

To foam or not to foam: A survival analysis of the foam head that forms when a soda is poured

Kate Phillips, Kennesaw State University

ABSTRACT

The goal of this study is to determine which factors influence the dissolve time of the foam head that forms after a soda is poured. This study proposes a logistic model in order to estimate a particular soda's probability of being a "small fizzer" (the desired outcome) as opposed to a "big fizzer," with the median dissolve time of 12 seconds serving as the cut point for the binary outcome. A standard procedure for testing foam head dissolve time was developed in order to collect the study data. A sample of 80 Coca-Cola products was then tested during fall 2013; characteristics of each product sampled were also recorded. All analyses were then conducted using SAS® 9.3 software. After conducting a univariate analysis for each factor of interest, the original continuous response variable was then dichotomized into the binary outcome of interest. A bivariate analysis was then conducted; odds ratios with their confidence intervals were examined in order to determine a predictor's significance with respect to the binary outcome. Table row percentages were examined for factors where odds ratios were not given by SAS 9.3. It was discovered that the most significant factors were sweetener type and a previously undiscovered (to the author's best knowledge) interaction between test container material and the presence/absence of caffeine ("test container material" refers to the material that the beverage was poured into for testing). According to the study results, this interaction was the most influential factor with respect to foam head dissolve time. The odds ratio for sweetener type was 2.25 (95% CI: 0.91, 5.54). With caffeine present, the odds ratio for test container material was 0.76 (95% CI: 0.23, 2.53). With caffeine absent, the odds ratio for test container material jumped to 11.70 (95% CI: 1.85, 74.19). The final logistic model retains the factors "sweetener type" and "test container material," as well as the interaction between test container material and the presence/absence of caffeine.

INTRODUCTION

To the avid soda drinker, it appears that different factors influence the length of time that it takes for a soda to go flat. These factors may include storage temperature, bottling size, and the addition of ice, among other things. However, when conducting a scientific analysis of which factors are most influential, the exact point of "flatness" of a soda is difficult both to observe and to quantify. A related phenomenon that can easily be observed is the foam head that forms after a soda is poured quickly. Since it is believed that foam head dissolve time is inversely related to the time that it takes a soda to go flat¹, a cohort study was conducted to investigate the factors that influence the dissolve time of the foam head. If factors can be found that are influential in decreasing foam head dissolve time, in theory, these factors can then be manipulated to increase the length of time that a soda stays fizzy.

In this study, which was conducted as a project for a master's level statistics course at Kennesaw State University, the following factors were investigated: place of storage, storage temperature (°F), bottling size (fl oz.), bottling material, flavor, sweetener type, test container size, test container material, the presence/absence of caffeine, the presence/absence of ice, days until expiration, and class. For this study, "class" refers to whether a particular item tested was Coke, Sprite, Fanta, Pibb Xtra, Mello Yello, or Barq's Root Beer (only Coca-Cola products were tested so that the study could be simplified and because the author is a Coke drinker). Also, "test container" refers to the container that a particular beverage unit was tested in. The null hypothesis of the study was that a particular factor is not related to foam head dissolve time, while the alternative hypothesis was that a particular factor is related to foam head dissolve time. Pilot data were collected to determine the sample size needed to have 90% power for detecting a minimum difference of 3 seconds between the mean dissolve times of the "ice present" and "ice absent" groups. (Exploration of the pilot data using PROC POWER, a procedure from the SAS/STAT® module within SAS 9.3, revealed that these groups would require the largest sample size to detect the desired minimum difference in their means at 90% power). Since the needed sample size was between 76 and 1312 units, a sample size of 80 units was chosen for collecting the main study data. Materials were randomized for pilot data collection and again for collecting the main study data; the methods of randomization are described in the following section.

¹ This is implied from the information on the following websites (cited in the REFERENCES section), as well as from the author's personal experience:

- http://www.wired.com/magazine/2011/04/st_equation_sodapop/
- <http://www.physics.org/food-physics/text-only/>

METHODS

For this investigation, Coca-Cola products were sampled from three local grocery stores. At the first store visited, one selling unit was purchased of every bottling size, bottling material, sweetener used, amount of caffeine, and flavor of every desired class of Coca-Cola product available. For items sold individually, two selling units were purchased. Items not available at the first store were purchased at the second store; items available at neither the first store nor the second store were purchased at the third store.

Bottling units were then unpackaged and arranged in a rectangular array. Starting in the back left corner and counting from left to right and from back to front, every fourth unit counted was designated for storage on the top shelf of a refrigerator. This counting method was also used to designate units for storage on both the middle and bottom shelves of the refrigerator, as well as for storage at room temperature. Sizes that were designated for storage on the refrigerator shelves but were too large to fit on the shelves were placed in the refrigerator door. Sizes too large for the middle shelf, but small enough for either the top or bottom shelf, were placed on either the top or bottom shelf. Bottling units were arranged in rectangular arrays in each storage area. After units were stored and allowed to chill for at least ten minutes, every third unit was counted off in each storage area (this was done by starting in the back left corner and by working from left to right and from back to front; units from the middle shelf were selected by counting from left to right and from front to back). These third units that were counted off were the units that were tested. Storage areas were sampled in the following order: room, top shelf, refrigerator door, middle shelf, and bottom shelf.

For each test unit selected, data were collected for each of the factors of interest. Storage temperature was measured using thermometers placed in each storage area. Both a calendar and the printed expiration date on each test unit were used to determine days until expiration. A coin was flipped to determine whether or not ice would be present during a particular test. To determine which test container (i.e., the container that a test unit was tested in) would be used in a particular test, nine containers selected for the study (one small, medium, and large container each of glass, plastic, or ceramic) were arranged in a 3 x 3 rectangular array. Starting in the back left corner and working from left to right and from back to front, every other container was selected to be the test container for a particular test.

After a test unit and a test container were selected, tests were conducted on a kitchen counter in a well-lit area. The test unit, a liquid measure measuring cup, and the test container were lined up in a row on the counter with the measuring cup in the middle. A metronome was then turned on at 60 beats per minute. The test unit was opened, and 4 fluid ounces of liquid were carefully poured down the side of the measuring cup. The measuring cup was then dumped into the test container between two clicks of the metronome. The moment that the sample was dumped, the experimenter said, "Time." The experimenter then observed the foam head dissolving in the test container. When only one ring of bubbles was left along the edge of the test container, the experimenter said, "Time," again. Each test was recorded using a video camera.

After tests were filmed, each video was reviewed. Foam head dissolve time was measured using a stopwatch; measurement began at the first instant that the experimenter said "time" and ceased the second time that the experimenter said "time" on each video.

ANALYSES

After the data were collected, a univariate analysis was conducted for each predictor of interest, as well as for the response variable "foam head dissolve time." For the quantitative predictors and response variable, PROC MEANS, a procedure from the SAS/STAT module within SAS 9.3, was used to investigate the mean, median, and 95% confidence limits for the mean. For each class of the categorical predictors, PROC FREQ, another procedure from the SAS/STAT module within SAS 9.3, was used to find the counts, frequencies, and 95% confidence limits for the frequencies. Confidence limits for the frequencies were found using the BINOMIAL option of PROC FREQ.

Next, the continuous response variable was dichotomized into a binary outcome of "small fizzer/big fizzer." This was done in preparation for a bivariate analysis and for fitting a final logistic model (one of the stipulations of this project was that the final outcome of interest had to be binary). The median dissolve time of 12 seconds, found during the univariate analysis, served as the cut point for the dichotomization because it was believed that the response variable would not be normally distributed. A histogram produced by PROC UNIVARIATE, yet another procedure from the SAS/STAT module within SAS 9.3, revealed that the response variable was somewhat right-skewed. This histogram can be seen in Figure 1 on the following page.

A bivariate analysis with respect to the outcome of interest was then conducted for each predictor of interest. Scatterplots were generated for each quantitative predictor using PROC GLM, a procedure from the SAS/STAT module within SAS 9.3; these were used to determine whether or not quantitative predictors were significant predictors of foam head dissolve time. AICs, the c statistic, and odds ratio estimates with their 95% confidence limits were also used to determine the significance of quantitative predictors; these were obtained using PROC LOGISTIC, another procedure from the SAS/STAT module within SAS 9.3. To determine the significance of categorical variables,

two-way tables were produced using PROC FREQ. These tables were also used to determine which levels of each categorical predictor were similar enough to be combined for further analysis. Wherever possible, odds ratios with their 95% confidence limits were also examined to determine categorical predictor significance; these were produced using the CMH option of PROC FREQ. For categorical predictors that were not binary, table row percentages from the two-way tables generated by PROC FREQ were examined to determine predictor significance.

Also, the presence or absence of a dose response was investigated for the categorical ordinal variable “test container size.” This was done by investigating tables, odds ratios, and confidence intervals produced using PROC FREQ with the CMH option enabled. Pairs of variables were also assessed for both interaction and confounding; this was done by investigating contingency tables, odds ratios, and confidence intervals produced by PROC FREQ with the CMH, RISKDIFF, and MEASURES options enabled.

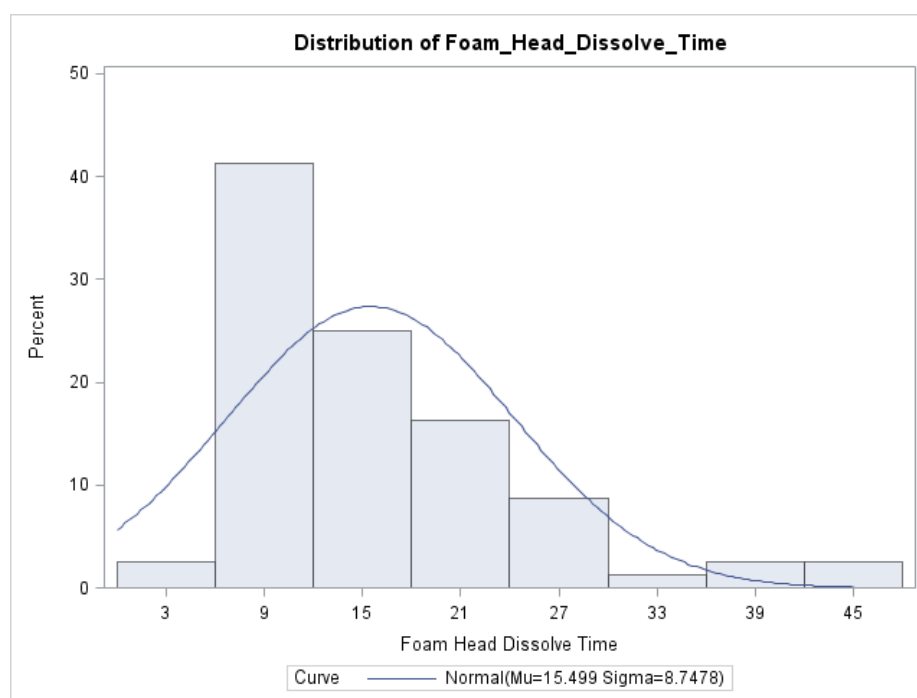


Figure 1. Histogram of foam head dissolve time plotted against a normal curve

A final logistic model was fit based on the results of all these analyses; this was done using PROC LOGISTIC. See Appendix IV for the SAS 9.3 code used for all analyses.

RESULTS

The results of the univariate analysis are displayed in Table 1, found in Appendix I. Most of the test units were classified as “Coke” under the variable “Class” and were bottled in 12 fluid ounce aluminum cans. Most of the items tested were also classified as “Regular” under the variable “Flavor” and were sweetened with high fructose corn syrup; 63.75% of the units tested also had caffeine. The percentages for the levels of the remaining categorical predictors were more evenly distributed due to the design of the study. The mean storage temperature was 47.0°F [95% CI: 44.6, 49.4], while the mean number of days until expiration was 94 [95% CI: 73, 115]. The mean foam head dissolve time was 15.50 seconds [95% CI: 13.55, 17.45]. The median foam head dissolve time was 12.41 seconds; to get the cut point for the outcome of interest for the bivariate analyses, this value was rounded to the nearest integer value. All tables containing the results of the univariate analyses can be found in Appendix I.

The results of the bivariate analyses for the continuous predictors are given in Table 2, found in Appendix II. Neither of the continuous predictors was found to be significant. The Table 3 series, also found in Appendix II, shows the results of the bivariate analyses for the categorical predictors. For the variable “class” in Table 3a, it can be seen from the row percentages that Barq’s Root Beer, Mello Yello, and Sprite were significantly different from the other classes; for this variable, no other classes were found to be significantly different from the others. After collapsing categories into similar groups, it was found that place of storage, bottling size, bottling material, flavor, test container size, test container material, the presence/absence of caffeine, and the presence/absence of ice were not significant predictors of foam head dissolve time, according to the results of this study. This can be seen from the row percentages, odds ratios, and confidence intervals for the odds ratios in the Appendix II table series. However, after collapsing categories into “artificial” and “natural,” the variable “sweetener” was found to be significant (see Table 3f), with an

odds ratio of 2.25 [95% CI: 0.91, 5.54]. This confidence interval is mostly positive and is consistent with a weak to moderate relationship.

Table 4, found in Appendix III, shows the results of checking for a dose response for the variable “test container size.” The categories “small” and “medium” were combined because they had almost identical “small fizzer/big fizzer” counts. However, a significant dose response was not found (OR = 0.83, with 95% CI: 0.32, 2.12).

Table 5, also in Appendix III, shows that a significant interaction between caffeine and test container material was found. Without caffeine, the odds ratio with respect to test container material was 11.70 [95% CI: 1.85, 74.19]. In other words, caffeine-free beverages tested in plastic or ceramic were 11.7 times more likely to be a “small fizzer” than those tested in glass. With caffeine, the odds ratio for test container material was 0.76 [95% CI: 0.23, 2.53].

Table 6, also in Appendix III, shows the results of checking for confounding between test container size and the presence/absence of ice, but no significant confounding was discovered for this pair of variables.

Based on the results of both the univariate and the bivariate analyses conducted in this study, a final logistic model was fit to the data using PROC LOGISTIC. The variables “sweetener” and “test container material” were retained, along with the test container material/caffeine interaction, resulting in the following model:

$$P(\text{Small Fizzer}) = \frac{1}{1 + e^x}$$

where

$$x = 0.3578 + 0.3816 * \text{sweetener} + 0.5823 * \text{test container material} + 0.6565 * \text{caffeine} * \text{test container material}$$

and

$$\text{sweetener} = \begin{cases} 1 & \text{if artificial} \\ -1 & \text{if natural} \end{cases}$$

$$\text{test container material} = \begin{cases} 1 & \text{if glass} \\ -1 & \text{if not glass} \end{cases}$$

$$\text{caffeine} = \begin{cases} 1 & \text{if absent} \\ -1 & \text{if present} \end{cases}$$

This model gives the probability that a foam head will dissolve in 12 seconds or less given the sweetener type, the test container material, and the presence/absence of caffeine. According to this model, the most influential factor was the interaction between caffeine and the test container material. The coefficients in the final model are from the Analysis of Maximum Likelihood Estimates table produced by PROC LOGISTIC. The factor levels are from the Class Level Information table produced by the same invocation of PROC LOGISTIC.

The odds ratio for sweetener was 2.15 (95% CI: 0.815, 5.645), while the odds ratio for testing in plastic or ceramic when caffeine was present was 1.79. The odds ratio for testing in plastic or ceramic when caffeine was absent was 3.45. These odds ratios were controlling for the type of sweetener used.

Figure 2 shows the bar chart comparing the frequencies of each fizzer type with respect to sweetener type. Clearly, test units with artificial sweeteners were more likely to be big fizzers than small fizzers. Figure 3 shows the bar charts comparing the frequencies of each fizzer type with respect to test container material; Figure 3a shows the bar chart for when caffeine is absent, while Figure 3b shows the bar chart for when caffeine is present. Clearly, different results were obtained depending on the presence/absence of caffeine.

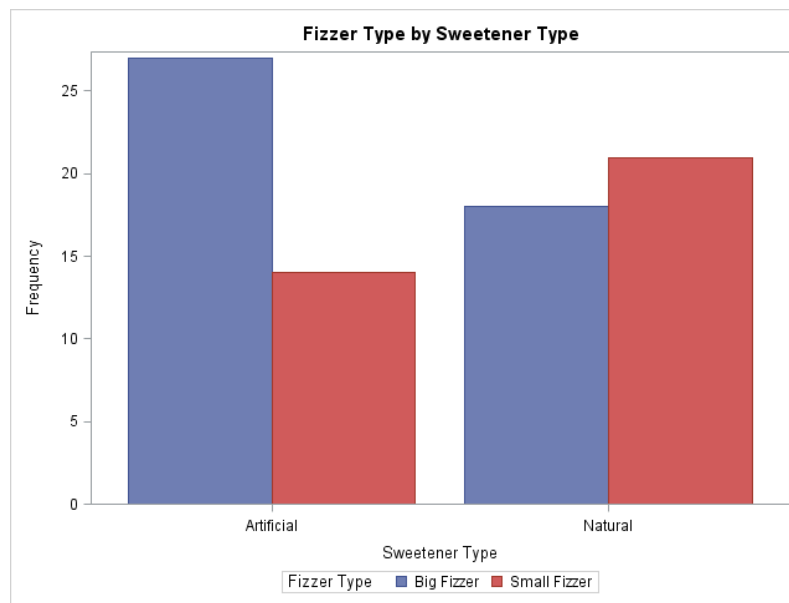


Figure 2. Bar Chart: Fizzer type according to sweetener type

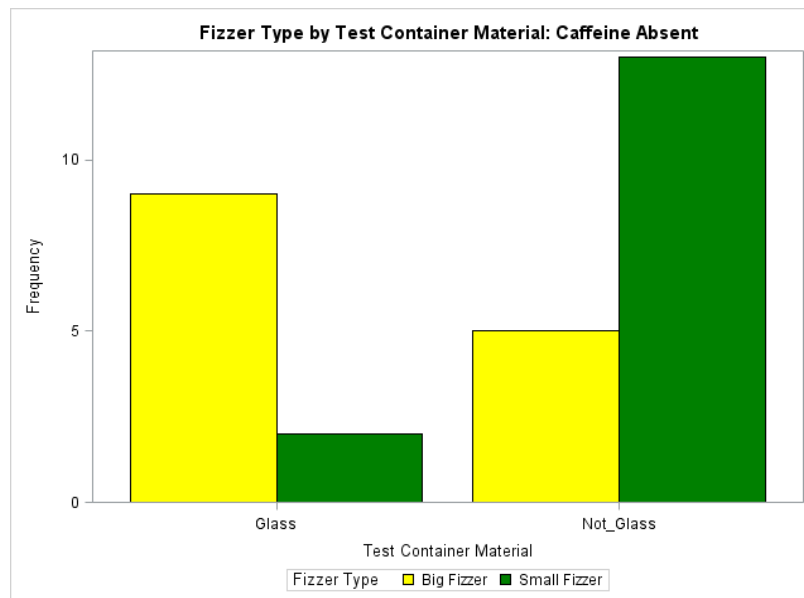


Figure 3a. Bar Chart: Fizzer type according to test container material – caffeine absent

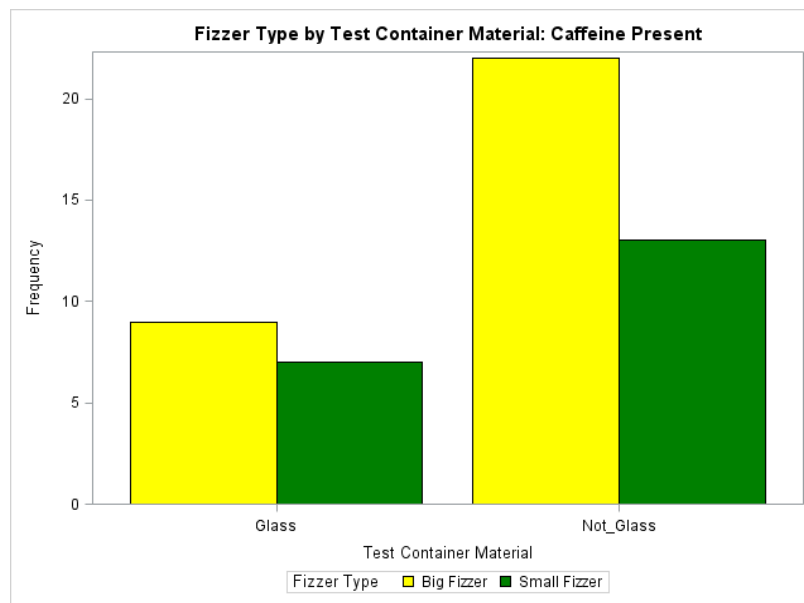


Figure 3b. Bar Chart: Fizzer type according to test container material – caffeine present

CONCLUSIONS

Out of all the factors investigated, only a few were significant predictors of whether or not foam head dissolve time would be 12 seconds or less, according to the results of this study. These factors were sweetener type and the test container material/caffeine interaction, with the interaction being the most influential factor. However, these results may be due to both the sample size tested and the particular sample, itself. If a larger sample had been used, perhaps different results would have been obtained. For example, the presence/absence of ice may have come out to be significant. Still, the knowledge gained from this study can be useful both to marketers and consumers. For example, Diet Coke has an artificial sweetener in it and is more likely to be a “big fizzer” than regular Coke. Perhaps, to increase sales of Diet Coke, marketing campaigns for Diet Coke should target people who like a large foam head when they pour their sodas. Also, people who do not want their caffeine-free sodas to lose a lot of fizz when they first pour them could be advised to pour them into something other than glass.

The results of this study suggest many possibilities for future studies. As mentioned earlier, this study could be repeated with a larger sample size in order to verify its results. This study could also be repeated using Pepsi products instead of Coke products, or Coke and Pepsi products could be compared in the same study. Perhaps a Cox proportional hazards model could be fit to analyze the data with respect to the original continuous response variable. The possibilities are many.

ACKNOWLEDGEMENT

Special thanks to Dr. Louise Lawson at Kennesaw State University for her much appreciated assistance with the analysis portion of this study and for coming up with the “big fizzer/small fizzer” naming convention.

REFERENCES

- IOP Publishing. "A Fizzy Drink Is like the Bends." Physics.org. Accessed July 31, 2014.
<http://www.physics.org/food-physics/text-only/>. Scroll down the page to find different articles.
- Rehmeyer, Julie. "Equation: Why Soda Goes Pop." Wired. Last modified April 26, 2011. Accessed July 31, 2014. http://www.wired.com/magazine/2011/04/st_equation_sodapop/.

CONTACT INFORMATION

Kate Phillips can be contacted via email at k8phill43@gmail.com. She can also be reached via cell phone at 770-596-6658. Mailing address: 3782 Shiloh Church Rd NW, Kennesaw, GA 30152.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are registered trademarks or trademarks of their respective companies.

APPENDIX I: UNIVARIATE ANALYSIS TABLES

Univariate Analysis: Continuous Variables				
Variable	Mean	Median	Lower Limit: 95% CI for Mean	Upper Limit: 95% CI for Mean
Storage Temperature (°F)	47.0	45.0	44.6	49.4
Days until Expiration	94	45	73	115
Foam Head Dissolve Time (sec.)	15.50	12.41	13.55	17.45

Table 1a. Univariate Analysis of Continuous Variables

Univariate Analysis: Categorical and Binary Variables				
Variable: Class	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
Barq's	3.75%	3	0.00%	7.91%
Coke	61.25%	49	50.57%	71.93%
Fanta	12.50%	10	5.25%	19.75%
Mello Yello	3.75%	3	0.00%	7.91%
Pibb Xtra	3.75%	3	0.00%	7.91%
Sprite	15.00%	12	7.18%	22.82%
Variable: Place of Storage	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
Bottom Shelf	20.00%	16	11.23%	28.77%
Door	18.75%	15	10.20%	27.30%
Middle Shelf	20.00%	16	11.23%	28.77%
Room	21.25%	17	12.29%	30.21%
Top Shelf	20.00%	16	11.23%	28.77%
Variable: Bottling Size (fl oz.)	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
7.5	3.75%	3	0.00%	7.91%
8	2.50%	2	0.00%	5.92%
12	60.00%	48	49.26%	70.74%
16.9	8.75%	7	2.56%	14.94%
42.2	5.00%	4	0.22%	9.78%
67.6	20.00%	16	11.23%	28.77%
Variable: Bottling Material	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
Aluminum	51.25%	41	40.30%	62.20%
Glass	3.75%	3	0.00%	7.91%
Plastic	45.00%	36	34.10%	55.90%
Variable: Flavor	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
Cherry	6.25%	5	0.95%	11.55%
Cranberry	2.50%	2	0.00%	5.92%
Grape	2.50%	2	0.00%	5.92%
Lime	1.25%	1	0.00%	3.68%
Orange	7.50%	6	1.73%	13.27%
Pineapple	1.25%	1	0.00%	3.68%
Regular	76.25%	61	66.92%	85.58%
Strawberry	1.25%	1	0.00%	3.68%
Vanilla	1.25%	1	0.00%	3.68%

Table 1b. Univariate Analysis of Categorical and Binary Variables

Univariate Analysis: Categorical and Binary Variables				
Variable: Sweetener	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
Aspartame	21.25%	17	12.29%	30.21%
Aspartame/ Acesulfame Potassium	26.25%	21	16.61%	35.89%
High Fructose Corn Syrup	47.50%	38	36.56%	58.44%
Sucralose/ Acesulfame Potassium	3.75%	3	0.00%	7.91%
Sugar	1.25%	1	0.00%	3.68%
Variable: Test Container Material	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
Ceramic	33.75%	27	23.39%	44.11%
Glass	33.75%	27	23.39%	44.11%
Plastic	32.50%	26	22.24%	42.76%
Variable: Test Container Size	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
Small	33.75%	27	23.39%	44.11%
Medium	32.50%	26	22.24%	42.76%
Large	33.75%	27	23.39%	44.11%
Variable: Caffeine Present	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
No	36.25%	29	25.72%	46.78%
Yes	63.75%	51	53.22%	74.28%
Variable: Ice Present	Percent	Count	Lower Limit: 95% CI for Percent	Upper Limit: 95% CI for Percent
No	53.16%	42	42.16%	64.17%
Yes	46.84%	37	35.83%	57.84%

Table 1b (cont.). Univariate Analysis of Categorical and Binary Variables

APPENDIX II: BIVARIATE ANALYSIS TABLES

Bivariate Analysis: Continuous Variables; "Big Fizzers"				
Variable	Mean	Median	Lower Limit: 95% CI for Mean	Upper Limit: 95% CI for Mean
Storage Temperature (°F)	47.5	45.0	44.4	50.7
Days until Expiration	82.73	42	56	110
Foam Head Dissolve Time (sec.)	20.65	18.94	18.08	23.22
Bivariate Analysis: Continuous Variables; "Small Fizzers"				
Variable	Mean	Median	Lower Limit: 95% CI for Mean	Upper Limit: 95% CI for Mean
Storage Temperature (°F)	46.2	41.5	42.4	50.1
Days until Expiration	109	49	75	142
Foam Head Dissolve Time (sec.)	8.88	8.82	8.29	9.46

Table 2a. Bivariate Analysis of Continuous Variables: Mean, Median, and 95% Confidence Limits for the Mean

Bivariate Analysis: Continuous Variables			
Variable	AIC	c statistic	Odds Ratio (95% CI)
Storage Temperature (°F)	113.352	0.545	1.012 (0.970, 1.055)
Days until Expiration	112.130	0.556	0.997 (0.992, 1.002)

Table 2b. Bivariate Analysis of Continuous Variables: AIC, c statistic, and Odds Ratio with 95% Confidence Limits for the Odds Ratio

Bivariate Analysis: "Class" by "Small Fizzer"			
Class	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
Barq's Root Beer	3 (100.00%)	0 (0.00%)	3
Coke	32 (65.31%)	17 (34.69%)	49
Fanta	7 (70.00%)	3 (30.00%)	10
Mello Yello	1 (33.33%)	2 (66.67%)	3
Pibb Xtra	2 (66.67%)	1 (33.33%)	3
Sprite	0 (0.00%)	12 (100.00%)	12
Total	45 (56.25%)	35 (43.75%)	80

Table 3a. Bivariate Analysis of Categorical Variables: "Class" by "Small Fizzer"

Bivariate Analysis: "Place of Storage" by "Small Fizzer"			
Place of Storage	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
Bottom Shelf	11 (68.75%)	5 (31.25%)	16
Door	9 (60.00%)	6 (40.00%)	15
Middle Shelf	7 (43.75%)	9 (56.25%)	16
Room	10 (58.82%)	7 (41.18%)	17
Top Shelf	8 (50.00%)	8 (50.00%)	16
Total	45 (56.25%)	35 (43.75%)	80

Table 3b. Bivariate Analysis of Categorical Variables: "Place of Storage" by "Small Fizzer"

Bivariate Analysis: "Bottling Size" by "Small Fizzer"			
Bottling Size (fl oz.)	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
7.5, 8, 12, or 16.9	33 (55.00%)	27 (45.00%)	60
42.2 or 67.6	12 (60.00%)	8 (40.00%)	20
Total	45 (56.25%)	35 (43.75%)	80
Odds Ratio (95% CI)			
0.81 (0.29, 2.28)			

Table 3c. Bivariate Analysis of Categorical Variables: "Bottling Size" by "Small Fizzer"

Bivariate Analysis: "Bottling Material" by "Small Fizzer"			
Bottling Material	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
Not Plastic	27 (61.36%)	17 (38.64%)	44
Plastic	18 (50.00%)	18 (50.00%)	36
Total	45 (56.25%)	35 (43.75%)	80
Odds Ratio (95% CI)			
1.59 (0.65, 3.87)			

Table 3d. Bivariate Analysis of Categorical Variables: "Bottling Material" by "Small Fizzer"

Bivariate Analysis: "Flavor" by "Small Fizzer"			
Flavor	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
Extra Flavoring	10 (52.63%)	9 (47.37%)	19
Regular	35 (57.38%)	26 (42.62%)	61
Total	45 (56.25%)	35 (43.75%)	80
Odds Ratio (95% CI)			
0.83 (0.29, 2.32)			

Table 3e. Bivariate Analysis of Categorical Variables: "Flavor" by "Small Fizzer"

Bivariate Analysis: "Sweetener" by "Small Fizzer"			
Sweetener	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
Artificial	27 (65.85%)	14 (34.15%)	41
Natural	18 (46.15%)	21 (53.85%)	39
Total	45 (56.25%)	35 (43.75%)	80
Odds Ratio (95% CI)			
2.25 (0.91, 5.54)			

Table 3f. Bivariate Analysis of Categorical Variables: "Sweetener" by "Small Fizzer"

Bivariate Analysis: "Test Container Material" by "Small Fizzer"			
Test Container Material	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
Glass	18 (66.67%)	9 (33.33%)	27
Not Glass	27 (50.94%)	26 (49.06%)	53
Total	45 (56.25%)	35 (43.75%)	80
Odds Ratio (95% CI)			
1.93 (0.73, 5.05)			

Table 3g. Bivariate Analysis of Categorical Variables: "Test Container Material" by "Small Fizzer"

Bivariate Analysis: "Test Container Size" by "Small Fizzer"			
Test Container Size	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
L	16 (59.26%)	11 (40.74%)	27
M	14 (53.85%)	12 (46.15%)	26
Sm	15 (55.56%)	12 (44.44%)	27
Total	45 (56.25%)	35 (43.75%)	80

Table 3h. Bivariate Analysis of Categorical Variables: "Test Container Size" by "Small Fizzer"

Bivariate Analysis: "Caffeine Present" by "Small Fizzer"			
Caffeine Present	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
No	14 (48.28%)	15 (51.72%)	29
Yes	31 (60.78%)	20 (39.22%)	51
Total	45 (56.25%)	35 (43.75%)	80
Odds Ratio (95% CI)			
0.60 (0.24, 1.51)			

Table 3i. Bivariate Analysis of Categorical Variables: "Caffeine Present" by "Small Fizzer"

Bivariate Analysis: "Ice Present" by "Small Fizzer"			
Ice Present	Small Fizzer		
	0 = No	1 = Yes	Total
	Count (Row Percentage)	Count (Row Percentage)	
No	24 (57.14%)	18 (42.86%)	42
Yes	21 (56.76%)	16 (43.24%)	37
Total	45 (56.96%)	34 (43.04%)	79 (n missing = 1)
Odds Ratio (95% CI)			
1.02 (0.42, 2.48)			

Table 3j. Bivariate Analysis of Categorical Variables: "Ice Present" by "Small Fizzer"

APPENDIX III: DOSE RESPONSE CHECK, INTERACTION CHECK, AND CONFOUNDING CHECK

Dose Response Check: "Test Container Size" versus "Small Fizzer" - Original				
Small Fizzer	Test Container Size			
	Large	Medium	Small	Total
	Count (Column %)	Count (Column %)	Count (Column %)	
No	16 (59.26%)	14 (53.85%)	15 (55.56%)	45 (56.25%)
Yes	11 (40.74%)	12 (46.15%)	12 (44.44%)	35 (43.75%)
Total	26	27	26	80

Table 4a. Primary Dose Response Check: "Test Container Size" versus "Small Fizzer"

Dose Response Check: "Test Container Size" versus "Small Fizzer" – Combined Categories			
Small Fizzer	Test Container Size		
	Small or Medium	Large	Total
	Count (Column %)	Count (Column %)	
No	29 (54.72%)	16 (59.26%)	45 (56.25%)
Yes	24 (45.28%)	11 (40.74%)	35 (43.75%)
Total	53	27	80
Odds Ratio (95% CI)			
0.83 (0.32, 2.12)			

Table 4b. Dose Response Check after Combining Categories: "Test Container Size" versus "Small Fizzer"

Check for Interaction: “Test Container Material” and “Caffeine Present”								
No Caffeine					With Caffeine			
Success	Test Container Material				Success	Test Container Material		
	Glass	Not Glass	Total			Glass	Not Glass	Total
No	9	5	14		No	9	22	31
Yes	2	13	15		Yes	7	13	20
Total	11	18	29		Total	16	35	51
Odds Ratio (95% CI:)					Odds Ratio (95% CI:)			
11.70 (1.85, 74.19)					0.76 (0.23, 2.53)			

Table 5. Testing for Interaction: "Test Container Material" and "Caffeine Present"

Check for Confounding: "Ice Present" and "Test Container Size"		
Ice Absent	OR for Test Container Size: 0.30	95% CI: (0.03, 2.76)
Ice Present	OR for Test Container Size: 0.28	95% CI: (0.02, 3.58)
	Crude OR: Test Container Size: 0.36	95% CI: (0.08, 1.72)

Table 6. Check for Confounding: "Ice Present" and "Test Container Size"

APPENDIX IV: SAS CODE

```
*****
STAT 8125 Project Analysis - The Coke Project
*****;

*****
Using pilot data, determine the needed sample size for the main study data.
*****;

*set up SAS library; LIBNAME coke "\\client\C$\Users\Kate\Documents\Graduate Files,
The\STAT 8125 (Epidem class)\Coke Project";

*import pilot data; PROC IMPORT DATAFILE= "\\client\C$\Users\Kate\Documents\Graduate
Files, The\STAT 8125 (Epidem class)\Coke Project\Excel database- with pilot data for
samp size.csv" OUT= cokeproject DBMS= CSV; DATAROW= 2; DELIMITER= ","; GETNAMES= YES;
RUN;

*****
Two-sample t-test: Caffeine vs. No Caffeine
*****;

PROC TTEST DATA= cokeproject; VAR Time_for_Foam_Head_to_go_Down__s; CLASS
Caffeine_Present; RUN;

*Determine sample sizes needed for 90% power when detecting 1, 3, and 5 second
differences between group mean dissolve times;
```

```

PROC POWER; TWOSAMPLEMEANS TEST= DIFF MEANDIFF= 1 3 5 STDDEV= 3.2959 8.2782 ALPHA=
0.05 NTOTAL= . POWER= .9; RUN;

*****
Two-sample t-test: Ice vs. No Ice
*****;
PROC TTEST DATA= cokeproject; VAR Time_for_Foam_Head_to_go_Down__s; CLASS Ice_Present;
RUN;

*Determine sample sizes needed for 90% power when detecting 1, 3, and 5 second
differences between group mean dissolve times;
PROC POWER; TWOSAMPLEMEANS TEST= DIFF MEANDIFF= 1 3 5 STDDEV= 3.9292 16.7445 ALPHA=
0.05 NTOTAL= . POWER= .9; RUN;

*For the following, determine the needed sample size for detecting (at 90% power) a
standard deviation of 3 seconds between mean dissolve times for each factor level.
*****
ANOVA: Class
*****;
*with randomization of materials, only two classes got tested during pilot data
collection (Coke and Spite);
PROC POWER; ONEWAYANOVA GROUPMEANS= 7.193 |12.936 STDDEV= 3 NPERGROUP= . POWER= .90;
RUN;

*****
ANOVA: Place of Storage
*****;
PROC POWER; ONEWAYANOVA GROUPMEANS= 13.777 |9.51| 12.87|9.843 STDDEV= 3 NPERGROUP= .
POWER= .90; RUN;

*****
ANOVA: Bottling Container Size
*****;
PROC POWER; ONEWAYANOVA GROUPMEANS= 11.138 |7.15| 17.15|11.16|7.4 STDDEV= 3 NPERGROUP=
. POWER= .90; RUN;

*****
ANOVA: Bottling Material
*****;
PROC POWER; ONEWAYANOVA GROUPMEANS= 10.47 |13.125|11.16 STDDEV= 3 NPERGROUP= . POWER=
.90; RUN;

*****
ANOVA: Flavor
*****;
PROC POWER; ONEWAYANOVA GROUPMEANS= 11.813 |8.06 STDDEV= 3 NPERGROUP= . POWER= .90;
RUN;

*****
ANOVA: Sweetener
*****;
PROC POWER; ONEWAYANOVA GROUPMEANS= 9.549 |12.56| 23.93 STDDEV= 3 NPERGROUP= . POWER=
.90; RUN;

*****
ANOVA: Test Container Material
*****;
PROC POWER; ONEWAYANOVA GROUPMEANS= 9.06 |12.198| 12.458 STDDEV= 3 NPERGROUP= . POWER=
.90; RUN;

*****
ANOVA: Test Container Size
*****;
PROC POWER; ONEWAYANOVA GROUPMEANS= 9.823 |16.128| 7.275 STDDEV= 3 NPERGROUP= . POWER=
.90; RUN;

```

```

/* Results: detecting a difference of 3 seconds between mean dissolve times for the
"ice present" and "ice absent" groups requires the largest sample size. Required
sample size is between 76 and 1312 units. A sample size of 80 will be used for the
main study data.*/

*****
Main Study Data Analyses
*****;
/*main study data imported using Import Wizard: import "Excel Database- Carbonated
Beverages Project with data.xlsx". Name the imported data set "cokes".*/

*Univariate analysis - quantitative variables. Look at mean, median, upper and lower
quantiles, standard deviation, minimum, maximum, and the mean's 95% confidence
limits.;
ODS RTF FILE= "\\client\C$\Users\Kate\Documents\Graduate Files, The\STAT 8125 (Epidem
class)\Coke Project\SAS output.rtf"; PROC MEANS DATA= cokes N MEAN MEDIAN Q1 Q3 STDDEV
MIN MAX CLM; VAR Storage_Temperature___F_ Days_Until_Expiration
Time_for_Foam_Head_to_go_Down; RUN;

*Shorten name of response variable; DATA cokes4; SET cokes; Foam_Head_Dissolve_Time =
Time_for_Foam_Head_to_go_Down; RUN;

*Checking distribution of Foam Head Dissolve Time;
PROC UNIVARIATE DATA= cokes4 NOPRINT; VAR Foam_Head_Dissolve_Time; HISTOGRAM
Foam_Head_Dissolve_Time / NORMAL; LABEL Foam_Head_Dissolve_Time= "Foam Head Dissolve
Time"; RUN;
/* Result: distribution is right-skewed. Use median Time_for_Foam_Head_to_go_Down (12
seconds) as cut point for binary outcome*/

*Univariate analysis - categorical variables. Look at counts, percents, and 95% CIs
for percents;
PROC FREQ DATA= cokes; TABLES Class / BINOMIAL (LEVEL= "B"); RUN;
PROC FREQ DATA= cokes; TABLES Class / BINOMIAL (LEVEL= "Co"); RUN;
PROC FREQ DATA= cokes; TABLES Class / BINOMIAL (LEVEL= "F"); RUN;
PROC FREQ DATA= cokes; TABLES Class / BINOMIAL (LEVEL= "MY"); RUN;
PROC FREQ DATA= cokes; TABLES Class / BINOMIAL (LEVEL= "PX"); RUN;
PROC FREQ DATA= cokes; TABLES Class / BINOMIAL (LEVEL= "Sp"); RUN;
PROC FREQ DATA= cokes; TABLES Place_of_Storage / BINOMIAL (LEVEL= "Bottom Shelf");RUN;
PROC FREQ DATA= cokes; TABLES Place_of_Storage / BINOMIAL (LEVEL= "Door");RUN;
PROC FREQ DATA= cokes; TABLES Place_of_Storage / BINOMIAL (LEVEL= "Middle Shelf");RUN;
PROC FREQ DATA= cokes; TABLES Place_of_Storage / BINOMIAL (LEVEL= "Room");RUN;
PROC FREQ DATA= cokes; TABLES Place_of_Storage / BINOMIAL (LEVEL= "Top Shelf");RUN;
PROC FREQ DATA= cokes; TABLES Bottling_Container_Size_fl_oz__ / BINOMIAL (LEVEL=
"7.5");RUN;
PROC FREQ DATA= cokes; TABLES Bottling_Container_Size_fl_oz__ / BINOMIAL (LEVEL=
"8");RUN;
PROC FREQ DATA= cokes; TABLES Bottling_Container_Size_fl_oz__ / BINOMIAL (LEVEL=
"12");RUN;
PROC FREQ DATA= cokes; TABLES Bottling_Container_Size_fl_oz__ / BINOMIAL (LEVEL=
"16.9");RUN;
PROC FREQ DATA= cokes; TABLES Bottling_Container_Size_fl_oz__ / BINOMIAL (LEVEL=
"42.2");RUN;
PROC FREQ DATA= cokes; TABLES Bottling_Container_Size_fl_oz__ / BINOMIAL (LEVEL=
"67.6");RUN;
PROC FREQ DATA = cokes; TABLES Bottling_Material / BINOMIAL (LEVEL= "Al");RUN;
PROC FREQ DATA = cokes; TABLES Bottling_Material / BINOMIAL (LEVEL= "G");RUN;
PROC FREQ DATA = cokes; TABLES Bottling_Material / BINOMIAL (LEVEL= "P");RUN;
PROC FREQ DATA= cokes; TABLES Flavor / BINOMIAL (LEVEL= "Ch");RUN;
PROC FREQ DATA= cokes; TABLES Flavor / BINOMIAL (LEVEL= "Cr");RUN;
PROC FREQ DATA= cokes; TABLES Flavor / BINOMIAL (LEVEL= "Gr");RUN;
PROC FREQ DATA= cokes; TABLES Flavor / BINOMIAL (LEVEL= "Li");RUN;
PROC FREQ DATA= cokes; TABLES Flavor / BINOMIAL (LEVEL= "O");RUN;
PROC FREQ DATA= cokes; TABLES Flavor / BINOMIAL (LEVEL= "Pin");RUN;
PROC FREQ DATA= cokes; TABLES Flavor / BINOMIAL (LEVEL= "R");RUN;
PROC FREQ DATA= cokes; TABLES Flavor / BINOMIAL (LEVEL= "Str");RUN;
PROC FREQ DATA= cokes; TABLES Flavor / BINOMIAL (LEVEL= "V");RUN;
PROC FREQ DATA = cokes; TABLES Sweetener / BINOMIAL (LEVEL= "A");RUN;

```

```

PROC FREQ DATA = cokes; TABLES Sweetener / BINOMIAL (LEVEL= "A/AP");RUN;
PROC FREQ DATA = cokes; TABLES Sweetener / BINOMIAL (LEVEL= "HFCS");RUN;
PROC FREQ DATA = cokes; TABLES Sweetener / BINOMIAL (LEVEL= "S/AP");RUN;
PROC FREQ DATA = cokes; TABLES Sweetener / BINOMIAL (LEVEL= "Su");RUN;
PROC FREQ DATA= cokes; TABLES Testing_Container_Material / BINOMIAL (LEVEL= "C");RUN;
PROC FREQ DATA= cokes; TABLES Testing_Container_Material / BINOMIAL (LEVEL= "G");RUN;
PROC FREQ DATA= cokes; TABLES Testing_Container_Material / BINOMIAL (LEVEL= "P");RUN;
PROC FREQ DATA= cokes; TABLES Testing_Container_Size / BINOMIAL (LEVEL= "L");RUN;
PROC FREQ DATA= cokes; TABLES Testing_Container_Size / BINOMIAL (LEVEL= "M");RUN;
PROC FREQ DATA= cokes; TABLES Testing_Container_Size / BINOMIAL (LEVEL= "Sm");RUN;
PROC FREQ DATA= cokes; TABLES Caffeine_Present_ / BINOMIAL (LEVEL= "0");RUN;
PROC FREQ DATA= cokes; TABLES Caffeine_Present_ / BINOMIAL (LEVEL= "1");RUN;
PROC FREQ DATA= cokes; TABLES Ice_Present_ / BINOMIAL (LEVEL= "0");RUN;
PROC FREQ DATA= cokes; TABLES Ice_Present_ / BINOMIAL (LEVEL= "1");RUN;

*Preparing variables for bivariate analysis;
DATA cokes2; SET cokes;
/*Dichotomizing outcome (Success = 1 : Time_for_Foam_Head_to_go_Down less than or
equal to 12 seconds. Success = 0 : Time_for_Foam_Head_to_go_Down greater than 12
seconds). Also, "success" = "small fizzer."*/
IF Time_for_Foam_Head_to_go_Down LE 12 THEN success = 1; ELSE success = 0;
*Collapsing predictors for analysis;
IF Storage_Temperature__F_ LE 45 THEN Storage_Temperature = 1; ELSE IF
Storage_Temperature__F_ LE 55 THEN Storage_Temperature = 2; ELSE Storage_Temperature
= 3; IF Bottling_Container_Size_fl_oz__ IN(7.5,8,12,16.9) THEN bottle_size = 1; ELSE
bottle_size = 2; IF Bottling_Material IN("A1","G") THEN bottle_material =
"not_plastic"; ELSE bottle_material = "plastic"; IF Flavor IN("R") THEN flavor =
"regular"; ELSE flavor = "extra"; IF Sweetener IN("HFCS", "Su") THEN sweetener =
"Natural"; ELSE sweetener = "Artificial"; IF Testing_Container_Material IN("C","P")
THEN test_material = "Not_Glass"; ELSE test_material = "Glass"; Dissolve_Time =
Time_for_Foam_Head_to_go_Down; RUN;

*Sorting data by outcome variable "success";
PROC SORT DATA = cokes2; BY success; RUN;
*****
*This is bivariate analysis for numeric predictors
*****;
*Get mean, median, min, max and 95% Confidence Intervals for continuous variables with
respect to the outcome of interest. Since median is the source of the cut point for
small fizzer/big fizzer, focus on median, min, and max;
PROC MEANS DATA= cokes2 N MEAN MEDIAN Q1 Q3 STDDEV MIN MAX CLM; VAR
Storage_Temperature__F_ Days_Until_Expiration Time_for_Foam_Head_to_go_Down; BY
success; RUN;

*Obtain AIC, c statistic, and odds ratio (with 95% CI for odds ratio) for each
continuous predictor;
PROC LOGISTIC DATA= cokes2; MODEL success = Storage_Temperature__F_ ;RUN;
PROC LOGISTIC DATA= cokes2; MODEL success = Days_Until_Expiration;RUN;
*Results: neither of these predictors is significant by itself;

*Getting scatterplots - storage temperature;
PROC GLM DATA= cokes2; WHERE success = 0; MODEL Dissolve_Time =
Storage_Temperature__F_ / EST; RUN; *not a good predictor of response variable;
PROC GLM DATA= cokes2; WHERE success = 1; MODEL Dissolve_Time =
Storage_Temperature__F_ / EST; RUN; *not a good predictor of response variable;
*Getting scatterplots - days until expiration;
PROC GLM DATA= cokes2; WHERE success = 0; MODEL Dissolve_Time = Days_Until_Expiration
/ EST; RUN; *also not a good predictor of response variable;
PROC GLM DATA= cokes2; WHERE success = 1; MODEL Dissolve_Time = Days_Until_Expiration
/ EST; RUN; *also not a good predictor of response variable; QUIT;

*****
This is bivariate analysis for categorical and binary predictors
*****;
/*In the following PROC FREQ, look at tables, ORs and CIs where available. Compare row
percentages where tables, ORs and CIs not available. Focus on case-control ORs and CIs
since primary focus is now on small fizzer/big fizzer outcome*/

```

```

PROC FREQ DATA= cokes2; TABLES(Class Place_of_Storage bottle_size bottle_material
Flavor Sweetener test_material Testing_Container_Size Caffeine_Present_
Ice_Present_)*success/ CMH; RUN;
*Result - sweetener type has the only significant OR;

*Check for dose response based on test container size;
PROC FREQ DATA= cokes2; TABLE success*Testing_Container_Size/ CMH; RUN;
PROC FREQ DATA= cokes2; TABLE success*Testing_Container_Size/ CMH; WHERE
Testing_Container_Size IN("L","M"); RUN;
PROC FREQ DATA= cokes2; TABLE success*Testing_Container_Size/ CMH; WHERE
Testing_Container_Size IN("M","Sm"); RUN;

*Combine categories - Small and Medium (data set "cokes3" created to combine
categories for testing dose response);
DATA cokes3; SET cokes2; IF Testing_Container_Size IN("M","Sm") THEN Test_Size = 1;
ELSE Test_Size = 2; RUN;

*Recheck for dose response;
PROC FREQ DATA= cokes3; TABLE success*Test_Size/ CMH; RUN; *no significant dose
response based on test container size;

*Check for interaction. In TABLE statement, put the response variable between the two
factors of interest. Also, list the factor that you want to "control for" first.;
PROC FREQ DATA= cokes2; TABLE Caffeine_Present_*success*test_material / CMH RISKDIFF
MEASURES; RUN; *interaction present!;

*Check for confounding 1: Ice present and Test container size;
PROC FREQ DATA= cokes3; WHERE Caffeine_Present_ = 0; TABLE
Ice_Present_*success*Test_Size / CMH RISKDIFF MEASURES; RUN; *no confounding here;

*Getting crude OR (to check for confounding): Test container size vs. success;
PROC FREQ DATA= cokes3; WHERE Caffeine_Present_ = 0; TABLE Test_Size*success / CMH
RISKDIFF MEASURES; RUN;

*Check for confounding 2: Ice present and Sweetener type (Natural vs. Artificial);
PROC FREQ DATA= cokes3; WHERE Caffeine_Present_ = 0; TABLE
Ice_Present_*success*Sweetener / CMH RISKDIFF MEASURES; RUN; *no confounding here;

*Getting crude OR (to check for confounding): Sweetener type (Natural vs. Artificial)
vs. success;
PROC FREQ DATA= cokes3; WHERE Caffeine_Present_ = 0; TABLE Sweetener*success / CMH
RISKDIFF MEASURES; RUN;
/*Results: no significant confounding present with these pairs of variables*/

*Fitting logistic regression model;
PROC LOGISTIC DATA= cokes2; CLASS Sweetener Caffeine_Present_ test_material; MODEL
success = Sweetener test_material Caffeine_Present_*test_material; RUN; ODS RTF CLOSE;

*****
Getting bar charts for significant factors
*****;
*Changing value of success variable from "0/1" to "Big Fizzer/Small Fizzer";
DATA varchange; SET cokes2; IF success = 1 THEN new_success = "Small Fizzer"; ELSE
new_success = "Big Fizzer"; RUN;

*Bar chart for sweetener type;
PROC SGPLOT DATA= varchange; VBAR Sweetener / GROUP= new_success GROUPDISPLAY=
CLUSTER; LABEL Sweetener = "Sweetener Type"; LABEL new_success = "Fizzer Type"; TITLE
"Fizzer Type by Sweetener Type"; RUN;

*****
Bar charts for test container material/caffeine interaction
*****;
*Creating data set: caffeine absent;
DATA nocaffchart; SET varchange; IF Caffeine_Present_ = 0 THEN OUTPUT; RUN;

*Changing colors of bar chart;

```

```

DATA attrmap; INPUT ID $11. @13 VALUE $12. @29 FILLCOLOR $8. LINECOLOR $8.;
    DATALINES;
new_success Small Fizzer      GREEN    BLACK
new_success Big Fizzer       YELLOW    BLACK
;
RUN;

*Creating bar chart for test container material vs. fizzer type (caffeine absent);
PROC SGPLOT DATA= nocaffchart DATTRMAP= attrmap; VBAR test_material / GROUP=
new_success GROUPDISPLAY= CLUSTER ATTRID= new_success; LABEL test_material = "Test
Container Material"; LABEL new_success = "Fizzer Type"; TITLE "Fizzer Type by Test
Container Material: Caffeine Absent"; RUN;

*Creating data set: caffeine present;
DATA caffchart; SET varchange; IF Caffeine_Present_ = 1 THEN OUTPUT; RUN;

*Creating bar chart for test container material vs. success (caffeine present);
PROC SGPLOT DATA= caffchart DATTRMAP= attrmap; VBAR test_material / GROUP= new_success
GROUPDISPLAY= CLUSTER ATTRID= new_success; LABEL test_material = "Test Container
Material"; LABEL new_success = "Fizzer Type"; TITLE "Fizzer Type by Test Container
Material: Caffeine Present"; RUN;

```