

... QUEM SABE ESTATÍSTICA É REI *

DINIS PESTANA

Departamento de Estatística e Investigação Operacional
Faculdade de Ciências da Universidade de Lisboa

SÍLVIO VELOSA

Centro de Estatística e Aplicações da Universidade de Lisboa
Faculdade de Ciências da Universidade de Lisboa

* Investigação financiada por FCT — PRAXIS XXI, FEDER. Conferência convidada no MATVISEU 98.

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.

H. G. Wells.

A witty statesman said you might prove anything by figures, but a judicious man looks at statistics, not to get knowledge, but to save himself from having ignorance foisted on him.

Thomas Carlyle, Chartism.

An embarrassing proportion of key decisions in the Government, from the negotiation of treaties to the management of resources, are made on the basis of insufficient information and unproved assumptions. And this is not so different from the way we conduct our private affairs.

Editorial, Science.

1. O que se pretende.

O título não há-de ser muito enigmático — procura ecoar, evidentemente, o ditado popular "em terra de cegos quem tem um olho é rei".

Diversos ensaístas se têm apercebido, desde meados do século XIX, da importância crescente da Estatística. H. G. Wells, mais conhecido pela sua ficção científica, mas que também reflectiu e escreveu

muito seriamente sobre a evolução das sociedades, afirmou que, num futuro não muito distante, a capacidade de entender informação estatística seria tão essencial como saber ler e escrever. Mais acutilantemente ainda, porventura, o pensador político Carlyle afirmou que a Estatística constituía a defesa mais segura contra quem [os políticos] nos quisesse enganar com pseudo-verdades alicerçadas em factos deturpados ou usados a contra-senso ou fora de contexto. Muito recentemente, um exemplar editorial do *Science* lamentava que os políticos tomassem tantas vezes decisões com base em informação errada e/ou insuficiente; para logo comentar "tal e qual como nós, na nossa vida".

O programa liberal para a educação dos povos, que ganhou ímpeto no século XIX, apresentava como palavra de ordem que "todos devem saber ler, escrever e contar". Sempre nos pareceu curioso que fosse distinguida a capacidade de ler e de escrever (não nos parece que com um sentido pleno, com a percepção de que saber ler não corresponde necessariamente a inteligir o sentido do que se está lendo), e que não se desse qualquer relevo à capacidade de saber "ler números", descodificar informação quantitativa.

A capacidade de saber "ler" informação quantitativa é o cerne da Estatística. E tudo quanto é bom contém em si mesmo a sua própria perversão. É fácil mentir com Estatística — todos os dias assistimos a isso, de forma bem criativa, em anúncios da televisão, ou (em geral de forma menos criativa) quando os líderes políticos procuram vender-nos as suas ideias.

A Estatística cumpriu as promessas dos seus apologistas do passado: vem invadindo a nossa vida quotidiana e ganhou direito de cidade na metodologia da investigação científica, tornando-se um auxiliar precioso de todas as ciências. Infelizmente, a cultura estatística é em geral deprimentemente baixa, o que permite que a Estatística seja usada contra nós, por malícia ou simplesmente por mau uso.

Procuramos ilustrar com o exemplo abaixo como uma utilização maliciosa de padrões que não são imediatamente evidentes pode viciar a investigação de um problema científico (ou mais directamente servir para comer dinheiro a incautos).

2. A Percepção Extra-Sensorial Existe?

— Acreditam na percepção extra-sensorial? A senhora não? Então venha aqui por favor ajudar-me.

...

— Ora muito obrigado. Faça então o favor de escolher um destes envelopes. Este? Então abra-o por favor, deixe cair as quatro moedas que contém. Vou pedir-lhe agora que confirme que em cada face de cada uma das moedas há um número escrito, e que os oito números escritos são distintos.

...

— Atire por favor uma das moedas ao ar, várias vezes. Umas vezes saíu um dos números e outras vezes o outro, não foi? A moeda parece-lhe equilibrada, no sentido em que não parece ter mais propensão para sair um dos lados do que o outro? Fico muito satisfeito por confirmar este ponto!

...

— Então, se os números são todos distintos, concorda que há $2 \times 2 \times 2 \times 2 = 24$ possibilidades quando se atiram as quatro moedas?

...

— Ora faça o favor de atirar as moedas ao ar e de me ditar os números que ficam à vista.

...

— Disse 32, 69, 60, 39? Então a soma é 200, não é? Faça então o favor de retirar de dentro do envelope o papel que lá está dobrado, e ler o que está escrito.

...

— Vou então repetir o que estava lá escrito por mim há vários dias: "A soma é 200".

...

- Bem, quem é que continua a não acreditar na percepção extra-sensorial? Ainda há cépticos? Sempre há gente muito pirrónica! Então escolha lá outro envelope e deixe cair as moedas. Os oito números são todos diferentes, não são? E são diferentes dos do envelope anterior, não são? Vá, não hesite, atire-os ao ar. Muito bem! Os números que saíram são agora 37, 55, 53 e 42, e, deixe cá ver, a soma é 187, não é? Ora leia lá o papelinho: "A soma é 187". Ora esta!

...

— Vamos a mais uma tentativa, a ver se desta vez eu me engano. Atire lá as quatro moedas ao ar. Posso aleatorizar um pouco mais, lançar eu próprio esta moeda ao ar? Muito bem! Ora os números desta vez são 15, 56, 8, 55, e puxe lá de uma máquina de calcular e diga-me o produto — é 369 600. E o papelinho diz "O produto é 369 600"!

...

— Bem minha senhora, agora já acredita em percepção extra-sensorial? Menos ainda do que no princípio? Isto é que a senhora me saíu uma finória com o faro bem apurado, percebeu logo que havia tramóia na questão!

Pois é, os números são todos diferentes, mas há uma estrutura — que não é imediatamente aparente. Os números escritos nas quatro moedas do primeiro envelope são

Face: 32 37 39 46

Corôa: 55=32+23 60=37+23 62=39+23 69=46+23

e conseqüentemente, em vez de haver $16 = 2^4$ somas diferentes (o que seria de esperar com 4 pares de números distintos, caso não houvesse nenhum padrão nos números), há apenas 5 somas distintas:

- $32+37+39+46 = 154$ (de uma única forma possível, os 4 números de face e nenhum de corôa, $\binom{4}{0} = 1$);
- $(32+37+39+46)+23 = 177$ (de 4 formas possíveis, três números de face e um de corôa, $\binom{4}{1} = 4$);
- $(32+37+39+46)+23+23 = 200$ (de 6 formas possíveis, dois números de face e dois de corôa, $\binom{4}{2} = 6$);
- $(32+37+39+46)+23+23+23 = 223$ (de 4 formas possíveis, um número de face e três de corôa, $\binom{4}{3} = 4$);
- $(32+37+39+46)+23+23+23+23 = 246$ (de 1 forma possível, zero números de face e quatro de corôa, $\binom{4}{4} = 1$).

Por outras palavras, a soma é a variável aleatória

$$S = \begin{bmatrix} 154 & 177 & 200 & 223 & 246 \\ \frac{1}{16} & \frac{4}{16} & \frac{6}{16} & \frac{4}{16} & \frac{1}{16} \end{bmatrix}$$

Desta forma, a soma 200 é a mais provável, e nenhum casino desdenharia este jogo, sendo 200 a aposta da casa, mesmo que os ganhos jogo a jogo fossem diminutos. Nesta brincadeira, usámos evidentemente um truque: nas faces pusémos números inferiores a 50, nas corôas pusémos números superiores a 50, e consequentemente sabíamos que a "predição" bateria certo se saíssem dois valores inferiores a 50 e dois superiores a 50 (no caso do produto, nas faces havia números terminados em 5, e nas corôas esses números multiplicados por 1.4). No caso em que não saía logo o que nos interessava, sugeríamos que uma (ou duas) das moedas — a(s) que não nos convinha(m) — fosse(m) de novo lançada(s) ao ar, e levaríamos essa pseudo-aleatorização até sair o que de facto nos interessava.

Note que num casino real este tipo de batota é totalmente impensável: ninguém estaria interessado num jogo que não é possível ganhar. O que faz a fortuna dos casinos é a capacidade de "gerir" pequenos ganhos em grande quantidade — e gerir as esperanças insensatas e o gosto pelo risco dos jogadores. Um jogo inspirado no exemplo acima é excessivamente matemático para atrair o público usual de casinos, mas se, por um momento, imaginarmos que por cada unidade monetária arriscada pelo jogador o casino

- arrecada essa propina no caso de o jogador perder;
- paga 2.5 vezes o valor da aposta no caso de o jogador ganhar apostando no 200 (devolve a propina e paga 1.5 vezes a aposta, como prémio);
- paga 3.5 vezes o valor da aposta no caso de o jogador ganhar apostando no 177 ou no 223;
- paga 13 vezes o valor da aposta no caso de o jogador ganhar tendo apostado no 154 ou no 246,

então:

1. O ganho correspondente a uma unidade monetária é, para um jogador que aposte no 200

$$G = \begin{bmatrix} -1 & 1.5 \\ \frac{10}{16} & \frac{6}{16} \end{bmatrix}$$

com valor esperado $E[G] = -1\cancel{\text{€}}\frac{10}{17} + 1.5\cancel{\text{€}}\frac{6}{17} = -\frac{1}{17}$.

2. O ganho correspondente a uma unidade monetária é, para um jogador que aposte no 177 ou no 223

$$G_{177} = G_{223} = \begin{bmatrix} -1 & 2.5 \\ \frac{12}{16} & \frac{4}{16} \end{bmatrix}$$

com valor esperado $E[G_{177}] = E[G_{223}] = -1\cancel{\text{€}}\frac{12}{17} + 2.5\cancel{\text{€}}\frac{4}{17} = -\frac{2}{17}$.

3. O ganho correspondente a uma unidade monetária é, para um jogador que aposte no 154 ou no 246

$$G_{154} = G_{246} = \begin{bmatrix} -1 & 12 \\ \frac{15}{16} & \frac{1}{16} \end{bmatrix}$$

com valor esperado $E[G_{154}] = E[G_{246}] = -1\cancel{\text{€}}\frac{15}{16} + 12\cancel{\text{€}}\frac{1}{16} = -\frac{3}{16}$.

Assim, em termos médios o jogador perde sempre (sendo mais penalizado, em termos médios, quanto maior é o risco para o casino de um jogador ocasional ter uma jogada de sorte e abandonar logo o jogo, dando prejuízo à casa), mesmo que ocasionalmente ganhe um prémio apetitoso relativamente ao que apostou.

Por outro lado o casino, numa noite em que haja 1600 apostas no 200, 3200 apostas no 177 ou no 223, 8000 apostas no 154 ou no 246, espera ganhar com este jogo 2000 unidades monetárias — não é mesmo tentador ter um casino?

3. Comentários finais.

O exemplo apresentado ilustra vários pontos, que desafiamos o leitor a aprofundar:

- a investigação de situações dúbias (como a existência de percepção extra-sensorial) pode ser feita com o auxílio da Estatística: se for detectado um padrão favorável ou desfavorável a uma determinada hipótese, as nossas convicções podem ser abaladas (ou pelo contrário reforçadas). A função do estatístico é um pouco como o papel do juiz, que tem que avaliar a acumulação de evidência num ou noutro sentido — não perdendo de vista, obviamente, que a apresentação da evidência pode ser enganosa, ou o nosso ajuizar dela imperfeito.

De facto, a ocorrência do insólito não deve abalar-nos.

Acontecimentos de probabilidade baixa estão constantemente a acontecer; por exemplo, todas as semanas na extracção da lotaria sai um número, cuja probabilidade não passava, antes da extracção, de $\frac{1}{300000}$ (estamos a supor que tinham sido emitidos 500 000 bilhetes). Mais impressionante ainda, se aceitarmos que a vida é uma sucessão de filtros, ainda que a probabilidade de sucesso em cada caso seja P_i "grande",

próxima de 1, a probabilidade de sucesso na sucessão de filtros que enfrentamos (admitindo, por

simplicidade, independência) é $\prod_{i=1}^n P_i$, que decresce vertiginosamente para 0 à medida que n aumenta.

De forma análoga, mesmo que a probabilidade de um acontecimento seja praticamente nula (por exemplo, a probabilidade de uma profecia vir a realizar-se), $P_i \approx 0$, $1 - P_i \approx 1$, mas se houver um

número muito grande de repetições da experiência $\prod_{i=1}^n (1 - P_i)$ converge para 0, e a probabilidade de

vir a realizar-se pelo menos uma vez esse acontecimento altamente improvável é $1 - \prod_{i=1}^n (1 - P_i) \approx 1$, altamente provável.

É por isso que não devemos impressionar-nos excessivamente com pressentimentos que vêm a ser confirmados pela sua realização — por cada pressentimento que corresponde à sua realização há milhares que abortam. A "lei do zero ou 1" têm várias consequências aparentemente inesperadas: por exemplo, diz-se que o inferno dos probabilistas é habitado não por demónios mas antes por infinitos macacos, a bater ao acaso nas teclas de infinitas máquinas de escrever, infinitamente; de acordo com a

lei do 0 ou 1, algum deles acabará por escrever, quase certamente, a Guerra e Paz (ou qualquer outro calhamaço), sem qualquer erro, sequer de pontuação!

Há assim que ter a maior das prudências no uso da Estatística na investigação científica de questões em relação às quais há uma natural desconfiança, mesmo que não se faça batota como no exemplo apresentado.

Livros como O Código da Bíblia, que baseiam os raciocínios na estranheza de observação de coincidências pouco prováveis, estão a fazer um mau uso da Estatística, e a confundir os incautos quando apelam a que em testes de hipóteses usuais se arredam as que têm uma probabilidade inferior a 5%, enquanto as coincidências que eles observam têm uma probabilidade muito menor, e por isso devem deixar de ser encaradas como coincidências e ganhar o estatuto de "profecia válida". Esquecem deliberadamente que os testes de hipóteses investigam hipóteses plausíveis no nosso modelo do Mundo, e que assim a observação de acontecimentos de probabilidade baixa tem um sentido totalmente diferente da observação de acontecimentos pouco prováveis que não estão alicerçados num modelo. Era um pouco como se baseados numa experiência desgarrada — uma vez um dos autores saiu de um combóio quando este fechou as portas e arrancou, ficando com o braço esquerdo e a mala presos dentro do combóio, e viajou do lado de fora durante cerca de meia-hora —, mas decerto não única na sua essência, construíssemos uma teoria em que viajar fora dos veículos ganhasse de repente foros de respeitabilidade e fosse questão a ponderar maduramente cada vez que se entra ou sai de um combóio!

- a realidade é complexa, e pode conter padrões inesperados que distorçam a nossa apreensão (gostaríamos que as nossas investigações tivessem sido devidamente planeadas por forma a eliminar a variabilidade espúria, que nada tem que ver com o que queremos investigar, de variáveis confounding). Neste caso, os oito números em causa são todos distintos, e parecem não ter relação uns com os outros; mas de facto não são independentes, e isso altera radicalmente a avaliação do problema por um leigo.

Note que a "regularidade", a existência de um padrão, em geral está muito mais camuflada. Por exemplo, muitos matemáticos seriam levados a dizer que não há um padrão definido nos números primos (para além do óbvio: serem, à excepção de 2, ímpares). No entanto, se considerarmos números primos consecutivos — por exemplo 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, e depois os seus quadrados, 1369, 1681, 1849, 2209, 2809, 3481, 3721, 4489, 5041, 5329, 6241, e contarmos o número de números primos que existe em cada intervalo definido por pares consecutivos daqueles quadrados de primos obtemos respectivamente

intervalo	número de números primos
(1369, 1681)	44
(1681, 1849)	20
(1849, 2209)	46
(2209, 2809)	80
(2809, 3481)	78
(3481, 3721)	32
(3721, 4489)	90
(4489, 5041)	66
(5041, 5329)	30
(5329, 6241)	106

Até aqui nada de particularmente notável.

Consideremos, agora, 8 subintervalos iguais em (1369, 1681), e contemos o número dos 44 primos do referido intervalo que cai em cada um dos 8 subintervalos:

(1369,1408)	(1408,1447)	(1447,1486)	(1486,1525)	(1525,1564)	(1564,1603)	(1603,1642)	(1642,1681)
3	7	6	6	5	6	7	4

que já parece uma distribuição bastante regular. De facto, se testarmos a hipótese nula de uniformidade discreta, o valor esperado em cada uma das classes é $44:8 = 5.5$, e a clássica estatística do

Qui quadrado é $X^2 = 2.55$, enquanto $\chi^2_{(7)} = 14.07$ (o valor esperado sob validade da hipótese nula é 7, pelo que o valor observado é surpreendentemente baixo!) — e consequentemente nesta situação devemos manter a hipótese nula.

Procedendo de forma análoga para os outros intervalos considerados, a tabela junta mostra que em todos os caso é de manter a hipótese nula de distribuição uniforme discreta nos subintervalos.

início.	1681	1702	1723	1744	1765	1786	1807	1828
subint.								
nº	3	3	2	3	2	3	2	2
primos								
	valor	observado	da	estatística	qui-	0.80		
	quadrado:							
início.	1849	1894	1939	1984	2029	2074	2119	2164
subint.								
nº	7	5	4	8	5	7	7	3
primos								
	valor	observado	da	estatística	qui-	3.74		
	quadrado:							
início.	2209	2284	2359	2434	2509	2584	2659	2734
subint.								
nº	10	11	10	8	8	9	14	10
primos								
	valor	observado	da	estatística	qui-	2.60		
	quadrado:							
início.	2809	2893	2977	3061	3145	3229	3313	3397
subint.								
nº	9	11	9	8	11	9	12	9
primos								
	valor	observado	da	estatística	qui-	1.38		
	quadrado:							
in.	3481	3511	3541	3571	3601	3631	3661	3691
subint.								
nº	3	6	4	3	5	3	4	4
primos								
	valor	observado	da	estatística	qui-	2.00		
	quadrado:							
início.	3721	3817	3913	4009	4105	4201	4297	4393
subint.								
nº	10	12	12	12	10	14	9	11
primos								
	valor	observado	da	estatística	qui-	1.56		
	quadrado:							
início.	4489	4558	4627	4696	4765	4834	4903	4972
subint.								
nº	8	7	10	7	9	5	10	10
primos								
	valor	observado	da	estatística	qui-	2.85		
	quadrado:							
início.	5041	5077	5113	5149	5185	5221	5257	5293
subint.								

nº 3 5 3 4 3 4 4 4
 primos
 valor observado da estatística qui- **0.93**
 quadrado:

início. 5329 5443 5557 5671 5785 5899 6013 6127
 subint.

nº 15 13 14 12 17 9 13 13
 primos
 valor observado da estatística qui- **2.83**
 quadrado:

Números primos:

2	277	643	1039	1471	1901	2351	2797	3307	3761	4241	4733
3	281	647	1049	1481	1907	2357	2801	3313	3767	4243	4751
5	283	653	1051	1483	1913	2371	2803	3319	3769	4253	4759
7	293	659	1061	1487	1931	2377	2819	3323	3779	4259	4783
11	307	661	1063	1489	1933	2381	2833	3329	3793	4261	4787
13	311	673	1069	1493	1949	2383	2837	3331	3797	4271	4789
17	313	677	1087	1499	1951	2389	2843	3343	3803	4273	4793
19	317	683	1091	1511	1973	2393	2851	3347	3821	4283	4799
23	331	691	1093	1523	1979	2399	2857	3359	3823	4289	4801
29	337	701	1097	1531	1987	2411	2861	3361	3833	4297	4813
31	347	709	1103	1543	1993	2417	2879	3371	3847	4327	4817
37	349	719	1109	1549	1997	2423	2887	3373	3851	4337	4831
41	353	727	1117	1553	1999	2437	2897	3389	3853	4339	4861
43	359	733	1123	1559	2003	2441	2903	3391	3863	4349	4871
47	367	739	1129	1567	2011	2447	2909	3407	3877	4357	4877
53	373	743	1151	1571	2017	2459	2917	3413	3881	4363	4889
59	379	751	1153	1579	2027	2467	2927	3433	3889	4373	4903

61	383	757	1163	1583	2029	2473	2939	3449	3907	4391	4909
67	389	761	1171	1597	2039	2477	2953	3457	3911	4397	4919
71	397	769	1181	1601	2053	2503	2957	3461	3917	4409	4931
73	401	773	1187	1607	2063	2521	2963	3463	3919	4421	4933
79	409	787	1193	1609	2069	2531	2969	3467	3923	4423	4937
83	419	797	1201	1613	2081	2539	2971	3469	3929	4441	4943
89	421	809	1213	1619	2083	2543	2999	3491	3931	4447	4951
97	431	811	1217	1621	2087	2549	3001	3499	3943	4451	4957
101	433	821	1223	1627	2089	2551	3011	3511	3947	4457	4967
103	439	823	1229	1637	2099	2557	3019	3517	3967	4463	4969
107	443	827	1231	1657	2111	2579	3023	3527	3989	4481	4973
109	449	829	1237	1663	2113	2591	3037	3529	4001	4483	4987
113	457	839	1249	1667	2129	2593	3041	3533	4003	4493	4993
127	461	853	1259	1669	2131	2609	3049	3539	4007	4507	4999
131	463	857	1277	1693	2137	2617	3061	3541	4013	4513	5003
137	467	859	1279	1697	2141	2621	3067	3547	4019	4517	5009
139	479	863	1283	1699	2143	2633	3079	3557	4021	4519	5011
149	487	877	1289	1709	2153	2647	3083	3559	4027	4523	5021
151	491	881	1291	1721	2161	2657	3089	3571	4049	4547	5023
157	499	883	1297	1723	2179	2659	3109	3581	4051	4549	5039
163	503	887	1301	1733	2203	2663	3119	3583	4057	4561	5051
167	509	907	1303	1741	2207	2671	3121	3593	4073	4567	5059
173	521	911	1307	1747	2213	2677	3137	3607	4079	4583	5077
179	523	919	1319	1753	2221	2683	3163	3613	4091	4591	5081
181	541	929	1321	1759	2237	2687	3167	3617	4093	4597	5087
191	547	937	1327	1777	2239	2689	3169	3623	4099	4603	5099

193	557	941	1361	1783	2243	2693	3181	3631	4111	4621	5101
197	563	947	1367	1787	2251	2699	3187	3637	4127	4637	5107
199	569	953	1373	1789	2267	2707	3191	3643	4129	4639	5113
211	571	967	1381	1801	2269	2711	3203	3659	4133	4643	5119
223	577	971	1399	1811	2273	2713	3209	3671	4139	4649	5147
227	587	977	1409	1823	2281	2719	3217	3673	4153	4651	5153
229	593	983	1423	1831	2287	2729	3221	3677	4157	4657	5167
233	599	991	1427	1847	2293	2731	3229	3691	4159	4663	5171
239	601	997	1429	1861	2297	2741	3251	3697	4177	4673	5179
241	607	1009	1433	1867	2309	2749	3253	3701	4201	4679	5189
251	613	1013	1439	1871	2311	2753	3257	3709	4211	4691	5197
257	617	1019	1447	1873	2333	2767	3259	3719	4217	4703	5209
263	619	1021	1451	1877	2339	2777	3271	3727	4219	4721	5227
269	631	1031	1453	1879	2341	2789	3299	3733	4229	4723	5231
271	641	1033	1459	1889	2347	2791	3301	3739	4231	4729	5233

Números primos:

5237	5741	6257	6763	7283	7793	8329	8849	9391	9883	10433
5261	5743	6263	6779	7297	7817	8353	8861	9397	9887	10453
5273	5749	6269	6781	7307	7823	8363	8863	9403	9901	10457
5279	5779	6271	6791	7309	7829	8369	8867	9413	9907	10459
5281	5783	6277	6793	7321	7841	8377	8887	9419	9923	10463
5297	5791	6287	6803	7331	7853	8387	8893	9421	9929	10477
5303	5801	6299	6823	7333	7867	8389	8923	9431	9931	10487
5309	5807	6301	6827	7349	7873	8419	8929	9433	9941	10499
5323	5813	6311	6829	7351	7877	8423	8933	9437	9949	10501

5333	5821	6317	6833	7369	7879	8429	8941	9439	9967	10513
5347	5827	6323	6841	7393	7883	8431	8951	9461	9973	10529
5351	5839	6329	6857	7411	7901	8443	8963	9463	10007	10531
5381	5843	6337	6863	7417	7907	8447	8969	9467	10009	10559
5387	5849	6343	6869	7433	7919	8461	8971	9473	10037	10567
5393	5851	6353	6871	7451	7927	8467	8999	9479	10039	10589
5399	5857	6359	6883	7457	7933	8501	9001	9491	10061	10597
5407	5861	6361	6899	7459	7937	8513	9007	9497	10067	10601
5413	5867	6367	6907	7477	7949	8521	9011	9511	10069	10607
5417	5869	6373	6911	7481	7951	8527	9013	9521	10079	10613
5419	5879	6379	6917	7487	7963	8537	9029	9533	10091	10627
5431	5881	6389	6947	7489	7993	8539	9041	9539	10093	10631
5437	5897	6397	6949	7499	8009	8543	9043	9547	10099	10639
5441	5903	6421	6959	7507	8011	8563	9049	9551	10103	10651
5443	5923	6427	6961	7517	8017	8573	9059	9587	10111	10657
5449	5927	6449	6967	7523	8039	8581	9067	9601	10133	10663
5471	5939	6451	6971	7529	8053	8597	9091	9613	10139	10667
5477	5953	6469	6977	7537	8059	8599	9103	9619	10141	10687
5479	5981	6473	6983	7541	8069	8609	9109	9623	10151	10691
5483	5987	6481	6991	7547	8081	8623	9127	9629	10159	10709
5501	6007	6491	6997	7549	8087	8627	9133	9631	10163	10711
5503	6011	6521	7001	7559	8089	8629	9137	9643	10169	10723
5507	6029	6529	7013	7561	8093	8641	9151	9649	10177	10729
5519	6037	6547	7019	7573	8101	8647	9157	9661	10181	10733
5521	6043	6551	7027	7577	8111	8663	9161	9677	10193	10739
5527	6047	6553	7039	7583	8117	8669	9173	9679	10211	10753

5531	6053	6563	7043	7589	8123	8677	9181	9689	10223	10771
5557	6067	6569	7057	7591	8147	8681	9187	9697	10243	10781
5563	6073	6571	7069	7603	8161	8689	9199	9719	10247	10789
5569	6079	6577	7079	7607	8167	8693	9203	9721	10253	10799
5573	6089	6581	7103	7621	8171	8699	9209	9733	10259	10831
5581	6091	6599	7109	7639	8179	8707	9221	9739	10267	10837
5591	6101	6607	7121	7643	8191	8713	9227	9743	10271	10847
5623	6113	6619	7127	7649	8209	8719	9239	9749	10273	10853
5639	6121	6637	7129	7669	8219	8731	9241	9767	10289	10859
5641	6131	6653	7151	7673	8221	8737	9257	9769	10301	10861
5647	6133	6659	7159	7681	8231	8741	9277	9781	10303	10867
5651	6143	6661	7177	7687	8233	8747	9281	9787	10313	10883
5653	6151	6673	7187	7691	8237	8753	9283	9791	10321	10889
5657	6163	6679	7193	7699	8243	8761	9293	9803	10331	10891
5659	6173	6689	7207	7703	8263	8779	9311	9811	10333	10903
5669	6197	6691	7211	7717	8269	8783	9319	9817	10337	10909
5683	6199	6701	7213	7723	8273	8803	9323	9829	10343	10937
5689	6203	6703	7219	7727	8287	8807	9337	9833	10357	10939
5693	6211	6709	7229	7741	8291	8819	9341	9839	10369	10949
5701	6217	6719	7237	7753	8293	8821	9343	9851	10391	10957
5711	6221	6733	7243	7757	8297	8831	9349	9857	10399	10973
5717	6229	6737	7247	7759	8311	8837	9371	9859	10427	10979
5737	6247	6761	7253	7789	8317	8839	9377	9871	10429	10987

Como se vê, em nenhum dos casos é rejeitada a hipótese nula de uniformidade do número de primos nos subintervalos! Podemos mesmo observar que existe um sobre-ajustamento estranho, uma vez que o valor observado da estatística de teste é sempre um quantil de probabilidade muito baixo, muito inferior ao valor esperado 7.

Juntámos uma tabela dos 1334 primeiros números primos, para o leitor ter oportunidade de experimentar com outros primos sequenciais, e dividindo cada intervalo em 6, 7, 9 ou 10 subintervalos iguais, em vez de 8.

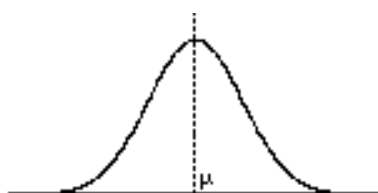
Se calhar, se procurarmos com suficiente afinco, acabamos sempre por encontrar padrões escondidos. O nosso bom-senso, em última análise, é que avalia se esses padrões têm alguma relevância ou são padrões cujo significado é irrelevante. A esta luz, será que vale a pena tirar tantas conclusões das combinações diversas das dimensões da grande pirâmide do Egípto, ou da proximidade das letras da Bíblia quando se escreve a palavra sem separação e sem vogais?

Deixem-nos citar de novo Carlyle: "Conclusive facts are inseparable from inconclusive except by a head that already understands and knows".

- A soma tem um papel regularizador (que neste caso é reforçado pelo padrão que existe nas parcelas).

O cálculo do ganho esperado do casino é relevante (como é relevante o débito de uma fonte, ou de uma torneira, e temos uma ideia suficientemente precisa para propósitos práticos de quanto tempo podemos deixar a água a correr para a banheira sem inundar a casa) devido a um resultado maravilhoso da Teoria da Probabilidade, a "lei dos grandes números", que estabelece que, sob condições muito gerais, a média de observações converge para o valor médio populacional.

O papel de média e desvio padrão é realmente notável; se não precisarmos de um modelo muito preciso, estas duas características amostrais (que são um "sinal" dos correspondentes parâmetros populacionais) permitem estabelecer limites para a probabilidade de observar a variável em intervalos centrados — é a desigualdade de Chebycheff, uma das pérolas da Probabilidade. Mais geralmente, admitindo a existência daqueles parâmetros e sob hipóteses muito gerais, o teorema limite central, cujo nome advém do papel central que ocupa na Teoria da Probabilidade, estabelece que sob condições de regularidade muito gerais, a soma de variáveis se vai aproximando de uma variável aleatória "normal" ou gaussiana, cuja função densidade de probabilidade é unimodal, em forma de sino,



evidenciando que os valores mais prováveis da soma são os valores "centrais". O teorema limite central vai assim um pouco mais longe do que a lei dos grandes números (o que não é de todo de estranhar, uma vez que assenta na existência de valor médio e de variância, enquanto a lei dos grandes números apenas requer valor médio).

De facto, expressando aqueles dois resultados limites no que respeita a média, a lei dos grandes números expressa que a média \bar{X}_n se aproxima de $m = E(X)$, no caso de o valor médio existir. De facto, $\bar{X}_{n+1} = \frac{n}{n+1} \bar{X}_n + \frac{X_{n+1}}{n+1}$; $\frac{n}{n+1} \rightarrow 1$, e se existir valor médio as caudas da distribuição são leves, pelo que $\frac{X_{n+1}}{n+1} \rightarrow 0$, e consequentemente $\bar{X}_{n+1} \approx \bar{X}_n$, ou seja \bar{X}_n vai estabilizando.

O teorema limite central, ao reforçar as hipóteses com existência de variância, vai mais longe: afirma que \bar{X}_n se vai aproximando de uma gaussiana com valor médio m (coincidente com o valor médio populacional) e com desvio padrão $\frac{\sigma}{\sqrt{n}}$ evanescente (ainda que esteja a convergir para 0 lentamente). Por outras palavras, $P[\mu - 1.96 \frac{\sigma}{\sqrt{n}} < \bar{X}_n < \mu + 1.96 \frac{\sigma}{\sqrt{n}}] = 0.95$, e $P[\mu - 3 \frac{\sigma}{\sqrt{n}} < \bar{X}_n < \mu + 3 \frac{\sigma}{\sqrt{n}}] = 0.997$. Assim, o teorema limite central vai mais longe do que a lei dos grandes números, quantificando a que ponto é provável \bar{X}_n afastar

se do seu limite m . Um terceiro resultado limite, a lei do logaritmo iterado de Khinchine, estabelece limites para as grandes flutuações.

Hoje é porventura mais compensador encarar a média como soma ponderada de estatísticas ordinais, e considerar a média aritmética apenas como uma das possíveis características amostrais de localização, a que porventura usa mais plenamente a informação disponível, mas por isso mesmo a menos resistente a erros grosseiros nos dados (pois usar eficientemente má informação conduz fatalmente a maus resultados). Neste sentido a preponderância de estatísticas ordinais centrais ou de estatísticas ordinais extremas caracteriza o tipo de teorema limite que se obtém.

Nos anos 50, Gnedenko e Kolmogoroff afirmavam que o verdadeiro valor epistemológico da Teoria da Probabilidade advém dos teoremas limites. É uma afirmação incontroversa à data em que foi publicada — hoje há que moderá-la, pois o advento da estatística computacional, e a necessidade de tratamento de pequenas amostras veio alterar os rumos da Estatística. E por outro lado os grandes teoremas limites clássicos são omissos no que respeita a velocidades de convergência, tornando a sua utilização controversa quando apenas dispomos de amostras de dimensão moderada. Uma nota de esperança: também neste campo muito se tem avançado, e continua a ser área fértil de investigação.

Bibliografia:

Gaither, C. C. and Cavazos-Gaither, A. E. (1996) "Statistically Speaking", Institute of Physics Publishing, Bristol and Philadelphia.

Gnedenko, B. V. and Kolmogoroff, A. N. (1954) Limit Theorems for Sums of Independent Random Variables. Addison-Wesley.

Hollander, M. and Proschan, F. (1984) The Statistical Exorcist — Dispelling Statistical Anxiety. M. Dekker, Inc., New York and Basel.

Mosteller, F. and Rourke, R. K. E. (1973) Sturdy Statistics. Nonparametrics and Order Statistics. Addison-Wesley, Philadelphia. (Tradução portuguesa: Estatísticas Firmes, Salamandra, Lisboa, 1993)

Pestana, D. (1990) "Como Mentir com Estatística — Um Curso Breve (Mas Intensivo). in Armadilhas dos Métodos Quantitativos, p. 27-35, FNE, Porto.