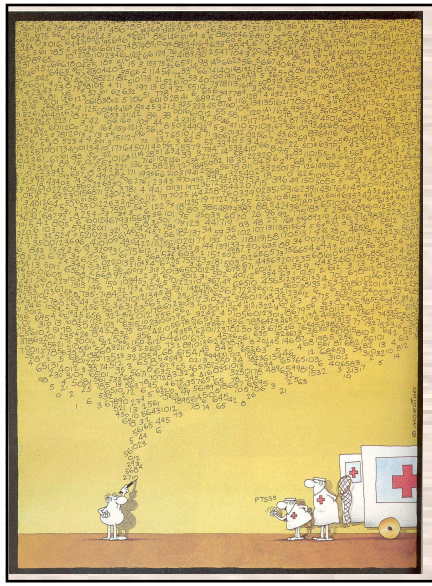


INTRODUÇÃO À ESTATÍSTICA DESCRITIVA

Rui Carvalho Oliveira
roliv@ist.utl.pt

A Estatística é a ciência que trata:

- da recolha
- tratamento
- síntese
- apresentação

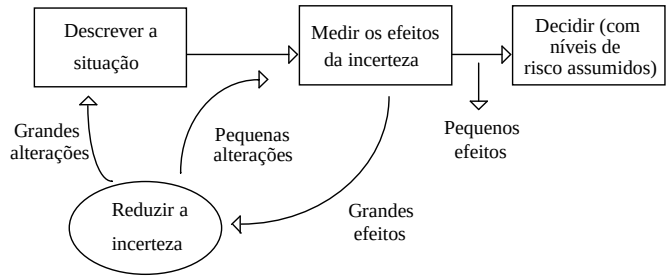
… da informação

A Estatística é a menos exacta das Ciências Exactas!

Incertezas... Incertezas...
Só incertezas!

- A Estatística é uma Ciência Exacta
- Instrumento de
  - ✓ modelação da incerteza
  - ✓ apoio à decisão

A incerteza (como nível de conhecimento) não é atributo da situação decisional mas sim do estado de conhecimento (do decisor) acerca dessa situação



➢ Redução da incerteza
- Reprocessamento da informação disponível
- Aquisição de nova informação

A Estatística é uma ciência para ignorantes !

- ✓ Representação da variabilidade de resultados quando não é conhecida "explicação" dessa variabilidade
- ✓ Utilizada em todos os domínios para os quais o conhecimento dos fenómenos não é suficiente para produzir modelos explicativos

1) "A pluviosidade diária média durante os três últimos meses foi de 13 mm."
2) "Atendendo à pluviosidade registada durante os três últimos meses, este vai ser um ano de seca."

1) Afirmação descritiva, sumarizando a informação disponível
2) Afirmação que transcende o observado, partindo deste para inferir sobre o futuro

Amostra → **População**

- **Estatística Descritiva**
  - ✓ Sumarizar
  - ✓ Descrever a amostra
- **Inferência Estatística**
  - ✓ Generalizar (inferir sobre a população a partir da informação da amostra)
  - ✓ Quantificar e controlar erros
  - ✓ Regular a recolha de informação adicional

Estatística Descritiva

- **Estatística Descritiva:** recolha, organização, apresentação, análise e interpretação de conjuntos de dados

  ↓

  (observações que constituem uma **amostra** extraída de uma fonte, que se designa por **população**)

- **Tratamento e descrição de dados: representações sintéticas visando**
  - i) simplificar e tornar mais atraente a apresentação
  - ii) facilitar a leitura e interpretação
- **Transformar dados em informação**

Estatística Descritiva

- Descrição de resultados das observações:

Variáveis
- Categorias
  - Nominal
  - Ordinal
- Quantidades
  - Discretas (contagem)
  - Contínuas (Medição)

Estatística Descritiva

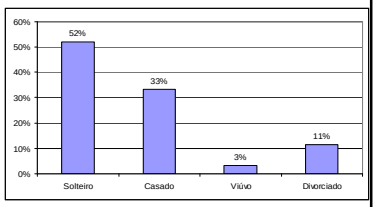
- Tratamento e apresentação de resultados das observações:
  - É habitual sintetizar e apresentar os dados recolhidos através de **Tabelas (de frequência) e Gráficos**

Exemplo: dados sobre estado civil (amostra de 150 indivíduos)

Tabela de frequências

Classes	freq. abs.	freq. rel.
Solteiro	78	0.52
Casado	50	0.33
Viúvo	5	0.03
Divorciado	17	0.12
Total	150	1.00

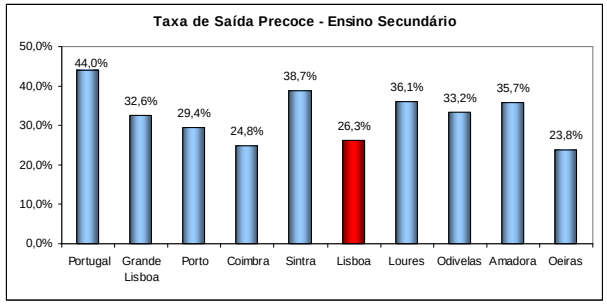
Gráfico de barras



Estatística Descritiva

Apresentação de dados: representações gráficas

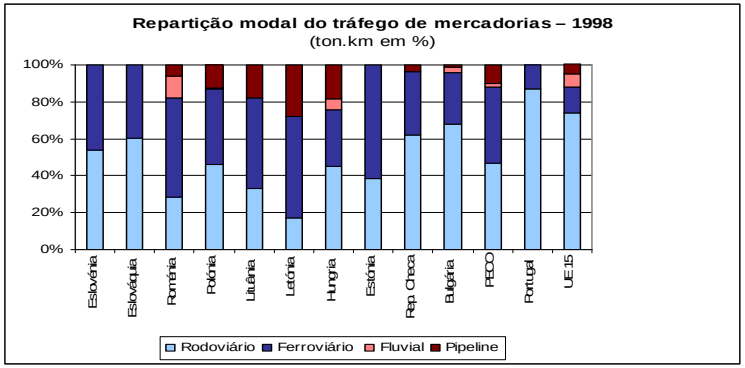
- **Gráficos de barras:**



Estatística Descritiva

Apresentação de dados: representações gráficas

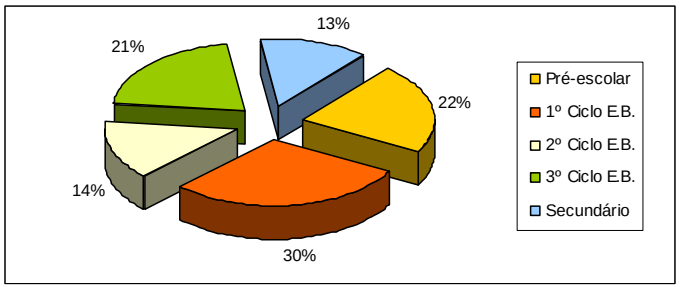
- **Gráficos de barras (acumulado):**



Estatística Descritiva

Apresentação de dados: representações gráficas

- Gráficos de sectores circulares

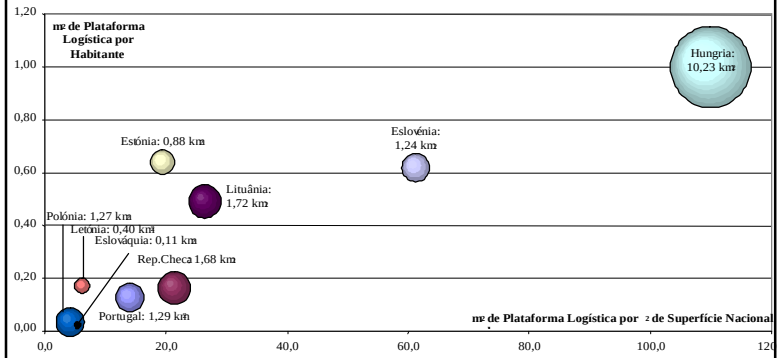


Repartição de alunos por nível e ciclo de ensino

Estatística Descritiva

Apresentação de dados: representações gráficas

- Gráficos de "bolhas"



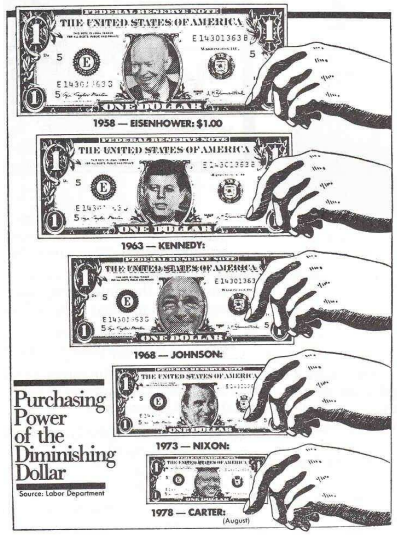
Nota: dimensão da "bolhas" proporcional à área total de Plataformas Logísticas (em Portugal e países do Alargamento -PECO)

Estatística Descritiva

- Uso (e abuso) de gráficos (pictogramas)

  - Kennedy: 94 c
  - Johnson: 83 c
  - Nixon: 64 c
  - Carter: 44 c

  Convenção: poder de compra proporcional ao "comprimento" da nota representada



Estatística Descritiva - Distribuições de frequência

- **Amostra: conjunto de observações (valores numéricos) de uma grandeza**

  Exemplo: observações do nº de acidentes ao longo de 16 semanas:

Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Nº de acidentes	16	15	19	16	17	18	15	16	17	16	17	15	16	17	16	18

  ✓ Variabilidade (aleatória, logo inexplicável e imprevisível) é inerente a muitos fenómenos

- **Distribuição de frequências**: caracterização e descrição sintética do comportamento de uma variável aleatória tirando partido de observações (numéricas) dessa grandeza (amostra)

Estatística Descritiva - Distribuições de frequência

Observações (nº de acidentes ao longo de 16 semanas):

Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Nº de acidentes	16	15	19	16	17	18	15	16	17	16	17	15	16	17	16	18

Nº de acidentes	Frequência simples	
	Absoluta	Relativa
15	3	0,19
16	6	0,38
17	4	0,25
18	2	0,13
19	1	0,06

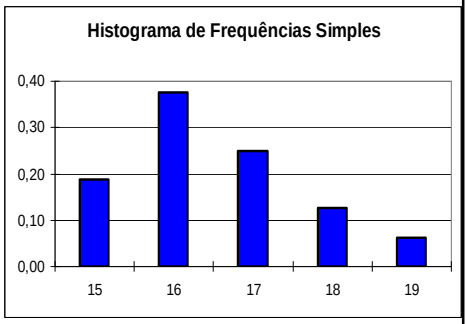
- **Frequência simples** (para um dado valor x da grandeza em estudo):
  - **Absoluta (f'):** número de vezes que o valor x foi observado
  - **Relativa (f):** quociente entre o número de vezes que o valor x foi observado e o número total de observações

Estatística Descritiva - Distribuições de frequência

Observações (nº de acidentes ao longo de 16 semanas):

Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Nº de acidentes	16	15	19	16	17	18	15	16	17	16	17	15	16	17	16	18

Nº de acidentes	Frequência simples	
	Absoluta	Relativa
15	3	0,19
16	6	0,38
17	4	0,25
18	2	0,13
19	1	0,06



Estatística Descritiva - Distribuições de frequência

Observações (nº de acidentes ao longo de 16 semanas):

Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Nº de acidentes	16	15	19	16	17	18	15	16	17	16	17	15	16	17	16	18

Nº de acidentes	Frequência simples		Frequência acumulada	
	Absoluta	Relativa	Absoluta	Relativa
15	3	0,19	3	0,19
16	6	0,38	9	0,56
17	4	0,25	13	0,81
18	2	0,13	15	0,94
19	1	0,06	16	1,00

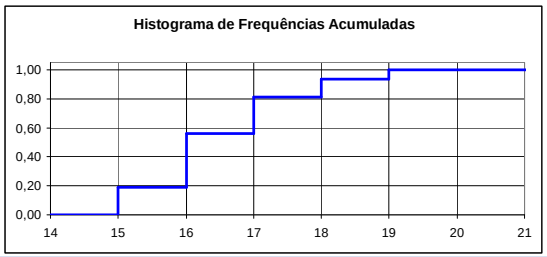
Frequência acumulada (para um dado valor x da grandeza em estudo):

- **Absoluta (F'):** número de vezes que o se observou um valor menor ou igual a x
- **Relativa (F):** quociente entre a frequência acumulada absoluta e o número total de observações

Estatística Descritiva - Distribuições de frequência

Observações (nº de acidentes ao longo de 16 semanas)

Nº de acidentes	Frequência simples		Frequência acumulada	
	Absoluta	Relativa	Absoluta	Relativa
15	3	0,19	3	0,19
16	6	0,38	9	0,56
17	4	0,25	13	0,81
18	2	0,13	15	0,94
19	1	0,06	16	1,00



Estatística Descritiva - Distribuições de frequência

Amostra: medições do tempo de realização de uma tarefa (100 observações)

386	370	371	374	377	435	383	383	422	418
363	388	390	391	392	412	395	395	408	408
396	397	397	397	398	404	403	402	401	400
404	405	405	406	407	396	411	409	394	394
412	414	415	416	416	385	428	424	382	381
413	414	406	416	407	384	426	423	381	380
404	405	415	398	416	396	410	409	394	393
396	397	397	407	399	403	402	402	400	399
387	389	390	391	393	411	395	395	408	407
369	371	373	376	379	429	383	382	420	417

Observações ordenadas

363	369	370	371	371	373	374	376	377	379
380	381	381	382	382	383	383	383	384	385
386	387	388	389	390	390	391	391	392	393
393	394	394	394	395	395	395	395	396	396
396	396	397	397	397	397	397	398	398	399
399	400	400	401	402	402	402	403	403	404
404	404	405	405	405	406	406	407	407	407
407	408	408	408	409	409	410	411	411	412
412	413	414	414	415	415	416	416	416	416
417	418	420	422	423	424	426	428	429	435

Estatística Descritiva - Distribuições de frequência

Observações ordenadas

363	369	370	371	371	373	374	376	377	379
380	381	381	382	382	383	383	383	384	385
386	387	388	389	390	390	391	391	392	393
393	394	394	394	395	395	395	395	396	396
396	396	397	397	397	397	397	398	398	399
399	400	400	401	402	402	402	403	403	404
404	404	405	405	405	406	406	407	407	407
407	408	408	408	409	409	410	411	411	412
412	413	414	414	415	415	416	416	416	416
417	418	420	422	423	424	426	428	429	435

Tratamento de dados:

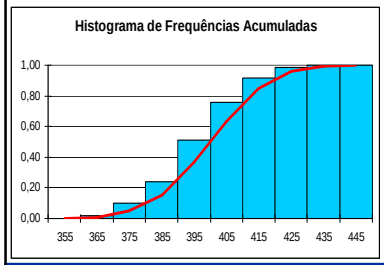
8 classes de amplitude 10 (abertas à direita)

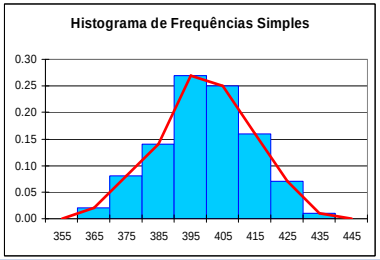
Muitas vezes (nomeadamente quando a grandeza é contínua) as frequências são apuradas não para valores singulares da grandeza mas sim para intervalos de variação (classes)

Limites das classes		Freq.ia simples		Freq.ia acumulada	
Inferior	Superior	Absoluta	Relativa	Absoluta	Relativa
360	369.9	2	0,02	2	0,02
370	379.9	8	0,08	10	0,10
380	389.9	14	0,14	24	0,24
390	399.9	27	0,27	51	0,51
400	409.9	25	0,25	76	0,76
410	419.9	16	0,16	92	0,92
420	429.9	7	0,07	99	0,99
430	439.9	1	0,01	100	1,00

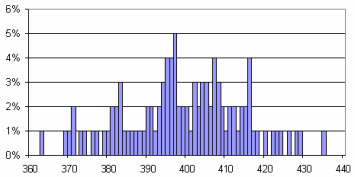
Estatística Descritiva Distribuições de frequência

Limites das classes		Freq.ia simples		Freq.ia acumulada	
Inferior	Superior	Absoluta	Relativa	Absoluta	Relativa
360	370	2	0,02	2	0,02
370	380	8	0,08	10	0,10
380	390	14	0,14	24	0,24
390	400	27	0,27	51	0,51
400	410	25	0,25	76	0,76
410	420	16	0,16	92	0,92
420	430	7	0,07	99	0,99
430	440	1	0,01	100	1,00

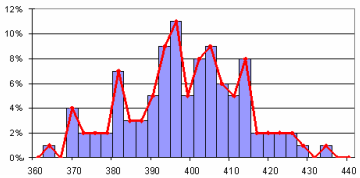




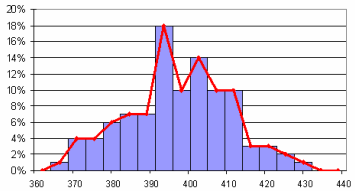
Estatística Descritiva - Histogramas



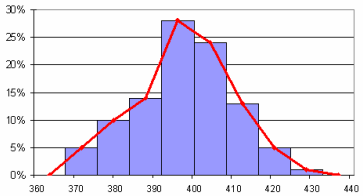
Amplitude: 1 **nº classes: 71**
Média: 398,74 **desvio padrão: 14,77**



Amplitude: 3 **nº classes: 25**
Média: 398,73 **desvio padrão: 14,79**

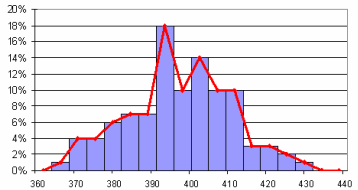


Amplitude: 5 **nº classes: 15**
Média: 398,70 **desvio padrão: 14,88**

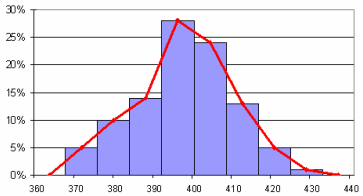


Amplitude: 10 **nº classes: 8**
Média: 399,6 **desvio padrão: 14,52**

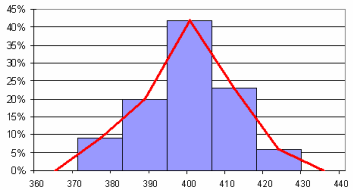
Estatística Descritiva - Histogramas



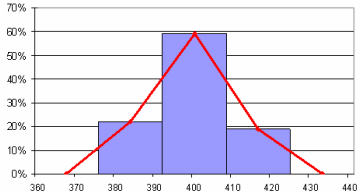
Amplitude: 5 **nº classes: 15**
Média: 398,70 **desvio padrão: 14,88**



Amplitude: 10 **nº classes: 8**
Média: 399,6 **desvio padrão: 14,52**

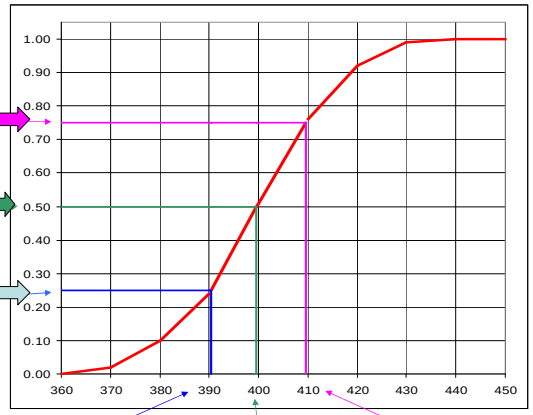


Amplitude: 15 **nº classes: 5**
Média: 399,00 **desvio padrão: 15,44**



Amplitude: 25 **nº classes: 3**
Média: 398,75 **desvio padrão: 16,01**

Estatística Descritiva



➢ **Quartis**
valores da variável que dividem a distribuição de frequências em 4 partes iguais

Quartil de 25% : $q_{25\%}$= 390.4

Mediana = Quartil de 50% : $q_{50\%}$= 399.6

Quartil de 75% : $q_{75\%}$= 409.6

Estatística Descritiva

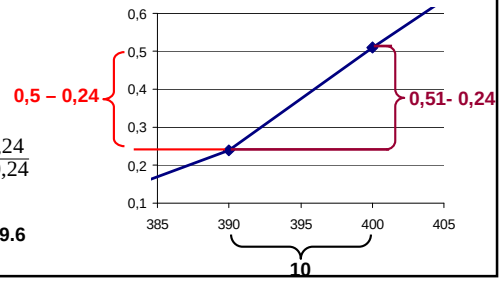
Limites das classes		Freq.ia simples		Freq.ia acumulada	
Inferior	Superior	Absoluta	Relativa	Absoluta	Relativa
360	370	2	0,02	2	0,02
370	380	8	0,08	10	0,10
380	**390**	14	0,14	24	**0,24**
390	**400**	27	0,27	51	**0,51**
400	410	25	0,25	76	0,76
410	420	16	0,16	92	0,92
420	430	7	0,07	99	0,99
430	440	1	0,01	100	1,00

Quartil de 50% está entre 390 e 400

Interpolação linear:

$$q_{50\%} = 390 + 10 \times \frac{0,5 - 0,24}{0,51 - 0,24}$$

Quartil de 50% : $q_{50\%}$ **= 399.6**



Estatística Descritiva – Medidas de tendência não central

➢ **Quartis:** valores da variável que dividem a distribuição de frequências em 4 partes iguais

➢ **Q_1** (quartil de 25%) – valor da variável para o qual a frequência acumulada atinge 25%

Q_1 = 37 ½

➢ **Q_2** (quartil de 50%) – idem, para 50%

Q_2 = 38 ½

➢ **Q_3** (quartil de 75%) – idem, para 75%

Q_3 = 39

Tamanho dos sapatos	Frequências absolutas		Frequências absolutas	
	Simples	Acum.	Simples	Acum.
35	**30**	**30**	**0,7%**	**0,7%**
35 1/2	**40**	**70**	**0,9%**	**1,5%**
36	**50**	**120**	**1,1%**	**2,6%**
36 1/2	**150**	**270**	**3,3%**	**5,9%**
37	**300**	**570**	**6,5%**	**12,4%**
37 1/2	**600**	**1170**	**13,0%**	**25,4%**
38	**950**	**2120**	**20,6%**	**46,0%**
38 1/2	**820**	**2940**	**17,8%**	**63,8%**
39	**750**	**3690**	**16,3%**	**80,1%**
39 1/2	**440**	**4130**	**9,5%**	**89,6%**
40	**250**	**4380**	**5,4%**	**95,0%**
40 1/2	**150**	**4530**	**3,3%**	**98,3%**
41	**40**	**4570**	**0,9%**	**99,2%**
41 1/2	**39**	**4609**	**0,8%**	**100,0%**

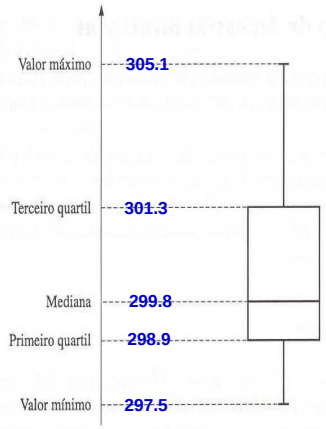
Estatística Descritiva – Medidas de tendência não central

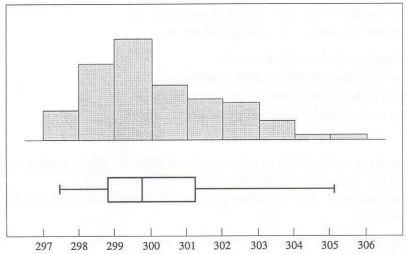
• **Decis:** valores da variável que dividem a distribuição de frequências em 10 partes iguais

• **Percentil (ou quantil) de α %:** valor da variável para o qual a frequência acumulada atinge α%

Tamanho dos sapatos	Frequências absolutas		Frequências relativas	
	Simples	Acum.	Simples	Acum.
35	**30**	**30**	**0,7%**	**0,7%**
35 1/2	**40**	**70**	**0,9%**	**1,5%**
36	**50**	**120**	**1,1%**	**2,6%**
36 1/2	**150**	**270**	**3,3%**	**5,9%**
37	**300**	**570**	**6,5%**	**12,4%**
37 1/2	**600**	**1170**	**13,0%**	**25,4%**
38	**950**	**2120**	**20,6%**	**46,0%**
38 1/2	**820**	**2940**	**17,8%**	**63,8%**
39	**750**	**3690**	**16,3%**	**80,1%**
39 1/2	**440**	**4130**	**9,5%**	**89,6%**
40	**250**	**4380**	**5,4%**	**95,0%**
40 1/2	**150**	**4530**	**3,3%**	**98,3%**
41	**40**	**4570**	**0,9%**	**99,2%**
41 1/2	**39**	**4609**	**0,8%**	**100,0%**

Estatística Descritiva - Diagramas de “caixa e bigodes”





Exemplo: selecção de fornecedor de baterias

Dados: distribuição da duração (vida útil, em horas) das baterias

Fornecedor A

Média=201h
D. padrão= 60h

Vida útil (hoas)		Frequências relativas	
de [	a [	Simples	Acum.
100	120	1,6%	1,6%
120	140	7,8%	9,4%
140	160	15,8%	25,2%
160	180	19,6%	44,8%
180	200	16,4%	61,2%
200	220	9,8%	71,0%
220	240	8,6%	79,6%
240	260	5,8%	85,4%
260	280	4,2%	89,6%
280	300	2,8%	92,4%
300	320	2,6%	95,0%
320	340	1,4%	96,4%
340	360	1,2%	97,6%
360	380	0,8%	98,4%
380	400	0,6%	99,0%
400	420	0,4%	99,4%
420	440	0,4%	99,8%
440	460	0,2%	100,0%

Fornecedor B

Média= 199h
D. padrão=30h

Vida útil (horas)		Frequências relativas	
de [	a [	Simples	Acum.
100	110	0,2%	0,2%
110	120	0,6%	0,8%
120	130	0,6%	1,4%
130	140	1,6%	3,0%
140	150	2,2%	5,2%
150	160	4,2%	9,4%
160	170	8,0%	17,4%
170	180	11,2%	28,6%
180	190	10,6%	39,2%
190	200	12,2%	51,4%
200	210	12,4%	63,8%
210	220	11,6%	75,4%
220	230	10,0%	85,4%
230	240	5,2%	90,6%
240	250	5,2%	95,8%
250	260	2,6%	98,4%
260	270	1,2%	99,6%
270	280	0,2%	99,8%
280	290	0,2%	100,0%

Exemplo: selecção de fornecedor de baterias

a) Política de manutenção: substituição da bateria cada 200 horas. Risco de falha?

Fornecedor A

Risco de falha =
61.2 %

Vida útil (hoas)		Frequências relativas	
de [	a [	Simples	Acum.
100	120	1,6%	1,6%
120	140	7,8%	9,4%
140	160	15,8%	25,2%
160	180	19,6%	44,8%
180	200	16,4%	61,2%
200	220	9,8%	[illegible]
220	240	8,6%	[illegible]
240	260	5,8%	85,4%
260	280	4,2%	89,6%
280	300	2,8%	92,4%
300	320	2,6%	95,0%
320	340	1,4%	96,4%
340	360	1,2%	97,6%
360	380	0,8%	98,4%
380	400	0,6%	99,0%
400	420	0,4%	99,4%
420	440	0,4%	99,8%
440	460	0,2%	100,0%

Fornecedor B

Risco de falha =
51.4%

Vida útil (horas)		Frequências relativas	
de [	a [	Simples	Acum.
100	110	0,2%	0,2%
110	120	0,6%	0,8%
120	130	0,6%	1,4%
130	140	1,6%	3,0%
140	150	2,2%	5,2%
150	160	4,2%	9,4%
160	170	8,0%	17,4%
170	180	11,2%	28,6%
180	190	10,6%	39,2%
190	200	12,2%	51,4%
200	[illegible]	12,4%	[illegible]
210	220	11,6%	75,4%
220	230	10,0%	85,4%
230	240	5,2%	90,6%
240	250	5,2%	95,8%
250	260	2,6%	98,4%
260	270	1,2%	99,6%
270	280	0,2%	99,8%
280	290	0,2%	100,0%

Exemplo: selecção de fornecedor de baterias

b) Risco de falha: max. de 10%. Intervalo de tempo entre substituições?

Fornecedor A

Intervalo =
140 horas

Vida útil (hoas)		Frequências relativas	
de [	a [	Simples	Acum.
100	120	1,6%	1,6%
120	140	7,8%	9,4%
140	160	15,8%	25,2%
160	180	19,6%	44,8%
180	200	16,4%	61,2%
200	220	9,8%	71,0%
220	240	8,6%	79,6%
240	260	5,8%	85,4%
260	280	4,2%	89,6%
280	300	2,8%	92,4%
300	320	2,6%	95,0%
320	340	1,4%	96,4%
340	360	1,2%	97,6%
360	380	0,8%	98,4%
380	400	0,6%	99,0%
400	420	0,4%	99,4%
420	440	0,4%	99,8%
440	460	0,2%	100,0%

Fornecedor B

Intervalo =
160 horas

Vida útil (horas)		Frequências relativas	
de [	a [	Simples	Acum.
100	110	0,2%	0,2%
110	120	0,6%	0,8%
120	130	0,6%	1,4%
130	140	1,6%	3,0%
140	150	2,2%	5,2%
150	160	4,2%	9,4%
160	170	8,0%	17,4%
170	180	11,2%	28,6%
180	190	10,6%	39,2%
190	200	12,2%	51,4%
200	210	12,4%	63,8%
210	220	11,6%	75,4%
220	230	10,0%	85,4%
230	240	5,2%	90,6%
240	250	5,2%	95,8%
250	260	2,6%	98,4%
260	270	1,2%	99,6%
270	280	0,2%	99,8%
280	290	0,2%	100,0%

Exemplo: selecção de fornecedor de baterias

c) Baterias "Muito Boas": duram pelo menos 260 horas.

Fornecedor A

Vida útil (hoas)		Frequências relativas	
de [	a [	Simples	Acum.
100	120	1,6%	1,6%
120	140	7,8%	9,4%
140	160	15,8%	25,2%
160	180	19,6%	44,8%
180	200	16,4%	61,2%
200	220	9,8%	71,0%
220	240	8,6%	79,6%
240	260	5,8%	85,4%
260	28[illegible]	4,2%	[illegible]
280	300	2,8%	92,4%
300	320	2,6%	95,0%
320	340	1,4%	96,4%
340	360	1,2%	97,6%
360	380	0,8%	98,4%
380	400	0,6%	99,0%
400	420	0,4%	99,4%
420	440	0,4%	99,8%
440	460	0,2%	100,0%

% de baterias "muito boas" = 1-0.854

14.6%

Fornecedor B

Vida útil (horas)		Frequências relativas	
de [	a [	Simples	Acum.
100	110	0,2%	0,2%
110	120	0,6%	0,8%
120	130	0,6%	1,4%
130	140	1,6%	3,0%
140	150	2,2%	5,2%
150	160	4,2%	9,4%
160	170	8,0%	17,4%
170	180	11,2%	28,6%
180	190	10,6%	39,2%
190	200	12,2%	51,4%
200	210	12,4%	63,8%
210	220	11,6%	75,4%
220	230	10,0%	85,4%
230	240	5,2%	90,6%
240	250	5,2%	95,8%
250	260	2,6%	98,4%
260	2[illegible]	1,2%	[illegible]
270	280	0,2%	99,8%
280	290	0,2%	100,0%

% de baterias "muito boas" = 1-0.984

1.6%

Síntese de Informação

Parâmetros: números singulares que evidenciam determinadas propriedades das distribuições de frequência

i. **Medidas de tendência central** (localização do "centro da distribuição");

ii. **Medidas de dispersão** (grau de variação dos valores em torno do ponto central);

iii. **Medidas de assimetria** (grau de simetria em relação ao ponto central).

Medidas de tendência central - **Média**

Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Nº de acidentes	16	15	19	16	17	18	15	16	17	16	17	15	16	17	16	18

Nº de acidentes	Frequência simples		Frequência acumulada	
	Absoluta	Relativa	Absoluta	Relativa
15	**3**	**0,1875**	**3**	**0,19**
16	**6**	**0,375**	**9**	**0,56**
17	**4**	**0,25**	**13**	**0,81**
18	**2**	**0,125**	**15**	**0,94**
19	**1**	**0,0625**	**16**	**1,00**

Média = $\overline{X} = (16+15+19+16+17+18+15+16+17+16+17+15+16+17+16+18)/16 = 16.5$

(usando dados originais, sem tratamento)

$\overline{X}$ **= (3 x 15 + 6 x 16 + 4 x 17 + 2 x 18 + 1 x 19) / 16**

(usando frequências absolutas, após tratamento estatístico dos dados)

$\overline{X}$ **= (0.1875 x 15 + 0.375 x 16 + 0.25 x 17 + 0.125 x 18 + 0.0625 x 19)**

(usando frequências relativas, após tratamento estatístico dos dados)

Medidas de tendência central - Média

Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Nº de acidentes	16	15	19	16	17	18	15	16	17	16	17	15	16	17	16	18

Média = $\overline{X} = (16+15+19+16+17+18+15+16+17+16+17+15+16+17+16+18)/16 = 16.5$

(usando dados originais, sem tratamento)

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

n – número de observações

X_i – valor da i-ésima observação (nº de acidentes na semana i, neste exemplo)

Medidas de tendência central - Média

Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Nº de acidentes	16	15	19	16	17	18	15	16	17	16	17	15	16	17	16	18

Nº de acidentes	Frequência simples		Frequência acumulada	
	Absoluta	Relativa	Absoluta	Relativa
15	3	0,1875	3	0,19
16	6	0,375	9	0,56
17	4	0,25	13	0,81
18	2	0,125	15	0,94
19	1	0,0625	16	1,00

$\overline{X}$ = (0.1875 x 15 + 0.375 x 16 + 0.25 x 17 + 0.125 x 18 + 0.0625 x 19)

(usando frequências relativas, após tratamento estatístico dos dados)

$$\overline{X} = \sum_{k=1}^{c} f_k \cdot X_k$$

c – número de classes

f_k – frequência relativa da k-ésima classe

X_k – valor (médio) da k-ésima classe

Nota: quando classes contêm mais do que um valor da grandeza, esta expressão produz uma aproximação

Medidas de tendência central - Média

Cálculo da média para dados tratados (usando frequências relativas de ocorrência)

$$\overline{X} = \sum_{k=1}^{c} f_k \cdot X_k$$

c – número de classes

f_k – frequência relativa da k-ésima classe

X_k – valor (médio) da k-ésima classe

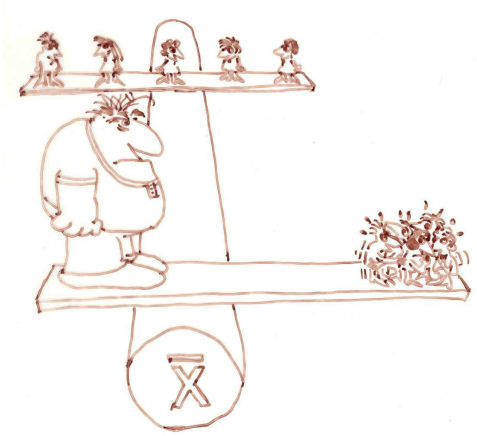
Limites das classes		Marca da classe	Freq.ia simples		Freq.ia acumulada	
Inferior	Superior		Absoluta	Relativa	Absoluta	Relativa
360	370	365	2	0,02	2	0,02
370	380	375	8	0,08	10	0,1
380	390	385	14	0,14	24	0,24
390	400	395	27	0,27	51	0,51
400	410	405	25	0,25	76	0,76
410	420	415	16	0,16	92	0,92
420	430	425	7	0,07	99	0,99
430	440	435	1	0,01	100	1

Média = 0.02 x 365 + 0.08 x 375 + 0.14 x 385 + 0.27 x 395 + 0.25 x 405 +
+ 0.16 x 415 + 0.07 x 425 + 0.01 x 435 **= 399.6**

(Nota: Média = 399.17 para dados originais, não tratados**)**

Estatística Descritiva – **Medidas de centralidade**

A média corresponde ao "centro de gravidade" da distribuição de frequências



Medidas de tendência central - **Mediana**

Distribuição de salários numa organização com 25 colaboradores:

Salário	Nº de colaboradores	Frequência relativa (%)	Frequência acumulada
600	12	48%	48%
800	1	4%	52%
1.200	5	20%	72%
2.000	3	12%	84%
5.000	2	8%	92%
10.000	1	4%	96%
20.000	1	4%	100%

- **Média:** $\overline{X} = 2400$

- **Mediana (Med): valor da grandeza tal que há um número igual de observações abaixo e acima desse valor mediano (divide a distribuição de frequência em duas partes iguais)**

Med = 800

Medidas de tendência central - Mediana

Notas:

1) **Quando o nº de observações é par, há dois valores medianos, convencionando-se que a mediana é a média aritmética desses dois valores**

Dados : 5 ; 8 ; 14 ; 21 ⟶ Med = (8+14)/2 = 11

2) **Para dados agrupados em classes, pode falar-se de "classe mediana" que corresponde à primeira classe para a qual a frequência acumulada é igual ou superior a 50%**

Limites das classes		Freq.ia simples		Freq.ia acumulada		
Inferior	Superior	Absoluta	Relativa	Absoluta	Relativa	
360	370	2	0,02	2	0,02	
370	380	8	0,08	10	0,10	
380	390	14	0,14	24	0,24	
390	**400**	27	0,27	51	**0,51**	⇨ Classe mediana
400	410	25	0,25	76	0,76	
410	420	16	0,16	92	0,92	
420	430	7	0,07	99	0,99	
430	440	1	0,01	100	1,00	

Medidas de tendência central - Moda

Salário	Nº de colaboradores	Frequência relativa (%)	Frequência acumulada
600	12	48%	48%
800	1	4%	52%
1.200	5	20%	72%
2.000	3	12%	84%
5.000	2	8%	92%
10.000	1	4%	96%
20.000	1	4%	100%

Moda (Mod): corresponde ao valor da grandeza com maior frequência de ocorrência

Mod = 600

Nota: para dados tratados (agrupados em classes), pode falar-se de "classe modal"

Medidas de tendência central

- **Propriedades da Média** $(\overline{X})$

$$\sum_{i=1}^{n}(X_i - \overline{X}) = 0$$ → Soma dos desvios (em relação à média) é nula (desvios positivos "compensam" desvios negativos)

$$(X_1 - \overline{X}) + (X_2 - \overline{X}) + \ldots + (X_n - \overline{X}) = 0$$

$\overline{X}$ corresponde ao "centro da gravidade" dos valores observados

Requer dados quantitativos

- **Propriedades da Mediana :**
  - Mediana corresponde a um "**centro posicional**", quando as observações são ordenadas do menor para o maior valor, não interessando o valor numérico de cada observação mas apenas a sua posição nessa ordenação.
  - É, portanto, aplicável a dados qualitativos (sem métrica), desde que expressos numa escala ordinal.

Medidas de tendência central

- **Propriedades da moda (Mod):**

• É aplicável a todos os tipos de dados (qualitativos ou quantitativos), mesmo quando estes são puramente nominais.

• Quando os dados são quantitativos e estão classificados em classes (ou são qualitativos), usa-se o termo classe (ou categoria) modal.

• Pode haver mais de uma moda ou classe (ou categoria) modal

Comparação das medidas de tendência central

➢**Média** requer dados quantitativos, enquanto a **mediana** pode ser aplicada a dados qualitativos ordinais e a **moda** é até aplicável a simples categorias

➢**Média** é sensível aos valores numéricos das observações, sendo afectada por alterações dos casos extremos, o que não sucede com a **mediana**:

57 ; 69 ; 72 ; 81 ; 86 $\rightarrow \begin{array}{l} \overline{X} = 73 \\ Med = 72 \end{array}$ 57 ; 69 ; 72 ; 81 ; 961 $\rightarrow \begin{array}{l} \overline{X} = 248 \\ Med = 72 \end{array}$

✓No caso de distribuições muito assimétricas, a média é "puxada" para o lado em que a "cauda" da distribuição é mais estendida (com valores extremos mais afastados dos "casos típicos" ou mais frequentes), podendo dar indicações enganadoras quanto à localização do "centro" da distribuição.

➢ **Média** usa toda a informação disponível (nomeadamente, de carácter numérico), enquanto que à **mediana** apenas importa a posição relativa das observações.

➢ **Mediana** é mais sensível aos resultados amostrais, apresentando maiores variações de amostra para amostra

Comparação das medidas de tendência central

Salário	Nº de colaboradores	Frequência relativa (%)	Frequência acumulada
600	12	48%	48%
800	1	4%	52%
1.200	5	20%	72%
2.000	3	12%	84%
5.000	2	8%	92%
10.000	1	4%	96%
20.000	1	4%	100%

Média = 2400 Mediana=800 Moda=600

Salário	Nº de colabor.	Freq. Relat. (%)	Freq. Acumul.
600	12	48%	48%
800	1	4%	52%
1.200	5	20%	72%
2.000	3	12%	84%
3.000	2	8%	92%
4.000	1	4%	96%
5.000	1	4%	100%

Média = 1400 Mediana=800 Moda=600

➢ **Média é sensível aos valores (numéricos) das observações, particularmente a "casos extremos"**

➢ O mesmo não sucede com a **mediana (desde que o "valor central" na ordenação das observações não sofra alteração)**

Estatística Descritiva – **Medidas de dispersão**

➢**Domínio** (diferença entre o máximo e o mínimo das observações)

✓Crítica: demasiado sensível a valores extremos (e eventualmente "atípicos")

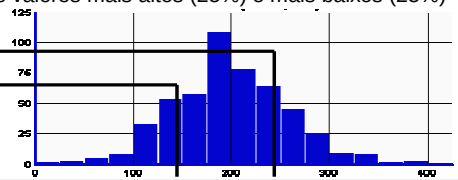
Nº de ordem ->	1º	2º	3º	.....	495º	496º	497º	498º	499º	500º
Observações ->	18	41	63	.....	335	338	352	380	385	421

➢ **Diferença inter-quartílica** : diferença entre os quartis de 75% ($q_{75\%}$) e de 25% ($q_{25\%}$)

✓Corresponde a ignorar os valores mais altos (25%) e mais baixos (25%)

$q_{75\%}$ = 274

$q_{25\%}$ = 142



Estatística Descritiva – **Medidas de dispersão**

VARIÂNCIA: **desvio quadrático médio**

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

a)

Observações →	6	24	37	49	64
Desvio ⟶	-30	-12	+1	+13	+28
Desvio quadrático →	900	144	1	169	784

Média: $\overline{X}$ = (6+24+37+49+64) / 5 = 36

Variância: $S^2 = \frac{900+144+1+169+784}{5} = 399.6$

Desvio padrão: $S = \sqrt{399.6} = 20$

Coeficiente de variação: $\frac{\text{Desvio padrão}}{\text{Média}} = \frac{20}{36} = 0.556$

Estatística Descritiva – **Medidas de dispersão**

b)

Observação →	111	114	117	118	120
Desvio ⟶	-5	-2	+1	+2	+4
Desvio quadrático →	25	4	1	4	16

Média: $\overline{X}$ = (111+114+117+118+120) / 5 = 116

Variância: $S^2 = \frac{25+4+1+4+16}{5} = 10$

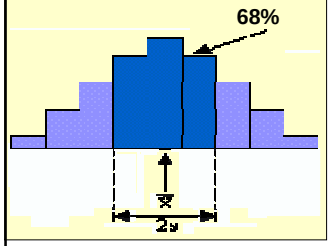
Desvio padrão: $S = \sqrt{10} = 3.16$

Coeficiente de variação: $\frac{\text{Desvio padrão}}{\text{Média}} = \frac{3.16}{116} = 0.027$

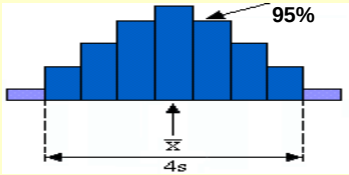
Desvio Padrão

➢ **Para uma distribuição normal…**

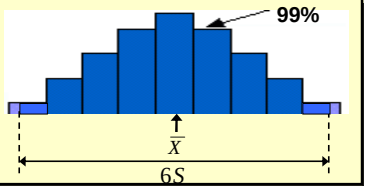
…aproximadamente **68%** das observações estão no intervalo $[\overline{X} - S, \overline{X} + S]$



…aproximadamente **95%** das observações estão no intervalo $[\overline{X} - 2S, \overline{X} + 2S]$



…aproximadamente **99%** das observações estão no intervalo $[\overline{X} - 3S, \overline{X} + 3S]$



Estatística Descritiva – medidas de dispersão

Salário mensal	Número de trabalhadores		
	Empresa A	Empresa B	Empresa C
50	13		7
80	70	49	32
120	30	68	92
160	20	20	12
200	10	13	4
350	5		2
500	2		1
Média	**119**	**119**	**119**
Variância	**5498**	**1290**	**2536**
Desvio padrão	**74**	**36**	**50**
Coeficiente de variação	**0.62**	**0.3**	**0.42**

Medidas de dispersão - Variância

Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Nº de acidentes	16	15	19	16	17	18	15	16	17	16	17	15	16	17	16	18

Média = $\overline{X} = 16.5$

(usando dados originais, sem tratamento)

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

n – número de observações

X_i – valor da i-ésima observação (nº de acidentes na semana i, neste exemplo)

$$S^2 = \left[(16-16.5)^2 + (15-16.5)^2 + (19-16.5)^2 + \ldots + (18-16.5)^2\right] / 16$$

Medidas de dispersão - Variância

Semana	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Nº de acidentes	16	15	19	16	17	18	15	16	17	16	17	15	16	17	16	18

Nº de acidentes	Frequência simples		Frequência acumulada	
	Absoluta	Relativa	Absoluta	Relativa
15	3	0,1875	3	0,19
16	6	0,375	9	0,56
17	4	0,25	13	0,81
18	2	0,125	15	0,94
19	1	0,0625	16	1,00

Média = $\overline{X} = 16.5$

S^2 = 0.1875 x $(15-16.5)^2$ + 0.375 x $(16-16.5)^2$ + 0.25 x $(17-16.5)^2$ +

+ 0.125 x $(18-16.5)^2$ + 0.0625 x $(19-16.5)^2$

(usando frequências relativas, após tratamento estatístico dos dados)

$$S^2 = \sum_{i=1}^{c} f_i \left(X_i' - \overline{X}\right)^2$$

c – número de classes

f_i – frequência relativa da i-ésima classe

X_i – valor (médio) da i-ésima classe

Nota: quando classes contêm mais do que um valor da grandeza, esta expressão produz uma aproximação

Medidas de dispersão - **Variância**

• Variância (desvio quadrático médio): cálculo para dados agrupados

$$S^2 = \sum_{i=1}^{c} f_i \left(X_i' - \overline{X}\right)^2$$

marca da classe

frequência relativa da classe

Limites das classes		Marca da classe	Freq.ia simples		Freq.ia acumulada	
Inferior	Superior		Absoluta	Relativa	Absoluta	Relativa
360	370	365	2	0,02	2	0,02
370	380	375	8	0,08	10	0,1
380	390	385	14	0,14	24	0,24
390	400	395	27	0,27	51	0,51
400	410	405	25	0,25	76	0,76
410	420	415	16	0,16	92	0,92
420	430	425	7	0,07	99	0,99
430	440	435	1	0,01	100	1

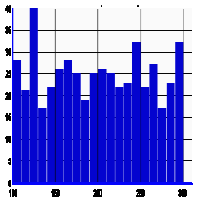
Média= 399.6

Variância: S^2 = 0.02 $(365-399.6)^2$ + 0.08 $(375-399.6)^2$ + 0.14 $(385-399.6)^2$ +

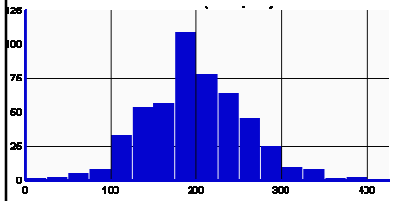
+ ... + 0.07 $(425-399.6)^2$ + 0.01 $(435-399.6)^2$ **= 210.84**

Desvio padrão: S = 14.52 **(Nota: S= 14.82** para dados originais, não tratados**)**

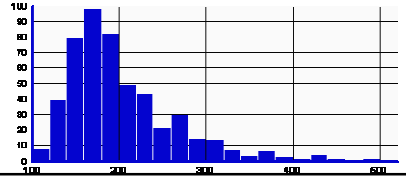
Comparação de distribuições de frequência



Média	**199,7**
Desvio padrão	**58,6**
Mínimo	**101,0**
Máximo	**300,0**



Média	**199,0**
Desvio padrão	**56,9**
Mínimo	**18,0**
Máximo	**385,0**

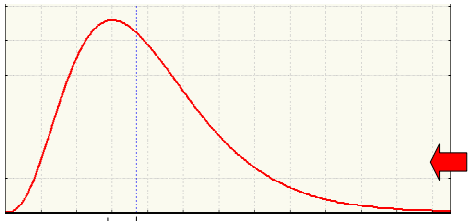


Média	**202,5**
Desvio padrão	**60,6**
Mínimo	**114,0**
Máximo	**491,0**

Medidas de Assimetria

➢ **Coeficientes de assimetria de Pearson:**

$$Ap_1 = \frac{\overline{X} - \text{Mod.}}{s} \qquad Ap_2 = \frac{3\left(\overline{X} - \text{Med.}\right)}{s}$$



Ap < 0 – distribuição com **assimetria esquerda** (ou negativa)

Ap = 0 – distribuição **simétrica**

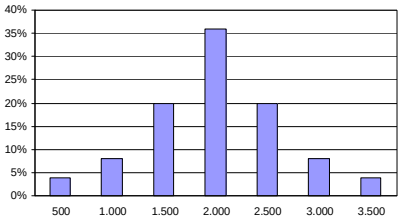
Ap > 0 - distribuição com **assimetria direita** (ou positiva)

➢ **Coeficiente assimetria**

$$C_1 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{s}\right)^3$$

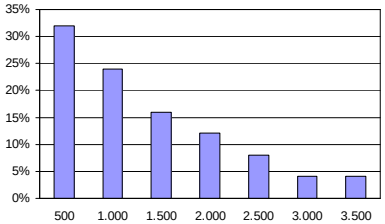
Estatística Descritiva - Distribuições de frequência e parâmetros de síntese

Salário	Nº de colab.	Freq.ia relativa	Freq.ia acumulada
500	1	4%	4%
1.000	2	8%	12%
1.500	5	20%	32%
2.000	9	36%	68%
2.500	5	20%	88%
3.000	2	8%	96%
3.500	1	4%	100%



Média = Mediana = Moda = 2000 **Desvio padrão = 663** **Coef. assimetria = 0**

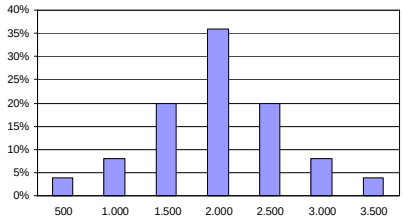
Salário	Nº de colab.	Freq.ia relativa	Freq.ia acumulada
500	8	32%	32%
1.000	6	24%	56%
1.500	4	16%	72%
2.000	3	12%	84%
2.500	2	8%	92%
3.000	1	4%	96%
3.500	1	4%	100%



Média=1340 Mediana=1000 Moda=500 **Desvio padrão=845** **Coef. assimetria= 0.91**

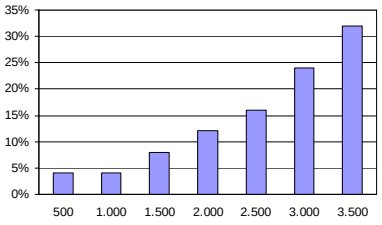
Estatística Descritiva - Distribuições de frequência e parâmetros de síntese

Salário	Nº de colab.	Freq.ia relativa	Freq.ia acumulada
500	1	4%	4%
1.000	2	8%	12%
1.500	5	20%	32%
2.000	9	36%	68%
2.500	5	20%	88%
3.000	2	8%	96%
3.500	1	4%	100%



Média = Mediana = Moda = 2000 **Desvio padrão = 663** **Coef. assimetria = 0**

Salário	Nº de colab.	Freq.ia relativa	Freq.ia acumulada
500	1	4%	4%
1.000	1	4%	8%
1.500	2	8%	16%
2.000	3	12%	28%
2.500	4	16%	44%
3.000	6	24%	68%
3.500	8	32%	100%

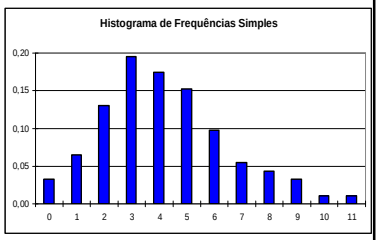


Média=2660 Mediana=3000 Moda=3500 **Desvio padrão=845** **Coef. assimetria= - 0.91**

Distribuições de frequência – exemplo de aplicação

Dimensionamento de uma frota de veículos: distribuição do nº de pedidos diários (requisições de veículos)

Nº de pedidos	Frequência simples		Frequência acumulada	
	Absoluta	Relativa	Absoluta	Relativa
0	3	0,03	3	0,03
1	6	0,07	9	0,10
2	12	0,13	21	0,23
3	18	0,20	39	0,42
4	16	0,17	55	0,60
5	14	0,15	69	0,75
6	9	0,10	78	0,85
7	5	0,05	83	0,90
8	4	0,04	87	0,95
9	3	0,03	90	0,98
10	1	0,01	91	0,99
11	1	0,01	92	1,00



Média = 4,2 pedidos/dia

Variância = 4,7

Distribuições de frequência – exemplo de aplicação

Dimensionamento de uma frota de veículos

- **Nº de veículos da frota própria = 4**

Frequência com que a frota própria é insuficiente para satisfazer todos os pedidos (rotura)?

Nº de pedidos	Frequência simples	Frequência acumulada
0	0,033	0,033
1	0,065	0,098
2	0,130	0,228
3	0,196	0,424
4	0,174	0,598
5	0,152	0,750
6	0,098	0,848
7	0,054	0,902
8	0,043	0,946
9	0,033	0,978
10	0,011	0,989
11	0,011	1,000

Freq (Rotura) = Freq (Nº de pedidos > 4) =
= Freq (X > 4) =
= 1 - Freq (X ≤ 4) =
= 1 - 0.598 = 40%

Distribuições de frequência – exemplo de aplicação

Dimensionamento de uma frota de veículos

- **Quantos veículos deveria ter a frota própria para que esta possa satisfazer todos os pedidos em 95% dos dias?**

N – nº de veículos da frota própria

Freq (Nº de pedidos ≤ N) = 95%

Nº de pedidos	Frequência simples	Frequência acumulada
0	0,033	0,033
1	0,065	0,098
2	0,130	0,228
3	0,196	0,424
4	0,174	0,598
5	0,152	0,750
6	0,098	0,848
7	0,054	0,902
8	0,043	0,946
9	0,033	0,978
10	0,011	0,989
11	0,011	1,000

N = 8

Freq (X ≤ 8) = 0,946

Distribuições de frequência – exemplo de aplicação

Dimensionamento de uma frota de veículos

- **Nº de veículos da frota própria = 4**

Em média, quantos veículos são utilizados (por dia)?

Nº de pedidos	Frequência simples	Frequência acumulada	Veículos usados
0	0,033	0,033	0
1	0,065	0,098	1
2	0,130	0,228	2
3	0,196	0,424	3
4	0,174	0,598	4
5	0,152	0,750	4
6	0,098	0,848	4
7	0,054	0,902	4
8	0,043	0,946	4
9	0,033	0,978	4
10	0,011	0,989	4
11	0,011	1,000	4

Média = 4,2 pedidos/dia

Y – nº de veículos utilizados

Média = 0x0.033 + 1x0.065 + 2x0.13 +
+ 3x0.196 +4x(1-0.424)
= 3.2

Taxa de utilização = 3.2 / 4 = 80.4%

Distribuições de frequência – exemplo de aplicação

Dimensionamento de uma frota de veículos

- **Nº de veículos da frota própria = 4**

Em média, quantos pedidos não são satisfeitos (por dia)?

Nº de pedidos	Frequência simples	Frequência acumulada	Pedidos não sat.
0	0,033	0,033	0
1	0,065	0,098	0
2	0,130	0,228	0
3	0,196	0,424	0
4	0,174	0,598	0
5	0,152	0,750	1
6	0,098	0,848	2
7	0,054	0,902	3
8	0,043	0,946	4
9	0,033	0,978	5
10	0,011	0,989	6
11	0,011	1,000	7

Média = 4,2 pedidos/dia

Z – nº de pedidos não satisfeitos

Média = 0x0.598 + 1x0.152 + 2x0.098 +
+ 3x0.054 + … + 7x0,011
= 0.99

Pedidos satisfeitos = 4.2 – 0.99 = 3.21

Nível de serviço = 3.21 / 4.2 = 76.4%

(% de pedidos satisfeitos pela frota própria)

Distribuições de frequência – exemplo de aplicação

Dimensionamento de uma frota de veículos

- **Nº de veículos da frota própria = 8**

(Resolvendo como para frota de 4 veículos)

Nº de pedidos	Frequência simples	Frequência acumulada	Veículos usados	Pedidos não sat.
0	0,033	0,033	0	0
1	0,065	0,098	1	0
2	0,130	0,228	2	0
3	0,196	0,424	3	0
4	0,174	0,598	4	0
5	0,152	0,750	5	0
6	0,098	0,848	6	0
7	0,054	0,902	7	0
8	0,043	0,946	8	0
9	0,033	0,978	8	1
10	0,011	0,989	8	2
11	0,011	1,000	8	3
4,2	**Médias**		4,1	0,09

Taxa de utilização = 4.1 / 8
= 51.5%

Pedidos satisfeitos =
= 4.2 – 0.09 = 4.11

Nível de serviço =
= 4.11 / 4.2 = 97.9%

(% de pedidos satisfeitos pela frota própria)

Distribuições de frequência – exemplo de aplicação

Dimensionamento de uma frota de veículos

➢ **Comparação de alternativas de frota própria** (4 ou 8 veículos)

Indicadores de desempenho

Frota (nº veic.)	Risco rotura	Veículos usados	Taxa utilização	Pedidos não satisfeitos	Nível serviço
4	40%	3.2	80,4%	0,99	76,5%
8	5%	4.1	51,5%	0,09	97,9%

- ✓ Quando há rotura (frota insuficiente), recorre-se a veículos alugados
- ✓ Aluguer de um veículo custa 10 000 dobrões/dia
- ✓ Custo de um veículo da frota própria: 1 000 dobrões/dia

➢ **Custo total = custo da frota própria + custo de alugueres**
= 1000 x Nº veículos da frota própria + 10 000 x Nº veículos alugados

✓ **4 veículos**: Custo médio diário: 4x1000 + 0.99x10 000 = 13 900 dobrões

✓ **8 veículos**: Custo médio diário: 8x1000 + 0.09x10 000 = 8 900 dobrões

Distribuições de frequência – exemplo de aplicação

Dimensionamento de uma frota de veículos

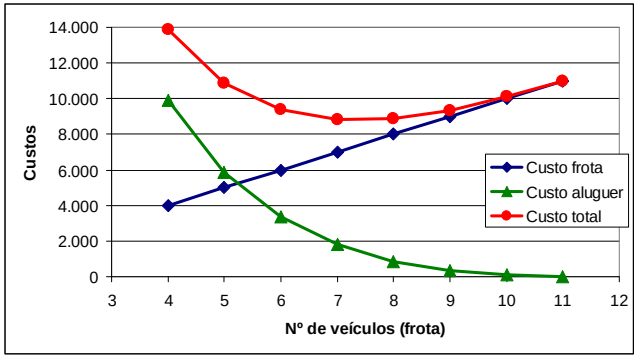
➢ **Comparação de alternativas de frota própria**

Frota (nº veic.)	Custo frota	Nº veic. alugados	Custo aluguer	Custo total	Risco rotura	Nível serviço	Taxa utilização
4	4.000	0,9891	9.891	13.891	40%	76%	80,4%
5	5.000	0,5870	5.870	10.870	25%	86%	72,4%
6	6.000	0,3370	3.370	9.370	15%	92%	64,5%
7	7.000	0,1848	1.848	8.848	10%	96%	57,5%
8	8.000	0,0870	870	8.870	5%	98%	51,5%
9	9.000	0,0326	326	9.326	2%	99%	46,4%
10	10.000	0,0109	109	10.109	1%	100%	42,0%
11	11.000	0,0000	0	11.000	0%	100%	38,2%

Distribuições de frequência – exemplo de aplicação

Dimensionamento de uma frota de veículos

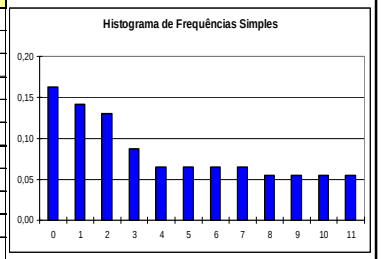
➢ **Comparação de alternativas de frota própria**



Distribuições de frequência – exemplo de aplicação (caso 2)

➢ Dimensionamento de uma frota de veículos: distribuição do nº de pedidos diários

Nº de pedidos	Freq. simples Absoluta	Freq. simples Relativa	Freq. acumulada Absoluta	Freq. acumulada Relativa
0	15	0,16	15	0,16
1	13	0,14	28	0,30
2	12	0,13	40	0,43
3	8	0,09	48	0,52
4	6	0,07	54	0,59
5	6	0,07	60	0,65
6	6	0,07	66	0,72
7	6	0,07	72	0,78
8	5	0,05	77	0,84
9	5	0,05	82	0,89
10	5	0,05	87	0,95
11	5	0,05	92	1,00



Média = 4,2 pedidos/dia

Variância = 9.4

➢ Para risco de rotura de 5% ➔ **Nº de veículos = 10**

Distribuições de frequência – exemplo de aplicação (caso 2)

Dimensionamento de uma frota de veículos (caso 2)

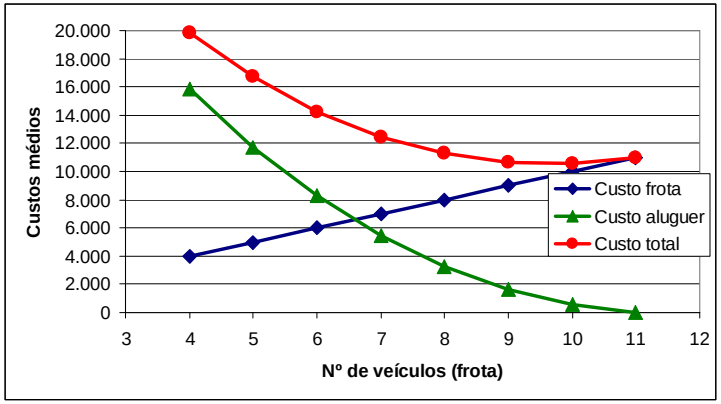
➢ Comparação de alternativas de frota própria

Frota (nº veic.)	Custo frota	Nº veic. alugados	Custo aluguer	Custo total	Risco rotura	Nível serviço	Taxa utilização
4	4.000	1,5870	15.870	19.870	41%	62%	64,4%
5	5.000	1,1739	11.739	16.739	35%	72%	59,8%
6	6.000	0,8261	8.261	14.261	28%	80%	55,6%
7	7.000	0,5435	5.435	12.435	22%	87%	51,7%
8	8.000	0,3261	3.261	11.261	16%	92%	48,0%
9	9.000	0,1630	1.630	10.630	11%	96%	44,4%
10	10.000	0,0543	543	10.543	5%	99%	41,1%
11	11.000	0,0000	0	11.000	0%	100%	37,8%

Distribuições de frequência – exemplo de aplicação (caso 2)

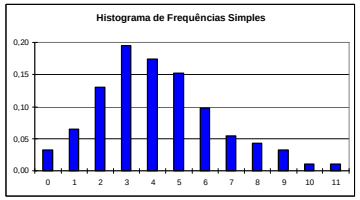
Dimensionamento de uma frota de veículos

➢ Comparação de alternativas de frota própria

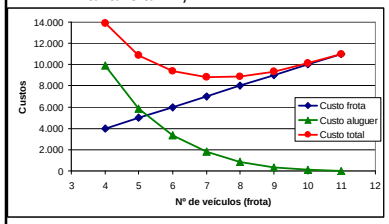


Distribuições de frequência – exemplo de aplicação

Caso 1

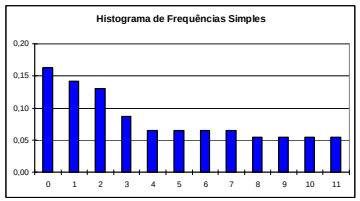


Média = 4,2 pedidos/dia
Variância = 4,7

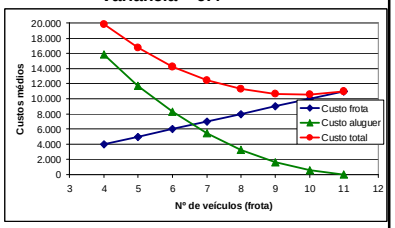


Frota óptima= 7 veículos

Caso 2



Média = 4,2 pedidos/dia
Variância = 9.4



Frota óptima= 10 veículos

Distribuições "empíricas" *vs* modelos "teóricos"

Distribuição "empírica" (frequências observadas) – Média = 4.2

Nº de pedidos	Frequência simples		Frequência acumulada	
	Absoluta	Relativa	Absoluta	Relativa
0	3	0,033	3	0,033
1	6	0,065	9	0,098
2	12	0,130	21	0,228
3	18	0,196	39	0,424
4	16	0,174	55	0,598
5	14	0,152	69	0,750
6	9	0,098	78	0,848
7	5	0,054	83	0,902
8	4	0,043	87	0,946
9	3	0,033	90	0,978
10	1	0,011	91	0,989
11	1	0,011	92	1,000
12	0	0,000	92	1,000
13	0	0,000	92	1,000

Distribuição "teórica" - Poisson (λ = 4.2)

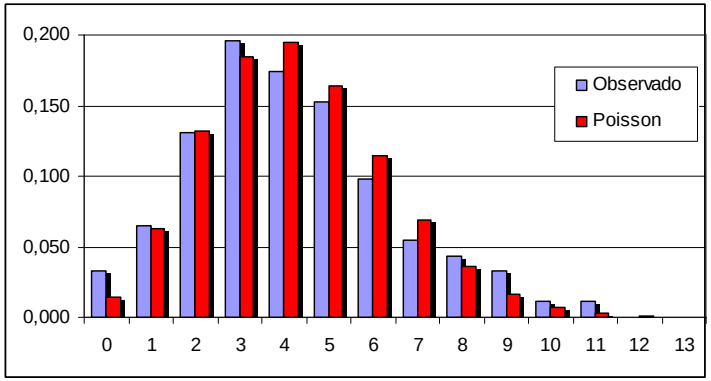
Probabilidade	
simples	acum.
0,015	0,015
0,063	0,078
0,132	0,209
0,185	0,394
0,194	0,589
0,164	0,752
0,115	0,867
0,069	0,936
0,036	0,972
0,017	0,989
0,007	0,996
0,003	0,999
0,001	1,000
0,000	1,000

$$p(x) = \frac{e^{\lambda}\,\lambda^{x}}{x!}$$

Distribuições “empíricas” *vs* modelos “teóricos”

Distribuição “empírica” (frequências observadas) – Média = 4.2

Distribuição “teórica” - Poisson (λ = 4.2) $\Longrightarrow p(x) = \dfrac{e^{\lambda}\,\lambda^{x}}{x!}$

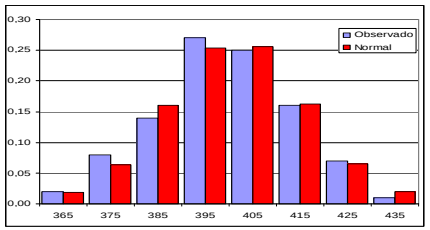


Distribuição “empírica” (frequências observadas)

Distribuição “teórica” - Normal (400, 14.5)

Limite das classes		Freq.ia simples		Freq.ia acumulada		Probabilidades	
Inferior	Superior	Absoluta	Relativa	Absoluta	Relativa	Simples	Acum.
360	370	2	0,02	2	0,02	0,019	0,019
370	380	8	0,08	10	0,10	0,064	0,083
380	390	14	0,14	24	0,24	0,160	0,243
390	400	27	0,27	51	0,51	0,254	0,497
400	410	25	0,25	76	0,76	0,255	0,753
410	420	16	0,16	92	0,92	0,162	0,915
420	430	7	0,07	99	0,99	0,065	0,980
430	440	1	0,01	100	1,00	0,020	1,000

Distribuições “empíricas” *vs* modelos “teóricos”



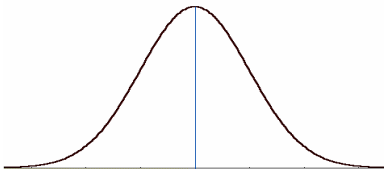
Modelos “teóricos” - Distribuição normal (de Gauss)

Densidade de probabilidade:

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ – média

σ – desvio padrão

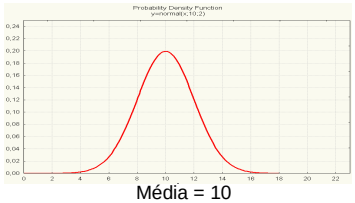


μ (média)

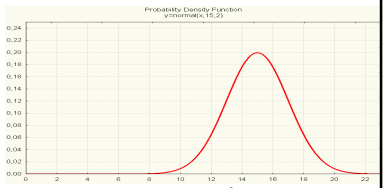
- Distribuição simétrica
  - ✓ centrada na média
  - ✓ 50% dos valores abaixo da média e 50% acima da mesma
  - ✓ Moda = mediana = média
- Maiores probabilidades na vizinhança da média, decaindo à medida que os valores se vão afastando da média (quer para esquerda, quer para a direita)

Modelos “teóricos” - Distribuição normal (de Gauss)

Efeito da alteração da média (localização do “centro” da distribuição)

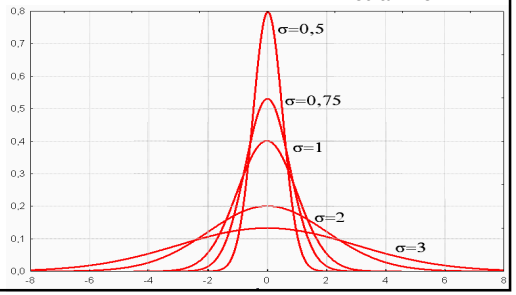


Média = 10



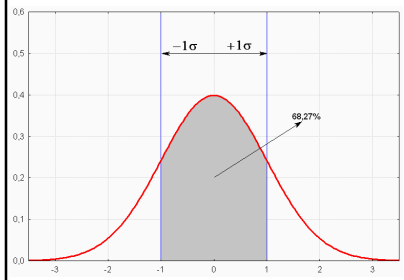
Média = 15

Efeito da alteração da variância (grau de dispersão)



Distribuição normal (de Gauss)

➢ **Cerca de 68% das probabilidades concentram-se no intervalo (média – desvio padrão) ; (média + desvio padrão)**



Exemplo 1:

X ~ Normal (Média=100, Desvio padrão=5)

$$X \sim \text{Normal}(100,\ 5)$$

$$P[95 \le X \le 105] \approx 68\%$$

Exemplo 2:

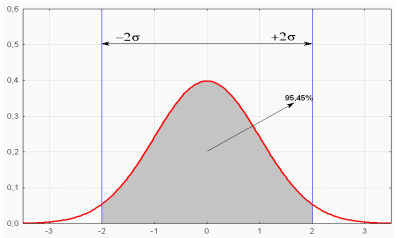
$$Y \sim \text{Normal}(100,\ 30)$$

$$P[70 \le Y \le 130] \approx 68\%$$

Maior dispersão ➡ Intervalo mais largo

Distribuição normal (de Gauss)

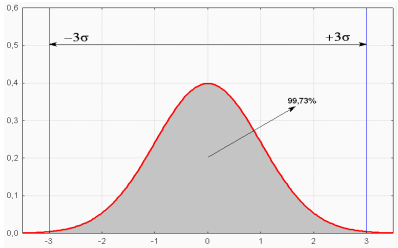
X ~ Normal (100, 5)
Y ~ Normal (100, 30)



➢ **Cerca de 95.5% das probabilidades concentram-se no intervalo (média – 2 d. padrão) ; (média + 2 d. padrão)**

$$P[90 \le X \le 110] \approx 95.5\%$$

$$P[40 \le Y \le 160] \approx 95.5\%$$



➢ **Cerca de 99.7% das probabilidades concentram-se no intervalo (média – 3 d. padrão) ; (média + 3 d. padrão)**

$$P[85 \le X \le 115] \approx 99.7\%$$

$$P[10 \le Y \le 190] \approx 99.7\%$$

Distribuição normal padrão (ou reduzida, ou standardizada)

$$X \sim Normal\ (\mu, \sigma) \quad \rightarrow \quad Z = \frac{X - \mu}{\sigma}$$

Padronização da variável X

Z – variável normal padrão
- média nula
- variância unitária

$$X \sim Normal\ (100, 15)$$

$$P[X \le 130] = P\left[Z \le \frac{130 - 100}{15}\right] = P[Z \le 2]$$

$$= \varnothing(2) \approx 0.977$$

Função cumulativa da lei normal padrão (estão disponíveis tabelas)

Teorema do Limite Central (TLC)

➢ Seja *Y* uma variável que resulta da soma de *n* variáveis (X_i) independentes e com idêntica distribuição:

$$Y = X_1 + X_2 + \ldots + X_n$$

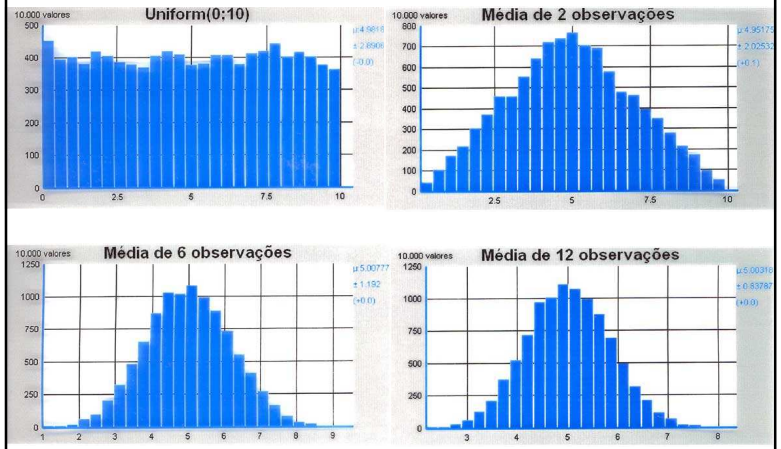
➢ **À medida que *n* cresce, a distribuição de *Y* tende assimptóticamente para a distribuição normal** (qualquer que seja a distribuição dos X_i)

$$Y = \sum_{i=1}^{n} X_i \xrightarrow[n \to \infty]{} Normal\left(\mu_y, \sigma_y\right)$$

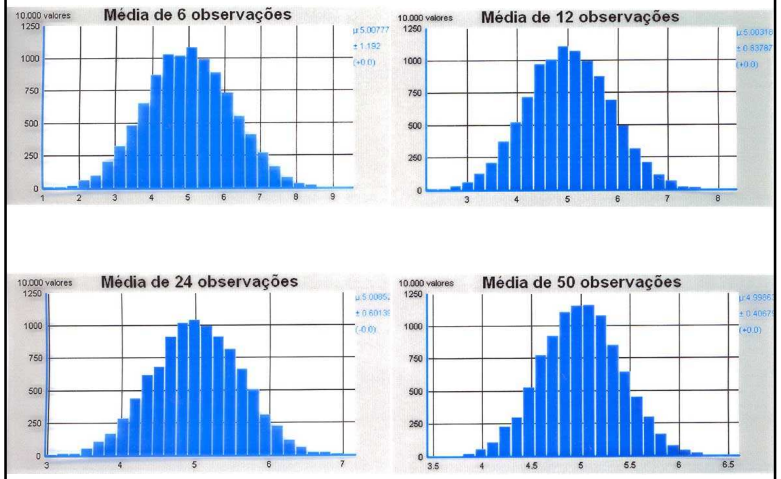
Média de Y = soma das médias das variáveis X_i

Variância de Y = soma das variâncias das variáveis X_i

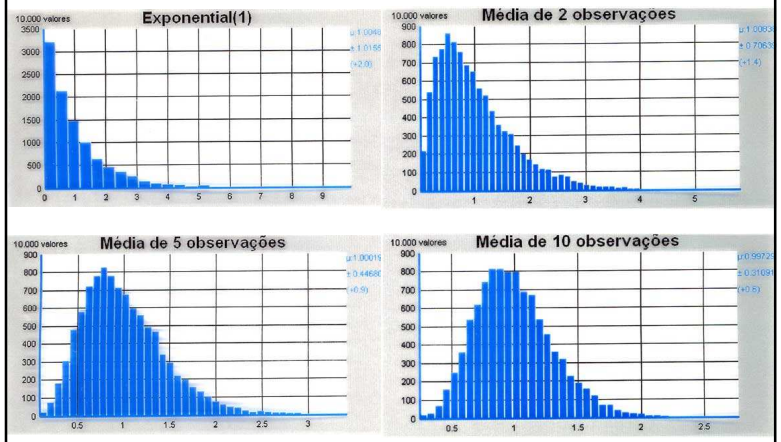
TLC - População Uniforme



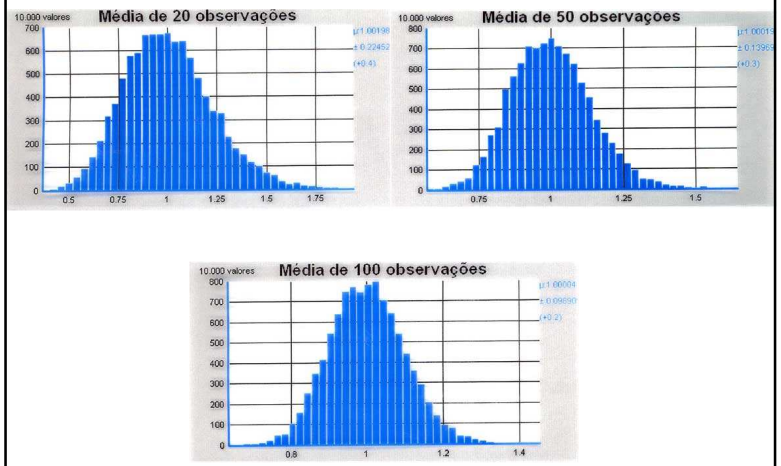
TLC - População Uniforme



TLC - População Exponencial



TLC - População Exponencial



Teorema do Limite Central (TLC)

- **A distribuição da soma de *n* variáveis aleatórias (independentes e identicamente distribuídas) tende para a distribuição normal à medida que *n* cresce** (qualquer que seja a distribuição das parcelas!)

- Note-se que:

1. Rapidez da convergência depende da distribuição dos X_i, sendo necessárias menos parcelas se esta distribuição for simétrica. Por exemplo:
   - Se X_i ~ Uniforme, a aproximação normal será "razoável" se $n \geq 12$
   - Se X_i ~ exponencial negativa (muito assimétrica), poderão ser necessárias algumas dezenas de parcelas para a aproximação à normal ser "aceitável"

2. Convergência para a normal verifica-se mesmo quando as parcelas (X_i) não têm a mesma distribuição!