

## CONJUNTOS, MATRIZES E PROBABILIDADES

### 6.<sup>a</sup> Sessão — A Optimização da Informação (1)

por LUÍS ALVES

Professor do Instituto Superior Técnico

RESUMO: Pretende-se mostrar a importância fundamental que a informação cientificamente organizada, correcta e actualizada tem na resolução dos problemas em qualquer campo da actividade.

The author tries to show the fundamental importance that information scientifically organized, accurate and up to date has in the solution of the problems in any field of activity.

Na 2.<sup>a</sup> Sessão, que dedicámos à «Anatomia» da Lógica, lembrou-se o conceito de «pessoa inteligente», o qual se definiu como sendo aquela que é capaz de:

- I — Reunir toda a informação necessária para o equacionamento correcto dos problemas;
- II — Utilizar uma técnica correcta no manuseamento dessa informação.

Nessa sessão dedicámo-nos ao 2.<sup>o</sup> ponto; nesta, vamos tratar do primeiro, o que aliás constitui um problema fundamental dos tempos modernos.

Na 5.<sup>a</sup> Sessão de «O Diálogo com os Computadores» expôs-se com relativo pormenor o problema da informação, tendo-se apresentado o conceito de quantidade de informação.

Afirmou-se aí que a quantidade de informação pode ser expressa através da seguinte expressão:

$$H = - \sum_{i=1}^n p_i \log_2 p_i$$

na qual  $p_i$  representa a probabilidade  $i$  de um acontecimento, sendo

$$\sum_{i=1}^n p_i = 1,$$

desde que se considerem as probabilidades de todos os acontecimentos possíveis.

(1) Lição proferida em 25 de Março do ano corrente e integrada num Curso para o pessoal da CUF, sobre «Conjuntos, Matrizes e Probabilidades».

Suponhamos, por exemplo, que temos uma parte de um baralho de cartas com 6 copas, 3 paus, 5 ouros e 8 espadas.

Suponhamos ainda que se tira uma carta ao acaso e se pergunta que naipe é que saiu.

Admitindo que a resposta foi copas, pretende-se determinar qual foi a quantidade de informação que se deu.

Será:

$$\begin{aligned}
 H &= - \sum_{i=1}^4 p_i \log_2 p_i = \\
 &= - [p_1 \log_2 p_1 + p_2 \log_2 p_2 + p_3 \log_2 p_3 + p_4 \log_2 p_4] = \\
 &= - [0,28 \log_2 0,28 + 0,14 \log_2 0,14 + 0,22 \log_2 0,22 + 0,36 \log_2 0,36] = \\
 &= - [0,28 \times 1,78 - 0,14 \times 2,83 - 0,22 \times 2,20 - 0,36 \times 1,48] = \\
 &= 0,50 + 0,40 + 0,48 + 0,53 = 1,91 \text{ bits}
 \end{aligned}$$

Estas contas foram feitas à tração e por isso vamos esclarecer o que fizemos.

Em primeiro lugar, as probabilidades  $p_1$ ,  $p_2$ ,  $p_3$  e  $p_4$  são as probabilidades de sair respectivamente uma copa, um pau, um ouro ou uma espada.

Como o número total de cartas é 22, é evidente que as probabilidades de sair uma carta de cada um dos naipes serão evidentemente  $p_1 = \frac{6}{22} = 0,28$ ,  $p_2 = \frac{3}{22} = 0,14$ ,  $p_3 = \frac{5}{22} = 0,22$  e  $p_4 = \frac{8}{22} = 0,36$ .

Em segundo lugar, os logaritmos na base 2 calculam-se a partir dos logaritmos na base 10 multiplicando-os por 3,32, sendo  $\log 0,28 = \bar{1},464 = -0,536$ ,  $\log 0,14 = \bar{1},146 = -0,854$ ,  $\log 0,22 = \bar{1},342 = -0,658$  e  $\log 0,36 = \bar{1},555 = -0,445$  (e  $0,536 \times 3,32 = 1,78$ ,  $0,854 \times 3,32 = 2,83$ ,  $0,658 \times 3,32 = 2,20$  e  $0,445 \times 3,32 = 1,48$ ).

Em terceiro lugar, o resultado vem expresso em bits, porque é o nome que se dá à unidade de informação e que vem de *binary digit*.

Essa unidade corresponde à informação colhida quando se lança a «cara» ou «coroa». Neste caso as probabilidades  $p_1$  e  $p_2$ , se sair cara ou coroa, são iguais, isto é,

$$p = p_1 = p_2 = \frac{1}{2}$$

e, nestas condições, será:

$$\begin{aligned}
 H &= - \sum_{i=1}^2 p_i \log_2 p_i = - [p_1 \log_2 p_1 + p_2 \log_2 p_2] = \\
 &= - 2 p \log_2 p = - 2 \frac{1}{2} \log_2 \frac{1}{2} = \log_2 2 = 1
 \end{aligned}$$

Este conceito de quantidade de informação mostra que essa quantidade é tanto menor quanto mais se conhece o que vai acontecer.

Por exemplo, se nós tivermos um conjunto de 22 cartas que sabemos só ter copas e formos informados de que se tirou uma copa, a quantidade de informação que recebemos é zero. E, se fizermos o cálculo, chegamos ao mesmo resultado.

Com efeito, como a probabilidade de sair uma copa é  $p = 1$ , a quantidade de informação é

$$H = - \sum_{i=1}^1 p_i \log_2 p_i = - p \log_2 p = - 1 \log_2 1 = 0$$

E bate tudo certo.

Se calhar não bate tão certo como isso, como vamos ver a seguir.

Suponhamos que a informação relativa ao naipe que tinha saído era fornecida dentro de uma determinada orgânica de trabalho e, por consequência, tinha custado dinheiro.

E vamos supor que, com todas as alcavalas indispensáveis, essa informação tinha custado 14\$00, o que não se pode dizer pròpriamente que tenha sido «meter a unha».

Acontece então que no caso em que os naipes eram diversificados e em que se receberam 1,91 bits de informação, o preço por bit tinha sido 7\$30; no caso em que só havia copas e em que portanto a informação foi 0, o preço do bit foi  $\infty$ , o que é meter uma unha de perissodáctilo.

Ora, isto é um problema muito sério, porque quando se fala no preço da informação, não se tem em vista, na maioria dos casos, o número de bits que contém e apenas se explora o seu valor absoluto.

Mas o que é certo é que se uma informação custou 50 contos, mas contiver 50 000 bits de informação, custa 1\$00 por bit, mas os 50 contos causam muito mais escândalo do que a informação de 14\$00 sobre o naipe do baralho, que custou 7\$30 por bit.

É claro que a medição do número de bits de uma informação não é um problema muito fácil, mas o que é verdade é que é preferível uma informação cara mas útil a uma informação barata mas inútil.

Também se pode dizer que uma informação útil e barata é o ideal, mas o «amigo Banana» já dizia o mesmo.

Tentemos analisar este problema com todo o realismo e vamos por agora circunscrever-nos à literatura técnica publicada.

Pode ser que nem todos os presentes saibam que existe um manual internacional que tem o nome de «Ulrich's International Periodicals Directory», que apresenta uma lista das revistas que se publicam em todo o Mundo sobre os diferentes ramos de conhecimento.

O número dessas revistas resume-se como se indica a seguir.

Tendo em linha de conta as revistas que não fazem parte desse manual, calcula-se que o número total de revistas publicadas é da ordem de grandeza de 50 000 — 60 000.

Admitindo que o preço médio de uma revista é de 500\$00/ano, a aquisição de todas as revistas publicadas custaria à volta de 30 000 contos.

É evidente que nenhuma empresa vai comprar as 60 000 revistas que se publicam, mas isso não simplifica totalmente a questão.

Consideremos por exemplo uma empresa que se dedique a indústrias da alimentação.

### 1 — Revistas científicas técnicas e médicas

— Abstracts . . . . .	490	
— Aeronáutica e Astronáutica . . . . .	270	
— Agricultura . . . . .	550	
— Arquitectura e Construção . . . . .	500	
— Energia Atómica . . . . .	100	
— Bioquímica e Biologia . . . . .	400	
— Botânica e Zoologia . . . . .	430	
— Química e Engenharia Química . . . . .	420	
— Electricidade e Eng. <sup>a</sup> Electrónica . . . . .	210	
— Engenharia Mecânica . . . . .	570	
— Pesca e Alimentação . . . . .	210	
— Geologia e Geofísica . . . . .	240	
— Matemática e Física . . . . .	500	
— Ciências Médicas e Farmacêuticas . . . . .	1 900	
— Metalurgia e Minas . . . . .	400	
— Ciências Naturais . . . . .	470	
— Patentes e Marcas Comerciais . . . . .	90	
— Papel, Florestas e Madeira . . . . .	150	
— Petróleo . . . . .	100	
— Plásticos e Têxteis . . . . .	250	
— Psicologia . . . . .	220	
— Navios e Construção Naval . . . . .	130	
— Outras . . . . .	3 400	12 000

**2 — Artes, Humanidades, Negócios e Ciências Sociais**

— Abstracts . . . . .	100	
— Bancos e Finanças . . . . .	260	
— Negócios e Indústria . . . . .	600	
— Economia . . . . .	400	
— Educação . . . . .	640	
— Trabalho e Relações Industriais . . . . .	250	
— Legislação. . . . .	410	
— Ciências Políticas . . . . .	430	
— Sociologia . . . . .	310	
— Outras . . . . .	12 600	16 000
Total . . . . .		<u>28 000</u>

Se comprar as 210 revistas que são mencionadas na lista no domínio da «Pesca e Alimentação», pode afirmar de forma simplista que, se estiver disposta a gastar 100 contos por ano, fica com conhecimento total de tudo o que se publica sobre o seu ramo de actividade.

Uma afirmação deste género seria profundamente falsa porque, admitindo mesmo que tudo o que interessa conhecer se encontra nas revistas (o que evidentemente não é verdade), uma revista só por si não dá qualquer conhecimento a não ser que se dêem duas condições:

a) — Que seja lida

b) — Que os resultados da leitura sejam utilizados

Ora, o preço destas duas operações é muito mais elevado do que o valor das próprias revistas e, por isso, o aproveitamento de todas as revistas publicadas num dado domínio traduz-se num custo que se torna inoportável.

Nestas condições, qualquer gestor tem de fazer face ao problema de seleccionar as publicações que deve comprar com a certeza antecipada de que irá perder informação por não as poder comprar todas.

Esta situação tem um «approach» científico através da chamada Lei de Bradford.

Esta lei estabelece que há uma relação logarítmica entre o número  $n$  de revistas e o número  $N$  de artigos úteis publicados sobre um determinado assunto, isto é

$$N = A \log n,$$

sendo  $A$  um coeficiente experimental que depende do tipo do assunto.

Se nós fixarmos o valor  $n_t$  do número total das publicações existentes (com o número  $N_t$  de artigos) e dispusermos de uma percentagem  $P$  deste número total (com um número  $N_p$  de artigos), a percentagem  $P_1$  de artigos de que podemos dispor será evidentemente:

$$P_1 = \frac{N_p}{N_t} = \frac{\log P n_t}{\log n_t} = \frac{\log n}{\log n_t}$$

sendo  $n$  o número de publicações adquiridas.

Tomando por exemplo  $n_t = 10\,000$  e  $P = 10\%$ , será  $n = 0,1 \times 10\,000 = 1\,000$  e  $P_1 = \frac{\log 1\,000}{\log 10\,000} = \frac{3}{4} = 75\%$  o que significa que apenas em  $10\%$  das revistas publicadas nós obtemos  $75\%$  dos artigos de interesse.

No quadro I indicam-se os valores  $P_1$  correspondentes a vários valores de  $n_t$  e de  $P$ .

QUADRO I  
Percentagens de artigos úteis

$n_t \backslash P$	5%	10%	15%	20%	25%	30%
50	23%	41%	52%	59%	64%	69%
100	35%	50%	58%	65%	70%	74%
1 000	56%	67%	72%	77%	80%	83%
10 000	64%	75%	79%	82%	85%	87%

Deve notar-se que a Lei de Bradford só é válida quando as publicações são classificadas de acordo com uma ordem de relevância especificada pelo utilizador.

Os valores do Quadro I podem ser entendidos em termos de probabilidade.

Por exemplo, se o número total de publicações existentes sobre um dado assunto for  $n_t = 100$  e nós comprarmos uma percentagem  $P = 0,30$ , temos uma probabilidade  $p_1 = 0,74$  de encontrarmos todos os artigos úteis ou seja uma percentagem  $p_2 = 0,26$  de perder artigos úteis.

E então poderá pôr-se uma pergunta curiosa que é a de saber qual o número óptimo de publicações  $n_{op}$  que se deve comprar sobre um determinado assunto.

Para respondermos a essa pergunta, podemos dizer que  $n_{op}$  é o número que torna mínima a seguinte expressão:

$$C_T = C_m n + C_p N_p$$

sendo

$C_m$  = preço médio de uma publicação por ano.

$C_p$  = custo médio dos artigos perdidos por ano.

$N_p$  — número de artigos perdidos por ano.

Mas,

$$\begin{aligned} N_p &= N_t (1 - P_1) = A \log n_t (1 - P_1) = \\ &= A \log n_t \left( 1 - \frac{\log n}{\log n_t} \right) = A \log n_t - A \log n = \\ &= A \log \frac{n_t}{n} = \frac{A}{2,3} 1n \frac{n_t}{n} = -\frac{A}{2,3} 1n \frac{n}{n_t} \end{aligned}$$

sendo  $1n$  o logaritmo neperiano.

Substituindo, vem

$$C_T = C_m n - \frac{A C_p}{2,3} \log \frac{n}{n_t}$$

O mínimo desta expressão obtém-se derivando  $C_T$  em ordem a  $n$  e igualando a zero.

Virá:

$$\frac{d C_T}{d n} = C_m \frac{A C_p}{2,3} \cdot \frac{1}{n} \cdot \frac{1}{n_t} = C_m - \frac{A C_p}{2,3n} = 0$$

O valor óptimo será pois

$$n = n_{op} = \frac{A C_p}{2,3 C_m} = \frac{A R}{2,3} = B r$$

sendo

$$B = \frac{A}{2,3} \text{ e } r = \frac{C_p}{C_m}$$

O raciocínio é perfeitamente simples, mas se o quisermos concretizar com um exemplo numérico, ficamos imediatamente «entalados» porque não temos uma ideia intuitiva da ordem de grandeza dos valores  $A$ ,  $C_p$  e  $C_m$ .

Por isso, vamos simular uma situação com valores arbitrados, o que nos dará uma base para uma análise crítica subsequente.

Seja então

$$A = 50$$

$$C_p = 2\,000\$00$$

$$C_m = 3\,000\$00$$

Virá

$$n_{op} = \frac{50 \times 2\,000}{2,3 \times 3\,000} = 15 \text{ publicações}$$

A sensação que resulta desta aplicação é a de que o resultado é extraordinariamente frágil porque os valores de partida não inspiram qualquer confiança.

Em todo o caso, como a fórmula apresentada constitui o modelo matemático que descreve a situação de um gestor em relação à utilização da informação existente, não nos fica bem camuflar o problema, pelo que vamos «radiografar» os três parâmetros  $A$ ,  $C_m$  e  $C_p$ , que estão na base da optimização da informação e que representam respectivamente:

- I — A relevância da informação utilizada
- II — O custo de obtenção da informação
- III — O custo da informação perdida

A individualização destes três parâmetros pode parecer, até certo ponto, insólita às pessoas que têm frequentado cursos de organização, de gestão e de formação, porque não é vulgar fazerem parte do âmbito desses cursos.

No entanto, isto não é muito estranho, dado que é costume partir do pressuposto de que a definição dos dados é um problema que se considera resolvido e, por isso, toda a ênfase dos assuntos a tratar se faz incidir sobre o seu manuseamento.

Tal pressuposto está porém muito longe de ser correcto em muitos casos, porque o nosso País sofre de uma herança histórica que complica a resolução de muitos dos nossos problemas.

Com efeito, por razões que não interessa aprofundar, mas que constituem uma realidade, Portugal não acompanhou o movimento científico do Séc. XIX e do princípio do Século XX, pelo que não se ajustou à curva de evolução da maioria dos países europeus.

Basta dizer-se, como documentação desta afirmação, que há 30 anos se punha a ridículo quem pretendesse estudar qualquer problema técnico em bases científicas, saindo das soluções tradicionais assentes na experiência adquirida.

Daí resultou que, por vezes, nos encontramos na situação da «sopa de pedra» que se traduz em algumas pessoas se considerarem evoluídas apenas por aprenderem a «fazer a sopa», esquecendo-se das batatas, das couves e dos feijões.

Por outras palavras, aprende-se a organizar, a gerir, a planificar e a coordenar, sem se ter em linha de conta que estes três verbos são verbos transitivos e que, portanto, não funcionam sem complemento directo.

Isto explica a razão pela qual muitas tentativas falham ao nível nacional apenas porque não há o abastecimento conveniente em «complementos directos».

Ora, o que se pretende nesta sessão é precisamente falar nos «complementos directos».

Talvez seja, de certo modo, jogar contra a «corrente do jogo», mas o que é certo é que, apesar de muito boa vontade, ainda não se conseguiu que a «sopa de pedra» constituísse um prato turístico do «Abril au Portugal».

Este problema dos complementos directos já originou, com outro objectivo, um conjunto de considerações que se transcrevem parcialmente a seguir:

«1.º — O robustecimento dos complementos directos é um problema difícil porque obriga à criação de duas formas de mentalidade que, de certo modo, estão em contradição com o conceito clássico do empregado cumpridor:

I — O da imperfeição da nossa perfeição

II — O da ignorância da nossa ciência.

2.º — A necessidade de aquisição desta forma de mentalidade é uma das consequências de a indústria moderna ser um sistema em equilíbrio dinâmico e não em equilíbrio estático, como tradicionalmente era considerada.

E o aspecto circunscrito da organização apenas abarca o equilíbrio estático.

3.º — O conceito de equilíbrio estático é o que melhor se coaduna com a formação universitária tradicional que se baseia numa lógica dedutiva, na qual as consequências resultam naturalmente dos postulados de partida.

No aspecto psicológico, este tipo de lógica conduz à perfeição e à segurança, porque uma vez que os postulados sejam respeitados, as consequências vêm necessariamente correctas.

4.º — O conceito de equilíbrio dinâmico é incompatível com uma lógica de tipo dedutivo, porque como o progresso se faz no domínio do desconhecido, os raciocínios têm que ser do tipo indutivo partindo de hipóteses em vez de postulados.

E as hipóteses arrastam consigo iterações sucessivas que têm de ser conduzidas no máximo de quantidades e no mínimo de tempo.

No aspecto psicológico, perde-se o sentido da perfeição e da segurança.

5.º — No equilíbrio estático, as decisões são criadas de fora para dentro; no equilíbrio dinâmico, as decisões são criadas de dentro para fora. No primeiro caso, resultam de situações que aparecem no segundo, de situações que se provocam. Por outras palavras,

no primeiro caso a atitude de espírito é a de reduzir ao mínimo as decisões a tomar; no segundo, a de ter oportunidade de tomar o maior número de decisões possível.

6.º — No equilíbrio estático, não haver necessidade de tomar decisões é um índice de perfeição de funcionamento; no equilíbrio dinâmico, a ausência de decisões significa incapacidade de progresso.

7.º — O progresso das empresas resulta, portanto, da matéria prima para tomar decisões; mas essa matéria prima não aparece espontaneamente, mas tem de ser criada.

Deste modo, o problema das empresas modernas está na abundância da matéria prima para tomar decisões para o futuro.

8.º — Por outro lado, a matéria prima para as decisões tem de ser procurada com avidez e sem descanso, para não se correr o risco de a concorrência se antecipar nas decisões que gostaríamos de ter tomado.

9.º — É neste aspecto que se situa a grande batalha das empresas do mundo de hoje, na procura incessante de novos produtos e de novas tecnologias.

10.º — A maior dificuldade na resolução do problema reside na formação de lógica dedutiva a que atrás se aludiu, que cria em um grande número de pessoas a aversão de fazer coisas inúteis, considerando-se, neste caso, como coisa inútil a constituição de uma base de trabalho sem um objectivo directamente definido.

Simplemente, para fazer progresso não é possível esperar apenas «as ocasiões de golo feito», sendo necessário «rematar muito à baliza» para tentar «surpreender a barreira defensiva» do desconhecido.

11.º — «Rematar à baliza», neste caso, significa uma auto-disciplina de criação de ideias novas, de «brainstorming» para a geração de ideias, de aprofundamento da informação existente sobre as ideias geradas, de disseminação rápida da informação publicada e de investigação sobre a viabilidade das ideias incipientes».

A transcrição destes 11 pontos pode parecer à primeira vista «conversa fiada». Mas não é! E não é, porque toda ela é traduzível em regras de actuação.

Aliás, este problema das regras é um problema fundamental, como já se fez realçar na sessão anterior porque, na realidade, é totalmente deprimente assistir-se a qualquer debate sobre a filosofia da conduta profissional e ir-se para casa apenas com esta conclusão: «Aquilo foi muito interessante!»

E para ficarmos com a certeza de que o texto transcrito se pode transformar em regras, vamos analisar em pormenor os três parâmetros que condicionam o problema da informação dos gestores.

Comecemos pela relevância da informação e o parâmetro *A*.

É evidente que o tratamento deste ponto implica que se comece pelo princípio, isto é, pela identificação das fontes de informação à disposição.

Parece à primeira vista que é um «*approach*» muito infantil, mas se não seguirmos por este caminho, voltamos a cair no problema da «*sopa de pedra*».

Ora, as fontes de informação podem agrupar-se em 26 tipos diferentes:

- 1 — Jornais de informação corrente
- 2 — Jornais comerciais e financeiros
- 3 — Revistas científicas e técnicas
- 4 — Livros e monografias
- 5 — Publicações de congressos
- 6 — Revistas de compilação
- 7 — Revistas de abstracts e de índices
- 8 — Listas de traduções
- 9 — Listas de patentes
- 10 — Relatórios governamentais
- 11 — Relatórios não governamentais
- 12 — Instituições nacionais
- 13 — Instituições estrangeiras
- 14 — Universidades
- 15 — Embaixadas
- 16 — Directórios
- 17 — Manuais
- 18 — Livros de tabelas
- 19 — Almanques
- 20 — Enciclopédias
- 21 — Dicionários
- 22 — Atlas e mapas
- 23 — Reuniões
- 24 — Brainstorming
- 25 — Visitas, cursos e congressos
- 26 — Contactos pessoais

O facto de se gastar quase uma página a apresentar uma lista numerada de 26 fontes de informação, que são do conhecimento de toda a gente, pode parecer perfeitamente disparatado dentro da mentalidade muito generalizada de ridicularizar os processos simples de análise dos sistemas.

Ora, o que é certo é que ninguém faz troça das agulhas dos caminhos de ferro, mas uma agulha errada pode alterar um percurso de centenas de quilómetros e até, numa hipótese mais dramatizada, dar origem a acidentes mais ou menos complicados.

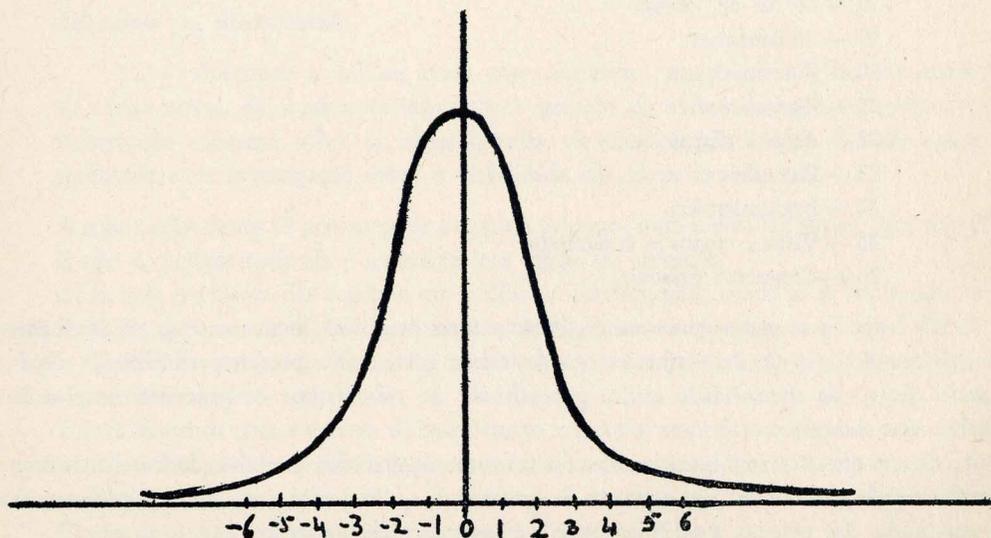
É por isso que, como se insistiu por várias vezes nas sessões sobre «O Diálogo com os Computadores», o gestor moderno tem de ter uma mentalidade analítica, que o leve ao contacto com os dados elementares em que baseia a sua acção porque, de outro modo, se arrisca a «fazer agulhas erradas» e ter depois de gastar grande parte da sua energia intelectual a justificar as razões lógicas por que isso aconteceu.

Aliás, como se fez notar na sessão anterior, justificar erros é um conceito que não faz sentido, porque um dos «slogans» que está a ser muito usado, e que parece ser totalmente correcto, é o de que um gestor deve proceder «como se o dinheiro fosse seu e não do Patrão».

Este «slogan» é particularmente significativo em relação ao problema da informação, porque se um gestor actuar «como se o dinheiro fosse seu», tende para se tornar um «profissional de utilização da informação»; se actuar «como se o dinheiro fosse do Patrão», bastar-lhe-á fornecer uma informação sensacional por semana ou por quinzena para convencer o chefe hierárquico de que «anda muito bem informado».

Acontece também que qualquer ente humano não tem capacidade para apreender toda a informação que se gera à sua volta e, por isso, precisa de começar por definir qual é o seu perfil de interesses em relação à informação que há disponível.

Esse perfil de interesses pode visualizar-se através de uma curva de Gauss em que em abcissas se marquem os assuntos com um número de código qualquer (como números positivos ou negativos) e, em ordenadas se indiquem os valores relativos  $Q_r$  das quantidades de informação requeridas:



Deste modo, a profissionalização dos gestores em relação à utilização da informação baseia-se em três regras:

- 1 — Definir o seu perfil de interesses.
- 2 — Conhecer todas as possibilidades em relação aos 26 tipos de fontes de informação no campo das informações específicas a colher.
- 3 — Seleccionar de entre essas possibilidades as que, por experiência ou por indicação alheia, considerar de maior utilidade dentro de um orçamento global a respeitar.

É na quantificação desta 3.<sup>a</sup> regra que aparece o parâmetro  $A$ .

Os estudos feitos por Bradford remontam a 1930 e deles se referem os casos da «Geofísica Aplicada» e da «Lubrificação» que foram apresentados como exemplos nesses estudos.

Assim, para a Geofísica Aplicada, tem-se:

- As 9 fontes de informação mais relevantes produzem 429 referências por ano.
- As 59 fontes seguintes, de relevância média, produzem 499 referências.
- As 258 fontes seguintes, de relevância menor, produzem 404 referências.

Partindo da expressão  $N = A \log n$ , os valores de  $A$ , que correspondem a cada um destes tipos de fontes de informação, são evidentemente:

$$A = \frac{429}{\log 9} = \frac{429}{0,948} = 460$$

$$A = \frac{429 + 499}{\log 59} = \frac{928}{1,771} = 520$$

$$A = \frac{928 + 404}{\log 258} = \frac{1\,332}{2,413} = 550$$

Tomando a média destes três valores, obtém-se:

$$A = 510$$

De igual modo, para a Lubrificação, tem-se:

- As 8 fontes de informação mais relevantes produzem 110 referências por ano.
- As 29 fontes seguintes, de relevância média, produzem 133 referências.
- As 127 fontes seguintes, de relevância menor, produzem 152 referências.

Procurando o valor de  $A$  obtém-se um valor médio de

$$A = 150$$

Apesar da importância normativa da Lei de Bradford, não existem estudos sistemáticos que nos indiquem os valores de  $A$  para cada caso particular e, por isso, o parâmetro  $A$  tem um certo quê de pessoal em relação a cada utilizador.

Há porém um aspecto perfeitamente objectivo, que é indiscutível, e que é o que se refere aos valores relativos de *A*, os quais serão tanto menores quanto mais especializados forem os assuntos do perfil de interesse dos utilizadores.

Deste modo, o valor *A* = 50 com que entrámos no exemplo que demos, pode considerar-se aceitável.

Passemos agora ao 2.º parâmetro do modelo de optimização da informação, que é o do custo da informação.

Utilizando a linguagem da Teoria da Documentação, as informações produzidas são «mensagens», as quais só chegam ao utilizador através de um sistema de comunicações.

Para que este sistema de comunicações seja efectivo e sistemático, tem de se apoiar num conjunto de 8 «operações unitárias» que se sintetizam no quadro II.

QUADRO II

Operação Unitária	Função desempenhada
I — AQUISIÇÃO DA INFORMAÇÃO	Localização, selecção, encomenda e recepção da informação.
II — ANÁLISE DA INFORMAÇÃO	Apreciação das informações recebidas e selecção dos aspectos que se consideram de importância suficiente para justificar o trabalho de as tornar acessíveis em qualquer momento.
III — CONTROLE DA TERMINOLOGIA	Estabelecimento de relações arbitrárias entre esses aspectos e um determinado sistema de arrumação.
IV — REGISTO DOS RESULTADOS DA ANÁLISE DE MODO A TORNAR A PESQUISA POSSÍVEL	Uso de um cartão, de uma fita, de um filme ou qualquer outro meio material, no qual os aspectos escolhidos são anotados.
V — ARMAZENAGEM DAS INFORMAÇÕES DE BASE	Colocação ou registo num dado local, quer sob a forma original quer transcrito para o novo documento.
VI — ANÁLISE DAS PERGUNTAS E DESENVOLVIMENTO DE UMA ESTRATÉGIA DE PESQUISA	Expressão da pergunta em termos de um mecanismo de pesquisa adequado.
VII — ORIENTAÇÃO DA PESQUISA	Manipulação do mecanismo de pesquisa, de modo a identificar os registos arquivados no ficheiro.
VIII — FORNECIMENTO DOS RESULTADOS DA PESQUISA	Remoção física ou cópia dos registos do ficheiro em arquivo, de modo a serem utilizados como resposta às perguntas.

A análise da informação pode ainda decompor-se nas 4 operações do quadro III.

QUADRO III

Operação	Função desempenhada
a) — PRODUÇÃO DE ABSTRACTS	Resumo dos aspectos essenciais de uma informação extensa acompanhado de uma descrição bibliográfica adequada, de modo a permitir detectar a informação original.
b) — PRODUÇÃO DE ÍNDICES	Seleção de palavras ou ideias, de modo a facilitar a identificação ou selecção futuras.
c) — CLASSIFICAÇÃO	Agrupamento por categorias de acordo com um código previamente estabelecido
d) — EXTRACÇÃO	Seleção das citações mais representativas da informação original.

O preço do conjunto de todas estas informações é muito superior ao custo das publicações em si mesmas.

Daqui resulta que o custo da informação é muito elevado, porque um gestor não pode viver à custa de informações isoladas, mas tem de estar ao corrente, sob forma sistemática, de toda a informação relevante que diz respeito ao seu ramo de actividade, como condição «sine-qua-non» de se poder orientar no sentido da evolução dessa actividade.

E ao nível mundial, o problema é tomado de tal modo a sério que, por exemplo, nos Estados Unidos começa a haver dificuldades com o pessoal de documentação, porque as empresas se vêem na necessidade de aumentar os seus ordenados, dentro do jogo normal da oferta e da procura.

Em termos numéricos, podemos aceitar que as operações indicadas são equivalentes a 5 vezes o valor das publicações, pelo que, admitindo que uma revista custa 500\$00 por ano «em bruto», custará  $C_m = 500\$00 + 5 \times 500\$00 = 3\,000\$00$ , depois de «trabalhada».

Isto explica o valor de  $C_m = 3\,000\$00$  com que entrámos no exemplo dado.

Para se perceber ainda melhor qual a incidência dos parâmetros  $A$  e  $C_m$ , vamos utilizar os valores já considerados para calcularmos o custo  $C$  por artigo.

Será evidentemente:

$$C = C_m \frac{n}{N} = \frac{C_m n}{A \log n} = \frac{3\,000}{50} \cdot \frac{n}{\log n} = 60 \frac{n}{\log n}$$

A curva  $n/\log n$  tem um mínimo para  $n = e$  (2,71828) e alguns valores são dados no quadro IV.

QUADRO IV

$n$	$n/\log n$
1	$\infty$
2	6,7
$e$	6,3
5	7,2
10	10,0
50	29,7
100	50,0
500	187,0
1 000	333,3
5 000	1 490,0

Este quadro mostra que o custo por artigo cresce muito rapidamente com  $n$ , o que aliás seria de esperar.

Para o valor de  $n = 15$  que utilizámos no exemplo dado, teremos por interpolação linear  $n/\log n = 12,5$  pelo que

$$C = 60 \times 12,5 = 750\$00/\text{artigo}$$

Como uma revista não conterà apenas artigos sobre um dado assunto, o valor de  $C$  será efectivamente menor, digamos 5 vezes mais pequeno, para termos um valor de referência.

$$\text{Será portanto } C = 150\$00/\text{artigo}.$$

O valor achado situa-se na ordem de grandeza de informações americanas, que dão

$$C = \text{US } \$ 6,5/\text{artigo} \text{ ou seja } = 190\$00/\text{artigo}$$

É evidente que nenhum destes resultados deve ser tomado em valor absoluto, o que não quer dizer que não tenham o maior interesse para se conhecer a ordem de grandeza do custo real da informação.

Resta-nos agora analisar o 3.º parâmetro que é o do preço da informação perdida.

O valor de uma informação perdida pode avaliar-se empiricamente pela seguinte fórmula:

$$C_p = \frac{0,001 (V + C)}{A \log n_t (1 - P_1)}$$

em que  $V$  é o volume de vendas por ano,  $C$  o capital e  $A$ ,  $n_t$  e  $P_1$ , têm os significados indicados atrás.

Antes de a analisarmos consideremos um exemplo, com os seguintes dados:

$$V = 50\,000 \text{ contos}$$

$$C = 30\,000 \text{ contos}$$

$$A = 50$$

$$n_t = 100$$

$$P_1 = 0,58 \text{ (tirado do quadro I para } n = 15 \text{ ou seja } P = 15\%)$$

Será portanto:

$$C_p = \frac{0,001 \times 80\,000}{50 \times 2 \times 0,42} = \frac{80}{42} = 1\,900\$00 \approx 2\,000\$00$$

O resultado obtido explica o valor de  $C_p = 2\,000\$00$  com que entrámos atrás.

O significado da fórmula é o seguinte:

No que se refere ao denominador, representa o número total  $N_t$  de artigos perdidos, como vimos atrás, e serve portanto para calcular o valor médio unitário.

Quanto ao numerador, constitui uma hipótese de trabalho segundo a qual o valor total dos artigos perdidos é 1% da soma do volume das vendas e do capital, sendo  $0,001 = 0,1 \times \times 0,1 = R_1 \times R_2 \times R_3$ , isto é o produto de três factores  $R_1$ ,  $R_2$  e  $R_3$  todos iguais a 10%.

Analisemos cada um destes três factores:

- $R_1$  exprime que só 10% do volume das vendas anuais pode ser influenciado por informações acidentais porque se admite que 90% está planificado; por outro lado, em relação ao capital, pressupõe uma renovação dos investimentos de 10% todos os anos.
- $R_2$  significa que se atribui um lucro de 10% quer às vendas suplementares quer aos novos investimentos.
- $R_3$  traduz a hipótese de que só 10% das informações de interesse podem ser concretizadas.

É evidente que a atribuição de valores concretos a  $R_1$ ,  $R_2$  e  $R_3$  é extraordinariamente frágil, mas tem o grande mérito de nos permitir «quantificar» os raciocínios qualitativos.

E agora que já «radiografámos» o problema da obtenção da informação, vamos de novo pôr uma pergunta que já não é inédita:

— O que é que isto tudo tem que ver com os Conjuntos, com as Matrizes e com as Probabilidades?

Tem tudo e não tem nada!

Não tem nada, na medida em que o texto apresentado não foi utilizado para aumentar quaisquer conhecimentos relacionados com o esquema das tais sessões.

Mas tem tudo porque, para além das técnicas de equacionamento dos dados, há no momento presente um problema muito mais angustioso, que é o da definição desses dados.

E os dados só se definem com informação, que tem de obedecer a 4 regras básicas:

- I — Numa estrutura estática, o indivíduo preocupa-se com o que sabe e ofende-se com o que lhe ensinam; numa estrutura dinâmica, o indivíduo preocupa-se com o que não sabe e agradece o que lhe ensinam.
- II — Não vale a pena ter a ilusão de que se tomam decisões cientificamente se os dados em que se apoiarem não forem completamente «reliable».
- III — O conhecimento correcto dos dados implica a existência de uma informação cientificamente organizada que não dê origem nem a «ruído» nem a «silêncio».
- IV — A informação existente é praticamente inútil se não for sistematicamente registada, correctamente classificada e rapidamente acessível.

Infelizmente a importância destas 4 regras está muito longe de ser uma opinião assente em Portugal.

Na realidade, o problema da obtenção da informação é considerado ainda por muita gente uma actividade secundária correspondente a um tipo de trabalho que «toda a gente sabe fazer» e que se entrega a quem «não sabe o que fazer».

Simplesmente, quando se envereda por esse caminho, não se está a integrar a informação na empresa, mas está-se simplesmente a «brincar à informação».

E o que é curioso é que a maioria das pessoas fica muito envergonhada quando faz raciocínios incorrectos, mas fica totalmente tranquilizada quando se pode justificar com o argumento de que se baseou numa informação errada, não se sentindo responsável por isso.

Ora, o que é certo, é que cada um de nós é responsável pelos erros dos dados que utiliza, mesmo que esses erros sejam simplesmente de omissão.

Por outro lado, a informação é essencialmente dinâmica e só é válida no momento exacto em que foi colhida.

Daí se infere que a informação pode ser também errada por desactualização.

Este aspecto particular conduz ao conceito de SDI (Selective Dissemination of Information) o qual significa na prática que a organização da informação não pode ser apenas um processo passivo mas tem de ser fundamentalmente um processo agressivo que inunde as pessoas com a informação que requerem através da definição do seu perfil de utilizador.

Isto levanta um problema psicológico que é o das pessoas se habituarem a serem bombardeadas com informações por pessoas estranhas ao seu campo de actividade.

Esta afirmação parece poder levar à conclusão de que, para evitar esse problema psicológico, a solução estaria em integrar os Serviços de informação dentro das próprias actividades.

Tal conclusão não é totalmente correcta porque, a menos que sejam os próprios gestores que andem a preparar a pesquisa da sua informação (o que parece não ser a sua função), são sempre serviços estranhos à actividade que efectuem essa pesquisa.

O que o gestor tem de ser é, como se disse atrás, um profissional de utilização da informação, que se saiba zangar quando não lhe dão a «informação de que precisa», mas isso implica, como condição prévia, que haja quem saiba de que informação é que ele precisa.

E quando em Portugal se ouve falar na falta de ideias para o desenvolvimento económico do País, o que acontece, na realidade, é haver falta de profissionais da utilização da informação; e enquanto a grande maioria dos gestores não se convencer de que é assim, não há «mesas redondas», nem quadradas nem bicudas, que dêem o incremento indispensável ao progresso nacional.

Mas isso transcende o âmbito desta Sessão e, por isso, por aqui nos ficamos...