Variáveis Aleatórias
e
Distribuições de Probabilidade

jlborges@fe.up.pt

Experiment

... an act of conducting a controlled test or investigation.

An experiment results in something

The possible results of an experiment may be one or more.

Based on the number of possible results we classify experiments as

DETERMINISTIC

PROBABILISTIC

http://www.futureaccountant.com/probability/study-notes/experiment-random-deterministic.php

Deterministic Experiment

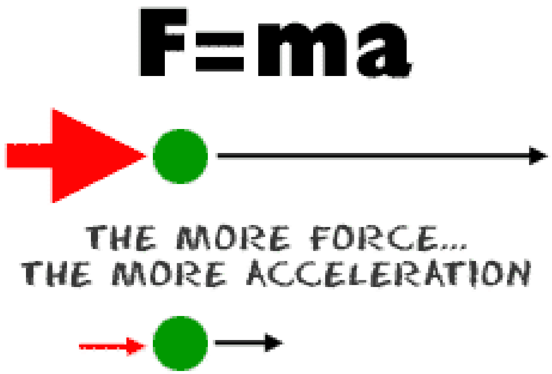
Experiments having **one possible result** i.e. whose result is certain

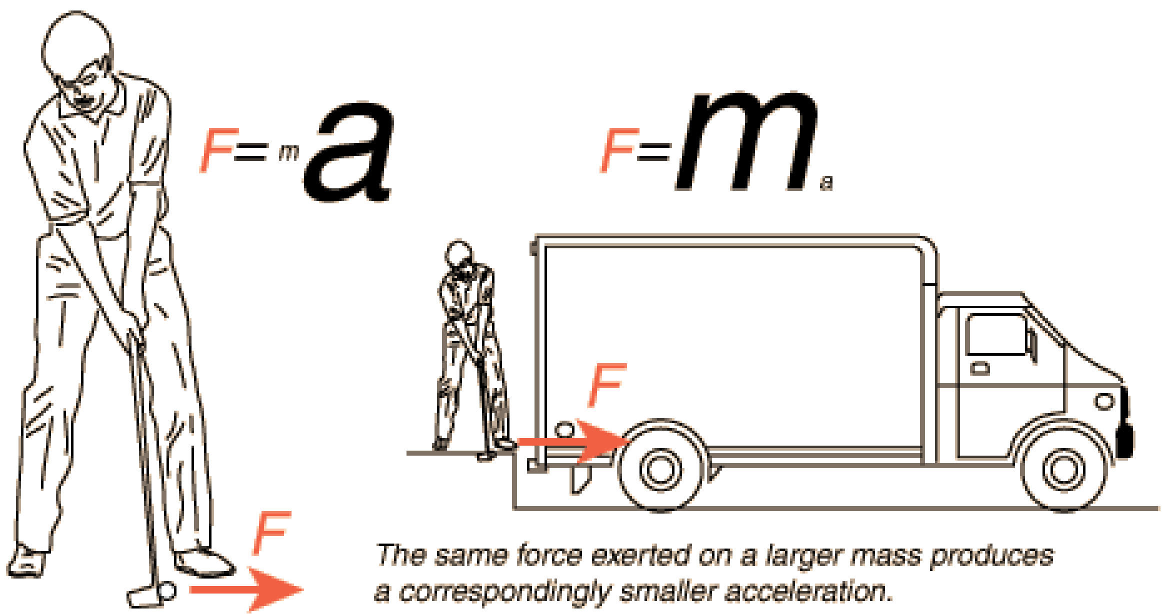
The result is **predictable with certainty** and is known prior to its conduct

This approach stipulates that the conditions under which the experiment is conducted would determine its result

The experiments that we conduct to verify **the laws of science** or established laws of other areas are the best examples for these

jlborges@fe.up.pt





Experiments whose result is uncertain are called **indeterministic** or **unpredictable** or **Probabilistic experiments**

There are **more than one possible result**

Whatever may be the conditions under which we conduct the experiment, **we can only say that the result would be one of the possible results**.

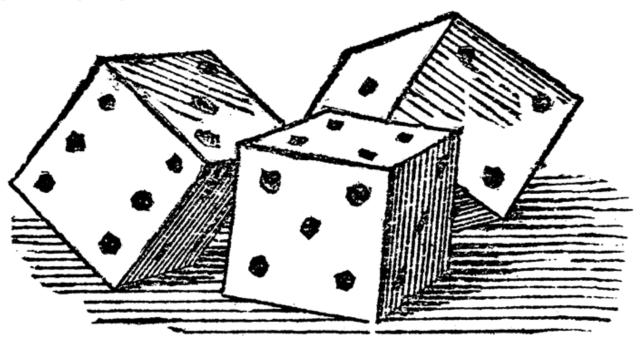
We cannot ensure a certain result by performing the experiment under certain conditions in a certain method

jlborges@fe.up.pt

Experiência aleatória

experiência cujo resultado depende do acaso

para a qual é possível caracterizar o conjunto de resultados possíveis

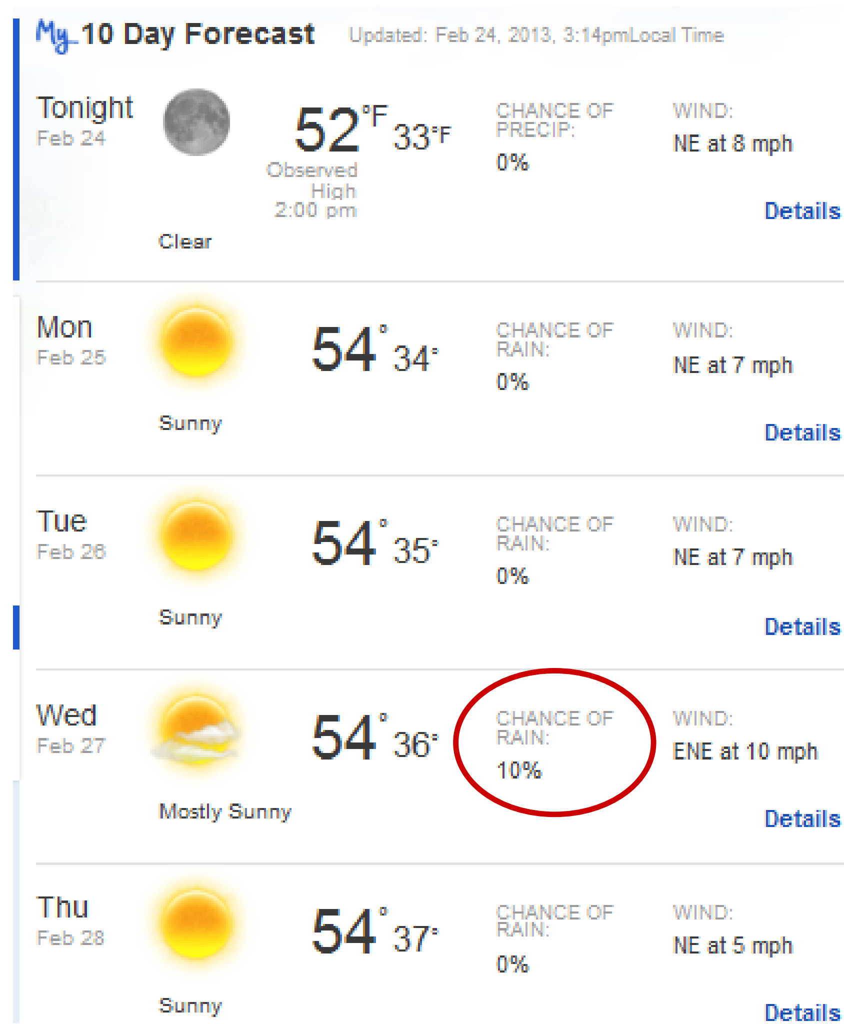


jlborges@fe.up.pt

Exemplos de Experiências Aleatórias:

- Se lançar uma moeda ao ar que resultado vou obter?
- Quem vai ganhar a Liga deste ano?
- Quanto tempo vou demorar hoje no meu percurso de regresso a casa?
- Quantos alunos vão aparecer para a aula de hoje?
- Qual a probabilidade de chover amanhã?

....



Uma variável aleatória é utilizada para estudar as experiências aleatórias

Se o número de resultados possíveis for finito temos uma variável aleatórias discreta

- Lançamento de um dado - 1,2,3,4,5,6

Se o número de resultados possíveis for infinito temos uma variável aleatória contínua

- o Peso exacto de um aluno da FEUP seleccionado ao acaso
- entre 50.0(0) e 100.0(0)

Variável Aleatória - expressar os resultados de uma experiência aleatória

Exemplo

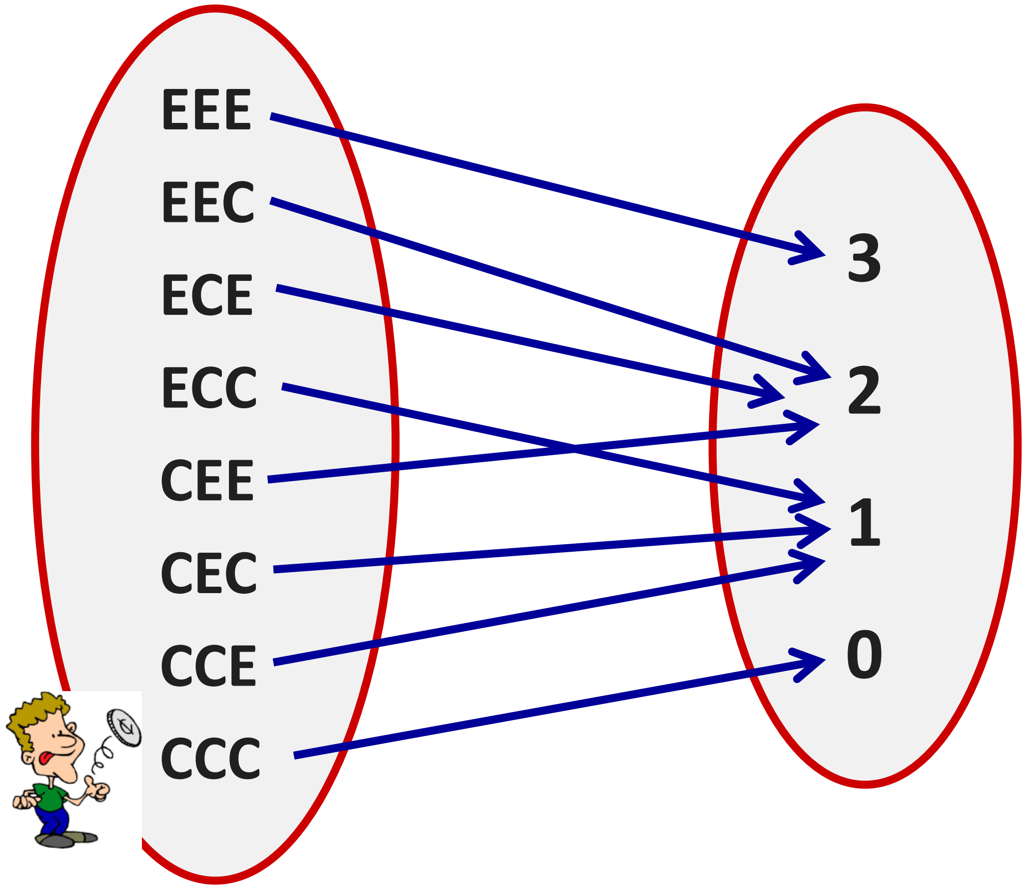
- Exp. Aleatória: medição da altura de uma pessoa ao acaso
  Esp. Amostral: conjunto de todas as alturas
  Var. Aleatória: altura

- Exp. Aleatória: lançamento da moeda E-C três vezes ao ar
  Esp. Amostral: sequências de E's e C's
  Var. Aleatória: nº de E's obtidos em cada sequência

Exemplo

Esp. Amostral: sequências de E's e C's obtidas no lançamento da moeda E-C três vezes ao ar

Var. Aleatória **Y**: **nº de E's obtidos em cada sequência**



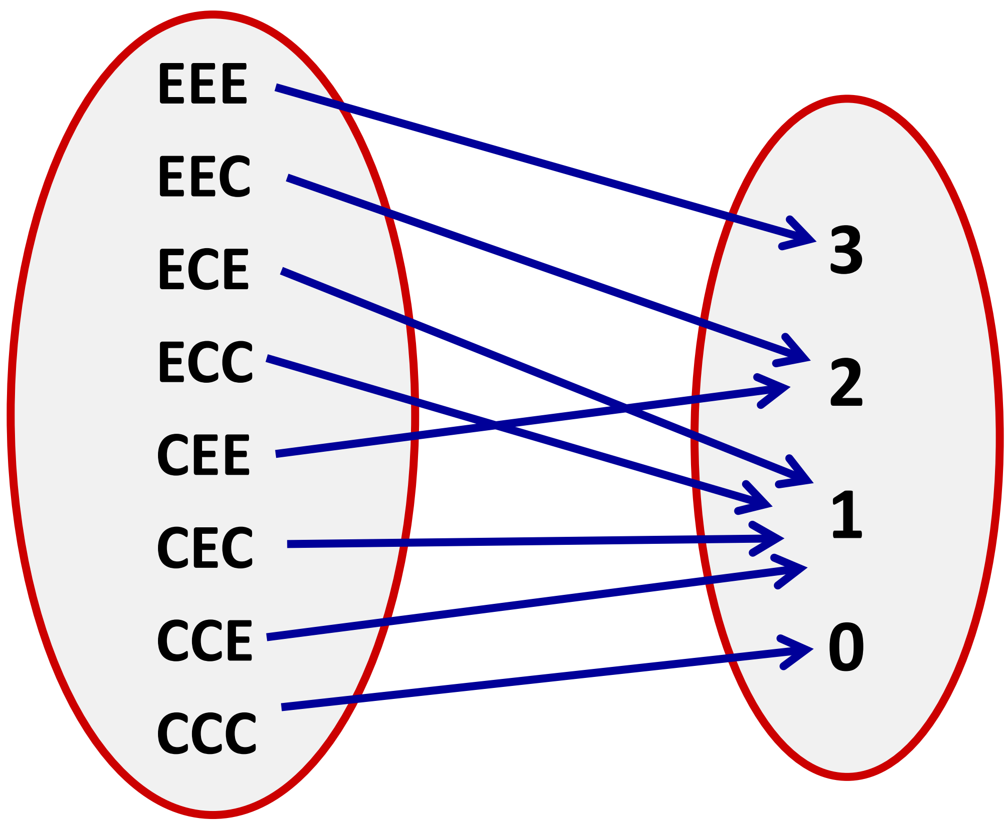
A cada resultado de uma variável aleatória associa-se uma probabilidade:

$$P(Y=0)=P(CCC)=1/8$$
$$P(Y=1)=P(ECC)+P(CEC)+P(CCE)=3/8$$
$$P(Y=2)=P(EEC)+P(ECE)+P(CEE)=3/8$$
$$P(Y=3)=P(EEE)=1/8$$

Exemplo de outra variável aleatória obtida a partir da mesma experiência

Esp. Amostral: sequências de E's e C's obtidas no lançamento da moeda E-C três vezes ao ar

Var. Aleatória **Y'**: nº máximo de E's em sequência



Variável Aleatória

- Agrupa resultados de uma experiência aleatória em acontecimentos

- Associa a cada acontecimento a respectiva probabilidade de ocorrência

Variáveis Aleatórias Discretas

Função de Probabilidade

associa a cada valor particular **y** a probabilidade de **P(Y= y)**

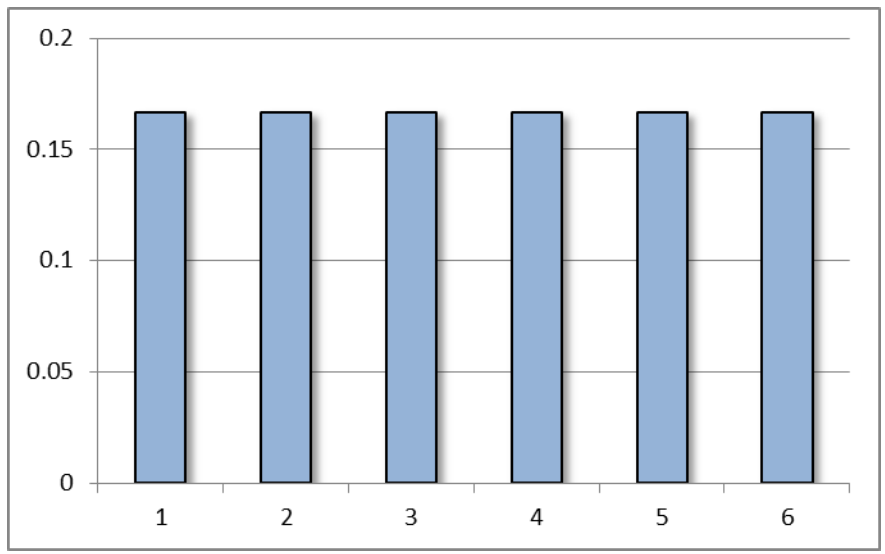
$$\boxed{p(y) = P(Y = y)}$$

Dos axiomas da probabilidade resulta

$$\forall y \in S': \quad 0 < p(y) \leq 1$$

$$\sum_{y \in S'} p(y) = 1$$

Um Dado	
y	P(y)
1	0.167
2	0.167
3	0.167
4	0.167
5	0.167
6	0.167



Função de Distribuição

F(y) - função de probabilidade acumulada
associa a cada valor particular **y** a probabilidade de **Y** ser menor ou igual a **y**

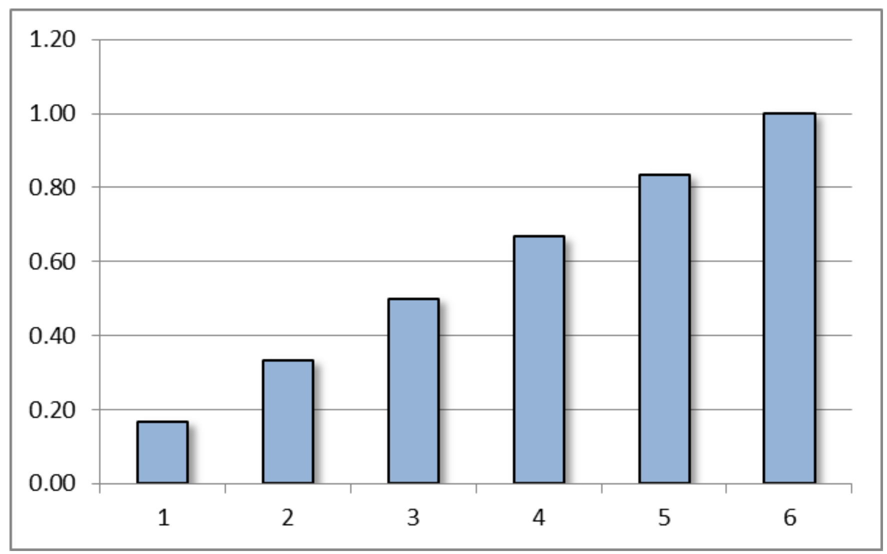
Um Dado	
y	F(y)
1	0.17
2	0.33
3	0.50
4	0.67
5	0.83
6	1.00

$$F(y) = P(Y \leq y) = \sum_{u \leq y} p(u)$$

Desta definição resulta

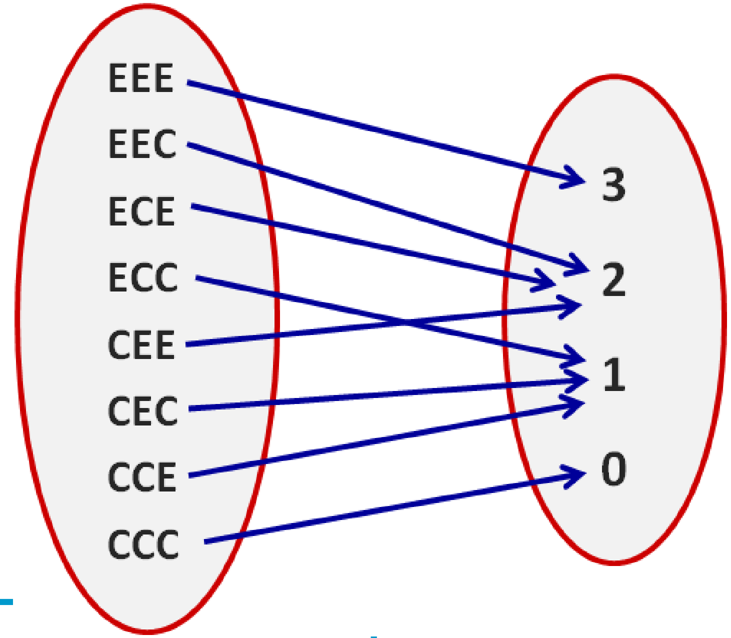
$$\forall y \in \mathrm{R}: \quad 0 \leq F(y) \leq 1$$

$$P(a < Y \leq b) = F(b) - F(a)$$



Exemplo

Var. Aleatória **Y**: nº de E's obtidos em cada sequência no lançamento de uma moeda E-C três vezes ao ar



Função de probabilidade:

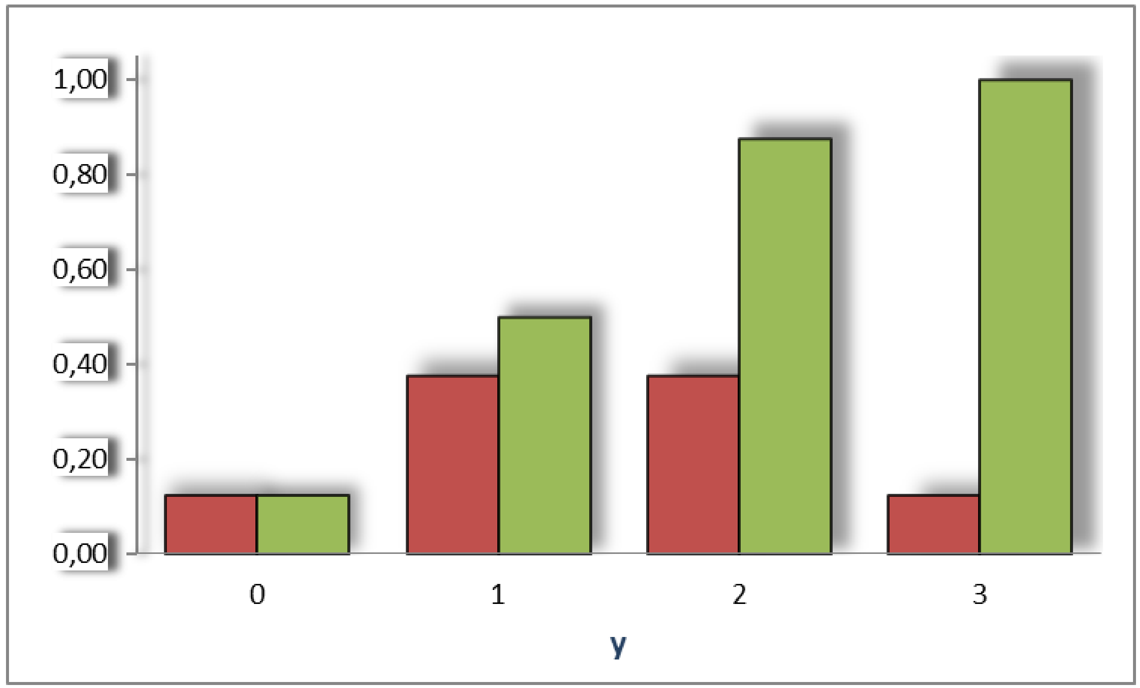
$$p(y=0)=p(0)=1/8$$
$$p(y=1)=p(1)=3/8$$
$$p(y=2)=p(2)=3/8$$
$$p(y=3)=p(3)=1/8$$

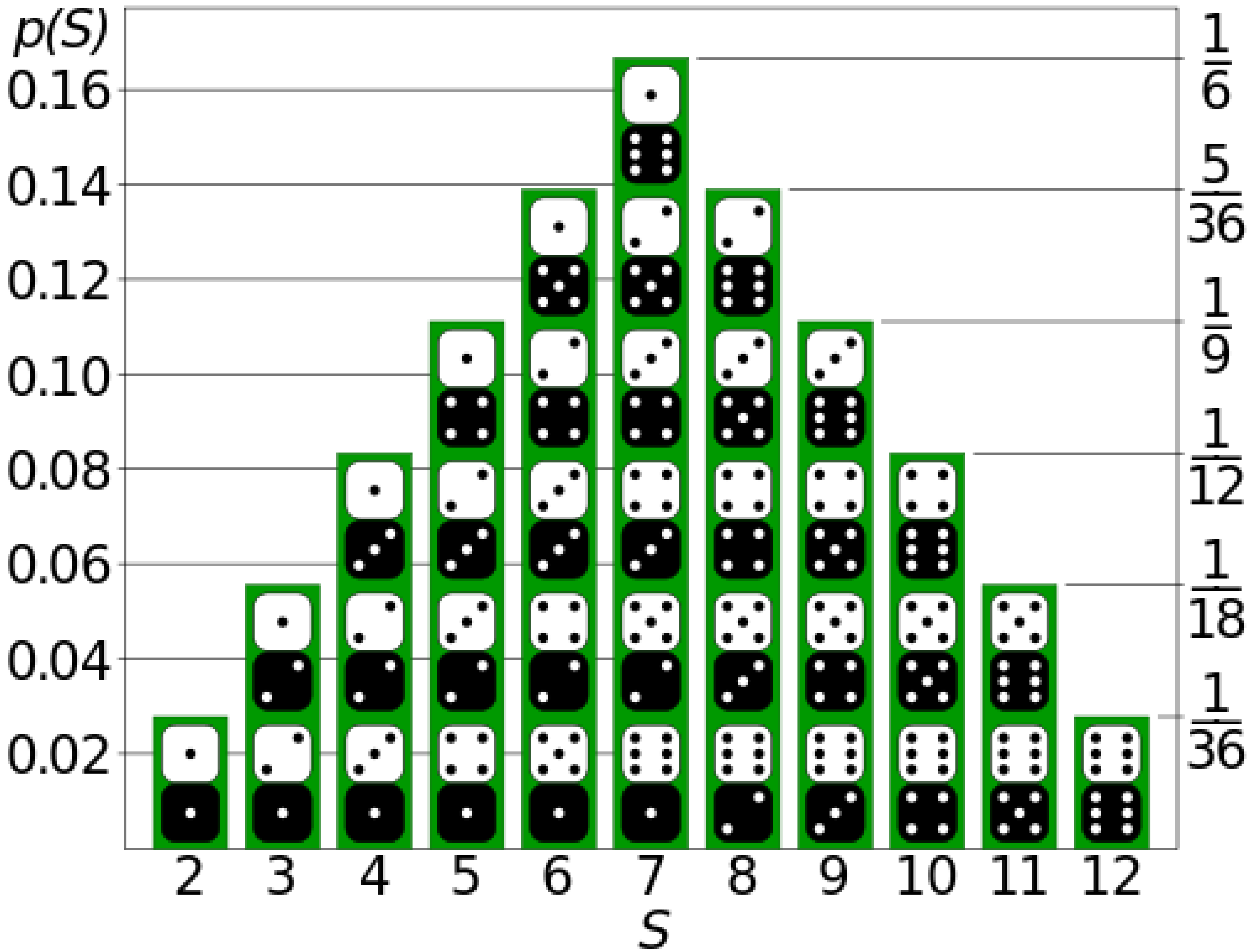
Função de distribuição:

$$F(y=0)=P(y\leq 0)=p(0)=1/8$$
$$F(y=1)=P(y\leq 1)=p(0)+p(1)=4/8$$
$$F(y=2)=P(y\leq 2)=p(0)+p(1)+p(2)=7/8$$
$$F(y=3)=P(y\leq 3)=p(0)+p(1)+p(2)+p(3)=1$$

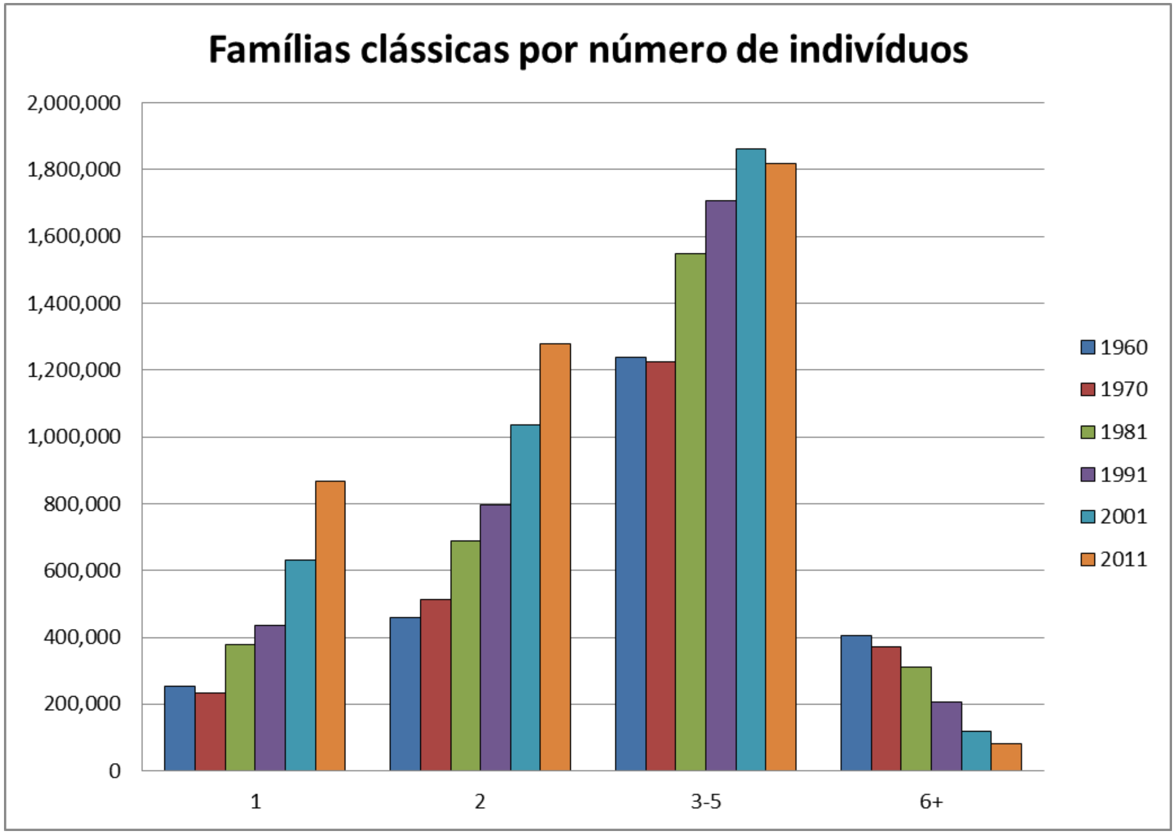
As funções de probabilidade e de distribuição podem ser representadas através de **tabelas** ou de **diagramas de barras**

Y	p(y)	F(y)
0	1/8	1/8
1	3/8	1/2
2	3/8	7/8
3	1/8	1




p(S)
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
2
3
4
5
6
7
8
9
10
11
12
S
1/6
5/36
1/9
1/12
1/18
1/36

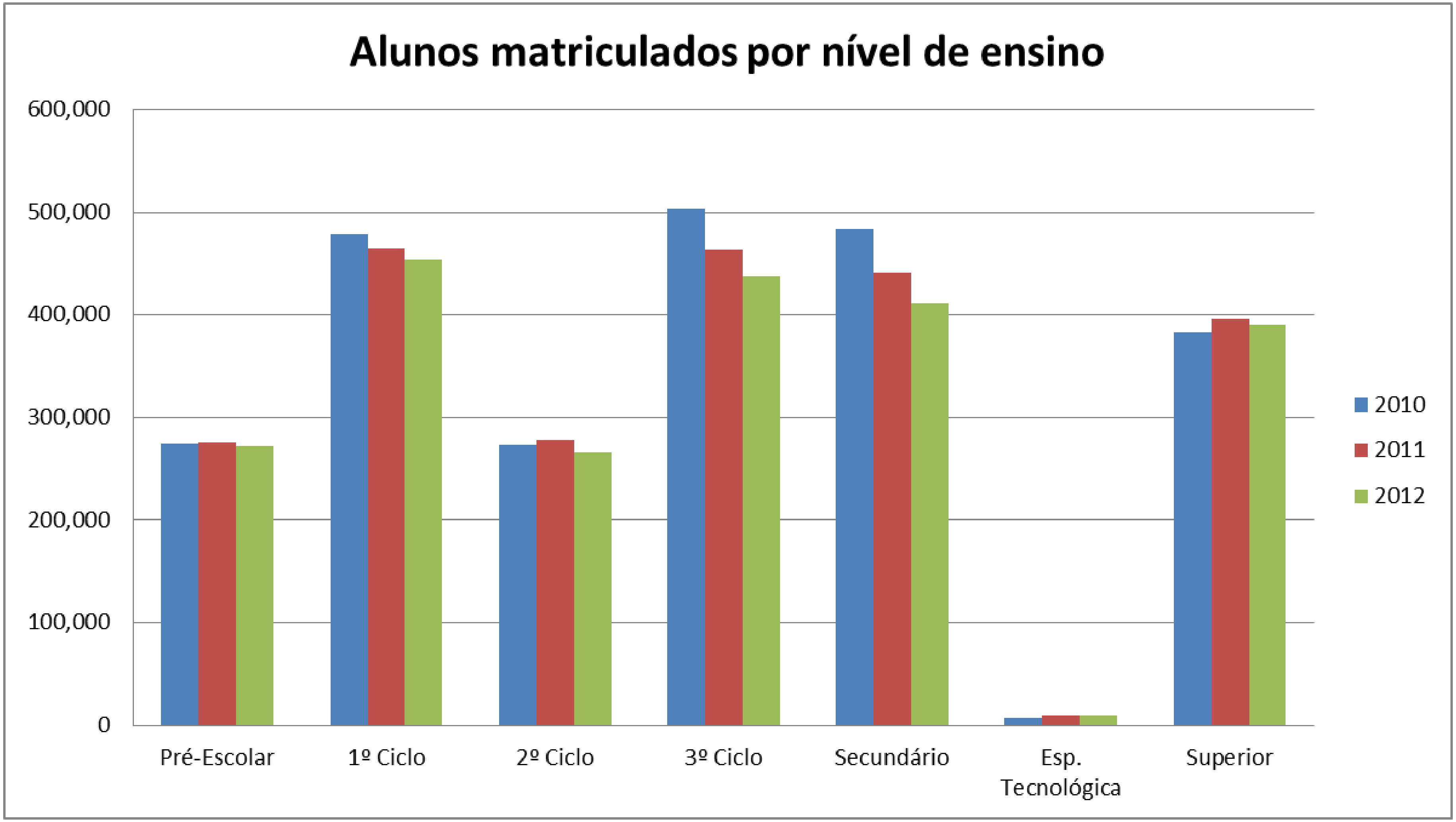
jlborges@fe.up.pt

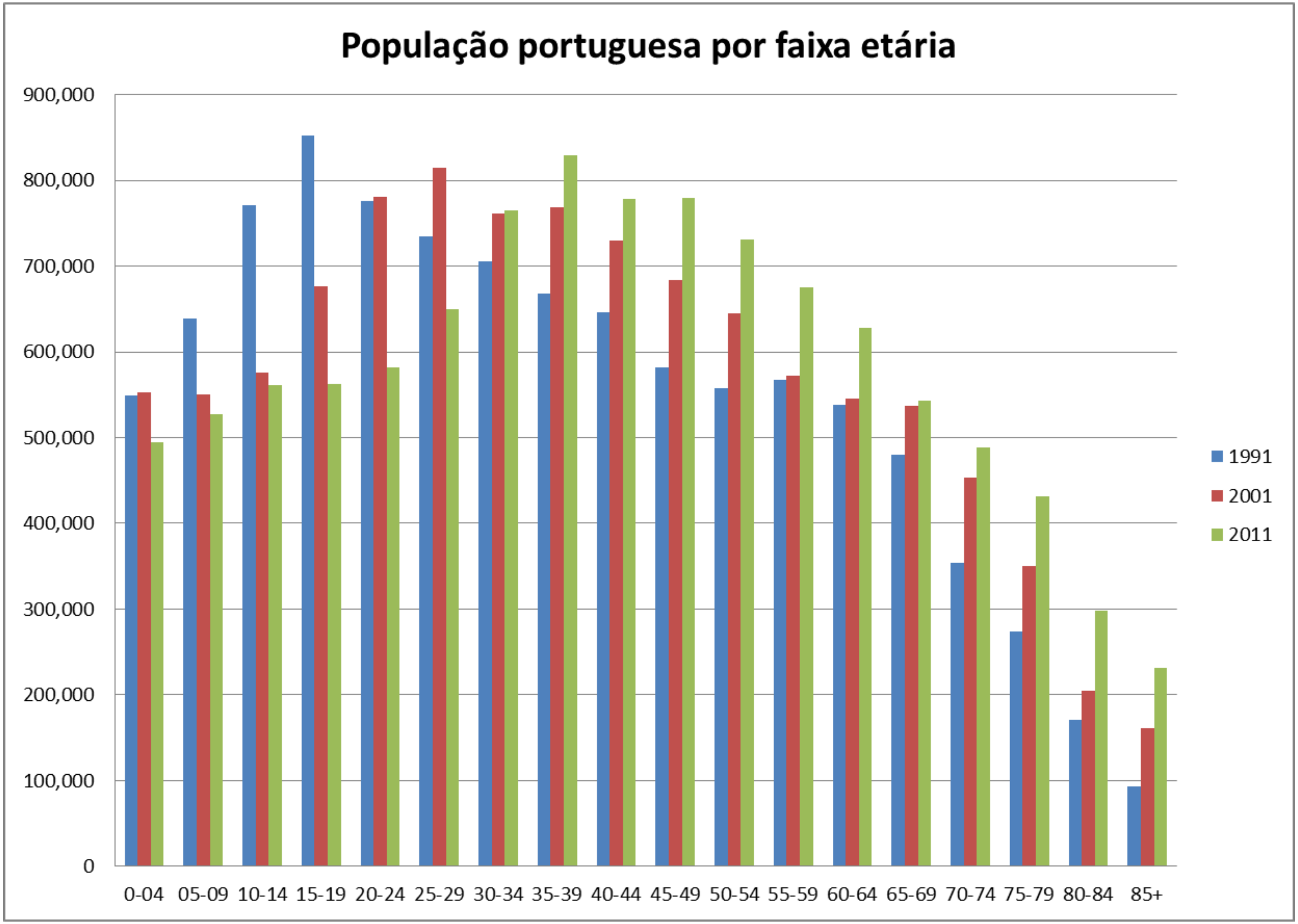


2011

Y: Num. Ind.	Num. Famílias	P(Y)	P(Y)
1	866.827	0,214	0,214
2	1.277.558	0,316	0,530
3-5	1.818.875	0,450	0,980
6+	80.466	0,020	1,000
Total	4.043.726	1,000	

jlborges@fe.up.pt





População portuguesa por faixa etária
900,000
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
0
0-04
05-09
10-14
15-19
20-24
25-29
30-34
35-39
40-44
45-49
50-54
55-59
60-64
65-69
70-74
75-79
80-84
85+
1991
2001
2011

jlborges@fe.up.pt

Golos do Cristiano Ronaldo por jogo

historial?

2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
0	0	2	2	1	0	2	0	3	0	1	0	2	0
0	0	0	2	1	0	0	1	0	1	0	0	2	1
0	0	0	2	2	0	0	1	1	0	1	2	0	1
0	0	0	0	0	0	2	2	2	1	1	1	4	2
0	1	0	1	1	0	2	0	1	1	3	3	1	0
0	0	0	0	2	0	1	2	2					
0	0	0	0	1	0	0	3	0					
0	1	0	1	0	2	0	0	3					
0	0	1	0	0	0	1	0	2					
1	0	0	0	2	0	0	0	0					
0	0	0	0	1	2	0	2	0					
0	0	0	1	1	2	0	4	3					
0	0	0	0	0	0	0	1	0					
0	0	0	0	2	0	1	2	0					
0	0	1	1	0	1	2	1	1					
0	2	0	0	1	0	0	1	0					
0	0	0	1	0	0	2	2	0					
0	0	0	0	2	0	0	0	1					
0	0	0	0	0	0	1	0	3					
0	0	0	2	1	1	0	1	0					
0	0	0	0	0	1	0	0	0					
	0	0	0	1	0	2	0	0					
	1	0	1	0	0	0	0	1					
	1	0	0	0	0	0	0	2					
	0	0	0	1	1	1	0	3					
	0	0	0	1	0	1	0	1					
	0	0	0	0	1	0	1	0					
	0	0	0	0	1	0	0	0					
	0	0	1	0	2	0	0	2					
	1	0	0	0	0	1	0	2					
	0	1	0	0	0	0	1	3					
	0	1	0	1	0	1	3	4	2	0	1	3	
	0	1	0	0	0	0	2	0	0	2	0	0	
	1	0	1	0	1	1	0	0	0	2	1	0	
	0	0	0	1	1	2	1	0	1	1	2	0	
	0	0	1	0	1	0	1	1	1	1	2	1	
	1	0	0	0	1	1	0	1	1	1	0	0	
	1	0	0	1	1	0	1	1	1	0	1	1	
	0	1	0	2	2	0	1	1	2	1	1	1	
	0	0	0	0	1	0	0	1	1	2	1	5	
	0	0	0	0	0	0	2	0	0	1	1	0	
	0	0	0	0	1			0	1	0	2	1	
	0	0	0	0	0			3	3	2	1	0	
	0	1	0	2	2			0	0	0	1	2	
	1	1	0	1	0			0	2	0	1	0	
	0	0	0	0	0			0	2	1	2	1	
	0	0	1	1	0			0	0	3	0	1	
	1	0	1	0	2			1	2	0	1	1	
	0	0	0	0	2			2	1	0	1	0	
	0	1	2	0	1			0	0	3	0	0	
	0	0	2	1	3			0	2	2	2	0	
	0	0	0	0	1			0	2	2		2	
	0	0	0	0	0			3	1			1	
	0	2	0	0				2	1			0	
	1	0	0	1					1			0	
	0	0	0	1					1			1	
	0	1		0					3				
	0	1		1					0				
	0	0		0					1				
	0	0		0					2				
		0							1				
		0							0				
		0							1				

Y	Frequência Observada	Função de Probabilidade	Função de Distribuição
0	369	0,542	0,542
1	189	0,278	0,819
2	86	0,126	0,946
3	31	0,046	0,991
4	4	0,006	0,997
5	2	0,003	1,000

fonte:
http://www.soccerbase.com

de 16-08-2003 a 27-02-2016

Variável Aleatória

- Agrupa resultados de uma experiência aleatória em acontecimentos

- Associa a cada acontecimento a respectiva probabilidade de ocorrência

Variáveis Aleatórias Contínuas

Variável aleatória contínua -> número de resultados associados à experiência aleatória é infinito

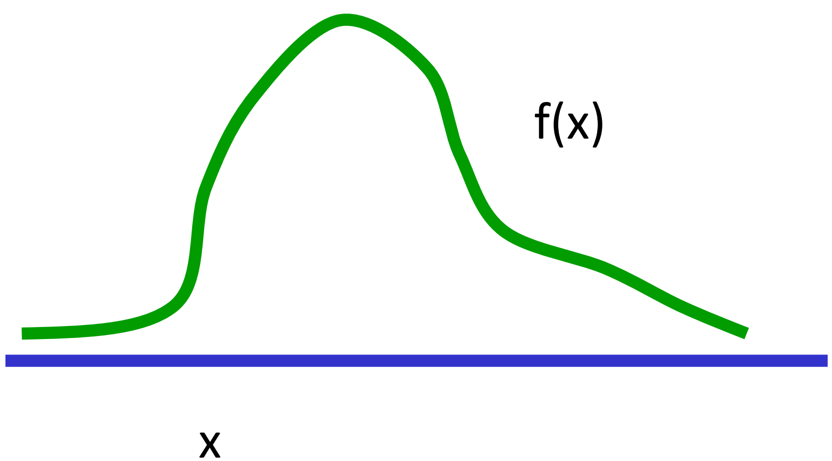
Peso, tempo, distância ...

O tempo exacto do vencedor da corrida de 100 m nos jogos olímpicos

Número de formigas que nascem por dia no planeta terra

Variáveis Aleatórias Contínuas

Sendo o número de resultados possíveis infimito é necessário definir uma função contínua que indica a frequência relativa dos diferentes acontecimentos



- f(x) não é uma probabilidade
- $\int f(x)dx = 1$
- f(x) > 0

Variáveis Aleatórias Contínuas

Admite-se que a probabilidade de uma variável aleatória contínua **X** tomar um valor qualquer particular é nula

P(altura = 1.800000(0) m)=0

Função Densidade de Probabilidade

função densidade de probabilidade de **X**, **f(x)** : associa a cada valor **x** a concentração de probabilidade por unidade em torno do ponto **X=x**

$$f(x)=\frac{\Delta P(x<X<x+\Delta x)}{\Delta x}=\frac{dP(x<X<x+dx)}{dx} \qquad (f(x)\geq 0)$$

jlborges@fe.up.pt

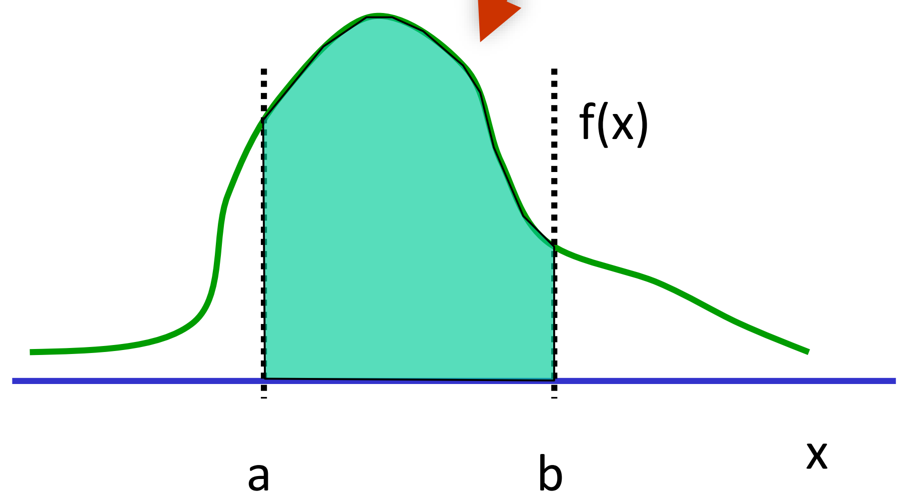
A probabilidade da variável X tomar um valor situado dentro do intervalo finito [a,b] vem

$$P\left(a \le X \le b\right) = \lim_{\Delta x \to 0} \sum_{n} f\left(x_n\right) \cdot \Delta x = \int_a^b f\left(x\right) \cdot dx$$

De acordo com a definição

$$P\left(-\infty < X < +\infty\right) = \int_{-\infty}^{+\infty} f\left(x\right) \cdot dx = 1$$

$$P\left(X = a\right) = \int_a^a f\left(x\right) \cdot dx = 0$$



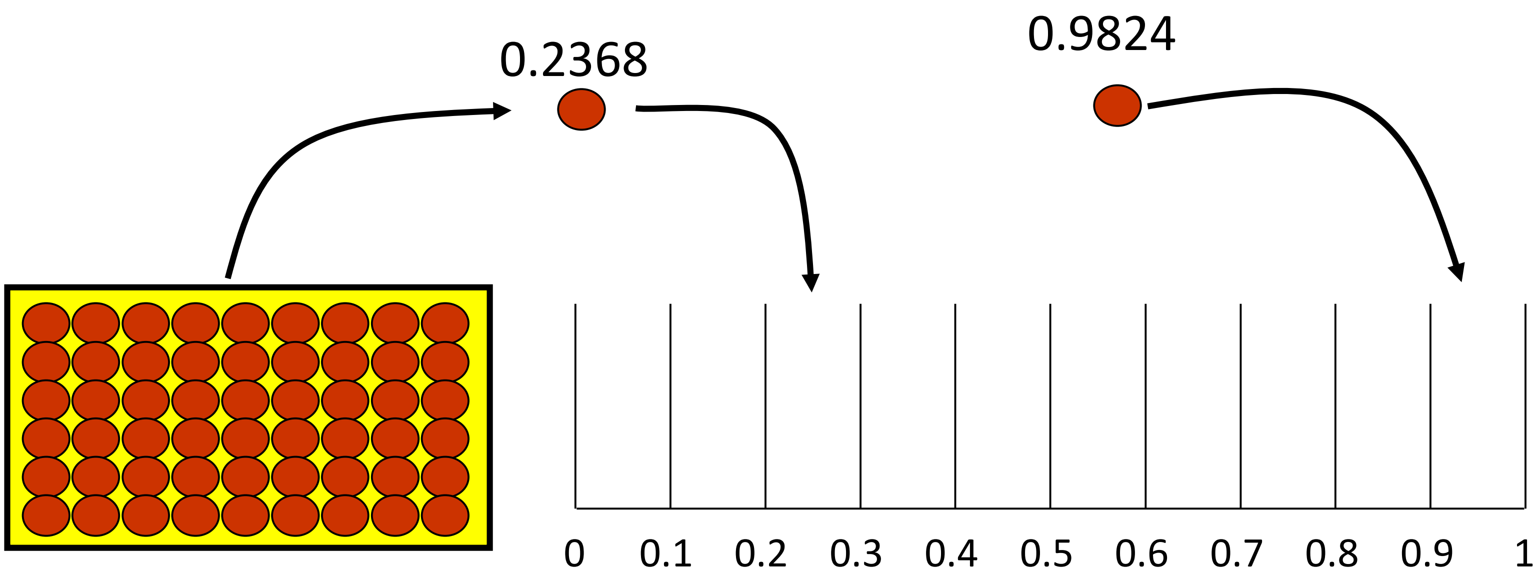
jlborges@fe.up.pt

Como se obtém a função densidade de probabilidade?

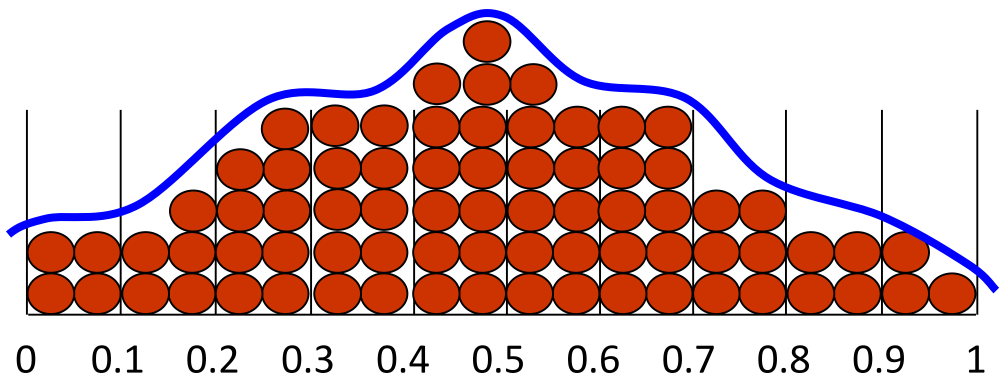
Considere uma caixa em que se encontram muitas bolas numeradas de 0 a 1.

Por exemplo, uma bola pode ter o número 0.2368 e outra 0.9824.

Para tentar compreender a variação dos números, vamos ordenar as bolas usando para isso um conjunto de caixa mais pequenas fazendo corresponder a cada uma um intervalo de valores.

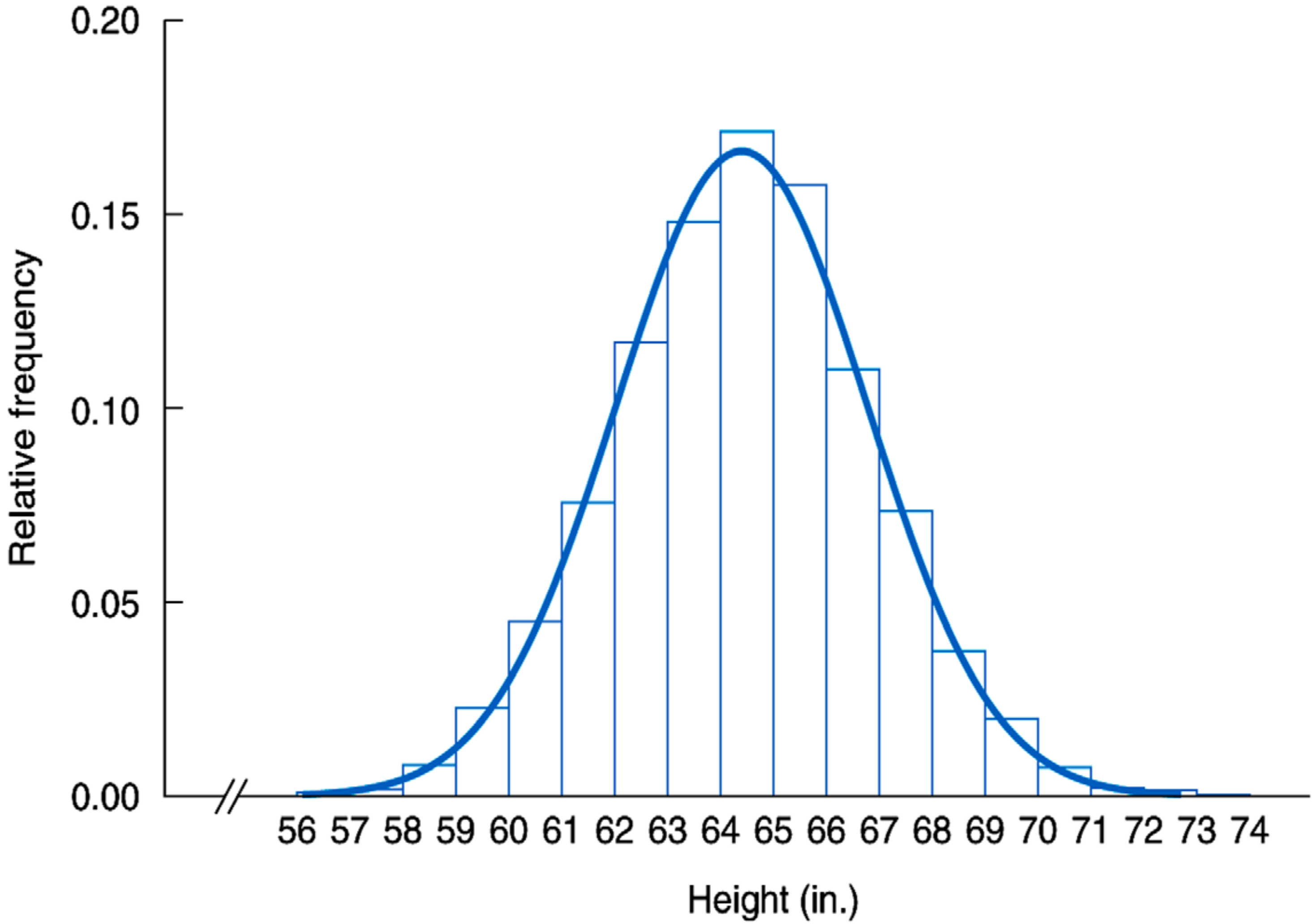


Após ter concluído o processo verificamos que há números que surgem com maior frequência (probabilidade) do que outros.



O passo seguinte seria tentar descrever a forma da variação nas probabilidades dos diferentes números por uma função contínua.

A essa função chama-se função densidade de probabilidade.

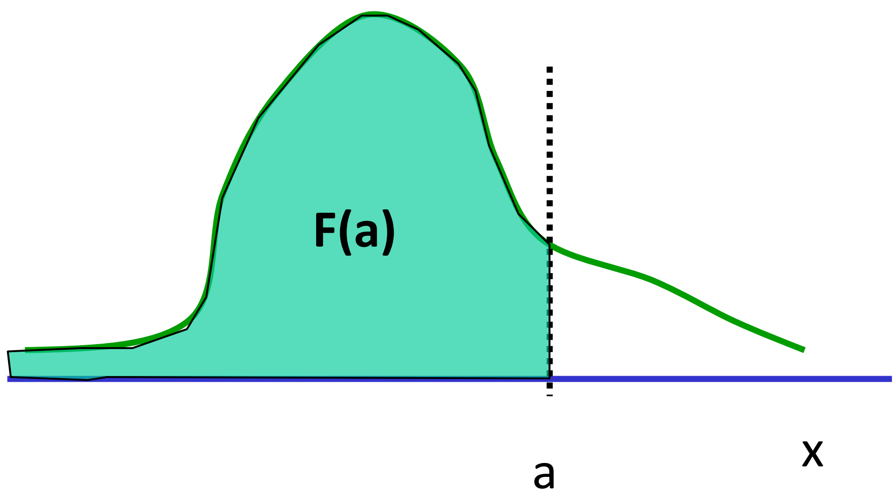

0.20
0.15
0.10
0.05
0.00
Relative frequency
56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74
Height (in.)

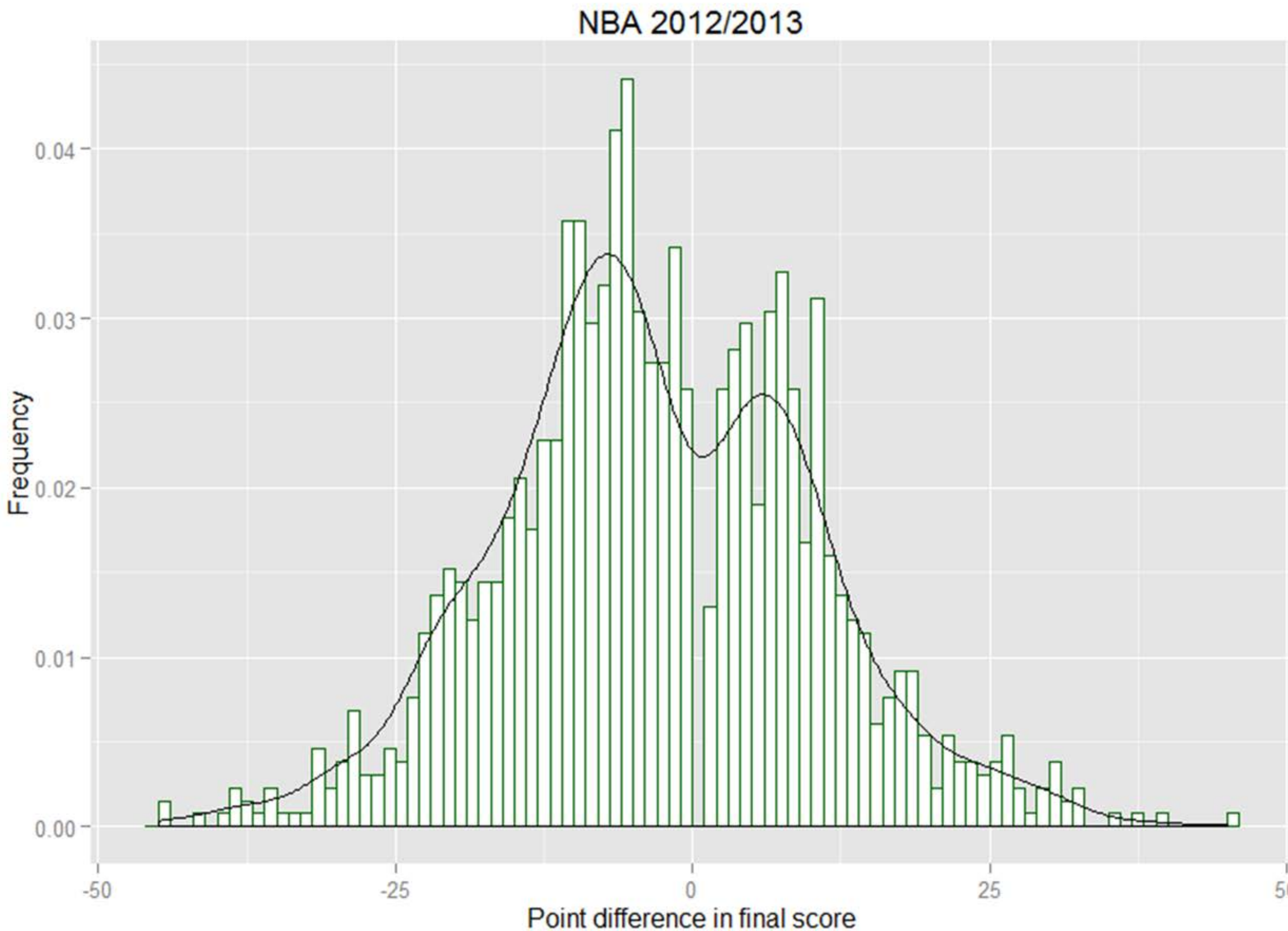
Função de Distribuição F(x)

(ou função de probabilidade acumulada)

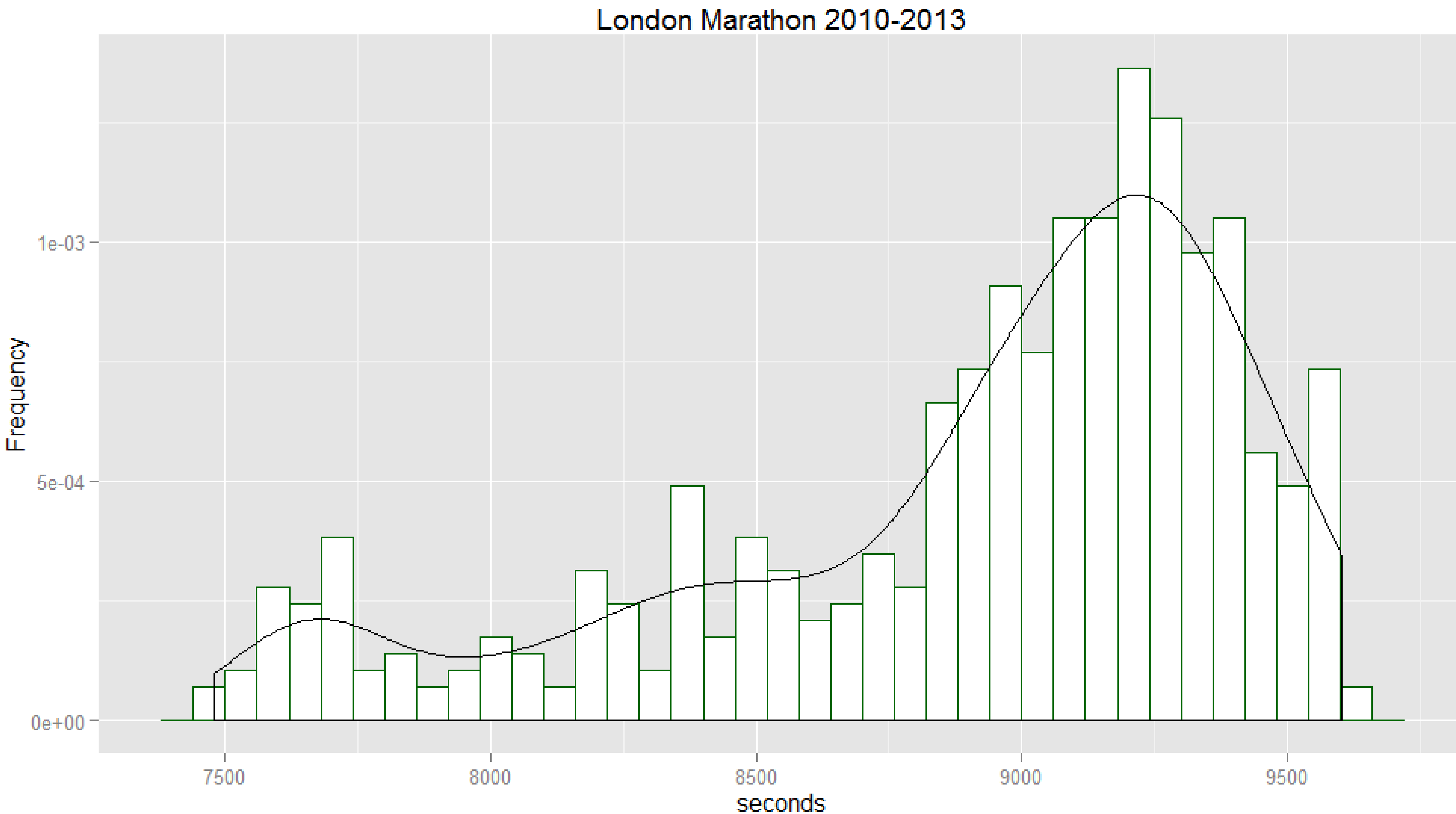
F(x): associa a cada valor particular **x** a probabilidade de **X** ser menor ou igual a **x**

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u) \cdot du \qquad \left(\lim_{x \to \infty} F(x) = 1\right)$$





NBA 2012/2013
Frequency
0.00
0.01
0.02
0.03
0.04
-50
-25
0
25
50
Point difference in final score



London Marathon 2010-2013
Frequency
1e-03
5e-04
0e+00
7500
8000
8500
9000
9500
seconds

As **funções de**

probabilidade / densidade de probabilidade

e de distribuição

podem ser encaradas como formas de representação de populações, recorrendo-se às **distribuições de frequências**,

obtidas a partir de amostras, para estimar essas funções.

Parâmetros das Distribuições

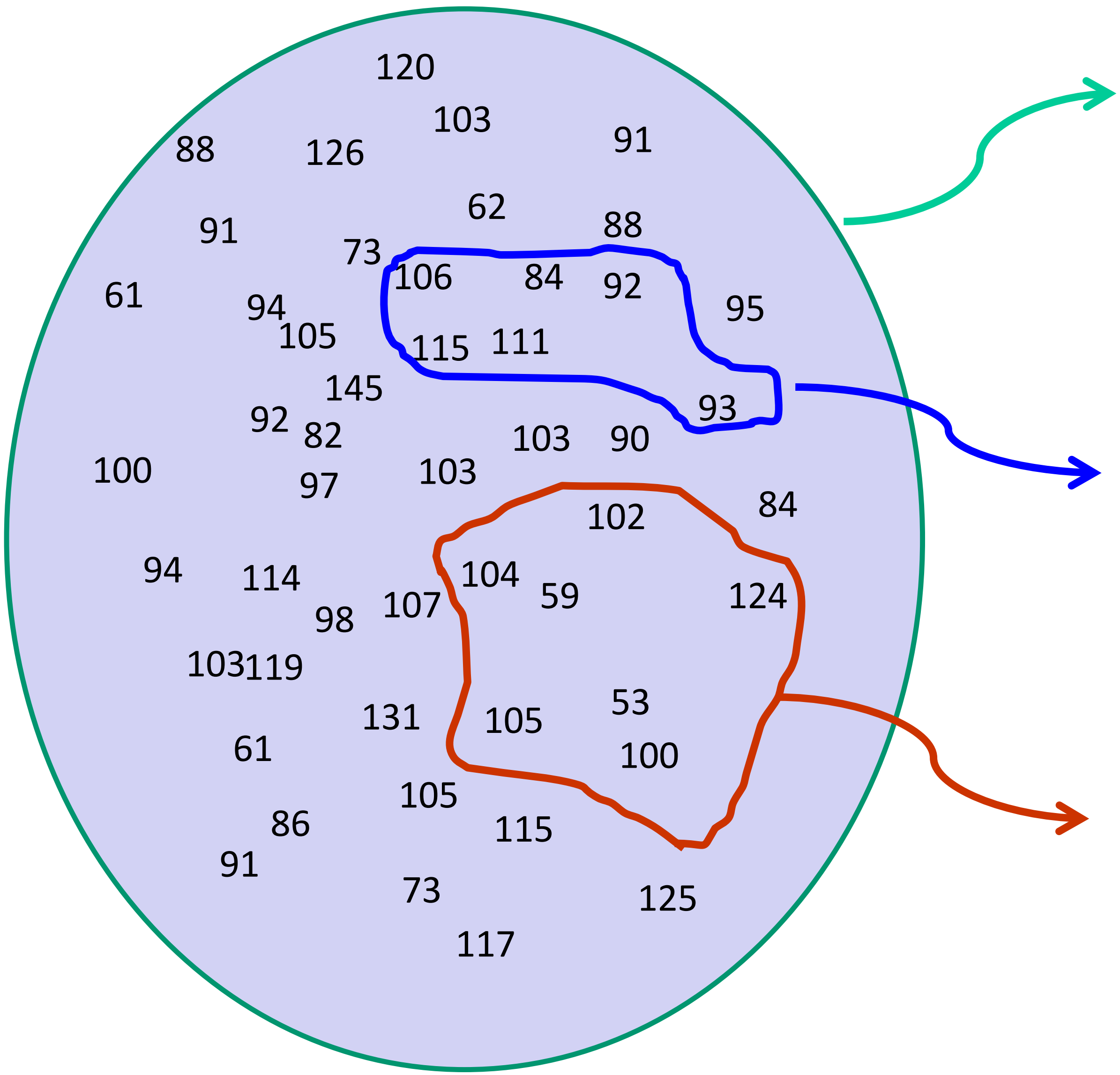
Os **parâmetros** desempenham em relação às distribuições **populacionais** um papel idêntico ao que as **estatísticas** desempenhavam em relação às distribuições **amostrais**

Amostra → **Estatísticas**

variáveis: *variam de amostra para amostra*

População → **Parâmetros**

fixos: *têm sempre o mesmo valor*



$\mu = 97.60$

$\sigma = 19.18$

Parâmetros

$n = 6$

$\overline{x} = 100.17$

$s = 12.25$

Estatísticas

$n = 7$

$\overline{x} = 92.43$

$s = 26.17$

jlborges@fe.up.pt

Parâmetros de localização

Valor Esperado da População ou média populacional

$$\mu_Y = E(Y) = \sum_y y \cdot p(y)$$

$$\mu_X = E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) \cdot dx$$

Exemplo:

Número de E's obtidos numa sequência de três lançamentos de uma moeda E-C.

Variável discreta Y	Função de probabilidade
0	1/8
1	3/8
2	3/8
3	1/8

$$\mu_Y = E(Y) = \sum_y y \cdot p(y) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = 1.5$$

Parâmetros de Dispersão

Desvio absoluto médio ou desvio médio

$$\delta_Y = \sum_y |y - \mu_Y| \cdot p(y)$$

$$\delta_X = \int_{-\infty}^{+\infty} |x - \mu_X| \cdot f(x) \cdot dx$$

Variância

$$\sigma_Y^2 = VAR(Y) = \sum_y (y - \mu_Y)^2 \cdot p(y)$$

$$\sigma_X^2 = VAR(X) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 \cdot f(x) \cdot dx$$

Desvio padrão

$$\sigma = \sqrt{\sigma^2}$$

Exemplo:

Variável discreta Y	Função de probabilidade
0	1/8
1	3/8
2	3/8
3	1/8

$$\mu_Y = 1.5$$

$$\sigma_Y^2 = (0 - 1.5)^2 \cdot \frac{1}{8} + (1 - 1.5)^2 \cdot \frac{3}{8} + $$
$$+ (2 - 1.5)^2 \cdot \frac{3}{8} + (3 - 1.5)^2 \cdot \frac{1}{8}$$

Outros Parâmetros

Coeficiente de assimetria

$$\gamma_1 = \frac{\sum_y (y - \mu_Y)^3 \cdot p(y)}{\sigma^3}$$

Coeficiente de kurtose

$$\gamma_2 = \frac{\sum_y (y - \mu_Y)^4 \cdot p(y)}{\sigma^4} - 3$$

jlborges@fe.up.pt

Cristiano Ronaldo

y: num golos marcados por jogo

Y	Função de Probabilidade	Função Distribuição
0	0,542	0,542
1	0,278	0,819
2	0,126	0,946
3	0,046	0,991
4	0,006	0,997
5	0,003	1,000

$$E(Y) = \mu = (0 \cdot 0.542 + 1 \cdot 0.278 + \cdots + 5 \cdot 0.003)=0.705$$

$$\eta=0$$

$$\xi = 0$$

$$var(Y) = \sigma^2 = 0.542(0 - 0.705)^2 + \cdots + 0.003(5 - 0.705)^2 = 0.863$$

$$\sigma = \sqrt{\sigma^2} = 0.929$$

$$\gamma_1 = \frac{0.542(0 - 0.705)^3 + \cdots + 0.003(5 - 0.705)^3}{0.929^3} = 1.353$$

$$\gamma_2 = \frac{0.542(0 - 0.705)^4 + \cdots + 0.003(5 - 0.705)^4}{0.929^4} - 3 = 1.628$$

Exemplo: apostar 1€ que vai sair ímpar na roleta.

Caracterizar variável aleatória lucro, X.

X	P(X)
0€	19/37
2€	18/37

Qual o ganho esperado a longo prazo?

$E(x)=(19/37) \cdot 0 + (18/37) \cdot 2 = 0.973€$

0.973€ ganho em média por jogada.

Lucro esperado = 0.973 –1 = - 0.027€.

E se jogar duas vezes 1€, qual é o valor esperado do lucro?

X	P(X)
0€	(19/37) * (19/37) = 0.264
2€	2 * (18/37) * (19/37) = 0.499
4€	(18/37) * (18/37) = 0.237
	0.26 + 0.50 + 0.24 = 1

Lucro = 2 – 0.499 * 2 – 0.237 * 4 = -0.054

jlborges@fe.up.pt

Max daily temp in degree Celsius (Porto 2011/2012)

Day	Max	Day	Max	Day	Max	Day	Max
03-Mar	16	29-Jan	14	26-Dez	16	22-Nov	15
02-Mar	15	28-Jan	14	25-Dez	15	21-Nov	17
01-Mar	16	27-Jan	12	24-Dez	14	20-Nov	17
29-Fev	15	26-Jan	15	23-Dez	16	19-Nov	16
28-Fev	19	25-Jan	18	22-Dez	16	18-Nov	18
27-Fev	19	24-Jan	15	21-Dez	13	17-Nov	16
26-Fev	16	23-Jan	17	20-Dez	12	16-Nov	16
25-Fev	19	22-Jan	17	19-Dez	12	15-Nov	17
24-Fev	21	21-Jan	17	18-Dez	13	14-Nov	23
23-Fev	19	20-Jan	16	17-Dez	16	13-Nov	24
22-Fev	15	19-Jan	15	16-Dez	16	12-Nov	22
21-Fev	16	18-Jan	15	15-Dez	14	11-Nov	18
20-Fev	15	17-Jan	13	14-Dez	16	10-Nov	18
19-Fev	15	16-Jan	12	13-Dez	15	09-Nov	18
18-Fev	15	15-Jan	11	12-Dez	13	08-Nov	17
17-Fev	16	14-Jan	14	11-Dez	14	07-Nov	17
16-Fev	16	13-Jan	14	10-Dez	15	06-Nov	16
15-Fev	15	12-Jan	13	09-Dez	16	05-Nov	16
14-Fev	13	11-Jan	15	08-Dez	17	04-Nov	16
13-Fev	11	10-Jan	15	07-Dez	16	03-Nov	19
12-Fev	13	09-Jan	17	06-Dez	16	02-Nov	18
11-Fev	12	08-Jan	16	05-Dez	16	01-Nov	15
10-Fev	12	07-Jan	16	04-Dez	14		
09-Fev	12	06-Jan	15	03-Dez	14		
08-Fev	15	05-Jan	16	02-Dez	16		
07-Fev	16	04-Jan	15	01-Dez	17		
06-Fev	15	03-Jan	15	30-Nov	16		
05-Fev	11	02-Jan	14	29-Nov	15		
04-Fev	10	01-Jan	13	28-Nov	16		
03-Fev	11	31-Dez	12	27-Nov	19		
02-Fev	13	30-Dez	14	26-Nov	18		
01-Fev	13	29-Dez	13	25-Nov	20		
31-Jan	14	28-Dez	15	24-Nov	21		
30-Jan	14	27-Dez	16	23-Nov	19		

raw data

Frequencies

Temp	Freq	%
10	1	0.01
11	4	0.03
12	8	0.06
13	11	0.09
14	13	0.10
15	25	0.20
16	31	0.25
17	11	0.09
18	7	0.06
19	7	0.06
20	1	0.01
21	2	0.02
22	1	0.01
23	1	0.01
24	1	0.01

Average 15.51

St. Dev 2.421

What is the average temp in Fahrenheit?

And the standard deviation?

$$°F = °C \cdot \frac{9}{5} + 32$$

$$\mu = \sum \frac{x}{n} = \sum y \cdot p(y) = 15.51$$

jlborges@fe.up.pt

Variáveis Transformadas

Para uma transformação de variável **W=ϕ(z)**,

podemos definir a partir de cada valor da variável original (**Z**) o valor correspondente da variável transformada (**W**) e especificar a distribuição da nova variável, sendo possível calcular em seguida os parâmetros desta distribuição

No entanto, o **valor esperado** e a **variância** podem ser calculados directamente (pelo menos de uma forma aproximada) a partir dos parâmetros relativos à variável original

Transformações lineares:

$$W = a + b \cdot Z$$

$$E(W) = \sum_w w \cdot p(w) = \sum_z (a + b \cdot z) \cdot p(z) =$$

$$= a \cdot \sum_z p(z) + b \cdot \sum_z z \cdot p(z) = a + b \cdot E(Z)$$

$$Var(W) = \sum_w (w - \mu_W)^2 \cdot p(w) = \sum_z [(a + b \cdot z) - (a + b \cdot \mu_Z)]^2 \cdot p(z) =$$

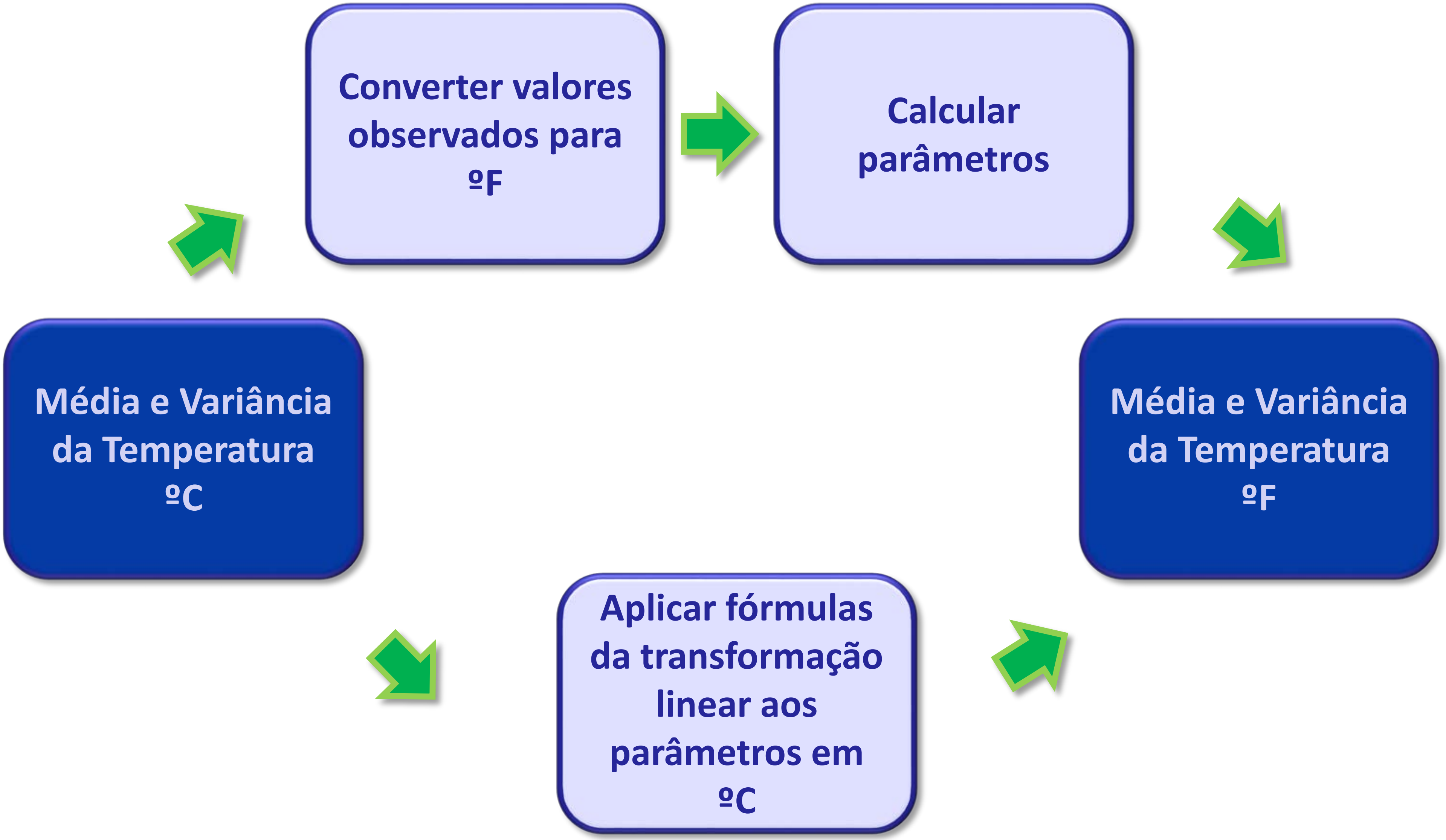
$$= b^2 \cdot \sum_z (z - \mu_Z)^2 \cdot p(z) = b^2 \cdot Var(Z)$$

$$W = a + b \cdot Z$$

$$\mu_W = a + b \cdot \mu_Z$$

$$\sigma_W^2 = b^2 \cdot \sigma_Z^2$$

Permite calcular os parâmetros da variável transformada a partir dos parâmetros da variável original


Média e Variância da Temperatura ºC
Converter valores observados para ºF
Calcular parâmetros
Média e Variância da Temperatura ºF
Aplicar fórmulas da transformação linear aos parâmetros em ºC

jlborges@fe.up.pt

ºC				ºF			
16	14	16	15	60.8	57.2	60.8	59
15	14	15	17	59	57.2	59	62.6
16	12	14	17	60.8	53.6	57.2	62.6
15	15	16	16	59	59	60.8	60.8
19	18	16	18	66.2	64.4	60.8	64.4
19	15	13	16	66.2	59	55.4	60.8
16	17	12	16	60.8	62.6	53.6	60.8
19	17	12	17	66.2	62.6	53.6	62.6
21	17	13	23	69.8	62.6	55.4	73.4
19	16	16	24	66.2	60.8	60.8	75.2
15	15	16	22	59	59	60.8	71.6
16	15	14	18	60.8	59	57.2	64.4
15	13	16	18	59	55.4	60.8	64.4
15	12	15	18	59	53.6	59	64.4
15	11	13	17	59	51.8	55.4	62.6
16	14	14	17	60.8	57.2	57.2	62.6
16	14	15	16	60.8	57.2	59	60.8
15	13	16	16	59	55.4	60.8	60.8
13	15	17	16	55.4	59	62.6	60.8
11	15	16	19	51.8	59	60.8	66.2
13	17	16	18	55.4	62.6	60.8	64.4
12	16	16	15	53.6	60.8	60.8	59
12	16	14		53.6	60.8	57.2	
12	15	14		53.6	59	57.2	
15	16	16		59	60.8	60.8	
16	15	17		60.8	59	62.6	
15	15	16		59	59	60.8	
11	14	15		51.8	57.2	59	
10	13	16		50	55.4	60.8	
11	12	19		51.8	53.6	66.2	
13	14	18		55.4	57.2	64.4	
13	13	20		55.4	55.4	68	
14	15	21		57.2	59	69.8	
14	16	19		57.2	60.8	66.2	

using the definitions

$$\mu_F = \sum \frac{x}{n} = 59.915\ °\mathrm{F}$$

$$\sigma_F{}^2 = \sum \frac{(x-\mu)^2}{n} = 4.385^2\ °\mathrm{F}$$

transforming the parameters in ºC

$$\mu_F = \mu_C \cdot \frac{9}{5} + 32 = 15.508 \cdot \frac{9}{5} + 32 = 59.915$$

$$\sigma_F = \sqrt{\sigma_C{}^2 \cdot \left(\frac{9}{5}\right)^2} = 4.358\ °\mathrm{F}$$

<u>Exemplo</u>

Sejam A e B duas variáveis que representam o lançamento de cada um de dois dados

<u>um dado</u>

s	p(s)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

$$\mu_A = \mu_B = (1+2+3+4+5+6) / 6 = 3.5$$

$$\sigma^2_A = \sigma^2_B = [(1-3.5)^2 + \ldots + (6-3.5)^2] \cdot (1/6)$$

$$= 2.9167$$

Se lançar um dado muitas vezes em média vou obter uma pontuação de 3.5.

Qual a pontuação média da soma total no **lançamento de dois dados?**

Res 1: definir variável pontuação total de dois dados e calcular parâmetros.

variável soma dois dados

s	p(s)	s.p(s)	$(s\text{-}x)^2.p(s)$
2	1/36	0.056	0.694
3	2/36	0.167	0.889
4	3/36	0.333	0.750
5	4/36	0.556	0.444
6	5/36	0.833	0.139
7	6/36	1.167	0.000
8	5/36	1.111	0.139
9	4/36	1.000	0.444
10	3/36	0.833	0.750
11	2/36	0.611	0.889
12	1/36	0.333	0.694
		$\mu_{s=}7$	σ^2_s=5,83

Res 2: obter valor esperado e variância da soma da pontuação (S) a partir dos parâmetros da variável correspondente ao lançamento de um dado (A)

dois dados

S = A + B terá

μ_S= E(A+B)=3.5 + 3.5 = 7

σ^2_S= VAR(A+B)= $\sigma^2_A + \sigma^2_B$

= 2 · 2.9167 = 5.83

E 5 dados?

Soma e diferença de variáveis:

$$E(X + Y) = E(X) + E(Y)$$

$$E(X - Y) = E(X) - E(Y)$$

$$V(X + Y) = V(X) + V(Y) + 2Cov(X,Y)$$

$$V(X - Y) = V(X) + V(Y) - 2Cov(X,Y)$$

Cov(X,Y) - covariância, = 0 no caso de variáveis independentes

Simpson's paradox (wikipedia)

- Simpson's paradox (or the Yule–Simpson effect) is a paradox in which

  a **trend** that appears in **different groups**

  of data disappears when these groups are combined,

  and the **reverse** trend appears **for the aggregate data**.

Simpson's paradox

- One of the best known real life examples occurred when the University of California, Berkeley was **sued for bias against women** who had applied for admission to graduate schools there.
- The admission figures for the fall of 1973 showed that:
  - **men applying** were **more likely** than women **to be admitted**,
  - the difference was so large that it was unlikely to be due to chance.

	Applicants	Admitted
Men	8442	**44%**
Women	4321	35%

Data from six large departments

	Applicants	Admited		Denied	
Men	2651	1158	44%	1493	56%
Women	1835	557	30%	1278	70%

	Men					Women				
Department	**Applicants**	**Admited**		**Denied**		**Applicants**	**Admited**		**Denied**	
A	825	512	62%	313	38%	108	89	82%	19	18%
B	520	313	60%	207	40%	25	17	68%	8	32%
C	325	120	37%	205	63%	593	202	34%	391	66%
D	417	138	33%	279	67%	375	131	35%	244	65%
E	191	53	28%	138	72%	393	94	24%	299	76%
F	373	22	6%	351	94%	341	24	7%	317	93%

Anytime that data is aggregated, watch out for this paradox to show up!

Dustan Mohr, Minnesota Twins

2013 Major League Baseball season

	Runners in Scoring Position		No Runners in Scoring Position	
	Mohr	Erstad	Mohr	Erstad
Hits	19	9	68	56
Number of times batted	97	50	251	208
Batting average	0.196	0.180	0.271	0.269

	Overall	
	Mohr	Erstad
Hits	87	65
Number of times batted	348	258
Batting average	0.250	0.252

Darin Erstad, San Francisco Giants

Muito importante

- Compreender muito bem o conceito de variável aleatória

- Khan Academy
  - Random variables and probability distributions
  - https://www.khanacademy.org/math/probability/random-variables-topic/random_variables_prob_dist/v/random-variables

Resultados de Aprendizagem

- Definir a função de probabilidade e distribuição de uma variável aleatória discreta
- Definir a função densidade de probabilidade e distribuição de uma variável aleatória contínua
- Calcular probabilidades de acontecimentos a partir das funções de probabilidade
- Calcular os parâmetros de localização e de dispersão de variáveis aleatórias
- Calcular a média e variância de uma transformação linear de uma variável aleatória
- Calcular a média e a variância de uma combinação linear de um conjunto de variáveis aleatórias