editorial

"When you have eliminated the impossible, whatever remains, however improbable, must be the truth." Sherlock Holmes's famous remark, from *The Sign of Four*, is perhaps the most famous statistical quotation in detective fiction. Alphonse Bertillon links factual detection with the fictional Holmes. The link is though one Harry Ashton-Wolfe, an Englishman who worked in Bertillon's Paris laboratory. Ashton-Wolfe was a friend of Sir Arthur Conan Doyle, and Doyle was the author and begetter of Sherlock Holmes. Hence, no doubt, the double-edged homage to Bertillon in *The Hound of the Baskervilles* on page 36 of this issue.

If you want to interest non-statisticians in statistics, there ought to be worse places to start than detective stories. Both, after all, are about the science of deduction. Start with six suspects, your priors, all equally likely; narrow down the probabilities in the light of further information until five of them become zero and one of them becomes certain (to within an acceptable confidence interval). It is Bayesian, in other words; and it is, more or less, the standard Agatha Christie/Hercule Poirot plot and technique. There is a crop of detective stories that mention statistics more specifically; perhaps not as abundant a crop as it could be, but it is there.

Take this, from *And Be a Villain* (UK title *More Deaths than One*) by Rex Stout, first published in 1948. (If you like detective stories and have not read Stout's Nero Wolfe series, you should.) Wolfe, the immensely fat detective who never leaves his house, is interviewing F. O. Savarese, an excitable Italian maths professor – who wants to help solve the crime. "His favorite branch of mathematics, he said, was the one that dealt with the objective numerical measurement of probability. Very well. What was any detective work, any kind at all, but the objective measurement of probability? All that he proposed to do was to add the word *numerical*, not as a substitute or replacement but as an ally and reinforcement.

"'I'll show you what I mean' he offered. 'May I have paper and pencil ?'" And the professor writes:

$$u = \frac{1}{\sqrt{2\pi D}}\left\{1 - \frac{1}{2}k\left(\frac{X}{D} - \frac{1}{3}\frac{X^3}{D^3}\right)\right\}e^{\frac{1}{2}X^2/D^2}$$

This equation, says Savarese, "is the second approximation of the normal law of error, sometimes called the generalized law of error", and he tries to apply it to solve a crime. Sadly, Wolfe solves the murder without involving statistics.

Coincidence is a staple of detective stories, and understanding coincidence is surely at least a quarter of the way to understanding statistics. It also separates good crime writers from bad ones. Perceptive readers rightly object if a denouement involves too much coincidence; some authors nevertheless sail blithely through.

Shifting genres slightly, Douglas Adams's comedy cult sci-fi novel *The Hitch-Hiker's Guide to the Galaxy* uses a spaceship with an Infinite Improbability Drive to cross interstellar space. (At the core of the drive is a source of Brownian motion – preferably a nice hot cup of tea.) Infinite improbability of course implies zero probability; and statisticians know well that a probability of zero does not mean that a thing cannot happen. (Somebody has to win a lottery, even if an infinite number of tickets has been sold.) One of the results of an infinite improbability generator is that events of zero probability happen: guided nuclear missiles turn in mid-space into a sperm whale and a bowl of petunias; at certain sorts of parties the molecules of the hostess's undergarments simultaneously all move a foot to the left (which, we are told, helps break the ice. Sadly, we are also told that statisticians rarely get invited to that sort of party.)

Raising the tone considerably, *Hamlet* can be dragged into statistical literature – at least through Tom Stoppard's 1966 play, *Rosencrantz and Guildenstern Are Dead*. It is *Hamlet* (without the Prince of Denmark) retold through the eyes of two very minor characters who cannot really make head or tail of what is going on. At intervals throughout the action our two heroes toss a coin, dozens and dozens of times – and it comes up heads 92 times in a row. Which proves that they are actors in a scripted play – or quite possibly proves nothing of the sort. It is, after all, perfectly possible, if deeply unlikely, in the real world – as any statistician can tell you.

Julian Champkin

Champagne and sympathy

The "champagne" chart of wealth distribution (*Significance*, February 2014) has a flaw that was not mentioned in the article. By presenting a curve instead of a five-bar histogram, it pretends to knowledge it does not have, and gets it wrong. Specifically, it conveys an incorrect impression of the distribution of wealth within quintiles, especially within the top quintile. Using US data (`http://en.wikipedia.org/wiki/Wealth_in_the_United_States`), we see that the top 20% of population held about 85% of the wealth, consistent with the champagne chart. The chart, however, suggests that the top 10% or so (top half of the quintile) are a homogeneous group of "haves" sharing approximately equal good fortunes. This is not the case. Of the wealth held by the top quintile, 73% (of that 85%) is held by its top quartile, the top 5% of the population. And within that very wealthy group, 56% of their wealth is held by their top quintile, the top 1% of the population. No doubt similar Paretian relationships hold within the top 1% as well; they are as heterogeneous as the other 99%. At all points on the wealth curve (except the last!) there are folks to the right doing very much better. A chart that got this detail correct would look more like a very wide trumpet than a champagne glass.

John Major
Burlington, Connecticut

Julian Champkin's article, "The Champagne Glass Effect" (*Significance*, February 2014), is incorrect in referring to "quintiles (fifths to non-statisticians)". It is non-statisticians who have spoilt the meaning of "quintiles" (and similar words ending in "ile") by thinking that it means "fifths". It is statisticians who know that the quintiles of a distribution are the four dividing lines between the fifths. To refer to the fifths as quintiles is similar to referring to the two halves of a distribution as medians.

The misuse has become so common nowadays that the battle to keep such words to their original meanings may have been lost, but I do not think that we should give up the fight yet.

David Hill
Chorleywood

Average differences

In a letter by Tom King (*Significance*, February 2014, p. 46), the writer states that he likes to ask his undergraduate students why we calculate averages. He complains that when an answer is offered, by his students or colleagues, "it comes as quite a technical notion of the value of an average as the first term in approximating the distribution of the data".

Here is a general definition of average that I convey to my undergraduate engineering students: "An average is a single value that can replace each of a set of values observed on a fluctuating phenomenon without changing the overall effect of that phenomenon". Examples are :

- The average income of a group of people is the single income value that, if replacing each income in the group, would result in

the same total income. Obviously, this is the algebraic average.

- The average annual inflation over a period of years is the single annual inflation rate that, if replacing each inflation rate in the said period, would result in the same total inflation over the period. Of course, this is the geometric average.
- The average resistance in a group of resistors, interconnected in parallel, is the single resistance value that, if replacing the resistance value of each resistor in the group, would result in the same overall resistance of the electric circuit. Obviously, this is the harmonic average.

Haim Shore
Ben-Gurion University, Israel

The comedian Jasper Carrott once said: "Did you know that the average human being has approximately one testicle and approximately one breast – which just goes to show how useless averages are!"

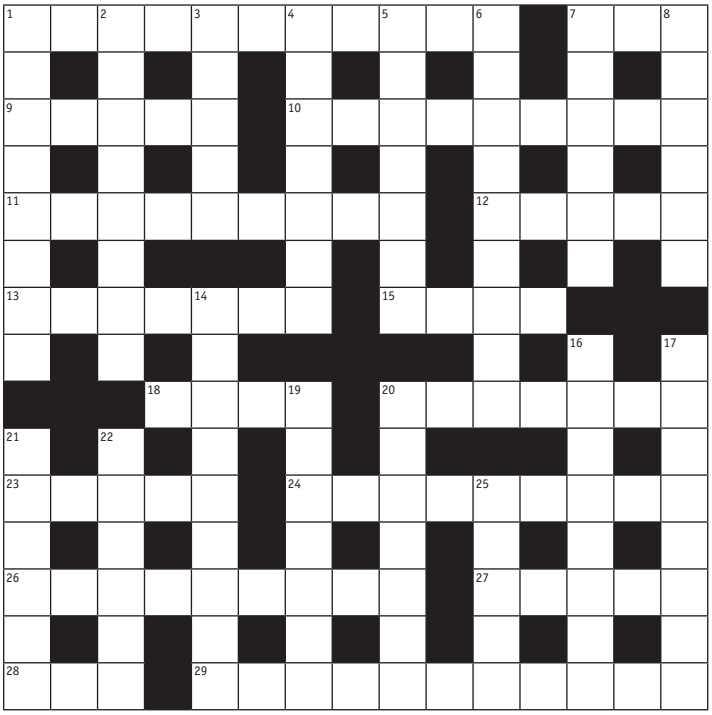
Since then, when teaching any statistical technique to my various teenage classes, I have always used this at some stage: the ensuing discussion would invariably lead towards the conclusion that there are situations when any particular technique is useful, and other situations when not.

Simon Cox
Cheltenham

Letters should be sent by e-mail to significance@rss.org.uk, or by post to: *Significance* Letters Page, Royal Statistical Society, 12 Errol Street, London, EC1Y 8LX. They should be short (preferably under 250 words), may be edited for length and should clearly indicate whether or not they are for publication. They must be received by May 22nd, 2014, in order to be considered for publication in the June issue.

puzzle

Wiley Prize Crossword
Kit by Goujeers



Send your solution to: Significance Crossword Competition, Royal Statistical Society, 12 Errol Street, London, EC1Y 8LX or scan it and email to significance@rss.org.uk. The competition is sponsored by Wiley (http://www.wiley.com/statistics), who will give the winner £100 or $150 to spend on Wiley books. Closing date: May 25th, 2014. The winner will be chosen randomly from the correct entries, and the correct solution published in a future issue. Photocopies are acceptable.

Italicised clues lead to the names of mathematicians or logicians. Grid entries are items of equipment named after them. For example, *Familiar French you call* might lead to Turing, and the entry would be MACHINE.

Across

1 Notice jurors' times to make changes (11)
7 *Perhaps dog state*
9 *No small volume in the morning*
10 Musical takes in South Sea Island's people practising magic (9)
11 Watch reel spinning in turn (9)
12 *Love muse's Scottish lords*
13 Clean up the French housing (7)
15 Jumper's test for a solution (4)
18 *Smell base*
20 A short measure of speech that's naughty (7)
23 *Disturb fish*
24 Again take up beer race, drunk with head of Mackeson (9)
26 Nurse has restrained tradesman (9)
27 River that's set out steering inaccurately (5)
28 Rear animal (3)
29 Perfection in nine points with law suit at first debatable (11)

Down

1 A short day may take a continental matter (8)
2 A short sleep, routine, taken by athlete - playing this? (4-4)
3 *Bow to the man sowing seed*
4 Hacker, a man of principle, has summer in France (7)
5 Mean most of light being hidden gets cloudy (7)
6 Where to buy beers in wild country returning (9)
7 As if may be set out in reflective study (6)
8 Mathematical collections rejected twice? Balls! (6)
14 John's seen last of bowels disturbed with this (9)
16 *Scalp a stranger*
17 Serb and Gael at war over symbol systems (8)
19 Get a tinker to do hair? (7)
20 Fierce melting mid-sea makes it this (3-4)
21 Current, current, current (amps). This is sensitive (6)
22 Veronica Russian, held an innovative high flier (6)
25 *Sleep? That's right*

Solution to December issue's crossword: *Rock Foundation* by Sam Buttrey

1 to 9 in the clues represented the first nine wedding anniversaries

In the adjustments to answers to italicised clues consecutive groups of letters making up the tonic solfa scale were made to share a single square each.

Notes

Across: 1. Anag. abbrev TEL(ephone); 3 2 defs; 5 ADO + (f)RES(h); 8: EGRET after last letter of mother; 9. CHARGE in RED; 10. pun on *tax*; 12. BUS rev + AM I in FLY; 14 last letters; 18. Anag; 20 anag; 22. Pun on *launch air*; 24 anag + initial letters; 27aAnag; 28 INK in (bea)STS; 29 anag SOME L9amb0 D(ressing); 30 hidden; 31 hidden;

Down: O in TART; 2. Anag; 3 BET A S; 4 L OR DS; 5 A +anag after A NAP; 6 RE GAL I A; 7 U,D inside S(ophomore)S; 10 BOO in CASE; 13 2 defs; 14. Hidden rev; 16 A in ANNULS; 17 anag; 19. Pun *on nude eel*; 21 2 defs (IN TENTS)' 23 anag; 25. UP rev +shy; 26 initial letters

Winner, December: Paul Gilbert, Hampshire