

# Quanto mais melhor?

DAVI NOBRE\*

Instituto de Física de São Carlos  
Universidade de São Paulo  
dv.nobr@gmail.com

---

## Resumo

*O imaginário popular diz que, quando tentamos estimar uma grandeza, a média de diversos palpites deve se aproximar do valor real quando se aumenta a quantidade de estimativas individuais. O que não era de se esperar é que esse mesmo imaginário faz com que grupos errem sistematicamente o palpite.*

---

## 1. A SABEDORIA DAS MASSAS

Sabedoria das massas é o termo utilizado para se referir ao fenômeno de agregação de informação em grupos que, de acordo com Surowiecki [1], são capazes de realizar estimativas melhores do que as que poderiam ser feitas por qualquer membro do grupo. A primeira pessoa a estudar esse fenômeno foi o britânico Francis Galton, em 1907, quando esteve em uma feira rural onde houve um concurso para se tentar acertar o peso de um boi que estava sendo exibido. Então, coletou os palpites dos diversos participantes e verificou que a mediana das estimativas acertava o valor exato com erro inferior a 1%. A partir disso, Galton introduziu o conceito de sabedoria das massas, ou, nas palavras dele, *Vox Populi*.

Até hoje não se sabe exatamente quais são as condições necessárias para que o fenômeno ocorra, apesar de algumas teorias apontarem que basta que o palpite coletivo seja feito por uma quantidade suficientemente grande de indivíduos diferentes entre si. Scott Page, por exemplo, defende que a estimativa coletiva pode aproximar-se arbitrariamente do valor real da grandeza, sendo necessário apenas que o grupo seja diversificado o bastante [3]. Para isso, Page define o palpite coletivo,  $\langle P \rangle$ , de um grupo com  $N$  indivíduos simplesmente como a média aritmética dos  $N$  palpites individuais, de modo que

$$\langle P \rangle = \frac{1}{N} \sum_{i=1}^N P_i. \quad (1)$$

Agora, considerando que  $T$  é o valor real da grandeza que está sendo estimada, Page define o erro quadrático médio como

$$\varepsilon = \frac{1}{N} \sum_{i=1}^N (P_i - T)^2, \quad (2)$$

---

\*Agradeço ao Prof. José Fontanari (IFSC-USP), pela oportunidade de participar do projeto, e ao Prof. Leandro Aurichi e à Profa. Marina Andretta pelo convite para escrever este artigo. Este projeto foi financiado pela FAPESP (processo 2017/08475-9).

e a diversidade dos palpites como

$$\delta = \frac{1}{N} \sum_{i=1}^N (P_i - \langle P \rangle)^2. \quad (3)$$

Já o erro quadrático coletivo é dado por

$$\gamma = (\langle P \rangle - T)^2, \quad (4)$$

de modo que

$$\gamma = \varepsilon - \delta. \quad (5)$$

O que Page propõe, portanto, é que aumentando arbitrariamente a diversidade dos palpites, que é uma medida da diferença dos indivíduos entre si, pode-se fazer com que o erro quadrático coletivo seja arbitrariamente próximo de 0, de modo que, pela equação (4),  $\langle P \rangle \rightarrow T$ . Entretanto, para que essa conclusão seja sempre verdadeira, é necessário que as variáveis  $\varepsilon$  e  $\delta$  sejam independentes, o que, como será mostrado adiante, não é verdade.

## 2. O MODELO

O modelo utilizado neste trabalho foi proposto por Nash [4, 5] que, ao contrário do que espera o senso-comum, diz que o palpite coletivo só é igual ao valor real da grandeza quando este corresponde exatamente à expectativa cultural do grupo para esta grandeza, que é simplesmente o valor que toda a população espera que a grandeza assuma caso não tenha conhecimento do objeto cuja grandeza se deseja estimar.

Por exemplo, quando o valor estimado é o peso de um boi, o modelo supõe que todo indivíduo diria um mesmo valor  $\mu_T$  caso não visse o boi, ou seja, que os indivíduos têm uma ideia comum do peso de um boi. Partindo desse valor comum a todos, cada indivíduo compara o objeto real ao objeto esperado, buscando o que se chama de regiões de interesse, que fornecem dicas se o valor estimado está acima ou abaixo da expectativa cultural. Por exemplo, para estimar o peso do boi, observa-se as dimensões do animal, como altura e comprimento, e algumas partes do corpo relevantes, como as costelas, com a intenção de comparar o peso real do boi com o peso esperado.

Então, por exemplo, digamos que as pessoas esperem que um boi pese 500kg; ao observar a altura, se o indivíduo julga que a altura real do boi é maior que o esperado, ele adiciona 50kg ao peso total do animal, se ele julga que a altura é menor que o esperado, ele subtrai 50kg do valor total. O mesmo acontece se ele julga que as costelas do boi estão mais ou menos visíveis que o esperado, adicionando ou subtraindo 50kg, respectivamente. Considerando, então, que o objeto com valor a ser estimado possui um total de  $C$  regiões de interesse, assumimos que há  $N_+$  regiões que contribuem para aumentar o valor da estimativa (chamadas de regiões acima da média, ou positivas), e  $N_-$  regiões que contribuem para diminuir o valor da estimativa (regiões abaixo da média, ou negativas), com o vínculo  $N_+ + N_- = C$ . Os parâmetros  $C$ ,  $N_+$  e  $N_-$  são intrínsecos ao objeto em estudo, sendo que  $C$  depende da classe do objeto enquanto  $N_+$  e  $N_-$  dependem do objeto individualmente. Por exemplo, para estimar o peso de todo boi, supostamente, sempre leva-se  $C$  regiões de interesse em consideração, mas os valores de  $N_+$  e  $N_-$  dependem do boi que

está sendo analisado, sujeitos ao vínculo  $C = N_+ + N_- = cte$ . Sendo  $\mu_T$  a expectativa cultural para o valor estimado, temos que o valor real da grandeza,  $T$ , será dado por

$$T = \mu_T + \alpha t, \quad (6a)$$

$$t = N_+ - N_- = \sum_{k=1}^C j_k, j_k = \begin{cases} +1, & \text{se a característica for positiva.} \\ -1, & \text{se a característica for negativa.} \end{cases} \quad (6b)$$

Veja que  $\alpha$  simplesmente determina quanto cada região de interesse influencia no valor final da estimativa, uma vez que se soma ou se subtrai  $\alpha$  dependendo se a região é positiva ou negativa, respectivamente ( $j_k = \pm 1$ ). Observe também que  $\alpha$  é igual para todas as  $C$  regiões de interesse.

Estabelecidos esses valores, cada indivíduo realiza o processo de categorização das regiões de interesse, ou seja, determina se cada característica é positiva ou negativa com relação ao esperado. Considera-se que um indivíduo tem uma probabilidade  $p$  de acertar a categoria de uma determinada região  $e$ , conseqüentemente, uma probabilidade  $1 - p$  de errar. Assim, o palpite individual é dado por

$$P_i = \mu_T + \alpha e_t, \quad (7a)$$

$$e_t = \sum_{k=1}^C i_k, i_k = \begin{cases} j_k, & \text{com probabilidade } p. \\ -j_k, & \text{com probabilidade } 1 - p. \end{cases} \quad (7b)$$

Perceba que a variável aleatória  $e_t$  é análoga à variável determinística  $t$ . Ambas são uma medida normalizada da diferença entre um valor e a expectativa cultural, mas, enquanto  $t$  determina a diferença entre o valor real da grandeza e a expectativa cultural,  $e_t$ , determina a diferença entre um palpite individual e a expectativa cultural, podendo assumir diversos valores por ser aleatória. Imaginemos, por exemplo, um objeto com  $C = 30$  regiões de interesse, sendo  $N_+ = 20$  positivas e  $N_- = 10$  negativas, e  $p = 0.7$ . Como  $t = N_+ - N_-$  sempre, ele é bem determinado e igual a 10 nesse caso, já  $e_t$  não tem um valor certo, podendo assumir qualquer valor inteiro entre  $-C$  e  $C$ , mas sabemos calcular o valor médio de  $e_t$ . Para isso, basta considerar o processo descrito acima, como  $p = 0.7$ , a população categoriza corretamente 70% das 30 regiões de interesse, ou seja, das 20 regiões positivas, 14 a população caracterizará como positivas (acertando), e 6 como negativas (errando), enquanto das 10 regiões negativas, 7 serão classificadas corretamente como negativas e 3 erroneamente como positivas. Somando as regiões categorizadas como positivas e subtraindo as categorizadas como negativas, obtém-se  $\langle e_t \rangle = 14 - 6 - 7 + 3 = 4$ , que é o valor médio de  $e_t$ , o valor real que essa variável assume depende de cada indivíduo.

Então, se o boi que se deseja acertar o peso é mais alto que a expectativa cultural, o  $j_k$  associado à altura é  $+1$  (lembre-se que as variáveis  $j_k$  são determinísticas), e, para obter  $T$ , deve-se somar  $\alpha$  ao se analisar essa característica. Contudo, como humanos são sujeitos a erro, existe uma probabilidade  $1 - p$  do indivíduo dizer que, na verdade, o boi em análise é mais baixo que o esperado, e atribuir um valor  $i_k = -j_k = -1$  para a altura, subtraindo  $\alpha$  ao invés de somar, e cometendo um erro. Seguindo esse processo de categorização, determina-se que a distribuição da variável  $e_t$  segue

$$P(e_t) = \sum_{j=\max\{0, \frac{e_t+t}{2}\}}^{\min\{\frac{C+e_t}{2}, \frac{C+t}{2}\}} \binom{\frac{C+t}{2}}{j} \binom{\frac{C-t}{2}}{\frac{C+e_t}{2}-j} p^{2j-\frac{e_t+t}{2}} (1-p)^{C+\frac{e_t+t}{2}-2j}. \quad (8)$$

A escolha das variáveis da equação (8) é arbitrária,  $C$  e  $t$  podem ser substituídas por  $(N_+ + N_-)$  e por  $(N_+ - N_-)$  sem perda de informação, enquanto  $e_i$  e  $p$  não têm variáveis análogas. Nas equações (7), a única variável aleatória é  $e_i$ , pois, para uma dada população,  $\mu_T$  e  $\alpha$  são constantes. Assim, para calcular o palpite coletivo (média dos palpites individuais) basta calcular a média da função massa de probabilidade dada por (8), que é

$$\langle e_i \rangle = (2p - 1)t. \quad (9)$$

Considerando que o erro individual é  $e_i = T - P_i$ , o erro coletivo será dado por

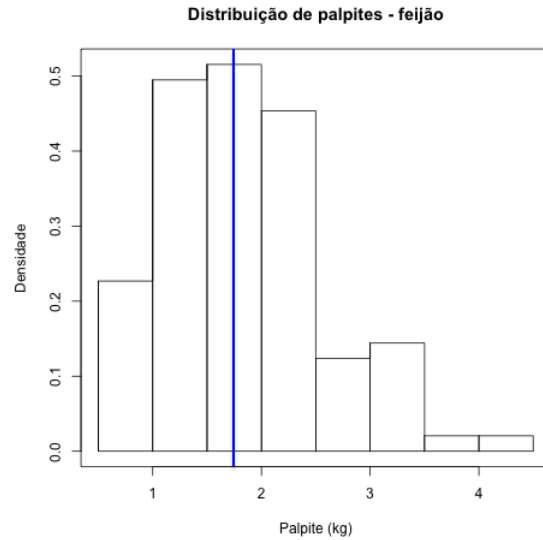
$$e_c = \langle e_i \rangle = 2\alpha t(1 - p). \quad (10)$$

A equação (10) nos diz que o erro coletivo só é zero em dois casos: ou quando  $p = 1$ , e todas as pessoas acertam exatamente o valor da grandeza, o que é quase impossível, ou quando  $t = 0$ , e o valor da grandeza corresponde exatamente à expectativa cultural do grupo para ele, o que, considerando uma grandeza contínua, também tem mínimas chances de acontecer.

### 3. EXPERIMENTOS

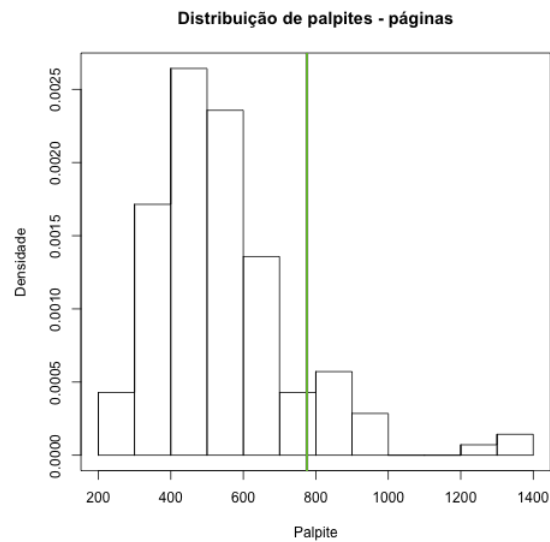
Nesta seção, serão apresentados os resultados de dois experimentos que foram feitos para testar a validade do modelo adotado. Em ambos os experimentos, pediu-se para um grupo de indivíduos, composto basicamente por alunos de graduação do Instituto de Física de São Carlos, estimar uma grandeza, sendo que os participantes davam palpites independentes entre si, ou seja, não conheciam os palpites anteriores.

No primeiro experimento, a grandeza a ser estimada era o peso de um pote de feijão (cilíndrico, de aproximadamente 20cm de diâmetro e 20cm de altura), e foram colhidos 97 palpites. A média dos palpites foi de  $\langle P \rangle = 1.9101kg$ , sendo que o valor real era  $T = (1.7452 \pm 0.0001)kg$ , o que representa um erro coletivo de 9,4%. A Figura 1 a seguir mostra o histograma dos palpites individuais para este caso.



**Figura 1:** Distribuição de estimativas individuais no experimento com o pote de feijões. A barra azul indica a posição do valor real do peso do pote.

Já no segundo experimento, perguntou-se a quantidade de páginas de um livro e colheu-se 140 palpites. A média dos palpites foi de  $\langle P \rangle = 560$ , sendo que o valor real era  $T = 784$ , levando a um erro coletivo de 28,6%. A Figura 2 a seguir mostra o histograma dos palpites individuais para este caso.



**Figura 2:** Distribuição de estimativas individuais no experimento com o livro. A barra verde indica a posição da quantidade real de páginas.

## 4. ANÁLISE DOS DADOS

O modelo de Nash nos fornece uma previsão para a distribuição de erros individuais, e não de palpites. Assim, para testá-lo, o primeiro passo é fazer o histograma de erros individuais para então ajustar a distribuição (8). Ao fazer isso, determina-se o valor de  $p$ ,  $N_+$  e  $N_-$  para cada problema. Na Figura 3 a seguir estão os histogramas dos erros individuais com o ajuste da função dada por (8).

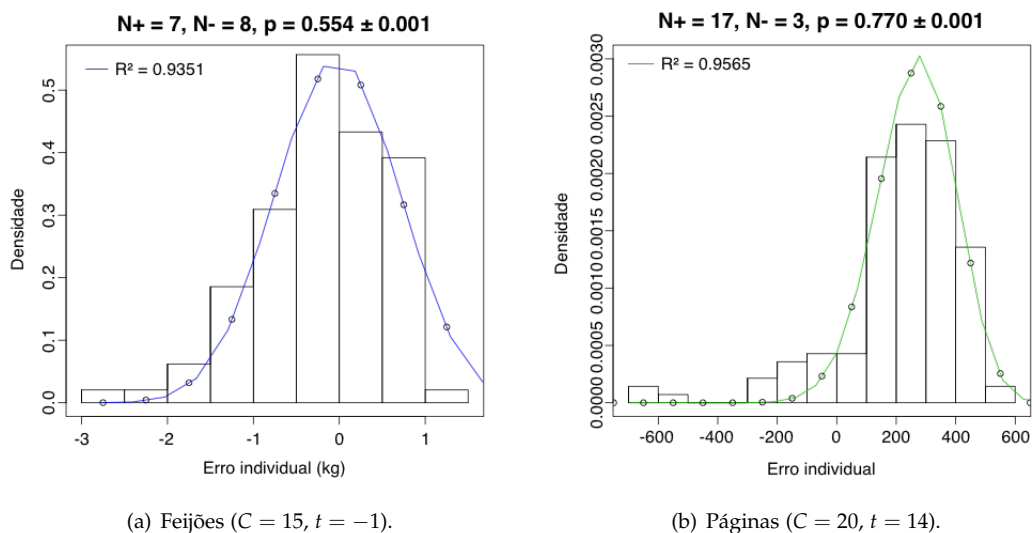


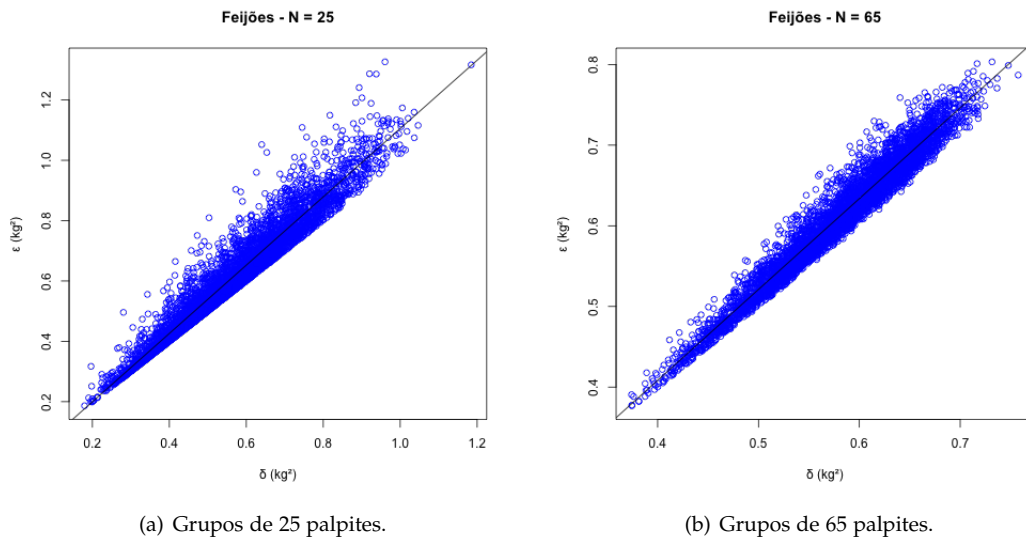
Figura 3: Distribuição de erros individuais.

Para ajustar a função massa de probabilidade (8), que descreve uma variável aleatória discreta, a erros individuais contínuos, ligou-se os pontos definidos por (8) por retas e renormalizou-se a função, de forma que a soma de todos os possíveis valores de erro individual permanecesse igual a 1.

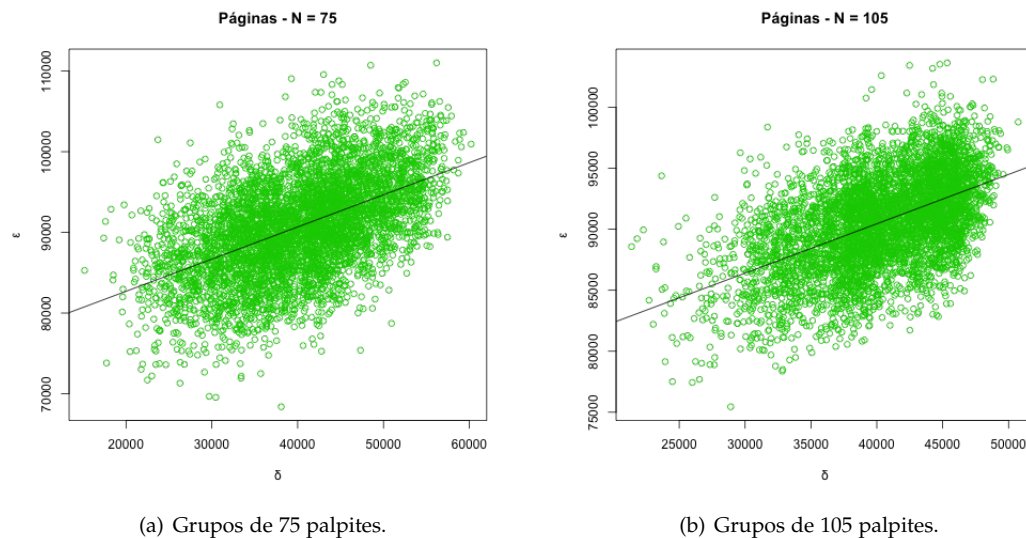
Atentando-se para os valores de  $p$  em cada um dos ajustes da Figura 3 e lembrando que  $p$  determina a competência do grupo em resolver o problema, é coerente que, no experimento dos livros este valor seja superior ao do experimento dos feijões, já que o grupo era composto por estudantes e livros estão mais presentes no seu dia a dia que potes de feijão. Contudo, pode parecer estranho que o erro coletivo no caso das páginas tenha sido maior do que no caso dos feijões, dado que o grupo resolve o problema das páginas do livro com uma probabilidade de acerto,  $p$ , maior que o problema do peso dos feijões. Isso acontece devido à variável  $t$ , que determina, a menos de um fator de escala e dimensão, a diferença entre o valor real da grandeza e a expectativa cultural para ele. Observando os valores de  $N_+$  e  $N_-$  percebe-se que, de fato, no caso dos feijões, o objeto mostrado ao grupo estava bem mais próximo dos que eles esperavam ( $t = -1$ ) que no caso das páginas ( $t = 14$ ), ou seja, mesmo que as pessoas não soubessem estimar o peso do pote de feijão, elas palpitavam próximo da expectativa cultural e erravam pouco.

Agora, o que queremos elucidar é quais mecanismos podemos usar para reduzir o erro coletivo. Segundo Page, isso pode ser feito simplesmente aumentando a diversidade de palpites ( $\delta$ ), o que

é facilmente obtido diversificando o grupo. Contudo, aumentar  $\delta$  para reduzir o erro coletivo,  $\gamma$ , pressupõe que o erro individual médio,  $\epsilon$ , não cresce quando se aumenta a diversidade dos palpites, ou pelo menos que cresce menos que  $\delta$ . Para verificar que isso não acontece, apresentamos as Figuras 4 e 5, que ilustram o erro quadrático médio em função da diversidade dos palpites para os experimentos dos feijões e do livro, respectivamente.



**Figura 4:** Erro quadrático médio em função da diversidade dos palpites para o experimento do pote de feijões.



**Figura 5:** Erro quadrático médio em função da diversidade dos palpites para o experimento do livro.

Das Figuras 4 e 5, percebemos que não podemos aumentar  $\delta$  sem aumentar  $\varepsilon$ , uma vez que essas variáveis não são independentes. Além disso, percebe-se que  $\varepsilon$ , na média, cresce mais rápido que  $\delta$ . Assim, seguir a proposta de Page, usualmente, aumentará o erro coletivo.

Uma vez refutada a hipótese de Page, investigou-se como o tamanho,  $N$ , do grupo afeta o erro coletivo. Para isso, foram gerados 100 mil grupos aleatórios variando o valor de  $N$ . Para cada valor de  $N$ , verificou-se quantos cometiam um erro coletivo inferior a 5%. A Figura 6 apresenta os resultados obtidos.

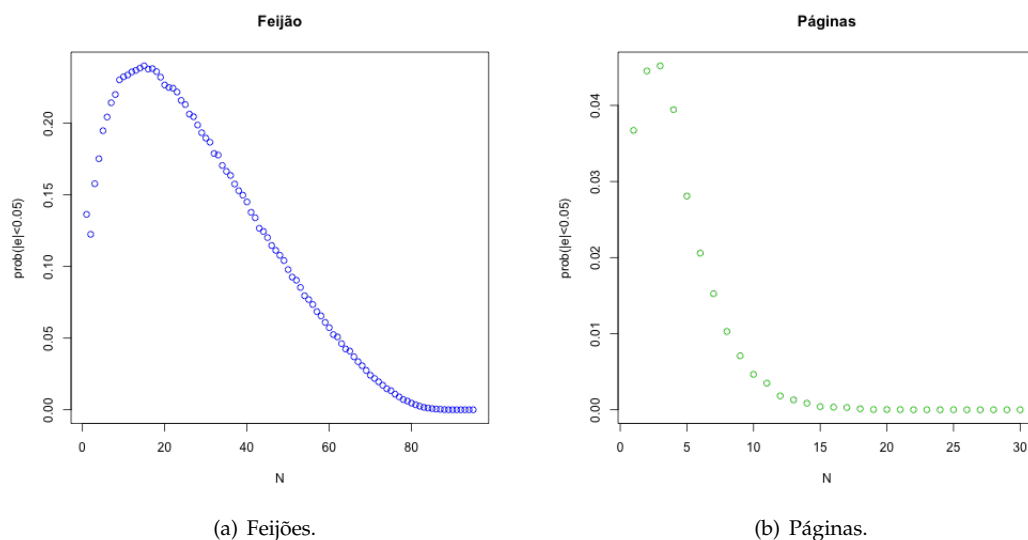


Figura 6: Probabilidade de erro coletivo inferior a 5% para grupos de tamanho  $N$ .

O primeiro ponto a se notar é que, à medida que  $N$  se aproxima do total de palpites coletados, a probabilidade de erro coletivo inferior a 5% tende a zero. Isso decorre do fato do erro coletivo cometido pelo grupo completo ser superior a 5% em ambos os casos. O mais interessante dos gráficos da Figura 6 é que a probabilidade não é estritamente decrescente e que há um tamanho específico de grupo que maximiza as chances do palpite coletivo ser mais preciso. Esse valor de  $N$  depende de diversas características do problema, incluindo o valor da grandeza estimada e a população que compõe o grupo.

## 5. CONCLUSÕES

Nem sempre grupos muito grandes fornecem palpites coletivos mais precisos, mas, em geral, há um tamanho específico de grupo que resolve o problema melhor que grupos de outros tamanhos. Isso se dá justamente porque o palpite coletivo sempre contém um erro sistemático, pois  $p \neq 1$  sempre, já que humanos falham, e  $t \neq 0$  sempre que se considera uma grandeza contínua. Em suma, nossa expectativa de que a média de muitos palpites leva a um valor muito preciso para a grandeza que está sendo estimada é errada, o que é mais um exemplo de erro na sabedoria das massas.



### REFERÊNCIAS

- [1] Surowiecki, J. *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations* (Anchor Books, 2005).
- [2] Galton, F. Vox Populi. *Nature* 75 (450)(1907).
- [3] Page, S. E. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton University Press, 2007).
- [4] Nash, U. W. The Curious Anomaly of Skewed Judgment Distributions and Systematic Error in the Wisdom of Crowds. *PLoS ONE* 9(11):e112386 (2014).
- [5] Nash, U. W. Sequential sampling, magnitude estimation, and the wisdom of crowds. *JMP* 77 (165-179) (2017).