How to Lie with Statistics

Noel Schutt

This is a summary of Darrell Huff's excellent 1954 book *How to Lie with Statistics*.

- Semiattachment: Use numbers that appear to be related and good measures, but aren't.

- Built-in sample bias:
  - Who is surveyed?
  - Who responds?

- Which type of average?

- Small sample sizes.

    e.g. Polio vaccine: 1130 children, 2 cases would have been expected, so how many more children should have been tested?

- Average without range.

- Missing numbers: range, graph labels, &c.

- Play up numbers (and differences in numbers) that are statistically insignificant.

- Cropped and poorly labeled graphs.

- Use picture graphs. The change in height causes a huge change (x^3) in volume.

- Post hoc conclusions:
  - the real correlation could be random
  - cause and effect could be reversed
  - the relationship could be due to a third factor

- False precision.

- Shifting base for percentages.

- Add numbers that don't really add to each other.

- Make predictions out of the range of the data. (Extrapolate too far.)

- Percentages are confusing: percentage vs. percentage points vs. percentiles.

- Who says so?
  - Look for unconscious and conscious biases.
  - The data may come from a good source, but what about the conclusion?
  - Watch out for "The survey shows...."

- Look out for biased sampling.
  - How was the sample selected?
  - Is the sample large enough?

- Without some other number for comparison, a number may be meaningless.

- What numbers are missing?

- Does the statistic relate to what it is claimed to, or to something else?

    e.g. A disease rate may go up because of better reporting, not a change in frequency. If you look at the death rate, it is still the same.

- What people say they do vs. what they do.

- Watch out for changing definitions.

- Confusing interest rates: 6% interest vs. $6 on the $100.