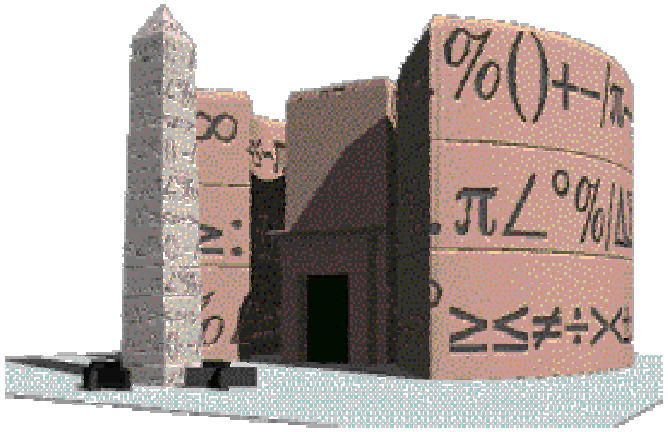


Estatística Descritiva 2

Prof. Lorí Viali, Dr.

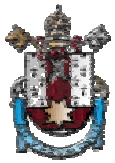
viali@mat.pucrs.br

http://www.pucrs.br/~viali/

- **Organização;**
- **Resumo;**
- **Apresentação.**

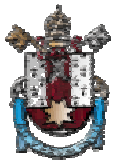
Amostra ou População

De Grande Conjuntos de Dados



Dados não organizados

Dados Brutos
Variável qualitativa

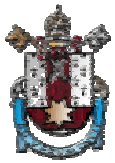


Defeitos em uma linha de produção

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

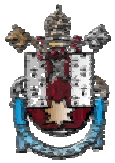
Dados organizados em uma distribuição de freqüências

* 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

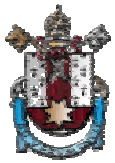
SIMPLES

- Absoluta
- Relativa
  - Decimal
  - Percentual

ACUMULADAS

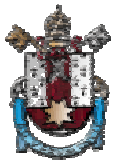
- Absoluta
- Relativa
  - Decimal
  - Percentual

Apresentação

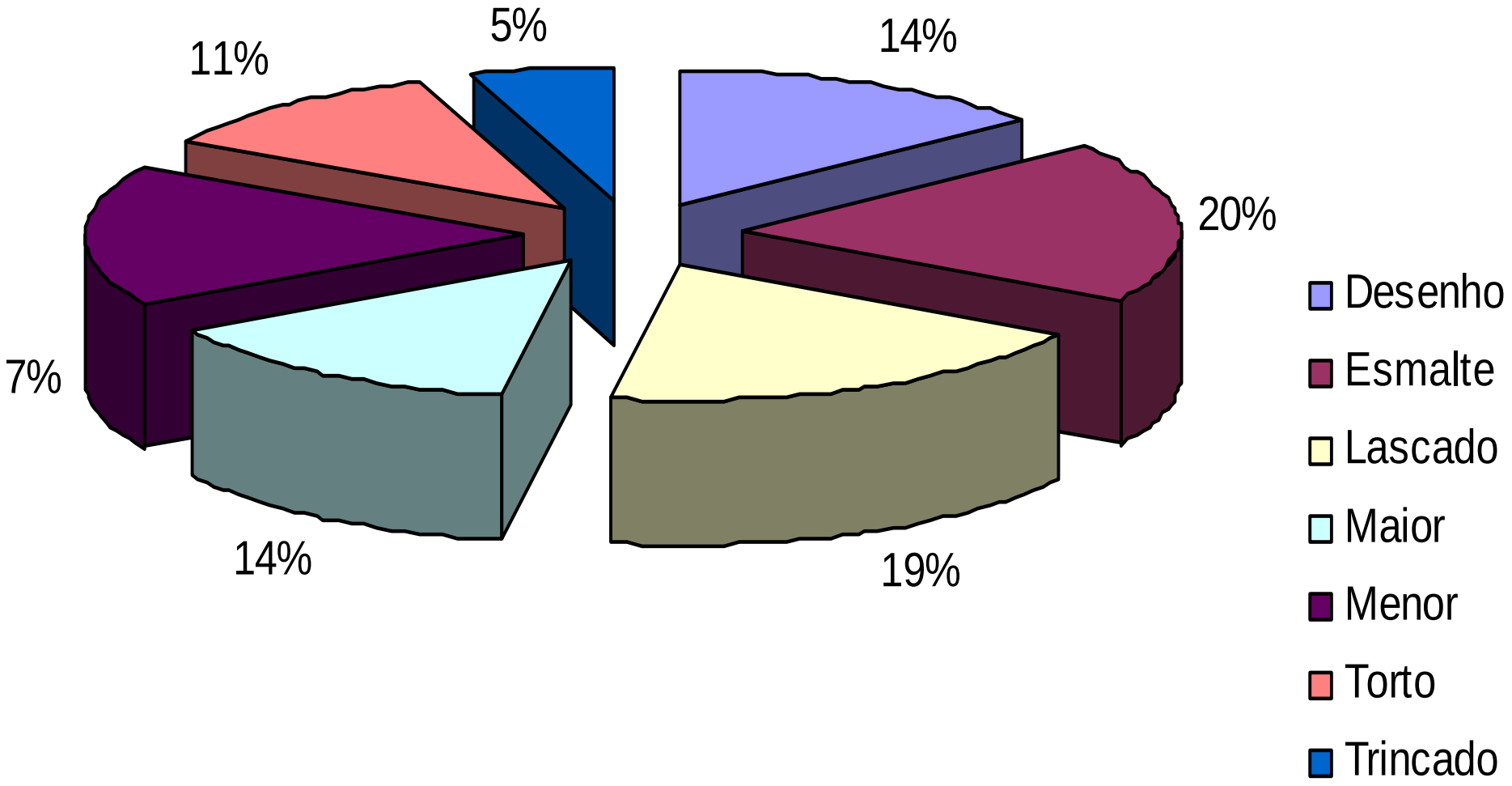


Freqüências - Representação

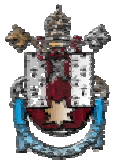
Valores	f_i	F_i	fr_i	fr_i	Fr_i
0	60	60	0,30	30	30
1	50	110	0,25	25	55
2	40	150	0,20	20	75
3	30	180	0,15	15	90
4	10	190	0,05	5	95
5	6	196	0,03	3	98
6	4	200	0,02	2	100
TOTAL	**200**	**—**	**1,00**	**100**	**—**



Representação gráfica
Diagrama de torta
ou pizza (Pie Chart)

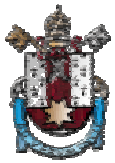

5%
14%
11%
20%
7%
14%
19%
Desenho
Esmalte
Lascado
Maior
Menor
Torto
Trincado

Dados Brutos
Variável discreta

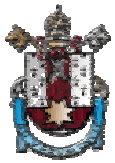


Número de irmãos dos alunos da turma 450 - Estatística - PUCRS - 2002/02

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

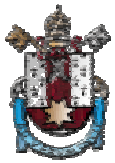


Distribuição de freqüências por ponto ou valores



Distribuição de freqüências **por ponto ou valores** da variável: "Número de irmãos dos alunos da turma 450" da disciplina: Probabilidade e Estatística PUCRS - 2002/01.

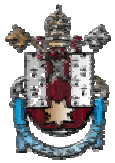
Nº de irmãos	Nº de alunos
0	7
1	21
2	8
3	5
4	4
5	3
6	2
∑	**50**

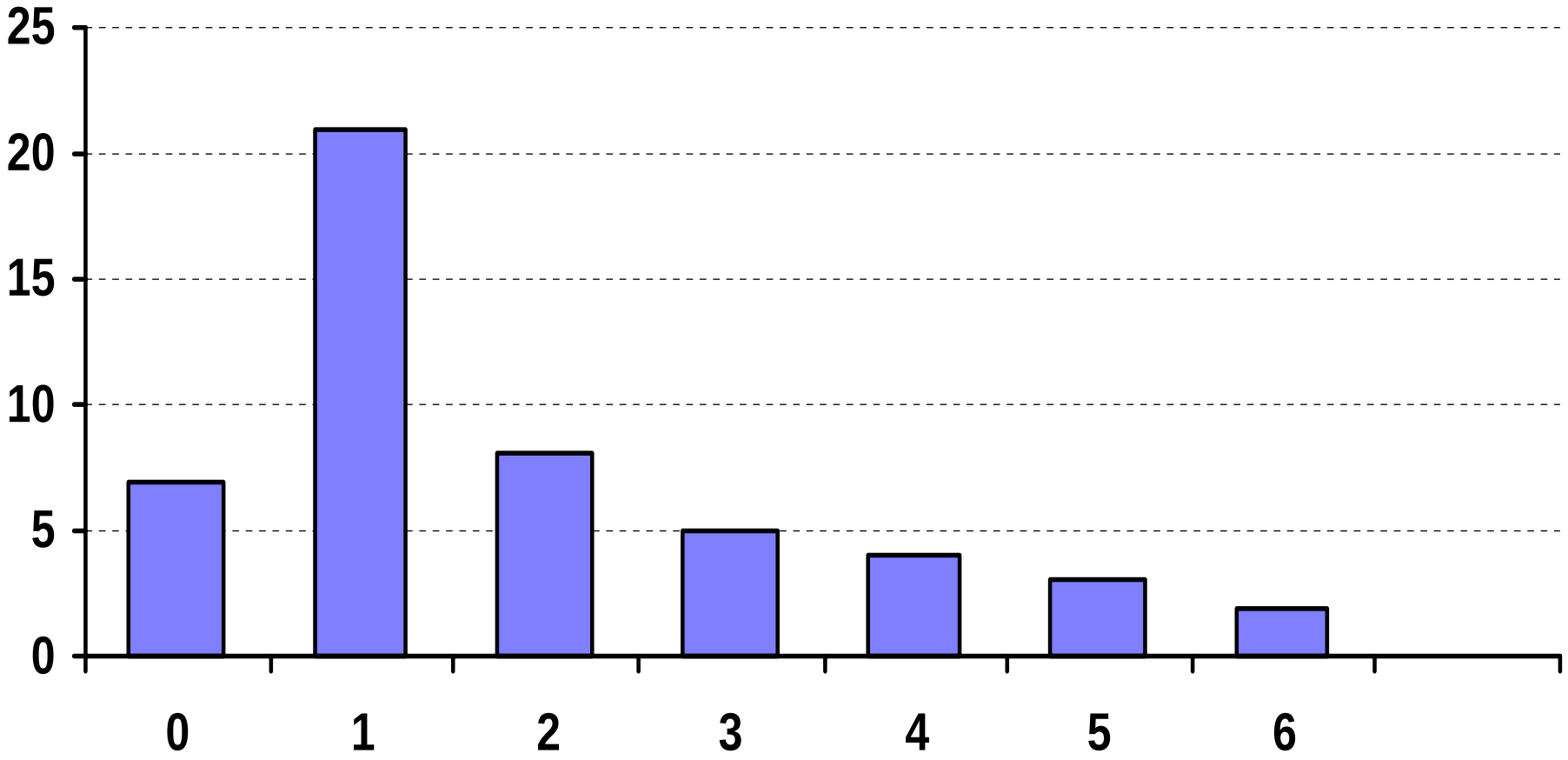


Representação gráfica

* Diagrama de colunas simples *

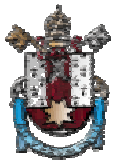
Diagrama de colunas simples da variável: Número de irmãos dos alunos da turma 450 Disciplina: Estatística, PUCRS - 2002/02




25
20
15
10
5
0
0
1
2
3
4
5
6

Resumo de uma Distribuição de freqüências por ponto ou valores

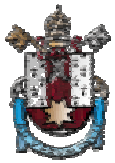
Medidas de tendência ou posição central



(A) A média Aritmética

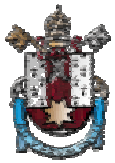
Neste caso, a média a dada por:

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



EXEMPLO

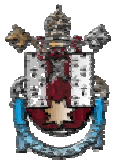
x_i	f_i	$f_i x_i$
0	7	0
1	21	21
2	8	16
3	5	15
4	4	16
5	3	15
6	2	12
$\sum$	**50**	**95**



EXEMPLO

A média será, então:

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

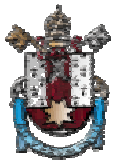


(B) A Mediana

Como n = 50 é par, tem-se:

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

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

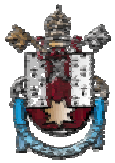


EXEMPLO

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
$\sum$	**50**	—

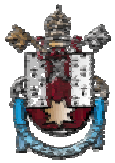
Total de dados n = 50 (par)

Metade dos dados n/2 = 25



(C) A Moda

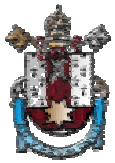
m_o = valor(es) que mais se repete(m)



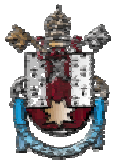
EXEMPLO

x_i	f_i
0	7
1	21
2	8
3	5
4	4
5	3
6	2
$\sum$	50

Pois ele se repete mais vezes



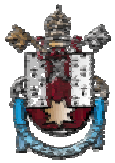
Medidas de dispersão ou variabilidade



(A) A Amplitude (h)

$$h = x_{máx} - x_{mín}$$

$$h = 6 - 0 = 6 \text{ irmãos}$$

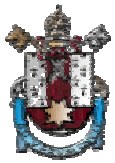


(B) O dma

Neste caso, o dma será dado por:

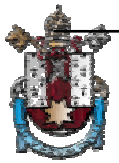
$$dma = \frac{f_1|x_1 - \bar{x}| + f_2|x_2 - \bar{x}| + \ldots + f_k|x_k - \bar{x}|}{f_1 + f_2 + \ldots + f_k} =$$

$$= \frac{\sum f_{i\cdot}|x_i - \bar{x}|}{n}$$



EXEMPLO

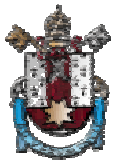
x_i	f_i	$f_i\|x_i - \overline{X}\|$
0	7	$7.\|0 - 1{,}90\| = 13{,}30$
1	21	$21.\|1 - 1{,}90\| = 18{,}90$
2	8	$8.\|2 - 1{,}90\| = 0{,}80$
3	5	$5.\|3 - 1{,}90\| = 5{,}50$
4	4	$4.\|4 - 1{,}90\| = 8{,}40$
5	3	$3.\|5 - 1{,}90\| = 9{,}30$
6	2	$2.\|6 - 1{,}90\| = 8{,}20$
$\sum$	**50**	**64,40**



EXEMPLO

O dma será, então:

$$\text{dma} = \frac{\sum f_{i\bullet} \left| x_i - \overline{x} \right|}{n} = \frac{64{,}40}{50} = 1{,}29 \text{ irmãos}$$

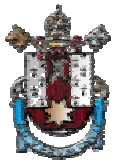


(C) A Variância (s^2)

Neste caso, a variância será:

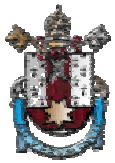
$$s^2 = \frac{f_1(x_1-\bar{x})^2 + f_2(x_2-\bar{x})^2 + ... + f_k(x_k-\bar{x})^2}{n} =$$

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



EXEMPLO

x_i	f_i	$f_i x_i^2$
0	7	$0^2.7 = 0$
1	21	$1^2.21 = 21$
2	8	$2^2.8 = 32$
3	5	$3^2.5 = 45$
4	4	$4^2.4 = 64$
5	3	$5^2.3 = 75$
6	2	$6^2.2 = 72$
$\sum$	**50**	**299**

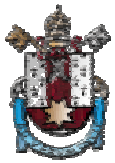


EXEMPLO

A variância será, então:

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

$$= 2{,}3700 \quad \text{irmãos}^2$$

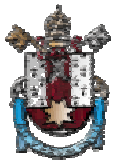


(D) O Desvio Padrão (s)

O desvio padrão será dado por:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \overline{x}^2} = \sqrt{2,3700} =$$

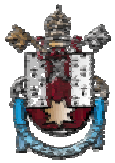
$$= 1,5395 \cong 1,54 \text{ irmãos}$$



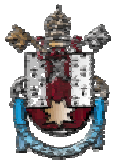
(E) O Coeficiente de Variação (g)

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

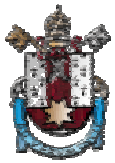
$$g = \frac{1,539480}{1,90} = 81,03\ \%$$



Dados Brutos
Variável contínua



Idade (em meses) dos alunos da turma 450 da disciplina: Probabilidade e Estatística PUCRS - 2002/02



276 245 345 240 270 310 368

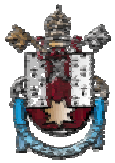
334 268 288 336 299 236 239 355 330

287 344 300 244 303 248 251 265 246

240 320 308 299 312 324 289 320 264

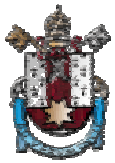
252 298 315 255 274 264 263 230 303

369 247 266 275 281 230 234

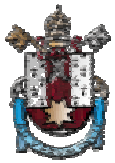


Distribuição de freqüências por classes ou intervalos

Distribuição por classes ou intervalos da variável “idade dos alunos da turma 450” da disciplina: Probabilidade e Estatística da PUCRS - 2002/02



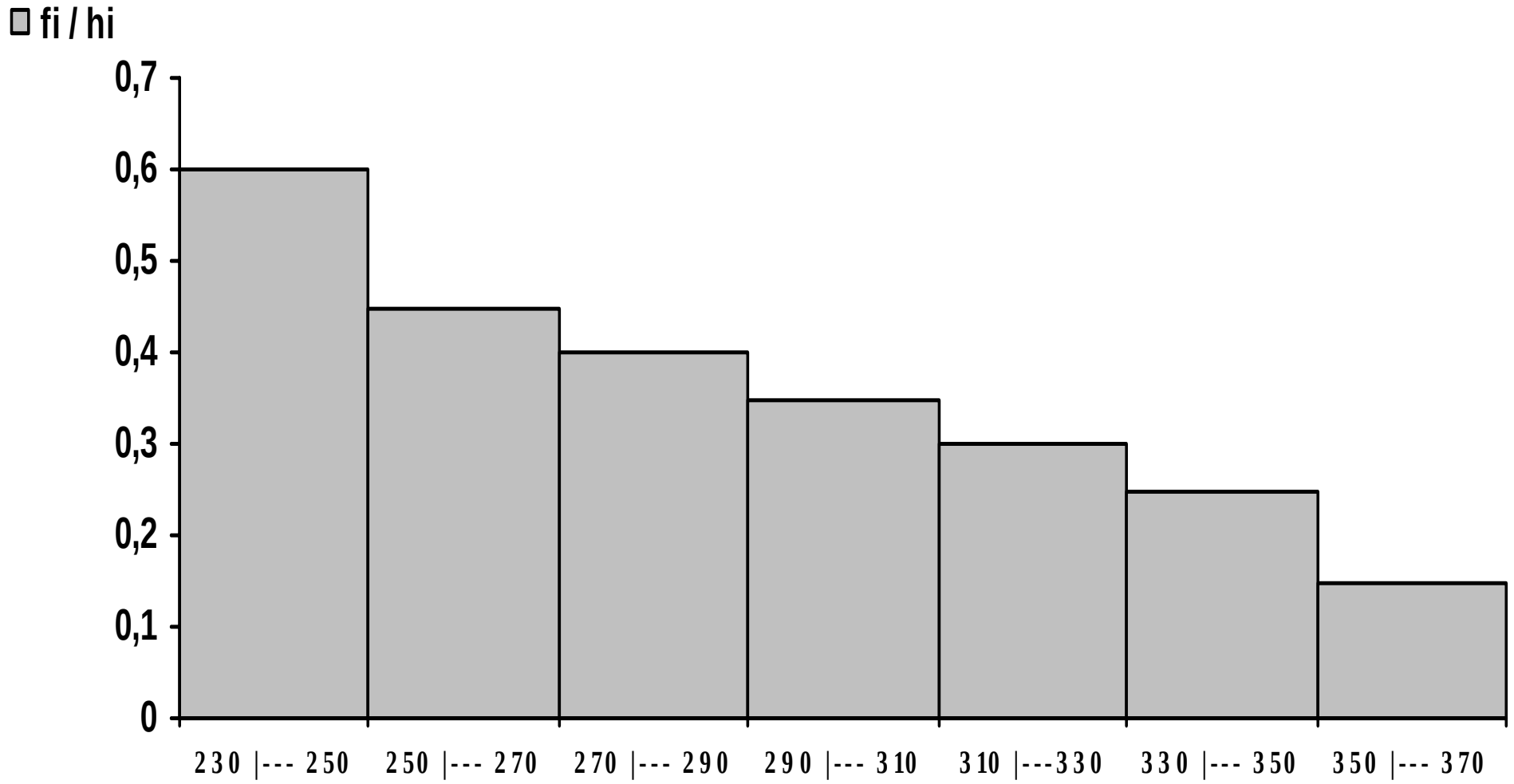
Idades	Número de alunos
230 |--- 250	12
250 |--- 270	9
270 |--- 290	8
290 |--- 310	7
310 |--- 330	6
330 |--- 350	5
350 |--- 370	3
Total	**50**



Representação gráfica

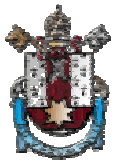
* Histograma *

Histograma de freqüências da variável "Idade dos alunos da turma 450" de Probabilidade e Estatística da PUCRS - 2002/02


fi / hi
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
230 |--- 250
250 |--- 270
270 |--- 290
290 |--- 310
310 |---330
330 |--- 350
350 |--- 370

Medidas

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



Simbologia

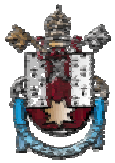
x_i = ponto médio da classe;

f_i = freqüência simples da classe;

li_i = limite inferior da classe;

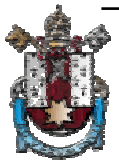
ls_i = limite superior da classe;

h_i = amplitude da classe.

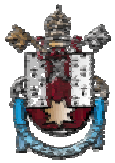


PONTO MÉDIO DA CLASSE

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
$\sum$	**50**	—

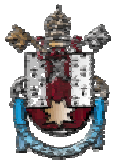


Medidas de tendência ou posição central



MÉDIA DA DISTRIBUIÇÃO

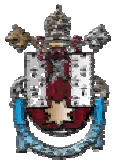
x_i	f_i	$f_i . x_i$
240	12	2880
260	9	2340
280	8	2240
300	7	2100
320	6	1920
340	5	1700
360	3	1080
$\sum$	**50**	**14260**



EXEMPLO

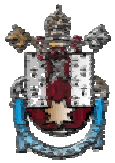
A média será:

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



(B) A Mediana

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

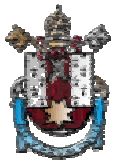


EXEMPLO

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
	$\sum$		**50**	—

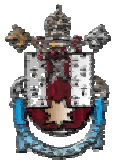
Total de dados n = 50 (par)

Metade dos dados n/2 = 25



EXEMPLO

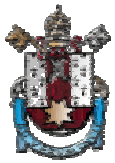
Portanto, a classe mediana é a terceira. Assim i = 3. A mediana será obtida através da seguinte expressão:



EXEMPLO

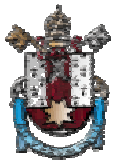
$$m_e = li_i + h_i \left[\frac{\frac{n}{2} - F_{i-1}}{f_i} \right] = 270 + 20 \left[\frac{\frac{50}{2} - 21}{8} \right] =$$

$$= 270 + 20 \left[\frac{\frac{50}{2} - 21}{8} \right] = 270 + 20 \frac{4}{8} = 280 \text{ meses}$$



(C) A Moda

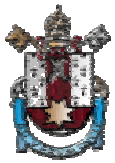
Neste caso é preciso inicialmente apontar a classe modal, i. é, a de maior freqüência. Neste exemplo é a primeira com f_i = 12. Assim i = 1.



EXEMPLO

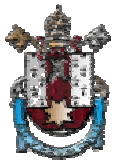
i	x_i			f_i
1	230	---	250	12
2	250	---	270	9
3	270	---	290	8
4	290	---	310	7
5	310	---	330	6
6	330	---	350	5
7	350	---	370	3
—		$\sum$		**50**

Classe modal, pois $f_i = 12$.



EXEMPLO

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

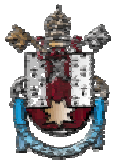


EXEMPLO

Critério de King:

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

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



EXEMPLO

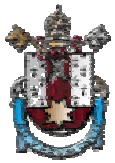
Critério de Czuber:

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

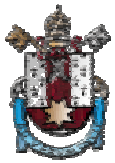
$$= 230 + 20 . \left[\frac{12 - 0}{2.12 - (0 + 9)} \right] =$$

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

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



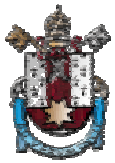
Medidas de dispersão ou variabilidade



(A) A Amplitude (h)

$$h = x_{máx} - x_{mín}$$

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

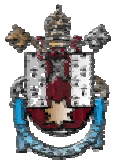


(B) O dma

Neste caso, o dma será dado por:

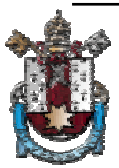
$$dma = \frac{f_1|x_1 - \bar{x}| + f_2|x_2 - \bar{x}| + ... + f_k|x_k - \bar{x}|}{f_1 + f_2 + ... + f_k} =$$

$$= \frac{\sum f_{i\bullet}|x_i - \bar{x}|}{n}$$



EXEMPLO

x_i	f_i	$f_i.\lvert x_i - \overline{X} \rvert$
240	12	$12.\lvert 240 - 285{,}20 \rvert = 542{,}40$
260	9	$9.\lvert 260 - 285{,}20 \rvert = 226{,}80$
280	8	$8.\lvert 280 - 285{,}20 \rvert = 41{,}60$
300	7	$7.\lvert 300 - 285{,}20 \rvert = 103{,}60$
320	6	$6.\lvert 320 - 285{,}20 \rvert = 208{,}80$
340	5	$5.\lvert 340 - 285{,}20 \rvert = 274{,}00$
360	3	$3.\lvert 360 - 285{,}20 \rvert = 224{,}40$
$\sum$	**50**	**1621,60**

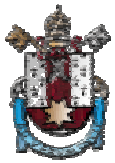


EXEMPLO

O dma será, então:

$$dma = \frac{\sum f_{i\bullet} \left| x_i - \overline{x} \right|}{n} = \frac{1621,60}{50} =$$

$$= 32,43 \text{ meses}$$

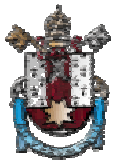


(C) A variância (s^2)

Neste caso, a variância será:

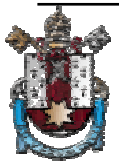
$$s^2 = \frac{f_1(x_1-\bar{x})^2 + f_2(x_2-\bar{x})^2 + ... + f_k(x_k-\bar{x})^2}{n} =$$

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



EXEMPLO

x_i	f_i	$f_i . x_i^2$
240	12	$12.240^2 = 691200$
260	9	$9.246^2 = 608400$
280	8	$8.280^2 = 627200$
300	7	$7.300^2 = 630000$
320	6	$6.320^2 = 614400$
340	5	$5.340^2 = 578000$
360	3	$3.360^2 = 388800$
$\sum$	**50**	**4 138 000**



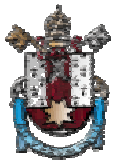
EXEMPLO

A variância será, então:

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

$$= \frac{4138000}{50} - 285,20^2 =$$

$$= 1420,96 \quad \text{meses}^2$$

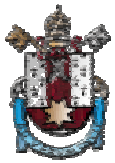


(D) O Desvio Padrão (s)

O desvio padrão será dado por:

$$s = \sqrt{\frac{\sum f_i x_i^2}{n} - \overline{x}^2} = \sqrt{1420\ ,96} =$$

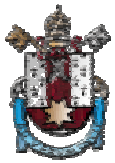
$$= 37,6956 \cong 37,70 \text{ meses}$$



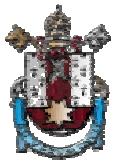
(E) O Coeficiente de Variação (g)

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

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



Medidas de Assimetria (Distorção)

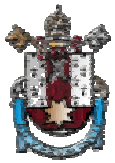


1. Primeiro coeficiente de Pearson;

2. Segundo coeficiente de Pearson;

3. Coeficiente quartílico;

4. Coeficiente do momento.

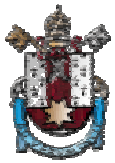


Primeiro coeficiente de Pearson

a_1 = (média - moda) / desvio padrão

Segundo coeficiente de Pearson

a_2 = 3(média - mediana) /desvio padrão

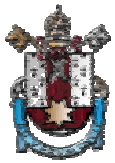


Coeficiente quartílico

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

Coeficiente do momento

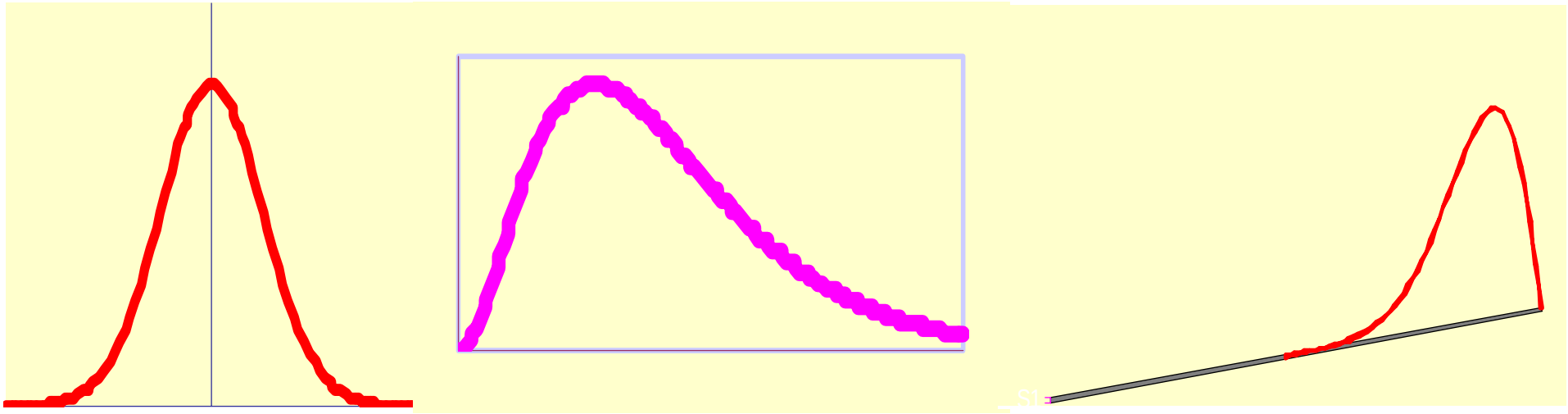
$$a_3 = m_3/s^3, \text{ onde } m_3 = \Sigma(X - \overline{X})^3/n$$



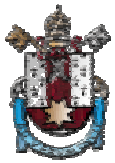
Coeficiente = 0 (Simétrica)

Coeficiente > 0 (Assimetria positiva)

Coeficiente < 0 (Assimetria negativa)



Medidas de Achatamento ou Curtose



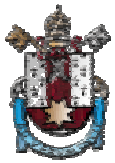
Coeficiente de curtose [*kurtosis*]

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

$a_4 = 3$ (Mesocúrtica)

$a_4 > 3$ (Leptocúrtica)

$a_4 < 3$ (Platocúrtica)



Até a próxima!