Unidade II - Estatística descritiva

2.1. Apresentação de dados
- 2.1.1 Séries estatísticas
- 2.1.2 Tabelas
- 2.1.3 Gráficos

2.2. Distribuições de frequências e gráficos
- 2.2.1 Tabelas de classificação simples
- 2.2.2 Tabelas de classificação cruzada

2.3. Medidas descritivas
- 2.3.1 Medidas de localização ou tendência central
- 2.3.2 Medidas separatrizes
- 2.3.3 Medidas de variação ou dispersão
- 2.3.4 Medidas de formato

2.4. Análise exploratória de dados

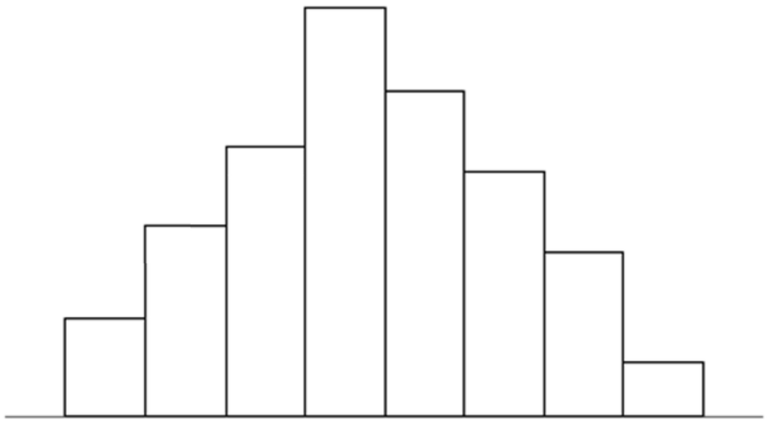
Distribuição de frequências e gráficos

- A distribuição de frequências é uma **forma de resumir a informação** sobre uma ou mais variáveis
- **Organiza um conjunto de dados em classes**, indicando a frequência de observações em cada classe
- Além de resumir a informação, tem por finalidade:
  1. Representar a **forma como se distribuem os valores** das variáveis (localização da maioria dos valores, simetria, número de picos e formato das caudas)
  2. Indicar qual **modelo de distribuição de probabilidade** poderia ser adequado para esses dados, pois fornece uma ideia empírica da distribuição da população
- Formato é muito **sensível ao número de observações** disponíveis

Dados

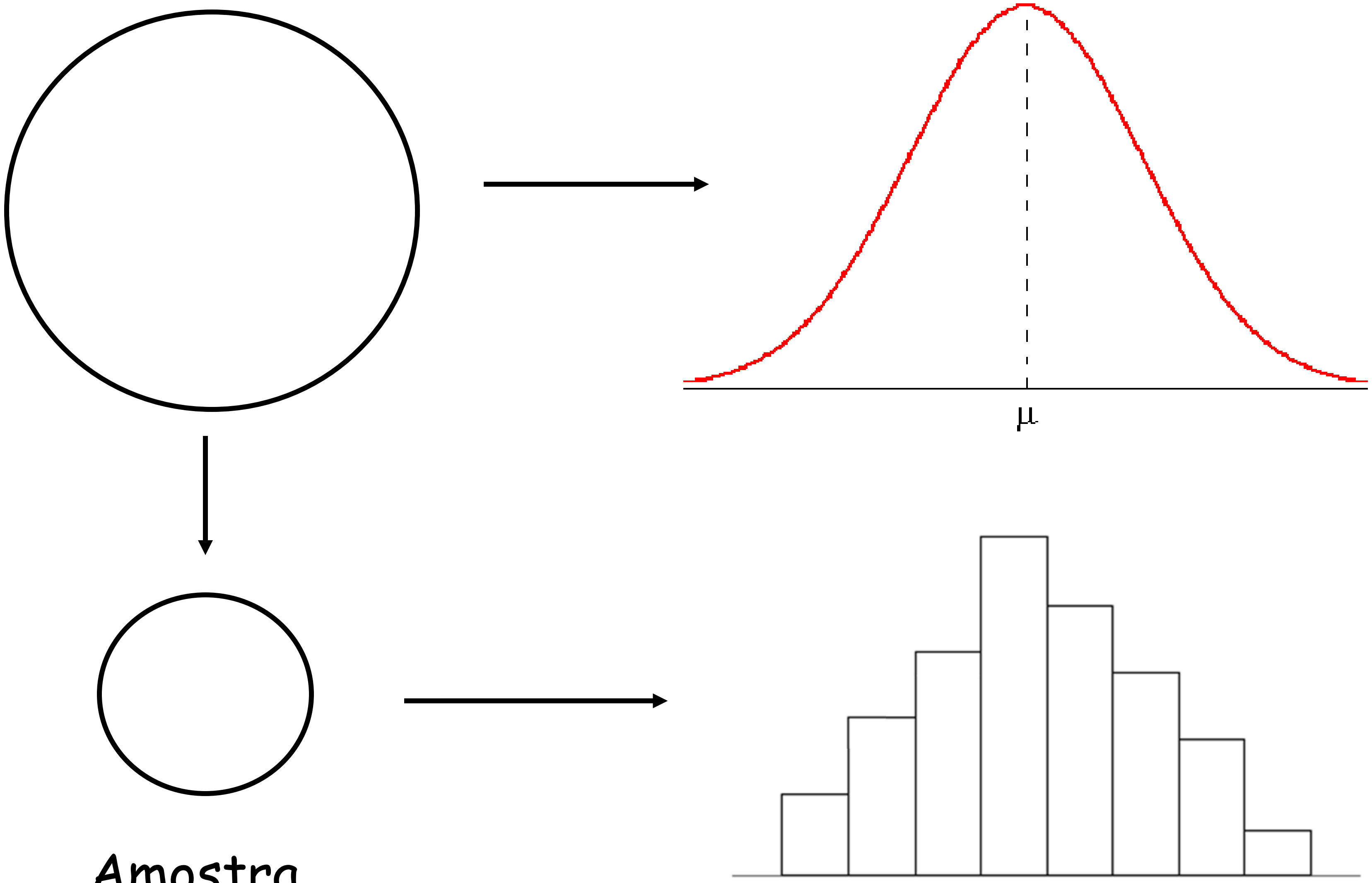
3,11	8,88	9,26	10,81	12,69	13,78	15,23	15,62	17,00	17,39
18,36	18,43	19,27	19,50	19,54	20,16	20,59	22,22	23,04	24,47
24,58	25,13	26,24	26,26	27,65	28,06	28,08	28,38	32,03	36,37
38,64	38,98	39,16	41,02	42,97	44,08	44,67	45,40	46,69	48,65
50,39	52,75	54,80	59,07	61,22	70,32	82,70	85,76	86,37	93,34

Gráfico da distribuição de frequências



População

Distribuição de probabilidade



Amostra

Distribuição de frequências

Tabelas de distribuição de frequências

- Tabelas de classificação simples

  → uma variável

- Tabelas de classificação dupla ou cruzada

  → duas variáveis

Tabelas de classificação simples

As características dessas tabelas variam de acordo com o tipo de variável em estudo.

- **Variáveis**
  - Categóricas
    - Qualitativas
      - Nominais
      - Ordinais
    - Quantitativas
  - Numéricas
    - Discretas
    - Contínuas

Tabelas de classificação simples

As características dessas tabelas variam de acordo com o tipo de variável em estudo.

- Se a variável é do tipo **categórica**, devemos obter as frequências para cada nível dessa variável.

- Se a variável é do tipo **numérica contínua**, devemos primeiro construir intervalos de mesma amplitude e depois obter as frequências para cada intervalo.

Distribuição de frequências para variáveis categóricas ou numéricas discretas

Exemplo 1: Variável categórica

Variável em estudo: **conceito na disciplina de Estatística**

Dados brutos: ruim, médio, bom, médio, ruim, médio, ruim, médio, ruim, bom, médio, médio, bom, médio, médio, médio, ótimo, médio, bom, ótimo, bom, ótimo, médio, ótimo, médio, ruim, médio, ótimo, médio, médio, bom, ruim, bom, bom, médio, ruim, médio, médio, ótimo, médio, bom, ruim, ruim, bom, médio, médio, ruim, bom, médio, médio, bom, bom, bom, médio, ruim, bom, médio, médio, ruim, médio

⇨ Quando a variável for categórica ou numérica discreta (com poucos valores), a tabela de distribuição de frequências apresentará a seguinte característica: cada valor da variável constituirá uma classe.

Construção da tabela

Para construir a tabela devemos seguir apenas dois passos:

1º passo. Ordenar as categorias ou valores da variável.
Cada categoria ou valor constituirá uma classe.

- O número da classe é representado por **j**, tal que **$j=1, 2, ..., k$**, onde **k** é o número total de classes.

2º passo. Contar o número de observações em cada classe, ou seja, contar quantas vezes o dado está repetido.

Exemplo 1. Variável categórica

Variável em estudo: **conceito na disciplina de Estatística**

Dados brutos: ruim, médio, bom, médio, ruim, médio, ruim, médio, ruim, bom, médio, médio, bom, médio, médio, médio, ótimo, médio, bom, ótimo, bom, ótimo, médio, ótimo, médio, ruim, médio, ótimo, médio, médio, bom, ruim, bom, bom, médio, ruim, médio, médio, ótimo, médio, bom, ruim, ruim, bom, médio, médio, ruim, bom, médio, médio, bom, bom, bom, médio, ruim, bom, médio, médio, ruim, médio

1º passo. Ordenar os níveis da variável

Número da classe (j)	Classe
1	Ruim
2	Médio
3	Bom
4	Ótimo

2º passo. Contar o número de observações em cada classe

j	Classe	F_j
1	Ruim	12
2	Médio	27
3	Bom	15
4	Ótimo	6
	Σ	60

$F_1 = 12$

$F_2 = 27$

Os valores provenientes desta contagem, denotados por F_j, são denominados **frequências absolutas das classes.**

Outras frequências importantes:

Frequência absoluta acumulada, denotada por F'_j , expressa o número de observações acumuladas em cada classe.

j	Classe	F_j	F'_j
1	Ruim	12	12
2	Médio	27	39
3	Bom	15	54
4	Ótimo	6	60
	Σ	60	-

$F'_1 = 12$

$F'_2 = 39$

Outras frequências importantes:

Frequência relativa, denotada por f_j , expressa a proporção de observações em cada classe.

j	Classe	F_j	F'_j	f_j
1	Ruim	12	12	0,2
2	Médio	27	39	0,45
3	Bom	15	54	0,25
4	Ótimo	6	60	0,1
	Σ	60	-	1

$f_1 = 0{,}2$

$f_3 = 0{,}25$

Outras frequências importantes:

Frequência relativa acumulada, denotada por f'_j, expressa a proporção de observações acumulada em cada classe.

j	Classe	F_j	F'_j	f_j	f'_j
1	Ruim	12	12	0,2	0,2
2	Médio	27	39	0,45	0,65
3	Bom	15	54	0,25	0,90
4	Ótimo	6	60	0,1	1
	$\sum$	60	-	1	-

$f'_1 = 0,2$

$f'_3 = 0,90$

Frequências importantes:

F_j: **frequência absoluta da classe j** → número de observações na classe j

F'_j: **frequência absoluta acumulada da classe j** → número de observações da classe 1 até a classe j

f_j: **frequência relativa da classe j** → proporção de observações na classe j

f'_j: **frequência relativa acumulada da classe j** → proporção de observações da classe 1 até a classe j

Interpretação:

Variável: **conceito na disciplina de Estatística**

Unidade de observação: **o aluno**

j	Classe	F_j	F'_j	f_j	f'_j
1	Ruim	12	12	0,2	0,2
2	Médio	27	39	0,45	0,65
3	Bom	15	54	0,25	0,90
4	Ótimo	6	60	0,1	1
	Σ	60	-	1	-

Interpretação:

proporção de alunos que obtiveram até conceito Médio

número de alunos que obtiveram até conceito Bom

j	Classe	F_j	F'_j	f_j	f'_j
1	Ruim	12	12	$\boxed{0,2}$	0,2
2	Médio	27	39	0,45	$\boxed{0,65}$
3	Bom	15	$\boxed{54}$	0,25	0,90
4	Ótimo	$\boxed{6}$	60	0,1	1
	Σ	60	-	1	-

proporção de alunos que obtiveram conceito Ruim

número de alunos que obtiveram conceito Ótimo

Variável categórica nominal

Variável em estudo: **estado de origem de estudantes da disciplina de Estatística**

j	Classe	F_j	F'_j	f_j	f'_j
1	Rio Grande do Sul	32	~~32~~	0,5333	~~0,5333~~
2	Santa Catarina	15	~~47~~	0,25	~~0,7833~~
3	São Paulo	6	~~53~~	0,1	~~0,8833~~
4	Paraná	4	~~57~~	0,0667	~~0,95~~
5	Rio de Janeiro	2	~~59~~	0,0333	~~0,9833~~
6	Amazonas	1	~~60~~	0,0167	~~1~~
	$\sum$	60	-	1,0000	-

Para variáveis categóricas qualitativas **nominais** as **frequências acumuladas** não têm sentido.

Exemplo 2. Variável numérica discreta

Pesquisa: Monitoramento de um canal de comunicação

Variável em estudo: **número de erros em um conjunto de caracteres (*string*) de 1.000 *bits*.** Foram avaliados 350 conjuntos.

Dados brutos:

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

Exemplo 2. Variável numérica discreta

Pesquisa: **Monitoramento de um canal de comunicação**

Variável em estudo: **número de erros em um conjunto de caracteres (*string*) de 1.000 *bits*.** Foram avaliados 350 conjuntos.

1º passo: Ordenar os valores da variável

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

2º passo: Contar o número de observações em cada classe

Profª. Clause Piana

Exemplo 2. Variável numérica discreta

Pesquisa: Monitoramento de um canal de comunicação

Variável em estudo: número de erros em um conjunto de caracteres (*string*) de 1.000 *bits*. Foram avaliados 350 conjuntos.

1º passo: Ordenar os valores da variável

2º passo: Contar o número de elementos em cada classe

j	Classe	F_j
1	0	55
2	1	60
3	2	112
4	3	82
5	4	31
6	5	8
7	6	2
	Σ	350

Interpretação:

Variável: número de erros

Unidade de observação: a string

j	Classe	F_j	F'_j	f_j	f'_j
1	0	55	55	0,1571	0,1571
2	1	60	115	0,1714	0,3286
3	2	112	227	0,32	0,6486
4	3	82	309	0,2343	0,8829
5	4	31	340	0,0886	0,9714
6	5	8	348	0,0229	0,9943
7	6	2	350	0,0057	1,0000
	Σ	350	-	1,0000	-

Exercício proposto:

Os dados a seguir se referem ao número de ovos danificados em uma inspeção feita em 30 embalagens de uma dúzia cada, em um carregamento para o mercado de Lavras.

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

Construa a distribuição de frequências para esses dados.

Resolução:

j	Classe	F_j	F'_j	f_j	f'_j
1	0	13	13	0,4333	0,4333
2	1	9	22	0,3	0,7333
3	2	3	25	0,1	0,8333
4	3	3	28	0,1	0,9333
5	4	1	29	0,03333	0,9667
6	5	1	30	0,03333	1,000
	Σ	30	-	1,000	-

Interpretação:

Variável: **número de ovos danificados**

Unidade de observação: **a embalagem**

j	Classe	F_j	F'_j	f_j	f'_j
1	0	13	13	0,4333	0,4333
2	1	9	22	0,3	0,7333
3	2	3	25	0,1	0,8333
4	3	3	28	0,1	0,9333
5	4	1	29	0,03333	0,9667
6	5	1	30	0,03333	1,000
	Σ	30	-	1,000	-

Tabelas de classificação simples

As características dessas tabelas variam de acordo com o tipo de variável em estudo.

- ⇨ Se a variável é do tipo **categórica**, devemos obter as frequências para cada nível dessa variável.

- ⇨ Se a variável é do tipo **numérica contínua**, devemos primeiro construir intervalos de mesma amplitude e depois obter as frequências para cada intervalo.

As **variáveis contínuas**, em geral, assumem muitos valores diferentes uns dos outros.

- Assim, as tabelas de distribuição de frequências são construídas de modo que **cada classe seja constituída por um intervalo de valores** da variável.

- Quando variáveis discretas assumem muitos valores diferentes **é usual agrupar os dados discretos em intervalos de classe**.

Distribuição de frequências para variáveis contínuas

Exemplo:

Variável em estudo: valores gastos (em reais) pelos primeiros 50 clientes que entraram num determinado Supermercado, no dia 01/01/2000.

Dados brutos:

32,03	19,54	45,40	25,13	46,69	18,36	13,78	15,23	36,37	15,62
17,00	27,65	85,76	38,64	86,37	24,58	20,16	93,34	48,65	22,22
23,04	42,97	28,06	52,75	3,11	8,88	9,26	10,81	12,69	28,38
18,43	61,22	41,02	44,67	19,50	17,39	39,16	44,08	38,98	19,27
26,24	28,08	59,07	82,70	26,26	24,47	54,80	70,32	50,39	20,59

Construção da tabela

1º passo. Ordenar o conjunto de dados: colocar os dados brutos em ordem crescente de grandeza.

Dados ordenados:

3,11	8,88	9,26	10,81	12,69	13,78	15,23	15,62	17,00	17,39
18,36	18,43	19,27	19,50	19,54	20,16	20,59	22,22	23,04	24,47
24,58	25,13	26,24	26,26	27,65	28,06	28,08	28,38	32,03	36,37
38,64	38,98	39,16	41,02	42,97	44,08	44,67	45,40	46,69	48,65
50,39	52,75	54,80	59,07	61,22	70,32	82,70	85,76	86,37	93,34

Construção da tabela

1º passo. Ordenar o conjunto de dados: colocar os dados brutos em ordem crescente de grandeza.

2º passo. Determinar o número de classes (k) da tabela.

A perda de informação é inevitável.

De modo geral, o k **não deverá ser inferior a 5, nem superior a 15**.

- **k muito grande** ⇨ maior precisão e menor eficiência no resumo
- **k muito pequeno** ⇨ resume demais e a precisão fica prejudicada

Existe uma grande variedade de propostas para determinação do número de classes ou da amplitude das classes.
Aqui serão apresentadas duas propostas de critérios para definir o número de classes.

Construção da tabela

1º passo. Ordenar o conjunto de dados: colocar os dados brutos em ordem crescente de grandeza.

2º passo. Determinar o número de classes (k) da tabela.

- Regras para determinação do número de classes:

Arredondar para cima

$$k = 1 + 3{,}32 \times \log n$$ ← Fórmula de Sturges ($30 \leq n \leq 40$)

$$k = \sqrt{n}$$ ← Regra empírica ($n \leq 100$)

onde: k: número de classes
n: número de observações

Outros métodos para determinar o número de classes ou a amplitude do intervalo:

- Fórmula de Scott
- Fórmula combinada de Terrel e Scott
- Método de Shimazaki e Shinomoto
- Método de Freedman-Diaconis

Ver Sistema Galileu. Disponível em http://www.galileu.esalq.usp.br/topico.html

3º passo. Determinar a amplitude do intervalo de classe.

Para isso utilizamos a expressão $i = \frac{a_t}{k}$ ← Arredondar para cima

onde: i: amplitude do intervalo

a_t: amplitude total = ES - EI $\begin{cases} X_{(1)} = \text{Extremo Inferior} \\ X_{(n)} = \text{Extremo Superior} \end{cases}$

4º passo. Construir os intervalos de classe.

j	Classe
1	$x_{(1)} \vdash x_{(1)} + i$
2	$x_{(1)} + i \vdash x_{(1)} + 2i$
3	$x_{(1)} + 2i \vdash x_{(1)} + 3i$
...	...
k	$x_{(1)} + (k-1)i \vdash x_{(1)} + ki$

5º passo. Contar o número de observações em cada classe.

Outra forma de determinar a amplitude do intervalo de classe.

Se o processo se inicia com a definição do número k de classes, faz-se necessário definir a amplitude das classes e quais serão as classes, o que também pode alterar o formato da distribuição. Há duas possibilidades:

1. Estabelecer $x_{(1)}$, o menor valor do conjunto, como extremo inferior da primeira classe. Neste caso, a amplitude do intervalo (i) é obtida pela expressão

$$i = \frac{a_t}{k},$$

onde: **i: amplitude do intervalo**

a_t: amplitude total = ES - EI

$$\begin{cases} x_{(1)} = \text{Extremo Inferior} \\ x_{(n)} = \text{Extremo Superior} \end{cases}$$

2. Considerando-se que na obtenção de um novo conjunto de dados nas mesmas circunstâncias, a possibilidade de que esse novo conjunto contenha o menor valor é pequena, ou ainda se o número de observações for aumentado, aumenta a possibilidade de se observar valores menores que $x_{(1)}$, pode-se iniciar a primeira classe com um limite menor que $x_{(1)}$, usualmente $x_{(1)} - \frac{i}{2}$.

Nesse caso, a amplitude do intervalo é obtida como $i = \frac{a_t}{k-1}$,

sendo k-1 um ajuste para que a primeira classe inicie antes do menor valor.

Na construção dos intervalos de classe é importante observar que:

- Recomenda-se o uso de intervalos de mesma amplitude, mas eventualmente uma amplitude variável poderá ser mais adequada ao contexto;
- Deve ser garantido que todas as observações sejam classificadas;
- As classes são mutuamente exclusivas, ou seja, uma observação pertence a uma única classe;
- Os intervalos são fechados à esquerda e abertos à direita, de modo que um valor que coincida com o extremo superior será classificado na classe seguinte. Somente a última classe será exceção, podendo ser fechada também à direita para garantir a classificação do maior valor.

Exemplo:

Os dados em rol abaixo (ordenação horizontal) se referem aos valores gastos (em reais) pelos primeiros 50 clientes que entraram em um determinado Supermercado, no dia 01/01/2000.

3,11	8,88	9,26	10,81	12,69	13,78	15,23	15,62	17,00	17,39
18,36	18,43	19,27	19,50	19,54	20,16	20,59	22,22	23,04	24,47
24,58	25,13	26,24	26,26	27,65	28,06	28,08	28,38	32,03	36,37
38,64	38,98	39,16	41,02	42,97	44,08	44,67	45,40	46,69	48,65
50,39	52,75	54,80	59,07	61,22	70,32	82,70	85,76	86,37	93,34

Faça a distribuição de frequências desses dados.

Resolução: Usando a fórmula empírica

$n = 50$

$k = \sqrt{n} = \sqrt{50} = 7{,}07 \cong 8$

$i = \dfrac{a_t}{k} = \dfrac{ES - EI}{k} = \dfrac{93{,}34 - 3{,}11}{8} = 11{,}27875 = 11{,}28$

Ponto médio ou centro de classe

j	Classe	F_j	F'_j	f_j	f'_j	c_j
1	3,11 |— 14,39	6	6	0,12	0,12	8,75
2	14,39 |— 25,67	16	22	0,32	0,44	20,03
3	25,67 |— 36,95	8	30	0,16	0,60	31,31
4	36,95 |— 48,23	9	39	0,18	0,78	42,59
5	48,23 |— 59,51	5	44	0,10	0,88	53,87
6	59,51 |— 70,79	2	46	0,04	0,92	65,15
7	70,79 |— 82,07	0	46	0,00	0,92	76,43
8	82,07 |— 93,35	4	50	0,08	1	87,71
	Σ	50	-	1	-	-

$c_j = \dfrac{LI_j + LS_j}{2}$

$c_2 = c_1 + i$

$c_3 = c_1 + 2i$

Interpretação:

Variável: valor gasto (R$)

Unidade de observação: o cliente

proporção de clientes que gastaram de 36,95 a 48,23 reais (exclusive)

j	Classe	F_j	F'_j	f_j	f'_j
1	3,11 |— 14,39	6	6	0,12	0,12
2	14,39 |— 25,67	16	22	0,32	0,44
3	25,67 |— 36,95	$\boxed{8}$	30	0,16	0,60
4	36,95 |— 48,23	9	39	$\boxed{0,18}$	0,78
5	48,23 |— 59,51	5	$\boxed{44}$	0,10	0,88
6	59,51 |— 70,79	2	46	0,04	$\boxed{0,92}$
7	70,79 |— 82,07	0	46	0,00	0,92
8	82,07 |— 93,35	4	50	0,08	1
	Σ	50	-	1	-

proporção de clientes que gastaram de 3,11 a 70,79 reais (exclusive)

número de clientes que gastaram de 25,67 a 36,95 reais (exclusive)

número de clientes que gastaram de 3,11 a 59,51 reais (exclusive)

Exercício proposto:

Os dados se referem às notas dos alunos dos cursos de Ciência e Engenharia da Computação da UFPel na primeira prova de Estatística Básica, no segundo semestre de 2013.

4,5	4,7	5,4	5,7	5,7	5,8	5,9	6,1
6,3	6,5	6,7	6,7	6,8	7,0	7,1	7,2
7,2	7,3	7,5	7,6	7,6	7,7	8,0	8,0
8,1	8,1	8,2	8,3	8,3	8,5	8,5	8,5
8,6	8,9	9,0	9,0	9,1	9,4	9,4	9,5

Faça a distribuição de frequências desses dados.

Resolução:

$$n = 40$$

$$k = 1 + 3{,}32 \times \log n = 1 + 3{,}32 \times 1{,}6 = 6{,}32 \cong 7$$

$$i = \frac{a_t}{k} = \frac{ES - EI}{k} = \frac{9{,}5 - 4{,}5}{7} = 0{,}8$$

j	Classe	F_j	F'_j	f_j	f'_j	c_j
1	4,5 |— 5,3	2	2	0,05	0,05	4,9
2	5,3 |— 6,1	5	7	0,125	0,175	5,7
3	6,1 |— 6,9	6	13	0,15	0,325	6,5
4	6,9 |— 7,7	8	21	0,2	0,525	7,3
5	7,7 |— 8,5	8	29	0,2	0,725	8,1
6	8,5 |— 9,3	8	37	0,2	0,925	8,9
7	9,3 |— 10,1	3	40	0,075	1	9,7
	Σ	40	-	1	-	-

Se o arredondamento for para duas casas decimais os limites dos intervalos e a frequências das classes mudam.

$$i = \frac{a_t}{k} = \frac{ES - EI}{k} = \frac{9{,}5 - 4{,}5}{7} = 0{,}72$$

j	Classe	F_j	c_j
1	4,5 |— 5,22	2	4,86
2	5,22 |— 5,94	5	5,58
3	5,94 |— 6,66	3	6,3
4	6,66 |— 7,38	8	7,02
5	7,38 |— 8,10	6	7,74
6	8,10 |— 8,82	9	8,46
7	8,82 |—| 9,54	7	9,18
	Σ	40	

Interpretação:

Variável: **nota na primeira prova na disciplina de Estatística**

Unidade de observação: **o aluno**

j	Classe	F_j	F'_j	f_j	f'_j	c_j
1	4,5 |— 5,3	2	2	0,05	0,05	4,9
2	5,3 |— 6,1	5	7	0,125	0,175	5,7
3	6,1 |— 6,9	6	13	0,15	0,325	6,5
4	6,9 |— 7,7	8	21	0,2	0,525	7,3
5	7,7 |— 8,5	8	29	0,2	0,725	8,1
6	8,5 |— 9,3	8	37	0,2	0,925	8,9
7	9,3 |— 10,1	3	40	0,075	1	9,7
	Σ	40	-	1	-	-

$F_2 = 5$

$F'_4 = 21$

$f_5 = 0,2$

$f'_3 = 0,325$

Representação gráfica

As distribuições de frequências podem ser representadas graficamente de duas formas distintas e exclusivas:

- Histograma
- Polígono de frequências

Histograma

j	Classe	F_j	F'_j	f_j	f'_j	c_j
1	3,11 |— 14,39	6	6	0,12	0,12	8,75
2	14,39 |— 25,67	16	22	0,32	0,44	20,03
3	25,67 |— 36,95	8	30	0,16	0,60	31,31
4	36,95 |— 48,23	9	39	0,18	0,78	42,59
5	48,23 |— 59,51	5	44	0,10	0,88	53,87
6	59,51 |— 70,79	2	46	0,04	0,92	65,15
7	70,79 |— 82,07	0	46	0,00	0,92	76,43
8	82,07 |— 93,35	4	50	0,08	1	87,71
	Σ	50	-	1	-	-

Variável contínua (dados de mensuração)

O histograma consiste de um conjunto de retângulos contíguos cuja base é igual à amplitude do intervalo e a altura proporcional à frequência das respectivas classes.

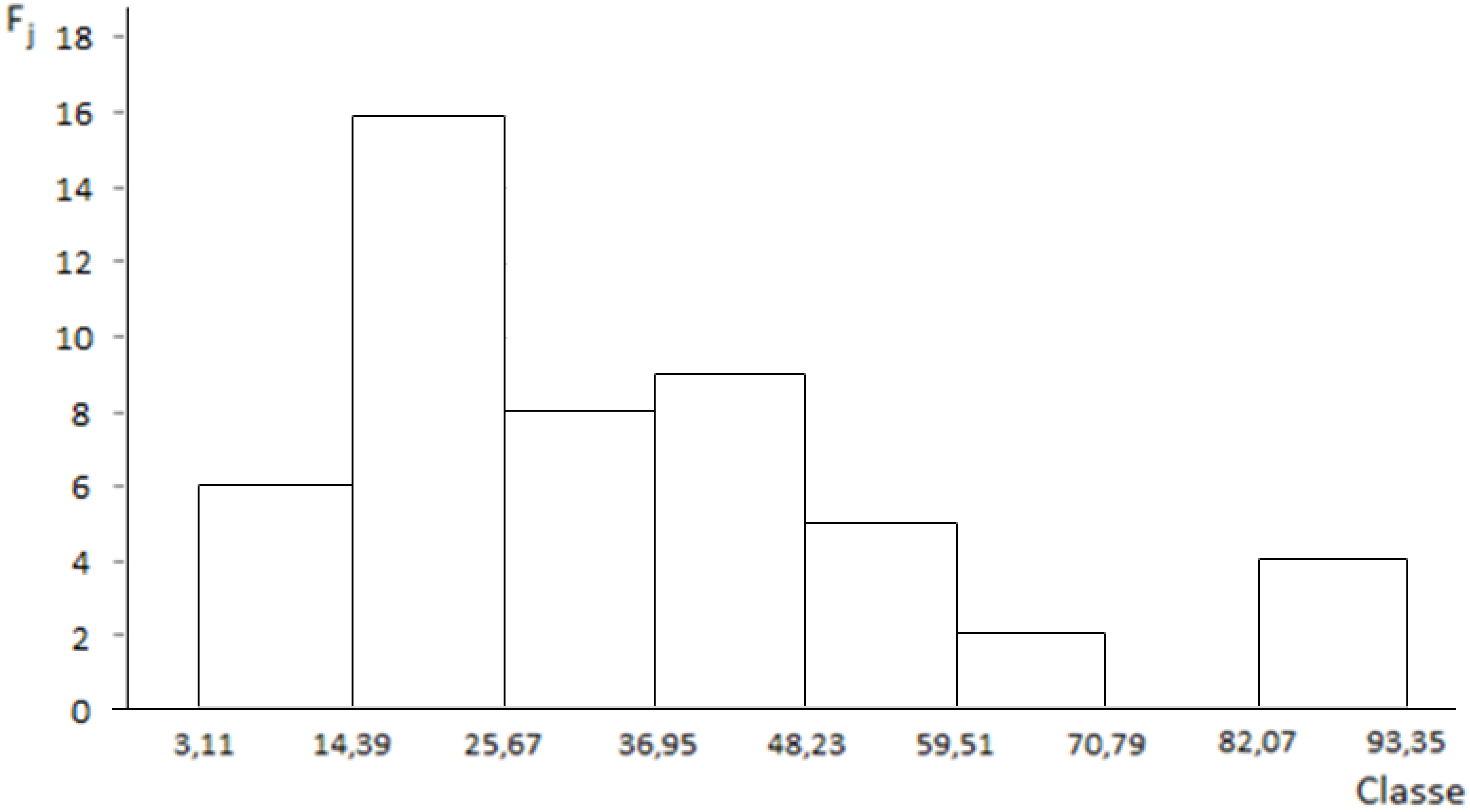


Figura 1. Frequência dos valores gastos (em reais) pelas primeiras 50 pessoas que entraram em um determinado Supermercado, no dia 01/01/2000.

Prof[a]. Clause Piana

Variável contínua (dados de mensuração)

O histograma consiste de um conjunto de retângulos contíguos cuja base é igual à amplitude do intervalo e a altura proporcional à frequência das respectivas classes.

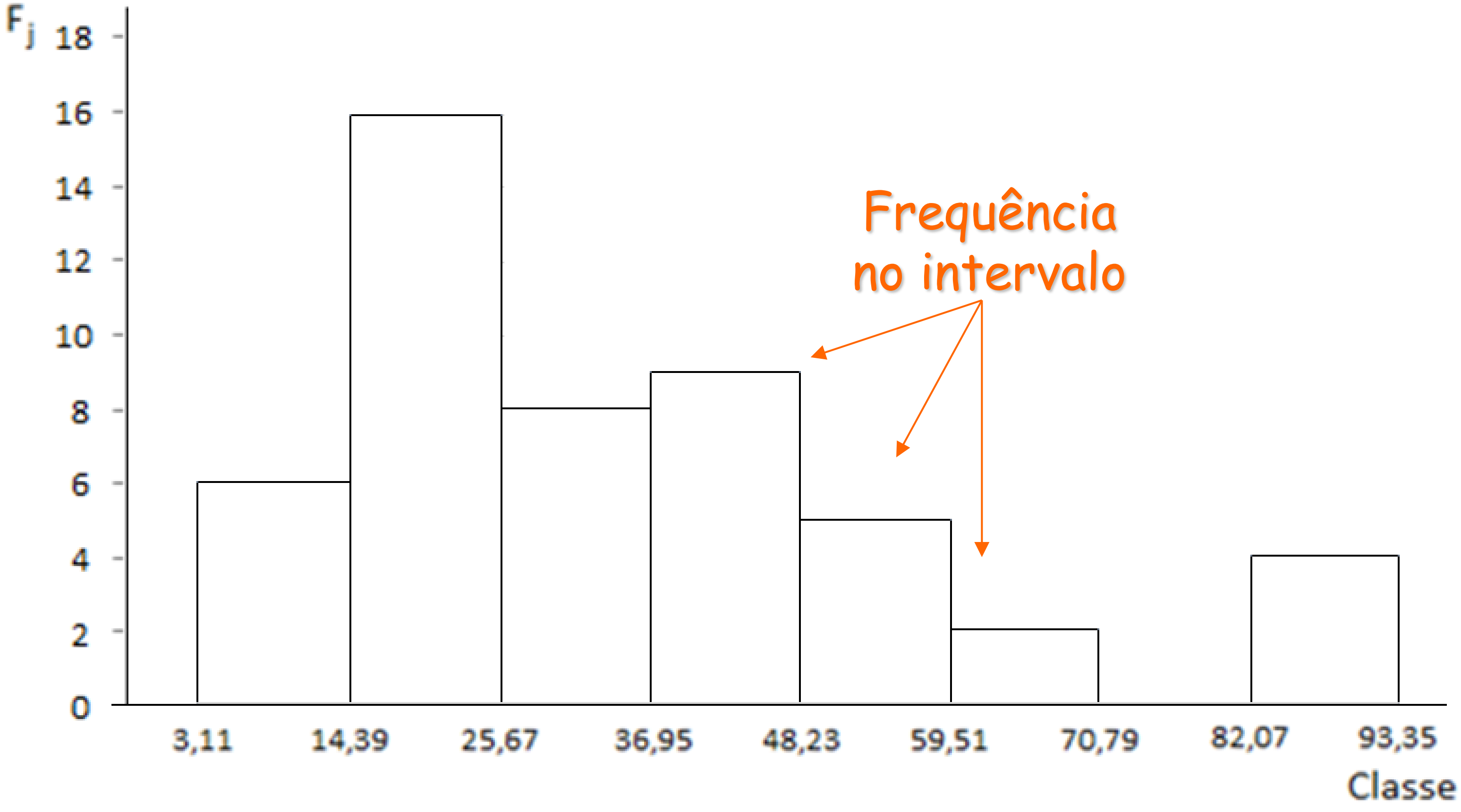


Figura 1. Frequência dos valores gastos (em reais) pelas primeiras 50 pessoas que entraram em um determinado Supermercado, no dia 01/01/2000.

Profª. Clause Piana

Histograma

Variável contínua (dados de mensuração)

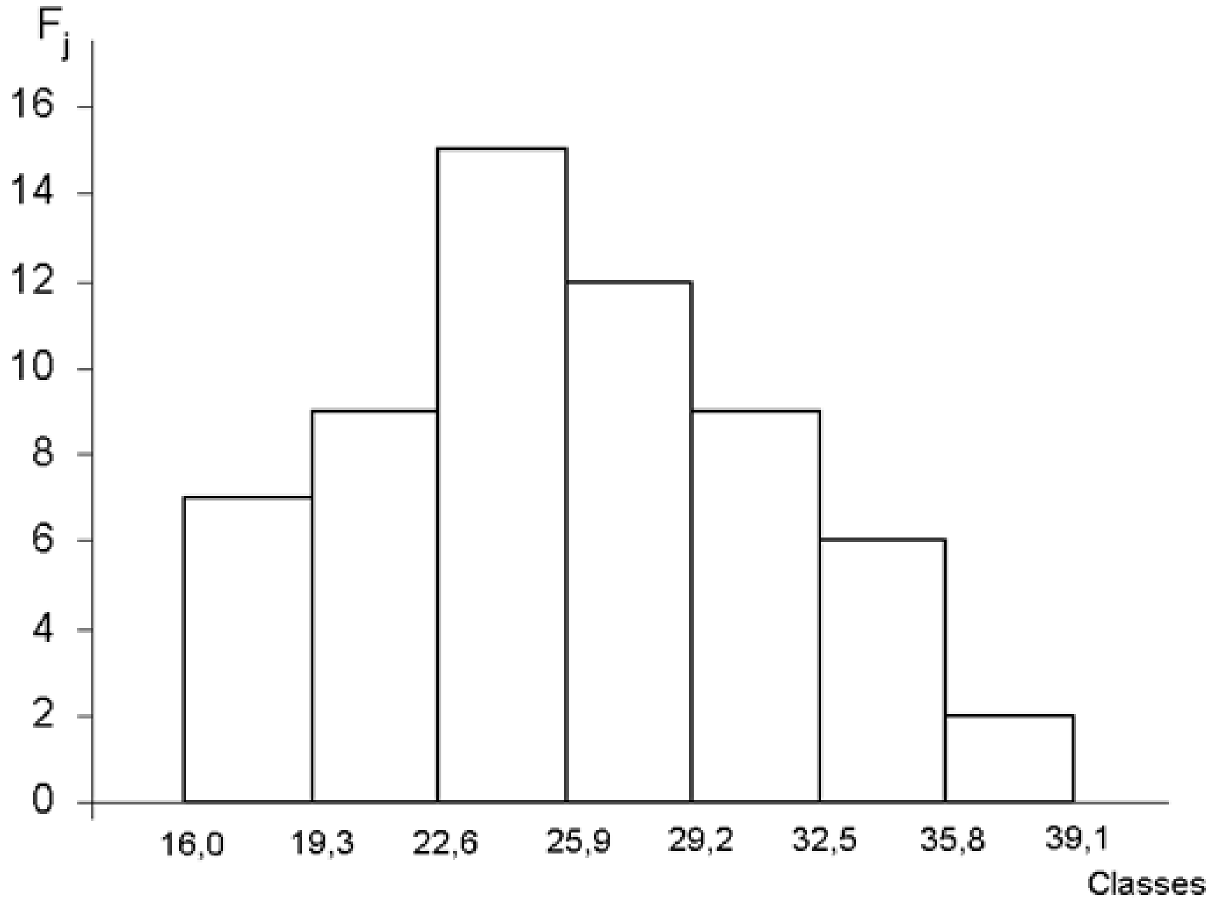


Figura 2. Frequência do peso ao nascer de 60 bovinos da raça Ibagé. UFPel, 2001.

Variável em estudo: **número de erros em *strings* de 1.000 *bits*.**

Histograma

j	Classe	F_j	F'_j	f_j	f'_j
1	0	55	55	0,1571	0,1571
2	1	60	115	0,1714	0,3286
3	2	112	227	0,32	0,6486
4	3	82	309	0,2343	0,8829
5	4	31	340	0,0886	0,9714
6	5	8	348	0,0229	0,9943
7	6	2	350	0,0057	1,0000
	Σ	350	-	1,0000	-

Histograma

Variável discreta (dados de enumeração)

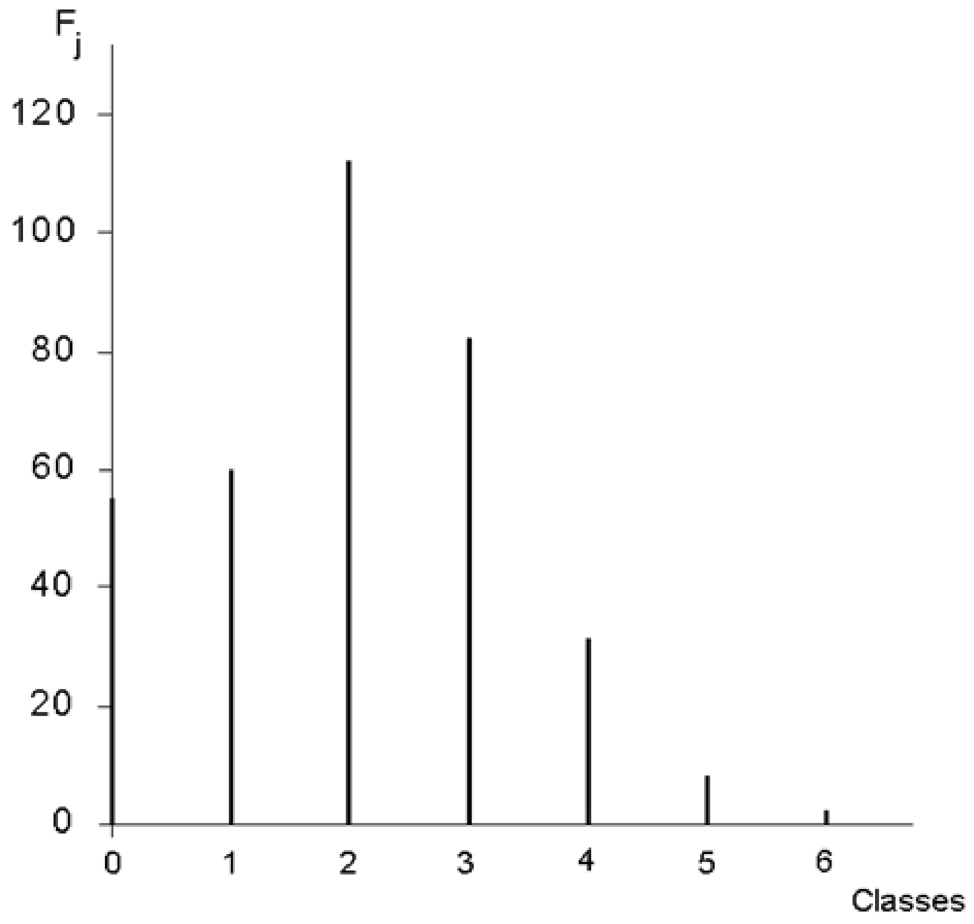


Figura 3. Frequência do número de erros em 350 conjuntos de caracteres (*strings*) de 1.000 *bits.*
Fonte: Dados fictícios.

Histograma

Variável discreta (dados de enumeração)

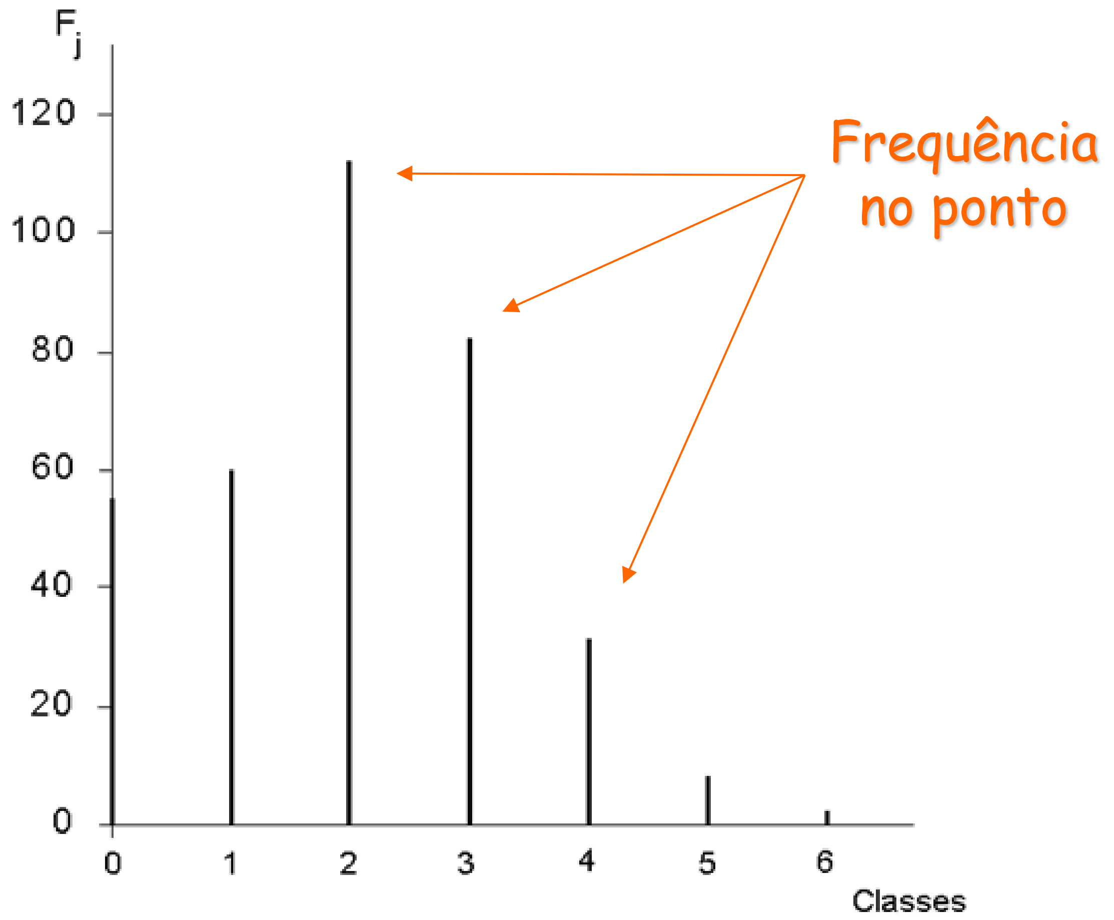


Figura 3. Frequência do número de erros em 350 conjuntos de caracteres (*strings*) de 1.000 *bits.*
Fonte: Dados fictícios.

Variável categórica

Distribuição de frequências do conceito dos alunos na disciplina de Estatística. UFPel, 2001.

j	Classe	F_j	F'_j	f_j	f'_j
1	Ruim	12	12	0,2	0,2
2	Médio	27	39	0,45	0,65
3	Bom	15	54	0,25	0,90
4	Ótimo	6	60	0,1	1
	Σ	60	-	1	-

Variável categórica

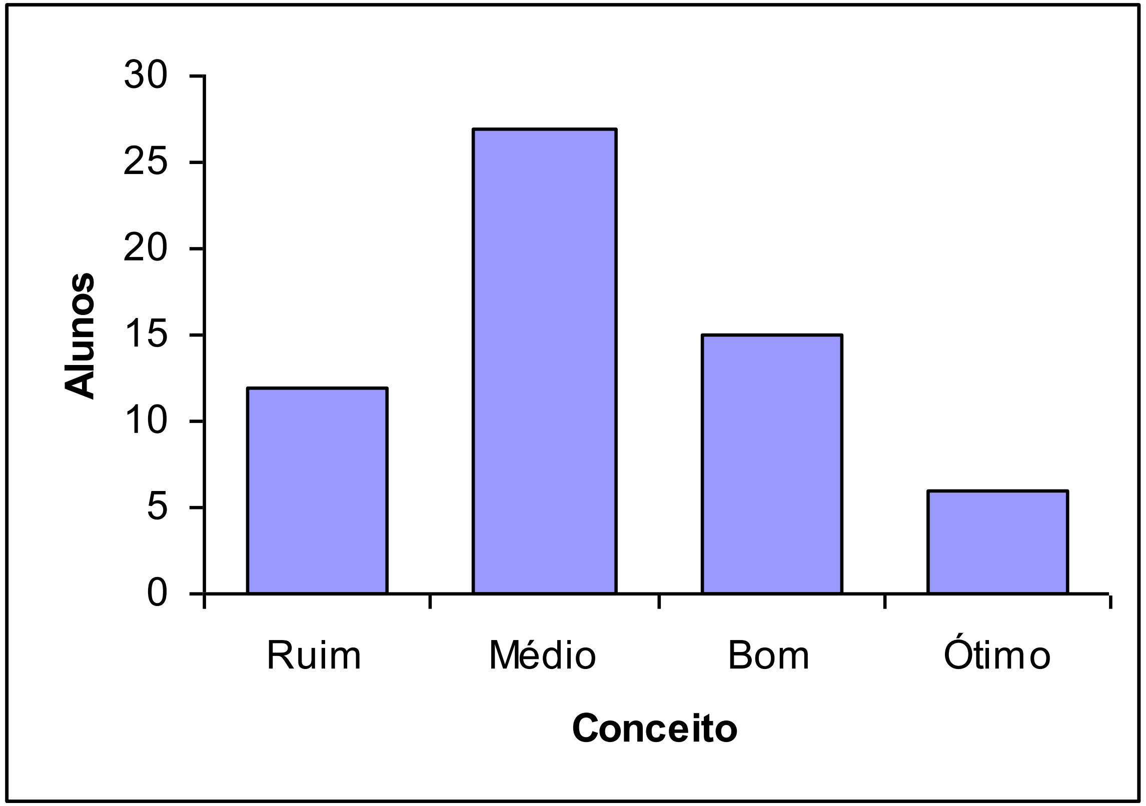


Figura 4. Conceito dos alunos na disciplina de Estatística. UFPel, 2001.

Variável categórica

Distribuição de frequências do conceito dos alunos na disciplina de Estatística. UFPel, 2001.

j	Classe	F_j	F'_j	f_j	f'_j
1	Ruim	12	12	0,2	0,2
2	Médio	27	39	0,45	0,65
3	Bom	15	54	0,25	0,90
4	Ótimo	6	60	0,1	1
	Σ	60	-	1	-

Variável categórica

Gráfico de setores

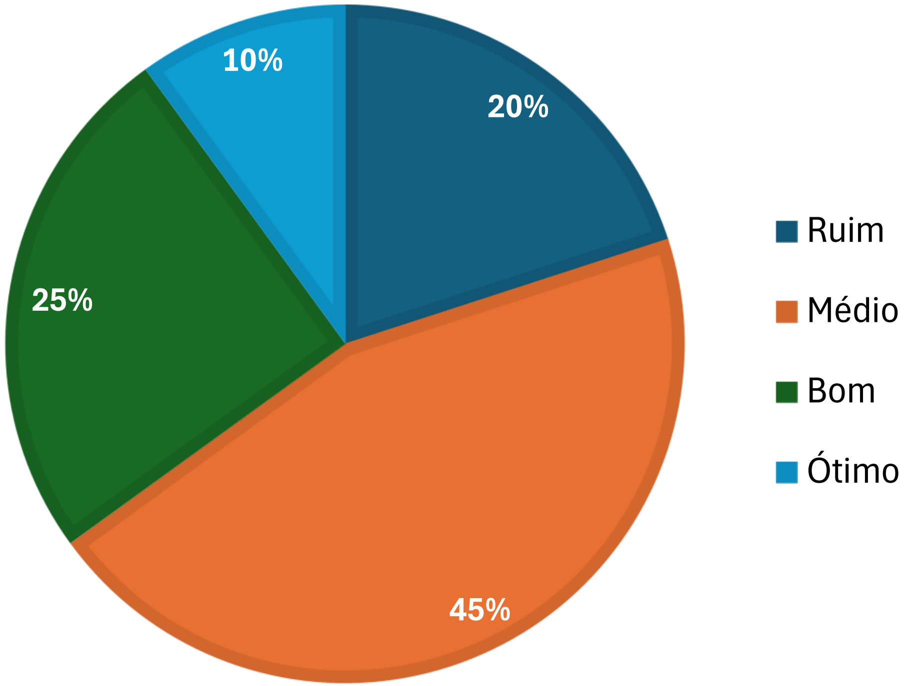


Figura 5. Conceito dos alunos na disciplina de Estatística. UFPel, 2001.

Variável categórica

Gráfico de colunas

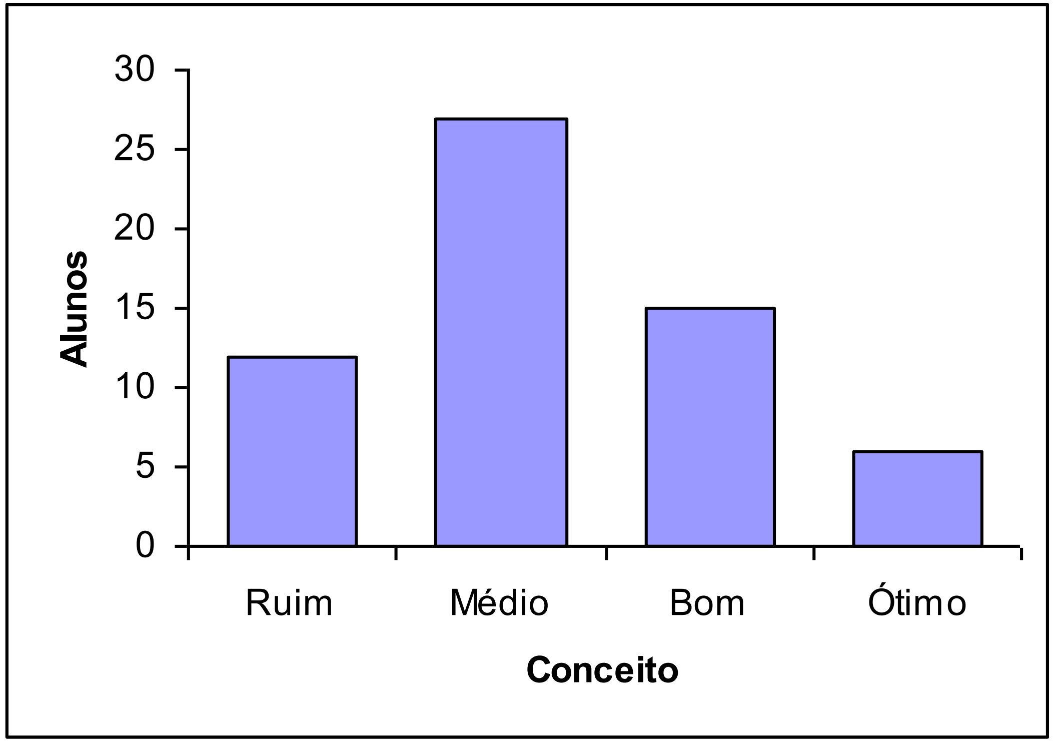


Figura 4. Conceito dos alunos na disciplina de Estatística. UFPel, 2001.

Variável numérica contínua

Histograma

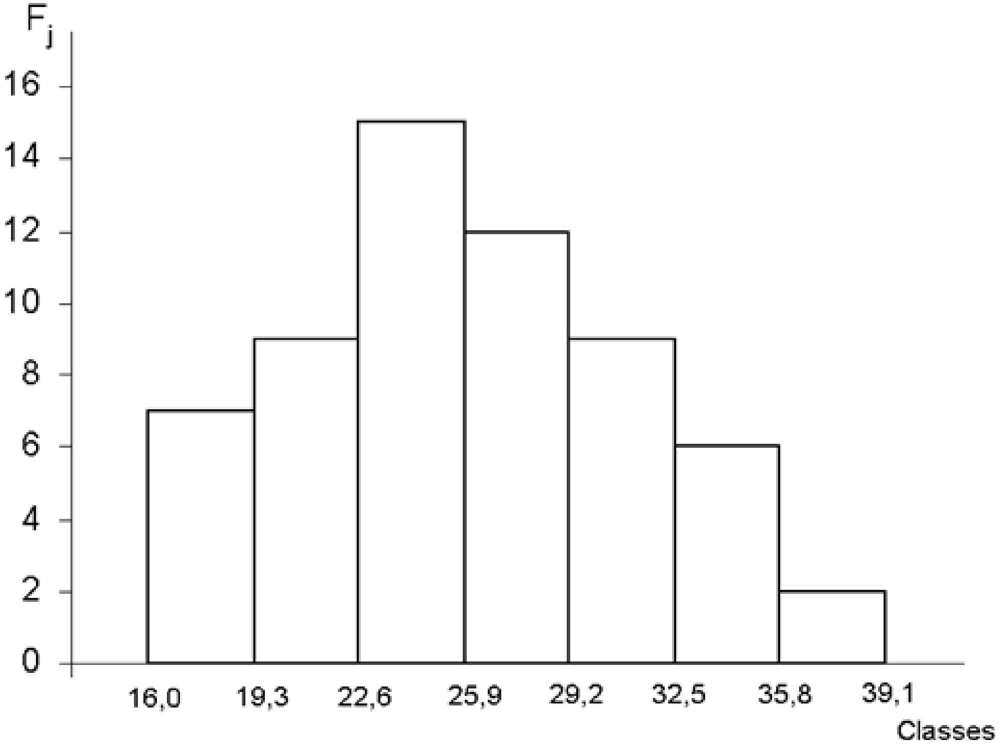


Figura 2. Frequência do peso ao nascer de 60 bovinos da raça Ibagé. UFPel, 2001.

Polígono de frequência

j	Classe	F_j	F'_j	f_j	f'_j	c_j
1	3,11 |— 14,39	6	6	0,12	0,12	8,75
2	14,39 |— 25,67	16	22	0,32	0,44	20,03
3	25,67 |— 36,95	8	30	0,16	0,60	31,31
4	36,95 |— 48,23	9	39	0,18	0,78	42,59
5	48,23 |— 59,51	5	44	0,10	0,88	53,87
6	59,51 |— 70,79	2	46	0,04	0,92	65,15
7	70,79 |— 82,07	0	46	0,00	0,92	76,43
8	82,07 |— 93,35	4	50	0,08	1	87,71
	Σ	50	-	1	-	-

Polígono de frequência

O polígono de frequências é constituído por segmentos de retas que unem os pontos cujas coordenadas são o ponto médio e a frequência de cada classe. Para fechá-lo toma-se uma classe anterior a primeira e uma posterior a última, uma vez que ambas possuem frequência zero.

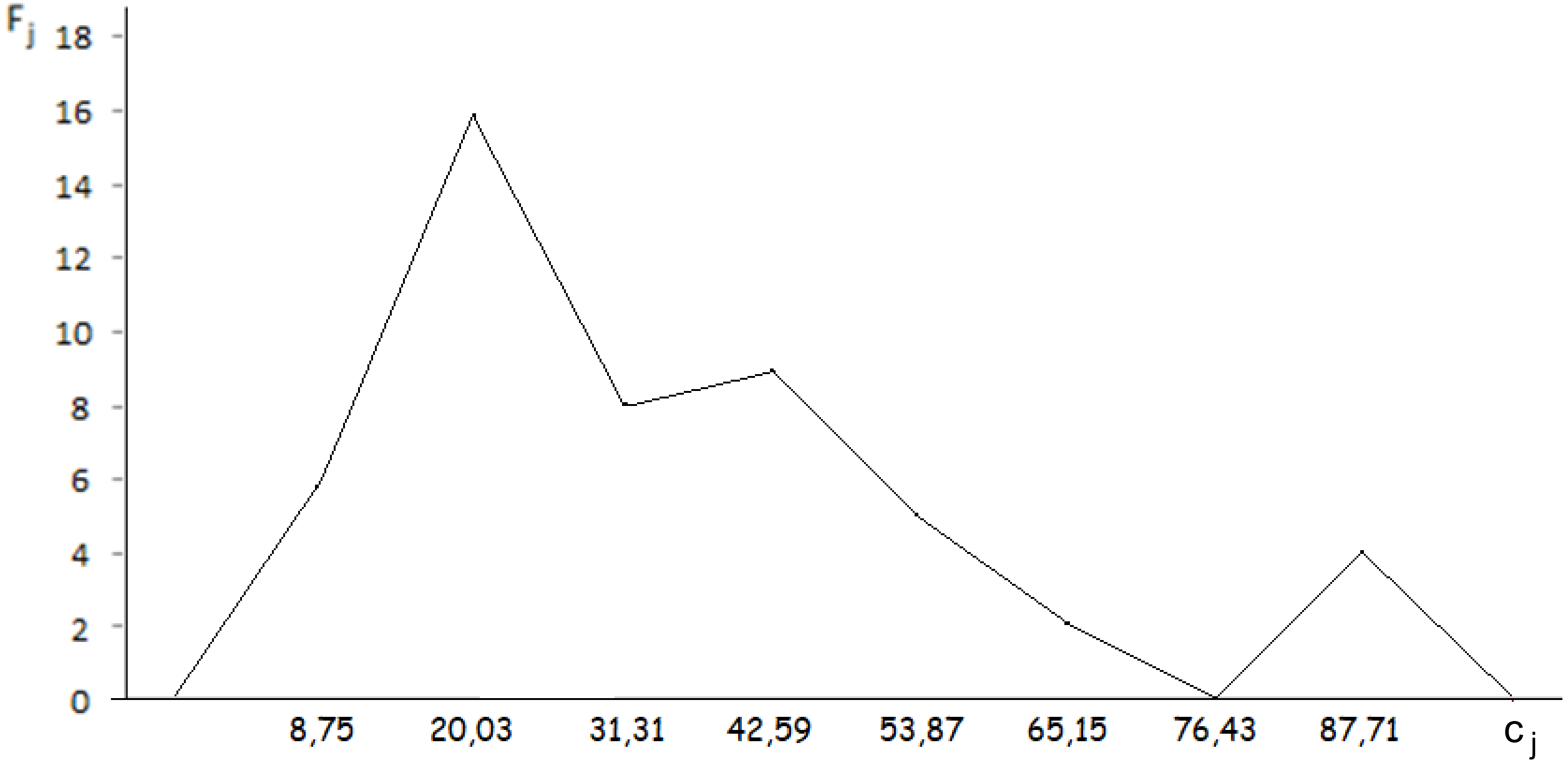


Figura 6. Notas dos alunos dos cursos de Ciência e Engenharia da Computação da UFPel na primeira prova de Estatística Básica, no segundo semestre de 2013.

Polígono de frequência

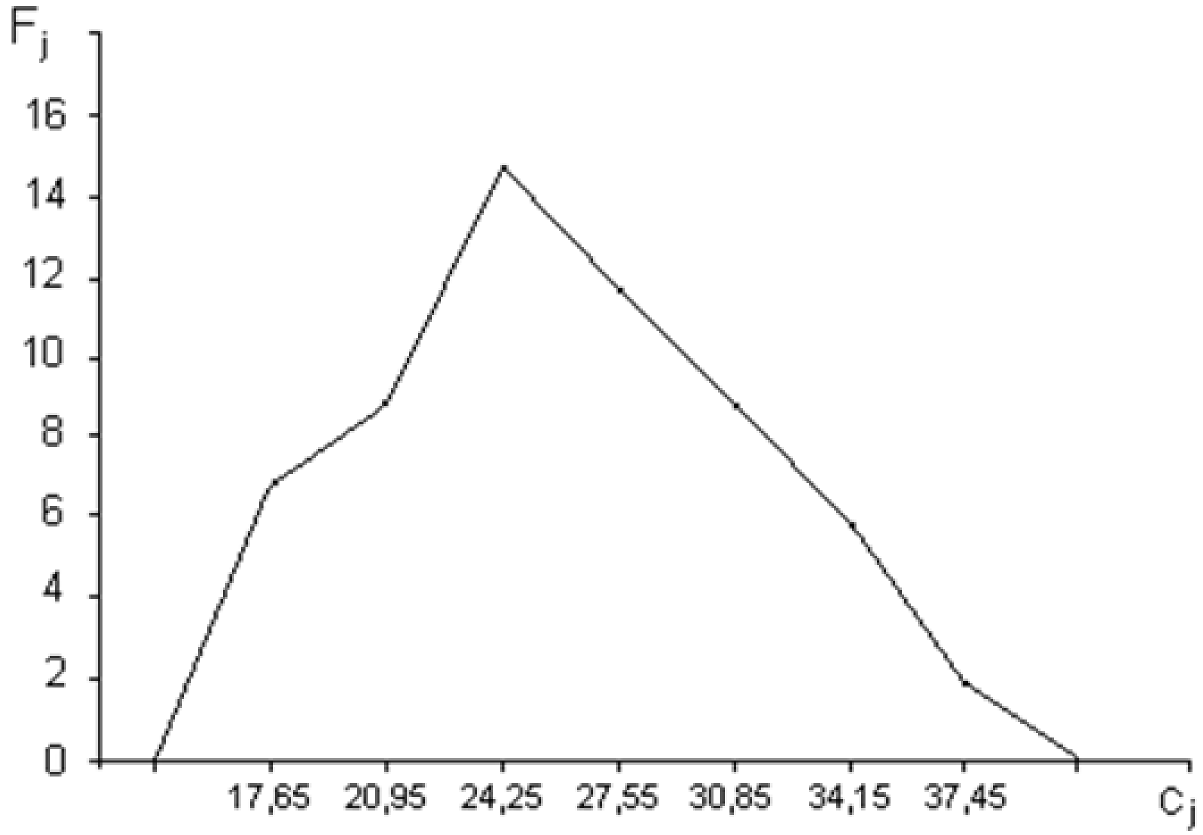
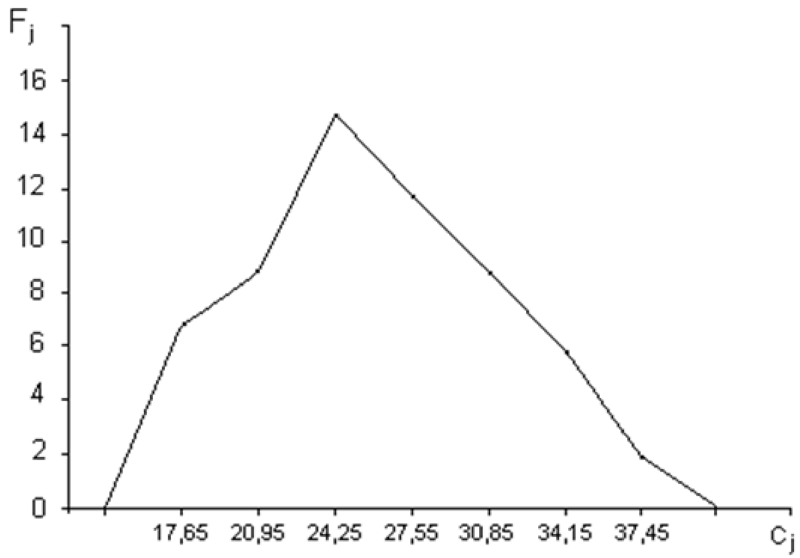
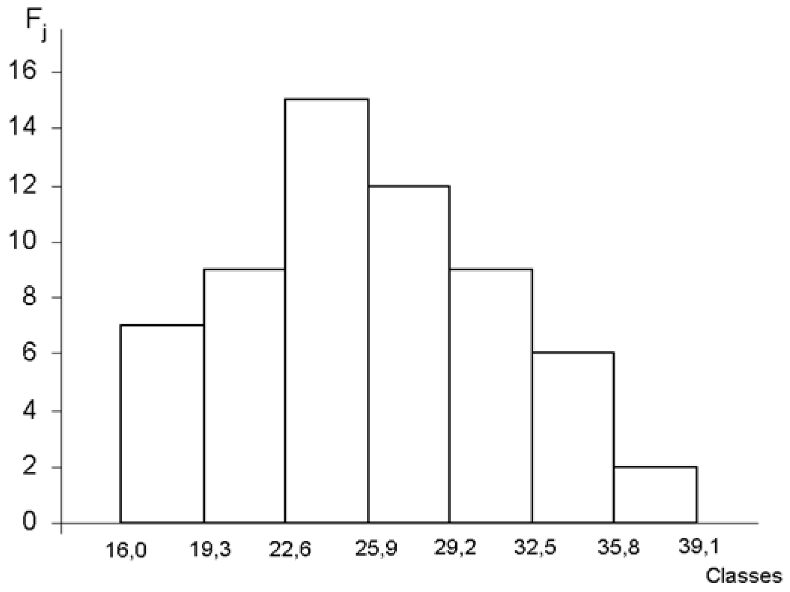


Figura 2. Frequência do peso ao nascer de 60 bovinos da raça Ibagé. UFPel, 2001.

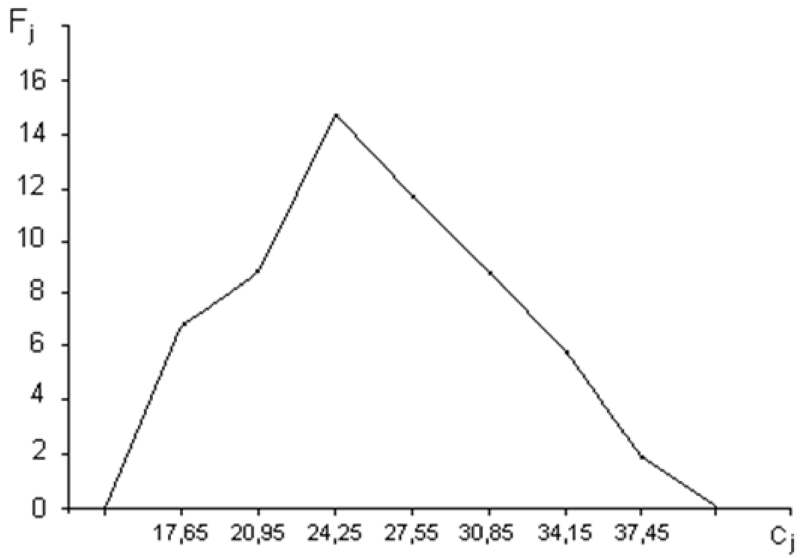
Polígono de frequências

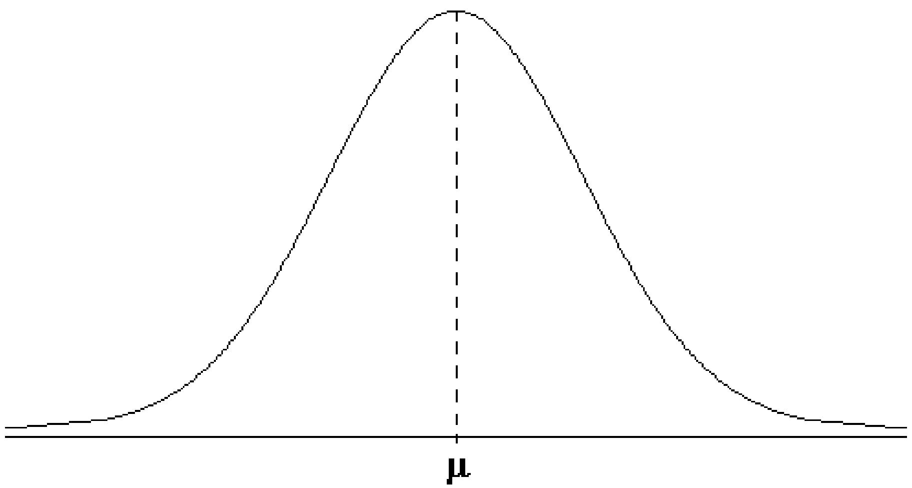


Histograma



Polígono de frequências





Ogivas – Frequências acumuladas

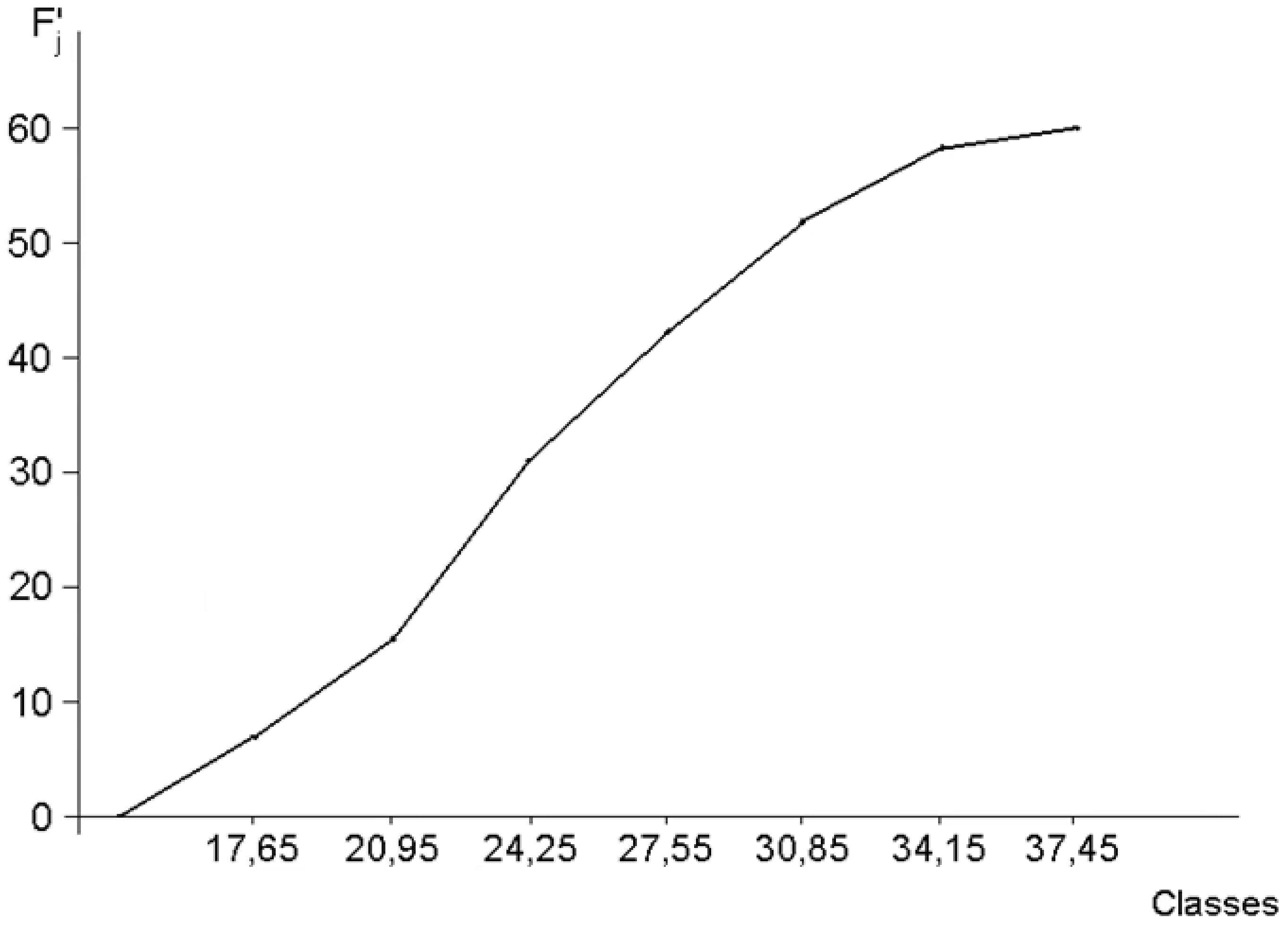


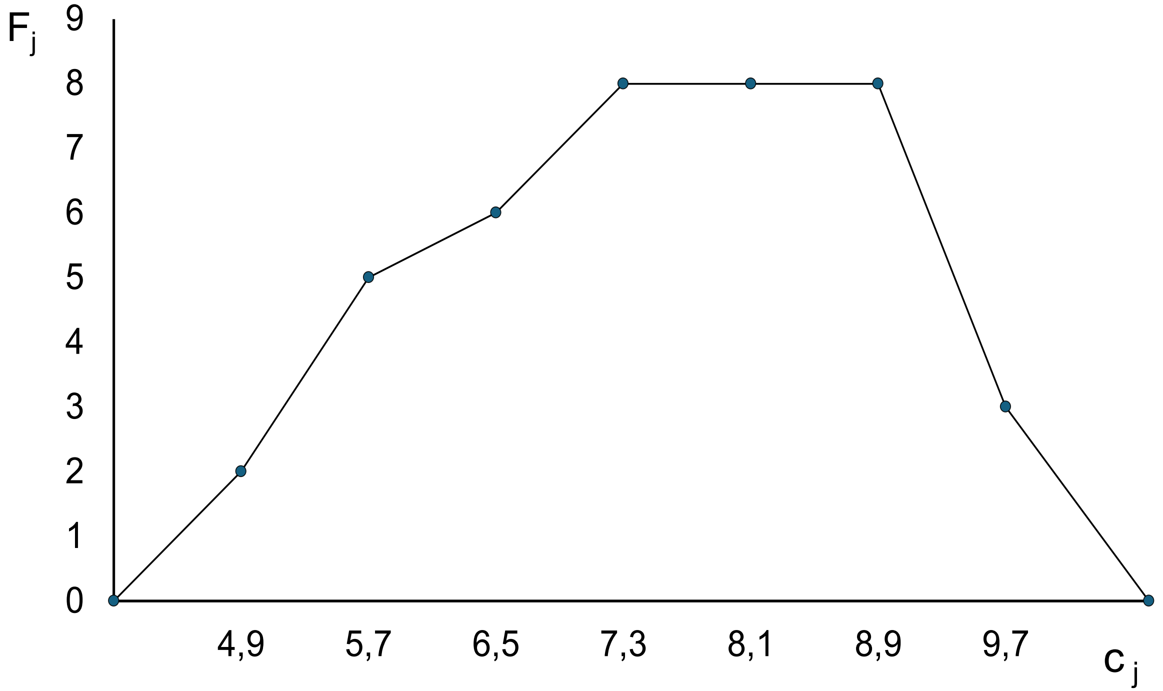
Figura 6. Frequências absolutas acumuladas do peso ao nascer de 60 bovinos da raça Ibagé. UFPel, 2001.

Exercício proposto:

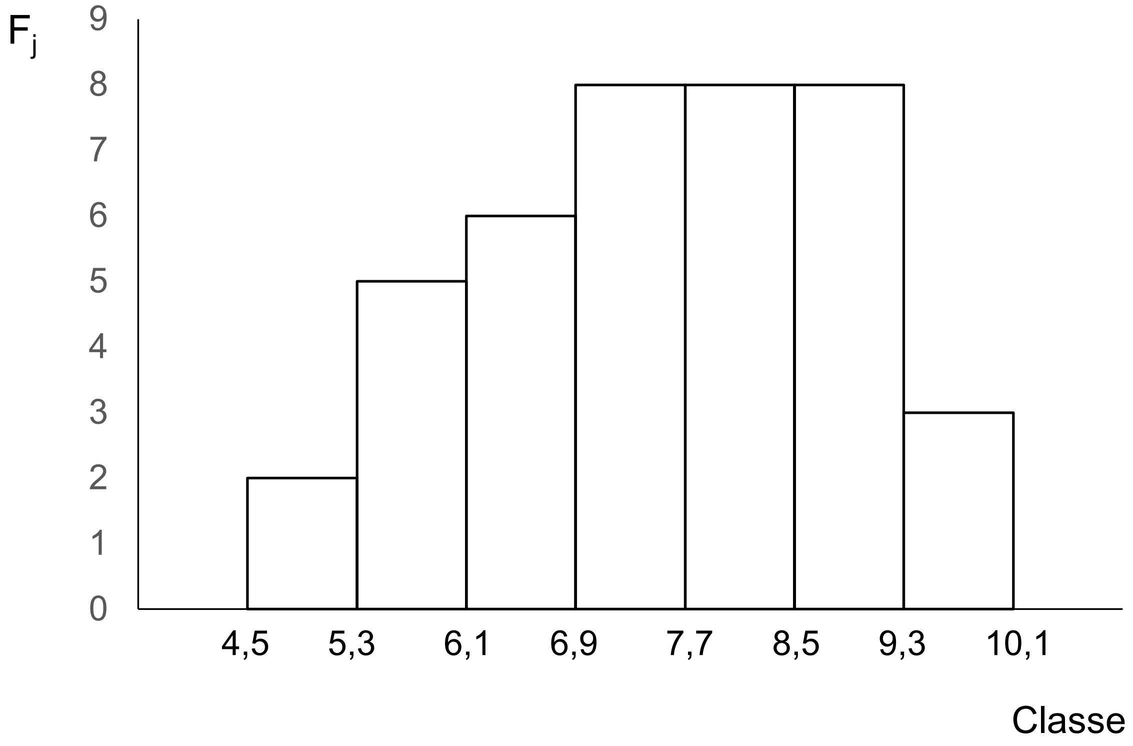
j	Classe	F_j	F'_j	f_j	f'_j	c_j
1	4,5 |— 5,3	2	2	0,05	0,05	4,9
2	5,3 |— 6,1	5	7	0,125	0,175	5,7
3	6,1 |— 6,9	6	13	0,15	0,325	6,5
4	6,9 |— 7,7	8	21	0,2	0,525	7,3
5	7,7 |— 8,5	8	29	0,2	0,725	8,1
6	8,5 |— 9,3	8	37	0,2	0,925	8,9
7	9,3 |— 10,1	3	40	0,075	1	9,7
	Σ	40	-	1	-	-

Construa um gráfico para as frequências absolutas das notas dos alunos dos cursos de Ciência e Engenharia da Computação da UFPel na primeira prova de Estatística Básica, no segundo semestre de 2013.

Polígono de frequências



Histograma



Tabelas de distribuição de frequências

- Tabelas de classificação simples

  → uma variável

- Tabelas de classificação dupla ou cruzada

  → duas variáveis

Tabela de classificação dupla ou cruzada

Tabela 2.14. Distribuição dos alunos da escola E, segundo o hábito de fumar e conceito em Estatística.

Conceito	Hábito de fumar		Totais
	Sim	Não	
Ruim	5	8	13
Médio	10	16	26
Bom	5	10	15
Ótimo	2	4	6
Totais	22	38	60

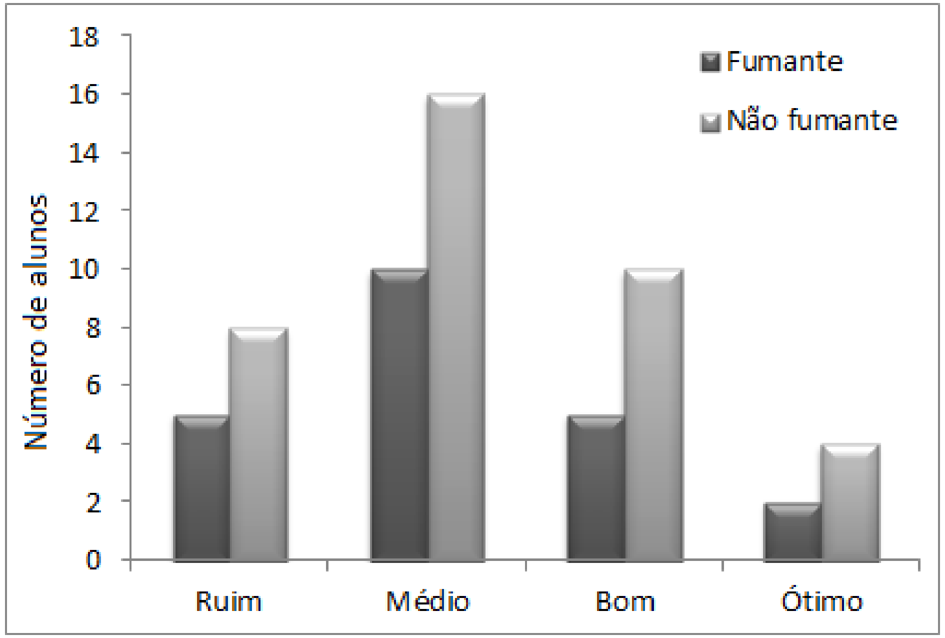


Figura 2.9. Distribuição dos alunos da escola E, segundo o hábito de fumar e conceito em Estatística.

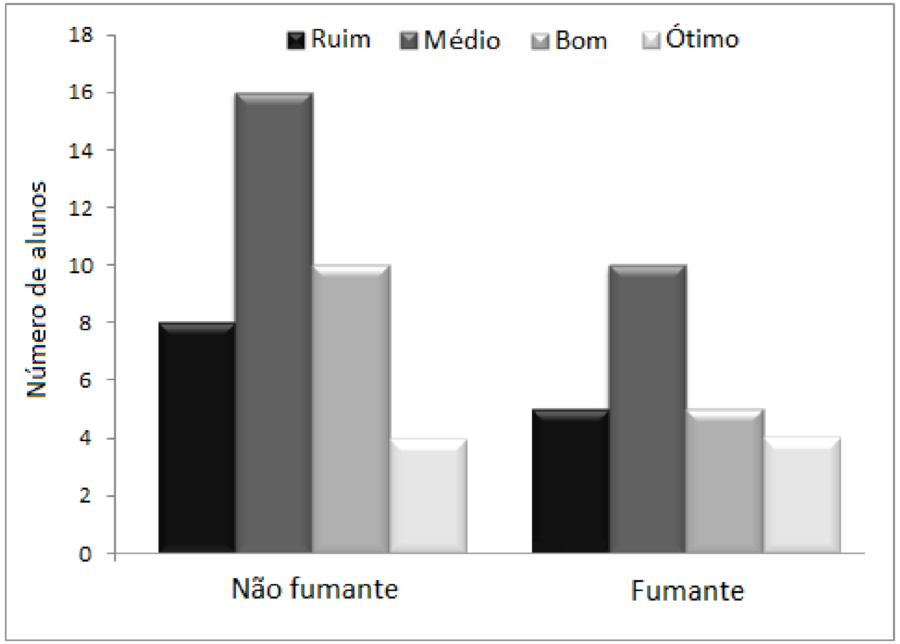


Figura 2.10. Distribuição dos alunos da escola E, segundo o hábito de fumar e conceito em Estatística.

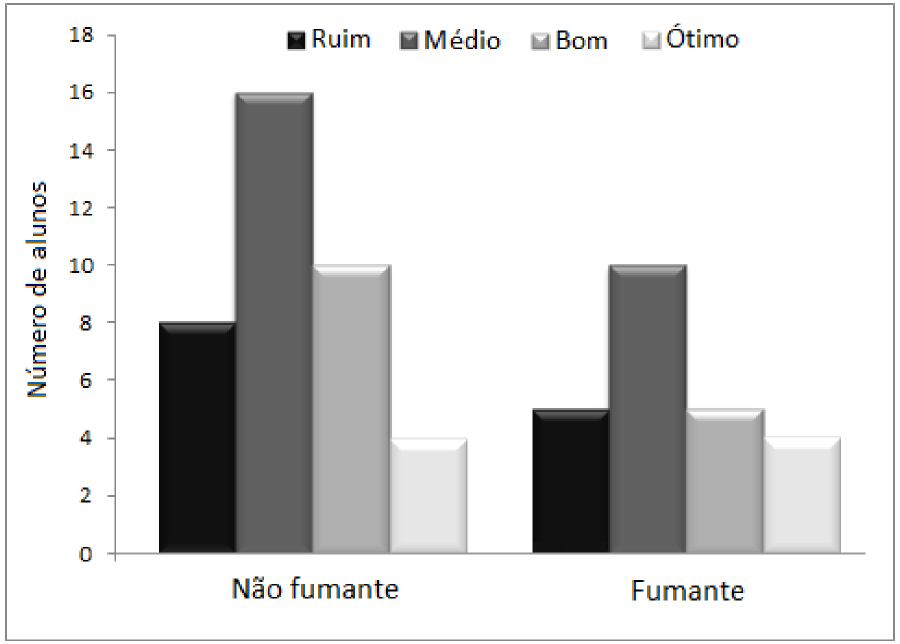


Figura 2.10. Distribuição dos alunos da escola E, segundo o hábito de fumar e conceito em Estatística.

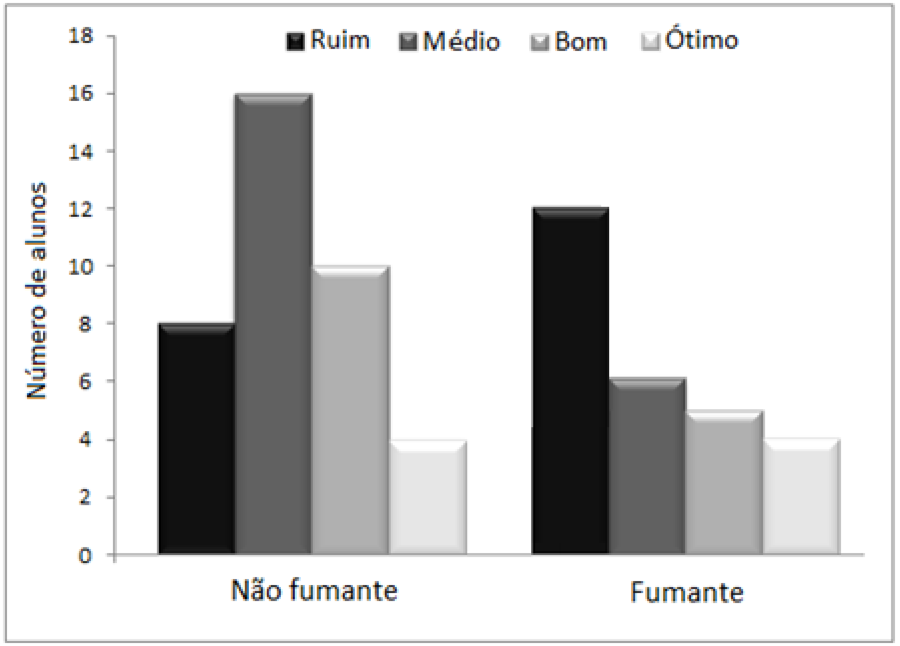


Figura 2.10. Distribuição dos alunos da escola E, segundo o hábito de fumar e conceito em Estatística.

Bibliografia utilizada

SILVA, J.G.C. da. **Estatística Básica (versão preliminar).** Universidade Federal de Pelotas.

Silveira Junior, P. ; Machado, A.A. ; Zonta, E.P.; Silva, J.B. da. **Curso de Estatística** v.1. Pelotas: Universidade Federal de Pelotas, 1992, 135p.

Sistema Galileu de Educação Estatística. Disponível em: http://www.galileu.esalq.usp.br/topico.html