9 (2): 285-294 (2025)

Journal of Aviation

https://dergipark.org.tr/en/pub/jav

e-ISSN 2587-1676

Assessment of Aircraft Fuel Efficiency in Domestic Flights using Multiple Regression Analysis

Ertuğrul Metehan Sertdemir[1], Hülya Kaftelen Odabaşı[2*], Ali Altınok[3]

[1]Firat University, Aviation Science and Technologies, 23000, Elazig, Türkiye. (ertugrulsertdemir2@gmail.com)
[2*] Firat University, Aircraft Maintenance and Repair, Elazig, Türkiye. (hkodabasi@firat.edu.tr)
[3] Istanbul Gelisim University, Aviation Management, Istanbul, Türkiye. (altinok777@gmail.com)

Article Info

Received: 22 March 2025
Revised: 22 May 2025
Accepted: 11 June 2025
Published Online: 22 June 2025

Keywords:
Fuel Efficiency
Multiple Linear Regression
Aviation Sustainability
Operational Optimization
Environmental Impact

Corresponding Author: *Hülya Kaftelen Odabaşı*

RESEARCH ARTICLE

https://doi.org/10.30518/jav.1662031

Abstract

The aviation industry has significantly evolved over the past century, playing a crucial role in global transportation, trade, and tourism. However, its reliance on fossil fuels has raised environmental concerns, necessitating sustainable practices to mitigate carbon emissions. This study examines the relationship between fuel consumption and various operational parameters for the Airbus A321 aircraft, utilizing multiple linear regression analysis to develop a predictive model for fuel efficiency.

The dataset, comprising 110 flight records from Istanbul Airport, includes independent variables such as the number of passengers, flight level, flight distance, average wind speed, airspeed, flight duration, aircraft takeoff weight, and total fuel load. Statistical tests, including normality checks, correlation analysis, and multicollinearity assessments, were conducted to ensure the validity of the model. Findings indicate that flight duration, aircraft takeoff weight, and total fuel load significantly influence fuel consumption, while variables such as flight level and wind speed have negligible effects.

The study highlights the importance of optimizing flight planning, weight management, and fuel policies to enhance operational efficiency and reduce environmental impact. The results provide valuable insights for the aviation industry, supporting data-driven decision-making in fuel efficiency and sustainability efforts. By integrating advanced statistical modeling and strategic operational planning, airlines can achieve cost optimization while promoting environmentally responsible practices.

This research contributes to aviation sustainability by offering a data-driven approach to fuel efficiency analysis, which can inform future innovations in aircraft design, air traffic management, and alternative fuel utilization.

1. Introduction

Since the early 20th century, the aviation industry has evolved beyond being merely a reflection of engineering advancements, becoming a pivotal sector that has profoundly reshaped global transportation, trade, and tourism (Sen et al., 2021). The widespread adoption of commercial air travel has significantly reduced travel times, fostering economic integration and cultural exchange across geographically distant regions (Esin et al., 2021). As aviation became more accessible, it not only enhanced connectivity but also played a crucial role in global economic growth by facilitating international business operations and expanding tourism markets. However, this rapid expansion has also resulted in substantial environmental challenges, primarily due to the industry's reliance on fossil fuels, which contributes to increasing carbon emissions. The adverse effects of climate change and global warming have intensified the urgency for the aviation industry to adopt sustainable practices, making environmental responsibility a key priority for its long-term viability (Bahadir et al., 2018).

In response to these challenges, the aviation sector has undertaken extensive initiatives to mitigate its ecological footprint through technological innovations and regulatory frameworks. International organizations such as the International Air Transport Association (IATA) have set ambitious sustainability targets, including the commitment to achieving net-zero carbon emissions by 2050 (IATA,2050). To support this objective, advancements in aircraft design have focused on optimizing aerodynamics, incorporating lightweight composite materials, and exploring alternative propulsion systems such as hybrid-electric and hydrogen-powered engines (Volkan, 2013). These technological improvements aim to enhance fuel efficiency, reduce operational costs, and minimize the overall environmental impact of air travel. For instance, the development of next-generation narrow-body aircraft like the A321 has demonstrated significant progress in fuel economy while maintaining high performance and passenger capacity, making them a model for future innovations in sustainable aviation (STM, 2021).

Scientific research and data-driven analysis play a fundamental role in this transition, enabling the precise evaluation of factors influencing fuel consumption and emissions. Through advanced statistical modeling techniques, such as regression analysis and machine learning-based simulations, researchers can identify optimization opportunities that contribute to the development of more efficient aircraft designs (Fenkli et al., 2023). Furthermore, sustainability efforts extend beyond aircraft engineering; operational strategies such as optimized flight planning, enhanced air traffic management, and the integration of sustainable aviation fuels (SAFs) are gaining traction as complementary approaches to reducing emissions.

The successful implementation of these measures requires coordinated efforts among industry stakeholders, including aircraft manufacturers, airlines, regulatory bodies, and research institutions. Government policies that incentivize sustainable technology investments, coupled with increased public awareness of eco-friendly travel choices, further support the industry's shift toward sustainability (Altinkeski et al., 2022). Additionally, collaborative projects between academia and the private sector continue to drive innovation in materials science, propulsion technology, and alternative energy sources, paving the way for a greener future in aviation.

Ultimately, achieving long-term sustainability in the aviation sector necessitates a holistic approach that integrates environmental, economic, and technological considerations. By leveraging cutting-edge research, policy-driven initiatives, and industry-wide collaboration, the aviation industry can continue to expand while mitigating its ecological impact, ensuring that future generations can benefit from the advancements of air travel without compromising environmental integrity.

2. Methodologies for measuring fuel efficiency

The dataset employed in this study comprises one dependent variable and eight independent variables. The dependent variable is defined as "fuel consumption" for the Airbus A321 aircraft model, which is a critical parameter in flight operations concerning energy efficiency and cost optimization (STM, 2021). The independent variables include "number of passengers," "flight level," "flight distance," "average wind speed," "average airspeed," "flight duration (minutes)," "aircraft takeoff weight," and "total fuel." These variables have been meticulously selected to model the relationship between flight performance and fuel consumption.

The data utilized for analysis originates from flight records obtained at Istanbul Airport. As one of Turkey's busiest aviation hubs, Istanbul Airport provides a comprehensive and reliable dataset (Kacar et al., 2025). The primary objective of this study is to model the impact of independent variables on fuel consumption and to develop a predictive model accordingly. To achieve this, flight data spanning three months were used for model training, while an additional month's data was employed to assess the predictive capability of the model. The decision to use a three-month dataset for training is based on its ability to capture sufficient data variability and provide reliable results.

The selection of dependent and independent variables was informed by prior studies in aviation operations and industry standards. "Number of passengers" directly influences the total payload, making it a significant determinant of fuel consumption (Ozturk, 2023). Similarly, "flight level" and "flight distance" are key operational parameters affecting flight dynamics and fuel efficiency (Kaltenecker et al ., 2022). "Average wind speed" and "average airspeed" reflect the environmental conditions experienced during flight, playing a crucial role in fuel consumption analysis. Additionally, "flight duration" and "aircraft takeoff weight" are fundamental factors in determining aircraft performance, while "total fuel" denotes the amount of fuel loaded before departure (FAA, 2023).

The dataset was rigorously analyzed throughout the model development and validation processes. A total of 110 flight records were incorporated into the modeling phase to account for potential seasonal variations and operational discrepancies. To evaluate the accuracy of the trained model and its applicability to future flights, one month of flight data was utilized for validation. This methodological approach enables an assessment of both the generalization capacity of the model and its alignment with real-world operational conditions.

In conclusion, this study provides an in-depth examination of the relationships between dependent and independent variables and develops a predictive model for estimating fuel consumption. The findings, based on flight data from Istanbul Airport, contribute valuable insights toward improving energy efficiency and operational performance in the aviation sector. The results offer practical implications for industry applications and serve as a reference for future academic research in this domain.

3. Data and Method

3.1. Data

In order to correctly analyze the relationship between the parameters affecting fuel efficiency, 110 flight data between Istanbul Airport and Elazığ Airport were taken into account. This data set, covering a flight period of approximately three months, was collected in a way that is ready for analysis in order to create a consistent and meaningful model.

3.2. Method

In this study, multiple linear regression analysis was employed to identify the impact of various parameters on fuel efficiency in flight operations. The analysis was conducted using 110 flight records obtained from Istanbul Airport. In the developed model, fuel consumption during the flight was designated as the dependent variable, while the following independent variables were considered: number of passengers, average wind speed, average airspeed, flight distance, flight level, flight duration (minutes), total aircraft takeoff weight and total fuel quantity.

The initial phase of the analysis involved assessing the normality distribution of variables to determine the suitability of the dataset for model development. Subsequently, preliminary tests for multiple linear regression analysis, including multicollinearity assessment and outlier detection, were performed. As part of the multicollinearity test, correlation coefficients among the independent variables were calculated to evaluate the presence of strong linear relationships.

Based on the findings from these preliminary tests and statistical evaluations, a multiple linear regression model was constructed to analyze the parameters influencing fuel efficiency in flight operations. This study contributes to the statistical modeling of fuel consumption in aviation operations, offering insights into energy efficiency and cost optimization through a data-driven approach.

3.2.1 Multiple linear regression analysis

Multiple linear regression analysis is a widely used statistical method for examining the relationship between a dependent variable and multiple independent variables (Bulut, 2024). It is particularly valuable in complex systems where multiple factors contribute to an outcome, providing a structured framework to assess both individual and combined effects.

In practical applications, this method is extensively utilized across various industries. For instance, an airline's financial performance depends on multiple factors such as fuel costs, load factor (LF), cost per available seat kilometer (CASK), break-even load factor (BELF), and market conditions (Kose, 2021). Evaluating these variables collectively through multiple linear regression allows for a more comprehensive understanding of their impact compared to analyzing them in isolation.

A fundamental assumption of multiple linear regression is linearity, which implies that changes in the independent variables lead to proportional variations in the dependent variable (Osborne et al., 2022). If this assumption is not met, the model may fail to produce reliable results. Therefore, before constructing the model, it is essential to assess whether the relationships among variables adhere to linearity. If nonlinearity is detected, appropriate transformations or alternative modeling techniques should be considered.

In conclusion, multiple linear regression is a crucial tool for analyzing multivariate relationships, offering insights for both academic research and industry applications. Ensuring that the assumptions are met and that variables are appropriately selected enhances the model's predictive accuracy, leading to more reliable and meaningful results.

Multiple linear regression analysis is a statistical method used to model the relationship between a dependent variable and multiple independent variables. It quantifies how changes in independent variables influence the dependent variable, making it useful for analyzing complex systems with multiple interacting factors. The general equation is (Karaca et al., 2016):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \qquad (1)$$

where y is the dependent variable, x_1, x_2, ..., x_k are independent variables, β_0 is the intercept, β_1, β_2, ..., β_k are regression coefficients, and ε represents the error term, assumed to follow a normal distribution.

The model relies on key assumptions: a linear relationship between variables, error terms with a mean of zero (E(ε) = 0) and constant variance (homoscedasticity), and no strong intercorrelation among independent variables (no multicollinearity) (Karaca et al., 2016). Violations of these assumptions can lead to biased estimates, requiring adjustments such as data transformations.

3.2.2 Linearity assumption

The linearity assumption posits that a linear relationship exists between the dependent and independent variables (Abebe, 2024). This assumption suggests that variations in the dependent variable can be consistently explained by proportional changes in the independent variables. As a fundamental determinant of model reliability, the linear relationship must accurately represent the underlying factors influencing the dependent variable (Abebe, 2024). When a linear association between the variables is absent, the predictive capability of the model diminishes, and its ability to accurately explain variations in the dependent variable becomes compromised. Consequently, the regression coefficients may lose their statistical validity, leading to misleading conclusions.

To assess the validity of this assumption, a correlation analysis is typically conducted between the dependent and independent variables (Anandhi, 2023). This analysis evaluates whether the dependent variable exhibits a linear trend, thereby determining whether the model should be expanded to incorporate nonlinear relationships. If no linear relationship is observed, logarithmic, polynomial, or other nonlinear transformations can be applied to better capture the association between variables (Anandhi, 2023). Such transformations enhance the model's predictive accuracy, ensuring more reliable and robust results.

3.2.3 Independence Assumption

The independence assumption states that independent variables should not exhibit intercorrelations (Chung, 2023). This assumption ensures that the effect of each independent variable on the dependent variable is evaluated separately, thereby preserving the accuracy and reliability of the model. If strong correlations exist among independent variables, a phenomenon known as multicollinearity arises. Multicollinearity reduces the reliability of regression coefficients, leading to inaccuracies in prediction and diminishing the explanatory power of the model (Chung, 2023). This issue complicates the differentiation of individual effects among independent variables, ultimately weakening the model's predictive capability.

Several statistical techniques are employed to detect multicollinearity, one of the most widely used being the Variance Inflation Factor (VIF). The VIF value measures the degree to which an independent variable is correlated with other independent variables in the model (Reid, 2020). If the VIF exceeds 10, it indicates a severe multicollinearity issue that can compromise the validity of the model's results. To address this problem, some independent variables may be removed, transformations may be applied, or alternative statistical methods may be considered (Reid, 2020). Implementing these adjustments helps maintain the independence assumption, thereby enhancing the model's statistical validity and predictive accuracy.

3.2.3 Normality Assumption of Error Terms

The normality assumption of error terms is a fundamental requirement for ensuring the reliability of regression models. This assumption states that error terms should follow a normal distribution, allowing predicted values to be symmetrically and consistently distributed (Schisterman et al., 2006). A violation of this assumption can reduce the predictive performance of the model and undermine the statistical validity of regression coefficients. Since this assumption is crucial for hypothesis testing, it is commonly evaluated using Shapiro-Wilk and Kolmogorov-Smirnov tests (Schisterman et al., 2006). If error terms deviate from normality, appropriate data transformations or non-parametric alternatives can be applied to enhance the model's accuracy.

3.2.4 Homoscedasticity Assumption of Error Terms

The homoscedasticity assumption requires that error terms exhibit constant variance across all levels of independent variables (Sahinler, 2020). When this assumption holds, the distribution of errors remains stable, ensuring consistent prediction accuracy. However, if error variance varies across different values of the independent variables, heteroscedasticity arises, which can distort regression estimates and reduce the model's reliability (Kilic, 2013).

Breusch-Pagan and White tests are commonly used to detect heteroscedasticity. If detected, corrective measures such as Weighted Least Squares (WLS) estimation or appropriate data transformations can be employed to stabilize error variance and improve model performance.

3.2.5 Advantages and Limitations of Multiple Linear Regression Analysis

Multiple Linear Regression (MLR) analysis is a robust statistical method that models the relationship between a dependent variable and multiple independent variables. This technique is widely used to quantify relationships between variables, make predictions, and provide a scientific basis for decision-making processes.

One of the primary advantages of MLR is its high predictive power. Incorporating multiple independent variables into the model enhances its ability to explain variations in the dependent variable, leading to more accurate predictions compared to univariate models (Karaca et al., 2016). This feature makes MLR particularly valuable in disciplines such as economics, engineering, and social sciences, where multiple factors influence the dependent variable (Kardes et al., 2024).

Furthermore, MLR enables the assessment of individual effects of independent variables. Through regression coefficients, the model quantifies the impact of each independent variable on the dependent variable, allowing researchers to determine which factors exert a more significant influence. This capability is crucial for strategic decision-making, particularly in fields where understanding variable interactions is essential (Li, 2014).

Additionally, MLR facilitates the evaluation of inter-variable relationships and statistical significance within the model. For instance, the Variance Inflation Factor (VIF) can be employed to detect multicollinearity among independent variables, ensuring that redundant predictors do not distort model accuracy (Kardes et al., 2024). This process enhances the reliability of the model by eliminating potential distortions caused by correlated predictors.

Multiple linear regression (MLR) was selected due to its suitability for interpreting linear relationships among multiple predictors and its transparent mathematical structure (Tranmer et al., 2020). Compared to more complex models such as support vector regression (SVR), decision trees, or neural networks, MLR provides interpretable coefficients, making it ideal for identifying the relative importance of operational variables like aircraft weight or flight duration (Tranmer et al., 2020). Additionally, MLR requires fewer computational resources and is less sensitive to overfitting when assumptions are met (Kang & Hansen, 2018). While advanced models can capture nonlinear dynamics, the current dataset exhibited linear tendencies, justifying the use of MLR for predictive modeling and hypothesis testing.

Finally, MLR is widely applicable across diverse research fields and industries. Its effectiveness has been demonstrated in various applications, from forecasting electricity consumption (Karaca et al., 2016) to predicting tourism demand. The adaptability of the model across different datasets enables its integration into decision support mechanisms in multiple domains.

The reliability of the Multiple Linear Regression (MLR) model depends on the fulfillment of fundamental assumptions. Violations of conditions such as the linear relationship between independent and dependent variables, the normal distribution of error terms, homoscedasticity, and the absence of high correlations among independent variables can negatively affect the validity of the model (Olden et al., 2000). In particular, when nonlinear relationships exist, the MLR model may fail to provide sufficient explanatory power and may not accurately capture the relationships between variables (Dinev et al., 2004).

High correlations among independent variables can lead to multicollinearity, a major issue that reduces the reliability of regression coefficients and weakens the model's predictive capacity (Daoud, 2017). In the presence of multicollinearity, the individual effects of independent variables cannot be accurately distinguished, resulting in a decline in the model's predictive performance. Although statistical measures such as the Variance Inflation Factor (VIF) are commonly used to detect multicollinearity, in some cases, it may not be possible to eliminate this issue entirely (Maxwell, 1975).

MLR analysis is also limited in modeling nonlinear relationships. If the relationship between the dependent and independent variables is not linear, the model may fail to adequately represent these associations, leading to reduced predictive accuracy. In such cases, alternative approaches such as nonlinear regression models or machine learning-based methods may be required to achieve more reliable results (Tezcan, 2011).

In conclusion, while MLR analysis is a powerful statistical tool, it has several limitations that must be carefully considered. To ensure reliable results, assumptions should be rigorously tested, relationships between independent variables should be thoroughly examined, and alternative methods should be employed when nonlinear patterns are detected. Failure to address these factors may negatively impact the predictive power and statistical validity of the model.

4. Results

In this study, multiple linear regression analysis (MLR) was employed to estimate fuel consumption during flight operations. To assess the robustness of the model, various statistical tests were conducted, including normality distribution, linearity, multicollinearity, correlation analysis, outlier detection, skewness, and kurtosis measures. These preliminary tests ensure the validity and reliability of the regression model.

The general mathematical expression of the multiple linear regression model is formulated as follows (Sezgin, 2013):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \tag{2}$$

where:

y represents the dependent variable,

$x_1, x_2, \ldots, x_p$ denote the independent variables,

β_0 is the intercept term,

$\beta_1, \beta_2, \ldots, \beta_p$ are the regression coefficients, indicating the effect of each independent variable on the dependent variable,

ε represents the error term.

The regression model, adapted to the dataset used in this study, is expressed as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \varepsilon \tag{3}$$

In this model, the dependent variable fuel consumption (y) is analyzed in relation to the following independent variables:

x_1: Number of passengers

x_2: Flight level

x_3: Flight distance

x_4: Average wind speed

x_5: Average airspeed

x_6: Flight duration (minutes)

x_7: Aircraft takeoff weight

x_8: Total fuel load

The H_1 hypothesis tested in this study posits that fuel consumption during flight is influenced by operational efficiency factors. These factors include the number of passengers, flight level, flight distance, average wind speed, average airspeed, flight duration, aircraft takeoff weight, and total fuel load. The significance level of the model was set at 0.05, and statistical analyses were performed within this confidence interval to evaluate the model's validity. The skewness and kurtosis values of the model are given in Table 1 and Table 2:

Table 1.

	Fuel Consumption	Number of Passengers	Flight Distance	Average Wind Speed
Skewness	1.14	0.27	0.94	0.94
Kurtosis	0.69	-0.34	1.15	1.37

Table 2. Skewness - Kurtosis values

	Average Speed	Flight Duration (Minutes)	Aircraft Takeoff Weight	Total Fuel Load
Skewness	0.22	0.65	0.59	0.94
Kurtosis	-0.81	0.33	-0.74	0.27

When the skewness and kurtosis values of the variables shown in Table 1 and Table 2 were examined, it was seen that all values were between -2 and +2 and were suitable for normal distribution.

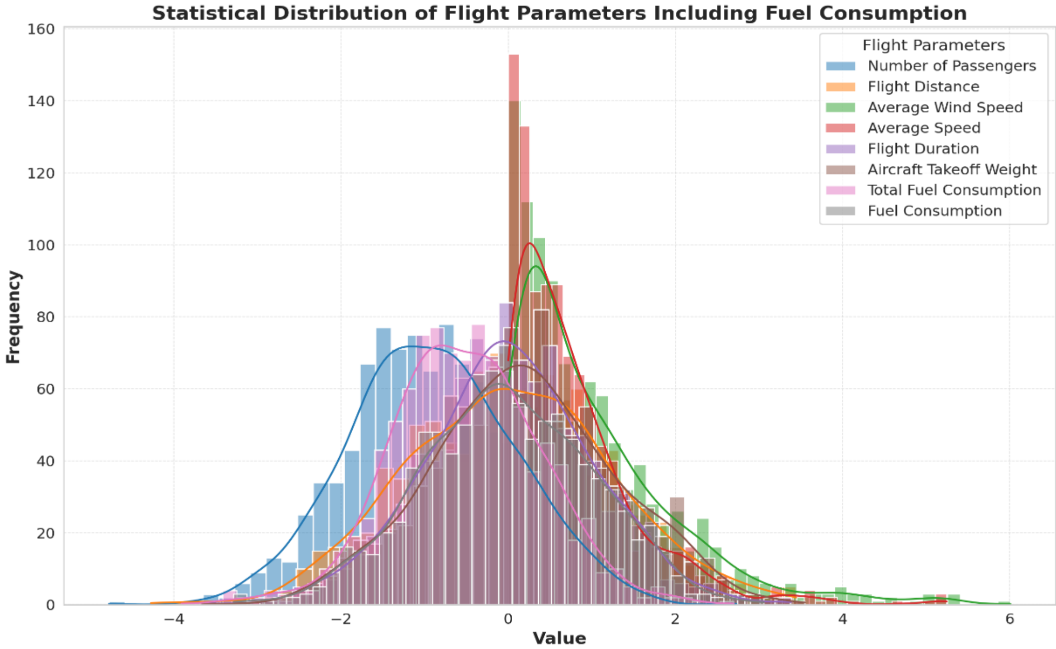

Figure 1. Statistical Distribution of Flight Parameters

Figure 1 illustrates the statistical distribution of various flight parameters, including fuel consumption, using both histograms and Kernel Density Estimation (KDE) curves. The visualization effectively compares the frequency distribution of multiple aviation-related parameters on a single plot, facilitating a detailed statistical assessment of flight performance and operational efficiency. The x-axis represents the values of different flight parameters, while the y-axis denotes their frequency, indicating how often specific ranges of values occur in the dataset. Each flight parameter is color-coded distinctly, with both its histogram and corresponding KDE curve displayed to enhance interpretability.

The y-axis represents frequency, indicating how often specific value ranges appear within each parameter's dataset. Higher bars correspond to more frequently observed values, while lower bars signify less common occurrences. The x-axis denotes the numerical range of different flight parameters, reflecting variations in flight performance. These values differ based on the parameter measured, such as distance in kilometers, speed in knots, or fuel consumption in kilograms, providing insight into the distribution and operational characteristics of each variable.

Number of Passengers (Blue): The number of passengers onboard an aircraft significantly affects operational efficiency and economic viability. As represented in the graph, variations in passenger load influence aircraft weight, fuel consumption, and overall performance. A higher passenger count increases the total weight, leading to greater fuel requirements, particularly during takeoff and climb phases. However, at optimal load factors, fuel efficiency per passenger improves, making load management a critical factor in airline profitability and environmental sustainability.

Flight Distance (Orange): Flight distance determines the total range covered from departure to arrival, directly impacting fuel consumption and operational planning. The graph indicates a distribution encompassing short-haul, medium-haul, and long-haul flights, each with distinct fuel efficiency characteristics. Short-haul flights experience higher per-kilometer fuel consumption due to the increased influence of takeoff and climb phases, while long-haul flights benefit from prolonged cruise efficiency. Route optimization plays a crucial role in minimizing unnecessary fuel burn.

Average Wind Speed (Green): Wind conditions encountered during flight play a crucial role in determining fuel efficiency and flight time. The distribution in the graph highlights variations in wind speeds, where stronger headwinds result in increased fuel consumption and extended flight durations, while favorable tailwinds contribute to reduced fuel burn and shorter travel times. Effective wind management through flight planning and real-time route adjustments can mitigate adverse effects and enhance operational efficiency.

Average Speed (Red): The average speed of an aircraft influences both fuel consumption and overall flight performance. The graph depicts a normal distribution, indicating a standard operational range. Higher speeds increase aerodynamic drag, necessitating greater thrust and fuel expenditure, whereas lower speeds may extend flight duration and reduce efficiency. Optimal cruise speed selection is essential to balance fuel efficiency with operational timelines, ensuring cost-effective and environmentally responsible flight operations.

Flight Duration (Purple): Total flight duration, from takeoff to landing, is a key parameter affecting airline scheduling, fuel efficiency, and passenger experience. The graphical representation (Fig.1) suggests variations in flight lengths, influenced by factors such as air traffic congestion, routing constraints, and weather conditions. Longer flights require careful fuel management and optimized altitude profiles to enhance efficiency, while shorter flights face proportionally higher fuel burn due to frequent altitude changes.

Aircraft Takeoff Weight (Brown): The total weight of an aircraft at departure, including passengers, cargo, and fuel, directly affects performance and efficiency. The graph illustrates a distribution that reflects variations in takeoff conditions across different flights. Higher takeoff weight demands increased thrust, leading to higher fuel consumption and potential performance limitations, particularly in high-altitude or short-runway airports. Effective weight management and fuel load optimization contribute to improved operational efficiency and safety.

Total Fuel Consumption (Teal): Fuel consumption is a fundamental parameter in aviation economics and environmental impact assessment. The graph's distribution reveals fluctuations in fuel burn across different flight operations, emphasizing the importance of fuel-efficient practices. Factors such as aircraft type, route length, wind conditions, and weight contribute to variations in total fuel usage. Airlines prioritize fuel efficiency strategies, including optimized routing, weight reduction, and aerodynamic improvements, to minimize operational costs and carbon emissions.

Fuel Consumption (Yellow-Green): The specific fuel consumption pattern in the graph provides insights into fuel burn trends under varying operational conditions. This parameter highlights the influence of different flight phases, including takeoff, cruise, and descent, on overall fuel efficiency. By analyzing these trends, aviation professionals can implement data-driven strategies to enhance performance, reduce environmental impact, and optimize fuel expenditure without compromising safety and reliability.

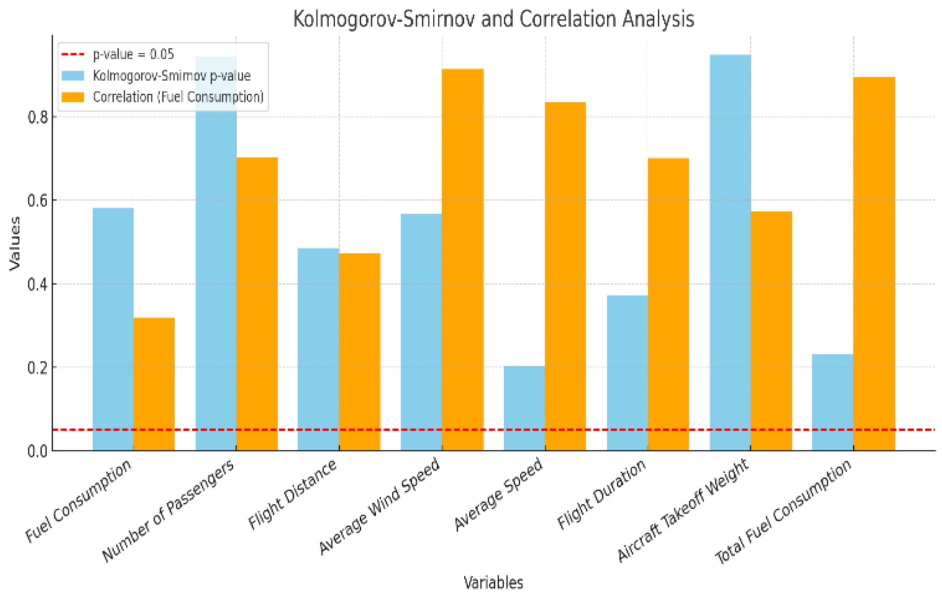

Figure 2. Kolmogorov-Smirnov and Correlation Analysis

Figure 2 provides a detailed statistical evaluation of key flight parameters using the Kolmogorov-Smirnov (K-S) test and correlation analysis with fuel consumption. The x-axis represents different flight parameters, including fuel consumption, number of passengers, flight distance, wind speed, average speed, flight duration, aircraft takeoff weight, and total fuel consumption, while the y-axis quantifies their respective Kolmogorov-Smirnov p-values and correlation coefficients. The Kolmogorov-Smirnov test assesses the normality of each parameter's distribution, depicted in blue bars, with higher values indicating stronger adherence to a normal distribution and lower values suggesting significant deviations. The correlation coefficients, illustrated in orange bars, represent the statistical relationship between each flight parameter and fuel consumption, providing insights into how these variables interact with operational efficiency. A red dashed line marks the statistical significance threshold of $p = 0.05$, below which a parameter's distribution significantly deviates from normality. The fuel consumption parameter serves as a benchmark for evaluating fuel efficiency, where a strong correlation with other parameters suggests key operational dependencies.

The results of the correlation analysis indicate that the number of passengers, flight distance, and takeoff weight exhibit a strong relationship with fuel consumption, highlighting the critical role of these variables in operational planning. The findings of the Kolmogorov-Smirnov (K-S) test further reveal that many variables do not follow a normal distribution, suggesting that operational factors do not fully conform to standard statistical models. Consequently, the prediction of fuel consumption requires the use of more flexible, data-driven approaches instead of strictly linear models. In conclusion, enhancing fuel efficiency in aviation operations necessitates careful management of weight optimization, flight duration, and speed control. Based on the results of the Kolmogorov-Smirnov test and correlation analysis, the following conclusions can be drawn:

Variables with a p-value ≥ 0.05: The null hypothesis cannot be rejected for these variables, implying that their distributions conform to normality. For instance, the p-value for "Fuel Consumption" is 0.533, indicating that it follows a normal distribution.

Variables with a p-value < 0.05: The null hypothesis is rejected for these variables, signifying that their distributions deviate from normality. For example, the p-values for "Flight Distance" and "Average Wind Speed" are both <0.001, suggesting that they do not follow a normal distribution.

The correlation analysis leads to the following interpretations:

Variables exhibiting a strong correlation with fuel consumption:

"Number of Passengers" (0.77)

"Aircraft Takeoff Weight" (0.85)

Since these correlation coefficients exceed 0.5 in absolute value, they suggest a strong association between fuel consumption and these parameters. Variable exhibiting a weak correlation with fuel consumption:

"Average Wind Speed" (0.26)

This low correlation value suggests that "Average Wind Speed" has no significant impact on fuel consumption.

For variables that do not conform to normality ("Flight Distance" and "Average Wind Speed"), data transformation techniques such as logarithmic transformation are recommended to approximate normal distribution.

Variables with strong correlations with fuel consumption should be considered in multiple linear regression models; however, potential multicollinearity issues among highly correlated independent variables must also be examined.

Including weakly correlated variables in the model may not enhance explanatory power, and therefore, careful assessment is required to determine their relevance. To achieve the most accurate predictions of fuel consumption, a comprehensive evaluation of the interactions among independent variables, their distribution properties, and potential multicollinearity issues is essential.

4.1 Multiple Linear Regression Analysis Test Results

The regression analysis results presented in Table 3 are utilized to assess the explanatory power and accuracy of the model in predicting the dependent variable. The multiple R value (0.92) indicates a strong positive correlation between the dependent variable and the independent variables, demonstrating a substantial linear relationship. The R^2 value (0.86) suggests that the model accounts for 86% of the total variance in the dependent variable, highlighting its strong predictive capability. The adjusted R^2 value (0.84), which considers the number of predictors in the model, implies a slight reduction in explanatory power due to model complexity; however, it still reflects a highly reliable and robust model. These findings indicate that the model is capable of accurately predicting operational variables such as fuel consumption, making it a valuable tool for decision-making in aviation management and fuel optimization strategies. Nevertheless, the standard error value (12.17) suggests that while the model demonstrates high accuracy, further refinements and additional predictor variables may enhance its predictive performance, reducing uncertainty and improving overall reliability.

Table 3. Regression Statistics

Quantity	Values
Multiple R	0.92
R Square	0.86
Adjusted R Square	0.84
Standard Error	12.17
Observation	110.00

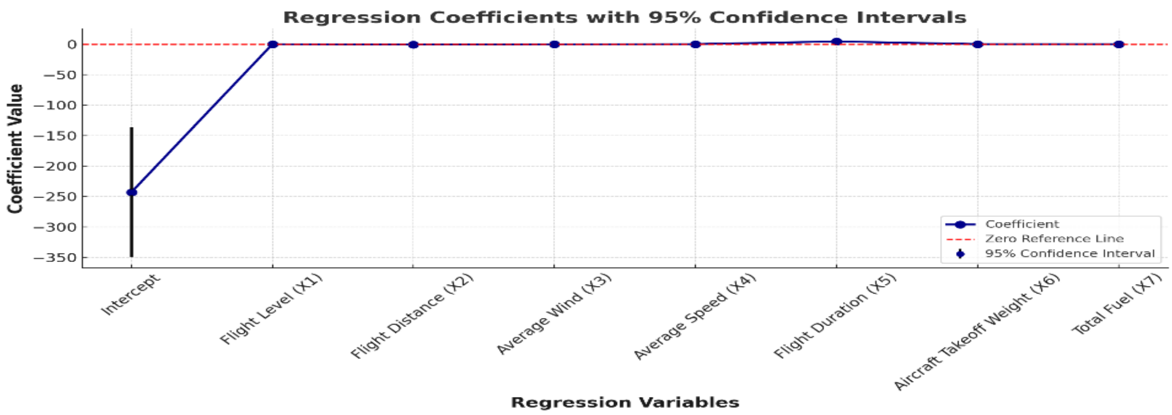

Figure 3. Regression Coefficients With 95% Confidence Intervals

Figure 3 visualizes the regression coefficients along with their 95% confidence intervals, providing insights into the influence of each independent variable on the dependent variable. The y-axis represents the coefficient values, while the x-axis lists the independent variables used in the regression model.

In Fig. 3. the intercept (-243.11) has the most substantial magnitude, with a wide confidence interval, indicating its significant role in baseline calculations. Among the independent variables, flight duration (X5, 4.37) and aircraft takeoff weight (X6, 0.01) exhibit positive coefficients, suggesting that an increase in these parameters leads to a rise in the dependent variable. Conversely, flight distance (X2, -0.69), average wind speed (X3, -0.18), and total fuel (X7, -0.01) have negative coefficients, indicating an inverse relationship with the dependent variable.

The red dashed reference line at zero is crucial for statistical interpretation. Variables with confidence intervals that do not cross this line are statistically significant, meaning their influence on the dependent variable is strong and reliable. In this case, flight duration (X5), aircraft takeoff weight (X6), and total fuel (X7) are statistically significant predictors, while flight level (X1) and average speed (X4) show negligible effects with high p-values, implying weak or no impact on the model.

Overall, the model demonstrates strong predictive capacity, with key variables like flight distance, flight duration, and takeoff weight significantly influencing the dependent variable. However, refinements, such as excluding insignificant predictors or adjusting for multicollinearity, could further enhance the model's explanatory power.

4.2 Model Performance Metrics

To further evaluate the accuracy and predictive capability of the multiple linear regression model, three commonly used error metrics were computed: Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). These metrics help quantify the magnitude of prediction errors and validate the model's generalization performance (Chai & Draxler, 2014). The values are presented in Table 4:

Table 4. Model Performance Metrics

Metric	Values
MSE (Mean Squared Error)	148.25
MAE (Mean Absolute Error)	9.65
RMSE (Root Mean Square Error)	12.17

The MSE represents the average of the squared differences between predicted and actual values, serving as a general indicator of overall prediction error. A lower MSE indicates a more accurate model.

The MAE reflects the mean of the absolute differences between predicted and observed values, offering a direct measure of average prediction error magnitude. It is particularly useful for operational decision-making due to its intuitive interpretation.

The RMSE, which is the square root of MSE, illustrates the standard deviation of prediction errors in the same units as the dependent variable—fuel consumption (kg). The value of 12.17 closely aligns with the model's reported standard error, confirming the consistency and reliability of the regression output.

Together, these metrics affirm the robustness of the model in estimating fuel consumption. The relatively low error values support the validity of the regression approach in practical aviation scenarios, particularly for route optimization and load management strategies.

5. Discussion

The findings of this study underscore the multifaceted nature of aircraft fuel efficiency and its dependency on operational and environmental parameters. Among the variables analyzed through multiple linear regression, flight duration, aircraft takeoff weight, and total fuel load were identified as statistically significant contributors to fuel consumption. These findings are consistent with prior studies that have demonstrated the critical role of weight management and flight planning in fuel efficiency (Seymour et al., 2020).

Notably, the inverse relationship between total fuel load and actual fuel consumption highlights an operational paradox: overfilling fuel tanks increases aircraft weight, thereby raising consumption levels during takeoff and climb. This observation aligns with the concept of "fuel penalty" found in airline operations literature, where carrying excess fuel to ensure safety margins leads to additional burn (Tabernier et al., 2021). The model's finding that average wind speed and flight level are not statistically significant suggests that environmental conditions may have a more complex or nonlinear relationship with fuel consumption. This contrasts with findings from studies on transatlantic and intercontinental routes, where wind optimization has shown significant fuel savings, especially in operations from major hubs such as JFK (USA), Heathrow (UK), and Frankfurt (Germany) (Hamdan et al., 2022).

Furthermore, the correlation between the number of passengers and fuel consumption emphasizes the importance of load factor optimization (He & Zhou, 2016). Higher passenger counts can improve per-passenger fuel efficiency if managed appropriately, as supported by empirical findings from studies on Lufthansa (Germany), Emirates (UAE), and Delta Airlines (USA) (Pinchemel et al., 2022).

Compared to previous works focused primarily on aircraft design and propulsion improvements, this study adds value by providing empirical evidence from operational data collected at Istanbul Airport. By integrating statistical modeling with practical performance indicators, the study offers actionable insights for both airline managers and policy makers concerned with reducing carbon emissions in domestic aviation.

Nevertheless, the study is not without limitations. The data is limited to one aircraft type (Airbus A321) and one route (Istanbul to Elazığ), which may affect the generalizability of the results. Future research could incorporate diverse aircraft models and a wider range of domestic and international routes to develop a more comprehensive model of fuel efficiency.

6. Conclusion

This regression analysis provides critical strategic insights into improving operational efficiency and reducing costs in the aviation industry. The impact of key variables such as fuel consumption, flight duration, takeoff weight, and flight distance on operational processes has been clearly identified. Notably, the significant positive effect of flight duration and aircraft takeoff weight on the dependent variable underscores the necessity of optimizing flight planning and load management. The negative coefficient of flight distance suggests that long-haul flights may offer operational advantages, while the negative coefficient of total fuel consumption highlights the adverse impact of excessive fuel loading on aircraft performance, emphasizing the need for more efficient fuel management strategies.

Furthermore, the lack of statistical significance for flight level, average wind speed, and average speed indicates that these factors may allow for greater flexibility in operational decision-making. These findings can support airlines in developing more data-driven strategies for air traffic management, fuel policies, and route optimization.

In conclusion, this study emphasizes the importance of science-based decision-making in enhancing fuel efficiency and promoting environmental sustainability in the aviation sector. The findings from this regression analysis serve as a foundation for minimizing the environmental impact of airline operations and optimizing costs. When supplemented with advanced analyses, these results will facilitate the development of comprehensive sustainability policies, guiding the aviation industry toward a more efficient and environmentally responsible future.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgement

This study was produced from the graduate student Ertugrul Metehan Sertdemir's MSc. thesis titled 'Multiple Linear Regression Analysis of Aircraft Fuel Efficiency in Domestic Flights'.

References

Abebe, T. H. (2024). Regression analysis: A theoretical approach. Journal of Statistical and Econometric Methods, 1(40).

Altınkeski, B. K., Özyiğit, O., & Çevik, E. (2022). The relationship among eco-friendly technologies, civil aviation and environmental quality: Panel threshold regression analysis. Gaziantep University Journal of Social Sciences, 21(3), 1162–1179. http://dergipark.org.tr/tr/pub/jss

Anandhi, P., & Nathiya, E. (2023). Application of linear regression with their advantages, disadvantages, assumptions, and limitations. International Journal of Statistics and Applied Mathematics, 8(6), 133–137.

Bahadır, E., Kalender, B., Yumuşak, F. N., & Karaman, G. (2018). A comparative study on modeling analyses: Structural equation modeling and regression. Kesit Akademi Dergisi(16), 410–420.

Bulut, Y. (2024). Measuring the determinants of CO2 emissions in Turkey: Regression analysis with an instrumental variable [Master's thesis, Sakarya University]. Sakarya University Social Sciences Institute.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7, 1247–1250.

Chung, W. (2012). Using the fuzzy linear regression method to benchmark the energy efficiency of commercial buildings. Applied Energy, 95, 45–49.

Daoud, J. I. (2017). Multicollinearity and regression analysis. Journal of Physics: Conference Series, 949(1), 012009.

Dinev, T., & Hart, P. (2004). Internet privacy concerns and their antecedents - Measurement validity and a regression model. Behaviour & Information Technology, 23(6), 413–422.

Gerede, E. (Ed.). (2015). Airline transportation and economic regulations: Theory and practice in Turkey. General Directorate of Civil Aviation Publications. ISBN: 978-975-493-067-2

Hamdan, S., Jouini, O., Cheaitou, A., Jemai, Z., Granberg, T. A., & Josefsson, B. (2022). Air traffic flow management under emission policies: Analyzing the impact of sustainable aviation fuel and different carbon prices. Transportation Research Part A: Policy and Practice, 166, 14–40.

He, M., & Zhou, Y. (2016, September). Analysis of power factor of capacitive pulse load in aviation variable frequency power supply system. In Proceedings of the 2016 4th International Conference on Advanced Materials and Information Technology Processing (AMITP 2016) (pp. 40–44). Atlantis Press.

IATA (International Air Transport Association). (2023). IATA Net Zero Roadmaps: Providing a strategic vision to achieve net zero CO2 emissions by 2050. 79th AGM and World Air Transport Summit, Istanbul, Türkiye, 19–21 June 2022.

Kaçar, E., & Kalkan, K. (2025). Investigation of the relationship between aircraft mobility and air quality at airports: Aerosol index and European example. Journal of Aviation Research, 7(1), 56–68.

Kaltenecker, S., & Leopold, K. (2022). Flight Levels: A short introduction. Loop Organisationsberatung GmbH.

Kang, L., & Hansen, M. (2018). Improving airline fuel efficiency via fuel burn prediction and uncertainty estimation. Transportation Research Part C: Emerging Technologies, 97, 128–146.

Karaca, C., & Karacan, H. (2016). Examining factors affecting electricity consumption demand using multiple regression method. Sujest, 4(3), 182–195.

Kilic, S. (2013). Linear regression analysis. Journal of Mood Disorders, 3(2), 90.

Köse, Y. (2021). Specific financial ratios in the aviation sector: Analysis and evaluation of airline companies in Turkey. Financial Research and Studies Journal, 13(25), 623–636.

Maxwell, A. E. (1975). Limitations on the use of the multiple linear regression model. British Journal of Mathematical and Statistical Psychology, 28(1), 51–62.

Olden, J. D., & Jackson, D. A. (2000). Torturing data for the sake of generality: How valid are our regression models? Écoscience, 7(4), 501–510.

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. Practical Assessment, Research & Evaluation, 8.

Öztürk, O. (2023). Comparative analysis of operational, financial, and environmental efficiency of major airlines worldwide [Doctoral dissertation, Hasan Kalyoncu University].

Pehlivanoğlu, Y. V. (2013). Introduction to aviation and space sciences. Air Force Academy Press.

Pinchemel, Alexandre & Caetano, Mauro & Rossi, Ricardo & Silva, Marco. (2022). Airline's business performance indicators and their impact on operational efficiency. Brazilian Business Review. 19. 642-665.

Reid, C. (2020). Technical appendix: The limitations of regression models.

Schisterman, E. F., Vexler, A., Whitcomb, B. W., & Liu, A. (2006). The limitations due to exposure detection limits for regression models. American Journal of Epidemiology, 163(4), 374–383.

Seymour, K., Held, M., Georges, G., & Boulouchos, K. (2020). Fuel estimation in air transportation: Modeling global fuel consumption for commercial aviation. Transportation Research Part D: Transport and Environment, 86, 102528.

Sezgin, E., & Çelik, Y. (2013). Comparison of data used for loss of data mining methods. In XV. Academic Informatics Conference Proceedings, January 23–25, 2013, Akdeniz University, Antalya. Retrieved from https://www.researchgate.net/publication/348787393

STM Defense Technologies Engineering and Trade Inc. (2021). Fuel efficiency in civil aviation: Research report, March 2021.

Şahinler, S. (2000). The basic principles of fitting linear regression model by least squares method. MKÜ Journal of Agriculture Faculty, 5(1–2), 57–73. Retrieved from https://www.researchgate.net/profile/Suat-Sahinler/publication/305711004_The_Basic_Principles_of_Fitting_Linear_Regression_Model_By_Least_Squares_Method/links/58ac442e92851cf0e3ccd32e/The-Basic-Principles-of-Fitting-Linear-Regression-Model-By-Least-Squares-Method.pdf

Şen, Y., & Erdağ, T. (2021). Evaluation of airline transportation sector development stages with PEST analysis: An investigation in the scope of five periods + Covid-19 pandemic process period. TroyAcademy International Journal of Social Sciences, 6(2), 422–461.

Tabernier, L., Calvo Fernández, E., Tautz, A., Deransy, R., & Martin, P. (2021). Fuel tankering: Economic benefits and environmental impact for flights up to 1500 NM (full tankering) and 2500 NM (partial tankering). Aerospace, 8(2), 37.

Tranmer, M., Murphy, J., Elliot, M., & Pampaka, M. (2020). Multiple linear regression (2nd ed.). Cathie Marsh Institute Working Paper 2020-01. https://hummedia.manchester.ac.uk/institutes/cmist/archive-publications/working-papers/2020/2020-1multiple-linear-regression.pdf

Cite this article: Sertdemir, E.M, Kaftelen Odabaşı, H. Altınok, A. (2025). Assessment of Aircraft Fuel Efficiency in Domestic Flights using Multiple regression analysis. Journal of Aviation, 9(2), 285-294