A Thurstonian Model for the Unspecified Hexad Test

Keith Eberhardt, Karen Robinson, Victoire Aubry

Background

- Discrimination tests are an indispensable tool to gauge and maintain product quality, where the sensory scientist has the possibility to customize test design to balance test power, sensory fatigue and protocol efficiency.

- When products are variable, there is interest in protocols involving multiple samples of each variant of the product.
  - Chocolate chip cookies
  - Trail mix
  - Cereals with fruit and nuts
  - Cottage cheese

Background

Several discrimination test methods are already available

- Triangle and tetrad have been extensively studied (published risk tables based on Thurstonian models)
- “2 out of 5” and “hexad” tests:
  - Guessing probability in both tests is 0.1
  - Intuitively, low guessing probability should make these tests more powerful
  - Psychometric functions have not been published
  - Risk profile calculations are so far not based on Thurstonian model
  - Possible disadvantage: task complexity may influence the quality of the results (Lawless and Heymann, 1999; O'Mahony et al. 1994)

Background

- Sorting: Recent research has capitalized on consumers' and experts' natural ability to group or sort samples according to their own criteria (Lelièvre et al. 2008; Ballester et al. 2005; Cartier et al. 2006).

- Internally, unspecified hexad (or "3 out of 6" test) had been used informally with trained panel to gauge difference between two product variants based on samples from multiple lots
  - Roundtable format
  - Feedback from panelists has been encouraging

Objective

- Develop a Thurstonian model for unspecified hexad test
- Use the model to derive the risk profile in terms of the inter-product distance, δ
- Report on an empirical test by sensory panel
  - Unspecified hexad ("3-out-of-6“) tests
  - Double triangle tests (sequential serving of 2 triads)
  - Comparison in terms of observed alpha and beta risks, and estimated d’ values

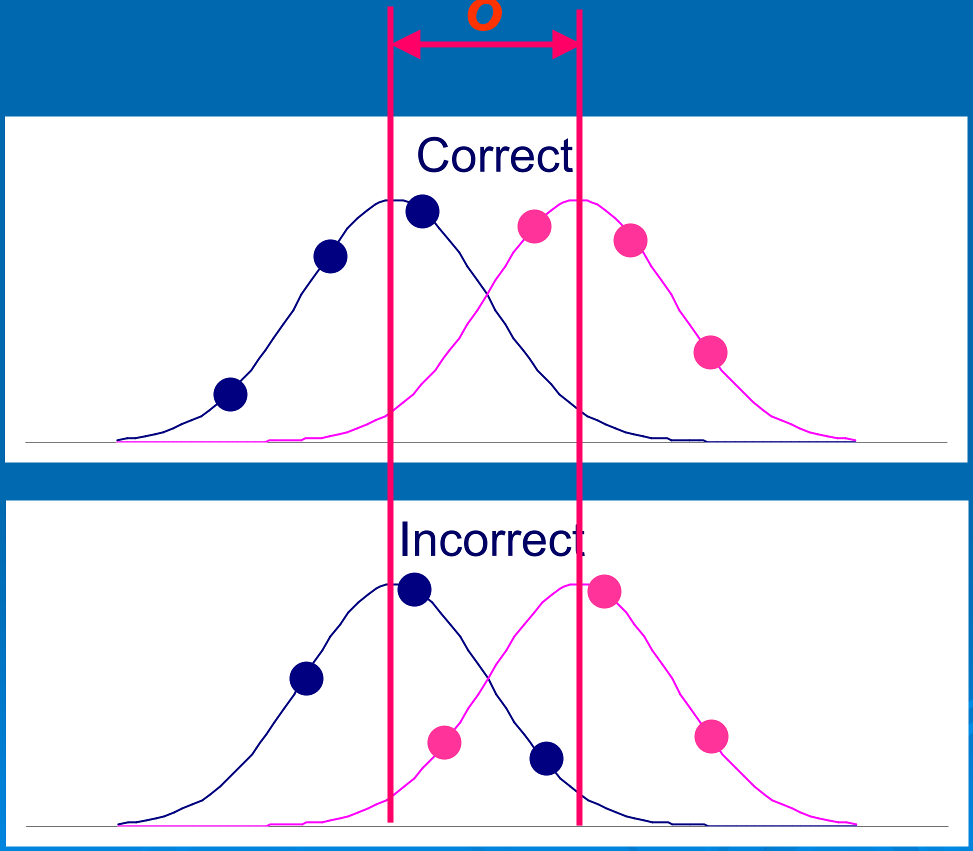
The Unspecified Hexad
or “3 out of 6” task

564 885 498 125 726 157

- Here are 6 samples: 3 are of one type and 3 are of another
- Sort these samples into 2 groups of 3
- For each group, specify the sensory criteria that helped you decide which sample belonged to which group

Statistical Model and Analysis

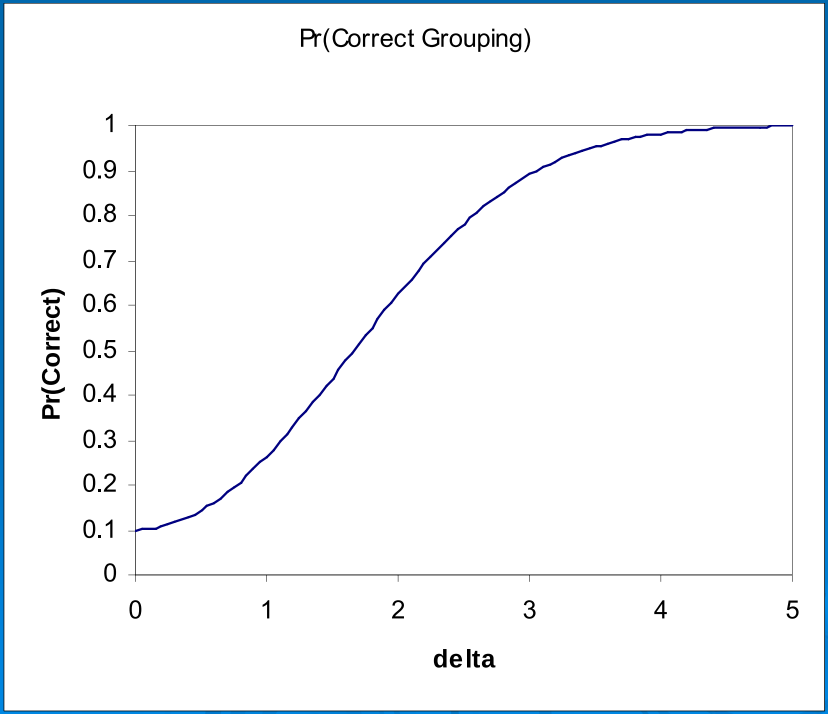
- Thurstonian Model:



Psychometric Function

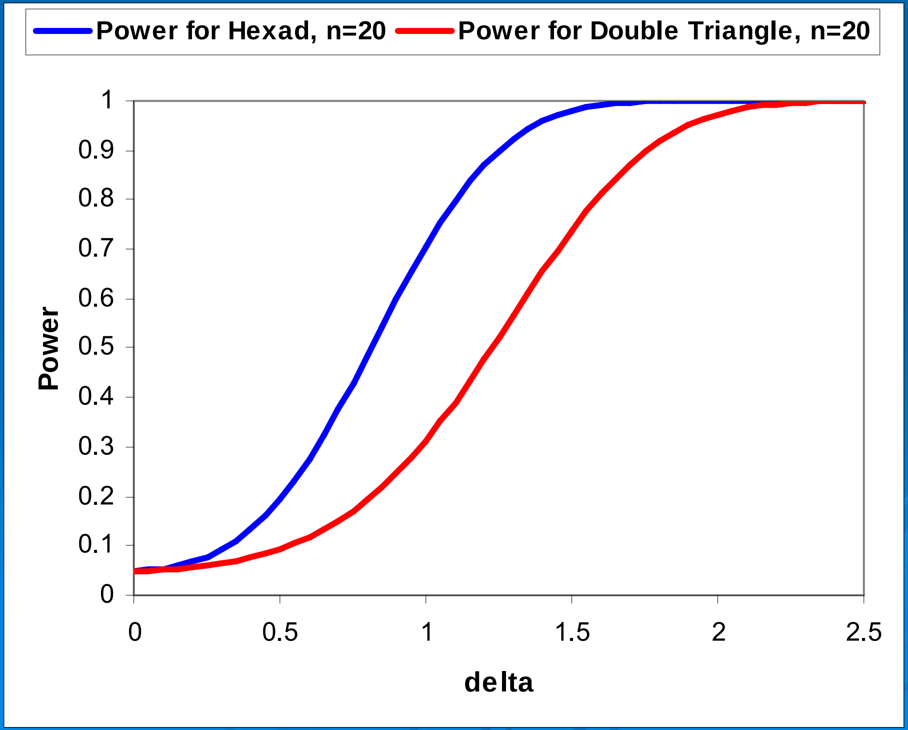
$$\Pr\left(\text{Correct } \middle| \delta \right) =$$

$$3\int_{-\infty}^{\infty}\Phi^3(t)(1-\Phi(t-\delta))^2\varphi(t-\delta)dt + 3\int_{-\infty}^{\infty}\Phi^3(t)(1-\Phi(t+\delta))^2\varphi(t+\delta)dt$$



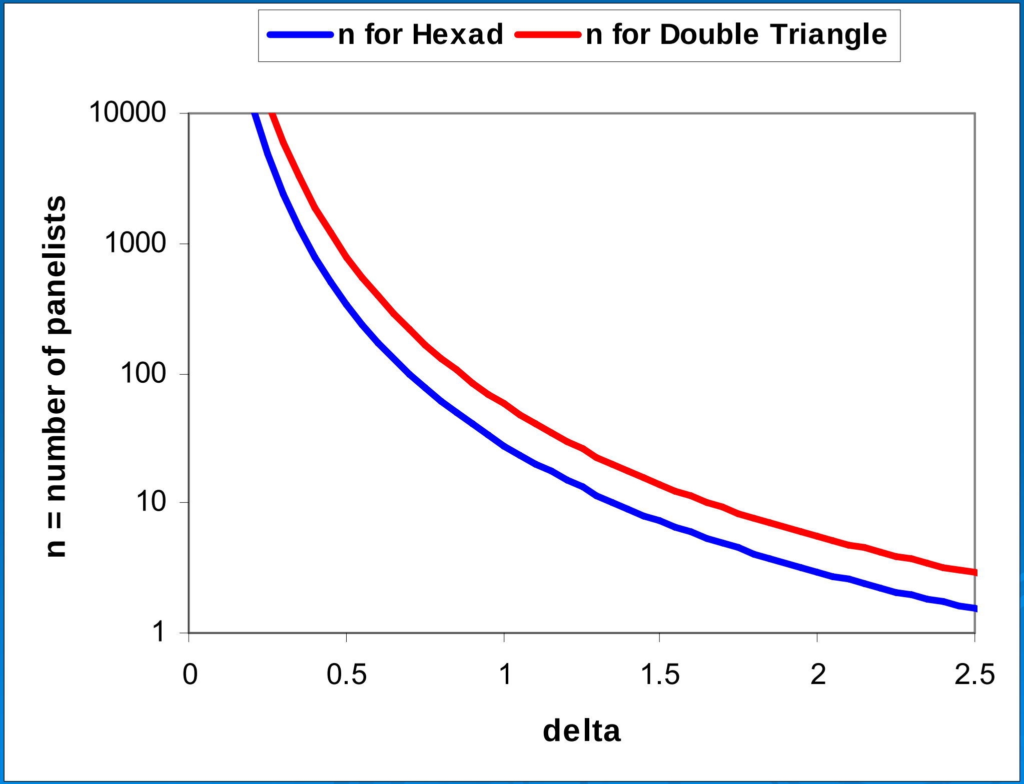
Power Curves for α=0.05

- Hexad test vs. double triangle, both with n=20 panelists
- Total of 120 samples, 6 per panelist, for each test
- Hexad has higher power for all δ values
  - Assuming independence of the two triangle tests per panelist
  - Power calculated by Normal approximation to binomial



Sample Size Comparison

- Comparison at $\alpha=0.05$, $\beta=0.2$ (power = 0.8)
  - Based on Normal approximation to binomial
- Hexad requires lower “n” compared to double triangle



Sensory Test Protocol

- A trained descriptive panel was used to differentiate pairs of confusable stimuli (Tropical Punch cold beverage solutions varying in concentration)
- The data were collected according to a balanced design in order to avoid learning effects and bias of test type (while half of the panel was doing hexad, the other half was doing double triangle)
- We used three pairs, which were designed to represent a range of task difficulty
  - Pair 1: 70g/L *vs.* 75g/L
  - Pair 2: 70g/L *vs.* 80g/L
  - Pair 3: 70g/L *vs.* 85 g/L

Sensory Test Protocol

- Hexad:
  - 6 samples per test
  - 3 reps per panelist (18 samples tasted)
  - 6 panelists → 18 hexad tests
- Double triangle:
  - 6 samples per test
  - 3 reps per panelist (18 samples tasted)
  - 6 panelists → 18 double triangles
    = 36 triangles (treated as independent)

Results: Observed α and β

$H_0: \delta = 0$ $H_1: \delta = 1$	Double triangle	Hexad
Pair 1 (70 g/L vs. 75 g/L)	Correct = 14 / 36 $\alpha = 0.29$ $\beta = 0.43$	Correct = 3 / 18 $\alpha = 0.27$ $\beta = 0.26$
Pair 2 (70 g/L vs. 80 g/L)	Correct = 16 / 36 $\alpha = 0.11$ $\beta = 0.69$	Correct = 6 / 18 $\alpha = 0.006$ $\beta = 0.83$
Pair 3 (70 g/L vs. 85 g/L)	Correct = 28 / 36 $\alpha < 0.001$ $\beta = 1.0$	Correct = 12 / 18 $\alpha < 0.001$ $\beta = 1.0$

These calculations assume replicates by same panelist are independent. Overdispersion was non-significant.

KRAFT

Results: Comparable d’

	Double triangle	Hexad
Pair 1 (70 g/L vs. 75 g/L)	Correct = 14 / 36 d' = 0.8 90% CI: 0 to 1.6	Correct = 3 / 18 d' = 0.63 90% CI: 0 to 1.1
Pair 2 (70 g/L vs. 80 g/L)	Correct = 16 / 36 d' = 1.2 90% CI: 0 to 1.9	Correct = 6 / 18 d' = 1.2 90% CI: 0.5 to 1.7
Pair 3 (70 g/L vs. 85 g/L)	Correct = 28 / 36 d’ = 3.0 90% CI: 2.3 to 4.0	Correct = 12 / 18 d’ = 2.1 90% CI: 1.6 to 2.8

These calculations assume replicates by same panelist are independent. Overdispersion was non-significant.

Discussion

- In our limited test of hexad *vs.* double triangle, results were consistent with predictions from Thurstonian model
  - Comparison of observed α and β values was consistent with higher power of hexad
  - Estimated d’ values were similar
- For practicality, the hexad method should be used primarily for testing products with limited carry-over effects, and requiring low preparation (due to simultaneous serving)
- Feedback from panelists regarding difficulty and speed of testing was encouraging

Next Steps

- When samples of each product represent multiple lots or batches, this two-distribution model assumes lot-to-lot variation is negligible
- If lot-to-lot variation is not negligible, should explore models that explicitly take this into account
  - E.g. use 6 distributions to model 3 lots of each product variant
- More data will be collected on other food categories with
  - Limited carry-over effects
  - Simple preparation

References

- Ballester, J., Dacremont, C., Le Fur, Y., Etiévant, P. The role of olfaction in the elaboration and use of the Chardonnay wine concept. *Food Quality and Preference*, 2005, 16 (4), 351-359.
- Cartier, R., Rytz, A., Lecomte, A., Poblete, F., Krystlik, J., Belin, E., et al. Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality and Preference*, 2006, 17, 562–571.
- FIZZ 1994-2007, Biosystemes, Couternon, France.
- Lawless and Heymann, Sensory Evaluation of Food: Principles and Practices; Springer, 1st edition, 1999.
- Lelièvre, M., Chollet, S., Abdi, H., Valentin, D. What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality and Preference*, 2008, in press. doi:10.1016/j.foodqual.2008.05.001
- O’Mahony, M., Masuoka, S., Ishii, R. A theoretical note on difference tests: Models, paradoxes and cognitive strategies. *Journal of Sensory Studies,* 1994, 9, 247-272.
- O'Mahony M., Rousseau B., Discrimination testing: a few ideas, old and new. *Food Quality and Preference*, Volume 14 (2), 2002, 157-164.