

Making Count-Based Claims from Sample Data

Daniel M. Ennis and Benoît Rousseau

Background: Numerical superiority claims such as “2 out of 3” or “4 out of 5 people prefer...” are powerful and commonly used to promote products. The National Advertising Division (NAD[®]) is the advertising industry’s self-regulatory body that adjudicates advertising claims in the USA. Examples of numerical superiority claims that were challenged at the NAD include a claim by General Mills that “In a national taste test, nearly 2 out of 3 Americans agree that Yoplait[®] Greek tastes better than Chobani[®]”¹. Another example is the claim that “Four out of five consumers prefer LG Cinema 3D over ... Samsung Active 3D in head-to-head comparison.”². These claims involved performance numbers based simply on the actual results of the experiments. These cases raise a number of issues in using sample data to support advertising claims. One issue is whether it is justified to report sample data results in a claim where clearly the advertiser does not wish to limit the message to the particular consumers tested on a particular occasion. Even if it is truthful to say that particular results were obtain in a particular experiment, the implication to a consumer is that he/she will experience the benefit at a similar level. In reality the advertiser wishes to communicate a benefit message that applies to a target population at large, rather than reporting what amounts to an interesting factoid about an experiment that they conducted. A second issue is whether there should be consistency in the application of statistical principles to superiority tests that are not numerical and those that are. In this report we consider these two issues using a scenario for illustration of the points we will make.

Scenario: You work for a major beverage company that markets pomegranate juice. A small competitor that wishes to establish its market position, claims that its product is preferred to yours by almost 3 out of 5 consumers. To support this claim they conducted a nationwide consumer test among users of the category. The test itself was competently conducted. It was double-blind, included demographic specifications and brand share that matched the target users, and was conducted in a central location setting. Two sets of codes were used and sample presentation order was balanced³, sample volume and temperatures were

carefully controlled. Data analysis involved the splitting of the “No preference” responses equally between the two brands⁴, resulting in significant superiority of their product over yours. The results also showed that 58% of consumers preferred the advertiser’s product, based on a sample of 250 consumers. The advertiser conducted no further statistical testing on the data and naïvely reported the test results in their claim, “almost 3 out of 5 consumers prefer...”. 60% corresponds to a count-based measure of 3 out of 5. You gleaned this information when your company initiated an advertising claim challenge.

Superiority Testing: It is a well-established practice when making superiority claims that the data is subjected to statistical testing⁵. This practice is based on the first issue mentioned in the background above – claim statements refer to a target population and are not strictly applicable to the sample of consumers tested. Assuming that the two products do not differ on the variable of interest (null hypothesis), it is typically necessary to show that this assumption does not hold with a reasonable level of confidence when applied to the population at large. Figure 1 shows the minimal preference percentage needed to reject the hypothesis of no preference for various sample sizes in a one-tailed test at the 95% level. It can be seen that for small sample sizes, large preference proportions in the sample are needed and that they become smaller as the sample size increases. The idea of statistical significance for preference, above the threshold of 50%, is well established and accepted, as few researchers would conclude that a product is superior to another simply because they obtain a ratio > 50/50 preference, such as 51/49, in their experiment.

Numerical claims such as “2 out of 3” or “3 out of 5” are simply alternative expressions of superiority which have been numerically specified and correspond to 67% and 60%, respectively⁶. Similarly, a traditional superiority claim could be stated numerically as greater than “1 out of 2”. Count-based comparisons also may include metrics such as “1 out of 5 people meet their future spouses on-line” (which may or may not be true.) Figure 2 is similar to Figure 1 but with the minimal preference percentage needed at

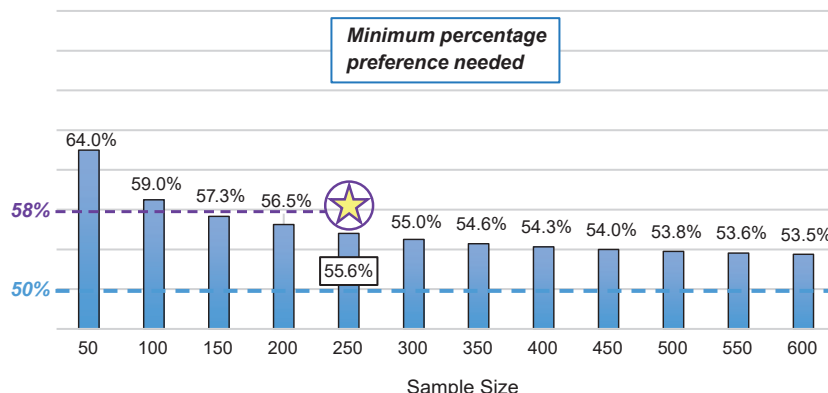


Figure 1. Minimum preference percentage needed to declare superiority ($p=0.05$) depending on the experiment’s sample size.

various sample sizes to reject hypotheses that are less than the numerical claim requirements of 60%. For instance, to establish a “3 out of 5” (60%) claim it is necessary to reject all hypotheses that include less than three out of five people preferring the advertiser’s product (59%, 58%, 57%, etc.). Similarly, for simple superiority, we reject the 50% point, but also 49%, 48%, 47%, etc. The underlying reasoning is that we must show that our reported result is unlikely to occur if the true preference probability is 50% or less. It is clear from the established standard expressed in Figure 1 that the same principles should also apply to other statements of superiority as shown in Figure 2. These two figures illustrate principles needed to address the issues raised in the Background above: a) Merely reporting test results is not sufficient to establish a claim because it is necessary to reject non superiority hypotheses, and b) In the case of a numerical superiority claim, such as “3 out of 5”, the same principles should apply.

Challenging the Advertiser: The advertiser’s data showed that 58% of consumers preferred their product to yours. While they rightfully concluded that their product was significantly preferred over yours, they went beyond their statistical evidence and, without any statistical testing, the advertiser claimed that “almost 3 out of 5” people preferred the advertiser’s product. You submit two objections to this conclusion: 1) There is no statistical standard for “almost” which can be arbitrarily set, and 2) the advertiser has not considered that the test results may be consistent with possible preference outcomes that are less than three out of five. To establish a “3 out of 5” claim it will be necessary to reject these hypotheses just as a hypothesis of 50:50 or less is rejected in a non-numerical preference claim. Even if the experimental outcome had been exactly “3 out of 5”, or 60%, it would still be likely that the true preference was below “3 out of 5”. With such high odds for an incorrect conclusion, it is easy to see how the “3 out of 5” claim cannot be made. Since 58% is in fact even less than three out of five or 60%, the advertiser cannot reject the hypothesis of less than 60% with any reasonable confidence. The advertiser’s result is shown as a star in Figure 1 and Figure 2. You argue that the claim should be discontinued.

Losing a case such as this by the advertiser at the NAD may lead to a recommendation to modify the claim and remove the numerical component. If this is done, the advertiser may still claim superiority without using the “3 out of 5” form of the claim. With their test result of 145 choices in favor of their product out of 250, they can claim superiority at the 99.3% confidence level (although they should have had a simple superiority test in their protocol). This challenge, although successful in removing the original claim, may become a Pyrrhic victory for your company as your competitor can continue to advertise superiority. You decide to now focus on the reason that your new competitor makes a preferred product, irrespective of the degree of superiority.

Conclusion: It is not uncommon for advertisers to base their claims on actual test results without considering the need to establish a basis for the generalization that their claims imply. Numerical superiority claims are extremely attractive in marketing the benefits of consumer products. It is not surprising that these claims are sometimes exaggerated or even simply naively reported based on a single product test. Consumers may be misled to conclude that the reported results from this single experiment apply to them, when in fact, they may belong to a substantial group who may never experience the benefit claimed.

References

1. General Mills Inc. (Yoplait Blended Yogurt), Report #5715, NAD/CARU Case Reports (May 2015). Challenger: Chobani, Inc.
2. LG Electronics USA, Inc. (Cinema 3D Television and 3D Glasses), Report #5416, NAD/CARU Case Reports (January 2012). Challenger: Samsung Electronics America, Inc.
3. Rousseau, B., Ennis, D. M. and Ennis, J. M. (2018). Identifying and removing sources of bias in product tests and surveys. In D. M. Ennis, J. M. Ennis, and B. Rousseau (Eds.), *Tools and Applications of Sensory and Consumer Science* (pp. 140-141). Richmond, VA: The Institute for Perception.
4. Ennis, D. M. and Ennis, J. M. (2012). Accounting for no difference/preference responses or ties in choice experiments. *Food Quality and Preference*, 23(1), 13-17.
5. ASTM E1958-18, Standard Guide for Sensory Claim Substantiation, ASTM International, West Conshohocken, PA, 2018, www.astm.org
6. Ennis, J. M. and Ennis, D. M. (2012). Justifying count-based comparisons. *Journal of Sensory Studies*, 27(2), 130-136.

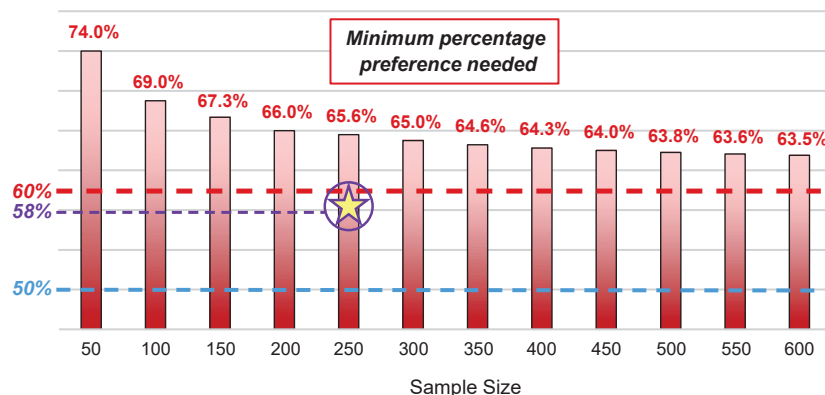


Figure 2. Minimum preference percentage needed to declare “3 out of 5 prefer” ($p=0.05$) depending on the experiment’s sample size.