

Estatística Descritiva

Prof. Lorí Viali, Dr.
viali@mat.ufrgs.br
<http://www.mat.ufrgs.br/~viali/>

2/2



Tratamento de grandes conjuntos de dados



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Grande Conjuntos de Dados

- ✚ Organização;
- ✚ Resumo;
- ✚ Apresentação.

Amostra
ou
População



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Dados não organizados



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Dados Brutos

Variável qualitativa

Defeitos em uma linha de produção

Lascado	Menor
Desenho	Maior
Torto	Lascado
Desenho	Esmalte
Torto	Esmalte
Lascado	Lascado
Torto	Desenho
Maior	Menor
Menor	Maior
Desenho	Torto
.....



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Prof. Lorí Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

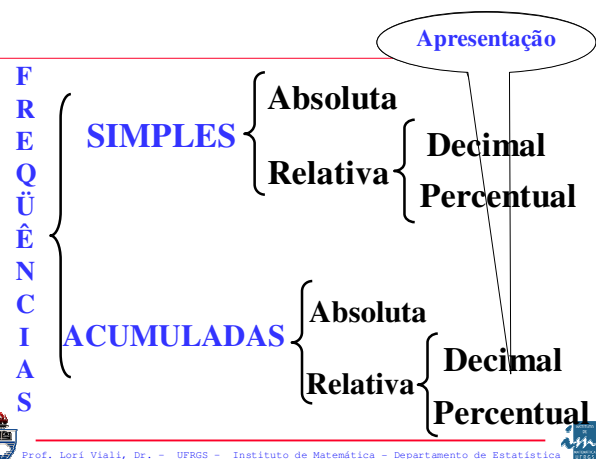


Dados organizados
em uma distribuição
de freqüências
* Variável qualitativa *

Distribuição de freqüências

Defeito	Freqüência	%
Desenho	71	14,20
Esmalte	95	19,00
Lascado	97	19,40
Maior	70	14,00
Menor	83	16,60
Torto	57	11,40
Trincado	27	5,40
TOTAL	500	100

Freqüências (Tipos)

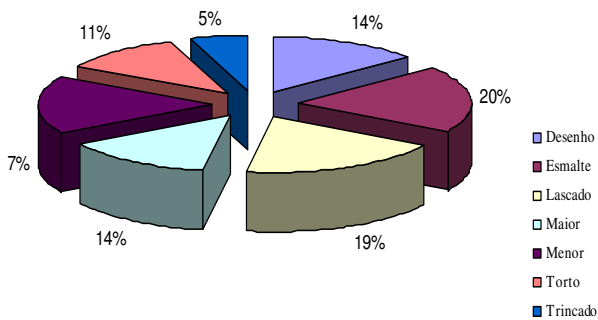


Freqüências: representação

Valores	f_i	F_i	fr_i	fr_i	Fr_i
0	60	60	0,30	30	30
1	50	110	0,25	25	55
2	40	150	0,20	20	75
3	30	180	0,15	15	90
4	10	190	0,05	5	95
5	6	196	0,03	3	98
6	4	200	0,02	2	100
TOTAL	200	—	1,00	100	—

Representação gráfica
Diagrama de torta
ou pizza (Pie Chart)

Defeitos em uma linha de produção



Dados Brutos Variável discreta

Número de irmãos dos alunos da turma G – Pro. & Estatística - UFRGS - 2009/01

0	1	1	6	3	1	3	1	1	0
4	5	1	1	1	0	2	2	4	1
3	1	2	1	1	1	1	5	5	6
4	1	1	0	2	1	4	3	2	2
1	0	2	1	1	2	3	0	1	0

Distribuição de freqüências por ponto ou valores

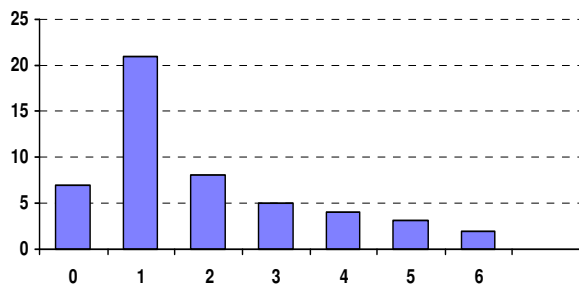
Distribuição de freqüências **por ponto ou valores** da variável:
“Número de irmãos dos alunos da turma G” da disciplina:
 Probabilidade e Estatística UFRGS
 - 2009/01.

Nº de irmãos	Nº de alunos
0	7
1	21
2	8
3	5
4	4
5	3
6	2
Σ	50

Representação gráfica

* Diagrama de colunas simples *

Diagrama de colunas simples da variável: **Número de irmãos dos alunos da turma G** Disciplina: Probabilidade e Estatística, UFRGS - 2009/01



Resumo de uma Distribuição de frequências por ponto ou valores

Medidas de tendência ou posição central

A média Aritmética

Neste caso, a média é dada por:

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + \dots + f_k \cdot x_k}{f_1 + f_2 + \dots + f_k} = \frac{\sum f_i \cdot x_i}{n}$$

Exemplo

x_i	f_i	$f_i x_i$
0	7	0
1	21	21
2	8	16
3	5	15
4	4	16
5	3	15
6	2	12
Σ	50	95

A média será, então:

$$\bar{x} = \frac{\sum f_i \cdot x_i}{n} = \frac{95}{50} = 1,90 \text{ irmãos}$$

A Mediana

Como $n = 50$ é par, tem-se:

$$m_e = \frac{X_{n/2} + X_{(n/2)+1}}{2} = \frac{X_{50/2} + X_{(50/2)+1}}{2} =$$

$$= \frac{X_{25} + X_{26}}{2} = \frac{1+1}{2} = 1 \text{ irmão}$$

Exemplo

x_i	f_i	F_i
0	7	7
1	21	28
2	8	36
3	5	41
4	4	45
5	3	48
6	2	50
Σ	50	—

Total de dados
 $n = 50$
(par)

Metade dos dados
 $n/2 = 25$

A Moda

$m_o =$ valor(es) que mais se repete(m)

Exemplo

x_i	f_i
0	7
1	21
2	8
3	5
4	4
5	3
6	2
Σ	50

Pois ele se repete mais vezes

Medidas de dispersão ou variabilidade

A Amplitude

$$h = x_{\text{máx}} - x_{\text{mín}}$$

$$h = 6 - 0 = 6 \text{ irmãos}$$

O Desvio Médio

Neste caso, o dma será dado por:

$$\begin{aligned} \text{dma} &= \frac{f_1|x_1 - \bar{x}| + f_2|x_2 - \bar{x}| + \dots + f_k|x_k - \bar{x}|}{f_1 + f_2 + \dots + f_k} = \\ &= \frac{\sum f_i \cdot |x_i - \bar{x}|}{n} \end{aligned}$$

Exemplo

x_i	f_i	$f_i x_i - \bar{x} $
0	7	$7 \cdot 0 - 1,90 = 13,30$
1	21	$21 \cdot 1 - 1,90 = 18,90$
2	8	$8 \cdot 2 - 1,90 = 0,80$
3	5	$5 \cdot 3 - 1,90 = 5,50$
4	4	$4 \cdot 4 - 1,90 = 8,40$
5	3	$3 \cdot 5 - 1,90 = 9,30$
6	2	$2 \cdot 6 - 1,90 = 8,20$
Σ	50	64,40

O dma será, então:

$$\text{dma} = \frac{\sum f_i \cdot |x_i - \bar{x}|}{n} = \frac{64,40}{50} = 1,29 \text{ irmãos}$$

A Variância

Neste caso, a variância será:

$$\begin{aligned} s^2 &= \frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_k(x_k - \bar{x})^2}{n} = \\ &= \frac{\sum f_i(x_i - \bar{x})^2}{n} = \frac{\sum f_i x_i^2}{n} - \bar{x}^2 \end{aligned}$$

Exemplo

x_i	f_i	$f_i x_i^2$
0	7	$0^2 \cdot 7 = 0$
1	21	$1^2 \cdot 21 = 21$
2	8	$2^2 \cdot 8 = 32$
3	5	$3^2 \cdot 5 = 45$
4	4	$4^2 \cdot 4 = 64$
5	3	$5^2 \cdot 3 = 75$
6	2	$6^2 \cdot 2 = 72$
Σ	50	299



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



A variância será, então:

$$s^2 = \frac{\sum f_i x_i^2}{n} - \bar{x}^2 = \frac{299}{50} - 1,90^2 = 2,3700 \text{ irmãos}^2$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O Desvio Padrão

O desvio padrão será dado por:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \bar{x}^2} = \sqrt{2,3700} = 1,5395 \cong 1,54 \text{ irmãos}$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O Coeficiente de Variação

Dividindo a média pelo desvio padrão, tem-se o coeficiente de variação:

$$g = \frac{1,539480}{1,90} = 81,03 \%$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Dados Brutos

Variável contínua

Idade (em meses) dos alunos
da turma G da disciplina:
Probabilidade e Estatística
UFRGS - 2009/01



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



276 245 345 240 270 310 368
334 268 288 336 299 236 239 355 330
287 344 300 244 303 248 251 265 246
240 320 308 299 312 324 289 320 264
252 298 315 255 274 264 263 230 303
369 247 266 275 281 230 234



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Distribuição de frequências por classes ou intervalos



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Distribuição por classes ou intervalos da variável “idade dos alunos da turma G” da disciplina: Probabilidade e Estatística da UFRGS - 2009/01



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Idades	Número de alunos
230 --- 250	12
250 --- 270	9
270 --- 290	8
290 --- 310	7
310 --- 330	6
330 --- 350	5
350 --- 370	3
Total	50



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Representação gráfica * Histograma *



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

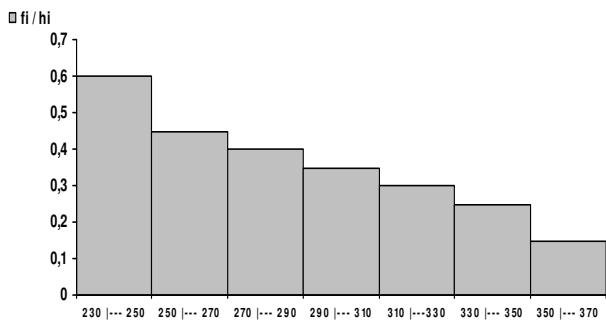


Histograma de frequências da variável “**Idade dos alunos da turma G**” de Probabilidade e Estatística da UFRGS - 2009/01



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística





Medidas

Antes de apresentar as medidas, i. é, representantes do conjunto, é necessário estabelecer uma notação para alguns elementos da distribuição.

Simbologia

- x_i = ponto médio da classe;
- f_i = frequência simples da classe;
- li_i = limite inferior da classe;
- ls_i = limite superior da classe;
- h_i = amplitude da classe.

O Ponto Médio da Classe

x_i	f_i	x_i
230 --- 250	12	240
250 --- 270	9	260
270 --- 290	8	280
290 --- 310	7	300
310 --- 330	6	320
330 --- 350	5	340
350 --- 370	3	360
Σ	50	—

Medidas de tendência ou posição central

A Média da Distribuição

x_i	f_i	$f_i \cdot x_i$
240	12	2880
260	9	2340
280	8	2240
300	7	2100
320	6	1920
340	5	1700
360	3	1080
Σ	50	14260

Exemplo

A média será:

$$\bar{x} = \frac{\sum f_i \cdot x_i}{n} = \frac{14260}{50} = 285,20 \text{ meses}$$

A Mediana

Neste caso, utilizam-se as frequências acumuladas para identificar a classe mediana, i. é, a que contém o(s) valor(es) central(is).

Exemplo

x_i	f_i	F_i
230 --- 250	12	12
250 --- 270	9	21
270 --- 290	8	29
290 --- 310	7	36
310 --- 330	6	42
330 --- 350	5	47
350 --- 370	3	50
Σ	50	—

Total de dados
 $n = 50$
(par)

Metade dos dados
 $n/2 = 25$

Portanto, a classe mediana é a terceira. Assim $i = 3$. A mediana será obtida através da seguinte expressão:

A Moda

Neste caso é preciso inicialmente apontar a classe modal, i. é, a de maior frequência. Neste exemplo é a primeira com $f_i = 12$. Assim $i = 1$.

$$m_e = l_i + h_i \left[\frac{\frac{n}{2} - F_{i-1}}{f_i} \right] = 270 + 20 \left[\frac{\frac{50}{2} - 21}{8} \right] =$$

$$= 270 + 20 \left[\frac{50 - 21}{8} \right] = 270 + 20 \frac{4}{8} = 280 \text{ meses}$$

Exemplo

i	x_i	f_i
1	230 --- 250	12
2	250 --- 270	9
3	270 --- 290	8
4	290 --- 310	7
5	310 --- 330	6
6	330 --- 350	5
7	350 --- 370	3
—	Σ	50

Classe modal, pois $f_i = 12$.

Portanto a moda poderá ser obtida através de uma das seguintes expressões:

Crítério de King:

$$m_o = l_i + h_i \left[\frac{f_{i+1}}{f_{i-1} + f_{i+1}} \right] = 230 + 20 \left[\frac{9}{0 + 9} \right] =$$

$$= 230 + 20 \left[\frac{9}{9} \right] = 250 \text{ meses}$$

Crítério de Czuber:

$$m_o = l_i + h_i \left[\frac{f_i - f_{i-1}}{2 \cdot f_i - (f_{i-1} + f_{i+1})} \right] =$$

$$= 230 + 20 \left[\frac{12 - 0}{2 \cdot 12 - (0 + 9)} \right] =$$

$$= 230 + 20 \left[\frac{12}{24 - 9} \right] =$$

$$= 230 + 16 = 246 \text{ meses}$$

Medidas de dispersão ou variabilidade

A Amplitude

$$h = x_{\text{máx}} - x_{\text{mín}}$$

$$h = 370 - 230 = 140 \text{ meses}$$

O Desvio Médio Absoluto

Neste caso, o dma será dado por:

$$\begin{aligned} \text{dma} &= \frac{f_1|x_1 - \bar{x}| + f_2|x_2 - \bar{x}| + \dots + f_k|x_k - \bar{x}|}{f_1 + f_2 + \dots + f_k} = \\ &= \frac{\sum f_i \cdot |x_i - \bar{x}|}{n} \end{aligned}$$

Exemplo

x_i	f_i	$f_i x_i - \bar{x} $
240	12	$12 \cdot 240 - 285,20 = 542,40$
260	9	$9 \cdot 260 - 285,20 = 226,80$
280	8	$8 \cdot 280 - 285,20 = 41,60$
300	7	$7 \cdot 300 - 285,20 = 103,60$
320	6	$6 \cdot 320 - 285,20 = 208,80$
340	5	$5 \cdot 340 - 285,20 = 274,00$
360	3	$3 \cdot 360 - 285,20 = 224,40$
Σ	50	1621,60

O dma será, então:

$$\begin{aligned} \text{dma} &= \frac{\sum f_i \cdot |x_i - \bar{x}|}{n} = \frac{1621,60}{50} = \\ &= 32,43 \text{ meses} \end{aligned}$$

A Variância

Neste caso, a variância será:

$$\begin{aligned} s^2 &= \frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \dots + f_k(x_k - \bar{x})^2}{n} = \\ &= \frac{\sum f_i (x_i - \bar{x})^2}{n} = \frac{\sum f_i x_i^2}{n} - \bar{x}^2 \end{aligned}$$

Exemplo

x_i	f_i	$f_i \cdot x_i^2$
240	12	$12 \cdot 240^2 = 691200$
260	9	$9 \cdot 260^2 = 608400$
280	8	$8 \cdot 280^2 = 627200$
300	7	$7 \cdot 300^2 = 630000$
320	6	$6 \cdot 320^2 = 614400$
340	5	$5 \cdot 340^2 = 578000$
360	3	$3 \cdot 360^2 = 388800$
Σ	50	4 138 000



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



A variância será, então:

$$\begin{aligned} s^2 &= \frac{\sum f_i x_i^2}{n} - \bar{x}^2 = \\ &= \frac{4138000}{50} - 285,20^2 = \\ &= 1420,96 \text{ meses}^2 \end{aligned}$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O Desvio Padrão

O desvio padrão será dado por:

$$\begin{aligned} s &= \sqrt{\frac{\sum f_i x_i^2}{n} - \bar{x}^2} = \sqrt{1420,96} = \\ &= 37,6956 \cong 37,70 \text{ meses} \end{aligned}$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



O Coeficiente de Variação

Dividindo o desvio padrão pela média, tem-se o coeficiente de variação:

$$g = \frac{37,695623}{285,20} = 13,22\%$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Medidas de Assimetria (Distorção)

Skewness



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Primeiro Coeficiente (de Pearson)

$$a_1 = (\text{Média} - \text{Moda}) / \text{Desvio Padrão}$$

Segundo Coeficiente (de Pearson)

$$a_2 = 3 \cdot (\text{Média} - \text{Mediana}) / \text{Desvio Padrão}$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



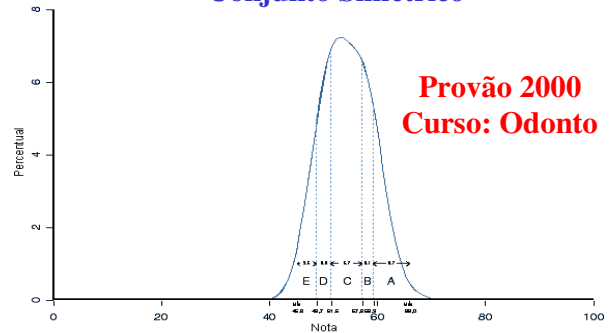
Coeficiente Quartílico

$$CQA = [(Q_3 - Q_2) - (Q_2 - Q_1)] / (Q_3 - Q_1)$$

Coeficiente do Momento

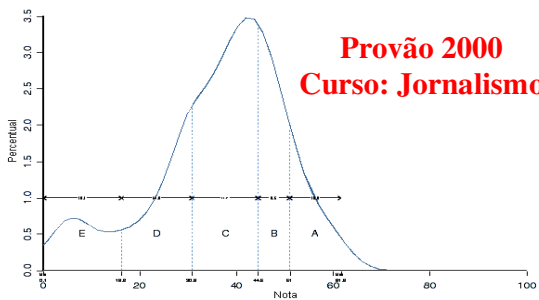
$$a_3 = m_3/s^3, \text{ onde } m_3 = \Sigma(X - \bar{X})^3/n$$

Coeficiente = 0 Conjunto Simétrico



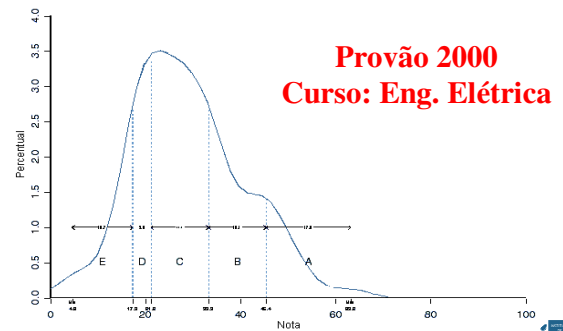
Coeficiente < 0

Conjunto: Negativamente Assimétrico



Coeficiente > 0

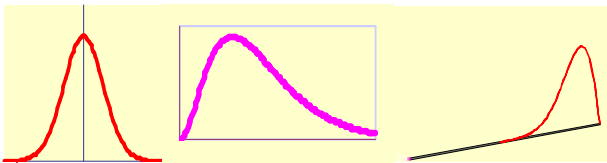
Conjunto: Positivamente Assimétrico



Coeficiente = 0 (Simétrica)

Coeficiente > 0 (Assimetria positiva)

Coeficiente < 0 (Assimetria negativa)



Medidas de Achatamento ou Curtose (Kurtosis)

Coeficiente de Curtose (momentos)

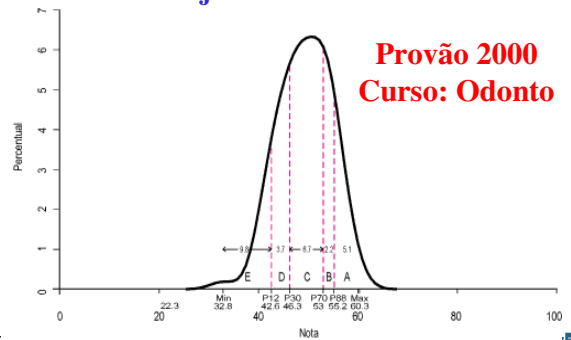
$$a_4 = m_4/s^4, \text{ onde } m_4 = \Sigma(X - \bar{X})^4/n$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



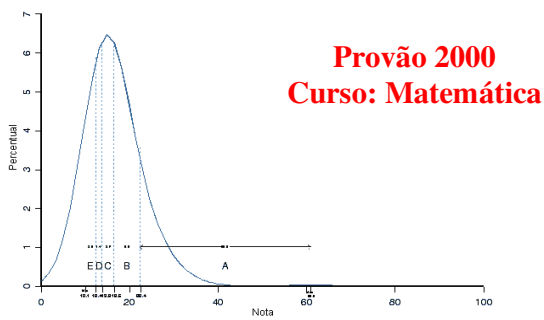
Coeficiente = 3 ou 0
Conjunto: Mesocúrtico



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



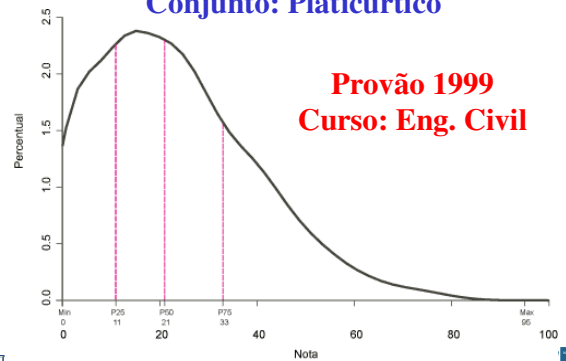
Coeficiente > 3 ou (> 0)
Conjunto: Leptocúrtico



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Coeficiente < 3 ou (< 0)
Conjunto: Platicúrtico



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Propriedades das Medidas

Se $y = ax + b$

Então:

$$\bar{y} = a\bar{x} + b$$

$$s_y^2 = a^2 s_x^2$$

$$s_y = |a| s_x$$



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Teorema de Chebyshev

O teorema de Chebyshev permite verificar qual é o percentual mínimo de valores de um conjunto de dados que deve estar um “certo número” de desvios em torno da média.



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Em qualquer conjunto de dados com desvio padrão “s”, pelo menos $(1 - 1/k^2)$ dos valores do conjunto devem estar entre “k” desvios em torno da média, onde “k” é um valor tal que $k > 1$.



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Exemplos:

Assim pelo menos:

75% dos valores estão dentro de **k = 2** desvios a partir da média;

89% dos valores estão dentro de **k = 3** desvios a contar da média;

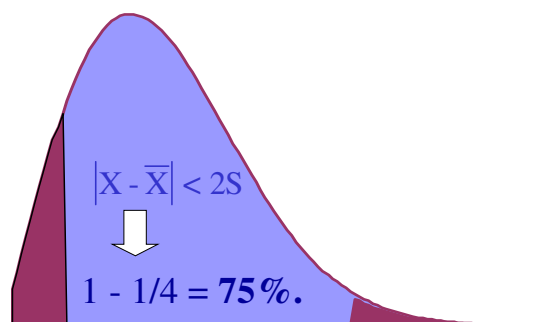
94% dos valores estão dentro de **k = 4** desvios a contar da média.



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística



Graficamente:



Prof. Lori Viali, Dr. - UFRGS - Instituto de Matemática - Departamento de Estatística

