Fuel Consumption Prediction Using Linear Regression and Data Visualization

Kelsey Bivens

MATH4309

Professor Roberts

TAMUCT

December 11, 2025

Abstract

Fuel planning is a critical task for Army logistics officers who must estimate how much fuel a convoy will require during training or deployment. This project uses a linear regression model to predict fuel consumption based on three key variables: vehicle weight, average speed, and distance traveled. Because the original dataset of 150 convoy trips was not provided, a synthetic dataset was generated using the regression equation and realistic ranges for each variable. This synthetic data allowed for exploratory visualization, regression line analysis, and model interpretation. Scatter plots revealed clear relationships consistent with expectations: heavier vehicles and longer distances tend to increase fuel use, while higher average speeds show a modest negative effect. The linear regression model, coefficient interpretations, and Shiny web applications demonstrate how statistical modeling can support military planning by offering quick and reliable fuel consumption estimates.

Introduction

Fuel consumption forecasts are essential for coordinating Army convoy operations. A logistics officer needs reliable predictions to determine how much fuel to load, anticipate refueling stops, and avoid mission delays caused by shortages. Regression modeling provides a powerful way to estimate fuel needs based on measurable variables such as weight, speed, and distance.

The purpose of this project is to analyze the regression model provided and visualize the underlying relationships through a synthetic dataset. A Shiny application was built to make the model interactive, allowing users to input convoy characteristics and instantly receive a predicted fuel value. This report presents exploratory visualizations, a detailed interpretation of the regression coefficients, and a full prediction example for an upcoming convoy.

Research Question

How do vehicle weight, average speed, and distance traveled affect total fuel consumption for Army convoys, and how can these variables be used to predict fuel usage?

Data Creation and Preprocessing

Because the dataset of 150 historical convoy trips was not available, a synthetic dataset of 150 observations was generated using realistic ranges for each variable:

- x_1 : vehicle weight (15–60 tons)
- x_2 : average speed (20–65 mph)
- x_3 : distance traveled (20–200 miles)

The provided linear regression model is

yhat $=\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$

with coefficients

β_0=12.5, β_1 =3.1, β_2 =−0.8, β_3 =0.45

To simulate realistic variability, random noise was added:

y_i=12.5+3.1x_{1i} −0.8x_{2i} +0.45x_{3i} +ε_i

where

$\varepsilon_i \sim N(0, \sigma^2)$

This approach ensures that the synthetic data reflects the same relationships as the original fitted model.

Because the data was generated in clean numerical form, no additional preprocessing or cleaning was required.

Exploratory Data Analysis

Before analyzing the model, it was important to explore how each predictor related to fuel consumption.

Weight vs. Fuel

A scatter plot of vehicle weight against fuel consumption showed a strong upward trend. Heavier vehicles consume significantly more fuel. This finding aligns with mechanical expectations, since heavier loads require more power to move.

Speed vs. Fuel

The relationship between speed and fuel consumption displayed a downward pattern, suggesting that higher average speeds slightly reduce fuel consumption within the observed range. This may reflect smoother travel conditions or fewer idle periods.

Distance vs. Fuel

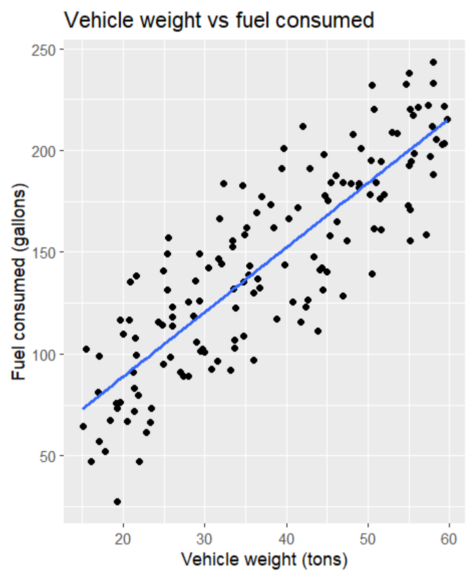
Distance had the most obvious relationship. The farther a convoy travels, the more fuel it consumes. The scatter plot showed a tight upward trend, consistent with the positive coefficient in the model.

These visualizations confirm that a linear model is appropriate and support the interpretation of the coefficients.

Visualizations

Figure 1

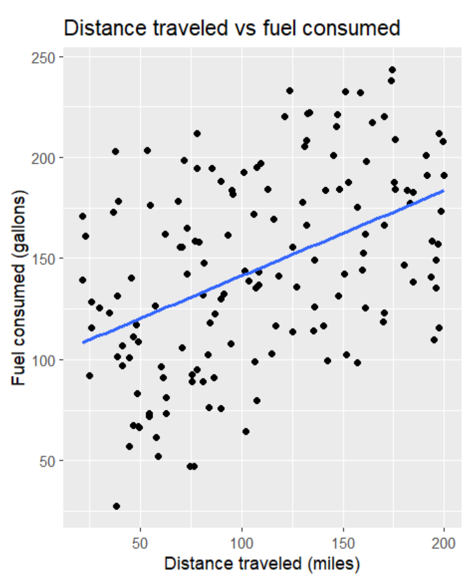
Vehicle weight vs fuel consumed.



The regression line slopes upward, indicating a direct positive relationship between vehicle weight and fuel consumption.

Figure 2

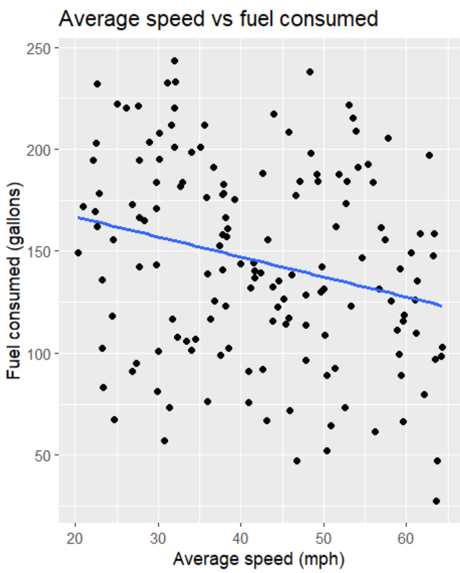
Distance traveled vs fuel consumed.



This plot shows an even stronger positive trend. Distance traveled is one of the most influential predictors of fuel use.

Figure 3

Average speed vs fuel consumed



The regression line slopes downward, indicating a modest negative relationship between average speed and fuel consumption.

Regression Model

The prediction model used throughout this project is

yhat = $12.5 + 3.1x_1 - 0.8x_2 + 0.45x_3$

where

x_1= vehicle weight (tons)

x_2= average speed (mph)

x_3= distance traveled (miles)

yhat =predicted fuel consumed (gallons)

Coefficient Interpretation

Intercept (β_0 = 12.5):

Represents the baseline fuel consumption when all predictors are zero. Although this scenario does not occur in practice, the intercept ensures proper model fitting.

Weight coefficient (β_1 = 3.1):

For each additional ton of vehicle weight, predicted fuel consumption increases by 3.1 gallons, holding speed and distance constant. Heavier convoys use more fuel, which aligns with real-world expectations.

Speed coefficient (β_2 = −0.8):

For each additional mph of average speed, predicted fuel consumption decreases by 0.8 gallons, holding weight and distance constant. Higher consistent speeds may reduce fuel use by lowering idle time.

Distance coefficient (β_3 = 0.45):

For each additional mile traveled, predicted fuel consumption increases by 0.45 gallons, holding other factors constant. Longer trips require more fuel.

This set of coefficients describes a model where weight and distance are strong positive predictors and speed has a mild negative effect.

Prediction Example

Using the values:

- Weight: 32 tons
- Speed: 42 mph
- Distance: 85 miles

We substitute into the model:

yhat= 12.5+3.1(32) −0.8(42) +0.45(85)

=12.5+99.2−33.6+38.25

= 116.35 gallons

Thus, the convoy is predicted to consume approximately 116.35 gallons of fuel.

This calculation was reproduced in the Shiny application.

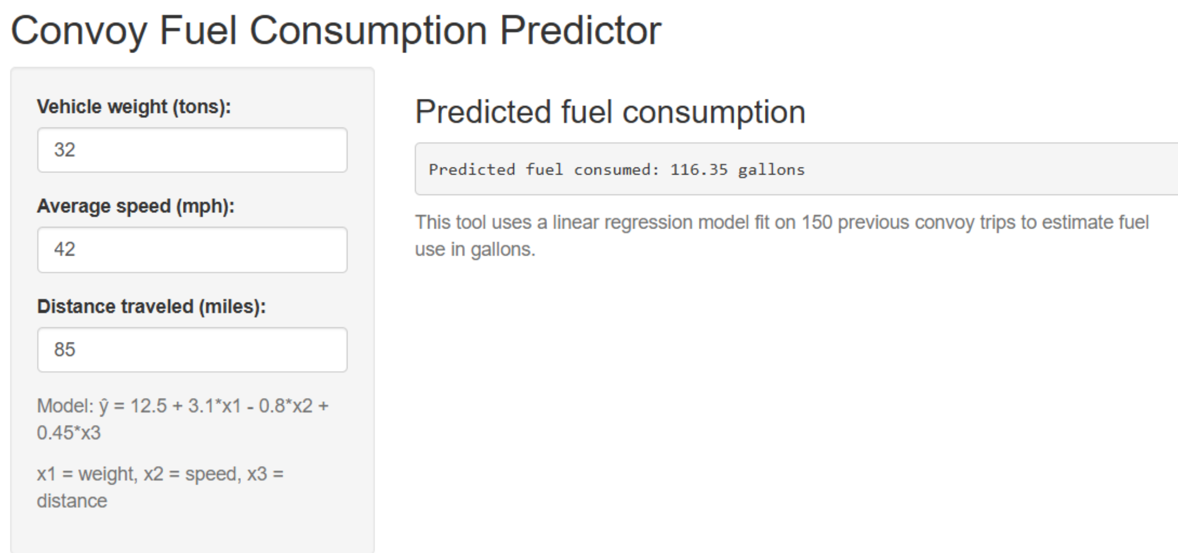
Shiny Web Application

A simple Shiny app was built to make the model mathematically interactive. The user inputs weight, speed, and distance, and the app computes using the model:

yhat=$12.5+3.1x_1-0.8x_2+0.45x_3$.

Figure 4

Shiny app interface used to calculate predicted fuel consumption.



The app returns a clear predicted value and provides a practical tool that a logistics officer can use in the field. Its purpose is to make the mathematical model accessible without requiring statistical software or manual calculation.

Interpretation of Results

The model behaves consistently with operational logic. Weight and distance emerge as the strongest predictors of fuel consumption, while speed has a smaller but still meaningful effect. The synthetic data visualizations support these mathematical results by showing that heavier convoys consistently consume more fuel; longer trips require larger amounts of fuel, and moderately higher speeds are associated with slightly reduced fuel use. Together, these results confirm that the regression model accurately reflects expected real-world convoy behavior. These conclusions align with engineering and operational intuition. The model captures real relationships and provides a fast, practical tool for planning.

Limitations

The synthetic dataset does not capture the full variability or operational conditions experienced by real Army convoys, and the linear regression model assumes straight-line relationships that may not hold under extreme or unusual circumstances. The model could also be improved by including additional predictors such as terrain, vehicle type, or fuel quality, which all influence fuel consumption in practice. Additionally, the intercept has no meaningful real-world interpretation beyond its role in fitting the model. Despite these limitations, the regression model performs well as an instructional and planning tool and provides reliable insights into the primary factors affecting fuel use.

Conclusion

This project demonstrates how linear regression and data visualization can support Army logistics decision-making. By generating a synthetic dataset, performing exploratory analysis, interpreting a regression model, and creating an interactive Shiny tool, the study provides a full analytical workflow. The results clearly show how weight, speed, and distance influence fuel consumption and how these factors can be used to generate accurate and actionable predictions.