APPLIED MASTERS EXAM SPRING 2024

Satoshi Ido

April 2024

Problem 1: Effects of Autonomous Driving Factors on Heart Rate Variability: A Crossover Design Study

1 Summary

This study aims to investigate the effects of three key factors associated with autonomous driving - transparency, recommended control mode, and system reliability - on heart rate variability (HRV). Each factor is dichotomized into 'high' and 'low' levels, which leads to a total of eight unique treatment combinations. Subjects will undergo two distinct simulated drives, each tailored to measure HRV across these various treatment combinations. The study's experimental design utilizes a crossover method combined with Latin Square arrangements, ensuring subjects are evenly exposed to all treatments. This strategy allows for a comprehensive assessment without the need for 100 unique simulation drives.

2 Introduction

The goal of this study is to explore how different combinations of transparency, recommended control mode, and system reliability impact HRV. By employing simulated drives and wearable HRV measurement devices, this research seeks to identify the most effective combinations of these factors in influencing HRV.

3 Methods

3.1 Participants

A total of 48 subjects will be recruited to participate in the experiment twice, ensuring a balanced and comprehensive examination of the treatment effects.

3.2 Procedure

Participants will receive a brief orientation before commencing the experimental trials. Following this orientation, each participant will be randomly assigned a number from 1 to 24, which determines the specific sequence of treatments they will receive. Participants will complete two simulated driving sessions (Drive 1 and Drive 2), each consisting of four periods where they will be exposed to different treatment conditions. During each drive, one factor remains constant, while the levels of the other two factors change with each period.

3.3 Experimental Design

The experiment is designed as a crossover study incorporating two 4x4 Latin Squares to manage the eight possible treatment combinations and to account for variations between subjects. This approach also facilitates the evaluation of first-order residual or carryover effects. *A carryover or residual effect is defined as the effect of the treatment from the previous time period on the response at the current time period.* The design guarantees an equal test frequency for each treatment combination across all subjects, albeit in varied sequences. Participants are randomly allocated a specific sequence of treatment combinations prior to their simulation sessions.

Table 1 simplifies the presentation by using letters A through H to represent specific treatment combinations. This ensures each participant will encounter each treatment combination once across the two simulated drives. The sequence of treatments and exposure to each combination for a participant are dictated by robust statistical design techniques, namely, the Crossover design and Randomization.

Table 1: Factor Combinations (Treatments)

	Transparency	Recommended Control Mode	System Reliability
A	High (+)	High (+)	High (+)
B	High (+)	High (+)	Low (-)
C	High (+)	Low (-)	High (+)
D	High (+)	Low (-)	Low (-)
E	Low (-)	High (+)	High (+)
F	Low (-)	High (+)	Low (-)
G	Low (-)	Low (-)	High (+)
H	Low (-)	Low (-)	Low (-)

These alphabetic designations are further elucidated in tables below, showcasing how two distinct Latin Squares are constructed for periods 1 to 4 and 5 to 8, respectively.

3.4 Crossover Design

For this study, a Crossover design incorporating Latin Squares is an appropriate methodology. A Crossover design is a type of repeated measurements design in which each participant (experimental unit) receives different treatments over several time periods. This allows each individual to serve as their own control, as they 'cross over' from one treatment to another during the course of the experiment. In our case, participants will encounter a series of periods within each of the two simulated drives, receiving multiple treatments over time. Specifically, each subject will be exposed to 8 different treatments at varied times, with the sequence randomized for each individual. This approach effectively addresses two potential sources of variation—differences among subjects and time periods—as blocking factors.

3.4.1 Latin Squares

Latin Squares are sophisticated designs that control for two sources of variability—subjects and periods—by treating them as blocking factors. This design has two key features: the frequency of each treatment is consistent across sequences and periods. For our experiment, the Latin Squares (depicted in **Figures 1 to 6**) are structured to account for a first-order carryover effect, ensuring that each treatment is preceded by every other treatment exactly once.

Participants are randomly assigned a number from 1 to 24, which corresponds to a specific sequence of the 8 treatments outlined in the Latin Square. For instance, if an individual is assigned number 5, their treatment schedule would unfold as follows: They would first undergo treatments E, F, G, H during Drive 1. Subsequently, in Drive 2, they would proceed with treatments A, B, C, D.

	Drive 1					Drive 2			
	Period 1	Period 2	Period 3	Period 4		Period 5	Period 6	Period 7	Period 8
1	A	B	C	D		E	F	G	H
2	C	A	D	B		G	E	H	F
3	B	D	A	C		F	H	E	G
4	D	C	B	A		H	G	F	E

Figure 1: Transparency being held ‘high‘ in Drive 1 and ‘low‘ in Drive 2

	Drive 1					Drive 2			
	Period 1	Period 2	Period 3	Period 4		Period 5	Period 6	Period 7	Period 8
5	E	F	G	H		A	B	C	D
6	G	E	H	F		C	A	D	B
7	F	H	E	G		B	D	A	C
8	H	G	F	E		D	C	B	A

Figure 2: Transparency being held ‘low‘ in Drive 1 and ‘high‘ in Drive 2

	Drive 1					Drive 2			
	Period 1	Period 2	Period 3	Period 4		Period 5	Period 6	Period 7	Period 8
9	A	B	E	F		C	D	G	H
10	E	A	F	B		G	C	H	D
11	B	F	A	E		D	H	C	G
12	F	E	B	A		H	G	D	C

Figure 3: Recommended Control Mode being held ‘high‘ in Drive 1 and ‘low‘ in Drive 2

	Drive 1					Drive 2			
	Period 1	Period 2	Period 3	Period 4		Period 5	Period 6	Period 7	Period 8
13	C	D	G	H		A	B	E	F
14	G	C	H	D		E	A	F	B
15	D	H	C	G		B	F	A	E
16	H	G	D	C		F	E	B	A

Figure 4: Recommended Control Mode being held ‘low‘ in Drive 1 and ‘high‘ in Drive 2

	Drive 1					Drive 2			
	Period 1	Period 2	Period 3	Period 4		Period 5	Period 6	Period 7	Period 8
17	A	C	E	G		B	D	F	H
18	E	A	G	C		F	B	H	D
19	C	G	A	E		D	H	B	F
20	G	E	C	A		H	F	D	B

Figure 5: System Reliability being held ‘high‘ in Drive 1 and ‘low‘ in Drive 2

	Drive 1					Drive 2			
	Period 1	Period 2	Period 3	Period 4		Period 5	Period 6	Period 7	Period 8
21	B	D	F	H		A	C	E	G
22	F	B	H	D		E	A	G	C
23	D	H	B	F		C	G	A	E
24	H	F	D	B		G	E	C	A

Figure 6: System Reliability being held 'low' in Drive 1 and 'high' in Drive 2

4 Analysis

Based on the Crossover design explained above, Mixed-effects model with the first-order residual effect is preferable. Mixed models are especially useful when working with a within-subjects design like Crossover design. The some of the main reasons are:

- It can include random effects to account for subject variability, allowing for more accurate estimates of treatment effects by treating the subject as a random effect.
- It can handle correlation within subjects since measurements from the same subject are more similar to each other than to those from different subjects.
- It analyzes data that includes both fixed effects (treatment effects) and random effects (e.g., subject).

4.1 Statistical Model

The analysis will involve an examination of the main effect of treatment combinations as well as interaction effect on the response variable, 'HRV'. Two blocking factors will be part of the model as well.

$$y_{ijk} = \mu + P_i + \tau_j + (\tau P)_{ij} + S_k + r_{j'} + \varepsilon_{ijk}$$

where:

y_{ijk} = HRV for the kth subject in the ith period receiving the jth treatment
μ = Overall mean HRV
P_i = Fixed effect of the ith period (=block)
τ_j = Fixed effect of the jth treatment
$(\tau P)_{ij}$ = Interaction effect of the jth treatment in the ith period
S_k = Random effect of the kth subject (=block)
$r_{j'}$ = First-order residual effect of j'th treatment, which is used in the previous period
$\epsilon_{ijk} \sim N(0, \sigma^2)$ (Random error term)

4.1.1 Sample Data Content

From this model, we aim to test the main effects, which are the direct influences of three variables of interest on HRV while addressing the effect of subject and period and interaction between treatment and period. To evaluate the combined impact of transparency, recommended control mode, and system reliability on HRV, we will use a method known as **Least Squares Means (LSMeans)**. **LSMeans** are essentially averages that have been adjusted for other variables in the model. They are calculated as a sum of the model's estimated effects, which allows us to make fair comparisons between different treatment combinations. In order to run the statistical model defined above, the specific contrast systems should pertain to the original data. It should look like **Figure 7** below:

J' th trt	J th trt	A(Resid1)	B(Resid2)	C(Resid3)	D(Resid4)	E(Resid5)	F(Resid6)	G(Resid7)	Response
0	A	0	0	0	0	0	0	0	**
A	B	1	0	0	0	0	0	0	**
B	C	0	1	0	0	0	0	0	**
C	D	0	0	1	0	0	0	0	**
D	E	0	0	0	1	0	0	0	**
E	F	0	0	0	0	1	0	0	**
F	G	0	0	0	0	0	1	0	**
G	H	0	0	0	0	0	0	1	**
H	any	-1	-1	-1	-1	-1	-1	-1	**

Figure 7: Sample Data Content

5 Conclusion

This research provides valuable insights into the combinations between transparency, recommended control mode, and system reliability on heart rate variability (HRV) within the context of autonomous driving. Through a methodical experimental design employing a crossover approach with Latin squares, we ensured comprehensive exposure to all treatment combinations, thereby facilitating a robust analysis of their effects on HRV. The analysis illuminates the significant impact of these factors on physiological responses during simulated driving tasks.

6 Appendix

6.1 Degrees of Freedom Explanation

This section provides an updated explanation for calculating the degrees of freedom for each factor in the model, reflecting the structure of the crossover study with two Latin squares.

1. **Period (P_i):**
   - $DF = i - 1$ where $i = 8$ (number of periods).
   - $DF_{P_i} = 8 - 1 = 7$.
2. **Treatment (τ_j):**
   - $DF = j - 1$ where $j = 8$ (number of treatments).
   - $DF_{\tau_j} = 8 - 1 = 7$.
3. **Treatment (τP_{ij}):**
   - $DF = (i - 1) * (j - 1)$ where $i = 8$ and $j = 8$ (number of treatments in number of periods).
   - $DF_{\tau P_{ij}} = 7 * 7 = 49$.
4. **Subject (S_k) (considered as blocks):**
   - $DF = k - 1$ where $k = 48$ (number of subjects).
   - $DF_{S_k} = 48 - 1 = 47$.
5. **First-order Residual Effect ($r_{j'}$):**
   - Each subject has 7 potential carryover effects for every treatment.
   - $DF_{r_{j'}} = j'$ where $j' = 7$ (number of potential carryover effects).
   - $DF_{r_{j'}} = 7$.
6. **Error (ϵ_{ijk}):**
   - Total observations $= n \times p = 48 \times 8 = 384$.
   - Error DF = Total observations - DF_{P_i} - DF_{τ_j} - $DF_{\tau P_{ij}}$ - DF_{S_k} - $DF_{r_{j'}}$.
   - $DF_{\epsilon_{ijk}} = 384 - 7 - 7 - 49 - 47 - 7 = 267$.

Problem 2: Analyzing Gas Consumption Patterns: An Analytical Study on Interval Mileage and Gallons with Key Visualizations and Statistical Modeling Insights

1 Summary

Analysis of the data indicates seasonal patterns in fuel efficiency, measured by miles per gallon (MPG), with variations observed across different times of the year, months, and weeks. Given the sample size, data characteristics, and the necessity for straightforward interpretations, multiple linear regression comes up as the fitting method to analyze what factors influence fuel efficiency.

2 Introduction

The dataset records a car's fuel efficiency, known as interval MPG, a key indicator of how well a car manages fuel: A higher MPG means better fuel efficiency. The objective of this study is to identify meaningful patterns within the data. This report illustrates these trends through visualizations and statistical modeling, specifically focusing on the actual fuel efficiency rather than the car's computer estimates. This focus is due to a significant discrepancy found through a statistical test (paired t-test) between the computer's estimates and the actual fuel usage figures. Thus, further analysis and visualizations are grounded in actual MPG data, painting a picture of the car's efficiency and examining how various factors might influence it.

3 Hybrid Car Characteristics - Prius Prime

The Prius Prime, the model of car from which our data originates, benefits from a hybrid engine that operates on both electricity and gas. It can travel 25-30 miles on electricity alone before switching to gas, typically at speeds of 15 mph or higher. Consequently, it is an area of interest in exploring any elements that might affect the car's fuel efficiency during regular use. Notably, the Prius relies on gasoline under certain conditions, such as driving at a consistent speed of 25 mph or more, needing to heat the interior on cooler days, or if the engine has not yet reached its standard operating temperature.

4 Methods

For ease of understanding, data representations have been categorized as 'Overall' and 'interval,' prefaced correspondingly in the analysis. Visualizations are generated considering 'Date,' 'Year,' 'Month,' and 'Day of Week,' giving a hint of utilizing seasonality eventually. Lag data for miles and gallons, both 'overall' and 'at fill-up,' have been formulated to assess the influence of historical data on current outcomes, integrating rolling averages and exponentially weighted moving averages (EWMAs) to underscore recent patterns. Some of the charts also display the Exponentially Weighted Standard Deviation (EWSD), underscoring the importance of recent data in gauging variability. Finally, the linear model we applied aims to pinpoint which factors have a significant role in the fuel efficiency (Actual interval MPG) of the car.

5 Analysis

5.1 Paired T Test

As we can see from , **Figure 8**, the histogram of difference between computer-based MPG and actual MPG below, it is almost obvious that the computer-generated number is overestimated. To determine statistically whether the computer-based MPG values are consistently overestimated, we can conduct a paired t-test.

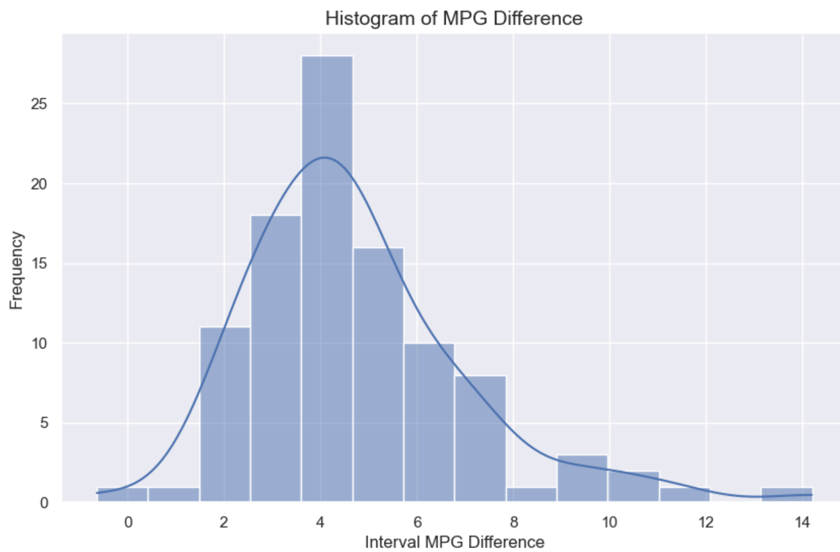

Figure 8: Histogram of MPG Differences (Computer Based Number - Actual Number)

The hypotheses for the test are defined as follows:

- Null hypothesis (H_0): The mean difference between the computer-generated MPG and the actual MPG is zero. Formally, $H_0 : \mu_{diff} = 0$.
- Alternative hypothesis (H_1): The mean difference between the computer-generated MPG and the actual MPG is greater than zero, indicating that the computer-generated MPG values are typically higher than the actual MPG values. Formally, $H_1 : \mu_{diff} > 0$.

Upon running the t-test, we obtained the following results:

Statistic	Value
Test Statistic (t)	20.434
p-value	$< 2.2 \times 10^{-16}$
Degrees of Freedom (df)	100
Mean Difference	4.728713

Table 2: Paired t-test Results

Given the extremely low p-value and the fact that the 95% confidence interval for the mean difference does not include 0 and is entirely above 0, we reject the null hypothesis in favor of the alternative hypothesis. This indicates significant statistical evidence to conclude that the computer-generated interval MPG is, on average, 4.73 MPG greater than the actual interval MPG.

5.2 Visualization

The histogram of interval MPG in **Figure 9**, shows a right-skewed distribution, suggesting that while there are instances where the Prius achieves over 100 MPG, the majority of MPG values are concentrated between 50 and 80 MPG.

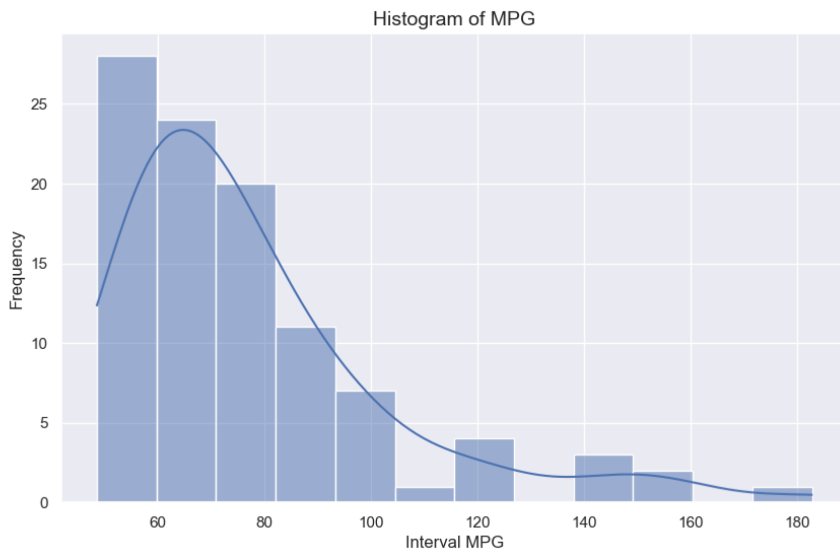

Figure 9: Histogram of Interval MPG

In **Figure 10**, An analysis of the average interval miles and average interval MPG by year, along with their Exponentially Weighted Moving Averages (EWMAs), reveals notable trends. In 2020, there was a noticeable decrease in both Mileage and MPG, likely attributable to the pandemic's impact, which included restrictions on movement and the closure of public spaces, resulting in less travel.

In contrast, 2021 saw a rebound with the highest mileage recorded over the studied period, despite a lower MPG. This suggests a mix of short and long-distance travel as activities began to resume more normal patterns. Beyond 2021, the data shows a gradual decline in mileage, while MPG has remained relatively stable. It will be important to continue monitoring these trends through 2024 to gain a full understanding of the long-term impacts on driving habits and vehicle efficiency.

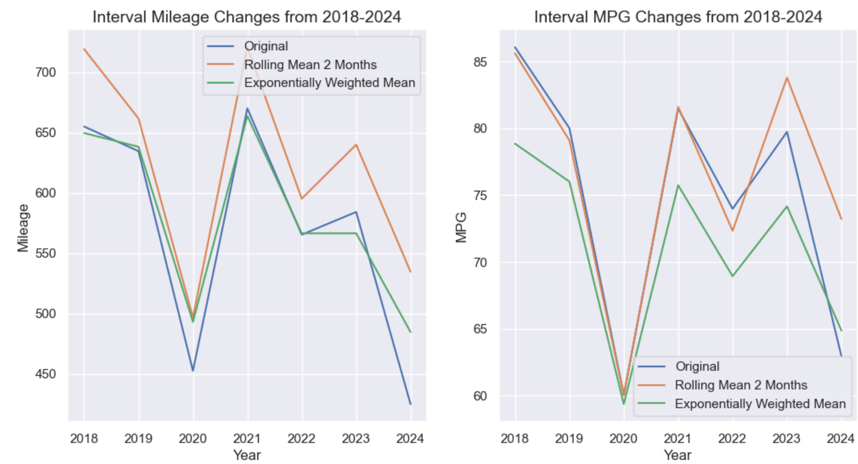

Figure 10: Interval MPG and Interval Mileage by Year

The correlation plot between interval miles and interval gallons in **Figure 11** illustrates a curved relationship rather than a strictly linear one. This suggests that, for a hybrid car, the relationship between the distance driven and fuel consumed is not directly proportional throughout; as the mileage increases, the gallons used do not increase at the same rate. This could indicate that other factors are influencing MPG performance, such as number of time the car ran over 25mph or under different driving conditions, reflecting the unique characteristics of hybrid vehicle fuel consumption.

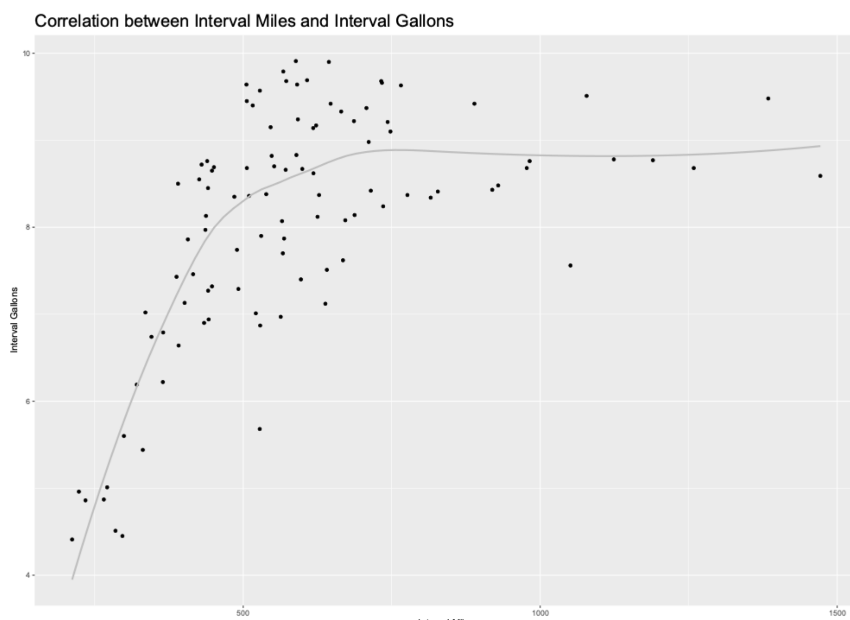

Figure 11: Correlation between Mileage Run and Gallons Consumption

Reviewing monthly data for mileage and gas consumption reveals distinct trends. Notably, long-distance driving and significant gas usage occur especially from May, and September through November, correlating with increased activities during warmer months. A distinctive feature between spring and fall activities is the higher EWSD in fall, suggesting a mix of short and long-distance driving.

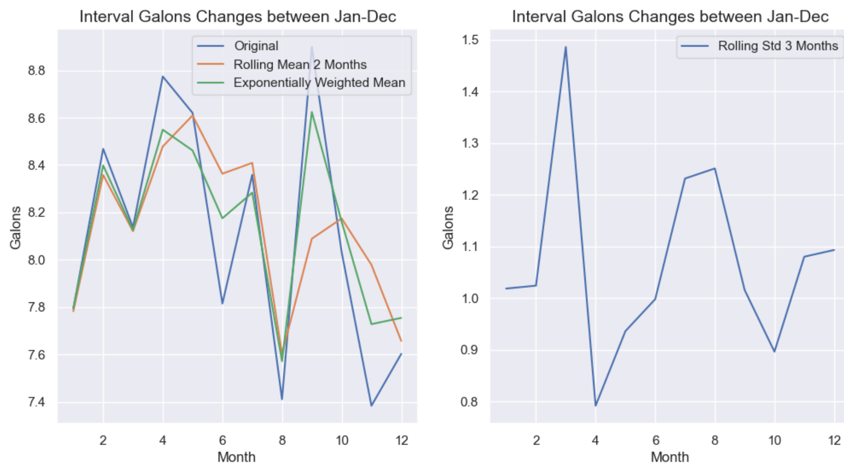

Figure 12: Interval Gallons by Month

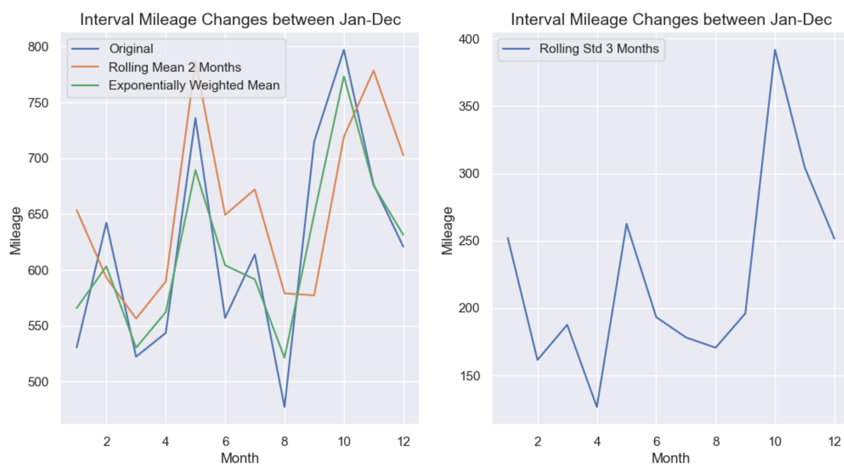

Figure 13: Interval Mileages by Month

The correlation plot generally indicates a linear trend between interval days and interval MPG, suggesting that as the number of days between intervals increases, there tends to be an improvement in MPG. If we discount the outlier, which can be seen in the right side of the plot, the relationship appears to be consistently linear, emphasizing a positive correlation where longer intervals may correspond to higher MPG efficiency.

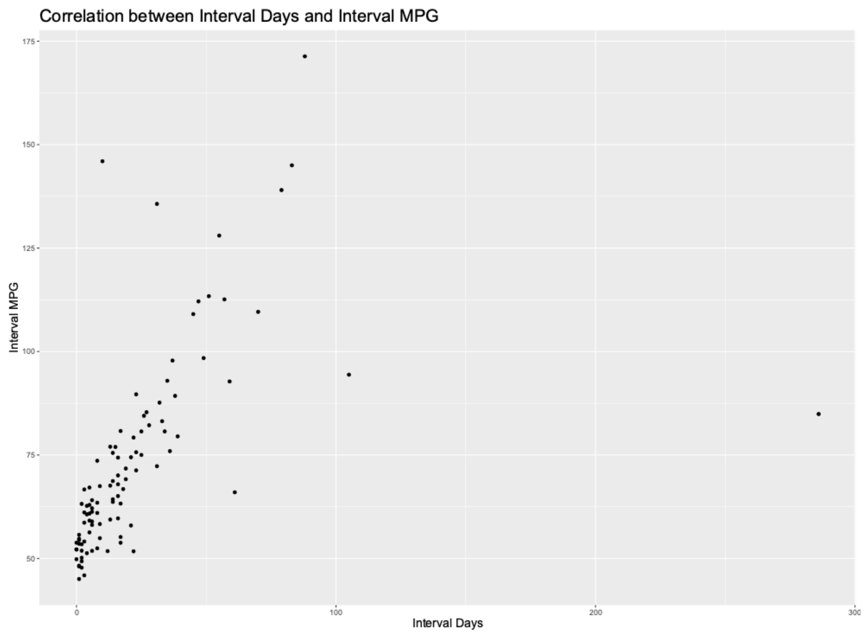

Figure 14: Correlation between Interval Days and Interval MPG

5.3 Multiple Linear Regression Analysis of Gas Consumption

We conducted a multiple linear regression analysis to examine the relationship between gas usage over time intervals (measured in gallons) and other associated factors. After meticulous analysis and comparison of models using stepwise model selection, the formulated model is delineated as follows. This model has $R^2 = 0.6571$ and AIC $= -792.01$, showing a moderately fitted model that can be used to predict *Interval MPG* given some information. The stepwise Comparison can be found in Appendix.

$$\begin{aligned}\text{Interval MPG}_i = &1.731 + (1.186 \times 10^{-3}) \cdot \text{interval\_days}_i \\ &- (1.777 \times 10^{-2}) \cdot \text{seasonSpring} - (4.056 \times 10^{-3}) \cdot \text{seasonSummer} \\ &- (1.300 \times 10^{-2}) \cdot \text{seasonWinter} \end{aligned} \tag{1}$$

- The intercept 1.731 represents the expected value of Interval MPG when all other variables are zero.
- The coefficient for interval_days 1.186×10^{-3} implies that for each additional day between intervals, the MPG is expected to increase by this coefficient's value, holding other factors constant.

- The seasonSpring coefficient -1.777×10^{-2} suggests that in spring, the MPG is expected to decrease by this value compared to the baseline season.
- Similarly, seasonSummer -4.056×10^{-3} and seasonWinter -1.300×10^{-2} coefficients indicate the expected decrease in MPG during these seasons compared to the baseline season.
- The higher MPG observed in fall compared to other seasons could be attributed to the prevalence of long-distance driving during this period. It's plausible that fall activities in Indiana lead to more extended travel distances, resulting in greater gas consumption. This seasonal variance underlines the impact of driving behavior on fuel efficiency.

In this model, the seasonal variables are most likely dummy variables, where each season is represented by a 0 or a 1. If the season is present, it's 1; if not, it's 0. So, for a given observation, only one of these seasonal dummies would be 1, and the rest would be 0.

6 Conclusion

The extensive analysis of interval mileage and gallons consumption for the Prius Prime—a hybrid car operating on both electricity and gas—yields several key insights into fuel efficiency. The study identifies clear seasonal patterns, with variations noted throughout the year. In particular, the year 2020 saw a decline in both mileage and MPG, attributed to the pandemic's impact on driving behavior. However, 2021 experienced a recovery, with the period's highest mileage despite lower MPG figures, suggesting diverse travel lengths as life returned closer to normal.

A paired t-test further reveals that the car's computer tends to overestimate MPG by an average of approximately 4.73 MPG when compared to actual measurements. Additionally, the relationship between interval miles and gallons consumed is not strictly linear but exhibits a complex, curved correlation, indicating that MPG is affected by more than mere distance traveled. This is reflected in the nuances captured by the multiple linear regression model, which shows that longer intervals between usage generally correspond with improved MPG, while seasonal shifts can lead to reduced efficiency. With an R^2 of 0.6571 and an AIC of -792.01 compared to other model which are shown in Appendix, the model is moderately fit, serving as a tool for predicting Interval MPG based on available data.

In conclusion, while the Prius Prime demonstrates efficient fuel management, its MPG is influenced by an array of factors including driving habits, seasonality, and specific vehicle characteristics. These findings highlight the need for a comprehensive approach when evaluating fuel efficiency and constructing a framework to understand hybrid vehicles' gas consumption patterns. As such, ongoing monitoring and analysis, especially beyond 2024, are essential for reaching more concrete conclusions and informing the advancement of hybrid vehicle technologies.

7 Appendix

The **Table 3** below shows the statistical summary of the model suggested above. It includes the t value, p value, and AIC, which is a mathematical method for evaluating how well a model fits the data from which it was generated.

Table 3: Linear Model Results for Interval MPG

Term	Estimate	Std. Error	t value	p value
(Intercept)	1.731	5.149×10^{-3}	336.120	<2e-16
interval_days	1.186×10^{-3}	8.995×10^{-5}	13.182	<2e-16
seasonSpring	-1.777×10^{-2}	5.917×10^{-3}	-3.003	0.00342
seasonSummer	-4.056×10^{-3}	5.669×10^{-3}	-0.715	0.47612
seasonWinter	-1.300×10^{-2}	5.879×10^{-3}	-2.211	0.02944

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0186 on 95 degrees of freedom
Multiple R-squared: 0.6709, Adjusted R-squared: 0.6571
F-statistic: 48.42 on 4 and 95 DF, p-value: <2.2e-16, AIC=-792.01

7.1 Model Selection via Stepwise AIC

The model's AIC of -792.01 suggests its utility in forecasting the efficiency of the car, with caution advised due to potential overfitting or underfitting.

Start: AIC=-688.87
Model: interval_mpg $\sim$ 1

Step	Df	Sum of Sq	RSS	AIC
+ interval_days	1	0.062585	0.037325	< -785.33
+ season	3	0.006899	0.093011	< -690.02
+ day_of_week	6	0.011904	0.088007	< -689.55
none			0.099910	< -688.87
+ overall_miles_lag	1	0.001585	0.098325	< -688.46
+ Month	11	0.017469	0.082442	< -686.08

Step: AIC=-785.33
Model: interval_mpg $\sim$ interval_days

Step	Df	Sum of Sq	RSS	AIC
+ season	3	0.004447	0.032878	< -792.01
none			0.037325	< -785.33
+ overall_miles_lag	1	0.000016	0.037309	< -783.37
+ day_of_week	6	0.003255	0.034070	< -782.45
+ Month	11	0.006459	0.030866	< -782.33
- interval_days	1	0.062585	0.099910	< -688.87

Step: AIC=-792.01
Model: interval_mpg $\sim$ interval_days + season

Action	Df	Sum of Sq	RSS	AIC
none			0.032878	< -792.01
+ day_of_week	6	0.003411	0.029467	< -790.97
+ interval_days:season	3	0.001537	0.031341	< -790.80
+ overall_miles_lag	1	0.000136	0.032742	< -790.43
- season	3	0.004447	0.037325	< -785.33
+ Month	8	0.002011	0.030866	< -782.33
- interval_days	1	0.060133	0.093011	< -690.02

7.2 Statistical Assumption Check for the Suggested Model

Upon applying Box-cox transformation, Normality, linearity, homoscedasticity, and independence checks, alongside GVIF calculations for predictors, are essential for validating the model's assumptions and applicability. Even though the QQ plot does not show a perfect linearity, it is still shown that non-linearity is mitigated. Also, a slight non-linearity would not be a huge problem.

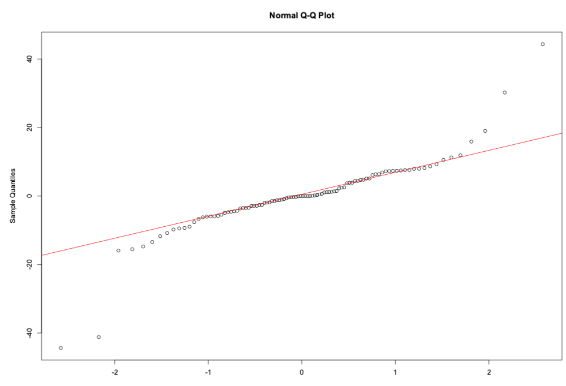

Figure 15: Normality Check - QQ Plot Before the Transformation

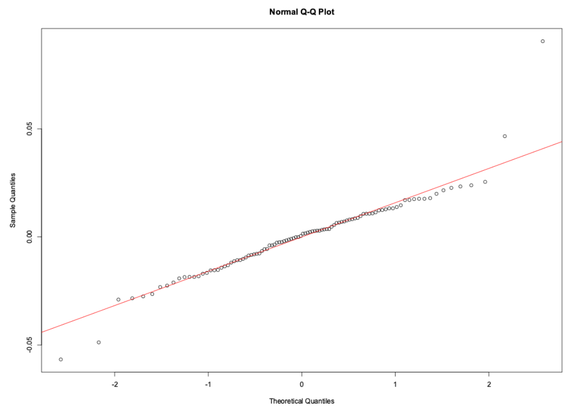

Figure 16: Normality Check - QQ Plot After the Transformation

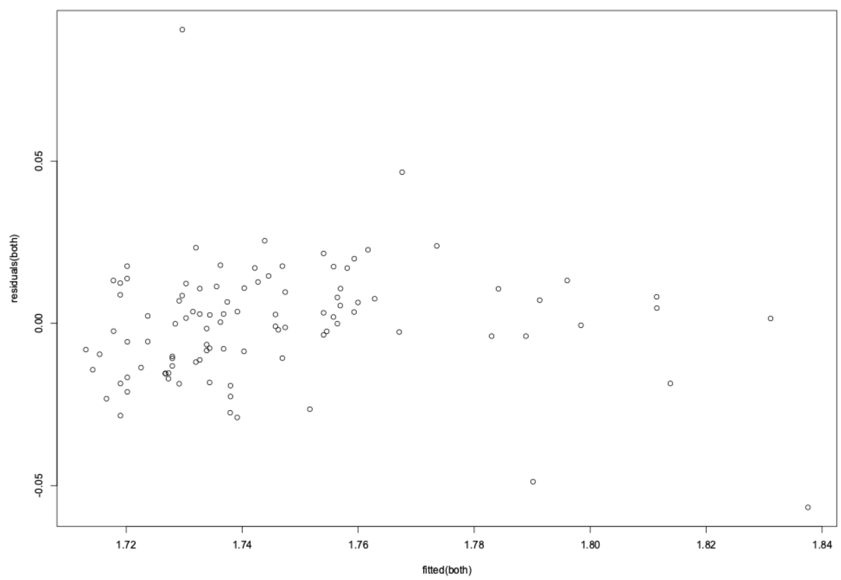

Figure 17: Linearity and Homoscedasticity Check - Residual Plot

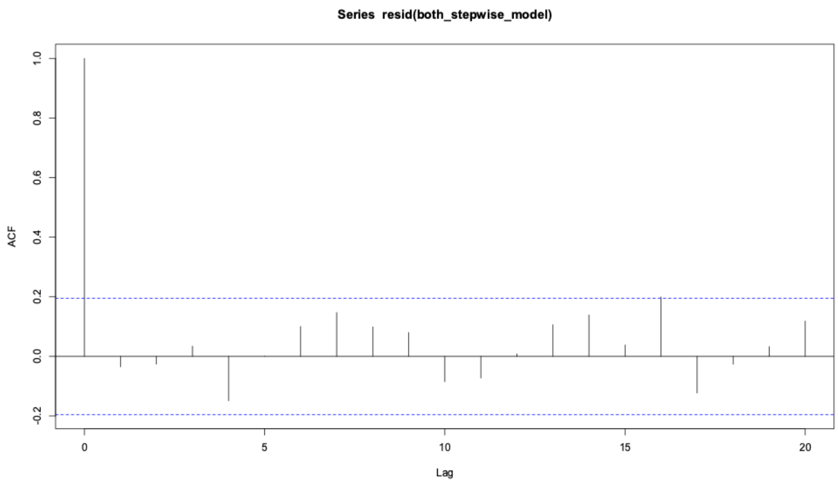

Figure 18: Independence Autocorrelation Check

The computed GVIF values for the predictors are as follows:

Predictor	GVIF Value
Interval Days	1.059535
season	1.059535

Table 4: Generalized Variance Inflation Factors (GVIF) for Each Predictor

To fit the model and want to interpret the expected future gallons in their original scale, it needs to be applied the inverse of the Box-Cox transformation to the predicted values.
The Box-Cox transformation is defined as:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \log(y) & \text{if } \lambda = 0 \end{cases}$$

The inverse transformation, which you apply to the predicted values, is:

$$y = \begin{cases} (\lambda y(\lambda) + 1)^{1/\lambda} & \text{if } \lambda \neq 0 \\ e^{y(\lambda)} & \text{if } \lambda = 0 \end{cases}$$

Where $y(\lambda)$ is the transformed value, y is the original value, and λ is the parameter used for the Box-Cox transformation.