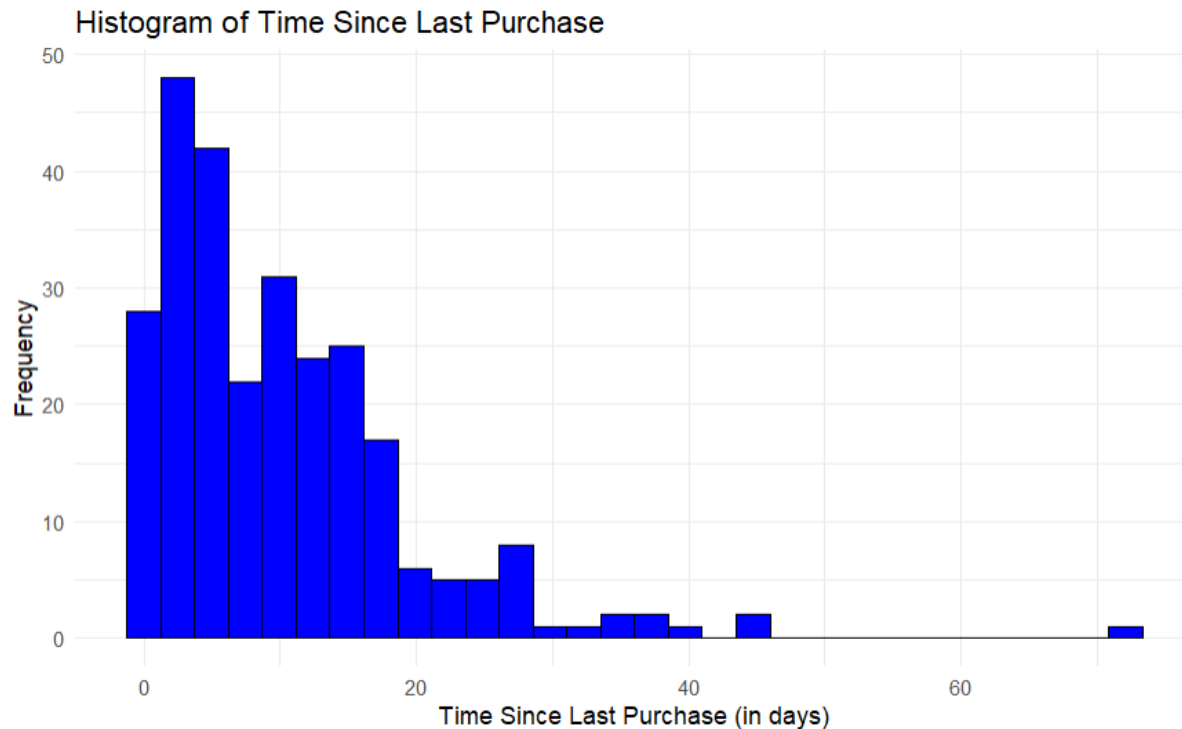**Q1**

Histogram of Time Since Last Purchase

The histogram of the "**Time_Since_Last_Purchase**" variable from the Walmart data is displayed above. Based on the shape of the histogram, we can make an educated guess about which distribution it most likely follows:

Exponential Distribution: Characterized by a rapid decrease in frequency as the value increases. This is typically used for modeling time until an event occurs and often has a long tail to the right.

Normal Distribution: This is a symmetric, bell-shaped curve. It is characterized by a single peak and tails that extend symmetrically on both sides.

Lognormal Distribution: This distribution is similar to the normal distribution but skewed, typically with a longer tail on the right side.

Gamma Distribution: A more generalized family of distributions, which includes the exponential distribution. It is often skewed to the right but can take on different shapes.

Based on the histogram, it appears that the "**Time_Since_Last_Purchase**" variable does not follow a symmetric bell-shaped curve, which rules out a normal distribution. The distribution seems to be right-skewed, which is typical of exponential, lognormal, or gamma distributions. Among these, the exponential distribution is often the simplest and is commonly used to model the time until an event (like a purchase) occurs.

However, without further statistical analysis, it's not possible to definitively conclude which distribution it follows. For a more accurate determination, you could perform statistical tests or fit different distribution models to the data.

Based on the histogram of the "**Time_Since_Last_Purchase**" variable and its right-skewed shape, the most likely distribution for this data is:

Exponential Distribution.

This conclusion is drawn because the exponential distribution is typically used to model the time until an event occurs and is characterized by a rapid decrease in frequency as the value increases, which seems to align with the shape of the histogram. However, it's important to note that this is a visual assessment, and a more precise determination would require statistical testing.

Q2

Histogram of Sample Means of Time Since Last Purchase
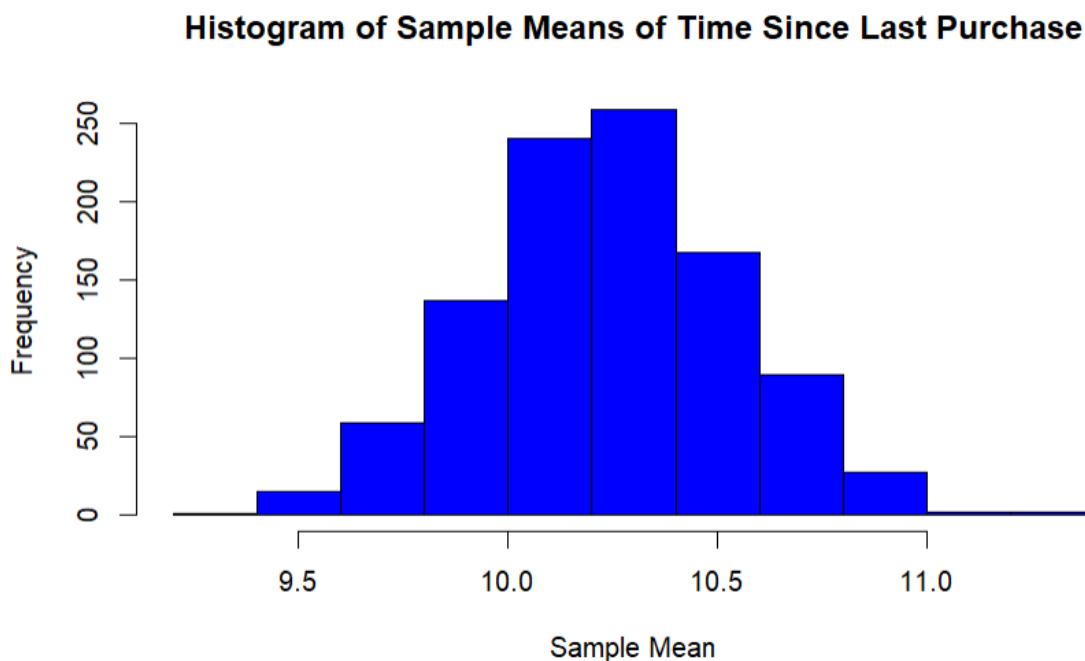
This code will create a histogram of the sample means of the "Time_Since_Last_Purchase" variable. According to the Central Limit Theorem, regardless of the original distribution of the dataset, the sampling distribution of the sample means tends to be normal (or Gaussian), especially as the sample size gets larger. Hence, the histogram you generate should resemble a normal distribution.

As for the options you've provided (exponential, normal, lognormal, gamma), the sampling distribution of the sample means most likely follows a normal distribution.

Q3

Histogram of Spending Last Year
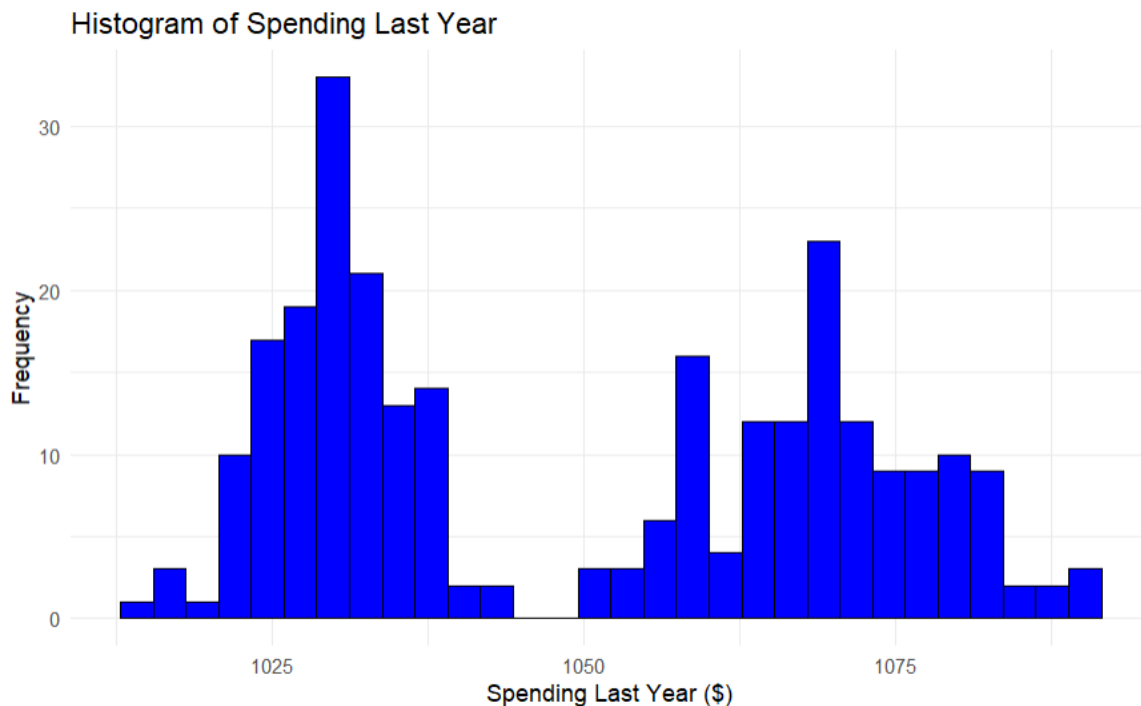
The histogram of the "Spending_Last_Year" variable from the Walmart data is displayed above. To determine which distribution it most closely follows, let's consider the characteristics of each option:

Bimodal (2 peaks): This distribution would show two distinct peaks in the histogram.

Exponential Distribution: Characterized by a rapid decrease in frequency as the value increases, with a longer tail on the right side.

Gamma Distribution: Similar to the exponential distribution but can have different shapes, typically right-skewed with a long tail.

Normal Distribution: Symmetrical, bell-shaped curve, indicating equal distribution around the mean.

Based on the shape of the histogram, we can visually assess which distribution the "Spending_Last_Year" variable is most likely to follow. However, the exact determination would depend on the visual characteristics of the histogram. If you can describe the shape of the histogram or provide more details, I can offer a more specific suggestion.

This code will create a histogram of the "Spending_Last_Year" variable using ggplot2, a popular package for data visualization in R. The geom_histogram() function creates the

histogram, and labs() and theme_minimal() add labels and a minimalistic theme to the plot.

Once you generate the histogram, you can visually assess which distribution it most closely resembles (bimodal, exponential, gamma, or normal) based on its shape. Remember, a bimodal distribution will show two peaks, an exponential distribution will show a rapid decrease with a long tail to the right, a gamma distribution will be skewed with a longer tail, and a normal distribution will appear as a symmetric, bell-shaped curve.
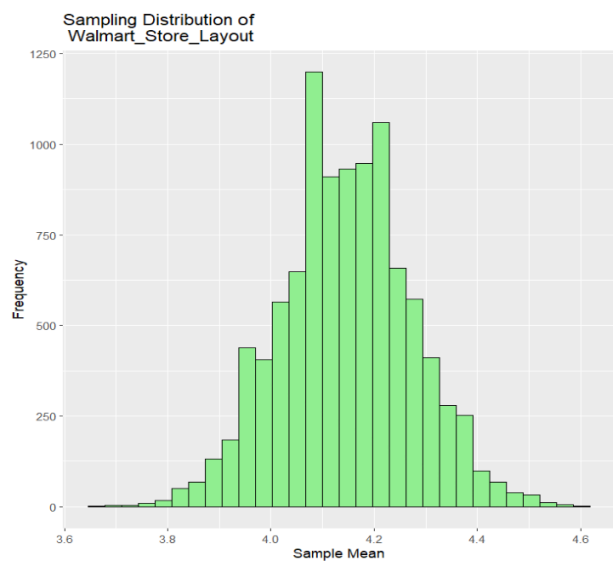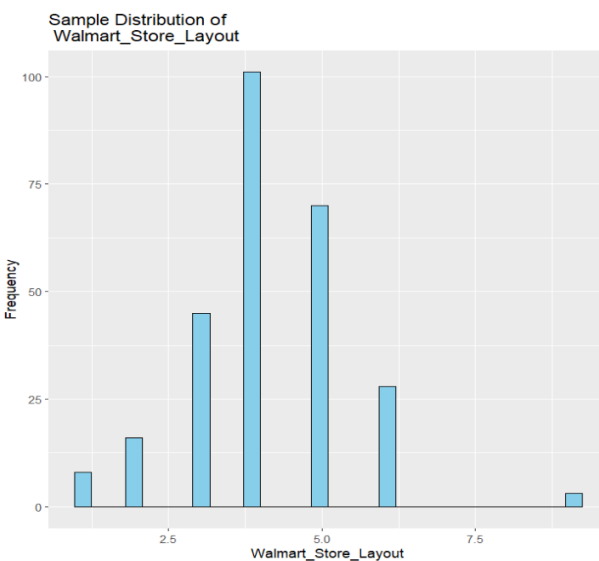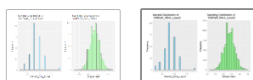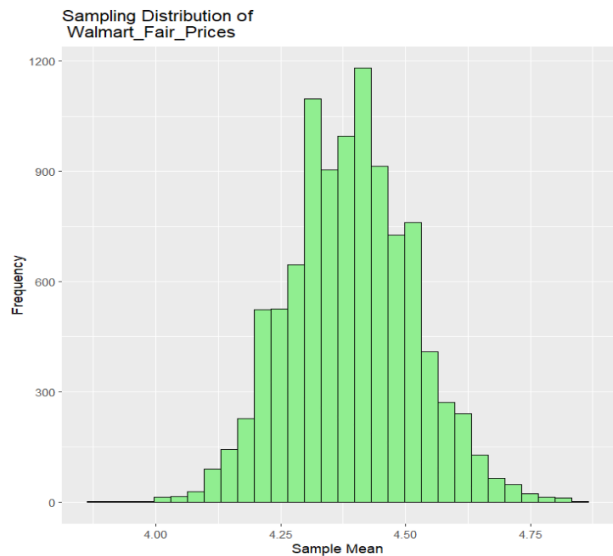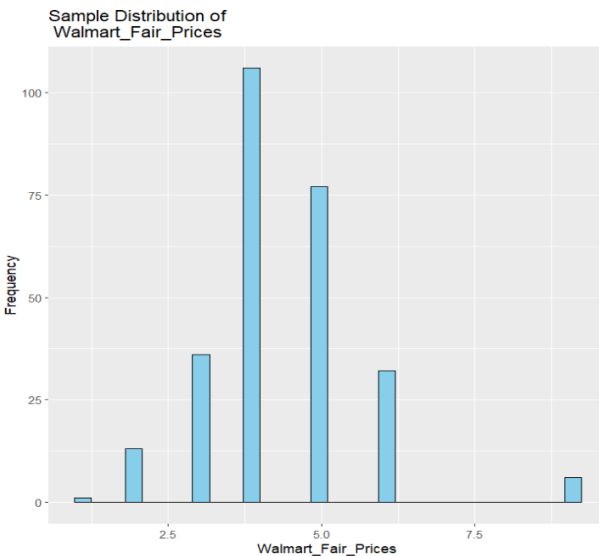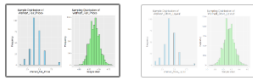
Q4

The histogram of the sample means for the "Spending_Last_Year" variable is displayed above. When analyzing the sampling distribution of sample means, the Central Limit Theorem (CLT) is a crucial consideration. According to the CLT, regardless of the original distribution of the dataset, the sampling distribution of the sample means will tend to be normal (Gaussian) as the sample size becomes larger.

Based on the histogram and keeping the CLT in mind, the sampling distribution of "Spending_Last_Year" most likely follows a:

Normal Distribution.

This is evident from the bell-shaped curve of the histogram, which is characteristic of a normal distribution. Remember, this conclusion is based on the visual appearance of the histogram and the principles of the CLT.

Q11

Sample Distribution of
Walmart_Fair_Prices

Sampling Distribution of
Walmart_Fair_Prices

Sample Distribution of
Walmart_Store_Layout

Sampling Distribution of
Walmart_Store_Layout

My Use of Confidence Intervals in Managing My Coffee Shop

As the owner of a neighborhood coffee shop, I'm constantly looking for ways to enhance our operations and profitability. Recently, I've been contemplating extending our shop's operating hours. While this decision could potentially increase our sales and customer reach, I'm keenly aware that it also might lead to higher operating costs and could impact my team's work-life balance. To make an informed decision, I decided to use statistical methods, specifically confidence intervals, to analyze various aspects of our business.

Key Variables I Considered:
Sales Revenue: I wanted to understand the possible increase in sales if we extended our hours. By estimating a confidence interval for projected sales, I could gauge the range of potential revenue enhancements.
Customer Foot Traffic: It was crucial to determine whether the extended hours would actually attract more customers. Analyzing customer visits during our current and proposed extended hours would offer insights into this.
Employee Workload: My team's well-being is paramount. I needed to assess if the longer hours would require hiring more staff or lead to overwork.
Customer Satisfaction: Maintaining high customer satisfaction is at the core of our values. I planned to measure this before and after any change in operating hours.
Operating Costs: Lastly, I looked at the potential increase in costs such as utilities and wages. It was vital to ensure that any increase in revenue would not be negated by these additional expenses.

Stakeholders Involved:
My staff, who are concerned about their schedules and job satisfaction.
Our loyal customers, who value both our service quality and availability.
Local suppliers, who might need to adjust their delivery schedules in response to our changes.

The local community, which benefits from a thriving local business.
By applying confidence intervals to these variables, I could estimate the potential impacts of extended hours with a certain level of statistical confidence. This approach offered a balanced view, helping me weigh the benefits against the risks from a data-driven perspective. It wasn't just about the numbers; it was about making a decision that aligned with our values and goals as a community-focused coffee shop.