

MP2 Technical Report:

Katy Cummings, Tulip Daaboul, Michel Ruiz-Fuentes

2023-05-04

Introduction

The dairy industry is a cornerstone of the United State's agricultural domain, even going as far as being the state drink for twenty-two out of thirty states with official state beverages (Davis, 2021); however, over the last decade or so, dairy farmers and corporations have struggled to maintain a strong clientele. This change in the industry is predominantly due to the increase in alternative milk options, like almond milk and oat milk, and varying beliefs about the health benefits of dairy milk and human consumption. An article published by the New York Times in April 2023 further discusses this generational change and the impact this change has on a historically prosperous industry (Severson, 2023). There is a changing trend in the processing and manufacturing of dairy products that relies on static consumption. In particular, there is an expansion and growth in the dairy consumption of physical products. Individuals are eating more dairy-based products than drinking dairy-based products (Durrow, 2023).

This study is interested in how different ad contents (dairy vs oat milk) and sentiment types (positive, negative, and neutral ads) influence people's consumption of dairy or plant-based products. The two types of milk are whole milk and oat milk, as well as health benefits/cons regarding the milk type. The article that inspired the study emphasizes that there are many misconceptions by the younger generations pertaining to the health benefits of dairy products, like milk being less healthy than alternatives (Severson, 2023). Therefore, we hypothesize that the alternative hypothesis that "there is at least one difference in the mean rate of participants' consumption depending on the treatment they received" will be supported by our findings, rejecting the null hypothesis that "there is no difference in the mean rate of participants' consumption depending on the treatment they received."

Methodology

Our project's experimental units consisted of thirty-five Smith College students. The study was a Qualtrics online survey that was distributed via an online link to current Smith students, predominantly from SDS 290. Due to this, the sample is a convenience sample, as those who

took the survey came majority from our Research Design and Analysis class, other classes, and some of our student organizations on campus. Because the respondents came from a very specific demographic, we cannot generalize results to the Smith population. If random sampling was used, then, we would have been able to generalize to Smithies.

Since participants were randomly assigned to the treatments, this study is an experiment. The two treatments are milk type and advertisement type. Participants were first assigned to either receive a dairy or plant based milk photo. Then, random assignment was used again and participants were either assigned to the positive, negative, or neutral group in respect to sentiment type. The type of sentiments were either positive or negative, and were presented through two bullet points for the product. The neutral group contains no bullet points, just the photo of the product.

The independent variables (experimental factors) in our study are categorical variables of milk type (two levels: whole or oat) and sentiment type (three levels: positive, negative, neutral). The dependent variable was average consumption, which is continuous. After being asked simple demographic questions, participants were asked if they were vegan or lactose intolerant. After seeing their treatment, participants rated on an ordinal scale from 1-5 (1 being least likely, 5 being most likely) on how many times a week they consume the type of milk they were shown. After completing this portion, participants rated on an ordinal scale their opinion on the dairy industry (1- dislike a great deal, 5 like a great deal). The purpose of the pre and post survey questions is to control for potential confounding variables. For our study, we are going to be using a two way ANOVA model to see if there is a true difference in average consumption based on ad contents and sentiment types.

Results

Load Data:

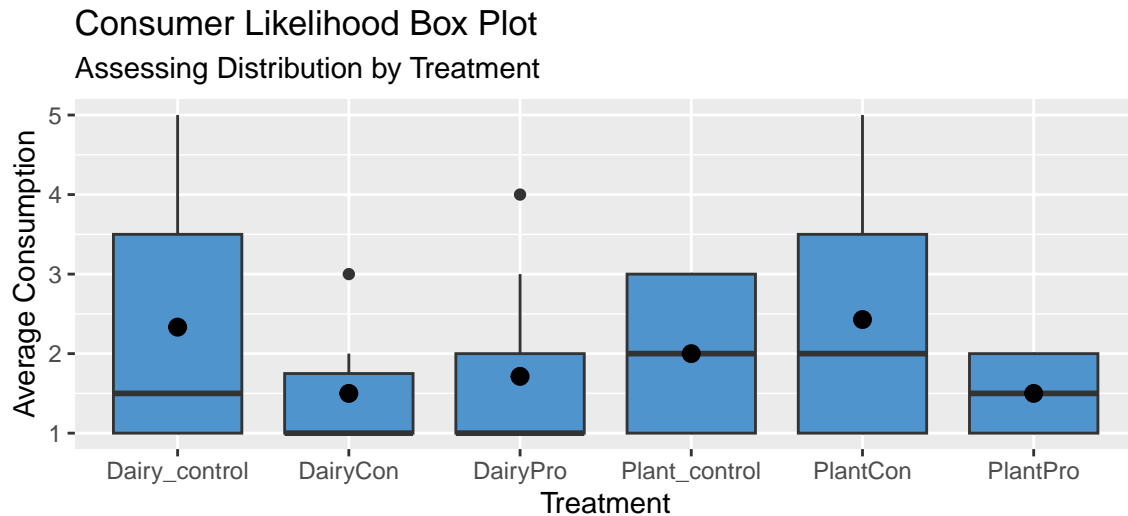
```
library(tidyverse)
library(rio)
library(asbio)
library(dplyr)
library(ggplot2)
library(Stat2Data)
library(car)

milk_df1 <- import("milk_cleaned_data.csv")

milk_df1$milk_type <- as.factor(milk_df1$milk_type)
milk_df1$ad_type <- as.factor(milk_df1$ad_type)
milk_df1$treatment <- as.factor(milk_df1$treatment)
```

Condition Check 1 Symmetrical Distribution around Group Means:

The box-plot below shows us the distribution of the average consumption on the different treatment combinations. We will compare the variation in the interquartile ranges of each treatment. By looking at the plot, we can see that the constant variance condition is violated. We see that we do not have constant variance across groups and that the spread is not normal. However, checking conditions with visualizations can be tricky, so we can refer to Levene's test to also check our constant variance condition.



source: Qualtrics Milk Data

Levene's Test:

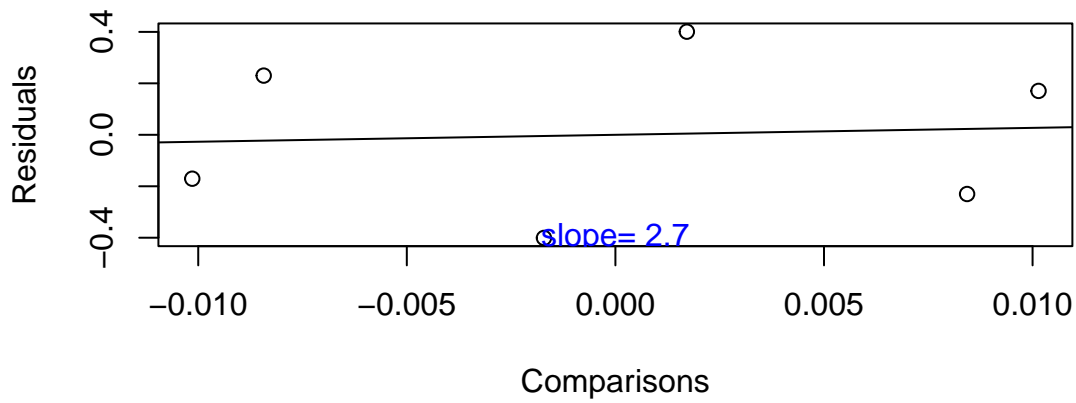
Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	5	0.8397	0.5326
	29		

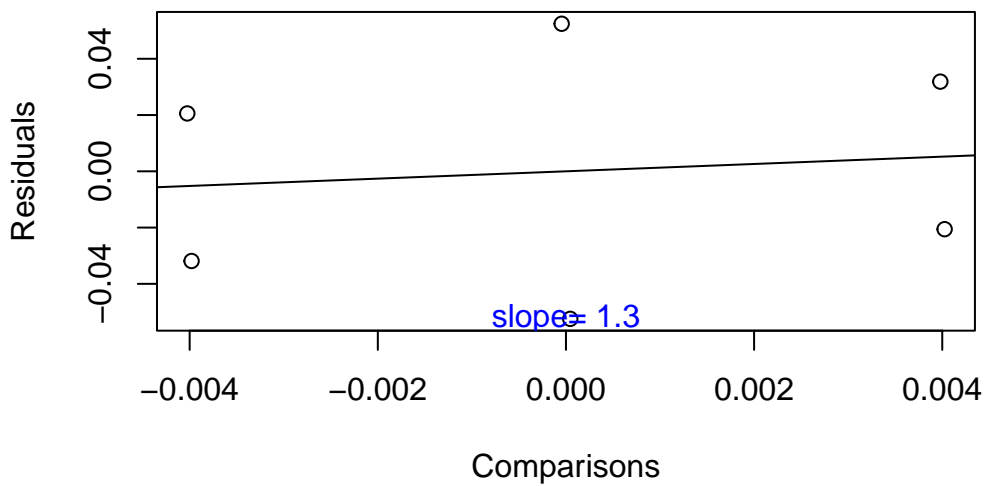
After conducting Levene's test of homogeneity of variance, we see that our p-value of 0.5326 is greater than 0.05. This means that we do not have statistically significant evidence that we have different variances, therefore constant variance condition is not violated.

Condition Check 2 Additive Effects & Transformation:

The Tukey Non-Additivity plot model helps us understand the relationship between the affect of both treatments over what we predicted. It also allows us to assess if the affects of treatments are additive or if there is some interaction. We want the relationship between the x-axis and y-axis to have a slope of 0 to indicate there is no relationship. However, because the data violates several assumptions and the slope is 2.7, so we must transform the data.



To diagnose the type of transformation we subtract 1-slope or 1-2.7 which gives the difference of -1.7. Because our $P = -1.7$, we must use a reciprocal transformation of $1/y$.



After transforming our data, we see that although we are getting closer to that slope of 0, we are not quite there yet. This means that an interaction might still be better for this model.

Two-Way ANOVA Model:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ad_type	2	1.70	0.8519	0.504	0.609
milk_type	1	0.28	0.2809	0.166	0.686
ad_type:milk_type	2	2.92	1.4624	0.866	0.431
Residuals	29	48.98	1.6888		

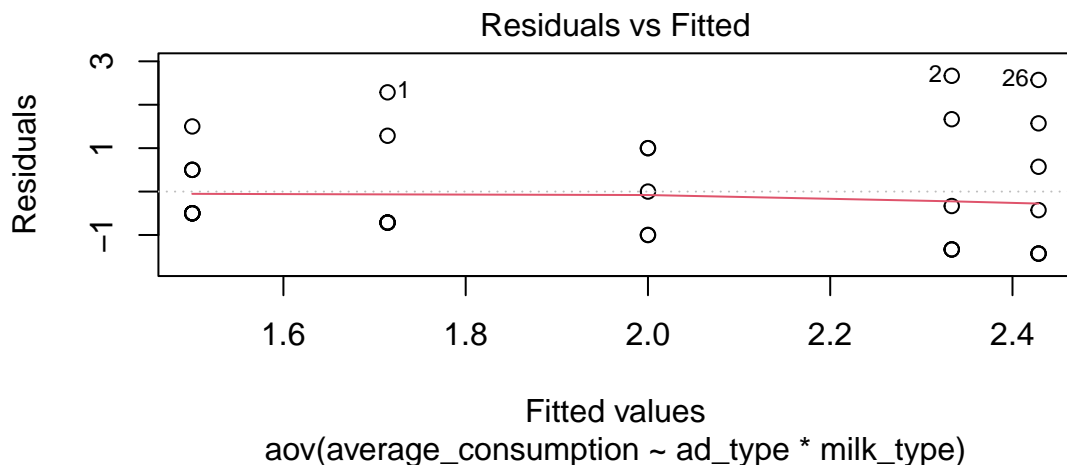
The interaction model above shows us whether we have main effects for ad type, milk type, and if we have an interaction present.

Because our p-value for ad_type is 0.609 (greater than 0.05), we do not have statistically significant evidence that average consumption is different based on ad type. We do not have a main effect for ad type.

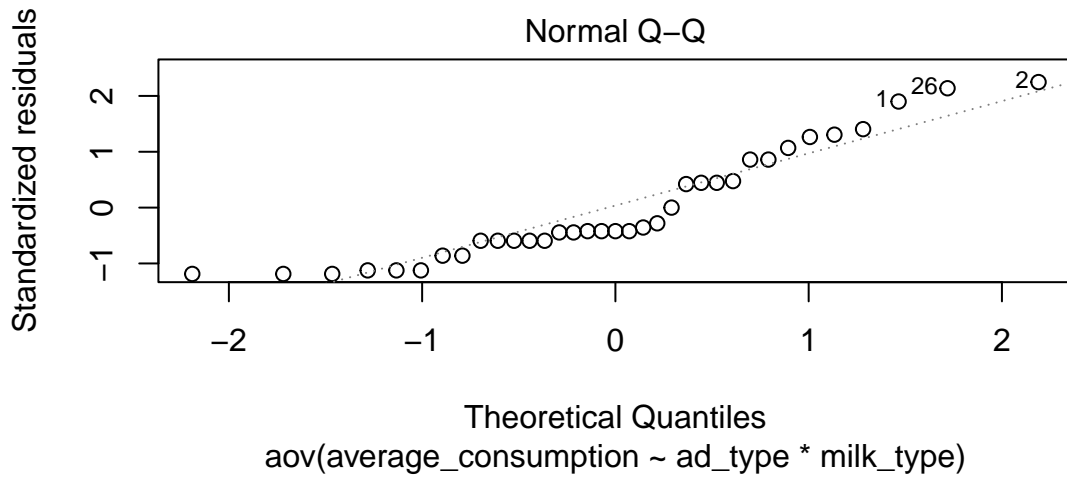
Because our p-value for milk_type is 0.686 (greater than 0.05), we do not have statistically significant evidence that average consumption is different based on milk type. We do not have a main effect for milk type.

When looking at our interaction effect, we see a p-value of 0.431. Because our p-value is greater than 0.05, we do not have statistically significant evidence that the effect of ad type is different based on the effect of milk type. We do not have statistically significant evidence that an interaction exists.

Independent residuals:



The residuals versus fitted plot can help us assess two patterns: non-linearity and non-constant spread. Because we see our residuals evenly spread out, we can say that our spread is constant. However, because our line does not fall on zero, non-linearity is present here.



We can use this normal quantile plot compares the standardized observed residuals with theoretical values from a normal distribution. This plot can help us determine whether our distribution is normal. Because we see that the residuals do not fall on the line, our distribution is not normal.

Confidence Intervals:

95% LSD confidence intervals

	LSD	Diff	Lower	Upper	Decision
muDairy.Con-muPlant.Con	1.47871	-0.92857	-2.40728	0.55014	FTR H0
muDairy.Con-muDairy.Control	1.53453	-0.83333	-2.36786	0.7012	FTR H0
muPlant.Con-muDairy.Control	1.47871	0.09524	-1.38347	1.57395	FTR H0
muDairy.Con-muPlant.Control	1.60943	-0.5	-2.10943	1.10943	FTR H0
muPlant.Con-muPlant.Control	1.5563	0.42857	-1.12772	1.98487	FTR H0
muDairy.Control-muPlant.Control	1.60943	0.33333	-1.27609	1.94276	FTR H0
muDairy.Con-muDairy.Pro	1.47871	-0.21429	-1.69299	1.26442	FTR H0
muPlant.Con-muDairy.Pro	1.4207	0.71429	-0.70641	2.13498	FTR H0
muDairy.Control-muDairy.Pro	1.47871	0.61905	-0.85966	2.09776	FTR H0
muPlant.Control-muDairy.Pro	1.5563	0.28571	-1.27058	1.84201	FTR H0
muDairy.Con-muPlant.Pro	1.71565	0	-1.71565	1.71565	FTR H0
muPlant.Con-muPlant.Pro	1.66592	0.92857	-0.73734	2.59449	FTR H0
muDairy.Control-muPlant.Pro	1.71565	0.83333	-0.88232	2.54899	FTR H0
muPlant.Control-muPlant.Pro	1.78296	0.5	-1.28296	2.28296	FTR H0

muDairy.Pro-muPlant.Pro	1.66592	0.21429	-1.45163	1.8802	FTR H0
	Adj. p-value				
muDairy.Con-muPlant.Con	0.20919				
muDairy.Con-muDairy.Control	0.27583				
muPlant.Con-muDairy.Control	0.89611				
muDairy.Con-muPlant.Control	0.53016				
muPlant.Con-muPlant.Control	0.57762				
muDairy.Control-muPlant.Control	0.67499				
muDairy.Con-muDairy.Pro	0.76905				
muPlant.Con-muDairy.Pro	0.31232				
muDairy.Control-muDairy.Pro	0.3989				
muPlant.Control-muDairy.Pro	0.71004				
muDairy.Con-muPlant.Pro	1				
muPlant.Con-muPlant.Pro	0.26362				
muDairy.Control-muPlant.Pro	0.32872				
muPlant.Control-muPlant.Pro	0.57069				
muDairy.Pro-muPlant.Pro	0.79435				

When conducting a pairwise comparison, Fishers LSD looks at the individual false alarm rate. This means it looks at the chance of a Type 1 error for a single test. Because our alpha level is 0.05, this means that for every error, those 0.05 error rates accumulate. Therefore, the Fishers LSD is the least conservative approach, meaning it is likely to reject the null hypothesis that the means of average consumption are the same. When looking at our pairwise comparisons, we see that all of the p-values are larger than 0.05. This means we do not have statistically significant evidence that the pairwise comparisons between the treatments are different using Fishers LSD.

95% Bonferroni confidence intervals

	Diff	Lower	Upper	Decision	Adj. p-value
muDairy.Con-muPlant.Con	-0.92857	-3.24092	1.38377	FTR H0	1
muDairy.Con-muDairy.Control	-0.83333	-3.23297	1.5663	FTR H0	1
muPlant.Con-muDairy.Control	0.09524	-2.21711	2.40758	FTR H0	1
muDairy.Con-muPlant.Control	-0.5	-3.01676	2.01676	FTR H0	1
muPlant.Con-muPlant.Control	0.42857	-2.0051	2.86224	FTR H0	1
muDairy.Control-muPlant.Control	0.33333	-2.18342	2.85009	FTR H0	1
muDairy.Con-muDairy.Pro	-0.21429	-2.52663	2.09806	FTR H0	1
muPlant.Con-muDairy.Pro	0.71429	-1.50734	2.93591	FTR H0	1
muDairy.Control-muDairy.Pro	0.61905	-1.6933	2.93139	FTR H0	1
muPlant.Control-muDairy.Pro	0.28571	-2.14796	2.71939	FTR H0	1
muDairy.Con-muPlant.Pro	0	-2.68287	2.68287	FTR H0	1

muPlant.Con-muPlant.Pro	0.92857	-1.67652	3.53366	FTR H0	1
muDairy.Control-muPlant.Pro	0.83333	-1.84954	3.5162	FTR H0	1
muPlant.Control-muPlant.Pro	0.5	-2.28812	3.28812	FTR H0	1
muDairy.Pro-muPlant.Pro	0.21429	-2.3908	2.81938	FTR H0	1

The Bonferroni confidence intervals focuses on family wise false alarm rate. Meaning we look at the 0.05 error on *all* the tests we conduct and have an overall false alarm rate of 0.05. This approach is the most conservative and least likely to reject the null hypothesis. Again, based on all of our pairwise comparisons, all of the p-values are larger than 0.05. This means we do not have statistically significant evidence that the pairwise comparisons between the treatments are different using Bonferroni confidence intervals.

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = average_consumption ~ ad_type * milk_type, data = milk_df1)

\$ad_type

	diff	lwr	upr	p adj
Control-Con	0.1818182	-1.133001	1.4966375	0.9378710
Pro-Con	-0.3636364	-1.678456	0.9511829	0.7751547
Pro-Control	-0.5454545	-1.913962	0.8230527	0.5924309

\$milk_type

	diff	lwr	upr	p adj
Plant-Dairy	0.1779306	-0.7239163	1.079778	0.6895292

\$`ad_type:milk_type`

	diff	lwr	upr	p adj
Control:Dairy-Con:Dairy	8.333333e-01	-1.453930	3.120596	0.8729870
Pro:Dairy-Con:Dairy	2.142857e-01	-1.989776	2.418348	0.9996576
Con:Plant-Con:Dairy	9.285714e-01	-1.275490	3.132633	0.7909276
Control:Plant-Con:Dairy	5.000000e-01	-1.898902	2.898902	0.9873020
Pro:Plant-Con:Dairy	4.440892e-16	-2.557238	2.557238	1.0000000
Pro:Dairy-Control:Dairy	-6.190476e-01	-2.823110	1.585014	0.9538115
Con:Plant-Control:Dairy	9.523810e-02	-2.108824	2.299300	0.9999938
Control:Plant-Control:Dairy	-3.333333e-01	-2.732235	2.065568	0.9980748
Pro:Plant-Control:Dairy	-8.333333e-01	-3.390571	1.723905	0.9163227
Con:Plant-Pro:Dairy	7.142857e-01	-1.403308	2.831880	0.9045235
Control:Plant-Pro:Dairy	2.857143e-01	-2.033994	2.605422	0.9989203
Pro:Plant-Pro:Dairy	-2.142857e-01	-2.697385	2.268814	0.9998091
Control:Plant-Con:Plant	-4.285714e-01	-2.748280	1.891137	0.9926666


```

Pro:Plant-Con:Plant      -9.285714e-01 -3.411671 1.554528 0.8605529
Pro:Plant-Control:Plant  -5.000000e-01 -3.157560 2.157560 0.9920274

```

Tukey HSD is the last pairwise comparison test that we will look at. Its level of strictness is somewhere in between Fishers LSD and Bonferonni. Similarly, when looking at our pairwise comparisons, we see that all of the p-values are larger than 0.05. We do not have statistically significant evidence that the pairwise comparisons between the treatments are different using Tukey HSD.

Descriptive Statistics & Effect Size:

```

# A tibble: 6 x 6
  treatment      group_mean group_n group_sd group_min group_max
  <fct>          <dbl>    <int>    <dbl>    <int>    <int>
1 Dairy_control    2.33        6     1.75        1        5
2 DairyCon         1.5         6     0.837        1        3
3 DairyPro         1.71        7     1.25        1        4
4 Plant_control    2           5     1           1        3
5 PlantCon         2.43        7     1.62        1        5
6 PlantPro         1.5         4     0.577        1        2

```

```
[1] 1.6888
```

```
[1] 1.299538
```

```
[1] 0.1615958
```

The output above shows us descriptive statistics of our milk data, and we can use it to calculate our effect size. To calculate effect size, we use the DairyPro (1.71) and DairyCon (1.5). This means $|1.71 - 1.5|$ are divided by $\text{rad}(1.6888) = 0.16$. This represents a trivial effect based on Cohen's table.

R-Squared:

Here, we see our different R squared values. An R squared value tells us the proportion of the variance in the dependent variable that can be explained by our independent variable. It tells us how well our data fits the model.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ad_type	2	1.70	0.8519	0.504	0.609
milk_type	1	0.28	0.2809	0.166	0.686
ad_type:milk_type	2	2.92	1.4624	0.866	0.431
Residuals	29	48.98	1.6888		

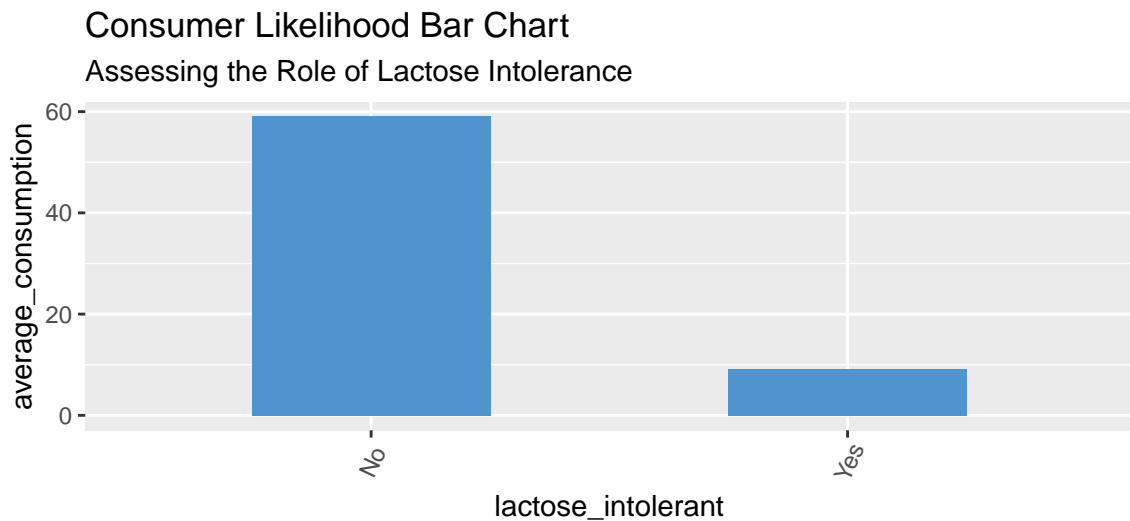
[1] 0.09094284

So for our model, the ad type, milk type, and interaction explain 9.09 % of the variability in the participants average consumption of dairy or plant based products.

Additional Analyses

Lactose Intolerant:

The bar chart demonstrates that there were more participants that are not lactose intolerant, than lactose intolerant. Therefore, although we accounted for this confounding variable pre-survey, the responses indicate that being lactose intolerant did not significantly impact our responses due less than half of students identifying with this.



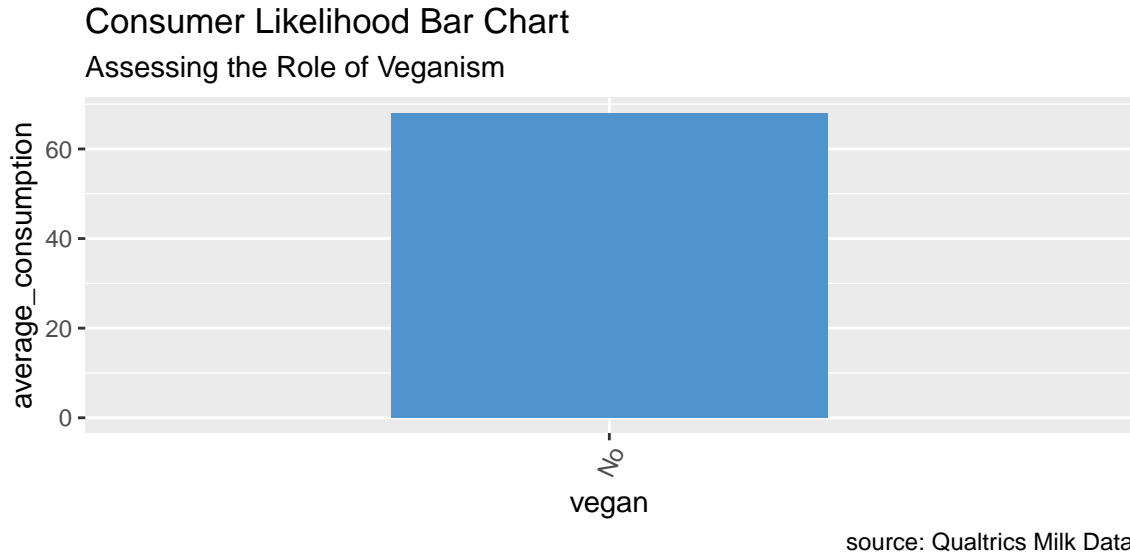
source: Qualtrics Milk Data

Vegan or Not:

The second bar chart shows that all students that responded to the survey were not vegan, so this is not a significant confounding variable impacting our results.

Conclusion

Throughout this experimental design, we investigated how different ad contents (dairy vs oat milk) and sentiment types (positive, negative, and neutral ads) influence people's consumption of dairy or plant-based products. We hypothesized that there would be at least one different mean rate of participant's consumption on the 1-5 Likert scale based on the ad content and



milk type. We supported the alternative hypothesis because in the NYT article that stated that Gen Z has the misconception that dairy milk is unhealthy (Severson, 2023).

We ran a two-way ANOVA via an interaction model to assess if there are main effects for the ad type and milk type, and whether there was an interaction between the impacts of both experimental factors on the mean rate of participant's consumption of the milk product. We do not have statistically significant evidence that mean rate of participant's consumption differs based on neither ad type nor milk type. Therefore both experimental factors do not have a main effect.

The interaction model demonstrates that we do not have statistically significant evidence that the effect of ad type on the mean rate of participant's consumption is different based on the effect of milk type, and vice versa. Additionally, the effect of milk type on the mean rate of participant's consumption is not different based on the effect of ad type. We do not have statistically significant evidence that an interaction exists between the ad type and milk type.

However, there are numerous reasons why we cannot accept these results as conclusive take-aways. Firstly, our data violates the constant variance and normality condition for the ANOVA model. Another limitation of our results is that we have yet to collect our data via random sampling to generalize our results.