

My title*

My subtitle if needed

First author

Another author

November 25, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023).... Our data (Toronto Shelter & Support Services 2024).... Following Alexander (2023), we consider...

Overview text

*Code and data are available at: https://github.com/RohanAlexander/starter_folder.

2.2 Measurement

The measurement process involves transforming real-world phenomena into data that can be analyzed. In this case, the real-world phenomenon is the sale of potatoes by various vendors over time. Each transaction recorded in the raw dataset represents an instance of this phenomenon.

To transition from real-world sales events to a structured dataset, we began by recording each transaction, which included details like the product name and brand, prices, and the vendor information. This research focus on the single product-potato. Therefore, `product_name` was used to identify the specific types of potatoes (yellow and white), allowing us to focus on these items exclusively. The `current_price` and `old_price` fields represent the financial aspect of these sales, capturing both the price at the time of the transaction and the historical price for comparison. The `month` variable was derived from the transaction timestamp (`nowtime`) to provide a temporal context for analysis, allowing us to observe trends and patterns over different periods.

By filtering and cleaning the raw data, the dataset reflects the aspects of potato sales that are of interest for this study. This process of measurement transforms abstract sales activities into quantifiable data points that can be used for statistical analysis, providing insights into vendor behavior and price dynamics over time.

2.3 Cleaning Process

The dataset used for this analysis comes from a combination of raw sales data and product-specific information, sourced from Project Hammer (<https://jacobfilipp.com/hammer/>). Project Hammer aims to drive more competition and reduce collusion in the Canadian grocery sector. The dataset, referred to as Canadian Grocery Price Data, includes sales data from eight vendors: Voila, T&T, Loblaws, No Frills, Metro, Galleria, Walmart, and Save-On-Foods. The available dates range from February 28, 2024, to the latest data load.

Initially, two raw datasets were downloaded from the Project Hammer website. These datasets were merged to create a comprehensive dataset containing all relevant columns, including `nowtime`, `vendor`, `product_id`, `product_name`, `brand`, `units`, `current_price`, `old_price`, `price_per_unit`, and `other`.

The cleaning process involved several key steps to refine the dataset for analysis. First, we selected only the columns relevant to our study shown in Table 1. The month of each transaction was extracted from the `nowtime` field to provide temporal context for the sales data.

Table 1: Summary of Selected Columns

Column	Description
nowtime	Timestamp indicating when the data was gathered
vendor	One of the 7 grocery vendors
current_price	Price at time of extract
old_price	An ‘old’ struck-out price, indicating a previous sale price
product_name	Product name, may include brand and units

Next, we filtered the dataset to include only the products of interest—yellow and white potatoes—by searching for these keywords in the product_name column, ensuring our analysis was focused specifically on these items. Additionally, any entries with missing values (NA) were removed to maintain data quality. Finally, the nowtime column was dropped after extracting the month, as it was no longer necessary for the analysis. Table 2 gives an preview of the cleaned dataset we will use in the following sections.

Table 2: Preview of Cleaned Data on Potato Sales

vendor	current_price	old_price	product_name	month
Loblaws	4.99	5.99	Small White Potatoes	10
Loblaws	4.99	5.99	Small White Potatoes	10
Loblaws	4.99	5.99	Small White Potatoes	10
Loblaws	4.99	5.99	Small White Potatoes	10
Loblaws	4.50	5.99	Small White Potatoes	10
Loblaws	4.50	5.99	Small White Potatoes	10

These steps resulted in a cleaned dataset containing the essential information needed for the analysis of potato sales trends.

2.4 Outcome variables

The cleaned dataset contains 685 rows, representing individual products sold, along with vendor information and price details. The summary statistics of the cleaned data indicate that the mean of current price is 2.811, and the mean of old price is 3.785. Figure 1 and Table 3 show that the cleaned dataset only contains five vendors instead of the eight in the raw dataset, and the month is from June to November. This is expected because data collection from February 28 to July 10/11 focused on a smaller set of products by the description of the raw dataset. After July, more products were added, and some data may be missing for certain vendors or days when extraction failed.

Table 3: Overview of Analysis Data

current_price	old_price	month
Min. :0.940	Min. :1.490	Min. : 6.000
1st Qu.:2.000	1st Qu.:2.490	1st Qu.: 8.000
Median :2.290	Median :2.490	Median :10.000
Mean :2.811	Mean :3.785	Mean : 9.064
3rd Qu.:3.990	3rd Qu.:5.990	3rd Qu.:10.000
Max. :7.990	Max. :8.990	Max. :11.000

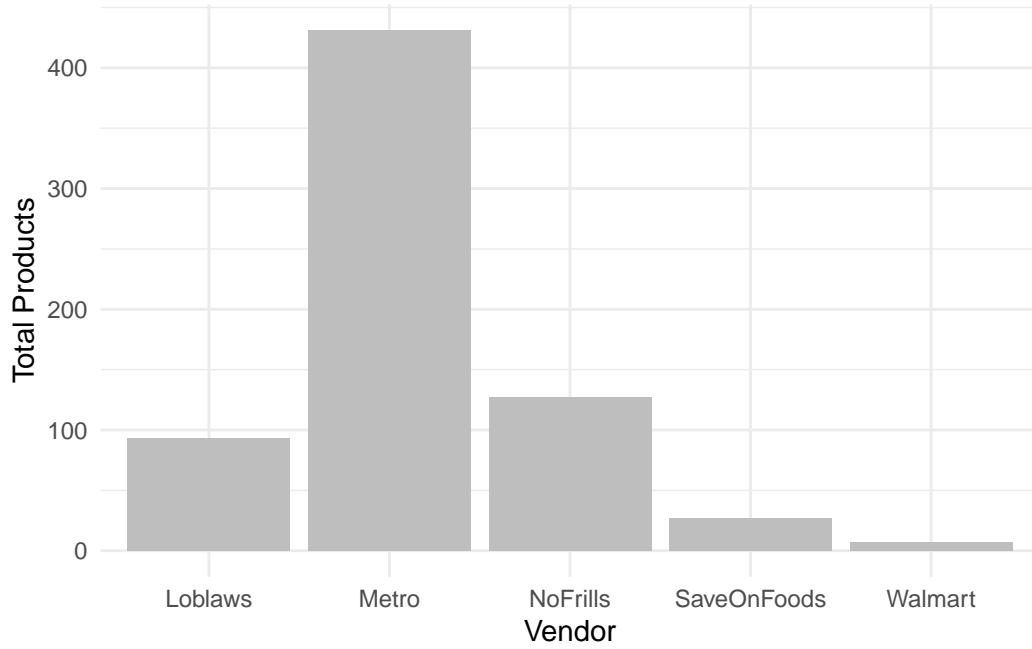


Figure 1: Total Number of Products by Vendor

3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix [B](#).

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$

$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$

$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ .

4 Results

Our results are summarized in `?@tbl-modelresults`.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com/>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “rstanarm: Bayesian applied regression modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. <https://open.toronto.ca/dataset/deaths-of-shelter-residents/>.