STAT5350 Assigment 7 Written Report

Statenfreude

Introduction

The goal of this survey was to determine which new shopping options are preferred by people in Laramie, Wyoming: Target, Costco, Whole Foods, and Trader Joe's. Our survey comprised three items which were intended to gauge respondent sentiment through ranking of options, rating of store brand characteristics, and monetary allocation decisions.

Questions

The questions used in the survey were:

Question 1: *Please rank the following stores in order of your preference (1-4) for a new store in Laramie.*

This was a forced ranking item with 4 rank options. Lower values indicate higher ranking.

Question 2: *Please rank the following characteristics of each store: price, quality, and variety.*

This item used a multiple rating scale asking respondents to rate the 4 store brand characteristics on a scale of 1 to 5 where higher ratings are better.

Question 3: If you had $100 to spend at the following stores, how much would you spend at each? (total must be $100)

This item used a fixed sum scale.

A priori hypotheses

The 4 hypotheses explored were:

1) For question 1, we expected Target and Costco to be rated the highest out of the four options.
2) There was an alternate hypothesis that Trader Joe's and Whole Foods might be more popular with students given that students are probably most interested in food options rather than general shopping.
3) Target has been experiencing push back on DEI business decisions and that may affect its ranking.
4) For question 2, we expected that the mean (combined price, quality, and variety scores) of overall approval of Whole Foods will be higher for individuals who rated Target higher than Costco in question 1. This was based on the impression that ranking Target above Costco implies cost may be less important to ranking, which would diminish one of the most salient cons of Whole Foods (price).

Methods

Focus group

We conducted a focus group with two teaching assistants and one university staff member. We selected this group because they were readily available and willing to provide quick feedback on our survey design. The focus group was conducted in the Center for Assistance with Statistics and Mathematics (CASM) and lasted about five minutes.

The focus group helped to quickly evaluate the clarity and effectiveness of our survey questions. Group discussion allowed us to identify areas where respondents might become confused, misinterpret instructions, or feel limited by the response options.

We looked primarily at whether participants could easily complete the ranking in question 1 without confusion, whether the rating scale in question 2 (price, quality, variety) captured meaningful differences across stores, and whether the fixed-sum allocation in question 3 (splitting $100) felt natural or confusing for respondents.

Overall, participants seemed comfortable with the survey questions. For question 1 (Forced Ranking), we assumed that "1" would represent the highest rating, rather than the lowest. We did not clarify this during the session, which highlights how easily both researchers and respondents can overlook important scale instructions. Participants also noted that some respondents might be interested in other store brands beyond the four we listed. This suggests that including an option such as "Other" could make the survey more inclusive. For the remaining questions, no major issues were raised, and participants indicated they were clear and straightforward.

Data Collection

Surveys were printed and data were collected in person. The sample was a convenience sample consisting of data collected at the University of Wyoming student union during a Friday lunch hour (37 of 40 respondents) and 3 family members sampled off campus. Of the 40 surveys collected, question 1 was filled out incorrectly in 3 of them and these data were recorded as NA as a result.

We did not capture demographic data, but since nearly all of our sampling was conducted in the student union, we expect that most of the survey respondents are very likely students at the university.

Data were first recorded in Google Sheets in wide format to assist with data entry and then imported in R for analysis in both long and wide formats.

Results

Question 1

Question 1 asked respondents to "rank the following stores in order of your preference (1-4) for a new store in Laramie." This question only had 37 usable responses.

The highest ranked store was Target, followed by Costco. The mean rankings and other descriptive statistics are listed in the table below.

Table 1: Descriptive statistics for store rankings, $n = 37$.

Store	Avg	Std dev	Min	Max	Median
Target	2.1	0.98	1	4	2
Whole Foods	3.1	1.03	1	4	3
Costco	2.3	1.05	1	4	2
Trader Joe's	2.5	1.19	1	4	2

As previously mentioned, this claim relies on the assumption that 1 on the scale represents highest preference and 4 represents lowest.

Question 2

The second survey question asked respondents to rate each store on 3 characteristics: price, quality, and variety. Each characteristic is measured on a multiple rating scale from 1 (poor) to 5 (excellent). The average characteristic score for each store is presented in the heat map below.

Figure 1: Heatmap of store characterics. Darker colours indicate higher rankings. The sample size for this question is $n = 40$.

Question 3

Question 3 asked respondents to answer "If you had \$100 to spend at the following stores, how much would you spend at each? (total must be \$100)" Descriptive statistics of the results are shown in the following table:

Table 2: Descriptive statistics for amount spent at each store, $n = 40$.

Store	Avg	Std dev	Min	Max	Median
Target	30.2	26.70	0	100	25
Whole Foods	13.6	21.96	0	100	0
Costco	35.0	29.64	0	100	30

Store	Avg	Std dev	Min	Max	Median
Trader Joe's	20.0	24.71	0	100	15

Discussion

Hypotheses 1-3:

In line with our first hypothesis, Target and Costco received the highest average ranking out of the four stores. The highest overall was Target with a mean ranking of 2.1. This is counter to the second hypothesis that Whole Foods and Trader Joe's would receive the highest ranking due to their focus on food and the habits of university students. It also may counter the third hypothesis, which proposed that Target may have a lower ranking due to recent negative publicity surrounding issues with the company's DEI policies. However, the survey was not designed to explore this question, so no firm conclusions cannot be drawn from this result.

The rankings of the four stores are presented in the box plot below.
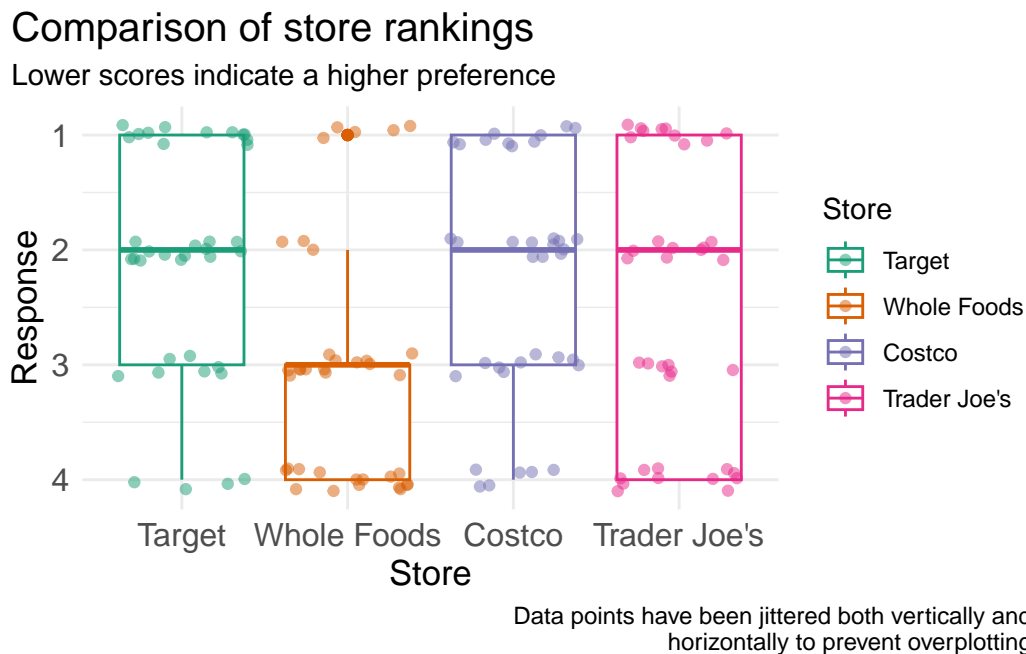

Figure 2: Comparison of store brand rankings. Trader Joe's exhibits a very wide distribution of rankings among respondents while Whole Foods have a very narrow range. Target and Costco show very similar patterns of rankings among respondents.

Hypothesis 4

For respondents who ranked Target above Costco, the aggregate rating score for Whole Foods was 2.95. For respondents who ranked Costco above Target, the aggregate rating score for Whole Foods was 3.32. This does not support the fourth hypothesis, which proposed that individuals who ranked Target above Costco in question 1 may rank Whole Foods higher in question 2 than individuals who ranked Costco above Target in question 1.

Other observations:

There was very little variability in quality ratings in question 2. This may indicate that quality is not a reliable differentiator of subject store brand preference. In a future survey design, this could be discussed in focus groups and potentially tested in a pilot study.

Contrary to the results of question 1, in question 3 Costco had the highest average amount spent (mean = \$35, median = \$30), while Whole Foods had the lowest (mean = \$13.60, median = \$0).

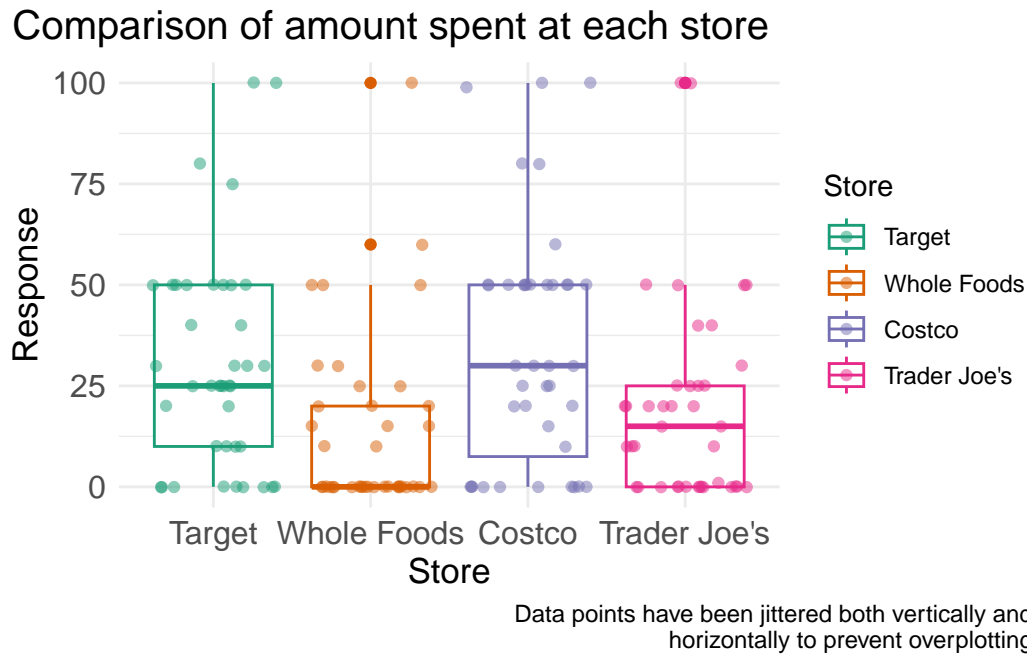The respective amounts spent at each store are displayed in the box plot below.

Figure 3: Comparison of amount spent on each store. Whole Foods has the most distinct distribution, with half of the respondents allocating \$0 to the store. In contrast, Target and Costco show similar high allocations from respondents for the hypothetical \$100. The wide range of responses for all stores indicates respondents often allocated the full amount to only one or two stores.

Additionally, item 3 had some interesting responses, as respondents seemed to prefer extreme

splits over more even distributions of the proposed $100. For example, many respondents chose to allocate $50 each to two stores or $100 to one store as opposed to distributing the $100 across all four. This may be because $100 is not quite enough to spend across multiple stores, or because the stores are similar enough that one or two stores are sufficient and there may not be need to go to all four.

One of the most interesting phenomena in this survey was the disagreement between responses for item 1 and items 2 and 3. Since they were intended to measure the same thing, we had expected that the highest ranking store in question 1 would have the highest rated characteristics in question 2 and the highest amount of money allocated in question 3. However, while Costco had the highest performance in both questions 2 and 3, Target had the highest ranking in question 1. A few things could account for this. First of all, transitivity issues can always be at play in a survey, and will result in illogical responses across items. Alternatively, the aforementioned scale use ambiguity of question 1 may have clouded the results of the question. Perhaps if the item 1 instructions were disambiguated, and the survey re-administered, the results of all three items would agree. Another potential source of this disagreement is the order of the items. Since item 2 prompts respondents to think more in-depth about specific characteristics of each store, a more thoughtful answer may disagree with a faster, first-impression based answer.

Challenges & Limitations

The most severe issues occurred with question 1, namely unexpected missingness and scale use ambiguity. Missingness occurred here via respondents misunderstanding what the item was asking. A small handful circled a single store instead of providing a ranking out of 4 for each. These responses were coded as `NA`s.

Additionally, the instructions for using the question 1 scale were ambiguous, which invalidates the results of the question. This item asked respondents to "…rank the following stores in order of [their] preference (1-4)…" Without clarification of whether 1 signifies the highest or lowest preference, the responses to this question are invalid. We did intuitively assume that 1 would be interpreted as the highest preference for analysis' sake.

An issue throughout question 1 and 2 (but particularly 2) was inapplicability. Several respondents indicated that they had never been to some of the stores, or that they had not been frequently enough to gauge the usual price, quality, or variety. This resulted in one respondent giving NA responses, and murky interpretability around the 'No opinion' response option. (I.e., do 'No opinion' responses mean the respondent feels neutral about the store, or that they do not know because it does not apply to them?)

An issue underlying the whole of the survey was an undefined population. In other words, we did not go into survey collection with a target population in mind, and sampled solely by convenience. As aforementioned, the majority of our respondents were people we approached in the student union and asked to complete a short survey. Since we had neither a target

population we were hoping to generalize our results to, nor demographic information about our respondents, our conclusions and hypotheses can only be general.

Conclusions

In line with our initial hypothesis for question 1, Target received the highest ranking across the four stores. However, there was some dissonance between this result and the results of question 2, which showed that Costco ranked the highest for each of price, quality, and variety.

Contrary to our second hypothesis, ranking Target above Costco in question 1 was not associated with a higher overall approval of Whole Foods. With price assumingly out of the question, it is unclear what may explain the opposite association observed in the data. Collection of demographic data and other opinions on store characteristics could help explain this.

These results are tentative due to the small sample size, undefined target population, and ambiguity in the question 1 wording.

Appendix

All R code used to clean, transform and visualise the data in this report are provided in this appendix.

Descriptive statistics for question 1 were generated with the following code.

```
df_proc |>
  filter(question == 1) |>
  group_by(brand) |>
  summarise(
    avg = mean(response, na.rm=TRUE),
    stderr = sd(response, na.rm=TRUE),
    min = min(response, na.rm=TRUE),
    max = max(response, na.rm=TRUE),
    md = median(response, na.rm=TRUE)
  ) |> kable(
    caption="Descriptive statistics for store rankings, $n=37$.",
    digits = c(0,1,2,0,0,1),
    col.names=c("Store", "Avg", "Std dev", "Min", "Max", "Median")
  )
```

Heat map for question 2.

```
df_proc |>
  filter(question == 2 & !is.na(response)) |>
  group_by(brand, category) |>
  summarise(
    avg = mean(response) |> round(2),
    stderr = sd(response) |> round(2),
    min = min(response) |> round(2),
    max = max(response) |> round(2)
  ) |> ggplot(aes(brand, category, fill= avg)) +
  geom_tile(colour="white") +
  geom_text(aes(label=avg), color="white") +
  scale_fill_viridis_b(direction=-1) +
  coord_flip() +
  labs(
    x="Store",
    y="Characteristic",
    fill="Avg Score"
  )
```

Box plot of question 1 data.

```
df_proc |>
  filter(question == 1 & !is.na(response)) |>
  ggplot(aes(x=brand, y=response, colour=brand)) +
  geom_boxplot() +
  geom_jitter(alpha=0.5, height=0.1, width=0.4) +
  scale_colour_brewer(palette="Dark2") +
  scale_y_reverse() +
  labs(
    y="Response",
    x="Store",
    colour="Store",
    title="Comparison of store rankings",
    subtitle="Lower scores indicate a higher preference",
    caption="Data points have been jittered both vertically and
            horizontally to prevent overplotting."
  )
```

The boxplot for question 3 fixed sum data was created with the following code.

```
df_proc |>
  filter(question == 3 & !is.na(response)) |>
  ggplot(aes(x=brand, y=response, colour=brand)) +
  geom_boxplot() +
  geom_jitter(alpha=0.5, height=0.1, width=0.4) +
  scale_colour_brewer(palette="Dark2") +
  labs(
    y="Response",
    x="Store",
    colour="Store",
    title="Comparison of amount spent at each store",
    caption="Data points have been jittered both vertically and
            horizontally to prevent overplotting."
  )
```

The code to examine hypothesis 4 included below.

```
df_pref_tgt <- df_wide |> filter(target_1 > costco_1)

wf_mean_tgt <- df_proc |> filter(question==2 & brand=="Whole Foods" &
  !is.na(response)) |>
  select(respondent, response) |>
  group_by(respondent) |>
  summarise(
    agg_response = mean(response)
  ) |> inner_join(df_pref_tgt, by="respondent") |>
  select(respondent, agg_response) |>
  summarise(
    wf_pref_avg = mean(agg_response)
  ) |> pull() |> round(2)

wf_mean_not_tgt <- df_proc |> filter(question==2 & brand=="Whole Foods" &
  !is.na(response)) |>
  select(respondent, response) |>
  group_by(respondent) |>
  summarise(
    agg_response = mean(response)
  ) |> anti_join(df_pref_tgt, by="respondent") |>
  select(respondent, agg_response) |>
  summarise(
    wf_pref_avg = mean(agg_response)
  ) |> pull() |> round(2)
```

The data cleaning and processing pipeline code for the survey data is provided below.

```r
library(tidyverse); library(janitor)
df <- read_csv("complete-data.csv") |> janitor::clean_names()

# perform transformation and field engineering.
df_processed <- df |>
  pivot_longer(cols=c(everything(), -respondent),
               names_to="category", values_to="response") |>
  mutate(
    question = case_when(
                 str_detect(category,"_1") ~ "1",
                 str_detect(category,"_2") ~ "2",
                 str_detect(category,"_3") ~ "3"),
    category = str_replace(category, "_\\d", ""),
    brand = case_when(
                 str_detect(category, "target") ~ "Target",
                 str_detect(category, "whole_foods") ~ "Whole Foods",
                 str_detect(category, "costco") ~ "Costco",
                 str_detect(category, "trader_joes") ~ "Trader Joe's",
    ),
    category = case_when(
                 str_detect(category,"price") ~ "Price",
                 str_detect(category,"quality") ~ "Quality",
                 str_detect(category,"variety") ~ "Variety",
                 .default=NA
    )
  ) |>
  select(respondent,question,brand,category,response) |>
  mutate(
    question = as_factor(question),
    brand = as_factor(brand)
  )

write_rds(df, "wide-data.rds")
write_rds(df_processed, "df_proc.rds")
```