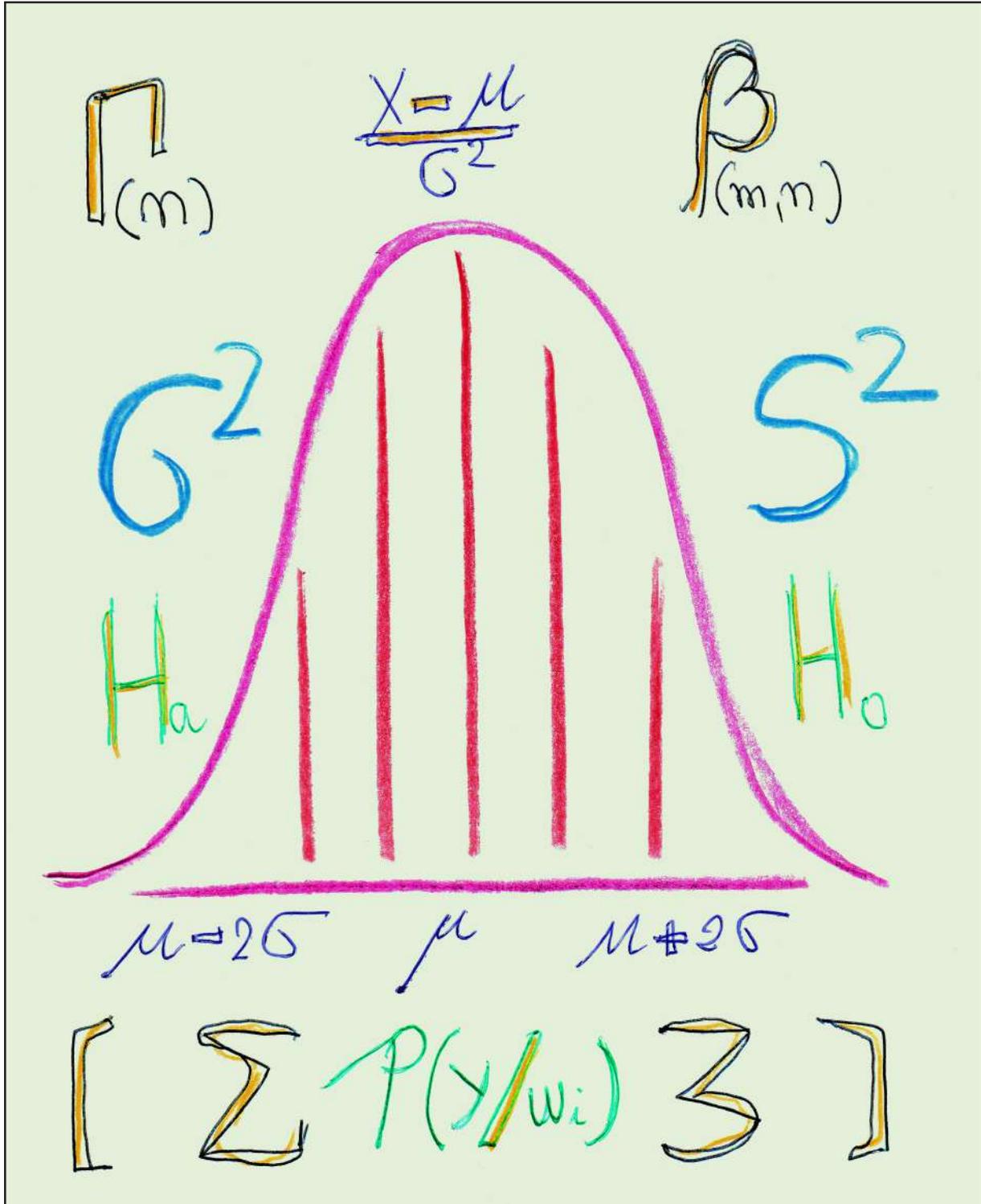




PRANDIANO®



EXPERIMENTANDO ESTATÍSTICA®

Prof. Aguinaldo Prandini Ricieri



12.0 Sintaxe e Relação de Concordância Formal

Segue uma estatística de duas palavras adjacentes (substantivo sucedido de adjetivo) contabilizados em 100000 frases – textos separados por um ponto final –, que vetorizados permite classificar um certo vetor/frase: $0,21\hat{i} + 0,69\hat{j}$.

Atributo	Banana (0,25)			Carro (0,30)			Mulher (0,45)		
Doce Nanica Madura	μ	0,12	σ	$\leq 0,03$			$\leq 0,04$		
	↓	0,65	↓						
	0,333	0,23	0,228	μ			$\leq 0,03$		
Novo Usado Vermelho	$\leq 0,02$								
				0,333	0,50	0,143			
Magra Grávida Inteligente	$\leq 0,01$			$\leq 0,02$			μ	0,60	σ
							↓	0,10	↓
							0,333	0,30	0,205

Banco de Dados Estatizados

Vetor $(\vec{S} = 0,21\hat{i} + 0,69\hat{j})$ Frase

<p>Banana (\vec{B})</p>	<p>Carro (\vec{C})</p>	<p>Mulher (\vec{M})</p>
<p>Numerização (\vec{B})</p>	<p>Numerização (\vec{C})</p>	<p>Numerização (\vec{M})</p>

Banco de Dados Vetorizados

Embedding Sentences With Dot Product



12.2 Extensão do Classificador Estatístico Sintático

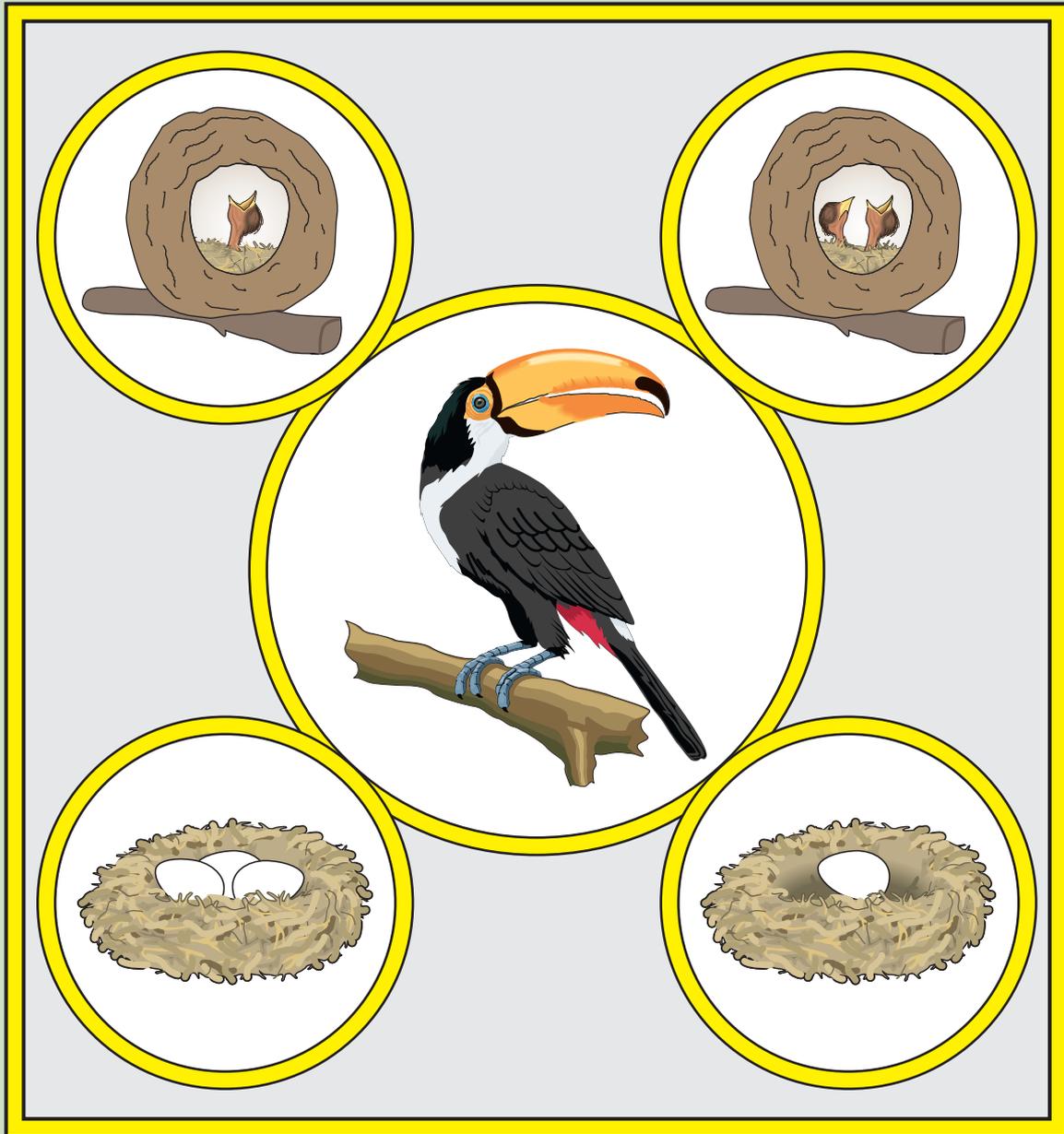
<p>Certo Rádio</p> $\begin{matrix} \vec{R}_1 & \vec{R}_2 & \vec{R}_3 \\ & \uparrow & \\ & \vec{R} & \end{matrix}$	<p>Alguma Panela</p> $\begin{matrix} & & \vec{P}_2 \\ & \swarrow & \\ \vec{P} & \rightarrow & \vec{P}_1 \\ & \searrow & \\ & & \vec{P}_3 \end{matrix}$	<p>A Jarra</p> $\begin{matrix} \vec{J}_2 & & \\ & \swarrow & \\ & \vec{J} & \\ & \swarrow & \\ \vec{J}_1 & \leftarrow & \\ & \searrow & \\ & & \vec{J}_3 \end{matrix}$
<p>Todo Gato</p> $\begin{matrix} \vec{G}_1 & & \\ & \swarrow & \\ & \vec{G} & \\ & \swarrow & \\ \vec{G}_2 & \leftarrow & \\ & \searrow & \\ & & \vec{G}_3 \end{matrix}$	<p>O Leão</p> $\begin{matrix} \vec{L} \\ \swarrow \quad \downarrow \quad \searrow \\ \vec{L}_1 \quad \vec{L}_2 \quad \vec{L}_3 \end{matrix}$	<p>Outra Onça</p> $\begin{matrix} & & \vec{O}_1 \\ & \swarrow & \\ \vec{O} & \rightarrow & \vec{O}_2 \\ & \searrow & \\ & & \vec{O}_3 \end{matrix}$
<p>Uma Banana</p> $\begin{matrix} & & \vec{B}_1 \\ & \swarrow & \\ \vec{B} & \rightarrow & \vec{B}_2 \\ & \searrow & \\ & & \vec{B}_3 \end{matrix}$	<p>Certo Carro</p> $\begin{matrix} \vec{C}_1 & \vec{C}_2 & \vec{C}_3 \\ & \uparrow & \\ & \vec{C} & \end{matrix}$	<p>Toda Mulher</p> $\begin{matrix} \vec{M}_1 & & \\ & \swarrow & \\ & \vec{M} & \\ & \swarrow & \\ \vec{M}_2 & \leftarrow & \\ & \searrow & \\ & & \vec{M}_3 \end{matrix}$
<p>Minha Andorinha</p> $\begin{matrix} \vec{A}_1 & \vec{A}_2 & \vec{A}_3 \\ & \uparrow & \\ & \vec{A} & \end{matrix}$	<p>Um Sabiá</p> $\begin{matrix} \vec{S} \\ \swarrow \quad \downarrow \quad \searrow \\ \vec{S}_1 \quad \vec{S}_2 \quad \vec{S}_3 \end{matrix}$	<p>Qualquer Tucano</p> $\begin{matrix} & & \vec{T}_1 \\ & \swarrow & \\ & \vec{T} & \\ & \swarrow & \\ & \vec{T}_2 & \\ & \searrow & \\ & & \vec{T}_3 \end{matrix}$
<p>Meu Nemo</p> $\begin{matrix} \vec{N} \\ \swarrow \quad \downarrow \quad \searrow \\ \vec{N}_1 \quad \vec{N}_2 \quad \vec{N}_3 \end{matrix}$	<p>Algum Urso</p> $\begin{matrix} & & \vec{U}_1 \\ & \swarrow & \\ \vec{U} & \rightarrow & \vec{U}_2 \\ & \searrow & \\ & & \vec{U}_3 \end{matrix}$	<p>Vosso Rino</p> $\begin{matrix} \vec{R}_1 & & \\ & \swarrow & \\ & \vec{R} & \\ & \swarrow & \\ \vec{R}_2 & \leftarrow & \\ & \searrow & \\ & & \vec{R}_3 \end{matrix}$

Sintaxe de Substantivos Precedidos de Pronome



TUCANAGEM

Alimentação de Tucanos



Iterações Aleatórias

Modelo Estatístico Otimizador

Prof. Aguinaldo Prandini Ricieri



Tucanagem Randômica

1.0 Heurística Sobre o Voo do Tucano

A tucanagem descreve uma trajetória formada por uma sucessão de mudanças aleatórias como é o deslocamento do tucano na busca do alimento (ovos e filhotes de pássaros). Trata-se de um processo estocástico onde a distribuição de probabilidade para o tucano alimentar-se depende da sua posição atual e não da sequência de supostos eventos precedentes – Cadeia de Andrei Markov (1856 - 1922).

O modelo matemático em questão tem como alicerce os chamados voos de Paul Pierre Lévy (1886 - 1971) que se caracterizam por trajetórias pertencentes à Geometria Fractal especificadas pela lei: $Sig = S \cdot U // |V|^{\frac{1}{b}}$ sendo $0 < b \leq 2$.

Esse tipo de busca aleatória/fractal – que tem relação com o movimento de diversos animais pela procura de alimento – permite que se explore de modo probabilístico o domínio da função que se pretende otimizar evitando mínimos e máximos locais. Trata-se de computação iterativa do tipo meta-heurística (MH).

Para que exista uma convergência para a solução com o menor número de iterações, a lei de Lévy exige uma modelagem matemática que associe os valores S e b como parâmetros que permitirão posicionar os tucanos nos pontos de alimentação em questão: simulações significativas (Sig) feitas numericamente.

①. Lévy Flights, Non-Local Search and Simulated Annealing
Pavlyukevich I.

Journal Computational Physics, v. 226, p. 1830, 2007.

②. Lévy Flights in Dobe Ju
Human Ecol v. 35, p. 124, 2007
Brown C., Liebovitch L.S, Glendin R.

③. Cuckoo Search Via Lévy Flights
Proceeding of Word Congress Computing, 2009
Xin-She Yang, Suash Deb

Alimentação Significativa dos Tucanos



6.0 Busca Aleatória pelas Facilidades de Alimento 2

L	U = Rx(-1,+1) • Su		
A	-0,8325	1,000	-0,8325
B	0,1487	1,000	0,1487
C	0,2645	1,000	0,2645
D	0,13364	1,000	0,1336
E	0,45547	1,000	0,4555
F	-0,3348	1,000	-0,3348
G	0,81646	1,000	0,8165
H	0,69412	1,000	0,6941
I	0,8524	1,000	0,8524
J	0,7625	1,000	0,7625
	RX	Su	U

Umidade

(X)	(Y)
-0,8325	0,4254
0,1487	0,0356
0,2645	0,4480
0,1336	0,3050
0,4555	0,0333
-0,3348	0,7756
0,8165	0,9314
0,6941	0,4821
0,8524	-0,2659
0,7625	-0,0146
Relativo ao U	

L	U = Ry(-1,+1) • Su		
A	0,4254	1,0000	0,4254
B	0,0356	1,0000	0,0356
C	0,4480	1,0000	0,4480
D	0,3050	1,0000	0,3050
E	0,0333	1,0000	0,0333
F	0,7756	1,0000	0,7756
G	0,9314	1,0000	0,9314
H	0,4821	1,0000	0,4821
I	-0,2659	1,0000	-0,2659
J	-0,0146	1,0000	-0,0146
	RY	Su	U

Umidade

L	V = Rx (-1,+1) • Sv		
A	0,6258	1,0000	0,6258
B	0,4468	1,0000	0,4468
C	0,2154	1,0000	0,2154
D	-0,4058	1,0000	-0,4058
E	0,93484	1,0000	0,9348
F	0,4985	1,0000	0,4985
G	0,4258	1,0000	0,4258
H	0,672	1,0000	0,6720
I	0,3725	1,0000	0,3725
J	0,55	1,0000	0,5500
	RX	Sv	V

Velocidade

(X)	(Y)
0,6258	0,8425
0,4468	-0,3900
0,2154	0,0459
-0,4058	0,3824
0,9348	-0,1287
0,4985	0,1720
0,4258	0,7501
0,6720	0,8354
0,3725	0,6748
0,5500	0,7531
Relativo ao V	

L	V = Ry (-1,+1) • Sv		
A	0,8425	1,0000	0,8425
B	-0,3900	1,0000	-0,3900
C	0,0459	1,0000	0,0459
D	0,3824	1,0000	0,3824
E	-0,1287	1,0000	-0,1287
F	0,1720	1,0000	0,1720
G	0,7501	1,0000	0,7501
H	0,8354	1,0000	0,8354
I	0,6748	1,0000	0,6748
J	0,7531	1,0000	0,7531
	RY	Sv	V

Velocidade

L	Sig (X)
A	-0,7926
B	0,1772
C	0,5128
D	0,1698
E	0,3319
F	-0,3710
G	1,0049
H	0,6302
I	1,1469
J	0,7913
Para U	

Umidade

b = 1,50

$$\text{Sig}() = \frac{S \circ U}{|V|^{\frac{1}{b}}}$$

$$\text{Sig}() = \frac{0,6966 \circ U}{|V|^{0,666}}$$

S = 0,6966

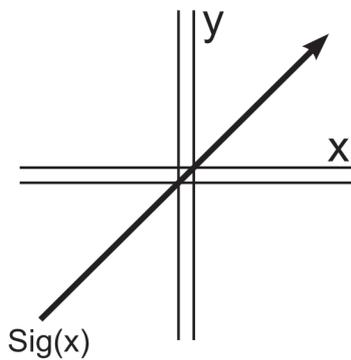
L	Sig (Y)
A	0,3322
B	0,0464
C	2,4358
D	0,4033
E	0,0911
F	1,7469
G	0,7859
H	0,3786
I	-0,2407
J	-0,0123
Para V	

Velocidade

Significação Associativa de Lévy para U e V



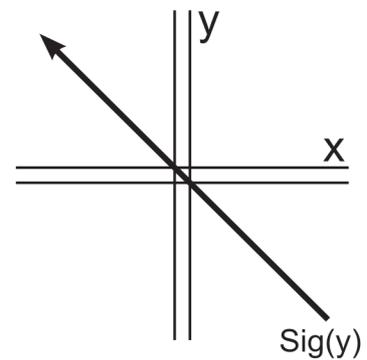
6.1 Mudanças das Posições dos Tucanos 2



Sig (X)
-0,7926
0,1772
0,5128
0,1698
0,3319
-0,3710
1,0049
0,6302
1,1469
0,7913

X

Sig (Y)
0,3322
0,0464
2,4358
0,4033
0,0911
1,7469
0,7859
0,3786
-0,2407
-0,0123



(Lt,p) X	Lt,a	e	Sig (Lt,a)	Lt,a	MF
3,0602	3,0602	0,10	-0,7926	3,0602	3,0602
2,9808	2,9822	0,10	0,1772	2,9822	3,0602
3,1351	3,1314	0,10	0,5128	3,1314	3,0602
3,0907	3,0901	0,10	0,1698	3,0901	3,0602
3,0932	3,0922	0,10	0,3319	3,0922	3,0602
3,8914	3,9234	0,10	-0,3710	3,9234	3,0602
4,3149	4,2003	0,10	1,0049	4,2003	3,0602
4,2498	4,1793	0,10	0,6302	4,1793	3,0602
3,0894	3,0863	0,10	1,1469	3,0863	3,0602
3,9149	3,8522	0,10	0,7913	3,8522	3,0602

$$(L_{t,p}) = L_{t,a} + e \cdot \text{Sig}(L_{t,a}) \cdot [L_{t,a} - \text{MF}]$$

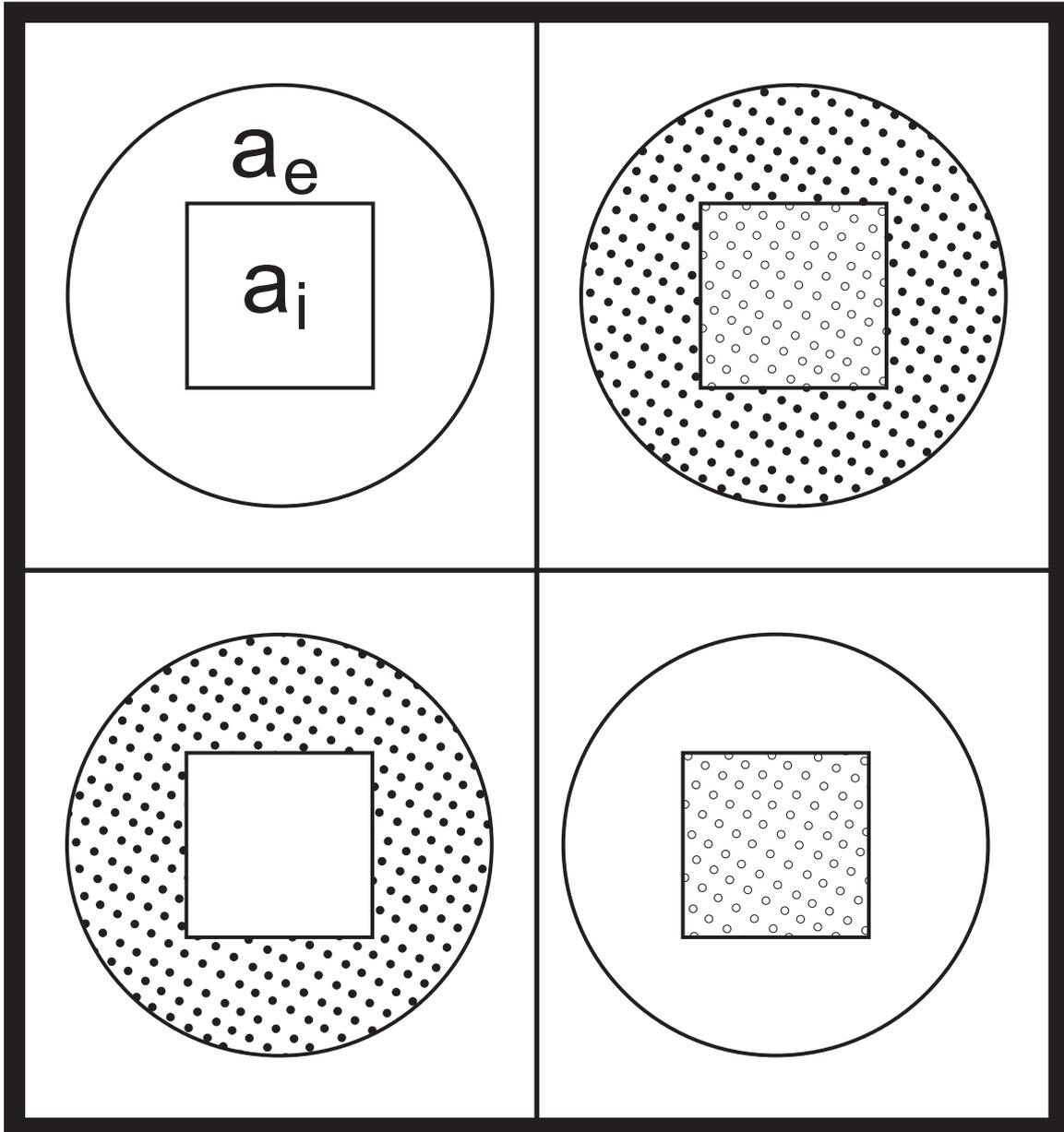
4,0407	4,0407	0,01	0,3322	4,0407	4,0407
1,8000	1,8010	0,01	0,0464	1,8010	4,0407
3,7831	3,7892	0,01	2,4358	3,7892	4,0407
2,3952	2,4018	0,01	0,4033	2,4018	4,0407
4,4984	4,4980	0,01	0,0911	4,4980	4,0407
4,4988	4,4909	0,01	1,7469	4,4909	4,0407
3,8503	3,8518	0,01	0,7859	3,8518	4,0407
3,4497	3,4519	0,01	0,3786	3,4519	4,0407
3,6526	3,6516	0,01	-0,2407	3,6516	4,0407
3,6134	3,6133	0,01	-0,0123	3,6133	4,0407
(Lt,p) Y	Lt,a	e	Sig (Lt,a)	Lt,a	MF

Iterações dos Tucanos por Facilidades 2



MONTE CARLO

Pontos Internos X Pontos Externos



Simulações Probabilísticas

Método Preditivo Numérico

Prof. Aguinaldo Prandini Ricieri



JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 247

SEPTEMBER 1949

Volume 44

THE MONTE CARLO METHOD

NICHOLAS METROPOLIS AND S. ULAM

Los Alamos Laboratory

We shall present here the motivation and a general description of a method dealing with a class of problems in mathematical physics. The method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences.

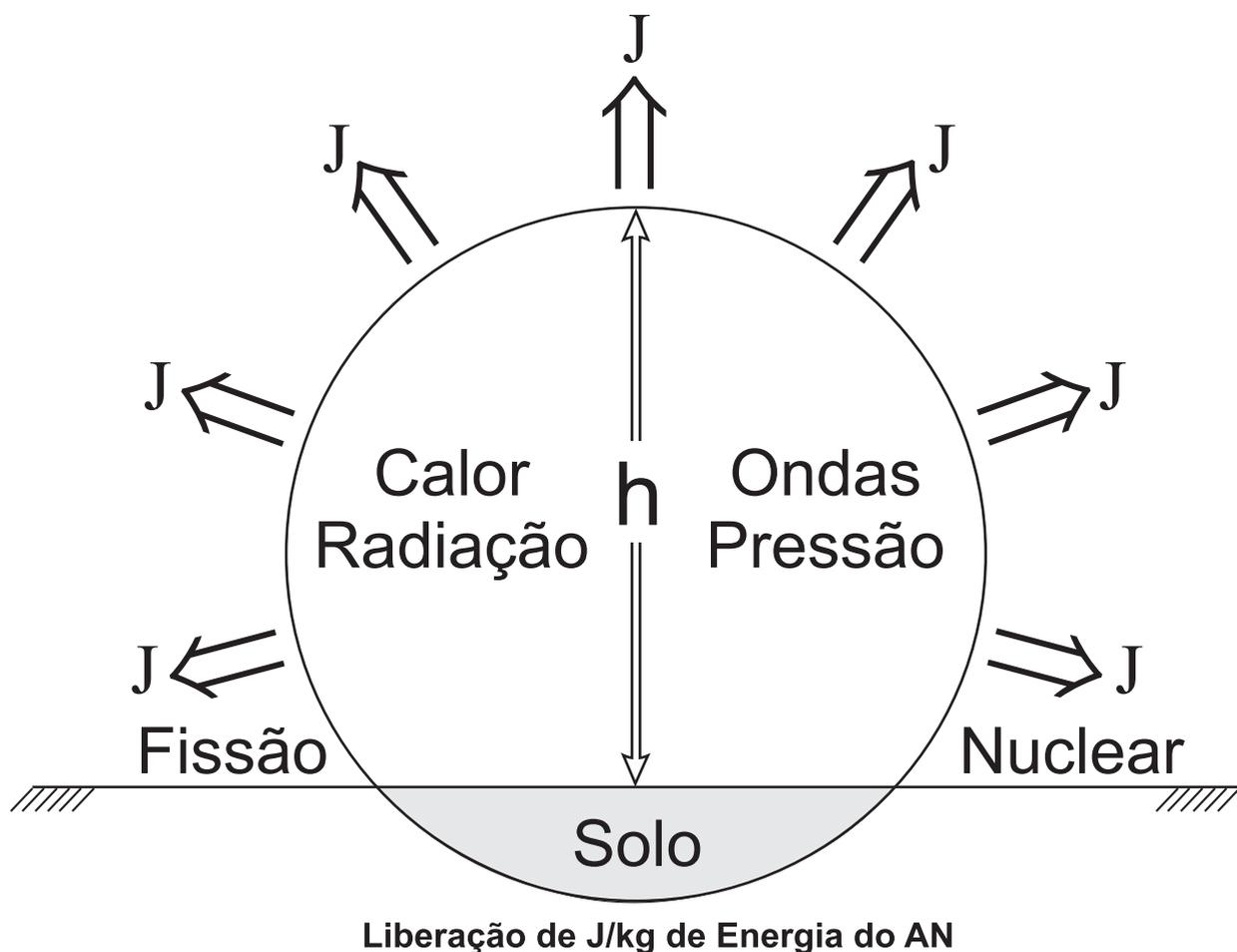
ALREADY in the nineteenth century a sharp distinction began to appear between two different mathematical methods of treating physical phenomena. Problems involving only a few particles were studied in classical mechanics, through the study of systems of ordinary differential equations. For the description of systems with very many particles, an entirely different technique was used, namely, the method of statistical mechanics. In this latter approach, one does not concentrate on the individual particles but studies the properties of *sets of particles*. In pure mathematics an intensive study of the properties of sets of points was the subject of a new field. This is the so-called theory of sets, the basic theory of integration, and the twentieth century development of the theory of probabilities prepared the formal apparatus for the use of such models in theoretical physics, i.e., description of properties of aggregates of points rather than of individual points and their coordinates.

Soon after the development of the calculus, the mathematical apparatus of partial differential equations was used for dealing with the problems of the physics of the continuum. Hydrodynamics is the most widely known field formulated in this fashion. A little later came the treatment of the problems of heat conduction and still later the field theories, like the electromagnetic theory of Maxwell. All this is very well known. It is of course important to remember that the study of the



2.0 Origem do Método de Monte Carlo

Esse método foi nominado por Nicholas Metropolis (1915 - 1999), estudado por John von Neumann (1903 - 1957) e aplicado por Stanislaw Ulam (1909 - 1984) no melhoramento de bombas atômicas (Projeto Manhattan / Los Alamos), isto é, foi desenvolvido para otimizar o poder explosivo de alguns artefatos nucleares (AN). Como exemplo, considere a simulação dos efeitos da explosão de uma bomba atômica lançada de um avião; como função da altura (h) em que ela explode.



Na fissão nuclear os núcleos do urânio/plutônio se “quebram” e liberam nêutrons, que dão início a uma reação em cadeia, desintegrando os núcleos atômicos próximos o que implica na liberação de uma grande quantidade de energia.

Simulação dos Efeitos Explosivos



7.1 Cálculo da Área da Distribuição Probabilística

1	-3	0,0044
2	-2,95	0,0051
3	-2,9	0,0060
4	-2,85	0,0069
5	-2,8	0,0079
6	-2,75	0,0091
7	-2,7	0,0104
8	-2,65	0,0119
9	-2,6	0,0136
10	-2,55	0,0154
11	-2,5	0,0175
12	-2,45	0,0198
13	-2,4	0,0224
14	-2,35	0,0252
15	-2,3	0,0283
16	-2,25	0,0317
17	-2,2	0,0355
18	-2,15	0,0395
19	-2,1	0,0440
20	-2,05	0,0488
21	-2	0,0540
22	-1,95	0,0596
23	-1,9	0,0656
24	-1,85	0,0721
25	-1,8	0,0789
26	-1,75	0,0863
27	-1,7	0,0940
28	-1,65	0,1023
29	-1,6	0,1109
30	-1,55	0,1200
31	-1,5	0,1295
32	-1,45	0,1394
33	-1,4	0,1497
34	-1,35	0,1604
35	-1,3	0,1714
36	-1,25	0,1826
37	-1,2	0,1942
38	-1,15	0,2059
39	-1,1	0,2178
40	-1,05	0,2299
41	-1	0,2419
42	-0,95	0,2540
43	-0,9	0,2661
44	-0,85	0,2780
45	-0,8	0,2897
46	-0,75	0,3011
47	-0,7	0,3122
48	-0,65	0,3229
49	-0,6	0,3332
50	-0,55	0,3429

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

$$\begin{matrix} \mu = 0 \\ \sigma = 1 \end{matrix}$$

$$P(x) = 0,3989 \cdot e^{-0,5x^2}$$

72	0,55	0,3429
73	0,60	0,3332
74	0,65	0,3229
75	0,70	0,3122
76	0,75	0,3011
77	0,80	0,2897
78	0,85	0,2780
79	0,90	0,2661
80	0,95	0,2540
81	1,00	0,2419
82	1,05	0,2299
83	1,10	0,2178
84	1,15	0,2059
85	1,20	0,1942
86	1,25	0,1826
87	1,30	0,1714
88	1,35	0,1604
89	1,40	0,1497
90	1,45	0,1394
91	1,50	0,1295
92	1,55	0,1200
93	1,60	0,1109
94	1,65	0,1023
95	1,70	0,0940
96	1,75	0,0863
97	1,80	0,0789
98	1,85	0,0721
99	1,90	0,0656
100	1,95	0,0596
101	2,00	0,0540
102	2,05	0,0488
103	2,10	0,0440
104	2,15	0,0395
105	2,20	0,0355
106	2,25	0,0317
107	2,30	0,0283
108	2,35	0,0252
109	2,40	0,0224
110	2,45	0,0198
111	2,50	0,0175
112	2,55	0,0154
113	2,60	0,0136
114	2,65	0,0119
115	2,70	0,0104
116	2,75	0,0091
117	2,80	0,0079
118	2,85	0,0069
119	2,90	0,0060
120	2,95	0,0051
121	3,00	0,0044

6/120

0,05	Área da Célula $2,5 \times 10^{-3}$ a_i a_e	0,40/8
0,05		
Elemento da Grade		

$$\frac{400 \cdot a_i}{1,00}$$

$$\frac{560 \cdot a_e}{1,40}$$

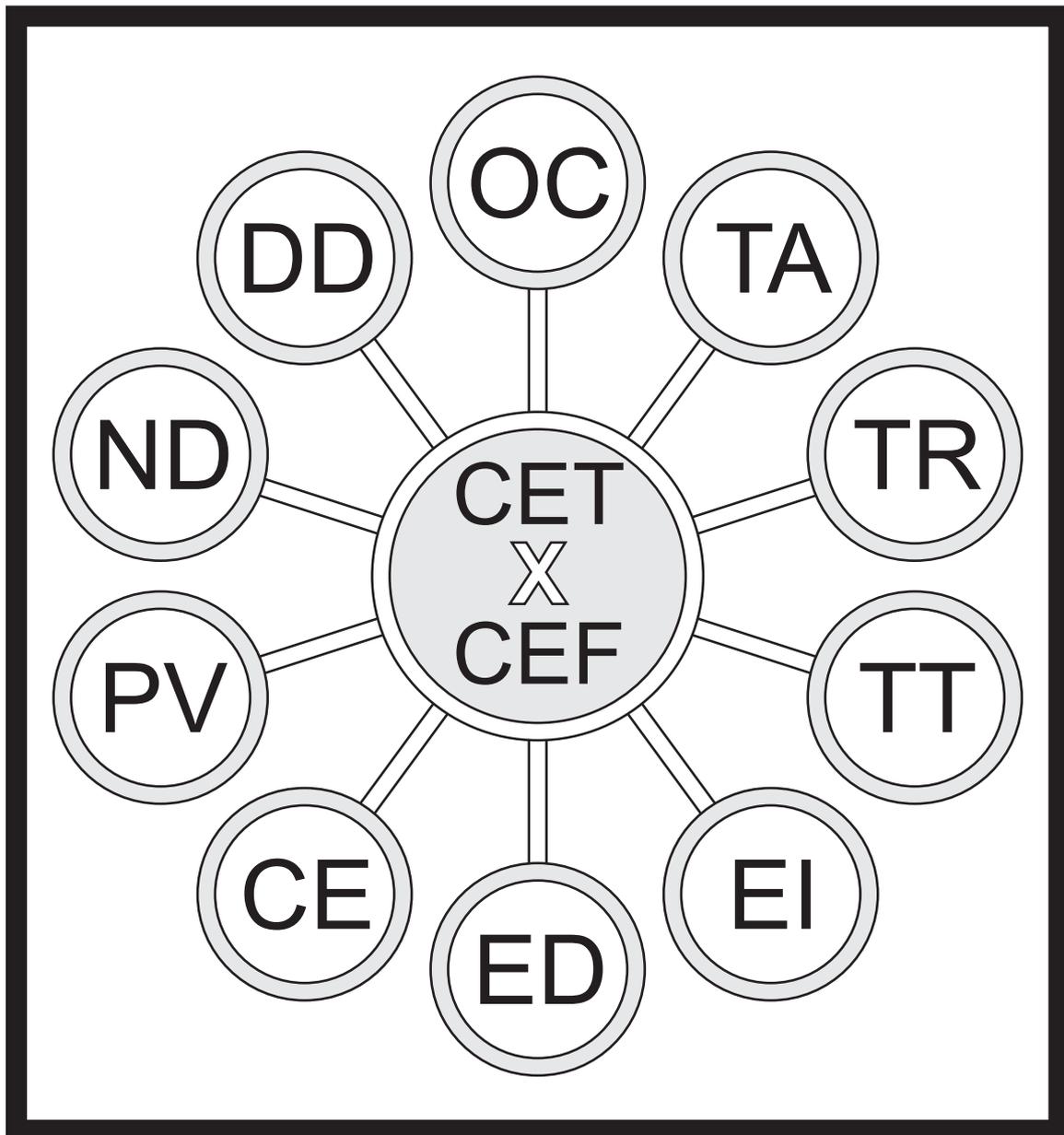
51	-0,5	0,3520
52	-0,45	0,3605
53	-0,4	0,3682
54	-0,35	0,3752
55	-0,3	0,3813
56	-0,25	0,3866
57	-0,2	0,3910
58	-0,15	0,3944
59	-0,1	0,3969
60	-0,05	0,3984
61	0	0,3989
62	0,05	0,3984
63	0,10	0,3969
64	0,15	0,3944
65	0,20	0,3910
66	0,25	0,3866
67	0,30	0,3813
68	0,35	0,3752
69	0,40	0,3682
70	0,45	0,3605
71	0,50	0,3520

Resultados da Probabilidade para $x \in [-3, +3]$



Fluxo de Produtos

Monte Carlo



Gestão de Estoque

Monitoração Estatística de Demanda

Prof. Aguinaldo Prandini Ricieri



4.3 Simulação Tertium do Estoque do Produto

DIA	DD	OC	TA	TR	TT	EI	ED	CE	CET	PV	CEF
00							0				
01	50	1	5	6	11	280	230	R\$ 2,00	R\$ 460,00	R\$ 23,00	R\$ -
02	37	2	4	6	16	0	193	R\$ 2,00	R\$ 386,00	R\$ 23,00	R\$ -
03	40	3	2	6	20	0	153	R\$ 2,00	R\$ 306,00	R\$ 23,00	R\$ -
04	45	4	0	6	24	0	108	R\$ 2,00	R\$ 216,00	R\$ 23,00	R\$ -
05	30	5	0	6	30	0	78	R\$ 2,00	R\$ 156,00	R\$ 23,00	R\$ -
06	32	6	1	6	37	0	46	R\$ 2,00	R\$ 92,00	R\$ 23,00	R\$ -
07	30	7	1	6	43	0	16	R\$ 2,00	R\$ 32,00	R\$ 23,00	R\$ -
08	30	8	0	6	48	0	-14	R\$ 2,00	R\$ -	R\$ 23,00	R\$ 322,00
09	47	9	1	6	55	0	-61	R\$ 2,00	R\$ -	R\$ 23,00	R\$ 1.403,00
10	30	10	0	6	60	0	-91	R\$ 2,00	R\$ -	R\$ 23,00	R\$ 2.093,00
11	43					280	146	R\$ 2,00	R\$ 292,00	R\$ 23,00	R\$ -
12	45					0	101	R\$ 2,00	R\$ 202,00	R\$ 23,00	R\$ -
13	35					0	66	R\$ 2,00	R\$ 132,00	R\$ 23,00	R\$ -
14	30					0	36	R\$ 2,00	R\$ 72,00	R\$ 23,00	R\$ -
15	36					0	0	R\$ 2,00	R\$ -	R\$ 23,00	R\$ -
16	40					280	240	R\$ 2,00	R\$ 480,00	R\$ 23,00	R\$ -
17	45					0	195	R\$ 2,00	R\$ 390,00	R\$ 23,00	R\$ -
18	45					0	150	R\$ 2,00	R\$ 300,00	R\$ 23,00	R\$ -
19	38					0	112	R\$ 2,00	R\$ 224,00	R\$ 23,00	R\$ -
20	50					280	342	R\$ 2,00	R\$ 684,00	R\$ 23,00	R\$ -
21	45					0	297	R\$ 2,00	R\$ 594,00	R\$ 23,00	R\$ -
22	37					0	260	R\$ 2,00	R\$ 520,00	R\$ 23,00	R\$ -
23	30					0	230	R\$ 2,00	R\$ 460,00	R\$ 23,00	R\$ -
24	40					280	470	R\$ 2,00	R\$ 940,00	R\$ 23,00	R\$ -
25	45					0	425	R\$ 2,00	R\$ 850,00	R\$ 23,00	R\$ -
26	43					0	382	R\$ 2,00	R\$ 764,00	R\$ 23,00	R\$ -
27	47					0	335	R\$ 2,00	R\$ 670,00	R\$ 23,00	R\$ -
28	40					0	295	R\$ 2,00	R\$ 590,00	R\$ 23,00	R\$ -
29	35					0	260	R\$ 2,00	R\$ 520,00	R\$ 23,00	R\$ -
30	30					280	510	R\$ 2,00	R\$ 1.020,00	R\$ 23,00	R\$ -
31	50					0	460	R\$ 2,00	R\$ 920,00	R\$ 23,00	R\$ -
32	40					0	420	R\$ 2,00	R\$ 840,00	R\$ 23,00	R\$ -
33	45					0	375	R\$ 2,00	R\$ 750,00	R\$ 23,00	R\$ -
34	48					0	327	R\$ 2,00	R\$ 654,00	R\$ 23,00	R\$ -
35	47					0	280	R\$ 2,00	R\$ 560,00	R\$ 23,00	R\$ -
36	50					0	230	R\$ 2,00	R\$ 460,00	R\$ 23,00	R\$ -
37	40					280	470	R\$ 2,00	R\$ 940,00	R\$ 23,00	R\$ -
38	43					0	427	R\$ 2,00	R\$ 854,00	R\$ 23,00	R\$ -
39	47					0	380	R\$ 2,00	R\$ 760,00	R\$ 23,00	R\$ -
40	35					0	345	R\$ 2,00	R\$ 690,00	R\$ 23,00	R\$ -
41	37					0	308	R\$ 2,00	R\$ 616,00	R\$ 23,00	R\$ -
42	40					0	268	R\$ 2,00	R\$ 536,00	R\$ 23,00	R\$ -
43	43					280	505	R\$ 2,00	R\$ 1.010,00	R\$ 23,00	R\$ -
44	40					0	465	R\$ 2,00	R\$ 930,00	R\$ 23,00	R\$ -
45	40					0	425	R\$ 2,00	R\$ 850,00	R\$ 23,00	R\$ -
46	38					0	387	R\$ 2,00	R\$ 774,00	R\$ 23,00	R\$ -
47	35					0	352	R\$ 2,00	R\$ 704,00	R\$ 23,00	R\$ -
48	42					280	590	R\$ 2,00	R\$ 1.180,00	R\$ 23,00	R\$ -
49	45					0	545	R\$ 2,00	R\$ 1.090,00	R\$ 23,00	R\$ -
50	40					0	505	R\$ 2,00	R\$ 1.010,00	R\$ 23,00	R\$ -
DADOS: MÉDIA da DEMANDA DIÁRIA = 40 un com $\sigma = 5$ un								Σ CET	R\$ 27.480,00	Σ CEF	R\$ 3.818,00



Monitoração de Ativos

Monte Carlo



Celesc
Distribuição S.A.



PETROBRAS

[B]³

Brasil - Bolsa - Balcão



bnf



sabesp

Previsões Futuras

Amostragem Estatística de Valores

Prof. Aguinaldo Prandini Ricieri



2.1 Simulações do iBovespa de Números 1, 2, 3

DATA	FECHAMENTO	X	Y = POLI(X)	DELTA	ALEA 1	SIMUL 1
06/07/2023	117426	0	118023,00	597,00	0,264	117671,78
07/07/2023	118898	1	118012,04	885,96	0,800	118460,18
10/07/2023	117942	2	118001,07	59,07	0,095	117397,25
11/07/2023	117220	3	117990,11	770,11	0,461	117932,56
12/07/2023	117666	4	117979,14	313,14	0,889	118559,61
13/07/2023	119264	5	117968,18	1295,82	0,093	117361,10
14/07/2023	117711	6	117957,22	246,22	0,877	118520,02
17/07/2023	118219	7	117946,25	272,75	0,174	117460,56
18/07/2023	117841	8	117935,29	94,29	0,053	117269,21
19/07/2023	117552	9	117924,32	372,32	0,799	118370,75
			DESVPAD	372,84	MÉDIA	117865,443

iBovespa

DATA	FECHAMENTO	X	Y = POLI(X)	DELTA	ALEA 2	SIMUL 2
06/07/2023	117426	0	118023,00	597,00	0,692	118309,29
07/07/2023	118898	1	118012,04	885,96	0,670	118265,24
10/07/2023	117942	2	118001,07	59,07	0,387	117832,85
11/07/2023	117220	3	117990,11	770,11	0,814	118457,89
12/07/2023	117666	4	117979,14	313,14	0,715	118299,05
13/07/2023	119264	5	117968,18	1295,82	0,228	117562,44
14/07/2023	117711	6	117957,22	246,22	0,123	117394,47
17/07/2023	118219	7	117946,25	272,75	0,195	117491,83
18/07/2023	117841	8	117935,29	94,29	0,378	117753,25
19/07/2023	117552	9	117924,32	372,32	0,683	118197,65
			DESVPAD	372,84	MÉDIA	118216,315

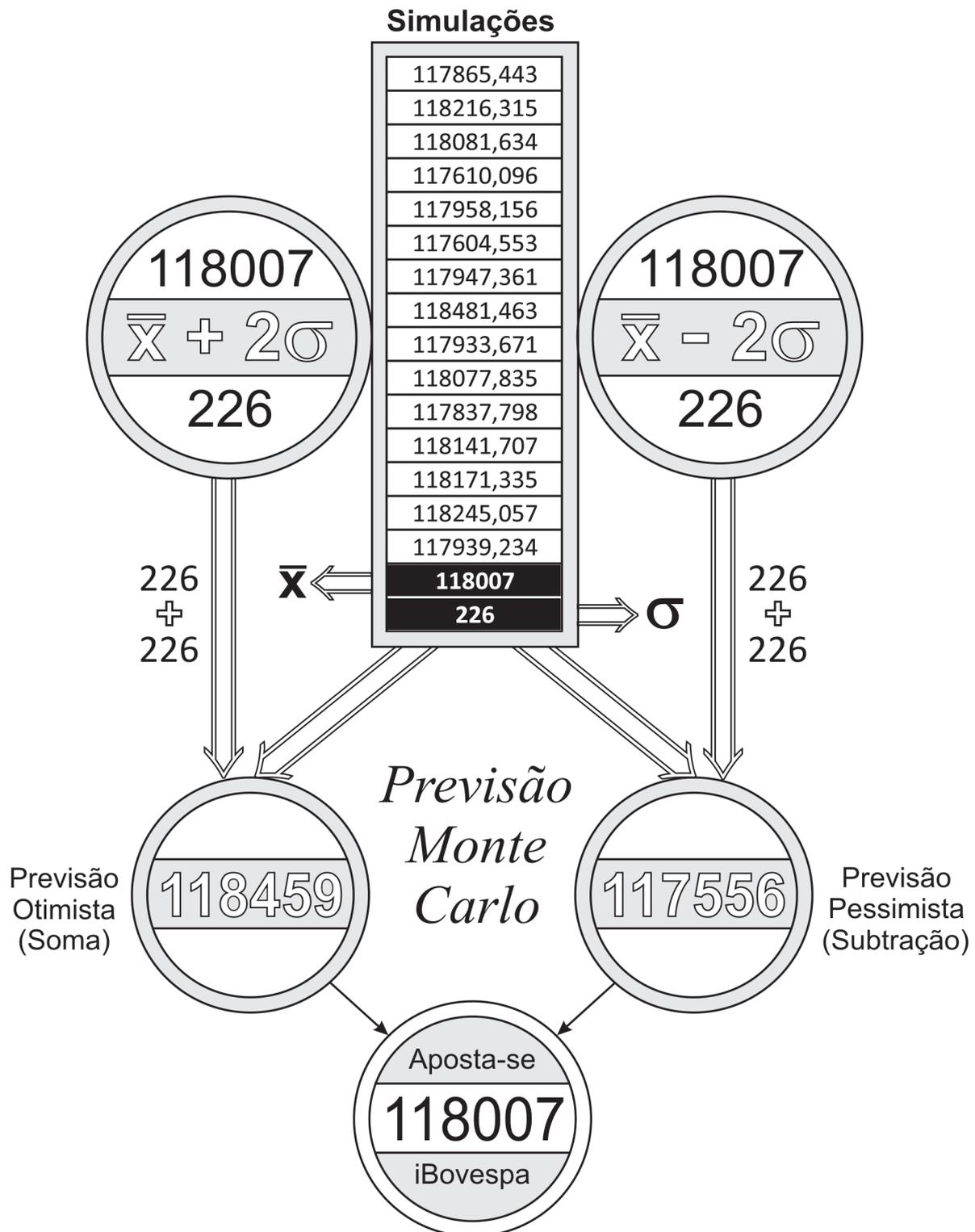
iBovespa

DATA	FECHAMENTO	X	Y = POLI(X)	DELTA	ALEA 3	SIMUL 3
06/07/2023	117426	0	118023,00	597,00	0,502	118025,59
07/07/2023	118898	1	118012,04	885,96	0,201	117566,26
10/07/2023	117942	2	118001,07	59,07	0,551	118077,57
11/07/2023	117220	3	117990,11	770,11	0,947	118657,12
12/07/2023	117666	4	117979,14	313,14	0,221	117562,94
13/07/2023	119264	5	117968,18	1295,82	0,135	117423,27
14/07/2023	117711	6	117957,22	246,22	0,517	117982,66
17/07/2023	118219	7	117946,25	272,75	0,154	117429,50
18/07/2023	117841	8	117935,29	94,29	0,815	118404,34
19/07/2023	117552	9	117924,32	372,32	0,163	117421,64
			DESVPAD	372,84	MÉDIA	118081,634



3.0 Coleção das Simulações de Monte Carlo

Dos resultados simulados obtêm-se uma aposta construída na forma:



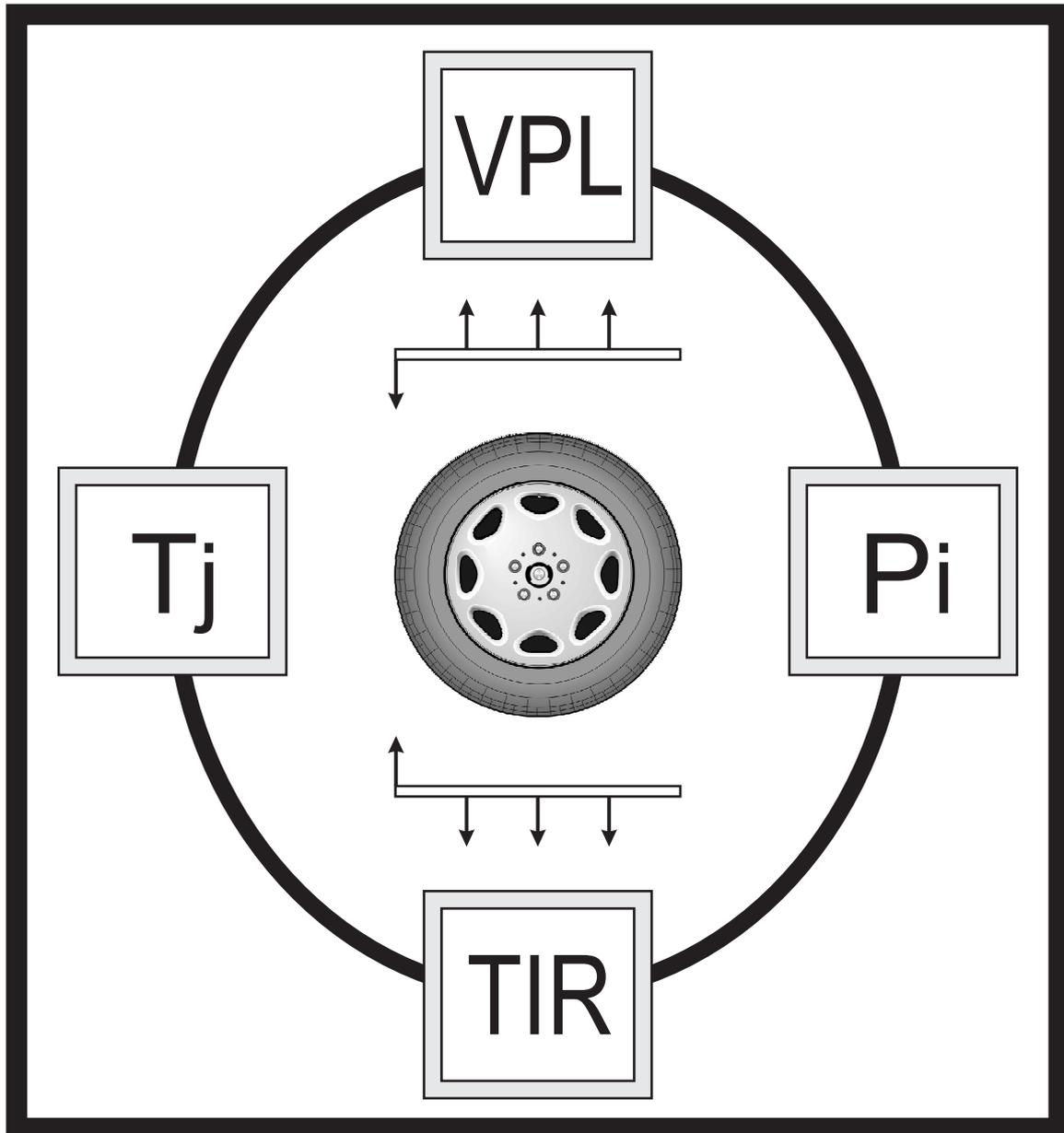
DATA	FECHAMENTO	X	Y = POLI(X)	DELTA	ALEA	SIMULAÇÃO
20/07/2023	118083					118007

Monte Carlo do Fechamento do iBovespa com ≠ 76 Pontos



PNEUMA

Borracha Cheia de Ar



Recauchutagem de Pneus

Viabilidade Comercial

Prof. Aguinaldo Prandini Ricieri

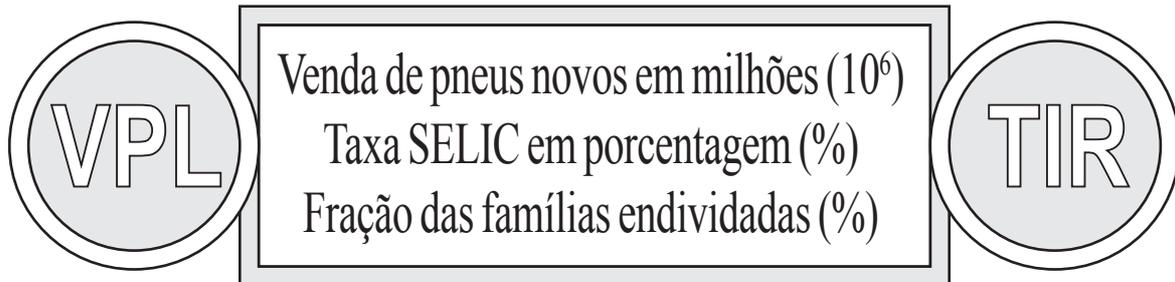


PNEUMA

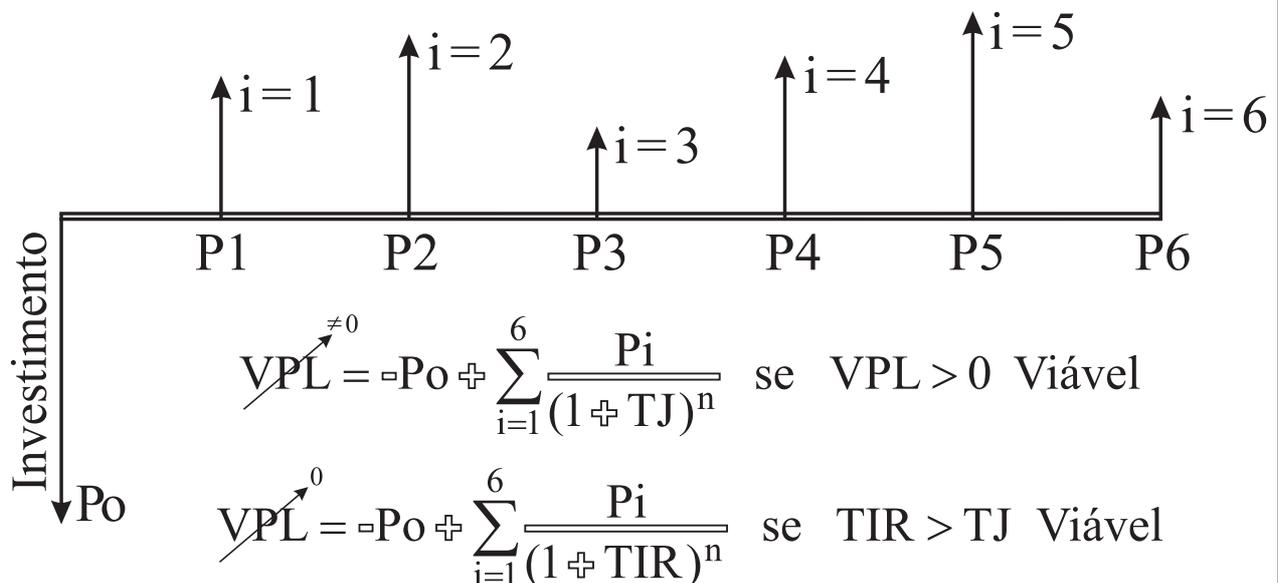
1.0 Estudo de Viabilidade Comercial da Empresa

Em 23 de dezembro de 2010, foi regulamentada a lei nº 12.305, referente a PNRS (Política Nacional de Resíduos Sólidos), que institui normas para o descarte de resíduos, especificamente legisla sobre uma possível reciclagem de pneus.

Assim, vislumbra-se o lançamento de uma empresa (PNEUMA); doze filiais instaladas em oito cidades do estado de São Paulo (SP), o que exige um estudo de viabilidade desse empreendimento com base nos três registros estatísticos:



Registros que ajudarão a formatar análise do VPL (Valor Presente Líquido) e análise do TIR (Taxa Interna de Retorno), segundo o esquema de fluxos de caixa simulados para seis semestres (P_i) anteriores ao início do ano base de 2021.

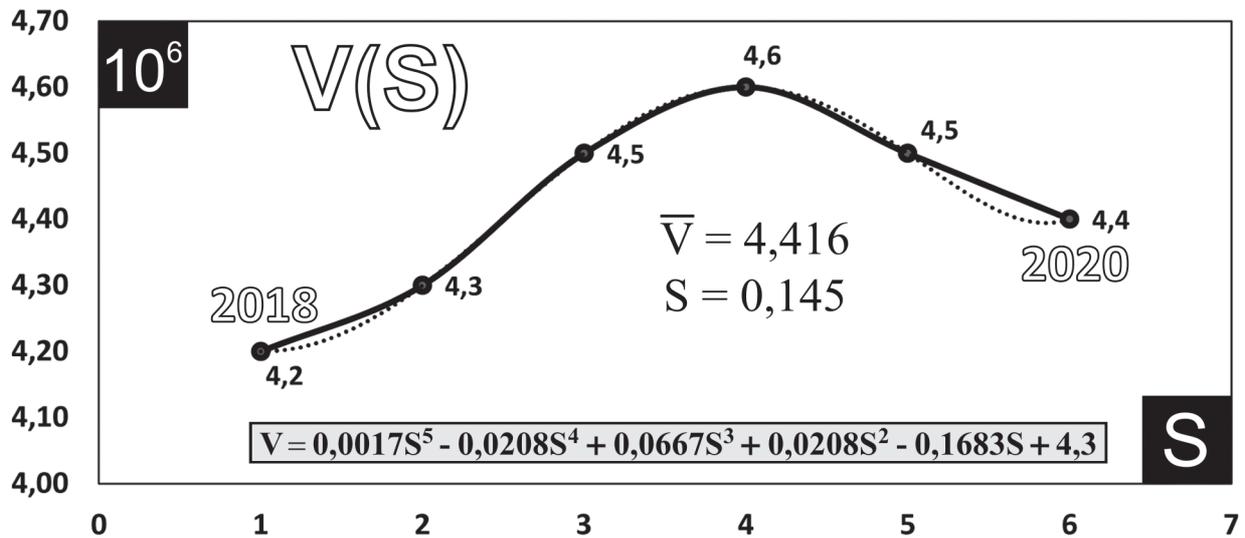


Fluxo de Caixa Simulador da Empresa Pneuma (P_i) para Seis Semestres

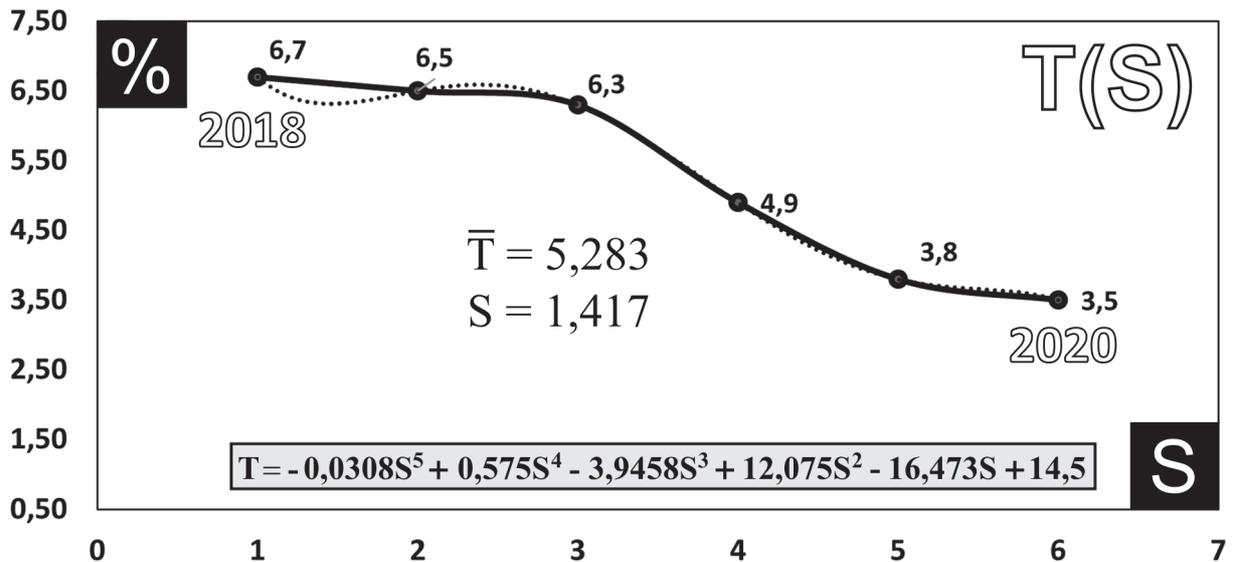


1.1 Estatísticas Referentes aos 6 Semestres das Estimativas

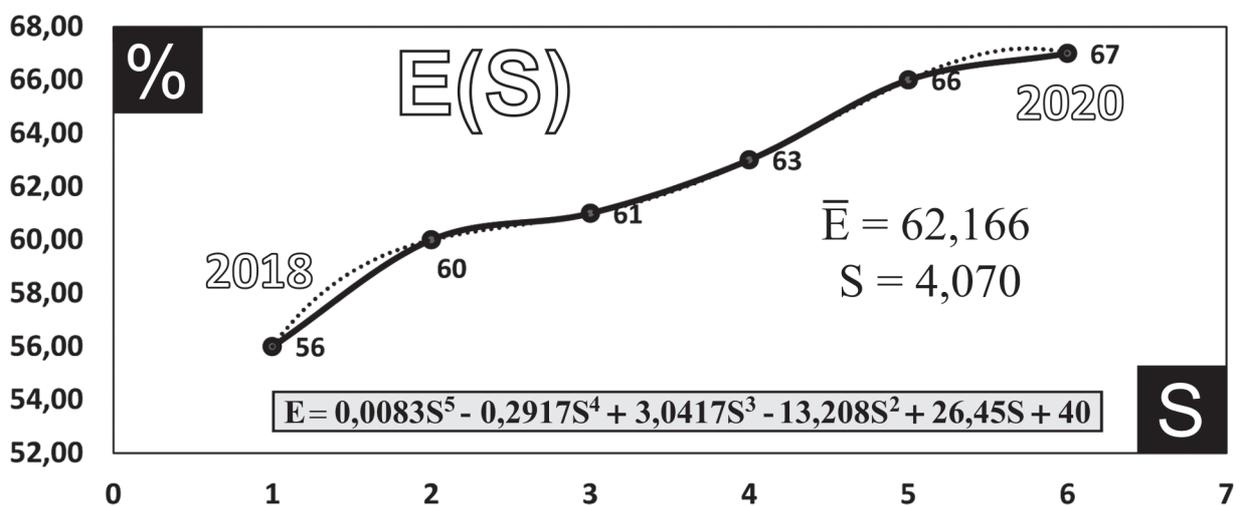
Vendas de Pneus (10⁶): 2018 (1 e 2), 2019 (3 e 4), 2020 (5 e 6)



Taxa SELIC (%): 2018 (1 e 2), 2019 (3 e 4), 2020 (5 e 6)



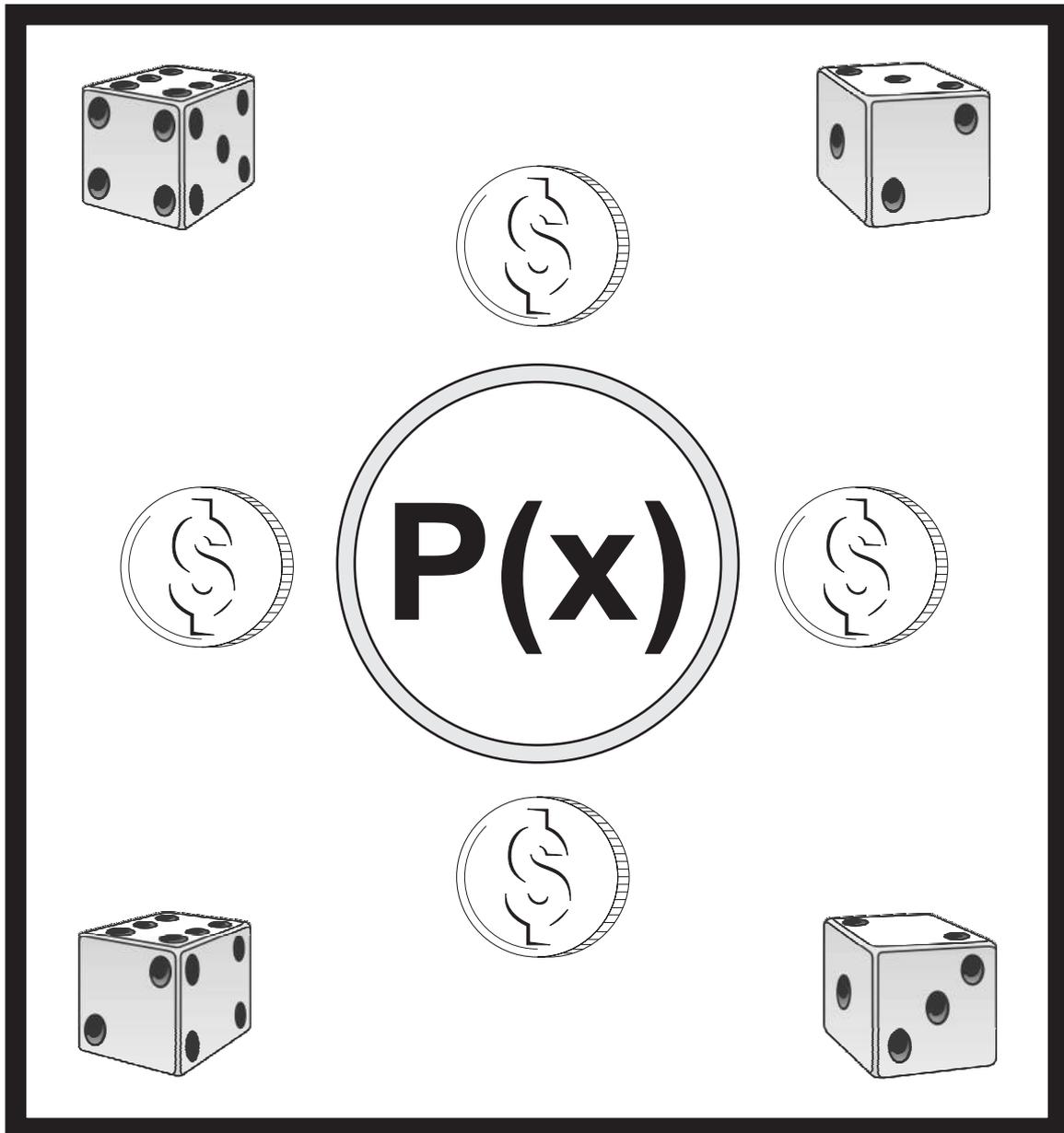
Famílias Endividadas (%): 2018 (1 e 2), 2019 (3 e 4), 2020 (5 e 6)





BINOMIAL

Sucesso **X** Fracasso



Eventos (**in**)Dependentes

Distribuição de Probabilidade

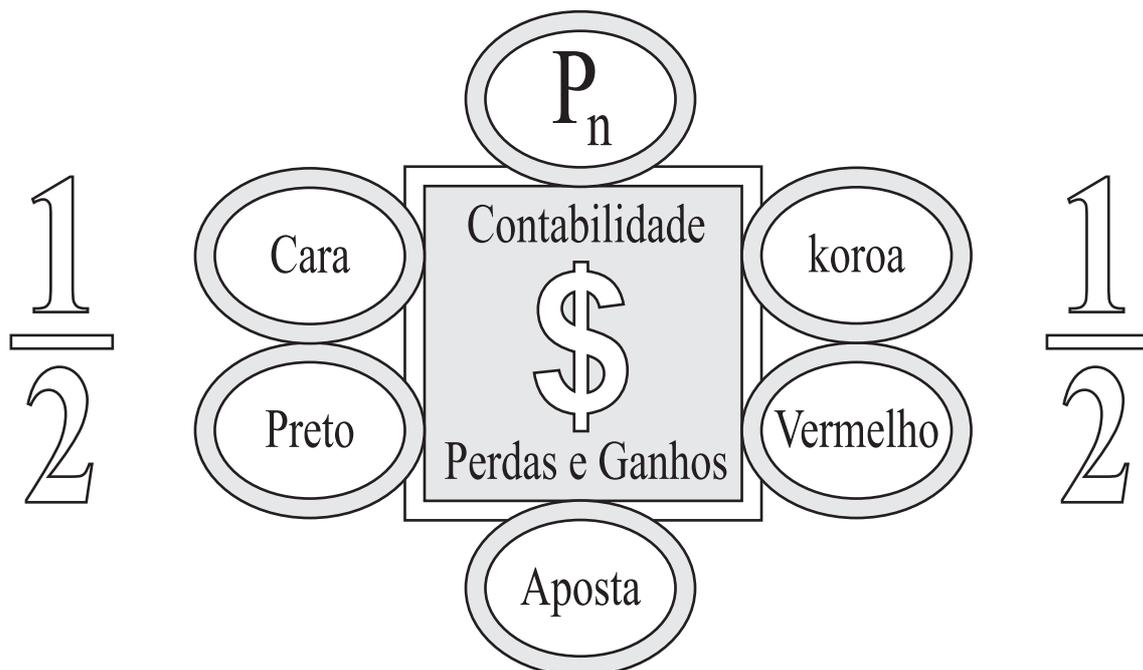
Prof. Aguinaldo Prandini Ricieri



6.0 Martingale como Estratégia nos Jogos

Martingale (conjunto de tiras de pano dobradas e fixadas uma sobre a outra), como teoria de jogos, foi trabalhada por Paul Lévy (1886 - 1971) no ano de 1934 com base nos trabalhos dos estatísticos Emile Borel (1871 - 1956), Serguey Berstein (1880 - 1968) e Andrei Kalmogorov (1903 - 1987). Enquanto o rótulo martingale propriamente dito foi dado no ano de 1939 por Jean André Ville (1910 - 1989) entendendo que “*um evento estatístico aleatório não ocorre para sempre*”. Isto é: o valor esperado da próxima ocorrência de um evento aleatório, dado todas as ocorrências anteriores verificadas, será igual (\equiv) à mais recente ocorrência, pois o valor da ocorrência ($n + 1$) não é afetado pela ocorrência (n).

É importante salientar que na sequência das jogadas de uma moeda honesta $P(p \in (0,1))$, os possíveis resultados (cara ou koroa) formam um passeio aleatório.



Apostas Martingale: c ou k, p ou v

Pensar um jogo onde a probabilidade de ganhar é $p = 1/2$, e que o jogador invista 1 R real na primeira tentativa, 2 R reais na segunda, 4 R na terceira, 8 R na quarta e assim até a (n) é-sima tentativa, sempre perdendo, e que na ($n + 1$) é-sima tentativa tenha sorte e ganhe, representa uma estratégia de risco martingale (M):

1 R \Rightarrow 2 R \Rightarrow 4 R \Rightarrow 8 R \Rightarrow 16 R \Rightarrow 32 R...



Markov e Chebyschev

Desigualdades

$$P[x \geq C] \leq \frac{E(x)}{C}$$

 σ X μ

$$P[|x - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

Limitantes Casuais

Estatísticas de Aproximações

Prof. Aguinaldo Prandini Ricieri

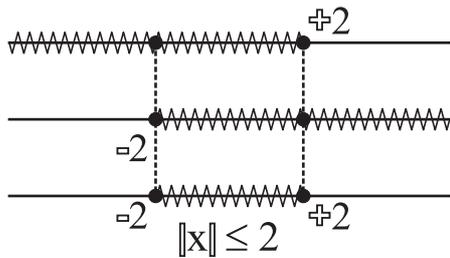


1.1 Módulos e Suas Operações Demarcatórias

$$|x| \leq 2$$

Se $x > 0 \Rightarrow +x \leq 2 \therefore x \leq 2$

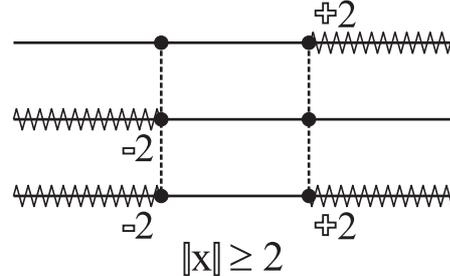
Se $x < 0 \Rightarrow -x \leq 2 \therefore x \geq -2$



$$|x| \geq 2$$

Se $x > 0 \Rightarrow +x \geq 2 \therefore x \geq 2$

Se $x < 0 \Rightarrow -x \geq 2 \therefore x \leq -2$



$$P[|x - \mu| \geq k\sigma] \leq \frac{1}{k^2}$$

Se $(x - \mu) > 0$

$+ (x - \mu) \geq k\sigma$

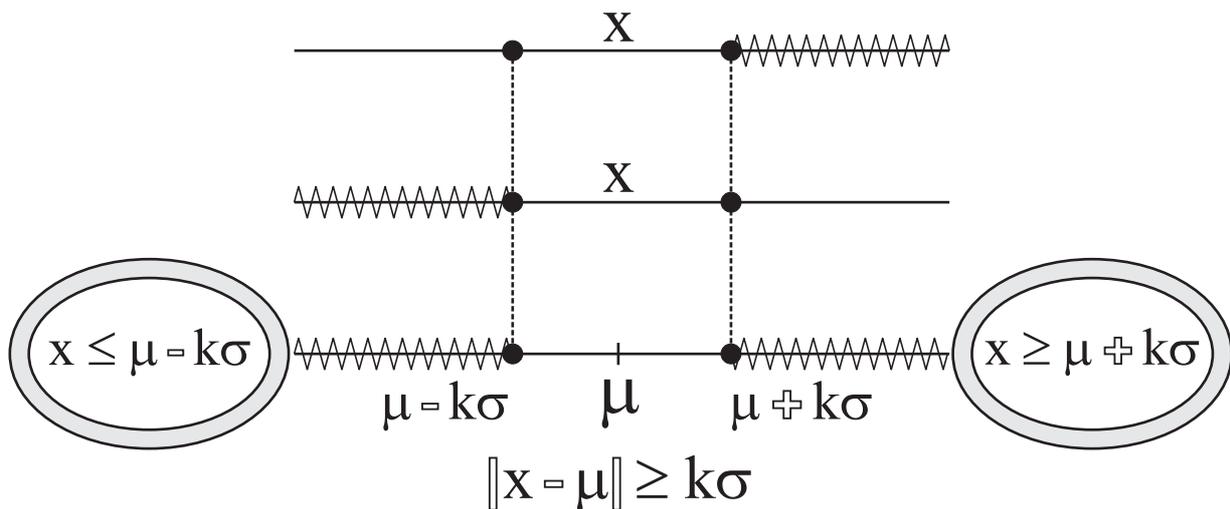
Então $x \geq \mu + k\sigma$



Se $(x - \mu) < 0$

$- (x - \mu) \geq k\sigma$

Então $x \leq \mu - k\sigma$



As regiões hachuradas têm uma probabilidade de ocorrer segundo a estatística de Chebyshev menor ou igual a $1/2k^2$

Demarcando Espaços de Domínios



5.1 Aplicação de Chebyshev no Lançamento de Moedas

Definida uma moeda honesta, pede-se o número de lançamentos que justifique 99% de certeza das percentualizações de cara e de koroa serem iguais $p = (1 - p) = 0,50$; com uma suposta margem de erro de $\epsilon = 0,01$.

P

$$P[|x - p| < \epsilon] > 1 - \frac{p \cdot (1 - p)}{n \cdot \epsilon^2}$$

$$P[|x - p| < \epsilon] > 1 - \frac{0,5 \cdot (1 - 0,5)}{n \cdot (0,01)^2} = 0,99$$

$$n = 250000 \text{ lançamentos da moeda}$$

n

Roteiro do Cálculo Estatístico de CHE e o Valor de n

5.2 Aplicação de Chebyshev nas Pesquisas Eleitorais

Certa pesquisa eleitoral feita com 2000 eleitores paranaenses revelou que entre os candidatos ao governo do estado A e B, 80% preferem serem governados pelo candidato A. Para que esse dado seja divulgado ao povo pede-se a margem de erro dessa pesquisa com 95% de confiança.

P

$$P[|x - p| < \epsilon] > 1 - \frac{p \cdot (1 - p)}{n \cdot \epsilon^2}$$

$$P[|x - p| < \epsilon] > 1 - \frac{0,8 \cdot (1 - 0,8)}{2000 \cdot \epsilon^2} = 0,95$$

$$\epsilon = 0,04 = 4\% \text{ de erro na pesquisa}$$

€

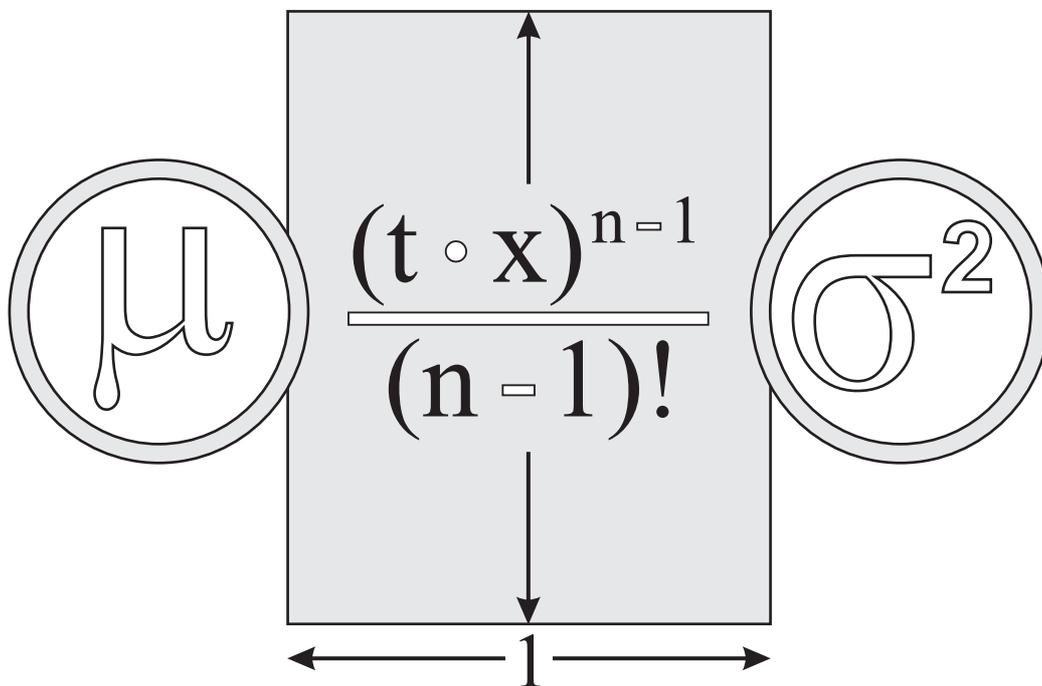
Roteiro do Cálculo Estatístico de CHE e o Valor de €



MOMENTOS

Estatística Gerativa

$$P(tx) = 1 + \frac{t^1 x^1}{1!} + \frac{t^2 x^2}{2!} + \frac{t^3 x^3}{3!} + \dots$$



$$E\left(\frac{\partial M}{\partial t}\right)$$

$$E\left(\frac{\partial^2 M}{\partial^2 t^2}\right)$$

Média e Variância

Ajuste feito por Karl Pearson

Prof. Aguinaldo Prandini Ricieri



[385]

prop. 1. the probability that some way or other it happens p times and fails q times in $p + q$ trials is $E a^p b^q$.

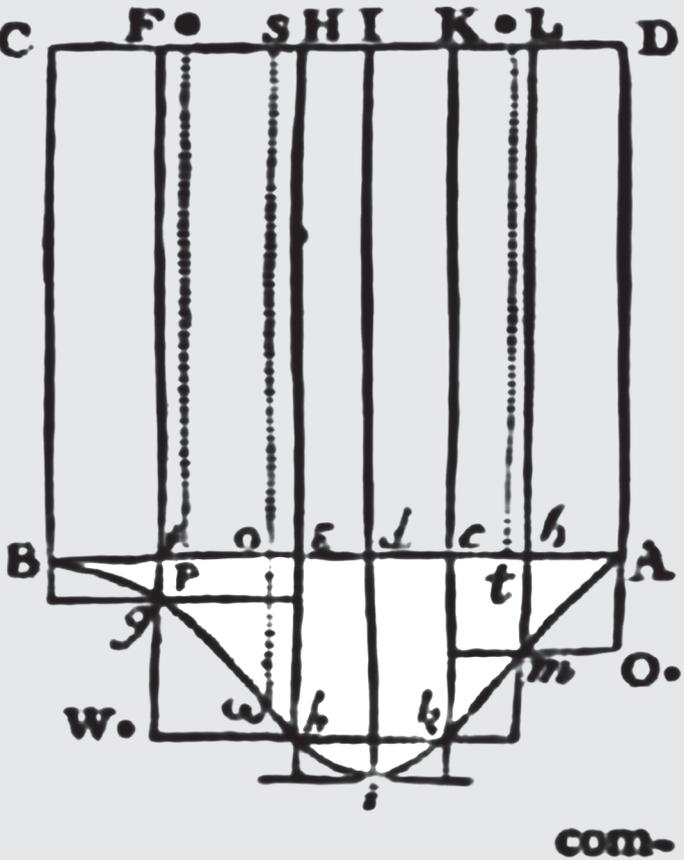
SECTION II.

Postulate. 1. I Suppose the square table or plane $ABCD$ to be so made and levelled, that if either of the balls o or W be thrown upon it, there shall be the same probability that it rests upon any one equal part of the plane as another, and that it must necessarily rest somewhere upon it.

2. I suppose that the ball W shall be 1st thrown, and through the point where it rests a line os shall be drawn parallel to AD , and meeting CD and AB in s and o ; and that afterwards the ball O shall be thrown $p + q$ or n times, and that its resting between AD and os after a single throw be called the happening of the event M in a single trial. These things supposed,

Lem. 1. The probability that the point o will fall between any two points in the line AB is the ratio of the distance between the two points to the whole line AB .

Let any two points be named, as f and b in the line AB , and through them parallel to AD draw fF , bL meeting CD in F and L . Then if the rectangles Cf , Fb , LA are



Philosophical Transactions, p. 385, 1764, Thomas Bayes (1702 - 1761).



1.3 Aplicação do Operador Esperança E

Para a função distribuição de probabilidades $P(x)$, o valor esperado de uma função $f(x)$ – sendo x a sua variável aleatória de controle de entrada –, é dada por $E[f(x)]$.



$$E[f(x)] = \int_{-\infty}^{+\infty} f(x) \circ P(x) dx$$



$$E[x] = \int_{-\infty}^{+\infty} x \circ P(x) dx = \mu$$



$$E[x - \mu]^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 \circ P(x) dx = \sigma^2$$



Para b, c constantes resultam:

$$E(c) = c$$

$$E(c + x) = c + E(x)$$

$$E(cx) = c \circ E(x)$$

$$E(b + cx) = b + c \circ E(x)$$

$$E(x \circ y) = E(x) \circ E(y)$$

Operador Esperança $E()$ Aplicado

$E(t)$: Espera-se por Um Certo Valor t



3.0 Esperança Normal da Potência (x^1)

Momento de ordem 1 da variável aleatória x^1 , operação de valor esperado $E(x^1)$, em termos dos dois parâmetros estatísticos de ajuste: μ e σ . Observe o resultado:

$$E(x^1) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \int_{-\infty}^{+\infty} x^1 \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \cdot dx$$

Distribuição de Probabilidade para x^1 e seu Momento $M(x^1)$

$$Z = \frac{x - \mu}{\sigma} \quad \therefore \quad x = \sigma \cdot z + \mu \quad \therefore \quad dx = \sigma dz$$

$$E(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \int_{-\infty}^{+\infty} (\sigma z + \mu)^1 \cdot e^{-\frac{z^2}{2}} \cdot \sigma dz$$

$$E(x) = \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{+\infty} \sigma z \cdot e^{-\frac{z^2}{2}} dz + \int_{-\infty}^{+\infty} \mu \cdot e^{-\frac{z^2}{2}} dz \right]$$

$$E(x) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \cdot e^{-\frac{z^2}{2}} dz + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz$$

$$E(x) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z \cdot e^{-\frac{z^2}{2}} dz + \mu \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} dz$$

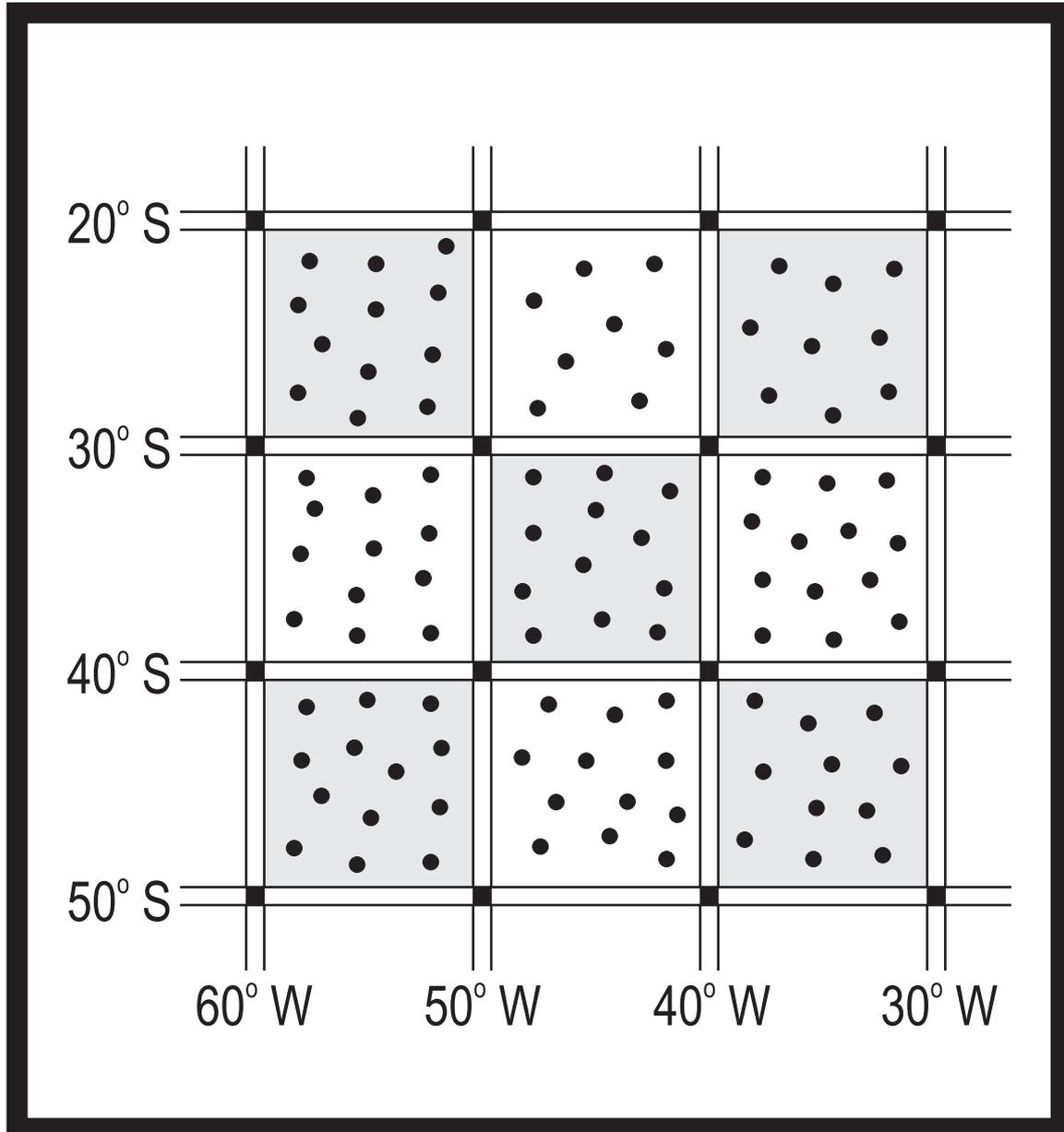
$$E(x) = \mu$$

Espera-se da Variável x^1 o Valor $E(x)$



KRIGAGEM

Estatística Espacial



Fenômenos Geoestatísticos

Inverse Distance Weighted

Prof. Aguinaldo Prandini Ricieri

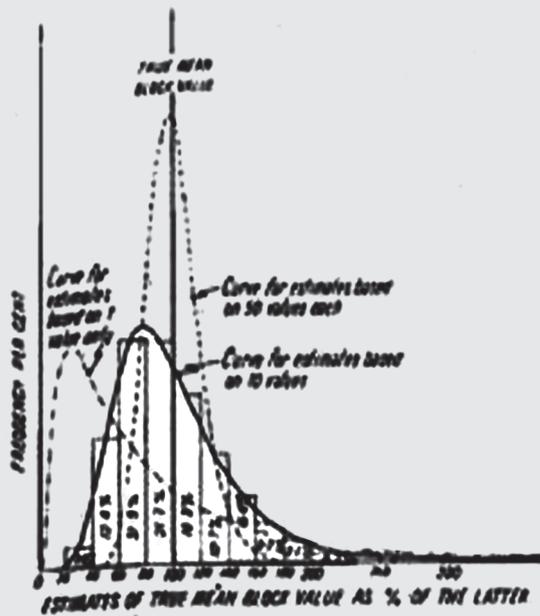


Diagram No. 3—Showing the frequency distributions of estimates of the true mean value of an average ore reserve block based on 1, 10 and 50 values per estimates respectively

represented by a smooth frequency curve. Statistical theory provides the method of determining the distribution curve of the estimates (to a first approximation in the case of a lognormal parent distribution).*

In the case of the so-called average ore reserve block considered above, the distribution of the estimates of the true mean value of the block if based on ten values in each case only, can be represented by the full curve on Diagram No. 3.† The curve for individual values (from Diagram No. 2)

* See Annexure and Ref. 5.

is shown dotted and it is immediately evident how the estimates based on sets of ten values each are clustered more closely on either side of the overall true mean block value.

From Diagram No. 3 it is evident that if a large number of sets of ten samples each are taken from this particular block of ore, the chances are that, e.g. 70.6 per cent of all the observed mean values of these sets of ten sample values each will be within the range of values from 60 per cent to 140 per cent of the true mean value of the block. There is therefore approximately a seven in ten chance of not exceeding an error of 40 per cent if the arithmetic mean of ten sample values is accepted as the true mean value of the block.

Similarly, the distribution curve for estimates based on any specific number of available sample values can be obtained, and Fig. 3 shows the curve (dotted) for sets of 50 values each. Based on a series of such curves, Table 1 was prepared to indicate the overall reliability of block valuation in the case of an average ore reserve block on the Rand.

It becomes clear from this table what 'chances' are taken by the valuator in the valuation of individual blocks and stope faces where conditions approach the

† Based on random sampling theory. For systematic sampling as practised on the Witwatersrand the curve will probably be slightly different, an aspect receiving attention at present.

TABLE 1

No. of samples per set	Maximum error either side of true mean value of block				
	10 per cent	20 per cent	30 per cent	40 per cent	50 per cent
	Probability percent				
1	7	14	22	30	39
5	14	28	42	56	68
10	19	38	55	71	82
20	27	51	71	85	92
50	41	72	90	97	99
100	55	87	97	99	100
200	71	97	100	100	100
500	91	100	100	100	100

E.g., there will be a 51 per cent probability, say, an even chance, of the observed mean of 20 sample values not being in error by more than 20 per cent either side of the true mean value.

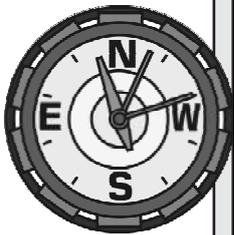


Estatística Espacial

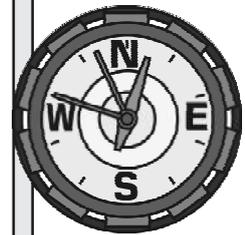
1.0 Fenômenos Distribuídos nas Geografias

A Krigagem ou Estatística Espacial se ocupa em coletar e analisar conjuntos de dados localizados/distribuídos em uma geografia fenomenológica – locus de cunho administrativo, epidemiológico, mineralógico, policial, econômico, florestal... –, com a finalidade de prever um fenômeno em certos pontos dessa mesma geografia. O princípio heurístico da Krigagem é que os locus avizinados tendem a ter valores do fenômeno expostos mais parecidos do que os locus distantes.

Como exemplos de ocorrências pertinentes à Estatística Espacial considere:



- a) Repartições de resíduos tóxicos nos aterros.
- b) Número mensal de assaltos em locus de cidades.
- c) Espalhamento do ouro nos solos aluviais.
- d) Distribuição de dengosos sobre uma região.
- e) Localização de mortos nas tragédias naturais.
- f) Locus de shoppings nas geografias de CPF's.
- g) Adensamento de pragas/insetos nas plantações.



O método geoestatístico de Krigagem, para estimar o valor de um fenômeno regional, requer que seja feito o semivariograma estatístico dos locus formadores da base de dados, que ali inicialmente foram coletados. Esse procedimento permite conhecer a homo(hetero)geneidade da variação do fenômeno analisado nas direções geográficas horizontais, verticais e inclinadas (45° , 135° , 225° e 315°).

Krigagem

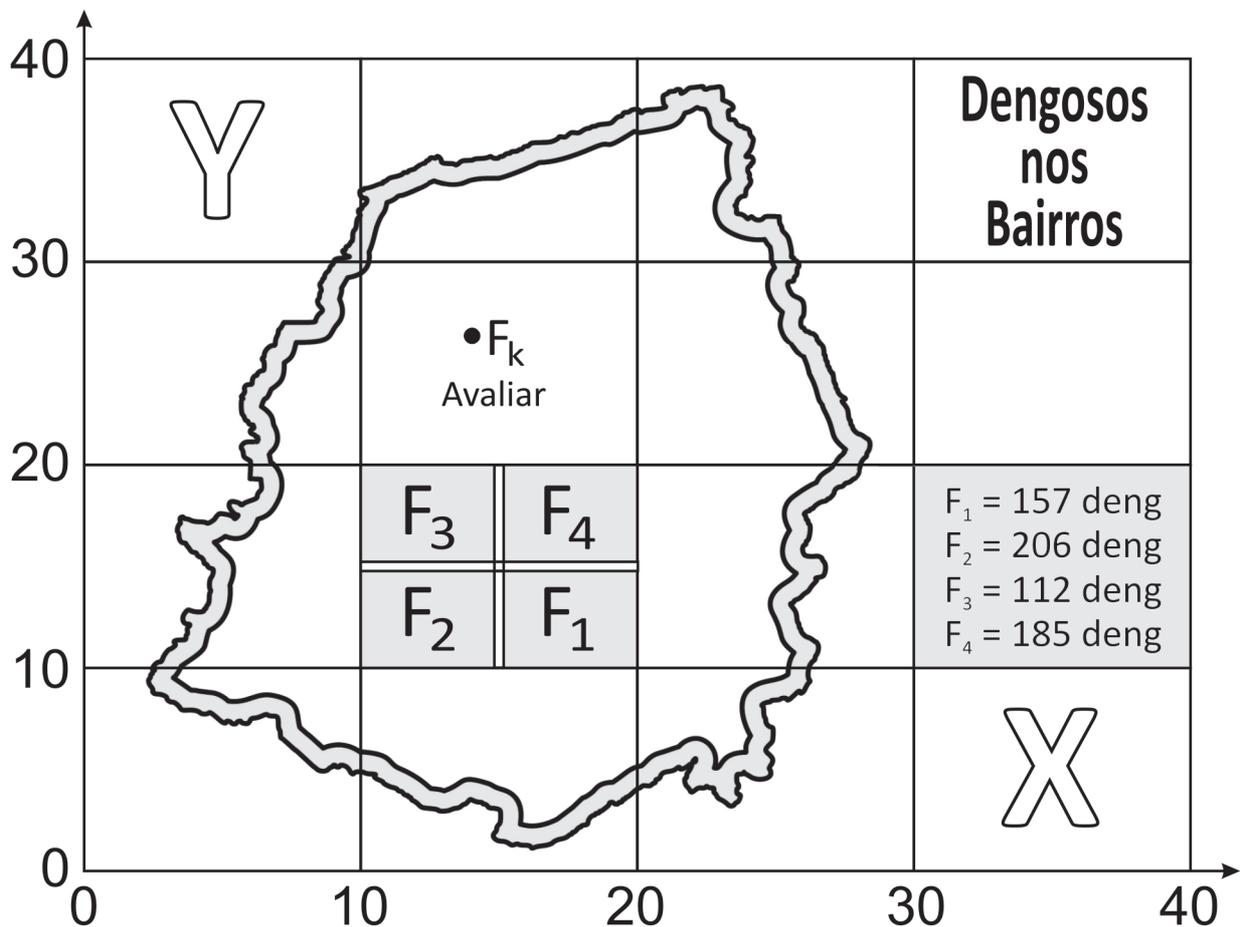
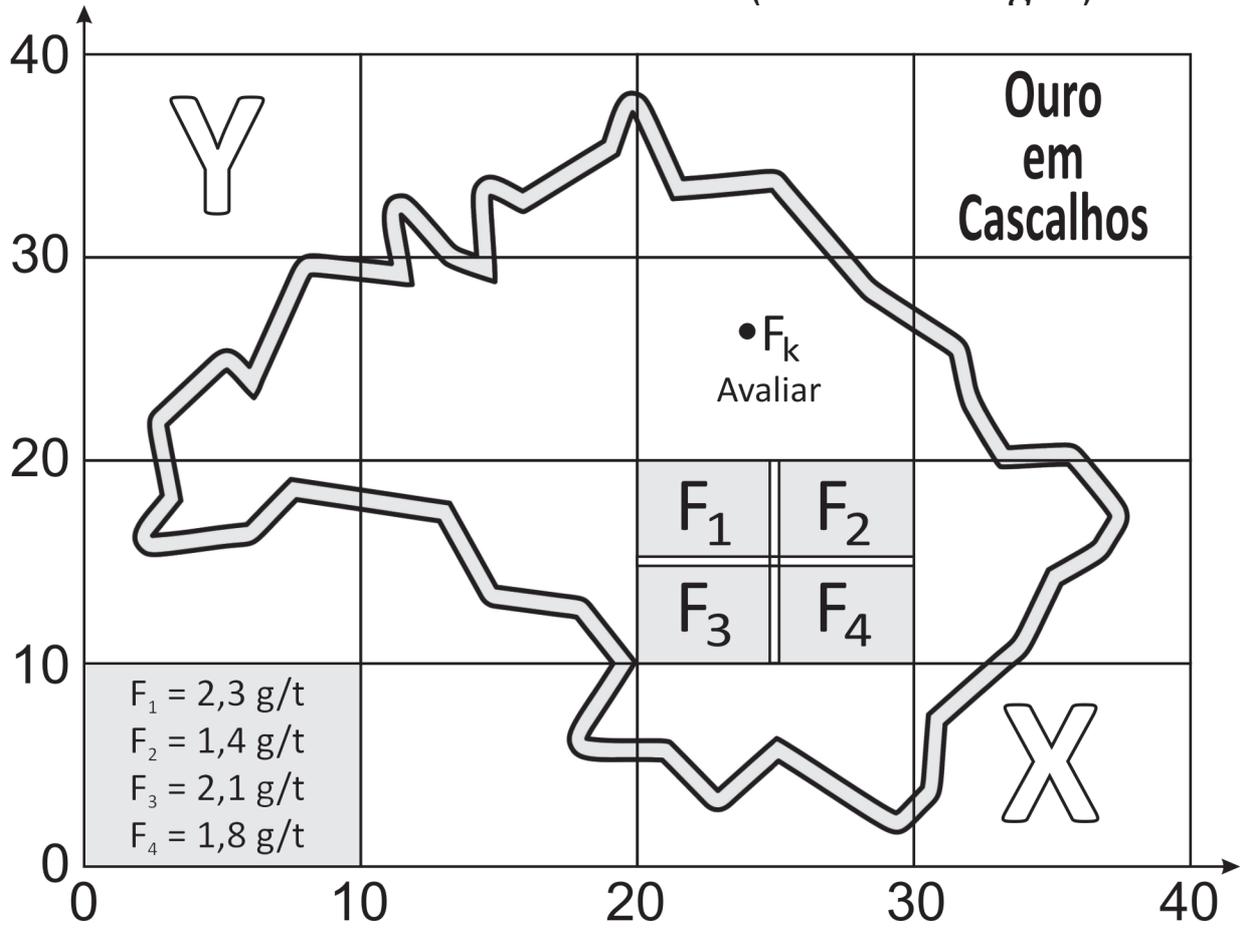
A Statistical Approach to Some Basic Mine Valuation Problem on the Witwatersrand

Daniel Gerhardus Krige (1919 - 2013)

Journal of the Chemical of South Africa, vol. 52, nº 6, p.119, 1951



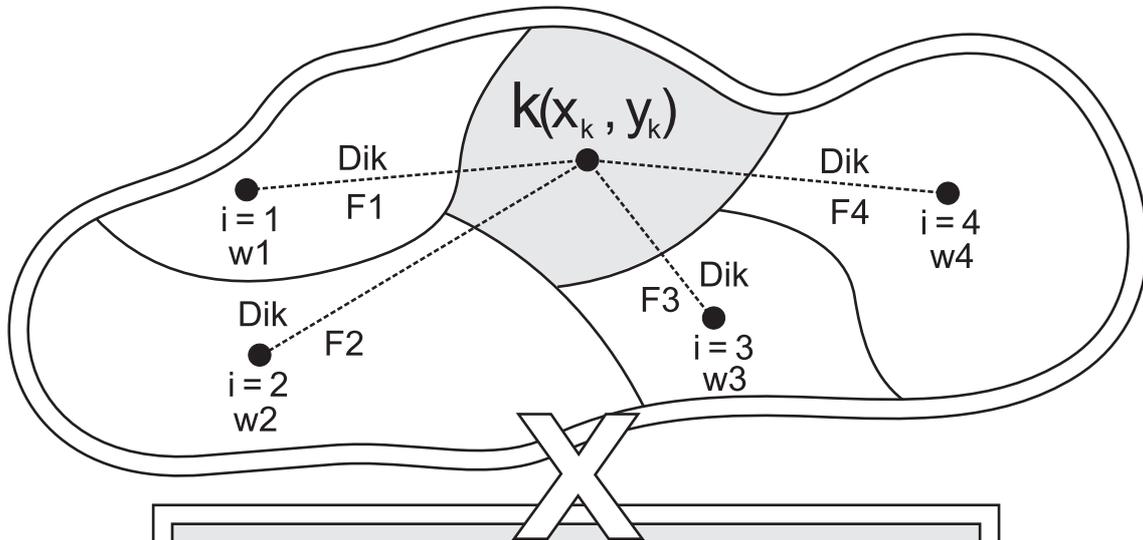
1.1 Fenômenos Distribuídos (Ouro e Dengue)





4.0 Krigagem Weighted Pontual

Dado o mapa abaixo e alguns dos seus locus (coordenadas x, y) onde foram coletadas e analisadas amostras pertinentes à ocorrência de um certo fenômeno, e seja uma coordenada $k = k(x_k, y_k)$ na qual será estimado o valor desse fenômeno:



W

- i = Pontos ou locus da geografia mensurada
- k = Pontos ou locus da geografia estimada
- D_{ik} = Distâncias dos locus (i) até o locus (k)
- W_i = Pesos dos pontos ou dos locus avaliados
- F_i = Valor do fenômeno medido nos locus (i)
- Z_k = Valor do fenômeno estimado no locus (k)
- n = Quantidade de amostras mensuradas

F

$$W_i = \frac{\frac{1}{D_{ik}}}{\sum_{i=1}^n \frac{1}{D_{ik}}} = \frac{\frac{1}{D_{ik}}}{\frac{1}{D_{1k}} + \frac{1}{D_{2k}} + \frac{1}{D_{3k}} + \dots + \frac{1}{D_{nk}}} \quad \circ \circ \quad \sum_{i=1}^n W_i = 1$$

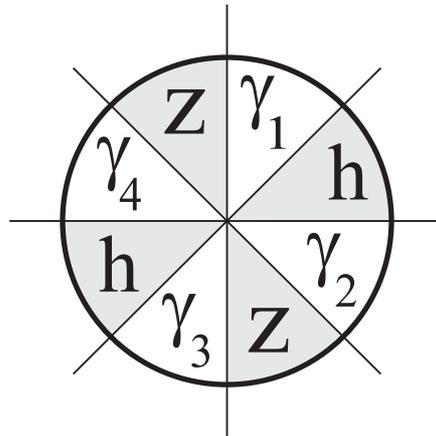
$$Z_k = \sum_{i=1}^n W_i \cdot F_i = W_1 \cdot F_1 + W_2 \cdot F_2 + \dots + W_n \cdot F_n$$

Parâmetros Estatísticos da Krigagem

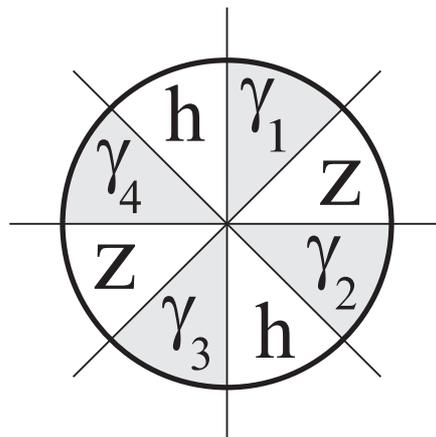


AGROKRIG

Estatística Agrográfica



$$\gamma(h) = \frac{1}{N(h)} \cdot \sum_{i=1}^{N(h)} \frac{[Z(x_i) - Z(x_i + h)]^2}{2}$$



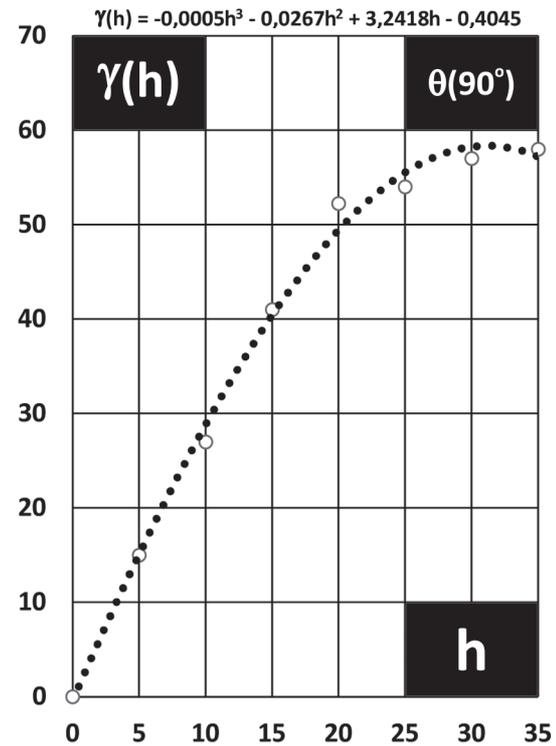
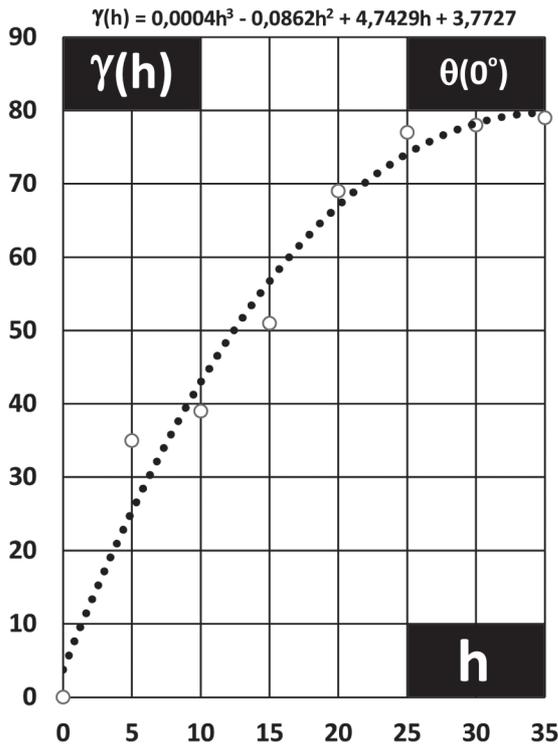
Agricultura de Precisão

Semivariograma Anisotrópico

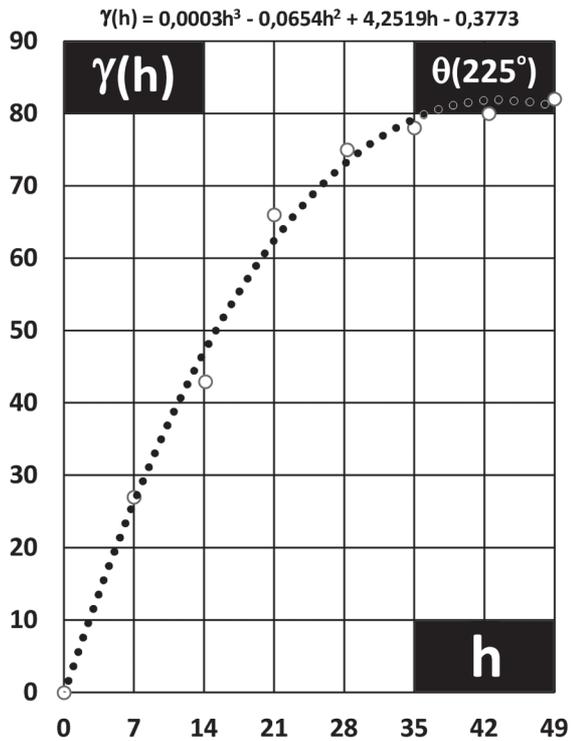
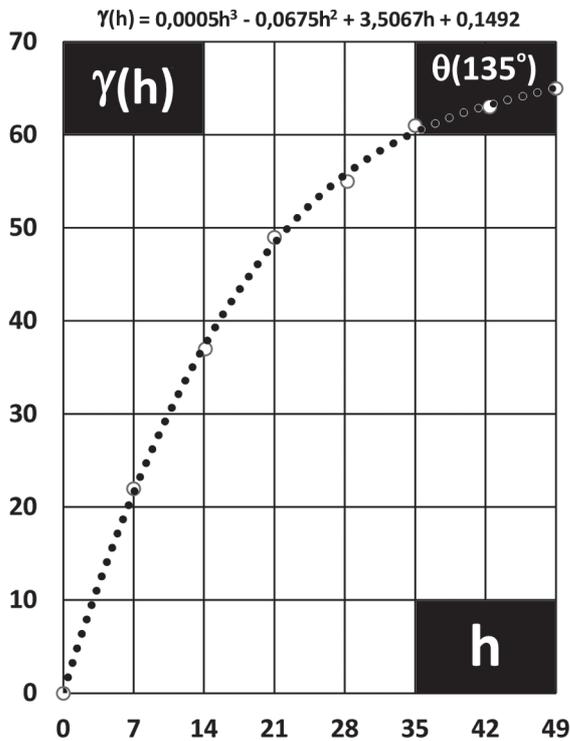
Prof. Aguinaldo Prandini Ricieri



3.6 Gráficos Melhorados das Segundas Medições dos Locus



$$\gamma(h) = \frac{1}{N(h)} \cdot \sum_{i=1}^{N(h)} \frac{[Z(x_i) - Z(x_i + h)]^2}{2}$$

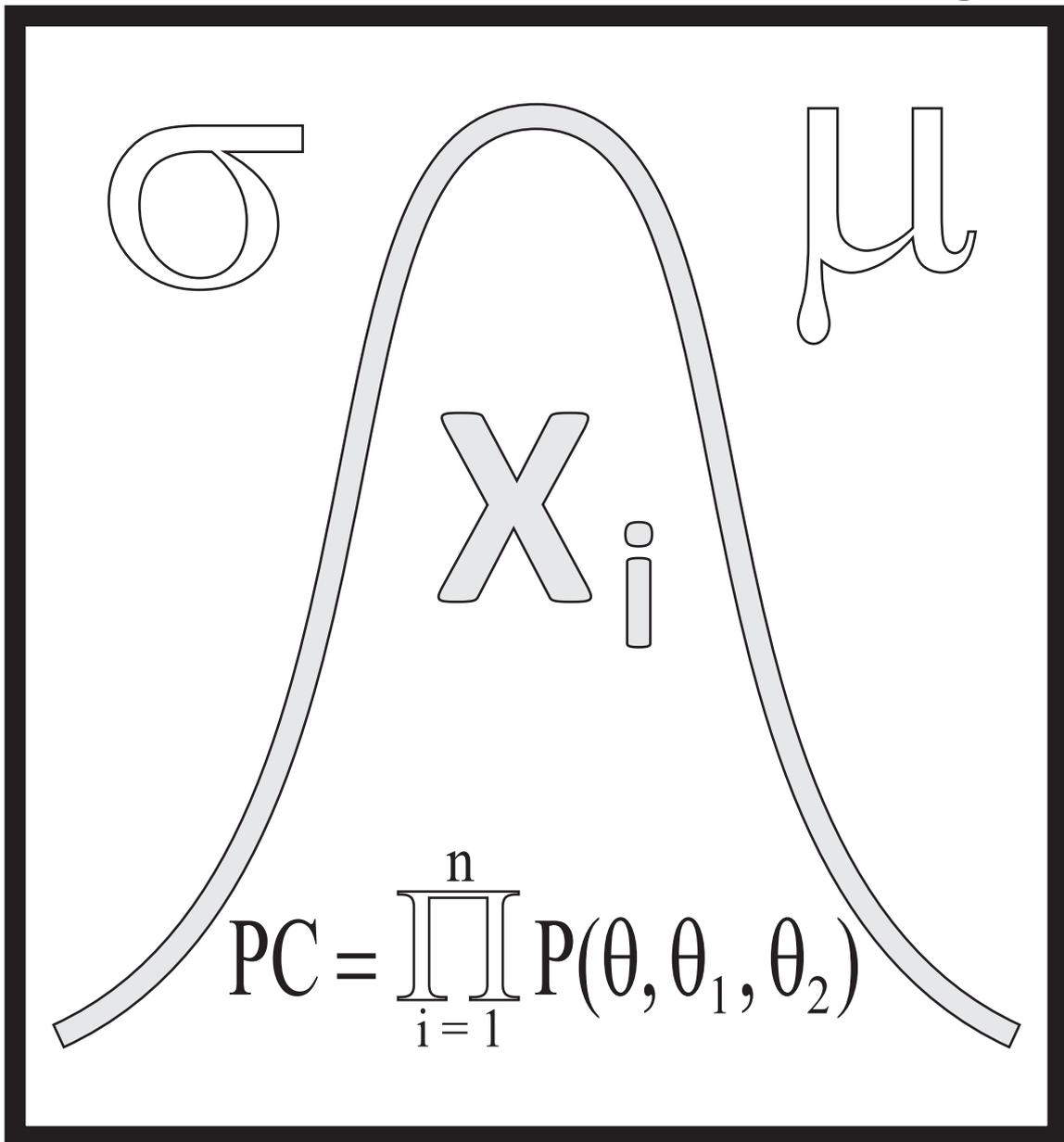


Semivariograma Anisotrópico Melhorado: Malha 5 x 5 // 7 x 7



VEROSSIMILHANÇA

Estatística de Confiança



Probabilidade Máxima

Likelihood Estimation

Prof. Aguinaldo Prandini Ricieri



4.0 Likelihood Matricial Estimation

4.1 Desenvolvimento Escalar

$$\ln PC = -\frac{n}{2} \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2$$

4.2 Desenvolvimento Pseudomatricial

$$" \ln PC " = \frac{1}{2\sigma^2} [y - x\theta] \circ [y - x\theta]$$

Para a matriz de entrada $[x]_{n,p}$ e a matriz de saída de dados $[y]_{n,k}$, qual matriz $[\theta]_{p,k}$ permite associar os valores:

$$[y] = [x] \circ [\theta] \quad \circ \circ \quad [x] \circ [\theta] = [y]$$

$$" \ln PC " = \frac{1}{2\sigma^2} [yy - 2yx\theta + \theta^2xx]$$

$$\frac{\partial}{\partial \theta} (" \ln PC ") = \frac{1}{2\sigma^2} \left(\frac{\partial}{\partial \theta} [yy - 2yx\theta + \theta^2xx] \right)$$

$$\frac{\partial}{\partial \theta} (" \ln PC ") = \frac{1}{2\sigma^2} [0 - 2yx + 2\theta xx] = 0$$

$$\theta_{xx} = yx$$

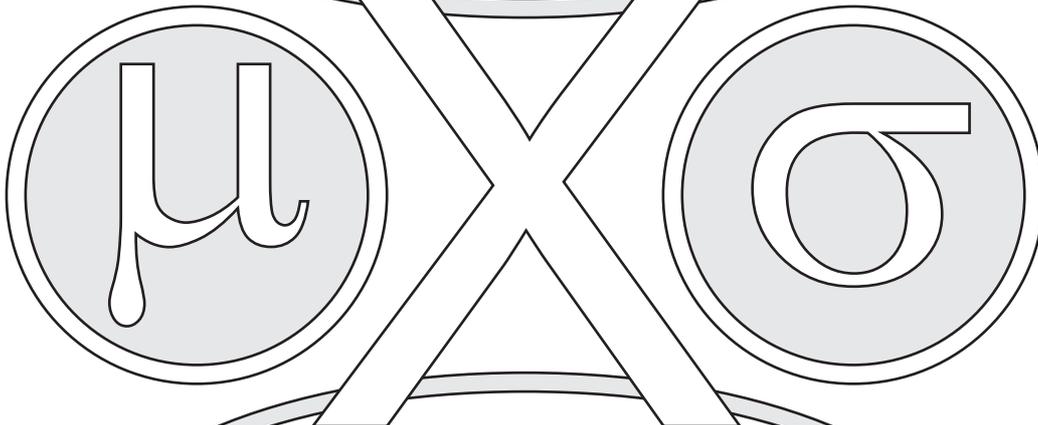




FILA

Transporte e Transbordo

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



$$E[x^n] = \int_a^b x^n \cdot P(x) dx$$

Estatística de Veículos em Fila

Descobrimos Distribuições

Prof. Aguinaldo Prandini Ricieri



4.0 Atraso de Veículos Canavieiros na Usina (2023)

Segundo os dados da primeira semana de agosto de 2023: foram 15 atrasos (F) de 3 minutos (x), 30 atrasos de 4 minutos, 54 atrasos de 5 minutos, 76 atrasos de 6 minutos, 102 atrasos de 8 minutos, ..., ... 4 atrasos de 18 minutos. Observe:

x(min)	Freq	TA	P(x)
0	0	0	0,00
1	1	1	0,001
2	5	10	0,006
3	15	45	0,019
4	30	120	0,038
5	54	270	0,069
6	76	456	0,097
7	94	658	0,120
8	102	816	0,130
9	99	891	0,126
10	90	900	0,115
11	75	825	0,096
12	54	648	0,069
13	36	468	0,046
14	24	336	0,031
15	15	225	0,019
16	8	128	0,010
17	3	51	0,004
18	4	72	0,005
171,00	785,00	6920	1,00

$P(x) = \frac{\text{Freq}}{785}$

P

$\lambda = \frac{\text{TA}}{\Sigma F}$

$P(x) = \frac{15}{785}$

X

$\lambda = \frac{6920}{785}$

Total de Atrasos (TA) e Frequência (Freq)

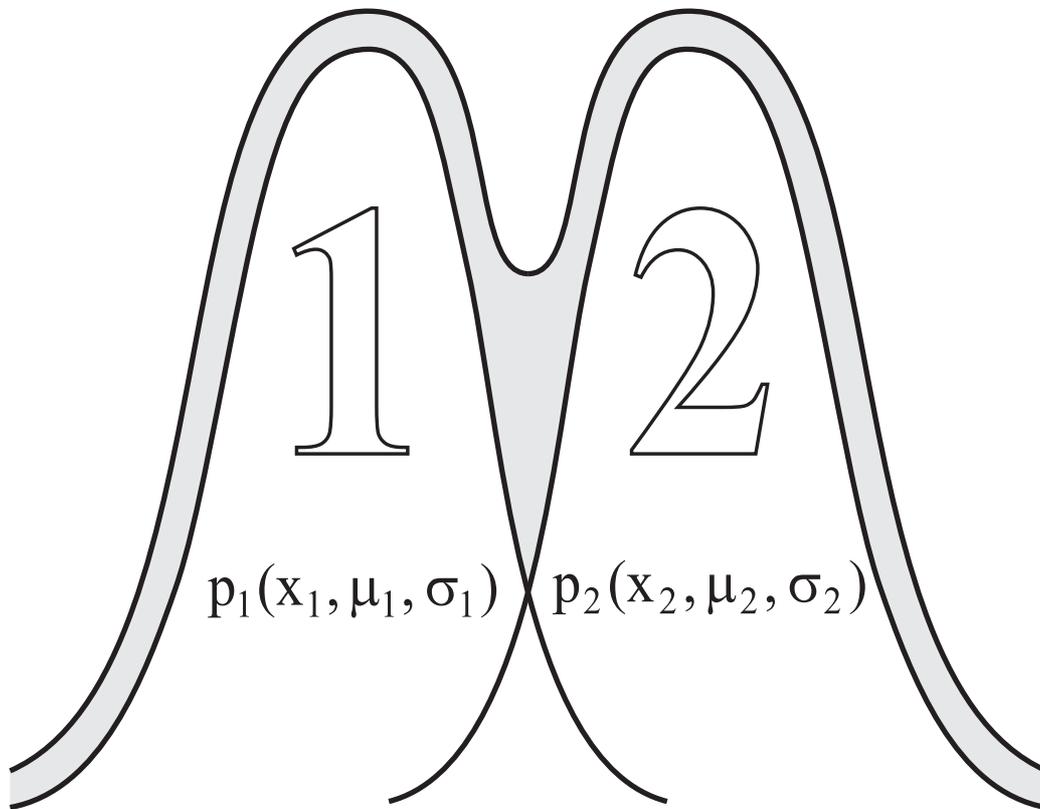


MIXTUS

Distributio $p_1(x)$ et $p_2(x)$ Distributio

$$E[x^1] = a \circ \mu_1 \oplus b \circ \mu_2$$

$$E[x^2] = a \circ (\mu_1^2 \oplus \sigma_1^2) \oplus b \circ (\mu_2^2 \oplus \sigma_2^2)$$



$$\sigma^2(x) = a\sigma_1^2 \oplus b\sigma_2^2 \oplus a\mu_1^2 \oplus b\mu_2^2 - [a\mu_1 \oplus b\mu_2]^2$$

Média e Variância

Distribuição de Probabilidade Bimodal

Prof. Aguinaldo Prandini Ricieri



Misturado

1.0 Bimodalidade Estatística de Dados

$$E[x^1] \quad \mathbf{X} \quad E[x^2]$$

Se uma amostra de dados bimodal e sua distribuição de probabilidade $p(x, \mu, \sigma)$ é composta de dados de duas populações independentes $p_1(x, \mu_1, \sigma_1)$ e $p_2(x, \mu_2, \sigma_2)$ – distribuições gaussianas unimodais –, e se não sendo possível conhecer o pertencimento de um dado x dessa amostra $p(x)$, faz-se necessário o uso dos dois momentos estatísticos ($E[x]$, $E[x^2]$) para descobrir os parâmetros (μ , σ) de ajustes dessa amostra, supostamente percentualizada: $(a \cdot p_1 + b \cdot p_2)$.

$$a + b = 1$$

$$p(x, a, b) = \frac{a}{\sigma_1 \sqrt{2\pi}} \exp\left[-\frac{(x - \mu_1)^2}{2\sigma_1^2}\right] + \frac{b}{\sigma_2 \sqrt{2\pi}} \exp\left[-\frac{(x - \mu_2)^2}{2\sigma_2^2}\right]$$

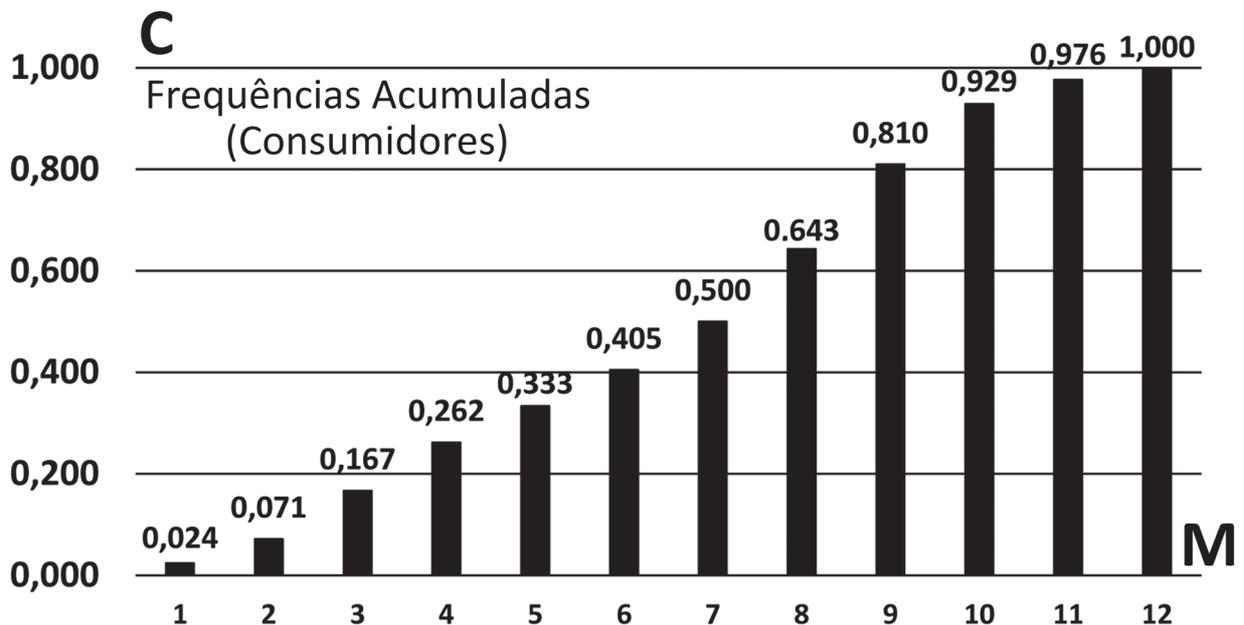
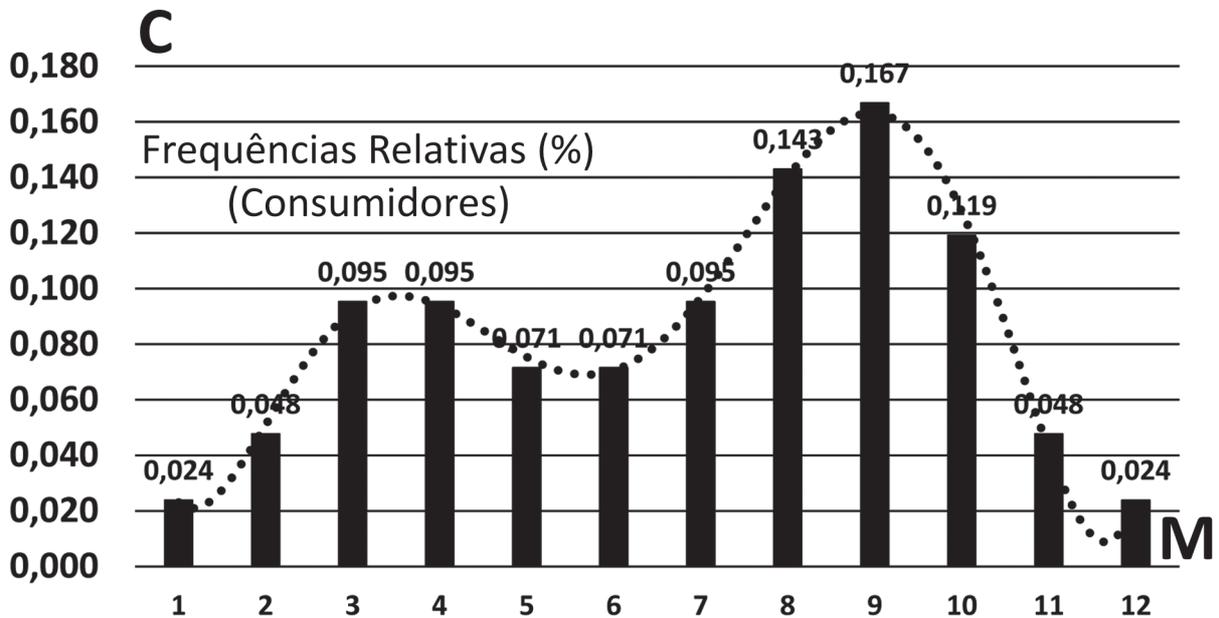
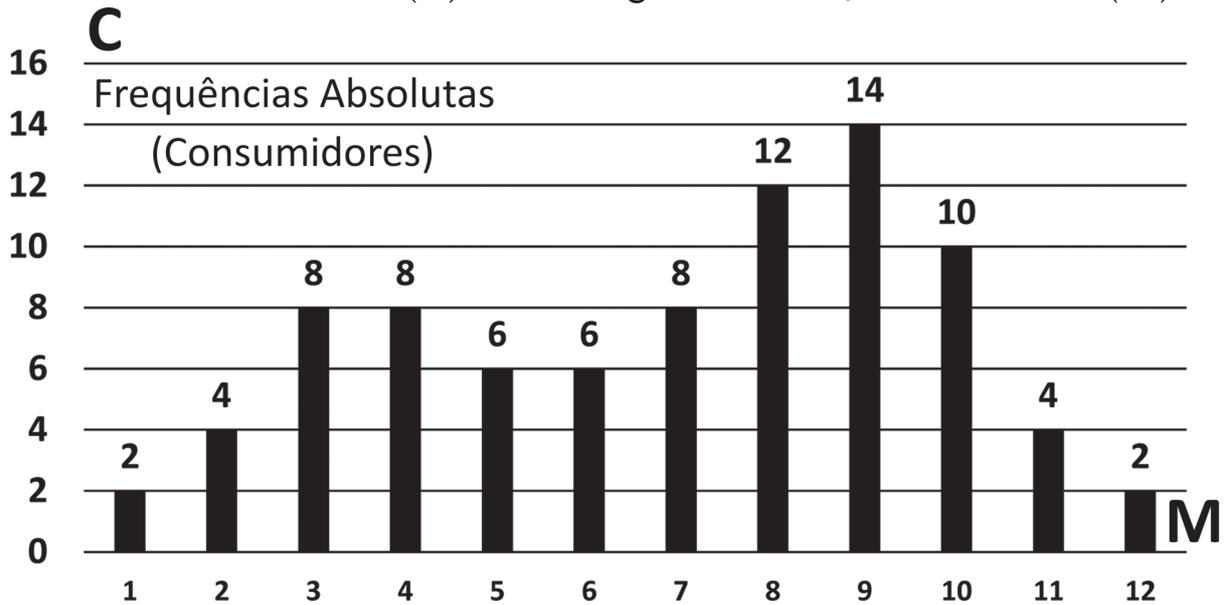
Distribuições de Dados Gaussianos Percentualizados com a% e b%

$$\sigma^1 [x] \quad \sigma^2 [x]$$

Estequiometria (a, b) da Distribuição de Gauss



2.1 Consumidores (C) de Energia em Função do Meses (M)





6.0 Particionamento Aleatório do Banco de Dados

$\begin{pmatrix} X \\ Y \end{pmatrix}$	$\begin{pmatrix} X1 \\ Y1 \end{pmatrix}$	$\begin{pmatrix} X2 \\ Y2 \end{pmatrix}$	$\begin{pmatrix} X3 \\ Y3 \end{pmatrix}$	$\begin{pmatrix} X4 \\ Y4 \end{pmatrix}$
AMO 1	p11	p21	p31	p41
AMO 2	p12	p22	p32	p42
Σp	1	1	1	1

p11: Probabilidade do dado (X1, Y1) pertencer à AMOstra 1

p12: Probabilidade do dado (X1, Y1) pertencer à AMOstra 2

p21: Probabilidade do dado (X2, Y2) pertencer à AMOstra 1

p22: Probabilidade do dado (X2, Y2) pertencer à AMOstra 2

p31: Probabilidade do dado (X3, Y3) pertencer à AMOstra 1

p32: Probabilidade do dado (X3, Y3) pertencer à AMOstra 2

p41: Probabilidade do dado (X4, Y4) pertencer à AMOstra 1

p42: Probabilidade do dado (X4, Y4) pertencer à AMOstra 2

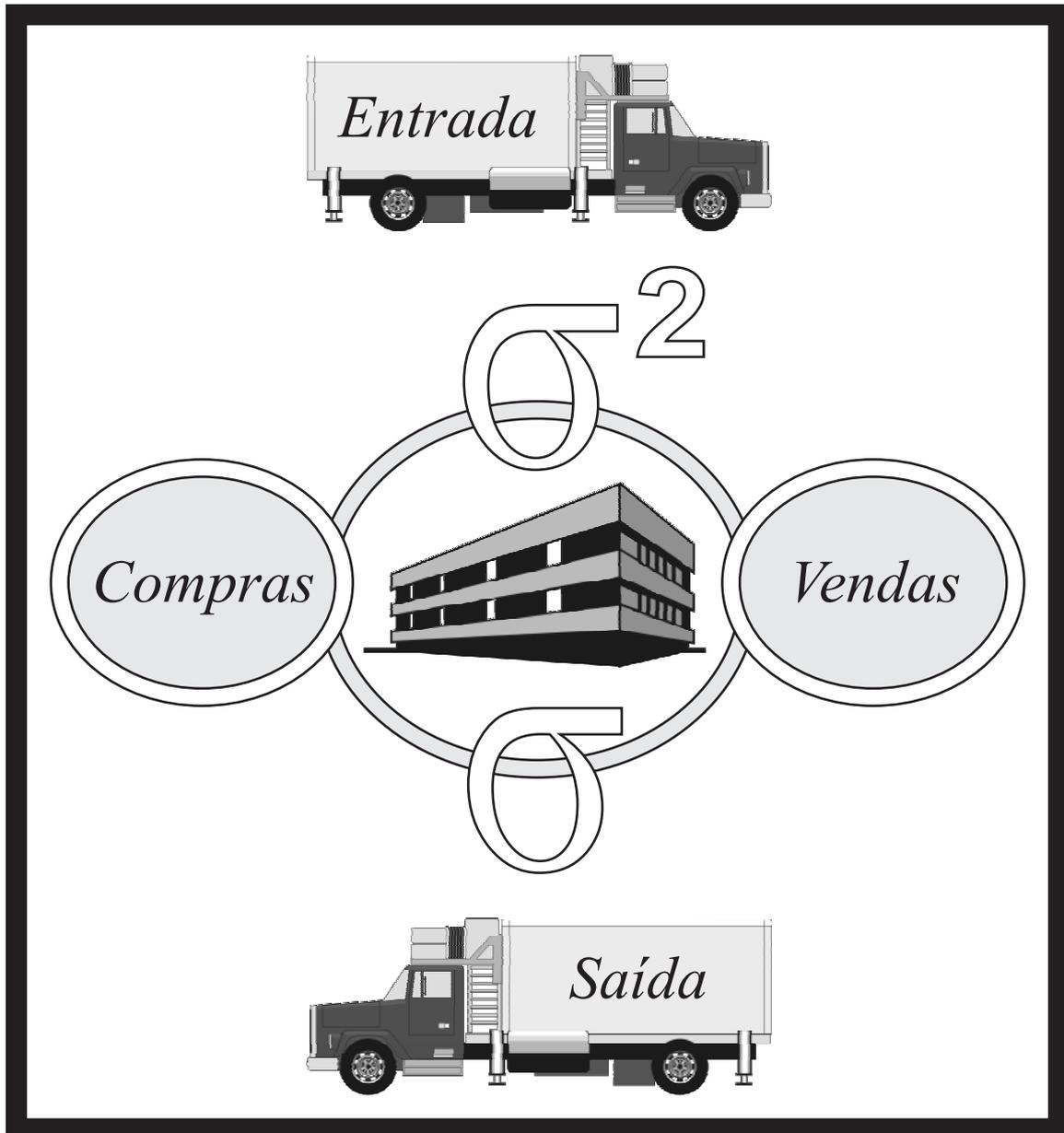
$\begin{pmatrix} X \\ Y \end{pmatrix}$	$\begin{pmatrix} X1 = 3 \\ Y1 = 2 \end{pmatrix}$	$\begin{pmatrix} X2 = 0 \\ Y2 = 3 \end{pmatrix}$	$\begin{pmatrix} X3 = 7 \\ Y3 = 6 \end{pmatrix}$	$\begin{pmatrix} X4 = 8 \\ Y4 = 7 \end{pmatrix}$
AMO 1	p11 = 0,70	p21 = 0,80	p31 = 0,10	p41 = 0,30
AMO 2	p12 = 0,30	p22 = 0,20	p32 = 0,90	p42 = 0,70
Σp	1	1	1	1

Probabilidade (p) para o Início das Iterações



LOTE IDEAL

Aprovisionamento



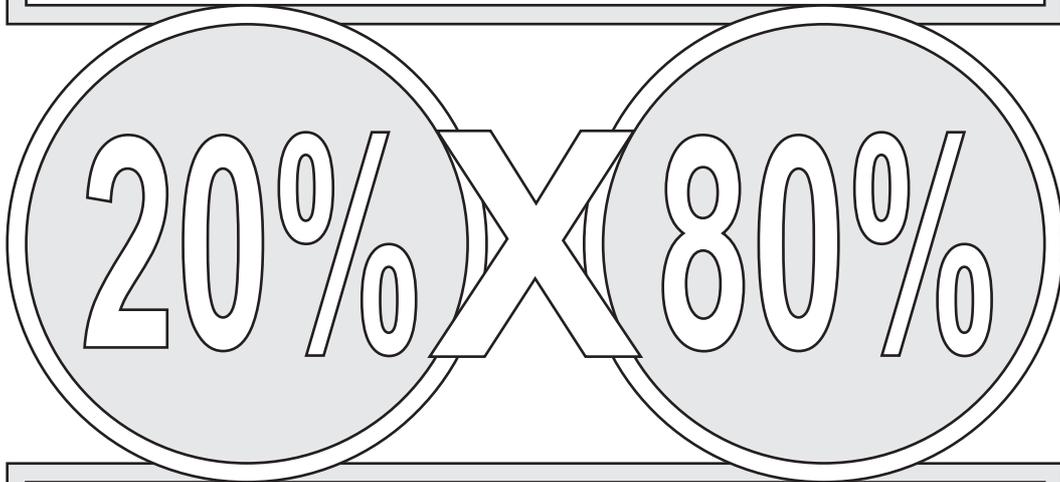
Reposição Otimizada
Quantidade de Produto e Demanda

Prof. Aguinaldo Prandini Ricieri



1.1 Princípio de Pareto na Estocagem de Produtos

O Fluxo de Produto tem seu início com a classificação ABC que caracteriza a importância econômica dos produtos pertinentes à atividade de uma empresa em foco; separando-os em três grupos ABC, por ordem decrescente de valores em relação ao valor total ou soma dos valores de cada produto analisado.



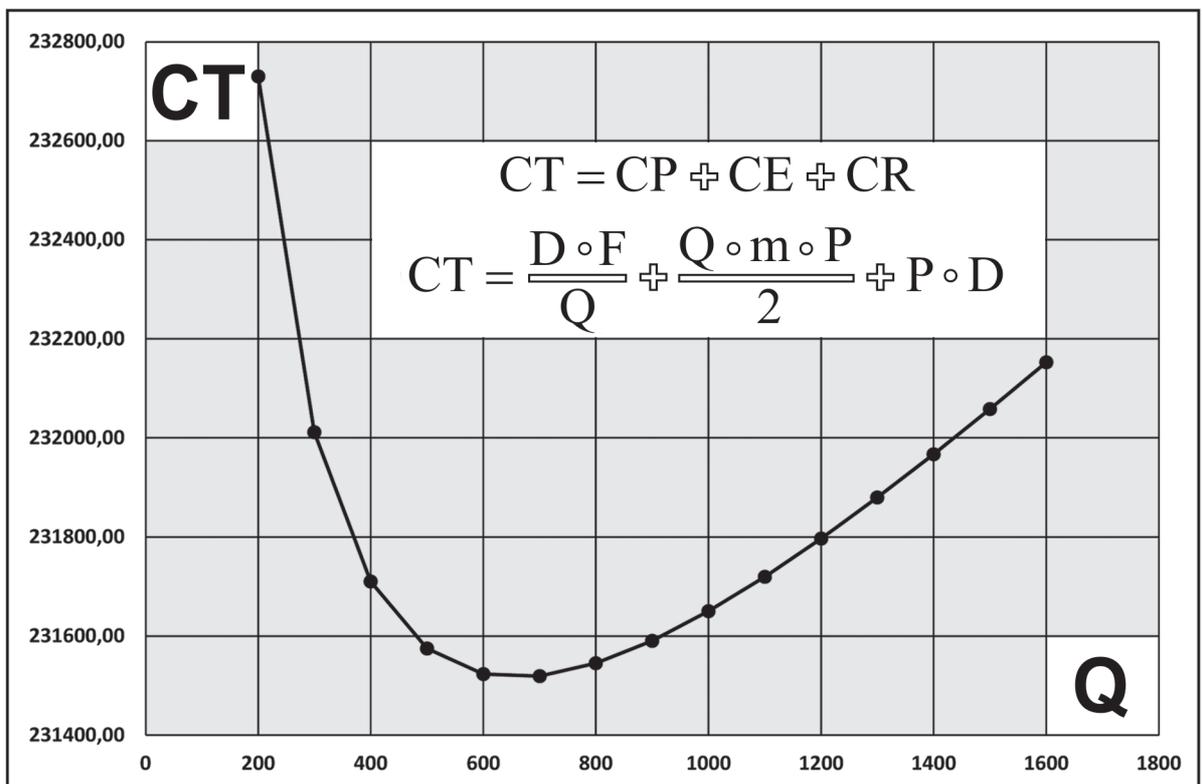
Essa classificação estatística que tem como base o Princípio de Vilfredo Pareto (1848 - 1923) – *20% das ações são responsáveis por 80% dos resultados* –, grosso modo aplicado no controle de estocagem de produtos ($\pm 20\%$ dos produtos “muito valiosos” representam $\pm 75\%$ do total, $\pm 30\%$ dos produtos “pouco valiosos” representam 20% do total e $\pm 50\%$ dos produtos “nada valiosos” representam $\pm 5\%$ do total), permite focar o modelo de otimização nos produtos que importam.

Focalização no que é Valioso para as Empresas



3.1 Simulação de Minimização do Custo Total (CT)

DEMANDA PRODUTO (D)	VALOR FIXO (F)	LOTE PEDIDO (Q)	PEDIDOS POSSIVÉIS (D/Q)	CUSTO PEDIDO (CP)	ESTOQUE CÍCLICO (Q/2)	TAXA MANUTEN (m)	PREÇO PRODUTO (P)	CUSTO ESTOQUE (CE)	CUSTO REPOSIÇÃO (CR)	CUSTO TOTAL (CT)
10000	50,00	200	50,00	2500,00	100	0,10	23,00	230,00	230000,00	232730,00
10000	50,00	300	33,33	1666,67	150	0,10	23,00	345,00	230000,00	232011,67
10000	50,00	400	25,00	1250,00	200	0,10	23,00	460,00	230000,00	231710,00
10000	50,00	500	20,00	1000,00	250	0,10	23,00	575,00	230000,00	231575,00
10000	50,00	600	16,67	833,33	300	0,10	23,00	690,00	230000,00	231523,33
10000	50,00	700	14,29	714,29	350	0,10	23,00	805,00	230000,00	231519,29
10000	50,00	800	12,50	625,00	400	0,10	23,00	920,00	230000,00	231545,00
10000	50,00	900	11,11	555,56	450	0,10	23,00	1035,00	230000,00	231590,56
10000	50,00	1000	10,00	500,00	500	0,10	23,00	1150,00	230000,00	231650,00
10000	50,00	1100	9,09	454,55	550	0,10	23,00	1265,00	230000,00	231719,55
10000	50,00	1200	8,33	416,67	600	0,10	23,00	1380,00	230000,00	231796,67
10000	50,00	1300	7,69	384,62	650	0,10	23,00	1495,00	230000,00	231879,62
10000	50,00	1400	7,14	357,14	700	0,10	23,00	1610,00	230000,00	231967,14
10000	50,00	1500	6,67	333,33	750	0,10	23,00	1725,00	230000,00	232058,33
10000	50,00	1600	6,25	312,50	800	0,10	23,00	1840,00	230000,00	232152,50

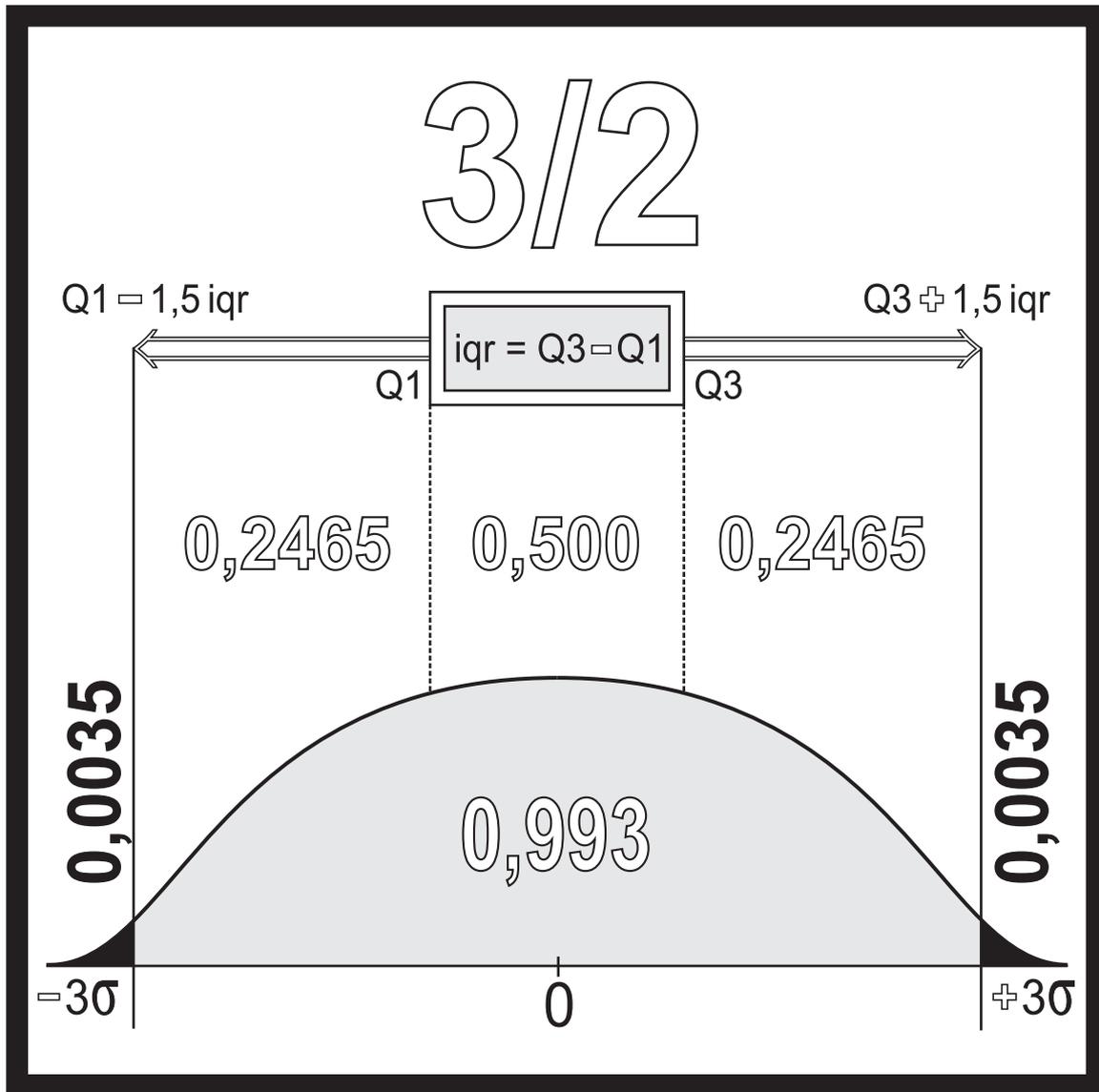


Custo Total CT(\$) Diário como Função do Lote Q(un)



OUTLIERS

Estatística Robusta



Interquartile - (**iqr**) - Range

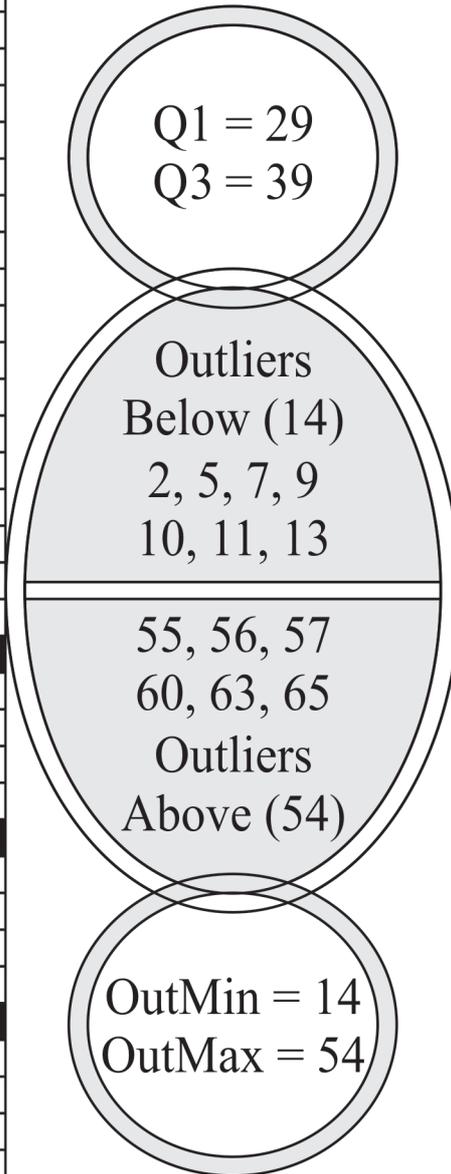
Minimum Outlier et Maximum Outlier

Prof. Aguinaldo Prandini Ricieri



3.0 Análise Estatística de Outliers Secunda

29	27	27	32	27
49	32	26	48	30
49	31	32	49	26
48	33	32	36	42
49	31	32	39	46
30	49	39	26	35
49	27	49	30	24
32	30	39	29	25
39	29	49	39	25
50	30	48	26	50
24	32	50	38	25
29	29	32	24	27
36	34	39	25	38
49	29	30	25	50
32	24	24	25	32
39	21	31	42	26
30	39	46	49	30
32	26	26	32	46
26	30	30	49	49
30	29	36	24	30
24	31	42	50	29
29	39	46	38	26
29	31	30	29	39
49	32	30	50	32
26	27	32	26	39
29	49	33	27	49
29	30	42	33	11
9	29	60	26	29
21	39	32	30	42
38	32	21	29	27
32	49	30	30	24
50	33	35	32	7
29	36	49	32	24
32	29	49	39	32
35	30	26	49	33
25	35	39	27	49
33	50	35	30	65
42	39	35	38	31
32	49	38	30	33
34	24	52	31	32
42	33	32	42	39
48	27	21	30	29
33	29	26	39	29
39	39	30	18	49
30	27	39	33	30
24	30	38	36	49
26	32	49	49	32
27	29	29	48	32
21	29	49	55	49
30	31	50	32	26

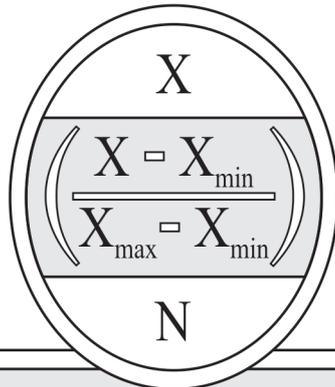


29	29	49	26	49
39	32	25	27	49
27	48	31	32	50
29	33	32	29	26
29	50	25	33	21
39	30	32	30	30
26	46	29	5	36
49	26	49	25	42
32	29	29	30	46
39	33	33	35	27
27	42	26	46	30
34	32	31	49	32
29	27	50	52	30
42	29	25	49	27
49	39	33	35	31
32	49	34	49	49
42	49	39	31	25
31	33	25	36	27
32	29	30	35	32
39	32	49	36	49
48	33	49	33	50
30	51	35	35	32
49	32	30	49	30
13	33	33	34	50
31	10	42	35	29
27	29	24	29	36
56	30	49	63	32
24	32	26	49	26
49	49	49	29	29
29	21	32	33	30
29	39	39	29	49
38	32	32	49	33
42	33	39	35	34
57	29	26	27	25
35	30	30	30	26
42	32	32	26	27
49	39	49	25	35
29	26	49	26	29
29	52	30	34	31
33	30	49	33	24
46	24	26	39	42
30	49	30	30	32
27	2	30	29	49
39	34	26	24	25
49	50	32	31	35
32	27	49	39	30
32	32	29	42	29
49	35	26	49	21
34	24	26	48	32
30	32	26	30	35



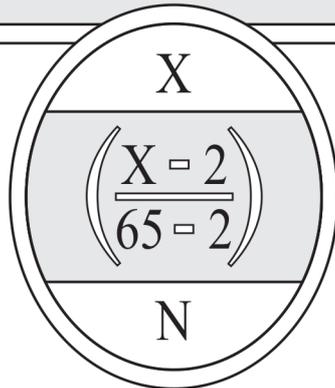
4.0 Normalização N(Max/Min) da Coleção de Dados Anterior

0,43	0,40	0,40	0,48	0,40
0,75	0,48	0,38	0,73	0,44
0,75	0,46	0,48	0,75	0,38
0,73	0,49	0,48	0,54	0,63
0,75	0,46	0,48	0,59	0,70
0,44	0,75	0,59	0,38	0,52
0,75	0,40	0,75	0,44	0,35
0,48	0,44	0,59	0,43	0,37
0,59	0,43	0,75	0,59	0,37
0,76	0,44	0,73	0,38	0,76
0,35	0,48	0,76	0,57	0,37
0,43	0,43	0,48	0,35	0,40
0,54	0,51	0,59	0,37	0,57
0,75	0,43	0,44	0,37	0,76
0,48	0,35	0,35	0,37	0,48
0,59	0,30	0,46	0,63	0,38
0,44	0,59	0,70	0,75	0,44
0,48	0,38	0,38	0,48	0,70
0,38	0,44	0,44	0,75	0,75
0,44	0,43	0,54	0,35	0,44
0,35	0,46	0,63	0,76	0,43
0,43	0,59	0,70	0,57	0,38
0,43	0,46	0,44	0,43	0,59
0,75	0,48	0,44	0,76	0,48
0,38	0,40	0,48	0,38	0,59
0,43	0,75	0,49	0,40	0,75
0,43	0,44	0,63	0,49	0,14
0,11	0,43	0,92	0,38	0,43
0,30	0,59	0,48	0,44	0,63
0,57	0,48	0,30	0,43	0,40
0,48	0,75	0,44	0,44	0,35
0,76	0,49	0,52	0,48	0,08
0,43	0,54	0,75	0,48	0,35
0,48	0,43	0,75	0,59	0,48
0,52	0,44	0,38	0,75	0,49
0,37	0,52	0,59	0,40	0,75
0,49	0,76	0,52	0,44	1,00
0,63	0,59	0,52	0,57	0,46
0,48	0,75	0,57	0,44	0,49
0,51	0,35	0,79	0,46	0,48
0,63	0,49	0,48	0,63	0,59
0,73	0,40	0,30	0,44	0,43
0,49	0,43	0,38	0,59	0,43
0,59	0,59	0,44	0,25	0,75
0,44	0,40	0,59	0,49	0,44
0,35	0,44	0,57	0,54	0,75
0,38	0,48	0,75	0,75	0,48
0,40	0,43	0,43	0,73	0,48
0,30	0,43	0,75	0,84	0,75
0,44	0,46	0,76	0,48	0,38



- X = 2 → 0,00 = N
- X = 5 → 0,05 = N
- X = 7 → 0,08 = N
- X = 9 → 0,11 = N
- X = 10 → 0,13 = N
- X = 11 → 0,14 = N
- X = 13 → 0,17 = N

- X = 55 → 0,84 = N
- X = 56 → 0,86 = N
- X = 57 → 0,87 = N
- X = 60 → 0,92 = N
- X = 63 → 0,97 = N
- X = 65 → 1,00 = N

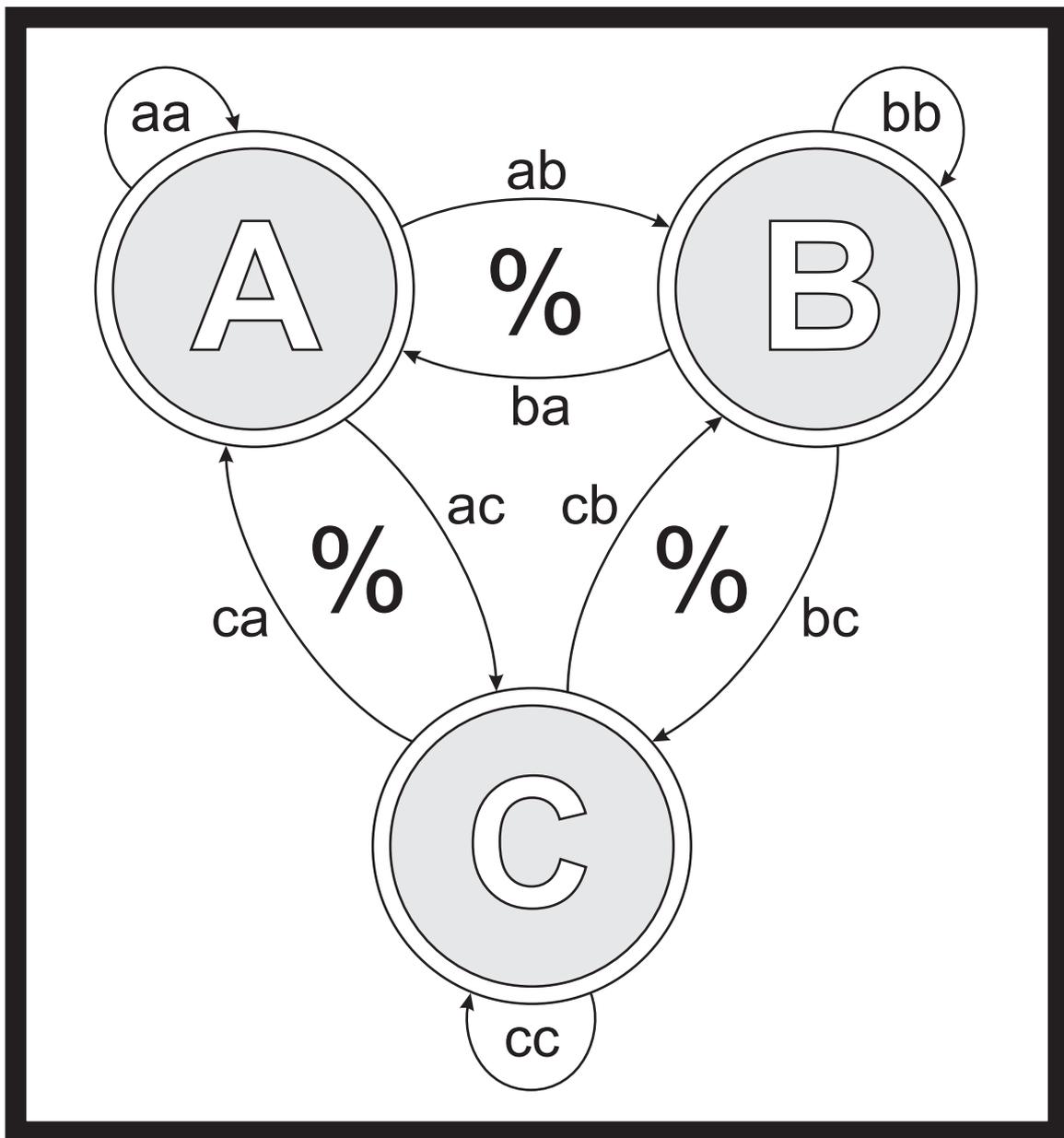


0,43	0,43	0,75	0,38	0,75
0,59	0,48	0,37	0,40	0,75
0,40	0,73	0,46	0,48	0,76
0,43	0,49	0,48	0,43	0,38
0,43	0,76	0,37	0,49	0,30
0,59	0,44	0,48	0,44	0,44
0,38	0,70	0,43	0,05	0,54
0,75	0,38	0,75	0,37	0,63
0,48	0,43	0,43	0,44	0,70
0,59	0,49	0,49	0,52	0,40
0,40	0,63	0,38	0,70	0,44
0,51	0,48	0,46	0,75	0,48
0,43	0,40	0,76	0,79	0,44
0,63	0,43	0,37	0,75	0,40
0,75	0,59	0,49	0,52	0,46
0,48	0,75	0,51	0,75	0,75
0,63	0,75	0,59	0,46	0,37
0,46	0,49	0,37	0,54	0,40
0,48	0,43	0,44	0,52	0,48
0,59	0,48	0,75	0,54	0,75
0,73	0,49	0,75	0,49	0,76
0,44	0,78	0,52	0,52	0,48
0,75	0,48	0,44	0,75	0,44
0,17	0,49	0,49	0,51	0,76
0,46	0,13	0,63	0,52	0,43
0,40	0,43	0,35	0,43	0,54
0,86	0,44	0,75	0,97	0,48
0,35	0,48	0,38	0,75	0,38
0,75	0,75	0,75	0,43	0,43
0,43	0,30	0,48	0,49	0,44
0,43	0,59	0,59	0,43	0,75
0,57	0,48	0,48	0,75	0,49
0,63	0,49	0,59	0,52	0,51
0,87	0,43	0,38	0,40	0,37
0,52	0,44	0,44	0,44	0,38
0,63	0,48	0,48	0,38	0,40
0,75	0,59	0,75	0,37	0,52
0,43	0,38	0,75	0,38	0,43
0,43	0,79	0,44	0,51	0,46
0,49	0,44	0,75	0,49	0,35
0,70	0,35	0,38	0,59	0,63
0,44	0,75	0,44	0,44	0,48
0,40	0,00	0,44	0,43	0,75
0,59	0,51	0,38	0,35	0,37
0,75	0,76	0,48	0,46	0,52
0,48	0,40	0,75	0,59	0,44
0,48	0,48	0,43	0,63	0,43
0,75	0,52	0,38	0,75	0,30
0,51	0,35	0,38	0,73	0,48
0,44	0,48	0,38	0,44	0,52



BREVIS ACTIO

Estatística de Markov



Memória Curta

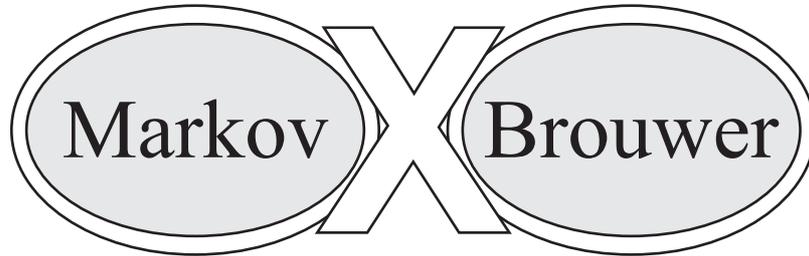
Desdobramento na Teoria do Ponto Fixo

Prof. Aguinaldo Prandini Ricieri



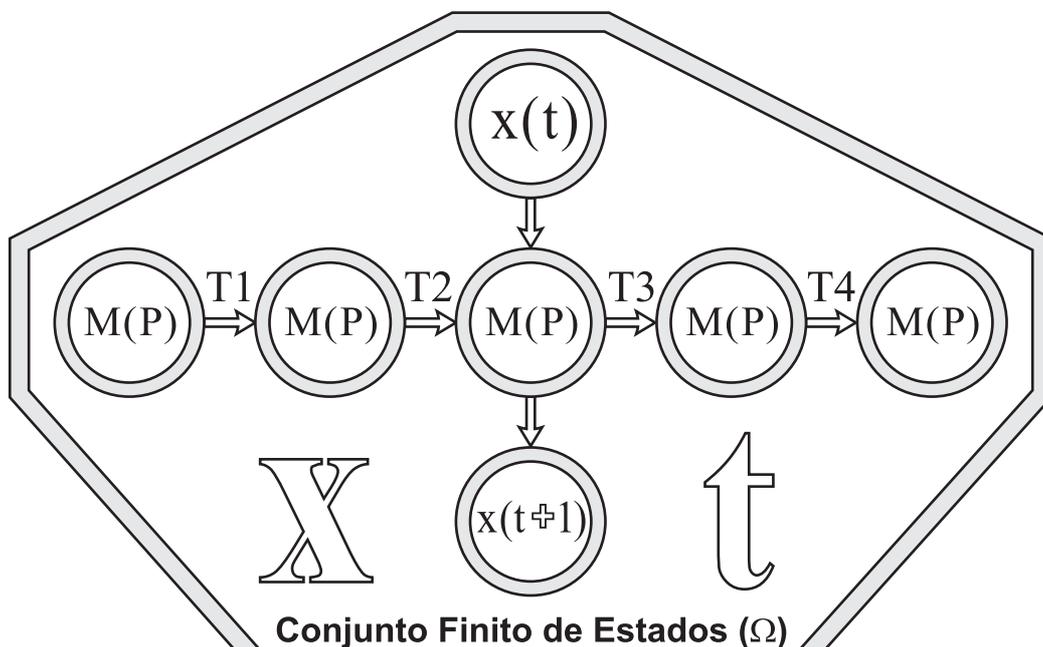
Brevis Actio

1.0 Cadeias ou Sequências de Andrei Markov



Uma “partícula” de informação (mensagem, molécula, bola de snooker, peteca, ..., meliante urbano) que ocupa uma posição $x(t)$ no espaço de um conjunto finito de estados, sofrendo à ação de um agente físico em certo tempo (t), ocupará uma outra posição $x(t \oplus 1)$ também pertencente a esse conjunto de estados (Ω).

Prever quando e onde estará essa “partícula” – dado o tempo questiona-se a sua posição – implica em uma questão de probabilidades (P) associada à matriz markoviana de transição de estado: $M(P)$. A mudança probabilística de um estado físico para um outro é quantificado em períodos de tempo T fixos (meses, semanas, dias, horas, segundos), o que permite relacionar os resultados estatísticos do russo Andrei Markov com os resultados matriciais do holandês Luitzen Egbertus Jan Brouwer (1881 - 1966) obtidos com a surpreendente Teoria do Ponto Fixo.





7.2 Cálculo Prima (A2) do Produto $x(t) \circ M = x(t + 1)$



0,000	0,500	0,250	0,333	0,250	0,500
0,200	0,000	0,250	0,000	0,000	0,000
0,200	0,500	0,000	0,333	0,250	0,000
0,200	0,000	0,250	0,000	0,250	0,000
0,200	0,000	0,250	0,333	0,000	0,500
0,200	0,000	0,000	0,000	0,250	0,000

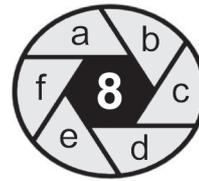
M

0,245
0,103
0,198
0,153
0,198
0,103

X(7)

0,253
0,099
0,201
0,148
0,201
0,099

X(8)



0,000	0,500	0,250	0,333	0,250	0,500
0,200	0,000	0,250	0,000	0,000	0,000
0,200	0,500	0,000	0,333	0,250	0,000
0,200	0,000	0,250	0,000	0,250	0,000
0,200	0,000	0,250	0,333	0,000	0,500
0,200	0,000	0,000	0,000	0,250	0,000

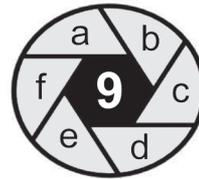
M

0,253
0,099
0,201
0,148
0,201
0,099

X(8)

0,248
0,101
0,199
0,151
0,199
0,101

X(9)



0,000	0,500	0,250	0,333	0,250	0,500
0,200	0,000	0,250	0,000	0,000	0,000
0,200	0,500	0,000	0,333	0,250	0,000
0,200	0,000	0,250	0,000	0,250	0,000
0,200	0,000	0,250	0,333	0,000	0,500
0,200	0,000	0,000	0,000	0,250	0,000

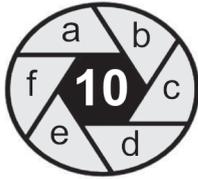
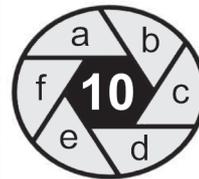
M

0,248
0,101
0,199
0,151
0,199
0,101

X(9)

0,251
0,100
0,200
0,149
0,200
0,100

X(10)



0,000	0,500	0,250	0,333	0,250	0,500
0,200	0,000	0,250	0,000	0,000	0,000
0,200	0,500	0,000	0,333	0,250	0,000
0,200	0,000	0,250	0,000	0,250	0,000
0,200	0,000	0,250	0,333	0,000	0,500
0,200	0,000	0,000	0,000	0,250	0,000

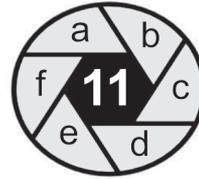
M

0,251
0,100
0,200
0,149
0,200
0,100

X(10)

0,249
0,100
0,200
0,150
0,200
0,100

X(11)



0,000	0,500	0,250	0,333	0,250	0,500
0,200	0,000	0,250	0,000	0,000	0,000
0,200	0,500	0,000	0,333	0,250	0,000
0,200	0,000	0,250	0,000	0,250	0,000
0,200	0,000	0,250	0,333	0,000	0,500
0,200	0,000	0,000	0,000	0,250	0,000

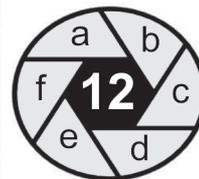
M

0,249
0,100
0,200
0,150
0,200
0,100

X(11)

0,250
0,100
0,200
0,150
0,200
0,100

X(12)



0,000	0,500	0,250	0,333	0,250	0,500
0,200	0,000	0,250	0,000	0,000	0,000
0,200	0,500	0,000	0,333	0,250	0,000
0,200	0,000	0,250	0,000	0,250	0,000
0,200	0,000	0,250	0,333	0,000	0,500
0,200	0,000	0,000	0,000	0,250	0,000

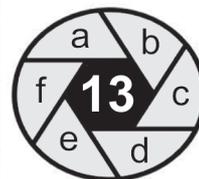
M

0,250
0,100
0,200
0,150
0,200
0,100

X(12)

0,250
0,100
0,200
0,150
0,200
0,100

X(13)



0,000	0,500	0,250	0,333	0,250	0,500
0,200	0,000	0,250	0,000	0,000	0,000
0,200	0,500	0,000	0,333	0,250	0,000
0,200	0,000	0,250	0,000	0,250	0,000
0,200	0,000	0,250	0,333	0,000	0,500
0,200	0,000	0,000	0,000	0,250	0,000

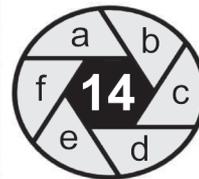
M

0,250
0,100
0,200
0,150
0,200
0,100

X(13)

0,250
0,100
0,200
0,150
0,200
0,100

X(14)



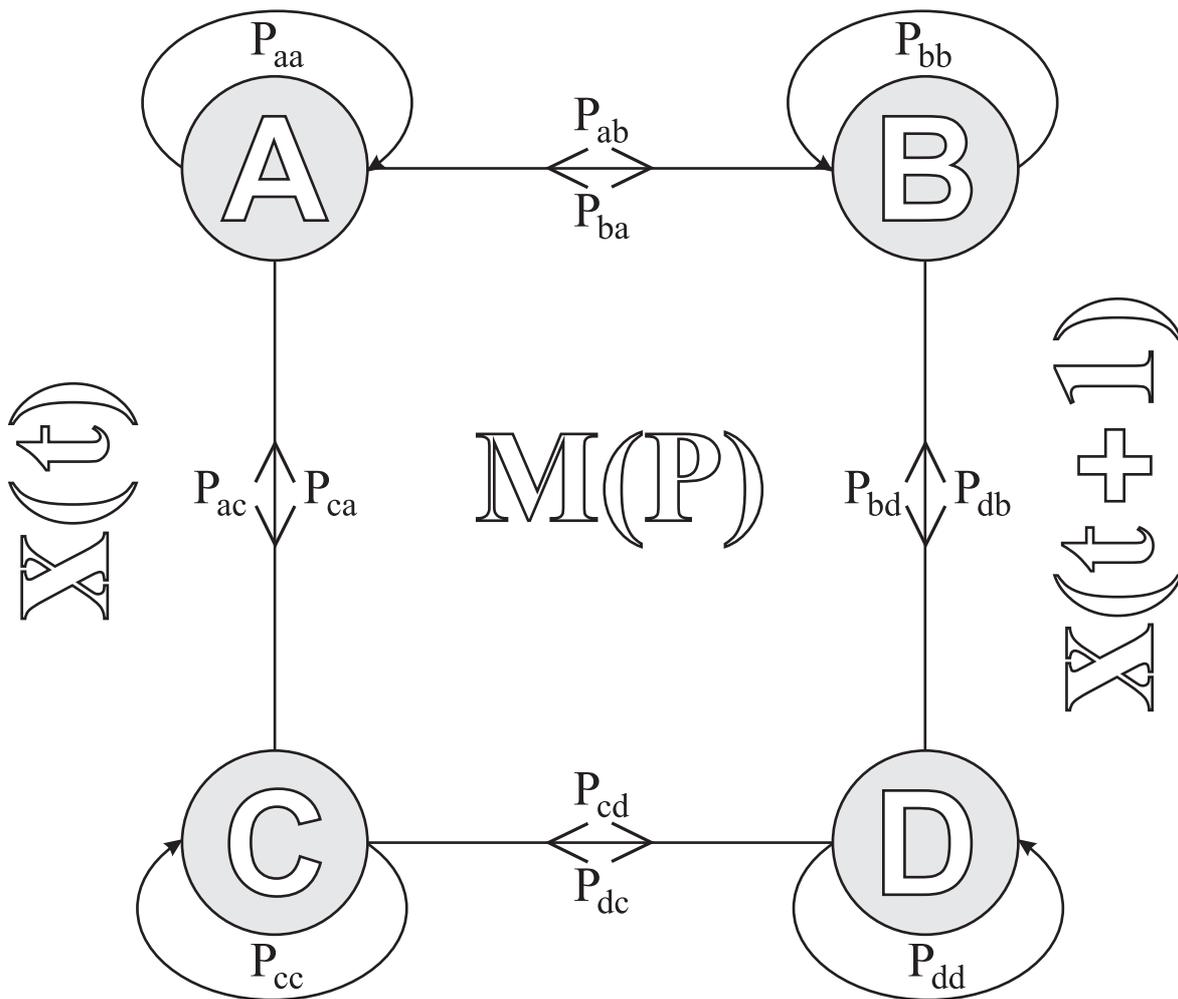
Probabilidades de Chegadas do Meliante



9.0 Movimento do Capital e Probabilidades M(P)

Sendo as arestas do grafo markoviano associadas as probabilidades de um investimento bancário mover-se no conjunto de aplicações/papéis, entre os pontos/papéis, por exemplo A, B, C e D, com o intuito de um melhor (%) rendimento.

Para as posições do capital/dinheiro exemplificado $X(t)$ e $X(t + 1)$, no tempo, associa-se uma matriz de probabilidade condicional segundo o próximo esquema:



$$X(t) \circ M(P) = X(t + 1)$$

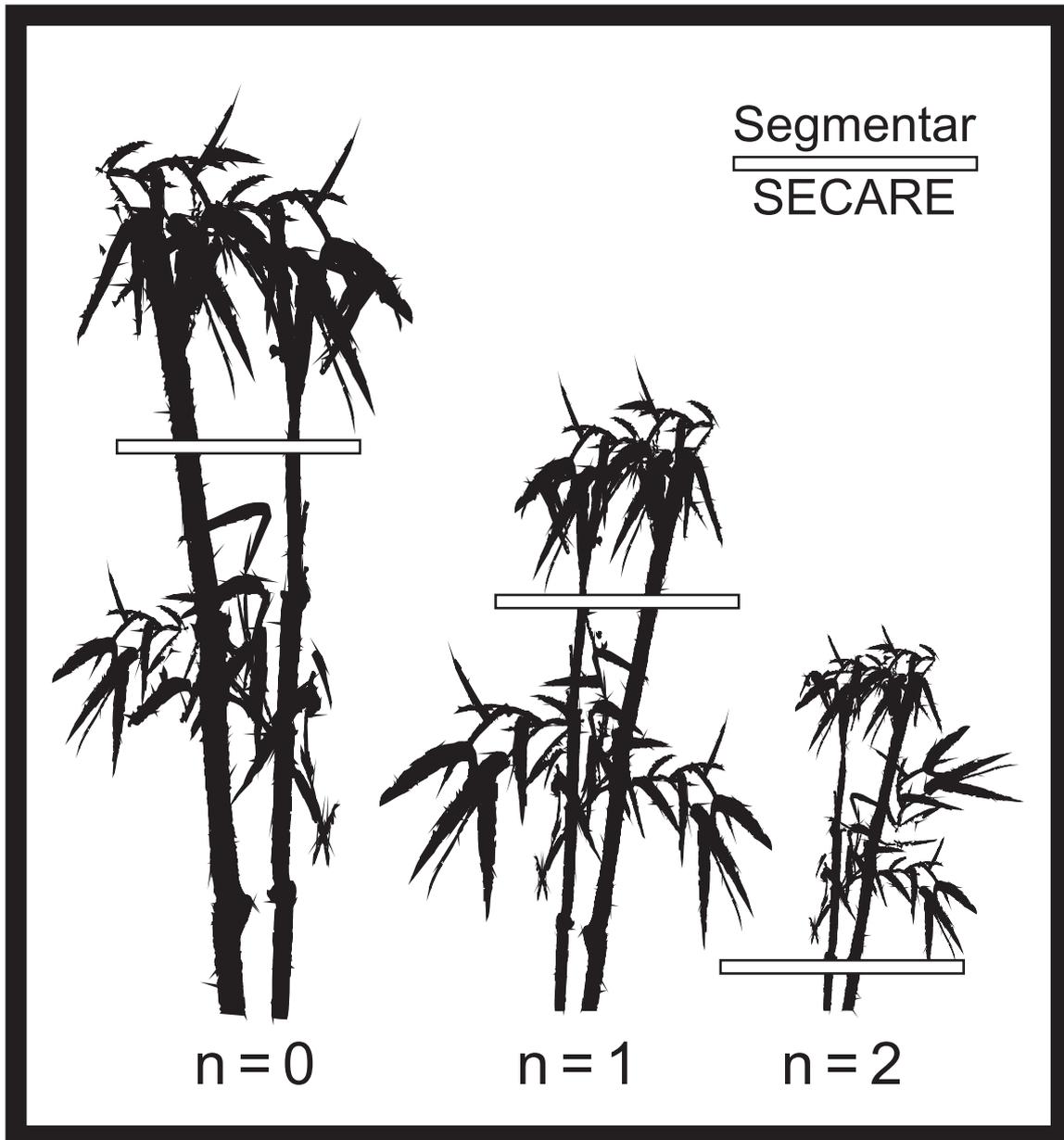
A técnica estocástica quadriangular de Markov, que associa a probabilidade de um investimento estar na posição diária $X(t + 1)$, dado uma matriz de probabilidade $M(P)$ de n elementos, o método matricial anunciado corrigirá a posição dada $X(t)$.

Poder Bancário e Previsão de Aplicações



SECTIONE

Sacchari Numerus Cuts



Variedade PR57213

Número de Cortes do Canavial

Prof. Aguinaldo Prandini Ricieri



2.0 Obtenção da Função Receita Média \bar{R}

$$R_n = [A] - n \circ [B]$$

R

$$R_0 = [A] - 0 \circ [B]$$

$$R_1 = [A] - 1 \circ [B]$$

$$R_2 = [A] - 2 \circ [B]$$

$$R_3 = [A] - 3 \circ [B]$$

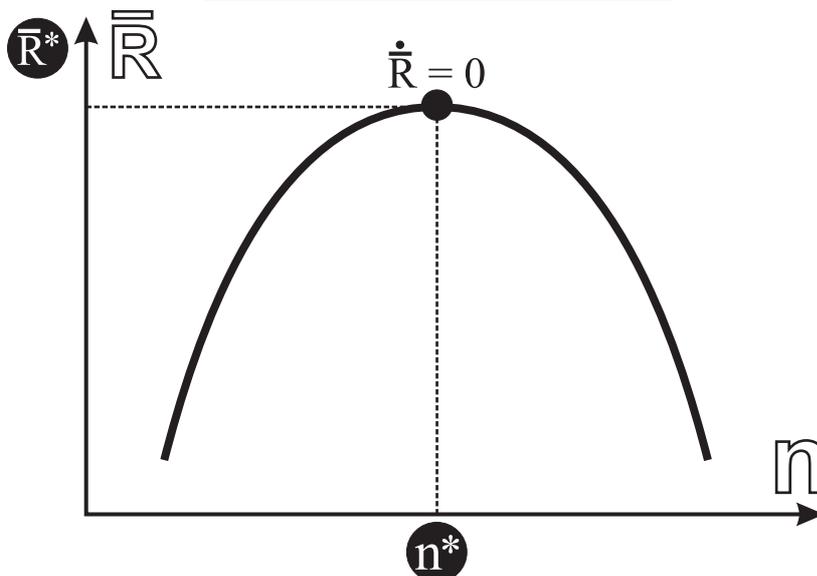
n

$$\bar{R} = \frac{-I_0 + A - 0 \circ B + A - 1 \circ B + A - 2 \circ B + \dots + A - n \circ B}{n + 1}$$

$$\bar{R} = \frac{-I_0 + (n + 1) \circ A - B \circ (1 + 2 + 3 + \dots + n)}{n + 1}$$

$$\bar{R} = A - \frac{I_0}{(n + 1)} - \frac{B}{(n + 1)} \circ \left(\frac{n(n + 1)}{2} \right)$$

$$\bar{R} = A - \frac{I_0}{n + 1} - \frac{B \circ n}{2}$$



Número de Cortes (n) e Investimento Inicial (I₀)



7.0 Ordenação Crescente das Médias das 1000 Reamostragem B

250,04	250,21	250,41	250,56	250,72	250,88	251,01	251,23	251,40	251,57
250,05	250,22	250,42	250,57	250,72	250,88	251,01	251,23	251,41	251,57
250,05	250,22	250,42	250,58	250,73	250,88	251,02	251,24	251,41	251,57
250,06	250,23	250,42	250,58	250,73	250,89	251,02	251,24	251,42	251,58
250,06	250,23	250,42	250,58	250,74	250,89	251,02	251,24	251,42	251,58
250,07	250,23	250,42	250,59	250,74	250,89	251,03	251,24	251,42	251,58
250,07	250,24	250,42	250,59	250,75	250,89	251,04	251,24	251,42	251,58
250,07	250,24	250,43	250,59	250,75	250,89	251,04	251,24	251,43	251,59
250,08	250,25	250,43	250,60	250,75	250,90	251,05	251,24	251,43	251,60
250,08	250,25	250,44	250,60	250,76	250,90	251,05	251,25	251,43	251,61
250,08	250,25	250,44	250,61	250,76	250,90	251,06	251,25	251,44	251,61
250,08	250,25	250,44	250,61	250,76	250,90	251,06	251,25	251,44	251,61
250,09	250,25	250,44	250,62	250,76	250,91	251,06	251,25	251,44	251,61
250,09	250,27	250,44	250,62	250,76	250,91	251,08	251,25	251,44	251,63
250,09	250,27	250,45	250,62	250,76	250,91	251,08	251,25	251,44	251,65
250,09	250,27	250,45	250,63	250,77	250,91	251,08	251,26	251,45	251,68
250,09	250,28	250,45	250,63	250,77	250,91	251,10	251,27	251,45	251,69
250,10	250,29	250,46	250,63	250,77	250,92	251,10	251,27	251,46	251,69
250,10	250,29	250,46	250,64	250,77	250,92	251,11	251,27	251,46	251,70
250,10	250,30	250,46	250,64	250,77	250,92	251,11	251,27	251,46	251,73
250,10	250,31	250,47	250,64	250,78	250,92	251,11	251,28	251,47	251,80
250,10	250,31	250,47	250,64	250,78	250,92	251,11	251,28	251,47	251,85
250,11	250,32	250,48	250,64	250,78	250,93	251,11	251,29	251,48	251,90
250,11	250,32	250,48	250,64	250,78	250,93	251,11	251,30	251,48	251,96
250,11	250,33	250,48	250,64	250,79	250,93	251,12	251,30	251,48	251,97
250,13	250,33	250,49	250,64	250,79	250,93	251,12	251,30	251,48	251,65
250,14	250,33	250,50	250,65	250,79	250,94	251,12	251,30	251,49	251,65
250,14	250,33	250,50	250,65	250,79	250,94	251,13	251,31	251,49	251,65
250,15	250,33	250,50	250,66	250,79	250,94	251,13	251,31	251,49	251,65
250,15	250,34	250,50	250,67	250,80	250,94	251,13	251,31	251,49	251,66
250,15	250,34	250,50	250,67	250,80	250,95	251,13	251,31	251,49	251,66
250,15	250,35	250,51	250,67	250,81	250,95	251,13	251,32	251,50	251,67
250,15	250,35	250,51	250,67	250,81	250,96	251,13	251,32	251,50	251,67
250,16	250,35	250,51	250,67	250,82	250,96	251,13	251,33	251,50	251,67
250,16	250,35	250,51	250,68	250,82	250,96	251,13	251,34	251,50	251,68
250,16	250,36	250,51	250,68	250,83	250,96	251,14	251,34	251,50	251,68
250,16	250,36	250,52	250,68	250,84	250,96	251,14	251,34	251,51	251,68
250,16	250,36	250,52	250,69	250,84	250,97	251,14	251,34	251,51	251,68
250,17	250,36	250,52	250,69	250,84	250,97	251,15	251,35	251,51	251,68
250,17	250,36	250,52	250,69	250,85	250,97	251,15	251,35	251,51	251,69
250,17	250,36	250,52	250,70	250,85	250,98	251,16	251,35	251,52	251,69
250,18	250,36	250,53	250,70	250,86	250,98	251,17	251,36	251,52	251,69
250,18	250,37	250,53	250,70	250,86	250,99	251,17	251,37	251,53	251,70
250,18	250,38	250,53	250,70	250,86	250,99	251,17	251,37	251,53	251,71
250,19	250,39	250,53	250,70	250,86	251,00	251,18	251,37	251,54	251,72
250,19	250,39	250,54	250,70	250,86	251,00	251,18	251,38	251,54	251,72
250,19	250,39	250,55	250,71	250,87	251,00	251,19	251,38	251,55	251,72
250,20	250,39	250,55	250,71	250,87	251,01	251,21	251,39	251,55	251,72
250,21	250,41	250,55	250,71	250,87	251,01	251,22	251,39	251,55	251,73
250,21	250,41	250,56	250,72	250,87	251,01	251,23	251,39	251,56	251,74

Invest = 250,04, Média



Jackknife e Bootstrap

Resampling

X01	X01	X01	X01	X01
X02	X02	X02	X02	X02
X03	X03	X03	X03	X03
X04	X04	X04	X04	X04
X05	X05	X05	X05	X05

X

	X01	X01	X01	X01
X02		X02	X02	X02
X03	X03		X03	X03
X04	X04	X04		X04
X05	X05	X05	X05	

Statistical Analysis

Estatística do Canivete e da Bota

Prof. Aguinaldo Prandini Ricieri



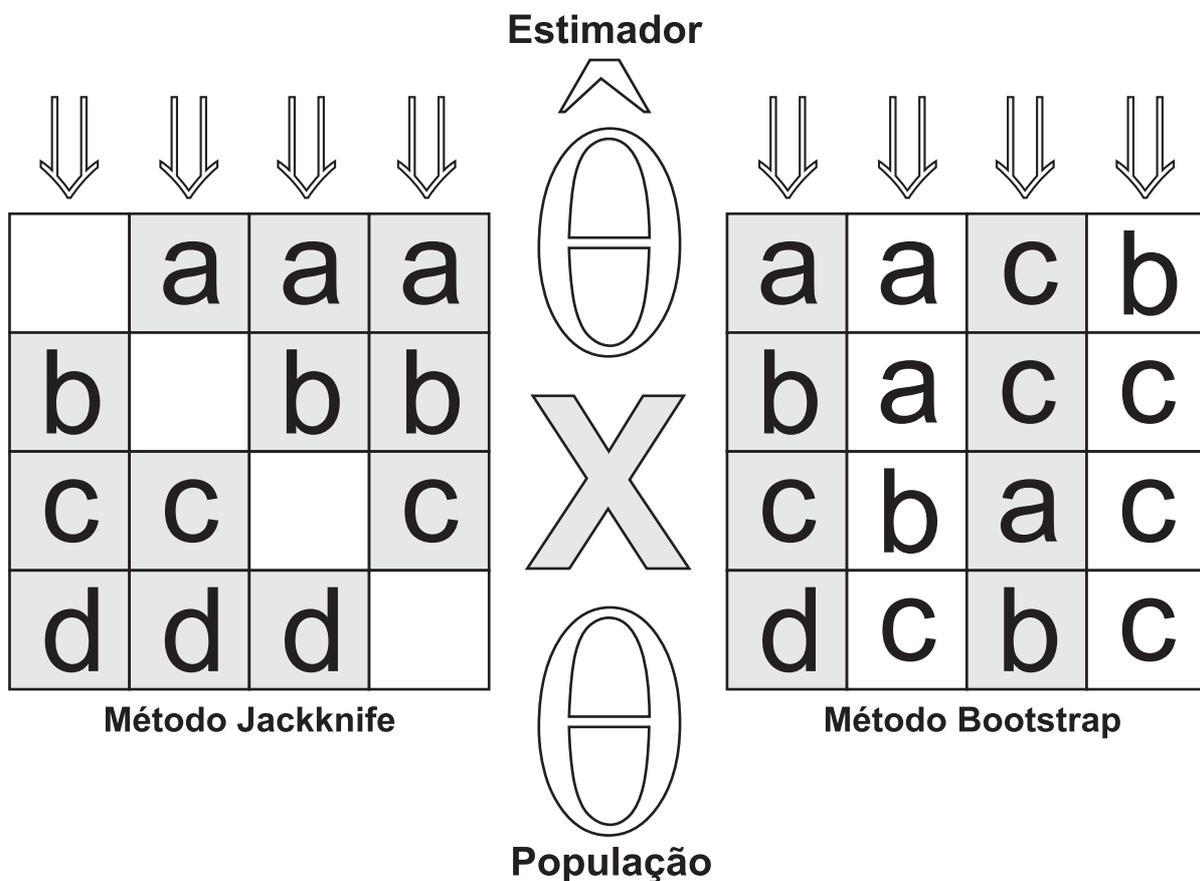
Canivetes e Botas

1.0 Jackknife e Bootstrap de “n” Dados

A estimativa de parâmetros estatísticos como média, variância e intervalo de confiança de populações, enviesadas por amostras extraídas dessas mesmas populações estudadas, implica em dois métodos de reamostragem das amostras.

Jackknife como estimador – reamostragem sistemática que subtrai um dado ($n - 1$) de cada vez da amostra original – foi aprimorado e sistematizado em 1956 por Maurice Quenoulli (1924 - 1973) no seu estudo *Notes on Day in Estimation*.

Bootstrap como estimador – reamostragens aleatórias com reposições de dados, retirados da amostra original de n dados – foi escrito e aplicado em 1979 por Bradley Efron (1938 -) no seu *Bootstrap Methods: Another Look at the Jackknife*.



Reamostragem/Resampling Estatística



2.0 Enviesamento no Cálculo da Variância

$$E(\hat{\theta}S^2) \neq E(\hat{\theta}\sigma^2)$$

Esperança do Estimador $\hat{\theta}$

$$E[\hat{\theta}S^2] = E\left[\frac{1}{n} \sum_{i=1}^n [x_i - \bar{x}]^2\right] = E\left[\frac{1}{n} \sum_{i=1}^n [x_i - \mu - \bar{x} + \mu]^2\right]$$

$$E[\hat{\theta}S^2] = E\left[\frac{1}{n} \sum_{i=1}^n ((x_i - \mu) - (\bar{x} - \mu))^2\right]$$

$$E[\hat{\theta}S^2] = E\left[\frac{1}{n} \sum_{i=1}^n ((x_i - \mu)^2 - 2(\bar{x} - \mu) \cdot (x_i - \mu) + (\bar{x} - \mu)^2)\right]$$

$$E[\hat{\theta}S^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - \frac{2}{n} \cdot (\bar{x} - \mu) \cdot \sum_{i=1}^n (x_i - \mu) + \frac{1}{n} \sum_{i=1}^n (\bar{x} - \mu)^2\right]$$

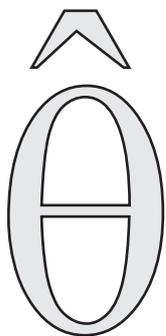
$$E[\hat{\theta}S^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - \frac{2}{n} (\bar{x} - \mu) \cdot (n) \cdot (\bar{x} - \mu) + (\bar{x} - \mu)^2\right]$$

$$E[\hat{\theta}S^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - 2(\bar{x} - \mu)^2 + (\bar{x} - \mu)^2\right]$$

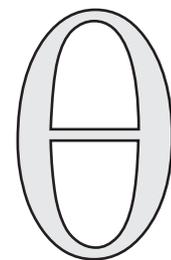
$$E[\hat{\theta}S^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - \bar{x}^2 + 2\bar{x}\mu - \mu^2\right]$$

$$E[\hat{\theta}S^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - (\bar{x} - \mu)^2\right]$$

$$E[\hat{\theta}S^2] = E\left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right] - E[(\bar{x} - \mu)^2]$$



Estimador



População

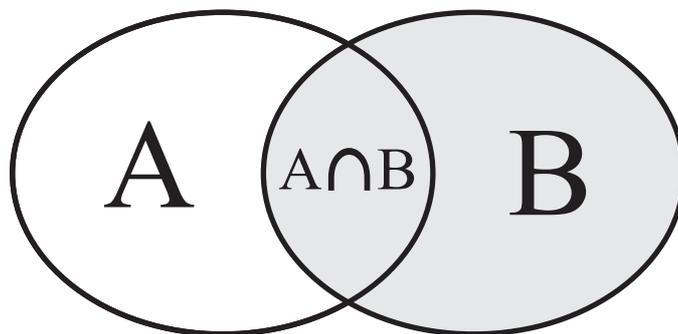
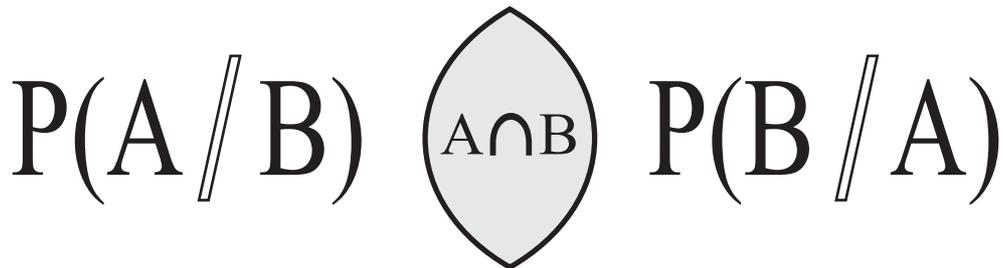
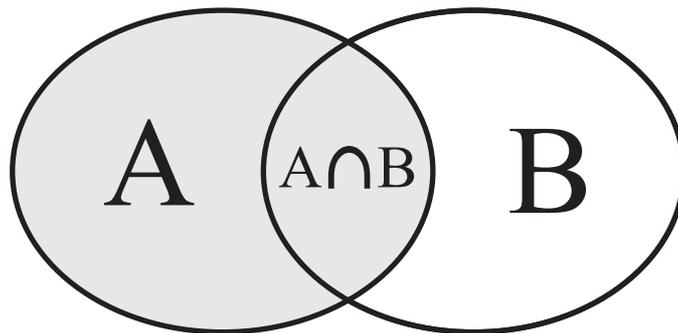
$$E[\hat{\theta}S^2] = \sigma^2 - \frac{\sigma^2}{n} = \left(\frac{n-1}{n}\right)\sigma^2$$

Estimador Jackknife da Variância



GENUS

Algoritmo Ingênuo



DataSet **X** SubSets

Prof. Aguinaldo Prandini Ricieri



2.0 Probabilidades Condicionadas e Aplicações

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

$P(A/B)$ = Probabilidade de ocorrer A tendo ocorrido o evento B.

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

$P(B/A)$ = Probabilidade de ocorrer B tendo ocorrido o evento A.

Como $P(A \cap B) = P(B \cap A)$ então: $\Rightarrow P(A/B) \circ P(B) = P(B/A) \circ P(A)$

(Thomas Bayes) $P(A/B) = \frac{P(B/A)}{P(B)} \circ P(A)$ (1701 - 1761)

$$P(A/B \circ C \circ \dots \circ E \circ F) = \frac{P(B/A) \circ P(C/A) \circ \dots \circ P(E/A) \circ P(F/A)}{P(B) \circ P(C) \circ \dots \circ P(E) \circ P(F)} \circ P(A)$$

Contabilometria

FRAUDES - IMPOSTOS	Y	FRAUDES - SEGUROS
Vendas sem Notas Fiscais	Y1	CEP Circulação ≠ Pernoite
Empresa de Laranjas	Y2	Provocante do Acidente ≠
Compra de Notas Fiscais	Y3	Acordo ≠ com Oficinas
Vendas com 1/2 Notas	Y4	Troca de Condutores
Lucro Contábil Disfarçado	Y5	Acidente que foi Simulado
Valor Vendido Incompatível	Y6	Roubo que foi Simulado

DADOS DE IMPOSTOS	X	DADOS DE SEGUROS
Anos de ICMS	X1	Idade do Segurado
CNPJ no Serasa	X2	CPF no Serasa
Receita Anual da Empresa	X3	Renda Mensal do Segurado
ICMS do Ano Anterior	X4	Tempo de Segurado
ICMS do Ano Atual	X5	Valor do Seguro Assinado
Número de Multas	X6	Número de Sinistros Feitos
Deduções de Valores do ICMS do Ano Atual	X7	Tempo Decorrido entre o Sinistro e seu Comunicado

Dados Fraudulentos

Banco de Dados do Cliente



6.0 Análise de Gastos com o Cartão de Crédito

Y1(Alimento)	Y2(Vestuário)	Y3(Remédio)	Y4(Curso)
--------------	---------------	-------------	-----------

X1(Renda): 1 a 9	X2(Ocupação): 1 a 8	X3(Gasto): 1 a 6
------------------	---------------------	------------------

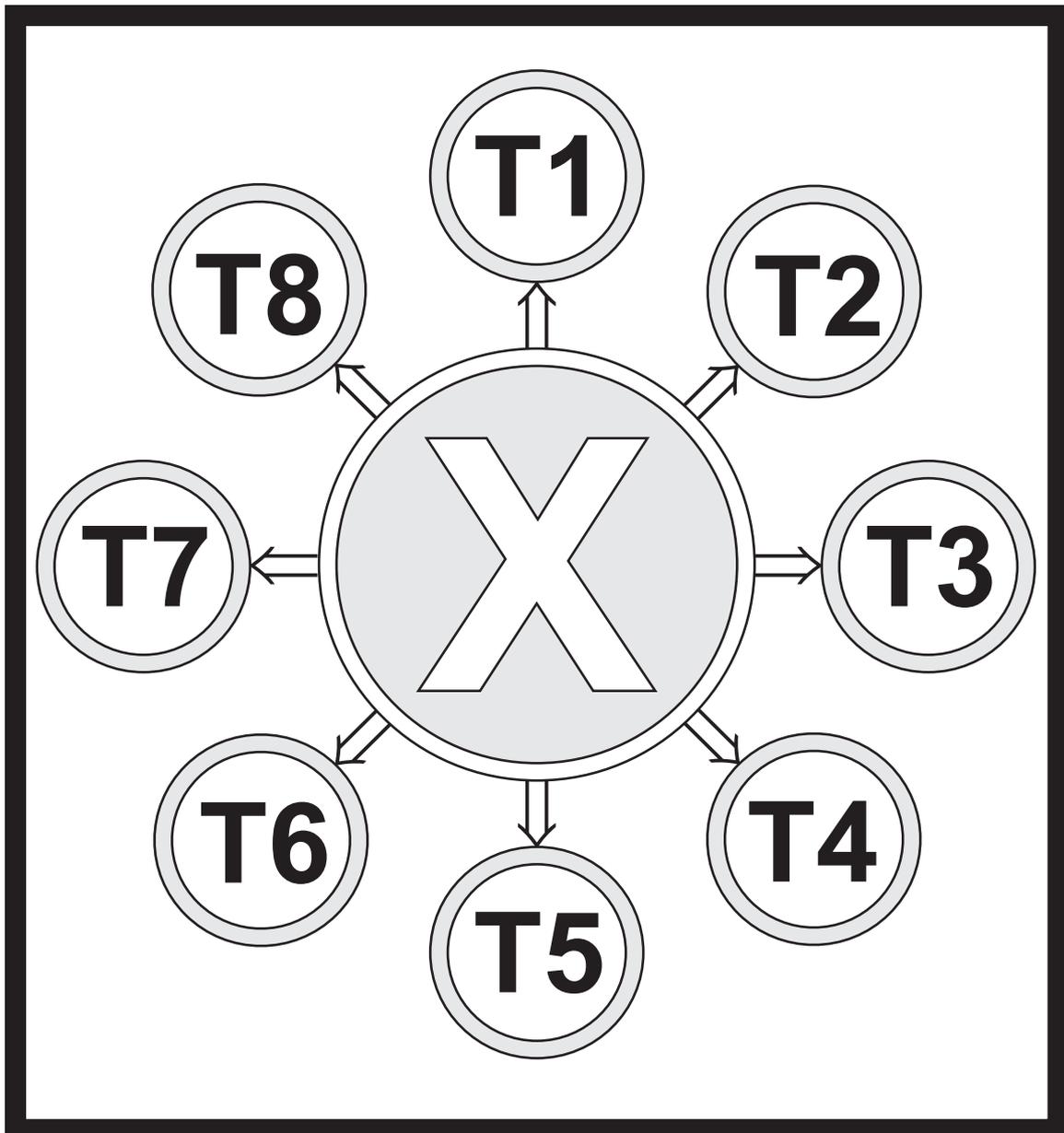
LÇ	C.CRÉD	IDADE	X1	X2	X3	Y
01	SARA	29	3	3	4	Y2 (V)
02	HUGO	43	2	7	5	Y1 (A)
03	ELSA	28	2	4	4	Y2 (V)
04	BINA	78	9	8	6	Y3 (R)
05	LARA	18	2	6	5	Y4 (C)
06	MILA	29	1	8	1	Y2 (V)
07	TIAGO	20	2	3	1	Y4 (C)
08	JADE	33	1	3	4	Y2 (V)
09	LISA	37	3	8	6	Y1 (A)
10	LUNA	18	2	6	1	Y4 (C)
11	YURI	85	3	3	6	Y3 (R)
12	ROGER	20	3	4	2	Y4 (C)
13	IVAN	51	1	3	4	Y1 (A)
14	LUIZ	69	2	3	5	Y3 (R)
15	FABIO	19	1	6	6	Y4 (C)
16	DENIS	19	9	1	1	Y4 (C)
17	PEDRO	43	3	4	6	Y1 (A)
18	AIDE	78	9	4	5	Y3 (R)
19	GABI	28	9	8	5	Y2 (V)
20	ERIC	85	9	3	4	Y3 (R)
21	ALDA	32	1	6	4	Y2 (V)
22	DORA	33	5	8	6	Y2 (V)
23	EDNA	37	3	6	1	Y1 (A)
24	DAVI	51	9	4	6	Y1 (A)
25	CARLA	18	2	8	4	Y4 (C)
26	LEDA	18	2	3	2	Y4 (C)
27	RAUL	69	1	6	1	Y3 (R)
28	DINA	32	4	8	1	Y2 (V)

DataSet do Agente de Crédito



DECERNERE

Gestão Estatística



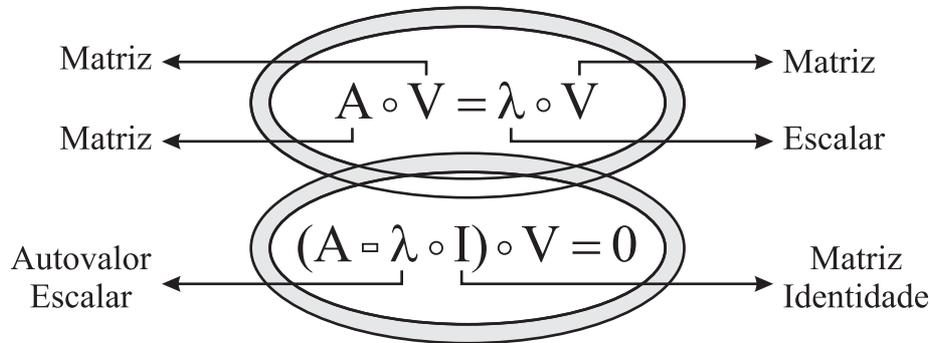
Verbo Escolher

Prof. Aguinaldo Prandini Ricieri



2.0 Problema Matricial Associado à Autovalores

Dada uma matriz quadrada $n \times n$, resolver o problema de Álgebra Linear proposto, associado ao conceito de autovalor (λ), segundo o próximo esquema:



Para $V \neq 0$ implica na solução do determinante matricial costumeiro:

$$\left(\begin{vmatrix} 5 & 4 \\ 1 & 2 \end{vmatrix} - \lambda \circ \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} \right) = 0$$

$$\lambda_1 \left(\begin{vmatrix} 5 & 4 \\ 1 & 2 \end{vmatrix} - \begin{vmatrix} \lambda & 0 \\ 0 & \lambda \end{vmatrix} \right) = 0 \quad \lambda_2$$

$$\begin{vmatrix} 5 - \lambda & 4 \\ 1 & 2 - \lambda \end{vmatrix} = 0$$

$$(5 - \lambda) \circ (2 - \lambda) - 1 \circ 4 = 0 \quad \therefore \lambda^2 - 7\lambda + 6 = 0$$

Equação dos Autovalores/Polinômio Característico

$$\lambda_1 = 6 \quad \sum_{i=1}^2 \lambda_i = 7 \quad \lambda_2 = 1$$

Somatória dos Autovetores

Vera Nikolaevna Kullanovskaya (1920 - 2012)



3.0 Percepto das Raízes Polinomiais de Verrier

Para o polinômio genérico $P(x)$ que segue na variável genérica x :

$$P(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_{n-1} x + a_n$$

Do qual, impondo a condição arbitrada $P(x) = 0$, suas raízes são dadas por:

$$x_1, x_2, x_3, x_4, x_5, \dots, x_n$$

E que a soma (S) dessas raízes, em termo do expoente k , que segue, $1 \leq k \leq n$, apresenta-se como o próximo objeto matemático indicial $i = 1, 2, 3, \dots, n$:

$$S_k = \sum_{i=1}^n x_i^k = x_1^k + x_2^k + x_3^k + \dots + x_n^k$$

Percebe-se os coeficientes (a_0, a_1, \dots, a_n) do polinômio, escritos segundo a soma das suas raízes (S_k), nas formas escalonadas, e recorrentes e semióticas:

$$\begin{array}{l}
 a_n \qquad 1a_1 = -a_0 \cdot S_1 \\
 \qquad 2a_2 = -a_0 \cdot S_2 - a_1 \cdot S_1 \\
 \qquad 3a_3 = -a_0 \cdot S_3 - a_1 \cdot S_2 - a_2 \cdot S_1 \\
 \qquad 4a_4 = -a_0 \cdot S_4 - a_1 \cdot S_3^2 - a_2 \cdot S_2 - a_3 \cdot S_1 \\
 \qquad 5a_5 = -a_0 \cdot S_5 - a_1 \cdot S_4^2 - a_2 \cdot S_3 - a_3 \cdot S_2 - a_4 \cdot S_1
 \end{array}
 \quad S_k$$

Perceptos de Urbain Le Verrier (1811 - 1877)



4.0 Obtenção das Matrizes e seus Elementos Avaliados

CI	T1	T2	T3
T1	1,0000	0,3333	0,2000
T2	3,0000	1,0000	0,5000
T3	5,0000	2,0000	1,0000

Conexão de Implementos

AP	T1	T2	T3
T1	1,0000	2,0000	2,0000
T2	0,5000	1,0000	2,0000
T3	0,5000	0,5000	1,0000

Ajuste de Patinagens

VF	T1	T2	T3
T1	1,0000	1,0000	2,0000
T2	1,0000	1,0000	1,0000
T3	0,5000	1,0000	1,0000

Valores de Financiamentos

VR	T1	T2	T3
T1	1,0000	2,0000	5,0000
T2	0,5000	1,0000	2,0000
T3	0,2000	0,5000	1,0000

Valores de Revendas

GM	T1	T2	T3
T1	1,0000	5,0000	1,0000
T2	0,2000	1,0000	0,3333
T3	1,0000	3,0000	1,0000

Gastos com Manutenções

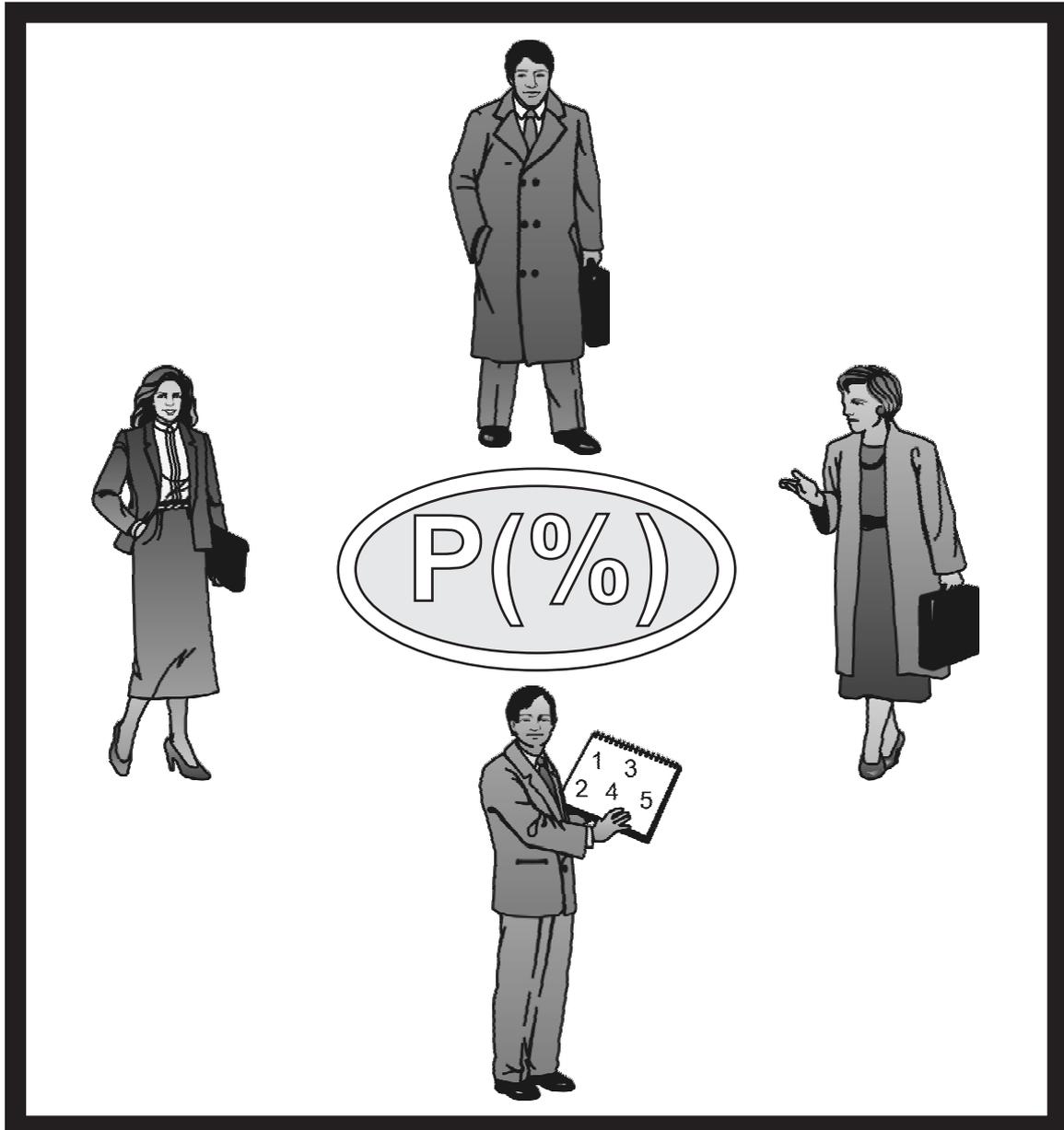
CTA	CI	AP	VF	VR	GM
CI	1,0000	0,1429	2,0000	0,1429	0,3333
AP	7,0000	1,0000	5,0000	3,0003	3,0000
VF	0,5000	0,2000	1,0000	0,3333	0,3333
VR	7,0000	0,3333	3,0000	1,0000	1,0000
GM	3,0000	0,3333	3,0000	1,0000	1,0000

Critério Total de Avaliação (CTA); Comparação Relativa



PROFESSUS

Profissionada Matricial

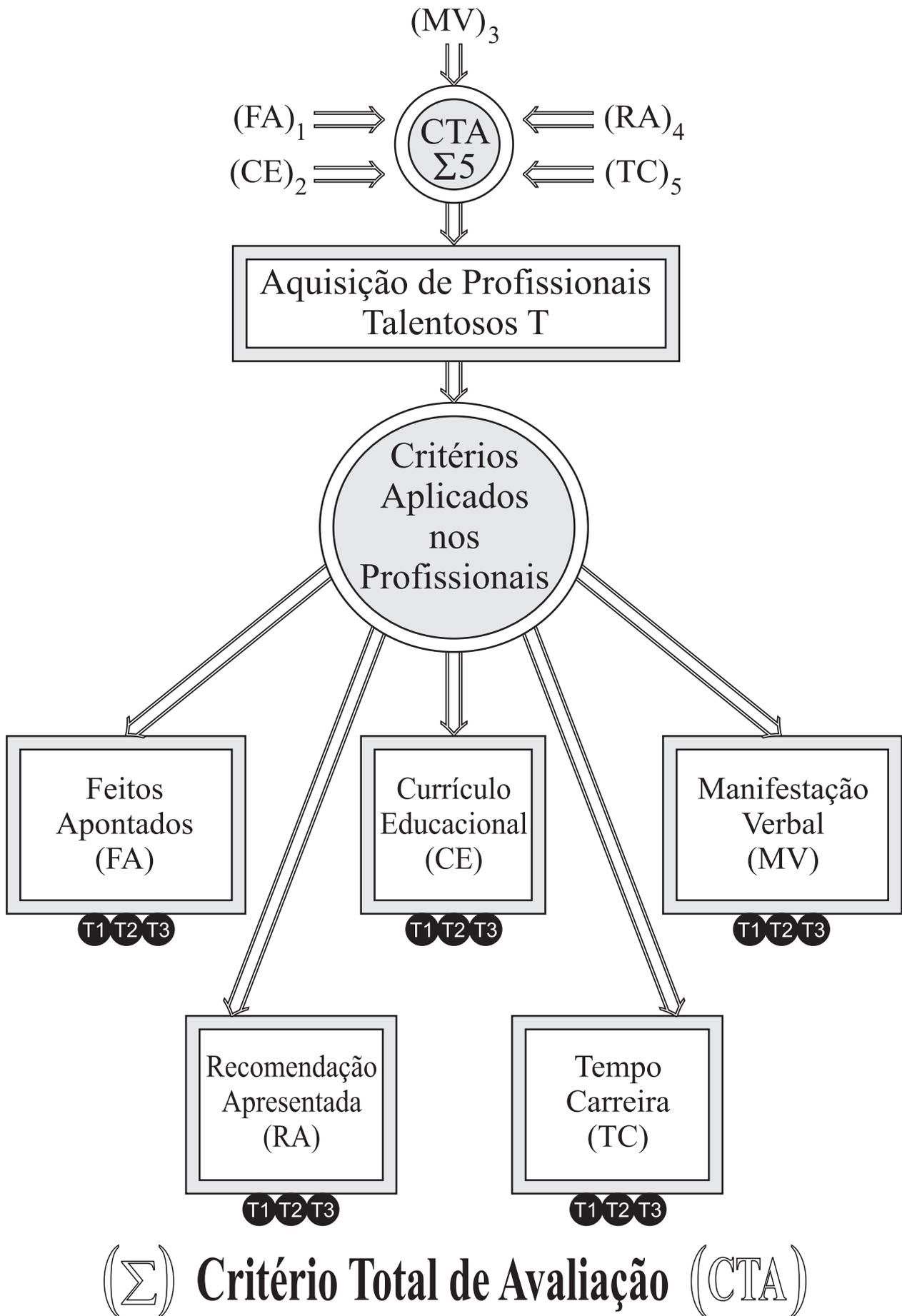


Escolha de Trabalhadores

Prof. Aguinaldo Prandini Ricieri



1.0 Fluxograma Computacional da Profissionalada

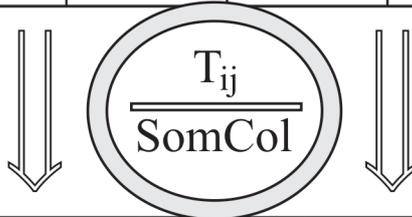




6.0 Resultados Referentes às Escolhas dos Profissionais

MATRIX PROFISSIONADA 5X5: CRITÉRIO TOTAL DE AVALIAÇÃO (CTA)					
CTA	FA	CE	MV	RA	TC
FA	1,0000	5,0000	3,0000	7,0000	9,0000
CE	0,2000	1,0000	0,3333	3,0000	5,0000
MV	0,3333	3,0000	1,0000	5,0000	3,0000
RA	0,1429	0,3333	0,2000	1,0000	5,0000
TC	0,1111	0,2000	0,3333	0,2000	1,0000
SomCol	1,7873	9,5333	4,8667	16,2000	23,0000

$$\frac{1,0000}{1,7873} = 0,5595$$



$$\frac{9,0000}{23,0000} = 0,5048$$

CTN	FA	CE	MV	RA	TC	MédLin
FA	0,5595	0,5245	0,6164	0,4321	0,3913	0,5048
CE	0,1119	0,1049	0,0685	0,1852	0,2174	0,1376
MV	0,1865	0,3147	0,2055	0,3086	0,1304	0,2291
RA	0,0799	0,0350	0,0411	0,0617	0,2174	0,0870
TC	0,0622	0,0210	0,0685	0,0123	0,0435	0,0415
SomCol	1,0000	1,0000	1,0000	1,0000	1,0000	

MATRIX PROFISSIONADA 5X5: CRITÉRIO TOTAL NORMALIZADO (CTN)

0,5040
X
0,2946
||
0,1487

MédLin CTG	MédLin (FA, CE ...TC)	Vezez
MédLin (FA) 0,5048	MédLin FA (T1) 0,2946	0,1487
MédLin (CE) 0,1376	MédLin CE (T1) 0,1549	0,0213
MédLin (MV) 0,2291	MédLin MV (T1) 0,7093	0,1625
MédLin (RA) 0,0870	MédLin RA (T1) 0,2674	0,0233
MédLin (TC) 0,0415	MédLin TC (T1) 0,5105	0,0212
CRITÉRIO DA PROFISSIONADA GERAL (T1)	SomCol:	0,3770

0,1376
X
0,1549
||
0,0213

0,2291
X
0,2141
||
0,0491

MédLin CTG	MédLin (FA, CE ...TC)	Vezez
MédLin (FA) 0,5048	MédLin FA (T2) 0,6486	0,3274
MédLin (CE) 0,1376	MédLin CE (T2) 0,0685	0,0094
MédLin (MV) 0,2291	MédLin MV (T2) 0,2141	0,0491
MédLin (RA) 0,0870	MédLin RA (T2) 0,6689	0,0582
MédLin (TC) 0,0415	MédLin TC (T2) 0,1001	0,0042
CRITÉRIO DA PROFISSIONADA GERAL (T2)	SomCol:	0,4482

0,0870
X
0,6689
||
0,0582

0,0870
X
0,0617
||
0,0055

MédLin CTG	MédLin (FA, CE ...TC)	Vezez
MédLin (FA) 0,5048	MédLin FA (T3) 0,0567	0,0286
MédLin (CE) 0,1376	MédLin CE (T3) 0,7766	0,1068
MédLin (MV) 0,2291	MédLin MV (T3) 0,0766	0,0176
MédLin (RA) 0,0870	MédLin RA (T3) 0,0637	0,0055
MédLin (TC) 0,0415	MédLin TC (T3) 0,3893	0,0162
CRITÉRIO DA PROFISSIONADA GERAL (T3)	SomCol:	0,1747

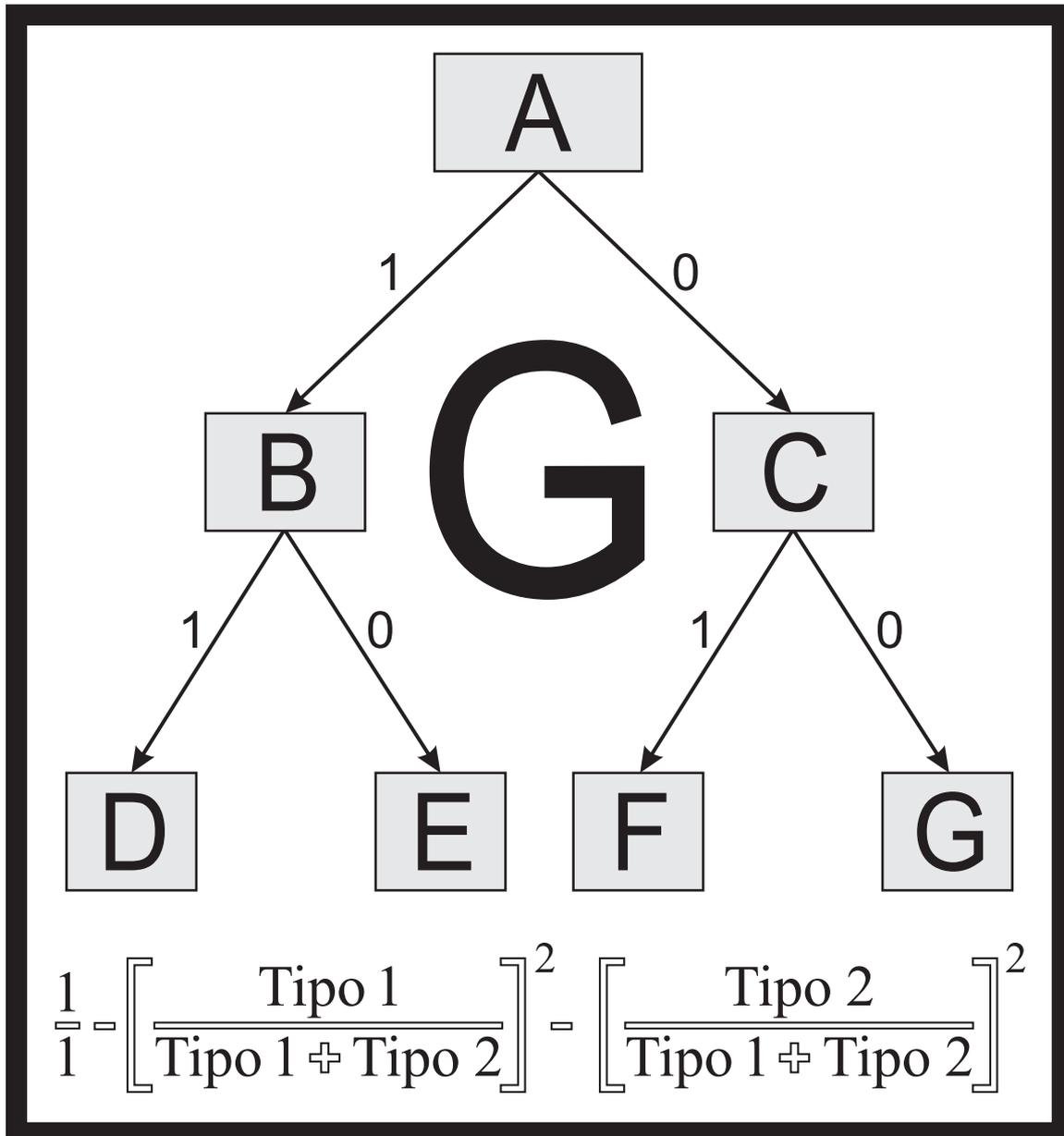
0,0415
X
0,3893
||
0,0162

MAXIMUM (CTG(T1), CTG(T2), CTG(T3)) = T2 = 0,4482 (EXECUTIVA ESCOLHIDA)



Classificador Gini

Estatística Entrópica



Numerização Probabilística

Árvore de Decisão

Prof. Aguinaldo Prandini Ricieri



Índice e Entropia Gini

1.0 (Im)pureza dos Nós Divisivos de Árvores

Proposto em 1912 pelo estatístico Conrado Gini (1884 - 1965) no seu estudo *Variabilità e Mutabilità: Contributo allo studio delle distribuzioni e delle relazioni statistiche*, o índice Gini mede o grau de heterogeneidade de classes de dados:

IMPUREZAS ESTATÍSTICAS DOS NÓS DAS ÁRVORES DE DECISÃO.

Para o número de classe (C) e a frequência relativa ou probabilidade (p_j) de uma classe acontecer em um nó divisor da árvore, o índice Gini (G) é dado por:

$$G = 1 - \sum_{j=1}^C p_j^2 = 1 - p_1^2 - p_2^2 - p_3^2 - \dots - p_C^2$$

Índice Gini da Soma Quadrada

Dado um nó (x) e duas classes de dados (j_1 e j_2), com as probabilidades de ocorrerem rotulados de $p(j_1)$ e $p(j_2)$, a entropização de Gini é conhecida na forma:

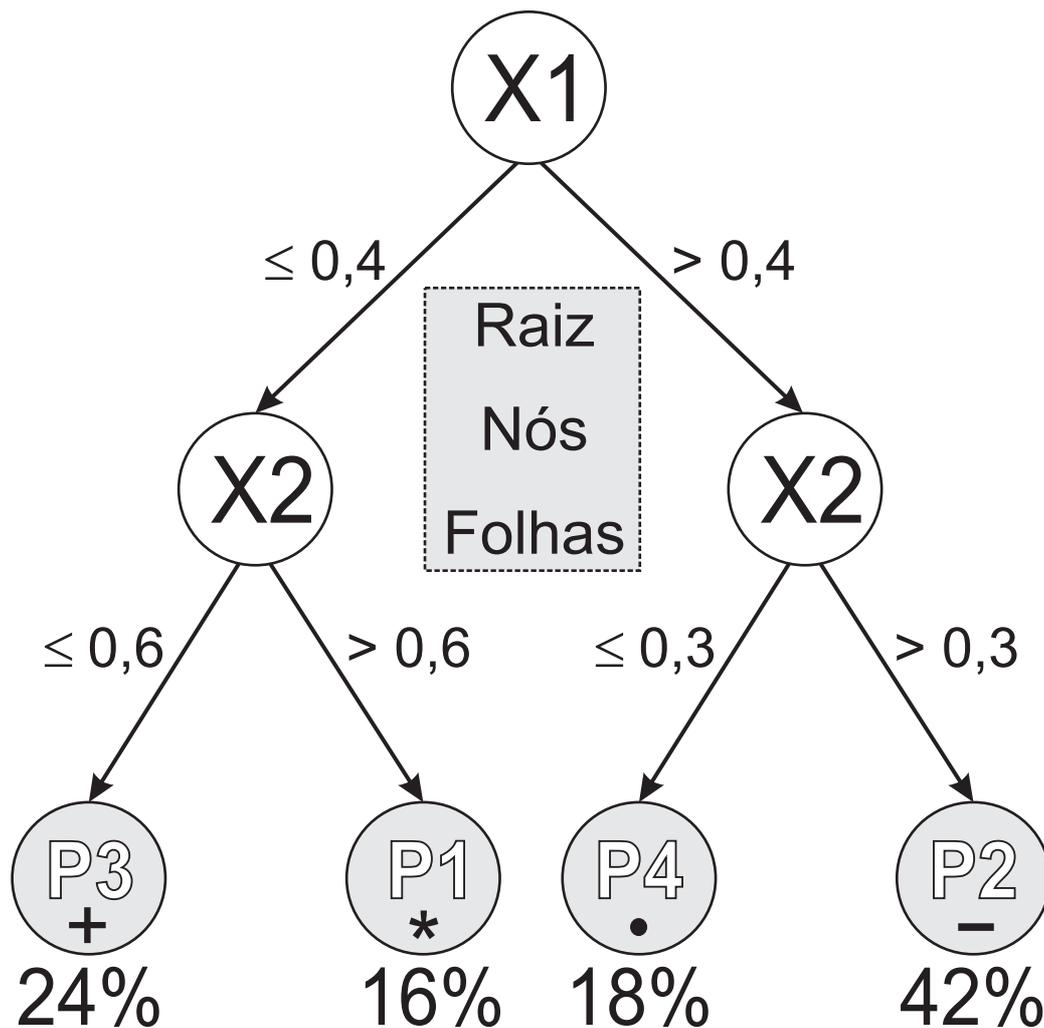
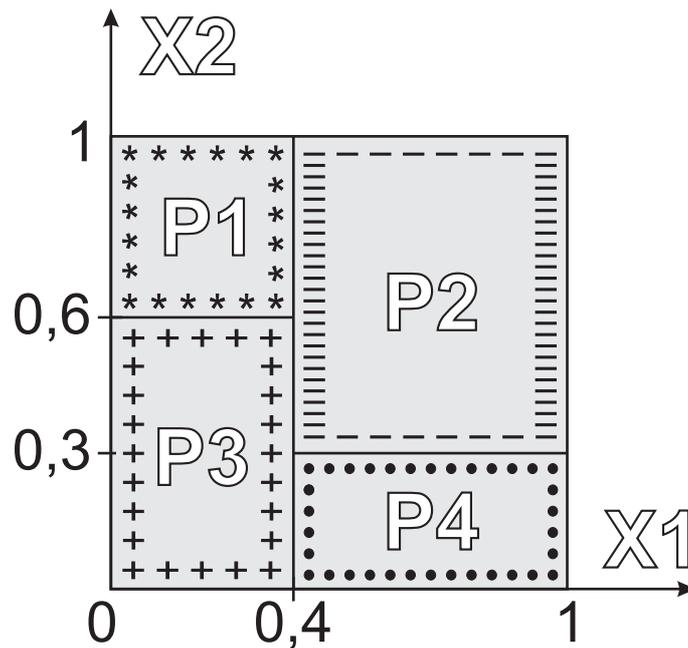
$$\text{ENT}(j_1, j_2)_x = G(j_1) \cdot p(j_1) + G(j_2) \cdot p(j_2)$$

ENTropização do Índice Gini (Grau de Desorganização)

Árvores de decisão são constructos que necessitam a priori de uma aprendizagem supervisionada. Assim, são capazes de classificarem dados médicos, financeiros, sociais..., com uma probabilidade de acerto. Esse tipo de possibilidade estatística é feita aplicando o conceito de entropia de Claude Shannon (1916 - 2001), que caracteriza a posteriori a (im)pureza das lições de um banco de dados.



1.4 Classificação Paletizada de Dados Arvorizados

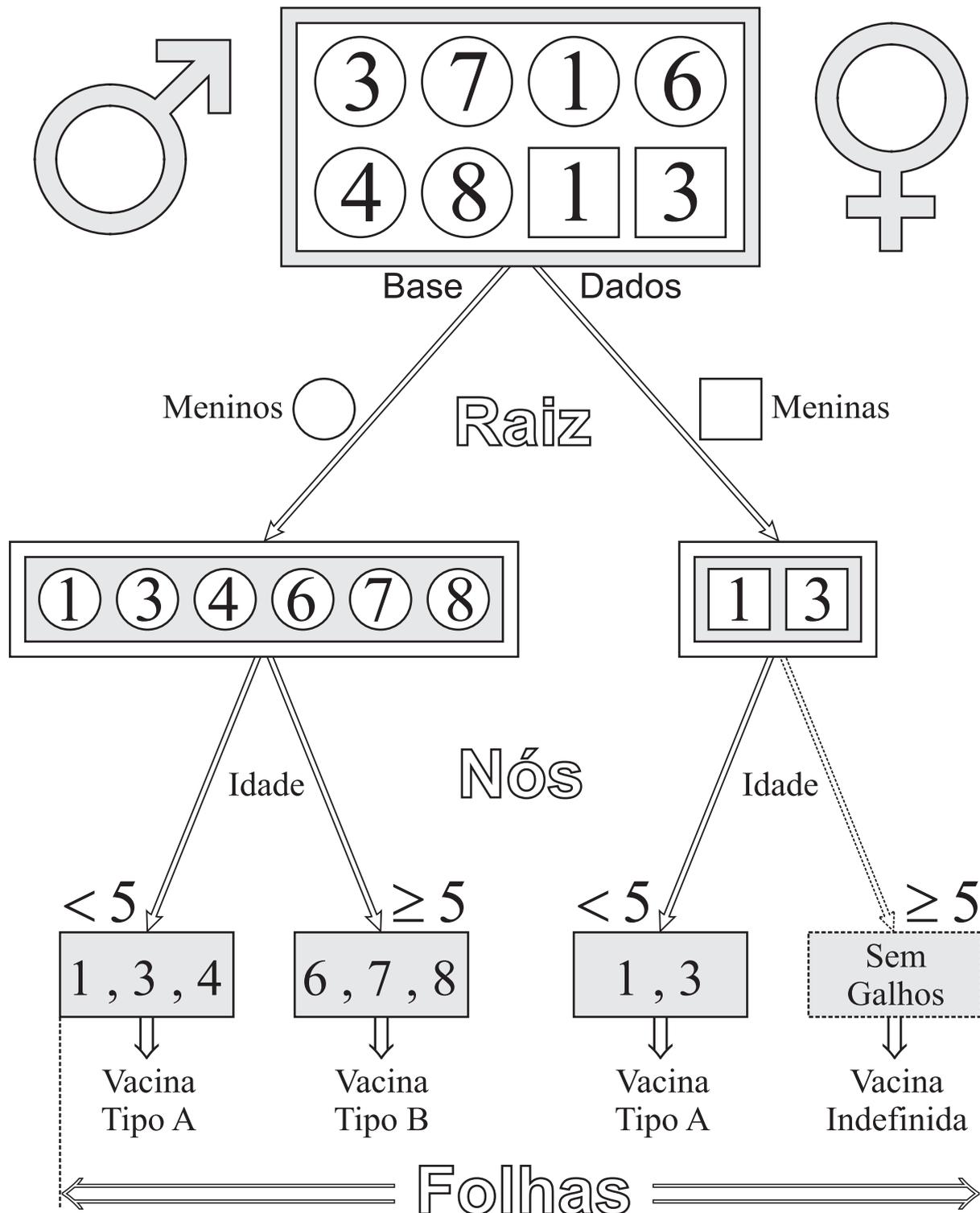


Estatística dos Percentuais Probabilísticos



2.1 Formatação da Árvore com Galhos Censitários

População de 1000 crianças monitoradas pelo Ministério da Saúde.



Árvore de Classificação Vacinal Infantil



Teste de Gosset

Médias

Variâncias **X** Parâmetros

H_0 σ H_a

Teste de Hipóteses

s/σ

Prof. Aguinaldo Prandini Ricieri



BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty:—(1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in a very large number of cases, this gives an approximation so close that a small sample will give no real information as to the manner in which the population deviates from normality: since some law of distribution must be assumed it is better to work with a curve whose area and ordinates are tabled, and whose properties are well known. This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that the deviation from normality must be very extreme to lead to serious error. We are concerned here solely with the first of these two sources of uncertainty.

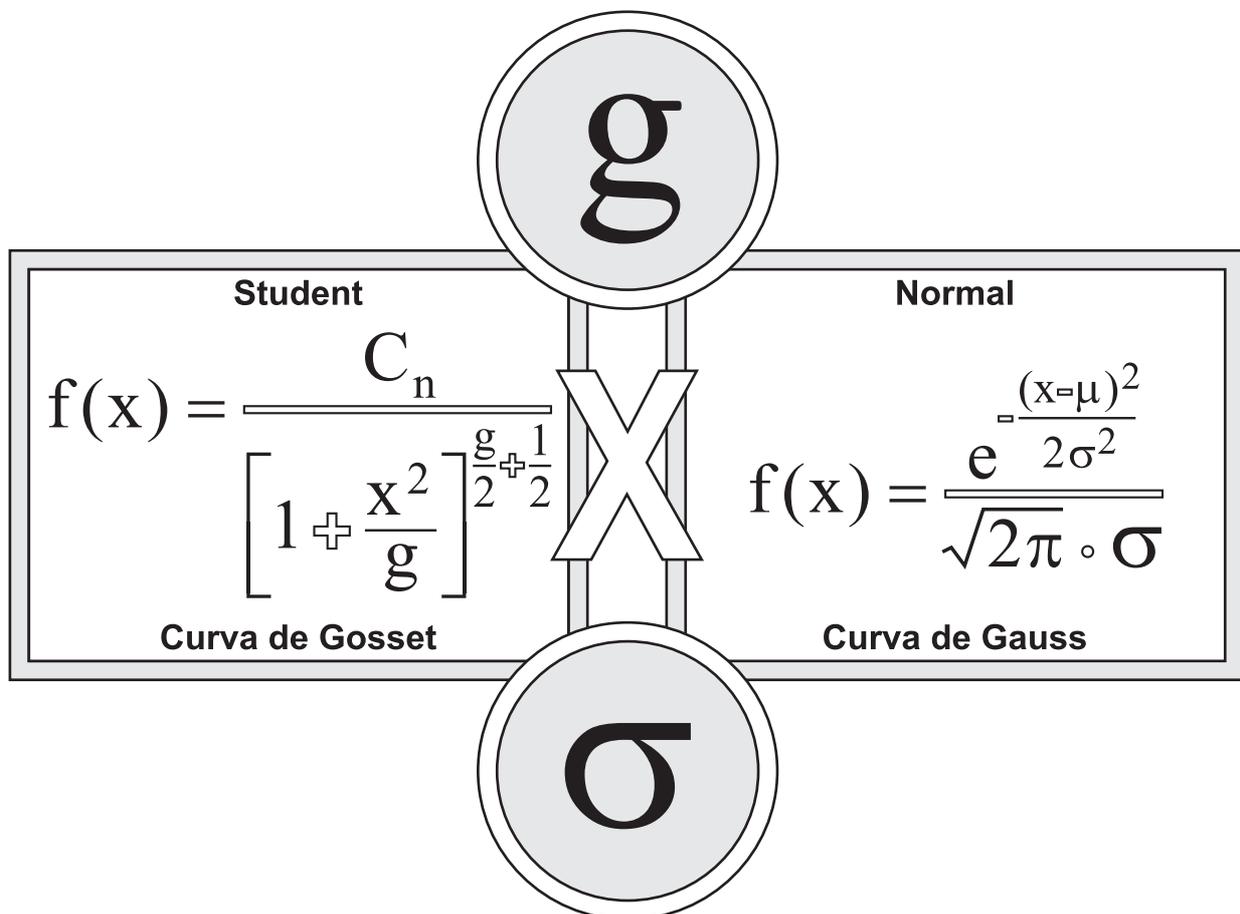
The usual method of determining the probability that the mean of the population lies within a given distance of the mean of the sample, is to assume a normal distribution about the mean of the sample with a standard deviation equal to s/\sqrt{n} , where s is the standard deviation of the sample, and to use the tables of the probability integral.



Teste de Student

1.0 Estatística Paramétrica de Gosset

O Teste t ou também nominado Teste de Student foi publicado em 1908 por William Sealy Gosset (1876 - 1937), com o pseudônimo *Student*, já que seu contrato de trabalho com a cervejaria Guinness proibia-lhe divulgar qualquer conteúdo científico de valor proveniente desse seu trabalho (segredos industriais - SI).



A distribuição t de probabilidade, caracterizada pelo parâmetro estatístico g (graus de liberdade $g = n - 1$), apresenta maior restrição que a distribuição normal quando g for pequeno (amostra pequena) e na medida que g for grande, a distribuição de probabilidade t, Student, se aproxima da curva normal, restrições iguais.

Graus de Liberdade $g = n - 1$



6.1.0 Aplicação em um Instrumento Computacional

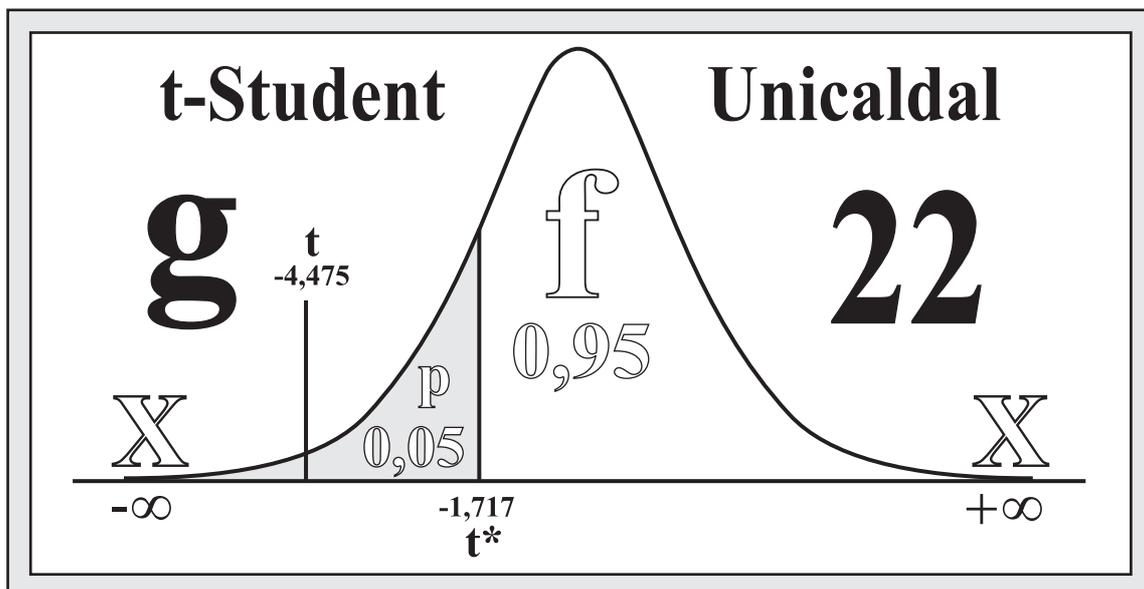
Como o tempo médio, por cirurgião cardíaco, para realizar operações, tem sido $\mu = 30$ minutos; um software foi comprado pelo hospital H = AP para *diminuir* esse tempo.

Sorteou-se então uma amostra de $n = 23$ médicos que usaram esse instrumento durante o segundo semestre de 2016 e o tempo médio da amostra obtida foi $\bar{X} = 28,04$ minutos com desvio padrão medido de $S = 2,10$ minutos.

Ao nível de significância estatístico $\alpha = 0,05$, aplicação na amostra do teste t-Student, esse instrumento COMP adquirido *diminuiu* o tempo das cirurgias realizadas?

$$\begin{array}{l} H_0 : \mu_1 = \mu_2 \text{ min} \Rightarrow H_0 : \mu = 30 \text{ min} \\ H_a : \mu_1 < \mu_2 \text{ min} \Rightarrow H_a : \mu < 30 \text{ min} \end{array}$$

$$\left. \begin{array}{l} t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{28,04 - 30,00}{\frac{2,10}{\sqrt{23}}} = -4,475 \\ t^* = \text{Tabela } (g = 22, p = 0,050) = -1,717 \end{array} \right\} t^*$$

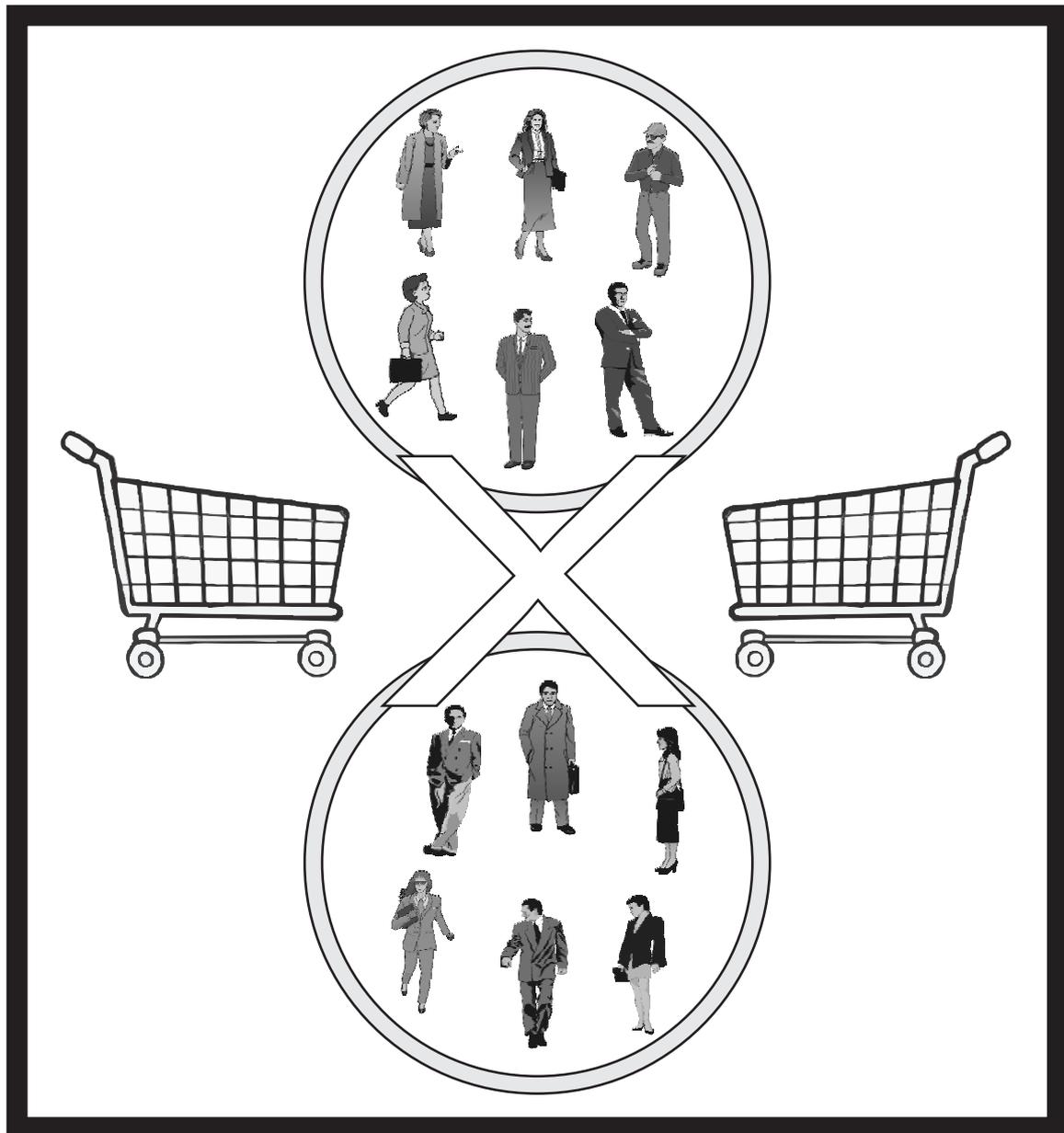


Como $(t = -4,475) < (t^* = -1,717)$, rejeita-se a hipótese H_0



PUBLIC

Efeito Publicitário



Vendas pela Internet

Teste de Gosset

Prof. Aguinaldo Prandini Ricieri



3.0 Aplicação Comparativa de Populações Distintas

Comparando: setembro (sem publicidade) e outubro (com publicidade); objetivando com o anúncio diminuir a idade dos clientes da empresa.

a₁
18
25
25
25
25
21
21
24
20
21
○ ○ ○
n ₁ = 1008
$\bar{a}_1 = 21,58$
S ₁ = 5,03
Setembro

$$S^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

S² = 5,06

$$g = n_1 + n_2 - 2$$

$$g = 1008 + 1080 - 2$$

g = 2086

$$t = \frac{\bar{a}_1 - \bar{a}_2}{\frac{S}{\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}}} = \frac{21,58 - 20,81}{\frac{2,25}{\sqrt{521,37}}}$$

t = +7,81

Ho : μ₁ = μ₂
 Ha : μ₁ ≠ μ₂
 α = 0,050
 Bicaldal : p = 0,025

t* = ± 1,961

t = +7,81 *está fora* t* = ± 1,961
 Aceita - se Ha: *Médias diferentes.*

a₂
18
24
24
23
24
20
20
23
19
21
○ ○ ○
n ₂ = 1080
$\bar{a}_2 = 20,81$
S ₂ = 5,09
Novembro



Teste de Wilcoxon

Numeração

Positiva X Negativa

H_0 SR H_a

Teste de Hipóteses

SR / SR

Prof. Aguinaldo Prandini Ricieri



INDIVIDUAL COMPARISONS BY RANKING METHODS

FRANK WILCOXON
American Cyanamid Co.

The comparison of two treatments generally falls into one of the following two categories: (a) we may have a number of replications for each of the two treatments, which are unpaired, or (b) we may have a number of paired comparisons leading to a series of differences, some of which may be positive and some negative. The appropriate methods for testing the significance of the differences of the means in these two cases are described in most of the textbooks on statistical methods.

The object of the present paper is to indicate the possibility of using ranking methods, that is, methods in which scores 1, 2, 3, . . . n are substituted for the actual numerical data, in order to obtain a rapid approximate idea of the significance of the differences in experiments of this kind.

Unpaired Experiments. The following table gives the results of fly spray tests on two preparations in terms of percentage mortality. Eight replications were run on each preparation.

Sample A		Sample B	
Percent kill	Rank	Percent kill	Rank
68	12.5	60	4
68	12.5	67	10
59	3	61	5
72	15	62	6
64	8	67	10
67	10	63	7
70	14	56	1
74	16	58	2
<hr/>		<hr/>	
Total	542	91	45
		494	

Rank numbers have been assigned to the results in order of magnitude. Where the mortality is the same in two or more tests, those tests are assigned the mean rank value. The sum of the ranks for B is 45 while for A the sum is 91. Reference to Table I shows that the probability of a total as low as 45 or lower, lies between 0.0104 and 0.021. The analysis of variance applied to these results gives an F value of 7.72, while 4.60 and 8.86 correspond

to probabilities of 0.05 and 0.01 respectively.

Paired Comparisons. An example of this type of experiment is given by Fisher (2, section 17). The experimental figures were the differences in height between cross- and self-fertilized corn plants of the same pair. There were 15 such differences as follows: 6, 8, 14, 16, 23, 24, 28, 29, 41, -48, 49, 56, 60, -67, 75. If we substitute rank numbers for these differences, we arrive at the series 1, 2, 3, 4, 5, 6, 7, 8, 9, -10, 11, 12, 13, -14, 15. The sum of the negative rank numbers is -24. Table II shows that the probability of a sum of 24 or less is between 0.019 and 0.054 for 15 pairs. Fisher gives 0.0497 for the probability in this experiment by the *t* test.

The following data were obtained in a seed treatment experiment on wheat. The data are taken from a randomized block experiment with eight replications of treatments A and B. The figures in columns two and three represent the stand of wheat.

Block	A	B	A-B	Rank
1	209	151	58	8
2	200	168	32	7
3	177	147	30	6
4	169	164	5	1
5	159	166	-7	-3
6	169	163	6	2
7	187	176	11	5
8	198	188	10	4

The fourth column gives the differences and the fifth column the corresponding rank numbers. The sum of the negative rank numbers is -3. Table II shows that the total 3 indicates a probability between 0.024 and 0.055 that these treatments do not differ. Analysis of variance leads to a least significant difference of 14.2 between the means of two treatments for 19:1 odds, while the difference between the means of A and B was 17.9. Thus it appears that with only 8 pairs this method is capable of giving quite accurate information about the significance of differences of the means.

Discussion. The limitations and advantages of ranking methods have been discussed by Fried-

Individual Comparisons... Biometrics Bulletin, 1945, Frank Wilcoxon (1892 - 1965).



Estatística de Wilcoxon

1.0 Teste de Hipótese Não Paramétrico

O Teste de Wilcoxon enquanto não paramétrico (não se conhece a variância populacional) permite a comparação de duas amostras de dados pareados: medidas dependentes (MD) feitas antes e feitas depois nos mesmos indivíduos.

Individual Comparisons by Ranking Methods
Biometrics Bulletin 1, (6), p. 80, 1945
 Frank Wilcoxon (1892 - 1965)

H₀

Ho: Amostras iguais
There is no difference in the data

Ha: Amostras desiguais
There is a difference in the data

H_a

Esse teste de hipóteses tem como fundamento a soma dos rankings (R) das diferenças absolutas dos seus elementos pareados ($X_A - X_B$), associados a uma numeração (NUM), sinalizadas (S) com mais (SR^+) e menos (SR^-), soma submetida ao operador Minimum (Min), como indica o próximo esquema:

$$\text{Min} \left[\left| \sum (SR)^+ \right|, \left| \sum (SR)^- \right| \right]$$

Esse valor do Min, comparado com os valores críticos (W^*), obtidas, com certo intervalo de confiança, permite as próximas duas conclusões:

$$\text{Min} \left[\left| \sum (SR)^+ \right|, \left| \sum (SR)^- \right| \right] > W^* \Rightarrow H_0$$

$$\text{Min} \left[\left| \sum (SR)^+ \right|, \left| \sum (SR)^- \right| \right] < W^* \Rightarrow H_a$$

Populações Iguais (H₀) X (H_a) Populações Diferentes



Teste de Pearson

Frequência

Registrada **X** Esperada

H_0 H_a

Teste de Hipóteses

Q/Q

Prof. Aguinaldo Prandini Ricieri

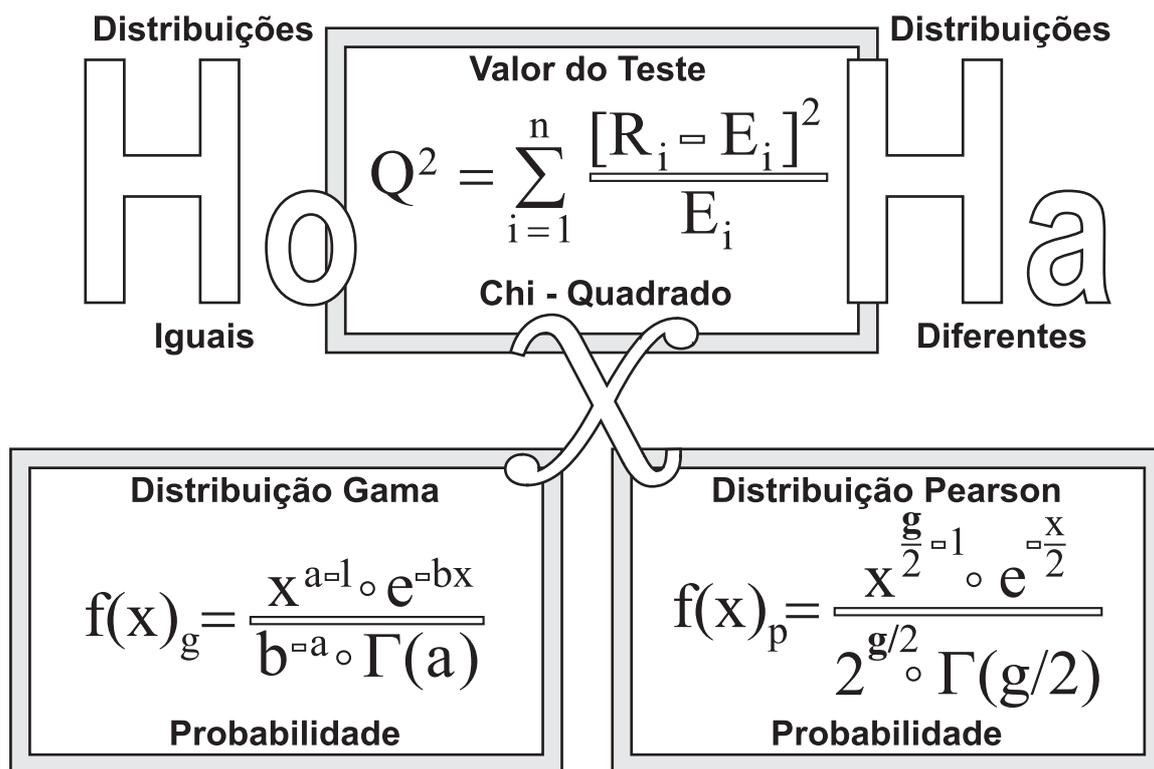


Estatística de Pearson

1.0 Teste de Qualidade das Distribuições

Dentre as inúmeras aplicações estatísticas do Q^2 destaca-se o importante e simples teste da qualidade de dados, que permite analisar, conhecer e distinguir: se a distribuição de frequências registradas (R) difere da frequência esperada (E).

Aceita-se a hipótese alternativa (H_a) – há diferença entre as distribuições –, se o valor do teste exceder o valor crítico (x^*) da distribuição Q^2 ; caso contrário, fica-se com a hipótese (H_o), isto é, não há diferença entre as distribuições analisadas.



Para $a = g/2$, $b = 1/2$, $x > 0 \Rightarrow F(x)_g = F(x)_p$

On the criterion that a given system of derivations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling

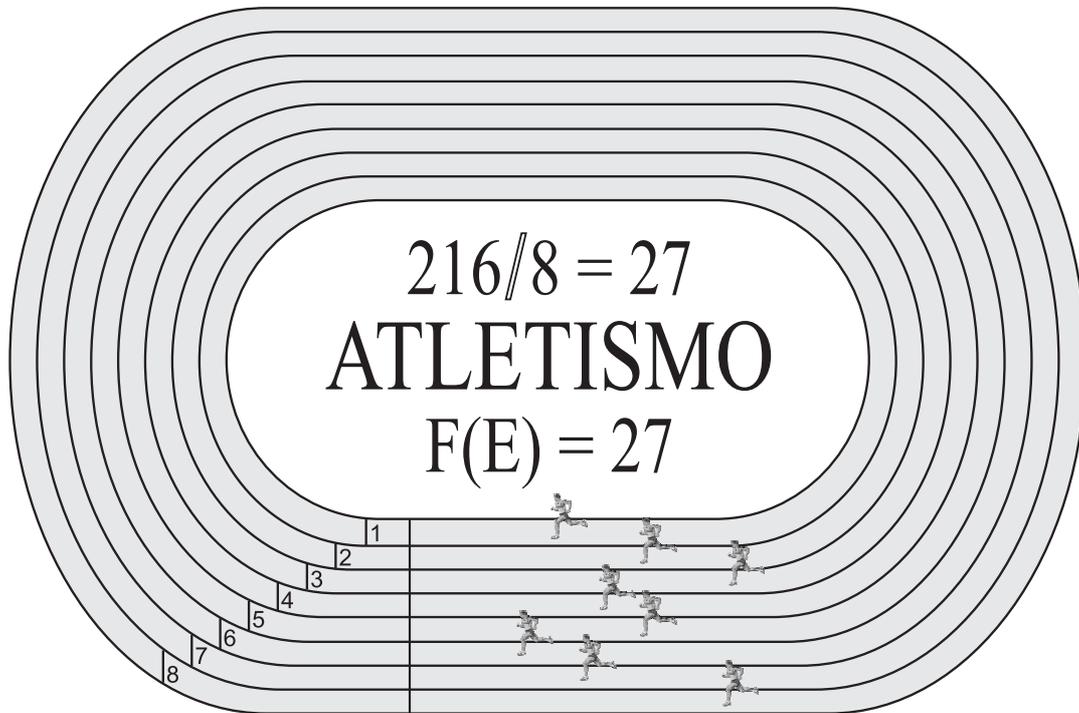
Philosophical Magazine, 50 (302), p. 157, 1900

Karl Pearson (1857 - 1936)



8.0 Efeitos Psicológicos das Raias no Atletismo ($p = 0,05$, $g = 7$)

Em corridas dos 800 m é comum acreditarem que a pista de raio 1 (interna) – por ser mais curta –, proporciona mais vitórias que a pista de raio 2 (externa) – por ser mais longa: cada corredor sai de um local diferente nas 8 raias indicadas: 1, ..., 8.



$g = 7$	Numeração das Raias da Pista								FE = 27
	1	2	3	4	5	6	7	8	
Vitórias	40	27	18	31	24	23	25	28	216

Vitórias Registradas F(R)

$$Q^2 = \frac{[40 - 27]^2}{27} + \frac{[27 - 27]^2}{27} + \frac{[18 - 27]^2}{27} + \dots + \frac{[28 - 27]^2}{27} = 10,96$$

$$Q^2 = 10,96 < x^* = 14,067 \Rightarrow H_0$$

Rejeita-se a hipótese H_a de que as raias são diferentes!

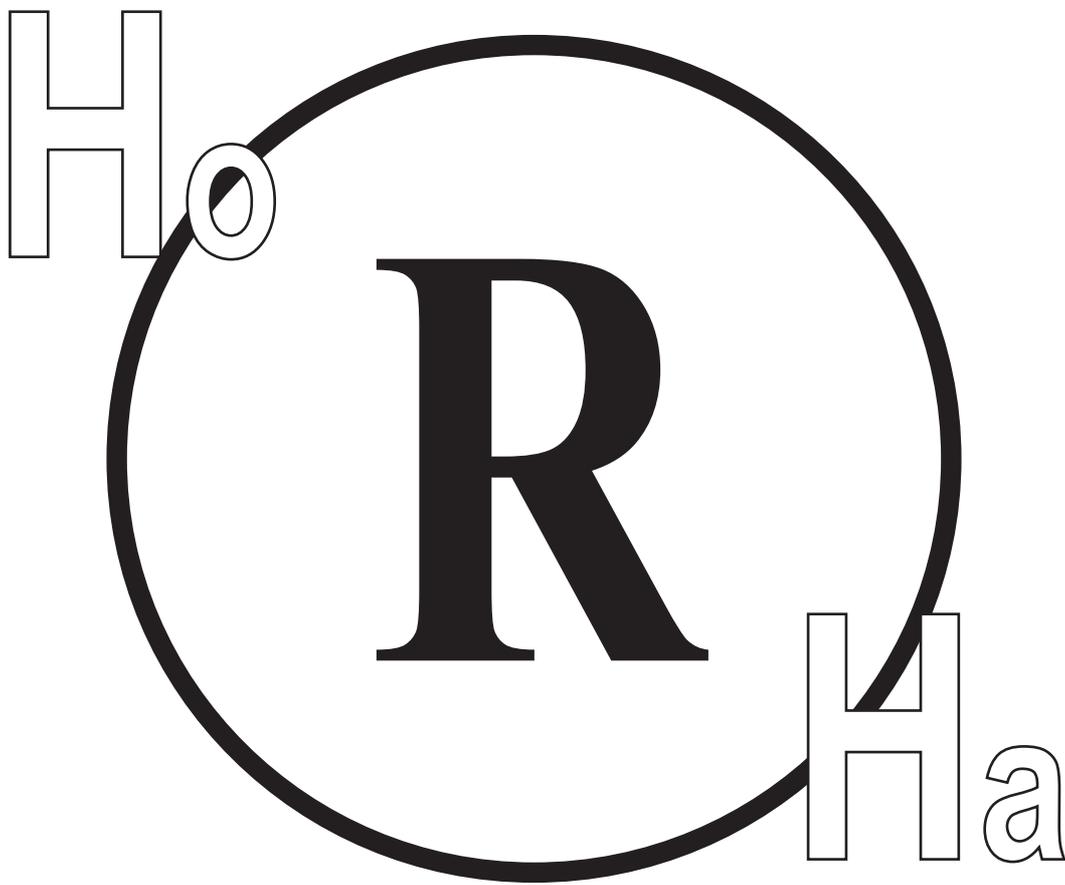
Análise para o Valor Crítico Tabelado $x^* = 14,067$



Teste de Friedman

Ranque

Registrado **X** Esperado



Teste de Hipóteses

R / R

Prof. Aguinaldo Prandini Ricieri



THE USE OF RANKS TO AVOID THE ASSUMPTION OF NORMALITY IMPLICIT IN THE ANALYSIS OF VARIANCE

BY MILTON FRIEDMAN
National Resources Committee

MOST projects involving the collection and analysis of statistical data have for one of their major aims the isolation of factors which account for variation in the variable studied. The statistical tool ordinarily employed for this purpose is the analysis of variance. Frequently, however, the data are sufficiently extensive to indicate that the assumptions necessary for the valid application of this technique are not justified. This is especially apt to be the case with social and economic data where the normal distribution is likely to be the exception rather than the rule. This difficulty can be obviated, however, by arranging each set of values of the variate in order of size, numbering them 1, 2, and so forth, and using these ranks instead of the original quantitative values. In this way no assumption whatsoever need be made as to the distribution of the original variate.

The utilization of ranked data is thus frequently a desirable device to avoid normality assumptions; in addition, however, it may be inescapable either because the data available relate solely to order, or because we are dealing with a qualitative characteristic which can be ranked but not measured.

The possibility of using ranked data in problems involving simple correlation and thereby avoiding assumptions of normality has recently been emphasized in an article by Harold Hotelling and Margaret Richards Pabst.¹ It is the purpose of the present article to outline a procedure whereby the analysis of ranked data can be employed in place of the ordinary analysis of variance when there are two (or more) criteria of classification. This procedure has two major advantages. As already indicated, it is applicable to a wider class of cases than the ordinary analysis of variance. In addition, it is less arduous than the latter technique, requiring but a fraction as much time. The loss of information through utilizing the procedure outlined below when the analysis of variance could validly be applied may thus be more than compensated for by its greater economy. This consideration is likely to be especially important with those large scale collections of social and economic data which have become increasingly frequent in recent years and for which the funds available for analysis are limited.

¹ "Rank Correlation and Tests of Significance Involving No Assumption of Normality," *Annals of Mathematical Statistics*, VII (1936) 29-43.



6.0 Ingestão de Cloridrato de Piridoxina/Cabelo

L	K1	K2	K3	R1	R2	R3	g	p = 0,05	p = 0,01
L01	1,0	2,0	3,0	1	2	3	01	3,841	6,635
L02	2,0	4,0	6,0	1	2	3	02	5,992	9,210
L03	3,0	5,0	7,0	1	2	3	03	7,815	11,345
L04	5,0	9,0	8,0	1	3	2	04	9,488	13,277
L05	4,0	6,0	8,0	1	2	3	05	11,071	15,086
L06	3,0	5,0	7,0	1	2	3	06	12,592	16,812
L07	1,0	2,0	3,0	1	2	3	07	14,067	18,475
L08	3,0	4,0	9,0	1	2	3	08	15,507	20,090
L09	4,0	5,0	6,0	1	2	3	09	16,919	21,666
L10	7,0	8,0	9,0	1	2	3	10	18,307	23,209
L11	8,0	6,0	9,0	2	1	3	11	19,675	24,725
L12	6,0	7,0	8,0	1	2	3	12	21,026	26,217
L13	1,0	2,0	3,0	1	2	3	13	22,362	27,688
L14	5,0	6,0	7,0	1	2	3	14	23,685	29,141
L15	4,0	8,0	9,0	1	2	3	15	24,996	30,578
L16	2,0	3,0	4,0	1	2	3	16	26,296	32,000
L17	5,0	6,0	7,0	1	2	3	17	27,587	33,409
L18	5,0	8,0	9,0	1	2	3	18	28,869	34,805
L19	3,0	7,0	8,0	1	2	3	19	30,144	36,191
L20	4,0	5,0	2,0	2	3	1	20	31,410	37,566
L21	3,0	1,0	4,0	2	1	3	21	32,671	38,932
L22	7,0	3,0	4,0	3	1	2	22	33,924	40,289
L23	8,0	5,0	2,0	3	2	1	23	35,172	41,638
L24	9,0	3,0	2,0	3	2	1	24	36,415	42,980
L25	2,0	7,0	6,0	1	3	2	25	37,652	44,314
L26	4,0	9,0	8,0	1	3	2	26	38,885	45,642
L27	6,0	3,0	2,0	3	2	1	27	40,113	46,963
L28	8,0	7,0	6,0	3	2	1	28	41,337	48,278
L29	9,0	3,0	4,0	3	1	2	29	42,557	49,588
L30	2,0	1,0	8,0	2	1	3	30	43,773	50,892
MED	4,467	5,000	5,933	1,533	1,967	2,500	TABELA QQ		

A ingestão de vitamina B6 com a finalidade de vigorar o cabelo, foi medido o comprimento dos fios, mês a mês (K1, K2 e K3), de trinta homens (40 anos).

$$F = 30 \cdot \{ [1,533 - 2]^2 + [1,967 - 2]^2 + [2,500 - 2]^2 \} = 14,067$$

A nível de $p = 0,05$ e $p = 0,01$, para $g = K = 1 = 2$, conclui-se:

$F > p$: nega-se a hipótese H_0 (são amostras diferentes).