Establishing a “meaningful difference” range using Cohen’s Delta effect size measure

P. Hongsoongnern, A. Cece, and J. Crudele
Unilever, United States

Introduction

Quality control is a critical element most manufacturing companies use to promote their products. Quality products must be produced within an established range of some measurements to be considered acceptably high quality. Thus, products need not be identical to be equally acceptable. Two products falling within the specification limit may be noticeably different but still considered acceptable. Minimal research has been done to understand “meaningful difference” from product specification aspect.

Why Effect Size?

- Effect Size is defined as “a standardized value that estimates the magnitude of the differences between groups” *(Thomas, Salazar, & Lander, 1991)*
- Unlike statistical difference, which relies primarily on sample size, Effect size is relative independent of sample size
- Effect Size thus enables us to compare current results to previous findings (historic data) to better understand “practical significance” of the results

Objectives

1) To determine if products made within an established range are rated significantly different in a traditional In-Home-Use consumer study, and
2) To determine variations in product ratings due to sampling replication (i.e. consumers evaluate same product twice).

Methodology

Test Design

- 5 Cell, 2-Product Sequential Monadic Tests, Branded HUT, followed by completing an online survey
- Order of presentation is balanced
- 10 day usage period per product

Consumer Sample

- N = 200 per cell, Females, Age:18-64 years old
- Current users of the brand
- Usage: at least 4x per week

Sample Structure

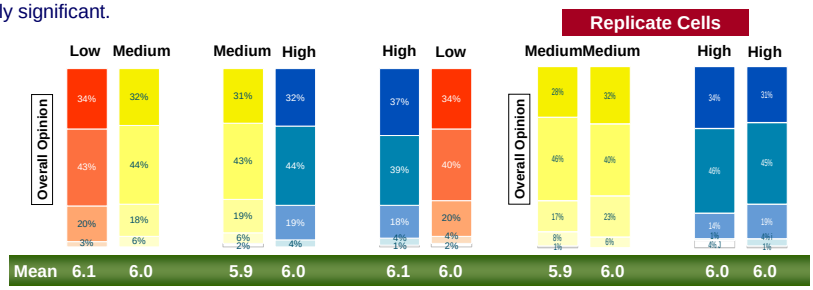
- Bar soap products, produced within an established specification, ranging low, medium and high in key ingredient were selected

Table 1: Sample structure

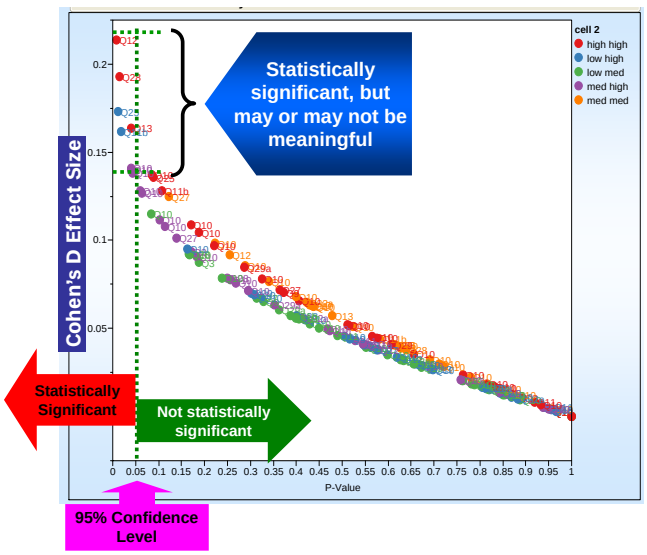
Variation / Cell	Low	Medium	High
Cell 1	**Low**	**Med**	
Cell 2		**Med**	**High**
Cell 3	**Low**		**High**
Cell 4 (replicate cell)		**Med/Med**	
Cell 5 (replicate cell)			**High/High**

Results

Results indicated that overall liking was not substantially impacted by the products ranging from low to high in key ingredient at 95% confidence level, however, differences in some product attributes were statistically significant.



Then Cohen’s Delta effect size were calculated for each cell and demonstrated in the chart below:



Conclusion/Discussion

1) This work has demonstrated that measurable variation due to test subject can be incurred as seen by the variation in responses to the replicate products. In fact these differences can be large enough to sometimes be detected as significantly different in standard consumer test situations. Understanding the magnitude of noise provides us with the ability to begin to define a lower boundary below which effects are likely to be not meaningful. Based on this work, Cohen’s D of ~ 0.15-0.2 is the possible range for meaningfulness consideration.
2) It is possible that products that consumers are typically exposed to can be rated statistically different in consumer evaluation, but this difference might be considered NOT meaningful or of practical importance.
3) It is a reasonable practice to consider this approach of meaningful difference, especially in cost reduction program.