

The use, misuse and abuse of significance

By: Patrick Baldasare, Vikas Mittel

Editor's note: Patrick M. Baldasare is president and CEO of The Response Center, a Philadelphia research and consulting firm. Vikas Mittel is a research analyst at The Response Center.

"The difference is significant!"

"This proportion is significantly greater than that!"

"The Results of the analysis are significant!"

Researchers often misuse and abuse the concept of significance. Many in research comb pile of crosstabulations and reams of analyses to find significant differences and formulate their decisions based on statistical significance. They

tend to associate statistical significance with the magnitude of the result. Their reasoning is something like this: "The more statistically significant a result, the bigger the difference between two numbers." In other words, the fact that one proportion is significantly different than another suggests to many that there is a big difference between the two proportions and statistical significance is often associated with the "bigness" of a result. People often think that if the difference between two numbers is significant it must be large and therefore must be considered in the analysis. We suggest that when comparing numbers, we should consider two types of significance: statistical significance and practical significance. By understanding the difference between statistical and practical significance, we can avoid the pitfall that many in the research industry make.

Statistical significance

What does statistical significance mean? A significance level of, say, 95 percent merely implies that there is a 5 percent chance of accepting something as being true based on the sample when, in fact, in the population it might be false. The statistical significance of an observed difference depends on two factors: the sample size and the magnitude of the difference observed in the samples.

For example, let's say we do a significance test between two groups of people who are exposed to a product concept and find a 20-point difference between Group A (65 percent acceptance) and Group B (45 percent acceptance). Is the difference statistically significant? Despite the large magnitude of the difference (20 points), its statistical significance will depend on the sample size. According to statistical theory, we need a sample size of about 50 or more people in each of the groups for the difference to be statistically significant at the 95 percent level of confidence. If, in fact, we meet the sample size requirement, then the difference of 20 points will be statistically significant at the 95 percent level of confidence.

What does this result mean? Many marketers will look at this result and conclude that since there is a 20-point difference and the difference is statistically significant, there must be a big difference between Groups A and

B. In reality, if we had done a census (i.e., surveyed the entire population) instead of surveying a sample, the difference between Group A and Group B may turn out to be smaller.

In other words, what this result tells us is merely this:

Given our particular sample size, there is a 5 percent chance that in the population represented by this sample, the proportions for Group A and Group B are not different.

That's all. Statistical significance does not tell us anything about how big the difference is. It only tells us the probability with which a difference found in the sample would not be found in the population. Thus, for this case statistical significance would allow us to conclude that there is only a 5 percent chance that in the population the proportion of Group A favoring the product is not higher than Group B; we are taking a 5 percent risk of concluding a difference exists when there may not be any such difference. If this difference were significant at the 99 percent level of confidence, it would not have become larger. It would only mean that there is a 1 percent chance that the difference observed in the sample would not be observed in the population. Thus, we are only taking a 1 percent risk.

Practical significance

From a marketing perspective, the statistically significant difference of 20 points may be meaningful or meaningless. It all depends on our research objectives and resources. If it costs millions of dollars to reach each additional percentage of the market, we may decide to funnel resources toward Group A since it has a higher acceptance rate. In this case, the difference may be termed a "big" difference because (a) we are reasonably sure (95 percent or 99 percent sure) that the difference observed in our sample also exists in the population and (b) each percentage of difference is worth millions of dollars to the client. Thus, statistical significance should not be used to decide how big a difference is, but merely to ascertain our confidence in generalizing the results from our sample to the population.

In another situation this same difference may be ignored despite the fact that it may be statistically significant. For instance, if the marketing costs are so low that it makes sense to market to both groups, we can ignore the difference (even though it is significant) and treat both groups as if they are the same. We may choose to market to both groups as if they had similar acceptance rates (even though our statistical test was significant).

Our logic is the following: Although we can be 95 percent sure that the difference observed here exists in the population, given the marketing scenario, the difference is not meaningful. Thus, the relevance of a statistically significant difference should be determined based on practical criteria including the absolute value of the difference, marketing objectives, strategy, and so forth. The mere presence of a statistical significance does not imply that the difference is large or that it is of noteworthy importance.

Implications

Statistical significance of a result is not a rule of thumb to ascertain how "big" a difference is, but a context dependent tool to assess the riskiness of the decisions we make based on a given sample. At most it can be used to ascertain that a difference actually exists in the population when we observe it in the sample.

One last thing: How can we avoid this trap whereby significance takes on a larger meaning? We recommend using the term statistically discernible instead of statistically significant when discussing results. While this cannot fully solve the problem, it certainly does not aggravate it either. We, as researchers can explicitly note in our reports: "While such and such result is statistically discernible, its practical significance will depend on..." In this way we can alert the end-user of our data to interpret the results realistically.