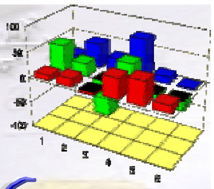


ECOP 16
Econometria Aplicada

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

Conceitos Básicos

Coleção de números = estatísticas

- ✓ O número de carros vendidos no país aumentou em 30%.
- ✓ A taxa de desemprego atinge, este mês, 7,5%.
- ✓ As ações da Telebrás subiram R$ 1,5, hoje.
- ✓ Resultados do Carnaval no trânsito: 145 mortos, 2430 feridos.

Estatística: uma definição

A ciência de coletar, organizar, apresentar, analisar e interpretar dados numéricos com o objetivo de tomar melhores decisões.

Estatística (divisão)

Descritiva — Os procedimentos usados para organizar, resumir e apresentar dados numéricos.

Indutiva — A coleção de métodos e técnicas utilizados para estudar uma população baseado em amostras probabilísticas desta população.

POPULAÇÃO

Uma coleção de todos os possíveis elementos, objetos ou medidas de interesse.

CENSO

Um levantamento efetuado sobre toda uma população é denominado de levantamento censitário ou simplesmente censo.

AMOSTRA

Uma porção ou parte de uma população de interesse.

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

AMOSTRAGEM

O processo de escolha de uma amostra da população é denominado de amostragem.

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

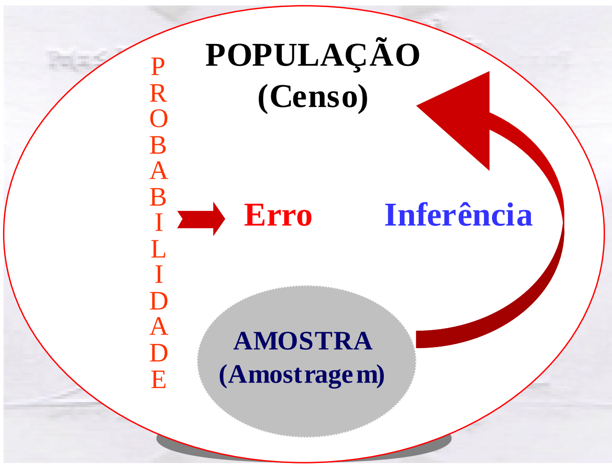
PROBABILIDADE (Matemática)

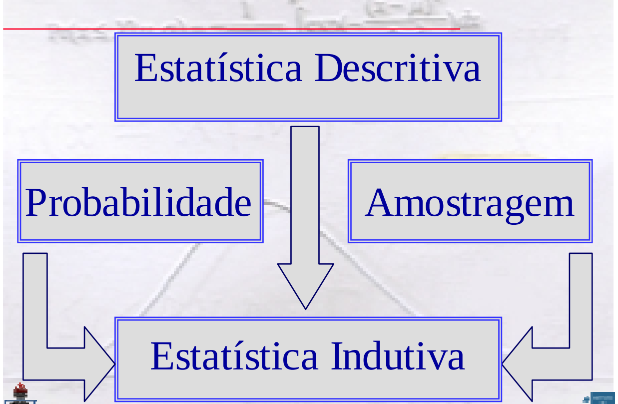
ESTATÍSTICA (Matemática Aplicada)

Univariada — *Trabalha com uma única característica dos dados*

Multivariada — *Trabalha com duas ou mais características dos dados*

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





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

Estatística x Probabilidade

Faces	Probabilidades
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
Total	**1**

Faces	Freqüências
1	15
2	18
3	23
4	25
5	22
6	17
Total	**120**

Arredondamento

Todo arredondamento é um erro.

O erro deve ser evitado ou então minimizado.

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

Arredondamento

Regra básica:

Arredondar sempre para o mais próximo.

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

Exemplos:

1,456 → 1,46

1,454 → 1,45

1,475 — É ímpar → Aumenta → 1,48

1,485 — É par → Não aumenta → 1,48

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

VARIÁVEIS

- **QUALITATIVAS**
  - NOMINAL
  - ORDINAL
- **QUANTITATIVAS**
  - DISCRETA
  - CONTÍNUA

Variável Qualitativa

NOMINAL: Sexo, Religião, Estado civil, Curso

ORDINAL: Conceito, Grau de Instrução, Mês, Dia da semana

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

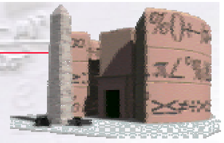
Variável Qualitativa

DISCRETA

- Número de faltas
- Número de irmãos
- Número de acertos

CONTÍNUA

- Altura
- Área
- Peso
- Volume



Estatística Descritiva 1

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

ESTATÍSTICA DESCRITIVA

- ☹ **Organização;**
- 😐 **Resumo;**
- ☺ **Apresentação.**

Conjunto de dados:

- **Amostra**

ou

- **População**

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

Um conjunto de dados é resumido de acordo com as seguintes características:

Amostra ou População

- **Tendência ou posição central**
- **Dispersão ou variabilidade**
- **Assimetria (distorção)**
- **Achatamento ou curtose**

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

Tendência ou Posição Central

(a) As médias — **Simples**

- **Aritmética**
- **Geométrica**
- **Harmônica**
- **Quadrática**
- **Interna**

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

As médias

Aritmética

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n}\sum x_i = \frac{\sum x_i}{n}$$

Geométrica

$$m_g = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\prod x_i}$$

Harmônica

$$m_h = \frac{1}{\dfrac{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}{n}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}}$$

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

A média Quadrática e Interna

$$m_q = \frac{x_1^2 + x_2^2 + \dots x_n^2}{n} = \frac{\sum x_i^2}{n}$$

Quadrática

É a média aritmética só que aplicada sobre o conjunto onde uma parte dos dados (extremos) é descartada.

Interna

Exemplo

Conjuntos		Médias $\bar{x}$	m_g	m_h
4	6	5	4,9	4,8
1	9	5	3	1,8

Relação entre as médias

Dado um conjunto de dados qualquer, as médias aritmética, geométrica e harmônica mantém a seguinte relação:

$$\bar{x} \geq m_g \geq m_h$$

Tendência ou Posição Central

(a) As médias — Ponderadas:

- Aritmética
- Geométrica
- Harmônica
- Quadrática

As médias Ponderadas

- Aritmética

$$m_{ap} = \frac{x_1 \cdot w_1 + x_2 \cdot w_2 + \dots + x_k \cdot w_k}{w_1 + w_2 + \dots + w_k} = \frac{\sum x_i \cdot w_i}{\sum w_i}$$

As médias Ponderadas

- Geométrica

$$m_{gp} = \sqrt[\sum w_i]{x_1^{w_1} . x_2^{w_2} . \dots x_k^{w_k}} = \sqrt[\sum w_i]{\prod x_i^{w_i}}$$

As médias Ponderadas

- **Harmônica**

$$m_{hP} = \frac{w_1 + w_2 + w_k}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \ldots + \frac{w_k}{x_k}} = $$

$$= \frac{\sum w_i}{\sum \frac{w_i}{x_i}}$$

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

(b) A mediana (*median*)

É o valor que separa o conjunto em dois subconjuntos do mesmo tamanho.

$m_e = [x_{(n/2)} + x_{(n/2)+1}]/2$ se "n" é par

$m_e = x_{(n+1)/2}$ se "n" é ímpar

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

(c) Separatrizes

A idéia de repartir o conjunto de dados pode ser levada adiante. Se ele for repartido em 4 partes tem-se os **QUARTIS**, se em 10 os **DECIS** e se em 100 os **PERCENTIS**.

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

(c) A moda (*mode*)

É o(s) valor(es) do conjunto que mais se repete(m).

Um conjunto pode ser **unimodal, multimodal ou amodal**.

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

Dispersão ou Variabilidade

(a) A amplitude (h)

(b) O Desvio Médio (dma)

(c) A Variância (s^2)

(d) O Desvio Padrão (s)

(e) A Variância Relativa (g^2)

(f) O Coeficiente de Variação (s)

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

(a) Amplitude (*range*)

$h = x_{máx} - x_{mín}$

Considere o conjunto:

-2 -1 0 3 5

h = 5 – (-2) = **7**

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

(b) DMA (*average deviation*)

$$dma = \frac{\sum |x_i - \bar{x}|}{n} =$$

$$= \frac{|-3| + |-2| + |-1| + |+2| + |+4|}{5} =$$

$$= \frac{12}{5} = 2{,}40$$

(b) A variância (*variance*)

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + ... + (x_n - \bar{x})^2}{n} =$$

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

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

O Desvio Padrão (*standard deviation*)

É a raiz quadrada da variância

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

Se extrairmos a raiz quadrada teremos do resultado anterior teremos o desvio padrão:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{6{,}80} = 2{,}61$$

A Variância Relativa

$$g^2 = s^2 / \bar{x}^2$$

O Coeficiente de Variação

$$g = s / \bar{x}$$

O coeficiente de variação do exemplo anterior, será:

$$g = \frac{s}{\bar{x}} = \frac{2{,}6077}{1} = 260{,}77\ \%$$

- Organização;
- Resumo e
- Apresentação.

Amostra
ou
População

De Grande Conjuntos de Dados

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)

Apresentação

Apresentação

FREQÜÊNCIAS

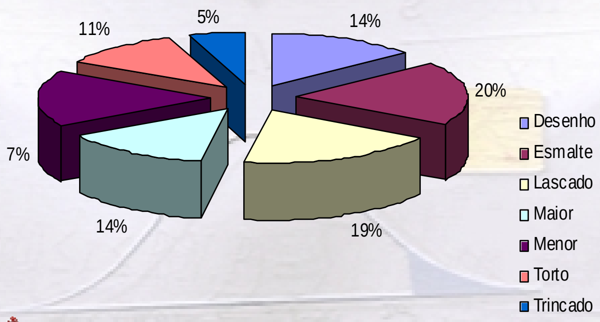
SIMPLES
- Absoluta
- Relativa
  - Decimal
  - Percentual

ACUMULADAS
- Absoluta
- Relativa
  - Decimal
  - Percentual

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)



Dados Brutos Variável discreta

Número de dependentes dos funcionários da Kapim S.A. – POA – 2004/01

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 dependentes dos funcionários da Kapim Ltda." Porto Alegre - 2004/01.

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

Dependentes	Funcionários
0	7
1	21
2	8
3	5
4	4
5	3
6	2
Σ	**50**

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

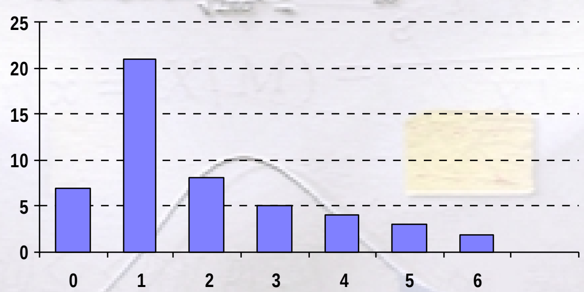
Representação gráfica

* Diagrama de colunas simples *

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

Diagrama de colunas simples da variável: "Número de dependentes dos funcionários da Kapim Ltda." Porto Alegre - 2004/01.

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



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

Resumo de uma Distribuição de freqüências por ponto ou valores

Medidas de tendência ou posição central

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
Σ	**50**	**95**

A média será, então:

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

Medidas de dispersão ou variabilidade

(A) 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	$6^2.2 = 72$
Σ	**50**	**299**

A variância será, então:

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

$$= 2{,}3700 \text{ dependente}^2$$

(B) O Desvio Padrão (s)

O desvio padrão será dado por:

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

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

(C) 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 G da disciplina: Probabilidade e Estatística UFRGS - 2004/01

		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		

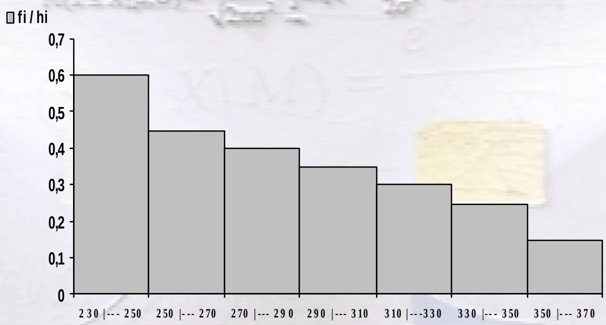
Distribuição de freqüências por classes ou intervalos

Distribuição por classes ou intervalos da variável "idade dos alunos da turma G" da disciplina: Probabilidade e Estatística da UFRGS - 2004/01

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 G" de Probabilidade e Estatística da UFRGS - 2004/01



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

Medidas

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

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

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

Simbologia

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

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.

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

PONTO MÉDIO DA CLASSE

x	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

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
Σ	50	14260

A média será:

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

Medidas de dispersão ou variabilidade

(A) A variância (s^2)

Neste caso, a variância será:

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

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

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	$3.360^2 = 388800$
Σ	50	4 138 000

A variância será, então:

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

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

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

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

(B) O Desvio Padrão (s)

O desvio padrão será dado por:

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

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

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

(C) O Coeficiente de Variação (g)

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

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

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

Medidas de Assimetria (Distorção)

Skewness

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

Primeiro Coeficiente (de Pearson)

a_1 = (Média - Moda) / Desvio Padrão

Segundo Coeficiente (de Pearson)

a_2 = 3.(Média - Mediana) / Desvio Padrão

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

Coeficiente Quartílico

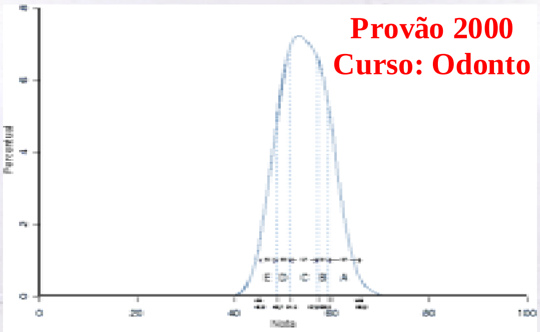
$CQA = [(Q_3 - Q_2) - (Q_2 - Q_1)]/(Q_3 - Q_1)$

Coeficiente do Momento

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

Coeficiente = 0

Conjunto Simétrico



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

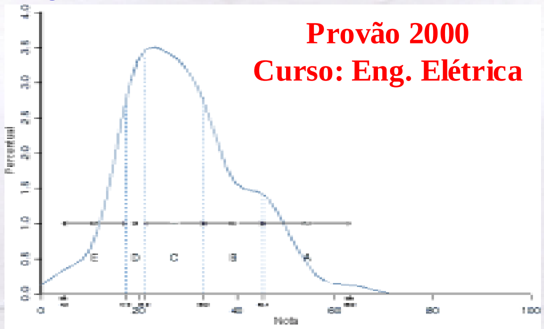
Coeficiente < 0

Conjunto: Negativamente Assimétrico

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

Coeficiente > 0

Conjunto: Positivamente Assimétrico



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

Medidas de Achatamento ou Curtose

(*Kurtosis*)

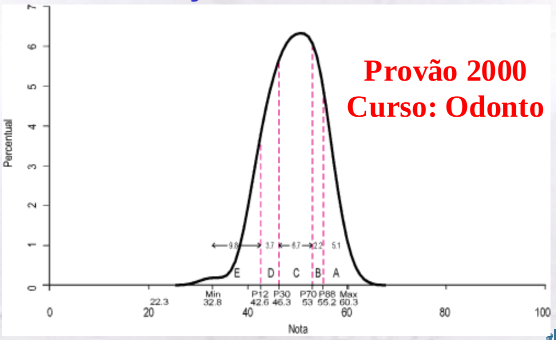
Prof. Lorí Viali, Dr. – UFRGS – Instituto de Matemática - Departamento de Estatística

Coeficiente de Curtose (momentos)

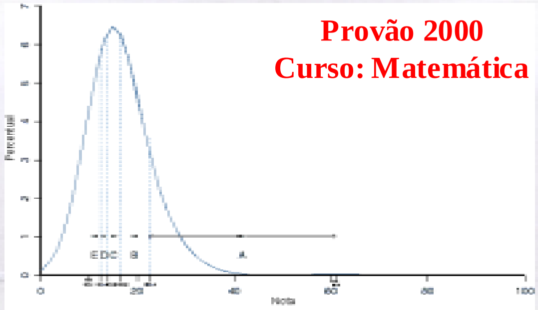
$\mathbf{a_4} = m_4/s^4$, onde $m_4 = \Sigma(X - \bar{X})^4/n$

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

Coeficiente = 3 ou 0
Conjunto: Mesocúrtico

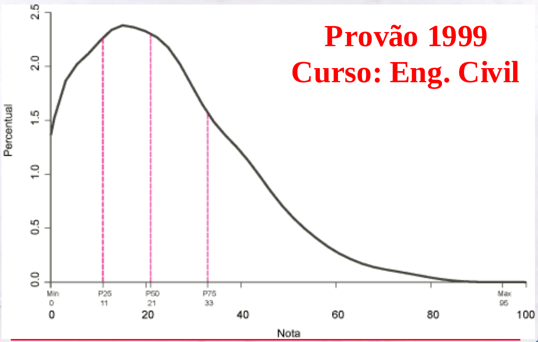


Coeficiente > 3 ou (> 0)
Conjunto: Leptocúrtico



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

Coeficiente < 3 ou (< 0)
Conjunto: Platicúrtico



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

Propriedades das Medidas

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

Seja "**x**" um conjunto de dados com média: $\bar{x}$ e desvio padrão s_x, e seja y = ax + b, então:

1. $\bar{y} = a.\bar{x} + b$
2. $s_y^2 = a^2.s_x^2$
3. $s_y = |a|.s_y$

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

Desigualdade de Chebyshev e Camp-Meidell

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

Seja "**x**" um conjunto de dados com média: $\bar{x}$ e desvio padrão s, então:

$$P(|x - \bar{x}| \geq k.s) \leq 1/k^2$$

ou

$$P(|x - \bar{x}| < k.s) \geq 1\text{-}1/k^2$$

Se "**x**" é um conjunto de dados simétrico e unimodal com média $\bar{x}$ e desvio padrão s, então:

$$P(|x - \bar{x}| \geq k.s) \leq 4/9k^2$$