



Damn lies

Ben Parker • 6 December, 2006 • 4 comments

This article is a runner-up in the general public category of the Plus new writers award 2006.

Whether it was Mark Twain or Benjamin Disraeli that first coined the idea that there are three types of falsehood — "Lies, damned lies, and statistics", the sentiment still persists today. Statisticians are manipulative, deceitful types, set to pollute our minds with meaningless and mendacious information that will make us vote for their favourite political party, use their demonstrably effective skin cream, or buy the cat food that their cats prefer. As a statistician, it's now time to debunk a few myths.

Exactly 96.4% of our modern world revolves around statistics, and although there are some shockingly bad statistics out there, I hope to be able to convince you that it's generally the presentation of them that is at fault.

Ear we go

I could make a confident bet that you, gentle reader, have more than the average number of ears. Why? Let's assume there are six billion people in this crowded world of ours, more than 99% of them have two ears. There are a few exceptional people, who due to injury or birth, may have one or even no ears. There are, to my knowledge, no three-eared people (Captain Kirk is unfortunately fictional, but he did have three ears: a left ear, a right ear, and a final front ear). When we take an average (add up the total number of ears that humanity possesses, and divide by the number of people), we get the sum



More than the average amount of ears?

Slightly less than 12 billion

6 billion

which is slightly less than two. This means that, as most people in the world have 2 ears, this is very slightly more than the average, so most times I would win my bet.

What does this mean?

Now of course, this is obviously just a statistician being pedantic. However, slightly less silly examples abound. Statistics about how one group of people earn less than a certain percentage of the national average income are used as political footballs. Commentaries are all too common in the newspapers about how shocking it is that people earn only a percentage of the national income, and it's all the fault of the Labour Government / previous Conservative administration / European Union / sunspots.

The distribution of incomes, according to the Department of Work and Pensions (see figure 1), is such that there are relatively few people that earn large sums of money (unfortunately statisticians do not fall into this high income group). Which means the average income, £427 per week from [these figures](#), is much more than the majority of people earn, in a parallel to the stupidity above about average ear count (a few extraordinary people, whether having less than two ears or earning large amounts of money, move the average from the situation for the majority of people).

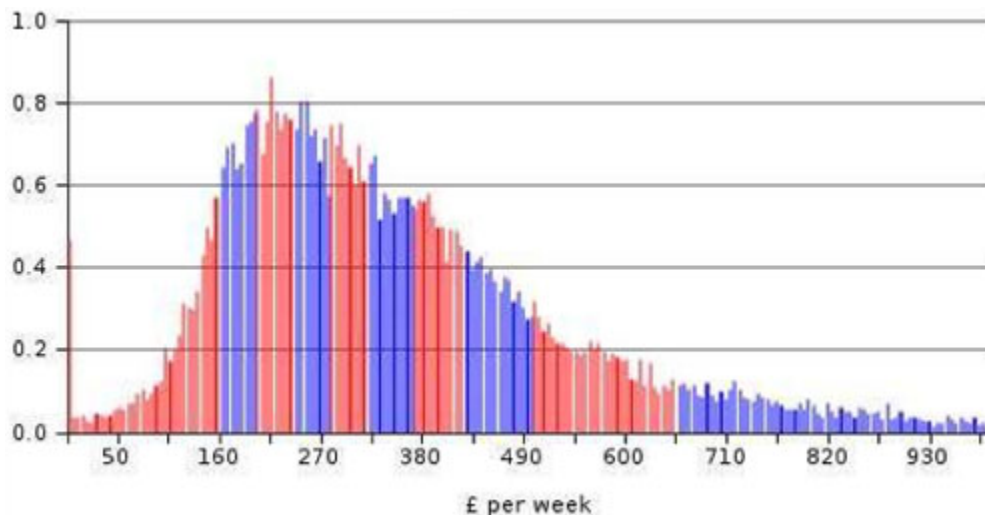


Figure 1: Distribution of weekly income 2002/2003
Number of individuals (millions), Great Britain

Source: [Households Below Average Income, Department for Work and Pensions](#)

Now of course, people soon realised that this commonly used average, where you add everything up and divide by the number of things you add up, more properly referred to as the *mean*, was likely to be misinterpreted. So the concept of the *median* is one that is often used in practice. If we were to put all the people in the UK in a line according to their income, the median salary would be that which the person standing in the middle of the line would have. The median, about £349 per week in this example, in practice often gives a better idea about what is typical.

At least we have reached some common sense — so we'd expect everyone to understand this fairly basic problem in conveying ideas with averages? After all, surely the role of a good journalist is to take ideas and present the truth in a way the public can understand?

Unfortunately, factual accuracy and putting a correct interpretation on data sometimes doesn't sell newspapers, or make the correct political point.

Ironing out the wrinkles



Only 0.2 of this customer was satisfied

Worse than the journalists, but not quite as bad as the politicians, are the advertisers. A recent television advert for a cosmetics company, claims that their latest wrinkle removing cream satisfies 8 out of 10 of their customers, based on a survey of 134. We can perhaps excuse the small sample size, even the rounding ($134 \times 8/10 = 107.2$), which means they must have found 0.2 of a customer to try out the cream, but the crucial question is not how many, but how did they do the survey?

It seems to me that asking 134 customers whether they like the product is dubious — if the people are already customers, and have bought the product voluntarily, perhaps it's not the fairest sample in the world. Why would anyone buy the product that doesn't like it? In most sensible scientific trials, one would hope to compare the performance of the cream objectively against a brand X cream, or a placebo, to see whether people chosen at random have had a positive effect with the cream.

There's no problem with advertising per se; philosophers argue that advertising is the most vital thing for a strong democracy. It's fine for advertisers to let people know about their product, and promote the benefits of it. However, what's not acceptable is putting a thin

veiner of science around the marketing; although cunningly worded, without explaining the method, the statistic "8 out of 10" is meaningless. It's just as bad saying "Our car has a top speed of 500 miles an hour" and not saying that this speed is only obtainable when measuring the downward speed when the car is dropped out of the aeroplane: it's true, but misleading. Using this kind of fake survey in advertising is paramount to lying.

Three come along at once

Maybe it's unfair to blame the conveyors instead of the statistics themselves. There are some real, difficult, non-intuitive facts that statistics throws up, which are very hard to work out why they're true. Let's say you're waiting for a bus, and you look at the little sign, which, assuming it hasn't been vandalised, tells you there are 5 buses an hour. How long would you expect to wait for a bus?

Sensible logic tells us that if there are 5 buses an hour, then the average (sorry, mean) time between buses is 12 minutes. So assuming you arrive at the bus stop at a random time within this 12 minute period, you'd expect to wait 6 minutes for a bus. Good logic: unfortunately in general wrong.

We know that buses don't run to the minute. They may leave the depot on time, but chance factors will alter their running in different ways, so we have to assume that the incoming pattern at our bus stop varies somewhat. What distribution exactly we choose might vary — we may for example, assume that times between arrivals of buses follow an exponential distribution — although the important fact is that the buses do not come at regular times. So let us now assume we arrive at the bus stop at some random point in time — what is our expected wait for a bus now?



Figure 2: You are more likely to arrive at the bus stop during a longer gap, when there is a longer than expected wait for the next bus.

When we turn up at the bus stop, we are more likely to pick a period where there is a big gap between buses than a small gap — the big gap occupies more time than the small gap, so picking at random we're more likely to get a big one. But given that we've picked a big gap, we know that the length of gap is more than 12 minutes (there are still 5 buses an hour) — so the average time to wait, given our exact arrival place is equally likely to be somewhere in the big gap, is more than 6 minutes.

This is known as the inspection paradox, and it's tricky to get your head round it. However, it's a real phenomenon that is used by traffic planners, and operational researchers, who are the people responsible for working out the most efficient method of arranging queues in post offices, and then ignoring it totally.

Are the bus companies wrong to advertise, then, that they have a bus approximately every 12 minutes? I think yes, although it's difficult to convey the whole gory details of the inspection paradox; perhaps in this case we can excuse a little statistical laxity.

Conclusions

In general statistics is fairly intuitive and difficult to conceptualise cases are rare. In general, a questioning reader must:

- Find out who is presenting the data, and what they are trying to achieve by it.
- If possible, find out the sample methodology. Whether the data comes from a suitably representative sample of the population that is being measured, and whether any testing is applied in a fair manner. Are fair comparisons used, and is the right question being asked?
- Question any averages or percentages and think about how extreme the statistics really are, and what you would expect. Particularly, don't assume that means are typical of the data.



Let the reader, and their cat, beware...

Statistics is a powerful and useful tool in the right hands, and we need to give people the ability to understand it. We also need to ensure that some basic education in statistics, particularly in relation to interpreting advertising, is something that every pupil learns at school. At the very least, until journalists, the marketing industry, and the people who regulate them learn some statistics and more importantly how to present them, then the world won't be buying the best skin cream and pet food for their cats, all of whom have an above average number of ears.

About the author



Ben Parker is studying for a PhD at Queen Mary, University of London, and investigating the statistics of data networks. He has an undergraduate mathematics degree from Cambridge, and a masters degree in statistics from Birkbeck, London. In his spare time Ben likes to act in pantomimes and drink tea.