Statistical Analysis for the Development of a Mathematical Model for Liquid Phase Color Reproduction

Wilson Rodríguez Rivera
Manufacturing Competitiveness
Rafael Nieves, Pharm D.
Industrial Engineering Department
Polytechnic University of Puerto Rico

Abstract — *This study aims to develop a mathematical model, capable to give a receipt of the dies concentrations to be used to develop a desire color. The model helps to minimize, and in some cases eliminate, the trial and error technique. This study investigates the effect of the color coordinates vector such as L, a and b in order to evaluate if there exist a correlation by using the FD&C dies such as FD&C Yellow 5 FD&C Yellow 6, FD&C Blue 1 and FD&C Red 40. A 2k factorial experiment was performed, to identify those dies that have a significant effect within the color coordinates. Results obtained revealed that of the four FD&C studied, three of them have significant effect within the color coordinates. The FD&C dies used, showed a significant effect when is used alone or when are mixed. A correlation of the color coordinates and the factors was found. Three linear formulas were obtained from the* ***L****,* ***a,*** *and* ***b*** *vectors. Those formulas were introduced in a matrix model to developing or matching colors.*

Key Terms *—Color space (L, a and B)[4], Lightness (L), a (position of green red axis), b (position of blue-yellow axis).*

BACKGROUND

The color reproduction is one the primary challenges that confront many industries were their final colored product play an important role. Those industries such as, Food, Drug, Cosmetics, and Printing industries, are always looking forward to improve or create colors were the result is to be attractive to the customer experience [1]. To create colored consumable products require using colorants or dies approved to the human consumption. Actually some of the consumable dies approved by the FDA are used in this project. They are FD&C Yellow 5, FD&C Yellow 6, FD&C Red 40 and FD&C Blue 1. Normally, the industries have a development group oriented to create new color. For example a development group intent to develop a specific green color to be used in an energetic drink. The development groups perform a series of experiment by mix different dies concentration to reach the color. Other groups enter to the trial and error practice. Those techniques are common at the time to develop the desired color.

Other issue when a color is developed is the color evaluation. Actually many industries use the human eyes as a quality auditor process, to determine if the color reach the customer expectative. There exist several studies of the human eyes limitations at the time to evaluate colors. Also the environment plays an important role at the time to the evaluation process. The health of the eyes as an important factor and may vary between persons. The Figure 1 [6] shows the factors that could affect the color evaluation.

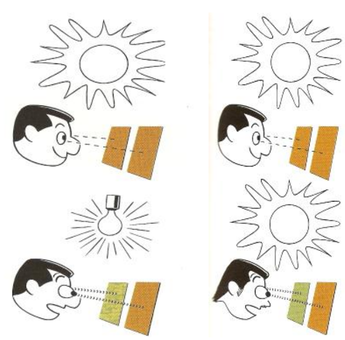

Figure 1
Color Evaluation

Munsell was one of the pioneers to study color, back in 1905. His goal was to reorganize the colors in numerical system and physical exemplification achieved via the Atlas of Munsell. He created a tridimensional color space where it separates by color as hue (change of color), lightness and chroma. See Figure 2 [6] for Munsell arrangement.

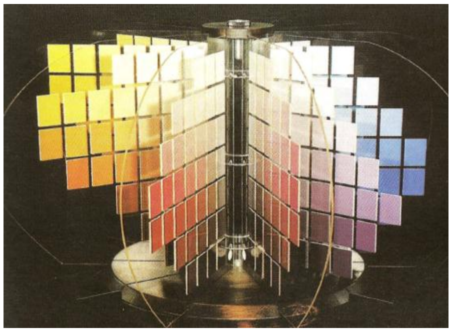

Figure 2
Munsell Color tree

Later, a committee was created to study and standardize the color interpretation, because the color interpretation was ambiguous and has different manner to be interpretive. This commission was established in 1931 and based in Vienna, Austria and named as the International Commission on Illumination (usually known as the CIE for its French name) their mission was to standardize all the themes and how could be manage the term colors. In 1973 the CIE standardize the term color as L, a, and b. This axis or coordinates are represented in the Figure 3 [6].

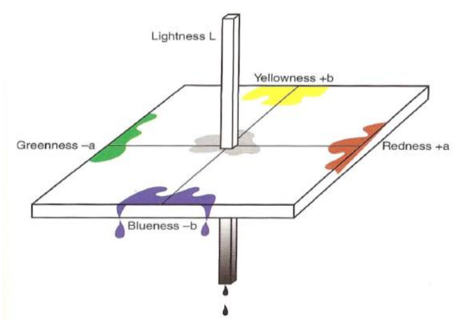

Figure 3
Color Coordinates

This study helps to the industries to develop or match color by using a mathematical model or matrix arrangement. This matrix has the capabilities to give the concentrations of each dies to be used at the time, to create or match color. By using this mathematical model the development groups could eliminate or minimize the trial and error technique. The trial and error practice and the creation of huge experiments by combining different color cause a lot cost to the Research and Development (R&D) department.

METHODOLOGY

Analytical techniques and statistical tools were implemented to identify the dies to be studied and been selected. The experimental plan was established as follow:

- Familiarize with the most popular FD&C dies used in the market of human consumption
- Evaluate concentrations to be studied, by study concentration of inks and juice used
- Experimental Plan and Strategy
- Statistical Analysis of the factors
- Mathematical model creation

The analysis performed shows that the dies to be studied are the FD&C Yellow 5, FD&C Yellow 6, FD&C Blue 1 and FD&C Red 40. The common concentrations in human consumption market (juice drinks) goes through 0.01 M to 0.1M. For this reason during the DoE the range to be evaluated will go through 0M (no presence of die) to 0.2 M. The word M, means Molar or concentration.

M = grams solute/ 1 Liter of solution (1)

The vehicle or solvent to be utilized is water. The water density is near of 1mg/1L. By take this in consideration the volume of the equation (1) can be changed by

M= grams solute/ 1000grams of solution (2)

The UV-vis spectrophotometer plays an important role during the experiment analysis. This instrument use a color calculation software, this software give the capabilities to get the values of the color coordinates (***L, a,*** and ***b***). Also this instrument in conjunction with the mathematical model will be used to get the color coordinates values of the samples and then be extrapolating through the formula in order to develop the desire color.

The design of experiment was a 2k factorial [2]. There four factors to be evaluated. The four factors studied were the FD&C dies (FD&C Blue 1, FD&C Red 40, FD&C Yellow 5 and FD&C Yellow 6). A full factorial design experiments (DOE) was developed in order to study the colorant interactions within the color coordinates [2]. The setting of the DoE is shown on Table 1. A total of twenty two (22) experimental trials are required to challenge all the factors.

Table 1
Design of Experiment (DOE)

Full Factorial Design		
Factors: 4	Base Design 4,16	Resolution with blocks: V
Runs: 22	Replicates 1	
Blocks: 2	Center Points (total) 6	

The full factorial design perform all the available combination between the factors. The Table 2 represent some of the runs to be performed. In the DoE the -1 value represent zero concentration or 0M, the value 0 represent a concentration of 0.1M this number represent the center points of the design. The number 1 value in the DoE represent 0.2 M. Table 2 is the representation of the available dies to be studied and the color obtained by their mixture.

Table 2
Representative Experiment design

Full Factorial Design							
StdOrder	RunOrder	CenterPt	Blocks	YELLOW 5	YELLOW 6	BLUE 1	RED 40
22	1	0	2	0	0	0	0
14	2	1	2	1	-1	1	-1
18	3	1	2	-1	-1	1	1
21	4	0	2	0	0	0	0
17	5	1	2	-1	1	-1	1

The Figure 4 is a representation of the primary colors used, such as Blue, Red and Yellow and a representative color when a mixture of the dies are performed. The right section of Figure 4 shows a representation of the sample preparation by using a stock solution and the different levels studied in the DoE.

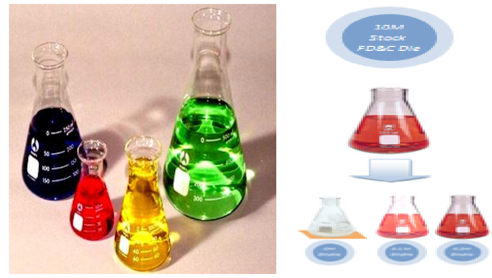

Figure 4
Representative color preparation

RESULTS AND DISCUSSION

After run the DoE, all the factors were evaluate within the color coordinates (***L***, ***a***, and ***b***). The factors concentrations will be evaluated up to 0.2M. By using a higher concentration, this study gets the flexibility to develop colors in other products such as candy, medicine, food, and inks application that could be covered in this design.

During the colorant evaluation a Pareto chart [8] for Lightness was developed. Figure 5, shows that the dies or factors that affect the lightness significantly. Can be observed that the FD&C Blue 1, is the primary factor that affects dramatically the lightness, in subsequence with FD&C Red 40 and FD&C Yellow 6. Other combinations affect the lightness too, like factors or mixture identify as BC, BD amount others, but the main factor that affects the lightness showed in the Pareto was the Blue 1 or factor C. Also can observe that the only mixture of the factor BCD, does not have any effect on the Lightness. The FD&C yellow 6 was considered to be taken out of the regression equation. This was concluded by the lightness behavior obtained when the FD&C Yellow 6 is mixture with FD&C Blue 1 and FD&C Red 40.

The Pareto chart result for ***a*** vector shows that the FD&C Blue 1 is the primary die statistical significant that affects this vector. The FD&C Red 40 and the FD&C Yellow 5 are other factors that affect the***a*** vector results from experimental analysis. The mixture of the factors named as BC and CD in the Figure 6 showed that those dies when combined together does not affect the ***a*** vector.

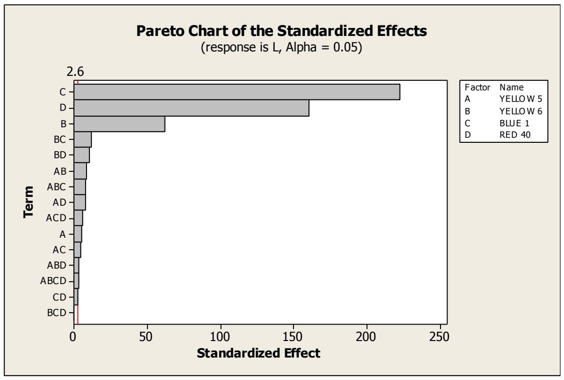

Figure 5
Pareto chart for Lightness

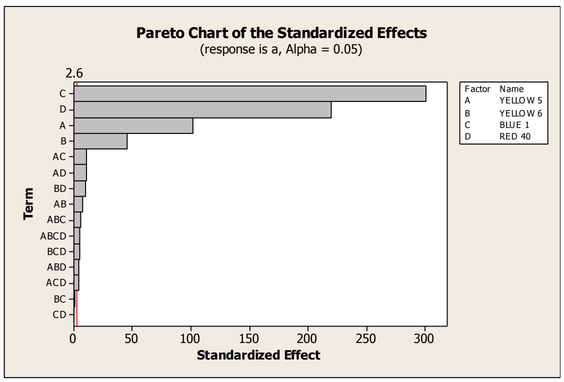

Figure 6
Pareto chart for a (greenness -a or redness +a)

The ***b*** vector showed in Figure 7 demonstrates that the FD&C Yellow 5 is statistically significant to affect the ***b*** vector, also follow by Blue 1 and FD&C Yellow 6. Also others colorants and mixture affect this vector.

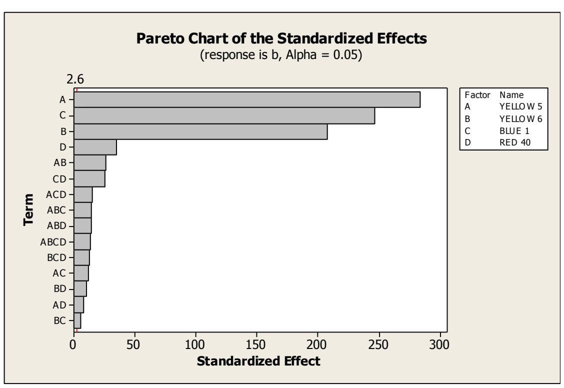

Figure 7
Pareto chart for b (blueness and yellowness)

Three dies were selected in order to study if the correlation of the factors within the color coordinates vectors. The selected dies were FD&C Yellow 5, FD&C Red 40 and FD&C Blue 1. FD&C Yellow 6 was eliminated because this die does not affect the Lightness when is combined with FD&C Blue 1 and FD&C Red 40. The three factors selected show a significant effect within the color coordinates. After confirmed this significantly effect, an analysis of variance was performed in order to evaluate a possible linear correlation. During the ANOVA analysis the main effect, 2 way interaction, 3 way interaction and 4 way interaction were included in the evaluation. The p value obtained for each color coordinate (***L***, ***a***, and ***b***) is less than alpha ($p_{value} < \alpha$)**.** The statistical value obtained indicates the possibility of a linear behavior. See Figure 8, Figure 9 and Figure 10.

The linear regression analysis for lightness, see Figure 8, show a linear correlation. The p value obtained in the regression analysis confirms a statistical linear model for lightness vector. Also the R-sq (adj) only has 10% of adjustment. Otherwise the lack of fit observation is significant in the regression analysis. This can be attributed to the FD&C Yellow 5, this die does not have a linear behavior were the p value obtained ($p_{value} > \alpha$) (0.759 > 0.05). The value obtained by this factor will be evaluated in the conclusion section.

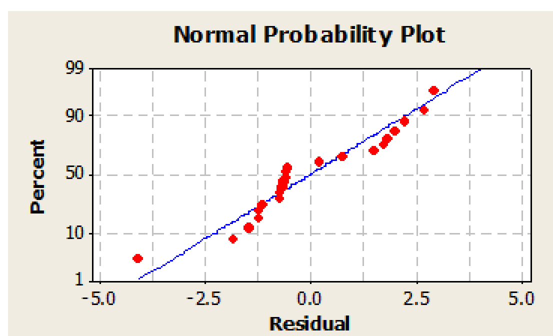

Figure 8
Linear regression for Lightness

The regression analysis for **a** vector can be seen in Figure 9. The linear regression analysis in the ANOVA model is significant. Also the R-sq (adj) only has 3.9% of adjustment. Otherwise the lack of fit observation is significant in the regression analysis. This could be attributed to the p value obtained in the constant section were the linear correlation is not significant. The p value obtained in the constant section in the analysis of variance, record a $p_{value} > \alpha$ (0.057 > 0.05). This factor could add variability in the regression equation.

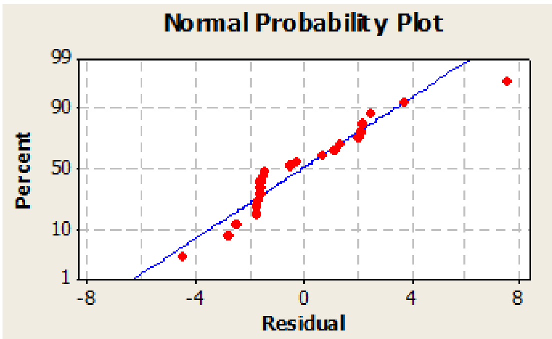

Figure 9
Linear regression for a vector

The graph obtained in the regression analysis for the vector ***b***, see Figure 10 shows a lack of linear behavior. However in the analysis of the variance section, the regression analysis show a significant linear correlation between the factors by obtained a ($p_{value} < \alpha$). The Figure 10 (***a*** vector) shows a linearity issue within the factors when is plotted. This issue also can be attributed to the p value obtained for the Factor FD&C Red 40. The p value obtained for this factor was $p_{value} > \alpha$ (0.519 > 0.05). In addition the R-sq (adj) shows a 71.5%. This means that regression is not completely linear. The lack of a linear behavior for ***b*** vector within the factor could be truth.

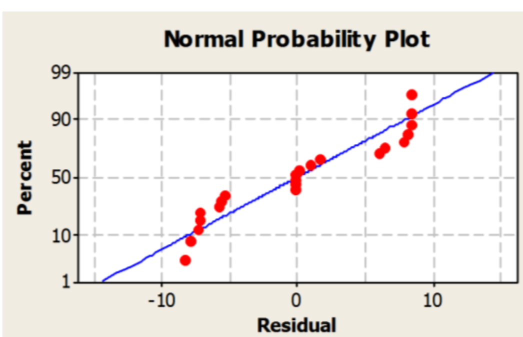

Figure 10
Linear regression for b vector

The three factors evaluated in the regression analysis show a linear behavior within each color coordinates, except for the ***b*** vector a lack of linear behavior is observed. The linear formula of the ***b*** vector was included in the matrix arrangement. This option affects the precision of the results obtained in the matrix arrangement (confirmed by the statistical results of the ANOVA obtained for the ***b*** vector). After this presumption, a corroboration of the mathematical model obtained was done. The importance to reach a mathematical model is that the model offers a guideline to the R&D department on how reproduce or match the desire color without the needs to calibrate a UV-vis instruments. By performing a dilution 1:50 of the sample and reading the color coordinates by using a spectrophotometer, provide all the information needed to be extrapolated in a matrix formula, in order to get the concentration needed of each die to match the color. The time to develop or match a desired color reduces dramatically and also the trial and error technique could be minimized.

Three linear regressions were obtained for each color coordinate. Each regression was multi-factorial and the factors were selected in the statistical analysis. By using the multi-factorial analysis and the three regression obtained, a matrix arrangement was developed to determine the concentration needed from each die, at the time to prepare a desired sample.

The following Table 2 shows the linear equation arrangement.

Table 2
Color Coordinates Linear equation

L = 97.7 - 1.46 YELLOW 5 - 53.7 BLUE 1 - 36.0 RED 40 a = 2.78 - 43.7 YELLOW 5 - 129 BLUE 1 + 88.9 RED 40 b = 7.96 + 93.8 YELLOW 5 - 82.0 BLUE 1 + 10.7 RED 40

By using the above formulas and the creation of a matrix arrangement for these three formulas, is a helpful tool to get the concentration of each die. To get the concentration of those dies is necessary to know the (***L***, ***a*** and ***b***) values of the sample. This value could be obtained by using a UV-vis spectrophotometer with a color calculation option. By getting the (***L***, ***a*** and ***b***) values, those values will be substitute to the matrix arrangement in order to obtain the concentration of each die to create the desired color. To get the concentration the color coordinates values will be substituted to be resolved in a 3x3 matrix. See the below Table 3 for the detailed matrix arrangement [5]. During this project, matrix software (MathCAD 14) was used to avoid manual calculation and minimize typing error to study the values obtained and analyze the matrix capabilities.

Table 3
Matrix arrangement for Linear equation

	Matrix		
Constants	**Y5**	**B1**	**R40**
L-97	**-1.46**	**-53.7**	**-36.0**
a-2.78	**-43.7**	**-129**	**88.9**
b-7.96	**93.8**	**-82**	**10.7**

Three samples were prepare with different concentrations and dies combination and then were treated to be read in the UV- Spectrophotometer. Table 4 shows a comparison of the values obtained in the mathematical model and the real concentration used of each sample.

Table 4
Samples vs. mathematical model values

		Dies Concentration		
Sample	**Model**	**R 40**	**B1**	**Y 5**
Sample 1	**Sample**	**0.0200**	**0**	**0.161**
	Ma. Model	**-0.034**	**0.0003**	**0.104**
Sample 2	**Sample**	**0**	**0.0299**	**0.169**
	Ma. Model	**-0.064**	**0.03**	**0.117**
Sample 2	**Sample**	**0.150**	**0.059**	**0.039**
	Ma. Model	**0.098**	**0.073**	**-0.031**

As it can be seen, the mathematical model is not reproducible. This is due to the lack of linear behavior of the color coordinates. Otherwise the mathematical model gives a better idea or a good reference of how the color will be prepared. For example the characteristic of the sample 1 is yellowness color with a very low concentration of FD&C Red 40. The model confirms a yellowness color by give a higher concentration to FD&C Yellow 5. The negatives values near to zero obtained in the mathematical model, will be attributed at zero concentration or near of them. The same behavior could be observed in the sample number 2 and 3. The mathematical model shows a good reference of the dies concentration to be used to reproduce a color.

CONCLUSIONS & RECOMMENDATIONS

The three factors evaluated were FD&C Yellow 5, FD&C Blue 1 and FD&C Red 40. The mathematical model show a drift response within the concentration obtained of each factor evaluated and is included in the matrix formula. This drift is attributed to the lack of linearity of the **b** vector. Otherwise ***b*** vector is included in the matrix within ***L*** and ***a*** vector. This drift is not severe but can cause methamerism issue explained in the background section. This drift could be acceptable for the customer. However by introducing ***b*** vector within the mathematical model the drift is not severe when a visual inspection occurs. Also the reading of the color coordinates in the UV-vis Spectrophotometer [3] validate that the statistical analysis and the mathematical model used eliminate the needs to create calibration curves at the time to develop a desire color.

Other advantage to use the mathematical model is that give a better idea or a reasonable start point at the time to prepare a specific color, without the needs to enter in a huge experimental analysis by mixing dies to reach the desired color This mathematical model could be useful tools to the R&D labs to match color by minimize experimental time and reduce cost of investigation. This mathematical model could be extrapolated to others industries where use the FD&C dies studied.

Further studies can be considered to evaluate other experimental studies [7] by using others statistical tool such as mixture analysis and surface analysis in order to correct the mathematical model by eliminate the drift issue.

REFERENCES

[1] Beatrice Trum Hunter (1999). *What are natural food colors,* 20-25. Retrieve August 2007, from EBSCO data base.

[2] Douglas C. Montgomery, George C. Runger, Norma Faris Hubele.(2006). Third Edition Engineering Statistics. United State of Americas: John Wiley and Sons, Inc.

[3] Ebbing D. Darrel, (1993). Fourth Edition General Chemistry. Houghton Mifflin Company.

[4] EBSCO data base, *When color maters, feb 15 2007*. (54-55),Retrieve August 2007.

[5] Hosteler P. Robert, Larson Ron, Edwards H. Bruce, (2001). Third Edition Precalculus a graphic approach. Houghton Mifflin Company.

[6] Roy S. Bearns. Principle of Color Technology. United State of Americas: John Wiley and Sons, Inc.

[7] Takayuki, Ogasahara . *Optimization of color dyes for spectra and colorimetric color reproduction*. (372-375).

[8] William Mendenhall, Robert J. Beaver, Barbara M. Beaver.(1999). Tenth Edition Introduction to Probability and Satistics. United States of Americas: ITP