Reviewer's report

Title: Interpretation of evidence in data by untrained medical students: a scenario-based survey

Version: 1 Date: 29 June 2010

Reviewer: Luis Silva

Reviewer's report:

General Comments

I enjoyed reviewing this interesting manuscript. It certainly has an original approach to address a rather elusive problem. To take into account student`s intuition in order to improve teaching methods is an important issue.

Major Compulsory Revisions

1. Background. First paragraph.

Interpreting the results of laboratory and the use of a null hypothesis significance test can be actually seen as similar processes (as explained in the Browner-Newman 1987 paper, cited by the authors). But to equate the application of a hypothesis significance test with the interpretation of “results reported in scientific articles” is hardly acceptable. Obviously, such an interpretation is a complex task that involves more than the mere value of a conditional probability (the p-value).

Besides, there are some important features of the hypothesis significance tests that are not present when deciding if a patient has a disease or not. For example, it is well known that, strictly speaking, null hypotheses of no difference are usually (almost always) known to be false (the decision about it depending on the sample size). See the old Bakan and Savage papers.

Box 1, introduced by the authors, does not seem very relevant when we take into account the rest of the paper. And there is another problem: the contents of Box 1 is correct when the Neyman Person Model (NPM) is considered, but the same paper under consideration uses the conventionally accepted version of this technique, which is not the NPM one.

Please, explicitly clarify these problems and consider again whether the inclusion of Box 1 is really necessary. Please, note that I am not saying that it has to be removed, but that some changes would be welcome and that to remove it is a possibility.

2. Background. Fourth paragraph.

It can be read: “What approach to inference is better is debated. P-values have come under fire [7-13], and the use of likelihood ratios has gained support [5-7].
Missing from this debate is the examination of how people interpret the world around them in everyday life.

One could deduce that no efforts to assess the way P-values are interpreted by students can be found. If this is the implied idea (if not, please, clarify), it is not entirely correct. For instance, there are several relevant contributions addressing this problem (outstandingly, those by Gigerenzer et al). Please, consider the possibility of citing their papers.

3. Methods. First paragraph

It reads: “The 2 versions of a given year were distributed in a haphazard, quasi-random way in the auditorium.” What do you mean by “quasi-random” in this context? This issue needs further clarification.

4. Results. Second paragraph

It reads: “Among those with a stated opinion (either for or against disease), the proportion who applied the likelihood ratio approach was …” However, as a matter of fact, students did not apply the likelihood ratio approach as such. They, perhaps, made a selection compatible with this approach. It is a rather subtle difference, but I think it is important to avoid this “language abuse” to make more understandable your method, which is not conventional at all.

5. Results. Last paragraph

It reads: “The only notable difference was that fewer respondents had a clear opinion, for or against the disease.” You do not mention which case is the one with fewer respondents. But the fact is that this result is rather puzzling, since one could expect an inverse relation instead of the one you obtained (one could expect fewer respondents with a clear opinion in the second case, because the likelihood ratio is larger in this case).

The importance of this is that one of the more significant limitations of your study could be the degree in which the students actually understood the scenarios they were faced to. These problems could be solved by a lot of them without really understanding the meaning of the data involving these problems. Your conclusions and considerations could be arguable if this were the case, especially because of the (apparently) not validated survey employed. And, precisely, the above commented result raises doubts in this sense. Please, elaborate this further on.

Minor Essential Revisions

1. Methods. Fourth paragraph

It reads: "The probability of the observed or more extreme results under the null hypothesis of no disease was 11%, a non-significant result.” Perhaps you could write instead: “The percentage of the observed or more extreme results under the null hypothesis of no disease was 11% (i.e., p=0.11), a non-significant result.”
which would be more clear to the reader, especially considering that your test is not a conventional one.

2. Methods. Sixth paragraph

“…comparing proportions that were compatible”… I am afraid that what were compatible were not “the proportions” but the opinions. Please, check this.

3. Methods. Third paragraph. First sentence

It reads “Sine the latter…” . Please, correct this.

4. Discussion. Third paragraph.

Of course, the criticism of statistical tests is not universal… but your reference is 10 years old. Along the last years (especially in the present century) it is every day more and more difficult to find statisticians supporting them. I think that this reality should be expressed in some way (the way you managed it gives the idea that the reaction is not so extended).

6. Percentages

Percentages can be rounded to the nearest integer throughout the manuscript.

References:


Gigerenzer G, Krauss S, Vitouch O (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. En David Kaplan (ed). The Handbook of Methodology for the Social Sciences. (Chap.21)

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

'I declare that I have no competing interests'