IFSCC 2025 full paper (IFSCC2025-642)

"The Science of Color: Iterative Clustering and Sensory Testing Shape the Future of Lipstick Visual Perception"

Franciane Cortez [1*], Vitoria Cavalheiro [1], Rebecca Graciolli [1], Ana Carolina Ribeiro [1], Dilene Bet [1], Jakceli Da Cruz [1], Juliana Amorim [1], Pedro Abreu [1], Fernanda Henning [1] and Desirée Schuck [1].

[1] Grupo Boticário, São José dos Pinhais, Paraná, Brazil.

1. Introduction

In the cosmetics industry, color transcends mere aesthetic appeal; it's a fundamental pillar of the consumer experience. Color influences brand perception, product appeal, perceived quality, loyalty, and ultimately, purchasing decisions [13]. Color perception is a complex process beginning with the interaction of light and an object. The human eye detects wavelengths between 400 and 750 nm, processed in the retina by photoreceptors (cones and rods) that convert light stimuli into electrical signals corresponding to primary colors (red, green, and blue). The brain's interpretation of these signals results in individual color perception [5, 6, 14].

Despite the human ability to discern color nuances, the subjectivity of perception necessitates reliable parameters for analysis. Colorimetry, through systems like CIE, aims to provide these parameters. The CIEDE2000 formula, a robust model considering lightness, chroma, and hue (CIELab), refines color assessment and approximates human perception, using Delta E (ΔE) as a metric [9, 16, 13]. Combining color scales with spectrophotometers, colorimeters, and digital cameras allows for more comprehensive analyses [3, 6].

However, color tolerances are generally broad, failing to consider the specifics of each context. To overcome this limitation, this study uses data clustering, specifically the K-means algorithm, to identify groups of similar colors, enabling more efficient analysis than individually analyzing each shade. K-means partitions the dataset by minimizing the distance of elements to centroids, grouping data with high internal similarity and low similarity between groups [14, 4, 11, 17].

This work presents a methodology combining instrumental analysis, K-means clustering, and sensory analysis to define color tolerances in lipsticks. The analysis of 40 lipstick shades, encompassing a wide range of hues, will yield Lab (CIELab) and LCh (CIELCh) values for each color. This innovative approach will integrate clustering algorithms, instrumental methods, and consumer testing, enabling the cosmetics industry to understand its color spectrum, identify key groups, and establish precise specifications for each.

2. Materials and Methods

2.1 Samples

A palette of 40 lipstick shades, encompassing a full range of hues from the lightest to the most intense, was analyzed. Using a color measurement system, Lab (CIELab) and LCh (CIELCh) values were obtained for each color. The K-means algorithm, through an iterative clustering process, classified these colors into four main groups (pinks/nudes, purples, reds, and browns) and eight subgroups. For each group, a centroid was defined, representing the color closest to the average of the group's colors. Nine of these centroids (Formulas S1 to S9) were then selected as a basis for producing nine to twelve variations each, resulting in a total of 101 samples. The variations maintained the basic formula structure, differing only in pigment concentration (See Tables 1-9 for detailed).

Table 1. Sample S1: Pigment Modifications

Base	Variation	% Change Key Dispersion	Key Dispersion
S1	S1	Standard	-
S1	S1A	+10%	Black Dispersion
S1	S1B	+19%	Blue Dispersion
S1	S1C	+5% and +10%	Black and Blue Dispersions
S1	S1D	+5%	White Dispersion
S1	S1E	+2,7% and +3%	White and Brown Dispersion
S1	S1F	+11%	Red Dispersion
S1	S1G	+19%	Red Dispersion
S1	S1H	>6000%	Yellow Dispersion
S1	S1I	+10%	All Dispersions
S1	S1J	+46%	Red Dispersion
S1	S1K	-10%	All Dispersions

Table 2. Sample S2: Pigment Modifications

Base	Variation	% Change Key Dispersion	Key Dispersion
S2	S2	Standard	-
S2	S2A	+3%	Red Dispersion
S2	S2B	+10%	Red Dispersion
S2	S2C	+6%	Yellow Dispersion
S2	S2D	+21%	Yellow Dispersion
S2	S2E	+9%	White Dispersion
S2	S2F	+16%	White Dispersion
S2	S2G	+100%	Blue Dispersion
S2	S2H	+200%	Blue Dispersion
S2	S2I	+300%	Blue Dispersion
S2	S2J	+48%	Yellow Dispersion

Table 3. Sample S3: Pigment Modifications

Base	Variation	% Change Key Dispersion	Key Dispersion
S3	S3	Standard	-
S3	S3A	+2% and +2%	White and Red Dispersion
S3	S3B	+3% and +3%	White and Red Dispersion
S3	S3C	>1600%	Yellow Dispersion
S3	S3D	+4000%	Yellow Dispersion
S3	S3E	+14% and +14%	Black and Blue Dispersion
S3	S3F	+467%	Blue Dispersion
S3	S3G	+11%	Red Dispersion
S3	S3H	+11% and +11%	Red, and Yellow Dispersion
S3	S3I	+2%	White Dispersion
S3	S3J	+4%	White Dispersion

Table 4. Sample S4: Pigment Modifications

Base	Variation	% Change Key Dispersion	Key Dispersion
S4	S4	Standard	-
S4	S4A	+2%	White Dispersion
S4	S4B	+4%	White Dispersion
S4	S4C	+2%	Red Dispersion
S4	S4D	+4%	Red Dispersion
S4	S4E	+36%	Blue Dispersion
S4	S4F	+91%	Blue Dispersion
S4	S4G	+13%	Yellow Dispersion
S4	S4H	+27%	Yellow Dispersion
S4	S4I	+30%	Yellow Dispersion

Table 5. Sample S5: Pigment Modifications

Base	Variation	% Change Key Dispersion	Key Dispersion
S5	S5	Standard	-
S5	S5A	+769%	Black Dispersion
S5	S5B	+3%	Blue Dispersion
S5	S5C	+2%	White Dispersion
S5	S5D	+5%	White Dispersion
S5	S5E	+385%, +2% and +5%	Black, Blue and White Dispersion
S5	S5F	+2%	Red Dispersion
S5	S5G	+5%	Red Dispersion
S5	S5H	>3800%	Yellow Dispersion
S5	S5I	+15%	All Dispersions
S5	S5J	+769%, +3% and +10%	Black, Blue and White Dispersion

Table 6. Sample S6: Pigment Modifications

Base	Variation	% Change Key Dispersion	Key Dispersion
S6	S6	Standard	-
S6	S6A	+29%	Black Dispersion
S6	S6B	+10%	Blue Dispersion
S6	S6C	+6%	White Dispersion
S6	S6D	+10%	White Dispersion
S6	S6E	+14%	Brown Dispersion
S6	S6F	+21%	Red Dispersion
S6	S6G	+2%	Red Dispersion
S6	S6H	+5% and +2%	Blue and Red Dispersion
S6	S6I	+10%	All Dispersions
S6	S6J	+10%	Red Dispersion

Table 7. Sample S7: Pigment Modifications

Base	Variation	% Change Key Dispersion	Key Dispersion
S7	S7	Standard	-
S7	S7A	+769%	Blue Dispersion
S7	S7B	+3%	White Dispersion
S7	S7C	+4%	Brown Dispersion
S7	S7D	+2% and +2%	White and Brown Dispersion
S7	S7E	+9%	Red Dispersion
S7	S7F	+2% and +6%	White and Red Dispersion
S7	S7G	+9% and >2300%	Red and Yellow Dispersion
S7	S7H	+357%	Yellow Dispersion

S7	S7I	+15%	All Dispersions
S7	S7J	+30%	Red Dispersion

Table 8. Sample S8: Pigment Modifications

Base	Variation	% Change Key Dispersion	Key Dispersion
S8	S8	Standard	-
S8	S8A	+14%	Red Dispersion
S8	S8B	+24%	Red Dispersion
S8	S8C	+9%	Brown Dispersion
S8	S8D	+17% and +1%	Brown and Red Dispersion
S8	S8E	+42%	Blue Dispersion
S8	S8F	+8% and +42%	Black and Blue Dispersion
S8	S8G	+2%	White Dispersion
S8	S8H	+4%	White Dispersion

Table 9. Sample S9: Pigment Modifications

Base	Variation	% Change Key Dispersion	Key Dispersion
S9	S9	Standard	-
S9	S9A	+2%	Red Dispersion
S9	S9B	+5%	Red Dispersion
S9	S9C	+7%	Red Dispersion
S9	S9D	>1500%	Blue Dispersion,
S9	S9E	>1500% and 7%	Blue and Red Dispersion
S9	S9F	'>3800%	Yellow Dispersion
S9	S9G	'>7600%	Yellow Dispersion
S9	S9H	'>3800% and +2%	Yellow and Red Dispersion
S9	S9I	+10%	All Dispersions
S9	S9J	'>7600%, >1500% and +9%	Yellow , Blue and Red Dispersion

2.2 Instrumental Color Measurement

The color of all samples (including variations and standards) was measured in triplicate using the DigiEye® colorimeter (with DigiProduction v3.2.4.6 software) and the CIEDE2000 color space to quantify color differences. To minimize the influence of the substrate, each sample was applied four times onto calibrated white paper, allowing for the evaluation of the applied lipstick color and not the color of the stick itself. The DigiProduction software allowed for the exclusion of the white background, ensuring that only the applied color was considered in the quantification of the color difference.

2.3 Sensory Evaluation

2.3.1 Participants

A sensory panel composed of 600 lipstick users—including men, women, and individuals with different skin tones—evaluated all samples previously analyzed by instrumental measurement.

2.3.2 Procedure

A discriminative sensory analysis experimental design, specifically a difference-from-control test, was employed using a 5-point categorical scale with complete randomization. In dry booths with controlled lighting and temperature, each participant:

1. Received detailed instructions on the procedure and the evaluation scale.
2. Applied the standard sample to the forearm with four uniform strokes.
3. Applied the remaining samples of the same shade, following the predetermined randomization, side-by-side with the standard sample, also with four strokes. An additional, unidentified standard sample was included between the samples for control.

4. Evaluated the color difference of each sample relative to the standard sample, using the following 5-point hedonic scale: 4: Extremely different from the standard; 3: Very different from the standard; 2: Moderately different from the standard; 1: Slightly different from the standard; 0: Identical to the standard.

2.4 Data Analysis

The data obtained from the sensory evaluation and instrumental color measurement were analyzed using ANOVA and Tukey’s test. The correlation between the two variables was performed using box plots and means. The objective of the statistical analysis was to determine which color variations were perceived by the participants and correlate them with the instrumental data.

2.5 Ethical Considerations

Participation in the study was voluntary. Participants were informed about the nature of the research and signed an informed consent form. The study followed ethical principles for research involving human subjects, as approved by an ethics committee.

3. Results

3.1 Instrumental Analysis

The objective of the instrumental color analysis was to determine the color difference (ΔE) between the samples, using the CIEDE2000 system. Figure 1 presents the distribution of the different lipsticks analyzed in the color space (a) as well as the standards and the variations of the chosen samples for the consumers test (b).

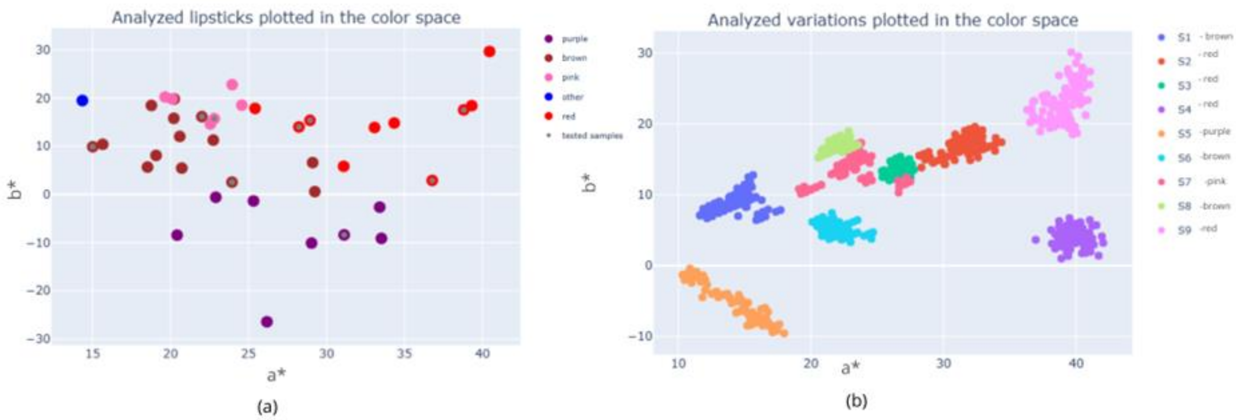


Figure 1. Distribution of samples in the color space (a) and standards and variations analyzed in the sensorial evaluation distributed in the color space (b).

Figure 2 presents the variability, measured by Delta E (ΔE), for the standard color of each shade analyzed (S1 to S9). It is observed that the ΔE values varied around less than one unit for S1, S5, S6, S7 and S8, and around 1 to 1.5 units for other samples, indicating the intrinsic variation within the production batch itself. Samples 1, 8 and 6 curiously belong to the same chromatic group in this study denominated as the “browns” group, and the same can be observed for the group named as “reds” containing the S2, S3, S4 and S9 samples.

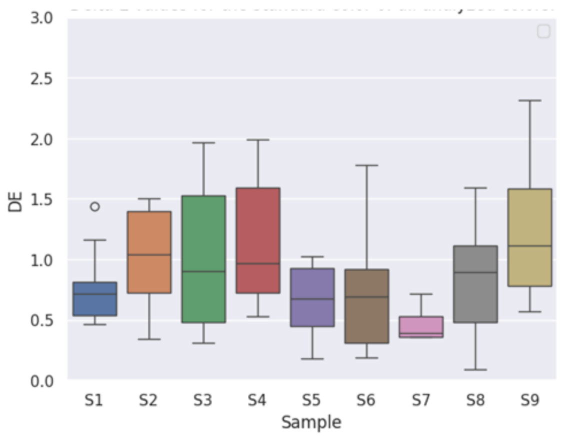
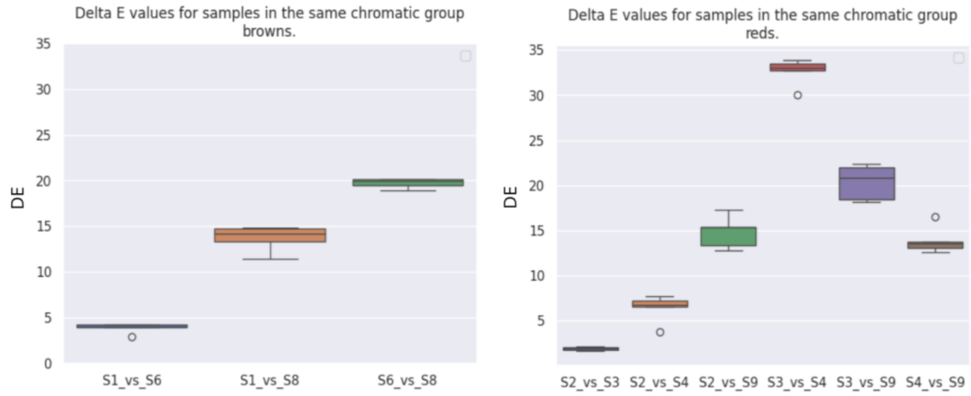


Figure 2. Distribution of Delta E (ΔE) among lipsticks from the same batch.

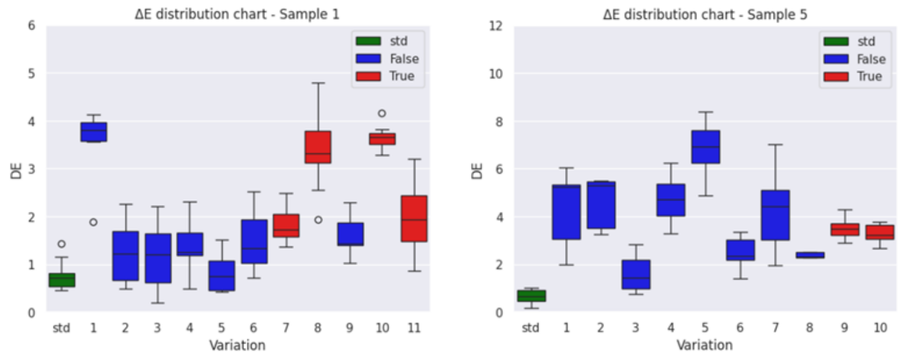
Figure 3 illustrates the color differences (ΔE) between the standards of two main groups of shades analyzed denominated as browns (A) and reds (B). The intra-group comparison is valuable by hinting thresholds for each group and demonstrating from which Delta E value a color can be considered a distinct one. This comparison was not evaluated in the other chromatic groups as they only contained one analysed sample each.



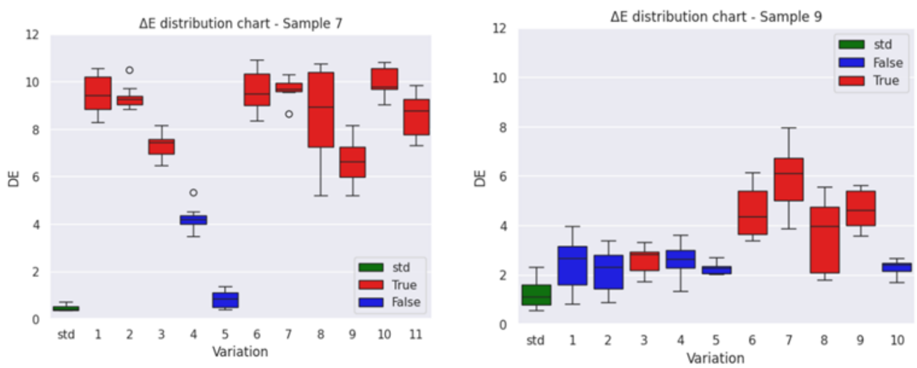
(a) (b)

Figure 3. Distribution of Delta E (ΔE) values among the different colors analyzed in the same chromatic groups named as browns (a) and reds (b).

Figure 4 presents the delta E values for the variations created for colors S1, S5, S7 and S9, respectively. The variations that were frequently identified by the research participants as being different from the standard color are represented in red, while in blue are the ones considered the most similar according to the perceptions from the lipstick users. Results in this format were not presented for other samples due to lack of statistical significance.



(a) (b)

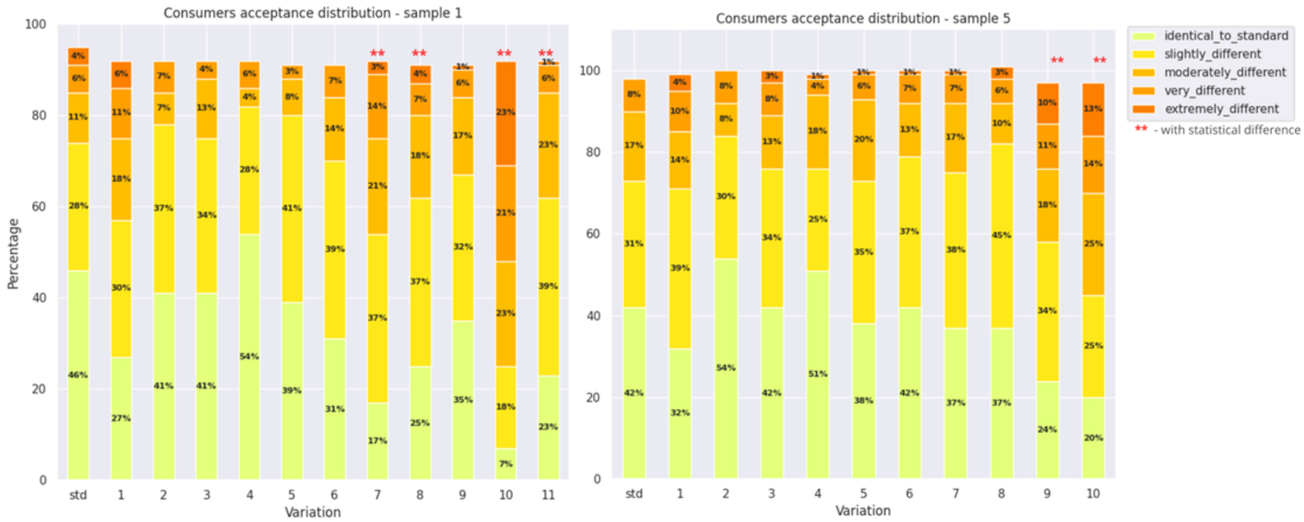


(c) (d)

Figure 4. Distribution of Delta E (ΔE) values for the different variations in relation to the standard in Sample 1 (a), Sample 5 (b), Sample 7 (c) and Sample 9 (d).

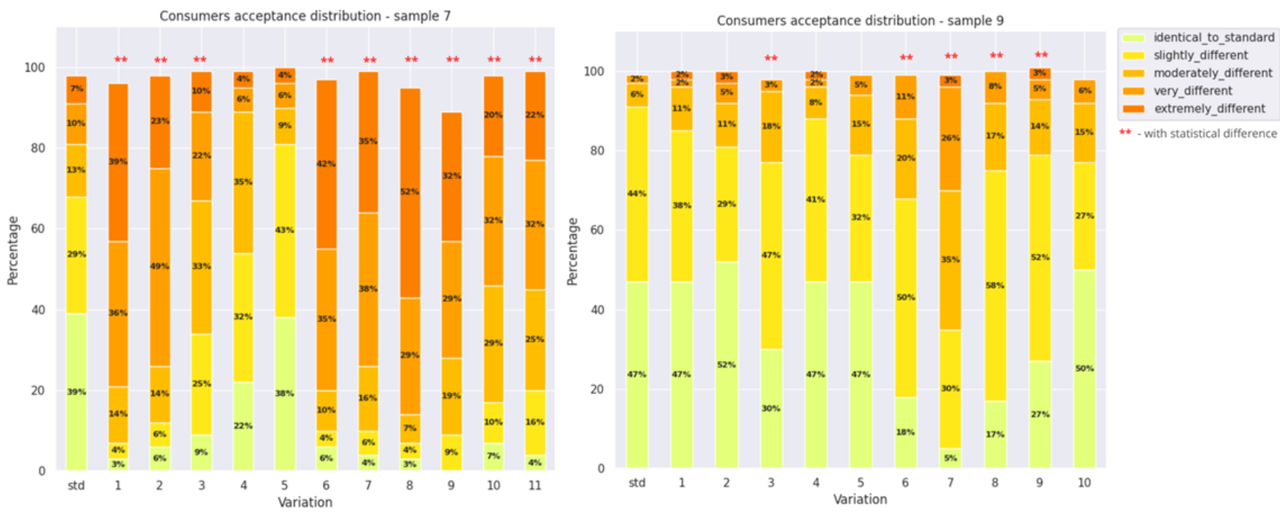
3.2 Sensorial Analysis

Figures 5 and 6 are demonstrating the distribution of answers given by the participants of the sensorial evaluation. Results in this format were not presented for other samples due to lack of statistical significance.



(a) (b)

Figure 5. Consumer sensorial identification for the different variations in relation to the standard in Sample 1 (a) and Sample 5 (b).



(a) (b)

Figure 6. Consumer sensorial identification for the different variations in relation to the standard in Sample 7 (a) and Sample 9 (b).

Table 10. Defined delta E thresholds and max acceptance for each sample.

Sample	Chromatic group	ΔE threshold	max ΔE acceptance
S1	brown	2	4
S2	red	4	5,6
S3	red	3	3
S4	red	4	4,2
S5	purple	4,8	6,8
S6	brown	3	4
S7	pink	2,5	4
S8	brown	3,5	4
S9	red	3,2	4

4. Discussion

To precisely quantify color differences, considering human visual perception, the CIEDE2000 formula (ΔE) was developed. This study investigates the applicability of this formula in the cosmetics industry, aiming to determine the perception threshold of different lipstick shades by consumers, through the correlation between instrumental analysis (CIEDE2000), interactive clustering, and sensory evaluation [1, 18].

The 40 lipstick shades studied underwent K-means clustering analysis, a technique that groups data based on its characteristics [14]. The K-means algorithm aims to form groups with high internal similarity and high dissimilarity between groups, using Euclidean distance to minimize the sum of squared distances between points and their respective centroids [4, 16]. This analysis identified four main color groups (pinks, purples, reds, and browns), whose centroids (midpoints) were defined as standards for creating 9 to 12 variations of each group, subsequently tested (Figure 1A).

Figure 2 illustrates the distribution of Delta E (ΔE) among lipsticks from the same batch, showing that the lowest variability was observed in samples S1, S5, S6, S7, and S8. Even so, these values are higher than the visual perception threshold of 1, frequently used as a parameter to indicate the absence of a perceptible color difference to an average observer [14]. This discrepancy suggests that adopting a ΔE below 1 as a criterion for detecting color differences, considering an average observer, may be inadequate. Samples from the red chromatic group (S1, S3, S4, and S9) showed Delta E values between 1 and 1.5 units, indicating a possibly inherently higher variability in this group of shades. This reinforces the need to review and possibly expand the ΔE limits, considering the specificities of the cosmetics industry and the tolerances for each color, the central objective of this work.

To define maximum color difference limits, we analyzed the ΔE distribution for the evaluated colors (Figure 3). The objective was to determine which ΔE values allow distinguishing two colors, considering the divergence in the literature regarding perception thresholds: Cabezas (2018) suggests values between 2 and 4 [2], while Mokrzycki et al. (2011) propose a narrower range, between 2 and 3.5 [14]. Analysis of the ΔE distribution (Figure 3B) indicates greater similarity between colors S2 and S3 (values between 2 and 2.5), suggesting possible overlap of their color spaces, as evidenced by the variations in Figure 1B.

In contrast, the colors in Figure 3A exhibit distinct behavior, even within the same chromatic group (e.g., S6 shows a ΔE of up to 5 compared to other colors). This distinction, also observed in Figure 3B, reinforces the hypothesis that ΔE values between 5 and 6 can define a maximum limit for the instrumental specification of colors, corroborating the observation of Mokrzycki et al. (2011) that ΔE values above 5 are frequently perceived as distinct colors [14].

Regarding sensory evaluations, samples S2, S3, S4, S6, and S8 showed no significant differences (all variations were considered equal to the standard), with the acceptable ΔE limit defined by the third quartile of the distribution of their ΔE values. Although we have maximum ΔE limits for each color cluster (4 for browns/pinks; 3-5.6 for reds; 6.8 for purples), for these samples, the average ΔE between variations and standard (Table 10) - 4, 3, 4, 3, and 3.5, respectively — is the most accurate calculation method. These results corroborate the findings of Khashayar et al. (2014) [7, 8], which considered variations up to 3.5 acceptable by patients in dental restorations, and the data from Mokrzycki et al. (2011) [14], which recorded a perception range between 2 and 3.5 in experienced observers.

For samples with accepted and rejected variations (Figures 5 and 6), the limit was defined by the median of the ΔE values of the accepted variations. Therefore, for S1, S7, and S9, a maximum acceptable limit of ΔE=4 was defined, while for sample S5 the limit was 6.8 (Table 10). Considering that for some evaluated color clusters there is more than one sample, for these cases, the average of the results was used to obtain the actual value of the color specification limit for the evaluated group. In this scenario, the average limit for clusters such as reds and browns is 4.2 and 4, respectively. With these findings, it can be observed that none of the evaluated clusters presented ΔE values below 2, demonstrating that the evaluated samples are indistinguishable in the skin.

Analysis of color S9 (Figure 6B) revealed interesting results. Variations 2 and 10 were perceived as closer to the standard, unlike variations 6 to 9, considered more divergent. Notably, even variations 6 to 9, with ΔE close to 6 (Figure 4D), were classified as “equal to the standard” by 17% and as “slightly different” by 47% of the participants, on average. This demonstrates that, despite significant color differences, a considerable portion of the population did not perceive drastic changes. The greater divergence in variations 6 to 9 was mainly due to the alteration in the yellow pigment, corroborating the findings of Na et al. (2013) on the lower tolerance to color variations in white and yellow tones [18].

In the sensory test of Sample 7 (Figure 6A), it was observed that most variations were considered “extremely different” or “very different” compared to the standard, although the changes in the formula were similar or smaller in percentage compared to the other colors studied. This result suggests that the acceptability limit for each color is dependent on its chromatic group, since, for this sample, a ΔE threshold of 2.5 was established, demonstrating that, for the pink color, the limit acceptable to the consumer is lower compared to the other color groups studied. Furthermore, these results corroborate the idea that the pink color, formed by white and red pigments, has lower tolerance, according to the study by Na et al. (2013), which indicates that the white pigment has the lowest acceptable tolerance among the colors studied [18].

5. Conclusion

This study investigated the applicability of the CIEDE2000 formula (ΔE) for determining the color perception threshold in lipsticks, combining instrumental analysis, K-means clustering, and sensory evaluation. Analysis of 40 lipstick shades, clustered into four main categories (pink, purple, red, and brown), revealed that the visually perceived ΔE < 1 threshold, frequently cited in the literature, is narrow and irrealistic for the cosmetics industry. Analysis of the ΔE distribution demonstrated significant variations between color groups, with red shades exhibiting higher intrinsic variability.

Analysis of the ΔE distribution in nine lipstick colors revealed important nuances in the perception of different colors within the distinct chromatic groups. Consumer tests indicated that ΔE values up to 3 were frequently perceived as identical, suggesting an acceptable limit for

color specification; however, in colors close to the neutral axis of the color space, smaller values could be identified by consumers. Generally, values between 3 and 4 represent a “gray zone” where color distinction becomes less clear. Above 4, the color difference was generally identified by consumers. Additionally, ΔE values up to 2 proved undetectable on the skin, highlighting the importance of colorimetric analysis in cosmetics.
The study proposes new guidelines for the industry, emphasizing the need to consider the diversity in color perception and the definition of specific ΔE limits for each chromatic group, considering pigment composition and the impact on perception by the skin. While the CIEDE2000 formula is useful, its application requires careful adjustment for each color group, as it indicates distance but not directional differences. Thus, the research demonstrates that even small ΔE values can generate perceptible discrepancies, justifying further research to improve colorimetric analysis in the cosmetics industry.

References

1. Baah KFM (2017) Defining Acceptable Colour Tolerances For Identity Branding In Natural Viewing Conditions. University Of The Arts London For The Degree Of Phd, 16-52.
2. Cabezas MJN (2018) Estudio Piloto Sobre La Influencia De La Luz Ambiental En La Toma Del Color Gingival. Universidad Complutense De Madrid - Facultad De Odontología, 60-70.
3. Chu SJ, Trushkowsky RD, Paravina RD (2010) Dental Color Matching Instruments And Systems . Review Of Clinical And Research Aspects. 2–16.
4. Everitt, B. S., Landau, S., and Leese, M., Cluster 55 Analysis, Arnold, 4th Edition, 2001.
5. Hofer H, Carroll J, Neitz J, Neitz M, Williams DR (2005) Organization Of The Human Trichromatic Cone Mosaic. Journal Neuroscience.
6. Johnston WM (2009) Color Measurement In Dentistry. Journal Of Dentistry 37, 2-6.
7. Kawakami DY (2018) Avaliação Da Perceptibilidade E Aceitabilidade Da Diferença De Cor Em Restaurações Estéticas Odontológicas. Universidade Estadual De Campinas - Faculdade De Odontologia De Piracicaba, 5-30.
8. Khashayar G, Bain PA, Salari S, Dozic A, Kleverlaan CJ, Feilzer AJ. Perceptibility and acceptability thresholds for colour diferences in dentistry. J Dent. 2014;42(6):637-644.
9. Kim A, Kim H, Park S (2011) Measuring Of The Perceptibility And Acceptability In Various Color Quality Measures. Journal Of The Optical Society Of Korea, 310-317.
10. Linden, Ricardo. Técnicas de Agrupamento - Revista de Sistemas de Informação da FSMA n. 4 (2009) pp. 18-36
11. Lloyd, S., Least squares quantization in PCM, IEEE transactions on information theory 28.2: 129-137, 1982.
12. Luo MR, Cui G, Rigg B (2000) The Development Of The Cie 2000 Colour-Difference Formula: Ciede2000. Colour & Imaging Institute,340-350.
13. Mohebbi B (2014) The Art Of Packaging: An Investigation Into The Role Of Color In Packaging, Marketing, And Branding. International Journal Of Organizational Leadership, 92-102.
14. Mokrzycki WS, Tatol M (2011) Colour Difference ∆E - A Survey. Faculty Of Mathematics And Informatics - University Of Warmia And Mazury, 1-25.
15. Na, Nooree, Kyungah Choi, Jeongmin Lee, and Hyeon-Jeong Suk. Color Tolerance Study on White in Practical Aspect: Perceptibility Versus Acceptability. Color Research & Application, vol. 39, no. 6, 2013, pp. 582-88.
16. Oliveira DCRS (2018) Esthetics Of Dental Composites. In: Miletic V. Dental Composite Materials For Direct Restorations. Switzerland: Springer Nature, 154-76.
17. Tan, P.-N., Steinbach, M., and Kumar, V., Introduction to Data Mining, AddisonWesley, 2006.
18. Wrolstad RE, Smith DE (2017) Color Analysis. Department Of Food Science And Technology, 545-554.