Introdução à Estatística

Estatística

- ✓ **É a ciência que se preocupa com:**

(i) Organização; **(ii) Descrição;**	**Estatística Descritiva**
(iii) Análises; **(iv) Interpretações.**	**Estatística Indutiva ou Estatística Inferencial**

Alguns Conceitos

- **População**
  - É o conjunto de elementos com pelo menos uma característica em comum.
  - Esta característica comum deve delimitar claramente quais os elementos que pertencem à população e quais os elementos que não pertencem.

- **Amostra**
  - É um subconjunto de uma população, onde todos os seus elementos serão examinados para efeito da realização do estudo estatístico desejado.

Alguns Conceitos

- ✓ **OBJETIVO DA ESTATÍSTICA: "tirar conclusões sobre populações com base nos resultados observados em amostras extraídas dessas populações".**

- ✓ **Variável**

  - É a característica dos elementos da amostra que nos interessa averiguar estatisticamente.

  - Ex.: variável **Idade** - se houver "*n*" elementos fisicamente considerados no estudo, esses elementos fornecerão "*n*" valores da variável idade, os quais serão tratados convenientemente pela Estatística Descritiva e/ou pela Estatística Inferencial.

Tipos de Variáveis

As variáveis de interesse podem ser classificadas em:

(i) **Qualitativas** => quando resultar de uma classificação por tipos ou atributos.

(ii) **Quantitativas** => quando seus valores forem expressos em números. Podem ser subdivididas:

(a) **Discretas;**

(b) **Contínuas.**

Tipos de Variáveis

(a) Variáveis Quantitativas Discretas

Assumem apenas valores pertencentes a um conjunto enumerável. São obtidos mediante alguma forma de contagem.

Exemplos de Discretas:

- População: Ovinos da raça Santa Inês da ASCCO;
  Variável: número de cordeiros ao parto (1, 2 ou 3).
- População: Bovinos Nelore da Agro-pecuária CFM Ltda.
  Variável: Escores de Musculosidade (1, 2, 3, 4 ou 5).
- População: Bovinos Nelore da Agro-pecuária CFM Ltda.
  Variável: Prenhez aos 14 meses de idade (0 ou 1).

Tipos de Variáveis

(b) Variáveis Quantitativas Contínuas

São aquelas, teoricamente, que podem assumir qualquer valor em um certo intervalo de variação. Resultam, em geral, de uma medição, sendo freqüentemente expressos em alguma unidade.

Exemplos de Contínuas:

- População: Bovinos Nelore da Agro-pecuária CFM Ltda.
  Variável: PN (28,0; 28,5; 30,2; 32,58)
- População: Bovinos Nelore da Agro-pecuária CFM Ltda.
  Variável: Peso aos 18 meses, em kg (250,0 até 415,0 kg)

Características Numéricas de uma Distribuição de Dados

Introdução

- ✓ As vezes é necessário resumir certas características das distribuições de dados (ou mesmo de freqüências dados) por meio de certas quantidades.

- ✓ Tais quantidades são usualmente denominadas de **MEDIDAS**, por quantificarem alguns aspectos de nosso interesse.

- ✓ Nosso objetivo é apresentar algumas das chamadas **MEDIDAS DE POSIÇÃO**, bem como, algumas **MEDIDAS DE DISPERSÃO**, consideradas mais importantes no campo da aplicabilidade prática do nosso dia a dia.

- ✓ Tais medidas servem para:

  (a) Localizar uma distribuição;

  (b) Caracterizar sua variabilidade.

Medidas de Posição (ou de Tendência Central)

- ✓ Servem para localizar a distribuição dos dados brutos (ou das freqüências) sobre o eixo de variação da variável em questão.

- ✓ Veremos os três tipos principais de medidas de posição:

  (a) **Média Aritmética;**

  (b) **Mediana;**

  (c) **Moda.**

Medidas de Posição (ou de Tendência Central)

✓ **Média (Aritmética)**

=> A notação internacional recomenda símbolos específicos para a Média:

(a) AMOSTRA:

Conjunto de Dados => $$\overline{\mathrm{x}} = \hat{\mu} = \hat{\mathrm{m}} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Tabelas de Freqüência => $$\overline{\mathrm{x}} = \hat{\mu} = \hat{\mathrm{m}} = \frac{\sum_{i=1}^{k} X_i f_i}{n} = \sum_{i=1}^{k} X_i p'_i$$

Medidas de Posição (ou de Tendência Central)

✓ **Média (Aritmética)**

(b) POPULAÇÃO:

Conjunto de Dados => $$\mu = m = \frac{\sum_{i=1}^{n} X_i}{n}$$

Tabela de Freqüência => $$\mu = m = \frac{\sum_{i=1}^{k} X_i f_i}{n} = \sum_{i=1}^{k} X_i p'_i$$

Medidas de Posição (ou de Tendência Central)

Exemplo 2: 50 determinações do tempo (em segundos) gasto por um funcionário

Tabela 2.7 Cálculo da média

Classes (limites reais)	f_i	x_i	$x_i f_i$
39,5 — 44,5	3	42	126
44,5 — 49,5	8	47	376
49,5 — 54,5	16	52	832
54,5 — 59,5	12	57	684
59,5 — 64,5	7	62	434
64,5 — 69,5	3	67	201
69,5 — 74,5	1	72	72
	50		2.725

$$\overline{\mathrm{x}} = \hat{\mu} = \hat{\mathrm{m}} = \frac{\sum_{i=1}^{k} X_i f_i}{n} = \frac{2.725}{50} = 54{,}5$$

Medidas de Posição (ou de Tendência Central)

✓ **Propriedades da Média**

(a) Multiplicando todos os valores de uma variável por uma constante, a média do conjunto fica multiplicada por essa constante.

(b) Somando-se ou subtraindo-se uma constante a todos os valores da variável, a média do conjunto fica acrescida ou subtraída dessa constante.

Medidas de Posição (ou de Tendência Central)

✓ **Mediana**

=> A mediana é uma quantidade que, como a média, também caracteriza o centro de uma distribuição pertencente a um conjunto de dados.

(a) **AMOSTRA:** $\hat{m}d$

(b) **POPULAÇÃO:** md

Medidas de Posição (ou de Tendência Central)

Conjunto de Dados: Para obtenção da estimativa de mediana de um conjunto de dados são necessários os seguintes passos:

1º Passo: Ordenar de forma crescente os "*n*" valores da variável em questão;

2º Passo: **(i)** Sendo "*n*" **ímpar**, a mediana será igual ao valor de ordem $\frac{(n+1)}{2}$;

(ii) Sendo "*n*" **par**, a mediana será o valor médio entre os valores de ordem $\frac{n}{2}$ e $\frac{n}{2}+1$.

Medidas de Posição (ou de Tendência Central)

✓ **Mediana**

Tabelas de Freqüência => $$\hat{md} = L_i + \frac{(n/2) - F_a}{f_{md}} h_{md}$$

L_i = limite inferior da classe que contém a mediana;

n = números de elementos do conjunto da dados;

F_a = soma das freqüências das classes anteriores que contém a mediana;

f_{md} = freqüência da classe que contém a mediana;

h_{md} = amplitude da classe que contém a mediana.

Medidas de Posição (ou de Tendência Central)

✓ **Mediana**

Exemplo 2: 50 determinações do tempo (em segundos) gasto por um funcionário

Tabela 2.7 Cálculo da média

Classes (limites reais)	f_i	x_i	$x_i f_i$
39,5 — 44,5	3	42	126
44,5 — 49,5	8	47	376
49,5 — 54,5	16	52	832
54,5 — 59,5	12	57	684
59,5 — 64,5	7	62	434
64,5 — 69,5	3	67	201
69,5 — 74,5	1	72	72
	50		2.725

$$\hat{m}d = L_i + \frac{(n/2) - F_a}{f_{md}} h_{md}$$

$L_i = 49,5$; $n = 50$; $F_a = 11$; $f_{md} = 16$; $h_{md} = 5$.

Medidas de Posição (ou de Tendência Central)

✓ **Mediana**

Exemplo 2: 50 determinações do tempo (em segundos) gasto por um funcionário

L_i = 49,5; n = 50; F_a = 11; f_{md} = 16; h_{md} = 5.

$$\hat{m}d = L_i + \frac{(n/2) - F_a}{f_{md}} h_{md}$$

$$\hat{m}d = 49{,}5 + \frac{(50/2) - 11}{16} .5 = 53{,}875$$

Medidas de Posição (ou de Tendência Central)

✓ **Moda**

=> A moda (ou modas) de um conjunto de valores é definida como o valor (ou valores) de máxima freqüência.

=> É uma quantidade que, como a média, também caracteriza o centro de uma distribuição, indicando a região das máximas freqüências.

(a) **AMOSTRA:** $\hat{m}_o$

(b) **POPULAÇÃO:** m_o

Medidas de Posição (ou de Tendência Central)

✓ **Moda**

Tabelas de Freqüência => $$\hat{m}_o = L_i + \frac{d_1}{d_1 + d_2} h$$

L_i = limite inferior da classe modal;

d_1 = diferença entre a classe modal e a da classe imediatamente anterior;

d_2 = diferença entre a classe modal e a da classe imediatamente seguinte;

h = amplitude das classes.

Medidas de Posição (ou de Tendência Central)

✓ **Moda**

Exemplo 2: 50 determinações do tempo (em segundos) gasto por um funcionário

Tabela 2.7 Cálculo da média

Classes (limites reais)	f_i	x_i	$x_i f_i$
39,5 — 44,5	3	42	126
44,5 — 49,5	8	47	376
49,5 — 54,5	16	52	832
54,5 — 59,5	12	57	684
59,5 — 64,5	7	62	434
64,5 — 69,5	3	67	201
69,5 — 74,5	1	72	72
	50		2.725

$$\hat{m}_o = L_i + \frac{d_1}{d_1 + d_2} h$$

$L_i = 49,5$; $d_1 = 16 - 8 = 8$; $d_2 = 16 - 12 = 4$; $h = 5$.

Medidas de Posição (ou de Tendência Central)

✓ **Moda**

Exemplo 2: 50 determinações do tempo (em segundos) gasto por um funcionário

L_i = 49,5; d_1 = 16 – 8 = 8; d_2 = 16 – 12 = 4; h = 5.

$$\hat{m}_o = L_i + \frac{d_1}{d_1 + d_2} h$$

$$\hat{m}_o = 49{,}5 + \frac{8}{8+4} .5 = 52{,}833$$

Medidas de Dispersão (ou de Variabilidade)

- ✓ A informação fornecida pelas Medidas de Posição em geral necessitam de ser complementas pelas Medidas de Dispersão.

- ✓ As Medidas de Dispersão servem para indicar o "**quanto os dados se apresentam dispersos em torno da região central**".

- ✓ Portanto caracterizam o **grau de variação** existente em um conjunto de valores.

- ✓ As Medidas de Dispersão que mais nos interessam são:
  - (a) **Amplitude;**
  - (b) **Variância;**
  - (c) **Desvio Padrão;**
  - (d) **Coeficiente de Variação.**

Medidas de Dispersão (ou de Variabilidade)

✓ **Amplitude**

=> A amplitude, já mencionada, é definida como a diferença entre o maior e o menor valores do conjunto de dados.

(a) **AMOSTRA:** $\hat{R} = X_{MAX} - X_{MIN}$

(b) **POPULAÇÃO:** $R = X_{MAX} - X_{MIN}$

=> Vantagem e Desvantagem.

=> Salvo aplicações de **Controle de Qualidade**, a amplitude não é muito utilizada como Medida de Dispersão.

Medidas de Dispersão (ou de Variabilidade)

✓ **Variância**

=> A variância é definida como a "**média dos quadrados das diferenças entre os valores em relação a sua própria média**".

(a) **AMOSTRA:** $S^2 = S_X^2 = S^2(X) = \hat{\sigma}^2 = \hat{\sigma}^2(X) = \hat{\sigma}_X^2$

(b) **POPULAÇÃO:** $\sigma^2 = \sigma^2(X) = \sigma_X^2$

=> Em se tratando de **Amostra**:

Conjunto de Dados => $S^2(X) = S_X^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N-1}$

Tabela de Freqüência => $S^2(X) = S_X^2 = \dfrac{\sum_{i=1}^{k}(X_i - \overline{X})^2 f_i}{N-1}$

Medidas de Dispersão (ou de Variabilidade)

✓ **Variância**

=> Em se tratando de **População**:

Conjunto de Dados => $\sigma^2 = \sigma^2(X) = \sigma^2_X = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N}$

Tabela de Freqüência => $\sigma^2 = \sigma^2(X) = \sigma^2_X = \dfrac{\sum_{i=1}^{k}(X_i - \overline{X})^2 f_i}{N}$

OBS:

(i) A variância calculada para **dados agrupados** deverá ser **superestimada** em relação à variância exata dos "N" **dados originais**.

Medidas de Dispersão (ou de Variabilidade)

✓ **Variância**

Exemplo: Executar o cálculo da variância de um conjunto pequeno de dados, formado pelos valores seguinte: {15, 12, 10, 17, 16}

É fácil ver que: $\overline{\mathrm{x}} = \hat{\mu} = \hat{\mathrm{m}} = \frac{\sum_{i=1}^{n} X_i}{N} = 14$

Logo: $S^2(X) = S_X^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N-1}$

Poderemos montar a seguinte **Tabela Auxiliar** nos cálculos:

Medidas de Dispersão (ou de Variabilidade)

✓ Variância

Exemplo: Cálculo da variância de um conjunto pequeno de dados: {15, 12, 10, 17, 16}

Tabela 2.8 Cálculo de $\Sigma(x_i - \bar{x})^2$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
15	1	1
12	−2	4
10	−4	16
17	3	9
16	2	4
		34

$$S^2(X) = S_X^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N-1}$$

$$S^2(X) = S_X^2 = \frac{34}{4} = 8{,}5$$

Nota-se que as expressões apresentadas **não são as mais apropriadas** para o cálculo da variância, pois a média é quase sempre um **valor fracionário**, o que viria a dificultar o cálculo dos desvios $(X_i - \overline{X})^2$.

Medidas de Dispersão (ou de Variabilidade)

✓ **Variância**

Note que o numerador pode ser trabalhado: $S^2(X) = S_X^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

$$\Sigma(X_i - \overline{X})^2 = \Sigma(X_i^2 - 2X_i\overline{X} + \overline{X}^2)$$

$$= \Sigma X_i^2 - 2\overline{X}\Sigma X_i + N\overline{X}^2$$

$$= \Sigma X_i^2 - 2\frac{\Sigma X_i}{N}\Sigma X_i + N\left(\frac{\Sigma X_i}{N}\right)^2$$

$$= \Sigma X_i^2 - 2\frac{(\Sigma X_i)^2}{N} + \frac{(\Sigma X_i)^2}{N}$$

$$\boxed{\Sigma(X_i - \overline{X})^2 = \Sigma X_i^2 - \frac{(\Sigma X_i)^2}{N}}$$

Medidas de Dispersão (ou de Variabilidade)

✓ **Variância**

Assim, para um conjunto com "N" dados:

$$S^2(X) = S_X^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{N-1} = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{N}}{N-1}$$

Da mesma forma, para dados agrupados em **Tabela de freqüência**, teremos:

$$S^2(X) = S_X^2 = \frac{\sum_{i=1}^{k}(X_i - \overline{X})^2 f_i}{N-1} = \frac{\sum_{i=1}^{k} X_i^2 f_i - \frac{\left(\sum_{i=1}^{k} X_i f_i\right)^2}{N}}{N-1}$$

Medidas de Dispersão (ou de Variabilidade)

✓ **Variância**

Exemplo: 50 determinações do tempo (em segundos) gasto por um funcionário

Tabela 2.9 Cálculo da variância

Classes (limites reais)	f_i	x_i	$x_i f_i$	$x_i^2 f_i$
39,5 — 44,5	3	42	126	5.292
44,5 — 49,5	8	47	376	17.672
49,5 — 54,5	16	52	832	43.264
54,5 — 59,5	12	57	**684**	38.988
59,5 — 64,5	7	62	434	26.908
64,5 — 69,5	3	67	201	13.467
69,5 — 74,5	1	72	72	5.184
	50		2.725	150.775

$$S^2(X) = S_X^2 \frac{\sum_{i=1}^{k} X_i^2 f_i - \frac{\left(\sum_{i=1}^{k} X_i f_i\right)^2}{N}}{N-1} = \frac{150.775 - \frac{(2.725)^2}{50}}{49} = 46{,}17$$

Medidas de Dispersão (ou de Variabilidade)

✓ **Propriedades da Variância**

(a) Multiplicando-se todos os valores de uma variável por uma constante, a variância do conjunto fica multiplicada pelo **quadrado dessa constante**.

(b) Somando-se ou subtraindo-se uma constante a todos os valores de uma variável, a **variância não se altera**.

OBS: **(i)** A variância é uma medida de dispersão **importante** na teoria estatística;

(ii) Do ponto de vista prático, ela tem o inconveniente de se expressar em **unidade quadrática** em relação a variável em questão.

Medidas de Dispersão (ou de Variabilidade)

✓ **Desvio Padrão**

=> Definimos desvio padrão como "**a raiz quadrada positiva da variância**".

=> O cálculo do desvio padrão é feito por meio da **variância**.

(a) **AMOSTRA:** $S = S_X = S(X) = \hat{\sigma} = \hat{\sigma}(X) = \hat{\sigma}_X$

(b) **POPULAÇÃO:** $\sigma = \sigma(X) = \sigma_X$

=> Em se tratando de **Amostra**: $S(X) = S_X = +\sqrt{S_X^2}$

Medidas de Dispersão (ou de Variabilidade)

✓ **Desvio Padrão**

OBS: **(i)** O desvio padrão **se expressa na mesma unidade da variável**, sendo por isso, de maior interesse que a variância nas aplicações práticas;

(ii) É mais realístico para efeito de comparação de dispersões.

Exemplo: 50 determinações do tempo (em segundos) gasto por um funcionário

$$S^2(X) = S_X^2 \frac{\sum_{i=1}^{k} X_i^2 f_i - \frac{\left(\sum_{i=1}^{k} X_i f_i\right)^2}{N}}{N-1} = \frac{150.775 - \frac{(2.725)^2}{50}}{49} = 46{,}17$$

$$\boxed{S(X) = S_X = \sqrt{46{,}17} = 6{,}79}$$

Medidas de Dispersão (ou de Variabilidade)

✓ **Coeficiente de Variação**

=> O coeficiente de variação é definido como “**o quociente entre o desvio padrão e a média”**, sendo frequentemente expresso em porcentagem.

(a) AMOSTRA: $\hat{CV}(X) = \hat{CV}_X$

(b) POPULAÇÃO: $CV\ (X) = CV_X$

=> Em se tratando de **Amostra**:

$$\hat{CV}(X) = \hat{CV}_X = \frac{S_X}{\overline{\overline{X}}}$$

Medidas de Dispersão (ou de Variabilidade)

✓ **Coeficiente de Variação**

Exemplo: 50 determinações do tempo (em segundos) gasto por um funcionário

$$\hat{CV}(X) = \hat{CV}_X = \frac{S_X}{\bar{X}}$$

$$\hat{CV}(X) = \hat{CV}_X = \frac{S_X}{\bar{X}} = \frac{6{,}79}{54{,}5} = 0{,}125 = 12{,}46\%$$

Medidas de Dispersão (ou de Variabilidade)

✓ **Coeficiente de Variação**

OBS: **(i)** A vantagem é caracterizar a dispersão dos dados **em termos relativos ao seu valor médio**;

(ii) Pequena dispersão absoluta pode ser, na verdade **considerável**, quando comparada com a ordem de grandeza dos valores da variável. Quando consideramos o CV, enganos de interpretações desse tipo não ocorrem;

(iii) Além disso, por ser adimensional, o CV fornece uma maneira de se compararem as dispersões de variáveis cujas medidas são irredutíveis.

Momentos de uma Distribuição de Dados

Momentos de uma Distribuição

✓ **Alguns conceitos**

Definimos **o momento de ordem "*t*"** de um conjunto de dados como:

$$\boxed{M_t = \frac{\sum_{i=1}^{n} X_i^t}{n}}$$

Definimos **o momento de ordem "*t*" *centrado*** em relação a uma constante **"*a*"** como:

$$\boxed{M_t^a = \frac{\sum_{i=1}^{n} (X_i - a)^t}{n}}$$

Momentos de uma Distribuição de Freqüências

✓ **Alguns conceitos**

Já vimos que temos interesse no caso de "**momento centrado em relação a média**", o qual designaremos simplesmente por "**momento centrado**", dado por:

$$\boxed{m_t = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^t}{n}}$$

Também sabemos que, nos casos da **média** e da **variância**, as expressões podem ser reescritas levando-se em consideração **Tabelas de freqüências** dos diferentes valores existentes.

Momentos de uma Distribuição de Freqüências

✓ **Alguns conceitos**

Assim, para dados agrupados em **Tabela de Freqüência**, teremos:

$$M_t = \frac{\sum_{i=1}^{k} X_i^t f_i}{n}$$

=> Para **momento de ordem "*t*"**

$$M_t^a = \frac{\sum_{i=1}^{k} (X_i - a)^t f_i}{n}$$

=> Para **momento de ordem "*t*" centrado** em relação a uma constante **"*a*"**

$$m_t = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^t f_i}{n}$$

=> Para **momento de ordem "*t*" centrado** em relação a uma constante **"*média*"**

Momentos de uma Distribuição de Freqüências

✓ **Alguns conceitos**

Nos interessa particularmente saber calcular os **momentos centrados de terceira e quarta** ordem.

$$\boxed{m_t = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^t}{n}}$$

$$m_3 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^3}{n} = \frac{\sum_{i=1}^{n} X_i^3}{n} - 3\overline{X}\,\frac{\sum_{i=1}^{n} X_i^2}{n} + 2\overline{X}^3$$

$$m_4 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^4}{n} = \frac{\sum_{i=1}^{n} X_i^4}{n} - 4\overline{X}\,\frac{\sum_{i=1}^{n} X_i^3}{n} + 6\overline{X}^2\,\frac{\sum_{i=1}^{n} X_i^2}{n} - 3\overline{X}^4$$

Momentos de uma Distribuição de Freqüências

✓ **Alguns conceitos**

Havendo **Tabelas de Freqüências com "*k*" classes** a considerar, as expressões equivalentes são:

$$\boxed{m_t = \frac{\sum_{i=1}^{k}(X_i - \overline{X})^t f_i}{n}}$$

$$m_3 = \frac{\sum_{i=1}^{k} X_i^3 f_i}{n} - 3\overline{X}\,\frac{\sum_{i=1}^{k} X_i^2 f_i}{n} + 2\overline{X}^3$$

$$m_4 = \frac{\sum_{i=1}^{n} X_i^4 f_i}{n} - 4\overline{X}\,\frac{\sum_{i=1}^{n} X_i^3 f_i}{n} + 6\overline{X}^2\,\frac{\sum_{i=1}^{n} X_i^2 f_i}{n} - 3\overline{X}^4$$

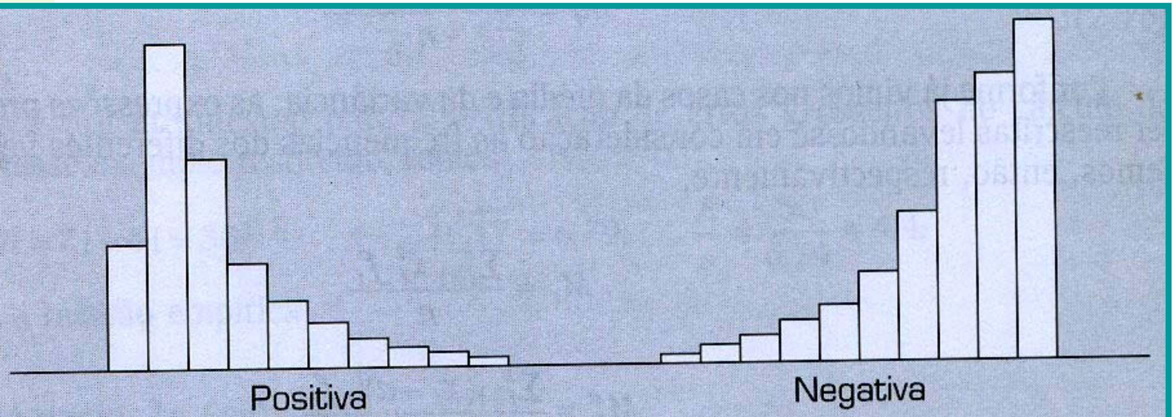
Medidas de Assimetria

Essas medidas procuram caracterizar **como** e **quanto** a distribuição dos Dados(ou freqüências) se afasta da condição de simetria.

Distribuições **alongadas a direita** são ditas **Positivamente Assimétricas**.

Distribuições **alongadas a esquerda** são ditas **Negativamente Assimétricas**.

Figura 2.12 *Assimetria positiva e assimetria negativa.*



Medidas de Assimetria

O **momento centrado de terceira ordem** pode ser usado como medida de assimetria.

Entretanto é mais conveniente a utilização de uma medida adimensional, definida como **Coeficiente de Assimetria**, dado por:

$$\boxed{a_3 = \frac{m_3}{(S_X)^3}}$$

Tabela 2.9 Cálculo da variância

Classes (limites reais)	f_i	x_i	$x_i f_i$	$x_i^2 f_i$
39,5 — 44,5	3	42	126	5.292
44,5 — 49,5	8	47	376	17.672
49,5 — 54,5	16	52	832	43.264
54,5 — 59,5	12	57	**684**	38.988
59,5 — 64,5	7	62	434	26.908
64,5 — 69,5	3	67	201	13.467
69,5 — 74,5	1	72	72	5.184
	50		2.725	150.775

Assim basta criamos uma nova coluna com $X_i^3 f_i$.

E utilizarmos **momento centrado de 3ª ordem**:

$$m_3 = \frac{\sum_{i=1}^{k} X_i^3 f_i}{n} - 3\bar{X}\frac{\sum_{i=1}^{k} X_i^2 f_i}{n} + 2\bar{X}^3$$

Medidas de Assimetria

Desta forma, poderemos classificar o **Coeficiente de Assimetria** (a_3) da seguinte forma:

(i) Se $a_3 = 0$ ➔ a distribuição é **Simétrica**;

(ii) Se $a_3 > 0$ ➔ a distribuição é **Assimétrica à direita** (Assimetria Positiva);

(iii) Se $a_3 < 0$ ➔ a distribuição é **Assimétrica à Esquerda** (Assimetria Negativa).

Fonte: Ferreira, D. F. Estatística Básica. Ed. UFLA, 2005. 664 p.

Medidas de Assimetria

Outra medida de assimetria mais simples pode ser obtido pelo **Índice de Assimetria de *Pearson***:

$$\boxed{A = \frac{\bar{X} - \hat{m}_0}{S_X}}$$

O **Índice de Assimetria de *Pearson*** também pode ser facilmente classificado:

$\boxed{|A| < 0{,}15}$ **=>** Distribuição praticamente **Simétrica**;

$\boxed{0{,}15 < |A| < 1{,}0}$ **=>** Distribuição moderadamente **Assimétrica**;

$\boxed{|A| > 1{,}0}$ **=>** Distribuição fortemente **Assimétrica**.

Medidas de Assimetria

Exemplo: 50 determinações do tempo (em segundos) gasto por um funcionário

$$\bar{x} = \hat{\mu} = \hat{m} = \frac{\sum_{i=1}^{k} X_i f_i}{n} = \frac{2.725}{50} = 54,5$$

$$\hat{m}_o = L_i + \frac{d_1}{d_1 + d_2} h$$

$$\hat{m}_o = 49,5 + \frac{8}{8+4} 5 = 52,833$$

$$S_X^2 = \frac{\sum_{i=1}^{k} X_i^2 f_i - \frac{\left(\sum_{i=1}^{k} X_i f_i\right)^2}{N}}{N-1} = 46,17$$

$$S_X = \sqrt{46,17} = 6,79$$

Tabela 2.7 Cálculo da média

Classes (limites reais)	f_i	x_i	$x_i f_i$
39,5 — 44,5	3	42	126
44,5 — 49,5	8	47	376
49,5 — 54,5	16	52	832
54,5 — 59,5	12	57	**684**
59,5 — 64,5	7	62	434
64,5 — 69,5	3	67	201
69,5 — 74,5	1	72	72
	50		2.725

$$A = \frac{\bar{X} - \hat{m}_0}{S_X} = \frac{54,5 - 52,833}{6,79} = 0,246$$

Medidas de Assimetria

Exemplo: 50 determinações do tempo (em segundos) gasto por um funcionário

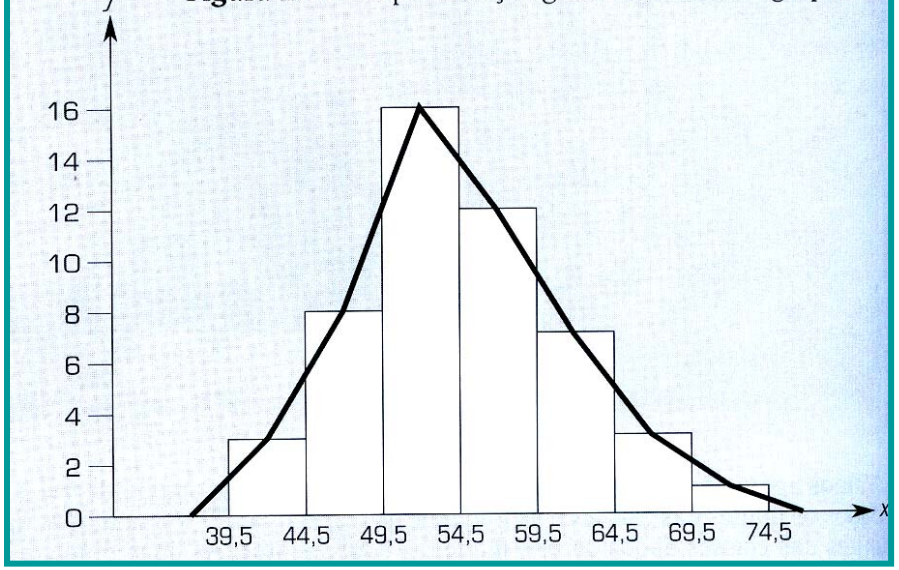
$$A = \frac{\overline{X} - \hat{m}_0}{S_X} = \frac{54{,}5 - 52{,}833}{6{,}79} = 0{,}246$$

Pelo **Índice de Assimetria de *Pearson*** essa distribuição seria classificada como "**Moderadamente Assimétrica**", pois

$$0{,}15 < |A| < 1{,}0 .$$

De fato isso ocorre, pois quando utilizados uma **Técnica de Descrição Gráfica** para Variáveis Quantitativas Contínuas, detectamos a **Assimetria Moderada**.

Figura 2.11 Representação gráfica dos dados agrupados.



Medidas de Achatamento ou Curtose

Essas medidas procuram caracterizar **a forma da distribuição quanto ao seu achatamento**.

O termo médio de comparação é dado pela **Distribuição Normal**, que é um modelo teórico de distribuição a ser estudado no capítulo relacionado à Probabilidades.

Quanto ao achatamento, podemos ter as seguintes situações: **Platicúrticas**, **Mesocúrticas** e **Leptocúrticas**.

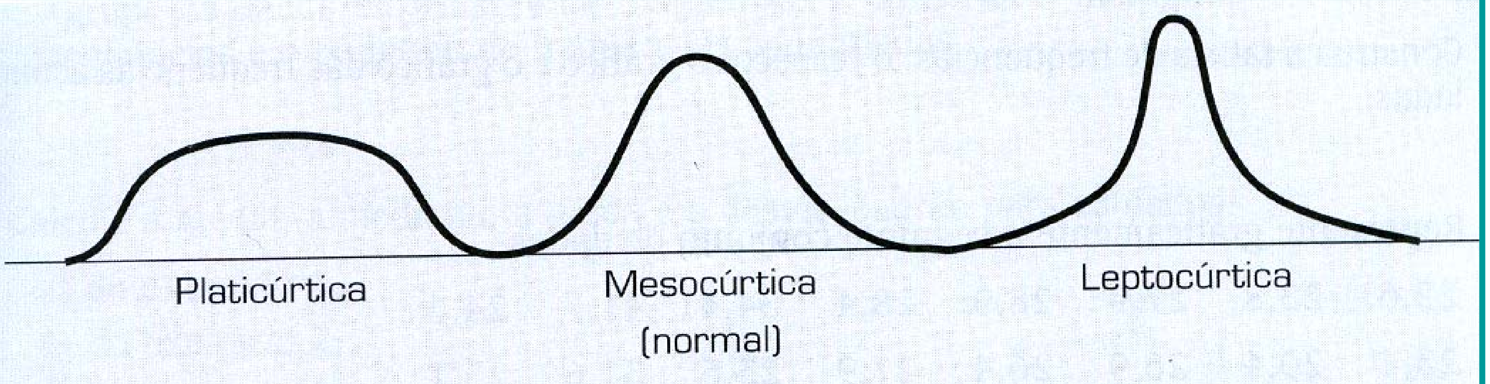


Figura 2.13 *Distribuições platicúrtica, mesocúrtica e leptocúrtica.*

Medidas de Achatamento ou Curtose

A caracterização do achatamento de uma distribuição só tem sentido, em termos práticos, se a distribuição for **aproximadamente Simétrica**.

Entre as possíveis medidas de achatamento, destacamos o **Coeficiente de Curtose**.

O **Coeficiente de Curtose** é obtido pelo quociente do momento centrado de 4ª ordem pelo quadrado da variância, ou seja:

$$\boxed{a_4 = \frac{m_4}{(S_X^2)^2} = \frac{m_4}{S_X^4}}$$

Medidas de Achatamento ou Curtose

Trata-se de coeficiente **adimensional**, permitindo a sua classificação:

$\boxed{a_4 < 3,0}$ => Distribuição **Platicúrtica**;

$\boxed{a_4 = 3,0}$ => Distribuição **Mesocúrtica**;

$\boxed{a_4 > 3,0}$ => Distribuição **Leptocúrtica**.

Medidas de Achatamento ou Curtose

Exemplo: 50 determinações do tempo (em segundos) gasto por um funcionário

Tabela 2.9 Cálculo da variância

Classes (limites reais)	f_i	x_i	$x_i f_i$	$x_i^2 f_i$
39,5 — 44,5	3	42	126	5.292
44,5 — 49,5	8	47	376	17.672
49,5 — 54,5	16	52	832	43.264
54,5 — 59,5	12	57	**684**	38.988
59,5 — 64,5	7	62	434	26.908
64,5 — 69,5	3	67	201	13.467
69,5 — 74,5	1	72	72	5.184
	50		2.725	150.775

Assim, basta criamos duas novas colunas com: $X_i^3 f_i$ e $X_i^4 f_i$.

E utilizarmos **momento centrado de 4ª ordem**:

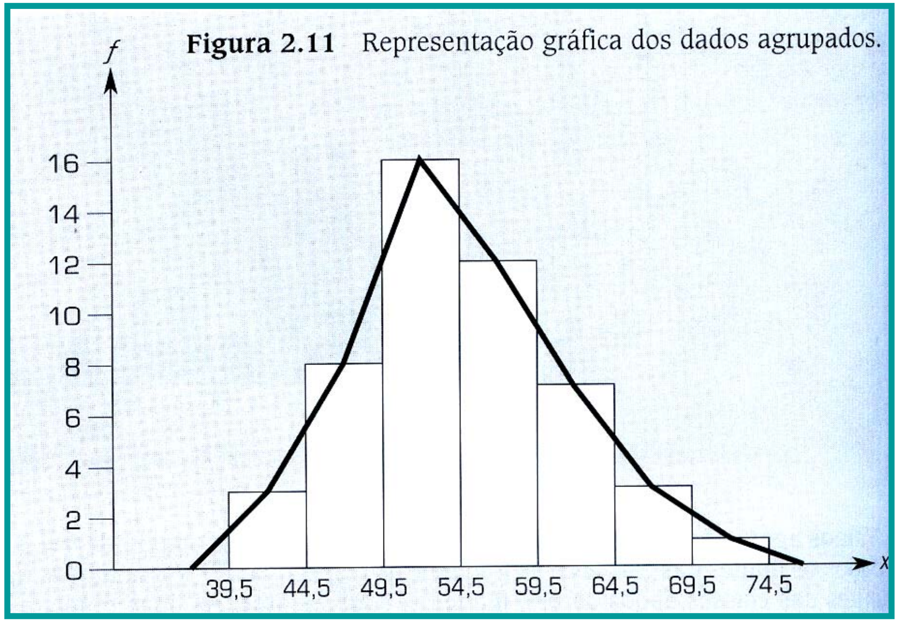
$$m_4 = \frac{\sum_{i=1}^{n} X_i^4 f_i}{n} - 4\overline{X}\frac{\sum_{i=1}^{n} X_i^3 f_i}{n} + 6\overline{X}^2\frac{\sum_{i=1}^{n} X_i^2 f_i}{n} - 3\overline{X}^4$$

Medidas de Achatamento ou Curtose

Exemplo: 50 determinações do tempo (em segundos) gasto por um funcionário

$$a_4 = \frac{m_4}{(S_X^2)^2} = \frac{m_4}{S_X^4} \cong 2{,}21$$

=> Distribuição ligeiramente **Platicúrtica**.



Medidas de Achatamento ou Curtose

Outra medida de achatamento mais simples pode ser obtido pelo **Grau de Curtose**, dado pelo coeficiente:

$$\boxed{K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}}$$

em que,

Q_3 = é o 3º Quartil;

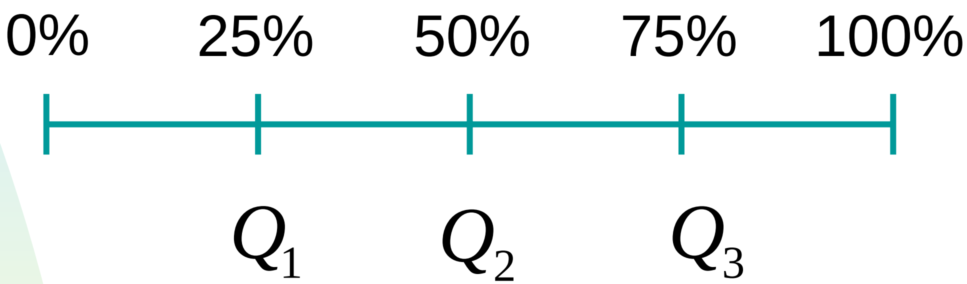
Q_1 = é o 1º Quartil;

P_{90} = é o 90º Percentil;

P_{90} = é o 10º Percentil.

Medidas de Achatamento ou Curtose

Quartis => dividem um conjunto de dados em quatro partes iguais.



em que,

Q_1 = o 1º Quartil deixa 25% dos elementos;

Q_2 = o 2º Quartil deixa 50% dos elementos e coincide com a **Mediana**;

Q_3 = o 3º Quartil deixa 75% dos elementos.

Medidas de Achatamento ou Curtose

- ✓ Fórmulas para cálculo de Q_1 e Q_3 para o caso de variáveis quantitativas contínuas

(a) Determinação de Q_1:

(i) Calcula-se: $\frac{N}{4}$;

(ii) Identifica-se a classe de Q_1 pela F_i (freq. acumulada);

(iii) Aplica-se a fórmula:

$$\boxed{Q_1 = L_{Q_1} + \frac{(n/4) - F_a}{f_{Q_1}} h}$$

Medidas de Achatamento ou Curtose

- ✓ Fórmulas para cálculo de Q_1 e Q_3 para o caso de variáveis quantitativas contínuas (continuação)

(b) Determinação de Q_3:

(i) Calcula-se: $\dfrac{3N}{4}$

(ii) Identifica-se a classe de Q_3 pela F_i (freq. acumulada);

(iii) Aplica-se a fórmula:

$$\boxed{Q_3 = L_{Q_3} + \frac{(3n/4) - F_a}{f_{Q_3}} h}$$

Medidas de Achatamento ou Curtose

Exemplo: Dada a distribuição, determinar os Quartis (Q_1 e Q_3) e a mediana.

Classes	f_i	F_i	
7 – 17	6	6	
17 – 27	15	21	Classe Q_1
27 – 37	20	41	Classe $\hat{m}d$
37 – 47	10	51	Classe Q_3
47 – 57	5	56	

$$Q_1 = L_{Q_1} + \frac{(n/4) - F_a}{f_{Q_1}} h$$

$$\hat{m}d = L_i + \frac{(n/2) - F_a}{f_{md}} h_{md}$$

$$Q_3 = L_{Q_3} + \frac{(3n/4) - F_a}{f_{Q_3}} h$$

Medidas de Achatamento ou Curtose

Exemplo: Dada a distribuição, determinar os Quartis (Q_1 e Q_3) e a mediana.

Classes	f_i	F_i
7 – 17	6	6
17 – 27	15	21
27 - 37	20	41
37 - 47	10	51
47 - 57	5	56

n = 56;

$$Q_1 = \frac{n}{4} = \frac{56}{4} = 14^\circ \text{ elemento}$$

$$Q_3 = \frac{3n}{4} = \frac{3.56}{4} = 42^\circ \text{ elemento}$$

$$\hat{m}d = \frac{\left(\frac{n}{2}\right) + \left(\frac{n}{2} + 1\right)}{2} = \frac{\left(\frac{56}{2}\right) + \left(\frac{56}{2} + 1\right)}{2} = 28^\circ \text{ e } 29^\circ \text{ elementos}$$

Medidas de Achatamento ou Curtose

Exemplo: Dada a distribuição, determinar os Quartis (Q_1 e Q_3) e a mediana.

Classes	f_i	F_i
7 – 17	6	6
17 – 27	15	21
27 - 37	20	41
37 - 47	10	51
47 - 57	5	56

$$\boxed{Q_1 = L_{Q_1} + \frac{(n/4) - F_a}{f_{Q_1}} h}$$

$$\boxed{\hat{m}d = L_i + \frac{(n/2) - F_a}{f_{md}} h_{md}}$$

$$\boxed{Q_3 = L_{Q_3} + \frac{(3n/4) - F_a}{f_{Q_3}} h}$$

Para Q_1 temos:

$L_{Q_1} = 17$; $n = 56$; $F_a = 6$;

$h = 10$; $f_{Q_1} = 15$

Para $\hat{m}d$ temos:

$L_i = 27$; $n = 56$; $F_a = 21$;

$h = 10$; $f_{\hat{m}d} = 20$

Para Q_3 temos:

$L_{Q_3} = 37$; $n = 56$; $F_a = 41$;

$h = 10$; $f_{Q3} = 10$

Medidas de Achatamento ou Curtose

Exemplo: Dada a distribuição, determinar os Quartis (Q_1 e Q_3) e a mediana.

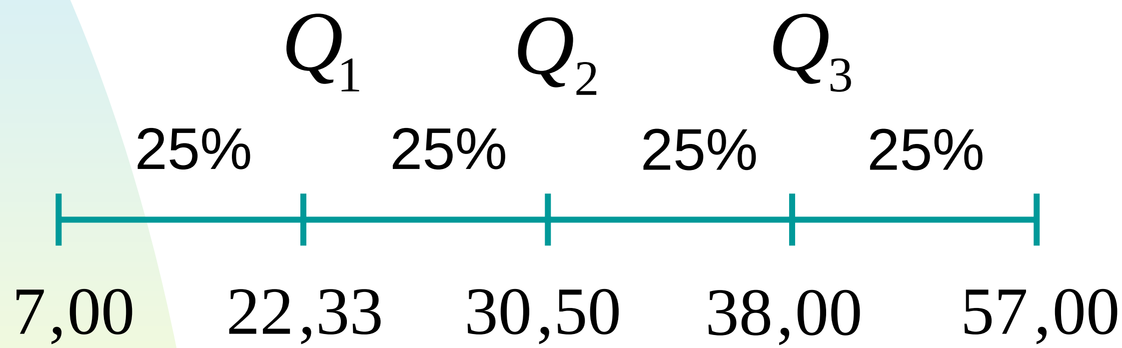
$$Q_1 = L_{Q_1} + \frac{(n/4) - F_a}{f_{Q_1}} h = 17 + \frac{\left(\frac{56}{2} - 6\right)}{15}.10 = 22{,}33$$

$$\hat{m}d = L_i + \frac{(n/2) - F_a}{f_{md}} h_{md} = 27 + \frac{\left(\frac{56}{2} - 21\right)}{15}.10 = 30{,}50$$

$$Q_3 = L_{Q_3} + \frac{(3n/4) - F_a}{f_{Q_3}} h = 37 + \frac{\left(\frac{3.56}{4} - 41\right)}{10}.10 = 38{,}00$$

Medidas de Achatamento ou Curtose

Exemplo: Dada a distribuição, determinar os Quartis (***Q_1*** e ***Q_3***) e a mediana.



Medidas de Achatamento ou Curtose

Decis **=>** são os valores que dividem um conjunto de dados em **10 partes iguais**.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	

em que,

D_1 = o 1º Decil deixa 10% dos elementos;

D_2 = o 2º Decil deixa 20% dos elementos;

... ...

D_9 = o 9º Decil deixa 90% dos elementos.

Medidas de Achatamento ou Curtose

Determinação de um Decil ***D_i***:

(i) Calcula-se: $\frac{i.N}{10}$ em que i = 1, 2, ..., 9**;**

(ii) Identifica-se a classe de ***D_i*** pela F_i (freq. acumulada);

(iii) Aplica-se a fórmula:

$$\boxed{D_i = L_i + \frac{(i.N/10) - F_a}{f_{Di}} h}$$

em que,

L_i = limite inferior da classe D_i ;

n = tamanho da amostra;

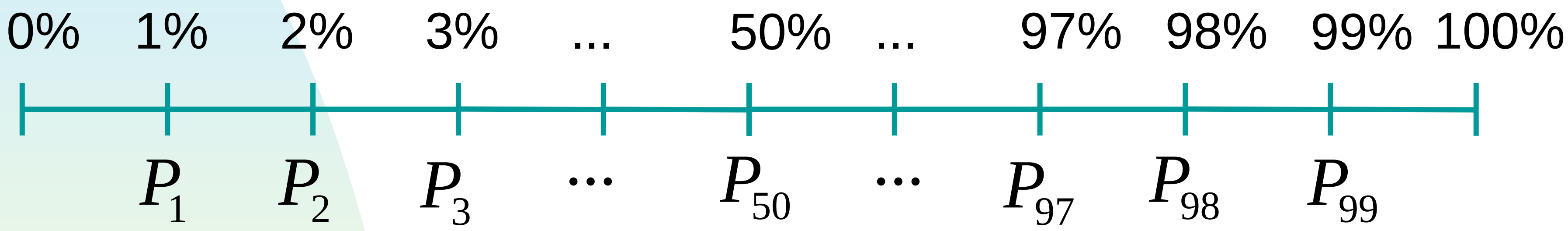
F_a = soma das freqüências das classes anteriores a que D_i

f_{Di}= freqüência da classe D_i;

h = amplitude da classe D_i.

Medidas de Achatamento ou Curtose

Percentis **=>** são os valores que dividem um conjunto de dados em **100 partes iguais**.



em que,

P_1 = o 1º Percentil deixa 1% dos elementos;

P_2 = o 2º Percentil deixa 2% dos elementos;

... ...

P_{99} = o 99º Percentil deixa 99% dos elementos.

Medidas de Achatamento ou Curtose

Determinação de um Percentil P_i:

(i) Calcula-se: $\frac{i.N}{100}$ em que i = 1, 2, ..., 98, 99;

(ii) Identifica-se a classe de P_i pela F_i (freq. acumulada);

(iii) Aplica-se a fórmula:

$$\boxed{P_i = L_i + \frac{(i.N/100) - F_a}{f_{P_i}} h}$$

em que,

L_i = limite inferior da classe P_i;

n = tamanho da amostra;

F_a = soma das freqüências das classes anteriores a que P_i;

f_{Pi}= freqüência da classe P_i;

h = amplitude da classe D_i.

Medidas de Achatamento ou Curtose

Exemplo: Dada a distribuição, determinar o Grau de Curtose (K).

Classes	f_i	F_i
7 – 17	6	6
17 – 27	15	21
27 - 37	20	41
37 - 47	10	51
47 - 57	5	56

$$\boxed{K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}}$$

Já tínhamos obtidos:

$Q_1 = 22{,}33$ e $Q_3 = 38{,}00$

$$\boxed{P_i = L_i + \frac{(i.N/100) - F_a}{f_{P_i}} h}$$

Para P_{10} temos:

$L_{P_{10}} = 7$; $n = 56$; $F_a = 0$;

$h = 10$; $f_{P_{10}} = 6$

$$\boxed{P_{10} = 16{,}33}$$

Para P_{90} temos:

$L_{P_{90}} = 37$; $n = 56$; $F_a = 41$;

$h = 10$; $f_{P_{90}} = 10$

$$\boxed{P_{90} = 46{,}40}$$

Medidas de Achatamento ou Curtose

Exemplo: Dada a distribuição, determinar o Grau de Curtose (*K*).

Classes	f_i	F_i
7 – 17	6	6
17 – 27	15	21
27 - 37	20	41
37 - 47	10	51
47 - 57	5	56

$$\boxed{K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}}$$

Agora temos tudo:

$Q_1 = 22{,}33$ e $Q_3 = 38{,}00$

$P_{10} = 16{,}33$ e $P_{90} = 46{,}40$

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{38{,}00 - 22{,}33}{2(46{,}40 - 16{,}33)} = 0{,}2606$$

Medidas de Achatamento ou Curtose

Assim o **Grau de Curtose**, de ser classificado da seguinte forma:

$$\boxed{K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}}$$

$\boxed{K = 0,263}$ **=>** Distribuição de freqüência **Mesocúrtica**;

$\boxed{K > 0,263}$ **=>** Distribuição de freqüência **Platicúrtica**;

$\boxed{K < 0,263}$ **=>** Distribuição de freqüência **Leptocúrtica**.