

Estatística Descritiva (II)

Exemplo:

Um experimento feito por alunos. Cada um deles registrou sua altura, peso, sexo, hábito de fumar e nível de atividade física. Depois, todos eles jogaram moedas e aqueles que tiraram “CARA” fizeram corrida estacionária por 1 minuto, registrando a pulsação antes de correr e a pulsação depois de correr. Os demais registraram a pulsação após 1 minuto, mesmo sem ter corrido.



Banco de dados:

Information of the worksheet

Column	Count	Name		
C1	92	Pulse1		
C2	92	Pulse2		
C3	92	Ran	1- fez corrida	2- não fez corrida
C4	92	Smokes	1- fuma	2- não fuma
C5	92	Sex	1- masculino	2- feminino
C6	92	Height		
C7	92	Weight		
C8	92	Activity	0- não tem	1- leve
			2- moderada	3- intensa

Planilha (parcial)

Row	Pulse1	Pulse2	Ran	Smokes	Sex	Height	Weight	Activity
-----	--------	--------	-----	--------	-----	--------	--------	----------

1	64	88	1	2	1	66,00	140	2
---	----	----	---	---	---	-------	-----	---

2	58	70	1	2	1	72,00	145	2
---	----	----	---	---	---	-------	-----	---

3	62	76	1	1	1	73,50	160	3
---	----	----	---	---	---	-------	-----	---

4	66	78	1	1	1	73,00	190	1
---	----	----	---	---	---	-------	-----	---

5	64	80	1	2	1	69,00	155	2
---	----	----	---	---	---	-------	-----	---

6	74	84	1	2	1	73,00	165	1
---	----	----	---	---	---	-------	-----	---

7	84	84	1	2	1	72,00	150	3
---	----	----	---	---	---	-------	-----	---

8	68	72	1	2	1	74,00	190	2
---	----	----	---	---	---	-------	-----	---

•

•

•

•

Variáveis qualitativas



Ran
Smokes
Sex
Activity

Nominal

Ordinal

Variáveis quantitativas



Pulse 1
Pulse 2
Height
Weight

Discreta

Contínua

Variáveis Quantitativas

Medidas de posição

Média ($\bar{x}$)

Mediana (md)

Quartis (Q1, Q3)

Máximo (máx)

Mínimo (min)

Medidas de dispersão

Variância (s^2)

Desvio padrão (s)

Intervalo-interquartil (Q3 – Q1)

Coeficiente de variação (CV)

Variáveis quantitativas no banco de dados

Medidas de posição e de dispersão

variável	N	Média	Mediana	DP	Erro médio
Pulse1	92	72,87	71	11,01	1,15
Pulse2	92	80,00	76	17,09	1,78
Height	92	68,72	69	3,66	0,38
Weight	92	145,15	145	23,74	2,48

Variável	Min	Max	Q1	Q3
Pulse1	48	100	64	80,0
Pulse2	50	140	68	87,0
Height	61	75	66	72,0
Weight	95	215	125	156,5

Descrevendo a pulsação em repouso segundo o sexo

Variável	Sex	N	Média	Mediana	DP	Erro médio
Pulse1	1	57	70,42	70	9,95	1,32
	2	35	76,86	78	11,62	1,96

Variável	Sex	Min	Max	Q1	Q3
Pulse1	1	48	92	63	75
	2	58	100	66	86

Os dados também podem ser resumidos construindo-se uma tabela de distribuição de frequências.

Distribuição de frequências de uma variável é uma lista dos valores individuais ou dos intervalos de valores que a variável pode assumir, com as respectivas frequências de ocorrência.

No banco de dados



Não há perda
de informação



Summary Statistics for Discrete Variables

Pulse1	Count	Percent
--------	-------	---------

48	1	1,09
----	---	------

54	2	2,17
----	---	------

58	3	3,26
----	---	------

60	4	4,35
----	---	------

61	1	1,09
----	---	------

62	9	9,78
----	---	------

64	4	4,35
----	---	------

66	5	5,43
----	---	------

68	11	11,96
----	----	-------

70	6	6,52
----	---	------

72	6	6,52
----	---	------

74	5	5,43
----	---	------

76	5	5,43
----	---	------

78	5	5,43
----	---	------

80	3	3,26
----	---	------

82	3	3,26
----	---	------

84	4	4,35
----	---	------

86	1	1,09
----	---	------

87	1	1,09
----	---	------

88	3	3,26
----	---	------

90	4	4,35
----	---	------

92	2	2,17
----	---	------

94	1	1,09
----	---	------

96	2	2,17
----	---	------

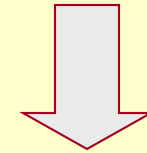
100	1	1,09
-----	---	------

N=	92	
----	----	--

Alternativa: construir intervalos de classe

Classe de pulsação	frequência
48 - 54	1
54 - 60	5
60 - 66	18
66 - 72	22
72 - 78	16
78 - 84	11
84 - 90	9
90 - 96	7
96 - 102	3

Informações mais resumidas



Perda de informação

Exemplo 2:

Variável: altura (*height*) $\Rightarrow$ contínua $\Rightarrow$

Construir
intervalos
de classe

Distribuição de frequência para altura

Classes de altura	f	fr
60,25 - 61,75	1	0,011
61,75 - 63,25	10	0,109
63,25 - 64,75	2	0,022
64,75 - 66,25	13	0,141
66,25 - 67,75	7	0,076
67,75 - 69,25	20	0,217
69,25 - 70,75	7	0,076
70,75 - 72,25	15	0,163
72,25 - 73,75	9	0,098
73,75 - 75,25	8	0,087
Total	92	1

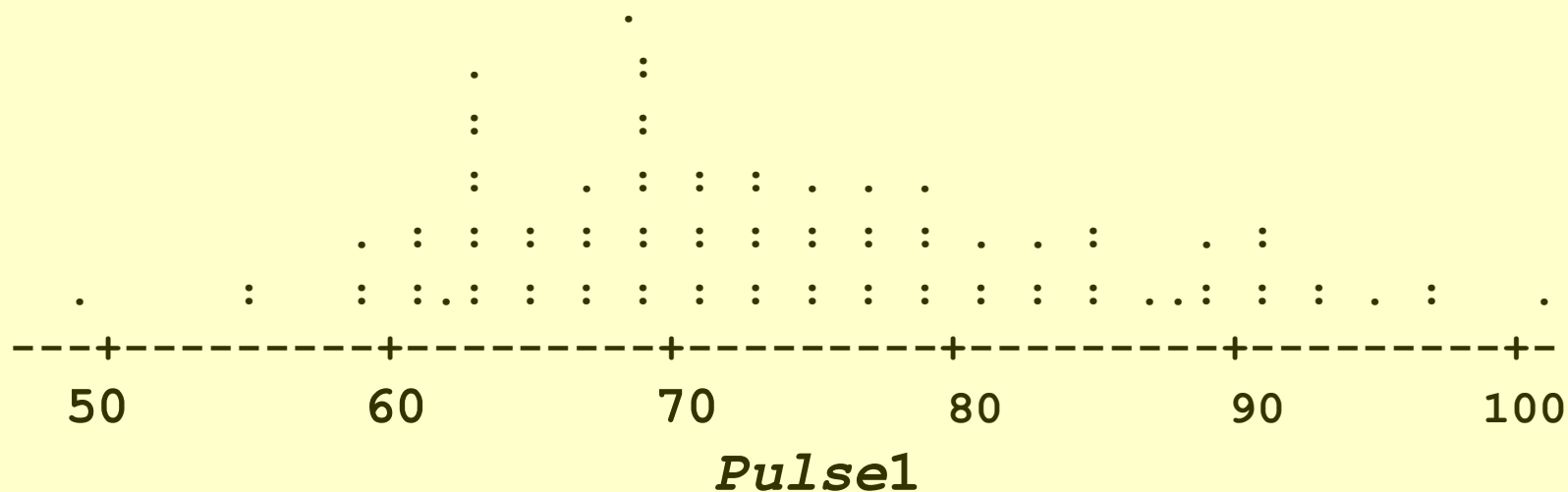
Variáveis Quantitativas

Gráficos

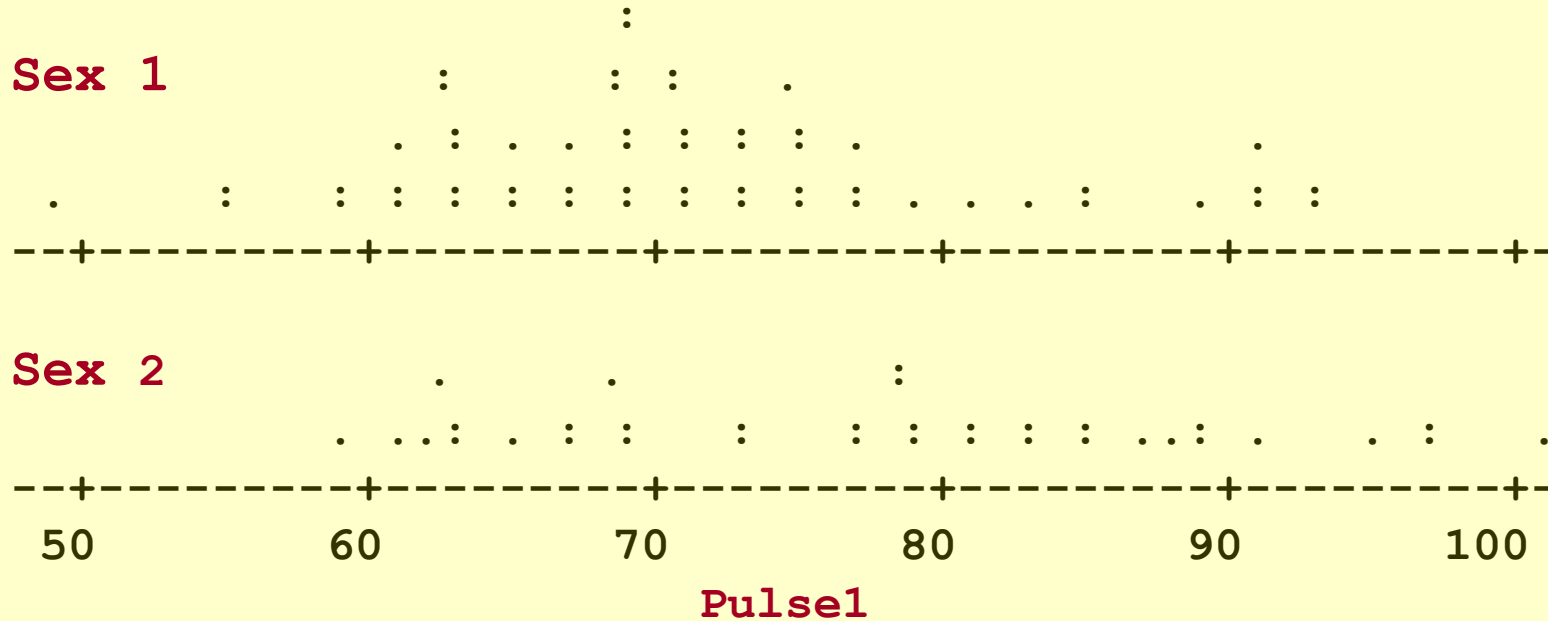
- “Dotplot ”
- “Boxplot ”
- Histograma

DOTPLOT

Dotplot da pulsação em repouso (*PULSE1*)



Dotplot da pulsação em repouso (*PULSE1*) segundo Sexo (*SEX*)

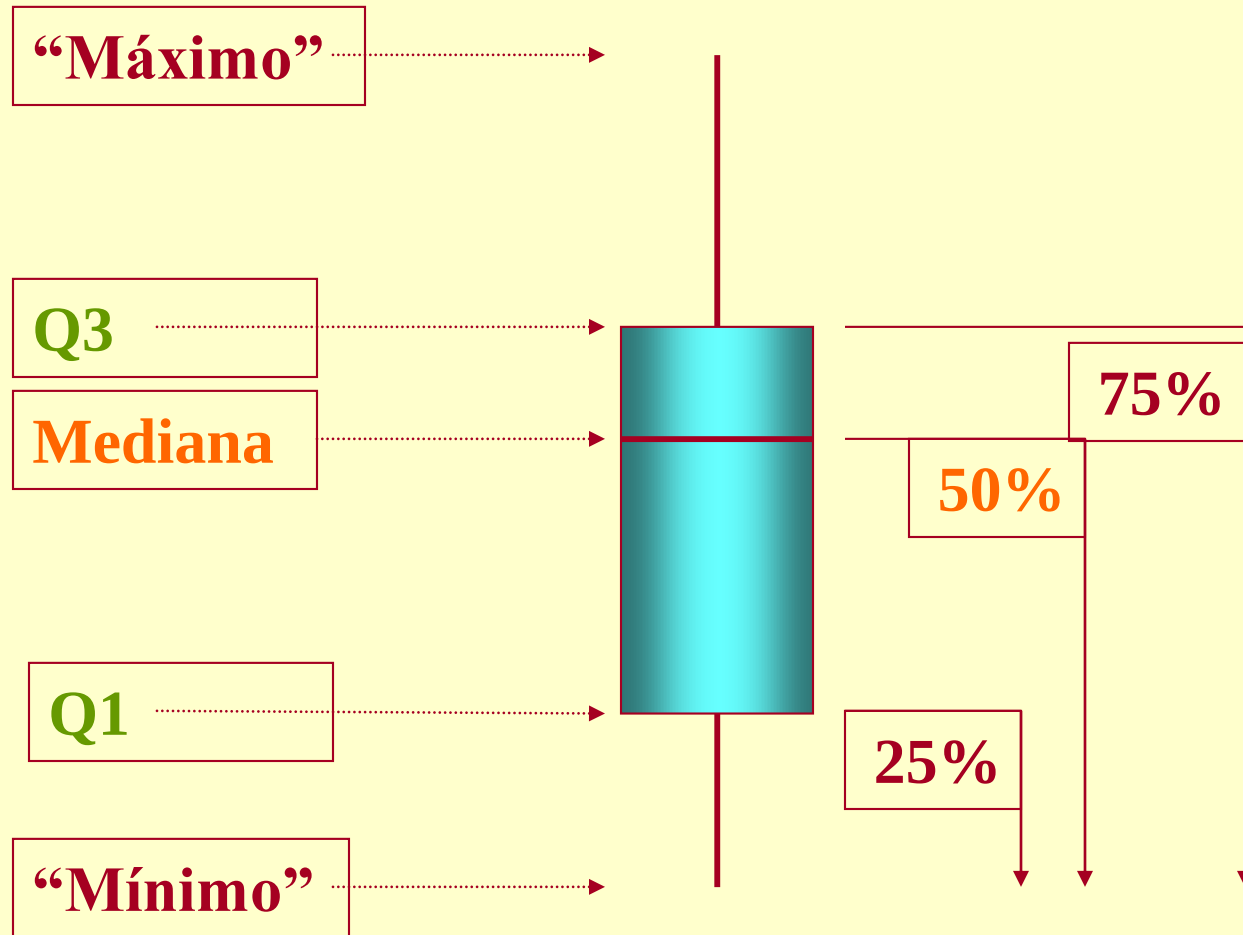


Boxplot

Representa os dados através de um retângulo construído com os **quartis** e fornece informações sobre os **valores extremos**.

Construção

$$LS = Q3 + 1,5(Q3 - Q1)$$



$$LI = Q1 - 1,5(Q3 - Q1)$$

“Máximo” é o maior valor menor que LS ;

“Mínimo” é o menor valor maior que LI .

Exemplo: Tempo de sobrevivência (dias)

Dados ordenados (n=36)

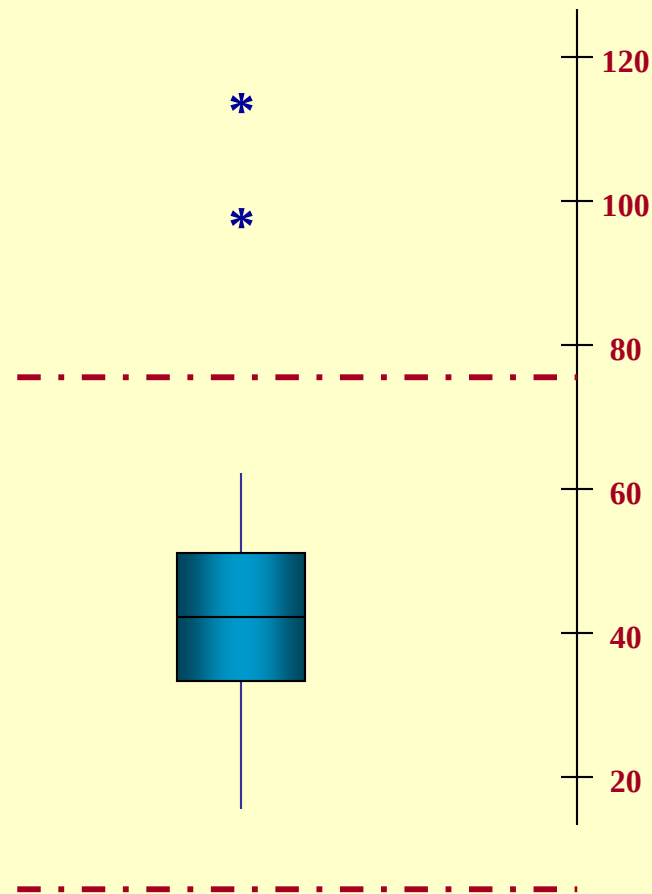
18	21	21	23	23	25
27	29	30	31	32	32
32	34	35	36	38	41
42	42	43	44	45	46
46	47	48	50	54	56
57	58	60	61	98	116

md = 41,5 Q1 = 30,25 Q3 = 49,5

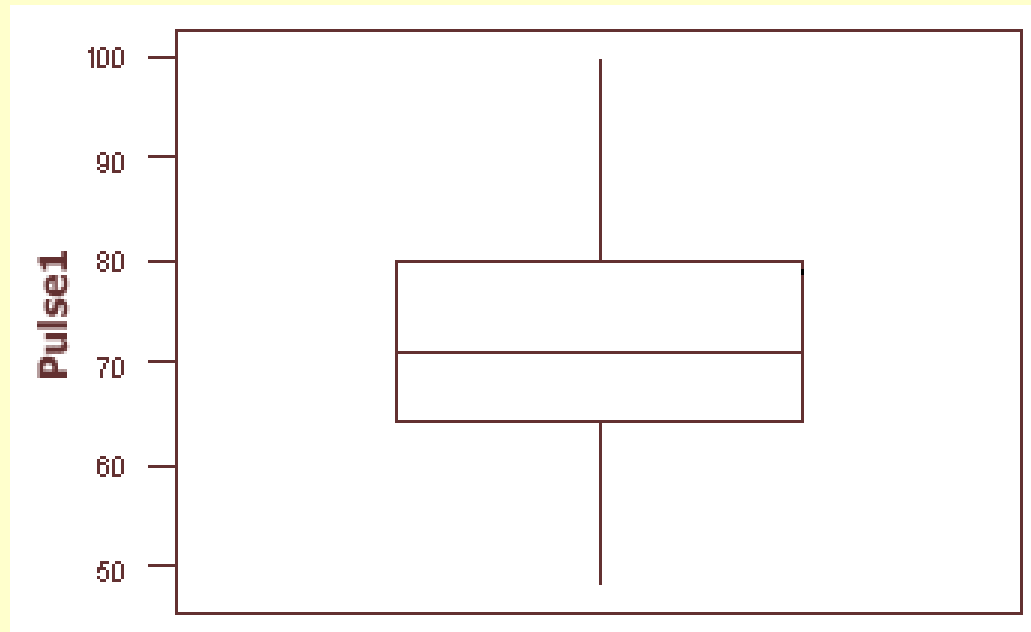
Observações discrepantes?

$$LI = Q1 - 1,5(Q3 - Q1) = 1,38$$

$$LS = Q3 + 1,5(Q3 - Q1) = 78,38$$



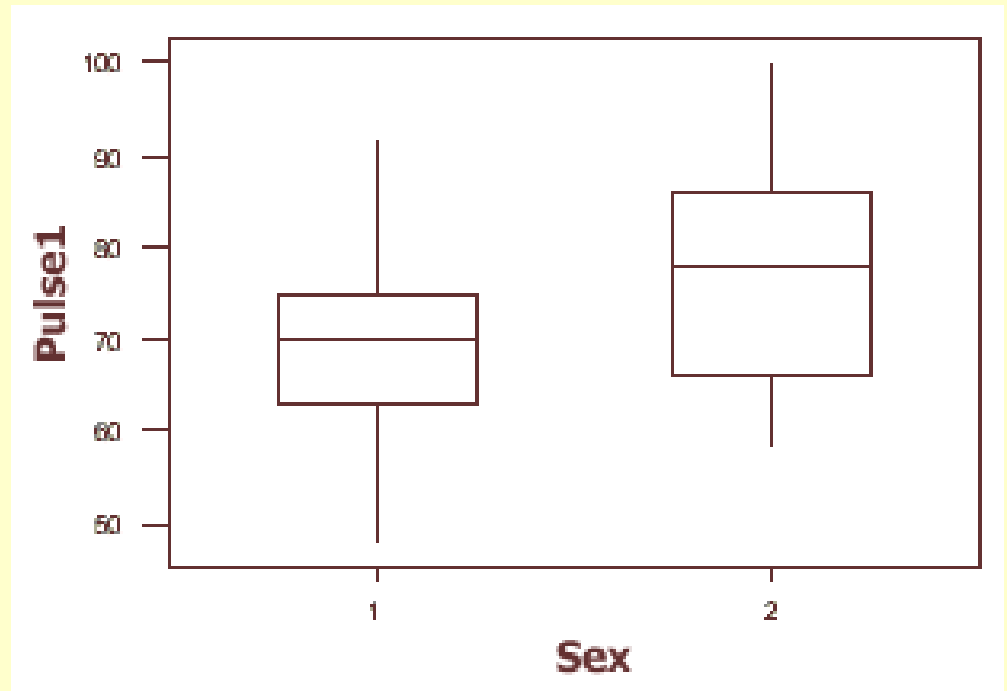
Boxplot da pulsação em repouso (PULSE1)



Alguns Comentários:

- não há observações discrepantes;
- a distribuição dos valores é aproximadamente simétrica.

Boxplots da pulsação em repouso (*PULSE1*) por sexo (*SEX*)



Alguns Comentários:

- não há observações discrepantes;
- as medidas de posição são maiores para o sexo feminino;
- não há fortes evidências de assimetria nos dois grupos.

Histograma

Agrupar os dados em intervalos de classes
(distribuição de frequências)

Bases iguais

Construir um retângulo para cada classe, com base igual ao tamanho da classe e altura proporcional à frequência da classe (f).

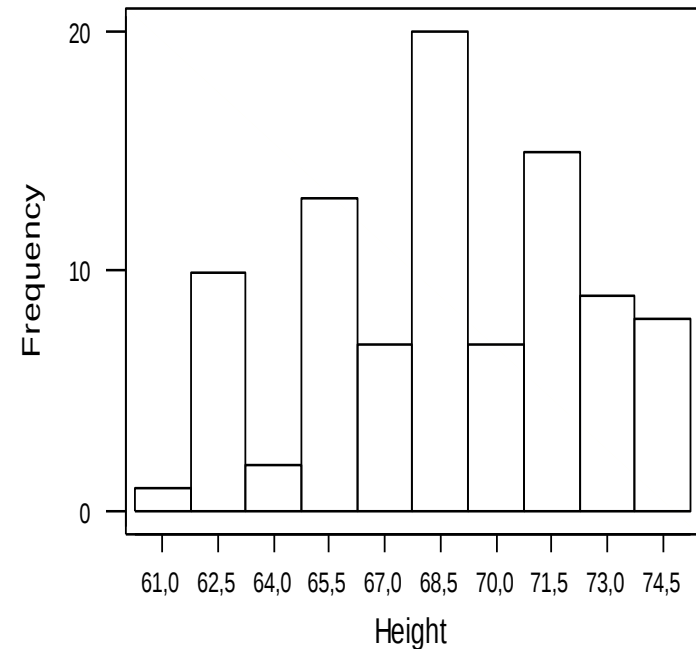
Bases diferentes

Construir um retângulo para cada classe, com base igual ao tamanho da classe e área do retângulo igual a frequência relativa da classe (fr). A altura será dada por $h = fr/base$ (densidade de frequência).

Histograma da altura (*Height*)

Distribuição de frequência para altura

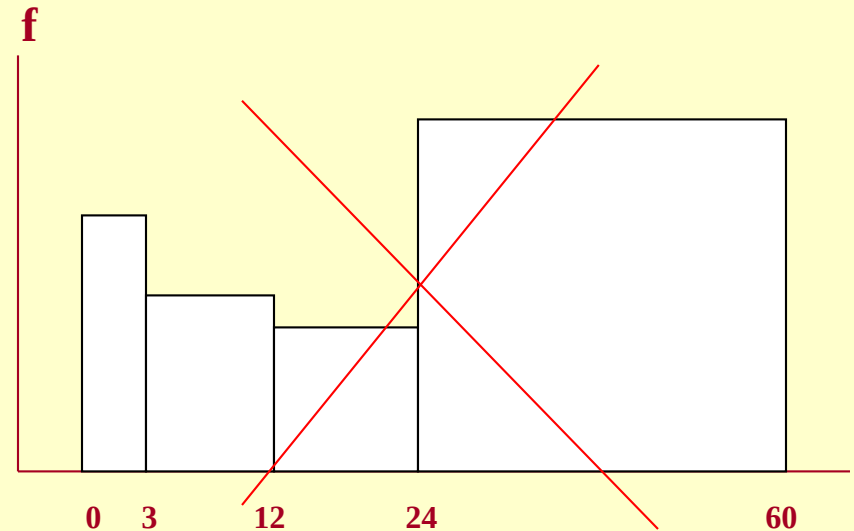
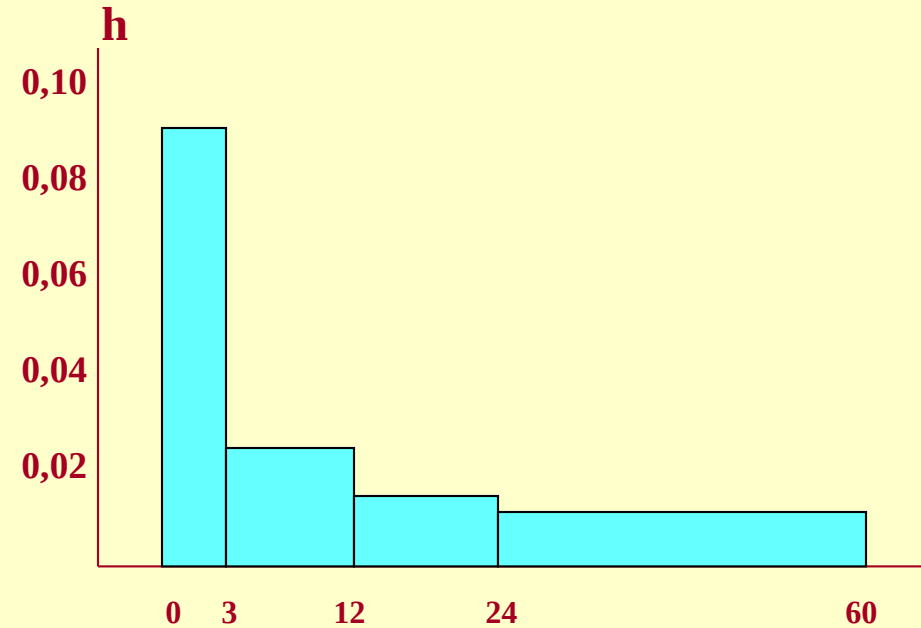
Classe de altura	f	fr
60,25 ┤ 61,75	1	0,011
61,75 ┤ 63,25	10	0,109
63,25 ┤ 64,75	2	0,022
64,75 ┤ 66,25	13	0,141
66,25 ┤ 67,75	7	0,076
67,75 ┤ 69,25	20	0,217
69,25 ┤ 70,75	7	0,076
70,75 ┤ 72,25	15	0,163
72,25 ┤ 73,75	9	0,098
73,75 ┤ 75,25	8	0,087
Total	92	1



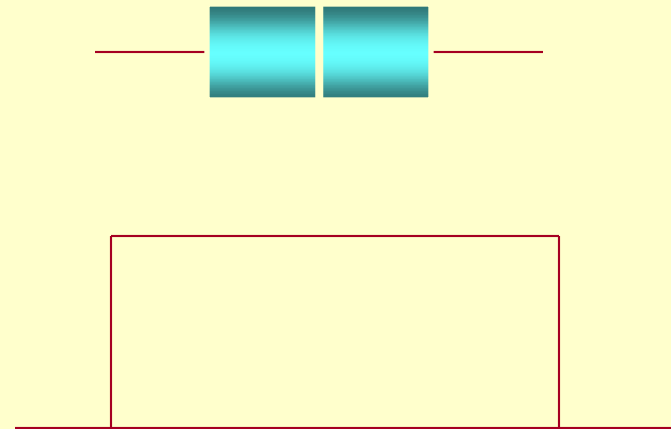
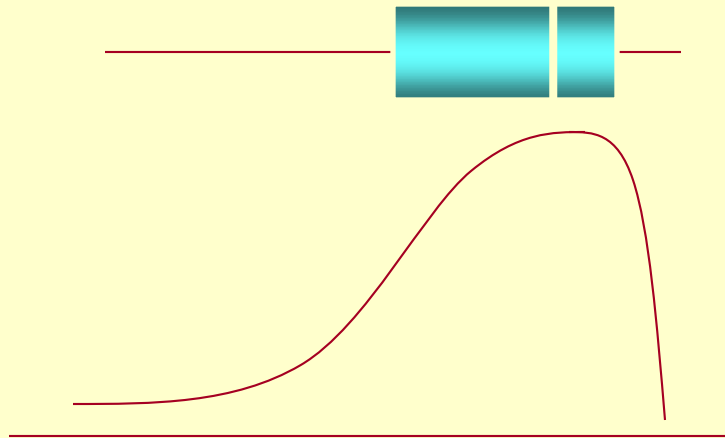
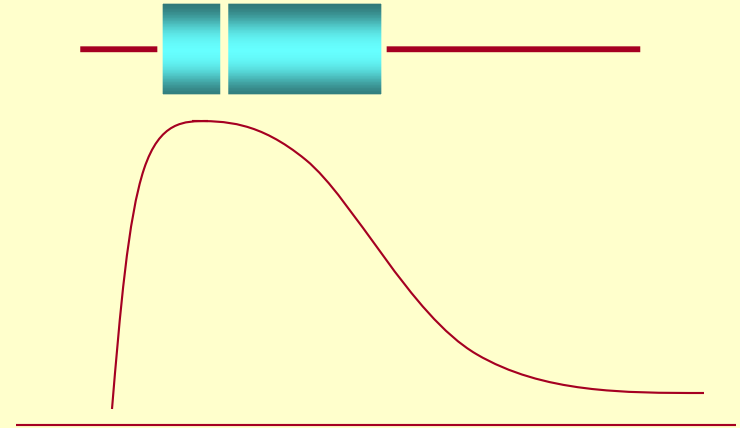
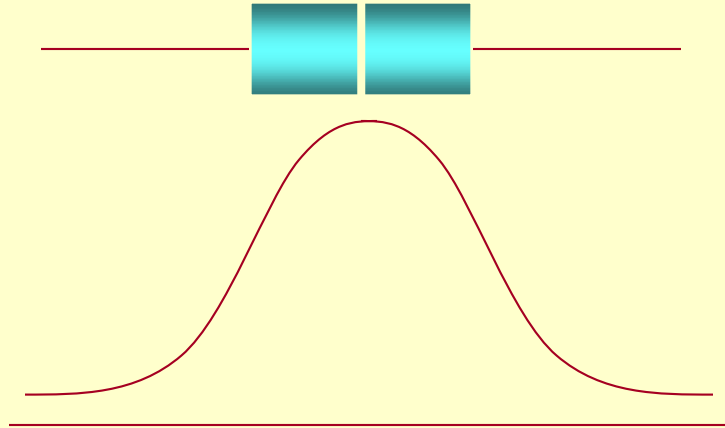
Exemplo: Classes desiguais

Vacinação Infantil

Classes (meses)	f	fr	h
0 - 3	140	0,28	0,093
3 - 12	100	0,20	0,022
12 -24	80	0,16	0,013
24 -60	180	0,36	0,010
Total	500	1,00	



Forma da Distribuição



Variáveis Qualitativas

Os dados podem ser resumidos construindo-se uma tabela de distribuição de frequências, que quantifica a frequência das distintas categorias.

Variáveis qualitativas no banco de dados

Ran

Smokes

Sex

Activity

Variáveis qualitativas no banco de dados

Summary Statistics for Discrete Variables

Sex	Count	Percent	Smokes	Count	Percent
1	57	61,96	1	28	30,43
2	35	38,04	2	64	69,57
N=	92		N=	92	

Activity	Count	Percent
0	1	1,09
1	9	9,78
2	61	66,30
3	21	22,83
N=	92	

Podemos também construir tabelas de frequências conjuntas (*tabelas de contingência*), relacionando duas variáveis qualitativas.

Exemplo 1: Há indícios de associação entre Sexo e Hábito de fumar?

	Hábito de Fumar		
Sexo	Fuma	Não Fuma	Total
Masculino	20	37	57
Feminino	8	27	35
Total	28	64	92

Qual o significado dos valores desta tabela?

Como concluir?

Verificar associação através da:

- porcentagem segundo as colunas, ou
- porcentagem segundo as linhas.

	Hábito de Fumar		
Sexo	Fuma	Não Fuma	Total
Masculino	71,43%	57,81%	61,96%
Feminino	28,57%	42,19%	38,04%
Total	100%	100%	100%

Qual o significado dos valores desta tabela?

Como concluir?

Exemplo 2: Dentre os que correram, qual a porcentagem de mulheres?

Ran\Sex			
	1	2	All
1	24	11	35
	68,57	31,43	100,00
2	33	24	57
	57,89	42,11	100,00
All	57	35	92
	61,96	38,04	100,00

Resposta:

31,43%

Variáveis Qualitativas

Gráficos

- Gráfico de setores
- Gráfico de barras

Gráfico de setores

Um círculo é dividido em tantos setores quantas forem as categorias da variável. A área de cada setor é proporcional à frequência da categoria

Gráfico de setores para a variável *Ran*

Pie Chart of Ran

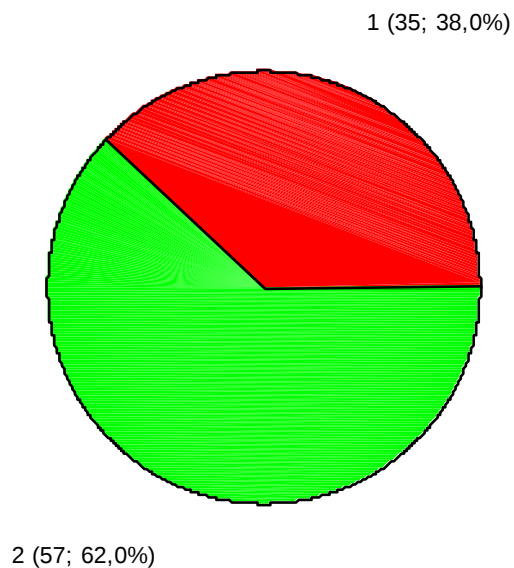


Gráfico de setores para a variável *Activity*

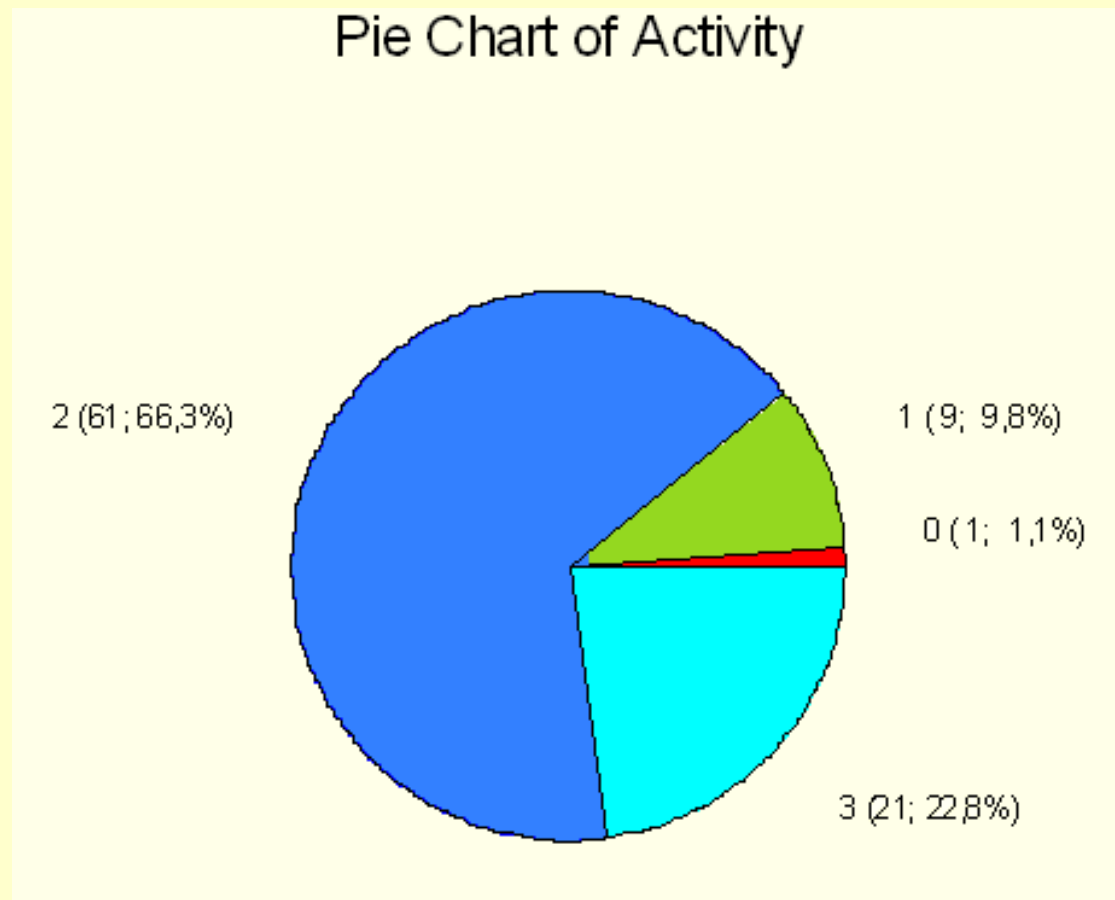


Gráfico de barras

Sobre um eixo, são representados retângulos, um para cada categoria da variável. A altura do retângulo é proporcional à frequência da categoria

Gráfico de barras para a variável *RAN*

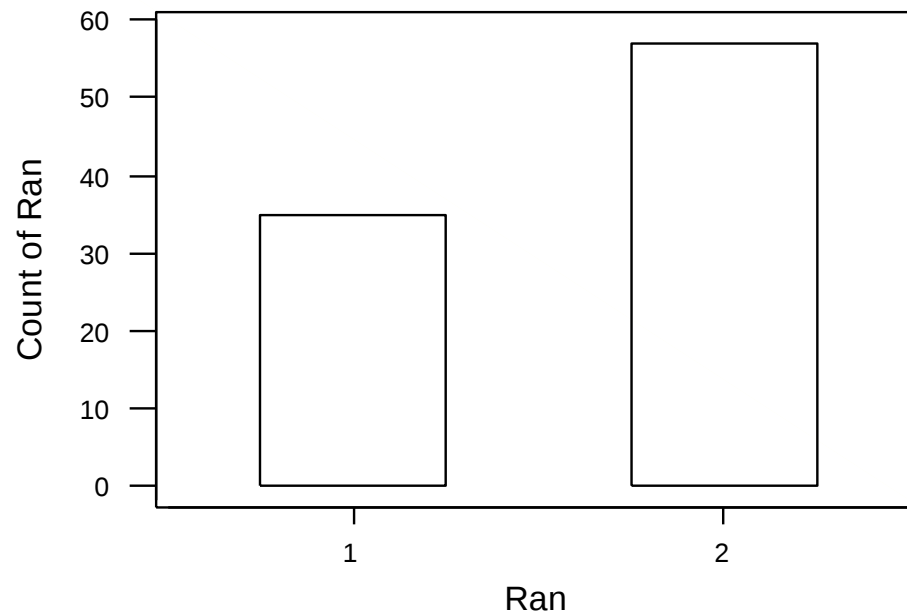


Gráfico de barras para a variável *Activity*

