A reprint from
American Scientist
the magazine of Sigma Xi, The Scientific Research Society

The Most Dangerous Equation

Ignorance of how sample size affects statistical variation has created havoc for nearly a millennium

Howard Wainer

What constitutes a dangerous equation? There are two obvious interpretations: Some equations are dangerous if you know them, and others are dangerous if you do not. The first category may pose danger because the secrets within its bounds open doors behind which lies terrible peril. The obvious winner in this is Einstein's iconic equation $e = mc^2$, for it provides a measure of the enormous energy hidden within ordinary matter. Its destructive capability was recognized by Leo Szilard, who then instigated the sequence of

Figure 1. Trial of the pyx has been performed since 1150 A.D. In the trial, a sample of minted coins, say 100 at a time, is compared with a standard. Limits are set on the amount that the sample can be over- or underweight. In 1150, that amount was set at 1/400. Nearly 600 years later, in 1730, a French mathematician, Abraham de Moivre, showed that the standard deviation does not increase in proportion to the sample. Instead, it is proportional to the square root of the sample size. Ignorance of de Moivre's equation has persisted to the present, as the author relates in five examples. This ignorance has proved costly enough that the author nominates de Moivre's formula as the most dangerous equation.

events that culminated in the construction of atomic bombs.

Supporting ignorance is not, however, the direction I wish to pursue—indeed it is quite the antithesis of my message. Instead I am interested in equations that unleash their danger not when we know about them, but rather when we do not. Kept close at hand, these equations allow us to understand things clearly, but their absence leaves us dangerously ignorant.

There are many plausible candidates, and I have identified three prime examples: Kelley's equation, which indicates that the truth is estimated best when its observed value is regressed toward the mean of the group that it came from; the standard linear regression equation; and the equation that provides us with the standard deviation of the sampling distribution of the mean—what might be called de Moivre's equation:

$$\sigma_{\bar{x}} = \sigma / \sqrt{n}$$

where $\sigma_{\bar{x}}$ is the standard error of the mean, σ is the standard deviation of the sample and n is the size of the sample. (Note the square root symbol, which will be a key to at least one of the misunderstandings of variation.) De Moivre's equation was derived by the French mathematician Abraham de Moivre, who described it in his 1730 exploration of the binomial distribution, *Miscellanea Analytica*.

Ignorance of Kelley's equation has proved to be very dangerous indeed, especially to economists who have interpreted regression toward the mean as having economic causes rather than merely reflecting the uncertainty of prediction. Horace Secrist's *The Triumph of Mediocrity in Business* is but one example listed in the bibliography. Other examples of failure to understand Kelley's equation exist in the sports world, where the expression "sophomore slump" merely describes the likelihood of an average season following an especially good one.

The familiar linear regression equation contains many pitfalls to trap the unwary. The correlation coefficient that emerges from regression tells us about the strength of the linear relation between the dependent and independent variables. But alas it encourages fallacious attributions of cause and effect. It even encourages fallacious interpretation by those who think they are being careful. ("I may not be able to believe the exact value of the coefficient, but surely I can use its sign to tell whether increasing the variable will increase or decrease the answer.") The linear regression equation is also badly non-robust, but its weaknesses are rarely diagnosed appropriately, so many models are misleading. When regression is applied to observational data (as it almost always is), it is difficult to know whether an appropriate set of predictors has been selected—and if we have an inappropriate set, our interpretations are questionable. It is dangerous, ironically, because it can be the most useful model for the widest variety of data when wielded with caution, wisdom and much interaction between the analyst and the computer program.

Yet, as dangerous as Kelley's equation and the common regression equations are, I find de Moivre's equation more perilous still. I arrived at this conclusion because of the extreme length of time over which ignorance of it has caused confusion, the variety of fields that have gone astray and the seriousness of the consequences that such ignorance has caused.

In the balance of this essay I will describe five very different situations in which ignorance of de Moivre's equation has led to billions of dollars of loss over centuries yielding untold hardship. These are but a small sampling; there are many more.

The Trial of the Pyx

In 1150, a century after the Battle of Hastings, it was recognized that the King of England could not just mint money and assign it to have any value he chose. Instead the coinage's value needed to be intrinsic, based on the amount of precious materials in its make-up. And so standards were set for the weight of gold in coins—a guinea, for example, should weigh 128 grains (there are 360 grains in an ounce). In the trial of the pyx—the pyx is actually the wooden box that contains the standard coins—samples are measured and compared with the standard.

It was recognized, even then, that coinage methods were too imprecise to insist that all coins be exactly equal in weight, so instead the king and the barons who supplied the London Mint (an independent organization) with gold insisted that coins when tested in the aggregate (say 100 at a time) conform to the regulated size plus or minus some allowance for variability. They chose 1/400th of the weight, which for one guinea would be 0.28 grains and so for the aggregate, 28 grains. Obviously, they assumed that variability increased proportionally to the number of coins and not to its square root, as de Moivre's equation would later indicate. This deeper understanding lay almost 600 years in the future.

The costs of making errors are of two types. If the average of all the coins was too light, the barons were being cheated, for there would be extra gold left over after minting the agreed number of coins. This kind of error is easily detected, and, if found, the director of the mint would suffer grievous punishment. But if the allowable variability was larger than necessary, there would be an excessive number of too heavy coins. The mint could thus stay within the bounds specified and still provide the opportunity for someone at the mint to collect these

Howard Wainer is Distinguished Research Scientist at the National Board of Medical Examiners and an adjunct professor of statistics at the Wharton School of the University of Pennsylvania. He has published 16 books, most recently, Testlet Response Theory and Its Applications *(Cambridge University Press). He is a Fellow of the American Statistical Association and was awarded the 2007 National Council on Measurement in Education Career Achievement Award for Contributions to Educational Measurement. Address: National Board of Medical Examiners, 3750 Market St., Philadelphia, PA 19105. Internet: hwainer@nbme.org*

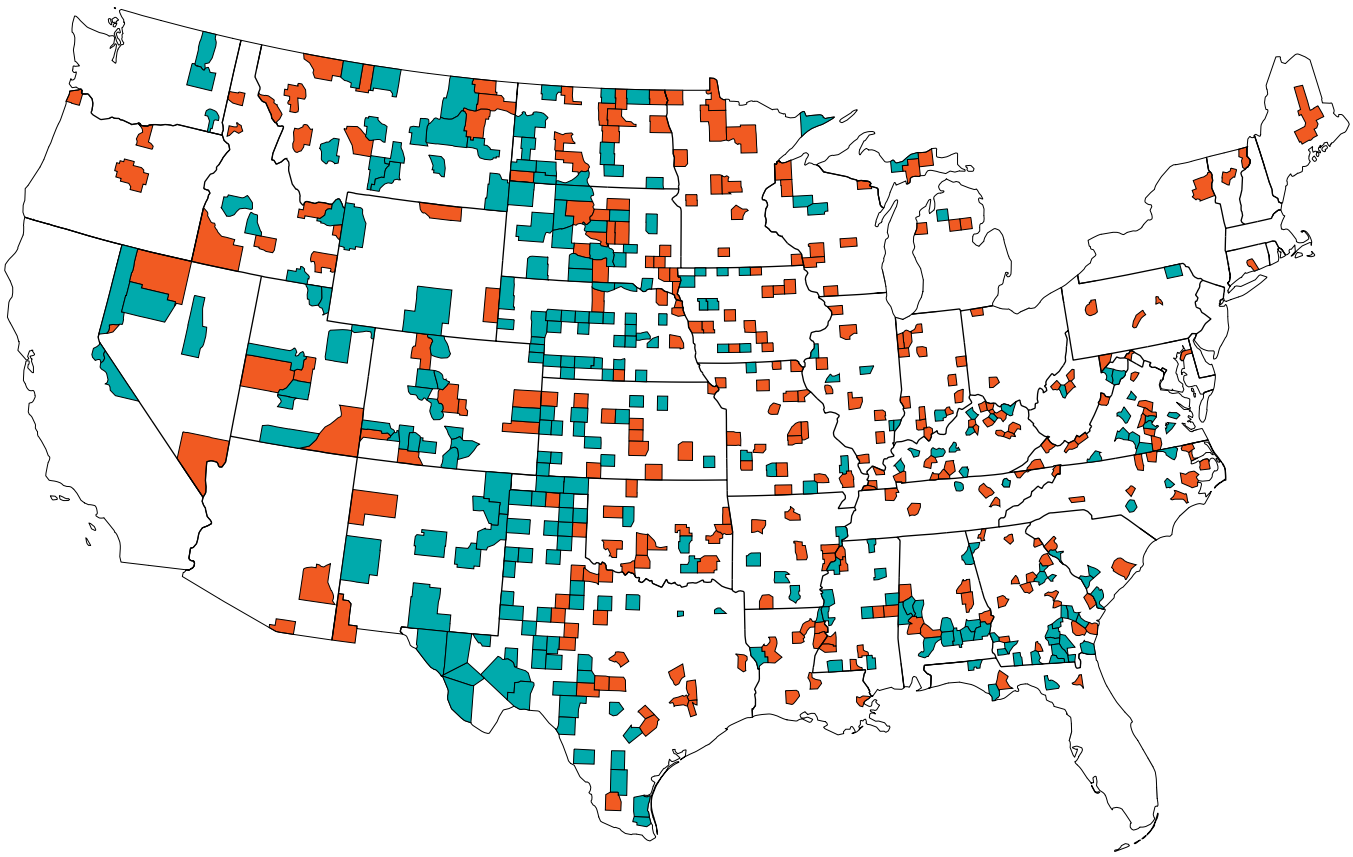

Figure 2. A cursory glance at the distribution of the U.S. counties with the lowest rates of kidney cancer *(teal)* might lead one to conclude that something about the rural lifestyle reduces the risk of that cancer. After all, the counties with the lowest 10 percent of risk are mainly Midwestern, Southern and Western counties. When one examines the distribution of counties with the highest rates of kidney cancer *(red)*, however, it becomes clear that some other factor is at play. Knowledge of de Moivre's equation leads to the conclusion that what the counties with the lowest and highest kidney-cancer rates have in common is low population—and therefore high variation in kidney-cancer rates.

overweight coins, melt them down and recast them at the correct lower weight. This would leave the balance of gold as an excess payment to the mint. The fact that this error continued for almost 600 years provides strong support for de Moivre's equation to be considered a candidate for the title of most dangerous equation.

Life in the Country: Haven or Threat?

Figure 2 is a map of the locations of of counties with unusual kidney-cancer rates. The counties colored teal are those that are in the lowest tenth of the cancer distribution. We note that these healthful counties tend to be very rural, Midwestern, Southern or Western. It is both easy and tempting to infer that this outcome is directly due to the clean living of the rural lifestyle—no air pollution, no water pollution, access to fresh food without additives and so on.

The counties colored in red, however, belie that inference. Although they have much the same distribution as the teal counties—in fact, they're often adjacent—they are those that are in the *highest* decile of the cancer distribution. We note that these unhealthful counties tend to be very rural, Midwestern, Southern or Western. It would be easy to infer that this outcome might be directly due to the poverty of the rural lifestyle—no access to good medical care, a high-fat diet, and too much alcohol and tobacco.

What is going on? We are seeing de Moivre's equation in action. The variation of the mean is inversely proportional to the sample size, so small counties display much greater variation than large counties. A county with, say, 100 inhabitants that has no cancer deaths would be in the lowest category. But if it has 1 cancer death it would be among the highest. Counties like Los Angeles, Cook or Miami-Dade with millions of inhabitants do not bounce around like that.

When we plot the age-adjusted cancer rates against county population, this result becomes clearer still *(see Figure 3)*. We see the typical triangle-shaped bivariate distribution: When the population is small (left side of the graph) there is wide variation in cancer rates, from 20 per 100,000 to 0; when county populations are large (right side of graph) there is very little variation, with all counties at about 5 cases per 100,000 of population.

The Small-Schools Movement

The urbanization that characterized the 20th century led to the abandonment of the rural

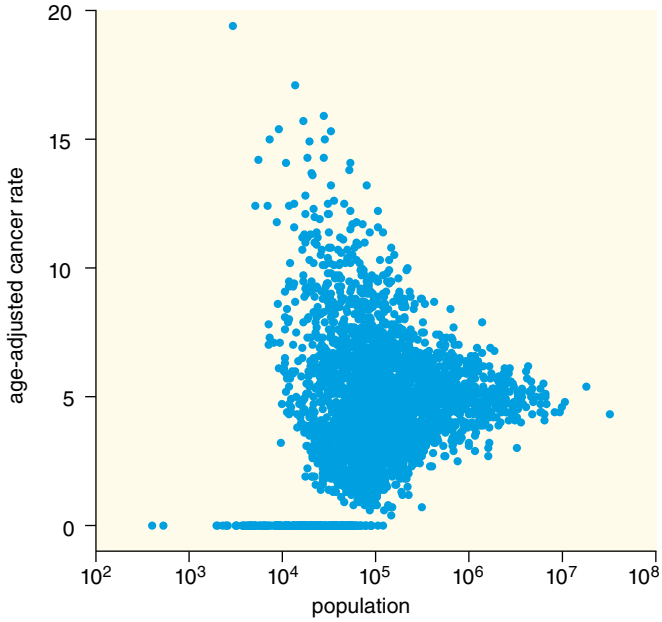

Figure 3. When age-adjusted kidney-cancer rates in U.S. counties are plotted against the log of county population, the reduction of variation with population becomes obvious. This is the typical triangle-shaped bivariate distribution.

lifestyle and, with it, an increase in the size of schools. The era of one-room schoolhouses was replaced by one with large schools—often with more than a thousand students, dozens of teachers of many specialties and facilities that would not have been practical without the enormous increase in scale. Yet during the last quarter of the 20th century, there were the beginnings of dissatisfaction with large schools and the suggestion that smaller schools could provide a better education. In the late 1990s the Bill and Melinda Gates Foundation began supporting small schools on a broad-ranging, intensive, national basis. By 2001, the Foundation had given grants to education projects totaling approximately $1.7 billion. They have since been joined in support for smaller schools by the Annenberg Foundation, the Carnegie Corporation, the Center for Collaborative Education, the Center for School Change, Harvard's Change Leadership Group, the Open Society Institute, Pew Charitable Trusts and the U.S. Department of Education's Smaller Learning Communities Program. The availability of such large amounts of money to implement a smaller-schools policy yielded a concomitant increase in the pressure to do so, with programs to splinter large schools into smaller ones being proposed and implemented broadly (New York City, Los Angeles, Chicago and Seattle are just some examples).

What is the evidence in support of such a change? There are many claims made about the advantages of smaller schools, but I will focus here on just one—that when schools are smaller, student achievement improves. The supporting evidence for this is that among high-performing schools, there is an unrepresentatively large proportion of smaller schools.

In an effort to see the relation between small schools and achievement, Harris Zwerling and I looked at the performance of students at all of Pennsylvania's public schools, as a function of school size. As a measure of school performance we used the Pennsylvania testing program (PSSA), which is very broad and yields scores in a variety of subjects and over the entire range of precollegiate school years. When we examined the mean scores of the 1,662 separate schools that provide 5th-grade-reading scores, we found that of the top-scoring 50 schools (the top 3 percent) six were among the smallest 3 percent of the schools. This is an over-representation by a factor of four. If size of school was unrelated to performance, we would expect 3 percent to be in this select group, yet we found 12 percent. The bivariate distribution of enrollment and test score is shown in Figure 4.

We also identified the 50 lowest-scoring schools. Nine of these (18 percent) were among the 50 smallest schools. This result is completely consonant with what is expected from de Moivre's equation—smaller schools are expected to have higher variance and hence should be over-represented at both extremes. Note that the regression line shown on the left graph in Figure 4 is essentially flat, indicating that overall, there is no apparent relation between school size and performance. But this is not always true.

The right graph in Figure 4 depicts 11th-grade scores in the PSSA. We find a similar over-representation of small schools on both extremes, but this time the regression line shows a significant positive slope; overall, students at *bigger* schools do better. This too is not unexpected, since very small high schools cannot provide as broad a curriculum or as many highly specialized teachers as can large schools. A July 20, 2005, article in the *Seattle Weekly* described the conversion of Mountlake Terrace High School in Seattle from a large suburban school with an enrollment of 1,800 students into five smaller schools, greased with a Gates Foundation grant of almost a million dollars. Although class sizes remained the same, each of the five schools had fewer teachers. Students complained, "There's just one English teacher and one math teacher … teachers ended up teaching things they don't really know." Perhaps this anecdote suggests an explanation for the regression line in Figure 4.

Not long afterward, the small-schools movement took notice. On October 26, 2005, *The Seattle Times* reported:

> [t]he Gates Foundation announced last week it is moving away from its emphasis on converting large high schools into smaller ones and instead giving grants

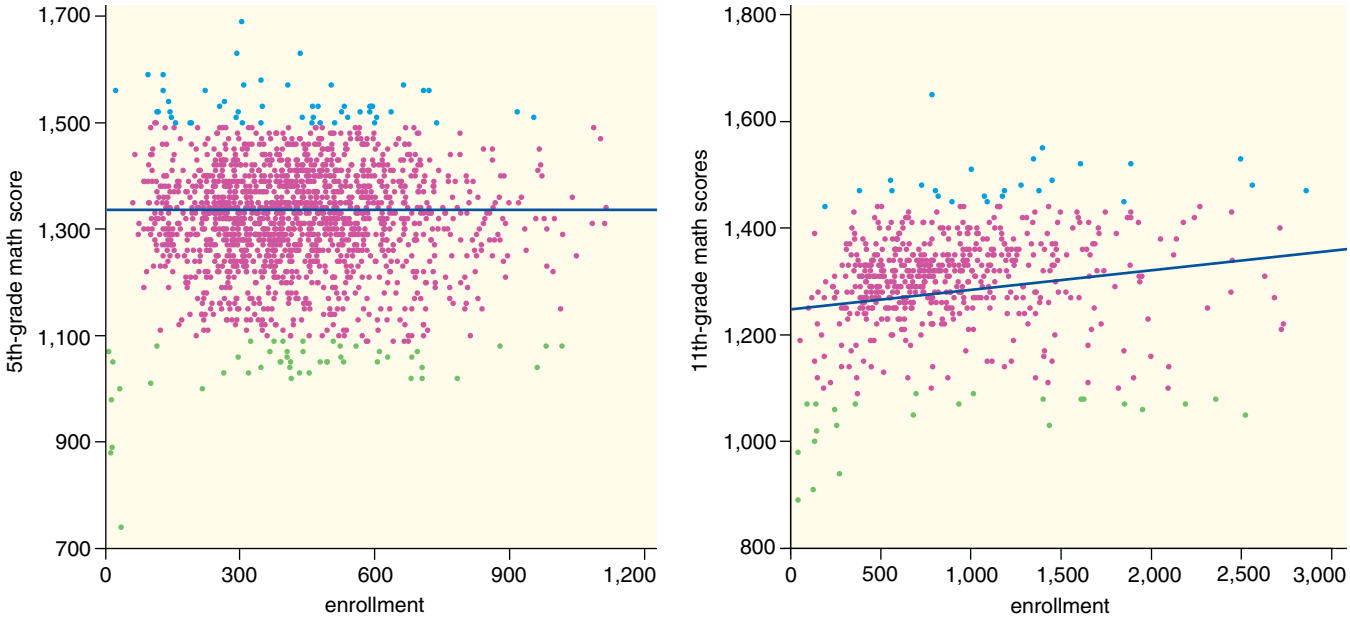

Figure 4. In the 1990s, it became popular to champion reductions in the size of schools. Numerous philanthropic organizations and government agencies funded the division of larger school based on the fact that students at small schools are over-represented in groups with high test scores. Shown here at left are math test scores from 1,662 Pennsylvania 5th-grade schools. The 50 highest-performing schools are shown in blue and the 50 lowest in green. Note how the highest- and lowest-performing schools tend to group at low enrollment—just what de Moivre's equation predicts. The regression line is nearly flat, though, showing that school size makes no overall difference to 5th-grade mean scores. Math scores for 11th-grade schools were also calculated *(right)*. Once again, variation was greater at smaller schools. In this case, however, the regression line has a significant positive slope, indicating that the mean score improved with school size. This stands to reason, since larger schools are able to offer a wider range of classes with teachers who can focus on fewer subjects.

to specially selected school districts with a track record of academic improvement and effective leadership. Education leaders at the Foundation said they concluded that improving classroom instruction and mobilizing the resources of an entire district were more important first steps to improving high schools than breaking down the size.

This point of view was amplified in a study presented at a Brookings Institution Conference by Barbara Schneider, Adam E. Wyse and Venessa Keesler of Michigan State University. An article in *Education Week* that reported on the study quoted Schneider as saying, "I'm afraid we have done a terrible disservice to kids."

Spending more than a billion dollars on a theory based on ignorance of de Moivre's equation—in effect serving only to increase variation—suggests just how dangerous that ignorance can be.

The Safest Cities

In the June 18, 2006, issue of the *New York Times* there was a short article that listed the ten safest United States. cities and the ten least safe based on an Allstate Insurance Company statistic, "average number of years between accidents." The cities were drawn from the 200 largest cities in the U.S. With an understanding of de Moivre's equation, it should come as no surprise that a list of the ten safest cities, the ten most dangerous cities and the ten largest cities have no overlap *(see Figure 5).*

Sex Differences in Performance

For many years it has been well established that there is an over-abundance of males at the high end of academic test-score distributions. About twice as many males as females received National Merit Scholarships and other highly competitive awards. Historically, some observers used such results to make inferences about differences in intelligence between the sexes. Over the past few decades, however, most enlightened investigators have seen that it is not necessarily a difference in level but a difference in variance that separates the sexes. Public observation of this fact has not always been greeted gently, witness the recent outcry when Harvard (now ex-) President Lawrence Summers pointed this out. Among other comments, he said:

> It does appear that on many, many, different human attributes—height, weight, propensity for criminality, overall IQ, mathematical ability, scientific ability—there is relatively clear evidence that whatever the difference in means—which can be debated—there is a difference in standard deviation/variability of a male and female population. And it is true with respect to attributes that are and are not plausibly, culturally determined.

city	state	population rank	population	number of years between accidents
ten safest				
Sioux Falls	South Dakota	170	133,834	14.3
Fort Collins	Colorado	182	125,740	13.2
Cedar Rapids	Iowa	190	122,542	13.2
Huntsville	Alabama	129	164,237	12.8
Chattanooga	Tennessee	138	154,887	12.7
Knoxville	Tennessee	124	173,278	12.6
Des Moines	Iowa	103	196,093	12.6
Milwaukee	Wisconsin	19	586,941	12.5
Colorado Springs	Colorado	48	370,448	12.3
Warren	Michigan	169	136,016	12.3
ten least safe				
Newark	New Jersey	64	277,911	5.0
Washington	DC	25	563,384	5.1
Elizabeth	New Jersey	189	123,215	5.4
Alexandria	Virginia	174	128,923	5.7
Arlington	Virginia	114	187,873	6.0
Glendale	California	92	200,499	6.1
Jersey City	New Jersey	74	239,097	6.2
Paterson	New Jersey	148	150,782	6.5
San Francisco	California	14	751,682	6.5
Baltimore	Maryland	18	628,670	6.5
ten biggest				
New York	New York	1	8,085,742	8.4
Los Angeles	California	2	3,819,951	7.0
Chicago	Illinois	3	2,869,121	7.5
Houston	Texas	4	2,009,690	8.0
Philadelphia	Pennsylvania	5	1,479,339	6.6
Phoenix	Arizona	6	1,388,416	9.7
San Diego	California	7	1,266,753	8.9
San Antonio	Texas	8	1,214,725	8.0
Dallas	Texas	9	1,208,318	7.3
Detroit	Michigan	10	911,402	10.4

Figure 5. Allstate Insurance Company ranked the ten safest and ten least-safe U.S. cities based on the number of years drivers went between accidents. The *New York Times* reported on this in 2006. By now the reader will not be surprised to find that none of the ten largest cities are among either group.

The males' score distributions are almost always characterized by greater variance than the females'. Thus while there are more males at the high end, there are also more at the low end.

An example, chosen from the National Assessment of Educational Progress (NAEP), is shown in Figure 7 NAEP is a true survey, so problems of self-selection (rife in college entrance exams, licensing exams and so on) are substantially reduced. The data summarized in the table are over 15 years and five subjects. In all instances the standard deviation of males is from 3 to 9 percent greater than females. This is true both for subjects in which males score higher on average (math, science, geography) and lower (reading).

Both inferences, the incorrect one about differences in level, and the correct one about differences in variability, cry out for explanation. The old cry would have been "why do boys score higher than girls?" The newer one should be "why do boys show more variability?" If one did not know about de Moivre's result and only tried to answer the first question, it would be a wild goose chase, a search for an explanation for a phenomenon that does not exist. But if we focus on greater variability in males, we may find pay dirt. Obviously the answer to the causal question "why?" will have many parts. Surely socialization and differential expectations must be major components—especially in the past, before the realization grew that a society cannot compete effectively in a global economy with only half of its workforce fully mobilized. But there is another component that is key—and especially related to the topic of this essay.

In discussing Lawrence Summers's remarks about sex differences in scientific ability, Christiane Nüsslein-Volhard, the 1995 Nobel laureate in physiology and medicine, said:

> He missed the point. In mathematics and science, there is no difference in the intelligence of men and women. The difference in genes between men and women is simply the Y chromosome, which has nothing to do with intelligence.

But perhaps it is Professor Nüsslein-Volhard who missed the point here. The Y chromosome is not the only genetic difference between the sexes, although it may be the most obvious. Summers's point was that when we look at either extreme of an ability distribution we will see more of the group that has greater variation. Mental traits conveyed on the X chromosome will have larger variability among males than females, for females have two X chromosomes, whereas males have an X and a Y. Thus, from de Moivre's equation we would expect, all other things being equal, about 40 percent more variability among males than females. The fact that we see less than 10 percent greater variation in NAEP scores demands the existence of a deeper explanation. First, de Moivre's equation requires independence of the two X chromosomes, and with assortative mating this is not going to be true. Additionally, both X chromosomes are not expressed in every cell. Moreover, there must be major causes of high-level performance that are not carried on the X chromosome, and still others that indeed are not genetic. But for some skills, perhaps 10 percent of increased variability is likely to have had its genesis on the X chromosome. This observation would be invisible to those, even those with Nobel prizes for work in genetics, who are ignorant of de Moivre's equation.

It is well established that there is evolutionary pressure toward greater variation within

species—within the constraints of genetic stability. This is evidenced by the dominance of sexual over asexual reproduction among mammals. But this leaves us with a puzzle. Why was our genetic structure built to yield greater variation among males than females? And not just among humans, but virtually all mammals. The pattern of mating suggests an answer. In most mammalian species that reproduce sexually, essentially all adult females reproduce, whereas only a small proportion of males do (modern humans excepted). Think of the alpha-male lion surrounded by a pride of females, with lesser males wandering aimlessly and alone in the forest roaring in frustration. One way to increase the likelihood of offspring being selected to reproduce is to have large variance among them. Thus evolutionary pressure would reward larger variation for males relative to females.

Conclusion

It is no revelation that humans don't fully comprehend the effect that variation, and especially differential variation, has on what we observe. Daniel Kahneman's 2002 Nobel prize in economics was for his studies on intuitive judgment (which occupies a middle ground "between the automatic operations of perception and the deliberate operations of reasoning"). Kahneman showed that humans don't intuitively "know" that smaller hospitals would have greater variability in the proportion of male to female births. But such inability is not limited to humans making judgments in Kahneman's psychology experiments.

Routinely, small hospitals are singled out for special accolades because of their exemplary performance, only to slip toward average in subsequent years. Explanations typically abound that discuss how their notoriety has overloaded their capacity. Similarly, small mutual funds are recommended, after the fact, by Wall Street analysts only to have their subsequent performance disappoint investors. The list goes on and on adding evidence and support to my nomination of de Moivre's equation as the most dangerous of them all. This essay has been aimed at reducing the peril that accompanies ignorance of that equation.

Bibliography

Baumol, W. J., S. A. B. Blackman and E. N. Wolff. 1989. *Productivity and American Leadership: The Long View.* Cambridge and London: MIT Press.

Beckner, W. 1983. *The Case for the Smaller School.* Bloomington, Indiana: Phi Delta Kappa Educational Foundation. ED 228 002.

Carnevale, A. 1999. Strivers. *Wall Street Journal,* August 31.

de Moivre, A. 1730. *Miscellanea Analytica*. London: Tonson and Watts.

Dreifus, C. 2006. A conversation with Christiane Nüsslein-Volhard. *The New York Times*, F2, July 4.

Jodi Hilton/Getty Images

Figure 6. Former Harvard President Lawrence Summers received some sharp criticism for remarks he made concerning differences in science and math performance between the sexes. In particular, Summers noted that variance among the test scores of males was considerably greater than that of females and that not all of it could be considered to be based on differential environments.

Dunn, F. 1977. Choosing smallness. In *Education in Rural America: A Reassessment of Conventional Wisdom,* ed. J. Sher. Boulder, Colo.: Westview Press.

Fowler, W. J., Jr. 1995. School size and student outcomes. In *Advances in Education Productivity: Vol. 5. Organizational Influences on Productivity,* ed. H. J. Walberg, B. Levin and W. J. Fowler, Jr. Greenwich, Conn.: Jai Press.

Friedman, M. 1992. Do old fallacies ever die? *Journal of Economic Literature* (30)2129–2132.

subject	year	mean scale scores		standard deviations		male:female SD ratio
		male	female	male	female	
math	1990	263	262	37	35	1.06
	1992	268	269	37	36	1.03
	1996	271	269	38	37	1.03
	2000	274	272	39	37	1.05
	2003	278	277	37	35	1.06
	2005	280	278	37	35	1.06
science	1996	150	148	36	33	1.09
	2000	153	146	37	35	1.06
	2005	150	147	36	34	1.06
reading	1992	254	267	36	35	1.03
	1994	252	267	37	35	1.06
	1998	256	270	36	33	1.09
	2002	260	269	34	33	1.03
	2003	258	269	36	34	1.06
	2005	257	267	35	34	1.03
geography	1994	262	258	35	34	1.03
	2001	264	260	34	32	1.06
U.S. history	1994	259	259	33	31	1.06
	2001	264	261	33	31	1.06

Figure 7. Data from the National Assessment of Educational Progress show just the effect that Lawrence Summers claimed. The standard deviation for males is from 3 to 9 percent greater on all tests, whether their mean scores were better or worse than those of females.

Geballe, B. 2005. Bill Gates' Guinea pigs. *Seattle Weekly*, 1–9.

Gelman, A., and D. Nolan. 2002. *Teaching Statistics: A Bag of Tricks*. Oxford: Oxford University Press.

Hotelling, H. 1933. Review of *The Triumph of Mediocrity in Business,* by H. Secrist. *Journal of the American Statistical Association* (28)463–465.

Howley, C. B. 1989. Synthesis of the effects of school and district size: What research says about achievement in small schools and school districts. *Journal of Rural and Small Schools* 4(1):2–12.

Kahneman, D. 2002. Maps of bounded rationality: A perspective on intuitive judgment and choice. Nobel Prize Lecture, December 8, 2002, Stockholm, Sweden. http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahneman-lecture.html

Kelley, T. L. 1927. *The Interpretation of Educational Measurements*. New York: World Book.

Kelley, T. L. 1947. *Fundamentals of Statistics*. Cambridge: Harvard University Press.

King, W. I. 1934. Review of *The Triumph of Mediocrity in Business,* by H. Secrist. *Journal of Political Economy* (4)2:398-400.

Laplace, Pierre Simon. 1810. Mémoire sur les approximations des formulas qui sont functions de trés-grand nombres, et sur leur application aux probabilities. *Mémoires do la classe des sciences mathématiques et physiques de l'Institut de France, année 1809*, pp. 353–415, Supplément pp. 559–565.

Larson, R. L. 1991. Small is beautiful: Innovation from the inside out. *Phi Delta Kappan,* March, pp. 550–554.

Maeroff, G. I. 1992. To improve schools, reduce their size. *College Board News* 20(3):3.

Mills, F. C. 1924. *Statistical Methods Applied to Economics and Business*. New York: Henry Holt.

Mosteller, F., and J. W. Tukey. 1977. *Data Analysis and Regression*. Reading, Mass.: Addison-Wesley.

Schneider, B., A. E. Wyse and V. Keesler. 2007. Is small really better? In *Brookings Papers on Education Policy 2006–2007*, ed. T. Loveless and F. Hess. Washington: Brookings Institution.

Secrist, H. 1933. *The Triumph of Mediocrity in Business*. Evanston, Ill.: Bureau of Business Research, Northwestern University.

Sharpe, W. F. 1985. *Investments*. Englewood Cliffs, N.J.: Prentice-Hall.

Stigler, S. 1997. Regression toward the mean, historically considered. *Statistical Methods in Medical Research* 6:103-114.

Stigler, S. M. 1999. *Statistics on the Table*. Cambridge, Mass.: Harvard University Press.

Viadero, D. 2006. Smaller not necessarily better, school-size study concludes. *Education Week* 25(39):12–13.

Wainer, H., and L. Brown. 2007. Three statistical paradoxes in the interpretation of group differences: Illustrated with medical school admission and licensing. In *Handbook of Statistics* (*Volume 27*) *Psychometrics,* ed. C. R. Rao and S. Sinharay. Amsterdam: Elsevier Science.

Wainer, H., and H. Zwerling. 2006. Logical and empirical evidence that smaller schools do not improve student achievement. *The Phi Delta Kappan* 87:300–303.

Williamson, J. G. 1991. Productivity and American leadership: A review article. *Journal of Economic Literature* 29(1):51–68.

For relevant Web links, consult this issue of *American Scientist Online:*

http://www.americanscientist.org/Issue TOC/issue/961