A lei do grande número, lei básica da estatística

Dr. H. Franke

Diretor do Escritório de Pesquisas da Comissão Permanente de Ação Social, São Paulo

ESTAMOS na época da estatística, a ponto de, às vezes, se confiar demasiadamente no seu valor. Irrefutaveis são, porem, as facilidades que ela nos cria para a compreensão dos fatos. Surge daí a seguinte pergunta: Qual é a base da certeza nos resultados estatísticos?

A natureza da estatística é muito discutida pelos cientistas que dela se ocupam. Uns (1) a consideram ciência própria, outros (2) método científico e outros (3) há que a classificam, na ordem científica, método e ciência ao mesmo tempo.

O que podemos afirmar, é, que, apesar das divergentes opiniões a respeito, há unanimidade entre as mesmas em reconhecer a base concreta da estatística na chamada lei do grande número. Isto quer dizer que há necessidade de uma quantidade máxima de meticulosas observações, base imprescindivel para se chegar a uma conclusão certa.

Tentaremos em seguida explicar, como se pode justificar a afirmação de que os resultados dos inquéritos estatísticos realizados de acordo com as exigências da citada lei do grande número, podem aproximar-se ao máximo da realidade.

É um jogo muito apreciado pela mocidade o de atirarem-se as moedas ao ar, afim de adivinhar se cairá "cara" ou "coroa", ganhando-se ou perdendo-se a moeda, conforme a aposta. Olhando para uma moeda e vendo-a igualmente cunhada em ambos os lados, supomos, que uma vez atirada ao ar, cairá uma vez de um lado, e outra vez do outro, formando dessa maneira os resultados a proporção de 1 por 1. Mas não é assim que realmente acontece. A moeda é capaz de cair cinco, dez vezes e mais, sempre do mesmo lado. Dizemos, então, que o resultado depende de casualidade. Mas que é casualidade neste sentido? A posição da moeda antes, a maneira de pegá-la, de atirá-la a de cair, etc. Cada um desses fatos depende ainda de uma série de outros, inverificaveis pelos sentidos humanos: particularidades da moeda e da mão que a pega, do ar atravessado e do chão. Casualidade aquí é por isso a cooperação de numerosos pequenos fatos indeterminaveis na sua totalidade. O acaso é considerado por isso em geral como uma cousa mística e sobrenatural. No entanto, não é tão misterioso e irregular como se pensa, mostrando até certas regularidades. Aquelas se provam facilmente por meio de seguinte experiência: atiremos simultaneamente dez moedas ao ar, e isso por várias vezes; marquemos os resultados obtidos depois da queda das moedas. Poderíamos supor que haveria igual probabilidade de cair tanto "cara" quanto "coroa". Na realidade, tal não se dá, pois se dão todas as combinações possiveis, isto é, 10 "caras" por 0 "coroas", 9 por 1, 8 por 2, etc. 0 "caras" por 10 "coroas". Numa experiência feita com cem lances de 10 moedas obtivemos o resultado que apresentamos na seguinte tabela:

---

(1) Filippo Virgilii: *Manual de Estatística*.

(2) Armand Julin: *Précis du cours de statistique générale et appliquée*.

(3) Adolf Wagner: *Begriff und Grensen der Statistik* (O conceito e os limites da Estatistica).

TABELA I

	1.º	2.º	3.º	4.º	5.º	6.º	7.º	8.º	9.º	10.º	
	LANCE COM 10 MOEDAS										
aparece	6	4	6	8	6	7	3	8	6	5	vezes "cara" na 1.ª dezena dos lances com 10 moedas
»	4	1	6	1	6	5	7	3	7	5	» » » 2.ª » » » » » »
»	5	5	6	4	4	4	6	8	4	4	» » » 3.ª » » » » » »
»	5	2	6	4	3	5	5	6	5	3	» » » 4.ª » » » » » »
»	4	5	4	5	5	4	6	5	3	5	» » » 5.ª » » » » » »
»	5	3	5	5	3	4	6	5	5	7	» » » 6.ª » » » » » »
»	4	4	3	5	4	5	4	6	8	6	» » » 7.ª » » » » » »
»	5	4	2	5	8	5	5	4	5	5	» » » 8.ª » » » » » »
»	5	7	6	6	7	4	5	3	3	4	» » » 9.ª » » » » » »
»	5	3	5	6	2	6	3	5	5	6	» » » 10.ª » » » » » »

A primeira impressão que suscitam tais algarismos é de uma irregularidade completa. Pelo exame, porem, da frequência com que obtivemos o resultado "cara", chegaremos aos resultados seguintes:

Combinações possiveis: 0 (10) (*), 1 (9), 2 (8), 3 (7), 4 (6), 5 (5), 6 (4), 7 (3), 8 (2), 9 (1), 10 (0).

Frequência na experiência: 2 3 12 20 33 19 6 5 — — vezes.

Como é facil de ver, a combinação cara/coroa 5 por 5 apresentou-se 33 vezes em 100 lances, portanto com frequência maior de que a dos outros casos. Vemos tambem que as outras combinações se apresentam tanto menos quanto mais se afastam da média de 5 por 5. Os desvios à esquerda e à direita dessa média são quase iguais, dando o aspecto da uma sucessão simétrica.

Essa última reflexão nos dá a idéia fundamental da estatística: considerando em lugar de lances separados de dez moedas, o conjunto de tais lances, essa forma fundamental deve exprimir-se muito mais claramente, pois os desvios acima e abaixo da média anular-se-ão tanto mais quanto maior for a quantidade dos lances. A prova deste fato, obteremos somando os lances acima indicados de tal maneira como se tivessem sido atirados 10, 20, 30..... 100 ...... 1.000 moedas em seguida. Usando o nosso exemplo anterior, se tomarmos 10 lances por base, teremos os seguintes desvios da média 5 por 5:

TABELA II

ATIRANDO-SE	CONFORME A TABELA N. 1 — NA MÉDIA DE CADA DEZENA DOS LANCES	APARECEU A CARA DAS MOEDAS	POR CONSEGUINTE, A PERCENTAGEM DO DESVIO (**) É
10 moedas........	—	6 vezes	20%
20 »	(6+4) : 2 =	5 »	—
30 »	(10+6) : 3 =	5,3 »	6%
40 »	(16+8) : 4 =	6 »	20%
50 »	(24+6) : 5 =	6 »	20%
60 »	(30+7) : 6 =	6,2 »	24%
70 »	(37+3) : 7 =	5,7 »	14%
80 »	48 : 8 =	6 »	20%
90 »	54 : 9 =	6 »	20%
100 »	59 : 10 =	5,9 »	18%
110 »	60 : 11 =	5,7 »	14%
120 »	64 : 12 =	5,3 »	6%
130 »	70 : 13 =	5,4 »	8%
140 »	71 : 14 =	5,1 »	2%
150 »	77 : 15 =	5,1 »	2%
160 »	82 : 16 =	5,1 »	2%
170 »	89 : 17 =	5,2 »	4%
180 »	92 : 18 =	5,1 »	2%
190 »	99 : 19 =	5,2 »	2%
200 »	104 : 20 =	5,2 »	4%
300 »	154 : 30 =	5,1 »	2%
400 »	198 : 40 =	4,9 »	4%
500 »	244 : 50 =	4,9 »	2%
600 »	292 : 60 =	4,8 »	4%
700 »	341 : 70 =	4,9 »	2%
800 »	389 : 80 =	4,9 »	2%
900 »	439 : 90 =	4.9 »	2%
1.000 »	485 : 100 =	4,9 »	2%

(*) Fóra do parentesis — "cara", dentro do parentesis — "coroa".

(**) Por percentagem do desvio se compreende aquí o afastamento da norma de 5 lances mostrando a "cara" quando se atiram 10 moedas ao ar. Quando ela aparece, por exemplo, 6 vezes, o desvio é 1; e 1 é 20% de 5.

Analisemos agora a última coluna da tabela n. 2, quer dizer as percentagens dos desvios, para chegarmos a entender o resultado obtido pelas adições. A coluna começa com grandes percentagens que vão diminuindo continuadamente, pois de 20% que era no princípio passou a 2% com 1.000 moedas. Entretanto, é sempre possivel haver interrupções casuais no declínio das percentagens, como bem mostra a tabela n. 2.

Assim se provou, que por causa da simetria dos desvios, os mesmos anulam uns aos outros, de maneira tanto mais perfeita, quanto maior o número das experiências. Isto demonstra que os resultados se aproximam sempre mais da proporção ideal, isto é, da proporção 5 por 5. Só o grande número de lances a faz aparecer, eliminando pela multidão de casos os desvios casuais.

Deixando o exemplo das moedas e voltando a olhar para o nosso mundo ambiente, verificamos tambem nele a grande importância do acaso, especialmente no tocante a fenômenos de ordem natural e social. Assim, como às vezes todas as moedas atiradas mostraram o mesmo lado e entre esses dois extremos houve todas as combinações possiveis, pode ser que numa família só nascerão meninos e noutra só meninas. Considerando, porem, a totalidade dos nascimentos em um país, o grande número obtido pela adição de todos os casos singulares eliminará a influência da casualidade. Reconheceremos, assim, a verdadeira proporção dos sexos entre os nascidos, que será quase de 1 por 1. — Consideremos agora a maneira pela qual famílias compostas de um determinado número de pessoas e de determinada classe social empregam a sua renda. Encontraremos famílias que fazem grandes despesas com a alimentação pois dão grande importância à sua boa qualidade, e outras que fazem despesas menores nesse sentido, preferindo bonitas moradias, boas roupas, livros, diversões, etc. Levando-se em conta muitos casos diferentes mostrar-se-á o gasto médio com a alimentação e manifestar-se-ão certas regularidades imperceptiveis em casos isolados, por exemplo, verse-á assim ser o gasto com a alimentação tanto menor quanto maior a renda da família, que se chama a regra de Engel (estatístico e economista alemão, que viveu em meiados do século passado em Berlim).

É de importância primordial para a compreensão da lei do grande número conhecer os "motivos casuais": A frequência nos meios de transportes, por exemplo, em um trem subúrbio de uma grande cidade, depende de motivos constantes e gerais, bem como de motivos casuais: pertencem aos primeiros a profissão dos viajantes, a influência de dias feriados, etc. Entretanto, os imprevistos da vida dos viajantes, por exemplo, enterros, casamentos de parentes e amigos representam os motivos casuais para as viagens. Aplicando-se a analogia à agricultura, teremos motivos gerais para uma má colheita na qualidade inferior da terra, no clima, no modo de plantar, na incapacidade dos agricultores. Os motivos casuais serão: tempo desfavoravel, greve dos trabalhadores, etc.

Só podemos falar em motivos casuais quando a probabilidade de oscilação não se limita a uma só direção. Efetuando-se um desvio da média sempre no mesmo sentido, como por exemplo a tendência frequente aos contribuintes de declararem às autoridades uma quantidade menor de que a sua renda, ou a tendência de um povo que corresponde à raça dominante de declarar pequeno o número dos indivíduos de outras raças dentro do território nacional em um país determinado, nesse caso não se fale de uma oscilação casual, mas sim de um desvio unilateral e deliberado.

Dessas reflexões tira-se a conclusão importante de que, *alem dos fatos isolados influenciados pela casualidade, existem ainda fatos gerais de importância mais decisiva baseados em motivos principais, que só se patenteam quando examinamos um grande número de casos. Nessa conclusão se baseia a chamada lei do grande número.*

Poder-se-ia perguntar, quando começaria o grande número, em que aparecem claramente os característicos da multidão. Para isso não se pode estabelecer um limite. Cada caso mais que se toma em conta, ajuda a aniquilar a influência da casualidade, conseguindo-se assim a verdadeira proporção, isto é, a média. Para saber o grau de exatidão dos resultados calculados e conhecer os limites dos erros do cálculo das probabilidades, foram criadas fórmulas. Usando-as, obteremos os resultados que apresentamos na tabela a seguir:

TABELA III

para	1	caso	de	10 casos, o desvio	máximo é até	95 %	do valor da média
»	1	»	»	100 » » »	» » »	30 %	» » » »
»	1	»	»	1.000 » » »	» » »	10 %	» » » »
»	1	»	»	10.000 » » »	» » »	3 %	» » » »
»	1	»	»	100.000 » » »	» » »	1 %	» » » »
»	1	»	»	1.000.000 » » »	» » »	0,3%	» » » »

Nesta tabela, as percentagens indicadas para os desvios são as maiores possiveis. Pode-se afirmar com uma probabilidade de 2:1, que, na realidade, o desvio máximo será só até um terço das percentagens indicadas nesta tabela. E isso aconteceu tambem (ver tabela n. 2) na nossa experiência com 10 e 1000 lances, porem, não com 100 lances (sendo 20% menos de que um terço do desvio máximo de 95%, 2% menos de que um terço de 10%, em quanto 18% é maior de que um terço de 30%).

Resta ainda algo a dizer sobre o descrescimento dos desvios. Eles não diminuem tão depressa quanto aumenta o número dos lances. Sendo o número das observações em nosso exemplo sempre o décuplo do número anterior, os desvios deveriam restringir-se cada vez à décima parte. Mas de fato, como resulta da tabela n. 3, essa diminuição importa não em nove décimos, mas somente em um terço ou mais: exato $\frac{1}{\sqrt{10}}$ (1 por raiz quadrada de dez) $= \frac{1}{3,17}$; de maneira que a exatidão não cresce com o número dos lances, mas sim com a raiz quadrada deste número. (No esquema dos desvios acima indicados, as percentagens foram arredondadas para número inteiro, razão porque esse cálculo tem só uma exatidão relativa) É claro tambem, que o erro somente poderia chegar a zero % o que significaria o aparecimento da verdadeira proporção e eliminação de influências casuais caso em que for infinito o denominador da fração premencionada o que realmente não pode suceder. Lembre-se que o número de baixo da raiz, indica a quantidade dos casos observados.

Essa incapacidade de realizar rigorosamente as condições do grande número esse "pecado original" da estatística, só nos permite falar de uma exatidão relativa de algarismos estatísticos. Entretanto, quando se trata de números muito grandes, de milhares e milhões, o máximo desvio possivel ficará tão pequeno que praticamente é desprezivel. Com muito mais razão se-lo-á o erro provavel isto é, o erro real, o qual, como já dissemos, não ultrapassa um terço do desvio calculado.

Precisa-se ainda acrescentar que os limites dos desvios são peculiares, em primeira linha, aos resultados devidos a motivos casuais, como lances de moedas e jogos em geral. Quando porem, em assuntos da vida humana, alem dos motivos casuais, a vontade do homem vem influenciar o resultado, os desvios podem ser maiores ou menores, conforme o caso. Os limites indicados na tabela n. 3 podem, entretanto, servir para a verificação da exatidão dos resultados obtidos.

A base dos resultados estatísticos é, como mostrámos, a multidão, fortuitamente formada. Daí deriva que o valor de tais observações da multidão deve ser diferente daquele que se obtem quando nos limitamos apenas a fatos singulares. Uma vez conhecidos os resultados de uma grande massa de observações de manifestações individuais de um fenômeno determinado, pode se predizer, até certo ponto, a maneira de sua manifestação num novo caso isolado.

As conclusões da estatística não levam a certezas por causa do carater particular da matéria, mas sim a probabilidades. A estatística, portanto, não estabelece regras, mas mostra as regularidades dos fatos. Estas já permitem conclusões muito valiosas, pois uma estatística baseada em um número muito elevado de observações não se afastará muito das circunstâncias reais, desde que se tenha em vista a influência do acaso sobre cada um dos elementos do inquérito.

As ciências físicas, naturais, políticas e sociais servem-se da estatística. Pode-se afirmar categoricamente que sem o concurso da estatística muita hipótese não se transformaria em fato, nem certas previsões poderiam converter-se em realidade. Eis aí a grande importância da lei do grande número.