1. ESTATISTICA DESCRITIVA

1.1 INTRODUÇÃO

Quem pretende tomar decisão ou efectuar estudos de natureza científica, começa normalmente por recolher informações que lhe parecem relevantes. Estas informações destinam – se a servir de base no estudo pretendido e são acumuladas de forma organizada, por isso, se designam de dados. Na maioria dos casos, estes dados são de natureza quantitativa, daí se chamarem de dados numéricos.

A importância da estatística bem como a sua utilização é cada vez mais acentuada em qualquer actividade profissional, isto é, em todas áreas da actividade sócio económica, como um método que permite indicar de forma sumária as características apresentadas por um conjunto de números.

Estatística é uma ciência cujo objectivo é a observação de fenómenos de mesma natureza, recolha, apresentação, análise e interpretação dos dados numéricos com o propósito de descrever o conjunto dos dados recolhidos e tomar decisões ou generalização das características tiradas na amostra, para todo o conjunto das unidades a estudar, população.

A estatística geralmente é dividida em duas partes: estatística descritiva e estatística indutiva ou inferencial.

Estatística descritiva é o ramo ou parte da estatística cujo objectivo é a observação de fenómenos de mesma natureza, recolha, organização, classificação, análise e interpretação dos dados sem deixar de calcular algumas medidas (estatísticas), que permitem resumidamente descrever o fenómeno estudado.

Estatística indutiva ou inferencial é o ramo ou parte da estatística que trata das condições de generalização das conclusões tiradas a partir da observação de uma parte das unidades estatísticas (amostra), para o todo o conjunto das unidades a estudar (população).

Técnicas elementares do método estatístico

Constituem técnicas elementares do método estatístico, os procedimentos utilizados na estatística descritiva desde a recolha, descrição dos dados, passando pela organização de tabelas, construção de gráficos, cálculo de medidas estatísticas que possam ajudar a análise e interpretação dos dados sem recorrer muitas vezes a juízos probabilísticos nem a inferência estatística.

Entre muitas fases do método estatístico destacam - se seis principais.

1. **Definição do problema** – definir correctamente o problema e objectivos.
2. **Planeamento** – definição dos procedimentos para a obtenção dos dados.
3. **Recolha dos dados** – é a fase operacional de registro dos dados com um objectivo definido. *Dados primários:* quando são publicados pela própria pessoa ou organização que os tenha

recolhido. Exemplo: as tabelas do censo demográfico do Instituto Nacional de Estatística – INE.

Dados secundários: quando são publicados por outra organização. Exemplo: quando determinado jornal publica estatísticas referentes ao censo demográfico extraídas do INE.

Observação: É mais seguro trabalhar com fontes primárias porque o uso de fontes secundárias traz o grande risco de erros de transcrição.

Recolha directa: quando é obtida directamente da fonte. Exemplo: Empresa que realiza uma investigação para saber a preferência dos consumidores de um dos seus produtos vendidos no mercado. A recolha directa pode ser: *contínua* (registos de nascimentos, óbitos, casamentos, etc.), *periódica* (recenseamento demográfico, censo industrial, agro-pecuário, etc.) e *ocasional* (registo de calamidades, e outras anomalias da natureza).

Recolha indirecta: é feita por deduções a partir dos elementos conseguidos pela recolha directa, por analogia, por avaliação, indícios ou proporcionalidade, ou ainda é feita em documentos já existentes.

4. **Apuramento dos dados** – é o resumo dos dados através de sua contagem e agrupamento.
5. **Apresentação dos dados** – exposição dos resultados obtidos por tabelas ou gráficos.
6. **Análise e interpretação dos dados** – é a última fase do trabalho estatístico, mais importante e delicada. Nesta fase essencialmente quase em paralelo com o processamento ou cálculo de algumas estatísticas que ajudam a interpretar ou descrever o fenómeno a investigar.

O conjunto das diversas fases para se preparar a informação estatística desde a recolha, produção e divulgação e sua posterior utilização, constitui o circuito da informação estatística. Os principais intervenientes deste circuito são: os utilizadores, os fornecedores e os produtores dos dados estatísticos.

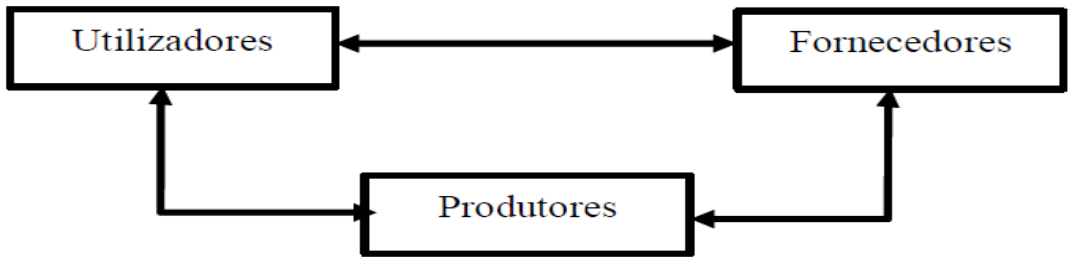


Figura 1.1. Circuito da informação estatística

Utilizadores – São governantes, homens de negócio, economistas, investigadores e técnicos em geral, estes têm a tarefa de seleccionar e analisar a informação de que necessitam para os seus trabalhos. Cabe aos utilizadores a crítica e colocação de opiniões para os produtores e fornecedores de forma a melhorar a informação estatística que recebem.

Fornecedores – prestam aos produtores informações de base, facultando dados que lhes são solicitados pelos utilizadores, sem se esquecer da cooperação e apoio que eles têm com os produtores.

Produtores – são as entidades especializadas, que têm por objectivo a obtenção e preparação da informação necessária segundo o conhecimento da realidade política, económica e social nos vários domínios de actuação de um determinado pais.

Para que a informação estatística, possa ser produzida, fornecida e utilizada de forma coerente pelos intervenientes acima apresentados e seja um instrumento ao serviço do desenvolvimento de uma realidade, ela deve possuir as seguintes características: utilidade, qualidade e actualidade.

Utilidade – a informação estatística serve de um meio para o desenvolvimento das sociedades, praticamente é um instrumento de tomada das decisões. É por isso, que ela deve ser compreendida pelos utilizadores, o que significa que ela deve ser acompanhada de todos aspectos metodológicos mais relevantes: gráficos, tabelas, notas explicativas, etc., para permitir a percepção do seu verdadeiro conteúdo e significado.

Qualidade – os dados estatísticos devem traduzir uma realidade de forma simples e clara. O controlo de qualidade dos dados é uma tarefa indispensável para a sua fiabilidade, por outro lado a informação estatística deve ser completa, pois, não tem sentido preparar e divulgar resultados baseados em apuramentos parciais como resultados finais.

Actualidade – a informação estatística deve estar disponível de forma atempada a pessoas que precisam utiliza –la no momento necessário. Isto implica um esforço na recolha de dados, processamento e divulgação dos resultados, permitindo que as bases de dados estejam continuamente actualizadas.

Definições básicas de estatística

Um **fenómeno estatístico** é qualquer evento que se pode analisar aplicando técnicas estatísticas. Os fenómenos estatísticos podem ser colectivos ou de massa por exemplo a evolução das exportações de uma empresa ou pais; individuais por exemplo a realização de um casamento; típicos ou regulares como festas de carnaval; atípicos ou irregulares como calamidades, etc.

Observação estatística é um processo sistemático cientificamente argumentado de dados em massa com uma característica comum objecto de estudo.

População ou Universo é o conjunto de todos os elementos que apresentam pelo menos, uma característica comum objecto de estudo. A população a ser estudada pode ser finita ou infinita, consoante seja finito ou não o número dos elementos considerados na observação.

Amostra é uma parte das unidades estatísticas seleccionadas da população para o estudo, muitas vezes quando não é possível ou é difícil estudar toda a população. A partir das conclusões tiradas da amostra, faz – se um juízo ou inferência destas para as características da população. As

características obtidas das amostras são chamadas estatísticas (medidas descritivas), enquanto as medidas populacionais são denominadas de parâmetros populacionais.

Unidade estatística é cada elemento que constitui a população observada, é precisamente sobre este elemento ou unidade que recai a observação estatística.

Um estatístico pode ser alguma pessoa que desenvolve funções oficiais utilizando dados numéricos, pode ser um analista especializado em metodologias estatísticas para a recolha, organização, análise, representação em forma de tabelas e gráficos e interpretação de dados numéricos; finalmente o termo, pode referir a um especialista que utiliza a matemática superior para desenvolver novos métodos de análise de dados quantitativos para tomar decisões. Os estatísticos são sempre necessários em qualquer nível em que desenvolvem suas funções.

Caracteres e modalidades das variáveis estatísticas

A estatística estuda fenómenos, a que se chamam de caracteres ou atributos estatísticos. Para descrever quantitativamente uma população é necessário que as unidades estatísticas sejam classificadas em subconjuntos adequados de acordo com os caracteres ou atributos mais relevantes. Por sua vez, estes atributos podem ser qualitativos ou quantitativos.

Chama – se **modalidade** de um atributo as diferentes variações ou valores que este atributo pode assumir.
Exemplos:

- O atributo **sexo** pode ter duas modalidades que são: masculino ou feminino;
- O atributo **estado civil** tem mais de 2 modalidades: solteiro, casado, divorciado, etc.;
- O atributo **idade** tem uma modalidade que varia com intensidade, pois uma pessoa pode ter 1 ano, 2 anos, 3 anos, ... ,35 anos, 70 anos, etc.

Um caractere é qualitativo, quando as modalidades não são numéricas ou não podem ser medidas mas podem ser apenas constatadas e será quantitativo quando ele for possível de medir. De um modo geral os atributos observados quando são qualitativos revestem – se em várias modalidades e quando são quantitativos apresentam uma modalidade com diferentes intensidades ou valores.

Tabela 1.1. Os caracteres associados aos seus universos.

Universo	Atributo / caracteres
População de pessoas	Sexo, idade, local de nascimento, estado civil, grau de instrução;
Economia de um pais	Produto interno bruto, consumo público, situação monetária;
Classificação de um aluno	Muito bom, bom, suficiente, mau, muito mau, etc.;
Actividade de uma empresa	Volume de vendas, despesas, lucros, investimentos, etc.

Variável estatística é o conjunto de resultados possíveis ou variações de um determinado atributo ou fenómeno.

Uma variável estatística é **qualitativa** quando se classifica em diversas modalidades ou categorias e é **quantitativa** quando tem uma modalidade com diferentes intensidades. Por outro lado, uma variável estatística pode ser discreta ou contínua.

Uma variável é discreta ou descontínua – quando seus valores são expressos através de números inteiros não negativos e resulta normalmente a partir de contagens. Exemplo: número de alunos presentes às aulas de introdução à estatística no 1º semestre de 2006.

Uma variável é contínua – quando resulta normalmente de uma medição ou quando a variável pode tomar qualquer valor dentro do conjunto dos números reais. Ou ainda dado um universo no intervalo $[a;b]$ com $a<b$, existe um valor x tal que: $x>a$ e $x<b$, isto é $a<x<b$. Por exemplo quando medimos a temperatura de corpo com um termómetro, o mercúrio ao dilatar-se passará por todas as temperaturas intermediárias até chegar a temperatura actual do corpo.

$$\text{variável}\begin{cases}\text{Qualitativa (categorias)}\begin{cases}\text{Nominais}\\ \text{Ordinais}\end{cases}\\ \text{Quantitativas (números)}\begin{cases}\text{Discretas}\\ \text{Continuas}\end{cases}\end{cases}$$

Escalas usadas para medir os atributos

Conforme a natureza dos atributos, existem quatro escalas principais usadas para medi –los:

1) **Escalas nominais** – são aquelas que separam os atributos em categorias diferentes não forçando uma ordenação em termos de hierarquia. Na utilização destas escalas, é preciso que se obedeçam três condições:
   a) A divisão deve ser coerente de acordo com um único critério;
   b) A divisão deve ser completa;
   c) As categorias que participam na divisão devem ser mutuamente exclusivas.

2) **Escalas Ordinais** – baseiam – se numa classificação hierárquica. Através desta escala os atributos são colocados em determinada ordem conforme um critério escolhido.

3) **Escalas de intervalo** – as escalas nominais separam os objectos em categorias distintas, as ordinais dispõem tais categorias numa certa ordem conforme um critério escolhido, as escalas de intervalo, para além de distinguirem categorias diferentes e ordenação, colocam as categorias a distâncias iguais. Uma propriedade importante nesta escala é a possibilidade de ser submetida as quatro operações aritméticas.

4) **Escalas de razão** – são um caso especial das escalas ordinais, as quais são também nominais hierárquicas. Assim, a escala de razão é também uma escala de intervalo dotada de zero absoluto.

Gráficos estatísticos e sua aplicação

Um **gráfico** é uma representação visual dos dados estatísticos sem de algum modo substituir as tabelas estatísticas. Normalmente as características dos gráficos são: uso de escalas, sistema de coordenadas, simplicidade, clareza e veracidade com a ocorrência do fenómeno.

Gráficos de informação*:* São gráficos destinados principalmente ao público em geral, objectivando proporcionar uma visualização rápida e clara. São gráficos tipicamente expositivos, dispensando comentários explicativos adicionais. As legendas podem ser omissas, desde que as informações desejadas estejam presentes.

Gráficos de análise*:* São gráficos que prestam-se melhor ao trabalho estatístico, fornecendo elementos úteis à fase de análise dos dados, sem deixar de ser também informativos. Os gráficos de análise frequentemente vêm acompanhados de uma tabela estatística. Inclui-se, muitas vezes um texto explicativo, chamando atenção ao leitor para ler e analisar os aspectos relevantes ilustrados pelo gráfico.

Os gráficos podem se classificar em Diagramas, Estereogramas, Pictogramas, Cartogramas, etc.

1. **Diagramas:** São gráficos geométricos dispostos em duas dimensões. São os mais usados na representação de séries estatísticas. Eles podem ser de barras horizontais, barras verticais, barras compostas, colunas superpostas, linhas, sectores circulares, etc.

2. **Estereogramas:** São gráficos geométricos dispostos em três dimensões, pois, representam o volume. São usados nas representações gráficas das tabelas de dupla entrada. Em alguns casos este tipo de gráfico fica difícil de ser interpretado dada a pequena precisão que oferece.

3. **Pictogramas:** São construídos a partir de figuras representativas da intensidade do fenómeno. Este tipo de gráfico tem a vantagem de despertar a atenção do público leigo, pois, sua forma é atraente e sugestiva. Os símbolos devem ser auto-explicativos. A desvantagem dos pictogramas é que apenas mostram uma visão geral do fenómeno e não os detalhes minuciosos.

4. **Cartogramas:** São ilustrações relativas a cartas geográficas (mapas). O objectivo desse tipo de gráficos é o de apresentar os dados estatísticos directamente relacionados com áreas geográficas ou políticas.

1.2 DISTRIBUIÇÃO DE FREQUÊNCIAS

Para que uma informação recolhida seja divulgada é necessário que esta seja organizada de modo a ser percebida pelos leitores ou outros investigadores.

De uma geral uma **distribuição de frequência** é um tipo de tabela que condensa uma colecção de dados conforme as repetições de seus valores.

Dados brutos - são os dados originais que ainda não estão numericamente organizados. É difícil formar uma ideia exacta do comportamento do fenómeno, a partir de dados não ordenados.

Exemplo 1.1 Os dados a seguir correspondem ao consumo de energia de uma determinada família num período de um ano (em Kw/h).
X = {145, 141, 142, 141, 142 143, 144, 141, 146, 150, 146, 150}.

Rol estatístico ou ROL – é uma lista em que os valores ou dados numéricos são ordenados de forma crescente ou decrescente.

X _Rol = {141, 141, 141, 142, 142, 143, 144, 145, 146, 146, 150, 150}

Como se pode observar da nova série de dados de consumo de energia, já pode-se distinguir qual é o valor mínimo bem como o consumo máximo.

Amplitude Total ou Range (At) – é a diferença entre o maior e o menor valor da série estatística ou a diferença positiva entre os extremos do Rol estatístico.

$$A_t = x_{\max} - x_{\min} \tag{1.1}$$

Assim, para os valores do exemplo 1.1, a amplitude total é: At = 150 – 141 = 9 kw/h

Quando os dados brutos se resumem, frequentemente costuma – se distribuir em classes ou categorias e determina – se o número de casos pertencentes a uma das classes ou o número de repetições de um determinado valor denominado frequência de classe ou do valor. Uma apresentação em forma de tabela de dados categorizados ou com intervalo de classe e as respectivas frequências é conhecida por distribuição de frequências.

Há dois tipos de distribuição de frequências: de dados não classificados ou não agrupados (tabela 1.2) e de dados ou valores classificados em classe (tabela 1.3).

As distribuições de frequências não agrupadas são utilizadas quando temos poucas observações ou dados e o número de valores ou modalidade apresenta repetições. Se o número de dados observados for elevado é conveniente agrupar os dados em classes.

As frequências absolutas, relativas e acumuladas

a) **Frequência absoluta (fi)** – é o número de repetições de um valor individual ou o número de valores pertencentes a uma classe da variável em estudo.

b) **Frequência relativa (fr)** – é a proporção de observações de um valor individual ou de valores pertencentes a uma classe em relação ao número total de observações.

$$fr = \frac{fi}{n} \text{ ou } fr = \frac{fi}{n} x100\% \quad \text{onde n} = \sum \text{fi} \tag{1.2}$$

Tabela 1.2. Distribuição de frequência de 40 trabalhadores segundo o seu salário.

i	Salário, X	frequências
1	2.500,00	2
2	3.000,00	4
3	3.400,00	9
4	4.000,00	11
5	4.500,00	6
6	5.000,00	5
7	5.500,00	3
Total	------	**40**

Tabela 1.3. Distribuição de frequência das notas do 1º teste de estatística de 50 estudantes

i	Classes de X			frequências
1	0	----	4	2
2	4	----	8	7
3	8	----	12	10
4	12	----	14	14
5	14	----	16	12
6	16	----	18	4
7	18	----	20	1
Total		----		**50**

c) **Frequência absoluta acumulada (Fi)** – é a soma de todas as frequências absolutas desde a primeira classe até a classe de ordem i ou desde a última classe até a classe de ordem i.

d) **Frequência relativa acumulada (Fr)** – pode ser calculada a partir da definição da frequência acumulada ou da definição de frequência relativa.

Exemplo 1.4. Calculo das frequências relativas, relativas percentuais, acumuladas, acumuladas relativas e acumuladas relativas percentuais.

i	xi	fi	fr	fr%	Fi	Fr	Fr%
1	2	3	0.12	12	3	0.12	12
2	6	4	0.16	16	7	0.28	28
3	10	10	0.40	40	17	0.68	68
4	14	6	0.24	24	23	0.92	92
5	18	2	0.08	08	25	1.00	100
Soma	---	25	1.00*	100	---	---	---

*Observação: a soma das frequências relativas deve ser sempre igual a unidade ou 100%

Exemplo 1.3. Num grupo de 40 alunos que foram reprovados em alguma disciplina do semestre anterior, perguntados sobre que disciplinas tinham sido reprovados, as suas respostas foram as seguintes.

Matemática	Matemática	Português	Álgebra	Estatística	Estatística	Matemática
Matemática	Matemática	Matemática	Estatística	Português	Estatística	Álgebra
Álgebra	Estatística	Matemática	Álgebra	Álgebra	Português	Português
Estatística	Matemática	Matemática	Matemática	Estatística	Português	Estatística
Matemática	Álgebra	Álgebra	Estatística	Português	Português	Matemática
Álgebra	Álgebra	Matemática	Álgebra	Estatística		

a) Elabore a tabela de distribuição de frequências.
b) Calcule as frequências relativas e relativas percentuais.

Resolução

Como temos uma variável qualitativa com diferentes categorias, os nomes das disciplinas constituem os nomes dos nossos valores individuais e as frequências absolutas serão a quantidade de repetição.

Tabela 1.5. Distribuição de frequências

Disciplinas	**fi**	**fr**	**fr%**
Matemática	13	0.325	32.5
Português	07	0.175	17.5
Álgebra	10	0.250	25.0
Estatística	10	0.250	25.0
Total	**40**	**1.000**	**100.0**

Distribuição de frequências para dados agrupados

Consideremos os seguintes dados que representam as alturas de 50 indivíduos medidos até aos centímetros.

162	188	173	168	170	183	186	177	187	174
164	174	159	177	173	163	180	196	171	184
170	190	181	166	181	182	176	169	172	162
175	192	178	177	200	191	188	168	165	193
175	160	180	187	176	170	156	174	179	167

Da forma como estão apresentadas as alturas e atendendo que o número de observações é muito elevado, vamos fazer uma tabela de distribuição de frequências com classes. A escolha do número de classes é variável consoante a conveniência e o número de casos observados.

Antes de apresentar a tabela de distribuição de frequências consideremos os elementos principais associados a uma tabela de distribuição de frequência com intervalos de classes.

Classe: são os intervalos de variação da variável e é simbolizada por ***i*** e o número total de classes simbolizado por ***k***.

Limite de classe: são os extremos de cada classe. O menor número é o limite inferior de classe e o maior número é limite superior de classe.

Amplitude do intervalo de classe: é obtida fazendo a diferença entre o limite superior e o inferior da classe. Para uma classe definida o intervalo é calculado pela fórmula:

$$i = x_{\max(i)} - x_{\min(i)} \tag{1.3}$$

Ponto médio de classe: é o ponto que divide o intervalo de classe em duas partes iguais. Este ponto representa a classe para efeitos de cálculo, denotando – se por: $\overline{x_i}$ ou simplesmente x .

$$\overline{x_i} = \frac{x_{\max(i)} + x_{\min(i)}}{2} \tag{1.4}$$

Para a elaboração de uma tabela de distribuição de frequências com dados agrupados e necessário no mínimo seguir algumas regras. Os procedimentos mais comuns têm os seguintes passos.

Passo 1. Organizar os dados brutos num rol, ver exemplo 1.1;

Passo 2. Calcular a amplitude total A_t , expressão (1.1);

Passo 3. Determinar ou escolher o número de classes *K*. Normalmente não existe um método exacto ou número fixo de classes a escolher. Existem apenas vários métodos ou procedimentos que são usados segundo a conveniência.
Método 1. Escolher arbitrariamente *K* entre 5 a 20 segundo a opção do investigador e a extensão dos dados.
Método 2. Calcular pela fórmula de Sturges. $k = 1 + 3.3 * \log N$, onde N é o número das observações. Deve-se salientar que o número de classes deve ser arredondado ao número inteiro mais próximo.
Método 3. Usar a fórmula $K = \sqrt{N}$

Passo 4. Determinar o intervalo de classe $i = \frac{A_t}{k}$. O valor de i não é necessariamente inteiro, ele pode ser arredondado assegurando que todas observações fiquem até ao limite superior da última classe.

Passo 5. Determinar os limites inferiores e superiores de classe.
Classe 1. $X\ \inf(1) = X\ \min; X\ \sup(1) = X\ \inf(1) + i$
Classe 2. $X\ \inf(2) = X\ \sup(1); X\ \sup(2) = X\ \inf(2) + i$
….
Classe k. $X\ \inf(k) = X\ \sup(k-1); X\ \sup(k) = X\ \inf(k-1) + i = X\ \max$

Exemplo 1.4. A partir dos dados correspondentes as alturas de 50 indivíduos construir a tabela de distribuição de frequências absolutas e determinar os pontos médios de cada classe.

Resolução:

Usando os passos anteriores:

1. Do Rol obtido: $x_{\max} = 200$, $x_{\min} = 156$
2. Amplitude total $At = X_{\max} - X_{\min} = 200 - 156 = 44$
3. Pelo método 2. $k = 1 + 3.3 * \log N = 1 + 3.3 * \log 50 = 6.6 \approx 7$ classes
4. Amplitude do intervalo de classe: $i = \frac{A_t}{k} = \frac{44}{7} = 6.28571$ ou (i = 6.3)

Tabela 1.6. Distribuição de frequências das alturas de 50 indivíduos

I	Classes (x, cm)	fi	xi
1	[156.0 -- 162.3 [	5	159.15
2	[162.3 -- 168.6 [	7	165.45
3	[168.6 -- 174.9 [	9	171.75
4	[174.9 -- 181.2 [	15	178.05
5	[181.2 -- 187.5 [	6	184.35
6	[187.5 -- 193.8 [	6	190.65
7	[193.8 -- 200.1 [	2	196.95
Total	--	**50**	--

Exemplo 1.5. Os dados que se encontram abaixo, correspondem a valores em meticais de 50 trabalhadores da Ministério da Saúde que pagaram o IRPS durante um período de um ano. Construir a tabela de distribuição de frequências, calcular os pontos médios de cada classe, as frequências relativas, acumuladas e relativas acumuladas.

964	1040	1358	968	1165	786	1416	1168	1080	1013
1007	1189	870	1385	1236	1325	1261	1103	1036	923
997	1130	1317	1312	1216	1051	1477	1151	1615	563
1037	1207	951	1102	1183	1300	1088	1090	1314	1233
1588	1347	1251	1592	1713	1120	948	1567	1104	1292

Resolução:

1. Do Rol obtido, $X_{max} = 1713$, $X_{min} = 563$
2. Amplitude total $At = 1713 - 563 = 1150$
3. Pelo segundo método. $k = 1 + 3.3 * \log N = 1 + 3.3 * \log 50 = 6.6 \approx 7$
4. Amplitude do intervalo de classe: $i = \frac{A_t}{k} = \frac{1150}{7} = 164.3$

Tabela 1.7. Distribuição de frequências dos valores de IRPS de 50 trabalhadores do MISAU

i	Classes (X, meticais)			fi	xi	fr	Fi	Fr
1	563.00	---	727.30	1	645.15	0.02	1	0.02
2	727.30	---	891.60	2	809.45	0.04	3	0.06
3	891.60	---	1055.90	12	973.75	0.24	15	0.30
4	1055.90	---	1220.20	15	1138.05	0.30	30	0.60
5	1220.20	---	1384.50	12	1302.35	0.24	42	0.84
6	1384.50	---	1548.80	3	1466.65	0.06	45	0.90
7	1548.80	---	1713.10	5	1630.95	0.10	50	1.00
soma		**---**		**50**	**---**	**1.00**	**---**	**---**

Representação gráfica de uma distribuição de frequências

Para representar uma informação resumida numa tabela de distribuição de frequências de dados não agrupados (variável discreta), basta apresentar um diagrama de barras, o polígono de frequências para as frequências absolutas ou relativas e uma ogiva para as frequências acumuladas.

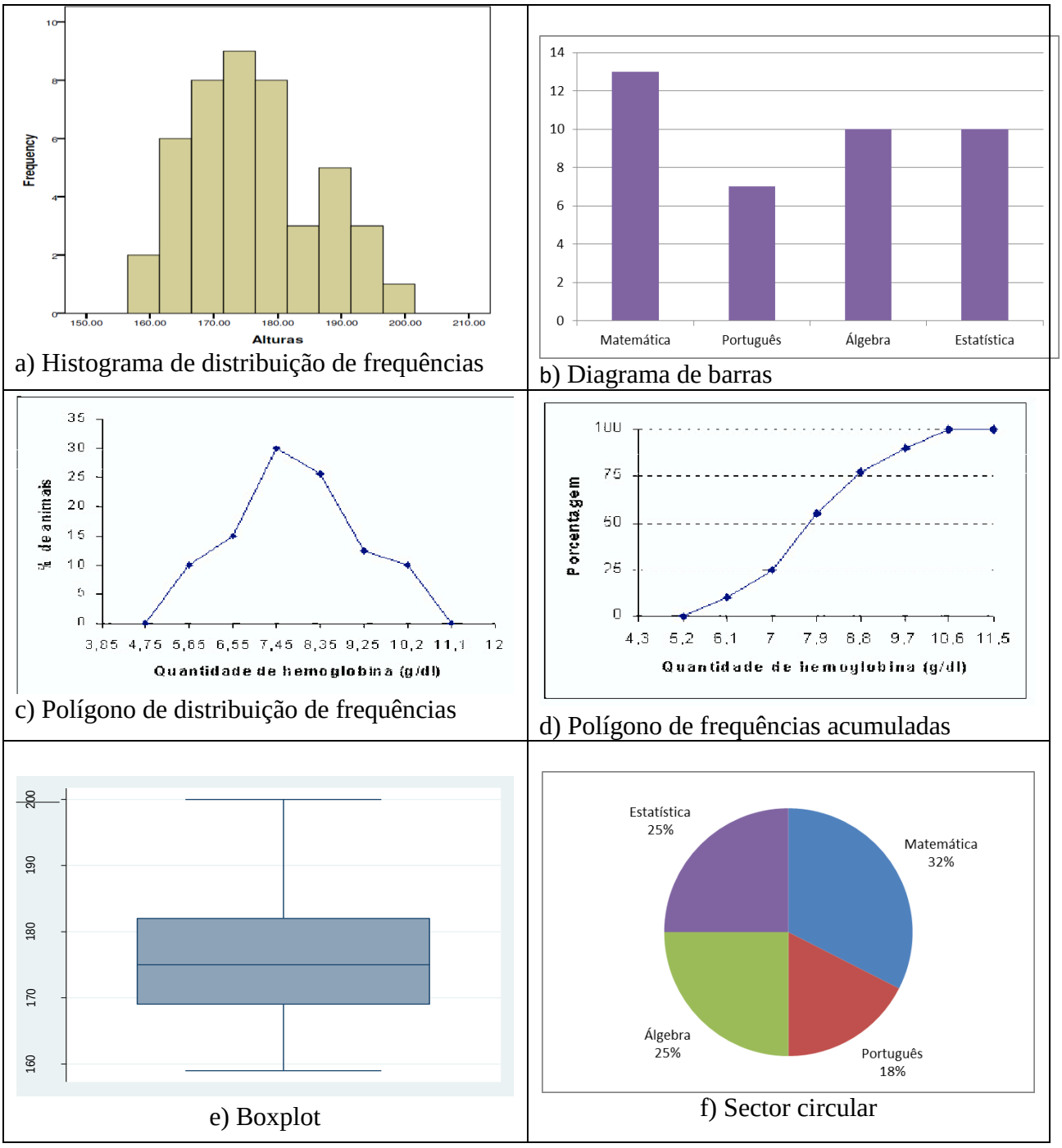
Para dados agrupados em classe, para além do histograma e da curva polida de frequências e as diferentes formas de representação de distribuições de frequências para dados não agrupados também podem ser feitas, neste caso basta considerar os pontos médios de classe como valores individuais.

Histograma: é um diagrama de áreas, formado por um conjunto de rectângulos justapostos, de tal modo que seus pontos médios coincidam com os pontos médios dos intervalos de classe. A área de um histograma é proporcional à soma das frequências simples ou absolutas.

Polígono de frequência: é um gráfico em linha, sendo as frequências marcadas sobre perpendiculares ao eixo horizontal, levantadas pelos pontos médios dos intervalos de classe. De notar que a altura de cada barra ou linha é proporcional a frequência da classe.

Enquanto o polígono de frequência nos dá a **imagem real** do fenómeno estudado, a **curva de frequência** nos dá a **imagem da tendência da distribuição**. O polimento de um polígono de frequência nos mostra o que seria tal polígono com um número maior de dados em amostras mais amplas.

Quando se estudam duas variáveis de certa forma relacionadas, frequentemente apresentam-se **diagramas de dispersão** para descobrir o tipo de relacionamento ou dependência entre elas.



a) Histograma de distribuição de frequências

b) Diagrama de barras

c) Polígono de distribuição de frequências

d) Polígono de frequências acumuladas

e) Boxplot

f) Sector circular

Figura 1.2. Diferentes formas de representação gráfica de distribuições de frequências

1.3 MEDIDAS DE TENDÊNCIA CENTRAL

O tratamento e interpretação torna – se difícil para quem pretende colocar uma informação como um todo, razão pela qual constuma – se calcular algumas medidas que resumidamente traduzem as importantes caracteristicas das distribuições de frequencias.
Existem três grupos de medidas que resumem um conjunto de dados observados. As medidas de tendência central, as medidas de dispersão e as características de forma de distribuição.

Os índices de centro de distribuição ou medidas de tendência central (medidas de localizaçõ), assim designados em vertude da tendência que todos os dados observados têm em torno deses valores centrais. Este grupo inclui as grandezas médias, a moda, a mediana, os quartis, decis e percentis.

A média – é uma grandeza ou valor representativo de um conjunto de dados como eles tendem a se localizar em torno do ponto central. A média aritmética é a mais usada nas investigações, e esta pode ser simples ou ponderada quando temos frequências.

Amostra: $\bar{x} = \frac{1}{n}\sum x_i$ simples; $\bar{x} = \frac{1}{n}\sum x_i * f_i$ ponderada (1.5)

População $\mu = \frac{1}{N}\sum x_i$ simples $\mu = \frac{1}{N}\sum x_i * f_i$ ponderada

Moda: define –se como o valor da variável que ocorre com mais frequência. Para casos em que o número das observações não é muito elevado esta medida pode ser lida no rol. De acordo com as frequências de cada valor observado, é possível não ter a moda (amodal), ter uma moda (unimodal), duas modas (bimodal) três modas (trimodal), muitas modas (plurimodal).

Quando temos dados agrupados em classe, a moda é calculada por aproximação, começando por localizar a classe modal, se necessário deve –se conhecer a moda bruta. A formula de Czuba é uma das alternativas para calcular o valor da moda com aproximação.

$$Mo = x_o + \frac{fmo - fa}{2 * fmo - (fa + fp)} * i \quad (1.6)$$

Onde

x_o - Limite inferior da classe modal

f_{mo} - Frequência da classe modal

fa – Frequência anterior a classe modal

fp – Frequência posterior a classe modal e

i – Intervalo de classe da classe modal

Mediana – é uma separactriz que divide a distribuição de frequência ou conjunto de dados em duas partes iguais. A mediana é o valor que está na posição do meio para n impar e é igual a média aritmética dos dois valores centrais para n par.

Para dados agrupados em classe, a mediana é determinada por aproximação, bastando conhecer a classe mediana que se localiza calculando as frequências acumuladas.

$$Me = x_o + \frac{\frac{n}{2} - Fa}{fme} * i \qquad (1.7)$$

Onde

x_o – Limite inferior da classe mediana

fme - Frequência da classe mediana

Fa – Frequência acumulada da classe anterior a classe mediana

i – Intervalo de classe da classe mediana

Exemplo 1.6. Os dados que se seguem representam as percentagens do valor total que 17 credores do Millennium Bim retornaram ao banco durante um ano.

32.2, 29.5, 29.9, 32.4, 30.5, 30.1, 32.1, 35.2, 10.0, 20.6, 28.6, 30.5, 38.0, 33.0, 29.4, 37.1, 28.6

Determine a média, moda e mediana dos 17 credores.

Resolução

Média. $\bar{x} = \frac{\sum x_i}{n} = \frac{507.7}{17} = 29.86$

Moda: Para localizar o valor que está na moda vamos colocar os valores na ordem crescente.

10.0 20.6 **28.6** **28.6** 29.4 29.5 29.9 30.1 **30.5** **30.5** 32.1 32.2 32.4 33 35.2 37.1 38

Do rol temos: Mo1 = 28.6, Mo2 = 30.5, temos uma distribuição bimodal.

Mediana. $E(me) = \frac{n+1}{2} = \frac{17+1}{2} = 9$, o elemento mediano está na posição 9, vejamos:

10.0	20.6	28.6	28.6	29.4	29.5	29.9	30.1	**30.5**	30.5	32.1	32.2	32.4	33	35.2	37.1	38
1	2	3	4	5	6	7	8	**9**	10	11	12	13	14	15	16	17

Assim, a mediana é 30.5. Isto significa que 50% das pessoas pagaram menos de 30.5% do valor que receberam quando pediram o empréstimo e outras 50% das pessoas devolveram igual ou maior que 30.5% do total.

Exemplo 1.7. A partir da tabela abaixo, determine a média, moda e mediana da distribuição.

i	Classes de X			fi	xi	xi*fi	Fi
1	10	-	20	10	15	150	10
2	20	-	30	20	25	500	30
3	30	-	40	35	35	1225	65
4	**40**	**-**	**50**	**40**	45	1800	**105**
5	50	-	60	25	55	1375	130
6	60	-	70	15	65	975	145
7	70	-	80	5	75	375	150
soma		**-**		**150**	**-**	**6400**	**-**

Resolução

Média: $\bar{x} = \frac{\sum x_i * f_i}{n} = \frac{6400}{150} = 42.67$

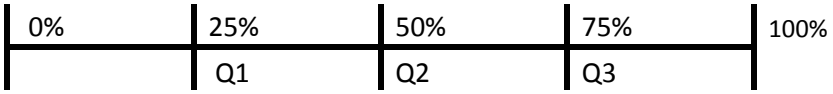
Moda: $Mo = x_o + \frac{fmo - fa}{2 * fmo - (fa + fp)} * i = 40 + \frac{40 - 35}{2 * 40 - (35 + 25)} * 10 = 42.5$

Mediana: $Me = x_o + \frac{\frac{n}{2} - Fa}{fme} * i = 40 + \frac{75 - 65}{40} * 10 = 42.5$

Quartis, Decis e Percentis

Como foi referido, para além da média, os quartis, decis e percentis são outros separa trizes utilizados como medidas de posição.

Os quartis dividem o conjunto de dados ou distribuições de frequências em 4 partes iguais.



Q1 – é um valor tal que 25% das observações são menores que este e 75% são superiores. Q2 coincide com a mediana e deixa 50% dos elementos em cada um dos subconjuntos e o Q3- é um valor tal que 75% das observações são menores e 25% são maiores que este.

Para dados não agrupados as posições onde se encontram os elementos pertencentes aos quartis podem ser determinados pelas fórmulas.

$$E(Q_1) = \frac{n+1}{4}; \quad E(Q_2) = \frac{n+1}{2} = E(me); \quad E(Q_3) = \frac{3*(n+1)}{4}; \qquad (1.8)$$

Para dados agrupados em classe, pode – se calcular com aproximação o quartil de ordem c, usando a fórmula.

$$Q_c = x_{oqc} + \frac{\frac{n*c}{4} - F_a}{f_{qc}} * i \; ; \quad \text{onde } c = 1,2,3 \tag{1.9}$$

Onde

c – é a ordem do quartil

x_{oqc} – limite inferior da classe onde existe o quartil

Fa – frequência acumulada até a classe anterior onde existe o quartil

f_{qc} – frequência absoluta da classe onde existe o quartil

Os decis são separactrizes que dividem as observações em 10 partes iguais.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100 %
	D1	D2	D3	D4	D5	D6	D7	D8	D9	

A fórmula básica para determinar a posição do elemento decil para dados não agrupados será:

$$E(D_i) = \frac{c*(n+1)}{10},$$

Quando os dados estão agrupados em classe a fórmula é semelhante as anteriores.

$$D_c = x_{odc} + \frac{\frac{n*c}{10} - F_a}{f_{dc}} * i \; ; \quad \text{onde } c = 1, 2,...,9 \tag{1.10}$$

Os percentis são separactriz que dividem as observações em 100 partes iguais.

0%	1%	2%	3%	.....	.....	.....	.....	98%	99%	100%
	P1	P2	P3	.....	.....	.....	.....	P98	P99	

Para calcular o percentil de ordem **c** basta usar a fórmula de aproximação quando temos dados classificados.

$$P_c = x_{opc} + \frac{\frac{n*c}{100} - F_a}{f_{pc}} * i \; ; \quad \text{onde } c = 1, 2,....., 99 \tag{1.11}$$

Exemplo 1.8. Registaram –se as alturas de 100 estudantes de uma Faculdade, tendo –se obtido a tabela de distribuição de frequências abaixo. Determine os valores das separactriz: Me, Q1, D3, P80 e comente os resultados.

i	Classes (xi, cm)			fi	Fi
1	1.40	---	1.46	7	7
2	1.46	---	1.52	9	16
3	1.52	---	1.58	5	21
4	1.58	---	1.64	12	33
5	1.64	---	1.70	26	**59**
6	1.70	---	1.76	20	79
7	1.76	---	1.82	11	90
8	1.82	---	1.88	10	100
soma		---		100	100

Resolução

Mediana: $Me = x_o + \frac{\frac{n}{2} - F_a}{f_{me}} * i = 1.64 + \frac{50 - 33}{26} * 0.06 = 1.68$

Primeiro Quartil: $Q_1 = x_{oq1} + \frac{\frac{n*1}{4} - F_a}{f_{q1}} * i = 1.58 + \frac{25 - 21}{12} * 0.06 = 1.60$

Terceiro Decil: $D_3 = x_{od3} + \frac{\frac{n*3}{10} - F_a}{f_{d3}} * i = 1.58 + \frac{30 - 21}{12} * 0.06 = 1.63$

Percentil 80: $P_{80} = x_{op80} + \frac{\frac{n*80}{100} - F_a}{f_{p80}} * i = 1.76 + \frac{80 - 79}{11} * 0.06 = 1.77$

Dos 100 estudantes, 25% destes possuem altura inferior a 1.60 cm, 30% possuem altura menor que 1.63 cm, 50% tem uma estatura inferior a 1.68 cm e só 20% dos 100 estudantes tem altura igual ou superior a 1.77 cm.

1.4 MEDIDAS DE DISPERSÃO

Como se sabe, nos fenómenos cuja análise intervêm o método estatístico, bem como nos dados estatísticos a eles referentes, caracterizam – se tanto pela sua semelhança quanto pela sua variabilidade. Assim, o cálculo de um promédio sobre um conjunto de dados têm sentido, caso contrário não teria sentido calcular a média de conjunto de dados onde não haja variação dos seus elementos.

Para medir o grau de variabilidade ou dispersão de um conjunto de valores são calculadas outras medidas estatísticas denominadas medidas de dispersão.

As medidas de dispersão permitem observar as flutuações das observações face as medidas de tendência central. Elas proporcionam um conhecimento mais completo do fenómeno a ser analisado em termos de um conjunto absoluto bem com relativo, estabelecendo comparações

entre fenómenos de mesma natureza, mostrando até que ponto os valores se distribuem acima e abaixo da medida de tendência central.

De um modo geral, as medidas de dispersão podem ser absolutas ou relativas. As medidas mais relevantes deste grupo das características de distribuições de frequências são a variância, o desvio padrão e o coeficiente de variação.

Variância ou dispersão de um conjunto de números, é a média aritmética dos quadrados dos desvios absolutos desses números em relação a sua média aritmética.

$$\sigma^2 = \frac{\sum \left(x_i - \bar{x}\right)^2}{N} = \frac{1}{N}\sum x_i^2 - (\bar{x})^2 \qquad (1.12)$$

Quando os dados observados estiverem agrupados em uma distribuição de frequências a fórmula da variância assume o seguinte aspecto.

$$\sigma^2 = \frac{\sum \left(x_i - \bar{x}\right)^2 * f_i}{N} = \frac{1}{N}\sum x_i^2 * f_i - (\bar{x})^2 \qquad (1.13)$$

Exemplo 1.9. Calcular a variância das notas obtidas num teste realizado para 50 estudantes.

i	x_i	*fi*	*xi * fi*	$(xi - \bar{x})$	$(xi - \bar{x})^2$	$(xi - \bar{x})^2 * fi$
1	8	7	56	-2.64	6.9696	48.7872
2	9	8	72	-1.64	2.6896	21.5168
3	10	9	90	-0.64	0.4096	3.6864
4	11	10	110	0.36	0.1296	1.296
5	12	8	96	1.36	1.8496	14.7968
6	13	4	52	2.36	5.5696	22.2784
7	14	4	56	3.36	11.2896	45.1584
soma	---	50	532	---	---	157.5200

Média: $\bar{x} = \frac{\sum x_i * f_i}{n} = \frac{532}{50} = 10.64$

Variância: $\sigma^2 = \frac{\sum \left(x_i - \bar{x}\right)^2 * f_i}{\sum f_i} = \frac{157.52}{50} = 3.15$

Desvio padrão - é a medida de dispersão mais usada ela é definida como a raiz quadrada positiva da média aritmética dos quadrados dos desvios dos valores observados em relação a grandeza média. As fórmulas para o cálculo estão abaixo conforme o tipo de dados.

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad \text{ou} \quad \sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2 * f_i}{\sum f_i}} \tag{1.14}$$

Se quisermos calcular o desvio padrão do exemplo anterior, bastará tirar a raiz quadrada de 3.15, assim $\sigma = \sqrt{3.15} = 1.77$ valores. Quando se trabalha com uma amostra e não com uma população, caso mais frequente na inferência estatística, ou quando o número das unidades observadas não é muito elevado (n <30), para obter uma melhor estimativa use –se o desvio padrão corrigido $(s \text{ ou } \sigma_{n-1})$.

Variância: $s^2 = \frac{1}{n-1}\sum \left(x_i - \bar{x}\right)^2 = \frac{1}{n-1}\left(\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right)$

Ou $s^2 = \frac{1}{n-1}\sum \left(x_i - \bar{x}\right)^2 * f_i = \frac{1}{n-1}\left(\sum x_i^2 * f_i - \frac{1}{n}\left(\sum x_i f_i\right)^2\right)$

Desvio padrão $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ ou $s = \sqrt{\frac{1}{n-1}(\sum x_i^2 - \frac{1}{n}(\sum x_i)^2)}$ (1.15)

Exemplo 1.10. calcular o desvio padrão e o seu valor corrigido nas seguintes alíneas.

i	x_i	*fi*	*xi * fi*	$(xi - \bar{x})^2 * fi$
1	2	4	8	10.50
2	3	9	27	3.46
3	4	8	32	1.16
4	5	3	15	5.71
5	6	2	12	11.33
soma	---	26	94	32.16

$\bar{x} = \frac{94}{26} = 3.62$

$\sigma = \sqrt{\frac{32.16}{26}} = 1.11$

$s = \sqrt{\frac{32.16}{25}} = 1.13$

Diferença = 0.02 ou $\rightarrow 2\%$

i	x_i	*fi*	*xi * fi*	$(xi - \bar{x})^2 * fi$
1	3	10	30	40.0
2	4	15	60	15.0
3	5	30	150	0.0
4	6	15	90	15.0
5	7	10	70	40.0
soma	---	80	400	110.0

$\bar{x} = \frac{400}{80} = 5.0$

$\sigma = \sqrt{\frac{110}{80}} = 1.17$

$s = \sqrt{\frac{110}{79}} = 1.18$

Diferença = 0.01 ou $\rightarrow 1\%$

Este exemplo mostra que quanto mais aumentamos o número dos elementos na amostra a diferença entre os desvios torna –se cada vez menor.

Na análise de dados, para determinadas situações, as medidas de variabilidade absolutas tais como o desvio padrão e a variância, não são muito expressivas, neste caso as medidas de dispersão relativas proporcionam uma avaliação mais apropriada do grau de dispersão da variável.

Os coeficientes de variação são medidas de variação relativas que muitas vezes são expressas em percentagem. Elas resultam do quociente entre uma medida de dispersão absoluta e uma medida de tendência central.

Quando a dispersão absoluta é igual ao desvio padrão e a medidas de tendência central é a média aritmética, a dispersão relativa é denominada coeficiente de variação de Pearson (Cv). Este é o coeficiente mais vulgar e é o mais utilizado designando –se simplesmente de coeficiente de variação ou coeficiente de dispersão.

$$C_v = \frac{\sigma}{\bar{x}} \quad \text{ou} \quad C_v = \frac{\sigma}{\bar{x}} * 100\% \tag{1.16}$$

Diz –se que a distribuição possui pequena variabilidade (dispersão) quando o coeficiente for até 10%, média dispersão ou variabilidade moderada quando estiver entre 10% a 20% e maior variabilidade ou grande dispersão quando superar a percentagem de 20%. Alguns autores consideram outras escalas como: $C_v < 15\%$ - baixa, $15\% \leq C_v \leq 30\%$ - moderada e $C_v > 30\%$ alta dispersão.

Exemplo 1.11. Considere que os alunos de uma turma realizaram dois testes. Os resultados estão apresentados abaixo.

Teste	Média	Desvio padrão	Coeficiente de variação
1	12 (escala de 0 – 20)	2.4 valores	0.20 → 20%
2	9 (escala de 0 – 14)	1.9 valores	0.20 → 20%

Comparando os valores da tabela entre os dois testes, conclui – se que o teste 1 embora aparenta maior dispersão absoluta, em geral a dispersão relativa dos resultados foi igual.

Em resumo, quando comparamos duas ou mais dispersões, podem ocorrer três tipos de situações:

1. As observações vêm expressas na mesma unidade de medida, e as médias são iguais ou muito próximas, é conveniente comparar com os valores de desvio padrão.
2. As observações vêm expressas na mesma unidade de medida, e as médias são significativamente diferentes, é conveniente comparar com as medidas de dispersão relativa como o coeficiente de variação de Pearson.
3. As observações vêm expressas em unidades de medida diferentes, neste caso, é totalmente conveniente comparar com as medidas de dispersão relativas como o coeficiente de variação de Pearson.

1.5 MEDIDAS DE FORMA DE DISTRIBUIÇÃO DE FREQUÊNCIAS

O terceiro grupo e último dos parâmetros que caracterizam uma distribuição de frequências de um conjunto de valores junto com as medidas de posição e de dispersão são os índices de forma de distribuição.

Assim, para complementar o estudo de uma distribuição no quadro da estatística descritiva a uma variável é necessário estudar as medidas de assimetria e curtose. As características mais importantes neste grupo são o grau de deformação ou assimetria e o grau de achatamento da curva de distribuição de frequências ou do histograma.

Chama – se **momento natural de ordem r**, de um conjunto de números ao valor dado pela formula (1.17).

$$m'_r = \frac{\sum x_i^r}{n} \quad \text{ou} \quad m'_r = \frac{\sum x_i^r * f_i}{\sum f_i} \tag{1.17}$$

O momento natural da primeira ordem (r = 1), é igual a média aritmética $m'_1 = \bar{x}$.

Chama –se **momento centrado na média de ordem r,** ao momento definido pela formula (1.18).

$$m_r = \frac{\sum (x_i - \bar{x})^r}{n} \quad \text{ou} \quad m_r = \frac{\sum (x_i - \bar{x})^r * f_i}{\sum f_i} \tag{1.18}$$

O segundo momento centrado na média $(r = 2)$ é igual a variância $m_2 = \sigma^2$.

Assimetria de uma distribuição e seus coeficientes

Chama – se **assimetria**, ao grau de desvio ou afastamento de uma curva de distribuição de frequências em relação a recta de simetria da distribuição normal.

- Uma curva de distribuição é **simétrica** quando a: $\vec{x} = Me = Mo$
- Uma curva de distribuição tem **assimetria negativa ou a esquerda** quando: $\vec{x} < Me < Mo$ ou $\vec{x} < Mo$
- Uma curva de distribuição tem **assimetria positiva ou à direita** quando: $\vec{x} > Me > Mo$ ou $\vec{x} > Mo$

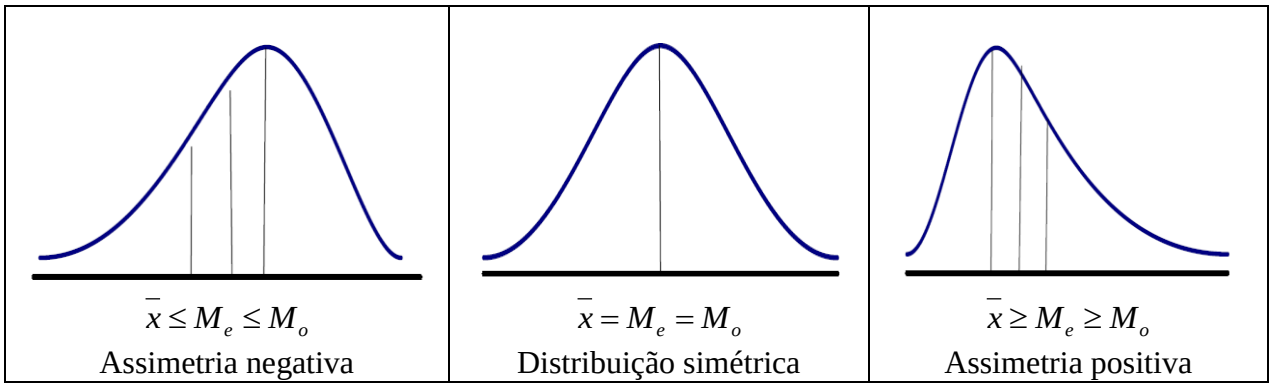


Figura 1.3. Posições relações entre as medidas de tendência central e a representação gráfica da assimetria de uma distribuição de frequências.

Coeficiente de assimetria

Karl Pearson, desenvolveu uma fórmula empírica da relação entre as três medidas de localização: a média (ponto de equilíbrio), a moda (ponto de máxima frequência) e a mediana (ponto do meio).

$$\bar{x} - M_o \approx 3*(\bar{x} - M_e) \tag{1.19}$$

Para medir ou avaliar o grau de assimetria são utilizados os coeficientes de assimetria de Pearson que podem ser obtidos partindo da expressão (1.19). O primeiro e segundo coeficientes de assimetria de Pearson são.

$$e_1 = \frac{\bar{x} - M_o}{\sigma} \;\; ; \;\; e_2 = \frac{3*(\bar{x} - M_e)}{\sigma} \quad \text{ou} \quad CS = \frac{\frac{1}{n}*\sum (x_i - \bar{x})^3 * f_i}{\sigma^3} \tag{1.20}$$

Calculado o coeficiente de assimetria, importa conhecer apenas o sinal do coeficiente quando não se necessita da extensão da assimetria apresentada pela curva.

a) Se CS < 0, temos assimetria negativa,
b) Se CS = 0, temos uma curva simétrica e
c) Se CS > 0, temos assimetria positiva.

Importa também referir que o grau de assimetria pode ser classificado conforme a escala apesentada independentemente de ser positiva ou negativa.

1. Se tivermos $-0.15 < CS < +0.15$ => pequena assimetria
2. Se tivermos $-1 \le CS \le +1$ => assimetria moderada
3. Se tivermos $CS < -1$ ou $CS > +1$ => assimetria elevada

Exemplo 1.12. Determine os tipos de assimetria das distribuições abaixo.

Distribuição A		
Classes	fi	xi
2 |----- 6	6	4
6 |---- 10	12	8
10 |--- 14	24	12
14 |--- 18	12	16
18 |--- 22	6	20
Total	60	---

Distribuição B		
Classes	fi	xi
2|---- 6	6	4
6 |--- 10	12	8
10 |-- -14	24	12
14 |--- 18	30	16
18 |--- 22	6	20
Total	78	---

Distribuição C		
Classes	fi	xi
2|--- 6	6	4
6 |--- 10	30	8
10 |-- -14	24	12
14 |--- 18	12	16
18 |--- 22	6	20
Total	78	--

Resolução

Calculadas as medidas de assimetria para as três distribuições pode se constatar

Dist.	média	mediana	moda	CS	d padrão	e1	e2	Tipo de assimetria
A	12.000	12.0	12.0	0.000	4.382	0.000	0.000	Curva simétrica
B	12.923	13.5	14.8	-0.471	4.196	-0.447	-0.413	Assimetria negativa
C	11.078	10.5	9.2	0.471	4.196	0.447	0.413	Assimetria positiva

Curtose de uma distribuição e seus coeficientes

Denomina – se **curtose** ao grau de achatamento de uma distribuição em relação a uma distribuição padrão, denominada curva normal (curva correspondente a uma distribuição teórica de probabilidade).

Quando a distribuição apresenta uma curva de frequência mais fechada que a normal (ou mais aguda ou afilada em sua parte superior), ela recebe o nome de **leptocúrtica**.

Quando a distribuição apresenta uma curva de frequência mais aberta que a normal (ou mais achatada em sua parte superior), ela recebe o nome de **platicúrtica**. E a curva normal, que é a nossa base de referência, recebe o nome de **mesocúrtica**.

Coeficiente de curtose

O coeficiente de curtose é calculado com base numa das fórmulas abaixo.

$$b_2 = \frac{m_4}{\sigma^4} \quad \text{ou} \quad CC = \frac{\frac{1}{n} * \sum (x_i - \bar{x})^4 * f_i}{\sigma^4} \quad \text{ou ainda} \quad C_1 = \frac{Q_3 - Q_1}{2 * (P_{90} - P_{10})} \tag{1.21}$$

Dependendo do coeficiente calculado a escala de comparação para se definir o tipo de curtose da distribuição de frequência na forma analítica poderá ser:

Para a expressão de C1	Para a expressão de CC
C1 < 0.263 => curva platicúrtica	CC < 3 => curva platicúrtica
C1 = 0.263 => curva mesocúrtica	CC = 3 => curva mesocúrtica
C1 > 0.263 => curva leptocúrtica	CC > 3 => curva leptocúrtica

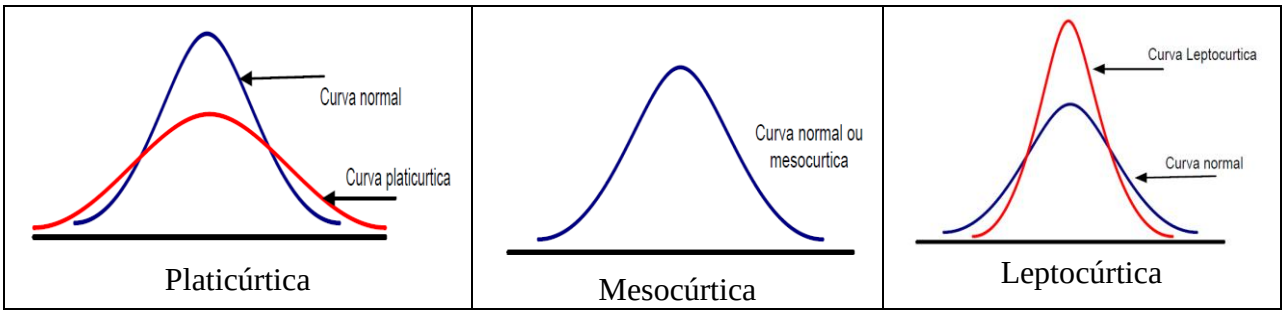


Figura 1.4. Representação gráfica do curtose de uma distribuição de frequências

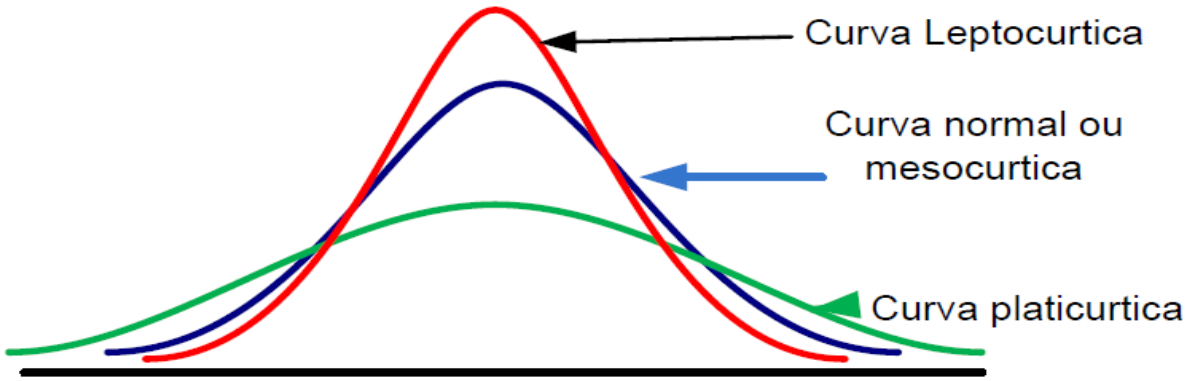


Figura 1.5. Curva da distribuição de frequências simétrica com diferentes graus de achatamento ou curtose.