HW2

Son Luu (id), Xihan Qian (54285556) and Javier Mtz.-Rdz. (94785938)

March 1, 2024

1 Introduction

Zero-calorie beverages offer an alternative to sugary drinks that can help to avoid the harmful effects of artificial sweeteners. Therefore, it is important to understand what actions can motivate people to switch to these products. The aim of this study is to assess the impact of messaging and discounts on the buying behaviour of zero-calorie and sugary drinks. In particular, it evaluates different interventions, such as discounts with and without explanation, messaging that displays the caloric content, messaging that shows the equivalent fiscal activity, and a combination of both.

The primary question under consideration is the effect of each intervention on the consumption of sugary and zero-calorie drinks. Specifically, the study will explore the direction, size, comparison, and impact of interventions on each site, as well as how different interventions interact and compare with each other. To answer these questions, this document discusses the characteristics of the data collected, explores its behaviour, and performs a statistical assessment.

2 Data Description and Summaries

To evaluate the effects of interventions, the study gathered data on beverages sold at four cafeterias and three convenience stores across three sites. The dataset records daily sales of these beverages for a period of 221 days, from October 27 to May 23 (#TODO: corroborate dates), and summarizes the data by site. Nevertheless, the observations for each site start on different dates: site A starts on day 1, site B on day 14, and site C on day 20. In total, there are 631 observations in the dataset.

The dataset includes variables related to time, sales, place, and intervention. The time variables are the count of days since the start of the study and the day of the week. The sales variables include zero-calorie, sugary, 100% juice, orange juice, sports, and total beverages sold, but only zero-calorie and sugary beverages are considered for this analysis. As for place and intervention, there is a variable for each one. Table 1 summarizes the variables available in the dataset, their classification, and how they are measured.

Table 1: **Description of variables**

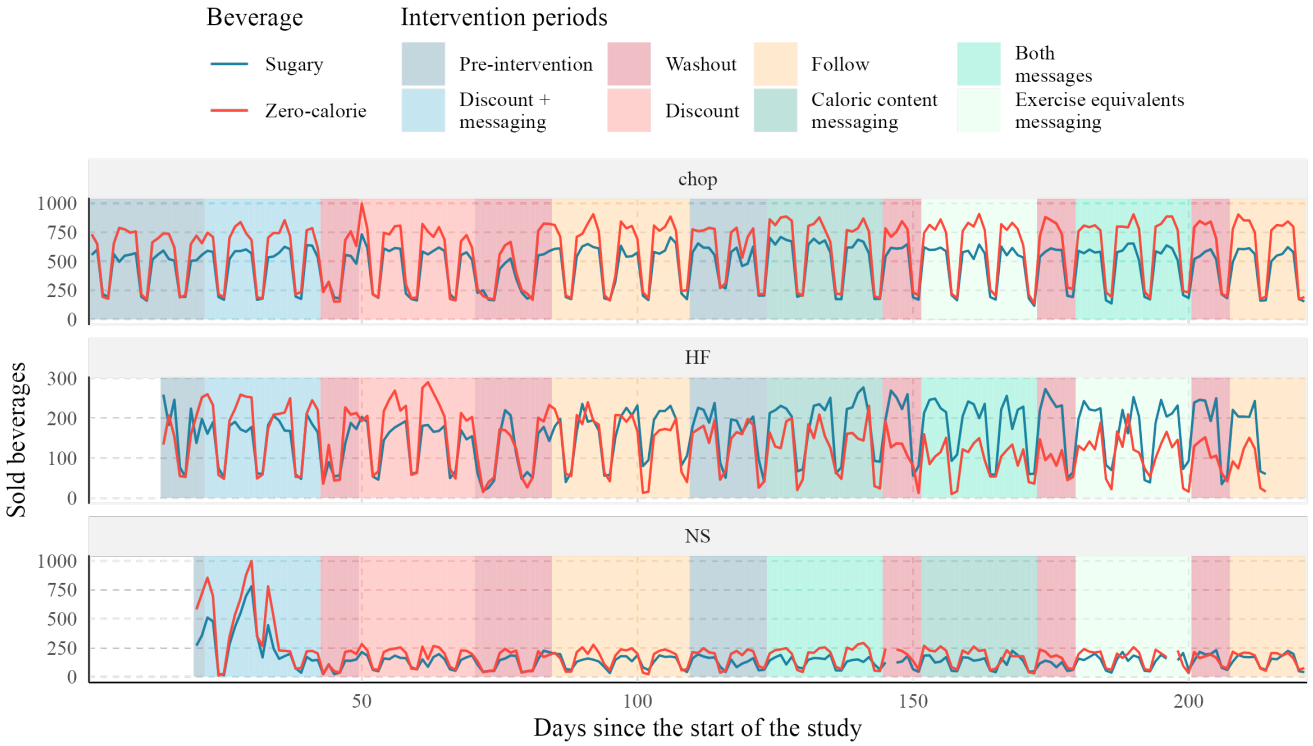
Variable	Type	Unit
Day of the quasi-experiment	Continuous	-
Day of the week	Continuous	-
Site	Categorical	-
Intervention	Categorical	-
Sugary beverages sold	Continuous	-
Zero-calorie beverages sold	Continuous	-
Other beverages sold	Continuous	-

In addition to the periods that were not recorded at the beginning of the study in sites B and C, there are nine missing values for sales of zero-calorie and sugary beverages. The missing observations correspond to the last week of site B and two days of site C. Aside from the missing information at the beginning and end of the study, the missing values are unrelated to any specific

factor. Furthermore, the sales data for other beverages and the total amount have several missing values, but they do not affect this analysis.

3 Exploratory Analysis

Given that the data consists of a time series of sales across three sites, it was necessary to carry out a time-based analysis. In that sense, Figure 1 helps visualize the beverages sold and the shadows behind the lines display the distinct intervention periods. Additionally, Section 7.1 visualization and tests about the relationship among variables, the effect from past observations in the new data points and the decomposition of sugary and zero-calorie sales series in the change by the mean level (trend), the periodicity of the data (seasonal variation) and factors that do not show a pattern (random variation).



Source: client submission.

Figure 1: **Sale of sugary and zero-calorie drinks by intervention**

In particular, Figure 1 shows some important characteristics of the dataset. Firstly, the measurements for each site began at different times. Secondly, the third site experienced a significant increase in sales during most of the first intervention, but afterwards, sales remained at a lower and more stable level. Thirdly, the order of the three calorie messaging interventions was different for each site. Lastly, it is evident from the data that there is a weekly seasonal effect.

4 Formal analysis

4.1 ANOVA/ANCOVA

ANOVA (Analysis of Variance) is a statistical method used to compare the means of three or more samples, determining if at least one sample mean significantly differs from the others. It's instrumental in experiments aimed at evaluating the effectiveness of different interventions, such as those designed to encourage the selection of zero-calorie beverages over sugary options in hospital settings. By examining the variance within and between intervention groups, ANOVA facilitates understanding whether observed differences in beverage consumption are attributable to the interventions or occur by chance.

This approach offers the advantage of simplicity and clear interpretation, allowing researchers to assess the relative effectiveness of each intervention. However, ANOVA's limitation lies in its inability to account for external variables that could influence outcomes, such as baseline consumption patterns or hospital site characteristics. Unlike ANCOVA, which adjusts for these covariates, ANOVA assumes all group differences result from the interventions themselves, which could lead to oversimplified conclusions. Despite this, ANOVA remains valuable for initial analysis, with its limitations highlighting the need for a comprehensive approach, possibly incorporating ANCOVA, to fully understand the interventions' impacts on beverage selection.

5 Conclusion

The study employs ANOVA/ANCOVA and LME models to analyze the effectiveness of interventions on beverage sales across hospitals, addressing four research questions. ANOVA models initially identify differences in intervention effects, while LME models are recommended for deeper analysis when significant variations are observed. This approach ensures a thorough investigation into how each intervention influences beverage choices, the impact of site characteristics, and the effectiveness of calorie information and exercise prompts. The study's methodology provides a clear path to understanding which interventions work best, promoting healthier beverage consumptions.

6 References

7 Appendices

7.1 Detailed Explorarory Analysis

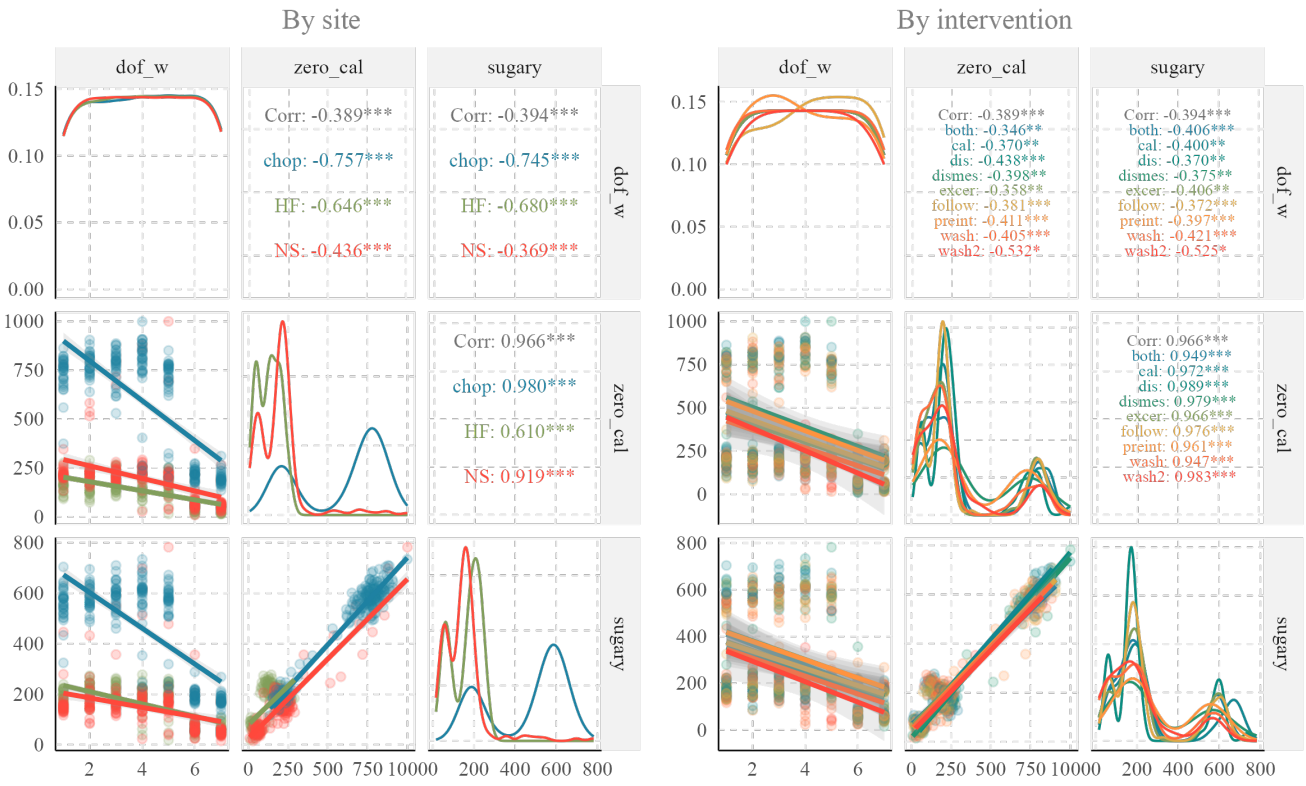
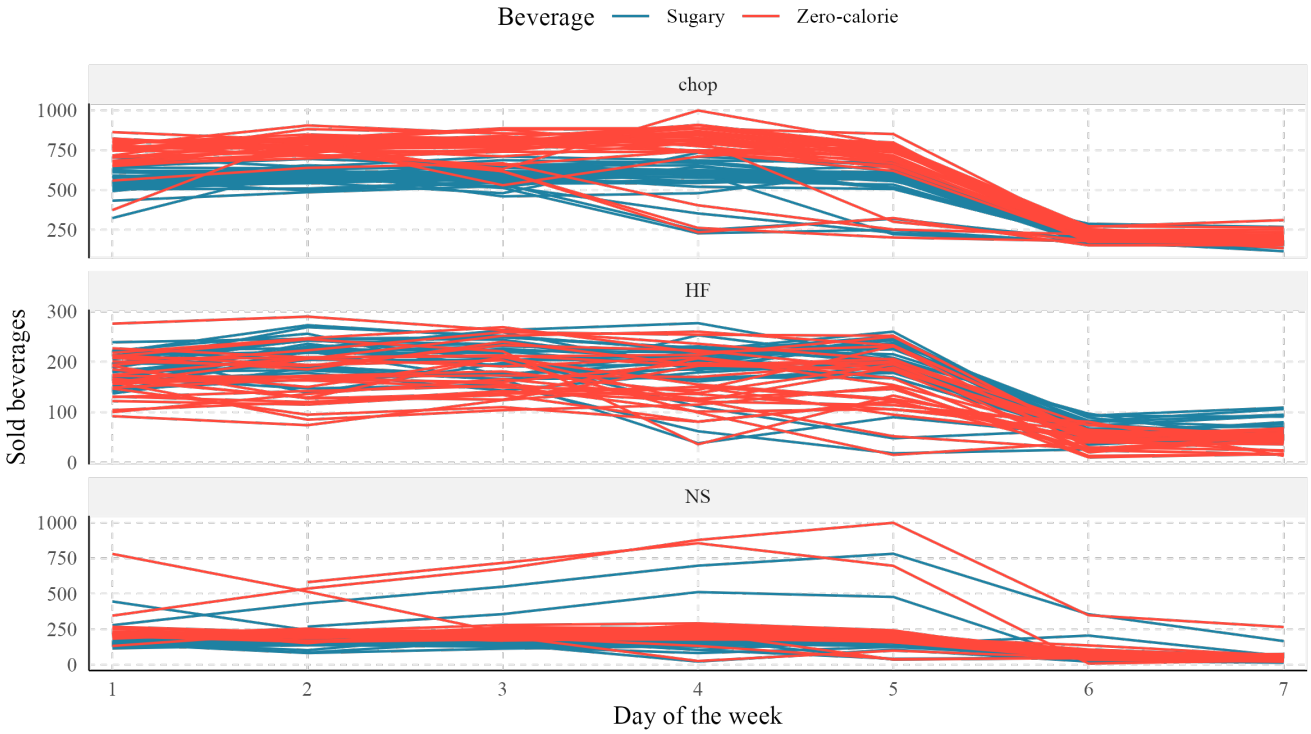


Figure A.1: **Correlation matrix of variables**



Source: client submission.

Figure A.2: **Sale of sugary and zero-calorie drinks by week and site**

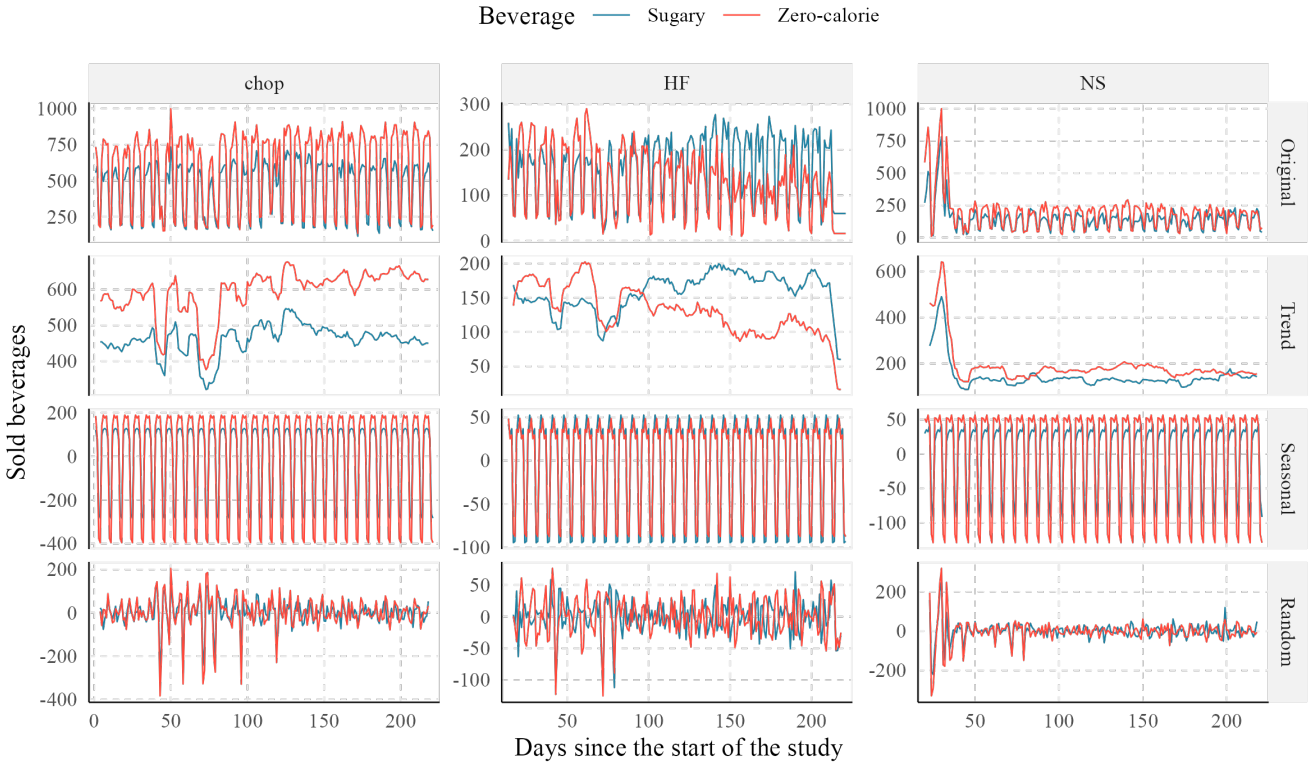


Figure A.3: **Decomposition Analysis of Sales for Sugary and Zero-Calorie Beverages**

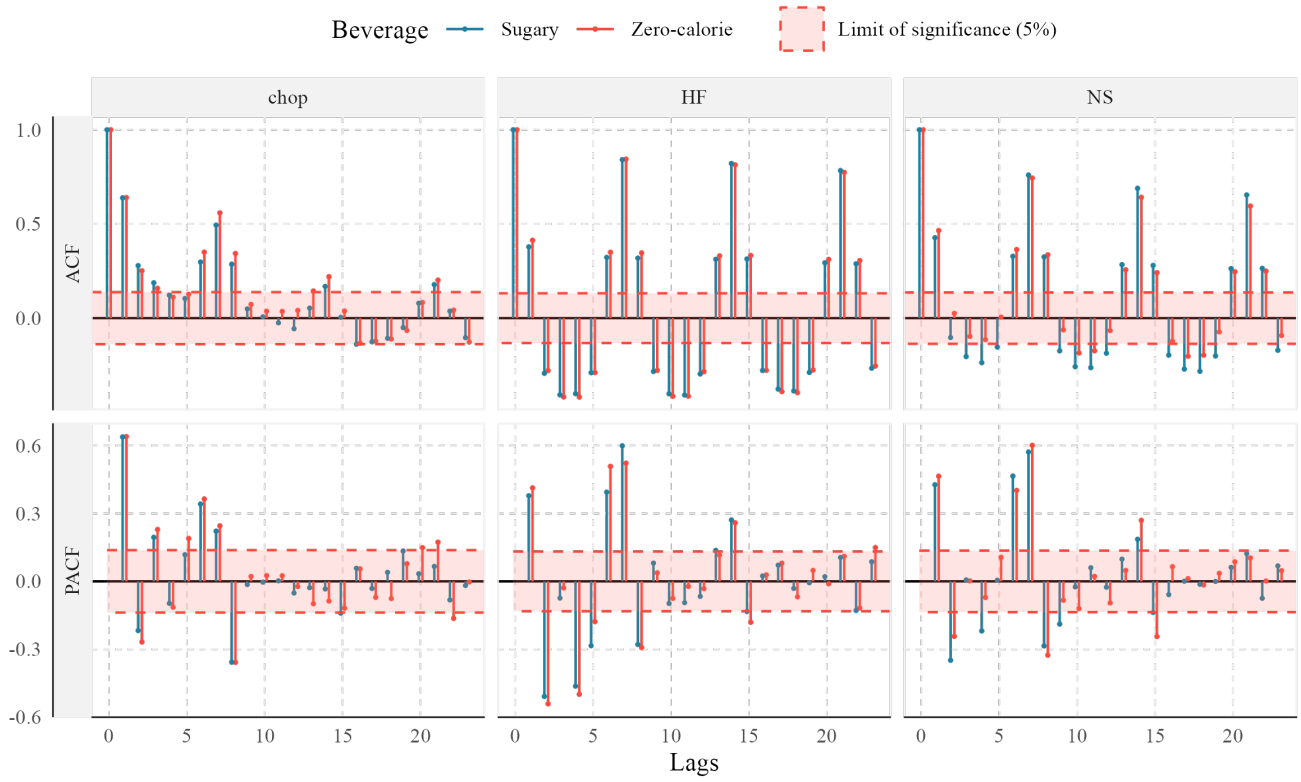


Figure A.4: **ACF and PACF by Beverage and Site**