Stat 158 Final Write-up

Jane Li, Isha Thapa, Jared Wilber

Spring 2016

1. Introduction

When it comes to competing brands, perhaps no rivalry is more popular than that between Coca-Cola and Pepsi-Cola. Although the two sodas are very similar in taste, many people have preferences for one or the other. Due to the similarity of taste of the two drinks, it is unclear whether people's preferences are solely based on the taste of the drinks or also due to bias caused by other factors. The motivation behind our experiment is to assess whether people's biases towards brand names and container types could potentially affect their soda preference. Specifically, our experiment seeks to examine the effect of brand label, bottle type, and tasting order on people's preference between Coke and Pepsi. We did a full factorial design where subjects were required to taste both types of soda, each subject receiving one treatment that had some combination of the factors listed above. Although the results of this experiment were inconclusive, this report explores the implications of these factors and the results that occurred as a consequence of changing them.

2. Methodology

This experiment tests whether people's biases towards brand names and bottle types affect their perception of soda quality. We decided to test three conditions of interest: container type (can, plastic bottle, or glass bottle), tasting order (pepsi first or coke first), and label (true label, no label, and false label) and see how these factors affect people's comparisons of the two sodas.

We had three levels for container type because we wanted to test all types of containers that these two sodas came in. Because we tested two soda brands, there were only two possible orders to give them in, which is why we had two levels for the order condition. We then also tested all possible combinations for labelling, which is why we had three levels for that condition.

Subjects were given two cups of soda in the order specified by the treatment. First they were asked which drink they liked better and then they were asked to rank by how much they preferred the better drink on a scale of 0 to 5, with 5 being that they strongly preferred the drink and 0 indicating no preference. This type of response is ordinal because it has an order, but no meaningful distance between the rankings. Our analysis later on considers the rankings of coke compared to pepsi. So if the subject preferred pepsi, we converted the ranking to a negative score to keep the responses consistent with respect preferring coke.

The units we used were the individual people we used as subjects. We did a full factorial design, so the units were the same for every condition - each person received only one treatment. 18 treatments were completed in each section, so we had two replications of the experiment. We used a completely randomized design. To randomize while accounting for the replication, we took two sets of the ordered list of 18 treatments and then numbered each treatment from 1-36. We then took a random permutation of the ordered list,, which we have listed in the appendix. Then subjects who came to us received the treatment in the order given by the permuted list.

We chose to do a full factorial design because we wanted to test all combinations of the conditions. We believe a full factorial design is the correct design for a variety of reasons. One reason is because it is feasible to do every crossing with our factors, so a full factorial model allowed us to do so and analyze all interactions between our conditions. Our sample size is not very large, so it may seem like blocking could have helped us gather more data. However, blocking can cause other problems. For example, if we show the label to the subject and then have them repeat the experiment where we switch the order, they may be biased by their first response. It is likely they would just give the negative of the original number they gave to keep their second response consistent with their first one. To avoid problems such as this one, we gave only one treatment to each person.

However, there were still other issues. One potential problem was if a subject could identify which soda was Coke and which soda was Pepsi. In that case, the treatments with false labels would be invalid because the subject would actually know which soda was which. There was no way for us to account for or prevent a problem such as that one.

Another potential problem was that some of the bottles had to be used during the whole section. We closed them immediately after opening them each time, but there was still a release of air and the change could have slightly altered the taste of the sodas. If one soda gets flat faster than the other, then having open bottles could have potentially changed the preferences. A way to account for this issue would be to open a new bottle for each subject, but we were unable to do so due to budget constraints.

One problem we actually ran into, which we were not expecting, was that one of the subjects refused to drink the soda. Eliminating people who dislike soda from the pool of subjects could potentially have an effect on the results. It is possible that those who hate soda have different biases and preferences than those who enjoy, so only testing those who are willing to participate may lead to results that are not representative of the whole population.

This experiment does have validity in the sense that the experiment gives results that help answer our overarching question of whether container type or brand name creates biases in perception of soda quality. However there are two issues. One is that we hope to answer the question regarding people in general, but we were only able to test a very small subset of the population, specifically the students in Professor Purdom's Stat 158 class, and this subset may not be representative of the population as a whole. This lessens our validity because the results cannot necessarily be extended to people in general. The validity can also be questionable if people are able to discern which soda is which because, in that case, the switched labels would not have an effect on the participant. Thus such subjects' rankings will not help answer the question of whether or not the label creates biases.

There are minor issues with replicability as well too. Although it is easy to repeat the experiment exactly the same way each time, it may be hard for subjects to give a consistent response due to the nature of the response type. When responses are ordinal, it is hard for subjects distinguish the difference between two rankings. For example, when asked the same question twice, the subject may give a four one time, but a three another time because the rankings have no fixed meaning.

3. Results

During the experiment, we had each individual subject input their response into a google form questionnaire. We used this form so that other subjects wouldn't hear a tested subject's response value. We then appended these response values to their corresponding treatment combinations in our R data.frame.

Once we had our data in an organized, usable format, we could begin our analysis. As a precursor to ANOVA, we performed some exploratory data analysis to get an idea of what our data looks like. The first thing we did was plot the obtained response values to get quick, basic idea about their distribution.

Figure 1.1

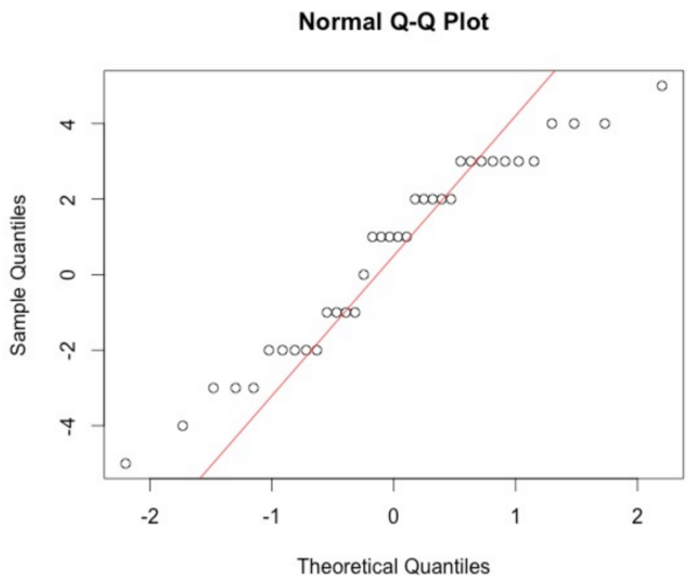
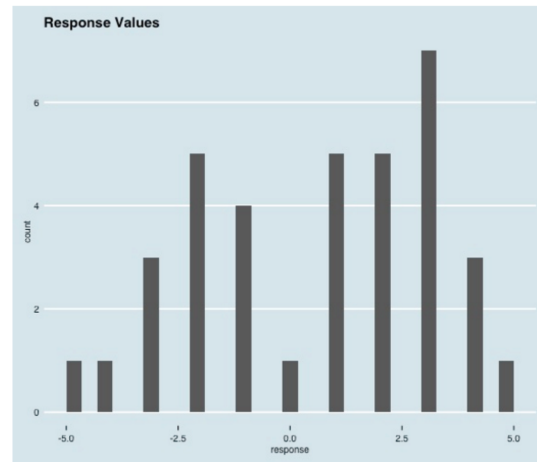


Figure 1.2



From the QQ plot (Figure 1.1) of the response values, we see the responses roughly follow a normal distribution, as the majority of the values fall on the 45 degree line in the QQ plot. The histogram (Figure 1.2) is also roughly bell shaped, suggesting a normal distribution.

Next, we created some box plots of our data. Box plots are useful for comparing distributions and identifying any outliers that may exist. Below we display the obtained box plots. Note that each box plot is grouped by both label type and order, as well as color-coded by bottle type.

Figure 2.1

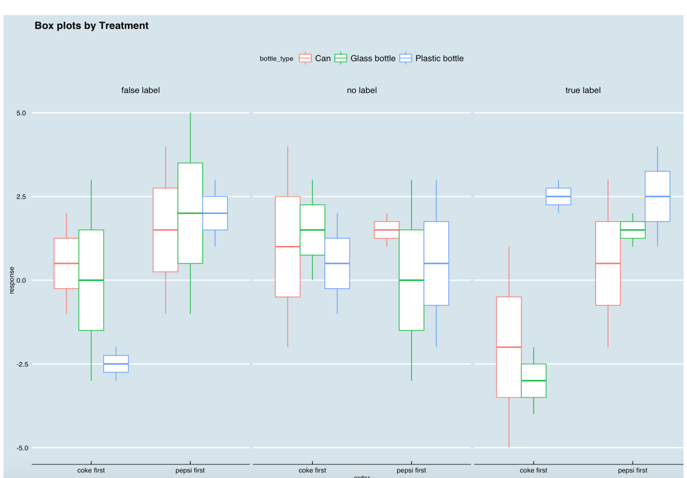
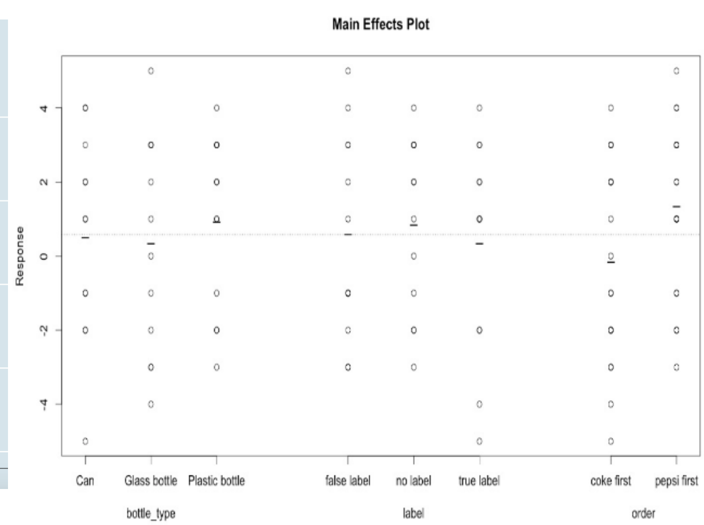


Figure 2.2



The above depicted box plots (Figure 2.1) provide information for all three treatment combinations. We can immediately see that most treatment combinations within the order treatment share relatively similar median values. However, this is violated starkly twice, once in the false label, coke first combination, and again in the true label, coke first combination. In both deviations, we observe that the outlier value belongs to the plastic bottle, indicating a possibility of an effect. While we observe that there is no apparent pattern of skewness in any of the treatment combinations, plastic bottles appear to have the smallest range of values across the board, and glass bottle appear to have the largest range (although this is violated for those treatment combinations with a true label, where cans possess the largest range of values). Our box plots depict no outlier values, which is not unexpected given our small scale of possible values.

The Main Effects plot shows that the 8 level means across 3 factors fall rather close to the overall mean, suggesting that the three factors might not have an significant effect on the average preference of Coke to Pepsi. All of the factors seem to have fairly symmetric data; none of the data seems skewed. The factors also have similar spreads, as the minimum and maximum points for each one are all very close to each other.

After getting an idea of the distribution, skew, and an overall feel for our data, we used interaction plots to evaluate any potential relationships and interactions between our treatment effects.

Figure 3.1

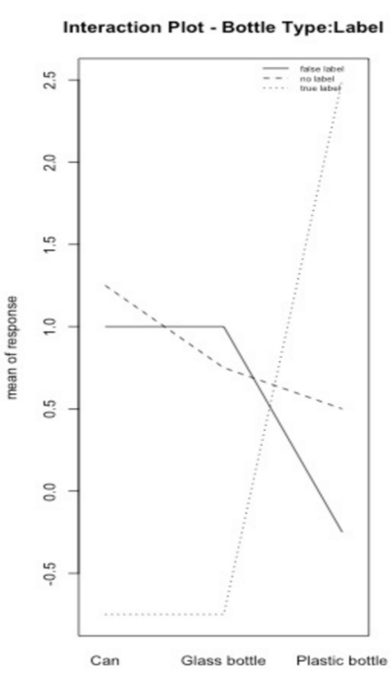


Figure 3.2

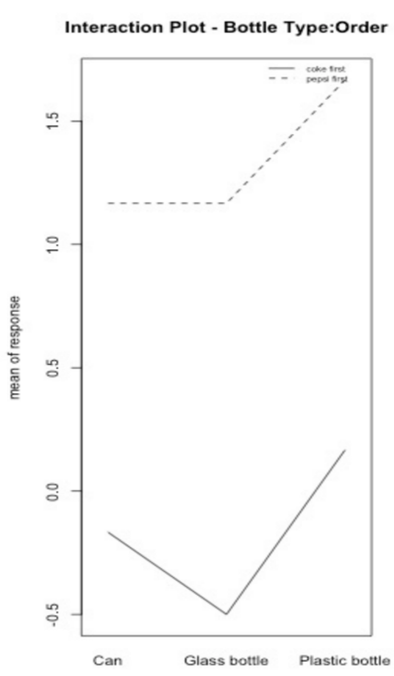
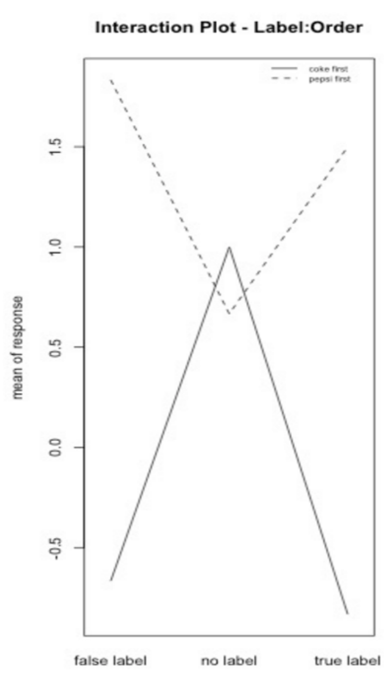


Figure 3.3



In terms of interactions among the factors, we see that there potentially could be interactions between bottle type and label, and between label and order, as the lines intersect in their interaction plots. Bottle type and order don't seem to have any meaningful interaction as the lines are almost parallel to each other on the interaction plot.

Figure 4.1

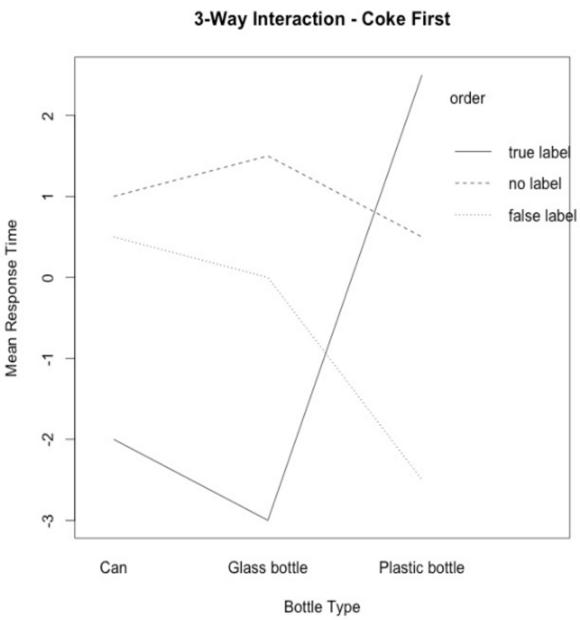
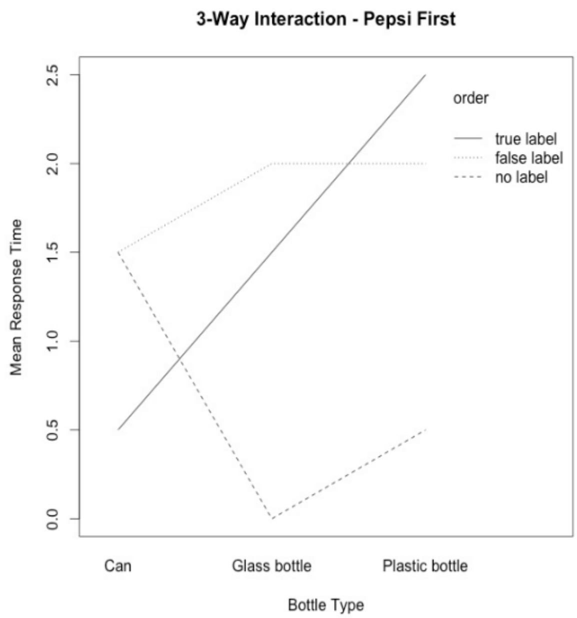


Figure 4.2



Since both the 3-way interactions (Figures 4.1 and 4.2) indicate lines that cross with one another, there could also be a potential 3-way interaction between the factors.

Now that we've acquired some intuition about our data, we're almost ready to begin our formal analysis. As our choice of design is a three-way factorial, we will use ANOVA to test for any significant effects among our treatments and their interactions. ANOVA relies on three primary assumptions: normally distributed residuals, homoscedastic residuals, and independence among residuals. Before performing our formal analysis, we must check whether our data satisfies these assumptions

Although ANOVA is generally robust to violations of normality, we deem it important to assess the assumption of normality. To accomplish this assessment, we produce a quantile-quantile plot for our residuals (Figure). The plot convincingly reveals that our data satisfies this assumption.

The next assumption to meet is that of homoscedasticity. That is, do our residuals share constant variance? This assumption is easily assessed by a residual vs. fitted plot, displayed below:

Figure 5.1

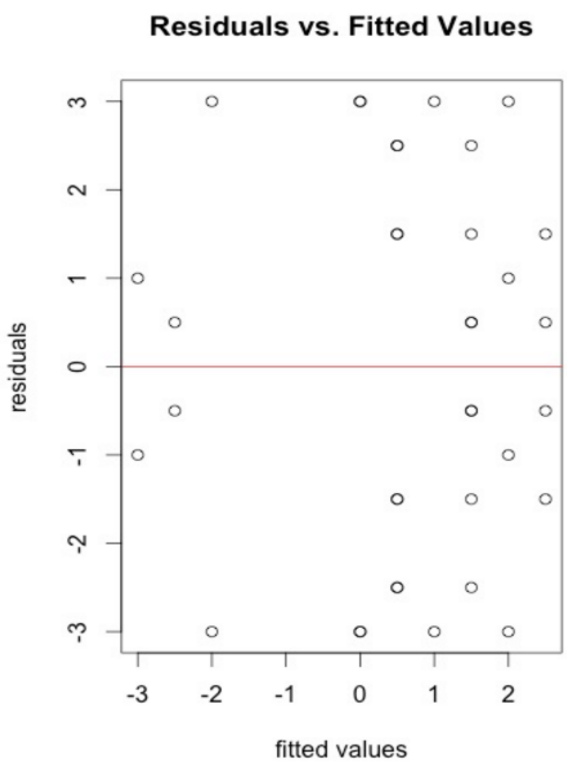
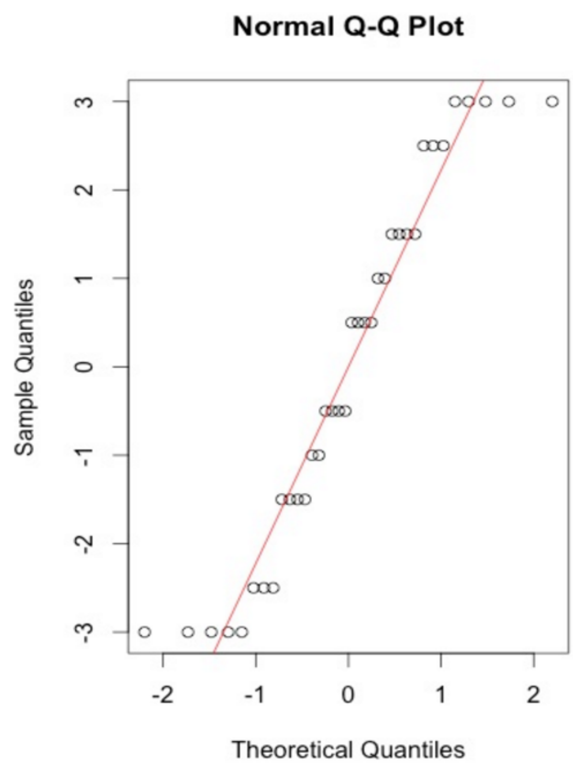


Figure 5.2



Since the residual points are randomly scattered around 0 and do not exhibit a relationship with the fitted values in the residuals vs. fitted plot, we see that the residuals do not vary in variance and it suggests that the errors are homoscedastic. The residuals also fall roughly on the 45 degree line in the QQ plot, hence the errors approximately follow a normal distribution. Assuming that the errors also are independent from one another with outside knowledge, all the assumptions are therefore satisfied for performing the parametric F-test, and we will be using it in our analysis.

For our analysis, we carried out a three-way factorial ANOVA. Our null hypothesis stipulated that there was no effect for any of the treatments or treatment combinations on a given subject's preference for Coca-Cola. Our alternative hypothesis specified that at least one of our treatments or treatment combinations did have an effect on a given subject's preference for Coca-Cola. For our test, we decided to assign our significance level to the conventional value of 0.05, as we wanted to retain a small type I error without being too stringent on our declarations for significance. The obtained ANOVA results are contained in the following table:

Figure 6.1

Treatment	Degrees of Freedom	Sum of Squares	Mean Square	F-Value	P-Value
Order	1	20.25	20.25	2.406	0.138
Label	2	1.50	0.75	0.089	0.915
Bottle Type	2	2.17	1.083	0.129	0.880
Order:Label	2	15.17	7.583	0.901	0.424
Order:Bottle Type	2	0.17	0.083	0.010	0.990
Label:Bottle Type	4	31.33	7.833	0.931	0.468
Order:Label:Bottle Type	4	18.67	4.667	0.554	0.698
Residuals	18	151.50	8.417		

As the ANOVA table displays, all of the p-values for the treatments and interactions have values less than 0.05. This result indicates that none of the treatments for this experiment had a significant effect. This holds true for the treatment combinations as well. We can observe that our lowest obtained p-value is associated with the "order" treatment, signifying that the order in which our subject drank the soda has potential to influence their preference. That said, the p-value is too large to draw any meaningful insights from the effect. The remaining treatments, aside from "order", all have very insignificant p-values, so we can confidently conclude that none of our treatments had significant effects on our response values. Thus, for our experiment, we fail to reject the null hypothesis; we don't possess enough evidence available to ascertain that the null hypothesis is false at a 95% confidence interval. That is, we conclude that our set of treatments and treatment combinations had no effect on a given subject's preference for either Pepsi or Cola.

4. Conclusion

Our experiment attempted to determine whether or not any of our selected treatments (or their combinations) had an effect on a subject's preference for Coca-Cola over Pepsi. The results from our ANOVA test reveal that none of treatments nor treatment combinations had significant effects on our response values. Thus we can conclude that our subjects' preferences for Pepsi or Coca-Cola are not influenced by the bottle type, label type, order received, nor any combination of the three. That said, it's still relevant to speculate any possible future treatments that could affect this preference. Naturally, this leads to the further question of what (if any) conditions affect a subject's preference for Pepsi over Cola (or vice-versa), or if this preference is completely determined by taste alone. Further considerations could involve including other beverages for comparison, implementing different ordering schemes, or including "influencers" (e.g. require the subject to watch an advert beforehand). Although our experiment yielded inconsequential results, significant findings in this enterprise could have very important applications in advertising, and thus further pursuance of factors affecting preference is advantageous.

5. Appendix

- TREATMENTS
- Combinations
- Code