Report: Analysis of US and European Cars

1. Introduction

This report analyzes a dataset containing information about cars from the USA, Europe, and Asia that have been driven for at least 100,000 miles. The dataset includes variables such as horsepower, region, and fuel efficiency measured as miles per gallon (MPG) before and after the cars reached 75,000 miles. The goal of this project is to:

1. Compare the performance and fuel efficiency of US and European cars.
2. Understand how fuel efficiency changes over time.
3. Estimate key parameters (e.g., mean horsepower, MPG) and assess their statistical significance.

The analysis involves data cleaning, visualizations, distribution fitting, and confidence interval estimation. The results provide insights into the differences between US and European cars and their fuel efficiency trends.

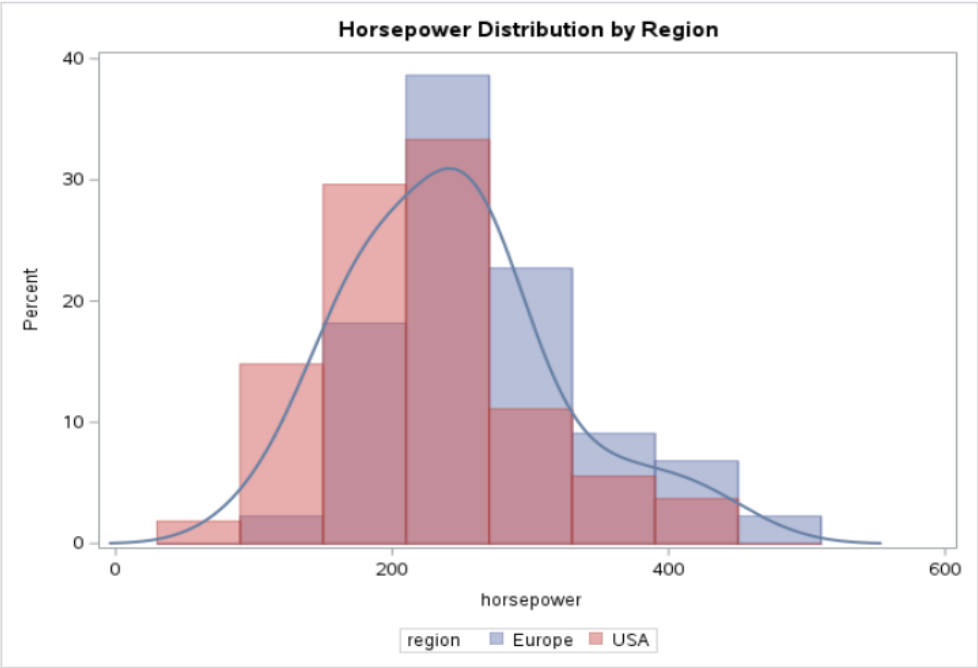
Discussion of Variables

The dataset includes the following variables:

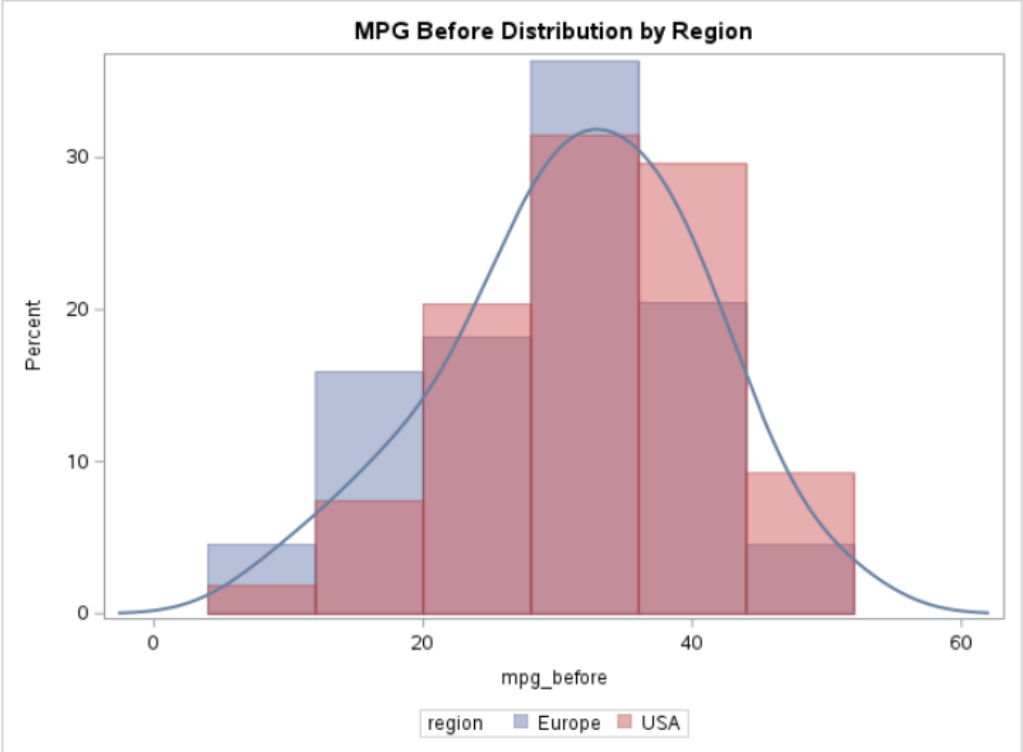
1. Horsepower: A measure of the car's engine power.
2. Region: The origin of the car (USA, Europe, or Asia).
3. MPG Before: The average miles per gallon before the car reached 75,000 miles.
4. MPG After: The average miles per gallon after the car reached 75,000 miles.
5. Average MPG: The average of mpg_before and mpg_after.

2. Key Observations from Visualizations

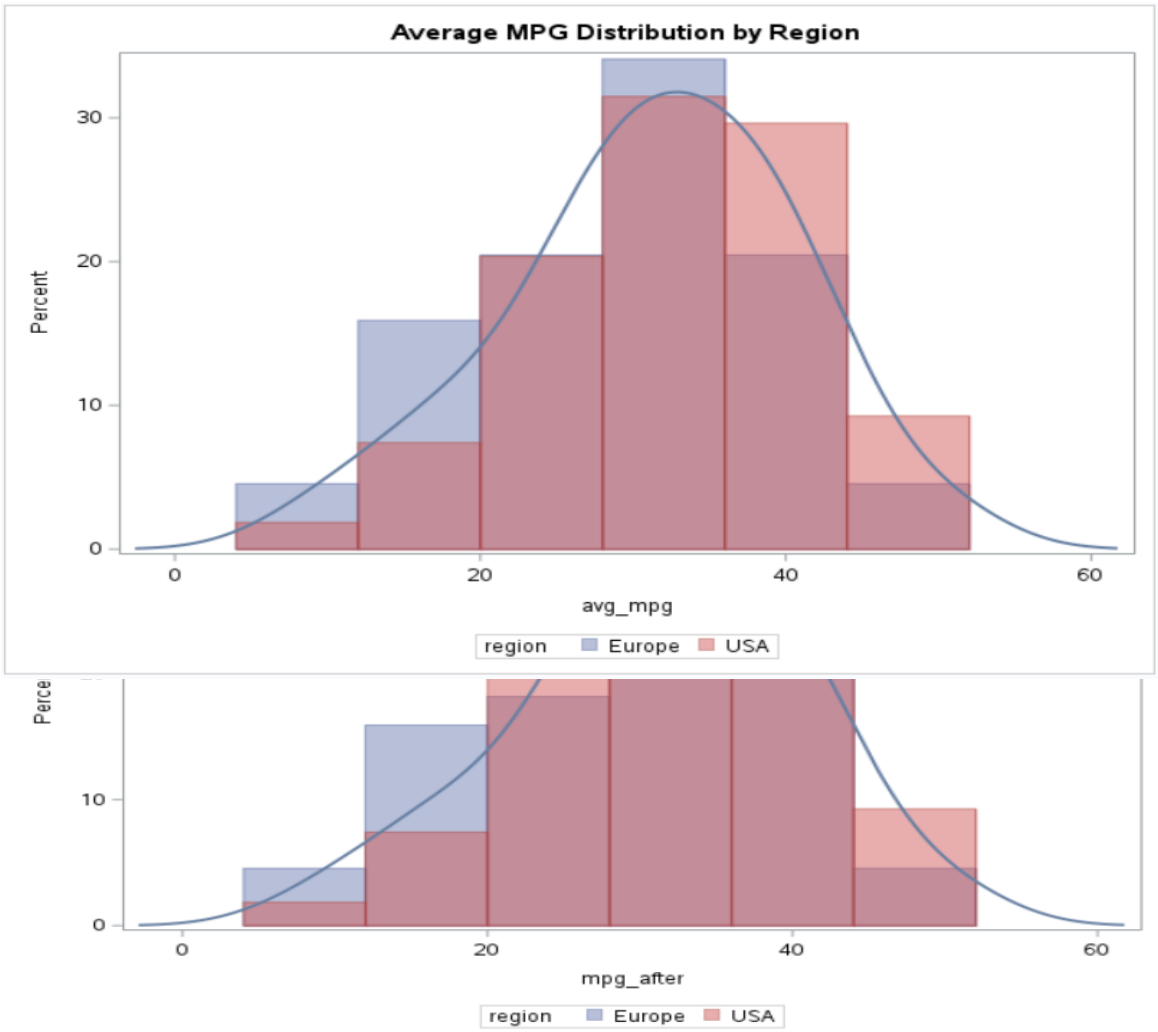
1. **Horsepower**:
    A. European cars are more likely to have higher horsepower, reflecting a focus on performance and luxury.
    B. US cars tend to have lower horsepower, aligning with a focus on fuel efficiency and affordability.
    C. The histogram clearly highlights the differences in engine performance between cars from these two regions.



2. **Fuel Efficiency (MPG)**:
    A. European cars are generally less fuel-efficient than US cars, reflecting a focus on performance and luxury.
    B. US cars tend to have higher MPG, indicating a focus on fuel efficiency and affordability.
    C. The histogram clearly highlights the differences in fuel efficiency between cars from these two regions.



3. **Fuel Efficiency(MPG After):**



A. European cars remain less fuel-efficient than US cars, even after 75,000 miles, reflecting their focus on performance and luxury.

B. US cars maintain their higher MPG, indicating a sustained focus on fuel efficiency and affordability.

C. The histogram highlights the **persistent differences** in fuel efficiency between cars from these two regions over time.

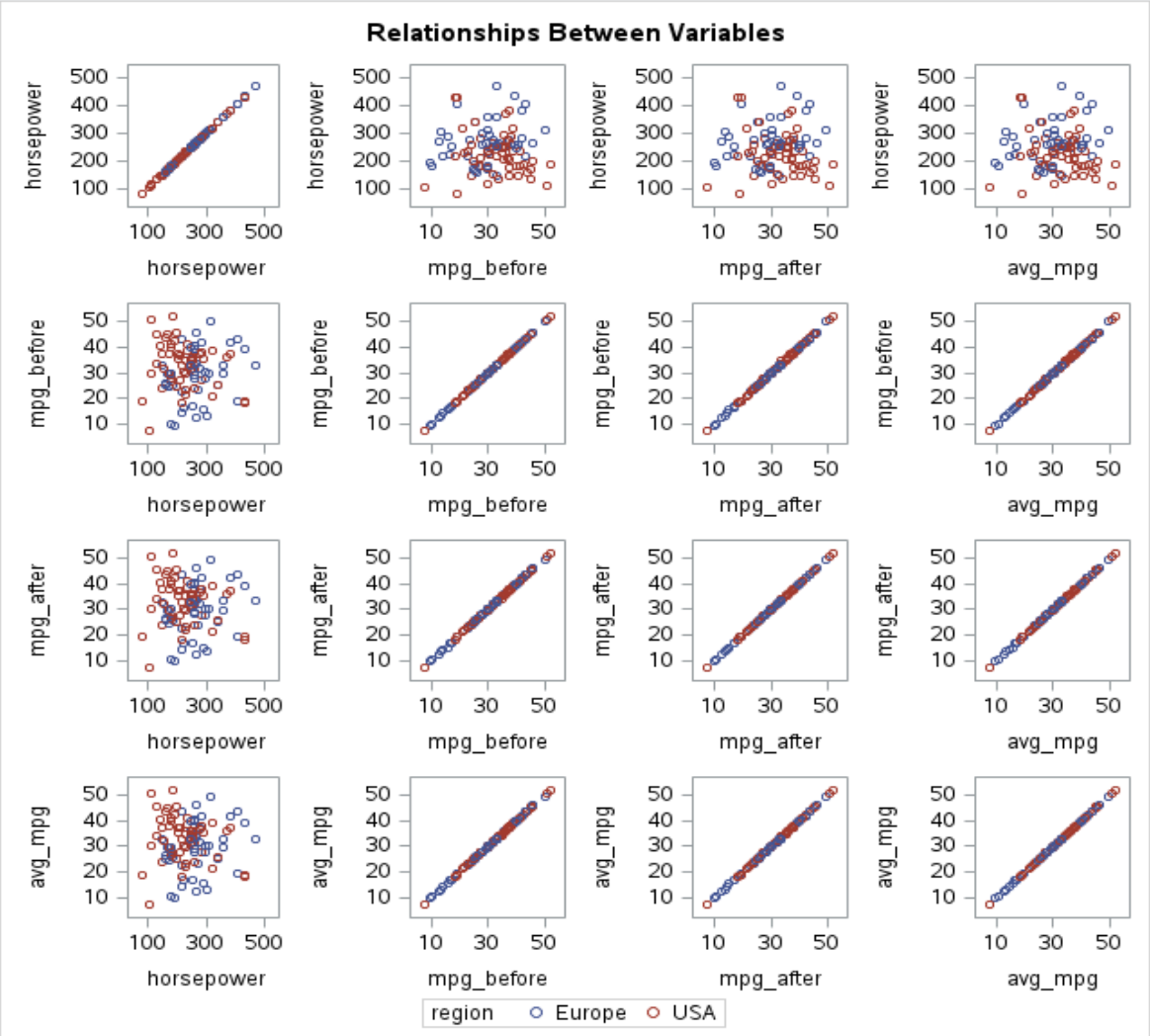
4. Fuel Efficiency(MPG Average):

A. European cars have lower average MPG, reflecting a focus on performance and luxury.

B. US cars have higher average MPG, indicating a focus on fuel efficiency and affordability.

C. The histogram clearly highlights the differences in **average fuel efficiency** between cars from these two regions.

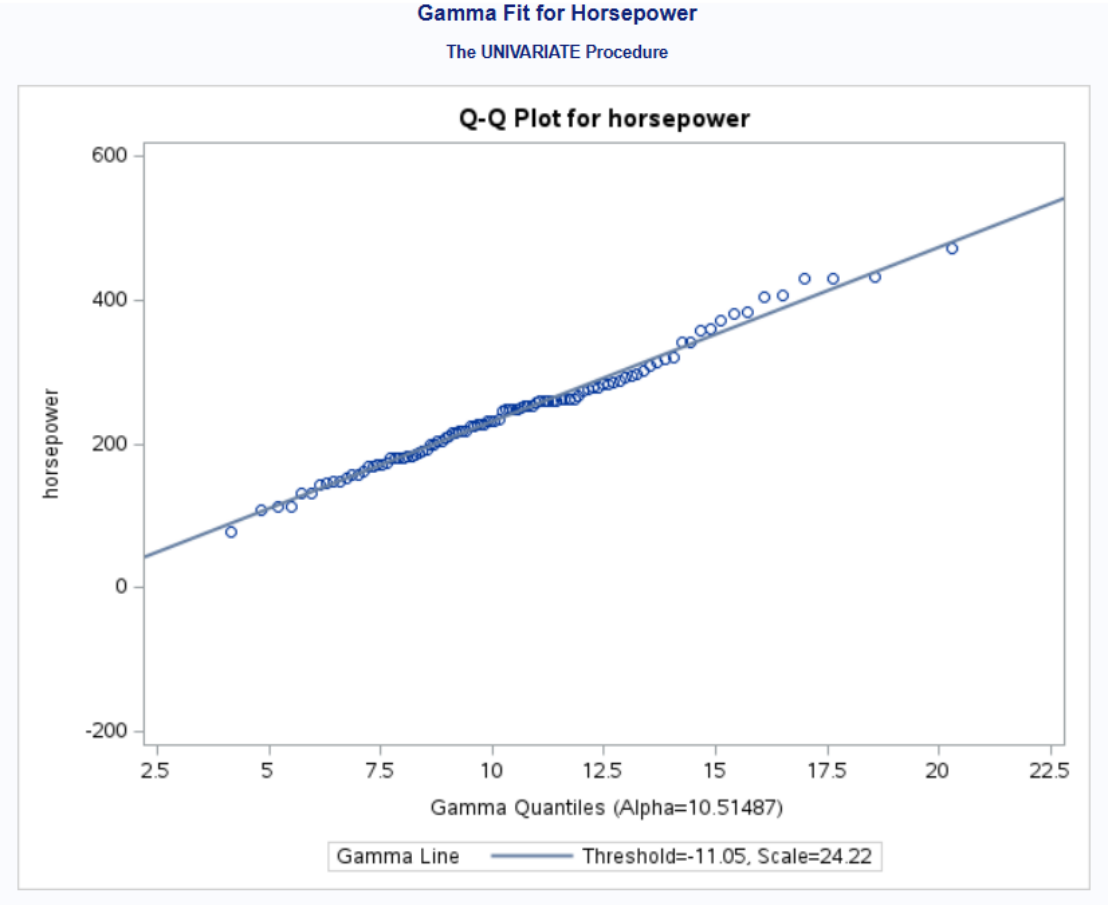
5. **Scatterplots**:
    A. **Horsepower vs. MPG (Before, After, and Average):** There is a negative correlation; higher horsepower leads to lower MPG, which is expected as more powerful engines consume more fuel.US and European cars form distinct clusters in the scatterplots, highlighting differences in performance and efficiency.
    B. **MPG Before vs. MPG After:** There is a strong positive correlation, meaning cars maintain similar fuel efficiency before and after 75,000 miles.
    C. **MPG Variables (Before, After, and Average) vs. Each Other:** These relationships are nearly perfectly linear since avg_mpg is derived from mpg_before and mpg_after.
    D. **Europe vs. USA Differences:** European cars (blue) have lower horsepower but better fuel efficiency, while USA cars (red) show a wider horsepower range and generally lower MPG.



3. Distribution fitting

1. Gamma distribution fitting for horsepower:

The Q-Q plot for horsepower shows how well the data fits a Gamma distribution. If the points closely follow the Gamma line, the distribution fits well. In our case, the points deviate, especially at the higher horsepower values, it means the Gamma distribution might not fully capture the extreme high-performance cars. This suggests the Gamma distribution is a decent fit but may not handle the very high horsepower values perfectly.



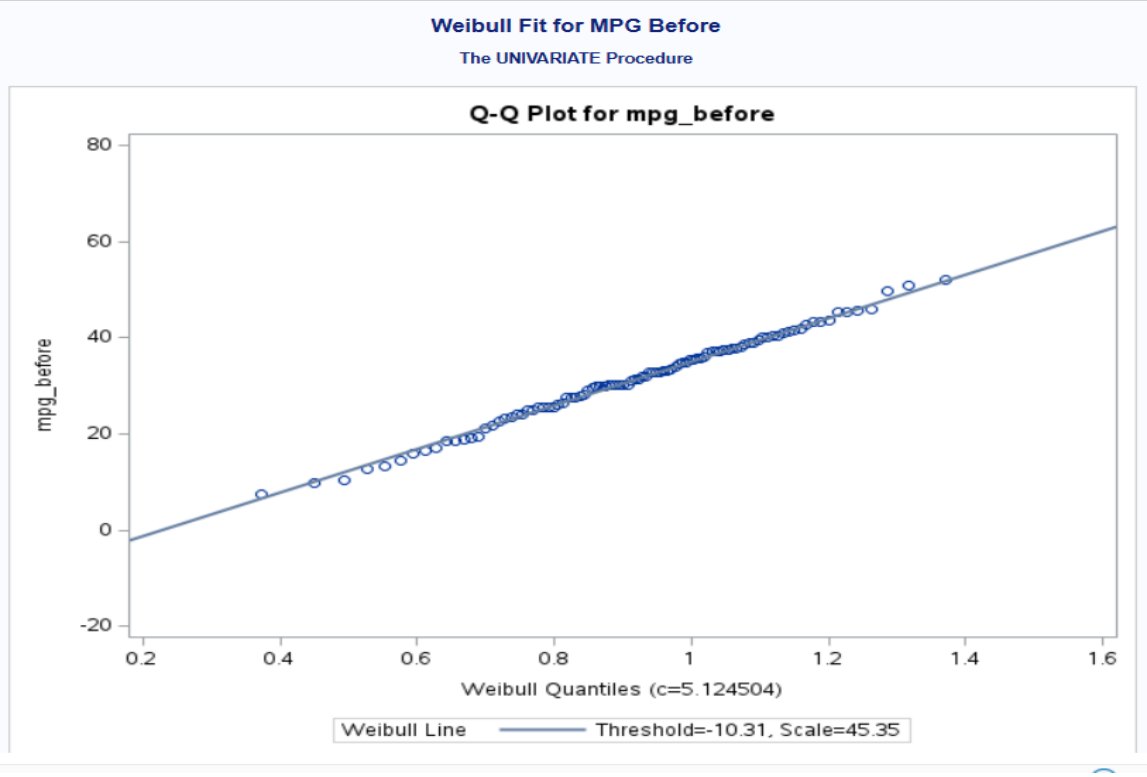
2. **Gamma distribution fitting for MPG After:**
   The Gamma distribution is not a good fit for the mpg_before data because the data is **left-skewed** (-0.3165), while Gamma distributions are typically **right-skewed**. Additionally, the kurtosis (-0.1679) shows the data has **lighter tails** than a Gamma distribution, which usually has heavier tails. The presence of extreme values, like very low MPG (7.40) and high MPG (51.95), also makes the Gamma distribution unsuitable. Overall, the Gamma distribution struggles to model this data well.

Gamma Fit for MPG BEFORE

The UNIVARIATE Procedure
Variable: mpg_before

Moments			
N	98	Sum Weights	98
Mean	31.3638776	Sum Observations	3073.66
Std Deviation	9.4731442	Variance	89.7404611
Skewness	-0.3165259	Kurtosis	-0.167856
Uncorrected SS	105106.721	Corrected SS	8704.82473
Coeff Variation	30.2039956	Std Error Mean	0.95693207

3. **Weibull distribution fitting for MPG After:**
   The Weibull distribution is a better fit for the mpg_before data compared to the Gamma distribution, as it can handle the **left-skewed nature** (-0.3165) and the wide range of MPG values (7.40 to 51.95). The Q-Q plot shows how well the data aligns with the Weibull distribution: if the points follow the Weibull line closely, the fit is good. However, deviations at the tails, especially for extreme low or high MPG values, suggest the Weibull distribution may not fully capture these extremes. Overall, the Weibull distribution is a reasonable choice, but further checks for outliers or alternative distributions like Log-Normal might improve the fit.



4. **Confidence Inteval:**
   1. **90% CI for mean horsepower :**
      The output shows that the **mean horsepower** is **245.89**, with a 90% confidence interval ranging from **232.23** to **259.54**. This means we are 90% confident that the true average horsepower of all cars in the population lies within this range. The analysis includes 99 observations, and the results suggest that the average horsepower is relatively high, with some variability across regions.

90% Confidence Interval for Mean Horsepower (All Regions)

The MEANS Procedure

Analysis Variable : horsepower			
N	Mean	Lower 90% CL for Mean	Upper 90% CL for Mean
99	245.8888889	232.2328591	259.5449187

2. **95% CI for avg_mpg (Europe vs. USA):**
   The output shows that the **mean average MPG** is **29.43** for Europe and **32.95** for the USA, with a **difference of -3.52**. The 95% CI for this difference ranges from **-7.29** to **0.26**, indicating that US cars are likely more fuel-efficient, but the difference is not statistically significant (p-value > 0.05). The analysis suggests that while US cars tend to have higher MPG, the difference is not strong enough to be conclusive.

95% CI for Average MPG: Europe vs. USA

The TTEST Procedure

Variable: avg_mpg

region	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Europe		44	29.4318	10.0107	1.5092	9.6700	49.6450
USA		54	32.9493	8.8108	1.1990	7.1800	51.9350
Diff (1-2)	Pooled		-3.5174	9.3673	1.9024		
Diff (1-2)	Satterthwaite		-3.5174		1.9275		

region	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Europe		29.4318	26.3883	32.4753	10.0107	8.2711	12.6838
USA		32.9493	30.5444	35.3541	8.8108	7.4066	10.8770
Diff (1-2)	Pooled	-3.5174	-7.2937	0.2588	9.3673	8.2091	10.9090
Diff (1-2)	Satterthwaite	-3.5174	-7.3489	0.3140			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	96	-1.85	0.0675
Satterthwaite	Unequal	86.465	-1.82	0.0715
Cochran	Unequal	.	-1.82	0.0745

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	43	53	1.29	0.3745

3. **99% CI for mean difference (mpg_after - mpg_before):**
   The code calculates a **99% confidence interval (CI)** for the mean difference between mpg_after and mpg_before. The output shows that the **mean difference** is **0.0122**, with a 99% CI ranging from **-0.0576** to **0.0821**. This means there is no significant change in fuel efficiency after 75,000 miles, as the interval includes zero. The p-value of **0.6461** further confirms that the difference is not statistically significant, suggesting that fuel efficiency remains relatively stable over time.

99% CI for MPG After-Before Difference

The TTEST Procedure

Difference: mpg_after - mpg_before

N	Mean	Std Dev	Std Err	Minimum	Maximum
98	0.0122	0.2632	0.0266	-0.8600	0.5500

Mean	99% CL Mean		Std Dev	99% CL Std Dev	
0.0122	-0.0576	0.0821	0.2632	0.2217	0.3218

DF	t Value	Pr > |t|
97	0.46	0.6461