Sensory equivalence testing – the reversed null hypothesis and the size of a difference that matters

Paul Arents, C.A.A. Duineveld, and Bonnie M. King

Quest International Nederland BV, P.O Box 2, 1400 CA Bussum, The Netherlands

Sensory Scientists are often confronted with the problem of proving equivalency, for example in situations involving product matching after process change. Standard difference tests use a null hypothesis of no difference. It has long been recognized that the absence of significant difference (failure to reject this null hypothesis) cannot be used to justify a claim of product equality. Power tests have been suggested by some authors as a way of getting around this problem (e.g. Cohen, 1988; Schlich, 1993) while others claim that higher observed power does not imply stronger evidence for a null hypothesis that is not rejected. (Hoenig and Heisey, 2001). MacRae (1995) has proposed confidence intervals as an alternative approach that requires only the choice of an alpha level to control the probability of the true answer lying outside the bounds. Schuirmann (1987) combines the idea of confidence interval with the intuitive ideal of the sensory scientist: a null hypothesis of nonequivalence. This approach, as the power approach, requires the sensory scientist to specify the smallest degree of difference detectability, *delta*, which matters.

The work presented in this poster shows how the Two One-Sided Test (TOST) as described by Schuirmann was applied to problems of sensory equivalency. Profiling is a more efficient method than difference testing for collecting sensory data. Given sufficient prior knowledge of how a panel uses an intensity scale, the sensory scientist can postulate a minimum *delta*. This value of *delta* can be used in the TOST and then later refined by comparing the rejection regions from TOST with the corresponding rejection region from a standard t-test. Additional information from difference tests, Weber ratios, or customer feedback can be used to confirm the choice of *delta*. The procedure is illustrated with several examples.

References

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hoenig, J. M., Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician, February 2001, Vol. 55, No. 1,* 19-24.
- MacRea, A. W. (1995). Confidence intervals for the triangle test can give reassurance that products are similar. *Food Quality and Preference, 6,* 61-67.
- Schlich, P. (1993). Risk tables for discrimination tests. *Food Quality and Preference, 4,* 141-151.
- Schuirmann, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics, Vol. 15, No. 6,* 657-680.