Use and misuse of quantitative and graphical Information in Statistics. An Approach in Teaching

Cécile Hardouin, Lucette Carter

▶ **To cite this version:**

Cécile Hardouin, Lucette Carter. Use and misuse of quantitative and graphical Information in Statistics. An Approach in Teaching. 10th Int. Conf. Mathematics Education into the 21st Century Project. Models in Developing Mathematics Education, Sep 2009, Dresden, France. pp.Carter. hal-00382120

HAL Id: hal-00382120
https://hal.science/hal-00382120

Submitted on 7 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Use and misuse of quantitative and graphical Information in Statistics. An Approach in Teaching.

Lucette Carter and Cécile Hardouin
University Paris X - Nanterre
lucette.carter@gmail.com hardouin@univ-paris1.fr

Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write. – H.G. Wells

Abstract
Miscellaneous examples of misleading statistical data or interpretation are presented in a form suitable for students in mathematics or Social Sciences during a first course of statistics. The aim is to promote critical thinking when confronted (mainly by the media or scientific papers) by information that is biased, incomplete, poorly defined, or deliberately oriented towards a preconceived target. Starting with the simple manipulation of Simpson paradox, the emphasis is put on the need for counfounding in the analysis of relationship between variables.

Introduction
Data and statistical analysis are often presented (particularly in the media or in politics) without the background information necessary for its proper interpretation (omission often due to carelessness, but sometimes with a deliberate intention to mislead). Several well known historical cases of misleading statistics are presented in statistical textbooks to stimulate students' critical thinking. The most famous example occured in the early days of opinion polling in United States (see (5)). A survey based on a sample of over two millions, collected mainly from telephone directories (thus biased in favour of relatively rich voters), predicted that Alf Landon would beat the incumbent democratic President Franklin Roosevelt. (It must be noted that, at the same time, a Gallup poll, based upon a representative sample, predicted rightly that Roosevelt would win).

Another example, presented in many textbooks (see, for example, (1)), is the bias towards expectation in Gregor Mendel's pea experiment, designed to confirm his hypothesis on genetical inheritance laws. More than 50 years later, the Mendel's experiment could be subjected to statistical analyses (applying mainly the chi-square test) which led to the conclusion that his data were suspisciously close to expectation. The controversy was instigated in 1936 by R. Fisher (2) who concluded that the ``data of most, if not all, of the experiments have been falsified so as to agree closely with Mendel's expectation " . Fischer attributed the falsification to an unknown assistant. The controversy continued with other hypotheses such as an unconscious data manipulation by Mendel himself. This example of *confirmation bias*, frequently encountered, encourages the students to apply critical scrutiny to evidence supporting a preconceived idea.

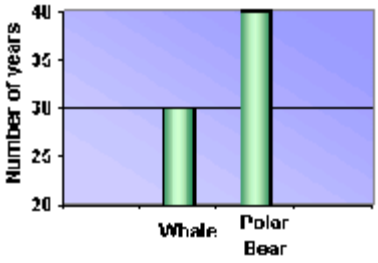
These two classical examples of biased sampling are often presented to the students at the stage of inferential statistics. But at the level of the first introduction to descriptive statistics, other kinds of examples should be presented such as numerical information given without error margin, lack of a precise definition of the variables involved in the study, omission of confounder variables to assess the relationship between two variables, confusion between causality and correlation....

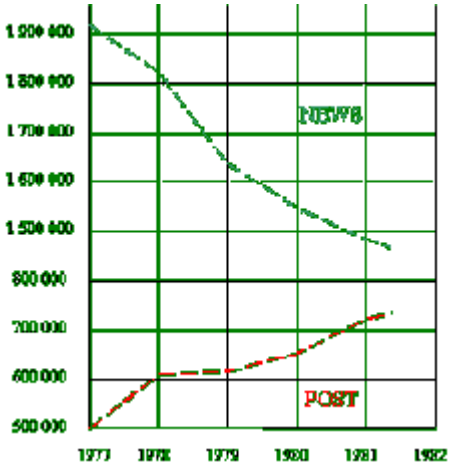
A few miscellaneous examples of such misuses are presented in this paper. They have been analysed by mathematics students, as well as students in Ethnology, in the course of Descriptive Statistics at University of Paris X (see (4) and (9)). It is hoped that these might lead to usefull discussions in teaching applications, and also to a better scrutiny of statistical informations in

various media. Simple examples could already be discussed at the secondary school level. Some of these examples are derived from *Penombre (3)*, a web site founded in 1993 by jurists and statisticians and extended to other academic fields. The authors attention focusses on the quality of quantitative and statistical information. Papers published by *Penombre* try to discern between methods and presentation (particularly in the media).

Misleading graphic Information.

The *start-at-zero-rule* has not been followed in the two graphs below. Moreover, in the second one, the bias in the vertical scale purposedly emphasizes the progression of the *New Yorker Post (from New Yorker Post, 1981*, see (5)) The pupils could be asked to reconstruct and comment the second graph with the right origin and the correct vertical scale.





The right construction of a histogram is unknown and misleading on the graphics hereafter.

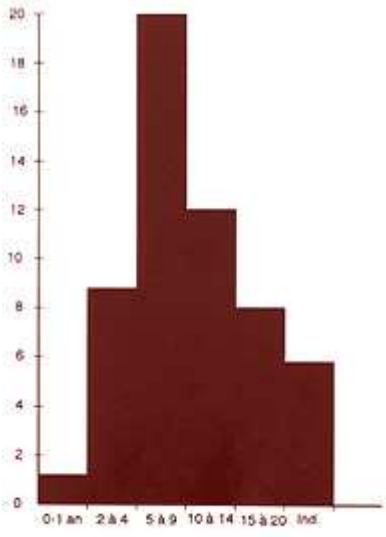


Fig. 1 : courbe de mortalité enfantine.

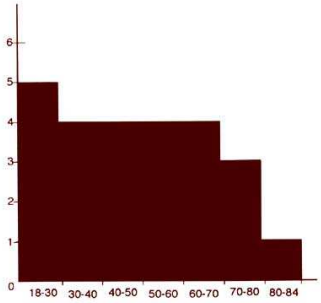


Fig. 2 : courbe démographique des individus masculins de la première période d'occupation de la nécropole (effectif réel).

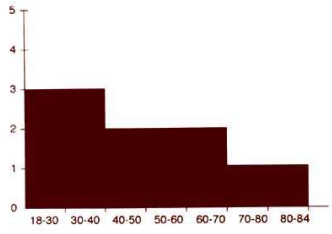


Fig. 3 : courbe démographique des individus féminins de la première période d'occupation de la nécropole (effectif réel).

These graphics are extracted from an archeological study, see (6); the one on the left hand side represents the age of the children found in a necropolis; the graphics on the right hand side

represents the age of adult bodies for men (high) and women (below). The author comments the graphics about modal classes, number of deaths, ignoring the lengths of the classes. Therefore, there are some mistakes which could be easily avoided on the basis of the a first course in statistics.

The role of confounding variables: Simpson Paradox

A common problem, particularly important in analysing clinical data, is that of confounding. This occurs when the association between exposure and outcome is investigated but the exposure and outcome are strongly associated with a third variable, called *confounder*.

The desirability of taking as many covariates as possible into account for assessing the association between two qualitative factors is illustrated by the Simpson's paradox which occurs when the direction of the relationship is reversed when a third factor confounder, sometimes also called *effect modifier factor*, is taken into account.

A classical case illustrating the Simpson's Paradox took place in 1975 (7), when UC Berkeley University was investigated for sex bias in graduate admission. Overall data showed a higher rate of admissions among male applicants, but broken down by departments, data showed a slight bias in favour of female applicants. The explanation is that the female applicants tended to apply to more competitive departments than males, and in these departments the rate of admission was low for both sexes.To analyse this paradox, we present to the students the following simplified version (with only 2 departments) similar to the Berkeley admission data

The table below shows the repartition of 400 students candidates in two schools (of Economics and Engineering) depending on the result of the candidacy and the sex of the candidate:

Result	*Admission*	*Non Admission*
Female	66	134
Male	96	104

The frequency of admissions for female candidates (66/200= 0,33) is smaller than the frequency of admissions for male candidates (96/200= 0,48)

Considering this table only, a question might arise: is the rate of admissions influenced by sex?

But another interpretation emerges when each school is taken separately!

For the school of Economics:

Result	*Admission*	*Non Admission*
Female	40	120
Male	12	48

In this case, the girls (40/160= 0,25 or 25 % admissions) are doing better than the boys (12/60 =0,20 or 20% admissions).)

For the school of Engineering:

Result	*Admission*	*Non Admission*
Female	26	14
Male	84	56

In this case, again, the girls (26/40= 0,65 so 65% admissions) are doing better than the boys (84/160= 0,60 or 60% admissions).

So when the data takes into account the type of School, the direction of the relationship between sex and result of admission is reversed!

This is an example of the Simpson paradox whereby the introduction of a third factor (the confounder or effect modifier factor) reverses the relationship between two factors.

To explain the paradox on this example, one can first compare the total percentages of admissions:
for the school of Economics, the percentage is equal to [(26+12)/220] x100= 24%;
for the school of Engeneering, the percentage is equal to [(26+84)/180] x100= 61%.

The next step is to compare the percentages of the candidacies for both sexes for each school:
for the girls, the percentage of the choice of the school of Economics is equal to [(40+120)/200]x100 = 80 % and for the school of Engeneering [(24+14)/200]x100=20%;
for the boys, the percentage of the choice of the school of Economics is equal to [(12+48)/200]x100 = 30 % and for the school of Engeneering [(84+56)/200]x100= 70 %.

The discrepancies between these various percentages can be summarized by observing that the girls tended to apply to the more competitive department with a low rate of admissions while men tended to apply to the less competitive department with a high rate of admission.

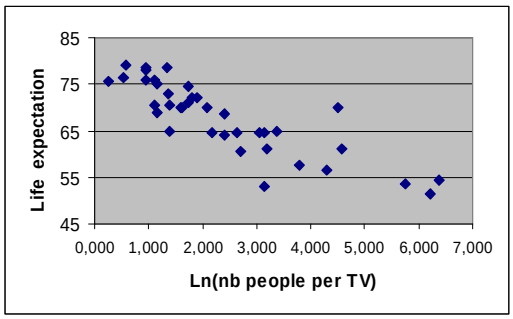
Examples of the Simpson's paradox occur in Epidemiology and clinical trials, particularly in the comparison of two treatments. For example, in a comparison of treatments for kidney stones (see (8)), treatment B is more effective. But when a lurking variable, the size of the stone, is introduced in the study, the conclusion on the two treatments effectiveness is reversed
The students could be asked to reconstruct the related contingency tables and try to explain the source of the paradox.

Confusion between causation and correlation. An example.
One of the most common errors in the press is the confusion between *correlation* and *causation* in scientific and health related studies. When the stakes are high, people are much more likely to jump to causal conclusions. This seems to be doubly true when it comes to public suspicion about chemicals and environmental pollution. There has been a lot of publicity in former years over the purported relationship between autism and vaccinations, for example. As vaccination rates went up across the United States, so did autism. However, this correlation (which led many to conclude that vaccination causes autism) has been widely dismissed by public health experts. The rise in autism rates is most probably to do with increased awareness and diagnosis, or one of many other possible factors that have changed over the past 50 years.

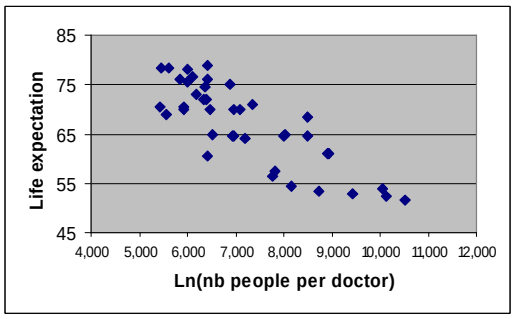
A very usefull dataset (drawn from *TheWorld Almanac and Book of Facts 1993)* helps the students to grasp that a strong association between two quantitative variables does not imply causation. The data for 40 countries (with populations of more than 20 millions as of 1990) gives life expectation at birth, the number of people per television set, the number of people per doctor (see (9)).

The strong negative association between life expectation and each of the two other variables is observed by inspecting the raw data, and confirmed by the correlation coefficients (equal to -0,606 and -0,666 respectively). To avoid the graphic clustering of the richest countries (Japan, USA, France..;), the two graphs below are constructed after a logarithmic transformation of the 2 variables "nb of people per tv" and "nb of people per doctor". Discussing the difference between the two observed associations, the students are led to conclude that in the case of the variable "nb of people per tv", the association with life expectancy is spurious, entirely explained by the effect of economic confounders (such as GNP). On the other hand, the association between the variables " nb of people per doctor " and life expectation is partly direct, partly explained by economic confounders.

This example could be presented at various levels of statistical teaching. A. Rossman (9) presents the analysis of these data for advanced students, including regression, role of outliers, and so on..



Correlation coefficient = -0,855



Correlation coefficient r = -0,832

Another example, which is as simple as easily understandable by students, is to compare the increasing evolution of the foot size with mental development for children. Obviously, the correlation is high, but explained by the fact that both variables are strongly dependent on age.

The following assertions, published in various Media, could be presented to the students to encourage them to judge whether the implicit or explicit interpretations are valid, and suggest some possible confounders:

* Eating a vegetarian diet is positively correlated with intelligence (USA media December 2006)

* Children evacuated to Texas during Hurricane Katrina do worse in school than children from Texas (USA media 2006)

* Watching a lot of television for a teen (A) is correlated negatively with SAT verbal score (B). Could reading ability be one confounder?

* 3 car accidents out of 4 take place at proximity of the drivers's home, implying that that the cause is a relaxation of vigilance (French TV news, 2006)

Unexpected or untractable confounders.

How, then, does one ever establish causality? This is one of the most daunting challenges for public health professionals and pharmaceutical companies. The most effective way of doing this is through a *controlled study*. In a controlled study, two groups of people who are comparable in almost every way are given two different sets of experiences (or treatments) and the outcome is compared. If the two groups have substantially different outcomes, then the different experiences may have "caused" the different outcome.

It is usefull to present some examples where the confounders are unexpected or difficult to detect. For example, in a criticism of a study on the risk of passive smoking, focused on the risk of the health of the non-smoking partner of a smoker, it has been noticed that a confounding missing factor is the inclination of smokers to have a rich diet (which is naturally shared by the non-smoking partner) (see 11).

Exagerated precision of ill defined data.

From an article published by *Populations et Sociétés de l'INED (1996):*At the global level, the FAO estimation for the average of food availability for 1992 is 2718 calories per day and person. A figure given with such precision (and without error margin) denotes a failure to appreciate the notion of confidence in estimation.

A even most flagrant example, occured during the heath wave of 2003 in France, with daily swarm of death numbers highjacked by the media (such as 42 deaths (July 26); 65 deaths (July 27); 10416 additional deaths for the first 3 weeks of August,...). A complete critical scrutiny of these figures is given in (12).
In both cases, the over-impressive figures lend an *aura of precision to the inexact.*

Information in terms of percentages.
In an era of increased quantitative information, such examples as the following could be analyzed. Does the announcement that there is 10% extra free of a certain product implies that the discount of the price is also of 10%?
To increase the impact of certain statistics informations (for example, concerning the improvement of security, or of a medical treatment), relative risk (which could be impressive) is often preferred by politicians or the media, since it may produce a spurious impression of great improvement. For example, if the chance of something happening is increased from 1 per thousand to 2 per thousand, the relative risk is 50%, which seems a lot. If it concerns the chance of survival after a treatment, the absolute chance of survival is increased from 99.98% to 99.99% which is relatively small.

Conclusion
The introduction of examples of misleading "statistics" encourages the students to become more discriminating consumers of political, economic or medical statistics. It seems particularly usefull at the time when one is assaulted through various media by the *hegemony of numbers* , statistical data, and often incomplete or biased interpretation (often in sensitive fields such as medicine, ecology, social sciences...).

Suspicion should not be systematic but one should be ready to ask relevant questions whether confronted by a media declaration or an "expert" article (see the critic on the passive smoking study (11)). It is also a way to enhance rigor in the students statistical applications or memoirs and meticulous presentation of their work: original aim of the study, definition of the variables, of the sample, possible biases induced by questionnaires....

References
(1) Hodges, J.L. e*t al. STATLAB,* Mc Graw-Hill, Inc.1975.
(2) Fisher, R.A., 1936. Has Mendel's work been rediscovered? *Ann. Sci.* ***1:*** 115-137.
(3) Pénombre www.penombre.org
(4) Carter, L. and Mougeot, M. Use of EXCEL in a first Course in Statistics. *Proceedings of the Fifth International Conference on Teaching Statistics.* ICOTS V, Singapore, 1998.
(5) Wonnacott, T.H. and Wonnacott, R.J. *Statistics for Business and Economics.* Wiley, 1994.
(6) Morazzani, C. Etude anthropologique de la nécropole de Goudelancourt-les-Pierrepont (Aisne). *Revue archéologique de Picardie, 1994.*
(7) Bickel, P.J., Hammel, E.A., O'Connell, J.W. Sex Bias in Graduate Admissions Data for Berkeley. *Science,* 187, 1975.
(8) Julious, S.A. and Mullee, M.A.. Confounding and Simpson's Paradox. *BMJ 309 (6967),* 1994.
(9) Rossman, A. Televisions, Physicians, and Life Expectancy. *Journal of Statistics Education vol 2. n2.* 1994.
(10) Carter, L. Multiple factor Interaction. The Role of effect modifier Factor. *The Mathematical Education into the 21st Century Project,* Conference Palermo. 2002.
(11) Molimard, R. Lifting the Smoke Screen. *Revue d'Epidémiologie et de Santé Publique, 56.* 2008.
(12) Cote-Colisson. Les chiffres de la canicule. *Lettre Blanche. Pénombre, n°35 bis.* 2003