

Want to Mislead and Confuse? Use Statistics!

Rhyd Lewis,

Senior Lecturer in Operational Research, School of Mathematics, Cardiff University UK.

www.rhydlewis.eu

This is a reprint of the article “Want to Mislead and Confuse? Use Statistics!” originally published in Scientists are Humans, Jan 2019 (<https://scientistsarehumans.com/2019/01/26/want-to-mislead-and-confuse-use-statistics/>).

In recent years, there has been a growing mistrust of data and statistics in public discourse. For example, figures from a recent University of Cambridge study found that 47% of the British people who voted “leave” in the 2016 Brexit referendum believed that the government had deliberately concealed the truth about how many immigrants were living in the UK. The same study also found that just under half of Donald Trump voters believe man-made global warming to be a hoax, compared with just 2.3% of Hillary Clinton voters [1].

When I teach statistics at Cardiff University, I like to emphasise to my students that they should always try to be “streetwise” when interpreting data that has been summarised by others. It is all very well understanding the mathematics underlying statistical error, significance tests and so on, but is this enough to help us spot some pernicious gobbledygook when reading the news, watching TV, or browsing social media?

As journalist Darrell Huff put it in his 1954 book *How to Lie with Statistics* - “The secret language of statistics, so appealing in a fact-minded culture is employed to sensationalise, inflate, confuse, and oversimplify.” He then went on to consider the harm that false information can bring to society, adding that “knowing nothing about a subject is frequently healthier than knowing what is not so.” To this I would like to add my own mantra: “bad information brings bad decisions”.

In this article I do not intend to explore the various political, sociological, and psychological reasons for the growing mistrust of data. Instead, I want to use several simple real-world examples to try to teach readers how to spot statistical shortcomings and encourage them to think a little bit more about what it is that a particular statistic is saying. What does it mean when a politician tells you that “record amounts” are being spent on something? Is that a useful statement or is valuable information being left unsaid? To these ends, over the past few years, I have made a bit of a hobby of collecting examples of misleading statistics that highlight many of the tricks that are frequently used in public discussion. Some of these examples are funny, some are tragic; some of them were stated to deceive, others were probably uttered through sheer incompetence. All of them are real.

Using numbers instead of proportions

In political debate it is common to hear elected officials saying things like “the number of people in work is higher than ever before”, or “the amount being spent on schools is at a record level.” But while such statements might well be true, are they actually telling us something useful?

In January 2018, the then Health Secretary of England Jeremy Hunt gave Channel 4 News an interview on Accident and Emergency (A&E) waiting times in England [2]. They were not good, but he still managed to sound positive: “Compared to seven years ago, we are treating 3,000 more people within the four-hour standard.” He was right that the number of people being treated within four hours had risen, but what he failed to mention was that the number of patients attending A&E had also risen steeply during that time. In addition, since being in government, the *proportion* of English patients being seen within four hours had actually fallen, and the total number of people waiting *more* than four hours had risen six-fold.

Another example, this time reported on the BBC Fact Checker series [3], is the statement from Prime Minister Theresa May that “the government is spending record amounts on education in England.” Although this might sound impressive, the problem with this statistic is that it again refers to a total, and not the amount spent per pupil. In fact, between 2009 and 2016, the school system in England expanded to take in an extra 470,000 pupils. So it is natural that total spending in schools should be going up, even if spending per pupil is falling. (One must assume that spending per pupil is indeed not increasing, otherwise the PM would have said so).

In these sorts of statements, we also need to consider the implications of inflation. In general, prices rise by around 2% (give or take) every year. So, in order to maintain levels of resources for a particular service, we

would expect them to rise in line with inflation every year. In fact, in 2015 the then Prime Minister David Cameron committed to freezing school spending per pupil. As a result, the Institute for Fiscal Studies warned that this would result in an 8% real terms cut in school spending per pupil in England.

In short, when the total number of people in a population is increasing, so too do other totals – the number of murders, the number of road accidents, the total collected in taxes, the number of babies born. When talking about money, the fact that we have inflation also means that, even if a population is stable, the total amount that we spend on things can go up, even if per-person they are being cut.

Use of the word “significance”

In statistics we often use the word “significance” to indicate that a statistic has been shown to be reliable according to some test; however, this word can be misleading because it is usually interpreted to mean that something is important or meaningful, when many “significant” statistics are anything but.

In his book “How Not to be Wrong”, Jordan Ellenberg considers the statistic that “a child is seven times more likely to die in the care of a nanny than in a day care centre.” On first glance this sounds shocking. But let us consider the numbers. Ellenberg finds that in the USA the number of infant deaths with nannies is 1.6 per 100,000 (0.0016%) compared to 0.23 per 100,000 (0.00023%) at day care centres. This is indeed around seven times the amount. But the thing to bear in mind here is that these are still both very, very small numbers. Nevertheless, we all love our children and want to do what is best for them, so maybe we should just sack the nanny and drive the kids to the nearest crèche instead. But the author then shows that the small benefit of doing this will be more than wiped out by the increased chances of dying in a car accident on the way to the day care centre.

Similar examples are often found in newspapers when talking about health. In October 2015 the UK’s *Sun* newspaper wrote on their front page that “eating just one-and-a-half sausages or two rashers of bacon a day could increase your risk of cancer by up to 18%” [4]. Note the use of the words “up to” in this quotation, meaning that the increases in risk could be less than this amount. More seriously, the offending article does not actually state what the underlying risk of getting bowel cancer is. So, what does an 18% increase actually amount to? According to Cancer Research UK, “1 in 15 UK males and 1 in 18 UK females will be diagnosed with bowel cancer in their lifetime” [5]. For males, this is a 6.6% probability. If I eat all this processed meat for the rest of my life with everything else remaining equal, an 18% increase on 6.6% leads to a 7.9% chance of getting bowel cancer overall. This increase might concern you or it might not. But the final figure is certainly less arresting than the original 18%.

Using negative numbers when calculating percentages

Jordan Ellenberg also notes a case where, in June 2011, the Republican Party of Wisconsin issued a news release boasting about their State Governor’s record of job creation. In that month just 18,000 jobs had been created across the USA, but in Wisconsin the numbers looked good: a net increase of 9,500 jobs. “Today,” the statement read, “we learned that over 50% of US job growth in June came from our state.” Another republican, Jim Sensenbrenner, put it even more simply: “The labour report that came out last week had an anaemic 18,000 created in this country, but half of them came here in Wisconsin. Something we are doing here must be working.”

Again, this might look fine on first glance, but let us consider the raw data. As we just saw, in this particular month Wisconsin added 9,500 jobs. But the neighbouring state of Minnesota, (controlled by the Democrats), added more than 13,000 in the same month. Texas, California, Michigan, and Massachusetts also saw higher numbers of jobs created than Wisconsin. On the other hand, various other US states saw net job losses, which more-or-less cancelled out these gains, leading to the original 18,000 total. So, if we really want to stretch things, we might agree that Wisconsin accounted for half the USA’s job growth, but we now also need to agree that Minnesota was responsible for 70% of job growth, which starts to sound a bit odd. Although “technically” correct, both statistics are very deceptive.

Making inferences on unseen data

Another common misuse of statistics is to talk about the “number of people who did not vote for something” and to then include non-voters in this figure. For example, around the turn of the century devolution was delivered, not without some controversy, to Wales, Scotland, and Northern Ireland. In Wales, this was off the

back of a referendum held in 1997 in which a very small majority (50.3%) voted in favour. Following this, a common complaint from anti-devolutionists was that it was not justified in Wales because “75% of people did not vote for it.” Strictly speaking this is true because turn-out for the referendum was itself only 50.2%. But such statements are unhelpful at best since they somehow suggest that we can second guess how all non-voters would have chosen if they had bothered to show up at the polling booth.

Let us try this technique on another example. The 2017 UK general election resulted in just under half of the 650 seats in Westminster going to Theresa May’s Conservative party. “How disgraceful,” I might shout, “when 72% of the people did not even vote Conservative.” To get this figure, I take the total number of Conservative votes (13.6 million) and then divide by the number of registered voters (46.8 million), which includes all those who did not vote at all. The statistic might be correct in some sense, but it is designed to mislead.

Confusing people with probabilities

Probabilities can often be used to misinform people, whether intentionally or through a lack of understanding. In the mid-1990s, US footballer OJ Simpson was accused of murdering his ex-wife Nicole Brown Simpson and her friend, Ron Goldman. There were several key pieces of evidence linking Simpson to the crime including the presence of a bloody glove behind Simpson’s house, blood in his car, and Simpson’s history of abuse towards his ex-wife. Despite this, Simpson was found not guilty. During the trial, Simpson’s defence lawyer Alan Dershowitz, tried to refute the fact of Simpson’s domestic violence record by citing a statistic that “only one in a thousand abusive husbands eventually murder their wives.”

Here, the suggestion was that the probability of Simpson’s guilt was very small – only 0.1% – therefore, it was unreasonable to convict him based on his history of spousal battery. Have you spotted the problem yet? The above statistic may well be true, but it is being used incorrectly in that it ignores some extra information: namely, that a murder has actually occurred. In the trial, Dershowitz considered the question “What is the probability that Person A will murder his wife given that he is known to beat her?”, for which 0.1% is the correct answer. However, the correct question in this case was actually “What is the probability that Person A murdered his wife, given that he is known to beat her *and given that she has been found murdered?*” In the USA in 1994, approximately 30% of female murder victims were killed by their husbands. Using this extra information in a correct application of Bayes’ probability rule, the answer to the latter question actually comes out as more than 95% [6].

An equally depressing example was used to convict a mother for two murders in 1999. Sally Clark’s first son died suddenly in December 1996 within a few weeks of his birth. Then in January 1998 her second child died in a very similar manner. Soon after she was found guilty in an English court of causing both deaths by smothering and was imprisoned for life. During the trial the prosecution case relied heavily on statistical evidence presented by the paediatrician Professor Sir Roy Meadow, who suggested that the likelihood of a single cot death in these circumstances was 1 in 8,543 and, consequently, the chance of two cot deaths could be found by squaring this figure, giving 1 in 73 million. He even went further to compare this probability to the chances of backing an 80-1 outsider in the Grand National four years running and winning each time [7]. The problem with this reasoning is that it assumes that two cot deaths in the same family are independent events. However, this is not the case. Indeed, if a family loses a baby to cot death, the chances of them losing another baby in the same way is usually more likely.

Even if the 1 in 73 million figure is correct (which it is not), many writers in the press also took this figure to represent the probability of Clark being innocent. But this is also wrong. To use a simple example, imagine that last weekend I won the UK lottery jackpot. The chances of doing this with a single ticket are about 1 in 45 million. This is very unlikely, so does it mean that I cheated? Possibly, but it does not mean that there is a 1 in 45 million chance that I am innocent of cheating. The moral here is that there are lots of people in the world so unlikely things, like winning the lottery fairly, will often happen to someone.

Eventually, Sally Clark was exonerated in 2003. A review into similar cases was subsequently ordered and two other women also had their convictions overturned. Sadly, Clark never recovered from her experiences, and died in her home in 2007 from alcohol poisoning.

Suspicious sampling methods

Many statistics rely on random samples being used to gain information about a population. For example, if we want to know whether the people of Belgium prefer Chocolate A over Chocolate B, the surest way would be to go around getting opinions from every Belgian person, presumably also giving them free chocolate samples along the way. However, such approaches are obviously very expensive and time consuming. Instead, responsible statisticians are more likely to carefully choose a random sample of Belgian participants, conduct taste tests with them, and then use these results to make inferences about the wider population. But are random samples always used?

Dave Gorman's book "Too Much Information" looks at the sorts of statistics used in adverts for beauty products. He cites an advert for Rimmel London's Lasting Finish who claim that it "feels hydrating all day", while at the same time running some small text at the bottom of the screen that whispers "73% of 34 women agree." If you think a sample of 34 people is rather small for a multinational company, you are probably not alone. And while we are being suspicious, we might also wonder whether this sample was selected at random. Which 34 women were asked? Many companies conduct research by sending questionnaires to their existing customers – presumably people who already view their products favourably. In addition, most people do not even respond to the questionnaires, so what you are left with is data collected from the small proportion of existing customers who have bothered to spend time filling out and returning a form.

It gets worse. Gorman then gives the example of Rimmel London's new Lasting Finish Minerals Foundation. While the voice-over states that it "lasts up to 12 hours", the following appears at the bottom of the screen: "20 out of 99 women agreed." I think we can all agree that the voice-over's claim is not a very helpful interpretation of their survey's results.

Leading questions and untruthful participants

Speaking of giving out questionnaires, we must also bear in mind that people are not always consistent in their responses. For example, it is known that people will often give different answers depending on whether the person asking the question is male or female. You may have also heard of the "shy Tory" factor, a phenomenon often seen in UK opinion polls, where people tell polling companies, they will not be voting Conservative, but then go ahead and do so anyway.

Results of surveys can also differ due to the gender of the participants. For example, in 1996 Swedish scientists asked 2,810 heterosexual people how many opposite-sex partners they had had. The mean results were consistent with previous studies: around seven for women and nearly twice as many, thirteen, for men. If we assume that the population of heterosexual people is equally split between men and women, a quick diagram should convince you that the means of each gender should be equal – after all, every time a man sleeps with a woman, a woman also sleeps with a man. So what is going on? Apparently, it comes down to a difference in how men and women remember their sex lives, with men tending to overestimate their number of partners and women tending to underestimate [8].

Even when respondents are telling the truth, we also need to trust that they are not being manipulated by leading questions. Take this example, paraphrased from the classic BBC comedy series "Yes, Prime Minister", regarding the reintroduction of compulsory military service.

- Q: Are you worried about the number of young people without jobs?
- Q: Are you worried about the rise in crime among teenagers?
- Q: Do you think there is lack of discipline in our Comprehensive Schools?
- Q: Do you think young people welcome challenges and leadership in their lives?
- Q: Would you be in favour of reintroducing National Service?

It is quite possible that your answers to all these questions will "yes". But now consider these:

- Q: Are you worried about the danger of war and the growth of arms sales?
- Q: Do you think there is a danger in giving young people guns and teaching them how to kill?
- Q: Do you think it is wrong to force people to take arms against their will?
- Q: Would you oppose the reintroduction of National Service?

If your answer to the last question is also "yes", you have just contradicted yourself.

If all else fails, just say something with conviction

As we know, politicians and other public figures are often able to sound authoritative, even if the words coming out of their mouths are anything but. A good example of this comes from Richard Nixon who, when campaigning for a second term as US President, boasted that under his leadership, “the rate of increase of inflation is decreasing.” To many ears, this might sound encouraging, especially as it is a sitting president saying it. However, this statement allows for the rate of inflation to rise; indeed, this was what was actually happening at the time and the economy was worsening. As was later wryly noted by Notices of the AMS editor Hugo Rossi, “this was the first time a sitting president used the third derivative to advance his case for re-election” [9].

Conclusions

The moral of this story, if there is one, is that people should always think carefully when someone in authority uses a statistic to back their arguments. Are they telling you what you need to know? Or are they telling you the version of events that they want you to know? It never pays to be a cynic, but there is certainly much merit in adopting a healthy scepticism about statistical claims, particularly if it encourages retractions, clarifications, and the production of evidence.

Despite the increasing pervasiveness of junk science and fake news, there are also grounds for optimism. Fact-checking services are now becoming more prevalent in public debate, which has recently culminated in the setting up of the International Fact-Checking Network [10]. The UK charity Full Fact (fullfact.org) is a current signatory of this network’s code of conduct and its main aim is to call out people (often politicians) who – willingly or otherwise – manage to pollute public debate with inaccuracies.

Recently, the UK Statistics Authority also wrote a public letter to the UK Department of Education regarding its repeated misuse of statistics, which they felt were “presented in such a way as to misrepresent changes in school funding”. One important example was the department’s repeated claims that the UK, very generously, is the third highest spender on education per pupil in the OECD. Again, as with many of the examples given in this article, this claim is actually true if you choose to do the numbers in a certain way. But this statistic also includes the money spent on UK universities, which now comes from the tuition fees paid by individuals, not the state [11]. I cannot help but wonder how my statistics students would feel about their future debts being used to back government claims such as this.

References

- [1] The Guardian, <https://www.theguardian.com/society/2018/nov/23/study-shows-60-of-britons-believe-in-conspiracy-theories>, accessed 07/01/2019.
- [2] Channel 4 News, <https://www.channel4.com/news/factcheck/the-number-of-people-waiting-over-four-hours-in-ac>, accessed 07/01/2019.
- [3] BBC News, <https://www.bbc.co.uk/news/education-39302746>, accessed 07/01/2019.
- [4] The Sun, <https://www.thesun.co.uk/archives/news/82002/banger-out-of-order-sausages-and-bacon-top-cancer-list/>, accessed 07/01/2019.
- [5] Cancer Research UK, <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer#heading-Three>, accessed 07/01/2019.
- [6] Crane, H. <http://www.stat.rutgers.edu/home/hcrane/Teaching/582/lectures/chapter5-Bayes.pdf>, accessed 07/01/2019.
- [7] Scheurer, V., <https://understandinguncertainty.org/node/545>, Accessed 22/01/2019
- [8] Fry, H. www.tipsonlifeandlove.com/love-and-relationships/do-men-average-more-sexual-partners-than-women, accessed 07/01/2019.
- [9] Notices of the AMS <http://www.ams.org/notices/199610/page2.pdf>, accessed 07/01/2019.
- [10] International Fact-Checking Network, <https://www.poynter.org/tag/international-fact-checking-network/>, accessed 07/01/2019.
- [11] BBC News, <https://www.bbc.co.uk/news/education-45678670>, accessed 07/01/2019.