"What is the significance of the significance level?" "It's the error costs, stupid!"

Haim Shore

Department of Industrial Engineering and Management

Ben-Gurion University of the Negev, Israel (shor@bgu.ac.il)

March, 2014

(from: "Professor Haim Shore Blog")

In a focus article that has appeared in Significance (October, 2013), the author Mark Kelly delivers an excellent review of what "luminaries have to say" regarding the proper significance level to use in statistical hypothesis testing. Among others, the author quotes Sir Ronald Fisher, Karl Pearson, and William Sealy Gosset. The author thence concludes:

"No one therefore has come up with an objective statistically based reasoning behind choosing the now ubiquitous 5% level, although there are objective reasons for levels above and below it. And no one is forcing us to choose 5% either."

I beg to differ. Personally, I was bewildered not to have found in this historical review reference to the only appropriate approach/prescription for determining significance levels:

Optimize the expected combined costs of sampling, Type I error and Type II error.

Or, to paraphrase President Clinton famous election campaign slogan:

"It's the error costs, stupid".

Of course, current lack of reference to error costs of hypothesis testing, as the latter is reported in the literature and as reflected in the review alluded to earlier, is none of the author's fault. Rather, the review faithfully reflects current state-of-the-art in reporting **academic** scientific research results, where costs associates with Type I and Type II errors are rarely considered. Things are different in industrial contexts (manufacturing or otherwise), where errors associated with hypothesis testing have huge consequences. Therein error costs of hypothesis testing are extremely significant and real and may mean a difference between profitability, or realized opportunity, and lack thereof.

In this short comment, I demonstrate with a single example how optimized error costs, associated with optimized Type I and Type II error rates, can be determined so as to reduce the penalty associated with the uncertainty of hypothesis testing. In other words, I demonstrate how significant the significance level may become.

Problem:

In a certain chemical manufacturing facility, the concentration in the final product of a certain ingredient determines the quality of the product. The current average concentration is 5%. An alternative chemical process is proposed that would increase the concentration to 6.5%. This would increase expected annual profit by 60 million dollars (this number already includes cost of switching to the new processing method). Management wishes to experiment with the new process prior to installing it in the production line. Therefore, an experimental production facility is planned at the cost of 700K$. To determine reliably the concentration of the chemical ingredient in the final product, the minimum required amount of the final product, which would produce a single concentration observation, is 1 ton. It costs an additional 1500 dollar (additional to the current cost) to produce one ton by the new method in the experimental line and an additional 500 dollars to produce in line (in the manufacturing process, once the new method is adopted). Current annual production is 50K tons (this will also be the production rate if the new method is adopted).

Determine optimal sample size, n (how many tons of the final product to produce in the experimental line) and optimal values of Type I and Type II error rates (α and β, respectively) that would minimize expected total error cost.

Solution:

To simplify the solution we make the rough assumption that all costs are loaded on a single year.

Null hypothesis is that the proposed new process is not better than the current one. The alternative hypothesis is that the new process increases concentration by 1.5% (from 5% to 6.5%). Denote by n the sample size needed to determine whether to adopt the new processing method.

If the null hypothesis is wrongly accepted (committing Type II error):

The conditional cost (given that the alternative hypothesis is true!) is the missed opportunity for additional annual profit of 60,000,000 $

(we assume that there is justification for the experiment and therefore building an experimental facility and conducting the experiment is not part of the cost of making Type II error).

If the null hypothesis is wrongly rejected (committing Type I error):

The conditional cost (given that the null hypothesis is true!) is the cost of conducting the experiment plus the additional annual cost of switching to production by the new method:

(Build-up of the experimental line) + (additional cost due to experimentally producing n tons of the product) + (annual additional cost of producing by the new method) =

= 700,000 + 1500n + (500)(50,000-n)

The combined *sampling cost* (1500n) and *expected error costs* is:

$$1500n + (700{,}000 + 1500*n + 500*(50{,}000 - n))*\alpha + (60{,}000{,}000)*\beta \qquad (1)$$

(I am aware that alternative cost equations can be built, under different no-less-justified arguments or assumptions).

Assuming that the standard deviation of the experimental process is the same as that of the current process and that it is known (denote it by σ=2%), we have the following criterion for conducting the experiment:

$$\text{If}: \bar{X} > \bar{X}_{cr.}: \text{Reject the null hypothesis;}$$
$$\text{Otherwise - Accept,}$$

where $\bar{X}_{cr.}$ is the critical minimal value of the sample average, $\bar{X}$, needed to accept the alternative hypothesis, namely:

$$\Phi[\frac{\bar{X}_{cr.}-5}{\sigma/\sqrt{n}}]=1-\alpha$$
$$\Phi[\frac{\bar{X}_{cr.}-6.5}{\sigma/\sqrt{n}}]=\beta, \qquad (2)$$

where Φ is the standard normal distribution function.

Minimizing (1), subject to (2), the optimal values of n*, α* and β* are easily obtained.

Doing just that, the optimized model suggests minimizing error rates to the optimal:

α* = 0.000398 ; β* = 0.000161,

with total optimal cost of 148771$.

While error rates are very low (expectedly, I might add, given the large costs involved), it would be instructive to study the sensitivity of total cost to deviations from the optimal values of the sampling plan decision variables:

$\bar{X}_{cr}^{*}$ =5.72% ; n* = 85.9 ≅ 86 tonnes.

The plots in Figure 1 display changes in total cost for each of the sampling plan decision variables, when the other is kept at its optimal value. We realize that deviating from optimal values can increase substantially the total cost. This demonstrates the

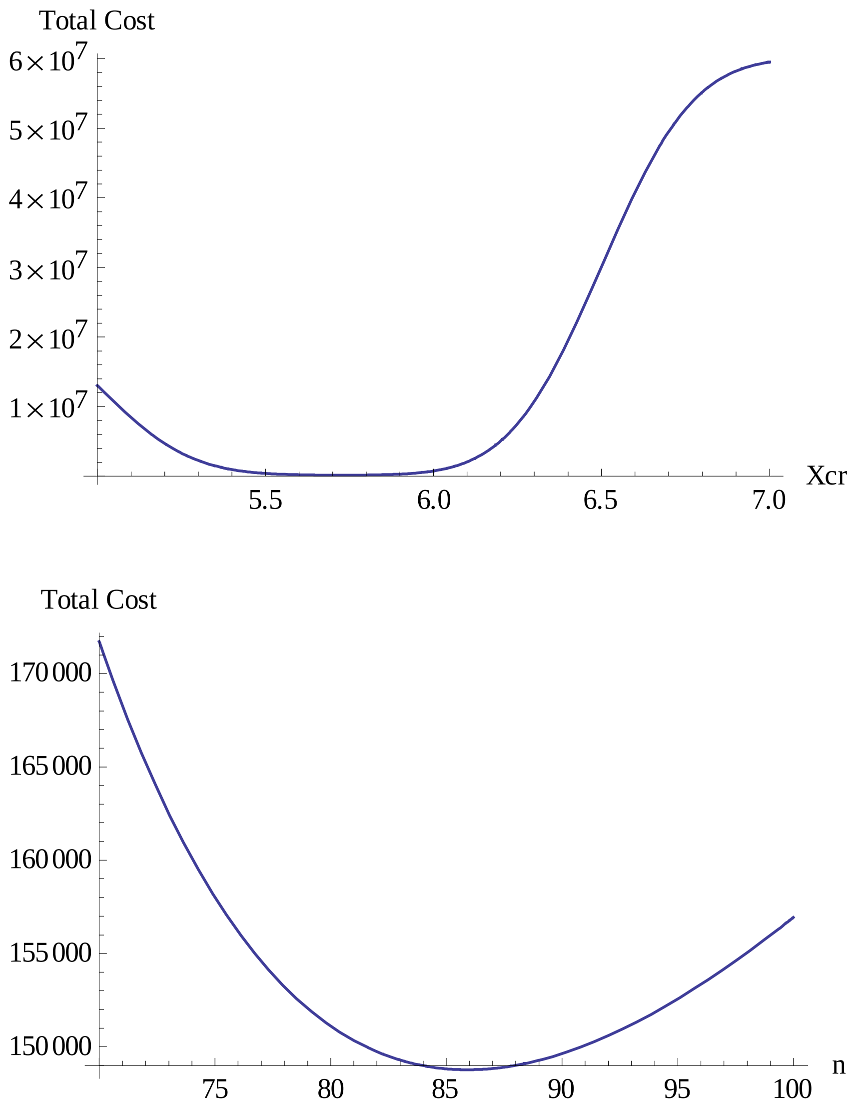

Figure 1. Sensitivity of total cost to deviation from optimal values:

Sensitivity of deviating from the optimal $\overline{X}_{cr}^{*}$ = 5.72% (n=n*; Upper plot) and deviating from n*= 86 tonnes ($\overline{X}_{cr} = \overline{X}_{cr}^{*}$).

importance and significance of choosing correctly the sampling decision variables, and consequently the associated error rates.

Conclusion

In this short comment I aimed to demonstrate that the answer to the logical question, raised and answered in the afore-cited article ("What is the significance of the significance level?"), is misplaced. The obvious and correct answer is:

An immense significance!!

Only make sure that the significance level is treated with the respect and the know-how that it deserves. In other words, stop determining it arbitrarily or based on some general principles.

References

Kelly, M. (2013). Emily Dickinson and monkeys on the stair; Or: What is the significance of the 5% significance level? *Significance*, 10(5), 21-22.