New Perspectives in Science Education
International Conference
NEW PERSPECTIVES in SCIENCE EDUCATION


Struggle against incompetent abuse of statistics as persuasive evidence

Tanja Van Hecke
Ghent University, faculty of Engineering and Architecture
Belgium
Tanja.VanHecke@ugent.be

With statistics we can prove everything! Due to misuse of statistics, this statement can become true. But it is a prominent duty of teachers to demonstrate the correct use of statistics and to convince students of the value of this discipline. For example the notion of correlation is often misused. Showing that two factors A and B are correlated, is sometimes inaccurately described as 'A causes B', although it is only spurious correlation. Another possible danger is hidden in graphics: pictures can mislead the reader due to inadequate scaling, inappropriate use of two-dimensional representations, etc. In everyday life, headlines of newspapers are full of astonishing conclusions based on 'scientific research'. If you are lucky, you will find more detailed information inside the article that reveals some information about the way the results were obtained. But in many cases not even the sample size is mentioned, nor is any confidence interval of the measured variable, so no notion of variability or statistical significance is included. Moreover, formulations of comparisons can be misleading, e.g. when a value is reduced by 50 %, we can say as well that we need an increase of 100 % to reach the original value… It is of major importance that we make our students, who are the future users of statistics, aware of these traps. Therefore teaching statistics consists of more than just proving formulas. This paper describes how we can ensure deeper insights into statistics to prevent the cultivation of the abuse of statistics by our students due to ignorance.

1. Introduction

Real life problems often have a level of uncertainty, while mathematical models assume certainty. Whiteley [1] stressed already the importance of probabilistic reasoning in modeling. The goal of statistics and science is facilitating the discovery, understanding, quantification, modeling, communication of facts about the world. Some regard statistical analysis as a gatekeeper, because statistical significance is the first requirement for publication in many social sciences.
As statistics is as a science not algorithmic or deterministic, the danger of misinterpretation and miscommunication arises [2] [3], especially in commercial environments, which can be the future working environment of our students, the impartiality of scientific research is not always guaranteed, because of financial interests.
In statistics three consecutive steps are followed: the gathering of data, the analysis of data and the interpretation of the results. In each of the stages it can go wrong.

2. Biased sampling

The problem starts with collecting data with a sample. As results will be generalized for the whole population, the sample should represent correctly this population. Biased sampling is often used to manipulate results:

- Self-selection bias: a participant's decision to participate may be correlated with traits that affect the study, making the participants a non-representative sample.

- Social desirability bias: if you ask people in a survey about how often they shower, or how often they recycle, your data is going to be biased by the fact that nobody wants to admit doing something socially undesirable.

- Leading question bias: the way of formulating questions in opinion polls, especially when the poll is not anonymous, can influence the answers of the person questioned, e.g. "Most people feel that €5.00 is way too much money to pay for a simple coffee. Would you pay €5.00 for a coffee?".

To our students it is important to introduce techniques as blind experiments or double-blind experiments.

3. Data analysis manipulation

Data analysis is a process of modeling and transforming data with the goal of suggesting conclusions and supporting decision making. But depending on the method you choose, different figures will be produced. Beware of formulations such as "Research showed that …" because you will always find people, who disagree with a commonly accepted opinion. If the tobacco industry wants to show that smoking doesn't affect health, they will gather a lot of metrics and then **only present the metrics that support their objectives.** If you gather enough data, you will always have metrics that confirm your statement.

3.1 Median, mean or mode

The choice between the arithmetic average or the median can be the result of different viewpoints. Consider the example of the trade-union in a company who wants to fight for the rights of the least-paid workers. Suppose the CEO of this company is rewarded with an extremely high salary. To motivate a necessary increase of the salary of the least-paid workers, the trade-union will choose to compare it with the average salary in the company. The CEO will spin these arguments by comparing their salary with the median value of the salaries or the mode which will be lower, as it is less influenced by the outliers.

But even if consensus is reached about the mean, different values can arise. Consider the example of indices to express the evolution of prices in time. Two products (bread and milk) are considered, where the price of bread has halved and the price of milk has doubled in 2009 compared with the previous year. If 2008 is chosen as the base period (see Table 1, left), the average index in 2009 is 125, so one can interpret this result as prices have gone up 25%. But if 2009 is chosen as base period (see Table 1, right), the average index in 2008 is 125, so prices in 2008 (!) are 25% higher than in 2009… These contradictory results can be avoided using the geometric average, where the geometric average of the values $x_1, x_2, \ldots, x_n$ is defined by $\sqrt[n]{\prod_{i=1}^{n} x_i}$. For the year 2008 is turns out to be $\sqrt{200 \times 50} = 100$, but in 2009 it is $\sqrt{100 \times 100} = 100$ as well, so the cost level hasn't changed.

	2008	2009
Bread	100	50
Milk	100	200

	2008	2009
Bread	200	100
Milk	50	100

Table 1: Indices of prices for bread and milk with different bases.

3.2 To reject or not to reject

In the case of hypothesis testing, things can go wrong if this theory is not properly applied.

If a null hypothesis is not rejected, this does not mean that this null hypothesis is true. Perhaps the sample was too small or the power was too small.

To explain the influence of the sample size on forming conclusions, we consider the situation where we have to judge about the null hypothesis $\mu = \mu_0$. Table 2 shows for different sample size dimensions *n*, the minimal difference between $\bar{x}$ and μ that is required to reject the null hypothesis if *x* is normally distributed.

If $|\bar{x} - \mu|$=1 (in case of $s = 1$), the null hypothesis is accepted for $n = 5$, but is rejected for $n = 10$. For small sample sizes we are more reluctant to reject.

n	H_0 is rejected ($\alpha = 0.05$), two-sided, if
5	$\lvert\bar{x} - \mu\rvert > 1.2433\ s$
10	$\lvert\bar{x} - \mu\rvert > 0.7147\ s$
20	$\lvert\bar{x} - \mu\rvert > 0.4673\ s$
30	$\lvert\bar{x} - \mu\rvert > 0.3725\ s$

Table 2: Decision-making with hypothesis testing for different sample sizes

3.3 Different viewpoints, different choices of methods

The choice of the significance level α influences hypothesis testing. Let's suppose that you sell milk in bottles of 1 liter. Consumers claim that there is less than 1 liter in their bottles. To counter the negative publicity, you take a sample of filled bottles and measure the net amount of milk. As $H_0: \mu = 1$, $H_1: \mu < 1$. As α is the chance that $H_0$is rejected although it is true, you will prefer the acceptance of the null hypothesis with the smallest possible α value. The consumers on the contrary want β to be small. As α and β are reversely proportional, it is clear that both sides will perform the test with different choices of α and β.
Juggling with figures and percentages can also be explained by different viewpoints. If a salesman wants to impress his chief, he will explain that he was able to sell at a profit of 200%. As the cost per piece is 1€, and the selling price is 3€ per piece, he based this percentage on the cost. It is likely that in case of a consumer's question, the same salesman will explain that the company is acting ethically as the profit is only 66.6%.

3.4 Spurious correlation

Newspapers like to make conclusions about surprising links between two factors not closely related at first sight. Headlines as "Cappuccino makers have healthier babies" sell well. This is an example of spurious correlation where two variables are not really linked to each other, but are both linked to a third, hidden variable. The hidden variable in this case is the wealth of the household. This explains the results of the sample used as alleged evidence, but does not justify the conclusion that drinking cappuccino will raise the chance of having a healthy baby. Correlation does not imply causality.

4. Interpretation of data

4.1 Misuse of graphical displays

The turnover of two companies is compared in Fig.1 during four periods. Although the right figure shows clearly that the turnover of company A exceeds the one of company B during period 1 and 3, the left figure hides this difference by its 3D representation.

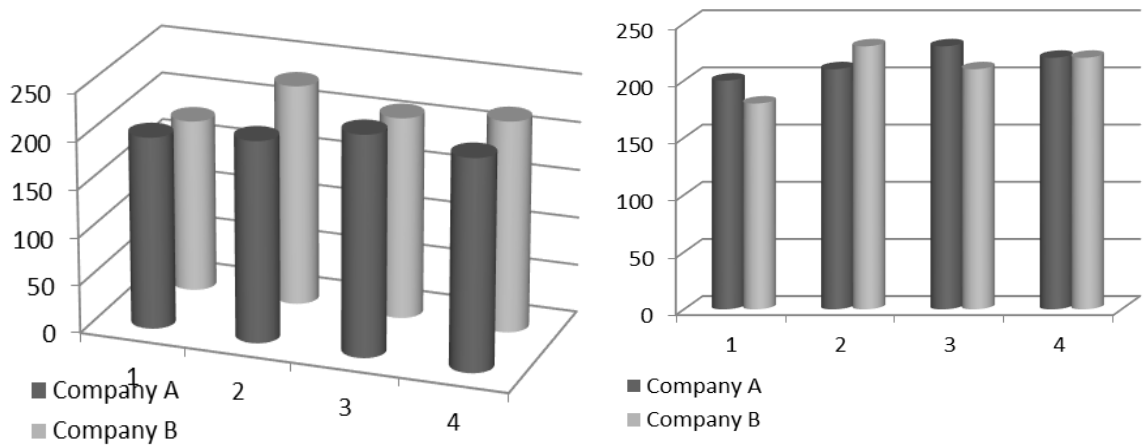

Fig.1: 3D and 2D presentation of data

Fig.2 shows what happens when the zero value is not marked on the vertical axis. The left chart creates the false impression that the democrat party is far ahead compared to the republican and other parties. The right chart is the most objective representation of the data (36 % for the democrats, 33 % for the republicans and 31 % for other parties).

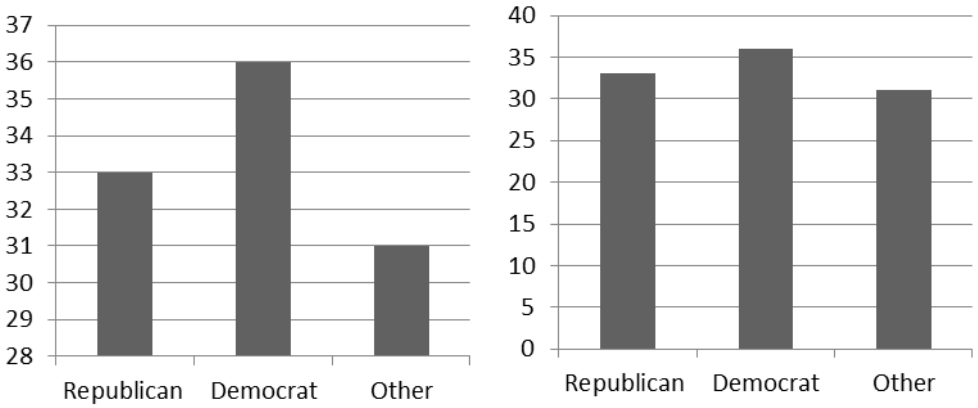

Fig.2: Effect of different scales on the vertical axis

Another type of misconception is created when two-dimensional figures are used to illustrate increase or decrease of a one-dimensional variable. When the evolution in time of e.g. the sheep population is reported by means of pictures of a sheep, false impressions can be created, as can be seen in Fig.3. The length and height of the illustration are doubled, but this makes that the surface is multiplied by four!

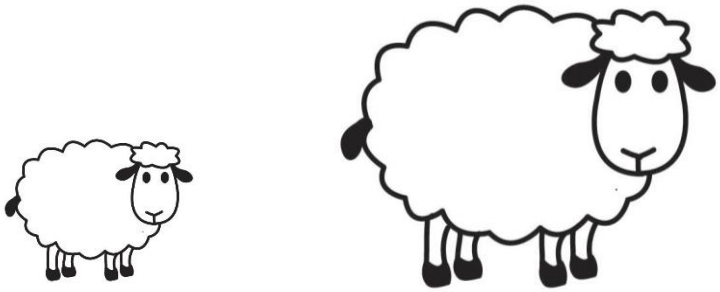

Fig.3: Amplification effect by two-dimensional images in case of doubling

4.2 Need for standard error

Fig.4 was published by the Pew Research Center giving information on the elections for presidency in 2016. The results suggest that Clinton is ahead of Trump. However, the text gives no details about the margin of error. Let us say that the measurement error is $\pm 7\%$. Then one can state that Clinton's support is really between 44% and 58%, while Trump's is between 35% and 49%. Clinton's lead is not statistically significant and this result suggests that it wouldn't be surprising to see another poll that shows that Trump is ahead.

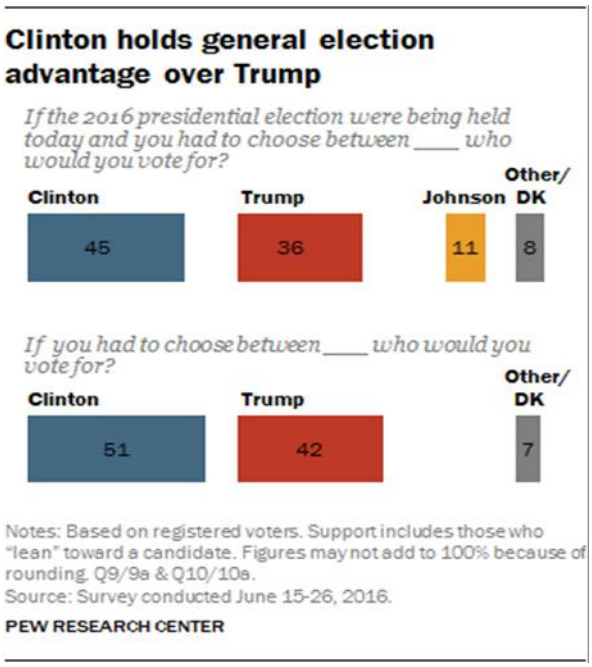

Fig.4: Presentation of poll data for elections

5. Juggling with chances

5.1 Percentiles

Sometimes there exist some misconceptions about percentiles. Consider the school results of children, where percentiles should give parents an idea about the performance of their child at school, weighed against the results of the classmates. Suppose the school results are normally distributed with mean 7 and variance 1. A common misconception is the idea that the difference in results is the same as long as the difference in percentage for the percentiles is the same. We do the calculations for the given example:

$$P(x < a) = 0.7 \Leftrightarrow a = 7.525$$

$$P(x < a) = 0.9 \Leftrightarrow a = 8.285.$$

This shows that an improvement of 0.525 for the result makes the child jump from the median to the 70^{th} percentile, while he/she has to raise his/her result with 0.76 to jump from the 70^{th} to the 90^{th} percentile.

6. Conclusions

Our students should be aware of the traps of incorrect use of statistics. The misused notion of correlation, inadequate scaling of figures, inappropriate use of two-dimensional representations, lack of a notion of variability are some examples. As teachers we must show the dangers of wrong conclusions when misusing statistics. However, this does not change the value of this discipline: statistics is a valuable tool to model, understand and quantify our world.

References

[1] Whiteley,W. “Modeling for life”, For the learning of mathematics **31**(2), 2011, 18-19.
[2] Huff, D. “How to lie with statistics”, New York, Norton, 1954.
[3] Hooke, R. “How to tell the liars from the statisticians”, New York, Marcel Dekker Inc., 1983.