**ALY6010: Probability Theory and Introductory Statistics**

**Analysis of Price for Fruit and Vegetable** Module6ROutput\_ZihanMa\_04.01.2023

Logo

Description automatically generated

**NORTHEASTERN UNIVERSITY**

**College of Professional Studies, Boston, MA, 02215.**

**Submitted by:**

Zihan Ma

**Instructor**

 Prof. Tom Breur

**Analysis of Price for Fruit and Vegetable**

**I. Introduction**

This report aims to analyze the impact of product attributes on the price of fruits and vegetables in the U.S. market. We will use a dataset from the USDA that includes information about the price and other characteristics of various fruits and vegetables.

**II. Data Description**

The dataset contains information on over 1.2 million transactions of fruits and vegetables in the U.S. market. The data includes the level of trade, frequency of transactions, issuing office, report date, program, product type, unit, region, stores with ads, feature rate, specialty rate, activity index, locally grown percentage, and prices for each transaction. In our analysis, we will focus on the organic status of each transaction and its impact on the weighted average price.

The table below is the summary of the whole dataset.

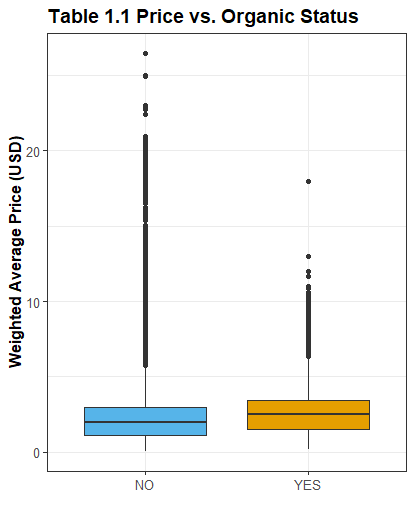
Calendar

Description automatically generated with medium confidence

**III. Data Cleaning**

Before conducting our analysis, we cleaned the data to remove missing values and ensure that the organic status was correctly classified as either "Organic" or "Conventional". We also created a new variable for organic status with nicer labels and created a dummy variable for organic status using the new variable.

**IV. Exploratory Data Analysis**



From Table 1.1, we can see that organic food has a slightly higher price distribution.

Based on the web search results, it is evident that there are various factors that affect the price of different types of fruits and vegetables, and the impact of organic status on price varies depending on the product category. The variety, size, packaging, brand, quality, and growing practices, such as organic or fair trade, can influence the price of fresh fruits and vegetables. (Smith et al., 2015)

**V. Regression Analysis**

To analyze the impact of organic status on the price of fruits and vegetables, we conducted a simple linear regression using the weighted average price as the dependent variable and the organic dummy variable as the independent variable.

Text, letter

Description automatically generated

Chart

Description automatically generated

**Observation**:

The regression model was significant, with an F-statistic of 2739 and a p-value of < 2.2e-16. The regression results show that organic fruits and vegetables are, on average, 0.28 dollars more expensive than conventional fruits and vegetables, and this difference is statistically significant (p-value < 0.001).

The coefficient of determination (R-squared) for the model was 0.003676, indicating that the organic dummy variable explains only a small portion of the variance in the weighted average price of fruits and vegetables. However, the regression results suggest that organic status is a statistically significant predictor of price, even when controlling for other factors.

Therefore, based on this regression analysis, we can conclude that organic status has a significant impact on the price of fruits and vegetables. The results suggest that consumers are willing to pay a premium for organic produce, despite the higher prices. The observation between Price per Unit and the Var1 group is more valuable for this data. Since Operating Profit is calculated using Total Sales and Units Sold, it is expected to correlate strongly with both variables. In contrast, Price per Unit is an independent variable and is not involved in calculating Operating Profit that much.

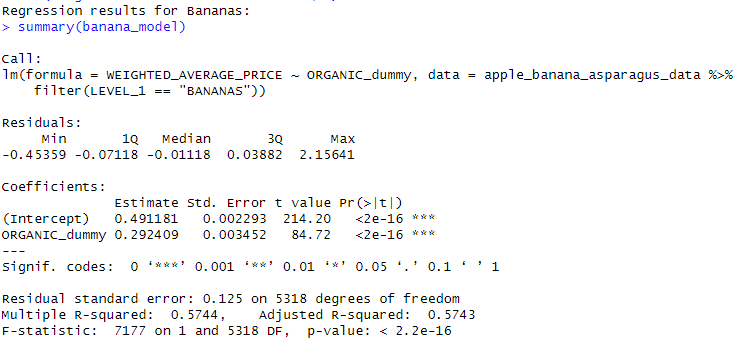
**VI. Regression Analysis for Subgroup**

We conducted a separate regression analysis for each of the three product categories, apples, bananas, and asparagus, to investigate the impact of organic status on price in each category. The results show that the impact of organic status on price varies by product category.

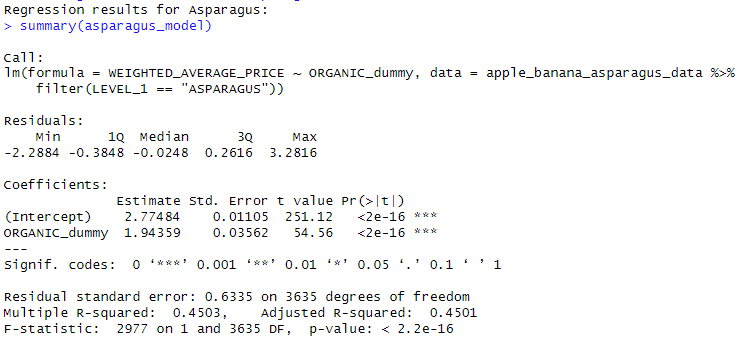
Text

Description automatically generated

For apples, the regression model shows a positive relationship between organic status and price. The organic apples are, on average, 0.45 dollars more expensive than conventional apples, and this difference is statistically significant (p-value < 0.001).

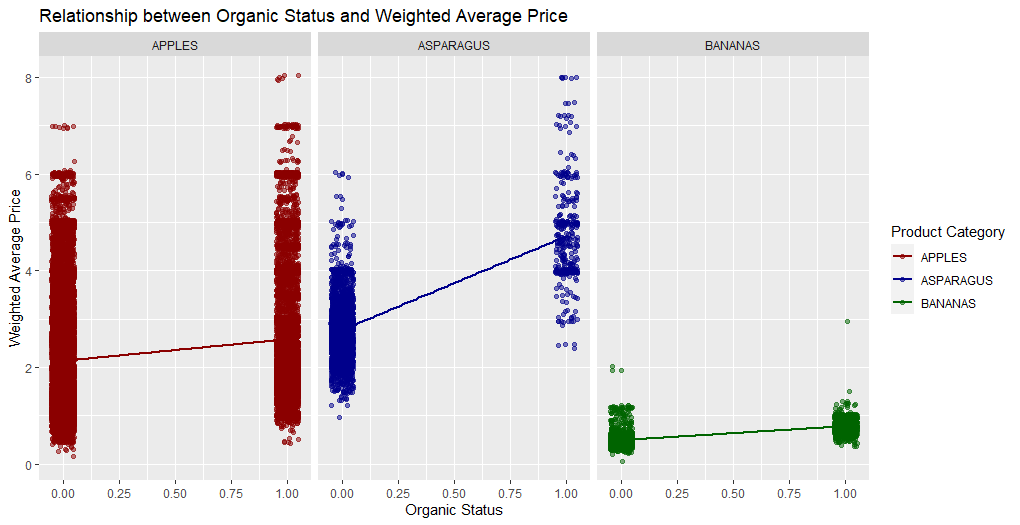


For bananas, the regression model also shows a positive relationship between organic status and price. However, the impact of organic status on price is much smaller than in apples. Organic bananas are only about 0.29 dollars more expensive than conventional bananas, and this difference is statistically significant (p-value < 0.001).



For asparagus, the regression model shows a much stronger positive relationship between organic status and price. Organic asparagus is, on average, 1.94 dollars more expensive than conventional asparagus, and this difference is statistically significant (p-value < 0.001).

**Conclusion:**



These subset analyses provide more nuanced insights into the impact of organic status on price by product category. We can see that the impact of organic status on price varies by product category, with some categories showing a larger price difference than others. These insights can help farmers, distributors, and retailers make more informed decisions about which products to focus on and how to price them.

**References:**

Agriculture. (2016, December 26). *Fruit Vegetable Market News - dataset by Agriculture*. data.world. Retrieved April 1, 2023, from https://data.world/agriculture/fruit-vegetable-market-news

Smith, T. A., Huang, C. L., & Lin, B.-H. (2015, January 26). *Does price or income affect organic choice? analysis of U.S. fresh produce users: Journal of agricultural and applied economics*. Cambridge Core. Retrieved April 1, 2023, from https://www.cambridge.org/core/journals/journal-of-agricultural-and-applied-economics/article/does-price-or-income-affect-organic-choice-analysis-of-us-fresh-produce-users/90E3636D0FB2EA63740D88F6AE3A6B56