How insignificant is statistical significance?

Luis Pericchi*

April 08, 2011

Abstract

The US Supreme Court has recently ruled that disclosure of side effects can not be based only on statistical significance. See these articles from the from the Wall Street Journal Online:

- Making a Stat Less Significant (http://on.wsj.com/dWcmGe)
- A Statistical Test Gets Its Closeup (http://on.wsj.com/hwp8Ms)

A set of evidence may not be statistically significant but still ought to be reported. A large number of statisticians have argued this for decades. To name only two groups that hold this opinion, there are the Bayesian statisticians and, I will argue here, the decision theorists, those convinced that statistical inference should depend on the purpose and consequences of the actions to be taken.

The Supreme Court with its timely ruling has opened a Pandora's box: If not in statistical significance only, then on what should the decision to disclose be based?

We argue here that the scientific community should open fresh thinking for reasonable **conventions** on relative losses over possible errors. The consequences of different mistakes are not the same; everybody would agree on that. But how different they are? This is the question that needs to be answered.

1 A simple motivating example

I have a data set of $n = 9$ observations from a normal (or nearly normal) population, standardized to have unit variance. The current wisdom (the null hypothesis H_0) is that the location is $\mu = 0$. However the production process may be out of hand, and it may be that $H_1 : \mu = 1$. We observe $\bar{x} = (x_1 + \ldots + x_9)/9 = 1/4$. Shall we report this data as hinting a potential danger that actually H_1 is producing the data?

Statistical Significance says NO: In fact $\bar{x} = 1/4$ is not that surprising:

$$Probability(\bar{X} > 1/4 \,|\, \mu = 0) = Probability(Z > 3/4) = \alpha = 1 - \Phi(3/4) = 0.227,$$

thus this is not a surprising result under the null H_0, since the p-value, that is the minimal significance level α to reject H_0, is 0.227. That is Type I error, the error of false rejection of the null.

*University of Puerto Rico, Rio Piedras. Facultad de Ciencias Naturales. Departamento de Matematicas. PO Box 70377, San Juan, PR, 00936-8377, USA. lrpericchi@uprrp.edu

On the other hand, we may ask about Type II (observed) error, the probability of the error of failing to reject the null when it is false:

$$Probability(\bar{X} < 1/4 \,|\, \mu = 1) = \beta = Probability(Z < -9/4 \,|\, \theta = 1) = \Phi(-9/4) = 0.0122,$$

thus the type II error β is far smaller. But what is the correct decision? Decision theory teaches us to weight the probabilities by the consequences, in the risk function, the expected value of the loss. In this case, if we call C_1 the cost of false rejection, that is to disclose when the null is correct (that is, to warn the public there is a danger when actually there is none), and C_2 the cost of NOT warning the public when actually the alternative hypothesis is correct.

The decision which is least risky is to disclose if and only if $C_1 \cdot \alpha < C_2 \cdot \beta$. In the example, even though $\alpha > \beta$, if $C_2/C_1 > 20$, say (which is quite reasonable if the consequences of the alternative hypothesis are serious), then the company should disclose the warning.

Incidentally a more incisive discussion may have been developed under a Bayesian framework, but for our purposes this discussion above is enough.

What is next? Now, the decision, which I judge very wise, opens up a plethora of problems. The most important one: If in general the cost of the different errors have to be clearly stated, what should be the equivalent convention of significance testing of $\alpha = 0.1$, 0.05, 0.01?