# Statistical literacy guide
## How to spot spin and inappropriate use of statistics

Last updated: July 2010
Author: Paul Bolton

Statistics can be misused, spun or used inappropriately in many different ways. This is not always done consciously or intentionally and the resulting facts or analysis are not necessarily wrong. They may, however, present a partial or overly simplistic picture:

> The fact is that, despite its mathematical base, statistics is as much an art as it is a science. A great many manipulations and even distortions are possible within the bounds of propriety.
>
> (*How to lie with statistics*, Darrell Huff)

Detailed below are some common ways in which statistics are used inappropriately or spun and some tips to help spot this. The tips are given in detail at the end of this note, but the three essential questions to ask yourself when looking at statistics are:

**Compared to what? Since when? Says who?**

This guide deals mainly with how statistics are used, rather than originally put together. The casual reader may not have time to investigate every methodological aspect of data collection, survey methods etc., but with the right approach they will be able to better understand how data have been used or interpreted. The same approach may also help the reader spot statistics that are misleading in themselves.

This guide is a brief introduction only. Some of the other guides in this series look at related areas in more detail. There are a number of books that go into far more detail on the subject such as *How to lie with statistics* by Darrell Huff, *Damned lies and statistics* and *Stat-Spotting. A Field Guide to Identifying Dubious Data*, both by Joel Best and *The Tiger That Isn't: Seeing Through a World of Numbers* by Michael Blastland and Andrew Dilnot. The following websites contain material that readers may also find useful

- Channel 4 FactCheck
- Straight Statistics
- NHS Choices –behind the headlines
- UK Statistics Authority
- STATS – US research organisation with a mission to 'improve the quality of scientific and statistical information in public discourse and to act as a resource for journalists and policy makers on scientific issues and controversies'

## Common ways in which statistics are used inappropriately or spun

**Lack of context.** Context is vital in interpreting any statistic. If you are given a single statistic on its own without any background or context then it will be impossible to say anything about what it means other than the most trivial. Such a contextual vacuum can be used by an author to help put their spin on a statistic in the ways set out in this guide. Some important areas of context are detailed below:

---

All statistical literacy guides are available on the Library Intranet pages:

This is one of a series of statistical literacy guides put together by the Social & General Statistics section of the Library. The series is available on the Library Intranet or at: http://www.parliament.uk/topics/Statistics-policyArchive.htm#SN

- *Historical* –how has the statistic varied in the past? Is the latest figure a departure from the previous trend? Has the statistic tended to vary erratically over time? The quantity in question may be at a record high or low level, but how long has data been collected for?
- *Geographical* –is the statistic the same as or different from that seen in other (comparable) areas?
- *Population if the statistic is an absolute value* –this is important even if the absolute value is very large or very small. How does the value compare to the overall population in question. What is the appropriate population/denominator to use to calculate a rate or percentage? The actual choice depends on what you want to use the rate/percentage to say. For instance, a road casualty rate based on the total distance travelled on the roads (casualties per 1,000 passenger km) is more meaningful than one based on the population of an area (casualties per 1,000 population). The rate is meant to look at the risk of travelling on the roads in different areas or over time and the total distance travelled is a more accurate measure of exposure to this risk than the population of an area.
- *Absolute value if the statistic is a rate/percentage* –what does this percentage (change) mean in things we can actually observe such as people, money, crimes operations etc? For instance, the statement "cases of the disease increased by 200% in a year" sounds dramatic, but this could be an increase in observed cases from one in the first year to three in the second.
- *Related statistics* –does this statistic or statistics give us the complete picture or can the subject be measured in a different way? (See also the section below on selectivity) Are there related areas that also need to be considered?
- *Definitions/assumptions* –what are the assumptions made by the author in drawing their conclusions or making their own calculations? Are there any important definitions or limitations of this statistic?

A related area is spurious comparisons that **do not compare like for like**. These are easier to pass off if there is minimal context to the data. Examples include, making comparisons at different times of the year when there is a seasonal pattern, using different time periods, comparing data for geographical areas of very different sizes, or where the statistic has a different meaning or definition. Comparisons over a very long time period may look at a broadly similar statistics, but if many other factors have changed a direct comparison is also likely to be spurious. For instance comparing the number of deaths from cancer now with those 100 years ago –a period when the population has increased greatly, life expectancy has risen and deaths from some other causes, especially infectious diseases, have fallen. These changes should be acknowledged and a more relevant statistic chosen.

**Selection/omission.** Selecting only the statistics that make your point is one of the most straightforward and effective ways in which statistics are spun. The author could be selective in the indicators or rates they choose, their source of data, the time period used for comparison or the countries, population groups, regions, businesses etc. used as comparators. The general principle applied by authors who want to spin by selection is that the argument/conclusion comes first and data is cherry picked to support and 'explain' this. Such an approach is entirely opposite to the 'scientific method' where observations, data collections and analysis are used to explore the issue and come before the hypothesis which is then tested and either validated or rejected.

Improvements in statistical analysis software and access to raw data (ie. from Government surveys) make the process of 'data mining' much easier. This is where a researcher subjects the data to a very large number of different analyses using different statistical tests, sub-groups of the data, outcome measures etc. Taking in isolation this can produce useful 'hidden' findings from the data. But, put alongside selective reporting of results, it increases the likelihood of one or more 'positive' findings that meet a preconceived aim, while other results can be ignored.

The omission of some evidence can be accidental, particularly 'negative' cases –studies with no clear findings, people who tried a diet which did not work, planes which did not crash, ventures which did not succeed etc.- as the 'positive' cases are much more attention grabbing –studies with clear results, people who lost weight on the latest diet, successful enterprises etc. Ignoring what Nassim Nicholas Taleb calls 'silent evidence'[2] and **concentrating on the anecdotal** can lead people to see causes and patterns where, if the full range of evidence was viewed, there are none.

It is highly unlikely that every piece of evidence on a particular subject can be included in a single piece of work whether it be academic research or journalism. All authors will have to select to some degree. The problem arises when selection results in a different account from one based on a balanced choice of evidence.

**Charts and other graphics**. Inappropriate or inaccurate presentation is looked at in detail in the charts guide. Charts can be used to hide or obscure trends in underlying data while purporting to help the reader visual patterns in complex information. A common method is where chart axes are 'adjusted' in one form or another to magnify the actual change or to change the time profile of a trend. Many charts in the print and visual media are put together primarily from a graphic design perspective. They concentrate on producing an attractive picture and simple message (something that will help their product sell) which can be at the expense of statistical integrity. These aims are compatible if there is input and consideration on both sides and there are examples of good practice in the media.[3]

**Sample surveys** are a productive source of spin and inappropriate use of statistics. Samples that are very small, unrepresentative or biased, leading questions and selective use by the commissioning organisation are some of the ways that this comes about. The samples and sampling guide gives more background.

**Confusion or misuse of statistical terms**. Certain statistical terms or concepts have a specific meaning that is different from that in common usage. A *statistically significant* relationship between variables means that the observation is highly unlikely to have been the result of chance (the likelihood it was due to chance will also be specified). In common usage significant can mean important, major, large etc. If the two are mixed up, by author or reader, then the wrong impression will be given or the meaning will be ambiguous. If an author wants to apply spin, they may use the statistical term to give an air of scientific impartiality to their own value judgement. Equally a researcher may automatically assume that a statistically significant finding has important implications for the relevant field, but this will not always be the case. The guide on statistical significance gives further background.

A (statistically significant) *correlation* between two variables is a test of *association*. An association does not necessarily mean causation, less still a particular direction of cause and effect. The guide on Regression gives some advice on the factors to consider when deciding whether an association is causal.

[2] Nassim Nicholas Taleb, *The Black Swan. The impact of the highly improbable*. (2007)
[3] See for instance some of the interactive and 'static' data graphics used by The New York Times. The finance sections of most papers tend to contain fewer misleading or confusing charts or tables than the rest of the paper.

*Uncertainty* is an important concept that can be lost, forgotten or ignored by authors. Say, for instance, research implies that 60-80% of children who were brought up in one particular social class will remain in the same class throughout their life. It is misleading to quote either end of this range, even phrases such as "up to 80%" or "as few as 60%" do not give the whole picture and could be the author's selective use of statistics. Quoting the whole range not only makes the statement more accurate it also acknowledges the uncertainty of the estimate and gives a measure of its scale. As statistician John W Tukey said:

> "Be approximately right rather than exactly wrong."

Much social science especially deals with relatively small differences, large degrees of uncertainty and nuanced conclusions. These are largely the result of complex human behaviour, motivations and interactions which do not naturally lead to simple definitive conclusions or rules. Despite this there is a large body of evidence which suggest that people have a natural tendency to look for simple answers, see patterns or causes where none exist and underestimate the importance of random pure chance. The uncertainty and risk guide looks at this more fully.

**Ambiguous definitions** are another area where language can impact on the interpretation of statistical facts. Ambiguous or incorrect definitions can be used, to make or change a particular point. For instance migration statistics have terms for different groups of migrants that use a precise definition, but the same terms are more ambiguous in common usage. This can be used to alter the meaning of a statistic. For instance, "200,000 *economic migrants* came to the UK from Eastern Europe last year" has very different meaning to "200,000 *workers* came to the UK from Eastern Europe last year". Similarly mixing up terms such as asylum seeker with refugee, migrant, economic migrant or illegal immigrant change the meaning of the statistic. George Orwell, writing just after the end of the Second World War, said of the misuse of the term 'democracy':[4]

> Words of this kind are often used in a consciously dishonest way. That is, the person who uses them has his own private definition, but allows his hearer to think he means something quite different.

**Averages.** The values of the mean and median will be noticeably different where the distribution is uneven (such as for incomes, wealth or a number of statistics relating to age). The term average generally refers to the mean. Ideally the type of average should be specified when the mean and median are known or thought to be different. This potential ambiguity can be used by authors to select the average that better makes their case. The measures of average and spread guide gives more detail.

An author may also use an ambiguous, subjective or personal definition of average to mean 'typical' –such as the typical family, country or school. Such subjectivity gives them scope to be selective and makes the definition of 'average' much less clear.

**Rounding** can help the reader better understand the quantities involved by not getting lost in unnecessary detail or over precise numbers. However, using rounded numbers to make further calculations can lead to incorrect answers if all or most of the rounding is in the same direction. Calculations should be made on unrounded data and the results rounded. Rounding can also be used intentionally to make a number seem smaller or larger than it is. Both types become more of a problem with higher degrees of rounding or rounding to fewer 'significant places'. For instance 0.5 rounded to the nearest whole number becomes 1; a 100% increase.

---

[4] George Orwell, *Politics and the English Language,* Horizon April 1946.

Similarly through rounding it is possible to make it look like two plus two equals five:

2.3 + 2.3 = 4.6

But when each element is rounded to the nearest whole number it becomes

2 + 2 =5

The guide on rounding and significant places gives more background.

**Reading too much into the data.** Authors may draw conclusions from data that are not fully supported by it. Their implication is that their conclusion naturally follows from the statistical evidence. Again, this could be due to spinning or a misunderstanding of the limits or meaning of the statistics and the logical steps involved in their argument. "Places at UK medical schools have been cut by 10% so in a few years time there will be a shortage of NHS doctors". This statement uses the statistic to present the conclusion as a fact. If one thinks logically about the statement it implies that the *only* factor that affects the supply of NHS doctors is UK medical school places and that demand for doctors will remain constant. In fact the supply of NHS doctors is affected by many more factors including foreign-trained doctors coming to the UK, the age profile and retirement rates of current doctors, the number leaving the profession (pre-retirement) or returning to it, drop-out rates at medical school, medical students training in the UK and working overseas, propensity to work in the private sector, changes in working hours, and so on.

It may be that the conclusion has some merit. But, if other contributory factors are ignored the reader may not be aware of the imbalance between assumption and fact. If assumptions about them are made but not mentioned by the author then the reader will not be able to make their own judgements about them. This situation, as with many other real world situations, is more complex. A (possibly unwitting) desire to view such factors in simple black and white terms with straightforward causes and effects can lead to inappropriate use of statistics.

A related inappropriate use of a statistic is to link data on the extent of a phenomenon to a dramatic or shocking anecdote. While the anecdote is related to the phenomenon it is selected as an example of the most shocking or dramatic. The most extreme cases are normally the rarest. By linking the extent data to the extreme example the author could be attempting to get the reader to imagine that all the cases of this phenomenon are as extreme as the example given. In reality a large majority of the cases will be less extreme and dramatic than the example given.

Many events and their related statistics vary around an average (mean) where the most likely events are close to average and their occurrence becomes less and less likely the further they are from average. Their distribution is said to be 'normal' or bell-shaped or approximate to these. When an observation or result is particularly high compared to the mean we expect the next one to be lower (closer to the mean) and *vice versa*. This is known as **regression to the mean**. It does not always happen, but the most likely follow-up to an extreme outcome is a less extreme one simply because most outcomes are less extreme. Ignoring this effect means reading too much into the data by viewing random variation as real change. Observations about regression to the mean have in part been adopted into common terms such as 'beginners luck', the 'difficult second album' or the 'sophomore slump'[5]. The particular problem with not recognising this

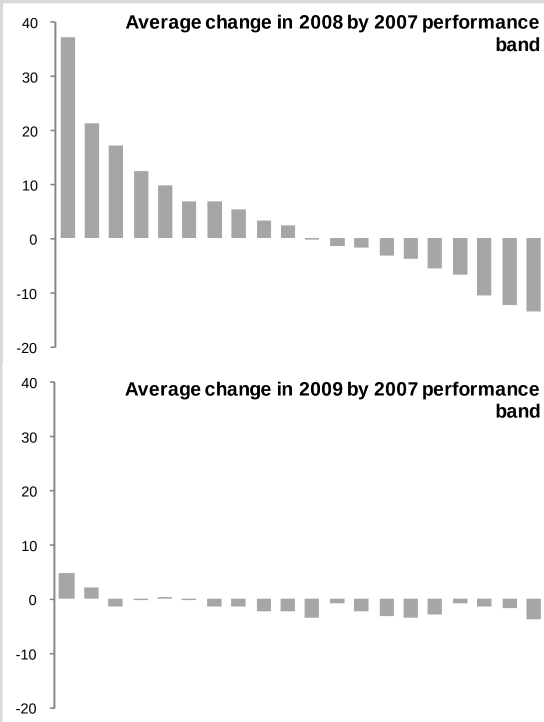
[5] An alternative statistical interpretation for all three is that we remember the extreme results which are much more 'available' –the band with the chart-topping debut album or the rookie footballer who sets a new scoring record- and ignore the large majority of more average results –the moderate selling debut albums,

phenomenon is that observed changes are viewed as being real and linked to a particular intervention where one has occurred.

Some observers have criticised analysis of the impact of speed cameras[6] and anecdotal evidence of the success of alternative therapies[7] because they do not take regression to the mean into account. On its own this does not completely invalidate any findings, it just means that like should be compared with like –accident black spots with and without speed cameras, people with similar illnesses who do and do not visit receive some sort of alternative therapy. Regression to the mean can be difficult to observe and disentangle from real changes to the mean. The box below illustrates an example where there has been no aggregate change in the mean, but variations show a strong indication of regression to the mean.

**Key stage 2 test results and regression to the mean**
Primary school attainment tables measure, among other things, the proportion of pupils reaching the expected level in English, maths and science. These percentages are added for each school and multiplied by 100 to give an aggregate score (out of 300). Nationally there was little change between the aggregate score between 2007 and 2009.

The charts opposite break schools down by 2007 aggregate score into twenty equal sized groups. The first chart looks at how the average performance changed for each group in 2008. Clearly the poorest performing schools in 2007 increased their aggregate score by most. This effect was smaller for each of the next ten groups. In all of the top prior performing ten groups average performance fell and it fell most for the best performing groups of schools. This strongly suggests regression to the mean. Results changed most where earlier results were most extreme. The next chart underlines this. It looks at the same groups of schools (2007 performance bands) and compares average change in results from 2008 to 2009. There is no clear pattern and average changes are small. There is still regression to the mean at an individual school level, but this is cancelled out for groups based on 2007 performance because these groups have already returned towards the mean.




Analysis of change in 2009 by 2008 performance bands shows a very similar pattern to that of the first chart. Regression to the mean occurs after extreme results and so is identifiable when results are organised that way. There are real differences between the performance of schools, but there is also substantial year to year variation in results which displays a clear element of random variation.

**Percentages and index numbers** are a step away from their underlying numbers and this can be used or forgotten by authors when interpreting them. Percentage changes in

bands who cannot get a deal or first year players who only make the reserves. When the next 'observation' is made for the successful debutants –sales of the next album or performances in the following year- it is much more likely that they will perform less well or closer to average. So success is labelled beginners' luck and the second album is viewed as much more difficult.

[6] Where a speed camera is placed in accident black spot its location has effectively been selected as the more extreme. As we cannot directly observe the true long term mean for each potential site one interpretation is that cameras are located in the 'unluckiest' places and we would expect fewer accidents in the next period anyway as through regression to the mean they are likely to be less unlucky next time.

[7] If someone visits a practitioner of alternative medicine when they are feeling unwell (an extreme event) and feel better soon after (a more likely and less extreme event).

percentages or index values and the effect of compounding will not be understood by all readers or authors. This can result in confusing or incorrect commentary by the author or be used by them to spin a point of view. In medicine the *relative risk reduction* looks at changes in the rate of mortality or morbidity from a particular treatment. This figure could be published because the value will be greater than the *absolute risk reduction* which looks at this change in the context of the entire population and is more meaningful. The guide on uncertainty and risk gives more background.

The choice of base year for percentage change figures will affect the size of the numbers involved. For instance, "this year's imports are 100% greater than last year" and "last year's imports were 50% less than those for this year" mean exactly the same thing. The same process can be used in index numbers to affect the change in index values. The guides on index numbers and percentages give more background.

Statistics on **money** are particularly prone to spinning or misuse. Underlying changes over time can be used or ignored selectively to give the author a more persuasive figure. Any comparison of amounts in different time periods should be converted to a common price base or real prices. Without this then any difference will be the result of underlying inflation as well as the 'real' difference. The guide How to adjust for inflation gives some practical advice on compiling a real terms price series and highlights some of the potential misunderstandings around falling levels of inflation and falling prices. Confusion between the two was common in late 2008 and early 2009 when inflation was falling, but still positive and there was a general expectation that inflation would soon turn negative and hence prices would actually fall. It is important to remember that prices are falling if and only if inflation is negative.

Converting a financial time series to real prices may not give a complete picture on its own. If it is funding for a specific service where the underlying demand is expected to change, then a rate based on an indicator of that demand will give a more complete picture. For instance "Funding is set to increase by 40% in real terms by 2010, this means expenditure can rise from £100 per head to £140".

Where the money covers a period of more than one year then a lack of clarity about the actual values for individual years and the start and end point used to calculate headline changes can leave room for ambiguity or spinning. For instance the phrase "there will be a cash increase of £6 billion over the period 2005 to 2008", could mean the 2008 level will be £6 billion more than in 2005. It could also mean that the sum of annual increases compared to 2005 is £6 billion (£1 billion in year 1, £2 billion in year 2 and £3 billion in year 3). In this latter case the 2008 figure is £3 billion less.

### *Tips on how spot spinning or misuse of statistics*

The list below gives a variety of ways to look at statistics to help identify spin or inappropriate use, or cases where a doubt remains that can only be answered by looking at the underlying data.

*General questions to ask yourself*

- What product or point of view is the author trying to 'sell'?
- Are there any statistics or background that is obviously missing?
- Do the author's conclusions logically follow from the statistics?
- Are comparisons made like-for-like?
- If there is any doubt about the original source of the statistic –Who created them and how, why and when were they created?

If a really simplified version is needed then try:

**Compared to what? Since when? Says who?**

*More specific points to look out for*

- Statistics without any context, background or comparisons
- Totals without rates or without any comparators
- Percentages without any absolute values
- A case that is made without any consideration of contrary or inconclusive evidence
- An overly simplistic view about cause and effect
- Very large or very small numbers where the author assumes importance or lack of it solely on this basis
- Records or hyperbole without any further context
- The term *significant* –assume it is the author's interpretation of what constitutes large/important unless it says *statistically significant*
- Ambiguous phrases such as 'could be', 'as high as', 'at least', 'includes', 'much more' etc.
- Unspecified averages (mean or median) where you expect them to be different.
- Use of average for 'typical', the definition of which is known only to the author.
- Lack of details of surveys (sample size, source, actual questions asked etc.)
- Cut-down, uneven or missing chart axes
- Percentage changes in percentages, rates or index numbers
- Statistics on money that compare amounts in different time periods without using real prices
- Statistics on money that do not spell out the time periods in question
- Over precision –intended to lend an air of authority
- Statistics that seem wildly unlikely or results that look too good to be true
- Data on things people are may want kept secret –the number of illegal immigrants, drug use, sexual relationships, extreme views etc.
- Where has the data/analysis been published? For anything even vaguely scientific, was it or the primary research published in a reputable peer-reviewed journal? This does not make the work infallible, just less likely to be spun or contain inappropriate use of data.
- Unsourced statistics

**Statistical benchmarks**

The author Joel Best[8] has suggested using statistical benchmarks to give the reader some mental context when looking at other statistics. This can help to identify statistics that seem wildly unlikely and those that appear to be questionable and where some further investigation may highlight their actual limitations. Some of the latest (rounded) benchmarks for the UK are given below:

| | |
|---|---|
| Population: | 61 million |
| Of whom | |
| *School age (5-15)* | *7.8 million* |
| *Working age* | *38 million* |
| *Pensioners* | *11.8 million* |
| *Minority ethnic group (2001)* | *4.6 million* |
| | |
| Live births: | 775,000 per year |
| Deaths: | 570,000 per year |
| Of which | |
| *Heart disease/stroke* | *190,000* |
| *Cancer* | *160,000* |
| *Respiratory illness* | *80,000* |
| *Accidents/suicide/ homicide* | *20,000* |
| | |
| GDP: | £1,400 billion |

**Other statistical literacy guides in this series:**

- What is a billion? and other units
- How to understand and calculate percentages
- Index numbers
- Rounding and significant places
- Measures of average and spread
- How to read charts
- How to spot spin and inappropriate use of statistics
- A basic outline of samples and sampling
- Confidence intervals and statistical significance
- A basic outline of regression analysis
- Uncertainty and risk
- How to adjust for inflation

[8] Joel Best, *Stat-Spotting. A Field Guide to Identifying Dubious Data* (2008)