

**CMB POV: Significance Testing—A Double-Edged Sword**

*Uses and Misuses in Market Research*

**By Rich Schreuer**

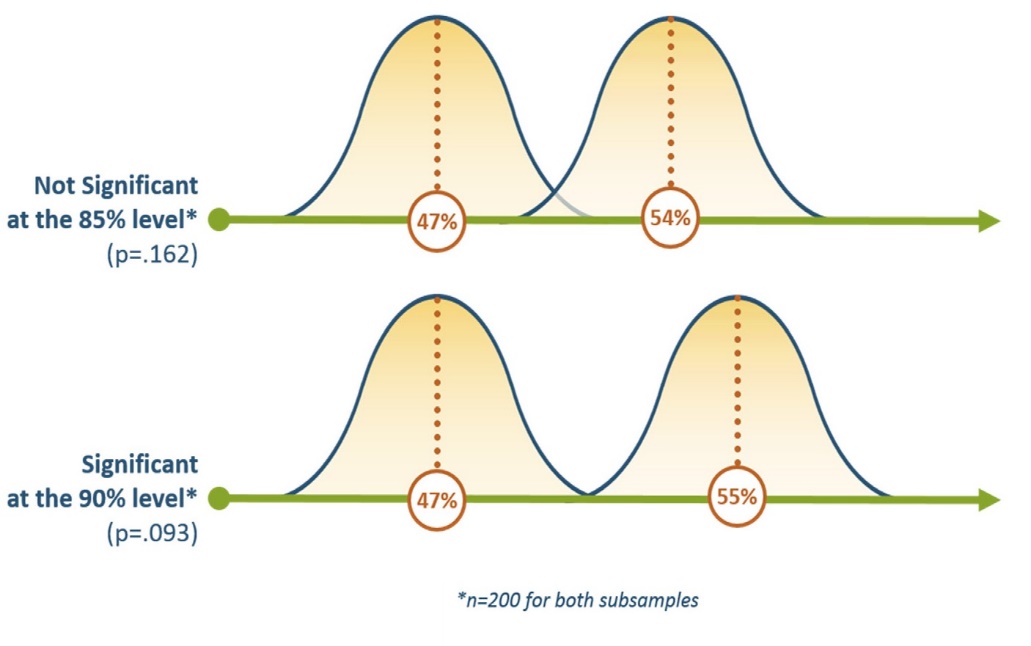


Significance testing plays a very important role in market research—it can help keep managers from overreacting to apparent changes that may not be real. However, significance testing can also lead to bad decisions when managers ignore insights because the supporting data doesn’t cross a statistical threshold.

The problem comes down to sample size. With larger samples, differences large enough to be meaningful are typi- cally significant. Not so with small samples, where large differences between data points (even those that make intuitive sense, or confirm hypotheses) are often dismissed because they are not statistically significant at a high degree of confidence (typically 90 or 95%).

Compounding the problem, there is a very fine line separating a significant finding from one that is not. The illustra- tion below makes this point: under these typical research conditions, an increase in the distance between scores of only 1 percentage point changes the finding from not significant at the 85% confidence to being significant at the 90% level.

In short, it’s important to remember



significance tests are double-edged

swords. They can help managers

avoid bad decisions, but they can

also keep them from making good

ones. Much of the problem lies in

how significance tests are used and

interpreted by decision-makers who

are not versed in statistics.

For example, if a manager gets the results of a concept test, in which the new creative outperforms the old by a score of 54% to 47%, she might not shift to the new creative because the improvement is not sig- nificant at the 90% confidence level.

The problem in the market research industry is that we typically assess significance at the 95% or 90% level, mean- ing that in this case we’d need to be 95% or 90% sure that the new creative outperformed the old for it to be judged “significant.” If the difference between scores doesn’t pass this rather strict threshold, then it is often assumed no difference exists. In fact the difference may well be real, but in this case we can only say so with 84% confidence.

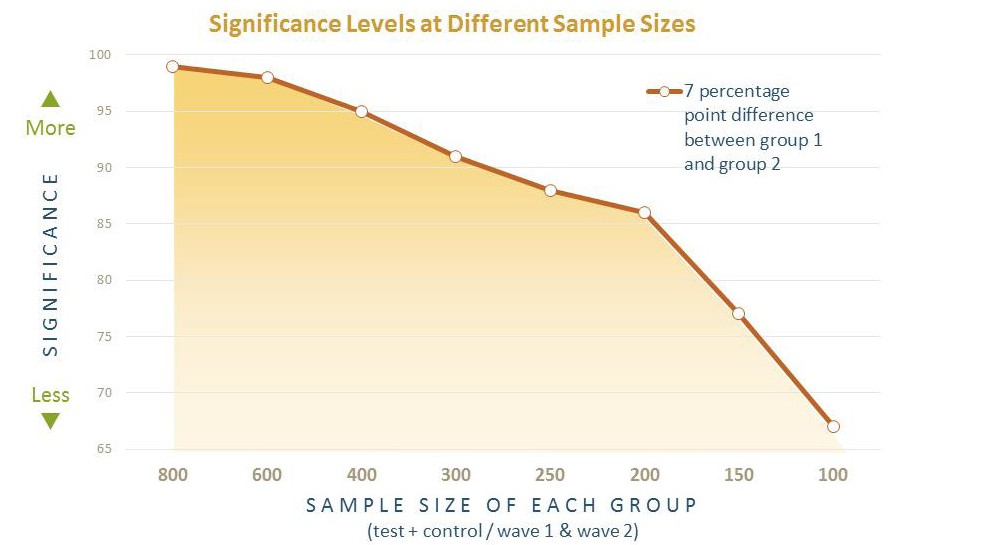
# The problem is not with significance testing per se; it’s with our industry’s inflexibility in weighing the need for statistical precision against the nature of the decisions at hand.

Continuing with the concept test example, researchers can only be 84% confident that the new creative outper- formed the old. However, they can be very sure that the new creative is not worse than the old (there’s only a 1% chance that the new creative’s score is below that of the old). So, the manager has an 84% chance of improving her advertising and a 1% chance of hurting it if she changes to the new creative—pretty good odds. The worst scenario is that the new creative will perform the same as the old. So, in this case, there is strong upside potential in go-

ing with the new creative and little downside (save the production expense). But if the manager were relying on industry-standard significance testing, she would likely have dismissed the creative immediately.

On the other hand, if a manager is facing a risky decision, and the negative consequence of making the wrong deci- sion is high, then it is completely appropriate to insist on high levels of statistical confidence. But it’s worth noting, most managers facing such a decision would probably insist on a large sample.

The table below shows how dramatically sample size affects significance levels. Considering a 7 percentage point difference between two groups (which in many cases would be large enough to be meaningful), the sample sizes would have to be 400 or greater (for each group) to be 95% sure the observed difference is real. With samples of 300, researchers could be 90% confident stating there is a 7 point difference between groups. As the sample size falls, confidence drops dramatically: at n=250, confidence in the difference falls below 90%; at n=200 it falls below 85%, and at n=150 it falls below 80%.



Given all of these considerations, CMB’s POV on significance testing is:

# Significance testing is a crucial tool, but we must use it intelligently and flexibly, never reflexively. An overreliance on significance testing can lead us to give poor recommendations. We owe it our- selves and our clients to have conversations with them when we believe a misapplication of significance testing will impede effective decision-making.