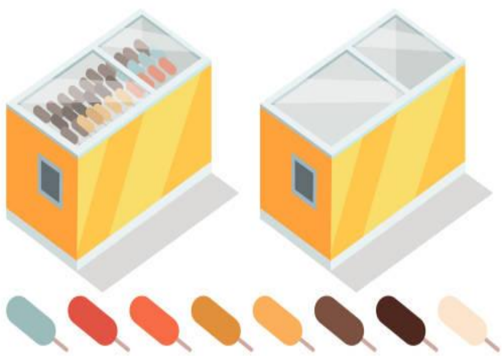

ANALYSIS OF INVENTORY RISKS IN ICE CREAM SALES AT A SUPERMARKET:

Assessing Sales Variability and Risk Management Through Bootstrapping and Simulation Modeling

ABSTRACT

This report analyzes the risk management of ice cream sales in a supermarket setting, using historical data to explore daily sales variability through bootstrapping and simulation modeling. We focus on the extreme risk scenarios at 5% and 1% probability levels to guide inventory decisions. The findings from both methods offer insights into managing inventory risks associated with sales peaks and troughs. Recommendations are provided for dynamic inventory strategies to optimize stock levels, prevent overstocking, and reduce the risk of shortages, thereby enhancing operational efficiency and customer satisfaction.

Wangjun Shen

Customer Analytics in Large Organisations

Content Table

Introduction

Effective inventory management is vital in the highly competitive retail sector, especially for fast-moving consumer goods. The challenge is to balance inventory to satisfy varying customer demands without incurring surplus, which can lead to product wastage and financial losses. This report focuses on the unique challenges of managing ice cream sales due to its perishability and high variability in demand, especially during warmer seasons.

The data used in this analysis comes from a large supermarket chain's transaction records spanning three years, from 2013 to 2015. Each trading day is represented by a separate file, with some days missing due to public holidays. The data, directly sourced from the supermarket's database, includes detailed sales data for the "ICE CREAMS AND ICE CONFECTIONS" category. However, only three variables, "Sale_Date" representing the time and date, "Quantity_Sold" representing the sales volume, and "UniSA_Receipt_No1" identifying the transaction, will be extracted as data for analysis.

The aim of this analysis is to understand market dynamics and consumer purchasing behaviors by looking at daily sales volumes, customer counts, and average purchase quantities per visit. The goal is to use this information to make more accurate forecasts for demand and inventory levels, improving operational efficiency and customer satisfaction.

To achieve these goals, we employ two main analytical methods: bootstrapping and simulation modeling. Bootstrapping is used to simulate potential sales outcomes based on empirical sales data distributions. Conversely, simulation modeling applies fitted probability distributions to predict sales peaks and potential inventory shortfalls at risk levels of 5% and 1%. This risk assessment is crucial for developing robust inventory strategies to prevent overstocking and wastage, supporting the supermarket's efforts to maintain supply adequacy, improve customer satisfaction, and gain a competitive edge in the market.

The aim of this report is to provide actionable insights to help the supermarket optimize its inventory management practices, reduce economic losses, and maximize profitability through data-driven decision-making.

Ice Cream Quantity Sold and Customer Behavior Analysis

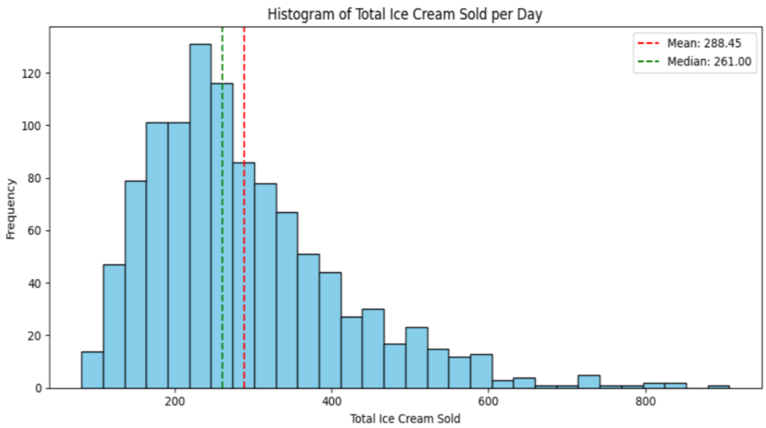

Figure 1: Frequency of Total Ice Sold per Day

Figure 1 illustrates the distribution and basic statistics of the daily quantity of ice cream sold. With an average daily sales volume of 288.45 units and a median of 261 units, the data suggests that while the mean sales volume is high, the typical daily sales are often lower. This discrepancy might be due to exceptionally high sales volumes on certain days, which could skew the average upwards. This is visually apparent in the histogram within Figure 1. The majority of daily ice cream sales volumes fall within the range of 180 to 400 units, indicating this as the most common sales quantity range. However, there exist outliers where daily sales exceed 600 units or fall below 100 units. These deviations from the general trend could potentially be influenced by external factors such as weather conditions or holiday periods.

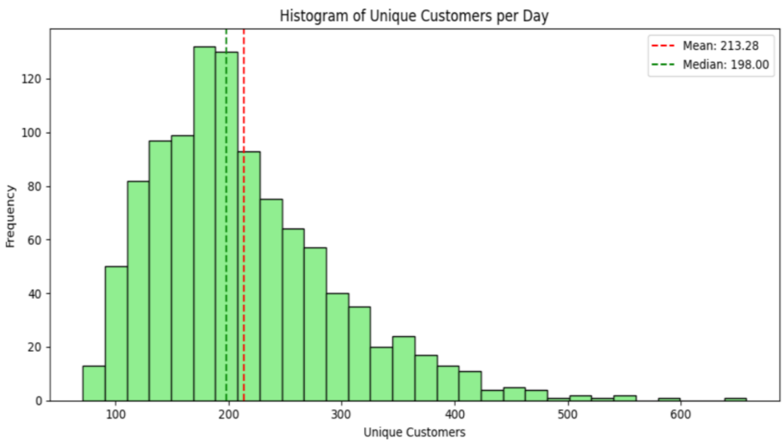

Figure 2: The Distribution of Unique Customers per Day

In the past three years, a total of 202,789 unique customers have purchased Ice Cream in this supermarket. To further understand customer behavior, the distribution of unique customers who purchase ice cream every day can be explored.

Figure 2 presents the distribution of unique daily visitors. Typically, the number of unique customers ranges from 150 to 250, averaging 213.28 with a median of 198. This pattern coincides with the usual daily total of items sold, suggesting a link between customer count and quantity sold on most days. However, there are peak days when customer numbers can surge to 350 or more, which may correspond with high quantities sold, emphasizing the influence of weather and seasonal events on ice cream buying behavior. Like the quantity sold, the histogram of daily unique customers is right-skewed, implying that while most days see fewer ice cream items sold, there are occasional days with significantly more customers.

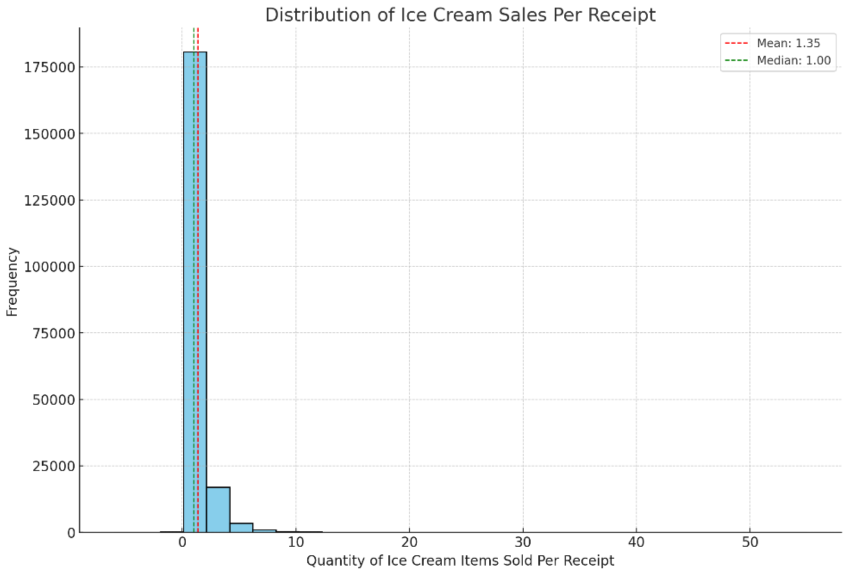

Figure 3: Distribution of Ice Cream Sales per Customer per Visit

Figure 3 shows the distribution of the number of ice creams purchased by customers each visit. As can be seen from Figure 3, most sales data are concentrated in a small number of items, especially in the range of 1 to 3, which indicates that most customers do not purchase a large quantity of ice cream each time. The median is exactly 1, which means that at least half of the customers only buy 1 ice cream per visit, but the mean is 1.53, which means that some customers buy larger orders. This drives up the average value, which also explains why there is an obvious long-tail distribution in Figure 3.

Our initial analysis suggests that the total daily quantity sold ranges from 180 to 400 units, with daily visitor numbers between 150 and 250 people. This provides us with a basic overview. However, to gather more precise data to guide our business activities, we need to perform additional risk scenario analyses at 5% and 1% probabilities.

Risk Scenario Analysis and Simulation

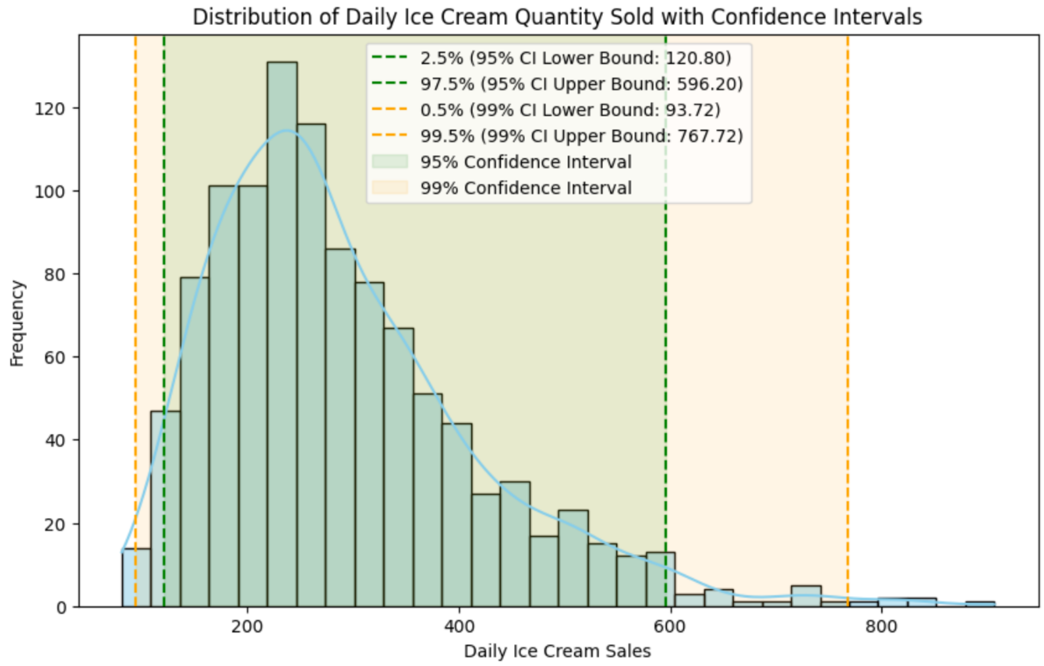

Figure 4: Distribution of Daily Ice Cream Quantity Sold with 95% and 99% Confidence Intervals

The 95% confidence interval for the quantity of ice cream sold ranges from 120.8 to 596.2. This suggests that during the observation period, the quantity of ice cream sold on 95% of days fell within this range. This interval can serve as a reliable benchmark for daily inventory management for the business. It allows the company to optimize its inventory levels, ensuring there is sufficient stock to meet demand on most days, while avoiding wastage due to overstocking.

On the other hand, the 99% confidence interval ranges from 93.72 to 767.72. This broader range indicates that during the observation period, the quantity of ice cream sold on 99% of days fell within this interval. This wider interval is crucial as it accounts for extreme situations, assisting the business in preparing for unusual peaks or troughs in sales. For instance, on hot summer days or special holidays, sales may approach or even reach the upper limit of 767.72. Conversely, on cold days or during market downturns, sales may plunge to the lower limit of 93.72.

While two-sided confidence intervals provide a comprehensive view of the data, including most of the range of expected values, one-sided detection can provide additional, targeted insights into potentially high-risk or low-probability business scenarios.

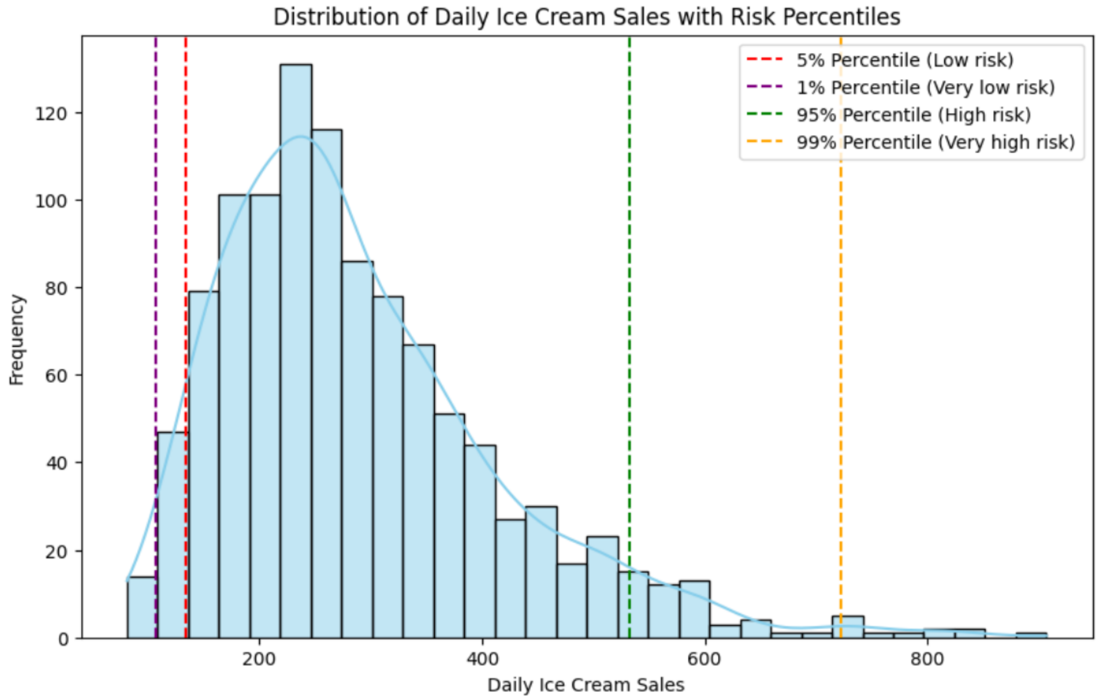

Figure 5: Original Distribution with Risk Percentiles

Figure 5 shows the daily ice cream sales distribution at four risk percentiles.

Low sales risk indicates overstocking danger. The 5% percentile (134 items, red dotted line) suggests sales will exceed this number 95% of the time. If sales drop below, it might mean fewer customers or off-peak seasons, indicating the need for inventory reduction or promotions. The 1% percentile (106.72 items, purple dotted line) warns of extremely low sales due to unusual circumstances like bad weather, requiring careful inventory management.

High sales risk signifies the risk of shortages. The 95% percentile (533 items, green dotted line) shows sales will exceed this number only 5% of the time. It serves as a safety margin to avoid shortages. Stock should increase for high sales predictions like holidays. The 99% percentile (722.4 items, orange dotted line) indicates an infrequent high sales situation, requiring extra stock or contingency plans for sudden demand spikes.

These analyses suggest that the shop should adjust its inventory strategy based on these percentiles, ensuring stock levels can meet a demand of at least 134 items most of the time and possibly 533 items during high sales events. This data-driven strategy can reduce inventory costs while maximizing customer satisfaction and sales opportunities.

However, this is only an analysis of risk based on past data. To deal with unknown situations, fitting the data to a known statistical distribution model can help us make more accurate predictions and risk assessments.

The Log-Normal distribution was chosen for analysis due to its alignment with the non-negativity of the data. This distribution matches the observed pattern where most values are concentrated in a lower range, and the tail is elongated due to a few higher values. Moreover, the Log-Normal distribution had the smallest sum of squared errors among all tested distributions, indicating its superior fit to the data.

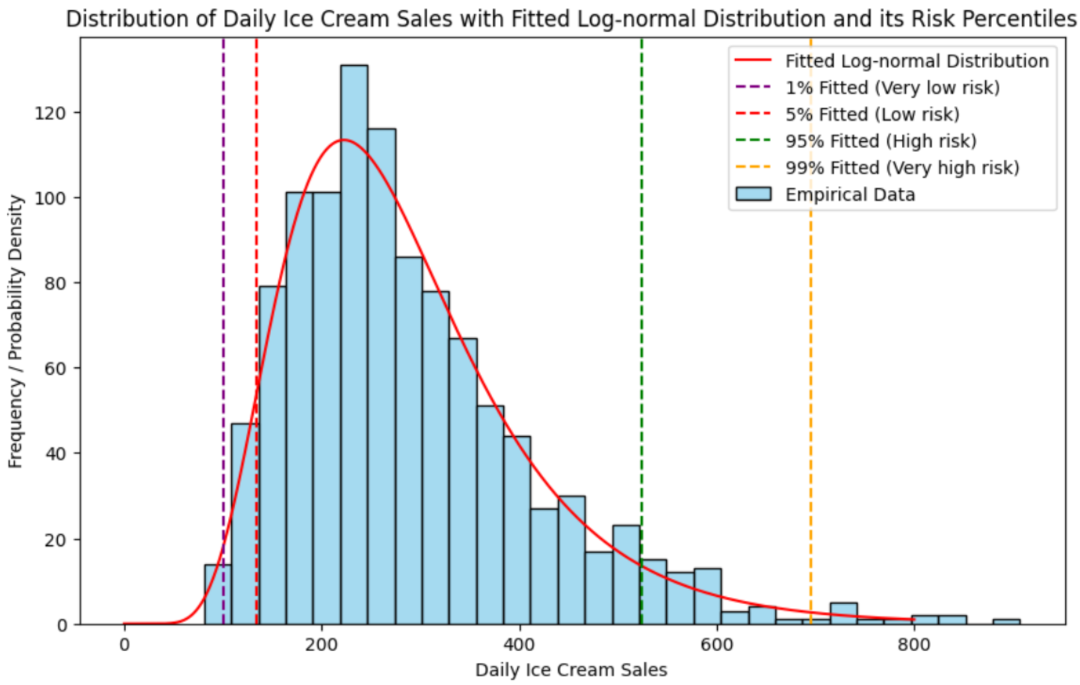

Figure 6: Risk Percentiles of Fitted Log-normal Distribution

After fitting a log-normal distribution, we obtain risk percentiles for daily ice cream sales from Figure 6. For 95% of the time, sales do not surpass 524 units, while for 5% of the time, sales fall below 134 units. In the 99% extreme case, sales could reach or surpass 695 units. Conversely, in the 1% extreme downturn, sales could drop to as low as 101 units. This data aids in accurately predicting sales volatility, thereby managing inventory risk more effectively.

The estimates derived from the log-normal distribution fitting closely align with the percentiles calculated from earlier empirical data. This confirms the appropriateness of our distribution choice and offers the supermarket a risk assessment based on theoretical models.

Bootstrapping

Bootstrapping is a statistical resampling technique that estimates the distribution of a statistic by repeatedly drawing samples (with replacement) from the original dataset. This method does not necessitate stringent assumptions about the underlying data distribution. Through extensive repetition, bootstrapping facilitates the exploration of extreme values, such as the best or worst scenarios represented by the 1% and 5% thresholds, which is critical for risk management and strategic planning.

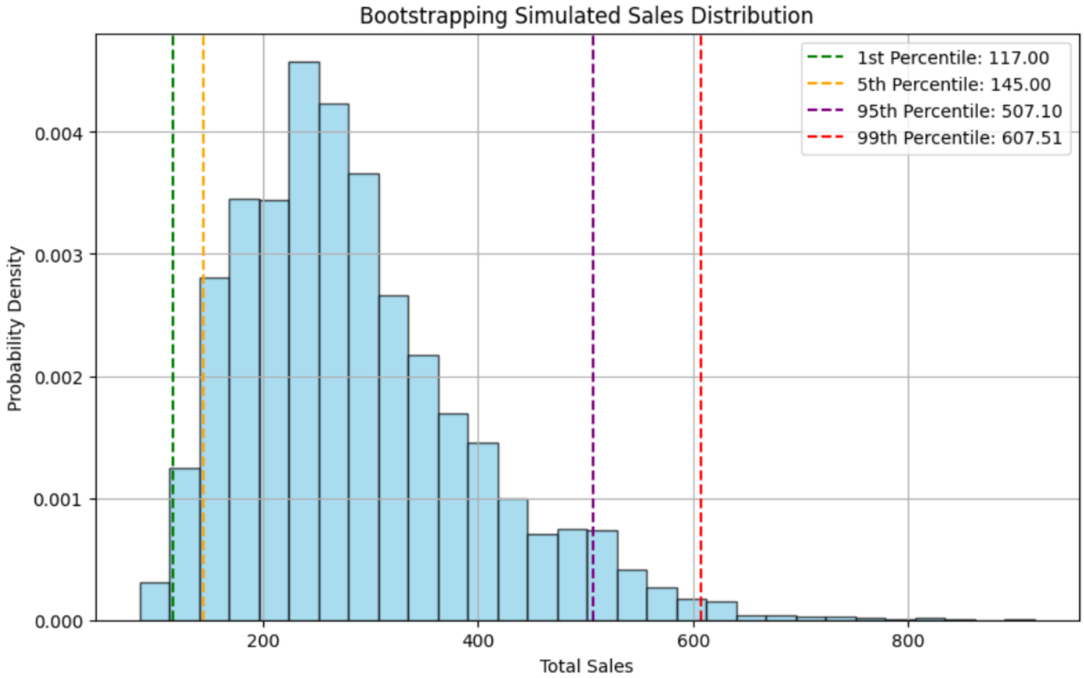

Figure 7: Bootstrapping Simulated Sales Distribution

The 1% and 5% probability scenarios shed light on the minimum sales levels during extremely adverse and moderately challenging market conditions, respectively. The 1st percentile, at 117 units, represents sales on the lowest 1% of days, likely due to significant market downturns or unfavorable weather. The 5th percentile at 145 units indicates that sales drop to this level on the lowest 5% of days, which could occur during mild seasonal lows or slight market fluctuations. These insights are crucial for businesses to devise strategies aimed at minimizing losses during low-demand periods by adjusting inventory, operational hours, or promotional efforts.

Conversely, the 95% and 99% probability scenarios provide benchmarks for the upper limits of sales. The 95th percentile at 507.10 units shows that 95% of days do not exceed this sales volume, suggesting a high sales threshold met during peak marketing or seasonal activities. The 99th percentile at 607.51 units indicates that sales surpass this level only 1% of the time, demonstrating potential for very high peaks, possibly during special promotions or unexpected surges in demand. Understanding these higher sales limits helps businesses ensure adequate resource allocation, including inventory and staffing, to meet potential high demand without facing shortages.

Simulation Modelling

Simulation modeling and Bootstrapping are techniques used to estimate statistics and understand data distribution but select and fit one or more appropriate probability distribution models based on the actual data and randomly sample from these fitted probability distributions instead of sampling directly from the actual data.

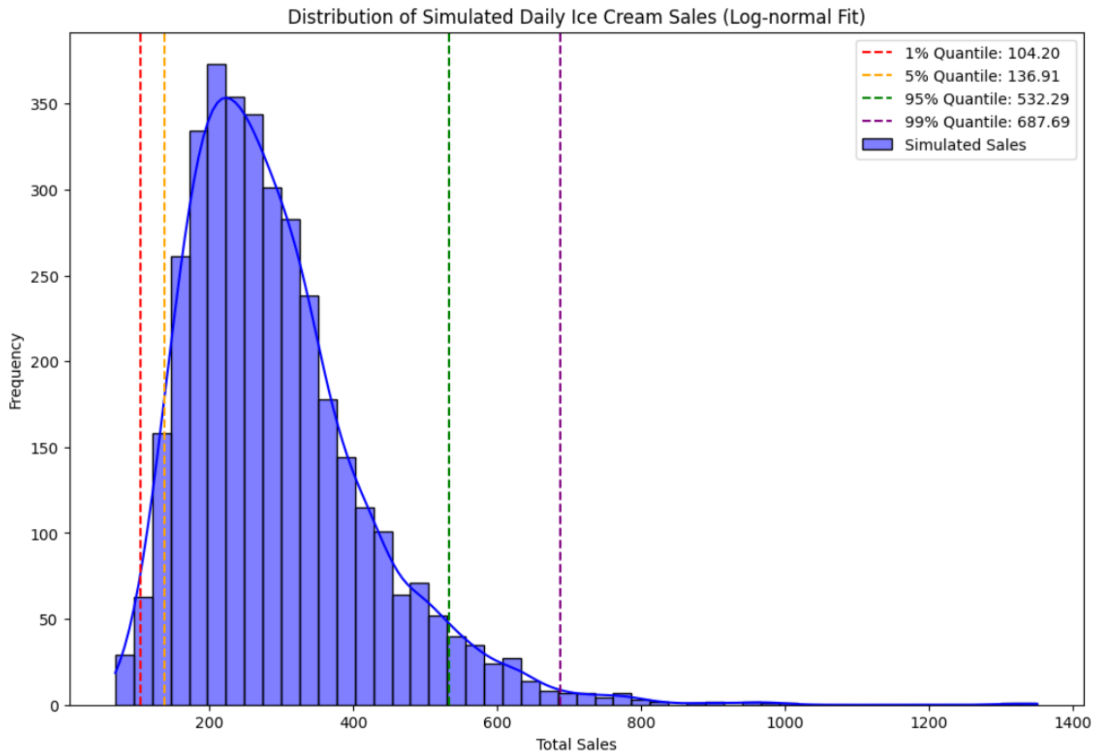

Figure 8: Simulation Modelling Distribution with Log-normal Fit

The Figure 8 illustrates the distribution of simulated daily ice cream sales based on a log-normal distribution, with key percentiles displayed, providing insights into potential sales extremes under different market conditions.

The 1st and 5th percentiles at 104.20 and 136.91 units, respectively, represent extremely unfavorable and minor market fluctuations or seasonal lows, enabling businesses to strategize inventory, operations, and promotions accordingly to minimize losses during low-sales periods. Conversely, the 95th and 99th percentiles at 532.29 and 687.69 units indicate the upper bounds of higher sales levels and the highest potential under extreme conditions. These values are critical for ensuring adequate resources during peak demands and preparing for possible expansions or large-scale promotions.

Compare Different Results from Approaches

In the comparative analysis of daily ice cream sales simulation results, two distinct methodologies were employed: Bootstrapping and Modeling Simulation. Here, we delve into their statistical characteristics and the implications these characteristics have on business decision-making processes, focusing particularly on how each method addresses the variability in sales data.

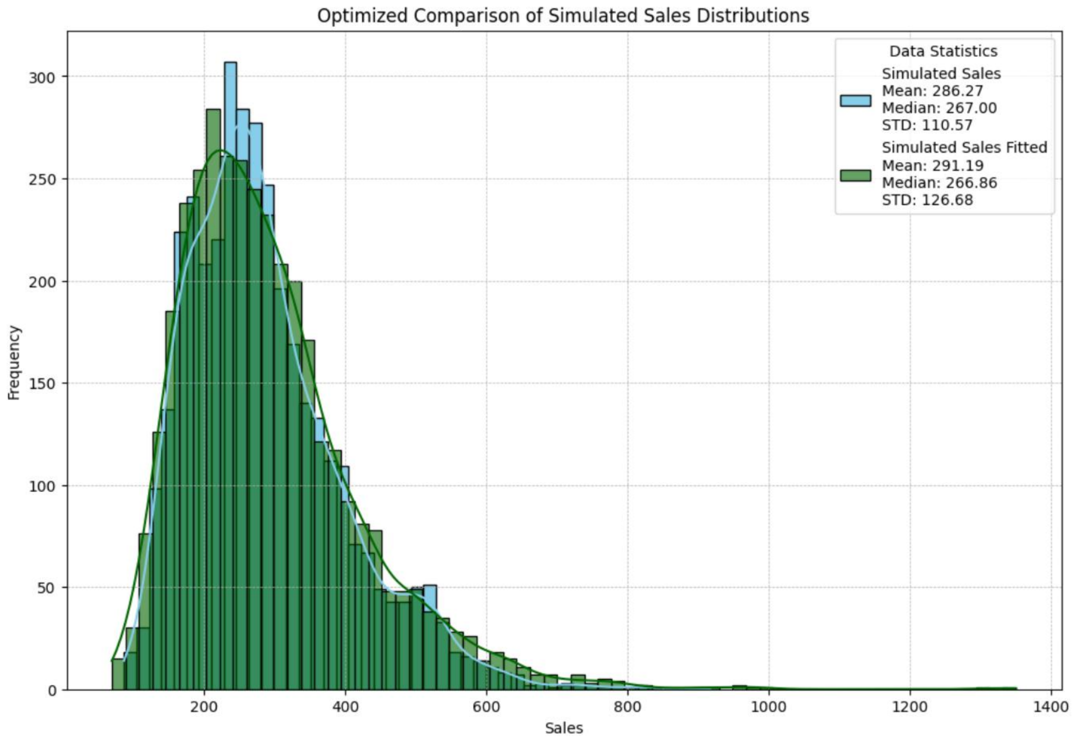

Figure 9: Compare Results of Two Different Approaches

The mean sales from the Bootstrapping simulation were observed at 286.27, whereas the Modeling Simulation reflected a slightly higher mean of 291.19. This increment in the Modeling Simulation suggests a better adaptation to higher sales volumes, likely attributed to the method's capability to fit the distribution tail more effectively. The medians of both simulations were closely aligned, registering at 267.00 for Bootstrapping and 266.86 for Modeling Simulation. This minimal discrepancy between the two medians indicates that both simulation methods stabilize around a similar central tendency, thus providing consistent central location estimations across methodologies.

A notable difference was observed in the standard deviations; Bootstrapping recorded a standard deviation of 110.57, contrasted with 126.68 for the Modeling Simulation. The increased standard deviation in the Modeling Simulation suggests a heightened sensitivity to extreme values in sales data, potentially introducing greater variability during the data generation phase. The Log-normal Fit Simulation demonstrated a pronounced long-tailed distribution, especially evident on the right side of the distribution curve. This characteristic suggests that while high sales days are rare, their potential maxima are significantly elevated compared to the central cluster of sales data. In contrast, the Bootstrapping simulation exhibited a more concentrated distribution, with shorter tails indicating a closer adherence to the median sales values observed in the raw data, and a less aggressive approach towards modeling extreme sales values.

In the Log-normal Fit Simulation, lower percentiles (1% at 104.20 and 5% at 136.91) and significantly higher upper percentiles (95% at 532.29 and 99% at 687.69) underscore the method's predictive strength for extreme sales outcomes. Conversely, the Bootstrapping method presented percentiles closer to the median, with upper percentiles (95% at 507.10 and 99% at 607.51) reflecting more moderate estimations, suggesting a conservative approach to forecasting extreme sales.

If business decisions depend on avoiding overstocking and capturing sales trends under general market conditions, the Bootstrapping method may be more suitable as it provides a more robust sales forecast and is less likely to overreact to extreme fluctuations. However, if the business strategy is to capture potential high-sales opportunities, such as during special promotions or holiday sales periods, the Log-normal Fit method would be more useful as it can simulate higher sales potential, helping enterprises prepare adequate inventory to meet possible sales peaks. The perspectives and information provided by the two methods have different emphases, and the choice of an appropriate simulation method should be based on specific business needs and risk preferences. For enterprises that prefer conservative operations, Bootstrapping provides more prudent data support, while enterprises pursuing maximum market opportunities may be more inclined to use the Log-normal Fit method.

Limitations of Analysis

From the perspective of the dataset, the analysis heavily relies on the completeness and accuracy of the historical sales data. Any inaccuracies in data collection, entry, or processing could introduce biases and skew the results. Furthermore, missing data for certain days, particularly those with significantly different sales patterns such as public holidays or days with extreme weather conditions, could introduce systematic biases.

From the methodological standpoint, both methods rely on specific assumptions regarding the data distribution. The bootstrapping approach assumes that the empirical data accurately represent the underlying distribution, which might not account for potential shifts in consumer behavior or market conditions over time. The simulation modeling approach, particularly when fitting data to a log-normal distribution, assumes that the chosen distribution is the best fit for the data, which might not always hold true, especially if the actual distribution undergoes changes.

Regarding external factors, both analytical methods may fail to adequately account for extraneous variables that could significantly impact ice cream sales, such as economic fluctuations, or competitive actions. Utilizing historical sales data to predict future trends does not account for potential new strategies or changes in the business environment. The static nature of historical data may not capture future shifts in consumer preferences or market conditions.

Conclusion

In the comprehensive analysis of inventory management for ice cream sales, our report has distilled significant insights from bootstrapping and simulation modeling methodologies. Our findings indicate a notable variability in sales volumes, crucially influenced by external factors like weather and holiday periods. From the bootstrapping analysis, the typical range of daily sales was found to oscillate between 180 to 400 units, with the 95% and 99% confidence intervals highlighting potential sales exceeding 507.10 units and 607.51 units only 5% and 1% of the time, respectively. This underscores a marked tendency for sales to spike significantly above average levels during peak periods.

The simulation modeling, particularly the log-normal fit, provided a nuanced perspective, predicting sales to not surpass 532.29 units 95% of the time, and 687.69 units 99% of the time. These predictive insights are instrumental for crafting robust inventory strategies. They suggest the necessity for the supermarket to maintain inventory levels that cater to typical sales volume but also accommodate occasional spikes, especially during favorable weather conditions or special promotions.

The strategic implications of these findings are profound. Firstly, inventory management must be dynamic, calibrated to buffer against potential sales volatility indicated by the upper confidence levels from both modeling approaches. For instance, maintaining an inventory that can cover up to 532 units would suffice for the majority of the year but preparing for as much as 687 units is prudent to avoid understocking during unexpected sales peaks. Additionally, integrating predictive analytics into pricing strategies could further optimize revenue, particularly by implementing dynamic pricing during anticipated high-demand periods.

However, it is crucial to acknowledge the limitations inherent in our analysis. The accuracy of our findings hinges on the quality and completeness of the historical sales data, which may not fully capture future market dynamics or shifts in consumer behavior. Furthermore, both the bootstrapping and simulation modeling techniques rely on assumptions that may not hold under all market conditions, and external factors like economic changes or competitor actions were not fully integrated into the predictive models.

To facilitate more accurate decision-making, it is recommended that the supermarket integrate real-time weather forecasting and local event tracking into their sales forecasting models. This integration can refine the predictive accuracy of our existing models and help tailor inventory and pricing strategies more closely to actual demand patterns.

Conclusively, the findings and models outlined in this report provide a solid foundation for the supermarket to enhance its operational efficiency, reduce economic losses due to overstocking or stockouts, and ultimately improve customer satisfaction. Implementing these recommendations requires a commitment to continuous monitoring and adjustment of inventory practices, ensuring that the business remains agile and responsive to market dynamics. This proactive approach will empower the supermarket to make informed business decisions that capitalize on predictable sales patterns and effectively manage the inherent unpredictability of retail environments.