

Análise de Dados

Estatística Descritiva e Probabilidades

Aplicadas à Investigação

- ❖ **Objectivos**
- ❖ **Conteúdos Programáticos**
- ❖ **Estratégias de Ensino-Aprendizagem**
- ❖ **Bibliografia**
- ❖ **Material (Máquina de calcular científica)**
- ❖ **Avaliação**



escola superior de
enfermagem
de coimbra

Prof. Manuel Gameiro

Objectivos:

- ❖ **Reconhecer a estatística como “ferramenta” (e “linguagem”) da investigação científica.**
- ❖ **Realizar análise descritiva / exploratória de dados.**
 - Organizar e apurar dados.
 - Calcular as medidas de tendência central, quantílicas, de dispersão, de simetria e de curtose.
 - Analisar a associação entre duas variáveis.
 - Apresentar os dados e resultados estatísticos.
- ❖ **Compreender os modelos probabilísticos básicos na previsão e inferência estatística.**
 - Compreender os conceitos fundamentais da teoria das probabilidades.
 - Conhecer os axiomas e teoremas essenciais das probabilidades.
 - Reconhecer as variáveis aleatórias discretas e contínuas.
 - Determinar o valor esperado $[E(X)]$ e a dispersão $[\sigma(X)]$ de uma variável aleatória.
 - Resolver problemas básicos de previsão baseados nos modelos probabilísticos básicos (Binomial, de Poisson, Normal).

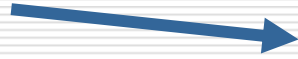


ESTATÍSTICA

DESCRITIVA

e Exploratória

Status



Estado

Estatística

Estatística:

Um repositório de técnicas para:

- ✓ Recolher
 - ✓ Organizar
 - ✓ Apresentar
 - ✓ Analisar conjuntos de dados
 - ✓ Retirar conclusões
- ...e Tomar decisões

com elevada confiança

**Também uma “linguagem” da
investigação quantitativa.**

Estatística ≠ Parâmetro

Estatística

❑ Estatística Metodológica / Estatística Aplicada

- Bioestatística
- Estatística Vital
- Estatística Aplicada às Ciências Humanas
- Estatística Aplicada à Administração

❑ Estatística Descritiva / Estatística Inferencial

Estatística Exploratória

“Objecto” de Estudo da Estatística

Agregado – Conjunto

População (N)

Amostra (n)

Elementos | Unidades

Variáveis

Dados

Informação

Descrição

Decisão



População | Agregado

Unidade Estatística

⊖ À estatística não interessa o estudo dos fenómenos isolados, dos casos considerados no individual. Estes só são objecto formal da estatística quando fazem parte de um **conjunto numeroso**.

POPULAÇÃO:

Conjunto de elementos (pessoas, animais, objectos, fenómenos) com características comuns e sobre a qual incide a atenção estatística.

* **Finita**

* **Infinita | Inumerável**

Bem definida no TEMPO e no ESPAÇO

△ Cada elemento do conjunto definido pelas características em comum é uma **UNIDADE ESTATÍSTICA**

Variáveis Dados Informação

Expressão | Nível de Medida

Variáveis:



Organização e Apuramento dos dados (nominais)

Tipo de Parto:

Caos

Normal	Normal	Normal	Cesariana	Normal
Normal	Normal	Normal	Normal	Normal
Ventosa	Normal	Normal	Cesariana	Normal
Normal	Normal	Normal	Fórceps	Normal
Normal	Cesariana	Normal	Cesariana	Normal
Cesariana	Normal	Normal	Normal	Normal
Normal	Fórceps	Normal	Normal	Normal

Série (Rol)

Cesariana	Cesariana	Cesariana	Cesariana
Cesariana	Cesariana	Fórceps	Fórceps
Normal	Normal	Normal	Normal
Normal	Normal	Normal	Normal
Normal	Normal	Normal	Normal
Normal	Normal	Normal	Normal
Normal	Normal	Normal	Normal
Normal	Normal	Ventosa	

Organização e Apuramento dos dados (ordinais, discretos)

Paridade:

Caos

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

Série (Rol)

x_1	x_2																
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	3	3	3	3	4	4	5
5																	
x_n																	

x_1 (Min.); x_n (Max.)

$AV = x_n - x_1 = 5 - 1 = 4$

Organização e Apuramento dos dados (nominais, ordinais, discretos)

Tabelas de Frequências

Tipo de parto

X_i	F_i	f_i	$\%_i$
Normal	28	0,76	75,68
Cesariana	6	0,16	16,22
Forceps	2	0,05	5,40
Ventosa	1	0,03	2,70
	$n = 37$	1,00	100,00

$$\sum_{i=1}^k F_i = n$$

Paridade

X_i	F_i	Fac_i	f_i	$\%_i$	$\%ac_i$
1	18	18	0,49	48,65	48,65
2	10	28	0,27	27,03	75,68
3	4	32	0,11	10,81	86,49
4	2	34	0,05	5,40	91,89
5	3	37	0,08	8,11	100,00
	$n = 37$		1,00	100,00	

$$f_i = \frac{F_i}{n}$$

$$\% i = f_i \times 100$$

Organização e Apuramento dos dados (Contínuos, discretos com f_i baixos)

Caos

2780	3620	3860	1975	2560	3810	3200
2880	2260	3400	3290	4100	1870	3120
3700	2600	3320	2210	2900	4320	2050
3280	3080	2190	2000	1980	2870	3140
3450	2340	2010	1630	3090	3920	2900
3470	3100					

Série (Rol)

$x_{i=1}$	$x_{i=2}$	x_3	x_4				
1630	1870	1975	1980	2000	2010	2050	
2190	2210	2260	2340	2560	2600	2780	
2870	2880	2900	2900	3080	3090	3100	
3120	3140	3200	3280	3290	3320	3400	
3450	3470	3620	3700	3810	3860	3920	
4100	4320						
		x_{37}					

$n=37$ Min. (x_1)=1630 grs Max. (x_n)=4320 grs AV=2690 grs



Apuramento e organização dos dados (contínuos, discretos com f_i baixos)

Tabelas de Frequências

K_i	F_i	Fac_i	f_i	$\%_i$	$\%ac_i$
1500 - 2000	4	4	0,11	10,81	10,81
2000 - 2500	7	11	0,19	18,92	29,73
2500 - 3000	7	18	0,19	18,92	48,65
3000 - 3500	12	30	0,32	32,43	81,08
3500 - 4000	5	35	0,14	13,51	94,58
4000 - 4500	2	37	0,05	5,41	100.00
	$n = 37$		1,00	100,00	

$5 \leq \# K \leq 15$

$$\frac{AV}{15} \leq h \leq \frac{AV}{5}$$

De preferência, de igual amplitude (h).

Limites Reais /
Limites Aparentes

K : Classe $I_s - I_i = h$ $P_m = X' = (I_i + I_s)/2$; $\sum Fx'(108250) \approx \sum Fx(108275)$

Organização e Apuramento dos dados (contínuos, discretos com f_i baixos)

Diagrama de caule e folhas

Peso ao Nascer Stem-and-Leaf Plot

Frequency	Stem & Leaf
4,00	1 . 6899
7,00	2 . 0001223
7,00	2 . 5678899
12,00	3 . 001112223444
5,00	3 . 67889
2,00	4 . 13

Stem width: 1000
Each leaf: 1 case(s)

Análise de Séries (univariadas)

- **Medidas de Tendência Central**
 - Média
 - Mediana
 - Moda
- **Medidas de Localização (Quantis)**
 - Quartis
 - Decis e Percentís
- **Medidas de Dispersão**
 - Centradas na Média
 - Centradas na Mediana
- **Medidas de Forma**
 - Assimetria
 - Curtose

△ Quando trabalhamos com séries muito extensas, as distribuições de frequências não são suficientes para uma análise satisfatória:

△ É necessário “reduzir” | “sumariar” os dados, calculando medidas (índices) que representem as características típicas da série (**Estatísticas / Parâmetros**)

Medidas de Tendência Central

Média Aritmética ($\bar{x}/\mu$)

❖ Pequenas séries:

Glicémias em jejum (mg/dl)

n=9

83 78 82 73 90 70 72 97 82

$$\bar{X} = \frac{\sum x_i}{n} = \frac{83+78+82+73+90+70+72+97+82}{9} = 80,78 \text{ mg / dl}$$

❖ Dados tabelados

n.º de irmãos

x	F	Fx
0	14	0
1	26	26
2	10	20
3	8	24
	58	70

$$\bar{X} = \frac{\sum Fx}{n} = \frac{70}{58} = 1,21 \text{ irmãos}$$

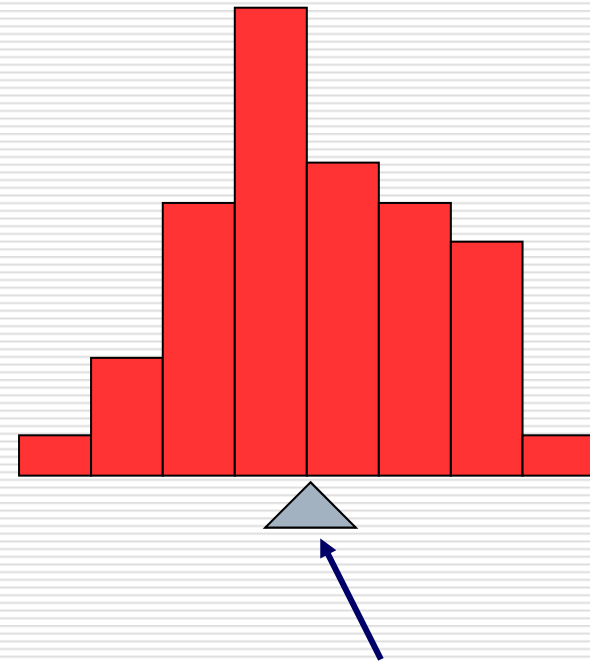
Medidas de Tendência Central

Média Aritmética ($\bar{x}/\mu$)

❖ Dados classificados

Doentes com Gonalgias Idiopáticas

Uricémia (mg/dl)	F
2 - 3	1
3 - 4	3
4 - 5	7
5 - 6	12
6 - 7	8
7 - 8	7
8 - 9	6
9 - 10	1
	45



$$\bar{X} = \frac{\sum Fx'}{n} = \frac{275,50}{45} = 6,12 \text{ mg / dl}$$

Medidas de Tendência Central

Médias de aplicação especial

❖ Média Ponderada

Quando aos diversos valores são atribuídas "importâncias" diferentes.

$$\bar{X}_p = \frac{\sum x_p}{\sum p}$$

❖ Média Geométrica

Quando o valor de uma variável evolui ao longo de um determinado período numa sequência geométrica.

$$G = \sqrt[n]{\prod_{i=1}^n x_i^{F_i}} \quad \text{ou} \quad \log G = \frac{\sum_{i=1}^n F_i \log x_i}{n}$$

❖ Média Harmónica

Quando os valores são inversamente proporcionais a um determinado critério.

$$H = \frac{n}{\sum_{i=1}^n \frac{F_i}{x_i}}$$

Medidas de Tendência Central

Mediana (Md)

Glicémias em jejum (mg/dl)

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
70	72	73	78	82	82	83	90	97

Md=82 mg/dl

n=9

$$Md = x_i \quad i = \frac{n+1}{2}$$



se n é par, estima-se a Md:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
63	70	72	73	78	82	82	83	90	97



Md = $x_{5,5} = (78+82)/2 = 80$ mg/dl

n=10

$$Md = \frac{x_i + x_{i+1}}{2}$$

Medidas de Tendência Central

Mediana (Md)

❖ Dados tabelados

nº de irmãos

x	F	Fac
0	14	14
1	26	40
2	10	50
3	8	58
	58	

$$Md = x_i \quad i = \frac{n+1}{2}$$

$$Md = X_{29,5} = 1 \text{ irmão}$$

Medidas de Tendência Central

Mediana (Md)

❖ Dados classificados

Peso ao nascer (grs)

K_i	F_i	Fac_i
1500 - 2000	4	4
2000 - 2500	7	11
2500 - 3000	7	18
3000 - 3500	12	30
3500 - 4000	5	35
4000 - 4500	2	37
$n = 37$		

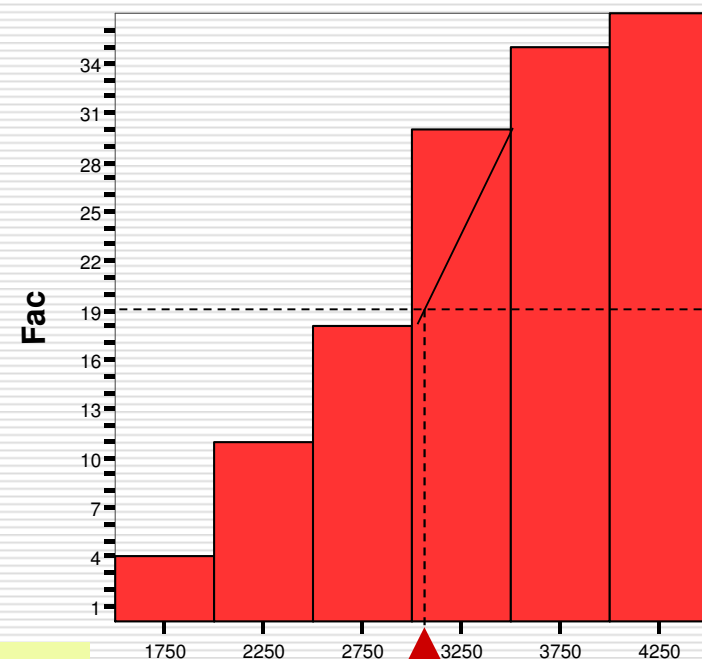
$$Md = x_{19}$$

$$Md = li + \frac{\frac{n}{2} - Fac}{F} \times h$$

$$Md = 3000 + \frac{18,5 - 18}{12} \times 500 = 3021 \text{ grs}$$

Intrapolação:

$$Md \in [3000, 3500[$$



Peso dos Recém-Nascidos (Classes)

Medidas de Tendência Central

Moda (Mo)

Tipo de parto

X_i	F_i	f_i	$\%_i$
Normal	28	0,76	75,68
Cesariana	6	0,16	16,22
Forceps	2	0,05	5,40
Ventosa	1	0,03	2,70
	n =37	1,00	100,00

Mo: Parto normal

Paridade

x_i	F_i	f_i	$\%_i$
1	18	0,49	48,65
2	10	0,27	27,03
3	4	0,11	10,81
4	2	0,05	5,40
5	3	0,08	8,11
	n =37	1,00	100,00

Mo=1 parto

Mo= x_i de maior F

Medidas de Tendência Central

Moda (Mo)

❖ Dados classificados

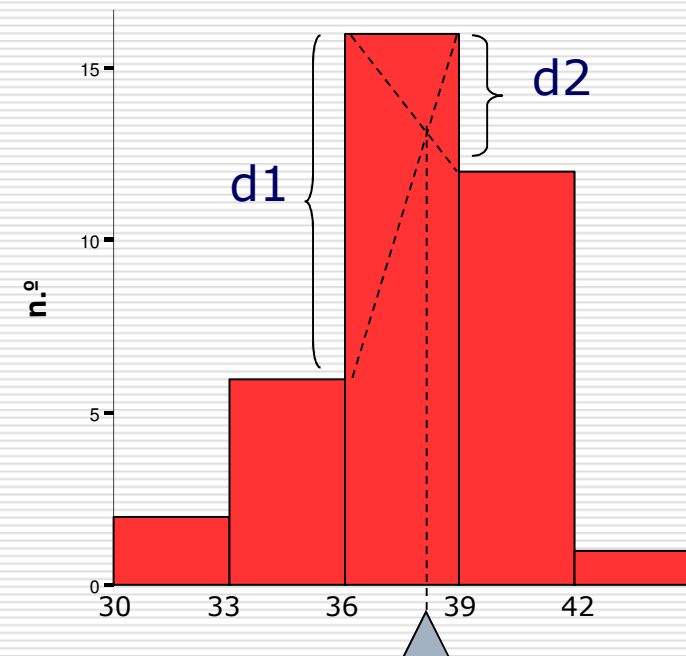
Tempo de Gestação	F	%
30 - 32	2	5,41
33 - 35	6	16,22
36 - 38	16	43,24
39 - 41	12	32,43
42 ou +	1	2,70
	37	100,00

Fórmula de Czyber:

$$Mo = li + \frac{d1}{d1 + d2} \times h$$

Intrapolação:

$$36 \leq Mo \leq 38$$



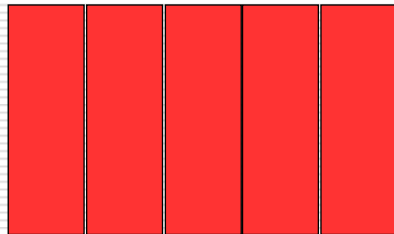
$$Mo = 36 + \frac{10}{10 + 4} \times 3 = 38,14 \text{ semanas}$$

Para classes de diferente h:
Fórmula de King

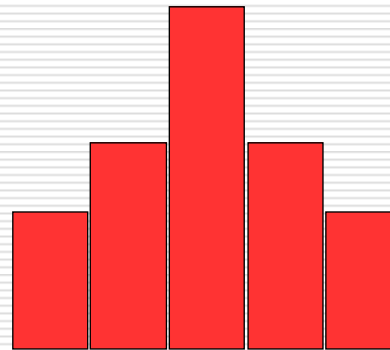
Medidas de Tendência Central

Moda (Mo)

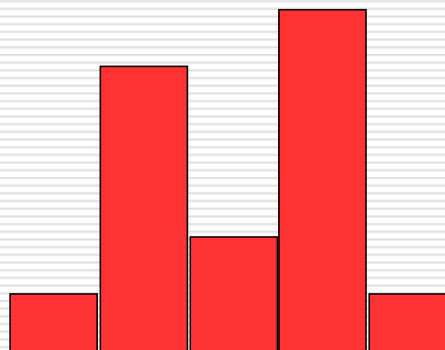
Distribuições



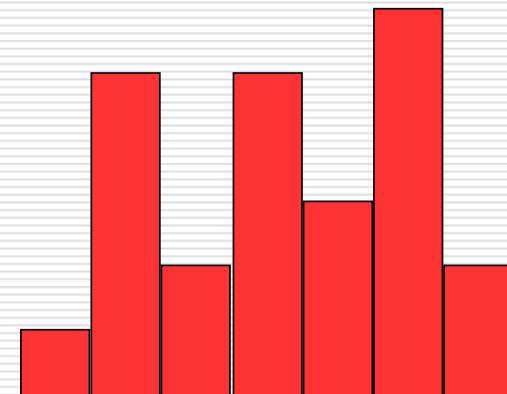
Amodal



Unimodal



Bimodal



Multimodal

Medidas de Localização (Quantis ou Separatrizes)

Quartis (Q_j) $j=1; j=2; j=3$

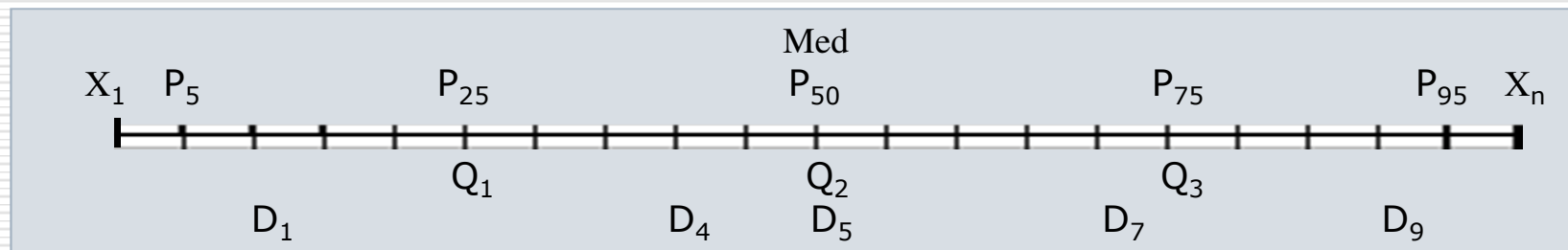
$$Q_j = x_i \rightarrow i = j \frac{n}{4}$$

Decis (D_j) $j=1; j=2; \dots; j=9$

$$D_j = x_i \rightarrow i = j \frac{n}{10}$$

Percentis (P_j) $j=1; j=2; \dots; j=99$

$$P_j = x_i \rightarrow i = j \frac{n}{100}$$



Medidas de Localização (Quantis ou Separatrizes)

❖ Dados classificados

Intrapolação:

$$S = li + \frac{i - 'Fac}{F} \times h$$

S – Separatriz a calcular.

i – Ordem da separatriz na série.

K	F
2 – 4	40
4 – 6	85
6 – 8	170
8 – 10	85
10 – 12	35
	415

Me=

Q₁=

Q₃=

P₁₀=

P₈₀=

Medidas de Localização (Quantis ou Separatrizes)

Localização quantílica de um valor (x) na série de dados classificados.

P: Determinar a localização percentílica (P) de um recém-nascido com 47 cm de estatura.

Estatura (cm)	F_i	Fac_i
30 - 35	14	14
35 - 40	21	35
40 - 45	32	67
45 - 50	68	135
50 - 55	36	171
	171	

$$J^{(P)} = \frac{i^{(x)}}{n} \times 100$$

$$i^{(x)} = 'Fac + \frac{(x - li)F_i}{h_i}$$

$$R: J^{(P)} = 55,09 \Rightarrow P_{55} < 47 < P_{56}$$

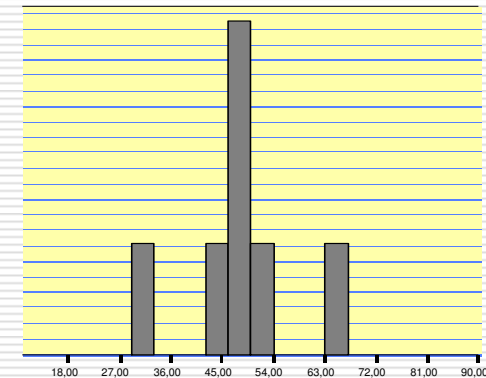
Medidas de Dispersão (de Variabilidade)

- ▲ As medidas de tendência central não são suficientes para caracterizar uma distribuição de dados.

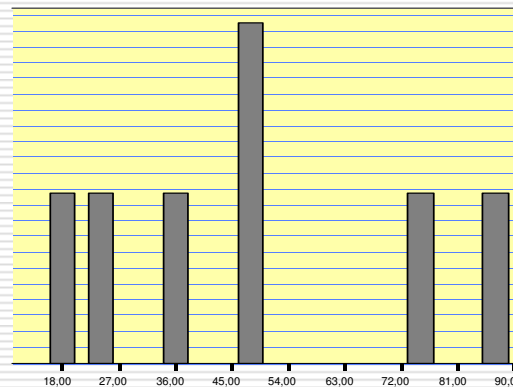
Grupo A: 31 44 48 48 48 52 65

Grupo B: 18 24 36 48 48 75 87

	Grupo A	Grupo B
n	7	7
Média	48,00	48,00
Moda	48,00	48,00
Mediana	48,00	48,00
AV	34,00	69,00



Grupo A



Grupo B

Medidas de Dispersão (de Variabilidade)

$$AV = x_n - x_1 \quad \text{Só considera os extremos}$$

❖ Medidas Centradas na Média

* **Desvio:** $d_i = x_i - \bar{X}$

* **Desvio Absoluto:** $DA_i = |x_i - \bar{X}|$

* **Quadrado dos Desvios:** $d_i^2 = (x_i - \bar{X})^2$

$$\sum d_i = \sum (x_i - \bar{X}) = 0$$

$$\sum DA_i = \sum |x_i - \bar{X}|$$

$$\sum d_i^2 = \sum (x_i - \bar{X})^2$$

Grupo A				Grupo B			
X_i	$X_i - \bar{X}$	$ X_i - \bar{X} $	$(x_i - \bar{X})^2$	X_i	$X_i - \bar{X}$	$ X_i - \bar{X} $	$(x_i - \bar{X})^2$
31	-17	17	289	18	-30	30	900
44	-4	4	16	24	-24	24	576
48	0	0	0	36	-12	12	144
48	0	0	0	48	0	0	0
48	0	0	0	48	0	0	0
52	+4	4	16	75	+27	27	729
65	+17	17	289	87	+39	39	1521
	0	42	610		0	132	3870

▲ Estas medidas não têm em conta diferenças de n

Medidas de Dispersão (de Variabilidade)

❖ Medidas Centradas na Média

x_i	F_i	$F_i x_i$	x_i^2	$F_i x_i^2$
1	18	18	1	18
2	10	20	4	40
3	4	12	9	36
4	2	8	16	32
5	3	15	25	75
	37	73		201

* Desvio Absoluto Médio:

$$DAM = \frac{\sum F_i |x_i - \bar{X}|}{n}$$

* Variância:

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

* Desvio Padrão (DP):

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

$$s^2 = \frac{201}{37} - 3,89 = 1,54 \text{ partos}^2$$

$$s = \sqrt{\frac{201}{37} - 3,89} = 1,24 \text{ partos}$$

$$\hat{s} = s \sqrt{\frac{n}{n-1}} = 1,26 \text{ partos}$$

Desvio Padrão (estimativa de σ):

$$\hat{s} = \sqrt{\frac{\sum F_i (x_i - \bar{X})^2}{n-1}} = s \times \sqrt{\frac{n}{n-1}}$$

Medidas de Dispersão (de Variabilidade)

❖ Dispersão Relativa em Relação à Média

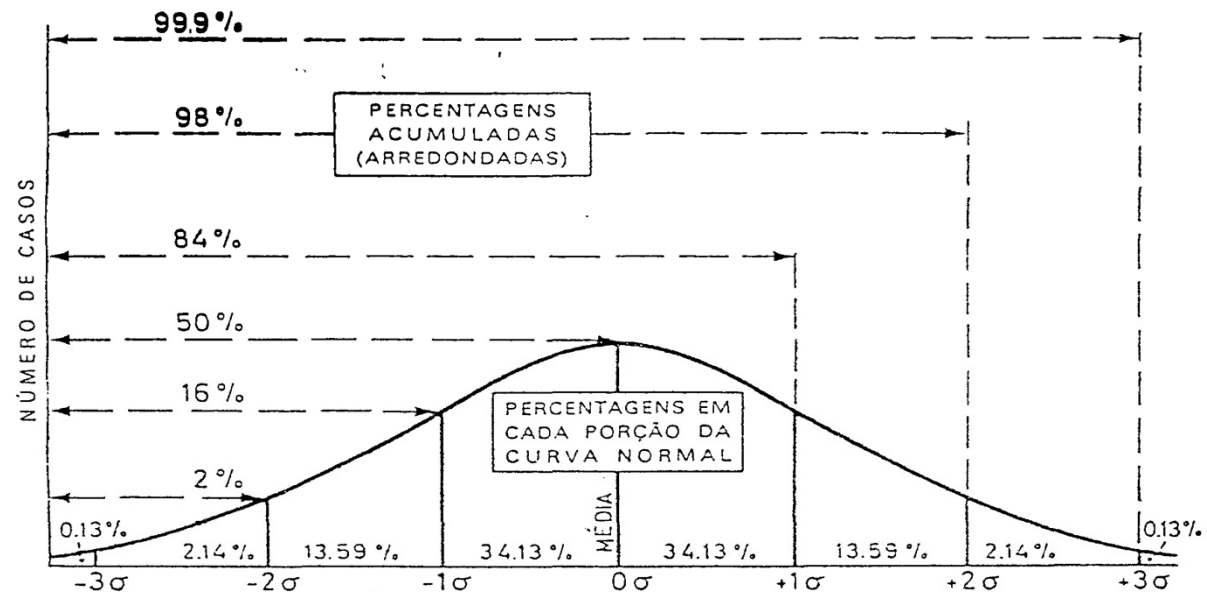
* Coeficiente de Variação (CV)

$$CV = \frac{s}{\bar{X}} \times 100$$

$CV \leq 15\%$	Dispersão fraca
$15\% < CV \leq 30\%$	Dispersão média
$CV > 30\%$	Dispersão elevada

$$CV = \frac{1,24}{1,97} \times 100 = 62,89\%$$

» » » »
**Relação entre a
média e o
desvio padrão
numa
distribuição
normal**

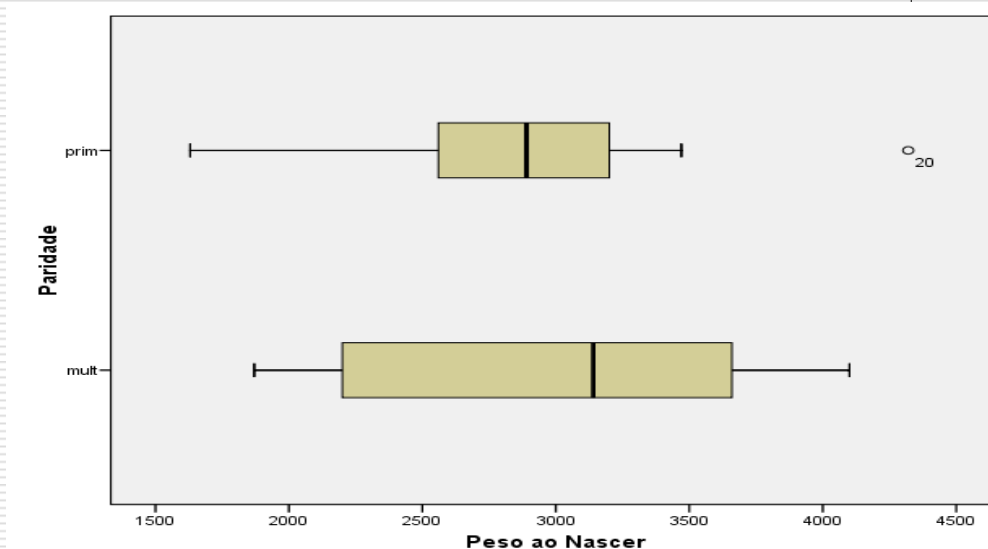


Medidas de Dispersão (de Variabilidade)

❖ Medidas Centradas na Mediana

* Amplitude Interquartilica:

* Amplitude Semi Interquartílica (AISQ):



$$Q = Q_3 - Q_1$$

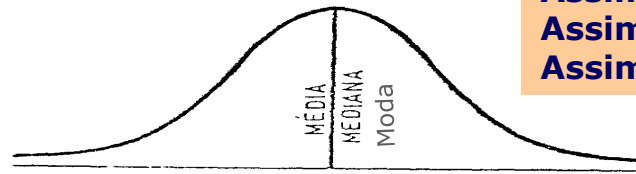
$$q = \frac{Q_3 - Q_1}{2}$$

Q contém os
50% dos
elementos mais
centrais.

* MAD (Mediana dos Desvios Absolutos em Relação à Mediana)

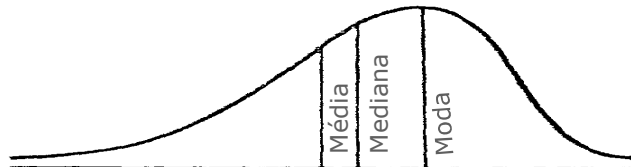
Medidas de Assimetria (enviesamento)

Assimetria Ligeira: $Ap < 0,5$
Assimetria Moderada: $0,5 < Ap < 1$
Assimetria Grave: $Ap > 1$



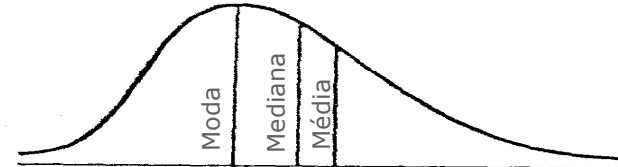
$$\bar{X} - Mo = 0$$

Assimetria nula



$$\bar{X} - Mo < 0$$

Assimetria negativa



$$\bar{X} - Mo > 0$$

Assimetria Positiva

* Coeficiente de Assimetria de Pearson:

$$Ap_1 = \frac{\bar{X} - Mo}{s}$$

$$Ap_2 = \frac{3(\bar{X} - Md)}{s}$$

* Coeficiente de Assimetria (SPSS):
(baseado no 3º Momento)

$$Sk = \frac{n \sum F_i (x_i - \bar{X})^3}{(n-1)(n-2) s^3}$$

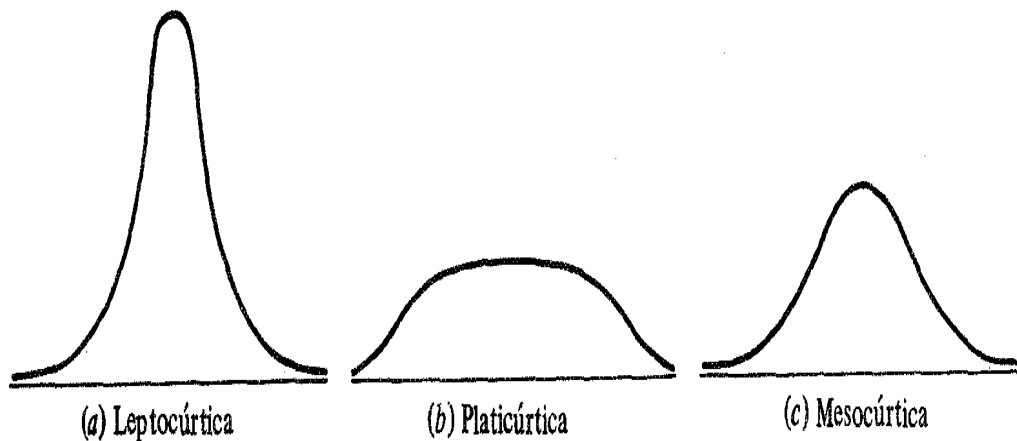
Medidas de Achatamento (Curtose)

* Coeficiente de Percentílico de Curtose:

$$K = \frac{0,5(Q_3 - Q_1)}{P_{90} - P_{10}}$$

* Coeficiente de curtose (SPSS):
(baseado no 4º Momento)

$$g_2 = \frac{n^2(n+1) \frac{\sum (xi - \bar{X})^4}{n}}{(n-1)(n-2)(n-3) s^3}$$



$$K > 0,263$$

$$g_2 > 0$$

$$K < 0,263$$

$$g_2 < 0$$

$$K = 0,263$$

$$g_2 = 0$$

Análise da Associação entre Duas Variáveis

- Tabelas de contingência e diagramas de dispersão
 - Coeficientes de associação / correlação entre duas variáveis
 - A recta de regressão [$y=f(x)$]
-

Análise da Associação entre Duas Variáveis (Qualitativas)

□ Tabelas de Contingência

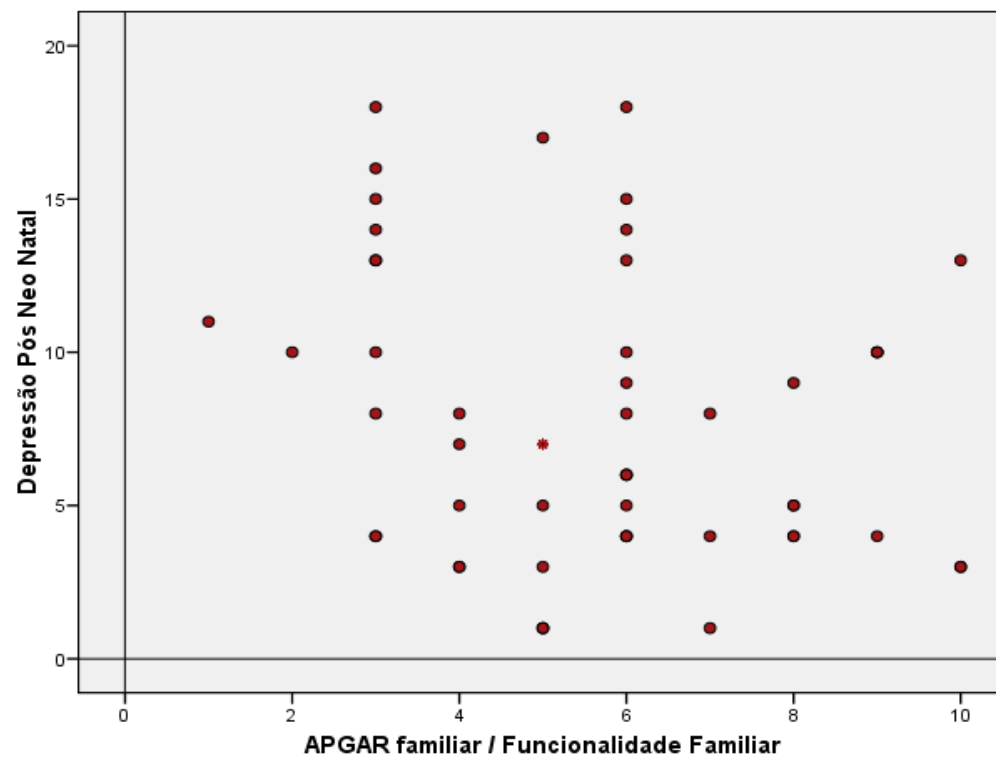
EST_SOCIO_ECONÓMICO	C_TABACO
médio/alto e alto	sim
médio/alto e alto	sim
médio/alto e alto	sim
médio	não
médio/baixo e baixo	sim
e alto	sim
e alto	sim
	sim
e baixo	não
	sim
e alto	sim
e alto	sim
e alto	sim
	...

Consumo de Tabaco * Nível Sócio-económico

Consumo de Tabaco	Nível Socio-económico						Total	
	médio/baixo e baixo		médio		médio/alto e alto			
	F	%	F	%	F	%	F	%
sim	22	55,00	70	53,03	33	64,71	125	56,05
não	18	45,00	62	46,97	18	35,29	98	43,95
Total	40	100,00	132	100,00	51	100,00	223	100,00

Análise da Associação entre Duas Variáveis (Quantitativas)

□ Diagramas de Dispersão



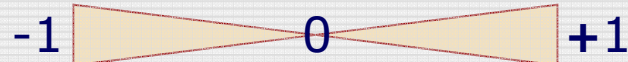
FUNCIONALIDADE FAMILIAR	DEPRESSÃO
3	13
5	1
4	3
6	4
3	4
9	10
6	18
5	7
3	13
6	6
6	10
1	11
6	8
5	1
6	4
...	...

Análise da Associação entre Duas Variáveis

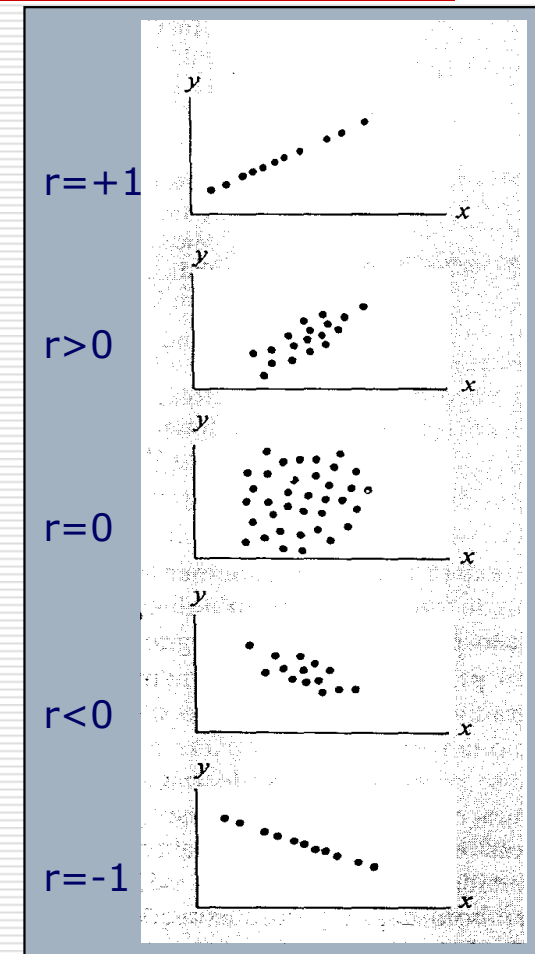
- ❖ **Medida de Associação entre Variáveis Intervalares ou Racionais (Correlação de Bravais-Pearson)**

▲ Relação Linear

$$r = \frac{S_{xy}}{S_x S_y} = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$



Coeficiente de Determinação: r^2
(variância explicada)



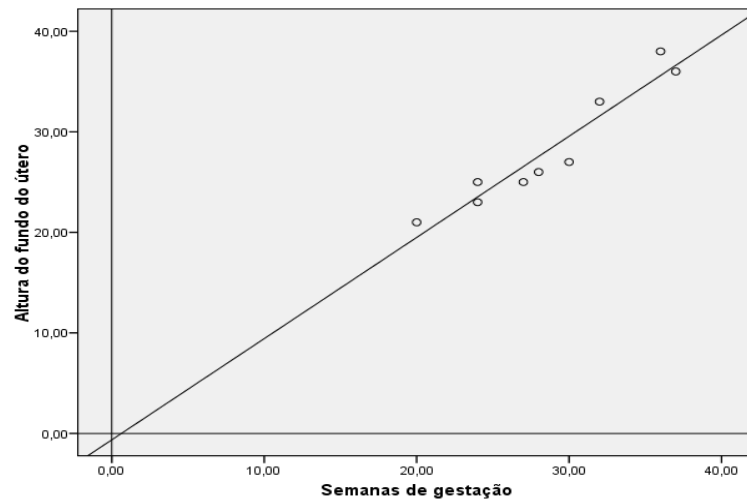
▲ Se os pontos se dispuserem de forma não linear (correlação curvilínea), o r resulta num valor inferior ao real.

Análise da Associação entre Duas Variáveis

❖ Correlação de Bravais-Pearson

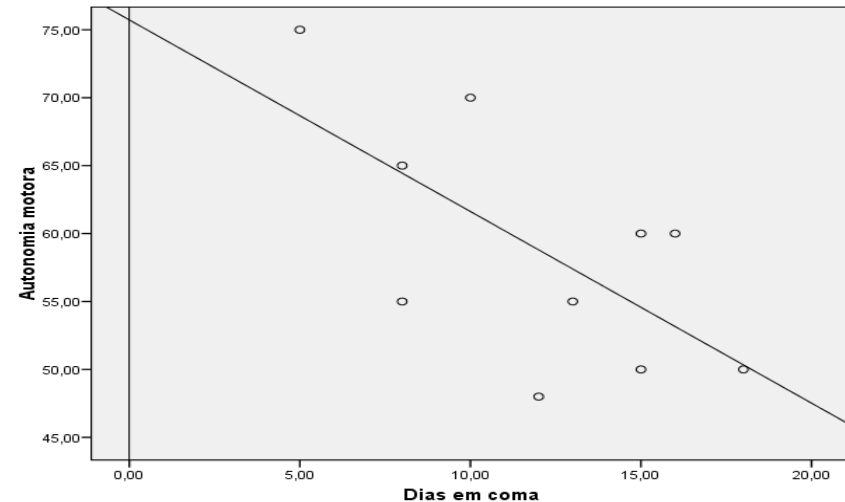
S. G.	A.F.U
20	21
24	23
24	25
27	25
28	26
30	27
32	33
36	38
37	36

$n = 9$
 $r = 0,96$
 $r^2 = 0,92$
 $(r^{2\%} = 91,52\%)$



D. Coma	A.M.
5	75
8	65
8	55
10	70
16	60
12	48
15	50
13	55
18	50
15	60

$n = 10$
 $r = -0,65$
 $r^2 = 0,43$
 $(r^{2\%} = 42,52\%)$



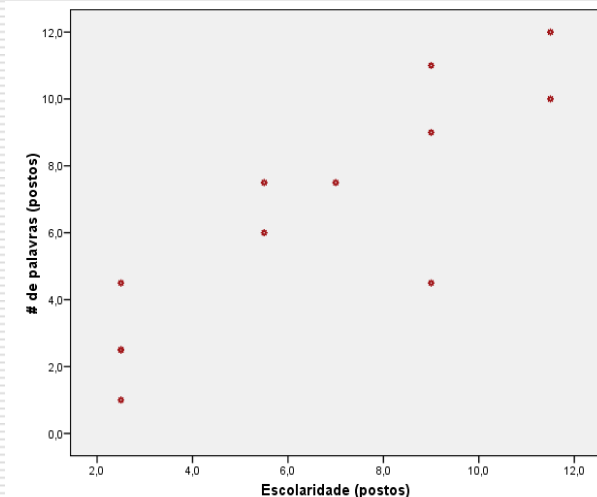
Análise da Associação entre Duas Variáveis

❖ Medida de Associação entre Variáveis Ordinais (Exemplo: Correlação de Spearman)

Doente	Escolaridade	Posto	Palavras	Posto
A	1º Ciclo	2,5	5	1,0
B	Secundário	9,0	8	4,5
C	3º Ciclo	7,0	10	7,5
D	Superior	11,5	12	10,0
E	1º Ciclo	2,5	6	2,5
F	1º Ciclo	2,5	8	4,5
G	2º Ciclo	5,5	10	7,5
H	Superior	11,5	15	12,0
I	Secundário	9,0	14	11,0
J	2º Ciclo	5,5	9	6,0
L	1º Ciclo	2,5	6	2,5
M	Secundário	9,0	11	9,0

$$r_s = 1 - \frac{6 \times 37,5}{12^3 - 12} = 0,87$$

$$r_s = 1 - \frac{6 \sum d^2}{n^3 - n}$$



▲ Com várias observações com ordens empatadas, recomenda-se a utilização de uma fórmula própria.

Análise da Associação entre Duas Variáveis Nominais

❖ Exemplos:

Características do parto * Paridade				
		Paridade		Total
		Múltipara	Primipara	
Características do parto	cesariana			
	Fo	2	8	10
	Fe	4,44	5,56	10,00
	%	8,33	26,67	18,52
via vaginal	Resíduos (Fo-Fe)	-2,44	2,44	
	Fo	22	22	44
	Fe	19,56	24,44	44,00
	%	91,67	73,33	81,48
Total	Resíduos (Fo-Fe)	2,44	-2,44	
	Fo	24	30	54
	Fe	24,00	30,00	54,00
	%	100,00	100,00	100,00

V de Cramér:

$$V = \sqrt{\frac{\chi^2}{n[\min.(C,L) - 1]}}$$

⊖

Coeficiente de Contingência:

$$C C = \sqrt{\frac{\chi^2}{n \chi^2}}$$

$$\chi^2 = \sum \frac{(Fo - Fe)^2}{Fe}$$

φ = 0,235

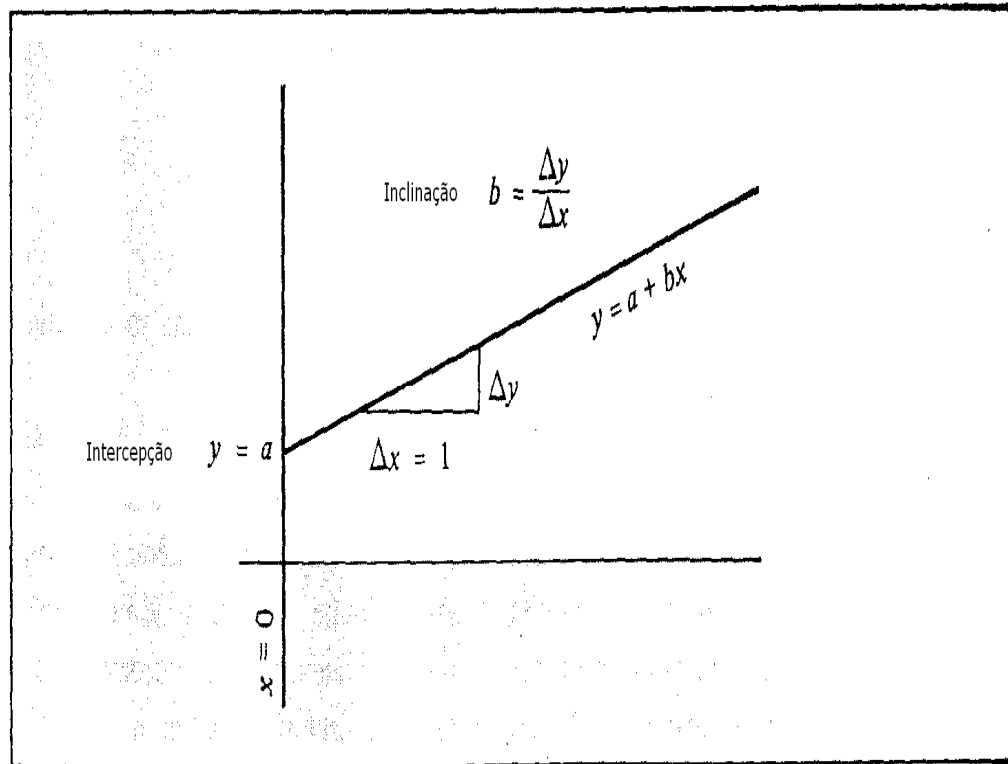
Odds Rácio = 0,25

⊖ Em tabelas de L=2 e C=2 (2x2), designa-se por φ (phi).

$$\Phi = \sqrt{\frac{\chi^2}{n}}$$

A recta de regressão [$y=f(x)$]

$$y = a + bx$$



▲ A recta dos mínimos quadrados (**recta de regressão**) é na grande maioria das relações bivariadas a linha que melhor se ajusta aos dados. Mas, outras situações se verificam em que uma linha curva será mais ajustada (**regressão quadrática; regressão cúbica**).

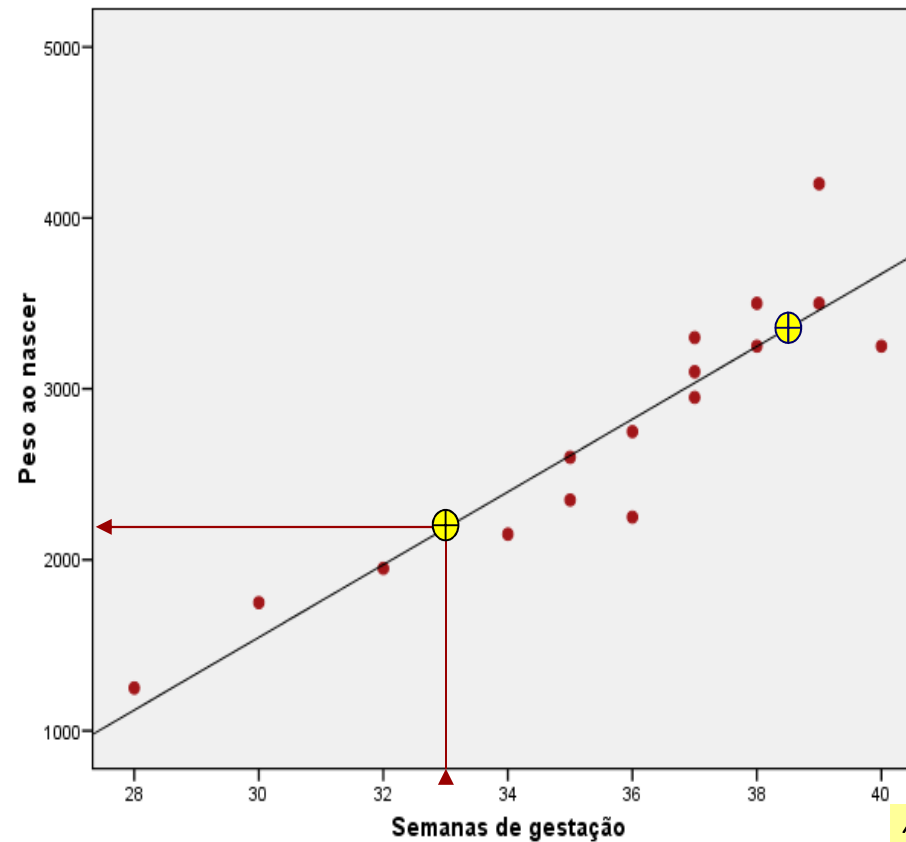
$$b = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$a = \bar{X} - b \bar{Y}$$

A recta $y = a + bx$ é diferente da recta $x = a + by$. Estas rectas cruzam-se no ponto $(\bar{X}, \bar{Y})$.

A recta de regressão [$y=f(x)$]

SG	Peso
28	1250
30	1750
32	1950
34	2150
35	2350
35	2600
36	2250
36	2750
37	3100
37	3300
37	2950
38	3500
38	3250
39	4200
39	3500
40	3250



$$r=0,92$$

$$r^2\%=84,37\%$$

$$a=-4833$$

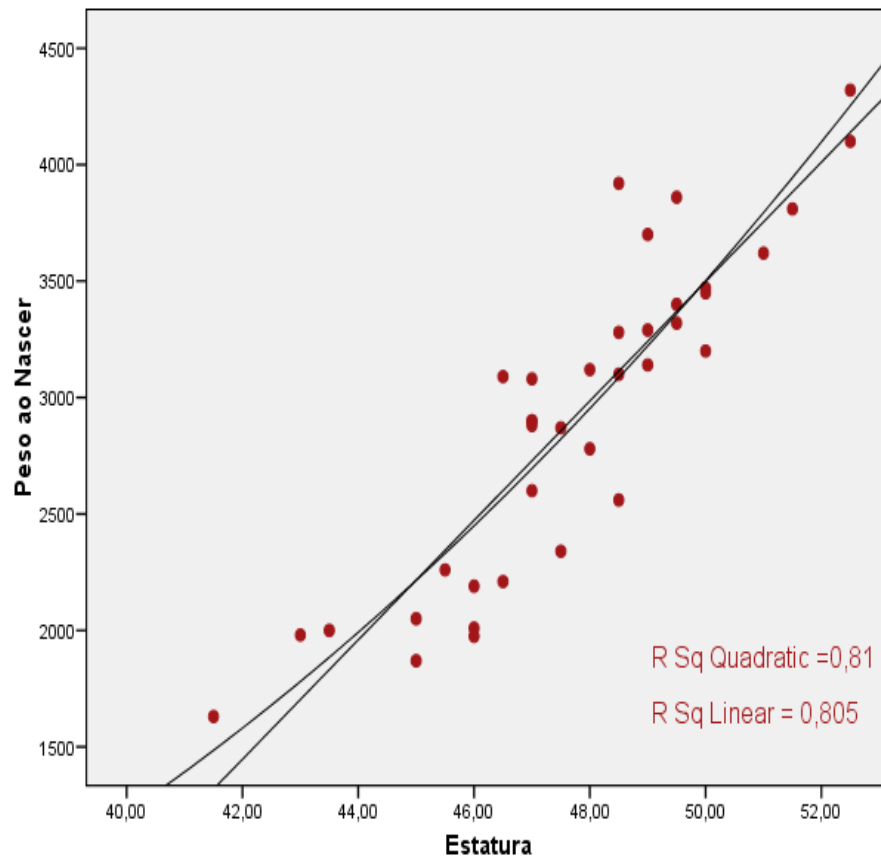
$$b=212$$

$$\hat{y} = -4833 + 212x$$

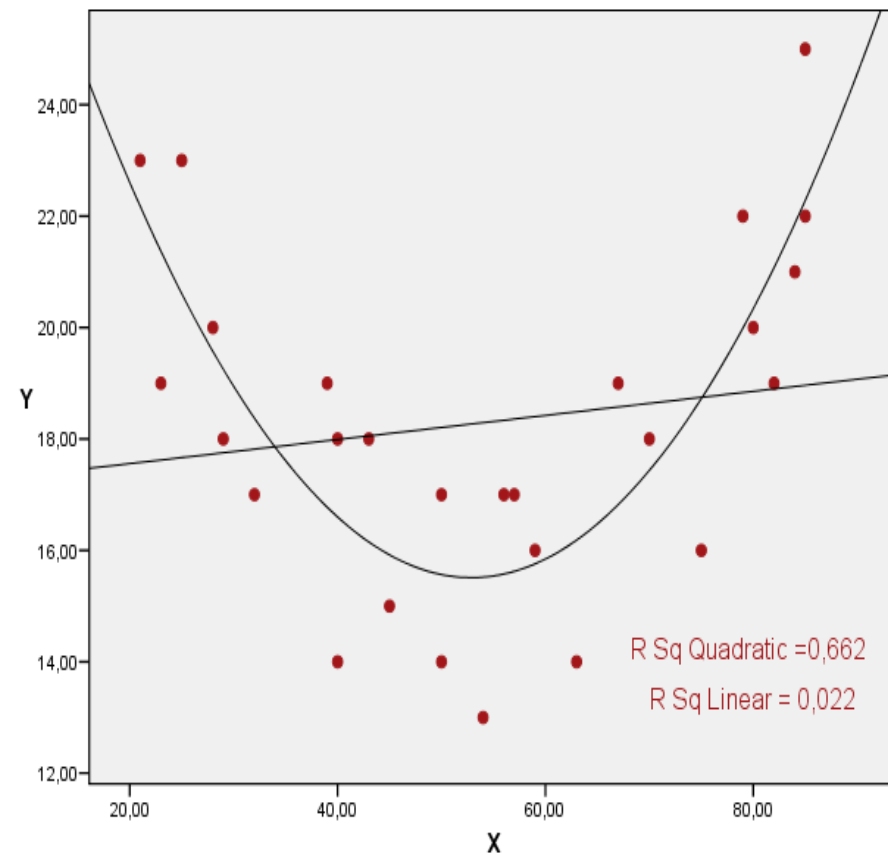
Q: Qual o peso previsível de um recém-nascido de 33 semanas de gestação?

R: 2189 gr

Regressão Curvilínea



Regressão linear



Regressão Quadrática

Apresentação dos Dados Estatísticos

- ❖ Tabelas
 - ❖ Quadros
 - ❖ Gráficos
-

Apresentação dos Dados Estatísticos

As **tabelas** são utilizadas para apresentar dados organizados. "Abertas" ao cálculo de medidas consideradas úteis.

"TABELAS DE DADOS"

Ex:

Tabela 1 – Distribuição das puérperas seguidas no centro de Saúde da Foz, no ano de 2007, segundo o grau de depressão às 20 semanas pós parto.

Depressão ^a	n. ^o	%
00 - 03	4	7,41
03 - 06	19	35,19
06 - 09	4	18,52
09 - 12	9	16,67
12 - 15	6	11,11
15 - 18	4	7,41
18 ou +	2	3,70
Total	54	100,00

a. Medida com a EDPNN de Edimburgo.

≠

Os **quadros** são reservados para a apresentação de resultados e de sínteses e de cálculos baseados nos dados. "Fechados" a novos cálculos.

"QUADRO DE RESULTADOS"

Ex:

Quadro 1 – Estatísticas de resumo relativas à funcionalidade familiar e depressão pós neonatal das puérperas seguidas no centro de Saúde da Foz, no ano de 2007, segundo o grau de depressão às 20 semanas pós parto.

		Paridade (v. numérica)		
		Múltipara	Primipara	Total
APGAR	n	24	30	54
familiar /	Média	5,96	5,37	5,63
Funcionalida	SD	1,829	2,470	2,209
de Familiar	Mediana	6,00	5,00	6,00
	Min.	2	1	1
	Max.	10	10	10
Depressão	n	24	30	54
Pós Neo	Média	6,17	9,03	7,76
Natal	SD	4,860	4,230	4,702
(EDPNN	Mediana	4,00	8,00	7,00
Edimburgo)	Min.	1	3	1
	Max.	18	18	18

Tabelas (elementos constituintes)

ELEMENTOS ESSENCIAIS

Titulo: Precede a tabela. Designa a(s) variável(eis) em estudo; indica as condições em que foram obtidos os dados (local, data).

Corpo: Conjunto de linhas e colunas onde os dados são representados.

Cabeçalho: É aparte superior da tabela. Especifica o conteúdo das colunas.

Coluna Indicadora: É a parte da tabela que especifica o conteúdo das linhas.

Fonte: Indicada imediatamente a seguir à tabela. Identifica a entidade que publicaram ou forneceram os dados. Omite-se se os dados são propriedade do autor.

ELEMENTOS COMPLEMENTARES (em rodapé da tabela)

Notas: Questões mais gerais. Indicadas com algarismos ou letras minúsculas entre parêntesis, à direita na coluna e à direita ou em baixo no cabeçalho.

Chamadas: Questões específicas. Indicadas à esquerda dos dados.

SINAIS CONVENCIONAIS

- O fenómeno não se registou (o dado é nulo).

? Quando há dúvidas sobre a precisão da observação.

X Quando a informação é sigilosa (secreta).

Tabela 2 – Distribuição dos casos de SIDA em Portugal, por ano de diagnóstico, no período de 1982 a 1986.

Anos	N.º de casos
1982	a)-
1983	a)1
1984	a)1
1985	27
1986	21
Total	50

Fonte: I.N.E. Grupo de Trabalho da Sida (1990).

a) Não inclui os casos de diagnóstico inseguro.

▲ Quando se incluem várias tabelas num relatório, estas devem ser numeradas para facilitar a sua referência e localização.

Tabelas (tipos)

Tabelas Simples (uma variável)

Tabela 3
Distribuição da amostra segundo a paridade.

Paridade	n. ^o	%
1	18	48,65
2	10	27,03
3	5	13,51
4	3	8,11
5	1	2,70
Total	37	100,00

Tabelas Compostas - Tabelas de Contingência - (duas ou mais variáveis)

Tabela 4

Distribuição das puérperas da amostra segundo os hábitos tabágicos e o tempo de gestação.

		Hábitos Tabágicos				Total	
		Não		Sim			
		n.º	%	n.º	%	n.º	%
Tempo de Gestação (Semanas)	30 - 32	2	7,69	0	,00	2	5,41
	33 - 35	4	15,38	2	18,18	6	16,22
	36 - 38	10	38,46	6	54,55	16	43,24
	39 - 41	9	34,62	3	27,27	12	32,43
	42 ou +	1	3,85	0	,00	1	2,70
Total		26	100	11	100	37	100

Quadros

A construção dos quadros obedece a regras idênticas às das tabelas, mas

- São delimitados (fechados) em todos os lados,
- O título pode ser colocado por baixo.

Quadro 2 – Matrix de correlações de Pearson entre a idade, paridade, apgar familiar e depressão pós neonatal das puérperas seguidas no centro de Saúde da Foz, no ano de 2007.

		Idade	Paridade	APGAR Familiar / Funcionalidade Familiar
Idade	r =	-		
Paridade	r =	,545	-	
	p =	,000		
APGAR familiar / Funcionalidade Familiar	r =	,025	,081	-
	p =	,857	,561	
Depressão Pós NeoNatal	r =	-,222	-,274	-,243
	p =	,107	,045	,077

Gráficos (Diagramas)

A apresentação gráfica dos dados é uma técnica estatística que:

- Evidencia a importância relativa das categorias ou classes em relação ao total;
- Facilita a comparação de valores relativos observados para a mesma variável, mas
 - Em amostras de n diferente;
 - Em grupos com características diferentes;
 - Colhidos em locais ou tempos diferentes.
- Facilita a comparação de “comportamentos” de variáveis afins.
- Permite em casos específicos visualizar a evolução de certas variáveis em determinado tempo ou espaço.

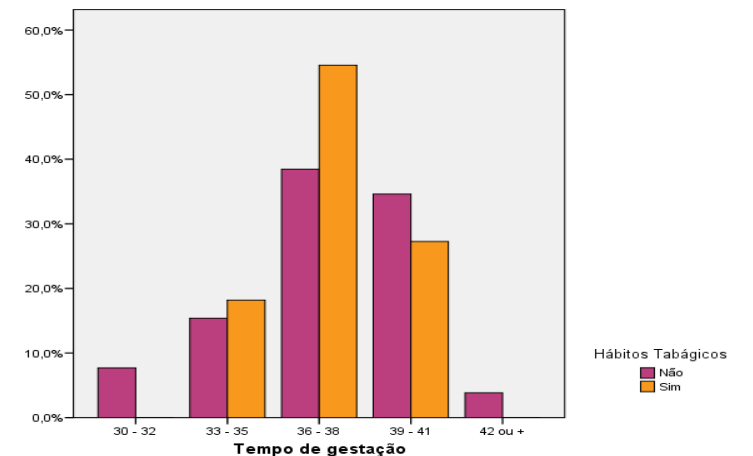


Gráfico 1 - Distribuição das puerperas da amostra segundo os hábitos tabágicos e o tempo de gestação.

▲ Não confundir técnica estatística com “marketing”.

Gráficos (Diagramas)

- ❑ Operacionalmente, salvo exceções, um gráfico é precedido da respectiva tabela (ou quadro).
- ❑ Só se elaboram gráficos quando esse tipo de representação acrescenta capacidade analítica.

A escolha do tipo de gráfico é feita em função de:

❖ **Características da variável**

(Nominal, Ordinal, Quantitativa Discreta, Quantitativa Contínua)

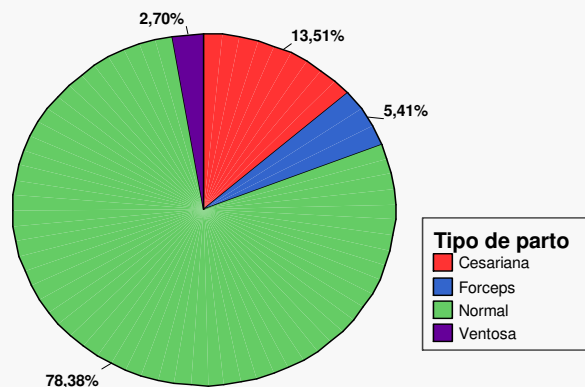
❖ **Características da série**

(de frequências simples, de frequências acumuladas, cronológicas, espaciais)

❖ **Objectivos da apresentação**

❖ - ...

Gráfico 2
Distribuição da amostra de acordo com o tipo de parto.



Gráficos (Diagramas)

❖ Elementos constituintes:

- **Titulo e Numeração:** Aplicam-se as mesmas considerações feitas para as tabelas (pode ser colocado por baixo do corpo do gráfico).
- **Escala:** Devem ser definidas sem ambiguidade. A percentagem deve ser a frequência preferida em amostras grandes ($n > 50$).
- **Legendas:** Se necessário, devem ser colocadas à direita ou em baixo do gráfico.
- **Fonte:** Como em relação às tabelas.

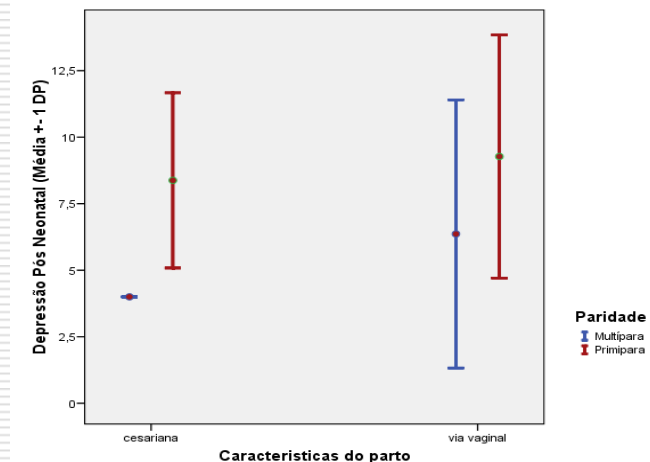


Gráfico 3 - Comparação do grau de depressão pós neonatal das puérparas seguidas no Centro de Saúde da Foz, segundo o tipo de parto e a paridade.

Gráficos Cartesianos

- Definidos num sistema de eixos cartesianos.

DE BARRAS (OU COLUNAS) SEPARADAS

- ⊗ [apropriados para variáveis qualitativas ordinais e quantitativas discretas]*

Gráfico 4 – Distribuição dos elementos da amostra de acordo com o nível socioeconómico.

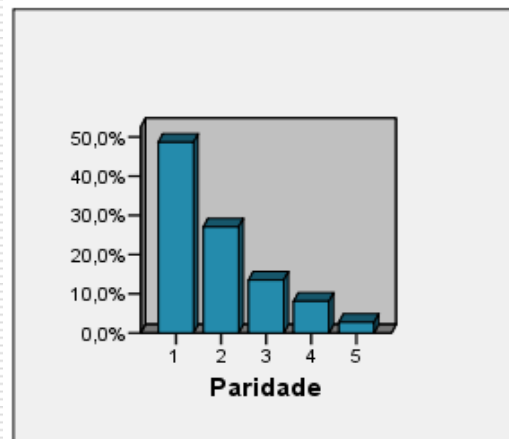
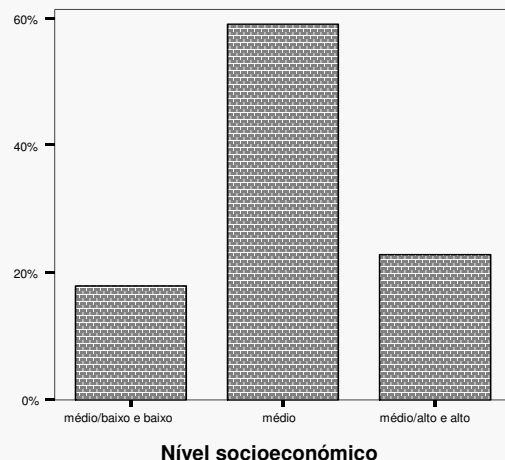
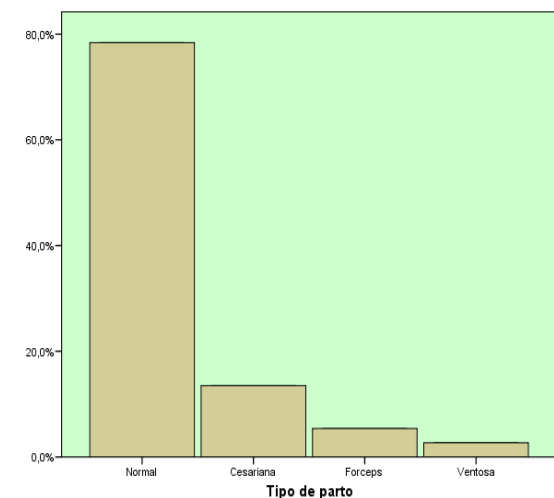


Gráfico 3 – Distribuição das puérperas da amostra segundo a paridade.

- * Também para variáveis nominais, destacando as diferenças de frequências entre as categorias (gráficos de Pareto)

Gráfico 5

Distribuição das puérperas da amostra por tipo de parto.



Gráficos Cartesianos

DE BARRAS (OU COLUNAS) EM CLUSTERS / SOBREPOSTAS

- ✕ **Apropriados para representar a relação entre variáveis qualitativas ou com os dados classificados.**

Gráfico 6 – Distribuição dos inquiridos segundo a ocupação, por sexo.

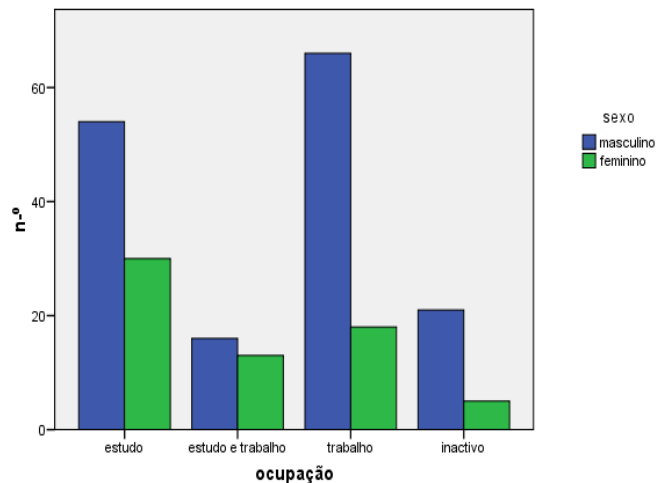
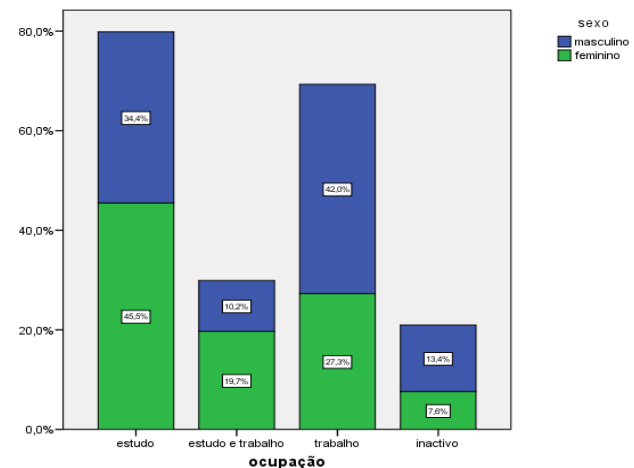


Gráfico 7 – Distribuição dos inquiridos segundo a ocupação, por sexo.



Gráficos Cartesianos

DE ÁREAS

☒ **Apropriados para variáveis quantitativas contínuas**

Histograma:

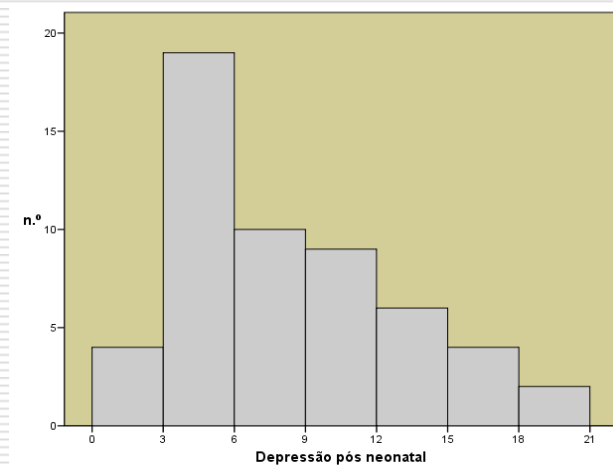


Gráfico 8 - Distribuição das puérperas da amostra segundo os valores da EDPNN de Edimburgo.

Polígono de Frequências:

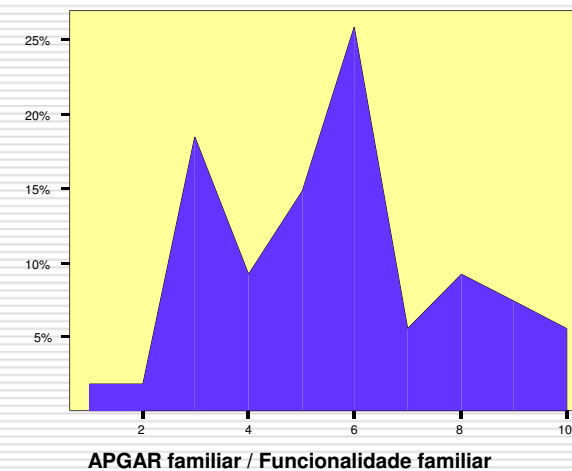


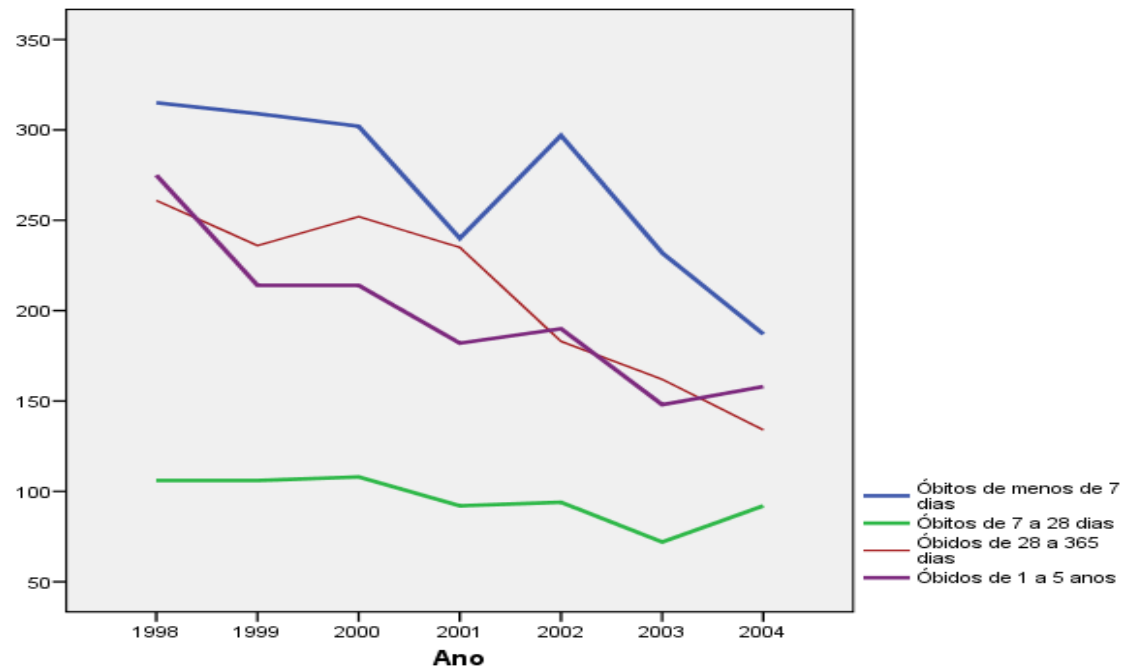
Gráfico 9 - Distribuição das puérperas da amostra segundo os valores do APGAR familiar-

Gráficos Cartesianos

LINEARES

☒ **Apropriados para séries temporais**

Gráfico 10 – Evolução dos óbitos infantis em Portugal, de 1998 a 2004



Fonte: INE – Estatísticas Demográficas.

Gráficos Circulares

DE SECTORES DE CÍRCULO

- ☒ **Apropriados para variáveis nominais**

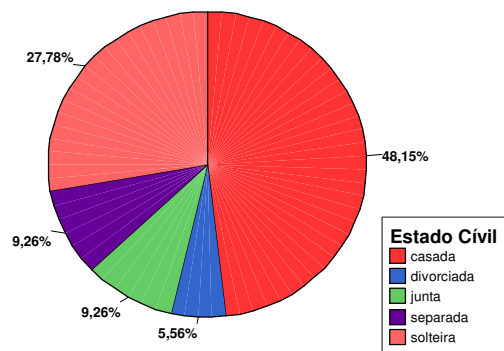
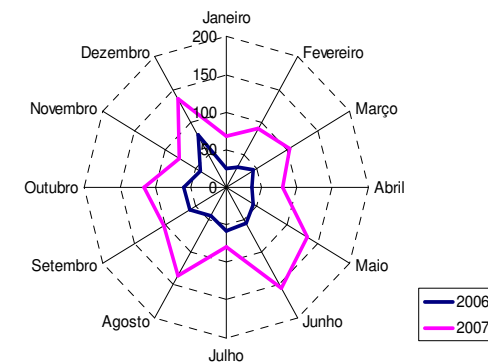


Gráfico 11 – Distribuição das puerperas da amostra segundo o estado civil.

RADIAIS (OU POLARES)

- ☒ **Apropriados para variáveis temporais cíclicas**

Gráfico 12 – Evolução do nº de PETs requisitados pelo Hospital Oncológico da Estrela nos anos de 2006 e 2007, por mês.

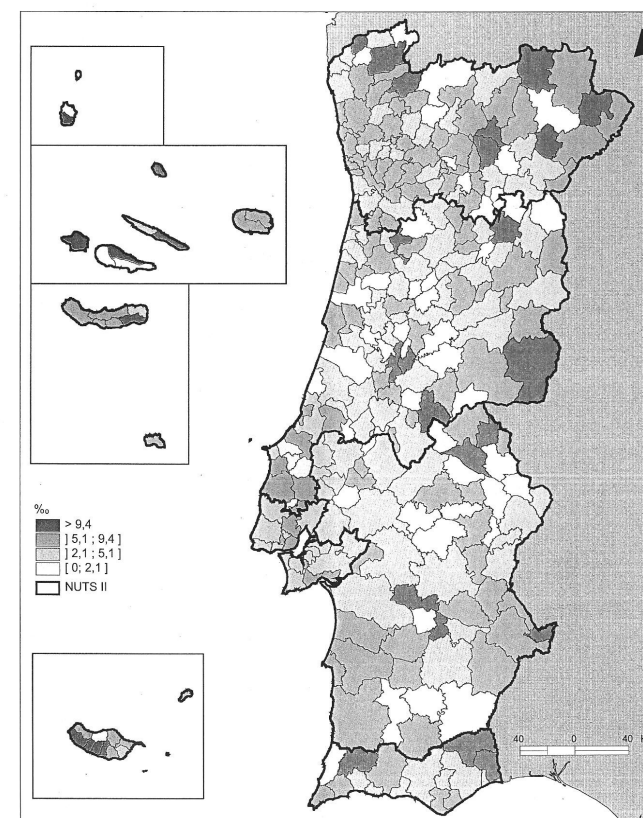


Fonte: Relatório de Actividades do HOE, 2007.

Cartogramas

- ⊗ **Apropriados para representar a frequência ou a densidade relativamente a uma variável espacial (geográfica)**

Gráfico 13 – Taxa média de mortalidade infantil, por concelho, 1999/2003.



Fonte: INE, 2004.

PROBABILIDADES

- NOÇÕES GERAIS

- Conceitos fundamentais; origens.
- Axiomática das probabilidades
- Teorema do produto e teorema da soma (acontecimentos independentes / dependentes; acontecimentos mutuamente exclusivos / não mutuamente exclusivos)

- DISTRIBUIÇÕES DE PROBABILIDADES

- Variáveis aleatórias (valor esperado - $E(X)$ -; dispersão - $\sigma(X)$ -)
 - Distribuições teóricas (binomial, de Poisson, normal)
-

Probabilidade:

Noção de previsibilidade fundamentada definida pelo grau de certeza / incerteza da ocorrência de um determinado acontecimento.

Teoria das Probabilidades

(Estudo dos fenómenos aleatórios)

Experiência aleatória *versus* Experiência determinística

- Incerteza** pontual
- Regularidade** a longo prazo

Origens das probabilidades:

Objectiva

Clássica (intuitiva: resultados equiprováveis e mutuamente exclusivos)

Empírica (dados históricos – observações / experiência – “fazendo fé em que o passado é representativo do futuro”)

Subjectiva

Axiomática das Probabilidades

Resultado de uma experiência singular:

- **Acontecimento elementar – (e) – Ponto amostral**

Conjunto de todos os resultados possíveis de uma experiência aleatória:

- **Universo da experiência – (Ω) – Espaço amostral**

Conjunto dos resultados possíveis que satisfazem as condições de A:

- **Acontecimento – (A) – Subconjunto do espaço amostral**

PROPRIEDADES:

Para todo o acontecimento A:

$$0 \leq p(A) \leq 1$$

Se A for um acontecimento certo:

$$p(A) = 1$$

Se A for um acontecimento impossível:

$$p(A) = 0$$

A probabilidade de que o acontecimento A não ocorra:

$$p(\tilde{A}) = 1 - p(A)$$

$$p(A) + p(\tilde{A}) = 1 \quad ::::: \quad p + q = 1$$

Operações com probabilidades

Origem intuitiva (clássica):
(baseada em eventos equiprováveis)

$$p(A) = \frac{\# A}{\# \Omega}$$

Origem empírica:
(baseada nas frequências relativas)

$$p(A) \approx f(A) \approx \frac{F(A)}{n}$$

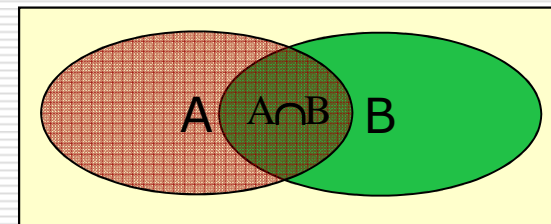
quando $n \rightarrow \infty$

Operações com probabilidades

TEOREMA DO PRODUTO:

Se os acontecimentos são independentes:

$$p(A \cap B) = p(A) \times p(B)$$



Se os acontecimentos não são independentes:

$$p(A \cap B) = p(A) \times p(B \setminus A)$$

ou

$$p(A \cap B) = p(B) \times p(A \setminus B)$$

Probabilidade condicional:

$$p(B \setminus A) = \frac{p(B \cap A)}{P(A)}$$

$$p(A \setminus B) = \frac{p(A \cap B)}{P(B)}$$

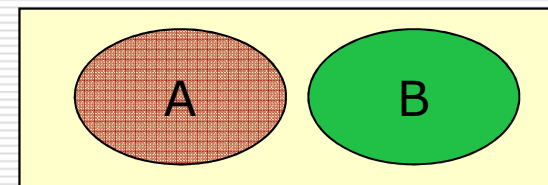


Operações com probabilidades

TEOREMA DA SOMA:

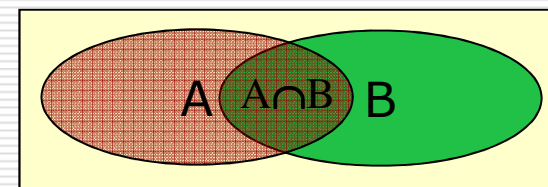
Se os acontecimentos são mutuamente exclusivos:

$$p(A \cup B) = p(A) + p(B)$$



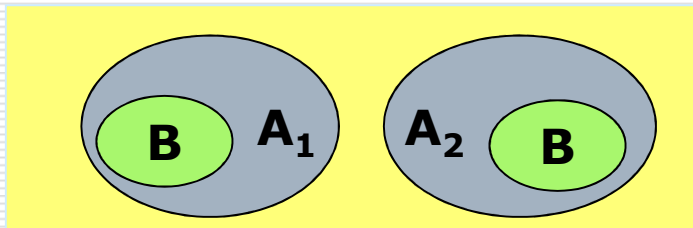
Se os acontecimentos são não mutuamente exclusivos:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$



Operações com probabilidades

TEOREMA DA PROBABILIDADE TOTAL:



$$p(B) = p(B \cap A_1) + p(B \cap A_2) + \dots + p(B \cap A_k)$$

$\Leftrightarrow$

$$p(B) = \sum_{k=1}^k p(A_k) \times p(B \setminus A_k)$$

TEOREMA DE BAYES:

(revisão das probabilidades à posteriori)

$$p(A_1 \setminus B) = \frac{p(A_1) \times p(B \setminus A_1)}{\sum_{k=1}^k p(A_k) \times p(B \setminus A_k)} \Leftrightarrow \frac{p(A_1) \times p(B \setminus A_1)}{p(A_1) \times p(B \setminus A_1) + p(\tilde{A}_1) \times p(B \setminus \tilde{A}_1)}$$

Variáveis Aleatórias

- ❖ Variáveis em que os resultados ou valores diferem de uma observação para a outra em razão de factores não controláveis – aleatórios –.
- ❖ Por razões práticas, é desejável que se defina uma variável aleatória associada a uma amostra ou a uma experiência de tal modo que os seus resultados possíveis sejam numéricos.

Por definição, a cada valor $x_1, x_2, \dots, x_n$, assumidos pela variável aleatória, correspondem as respectivas probabilidades $p(x_1), p(x_2), \dots, p(x_n)$.

Ex(1) – Seja a variável aleatória (V. A.) X o número de caras resultado do lançamento de uma moeda duas vezes:

x_i	0	1	2
$p(x_i)$	0,25	0,50	0,25

$$\sum_{i=1}^n p(x_i) = 1,00$$

Ex(2) – Seja a V. A. X o número de doentes com traumatismo vertebro-medular assistidos semanalmente no Serviço de Urgências de um determinado hospital:

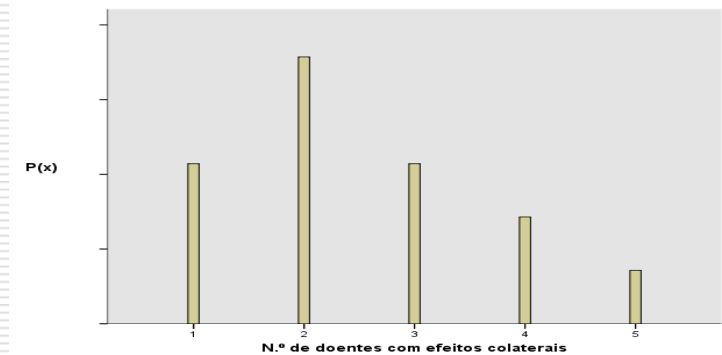
x_i	0	1	2	3	4
$p(x_i)$	0,58	0,31	0,08	0,02	0,01

$$\sum_{i=1}^n p(x_i) = 1,00$$

Variáveis Aleatórias

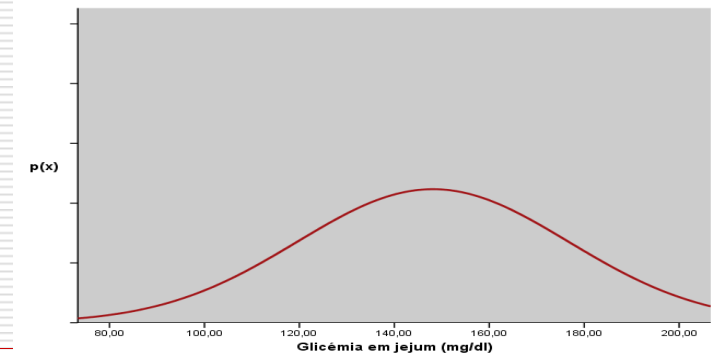
- Variáveis aleatórias discretas: Se os valores prováveis podem ser contados.

Ex.: n.º de doentes que sofrem um determinado efeito colateral negativo de um medicamento M.



- Variáveis aleatórias contínuas: Se são possíveis quaisquer valores num dado intervalo.

Ex.: Valores de glicémia em jejum de doentes com diabetes tipo II.



Variáveis Aleatórias

Valor esperado de uma experiência aleatória: esperança matemática $E(X)$.
Corresponde à média (μ) dos respectivos pontos amostrais:

$$E(X) = \mu = x_1p(x_1) + x_2p(x_2) + \dots + x_n p(x_n) = \sum_{i=1}^n x_i p(x_i)$$

Ex.(1): Lançamento de uma moeda 4 vezes.

P: Qual o número esperado de caras?

x_i	$p(x_i)$	$x_i p(x_i)$
0	0,0625	0,0000
1	0,2500	0,2500
2	0,3750	0,7500
3	0,2500	0,7500
4	0,0625	0,2500
	1,0000	2,0000

R: em quatro lançamentos, “espera-se” que dois (2) resultem cara.

Ex.(2): Acidentes nos primeiros três anos de apólice. **P:** Quantos acidentes é de esperar por apólice?

x_i	$p(x_i)$	$x_i p(x_i)$
0	0,58	0,00
1	0,31	0,31
2	0,08	0,16
3	0,02	0,06
4	0,01	0,04
	1,00	0,57

R: “Espera-se” em média 0,57 acidentes por apólice nos 1^{os} três anos.

Variáveis Aleatórias

Variância $\sigma^2(X)$ e Desvio Padrão de uma experiência aleatória $\sigma(X)$:

$$\sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 p(x_i) \quad \sigma = \sqrt{\sum_{i=1}^n (x_i - \mu)^2 p(x_i)} = \sqrt{\sum_{i=1}^n x_i^2 p(x_i) - \mu^2}$$

Ex.(1): Lançamento de uma moeda 4 vezes.

P: Qual o desvio padrão?

x_i	$p(x_i)$
0	0,0625
1	0,2500
2	0,3750
3	0,2500
4	0,0625
1,0000	

R: $\sigma = 1,00$

Ex.(2): Acidentes nos primeiros três anos de apólice. **P:** Qual desvio padrão?

x_i	$p(x_i)$
0	0,58
1	0,31
2	0,08
3	0,02
4	0,01
1,00	

R: $\sigma = 0,80$

Distribuição Binomial

➔ **Modelo probabilístico que representa o “comportamento” de grande parte das variáveis aleatórias discretas.**

Requisitos:

- 1 – Podem realizar-se n observações ou experiências;
- 2 – Cada experiência tem dois resultados possíveis (um chamado sucesso – ocorre o acontecimento A – e o outro de fracasso – não ocorre o acontecimento A);
- 3 – As probabilidades de sucesso (p) e as probabilidades de fracasso ($1-p = q$) permanecem constantes para todas as experiências;
- 4 – Os resultados das experiências são independentes uns dos outros.

Nestes casos, o número de vezes (X) que ocorre um acontecimento A num conjunto n de experiências é uma V. A. em que:

$$X \sim \beta(n, p)$$

$$\text{com } \begin{aligned} \mu &= np \\ \sigma &= \sqrt{np(1 - p)} \end{aligned}$$

A probabilidade de cada valor x - $p(x)$ - pode ser calculado através da seguinte fórmula:

$$P(X = x) = C_x^n \times p^x \times q^{n-x} \quad \text{em que} \quad C_x^n = \frac{n!}{x!(n-x)!}$$

Probabilidades binomiais acumuladas

x	p																		
	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65	.70	.75	.80	.85	.90	.95
0	.9500	.9000	.8500	.8000	.7500	.7000	.6500	.6000	.5500	.5000	.4500	.4000	.3500	.3000	.2500	.2000	.1500	.1000	.0500
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500	.2025	.1600	.1225	.0900	.0625	.0400	.0225	.0100	.0025
1	.9975	.9900	.9775	.9600	.9375	.9100	.8775	.8400	.7975	.7500	.6975	.6400	.5775	.5100	.4375	.3600	.2775	.1900	.0975
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250	.0911	.0640	.0429	.0270	.0156	.0080	.0034	.0010	.0001
1	.9928	.9720	.9393	.8960	.8438	.7840	.7183	.6480	.5748	.5000	.4253	.3520	.2818	.2160	.1563	.1040	.0608	.0280	.0073
2	.9999	.9990	.9966	.9920	.9844	.9730	.9571	.9360	.9089	.8750	.8336	.7840	.7254	.6570	.5781	.4880	.3859	.2710	.1426
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625	.0410	.0256	.0150	.0081	.0039	.0016	.0005	.0001	.0000
1	.9860	.9477	.8905	.8192	.7383	.6517	.5630	.4752	.3910	.3125	.2415	.1792	.1265	.0837	.0508	.0272	.0120	.0037	.0005
2	.9995	.9963	.9880	.9728	.9492	.9163	.8735	.8208	.7585	.6875	.6090	.5248	.4370	.3483	.2617	.1808	.1095	.0523	.0140
3	1.0000	.9999	.9995	.9984	.9961	.9919	.9850	.9744	.9590	.9375	.9085	.8704	.8215	.7599	.6836	.5904	.4780	.3439	.1855
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0313	.0185	.0102	.0053	.0024	.0010	.0003	.0001	.0000	.0000
1	.9974	.9185	.8352	.7373	.6328	.5282	.4284	.3370	.2562	.1875	.1312	.0870	.0540	.0308	.0156	.0067	.0022	.0005	.0000
2	.9988	.9914	.9734	.9421	.8965	.8369	.7648	.6826	.5931	.5000	.4069	.3174	.2352	.1631	.1035	.0579	.0266	.0086	.0012
3	1.0000	.9995	.9978	.9933	.9844	.9692	.9460	.9130	.8688	.8125	.7438	.6630	.5716	.4718	.3672	.2627	.1648	.0815	.0226
4	1.0000	1.0000	.9999	.9997	.9990	.9976	.9947	.9898	.9815	.9688	.9497	.9222	.8840	.8319	.7627	.6723	.5563	.4095	.2262
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156	.0083	.0041	.0018	.0007	.0002	.0001	.0000	.0000	.0000
1	.9672	.8857	.7765	.6554	.5339	.4202	.3191	.2333	.1636	.1094	.0692	.0410	.0223	.0109	.0046	.0016	.0004	.0001	.0000
2	.9978	.9842	.9527	.9011	.8306	.7443	.6471	.5443	.4415	.3438	.2553	.1792	.1174	.0705	.0376	.0170	.0059	.0013	.0001
3	.9999	.9987	.9941	.9830	.9624	.9295	.8826	.8208	.7447	.6563	.5585	.4557	.3529	.2557	.1694	.0989	.0473	.0159	.0022
4	1.0000	.9999	.9996	.9984	.9954	.9891	.9777	.9590	.9308	.8906	.8364	.7667	.6809	.5798	.4661	.3446	.2235	.1143	.0328
5	1.0000	1.0000	1.0000	.9999	.9998	.9993	.9982	.9959	.9917	.9844	.9723	.9533	.9246	.8824	.8220	.7379	.6229	.4686	.2649
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

...

Distribuição de Poisson

➔ **Modelo probabilístico que se aplica a variáveis discretas:**

- ① Do tipo “número de ocorrências num campo ou intervalo contínuo (em geral, espaço ou tempo)”;
- ② Do tipo binomial em que a ocorrência do acontecimento A é rara - $p(A) \leq 0,1$ -.

Exemplos:

A – Defeitos por metro quadrado; células por campo; acidentes por dia, utentes inscritos por hora, internamento por semana; consumo de um medicamento de utilização rara;...

▲ **As não ocorrências não são contáveis.**

B – Número de crianças com polidactilia; número de indivíduos com reações de hipersensibilidade a um medicamento; número de infecções nasocomiais numa enfermaria, ...

▲ **Distribuição de Poisson como aproximação à binomial.**

Nestes casos, o número de vezes (X) que ocorre um acontecimento A é uma V. A. em que:

$X \sim P(\lambda)$

$$\begin{array}{l} \mu = \lambda \\ \text{com} \\ \sigma = \sqrt{\lambda} \end{array}$$

A probabilidade de cada valor x - $p(x_i)$ - pode ser calculado através da seguinte fórmula:

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Probabilidades de Poisson acumuladas

		μ									
x		1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
0		.3679	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496
1		.7358	.6990	.6626	.6268	.5918	.5578	.5249	.4932	.4628	.4337
2		.9197	.9004	.8795	.8571	.8335	.8088	.7834	.7572	.7306	.7037
3		.9810	.9743	.9662	.9569	.9463	.9344	.9212	.9068	.8913	.8747
4		.9963	.9946	.9923	.9893	.9857	.9814	.9763	.9704	.9636	.9559
5		.9994	.9990	.9985	.9978	.9968	.9955	.9940	.9920	.9896	.9868
6		.9999	.9999	.9997	.9996	.9994	.9991	.9987	.9981	.9974	.9966
7		1.0000	1.0000	1.0000	.9999	.9999	.9998	.9997	.9996	.9994	.9992
8		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9999	.9998
9		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

		μ									
x		2.0	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9
0		.1353	.1225	.1108	.1003	.0907	.0821	.0743	.0672	.0608	.0550
1		.4060	.3796	.3546	.3309	.3084	.2873	.2674	.2487	.2311	.2146
2		.6767	.6496	.6227	.5960	.5697	.5438	.5184	.4936	.4695	.4460
3		.8571	.8386	.8194	.7993	.7787	.7576	.7360	.7141	.6919	.6696
4		.9473	.9379	.9275	.9162	.9041	.8912	.8774	.8629	.8477	.8318
5		.9834	.9796	.9751	.9700	.9643	.9580	.9510	.9433	.9349	.9258
6		.9955	.9941	.9925	.9906	.9884	.9858	.9828	.9794	.9756	.9713
7		.9989	.9985	.9980	.9974	.9967	.9958	.9947	.9934	.9919	.9901
8		.9998	.9997	.9995	.9994	.9991	.9989	.9985	.9981	.9976	.9969
9		1.0000	.9999	.9999	.9999	.9998	.9997	.9996	.9995	.9993	.9991
10		1.0000	1.0000	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9998	.9998
11		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	.9999
12		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

		μ									
x		3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
0		.0498	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202
1		.1991	.1847	.1712	.1586	.1468	.1359	.1257	.1162	.1074	.0992
2		.4232	.4012	.3799	.3594	.3397	.3208	.3027	.2854	.2689	.2531
3		.6472	.6248	.6025	.5803	.5584	.5366	.5152	.4942	.4735	.4532
4		.8153	.7982	.7806	.7626	.7442	.7254	.7064	.6872	.6678	.6484
5		.9161	.9057	.8946	.8829	.8705	.8576	.8441	.8301	.8156	.8006
6		.9665	.9612	.9554	.9490	.9421	.9347	.9267	.9182	.9091	.8995
7		.9881	.9858	.9832	.9802	.9769	.9733	.9692	.9648	.9599	.9546
8		.9962	.9953	.9943	.9931	.9917	.9901	.9883	.9863	.9840	.9815
9		.9989	.9986	.9982	.9978	.9973	.9967	.9960	.9952	.9942	.9931
10		.9997	.9996	.9995	.9994	.9992	.9990	.9987	.9984	.9981	.9977
11		.9999	.9999	.9999	.9998	.9998	.9997	.9996	.9995	.9994	.9993
12		1.0000	1.0000	1.0000	1.0000	.9999	.9999	.9999	.9999	.9998	.9998

...

Distribuição de Normal

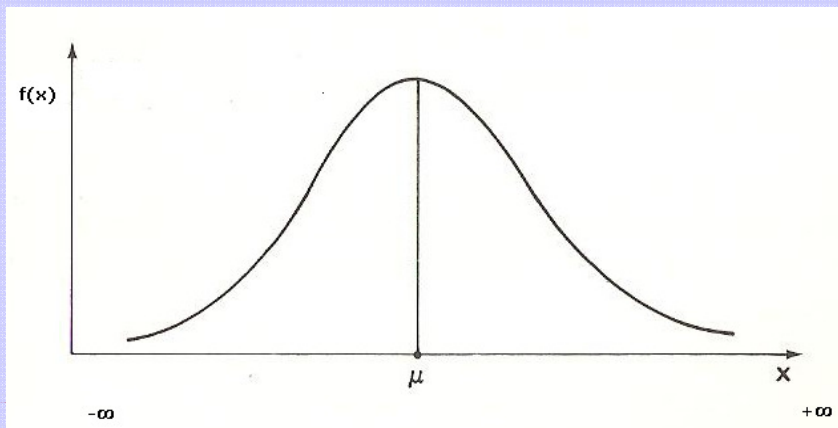
(de Gauss-Laplace)

➔ **Modelo probabilístico mais comum e de maior utilidade em estatística.**

- ❖ **Representa com boa aproximação, as distribuições de frequência observadas na maioria dos fenômenos da natureza**
- ❖ **Apropriado para V. A. contínuas, constitui ainda uma boa aproximação para muitas outras distribuições (binomial; de Poisson, etc.).**

A variável normal é contínua e a sua probabilidade é representada graficamente pela área debaixo da curva de função.

$$X \sim N(\mu, \sigma)$$



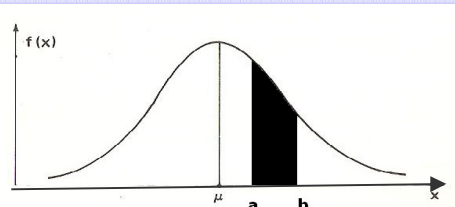
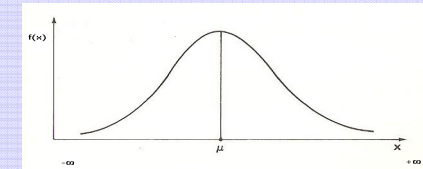
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Distribuição de Normal

(de Gauss-Laplace)

Características da distribuição normal:

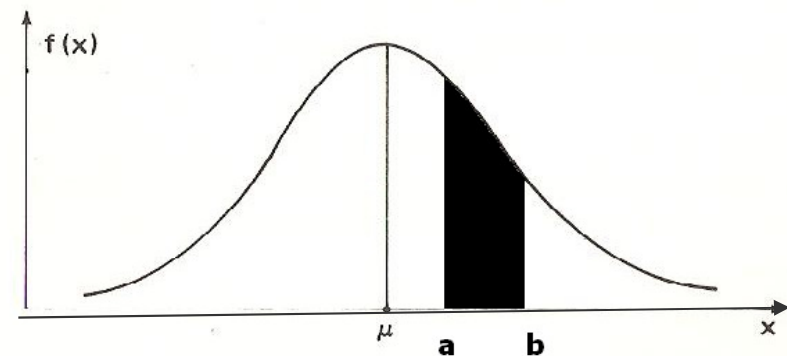
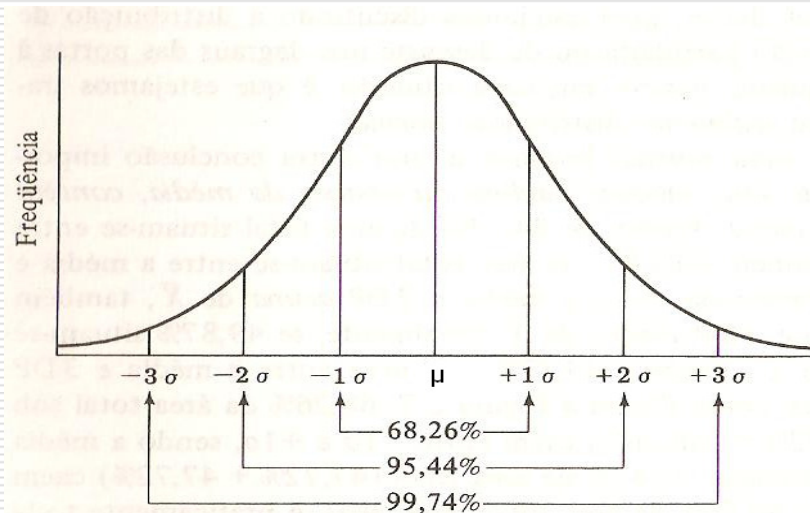
- 1 – A variável pode tomar qualquer valor real;
- 2 – A curva normal tem a forma de um sino;
- 3 – É simétrica em relação à média (média, moda e mediana coincidem);
- 4 – É totalmente especificada pela respectiva média e desvio padrão. A área sob a curva entre a média e um ponto arbitrário é função do número de desvios padrões entre a média e aquele ponto;
- 5 – A área sob a curva entre dois pontos corresponde à probabilidade de a V.A. distribuída normalmente tomar um valor nesse intervalo;
- 6 – A área total debaixo da curva corresponde à probabilidade da V.A. assumir qualquer valor real, sendo igual à unidade – $p(-\infty < X < +\infty) = 1$;
- 7 – A probabilidade de ocorrer um valor inferior à média é igual à probabilidade de ocorrer um valor superior à média – $p(x > \mu) = p(x < \mu) = 0,5$ -;
- 8 – Os valores mais afastados da média são menos prováveis do que os mais próximos;
- 9 – Dois intervalos iguais e à mesma distância da média têm iguais probabilidades;
- 10 – As probabilidades referem-se sempre a intervalos de valores de X e não a valores pontuais de X:



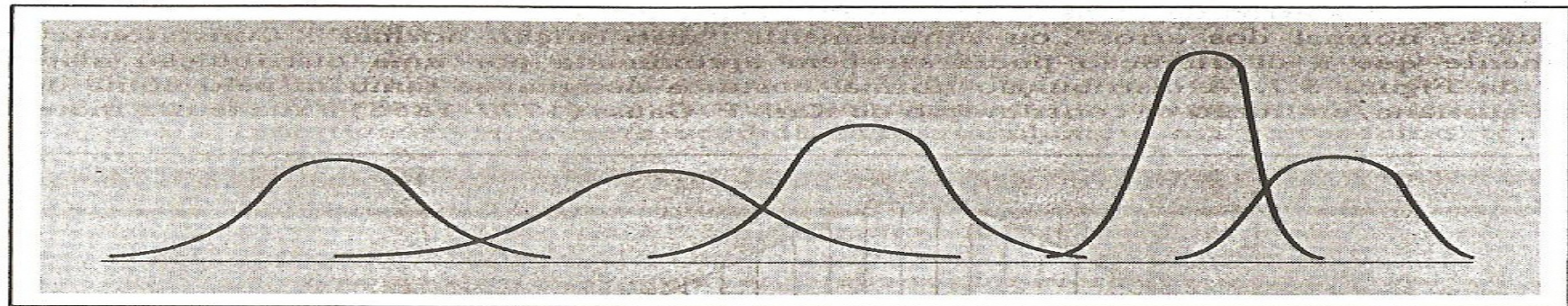
$$p(x=a) = p(x=b) = 0 \quad p(a < X < b) \quad (X > a) \quad p(X < b)$$

Distribuição de Normal

(de Gauss-Laplace)



$P(a < X < b) = \text{área sob a curva entre } \underline{a} \text{ e } \underline{b}$

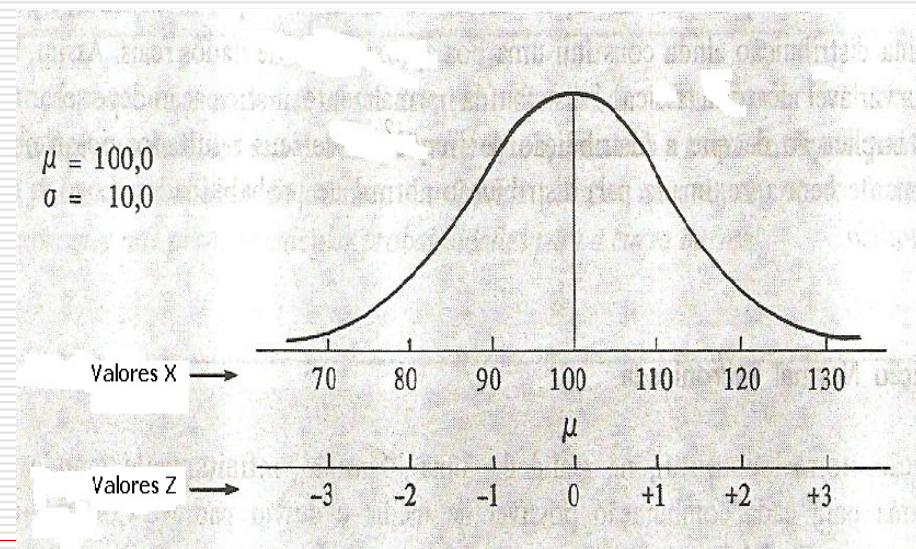
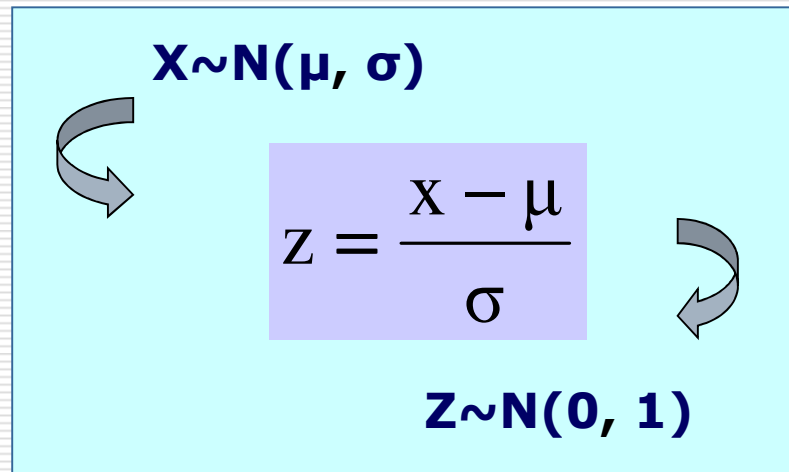


As combinações de médias e desvios padrões são em número ilimitado.

Distribuição de Normal Padronizada

O cálculo directo das probabilidades dos valores de X num determinado intervalo é complexo, exigindo o recurso a operações matemáticas sofisticadas.

- ❖ ... Mas, todas as V. A. normais – $X \sim N(\mu, \sigma)$ – podem ser transformadas numa V.A. padronizada centrada e reduzida – $Z \sim N(\mu=0 \text{ e } \sigma=1)$.
- ❖ Para a distribuição Z, dispomos e tabelas de distribuição de probabilidades.



A normal distribution curve is shown with a horizontal axis. The mean is labeled "Média" and marked with a vertical line. The area under the curve to the right of the mean is shaded gray. A point z is marked on the horizontal axis to the right of the mean.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4986	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.49							

Distribuições de Probabilidades

Aproximações

▲ Demonstra-se que a maioria das distribuições teóricas (mesmo as discretas) no seu limite tendem para a normal.

Binomial

Se $p \leq 0,1$ ou $p \geq 0,9$ e $n > 20^*$



De Poisson ($\lambda = np$)

Se λ muito grande*



Normal

$$z = \frac{(x \pm 0,5) - \lambda}{\sqrt{\lambda}}$$

Se $p > 0,1$ ou $p < 0,9$ e $n > 20^*$



$$z = \frac{(x \pm 0,5) - np}{\sqrt{npq}}$$

* Na prática, recorre-se às aproximações quando os parâmetros não são incluídos nas tabelas.

▲ As aproximações são tanto mais precisas quanto maior o valor de n .