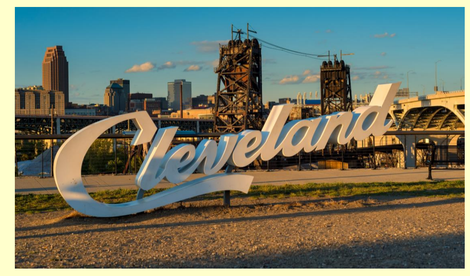
Planning for Business Relevance when Planning Sensory Difference Tests

John M. Ennis, Benoit Rousseau, & Daniel M. Ennis
The Institute for Perception, Richmond, VA
john.m.ennis@IFPress.com
www.IFPress.com



Presented at the 2018 **Society of Sensory Professionals** annual conference in Cleveland, OH, USA

Background

Sensory difference testing is a business relevant topic:

- Cost reductions
- Ingredient and process changes
- Compliance with health initiatives
- Quality control

Challenge

Sensory difference tests continue to rely on statistical techniques disconnected from the business-relevant issues difference tests are conducted to address

Scenario

Evaluation of a possible ingredient change to a prominent candy bar product:

- **Goal:** Ensure that there is a low probability that consumers will notice the change
- **Test:** Internal panel Tetrad test with 90 evaluations
- **Data:** 39 correct responses out of 90

The p-value associated with standard analysis of these data is 0.03, as shown in Figure 1, so that a standard analysis would reject this proposed change.

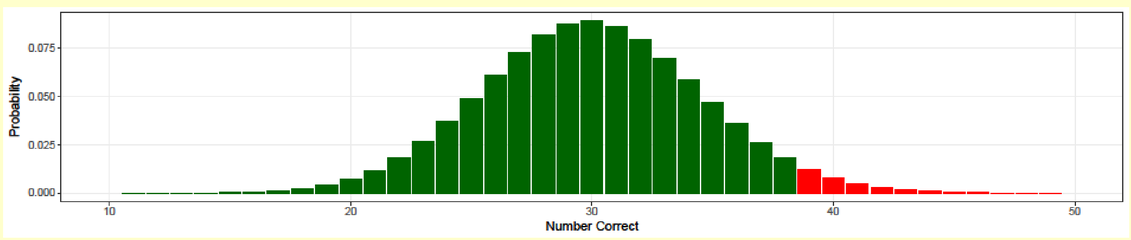


Figure 1. Probability of various outcomes in a sample of 90 independent Tetrad trials

Note: The estimate of Thurstonian δ, the underlying sensory effect size, is 0.768.

Criticisms of Standard Approach

Several criticisms exist of the standard approach to analyzing sensory difference testing data:

- **Power:** Several standard methods, including the Triangle test, require very large sample sizes to detect moderately-sized sensory differences reliably
- **Consumer Relevance:** Without considering the size of the difference that matters to consumers, it is not possible to know what size difference one wishes to detect

Interpreting Difference Testing Results

Even when power and consumer relevance are considered, there is still a need to know how to interpret significantly different results.

For example:

- Suppose we seek to detect a δ of 1 in a Tetrad test.
- 65 evaluations are required for 80% power
- 110 evaluations are required for 95% power

In both cases, Figure 2 shows that effect sizes between 0.5 and 0.75 are frequently rejected.

Figure 2. The probability of rejecting product changes at 95% or 80% power

Consumer Relevant Confidence

To conduct business relevant sensory difference tests, we must answer:

"How confident are we that the underlying sensory difference is not consumer relevant?"

If τ_R is the consumer relevant difference, the probability that the sensory difference is less than τ_R is called the **consumer relevant confidence**.

Given an estimate of τ_R, we propose:

- Difference tests should be designed to have high expected consumer relevant confidence when underlying difference sizes are acceptable
- Following a sensory difference test, consumer relevant confidence should be calculated

Benefit: Consumer relevant confidence is informative regardless of whether the test result is statistically significant

Accounting for Variability

Statistical Issue: Neither δ nor τ_R are known with certainty.

Solution: Given two normal random variables *X* and *Y,* the theory of multiplicative comparisons answers the question:

What is $P(X > Y \text{ and } X > 0)$?

Because τ_R is assumed to be well above 0, this theory can be applied to answer the question:

What is $P(\tau_R > \delta)$?

Thus multiplicative comparisons can be used to compute consumer relevant confidence

Illustration of Theory

Suppose:

- The true value of τ_R is 1
- τ_R is estimated via a Same-Different test with 100 same and 100 different pairs
- Tetrad testing occurs for various values of δ

Figure 3 illustrates how expected consumer relevant confidence compares with the percentage of tests rejected

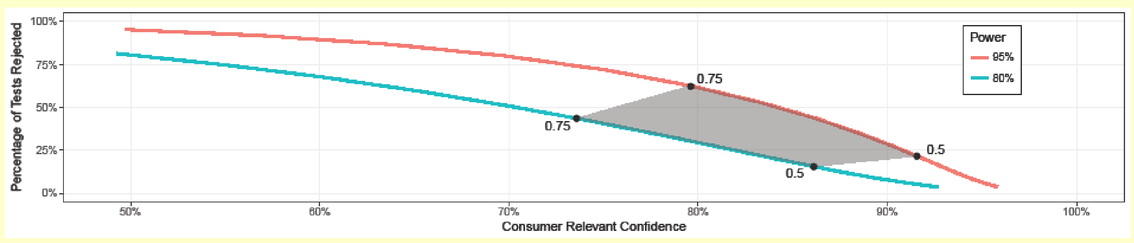


Figure 3. The highlighted region shows cases where a rejection would occur, although there is high confidence that the change is not consumer relevant

Extension of Scenario

Suppose that, in addition to the data collected in the original scenario, a Same-Different test is also conducted with consumers to estimate the consumer relevant difference

The data and results of Thurstonian analysis for this experiment are shown in Tables 1 and 2

Response	Same Pair	Different Pair
"Same"	54	50
"Different"	46	50

Table 1. Same-Different test results.

Parameter	Estimate	Variance in Estimate
δ	0.608	0.300
τ_R	1.040	0.013

Table 2. Thurstonian analysis of data in Table 1.

Suppose further that it has been previously established that internal panel testing typically yields δ estimates 25% larger than estimates obtained from consumer testing, because of increased panel sensitivity, so that both the estimate of τ_R and the variance in this estimate can be rescaled to values of 1.3 and 0.02, respectively

Using multiplicative comparisons, consumer relevant confidence would then be computed to be greater than 98%, **despite the fact the ingredient change would have been rejected**

Conclusion

Using the theory of consumer relevant confidence, we can:

- Design tests and analyze test results according to business goals rather than relying on *p*-values from difference tests
- We can compute confidence that sensory effect sizes are less than an experimentally determined consumer relevant action standard
- This calculation provides information regardless of the statistical significance of the test

Using consumer relevant confidence, we can make informed business decisions

References

1. Cumming, G. (2014). The new statistics: Why and how. Psychological Science, 25(1), 7-29.

2. Kruschke, J. K. and Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychonomic Bulletin & Review, 25(1), 178-206.

3. Ennis, J. M. and Jesionka, V. (2011). The power of sensory discrimination methods revisited. Journal of Sensory Studies, 26(5), 371-382.

4. Ennis, D. M. and Bi, J. (1998). The beta-binomial model: Accounting for inter-trial variation in replicated difference and preference tests. Journal of Sensory Studies, 13(4), 389-412.

5. Ennis, J. M., Ennis, D. M., Yip, D., and O'Mahony, M. (1998). Thurstonian models for variants of the method of tetrads. British Journal of Mathematical and Statistical Psychology, 51(2), 205-215.

6. Rousseau, B. and Ishii, R. (2016). How do perceived sensory differences and preferences relate? In The 2nd Asian Sensory and Consumer Research Symposium in Shanghai, China.

7. Rousseau, B. (2015). Sensory discrimination testing and consumer relevance. Food Quality and Preference, 43, 122-125.

8. Ennis, D. M. (1990). The relative power of difference testing methods in sensory evaluation. Food Technology, 44(4), 114, 116-117.

9. Ennis, J. M. and Ennis, D. M. (2011). Confidence bounds for multiplicative comparisons. Communications in Statistics-Theory and Methods, 40(17), 3049-3054.