January 6th, 2014/February 5th

From Jacques Ninio's web site : http://www.lps.ens.fr/~ninio
(see at the end of the index page)

STATISTICAL ILLUSIONISM : On misleading statistics in scientific publications.

(1) The "false positive" and other problems

(From J. Ninio : *The Science of Illusions* Cornell University Press, 2002, pp. 165-167)

Many techniques are used to give the illusion of objectivity as much in economic or political information as in the reporting of scientific work. Statistics are particularly effective to this end, and charlatans do not fail to make use of them to give a scientific polish to their frauds. The most common technique of statistical charlatanism consists of providing "significance levels." It is commonly said that a particular result is significant at one in a thousand or one in a million. By that, we understand that there is one chance in a million that the observation is due to chance, and since it is very unlikely, it must be due to what the authors think it is due to.

For example, if I toss a coin in the air under ideal conditions, theoretically it has one chance in two of landing tails up, and one chance in two of landing face up. Suppose I toss it a great many times and determine that the chance of finding "tails" occur in at least 51% of the cases. This probability is one in twenty for 10,000 tosses, one in two and a half billion for 100,000 tosses and falls vertiginously when the number of tosses increases further. We see how this applies when transposed to political life. Having had to reach a decision by a referendum on an interbank agreement that they had not read (1), the citizens of a banana republic approved the measure by 51 per cent, which earned them congratulations for their intelligence from the head of state. Statistically speaking, the resuly was highly significant—of the sort that, according to the calculation of probabilities, could not happen once in even fifty billion years throughout the universe, had it been due purely to chance alone. Nevertheless any person with common sense knew that the result would have been different had the weather been rainy instead of sunny on the voting day, or if the bank of a neighbouring country (2) had raised interest rates by a quarter of point instead of lowering them by a quarter of point, and so forth. In fact, all that statistics prove is that there is a disparity relative to a perfect theoretical ideal. But they in no way prove that the disparity is due to the factor that was supposedly being tested.

Often scientists lead themselves be deceived by statistics. In journals of experimental psychology we increasingly read articles in which the authors say that they compared one sitation with another and found a significant effect at a level of so much out of a million, without taking the trouble to specify what direction the effect was in. It is as if on the day after the referendum, the French found headlines such as : “A loud and clear result“ ; “The French did not vote at random“ ; “The turnout in areas is significantly correlated with voters’ preferences,“ and so forth, without saying whether the “yes“ or the “no“ had carried the day (3).

Another technique of statistical illusionism consists of making “significant correlations“ appear in large tables of data. For example, I examine a sample of two hundred inidividuals and I notice a number of their characteristics : the incidence of this or that illness, sexual behaviour, skin color, social success, and so forth. For each individual, I also determine the characteristic of a great many genes. Let's say that for each gene I have two possibilities, labeled + and -. Having done a systematic study of a great many genes, I can construct a double-entry table of the following kind :

	TYPE OF GENE PRESENT				
	1	2	3	4	5
Skill at pinochle					
Individual 1 (brilliant)	+	-	-	+	+
Individual 2 (brilliant)	-	+	+	+	-
Individual 3 (no-hoper)	-	-	+	-	+
Individual 4 (no-hoper)	+	+	-	-	+

From a table like this I can infer that there is a correlation between skill at pinochle and gene 4, since the skilled individuals have the + variant and the weak ones have the – variant. What must be understood is that even if the pluses and the minuses are distributed completely by chance, if you have a large enough number of columns you will indeed find one where the distribution of pluses and minuses is, by chance alone, correlated with one of the traits observed in the individuals. Thus it is that the journal of predigested science regularly announe, with statistics at the ready, the discovery of the gene for this or the gene for that. (….).

(1) Now I can reveal it : the Maastricht treatise.
(2) The Bundesbank.
(3) One will recognize here the misuse of ANOVA tests.

(2) More statistical insanities.

Non-independent correlations
The first statistical aberration that I encountered, as a reviewer, was in a biomedical manuscript. The authors wished to determine the efficiency of a potentially interesting drug X, against a number of macromolecules normally synthesized by the human organism, say a, b, c, d, e, ……. They found that X did lower very significantly the level of a, and also very significantly the level of d. They concluded that the drug X had a dual effect, it acted on both a and d. It did not occur to them that a and d where metabolically related, and in nearly constant proportions. Thus, if a and d are correlated, and X and a are correlated, then X and d are automatically correlated. The authors never admitted the point.

ANOVA tests
In psychophysics, we are plagued by the abuse of ANOVA tests. Users and reviewers as well do not seem to realize that the beloved ANOVA tests require that each contributing variable is normally distributed, which is practically never the case in psychophysical work.

On the use of correlation coefficients
Imagine two variables that are tied with an absolutely deterministic law, for instance $y = \sin(x)$. Assume that the measurements are perfectly accurate. Then the coefficient of the linear regression r is equal to zero. People who do not understand statistics will deduce that x and y are independent.

Sampling
However, psychophysicists do not invest their time in determining a law over a large number of x values on the abscissa axis. In most of their published work, they use a very small set of values, 2, 3 or 4. Then, of course, if they dared to make a plot of the measured y as a function of x, they could get anything, in the example $y = \sin(x)$. What they do, in fact, is repeating several times the experiments with exactly the same set of x values, take the average y values, and put them on a diagram with the standard deviations. Nobody seems to realize that these error bars do not make us progress. If instead of repeating 5 times the measures on a set of 4 x values, one makes 20 measures with 20 different x values, one will have a good chance to get a clearer picture of the underlying law — and of the experimental uncertainties as well.

On absurd mean values
Imagine a photographer who wishes to make a photo report on a city. He takes 120 snapshots that give a fair coverage of the atmosphere of the city, he then selects a particularly representative one for being reproduced on the cover of the magazine. Then, the editor of the magazine, may insist to remove this picture, and replace it with the “average“ picture, one in which each pixel takes the average value of the 120 pixels at the same position in the initial set of photographs (4). The pixels, perhaps in another diagram would take the value

of the standard deviations for the values of the pixels in the 120 original snapshots.

Similarly, assume that you have 120 different experiments that provide 120 different sinusoids with various amplitudes, periodicities, and phases. Taking their averages will lead to nohing. The lesson here is that one may loose a lot of information by taking the averages. In some situations, means and other statistical parameters are a good way to summarize the facts, but in other situations they conceal important structures in the data.
I found an extravagant example of the destructive effect of means in a scientific article, published in a highly respected scientific journal (The Journal of Molecular Biology, 1972). The authors had produced experimental results which they published in one article, then they made a theoretical treatment which they presented in a companion article. They showed in a figure the predicted theoretical curve, and the experimental results, averaged, and accompanied with their error bars. The points representing the results were not exactly on the theoretical curve, so everything looked fine, statistically. However, the “theoretical curve“ was in fact vacuous ! It was an identity of the kind $a^2 – b^2 = (a-b)(a+b)$. Any pair of a and b values should have fitted perfectly the equation. Had the authors taken their individual results, prior to averaging, their points would have been exactly on the theoretical curve. But of course, ou would have learnt nothing from this fit. After averaging, there is no reason that

$$< a^2> - <b^2>= (<a> - <b>)(<a> + <b>).$$

In this way, the authors got the statistically expected legitimate deviations (5).

----

(4) Inspired by a report for the journal “i-Perception“ by Casper Erkelens. He insisted to suppress the typical example on which all the discussions were based.

(5) The article was signed by 5 authors, and read by at least two reviewers and one editor. None of the 8 scientists noticed the enormous blunder!