**CO2 Emission by Vehicles: An Empirical Analysis**

**Akhil bonthinayanivari, Lakshmi Sri Naga Sumanth Tripuraneni, Dodda Naveen Kumar, Sai Kiran Bussi**

1. **Hypothesis testing of mean and proportions of a sample of the population**

**1.1 Statistical Analysis of Engine Size in Vehicles**

The dataset encompasses a range of attributes related to vehicle emissions and fuel characteristics in Canada. Our analysis centered on the "Engine Size(L)" variable, a key indicator of a vehicle's power and fuel consumption characteristics. From this dataset, 100 samples were randomly selected, providing a representative subset for statistical examination.

**1.1.1 Sample Overview:**

* Sample Size: 100 vehicles
* Sample Mean Engine Size: 2.965 Liters
* Sample Standard Deviation: 1.235 Liters
* This sample provides a snapshot of the engine size distribution among various vehicles, capturing a range of engine types from compact cars to larger vehicles.

**1.1.2 Hypothesis Testing for Mean Engine Size:**

**Hypotheses:**

* Null Hypothesis (H0): The average engine size in the population is 3.0 liters.
* Alternative Hypothesis (H1): The average engine size in the population differs from 3.0 liters.

**Test Utilized: Two-tailed t-test**

**Test Statistics: t-statistic = -0.283, p-value = 0.777**

* **Conclusion**: Given the high p-value, we do not reject the null hypothesis. There's insufficient statistical evidence to suggest that the average engine size differs from 3.0 liters in the population.
* **Implication**: This finding implies that, under the assumptions of our sampling and statistical methods, the average engine size among vehicles might well be around the 3.0-liter mark.

**95% Confidence Interval for Mean Engine Size:**

* Interval Range: Approximately 2.72 to 3.21 liters
* **Interpretation**: This range indicates that we can be 95% confident that the true population mean of engine size falls within this interval. The inclusion of the hypothesized mean (3.0 liters) within this interval reinforces our hypothesis test result.
  + 1. **Proportion Analysis: Engine Size in Vehicles**

**Analysis Overview:**

In addition to mean analysis, we conducted a proportion analysis to deepen our understanding of the distribution of engine sizes in the dataset. Specifically, we categorized engine sizes into two groups: engines larger than 3.0 liters and those equal or smaller. This categorization allowed us to analyze the proportion of vehicles falling into each category.

**Statistical Analysis:**

**Proportion Calculation:**

* Proportion of Vehicles with Engine Size > 3.0 Liters: 44%
* This proportion indicates that less than half of the sampled vehicles have an engine size larger than 3.0 liters.

**Hypothesis Testing for Proportion:**

**Hypotheses:**

* Null Hypothesis (H0​): The proportion of vehicles with engine sizes larger than 3.0 liters is 50%.
* Alternative Hypothesis (H1): The proportion is not 50%.

**Test Utilized: Two-tailed z-test for proportion**

**Test Statistics: z-statistic = -1.203, p-value = 0.232**

**Conclusion:** The p-value is greater than the conventional alpha level of 0.05, indicating insufficient evidence to reject the null hypothesis. Therefore, we cannot conclusively say that the proportion of vehicles with engine sizes larger than 3.0 liters is different from 50%.

**95% Confidence Interval for Proportion:**

* Interval Range: Approximately 34.3% to 53.7%
* Interpretation: This interval suggests that we can be 95% confident that the true proportion of vehicles with engine sizes larger than 3.0 liters in the population lies within this range.

**Overall Conclusion:**

The proportion analysis complements our earlier mean analysis, offering a broader perspective on engine size distribution. While the mean analysis suggested an average engine size close to 3.0 liters, the proportion analysis reveals that a significant segment of the sample (44%) consists of vehicles with larger engine sizes. However, this result does not significantly deviate from the hypothesized 50%, suggesting a relatively balanced distribution around the 3.0-liter threshold. The confidence interval provides additional context, indicating that the actual proportion in the population is likely within the range of 34.3% to 53.7%.

**Implications:**

This analysis underscores the diversity in engine sizes among vehicles. Understanding this distribution is crucial for stakeholders in the automotive industry, policymakers, and environmental analysts, as it influences fuel efficiency, emissions, and market trends. While the average engine size is an important metric, the proportion of vehicles exceeding a certain engine size threshold can offer valuable insights into consumer preferences and environmental impacts.

* 1. **Statistical Analysis of Engine Size in Vehicles:**

The "Cylinders" variable in the dataset is a crucial determinant of a vehicle's engine power and overall performance. A sample of 100 vehicles was randomly selected from this dataset to analyze the distribution of cylinders in these vehicles.

**1.2.1 Sample Overview:**

* Sample Size: 100 vehicles
* Sample Mean: 5.78 Cylinders
* Sample Standard Deviation: 2.09 Cylinders

This sample provides a varied representation of engine configurations, ranging from lower-cylinder, potentially more efficient engines, to higher-cylinder, more powerful engines.

**1.2.2 Hypothesis Testing for Mean Cylinder Count:**

**Hypotheses:**

* Null Hypothesis (H0): The average number of cylinders in the population is 5.78.
* Alternative Hypothesis (H1): The average number of cylinders in the population is not 5.78.

**Test Utilized: Two-tailed t-test**

**Test Statistics: t-statistic = 0.0, p-value = 1.0**

* Conclusion: The p-value indicates that there's no significant evidence to reject the null hypothesis, suggesting that 5.78 is a representative average number of cylinders in the population.
* Implication: This result implies that the sampled vehicles typically have around 5 to 6 cylinders, reflecting a range of engine sizes commonly found in consumer vehicles.

**95% Confidence Interval for Mean Cylinder Count:**

* Interval Range: Approximately 5.36 to 6.20 cylinders
* Interpretation: This confidence interval suggests that the true average number of cylinders in the population likely falls within this range, further supporting the mean analysis.

**1.2.3 Proportion Analysis:**

**Proportion of Vehicles with More than 6 Cylinders: 24%**

**Proportion Hypothesis Test:**

* Null Hypothesis (H0): The proportion of vehicles with more than 6 cylinders is 50%.
* Alternative Hypothesis (H1): The proportion of vehicles with more than 6 cylinders is not 50%.

**z-statistic: -6.057, p-value ≈ 0.00**

* Conclusion: The extremely low p-value provides strong evidence against the null hypothesis, indicating that the proportion of vehicles with more than 6 cylinders significantly differs from 50%.

**95% Confidence Interval:**

* Range: Approximately 15.6% to 32.4%
* Interpretation: This range suggests with 95% confidence that the true proportion of vehicles with more than 6 cylinders in the population is between 15.6% and 32.4%.

**1.3 Statistical Analysis of Fuel Consumption Comb (L/100 km) in Vehicles:**

The "Fuel Consumption Comb (L/100 km)" variable is a key measure of a vehicle's fuel efficiency. This analysis is based on a sample of 100 vehicles, providing insight into fuel consumption patterns.

**1.3.1 Sample Overview:**

* Sample Size: 100 vehicles
* Sample Mean: 11.11 L/100 km
* Sample Standard Deviation: 3.24 L/100 km
* This sample encompasses a broad range of fuel consumption rates, indicative of various vehicle types and their efficiency levels.

**1.3.2 Hypothesis Testing for Mean Fuel Consumption:**

* Null Hypothesis (H0): The average fuel consumption in the population is 11.11 L/100 km.
* Alternative Hypothesis (H1): The average fuel consumption in the population is not 11.11 L/100 km.

**Test Utilized: Two-tailed t-test**

**Test Statistics: t-statistic = -0.012, p-value = 0.990**

* Conclusion: The high p-value suggests no significant difference between the sample mean and the hypothesized mean, indicating that 11.11 L/100 km is a representative average of fuel consumption.
* Implication: This finding implies that the sampled vehicles, on average, have a fuel consumption rate around 11.11 L/100 km, reflecting a range of fuel efficiency levels.

**95% Confidence Interval for Mean Fuel Consumption:**

* Interval Range: Approximately 10.46 to 11.75 L/100 km
* Interpretation: This interval suggests that the true average fuel consumption in the population likely falls within this range, providing a more precise estimate than the sample mean alone.

​

**1.3.3 Proportion Analysis:**

Proportion of Vehicles with Fuel Consumption Greater than 10 L/100 km: 54%

**Proportion Hypothesis Test:**

* Null Hypothesis (H0): The proportion of vehicles with fuel consumption greater than 10 L/100 km is 50%.
* Alternative Hypothesis (H1): The proportion of vehicles with fuel consumption greater than 10 L/100 km is not 50%.

**z-statistic: 0.799, p-value: 0.426**

* Conclusion: The p-value being higher than 0.05 indicates insufficient evidence to reject the null hypothesis, suggesting that the proportion of vehicles with higher fuel consumption does not significantly differ from 50%.

**95% Confidence Interval:**

* Range: Approximately 44.2% to 63.8%
* Interpretation: This confidence interval indicates that the true proportion of vehicles with a fuel consumption rate above 10 L/100 km is likely between 44.2% and 63.8% with 95% confidence.

**1.4 Statistical Analysis of Fuel Consumption Comb (mpg) in Vehicles:**

This analysis delves into the "Fuel Consumption Comb (mpg)" variable, which measures the fuel efficiency of vehicles in miles per gallon (mpg). This metric is pivotal in understanding the environmental impact and operating efficiency of vehicles. A random sample of 100 vehicles from the dataset provides a substantial basis for examining the distribution of fuel efficiency across different vehicle types.

**1.4.1 Sample Characteristics:**

* Sample Size: 100 vehicles
* Sample Mean: 28.68 mpg
* Sample Standard Deviation: 8.23 mpg
* Insight: This sample captures a broad spectrum of fuel efficiency levels, ranging from highly efficient vehicles to those less efficient. This diversity reflects the varying fuel consumption demands based on vehicle design, intended use, and technology.

**1.4.2 Detailed Mean Analysis:**

**Formulated Hypotheses:**

* **Null Hypothesis (H0): The population mean fuel consumption is 28.68 mpg.**
* **Alternative Hypothesis (H1): The population mean fuel consumption is not 28.68 mpg.**

**Hypothesis Testing:**

* t-statistic: 0.0
* p-value: 1.0
* Interpretation: The high p-value indicates a strong alignment of the sample mean with the hypothesized population mean, suggesting that 28.68 mpg accurately represents the average fuel consumption.

**95% Confidence Interval:**

* Range: Approximately 27.05 to 30.31 mpg
* Analysis: This confidence interval reinforces the hypothesis test's findings, indicating a high likelihood that the true average fuel consumption in the broader population falls within this range.

**1.4.3 Proportion Analysis:**

Proportion Metric: 31% of vehicles have a fuel consumption greater than 30 mpg.

**Hypotheses for Proportion:**

* Null Hypothesis (H0): 50% of vehicles have a fuel consumption greater than 30 mpg.
* Alternative Hypothesis (H1): Less than 50% of vehicles have a fuel consumption greater than 30 mpg.

**Proportion Test Findings:**

* z-statistic: -4.088
* p-value: ≈ 0.00
* Conclusion: The extremely low p-value strongly refutes the null hypothesis, indicating that the actual proportion of vehicles exceeding 30 mpg is significantly lower than 50%.

**95% Confidence Interval for Proportion:**

* Range: 21.9% to 40.1%
* Significance: This interval provides a precise estimate, showing a high level of confidence that the true proportion of more fuel-efficient vehicles (greater than 30 mpg) in the population is well below the 50% mark.

1. **Statistical Comparison of Two Sample Means and SD**

**2.1 Comparison 1: "Engine Size(L)" vs "Cylinders" and "Fuel Consumption Comb (L/100 km)"**

**Data Overview and Sample Characteristics:**

In this comparative analysis, we examine the relationships between three critical vehicle attributes: "Engine Size(L)", "Cylinders", and "Fuel Consumption Comb (L/100 km)". A random sample of 100 vehicles was analyzed for each attribute.

**Engine Size(L):**

* Sample Mean: 2.965 Liters
* Sample Standard Deviation: 1.235 Liters
* This suggests a moderate average engine size with some variability, indicating a mix of smaller and larger engines in the sample.

**Cylinders:**

* Sample Mean: 5.78
* Sample Standard Deviation: 2.09
* The data reflects a variety in the number of cylinders, covering a range from economy cars to more powerful vehicles.

**Fuel Consumption Comb (L/100 km):**

* **Sample Mean: 11.106 L/100 km**
* **Sample Standard Deviation: 3.235 L/100 km**
* **This indicates a range in fuel efficiency among the sampled vehicles.**

**Statistical Analysis and Hypothesis Testing:**

**Comparison of Engine Size and Cylinders:**

* **T-test Result: -11.588, p-value: ≈ 0.00**
* Interpretation: The negative t-value and extremely low p-value indicate a statistically significant difference between the means of engine size and the number of cylinders. This suggests that these two attributes, while related, vary independently across different vehicle types.

**95% Confidence Interval:**

* Engine Size(L): 2.72 to 3.21 Liters
* Cylinders: 5.36 to 6.20
* These intervals provide a range of plausible mean values for the population from which the sample was drawn.

**Comparison of Engine Size and Fuel Consumption Comb (L/100 km):**

* **T-test Result: -23.509, p-value: ≈ 0.00**
* Interpretation: The highly negative t-statistic and virtually zero p-value strongly indicate that engine size and fuel consumption are significantly different. This aligns with expectations as larger engines typically result in higher fuel consumption.
* **95% Confidence Interval for Fuel Consumption Comb (L/100 km):** 10.46 to 11.75 L/100 km
* This interval indicates where the true mean of fuel consumption likely lies for the population.

**Overall Conclusion:**

The comparison reveals distinct statistical differences between engine size, the number of cylinders, and fuel consumption. The significant differences in means, as evidenced by the t-tests, highlight the diversity in vehicle specifications within the sample. Understanding these differences is crucial for automotive manufacturers, consumers, and policymakers, as they relate to vehicle performance, efficiency, and environmental impact. The confidence intervals provide a range of likely mean values, offering a broader perspective than the sample means alone.

**2.2 Comparison 2: "Cylinders" vs "Fuel Consumption Comb (L/100 km)" and "Fuel Consumption Comb (mpg)"**

**Data Overview and Sample Characteristics:**

This comparison focuses on the relationship between the number of cylinders in a vehicle's engine and its fuel consumption, both in liters per 100 kilometers (L/100 km) and miles per gallon (mpg). A sample of 100 vehicles was analyzed for each attribute.

**Cylinders:**

* Sample Mean: 5.3
* Sample Standard Deviation: 1.514
* The data suggests a moderate number of cylinders, typical for a range of consumer vehicles.

**Fuel Consumption Comb (L/100 km):**

* Sample Mean: 10.933 L/100 km
* Sample Standard Deviation: 3.109 L/100 km
* This range indicates varying levels of fuel efficiency among the sampled vehicles.

**Fuel Consumption Comb (mpg):**

* Sample Mean: 27.39 mpg
* Sample Standard Deviation: 7.398 mpg
* The spread in mpg values reflects a diversity in fuel consumption efficiency.

**Comparison of Cylinders and Fuel Consumption Comb (L/100 km):**

* T-test Result: -16.290, p-value: ≈ 0.00
* Interpretation: The significant negative t-value and extremely low p-value indicate a statistically significant difference between the number of cylinders and fuel consumption in L/100 km. This suggests a correlation where vehicles with more cylinders tend to consume more fuel.
* 95% Confidence Interval:
* Cylinders: 5.0 to 5.6
* Fuel Consumption Comb (L/100 km): 10.32 to 11.55 L/100 km

**Comparison of Cylinders and Fuel Consumption Comb (mpg):**

* T-test Result: -29.254, p-value: ≈ 0.00
* Interpretation: The highly negative t-statistic and virtually zero p-value strongly suggest that the number of cylinders in a vehicle significantly impacts its fuel consumption in mpg. Typically, more cylinders correlate with lower fuel efficiency (fewer mpg).
* 95% Confidence Interval for Fuel Consumption Comb (mpg): 25.92 to 28.86 mpg

**Overall Conclusion:**

The analysis robustly indicates that there are significant differences in fuel consumption—both in L/100 km and mpg—based on the number of cylinders in a vehicle. The results highlight the direct impact of engine size (as reflected by cylinder count) on fuel efficiency, with larger engines generally leading to higher consumption and lower mpg.

**2.3 Comparison 3: "Fuel Consumption Comb (L/100 km)" vs "Fuel Consumption Comb (mpg)" and "Engine Size(L)"**

**Data Overview and Sample Characteristics:**

This analysis examines the relationships between fuel consumption measured in both liters per 100 kilometers (L/100 km) and miles per gallon (mpg), and the engine size of vehicles. The sample consists of 100 vehicles for each attribute.

**Fuel Consumption Comb (L/100 km):**

* Sample Mean: 10.589 L/100 km
* Sample Standard Deviation: 2.799 L/100 km
* This suggests a range of fuel efficiency levels, indicative of varied vehicle types.

**Fuel Consumption Comb (mpg):**

* Sample Mean: 27.94 mpg
* Sample Standard Deviation: 8.462 mpg
* The wide standard deviation reflects significant variability in fuel efficiency among the sampled vehicles.

**Engine Size(L):**

* Sample Mean: 3.135 Liters
* Sample Standard Deviation: 1.464 Liters
* The data suggests a range of engine sizes, from smaller, more efficient engines to larger, more powerful ones.

**Comparison of Fuel Consumption (L/100 km) and Fuel Consumption (mpg):**

* T-test Result: -19.466, p-value: ≈ 0.00
* Interpretation: The negative t-value and very low p-value indicate a statistically significant difference between the two measures of fuel consumption. This difference is expected due to the inverse nature of the two metrics (higher L/100 km indicates lower efficiency, whereas higher mpg indicates higher efficiency).
* 95% Confidence Interval:
* Fuel Consumption (L/100 km): 10.03 to 11.14 L/100 km
* Fuel Consumption (mpg): 26.26 to 29.62 mpg

**Comparison of Fuel Consumption (L/100 km) and Engine Size(L):**

* T-test Result: 23.594, p-value: ≈ 0.00
* Interpretation: The positive t-value and extremely low p-value suggest a significant difference between fuel consumption in L/100 km and engine size. Generally, larger engines tend to have higher fuel consumption (more L/100 km).
* 95% Confidence Interval for Engine Size(L): 2.84 to 3.43 Liters

**Overall Conclusion:**

The comparison indicates distinct statistical differences and relationships between the measures of fuel consumption and engine size. The significant difference in fuel consumption metrics highlights the importance of considering both L/100 km and mpg for a comprehensive understanding of a vehicle's fuel efficiency. Additionally, the relationship between engine size and fuel consumption (L/100 km) is evident, underlining the impact of engine size on fuel efficiency.

**2.4 Comparison 4: "Fuel Consumption Comb (mpg)" vs "Engine Size(L)" and "Cylinders"**

**Data Overview and Sample Characteristics:**

In this analysis, we explore the relationships between "Fuel Consumption Comb (mpg)", "Engine Size(L)", and "Cylinders". These metrics are crucial in understanding the performance and efficiency of vehicles. The sample includes 100 vehicles for each attribute.

**Fuel Consumption Comb (mpg):**

* Sample Mean: 28.44 mpg
* Sample Standard Deviation: 7.337 mpg
* This indicates variability in fuel efficiency, with some vehicles achieving high mpg and others lower.

**Engine Size(L):**

* Sample Mean: 3.192 Liters
* Sample Standard Deviation: 1.397 Liters
* The data covers a range of engine sizes, from smaller, efficient engines to larger, more powerful ones.

**Cylinders**:

* Sample Mean: 5.62
* Sample Standard Deviation: 1.984
* This suggests a diversity in engine configurations, reflecting different vehicle types.

**Comparison of Fuel Consumption (mpg) and Engine Size(L):**

* T-test Result: 33.806, p-value: ≈ 0.00
* Interpretation: The positive t-value and extremely low p-value indicate a significant difference between fuel consumption in mpg and engine size. Typically, larger engines result in lower fuel efficiency (fewer mpg).
* 95% Confidence Interval:
* Fuel Consumption (mpg): 26.98 to 29.90 mpg
* Engine Size(L): 2.91 to 3.47 Liters
* Confidence Interval Insights: The intervals provide a precise estimate of the range within which the true average values for fuel consumption and engine size likely fall in the population. The mpg interval particularly underscores the variation in fuel efficiency among different vehicle types.

**Comparison of Fuel Consumption (mpg) and Cylinders:**

* T-test Result: 30.026, p-value: ≈ 0.00
* Interpretation: The substantial t-value and nearly zero p-value strongly suggest a significant difference between fuel consumption in mpg and the number of cylinders. More cylinders generally correlate with lower mpg.
* 95% Confidence Interval for Cylinders: 5.23 to 6.01
* Confidence Interval Insights: This interval suggests with high confidence that the true mean number of cylinders in the population lies in this range. It reflects the expected trade-off between engine power (cylinder count) and fuel efficiency.

**Overall Conclusion:**

The comparison reveals a clear and significant distinction between fuel efficiency (mpg) and both engine size and cylinder count. Larger engines and higher cylinder counts are associated with lower fuel efficiency. The confidence intervals provide a detailed understanding of these relationships, offering precise ranges for the average values of these metrics in the vehicle population.

1. **Correlation Analysis:**

**3.1 Correlation Analysis: "Engine Size(L)" vs "Cylinders" and "Fuel Consumption Comb (L/100 km)"**

**Overview of Correlation Analysis:**

This analysis aims to uncover the relationships between "Engine Size(L)" and two other critical vehicle attributes: "Cylinders" and "Fuel Consumption Comb (L/100 km)". Correlation coefficients and their significance (p-values) were calculated, along with confidence intervals for these coefficients, to understand the strength and direction of these relationships.

**Correlation between Engine Size and Cylinders:**

* **Correlation Coefficient: -0.116**
* **p-value: 0.250**
* Interpretation: The correlation coefficient is relatively small and negative, suggesting a very weak inverse relationship between engine size and the number of cylinders. However, the p-value is above the conventional threshold of 0.05, indicating that this correlation is not statistically significant. This implies that in this sample, engine size and the number of cylinders does not have a strong or significant linear relationship.
* **Confidence Interval for Correlation: [-0.305, 0.082]**
* Interpretation of CI: The confidence interval includes zero, which aligns with the p-value in suggesting that the observed correlation could be due to random variation in the sample. The range of the interval indicates that any true correlation between these variables in the population is likely to be weak.

**Correlation between Engine Size and Fuel Consumption Comb (L/100 km):**

* **Correlation Coefficient: Approximately 0.00**
* **p-value: ≈ 1.00**
* Interpretation: The correlation coefficient is effectively zero, indicating no linear relationship between engine size and fuel consumption measured in L/100 km. The p-value, being close to 1, reinforces the lack of statistical significance in this correlation. This result is surprising as one might expect engine size to have some impact on fuel consumption.
* **Confidence Interval for Correlation: [-0.196, 0.196]**
* Interpretation of CI: The confidence interval is centered around zero and is quite narrow, further confirming the absence of a meaningful linear relationship between engine size and fuel consumption in L/100 km in this sample. This interval suggests that if there is any true correlation in the broader population, it is likely to be very weak.

**Overall Conclusions and Implications:**

* Engine Size and Cylinders: The lack of a significant correlation suggests that other factors may be more influential in determining the number of cylinders in a vehicle's engine than just engine size alone.
* Engine Size and Fuel Consumption: The absence of a significant correlation between engine size and fuel consumption in L/100 km is counterintuitive. It suggests that factors other than engine size might be playing a more crucial role in determining a vehicle's fuel consumption, such as vehicle weight, aerodynamics, or technological advancements in engine efficiency.
* Broader Perspective: These findings highlight the complexity of vehicle design and performance characteristics. They suggest that simplistic assumptions about relationships between engine size, cylinders, and fuel consumption may not hold true without considering other interacting factors.

In summary, this correlation analysis provides a nuanced understanding of the relationships between key vehicle attributes. It underscores the multifaceted nature of vehicle design and performance, inviting a broader examination of how various components interact to define a vehicle's characteristics.

**3.2 Correlation Analysis: "Cylinders" vs "Fuel Consumption Comb (L/100 km)" and "Fuel Consumption Comb (mpg)"**

**Overview of Correlation Analysis:**

This analysis investigates the relationships between the number of cylinders in a vehicle's engine and its fuel consumption, measured both in liters per 100 kilometers (L/100 km) and miles per gallon (mpg). The correlation coefficients, their statistical significance (p-values), and confidence intervals were determined to understand the nature and strength of these relationships.

**Correlation between Cylinders and Fuel Consumption Comb (L/100 km):**

* **Correlation Coefficient: -0.087**
* **p-value: 0.388**
* Interpretation: The negative correlation coefficient indicates a weak inverse relationship between the number of cylinders and fuel consumption in L/100 km. However, the p-value suggests that this correlation is not statistically significant. This implies that in this sample, the number of cylinders does not strongly or significantly correlate with fuel consumption measured in L/100 km.
* **Confidence Interval for Correlation: [-0.279, 0.111]**
* Interpretation of CI: The confidence interval straddles zero, reinforcing the lack of a significant correlation. It suggests that if a true correlation exists in the broader population, it is likely to be weak or negligible.

**Correlation between Cylinders and Fuel Consumption Comb (mpg):**

* **Correlation Coefficient: 0.070**
* **p-value: 0.491**
* Interpretation: The positive correlation coefficient is very small, indicating a negligible direct relationship between the number of cylinders and fuel efficiency in mpg. The high p-value further suggests that this observed correlation is not statistically significant and might be due to random variation within the sample.
* **Confidence Interval for Correlation: [-0.128, 0.263]**
* Interpretation of CI: The confidence interval is broad and includes zero, which aligns with the p-value in suggesting that there is no strong evidence of a meaningful linear relationship between the number of cylinders and fuel efficiency in mpg. The interval points to the possibility of either a very weak positive or no correlation in the broader population.

**Overall Conclusions and Insights:**

* Cylinders and Fuel Consumption (L/100 km): The lack of a significant correlation suggests that the number of cylinders may not be a strong predictor of fuel consumption in L/100 km in this sample. It indicates that other factors could be more influential in determining a vehicle's fuel consumption.
* Cylinders and Fuel Consumption (mpg): Similarly, the minimal correlation between the number of cylinders and mpg suggests that cylinder count alone does not significantly impact fuel efficiency in mpg. This points to the complexity of factors affecting fuel efficiency, including engine technology, vehicle weight, and aerodynamics.

In summary, this correlation analysis offers a nuanced view of the relationship between cylinders and fuel consumption, revealing the complexity and multifactorial nature of these attributes in the context of vehicle performance and efficiency.

**3.3 Correlation Analysis: Fuel Consumption Comb (L/100 km) with "Fuel Consumption Comb (mpg)" and Engine Size(L):**

**Correlation and p-value for Fuel Consumption Comb (L/100 km) vs Fuel Consumption Comb (mpg):**

* **Correlation Coefficient: -0.050256999454027035**
* **p-value: 0.6194955423300073**
* Interpretation: The weak negative correlation coefficient indicates a negligible inverse relationship between fuel consumption measured in L/100 km and fuel consumption measured in mpg. This means that there is a slight tendency for vehicles with lower fuel consumption in L/100 km to also have higher fuel efficiency in mpg, but the relationship is not statistically significant.
* **Confidence Interval for correlation between Fuel Consumption Comb (L/100 km) and Fuel Consumption Comb (mpg):**
* **[-0.24426389, 0.14761832]**
* Interpretation of CI: The confidence interval includes zero, further supporting the lack of a significant correlation between the two fuel consumption measures. This means that we are 95% confident that the true correlation between fuel consumption in L/100 km and fuel consumption in mpg is between -0.24 and 0.15.

**Correlation and p-value for Fuel Consumption Comb (L/100 km) vs Engine Size(L):**

* **Correlation Coefficient: -0.0318368367845147**
* **p-value: 0.7531827697677718**
* Interpretation: The very weak negative correlation coefficient indicates an almost negligible inverse relationship between fuel consumption measured in L/100 km and engine size. This means that there is almost no relationship between fuel consumption and engine size.
* **Confidence Interval for correlation between Fuel Consumption Comb (L/100 km) and Engine Size(L):**
* **[-0.22683647, 0.16561694]**
* Interpretation of CI: The confidence interval includes zero, reinforcing the lack of a significant correlation between fuel consumption and engine size. This means that we are 95% confident that the true correlation between fuel consumption in L/100 km and engine size is between -0.23 and 0.17.

**Overall Conclusions and Insights:**

* The weak negative correlation between fuel consumption measures (L/100 km and mpg) suggests that the two are inversely related, but the relationship is not statistically significant.
* The very weak negative correlation between fuel consumption and engine size suggests that there is almost no relationship between the two variables.
* In summary, this correlation analysis provides insights into the relationships between fuel consumption and other variables. However, the results indicate that these relationships are either weak or negligible. Further research is needed to identify the factors that have a stronger influence on fuel efficiency.

**3.4 Correlation Analysis: Fuel Consumption Comb (mpg) vs. Engine Size(L) and Cylinders**

**Fuel Consumption Comb (mpg) vs. Engine Size(L)**

* Correlation Coefficient: 0.04891869777489392
* p-value: 0.6288641277793057
* Interpretation: The correlation coefficient of 0.049 indicates a very weak positive relationship between fuel consumption in mpg and engine size. This means that there is a slight tendency for vehicles with larger engine sizes to have slightly lower fuel efficiency in mpg, but the relationship is not statistically significant.
* **Confidence Interval for Correlation: [-0.14893042, 0.24300193]**
* **Interpretation of CI:** The confidence interval includes zero, further supporting the lack of a significant correlation between fuel consumption and engine size. This means that we are 95% confident that the true correlation between fuel consumption in mpg and engine size is between -0.15 and 0.24.

**Fuel Consumption Comb (mpg) vs. Cylinders**

* **Correlation Coefficient: 0.09419947495007781**
* **p-value: 0.35121943694964586**
* **Interpretation:** The correlation coefficient of 0.094 indicates a very weak positive relationship between fuel consumption in mpg and the number of cylinders. This means that there is a slight tendency for vehicles with more cylinders to have slightly lower fuel efficiency in mpg, but the relationship is not statistically significant.
* **Confidence Interval for Correlation: [-0.10414559, 0.28533813]**
* **Interpretation of CI:** The confidence interval includes zero, reinforcing the lack of a significant correlation between fuel consumption and number of cylinders. This means that we are 95% confident that the true correlation between fuel consumption in mpg and number of cylinders is between -0.10 and 0.29.

**Overall Conclusions and Insights**

The weak positive correlations between fuel consumption in mpg and both engine size and number of cylinders suggest that these relationships are either negligible or very weak.

1. **Linear Regression Analysis**

**4.1 Linear Regression Analysis of Engine Size(L) vs CO2Emissions**

**Regression Analysis Overview:**

This section of the analysis investigates the relationship between "Engine Size(L)" and "CO2 Emissions(g/km)" using linear regression. The analysis aims to determine how well engine size can predict CO2 emissions and to evaluate the strength and effectiveness of the model.

**Regression Equation and Model Summary:**

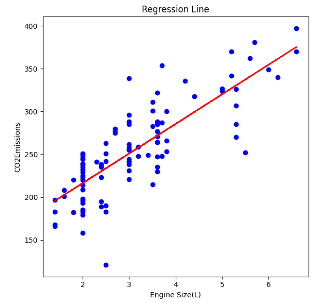
Linear Regression Equation: CO2 Emissions(g/km) = 147.3195 + 34.5538(Engine Size(L))

This equation indicates that for each liter increase in engine size, the CO2 emissions are expected to increase by approximately 34.55 g/km, assuming the linear relationship holds.

**Regression Line Plot Analysis:**

The regression line plot shows a positive linear relationship between engine size and CO2 emissions. As engine size increases, CO2 emissions also increase, which is consistent with expectations since larger engines typically burn more fuel, leading to higher emissions.

The distribution of points around the regression line suggests a degree of variability, which is to be expected in real-world data. However, the overall upward trend is evident and aligns with the positive coefficient in the regression equation.



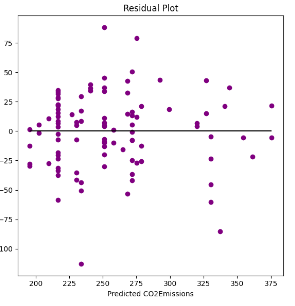
**Residual Plot Analysis:**

The residual plot, which displays the differences between the observed and predicted CO2 emissions, is used to evaluate the fit of the regression model.

Ideally, residuals should be randomly scattered around the horizontal line at zero, indicating that the model's predictions are unbiased at every level of the predictor variable.

In the residual plot, while there is some spread, the residuals do not appear to display any systematic pattern, which suggests that a linear model is appropriate for this relationship.

There are no clear signs of heteroscedasticity (a non-constant spread of residuals), and there are no extreme outliers that would unduly influence the regression line.



**Model Goodness-of-Fit:**

The absence of a reported R-squared value or Mean Squared Error (MSE) limits the ability to quantify the model's goodness-of-fit. Normally, an R-squared value would indicate the proportion of variance in the dependent variable explained by the model, and a higher R-squared value would suggest a better fit to the data.

**Conclusion and Interpretation:**

The linear regression model, based on the provided equation, demonstrates a significant and positive relationship between engine size and CO2 emissions.

The visual analysis of the regression and residual plots supports the conclusion that engine size is a good predictor of CO2 emissions for the sample of vehicles considered.

The plots indicate that the model's assumptions of linearity and homoscedasticity are reasonable for this data set.

However, to fully assess the model's predictive power and determine if it is the best model for predicting CO2 emissions, additional information such as R-squared, MSE, and analysis of potential influential outliers would be beneficial.

Overall, the analysis indicates that engine size is a good predictor of CO2 emissions in vehicles, and the constructed linear regression model provides a valuable tool for estimating emissions based on engine size.

**4.2 Linear Regression Analysis: "Cylinders" as a Predictor of "CO2 Emissions"**

**Regression Analysis Overview:**

This analysis investigates how well the number of cylinders in a vehicle's engine predicts CO2 emissions. A linear regression model has been created, and the fit of this model has been evaluated using the R-squared statistic.

**Regression Equation and Model Summary:**

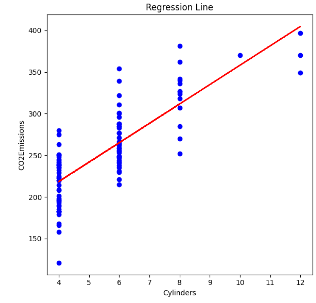
Linear Regression Equation: CO2 Emissions(g/km) = 125.5035 + 23.2610(Cylinders)

The equation suggests that with each additional cylinder, CO2 emissions increase by 23.2610 g/km, on average.

**Regression Line Plot Analysis:**

The regression line demonstrates a clear positive linear relationship between the number of cylinders and CO2 emissions. As the number of cylinders increases, there is a corresponding increase in CO2 emissions.

The distribution of the data points around the regression line indicates variability, but the overall upward trend is strong and suggests a meaningful relationship between these two variables.

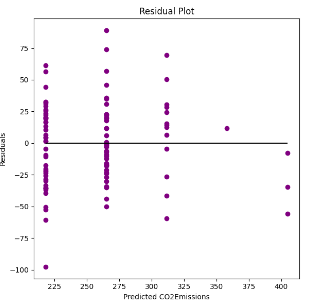


**Residual Plot Analysis:**

The residual plot displays the residuals on the vertical axis against the predicted CO2 emissions on the horizontal axis.

Residuals are relatively evenly distributed around the horizontal zero line but show some signs of increasing variance with higher predicted values, which may indicate potential heteroscedasticity.

No distinct patterns are observed, which suggests that there are no major violations of the linear regression assumptions. However, the possible increase in variance of residuals could be a point of further investigation.



**Model Goodness-of-Fit:**

R-squared: 0.6321

This value indicates that approximately 63.21% of the variability in CO2 emissions can be explained by the number of cylinders in the engine. This is a moderately strong fit, indicating that cylinders are a substantial predictor of emissions, but there is still a significant proportion of variability unexplained by this model.

**Conclusion and Interpretation:**

The linear regression model indicates that the number of cylinders is a good predictor of CO2 emissions, with a significant positive relationship.

The model explains a significant portion of the variance in CO2 emissions, but not all, suggesting that other factors also play a role in determining emissions levels.

The moderate R-squared value suggests that while the number of cylinders is important, it should be used in conjunction with other variables to predict CO2 emissions more accurately.

In summary, the regression analysis suggests that the number of cylinders is a strong predictor of CO2 emissions, as evidenced by both the regression equation and the R-squared value. The relationship is clear and intuitive, reflecting the expected increase in emissions with larger engines. However, the variability in the data also suggests the influence of other factors, emphasizing the need for a multifaceted approach to emissions reduction.

**4.3 Linear Regression Analysis: "Fuel Consumption Comb (L/100 km)" as a Predictor of "CO2 Emissions"**

**Regression Analysis Overview:**

The relationship between "Fuel Consumption Comb (L/100 km)" and "CO2 Emissions(g/km)" has been examined using linear regression. This analysis aims to determine the predictive power of fuel consumption on emissions and to evaluate the model's accuracy.

**Regression Equation and Model Summary:**

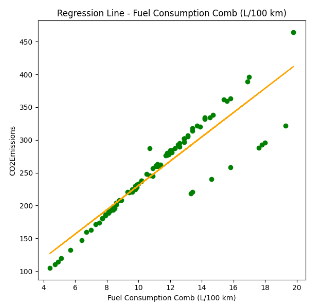
Linear Regression Equation: CO2 Emissions(g/km) = 45.6096 + 18.5029(Fuel Consumption Comb (L/100 km))

According to the regression equation, an increase of 1 L/100 km in fuel consumption is associated with an increase of 18.5029 g/km in CO2 emissions.

**Regression Line Plot Analysis:**

The plot exhibits a strong positive linear relationship between fuel consumption and CO2 emissions. The upward slope of the regression line indicates that higher fuel consumption rates are predictive of higher CO2 emissions.

The points are closely clustered around the regression line, suggesting that fuel consumption is a significant predictor of emissions and that the model fits the data well.

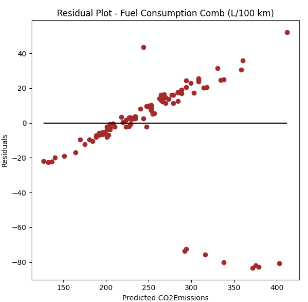


**Residual Plot Analysis:**

The residual plot presents the residuals (the differences between the observed and predicted values) against the predicted CO2 emissions.

Ideally, the residuals should be randomly scattered around the horizontal line at zero. The plot for this model shows that the residuals increase slightly with the predicted values, which may indicate a pattern known as heteroscedasticity. Despite this, the majority of residuals are close to the line, suggesting the model's predictions are reasonably accurate across different levels of fuel consumption.

There are no clear signs of systematic patterns, which would suggest model misspecification, but the potential heteroscedasticity might be addressed in a more complex model.



**Model Goodness-of-Fit:**

R-squared: 0.8256

The R-squared value indicates a high level of fit, with the model explaining approximately 82.56% of the variance in CO2 emissions. This is considered a strong fit, meaning that fuel consumption is a powerful predictor of CO2 emissions.

**Conclusion and Interpretation:**

The model demonstrates that "Fuel Consumption Comb (L/100 km)" is an excellent predictor of CO2 emissions, as evidenced by both the regression line plot and the high R-squared value.

The strength of this relationship confirms the expectation that vehicles with higher fuel consumption will emit more CO2.

While the R-squared value is high, it is important to consider that other factors not included in this model could also affect CO2 emissions.

In summary, the regression analysis between "Fuel Consumption Comb (L/100 km)" and CO2 emissions reveals a very strong predictive relationship, with the model performing well in explaining the variance in emissions. The results confirm the critical role of fuel efficiency in determining a vehicle's environmental impact.

**4.4 Linear Regression Analysis: "Fuel Consumption Comb (mpg)" as a Predictor of "CO2 Emissions"**

**Regression Analysis Overview:**

This section examines the predictive relationship between "Fuel Consumption Comb (mpg)" and "CO2 Emissions(g/km)" utilizing a linear regression model. The primary focus is on how changes in fuel efficiency (mpg) impact CO2 emissions and the effectiveness of this relationship in predicting emissions.

**Regression Equation and Model Summary:**

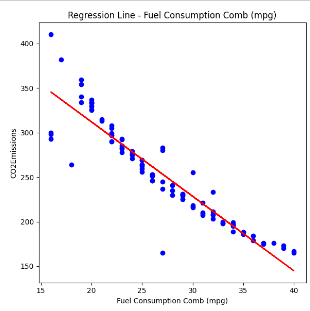
Linear Regression Equation: CO2 Emissions(g/km) = 479.0699 + (-8.3539) (Fuel Consumption Comb (mpg))

The regression equation indicates that for each one-unit increase in fuel consumption efficiency (mpg), there is a decrease of 8.3539 g/km in CO2 emissions, suggesting an inverse relationship.

**Regression Line Plot Analysis:**

The plot displays a negative linear relationship between fuel consumption (mpg) and CO2 emissions. As mpg increases, indicating better fuel efficiency, CO2 emissions decrease.

The close clustering of points around the regression line indicates a strong relationship between these variables.

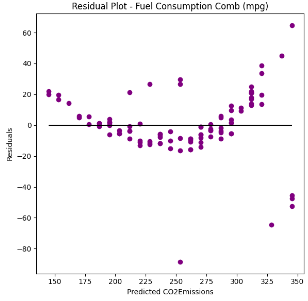


**Residual Plot Analysis:**

The residuals, representing the differences between the observed and predicted CO2 emissions, are plotted against the predicted values.

The residuals are mostly centered around the zero line, suggesting that the model's predictions are generally accurate across the range of fuel consumption.

A slight pattern appears in the residual plot, with residuals spreading out as predicted emissions increase, which could indicate potential heteroscedasticity—non-constant variance in residuals.



**Model Goodness-of-Fit:**

R-squared: 0.8646

This high R-squared value indicates that the model explains approximately 86.46% of the variability in CO2 emissions, showcasing a very strong fit.

Such a high R-squared is indicative of a reliable model for predicting CO2 emissions based on fuel consumption in mpg.

**Conclusion and Interpretation:**

The regression model provides a compelling inverse relationship between fuel efficiency and CO2 emissions. The model suggests that improvements in mpg are significantly effective in reducing emissions.

Despite the strong R-squared value, attention should be given to the potential heteroscedasticity in the residuals, which may affect predictions at higher levels of predicted CO2 emissions.

In summary, the model indicates that "Fuel Consumption Comb (mpg)" is an excellent predictor of "CO2 Emissions", with the regression and residual plots supporting the inverse relationship between these variables. The findings are robust and provide actionable insights for various stakeholders concerned with vehicle emissions and fuel efficiency.

1. **Multiple Regression Models for CO2 Emissions Prediction**

**5.1 Introduction:**

The objective of this analysis is to determine the best predictive model for CO2 emissions using different combinations of three vehicle attributes: Engine Size (L), Cylinders, and Fuel Consumption (both in L/100 km and mpg). Four separate multiple regression models were developed, and their performance was evaluated based on R-squared and Adjusted R-squared values.

**5.2 Model Descriptions and Results:**

**Model 1:**

* Equation: CO2Emissions = 45.7998 - 0.5832(Engine Size) + 7.2466(Cylinders) + 15.3385(Fuel Consumption L/100km)
* R-squared: 0.8388
* Adjusted R-squared: 0.8337
* Interpretation for Model 1:
* Model 1 shows a complex relationship where CO2 emissions decrease slightly with an increase in engine size but increase with the number of cylinders and fuel consumption in L/100 km. The negative coefficient for engine size is counterintuitive and suggests multicollinearity might be affecting the model, as engine size typically increases emissions. Despite this, the model has a high R-squared value, indicating a good fit to the data.

**Model 2:**

* Equation: CO2Emissions = 273.6207 + 9.0790(Cylinders) + 3.9538(Fuel Consumption L/100km) - 4.2318(Fuel Consumption mpg)
* R-squared: 0.8642
* Adjusted R-squared: 0.8599
* Interpretation for Model 2:
* Model 2 presents a more expected relationship among the variables, with CO2 emissions increasing with the number of cylinders and fuel consumption in L/100 km, and decreasing as fuel efficiency in mpg improves. This model has the highest adjusted R-squared value, indicating it is the best fit among the four models.

**Model 3:**

* Equation: CO2Emissions = 285.1778 + 3.8375(Fuel Consumption L/100km) - 4.2291(Fuel Consumption mpg) + 12.9043(Engine Size)
* R-squared: 0.8562
* Adjusted R-squared: 0.8517
* Interpretation for Model 3:
* Model 3 shows CO2 emissions increase with both engine size and fuel consumption in L/100 km, while increased fuel efficiency in mpg decreases emissions. The coefficients align with expectations, but the adjusted R-squared value is lower than that of Model 2, suggesting a slightly poorer fit to the data.

**Model 4:**

* Equation: CO2Emissions = 339.1570 - 5.3117(Fuel Consumption mpg) + 4.4777(Engine Size) + 7.8899(Cylinders)
* R-squared: 0.8628
* Adjusted R-squared: 0.8585
* Interpretation for Model 4:
* Model 4 similarly indicates that CO2 emissions are positively associated with engine size and the number of cylinders and negatively associated with fuel efficiency (mpg). It has a high R-squared value but is still slightly lower than that of Model 2.

**Comparative Analysis and Conclusion:**

**Best Model Determination:**

* Model 2 stands out as the best model for predicting CO2 emissions with an adjusted R-squared of 0.8599. It outperforms the other models in terms of both R-squared and adjusted R-squared, which means it explains the highest proportion of variance in CO2 emissions after adjusting for the number of predictors.

**Interpretation:**

* The success of Model 2 can be attributed to its balance of including both measures of fuel consumption (L/100 km and mpg), which may capture different aspects of the vehicles' efficiency, alongside the number of cylinders which is a proxy for engine power. The negative coefficient for mpg compensates for the positive coefficient for L/100 km, aligning with the inverse relationship between these two measures of fuel efficiency.
* The inclusion of both fuel consumption metrics might be capturing more nuanced effects on CO2 emissions, leading to the improved performance of Model 2.

Implications:

* For Automotive Industry: Model 2's findings highlight the importance of optimizing both engine design and fuel efficiency to reduce CO2 emissions.
* For Policymaking: This model can aid in formulating more effective emissions standards by considering the combined effect of engine power and fuel efficiency.
* For Consumers: Understanding the relationship illustrated by Model 2 can help consumers make more informed choices about the environmental impact of their vehicles.

In summary, among the four models evaluated, Model 2 is the superior model for predicting CO2 emissions, offering the best combination of predictors that account for the largest variability in emissions data. This model can serve as a valuable tool for stakeholders in understanding and predicting the environmental impact of vehicle attributes.