522,0
8,1	1,4	24	194,4	33,6
8,3	4,3	42	348,6	180,6
8,4	8,0	45	378,0	360,0
9,0	4,0	44	396,0	176,0
9,0	8,1	57	513,0	461,7

9,1	2,0	38	345,8	76,0
9,2	8,8	50	460,0	440,0
9,4	7,2	40	376,0	288,0
Somos		1102	6861,2	6365,5

Logo, o Centro Médio Ponderado tem coordenadas:

$$\bar{x}_p = \frac{6861,2}{1102} = 6,2261 \quad \text{e} \quad \bar{y}_p = \frac{6365,5}{1102} = 5,7763$$

Observe, no MAPA 7, que o Centro Médio Ponderado é “puxado” para a região com maior índice de frequência. Compare sua posição com a do Centro Médio, medida que leva em conta apenas as coordenadas dos pontos observados.



MAPA 7: Centro Médio, Centro Mediano, Centro Modal e Centro Médio Ponderado

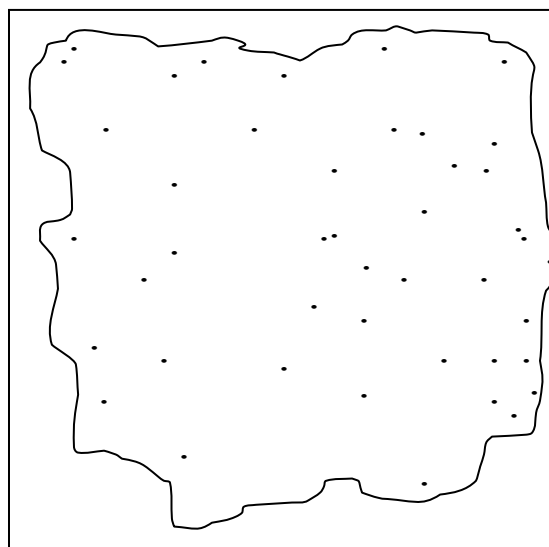
Pontos localizados em posições discrepantes podem atrair para si o Centro Médio ou o Centro Médio Ponderado, mas as maiores intensidades provocam atração apenas na localização do Centro Médio Ponderado.

É interessante acompanhar a evolução da posição do Centro Médio ao longo do tempo. Ao fazer isto é possível acompanhar o deslocamento do centro de equilíbrio da região, o que pode despertar o interesse de aprofundamento nos estudos deste fenômeno, tentando explicar as razões e fazer previsões.

4.2.5 - Centro Médio para dados agrupados

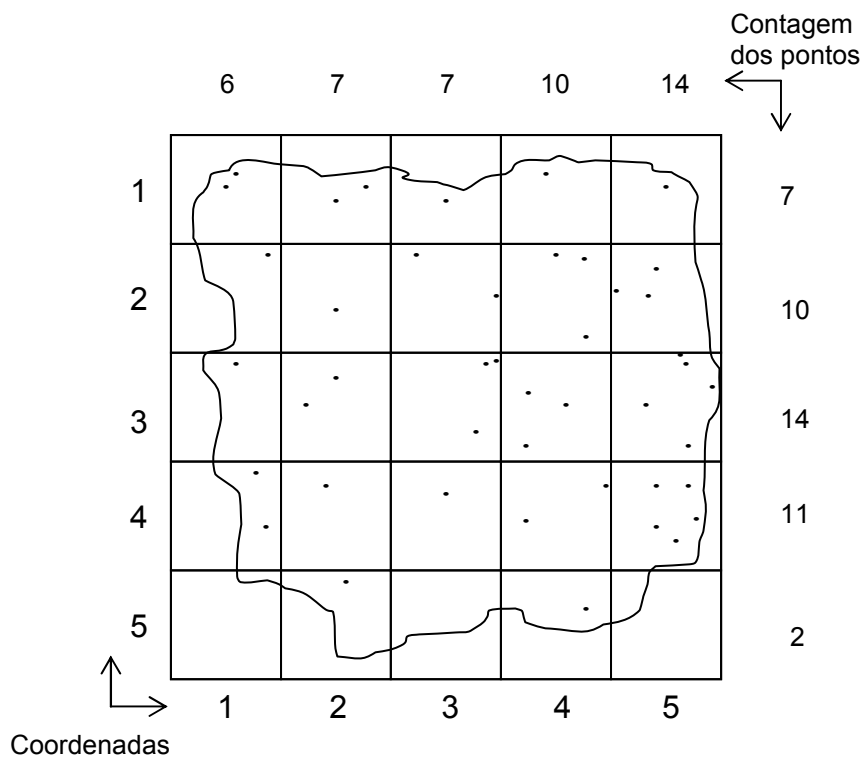
Como já foi discutido anteriormente, a Média aritmética presta-se a operações matemáticas de interesse, diferentemente da Mediana. Para calcular o Centro Médio de um grande conjunto de dados, toma-se “emprestada” a fórmula da Média Aritmética Ponderada .

Veja uma aplicação: Considere um conjunto de pontos localizados em uma determinada região do espaço, representado nas figuras abaixo, através de um fragmento de mapa (MAPA 8) sobreposto por um sistema de coordenadas ortogonais (MAPA 9):



MAPA 8: Distribuição da ocorrência de determinado

Fenômeno em uma região do espaço
Fonte: Situação hipotética



MAPA 9: Mapa 8 sobreposto a um sistema de coordenadas ortogonais

Neste caso, para obter o Centro Médio, usamos a fórmula de Média Ponderada, tomando como peso, a quantidade de pontos em cada coordenada x e y. Portanto,

$$\bar{x} = \frac{\sum x_i \cdot p_i}{n} = \frac{6 \cdot 1 + 7 \cdot 2 + 7 \cdot 3 + 10 \cdot 4 + 14 \cdot 5}{44} = 3,43 \quad (4.2.5.1)$$

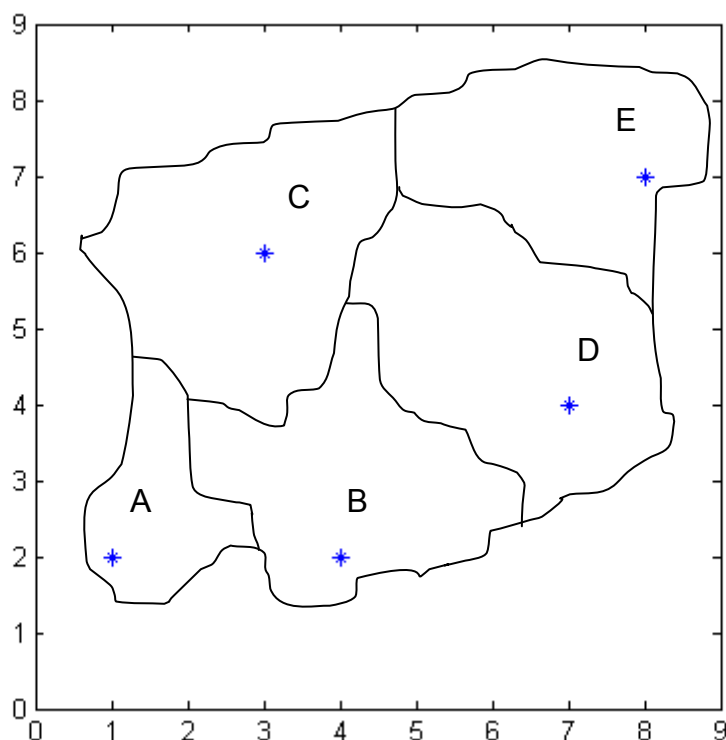
$$\bar{y} = \frac{\sum y_i \cdot p_i}{n} = \frac{7 \cdot 1 + 10 \cdot 2 + 14 \cdot 3 + 11 \cdot 4 + 2 \cdot 5}{44} = 2,80 \quad (4.2.5.2)$$

$$CM = (3,43 ; 2,80)$$

4.2.6 - Aplicação

A título de curiosidade e para fixação dos conceitos discutidos, serão resolvidos dois problemas: o primeiro deles, apresentado como exemplo 4.2, onde foram adotadas hipóteses simplificadoras, tem muito a ver com as situações enfrentadas pelas empresas de logística, transportadoras, supermercados, etc. O segundo problema, exemplo 4.3, apresenta o estudo da determinação do centróide do PIB do estado de Minas Gerais e sua evolução no período de 1990 a 1998.

Exemplo 4.2: Suponha que uma grande empresa de comércio varejista possua 5 supermercados distribuídos em alguns bairros de uma mesma cidade, conforme MAPA 10:



MAPA 10: Distribuição dos supermercados na cidade

Com o crescimento da empresa e da demanda de abastecimento de mercadorias a serem comercializadas pelas unidades, deseja-se estabelecer um local para instalação de um centro de distribuição, o mais econômico possível. Ou seja, busca-se uma localização que proporcione:

- economia de combustível;
- menor tempo de abastecimento das lojas;
- economia de pneus;
- economia na manutenção da frota de caminhões.

Com este objetivo, será determinada uma posição que minimize o deslocamento diário dos caminhões. Ou seja, busca-se um ponto que tenha a soma das distâncias de cada loja a ele como a menor possível.

Para tanto, algumas hipóteses serão assumidas:

- Homogeneidade do espaço geográfico;
- Facilidades de acesso idênticas para todas as lojas no trânsito local;

O ponto que respeita esta característica já foi estudado: o Centro Médio.

Para obtê-lo, é necessário verificar a localização de cada uma das lojas. Com a sobreposição de um mapa da região a um par de eixos cartesianos, conforme MAPA 10, as coordenadas podem ser determinadas:

TABELA 15
Coordenadas das Lojas de uma grande rede de supermercados

Loja	x	y
A	1	2
B	4	2
C	3	6
D	7	4

E	8	7
---	---	---

Fonte: Dados hipotéticos

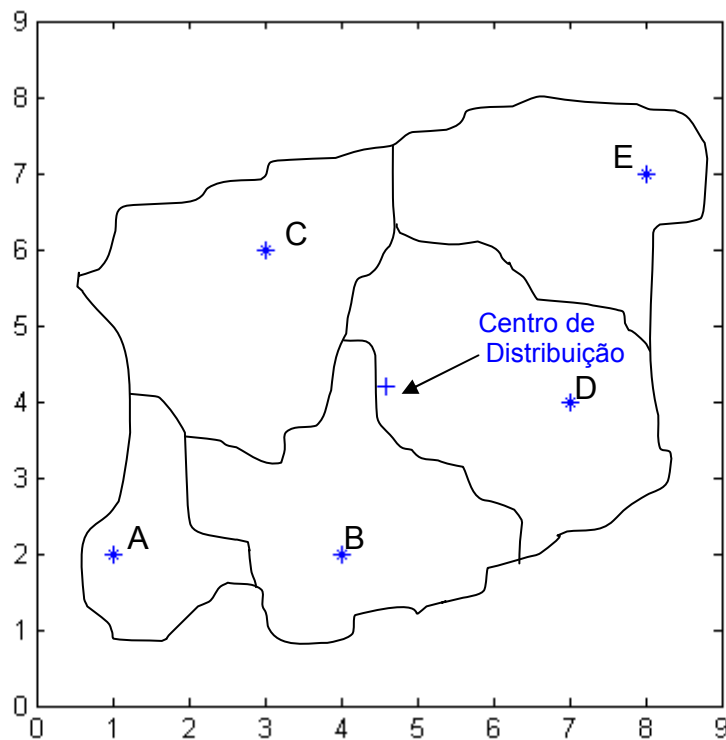
Portanto, o Centro Médio (ou Centro de Gravidade) será dado por:

$$\bar{x} = \frac{\sum_{i=1}^5 x_i}{5} = \frac{1+4+3+7+8}{5} = \frac{23}{5} = 4,6$$

e

$$\bar{y} = \frac{\sum_{i=1}^5 y_i}{5} = \frac{2+2+6+4+7}{5} = \frac{21}{5} = 4,2$$

De acordo com os dados do problema, o ponto mais adequado para instalação do centro de distribuição de mercadorias é o ponto (4,6 ; 4,2), indicado no MAPA 11:



MAPA 11: Centro de Distribuição

Mas, este problema foi resolvido considerando que todas as lojas apresentam o mesmo potencial para o comércio das mercadorias oferecidas. Este contexto nem sempre é respeitado. O potencial de cada loja depende do poder aquisitivo dos consumidores da região e de outros aspectos que não serão discutidos aqui. Mas, é possível melhorar um pouco esta discussão usando o valor total em Reais, movimentado por cada loja mensalmente e considerar que, quanto maior for este valor, maior será a necessidade de abastecimento desta loja.

Suponha conhecida a venda total de cada loja, conforme TABELA 16 :

TABELA 16
Localização das lojas e venda em milhares de Reais

Loja	x	Y	Venda (em milhares de Reais)
A	1	2	50
B	4	2	120
C	3	6	180
D	7	4	90
E	8	7	200

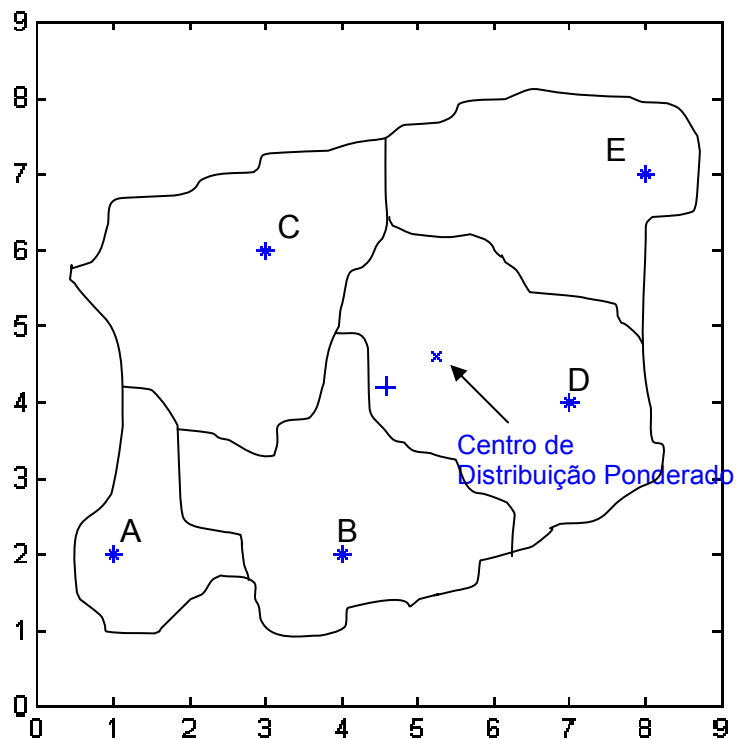
Fonte: Dados hipotéticos

Adotando o valor da venda como medidor da intensidade de cada ponto (cada loja) o Centro Médio Ponderado determinará o melhor ponto para instalação do Centro de distribuição de mercadorias. Ou seja, a venda assumirá a posição do peso na fórmula do Centro Médio Ponderado:

$$\bar{x}_p = \frac{\sum_{i=1}^5 p_i \cdot x_i}{\sum_{i=1}^5 p_i} = \frac{1 \cdot 50 + 4 \cdot 120 + 3 \cdot 180 + 7 \cdot 90 + 8 \cdot 200}{50 + 120 + 180 + 90 + 200} = 5,16 \quad (4.2.6.1)$$

$$\bar{y}_p = \frac{\sum_{i=1}^5 p_i \cdot y_i}{\sum_{i=1}^5 p_i} = \frac{2 \cdot 50 + 2 \cdot 120 + 6 \cdot 180 + 4 \cdot 90 + 7 \cdot 200}{50 + 120 + 180 + 90 + 200} = 4,97 \quad (4.2.6.2)$$

Portanto, levando-se em conta o volume de vendas de cada loja, o ponto mais adequado para instalação do centro de distribuição de mercadorias é o de coordenadas (5,25 ; 4,59), conforme indicado no MAPA 12:



MAPA 12: Centro de Distribuição Ponderado

Observe que o uso das vendas provocou um leve deslocamento do ponto sugerido anteriormente, sendo “puxado” um pouco para leste na direção horizontal e um pouco para norte na direção vertical.

Vale ressaltar que foram feitas algumas hipóteses simplificadoras para resolução deste exemplo. Para uma discussão mais adequada a este tipo de situação, sugere-se a aplicação de outros recursos da Análise Espacial e da Matemática, como por exemplo, a Teoria dos Grafos. Veja BARROSO (1998).

Exemplo 4.3: Neste exemplo apresenta-se a determinação da posição de alguns centróides do PIB do estado de Minas Gerais, no período de 1990 a 1998. Serão avaliados os PIB’s: Percapita, agropecuário, industrial e total, disponíveis na TABELA 17, incluída no ANEXO 9.3 deste trabalho.

Para a determinação do centróide de cada ano, usou-se as fórmulas do centro médio ponderado, tendo o PIB como peso. Ou seja, aplicou-se as fórmulas:

$$\bar{x}_p = \frac{\sum p_i \cdot x_i}{\sum p_i} \quad \bar{y}_p = \frac{\sum p_i \cdot y_i}{\sum p_i} \quad (4.2.4.1) \text{ e } (4.2.4.2)$$

$$CMP = (\bar{x}_p, \bar{y}_p) \quad (4.2.4.3)$$

Sendo:

x_i : Longitude do município

y_i : Latitude do município

p_i : PIB do município

Os resultados obtidos estão apresentados nas TABELAS 18, 19, 20 e 21. Após o cálculo das coordenadas dos centróides, usou-se o software MAPINFO para construção dos mapas.

TABELA 18
Centróides da variável PIB_percapita

Com Belo Horizonte			Sem Belo Horizonte		
Ano	Longitude	Latitude	Ano	Longitude	Latitude

1990	-44,7666	-20,0212	1990	-44,7699	-20,0220
1991	-44,7628	-20,0071	1991	-44,7659	-20,0079
1992	-44,7585	-20,0183	1992	-44,7618	-20,0191
1993	-44,7628	-20,0071	1993	-44,7659	-20,0079
1994	-44,8550	-19,9952	1994	-44,8580	-19,9959
1995	-44,7638	-20,0020	1995	-44,7671	-20,0028
1996	-44,7550	-20,0341	1996	-44,7582	-20,0350
1997	-44,6645	-19,8965	1997	-44,6672	-19,8968
1998	-44,6897	-19,8935	1998	-44,6924	-19,8938

FONTE: PUC-Minas



[5.18]

TABELA 19
Centróides da variável PIB_agropecuário

Ano	Longitude	Latitude
1990	-45,2361	-19,4865
1991	-45,2812	-19,4241
1992	-45,3420	-19,5070
1993	-45,2941	-19,4948
1994	-45,3846	-19,6252
1995	-45,3719	-19,5317
1996	-45,3794	-19,6563
1997	-45,4116	-19,4694
1998	-45,5033	-19,5475

FONTE: PUC-Minas

TABELA 20
Centróides da variável PIB_total

Com Belo Horizonte			Sem Belo Horizonte		
Ano	Longitude	Latitude	Ano	Longitude	Latitude
1990	-44,4521	-19,8711	1990	-44,6096	-19,8903

1991	-44,4650	-19,8612	1991	-44,6230	-19,8769
1992	-44,4598	-19,8672	1992	-44,6138	-19,8845
1993	-44,4671	-19,8606	1993	-44,6196	-19,8755
1994	-44,5258	-19,8543	1994	-44,6907	-19,8670
1995	-44,4845	-19,8493	1995	-44,6619	-19,8623
1996	-44,4806	-19,8690	1996	-44,6520	-19,8880
1997	-44,4902	-19,8598	1997	-44,6580	-19,8753
1998	-44,5115	-19,8561	1998	-44,6884	-19,8707

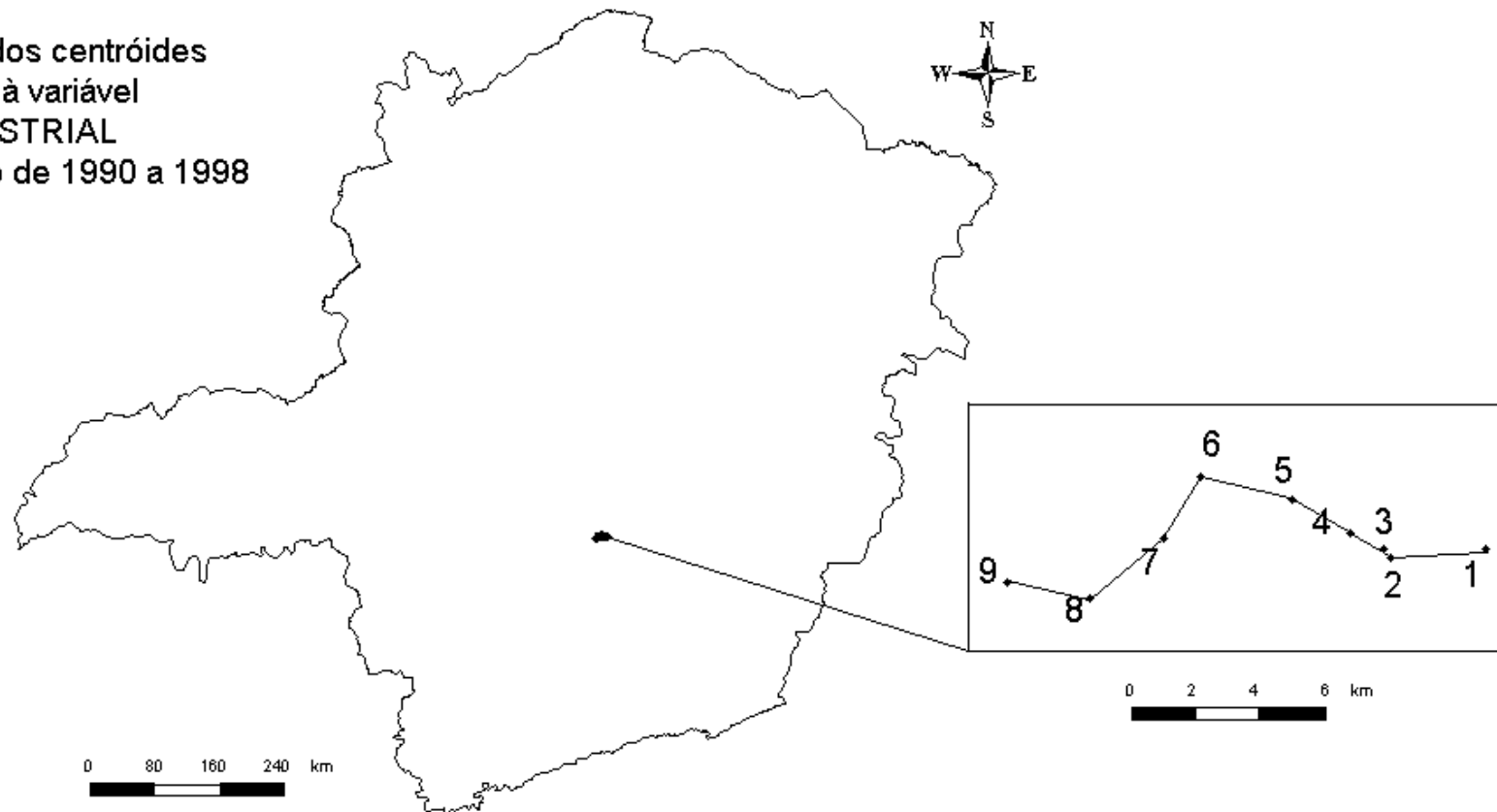
FONTE: PUC-Minas

TABELA 21
Centróides da variável PIB_industrial

Com Belo Horizonte			Sem Belo Horizonte		
Ano	Longitude	Latitude	Ano	Longitude	Latitude
1990	-44,2937	-19,9308	1990	-44,3665	-19,9564
1991	-44,3206	-19,9331	1991	-44,3953	-19,9579
1992	-44,3229	-19,9151	1992	-44,3991	-19,9366
1993	-44,3324	-19,9266	1993	-44,4073	-19,9494
1994	-44,3492	-19,9174	1994	-44,4319	-19,9396
1995	-44,3755	-19,9114	1995	-44,4663	-19,9330
1996	-44,3863	-19,9282	1996	-44,4733	-19,9518
1997	-44,4072	-19,9444	1997	-44,5030	-19,9725
1998	-44,4313	-19,9402	1998	-44,5341	-19,9680

FONTE: PUC-Minas

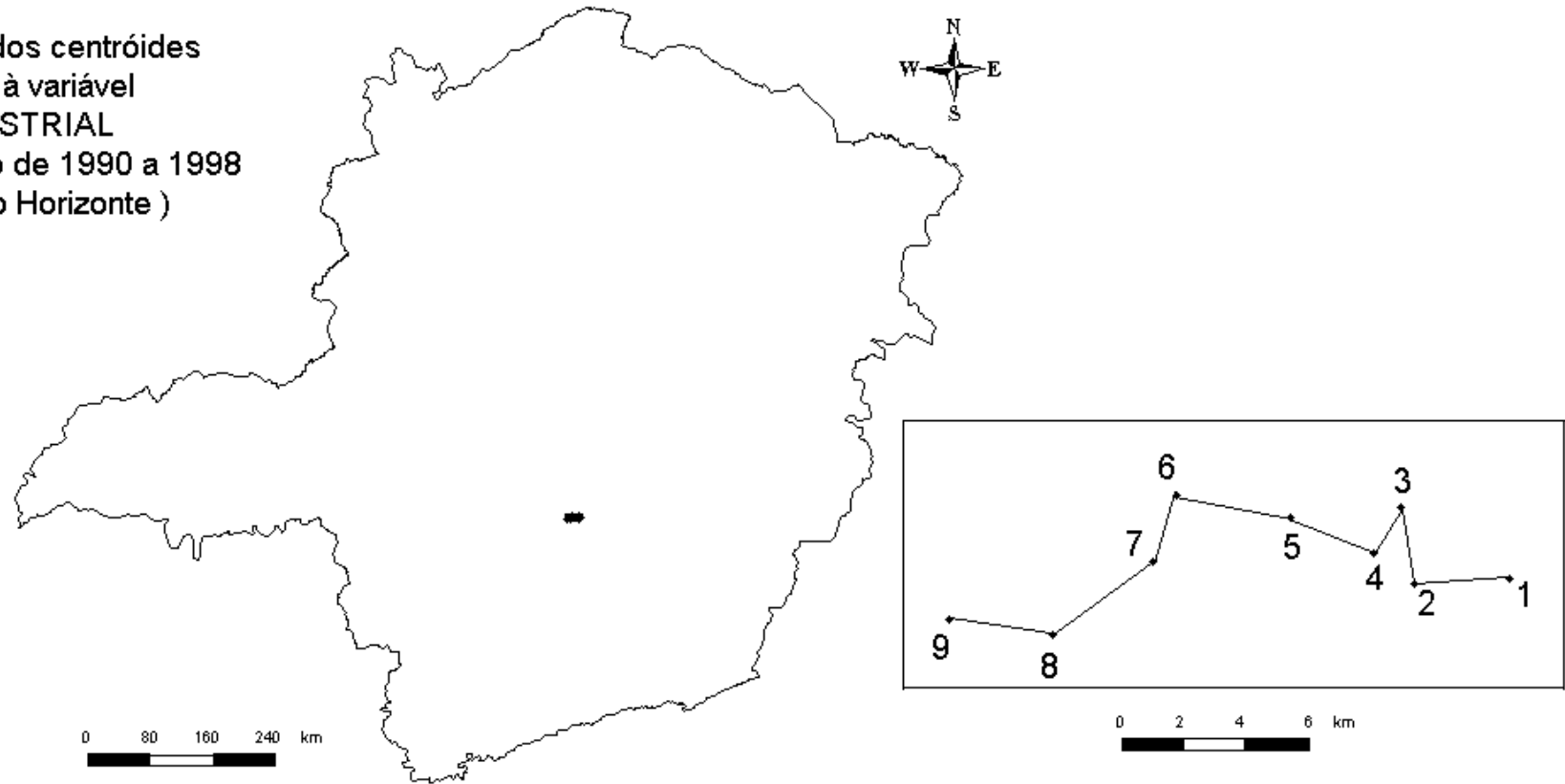
**Evolução dos centróides
referentes à variável
PIB_INDUSTRIAL
no período de 1990 a 1998**



Fonte: PUC-Minas

MAPA 13 – Evolução dos centróides referentes à variável PIB_industrial no período de 1990 a 1998

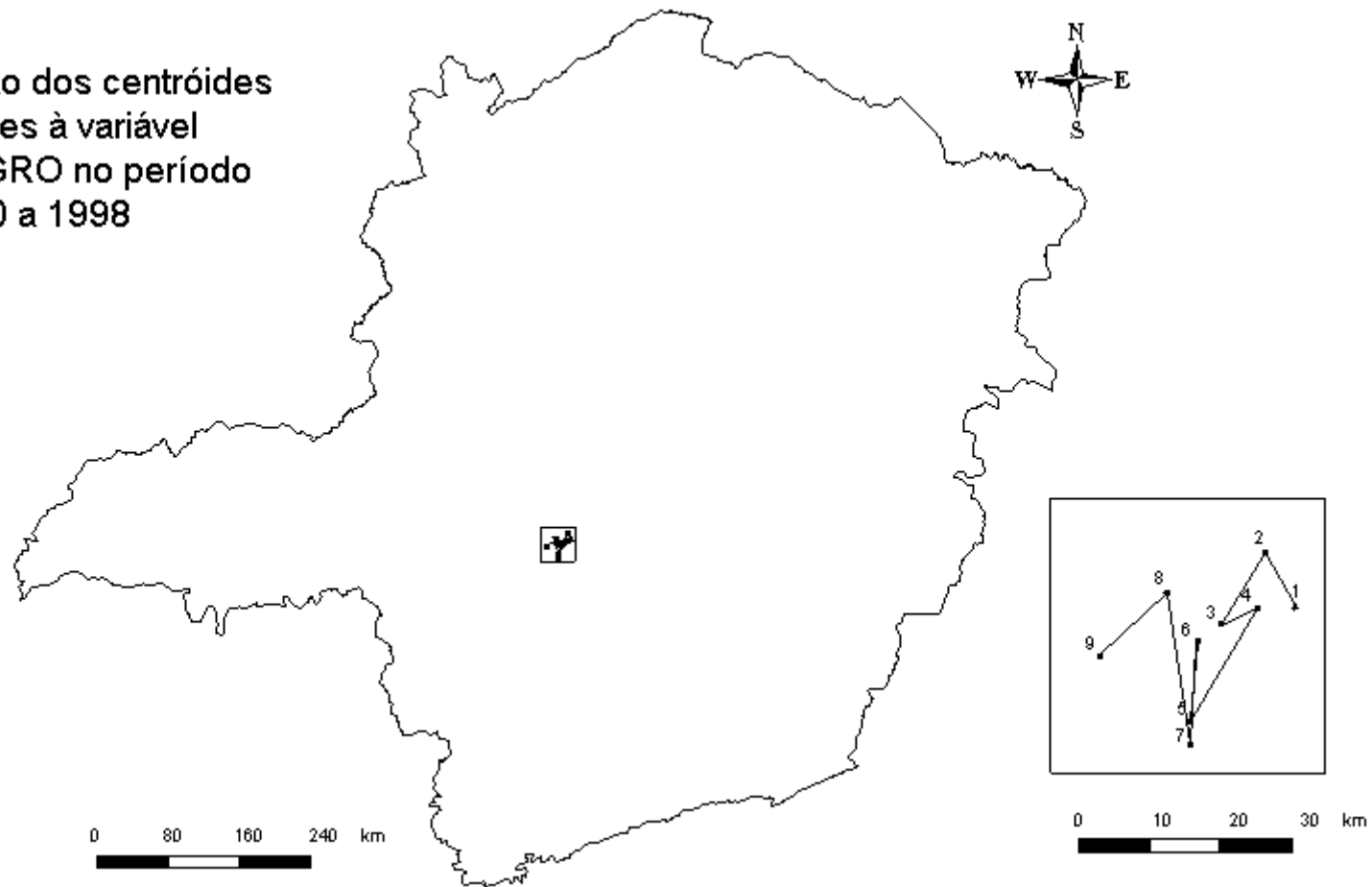
**Evolução dos centróides
referentes à variável
PIB_INDUSTRIAL
no período de 1990 a 1998
(Sem Belo Horizonte)**



Fonte: PUC-Minas

MAPA 14 – Evolução dos centróides referentes à variável PIB_indu no período de 1990 a 1998 (Sem Belo Horizonte)

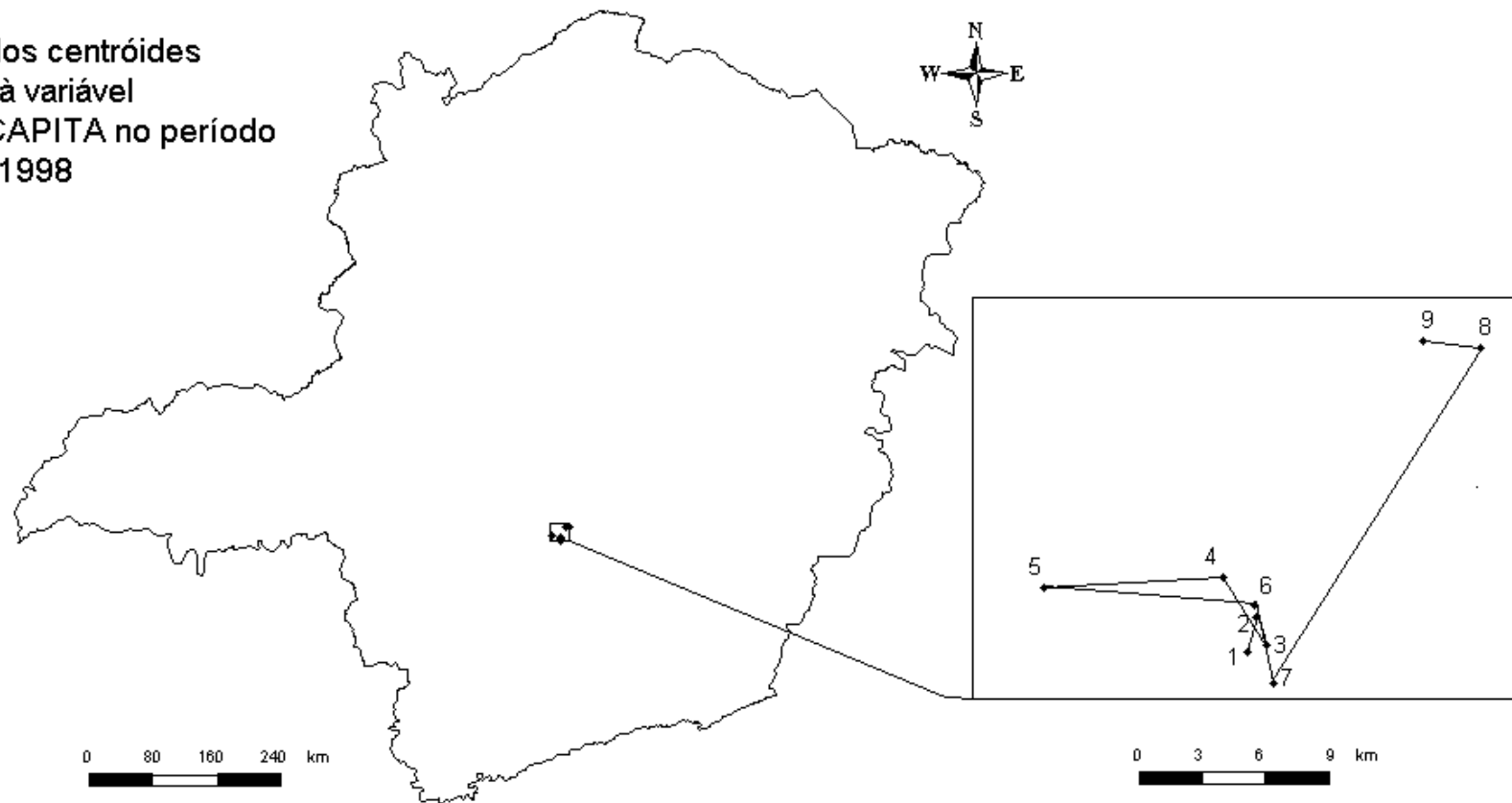
Evolução dos centróides
referentes à variável
PIB_AGRO no período
de 1990 a 1998



Fonte: PUC-Minas

MAPA 15 – Evolução dos centróides referentes à variável PIB_agro no período de 1990 a 1998

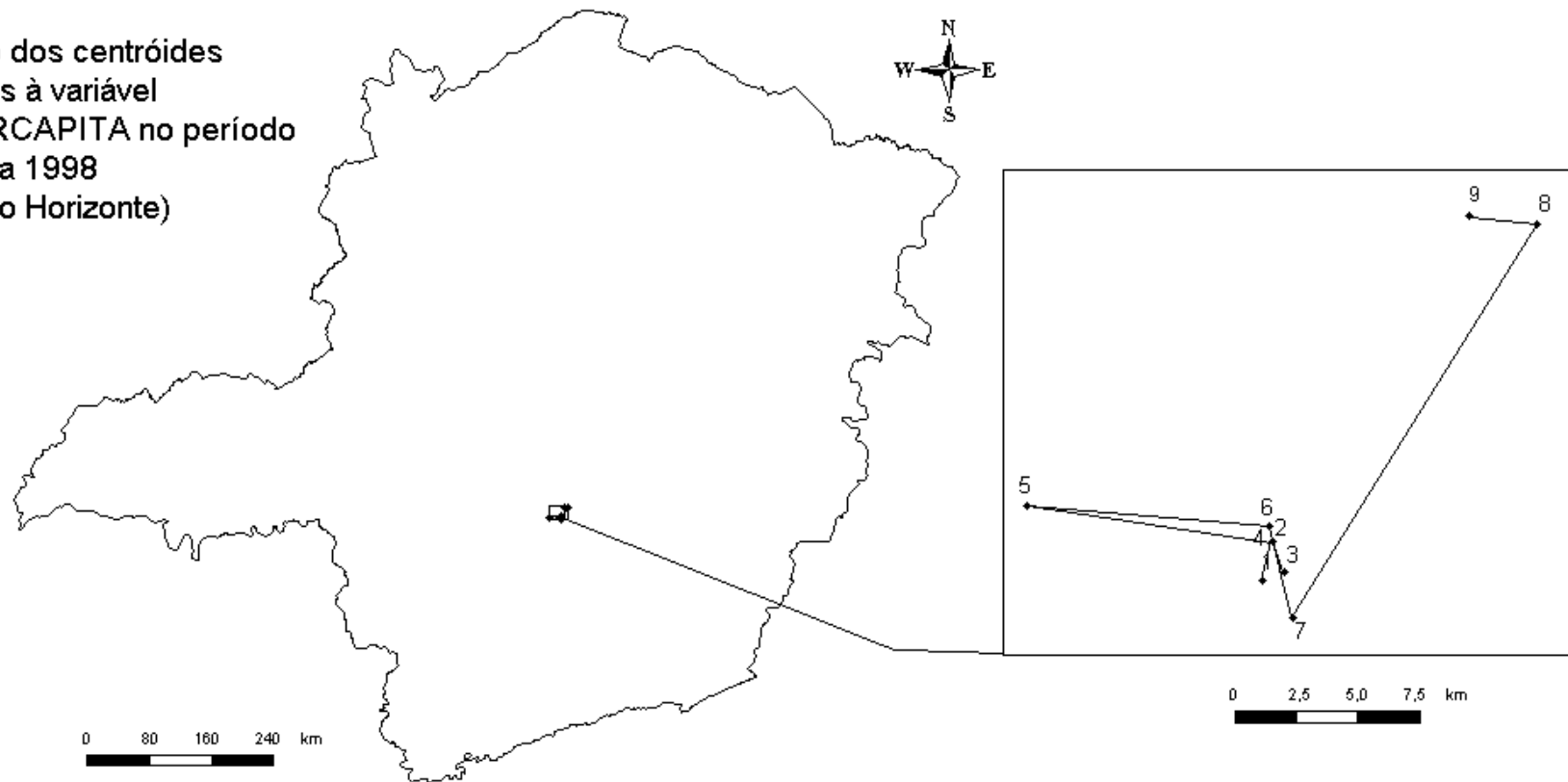
Evolução dos centróides
referentes à variável
PIB_PERCAPITA no período
de 1990 a 1998



Fonte: PUC-Minas

MAPA 16 – Evolução dos centróides referentes à variável PIB_percapta no período de 1990 a 1998

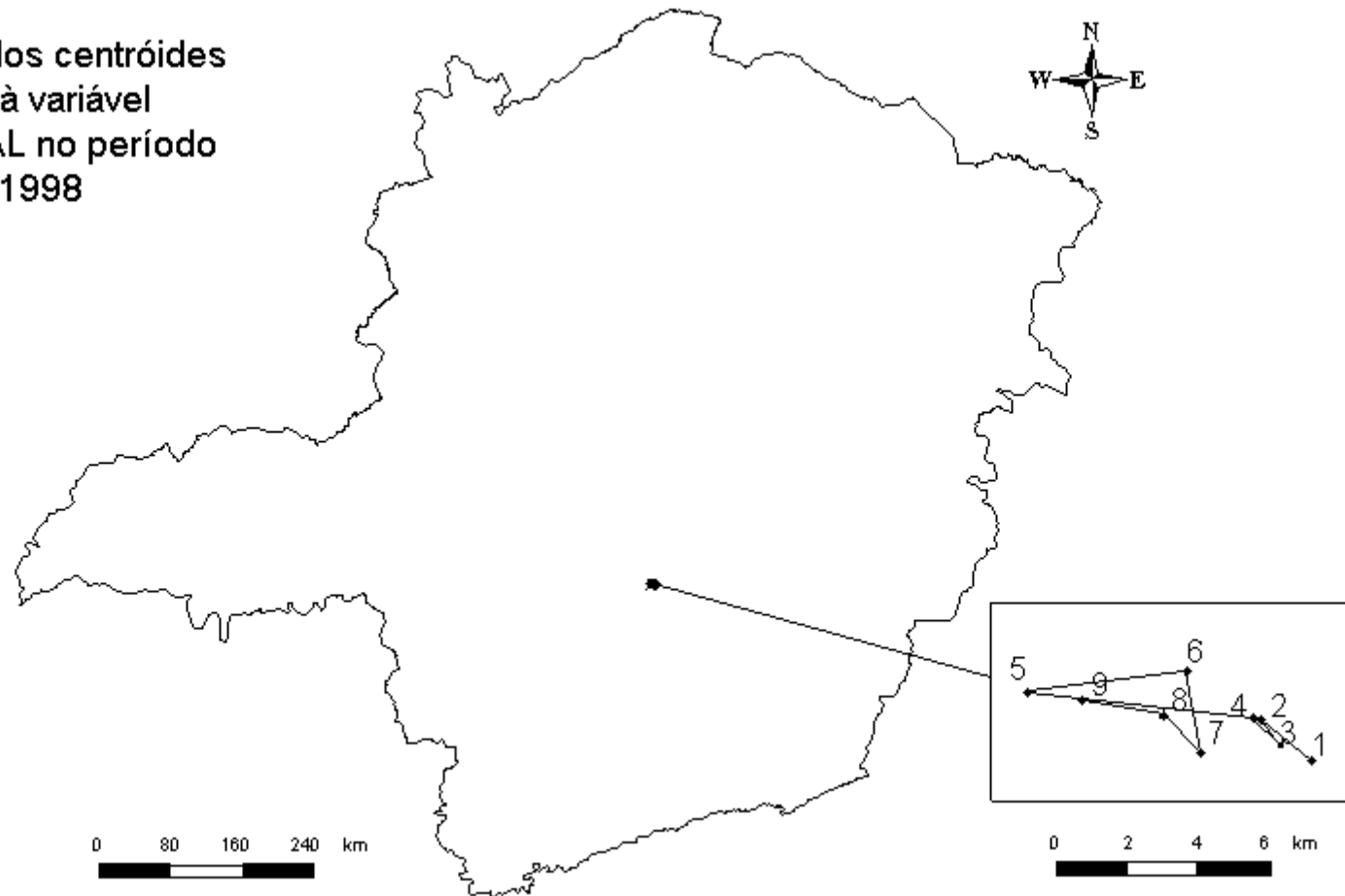
Evolução dos centróides
referentes à variável
PIB_PERCAPITA no período
de 1990 a 1998
(sem Belo Horizonte)



FONTE:PUC-Minas

MAPA 17 - Evolução dos centróides referentes à variável PIB_percapita no período de 1990 a 1998 (Sem Belo Horizonte)

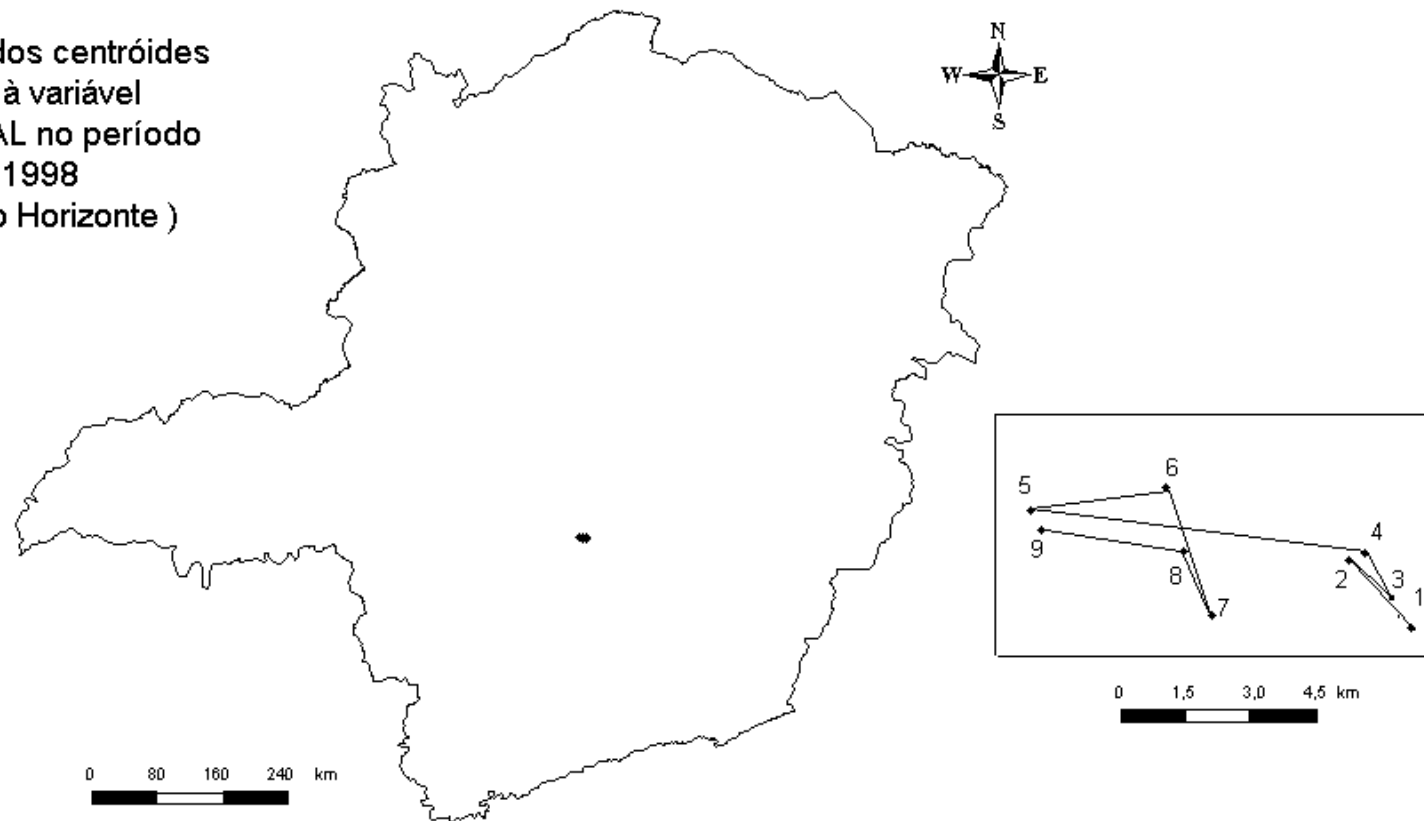
Evolução dos centróides
referentes à variável
PIB_TOTAL no período
de 1990 a 1998



Fonte: PUC-Minas

MAPA 18 – Evolução dos centróides referentes à variável PIB_total no período de 1990 a 1998

Evolução dos centróides
referentes à variável
PIB_TOTAL no período
de 1990 a 1998
(Sem Belo Horizonte)



Fonte: PUC-Minas

MAPA19 – Evolução dos centróides referentes à variável PIB_total no período de 1990 a 1998 (sem Belo Horizonte)

Ao analisar os mapas com os centróides dos PIB's agropecuário, industrial, percapita e total, do período de 1990 a 1998, percebe-se que não foram registradas grandes modificações no contexto geral do estado de Minas Gerais. Mas apesar disto, é possível constatar o aumento da representatividade da região oeste do estado.

Sabendo que Belo Horizonte possui altíssima influência no PIB do estado de Minas Gerais, fez-se a opção, quando necessário, de calcular os centróides, com e sem a presença dos dados relativos a esta capital. Com isso percebe-se com mais clareza a influência das demais regiões do estado no PIB regional.

Analisando a evolução do PIB industrial, apresentada nos mapas 13 e 14, verifica-se leve tendência para o norte de Minas Gerais, no período de 1991 a 1995, provavelmente provocada pelo fortalecimento dos municípios no entorno de Montes Claros. A partir de 1996, pode-se constatar que o centróide é atraído pelo oeste do estado, região onde estão localizadas Uberlândia e Uberaba. Veja mapas

Com relação ao PIB agropecuário, observa-se, no MAPA 15, a tendência do centróide para o oeste de Minas Gerais, provocada pela tradição agropecuária dos municípios desta região, como Uberlândia, Uberaba, Unaí, Paracatu, entre outros. No período de 1994 a 1996, este centróide apresenta breve tendência para o sul do estado, provavelmente ocasionada pelos municípios da região de Rio Claro, Três Corações e Alfenas.

Os centróides dos PIB's percapita e total mostram estabilidade na região central de Minas Gerais, sem deixar de apresentar a clara atração do "corredor" determinado pelos municípios da Grande Belo Horizonte, como Betim e Contagem, e do Vale do Rio Doce, principalmente nos anos de 1997 e 1998.

4.3 - Medidas de Variabilidade ou dispersão em distribuições espaciais de pontos

Assim como na Estatística univariada, conhecer os pontos de tendência central pode não significar muito na busca de uma boa análise de dados espaciais. Conhecer o centro da distribuição de um conjunto de pontos pode não significar muito, seja ele o Centro Médio, Centro Mediano ou outro ponto qualquer adotado como referência, se não sabemos como está a dispersão, no espaço estudado, dos demais pontos.

Portanto, merece atenção especial o estudo das medidas de dispersão de pontos espacialmente conhecidos. Diferentemente do que é feito na estatística univariada, quando mensura-se a variabilidade dos dados acima e abaixo, à direita ou à esquerda da medida central, ao trabalhar com dados espaciais, torna-se necessário mensurar esta variabilidade num entorno do ponto central adotado. Neste caso, pensa-se em raio de dispersão. Ou seja, a região centrada no Centro Médio, Centro Mediano, etc, deve transmitir uma boa idéia a respeito deste grau de dispersão dos pontos.

Buscando representar a variabilidade dos pontos, serão discutidos os conceitos de:

- Retângulo Quartílico;
- Índice de dispersão;
- Distância Padrão;
- Distância Padrão Ponderada;
- Dispersão relativa.

4.3.1 - Retângulo Quartílico

O retângulo quartílico está diretamente relacionado com a idéia do Percentil, assunto já abordado anteriormente. Em especial, aplicam-se os percentis 25, 50 e 75, também denominados quartis 1, 2 e 3.

A figura denominada retângulo quartílico é obtida quando é traçado o segmento que divide o conjunto de dados deixando $\frac{1}{4}$ dos dados abaixo e $\frac{3}{4}$ acima, ou seja o

segmento que passa pelo P_{25} (Quartil 1) e o segmento que deixa $\frac{3}{4}$ dos dados abaixo e $\frac{1}{4}$ acima dele, portanto o segmento deve passar pelo P_{75} (Quartil 3).

Isto será feito para os eixos x e y . As interseções destes segmentos determinarão o Retângulo Quartílico.

Veja tudo o que foi dito sendo aplicado aos dados do exemplo 4.1:

Primeiro vamos calcular $P_{25} = Q_1$ e $P_{75} = Q_3$, das coordenadas x e y . Para ajudar a visualização dos passos, na tabela abaixo, as coordenadas já estão ordenadas.

Mas, por este motivo, prestando atenção e voltando à TABELA 14, você perceberá que as colunas abaixo estão apresentadas como dados independentes. Por exemplo, na primeira linha encontramos as coordenadas $x = 1,3$ e $y = 1,0$, que não formam, juntas, as coordenadas de algum ponto.

TABELA 22
Modificação da tabela 14 para obtenção das medianas

i	x	y	i	x	y
1	1,3	1,0	15	6,1	5,3
2	1,8	1,3	16	6,3	5,8
3	1,9	1,3	17	7,0	6,0
4	2,4	1,4	18	7,0	7,2
5	2,7	1,8	19	7,4	7,9
6	3,0	2,0	20	7,8	8,0
7	3,2	2,1	21	8,1	8,0
8	4,0	2,3	22	8,3	8,1
9	4,1	3,2	23	8,4	8,1
10	4,2	4,0	24	9,0	8,7
11	5,6	4,0	25	9,0	8,7
12	5,8	4,2	26	9,1	8,8
13	5,9	4,3	27	9,2	9,0
14	5,9	4,3	28	9,4	9,0

Fonte: Dados Hipotéticos

Calculando as abscissas:

$$P_{25,x} = 3,2 \quad (25\% \text{ de } 28 = 7, \text{ logo } P_{25,x} \text{ é o } 7^{\text{o}} \text{ termo)} \text{ e}$$

$$P_{75,x} = 8,1.$$

Enquanto, para as ordenadas:

$$P_{25,y} = 2,1 \text{ e}$$

$$P_{75,y} = 8,0.$$

Traçando os segmentos por estas coordenadas, fica determinado o retângulo quartílico, apresentado na FIGURA 4.

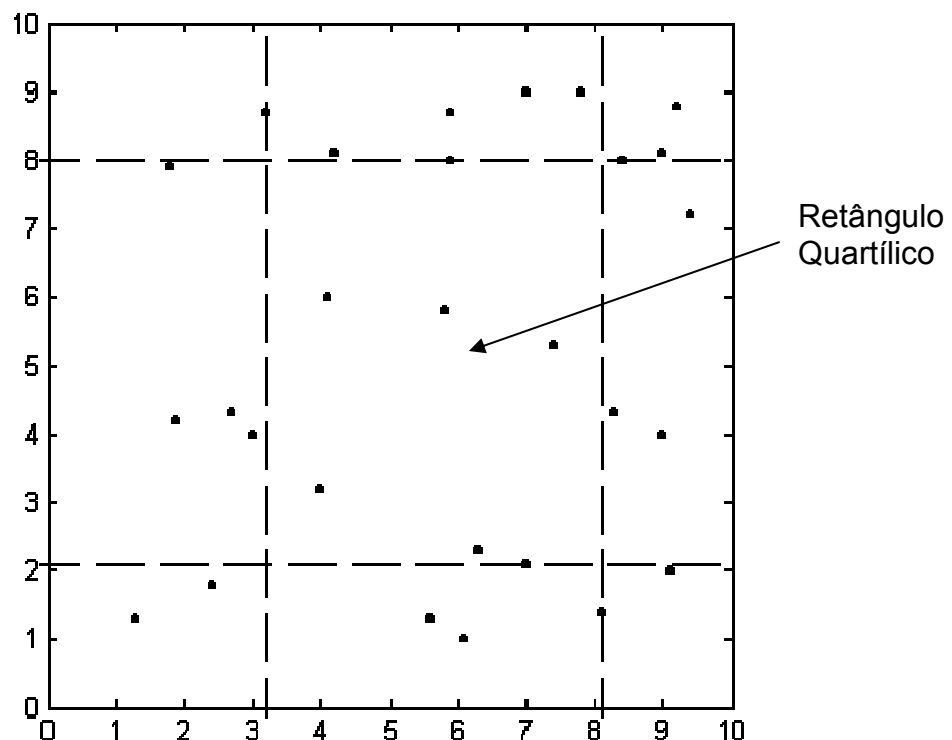


FIGURA 4: Retângulo Quartílico

Efetuada a contagem constata-se que 7 pontos pertencem ao Retângulo Quartílico. Mas, isto representa muito ou pouco? Na realidade, mais importante do que esta contagem é a razão existente entre a área do Retângulo Quartílico e o todo.

Sendo assim, define-se Índice de Dispersão.

4.3.2 - Índice de Dispersão

Na tentativa de dar significado para o Retângulo Quartílico, obtido no item anterior, calcula-se o Índice de Dispersão (ID), que indicará a magnitude deste retângulo em relação ao todo.

O índice de dispersão será calculado usando regra de três simples, considerando a área total como o inteiro, o um:

Usando a notação: Área total: A

Área do Retângulo Quartílico: A_{RQ}

Índice	Área
1	A
ID	A_{RQ}

$$\therefore \frac{1}{ID} = \frac{A}{A_{RQ}}$$

$$\therefore \boxed{ID = \frac{A_{RQ}}{A}} \quad (4.3.2.1)$$

Para o problema do exemplo 4.1:

$$A = 10 \cdot 10 = 100$$

$$A_{RQ} = (8,1 - 3,2) \cdot (8,0 - 2,1) = 28,91$$

Portanto,

$$ID = \frac{28,91}{100} = 0,2891.$$

Este valor indica que 28,91% da área total, onde os dados estão distribuídos, está contida no Retângulo Quartílico.

Observe que o Índice de Dispersão varia de 0 a 1. O ID será igual a zero se a área do Retângulo Quartílico for igual a zero, o que indica concentração máxima, ou seja, os pontos estão todos sobrepostos. O ID será igual a 1 se a área do Retângulo Quartílico for igual à área total, o que indica máxima dispersão. Mas, isto ocorrerá apenas se todos os pontos estiverem margeando a região total. O valor 0,25 indica uma distribuição uniforme dos pontos.

Para os dados do exemplo 4.1, foi obtido $ID = 0,2891$, logo os pontos apresentam uma distribuição aproximadamente uniforme.

4.3.3 - Distância Padrão

A medida estatística mais importante e mais útil para medir a variação de um conjunto de valores é o Desvio Padrão. Para chegar ao seu conceito, toma-se como referência as diferenças ao quadrado de cada valor em relação à média, veja o item 3.2.3. Na Estatística Espacial, a Distância Padrão é o seu equivalente.

A Distância Padrão ou Raio Padrão representa a variabilidade de um conjunto de pontos em torno do ponto central, que pode ser o Centro Médio, Centro Médio Ponderado, etc. Porém, trabalhando espacialmente, os pontos podem apresentar valores diversos para as distâncias, mas, com um novo aspecto, esta distância pode estar em todas as direções. Portanto, é necessário pedir ajuda à Geometria, ao usar o conceito de coordenadas retangulares e o teorema de Pitágoras, para determinar uma região circular que represente esta variabilidade do conjunto de pontos em torno do ponto central.

Isto será feito ao determinar a distância do ponto P ao Centro Médio de um conjunto de pontos, conforme FIGURA 5:

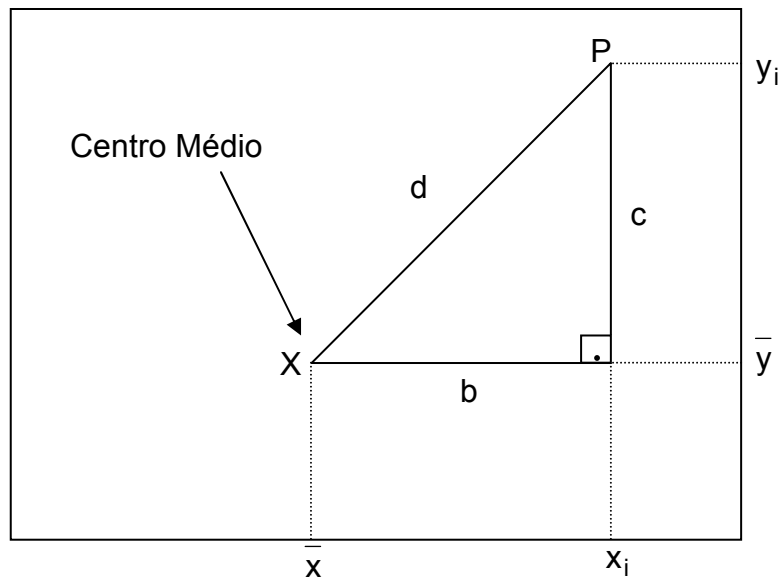


FIGURA 5: Construção auxiliar para obtenção da Distância Padrão

Dados: d : distância de P até X
 b : distância horizontal entre P e X
 c : distância vertical entre P e X

Medidas obtidas:

$$b = x_i - \bar{x}$$

$$c = y_i - \bar{y}$$

O teorema de Pitágoras diz que, num triângulo retângulo, o quadrado da hipotenusa é igual à soma dos quadrados dos catetos, logo:

$$d^2 = b^2 + c^2 \quad (4.3.3.1)$$

$$d^2 = (x_i - \bar{x})^2 + (y_i - \bar{y})^2 \quad (4.3.3.2)$$

$$d = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \quad (4.3.3.3)$$

Portanto, de modo semelhante ao que foi feito para o Desvio Padrão, tomando-o

como resultado da operação $\sqrt{\frac{\sum d^2}{n}}$, será definido o conceito de Distância Padrão:

$$\text{Distância Padrão} = D_p = \sqrt{\frac{\left(\sqrt{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}\right)^2}{n}} = \sqrt{\frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n}}.$$

(4.3.3.4)

Esta fórmula pode ser apresentada de modo simplificado, se aplicadas algumas propriedades matemáticas:

$$\text{Distância Padrão} = \sqrt{\left(\frac{\sum x_i^2}{n} - \bar{x}^2\right) + \left(\frac{\sum y_i^2}{n} - \bar{y}^2\right)}.$$

(4.3.3.5)

Como exemplo, será feito o cálculo da Distância Padrão para os dados da TABELA 14, acrescentando mais duas colunas, para facilitar os cálculos:

TABELA 23
Complementação da tabela 14 para cálculo da Distância Padrão

i	x	y	x ²	y ²
1	1,30	1,30	1,69	1,69
2	1,80	7,90	3,24	62,41
3	1,90	4,20	3,61	17,64
4	2,40	1,80	5,76	3,24
5	2,70	4,30	7,29	18,49
6	3,00	4,00	9,00	16,00
7	3,20	8,70	10,24	75,69
8	4,00	3,20	16,00	10,24
9	4,10	6,00	16,81	36,00
10	4,20	8,10	17,64	65,61
11	5,60	1,30	31,36	1,69
12	5,80	5,80	33,64	33,64
13	5,90	8,70	34,81	75,69
14	5,90	8,00	34,81	64,00
15	6,10	1,00	37,21	1,00
16	6,30	2,30	39,69	5,29
17	7,00	2,10	49,00	4,41
18	7,00	9,00	49,00	81,00
19	7,40	5,30	54,76	28,09
20	7,80	9,00	60,84	81,00
21	8,10	1,40	65,61	1,96
22	8,30	4,30	68,89	18,49
23	8,40	8,00	70,56	64,00
24	9,00	4,00	81,00	16,00
25	9,00	8,10	81,00	65,61
26	9,10	2,00	82,81	4,00
27	9,20	8,80	84,64	77,44
28	9,40	7,20	88,36	51,84
Somas	163,90	145,80	1139,27	982,16

Fonte: Dados hipotéticos

$$D_p = \sqrt{\left(\frac{1139,27}{28} - 5,8536^2\right) + \left(\frac{982,16}{28} - 5,2071^2\right)}$$

$$D_p = 3,7930.$$

A Distância Padrão, que possui a mesma unidade de medida utilizada para as coordenadas x e y , representa o raio do círculo centrado no Centro Médio, que é proporcional ao grau de dispersão dos pontos. Para conjuntos de pontos semelhantes, quanto maior o raio maior a dispersão. Observe que cada valor influencia o cálculo desta medida, inclusive os valores extremos, sejam eles superiores ou inferiores. Deve-se dedicar atenção especial a este fato, principalmente, porque as distâncias são elevadas ao quadrado.

4.3.4 - Distância Padrão Ponderada

Quando existe a preocupação em determinar um ponto de tendência central, para um conjunto de pontos, dedica-se atenção especial à intensidade com que determinado fenômeno ocorre em cada ponto do conjunto. Quando este interesse surgiu foi dada a definição de Centro Médio Ponderado.

De modo semelhante, será feita a definição de Distância Padrão Ponderada (D_{pp}), que busca transmitir o grau de dispersão dos dados.

Considerando que cada fenômeno ocorre com intensidade p_i num dado ponto (x_i, y_i) , definimos Distância Padrão Ponderada:

$$\text{Distância Padrão Ponderada} = D_{pp} = \sqrt{\frac{\sum p_i (x_i - \bar{x}_p)^2}{\sum p_i} + \frac{\sum p_i (y_i - \bar{y}_p)^2}{\sum p_i}} \quad (4.3.4.1)$$

Com o objetivo de facilitar o cálculo da D_{pp} para os pontos da TABELA 14, serão criadas algumas colunas auxiliares, obtendo a TABELA 24. Lembre-se que já foram calculados $\bar{x}_p = 6,2261$ e $\bar{y}_p = 5,7763$.

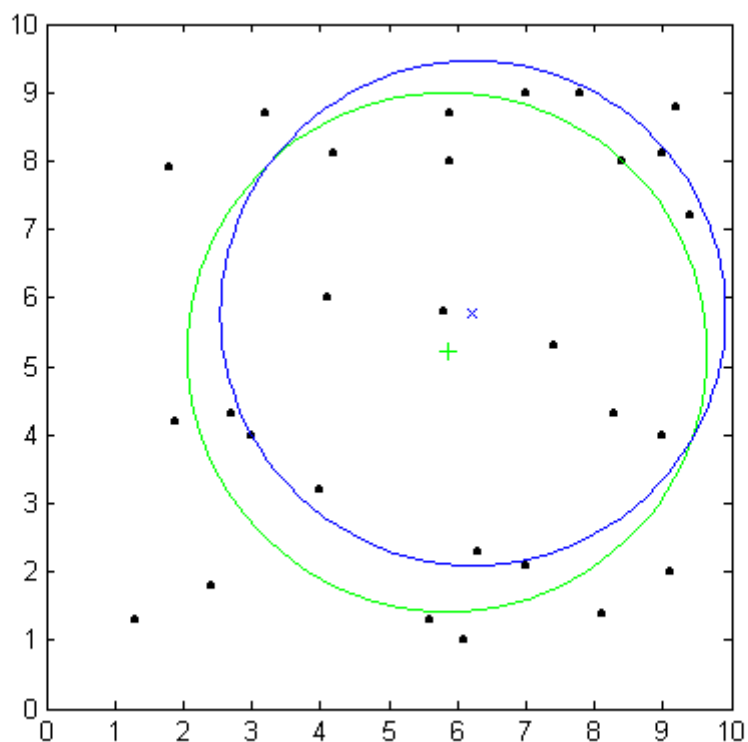
TABELA 24
Complementação da tabela 14 para cálculo da Distância Padrão Ponderada

i	p _i	x _i	x _i - \bar{x}_p	(x _i - \bar{x}_p) ²	p _i · (x _i - \bar{x}_p) ²	y _i	y _i - \bar{y}_p	(y _i - \bar{y}_p) ²	p _i · (y _i - \bar{y}_p) ²
1	21	1,30	-4,93	24,27	509,60	1,30	-4,48	20,04	420,78
2	30	1,80	-4,43	19,59	587,71	7,90	2,12	4,51	135,30
3	38	1,90	-4,33	18,72	711,18	4,20	-1,58	2,48	94,42
4	25	2,40	-3,83	14,64	365,98	1,80	-3,98	15,81	395,27
5	27	2,70	-3,53	12,43	335,70	4,30	-1,48	2,18	58,85
6	29	3,00	-3,23	10,41	301,82	4,00	-1,78	3,16	91,50
7	44	3,20	-3,03	9,16	402,92	8,70	2,92	8,55	376,11
8	32	4,00	-2,23	4,96	158,58	3,20	-2,58	6,64	212,39
9	38	4,10	-2,13	4,52	171,77	6,00	0,22	0,05	1,90
10	40	4,20	-2,03	4,11	164,20	8,10	2,32	5,40	215,98
11	31	5,60	-0,63	0,39	12,15	1,30	-4,48	20,04	621,16
12	41	5,80	-0,43	0,18	7,44	5,80	0,02	0,00	0,02
13	42	5,90	-0,33	0,11	4,47	8,70	2,92	8,55	359,02
14	46	5,90	-0,33	0,11	4,89	8,00	2,22	4,94	227,46
15	38	6,10	-0,13	0,02	0,60	1,00	-4,78	22,81	866,90
16	34	6,30	0,07	0,01	0,19	2,30	-3,48	12,08	410,88
17	29	7,00	0,77	0,60	17,37	2,10	-3,68	13,52	391,94
18	73	7,00	0,77	0,60	43,72	9,00	3,22	10,39	758,63
19	46	7,40	1,17	1,38	63,39	5,30	-0,48	0,23	10,44
20	58	7,80	1,57	2,48	143,68	9,00	3,22	10,39	602,75
21	24	8,10	1,87	3,51	84,28	1,40	-4,38	19,15	459,65
22	42	8,30	2,07	4,30	180,64	4,30	-1,48	2,18	91,54
23	45	8,40	2,17	4,73	212,66	8,00	2,22	4,94	222,52
24	44	9,00	2,77	7,69	338,56	4,00	-1,78	3,16	138,83
25	57	9,00	2,77	7,69	438,59	8,10	2,32	5,40	307,78
26	38	9,10	2,87	8,26	313,85	2,00	-3,78	14,26	541,90
27	50	9,20	2,97	8,84	442,20	8,80	3,02	9,14	457,14
28	40	9,40	3,17	10,07	402,95	7,20	1,42	2,03	81,08
		163,9				145,8			
Somas	1102	0	-10,43	183,76	6421,09	0	-15,94	232,03	8552,13

Fonte: Dados hipotéticos

$$D_{pp} = \sqrt{\frac{6421,09}{1102} + \frac{8552,13}{1102}} = 3,6861$$

A FIGURA 6 apresenta as duas últimas medidas encontradas:



- + Centro Médio
- x Centro Médio Ponderado
- Distância Padrão
- Distância Padrão Ponderada

FIGURA 6: Distância Padrão e Distância Padrão Ponderada

Observe que o raio da Distância Padrão é maior que o raio da Distância Padrão Ponderada, o que significa que a intensidade de ocorrência do fenômeno estudado está distribuída mais uniformemente que a distribuição dos pesos.

Estas medidas são de grande importância em estudos geográficos, quando se busca analisar a distribuição da ocorrência de determinado fenômeno em alguma região, como distribuição da produção agrícola, do atendimento escolar e de saúde, da vegetação, do transporte, dos serviços comerciais, etc.

4.3.5 - Dispersão Relativa

Uma Distância Padrão de 10km indica alto grau de dispersão do fenômeno estudado? Depende. Se o referencial for um país inteiro como o Brasil, provavelmente é possível considerá-la muito baixa, mas, se o referencial for uma cidade de porte médio, estes mesmos 10km de Distância Padrão, provavelmente, estarão indicando altíssimo grau de dispersão. Como fazer esta análise sem estar preso exclusivamente a uma interpretação subjetiva?

Para responder a este tipo de questão é dada a definição de Dispersão Relativa (Dr). Com este objetivo, será usada, mais uma vez, uma regra de três simples, buscando determinar qual a representatividade do círculo que ficou definido pela Distância Padrão e pelo ponto de tendência central, em relação ao todo ou em relação à Distância Padrão de algum outro fenômeno já avaliado na mesma região, ou seja, o país, estado, município, bairro, etc estudado.

No primeiro caso, comparando a distância padrão com o todo, é preciso conhecer a área da região estudada e descobrir o raio de uma circunferência com mesma área, o que não é difícil de ser feito. Veja um exemplo:

Exemplo 4.4 Determinar o raio de um círculo de área igual à 300 m².

Sabe-se que a área do círculo é dada por $A_{\text{círculo}} = \pi \cdot r^2$, onde π vale aproximadamente 3,14 e r é o raio da circunferência. Portanto,

$$A = \pi \cdot r^2 \Rightarrow r^2 = \frac{A}{\pi} \Rightarrow r = \sqrt{\frac{A}{\pi}}. \quad (4.3.5.1)$$

Usando a fórmula obtida, o exemplo proposto pode ser resolvido:

$$r = \sqrt{\frac{300}{3,14}} = 9,77 \text{ metros.}$$

Sendo assim, já sabendo como obter o raio de um círculo com área igual à região estudada, pode-se aplicar a regra de três que fornecerá a Dispersão do fenômeno avaliado em relação à área total da região estudada:

Adotando:

- r** raio da circunferência de área igual à da região estudada
- Dp** Distância padrão (também pode ser Distância Padrão Ponderada)
- Dr** Dispersão relativa

Observação: **r** e **Dp** devem estar na mesma unidade de medida

Raio	fração
r	1
Dp	Dr

$$\therefore \frac{r}{Dp} = \frac{1}{Dr}$$

$$\therefore r \cdot Dr = 1 \cdot Dp$$

$$\therefore \boxed{Dr = \frac{Dp}{r}}$$

(4.3.5.2)

Concluindo, define-se Dispersão relativa, para este caso, como a razão entre a Distância Padrão e o raio da circunferência com área igual à região em que se estuda o fenômeno. Quanto mais próximo de 1 maior a dispersão do fenômeno.

Mas, se o objetivo é comparar a Distância Padrão da distribuição do fenômeno estudado, com a de outro já conhecido, segue-se raciocínio análogo ao que foi feito na etapa anterior para definir:

$$Dr = \frac{\text{distância padrão da variável em estudo}}{\text{distância padrão da variável tomada como referencial}} \quad (4.3.5.3)$$

Neste caso, são avaliadas situações como a distribuição de farmácias na cidade em relação à distribuição da população nesta mesma cidade, ou a distribuição da população de uma sub-região de um estado com a distribuição da população no estado como um todo.

Assim como o Coeficiente de Variação estudado na unidade 3.2.4, que possui significado semelhante, a Dispersão Relativa também é adimensional, já que a Distância Padrão precisa estar na mesma unidade de medida de r.

Logo, com a Dispersão Relativa, é possível comparar as distribuições do fenômeno estudado em regiões diferentes, já que esta medida determina o grau de dispersão dos dados.

Exemplo 4.5 Comparar a dispersão de um fenômeno em duas regiões com áreas iguais a $A_1 = 985\text{km}^2$ e $A_2 = 1756\text{km}^2$, e com medidas de distância padrão, respectivamente iguais a $Dp_1 = 5 \text{ km}$ e $Dp_2 = 6,2 \text{ km}$.

Logo,

$$r_1 = \sqrt{\frac{A_1}{\pi}} = 17,7069 \Rightarrow DR_1 = \frac{5}{17,7069} = 0,2824$$

$$r_2 = \sqrt{\frac{A_2}{\pi}} = 23,6422 \Rightarrow DR_2 = \frac{6,2}{23,6422} = 0,2622$$

Observe que, apesar de Dp_1 ser menor que Dp_2 , a dispersão do fenômeno na segunda região é um pouco menor que na primeira. Ou seja, a distribuição do fenômeno está um pouco mais concentrada na segunda região.

Exemplo 4.6 Supondo que a Distância Padrão da distribuição da população em um estado seja de 45 km e de uma subregião I seja de 8,7 km, comparar a dispersão da distribuição da população de uma subregião II, com Distância Padrão de 4,3 km em outro estado que apresenta Distância Padrão igual 34km.

$$Dr_I = \frac{8,7}{45} = 0,1933$$

$$Dr_{II} = \frac{4,3}{34} = 0,1265$$

Ou seja, a primeira região possui dispersão maior que a segunda, em relação à distribuição da população em cada estado de origem.

5 - EXEMPLOS RESOLVIDOS USANDO O EXCEL

5.1 - Usando o arquivo indcrim.xls com os índices de criminalidade

- Criar uma pasta no diretório c:\ com o nome de “quantgeo”
- Fazer o download do arquivo indcrim_mg.xls, salvando-o na pasta quantgeo, criada no passo anterior. Veja figura 7.

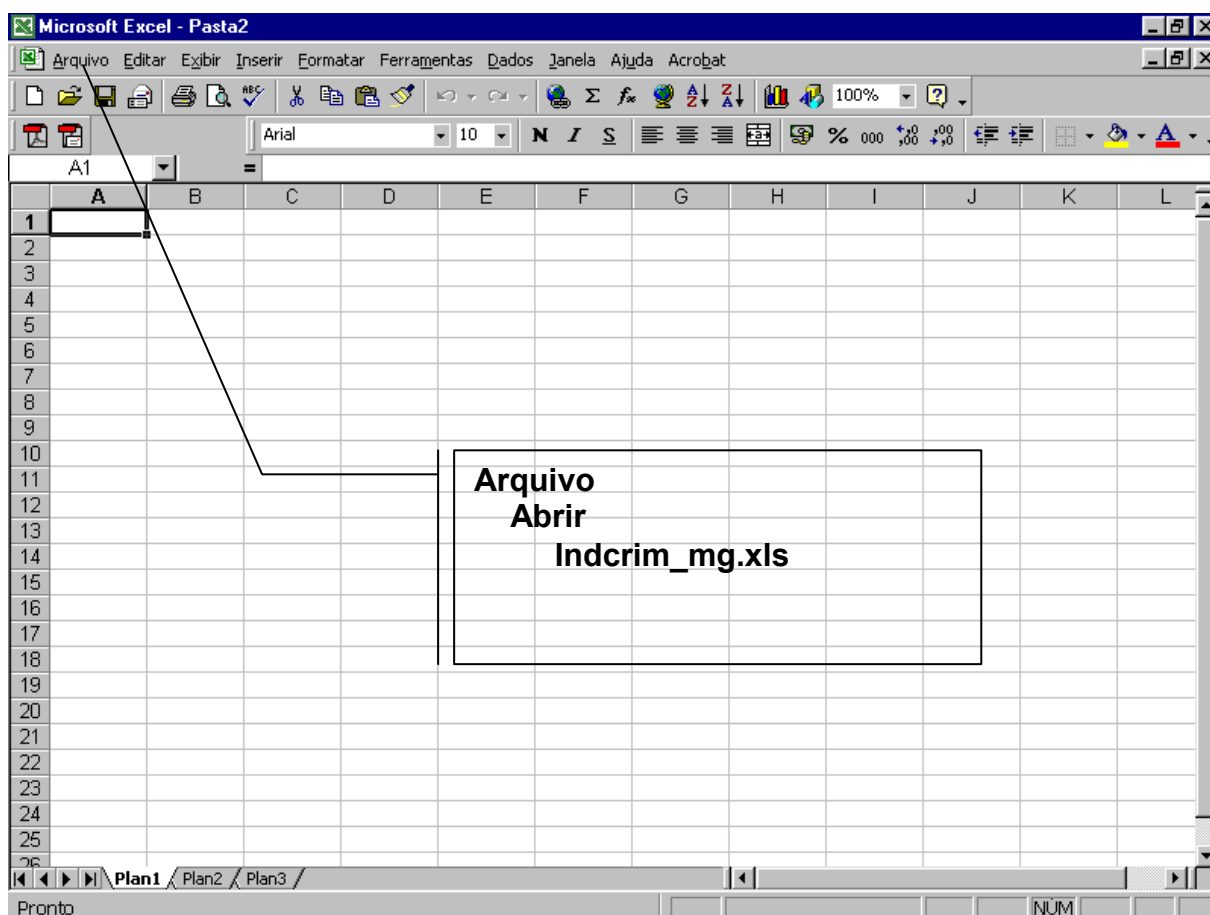


FIGURA 7: Abrindo o arquivo Indcrim_mg.xls

- Você verá os dados em uma tela como a da FIGURA 8:

i	Municípios	Índice de criminalidade de
1	Abadia dos Dourados	17,28
2	Água Comprida	31,72
3	Araguari	31,57
4	Araporá	29,23
5	Arapuá	8,73
6	Araxá	25,52
7	Cachoeira Dourada	43,86
8	Campina Verde	19,22
9	Campo Florido	15,59
10	Campos Altos	20,89
11	Canápolis	10,39
12	Capinópolis	18,14
13	Carmo do Paranaíba	15,8
14	Carneirinho	14,71
15	Cascalho Rico	13,24
16	Centralina	13,29
17	Comendador Gomes	7,13
18	Conceição das Alagoas	35,05
19	Conquista	6,66

FIGURA 8: Planilha dados

- Observe que os valores dos índices de criminalidade vão da célula C5 à C68, o que é indicado no Excel como C5:C68.

5.2 Cálculo da Amplitude

- Vá para a célula B70 e digite **Maior valor**
- Vá para a célula C70 e digite: **=MÁXIMO(C5:C68)**
 - Se quiser, você pode digitar **= MÁXIMO(** , selecionar as células com os valores dos índices de criminalidade, fechar os parênteses e teclar **ENTER**.
- Vá para a célula B71 e digite **Menor valor**
- Vá para a célula C71 e digite **=MÍNIMO(C5:C68)**
- Vá para a célula B72 e digite **=C70 – C71**

Você encontrará os resultados na FIGURA 9:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
58	54	Santa Rosa da Serra	4,5											
59	55	Santa Vitória	18,13											
60	56	São Francisco de Sales	30,01											
61	57	São Gotardo	9,87											
62	58	Serra do Salitre	17,08											
63	59	Tapira	13,63											
64	60	Tiros	9,59											
65	61	Tupaciguara	30,55											
66	62	Uberaba	42,34											
67	63	Uberlândia	35,96											
68	64	Veríssimo	30,52											
69														
70		Maior valor	43,86											
71		Menor valor	4,5											
72		Amplitude	39,36											
73														

FIGURA 9: Amplitude

*** Lembre-se: antes de qualquer fórmula é preciso digitar “ = ”

5.3 - Fórmula de Sturges

Em primeiro lugar, é preciso saber quantos valores existem (tamanho da amostra ou população), o valor de N. Para isto usaremos a função CONT.VALORES, que contará quantas células não-vazias existem no intervalo especificado:

- Em B73 digite **Quantidade de valores (N)**;
- Em C73 digite **=CONT.VALORES(C5:C68)**;
- Aplicação da fórmula de Sturges:
 - A função **LOG(CÉLULA;BASE)** calcula o logaritmo do valor contido em CÉLULA, na base BASE especificada (em nosso caso, base 2);
 - Em B74 digite **Número de classes**;
 - Vá para a célula C74 e digite **=1+LOG(C73;2)**

5.4 - Amplitude de classe até final da última classe

O comando **ARREDONDAR.PARA.CIMA(VALOR;CASAS)**, arredonda o conteúdo de VALOR, que pode ser uma fórmula, sempre para cima de acordo com o número de CASAS decimais solicitadas.

- Vá para a célula B75 e digite **Amplitude de classe (h)**
- Em C75, digite **ARREDONDAR.PARA.CIMA(C72/C74 ; 2)**

Para calcular o limite superior da última classe:

- Na célula B76, digite **Final da última classe**

- Na célula C76 digite $=C71+ C74*C75$ (equivalente a $=4,50+7*5,63$)

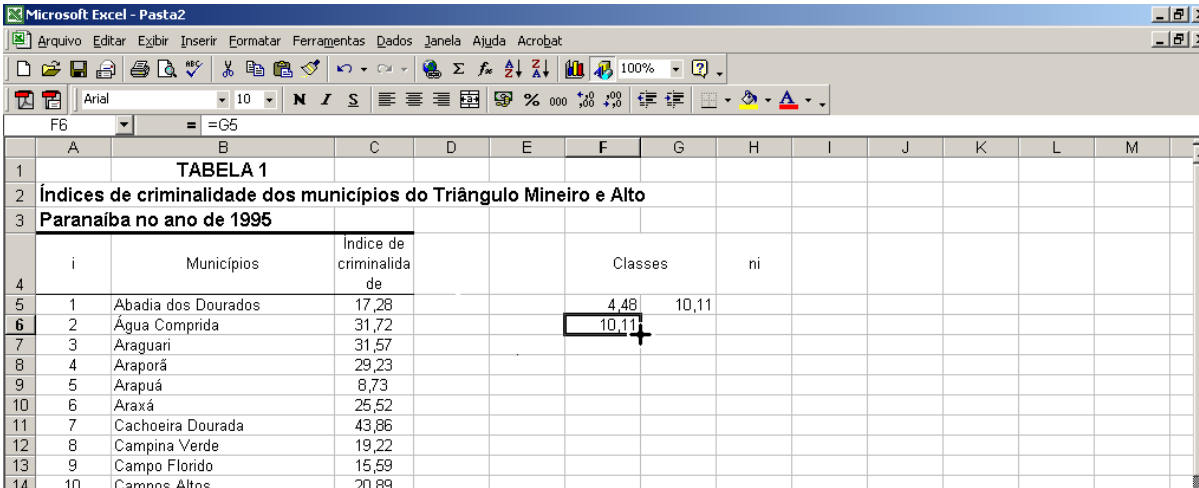
5.5 - Contagem de dados

Inicialmente, deve-se determinar as classes, definindo os limites inicial e final de cada classe. A determinação das classes será iniciada na célula F5. Pretende-se colocar na coluna F o início e na coluna G o final de cada classe. Lembre-se que, neste exemplo, a amplitude de cada classe é 5,63.

- Em F5, digite **4,48**, em G5 inclua a fórmula $=F5+5,63$.
- Como F6 deve ser igual ao final da classe anterior, digite $=G5$.
- Como o padrão de variação das classes já está determinado, pode-se apenas copiar a fórmula contida em F6 de F7 a F11 e a fórmula em G5 de G6 a G11.

ATENÇÃO: É IMPORTANTÍSSIMO SABER como copiar fórmulas de uma célula para outra, pois isto facilitará muito seu trabalho:

- Vá para a célula F6, aponte o cursor do mouse para o canto direito inferior desta célula, até que se transforme em uma pequena cruz em negrito, como na FIGURA 10.



i	Municípios	Índice de criminalidade	Classes	ni
1	Abadia dos Dourados	17,28	4,48	10,11
2	Água Comprida	31,72	10,11	
3	Araguari	31,57		
4	Araporá	29,23		
5	Arapuá	8,73		
6	Araxá	25,52		
7	Cachoeira Dourada	43,86		
8	Campina Verde	19,22		
9	Campo Florido	15,59		
10	Campos Altos	20,89		

FIGURA 10: Como copiar fórmulas

- Arraste o cursor do mouse até a célula F11, com o botão esquerdo do mouse pressionado. Veja o resultado na FIGURA 11.

i	Municípios	Índice de criminalidade	Classes	ni
1	Abadia dos Dourados	17,28	4,48	10,11
2	Água Comprida	31,72	10,11	
3	Araguari	31,57	0,00	
4	Araporá	29,23	0,00	
5	Arapuá	8,73	0,00	
6	Araxá	25,52	0,00	
7	Cachoeira Dourada	43,86	0,00	
8	Campina Verde	19,22		
9	Campo Florido	15,69		
10	Campos Altos	20,89		

FIGURA 11: Resultado parcial da cópia de fórmula de F6 para F7 a F11

- Faça o mesmo copiando a fórmula contida em G5 de G6 a G11.

Com as classes definidas, pode-se contar o número de valores em cada uma delas. Para isto, siga os seguintes passos:

- Vá para a célula H4 e digite **ni**, para título da coluna;
- selecione as células H5 até H11;
- digite a fórmula **=FREQUÊNCIA(C5:C68;G5:G11)**;
- tecla simultaneamente **Ctrl + shift + Enter**;
- Vá para a célula H12 e clique no ícone Σ e tecla **ENTER**, que apresentará a soma dos valores das célula H5:H11. Neste caso, igual a 64, pois existem 64 municípios.

Explicação da função FREQUÊNCIA:

C5:C68 indica quais são as células que contêm os dados

G5:G11 indica a matriz onde serão escritos os resultados

Ctrl + shift + Enter indica que está sendo efetuada uma fórmula matricial (preencherá várias células com um só comando).

Veja o resultado obtido na FIGURA 12:

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	TABELA 1													
2	ces de criminalidade dos municípios do Triângulo Mineiro e Alto													
3	anaíba no ano de 1995													
4	Municípios	Índice de criminalidade				Classes	ni							
5	Abadia dos Dourados	17,28			4,48	10,11	12							
6	Água Comprida	31,72			10,11	15,74	13							
7	Araguari	31,57			15,74	21,37	15							
8	Araporá	29,23			21,37	27	8							
9	Arapuá	8,73			27	32,63	9							
10	Araxá	25,52			32,63	38,26	2							
11	Cachoeira Dourada	43,86			38,26	43,89	5							
12	Campina Verde	19,22					64							
13	Campo Florido	15,69												
14	Campos Altos	20,89												
15	Canápolis	10,39												

FIGURA 12: freqüência da classe

5.6 - Cálculo das freqüências relativas

- Vá para a célula I4 e digite **fi**;
- Em I5 digite **=H5/H\$12**;
- Copie esta fórmula para as células I6 a I11;
- Faça o somatório da coluna, deixando o resultado na célula I12. Lembre-se, basta tornar a célula I12 ativa, clicar no ícone Σ e teclar **Enter**;

Veja o resultado na FIGURA 13.

5.7 - Freqüência acumulada da classe

- Vá para a H4 e digite **Fi**;
- Em J5 digite **=H5**;
- Em J6 digite **=H6 + J5**;
- Como o padrão da fórmula já está definido basta copiar esta fórmula para as células J7 a J11.

De modo análogo, construa a coluna K, para a Freqüência relativa acumulada da classe, Fi.

Veja o resultado na FIGURA 13.

5.8 - Ponto médio

- Vá para a célula L4 e digite xi;
- Em L5 digite $=(F5+G5)/2$;
- Copie a fórmula digitada em L5, nas células L6 a L11;

Veja o resultado na FIGURA 13.

	Municípios	Índice de criminalidade	Classes		ni	fi	Ni	Fi	xi
5	Abadia dos Dourados	17,28	4,48	10,11	12	0,19	12	0,19	7,295
6	Água Comprida	31,72	10,11	15,74	13	0,20	25	0,39	12,925
7	Araguari	31,57	15,74	21,37	15	0,23	40	0,63	18,555
8	Araporá	29,23	21,37	27	8	0,13	48	0,75	24,185
9	Arapuá	8,73	27	32,63	9	0,14	57	0,89	29,815
10	Araxá	25,52	32,63	38,26	2	0,03	59	0,92	35,445
11	Cachoeira Dourada	43,86	38,26	43,89	5	0,08	64	1,00	41,075
12	Campina Verde	19,22			64	1,00			
13	Campo Florido	15,59							
14	Campos Altos	20,89							
15	Canápolis	10,39							

FIGURA 13: Freqüências acumuladas e pontos médios das classes.

5.9 - Comandos básicos para construção dos gráficos

As colunas F e G já foram preenchidas com os limites de classe. Mas, para construção dos gráficos com uma melhor estética vamos criar algumas células auxiliares, denominadas rótulos das classes. Para tanto, siga os passos:

- Vá para a célula F15 e digite $=\text{CONCATENAR}(F5;"|---";G5)$
- Copie a fórmula da célula F15 de F16 a F22;

Veja o resultado na FIGURA 14.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

TABELA 1										
Índices de criminalidade dos municípios do Triângulo Mineiro e Alto Paranaíba no ano de 1995										
i	Municípios	Índice de criminalidade	Classes		ni	fi	Ni	Fi	xi	
1	Abadia dos Dourados	17,28	4,48	10,11	12	0,19	12	0,19	7,295	
2	Água Comprida	31,72	10,11	15,74	13	0,20	25	0,39	12,925	
3	Araguari	31,57	15,74	21,37	15	0,23	40	0,63	18,555	
4	Araporá	29,23	21,37	27,00	8	0,13	48	0,75	24,185	
5	Arapuá	8,73	27,00	32,63	9	0,14	57	0,89	29,815	
6	Araxá	25,52	32,63	38,26	2	0,03	59	0,92	35,445	
7	Cachoeira Dourada	43,86	38,26	43,89	5	0,08	64	1,00	41,075	
8	Campina Verde	19,22			64					
9	Campo Florido	15,59								
10	Campos Altos	20,89	Rótulos							
11	Canápolis	10,39	4,48 ---10,11							
12	Capinópolis	18,14	10,11 ---15,74							
13	Carmo do Paranaíba	15,80	15,74 ---21,37							
14	Carneirinho	14,71	21,37 ---27							
15	Cascalho Rico	13,24	27 ---32,63							
16	Centralina	13,29	32,63 ---38,26							
17	Comendador Gomes	7,13	38,26 ---43,89							
18	Conceição das Alagoas	35,05								
19	Conquista	6,66								

FIGURA 14: Rótulos de classes

Histograma:

- Selecione as células F15 a F21;
- Tecla **Ctrl** e matenha esta tecla pressionada;
- Use o mouse para selecionar as células H5 a H11;
- Clique no ícone de assistente de gráfico (veja FIGURA 12);
- Escolha as opções:
 - Tipo de gráfico: Colunas;
 - Subtipo de gráfico: O primeiro à esquerda no alto;
 - Clique em “Avançar”.

Veja a FIGURA 15.

ATENÇÃO: Nesta seqüência de comandos, a necessidade de manter a tecla **CTRL** pressionada se justifica pelo interesse de selecionar células não adjacentes.

Microsoft Excel - Indcrim_mg

Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela Ajuda Acrobat

={FREQÜÊNCIA(C5:C68;G5;G11)}

i	Municípios	Índice de criminalidade de	Classes	ni	fi	Ni	Fi	xi	
1	Abadia dos Dourados	17,28	4,48	10,11	12	0,19	12	0,19	7,295
2	Água Comprida	31,72	10,11	15,74	13	0,20	25	0,39	12,925
3	Araguari	31,57	15,74	21,37	15	0,23	40	0,63	18,555
4	Araporá	29,23					48	0,75	24,185
5	Arapuá	8,73					57	0,89	29,815
6	Araxá	25,52					59	0,92	35,445
7	Cachoeira Dourada	43,86					64	1,00	41,075
8	Campina Verde	19,22							
9	Campo Florido	15,59							
10	Campos Altos	20,89							
11	Canápolis	10,39							
12	Capinópolis	18,14							
13	Carmo do Paranaíba	15,80							
14	Carneirinho	14,71							
15	Cascalho Rico	13,24							
16	Centralina	13,29							
17	Comendador Gomes	7,13							
18	Conceição das Alagoas	35,05							
19	Conquista	6,66							
20	Coromandel	25,86							
21	Cruzeiro da Fortaleza	8,17							
22	Douradoquara	11,51							
23	Estrela do Sul	10,48							
24	Fronteira	32,45							
25	Frutal	24,67							
26	Grupiara	39,01							
27	Guimarânia	8,64							
28	Gurinhá	9,57							

Assistente de gráfico - etapa 1 de 4 - tipo de gráfico

Tipos padrão Tipos personalizados

Tipo de gráfico: Colunas

Subtipo de gráfico: Colunas agrupadas. Compara valores por categorias.

Manter pressionado para exibir exemplo

Cancelar < Voltar Avançar > Concluir

FIGURA 15: Assistente gráfico – escolha do tipo de gráfico

Na próxima janela:

- Clique em “Rótulo do eixo das categorias (X):” e selecione as células F15 a F21;
- Clique em “Avançar”;

Com mais algumas modificações, que são deixadas a cargo do leitor, pode-se chegar ao resultado da FIGURA 16.

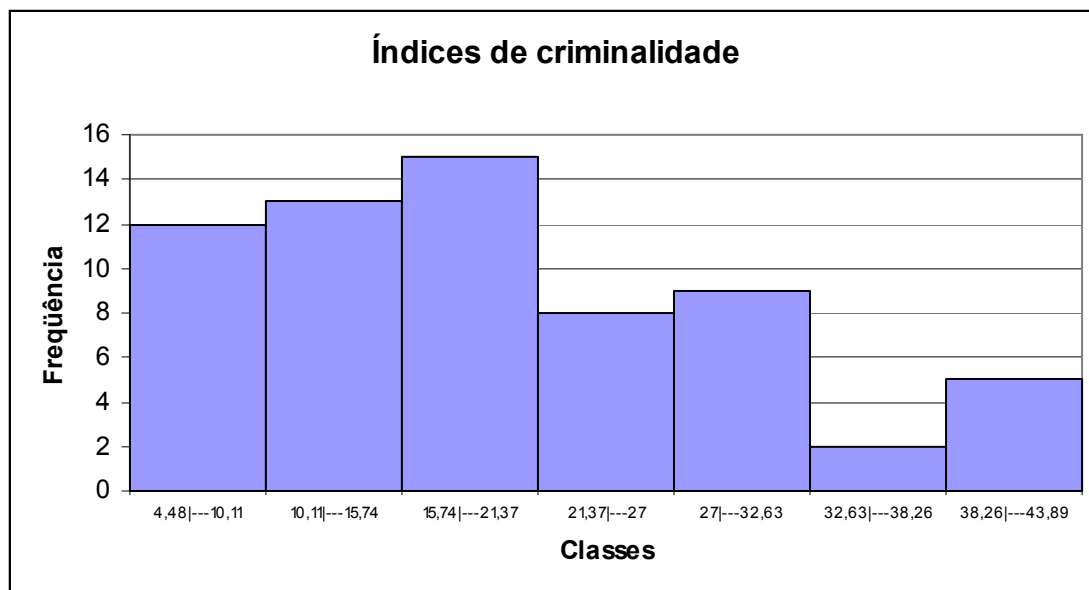


FIGURA 16: Histograma

Para construção do Polígono de Frequência, Ogiva e Gráfico de Linha escolha o gráfico do tipo dispersão ou em linha.

5.10 - Cálculo da média

- Em uma nova planilha, digite os dados na coluna A:
- Na célula A8 digite **Média**
- Em B8, digite **=MÉDIA(A1:A6)** , obtendo o resultado 15,1167

5.11 - Cálculo da média ponderada

- Em uma nova planilha, digite as notas e os pesos das provas, nas colunas A, B e C, conforme FIGURA 15;
- Na célula C7 digite **=(A2*C2+A3*C3+A4*C4)/C5**
- Na célula C8 digite **=(B2*C2+B3*C3+B4*C4)/C5**

Veja os resultados na FIGURA 17.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Candidato A	Candidato B	Pesos												
2	8	6	4												
3	7	8	3												
4	6	7	2												
5			9												
6															
7		Média A	7,222222												
8		Média B	6,888889												
9															

FIGURA 17: Média ponderada

5.12 - Mediana

- Digite os dados nas colunas A e B, conforme FIGURA 16.
- Em A10 digite **Mediana**;
- Vá para a célula B10 e digite **=MED(B2:B8)**

Acompanhe o resultado pela FIGURA 18.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	Cidade	Umidade relativa do ar (%)													
1															
2	A	83													
3	B	62													
4	C	81													
5	D	75													
6	E	94													
7	F	97													
8	G	75													
9															
10	Mediana	75													
11															

FIGURA 18: Mediana

5.13 – Moda

- Digite os dados na coluna A.
- Em A10 digite **Moda**;
- Vá para a célula B10 e digite **=MODO(A1:A8)**

5.14 - Desvio médio absoluto

- Use a função DESV.MÉDIO.

5.15 - Variância e Desvio Padrão

Para cálculo da variância de uma população, use a função VARPA e para cálculo da variância de uma amostra use a função VAR.

Para cálculo do desvio padrão use as funções DESVPADP ou DESVPAD, para população e amostra, respectivamente.

5.16 - Média para dados agrupados

Volte à planilha “dados”, na qual montamos a tabela distribuição de freqüência.

- Mova o gráfico para uma posição que não prejudique a visualização da tabela e de algumas linhas abaixo.
- Vá para a célula H14 e digite **Média**.

Cria-se uma coluna auxiliar, em M, para o cálculo da Média, contendo os resultados dos produtos $n_i \cdot x_i$.

- Digite **ni.xi**, na célula M4;
- Em M5 digite **=H5*L5**;
- Copie a fórmula para as células M6 a M11;
- Coloque o somatório dos resultado em M12. Lembre-se basta clicar no ícone Σ e selecionar as células M5 a M11, se necessário.
- Vá para a célula I14 e digite **=M12/H12**;

Confira os resultados na FIGURA 19.

Indice de criminalidade	Classes	ni	fi	Ni	Fi	xi	ni.xi
17,28	4,48 ---10,11	12	0,19	12	0,19	7,295	87,54
31,72	10,11 ---15,74	13	0,20	25	0,39	12,925	168,025
31,57	15,74 ---21,37	15	0,23	40	0,63	18,555	278,325
29,23	21,37 ---27,00	8	0,13	48	0,75	24,185	193,48
8,73	27,00 ---32,63	9	0,14	57	0,89	29,815	268,335
25,52	32,63 ---38,26	2	0,03	59	0,92	35,445	70,89
43,86	38,26 ---43,89	5	0,08	64	1,00	41,075	205,375
19,22		64	1,00				1271,97
15,59							
20,89	Rótulos	Média	19,8745				
10,39	4,48 ---10,11						
18,14	10,11 ---15,74						
15,80	15,74 ---21,37						
14,71	21,37 ---27						
13,24	27 ---32,63						
13,29	32,63 ---38,26						
7,13	38,26 ---43,89						
35,05							

FIGURA 19: Média para dados agrupados

5.17 - Desvio Padrão e Variância para dados agrupados

Continuando com a planilha “dados”, deve-se criar algumas colunas auxiliares, para o cálculo da variância e, posteriormente, cálculo do desvio padrão.

- Em N4 digite xi^2 , ou seja, será criada uma coluna com os quadrados de cada X_i ;
- Em N5, digite a fórmula **=L5^2**;
- Copie esta fórmula para as células L6 a L11;
- Em O4 digite $ni.xi^2$, indicando a coluna auxiliar com os resultados dos produtos da frequência absoluta pelo quadrado do ponto médio da classe;
- Em O5 digite a fórmula **=H5*N5**;
- Copie esta fórmula para as células O6 a O11;
- Em O12, obtenha o somatório de O5 a O11;
- Vá para a célula H15 e digite **Variância**;
- Em I15, digite a fórmula **=O12/H12-I14^2**
- Digite **Dv. Pad.** em H16 e, em I16 a fórmula **=RAIZ(I15)**;

5.18 – Cálculo do Centro Médio

Para resolução deste exemplo, deve-se abrir o arquivo do excel contendo os dados dos PIB'S dos municípios do estado de Minas Gerais: **mg_municipio_pib_90_98.xls**

Observe que as colunas I e J apresentam as longitudes e latitudes, respectivamente, dos municípios do estado de Minas Gerais. Da coluna O em diante, encontramos os diversos PIB'S disponíveis: Percapita, agropecuário, industrial e total.

Para obter o Centro Médio do PIB_percapta no ano de 1990 siga os passos abaixo:

- Insira um nova coluna após a coluna O. Para inserir uma única coluna, clique em uma célula na coluna imediatamente à direita da posição onde você deseja inserir a nova coluna. Por exemplo, para inserir uma nova coluna à esquerda da Coluna P, clique em uma célula na Coluna P;
- Em P1, digite **percapita.long.1990**;
- Em P2, digite **=I2*O2**;
- Agora, basta copiar esta fórmula para as demais linhas desta coluna. Faça isto clicando duas vezes com o mouse no canto direito inferior da célula P2. Observe que a fórmula já deve ter sido aplicadas a todas as células da coluna.
- Vá para a célula P855 e obtenha a soma dos valores obtidos. Basta clicar no ícone Σ e teclar **ENTER** (Resultado obtido para o somatório: -1407,05);
- Crie uma nova coluna à esquerda da coluna Q;
- Em Q1, digite **percapita.lat.1990**;
- Em Q2, digite **=J2*O2**;
- Copie esta fórmula para as outras células da coluna e obtenha o somatório, em Q855 (Resultado obtido para o somatório: -629,28);
- Como vamos precisar do somatório dos PIB, que representam o peso de cada longitude ou latitude no contexto, obtenha, em O55, o somatório da coluna O (Resultado obtido para o somatório: 31,43);
- Vá para a célula O859 e calcule a média ponderada para a longitude, aplicando a fórmula **=P855/O855**;
- Em P859, aplique a fórmula **=Q855/O855** para calcular a média ponderada para a latitude;

Veja os últimos resultados na FIGURA 20.

	L	M	N	O	P	Q	R
851	VARJÃO DE MINAS (MG)		7 Noroeste de Minas	0,00	0,00	0,00	
852	VERDELÂNDIA (MG)		8 Norte de Minas	0,00	0,00	0,00	
853	VEREDINHA (MG)		9 Jequitinhonha/Mucuri	0,00	0,00	0,00	
854	VERMELHO NOVO (MG)		2 Mata	0,00	0,00	0,00	
855			Somas	31,43	-1407,05	-629,28	
856							
857							
858			Centro Médio percapta	Longitude	Latitude		
859			1990	-44,7666	-20,0212		
860			1991				
861			1992				
862			1993				
863			1994				
864			1995				
865			1996				
866			1997				
867			1998				
868							

FIGURA 20: Centro Médio_percapta para o ano de 1990

- Siga os passos anteriores para calcular os Centros Médios para os anos de 1991 a 1998 e compare os resultados obtidos com os da TABELA 18 do texto.

Lembre-se: Esta tarefa é cansativa, mas poderia ser pior. Imagine-se fazendo tudo isto apenas com o auxílio de lápis, papel e calculadora. Com o tempo e com o aprendizado de outras ferramentas computacionais, você estará apto a utilizar alguns recursos menos laboriosos. Entretanto, para o momento atual, esta é uma boa alternativa.

6 - EXERCÍCIO PROPOSTO

Este exercício pode ser feito após a leitura completa do texto ou à medida que se caminha em sua leitura, depois do contato com os novos conceitos.

De acordo com os dados da tabela 21, construa uma tabela de distribuição de freqüência da taxa de crescimento populacional dos municípios da mesorregião do Triângulo Mineiro e Alto Paranaíba no período de 1996-2000.

Usando a tabela construída, calcule todas as medidas estatísticas estudadas, que sejam adequadas a esta situação.

TABELA 25

Taxa de crescimento populacional dos municípios da mesorregião do Triângulo Mineiro e Alto Paranaíba no período de 1996-2000

i	Município	Taxa de crescimento populacional	i	Município	Taxa de crescimento populacional
1	Abadia dos Dourados	0,08	34	Itapagipe	1,12
2	Água Comprida	0,67	35	Ituiutaba	0,30
3	Araguari	1,57	36	Iturama	2,57
4	Araporã	1,91	37	Lagoa Formosa	0,62
5	Arapuá	-0,76	38	Limeira do Oeste	-1,34
6	Araxá	1,53	39	Matutina	-0,24
7	Cachoeira Dourada	0,31	40	Monte Alegre de Minas	-0,54
8	Campina Verde	0,15	41	Monte Carmelo	2,38
9	Campo Florido	3,45	42	Nova Ponte	2,91
10	Campos Altos	2,21	43	Patos de Minas	2,35
11	Canápolis	0,29	44	Patrocínio	1,63
12	Capinópolis	-1,09	45	Pedrinópolis	-10,42
13	Carmo do Paranaíba	0,83	46	Perdizes	4,08
14	Carneirinho	-0,94	47	Pirajuba	-1,63
15	Cascalho Rico	3,77	48	Planura	0,64
16	Centralina	-0,24	49	Prata	3,01
17	Comendador Gomes	-0,15	50	Pratinha	1,50
18	Conceição das Alagoas	3,19	51	Rio Paranaíba	1,71
19	Conquista	-2,32	52	Romaria	1,98

20 Coromandel	1,23	53 Sacramento	0,62
21 Cruzeiro da Fortaleza	3,92	54 Santa Juliana	-5,90
22 Delta	-0,18	55 Santa Rosa da Serra	1,78
23 Douradoquara	1,57	56 Santa Vitória	-0,93
24 Estrela do Sul	-0,59	57 São Francisco de Sales	0,53
25 Fronteira	1,66	58 São Gotardo	5,32
26 Frutal	0,49	59 Serra do Salitre	-1,69
27 Grupiara	-1,69	60 Tapira	1,05
28 Guimarães	2,69	61 Tiros	-7,50
29 Gurinhatã	-0,94	62 Tupaciguara	1,95
30 Ibiá	1,60	63 Uberaba	1,96
31 Indianópolis	-0,63	64 Uberlândia	3,31
32 Ipiacu	0,27	65 União de Minas	1,77
33 Iraí de Minas	5,21	66 Veríssimo	-5,13

Fonte: Site do Geominas. Acesso em 04 jul 2001

TABELA 26

Distribuição de frequência dos dados do exercício proposto

Classes	x_i	n_i	F_i	N_i	F_i
-10,41 \mapsto -8,16	-9,285	1	0,02	1	0,02
-8,16 \mapsto -5,91	-7,035	1,00	0,02	2,00	0,03
-5,91 \mapsto -3,66	-4,785	2,00	0,03	4,00	0,06
-3,66 \mapsto -1,41	-2,535	4,00	0,06	8,00	0,12
-1,41 \mapsto 0,84	-0,285	26,00	0,38	34,00	0,52
0,84 \mapsto 3,09	1,965	25,00	0,38	59,00	0,89
3,09 \mapsto 5,34	4,215	7,00	0,11	66,00	1
Totais		66	1		

Fonte: Site do Geominas. Acesso em 04 jul 2001

7 – BIBLIOGRAFIA

AMORIM, Oswaldo Bueno Filho. **Reflexões sobre as tendências teórico-metodológicas da Geografia**. Belo Horizonte: 1985, [s.n]. 55p.

_____. **Percepção ambiental – contexto teórico e aplicações ao tema urbano.** Publicação Espacial. Belo Horizonte, n.5, p.9-20. 1987.

ASSUNÇÃO, Renato M. **Estatística espacial aplicada.** Belo Horizonte, 2001. Disponível em: <<http://www.est.ufmg.br/~assuncao/prima.html>>. Acesso em 20 set. 2002.

BARROSO, Magali M. de Araújo. **A Matemática na limpeza urbana: trajeto ótimo do caminhão de lixo.** In: CONGRESSO NACIONAL DE MATEMÁTICA APLICADA E COMPUTACIONAL, 21,1998, Caxambu. Minicurso. São José dos Campos: INPE, 1998, 95p.

BÍBLIA sagrada. A. T. Números. 4 ed. São Paulo: Paulinas, 1989. Vers. 1-3, 1631p.

BOLZAN, Regina de F. F. de A. **O conhecimento tecnológico e o paradigma educacional.** Florianópolis,1998. Dissertação (Mestrado em Engenharia) – Universidade Federal de Santa Catarina. Disponível em: <<http://www.eps.ufsc.br/disserta98/regina>>Acesso em 02 out. 2000.

BORBA, Marcelo de Carvalho, PENTEADO, Miriam Godoy. **Informática e educação matemática.** Belo Horizonte: Autêntica, 2001. 98p.

BORBA, Marcelo de Carvalho. Informática trará mudanças na educação brasileira? **Zetetiké**, Campinas, v.4,n.6,p.123-134, jul./dez. 1996.

BUESCU, Mircea. **Métodos quantitativos em História.** Rio de Janeiro: Livros Técnicos e Científicos, 1983. 127p.

CAMARA, Gilberto et al. **Análise espacial de dados geográficos.** São José dos campos, 2002. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/analise/>>. Acesso em 15 jul 2002.

CARLOS, Ana Fani Alessandri (Org.). **A Geografia na sala de aula.** São Paulo: Contexto, 1999.

CHORLEY, Richard J., HAGGET, Peter. **Modelos integrados em Geografia.** Trad. Arnaldo Viriato de Medeiros. Rio de Janeiro: Livros Técnicos e Científicos, 1974. 221p.

CLARK, David. **Introdução à Geografia Urbana.** Trad. Lúcia H. O. Gerardi & Silvana Maria Pintaudi.São Paulo: DIFEL, 1985. 286p.

COLE, John P. **Geografia Quantitativa.** Rio de Janeiro: Fundação IBGE, 1972. 120p.

COLE, John P., KING, Cuchlaine A.M. **Quantitative Geography; Techniques and Theories in Geography.** 3.ed. London: John Wiley & Sons, 1970. 692p.

COSTA, Sérgio Francisco. **Introdução Ilustrada à Estatística.** São Paulo: Harbra,

1992. 303p.

D'AMBROSIO, Ubiratan. **Etnomatemática: elo entre as tradições e a modernidade**. Belo Horizonte, Autêntica, 2001. 110p.

_____. **Etnomatemática; arte ou técnica de explicar e conhecer**. São Paulo, Ática, 1998. 88p.

_____. **A influência da tecnologia no fazer matemático ao longo da história**. São Paulo, 1999. Disponível em: <<http://sites.uol.com.br/vello/ubi.htm>>. Acesso em 16 abr.2002.

_____. **Etnomatemática: uma proposta pedagógica para a civilização em mudança**. São Paulo, 2000. Disponível em: <<http://sites.uol.com.br/vello/ubi.htm>>. Acesso em 16 abr. 2002.

_____. **Transdisciplinaridade e a proposta de uma nova universidade**. São Paulo, 1999. Disponível em: <<http://sites.uol.com.br/vello/ubi.htm>>. Acesso em 16 abr. 2002.

_____. **Informática, Ciências e Matemática**. São Paulo, 1999. Disponível em: <http://www.proinfo.gov.br/biblioteca/frame_de_baixo.htm> Acesso em 23 jul. 2000.

DIEESE-Departamento Intersindical de Estatística e Estudos Sócio-Econômicos. Disponível em: <<http://www.dieese.org.br/>>. Acesso em: 04 jan. 2002.

DINIZ, Flávio Roberto C. **Cálculo aplicado à nutrição; decifrando números na sua alimentação**. Belo Horizonte: AMEI, 2001. 91p.

FERREIRA, Conceição Coelho, SIMÕES, Natércia Neves. **Tratamento Estatístico e gráfico em Geografia**. Lisboa: Gradiva, 1987. 151p.

FERREIRA, Eduardo Sebastiani. **Etnomatemática**. In: Congresso Nacional de Matemática Aplicada e Computacional, 24, 2001, Belo Horizonte: Editora Gráfica Novo Milênio, 2001. 17p.

FIORENTINI, Dario. Alguns modos de ver e conceber o ensino da Matemática no Brasil. **Zetetiké**. Campinas, ano 3, v.4, pp.1-37, 1995.

GATRELL, Anthony C. et al. **Spatial point pattern analysis and its application in geographical epidemiology**. 1996. Disponível em: <<http://www.est.ufmg.br/%7Eassuncao/cursos/espacial/aplicado/programa.html>>. Acesso em 20 set. 2002.

GeoMINAS Geoprocessamento em Minas Gerais : Dados de Minas. Disponível em : < <http://www.geominas.mg.gov.br> >. Acesso em : 05 jul. 2001.

GERARDI, Lúcia Helena de Oliveira, SILVA, Bárbara-Christine Nentwig Silva. **Quantificação em Geografia**. São Paulo, Difel, 1981. 161p.

GRIFFITH, Daniel A., AMRHEIN, Carl G. **Multivariate Statistical Analysis for Geographers**. New Jersey: Prentice Hall. 1997, 345p.

NYSTUEN, John.D. **Identification of some fundamental spatial concepts**.

INMET- Instituto Nacional de Meteorologia. Disponível em : <<http://www.inmet.gov.br/>>. Acesso em : 19 out. 2001.

JOHNSTON, R.J. **Geografia e geógrafos**. Trad. Oswaldo Bueno Amorim Filho. São Paulo: 1986, 347p.

MORAES, Maria Candida. **Novas tendências para o uso das tecnologias da informação na educação**. Brasília, fev. 1998. Disponível em : <<http://www.sociologia.de/soc/index1.htm>> Acesso em 13 fev. 2002.

MOYSÉS, Lucia. **Aplicações de Vygotsky à Educação Matemática**. Campinas: Papirus, 1997. 176p.

O'BRIEN, Larry. **Introducing Quantitative Geography; Measurement, methods and generalised linear models**. London: Routledge, 1992. 356p.

OLIVEIRA, Marta Kohl. **Vygotsky: aprendizado e desenvolvimento um processo sócio-histórico**. São Paulo: Scipione, 1993. 111p.

PENTEADO, Miriam Godoy , BORBA, Marcelo de Carvalho, GRACIAS, Telma de Souza. **Informática como veículo para mudança**. Zetetiké, Campinas, v.6,n.10, p.77-86, Jul/Dez. 1998.

SILVA, Fernando César. **Informática, Educação e Trabalho, ideologias à parte**. Caruaru, 1998. Disponível em <http://www.proinfo.gov.br/biblioteca/frame_de_baixo.htm>. Acesso em 23 jul. 2000.

SIMONSEN, Mário Henrique. Prefácio. In: BUESCU, Mircea. **Métodos quantitativos em História**. Rio de Janeiro: Livros Técnicos e Científicos, 1983. 127p.

SOARES, José F., FARIAS, Alfredo A., CESAR, Cibele C. **Introdução à Estatística**. Rio de Janeiro: Guanabara Koogan, 1991. 378p.

STEVENSON, William J. **Estatística aplicada à administração**. São Paulo: Harbra, 1986,. 495p.

TAJRA, Sanmya Feitosa. **Informática na educação: Novas ferramentas pedagógicas para o professor da atualidade**. São Paulo: Érica, 2001. 182p.

VALENTE, José Armando. **Por quê o computador na educação?** Campinas, [199-] Disponível em: <http://www.proinfo.gov.br/biblioteca/frame_de_baixo.htm>. Acesso em 04 set. 2000.

VIEIRA, Fábila Magali Santos, MATOS, Maria de Lourdes. **A utilização das novas tecnologias na educação numa perspectiva construtivista.** Brasília, [199-]. Disponível em <http://www.proinfo.gov.br/biblioteca/frame_de_baixo.htm>. Acesso em 23 jul. 2000.

YEATES, Maurice. **An introduction to Quantitative Analysis in Human Geography.** McGraw-Hill, Inc. U.S.A., 1974.