Ethanol Analysis

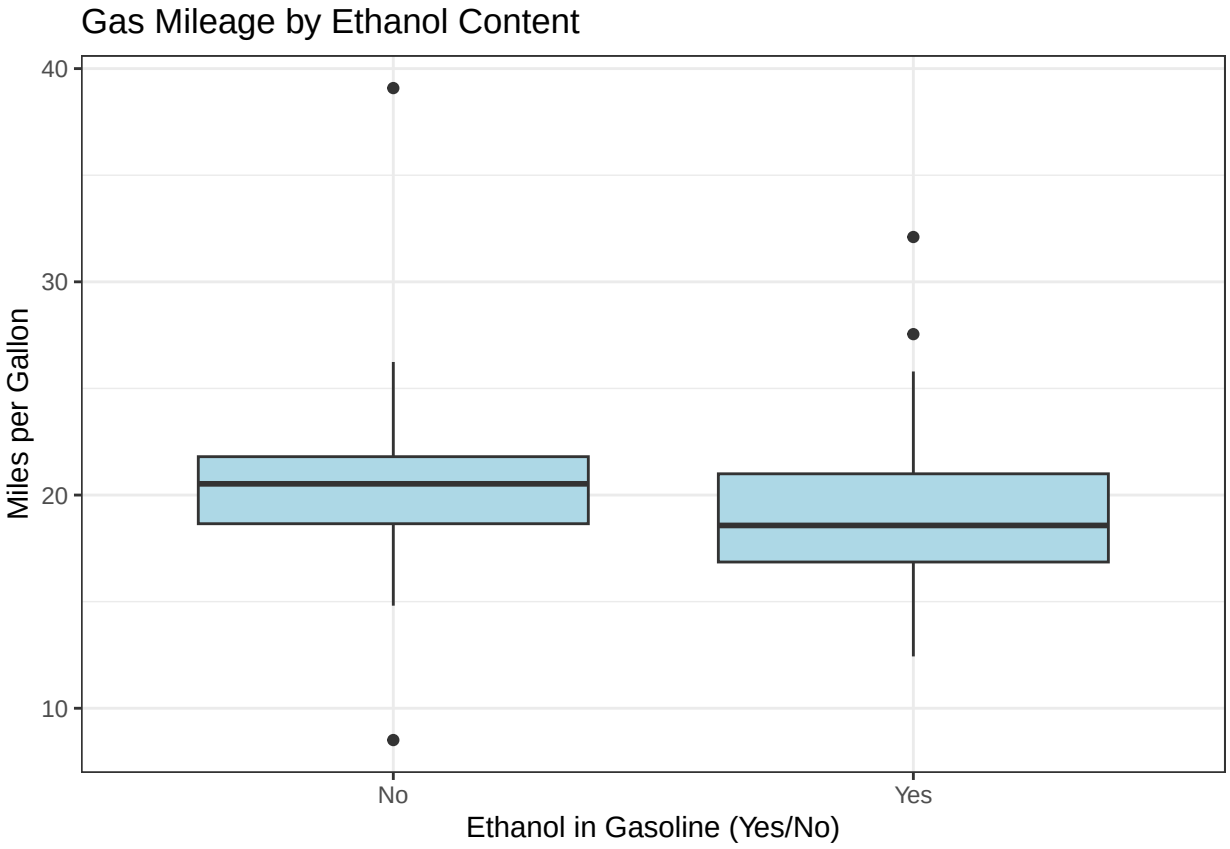
Alex Cory

2025-05-05

Introduction

The data was obtained by measuring gas mileage for vehicles using two types of gasoline: regular gasoline with no ethanol and gasoline containing 10% ethanol. The primary question is: How is gas mileage different when using gas with no ethanol compared to gas with 10% ethanol?

Data



The dataset contains a total of 347 observations with complete data on miles per gallon and ethanol content. The response variable, miles per gallon, measures vehicle fuel efficiency. The explanatory variable, ethanol, is a categorical variable indicating whether the gasoline contained 10% ethanol (Yes) or no ethanol (No).

Of the 347 total observations, 118 were recorded using gasoline with no ethanol, and 229 were recorded using gasoline containing 10% ethanol.

The mean gas mileage for the non-ethanol group was 20.5 mpg with a standard deviation of 3.05. For the ethanol group, the mean was 18.9 mpg with a standard deviation of 2.88.

A box plot comparing mpg between the two fuel types indicates that vehicles using gasoline without ethanol tend to have higher gas mileage. The variability in mileage is similar between groups, but the non-ethanol group exhibits a higher median and upper range.

These results suggest a modest reduction in fuel efficiency associated with ethanol-blended gasoline.

Results

To analyze the effect of ethanol on gas mileage, we fit a linear regression model with miles per gallon as the response variable and ethanol content (Yes/No) as the explanatory variable. The model can be described as:

$$\text{mpg}_i = \beta_0 + \beta_1 \cdot \text{ethanol}_i + \varepsilon_i$$

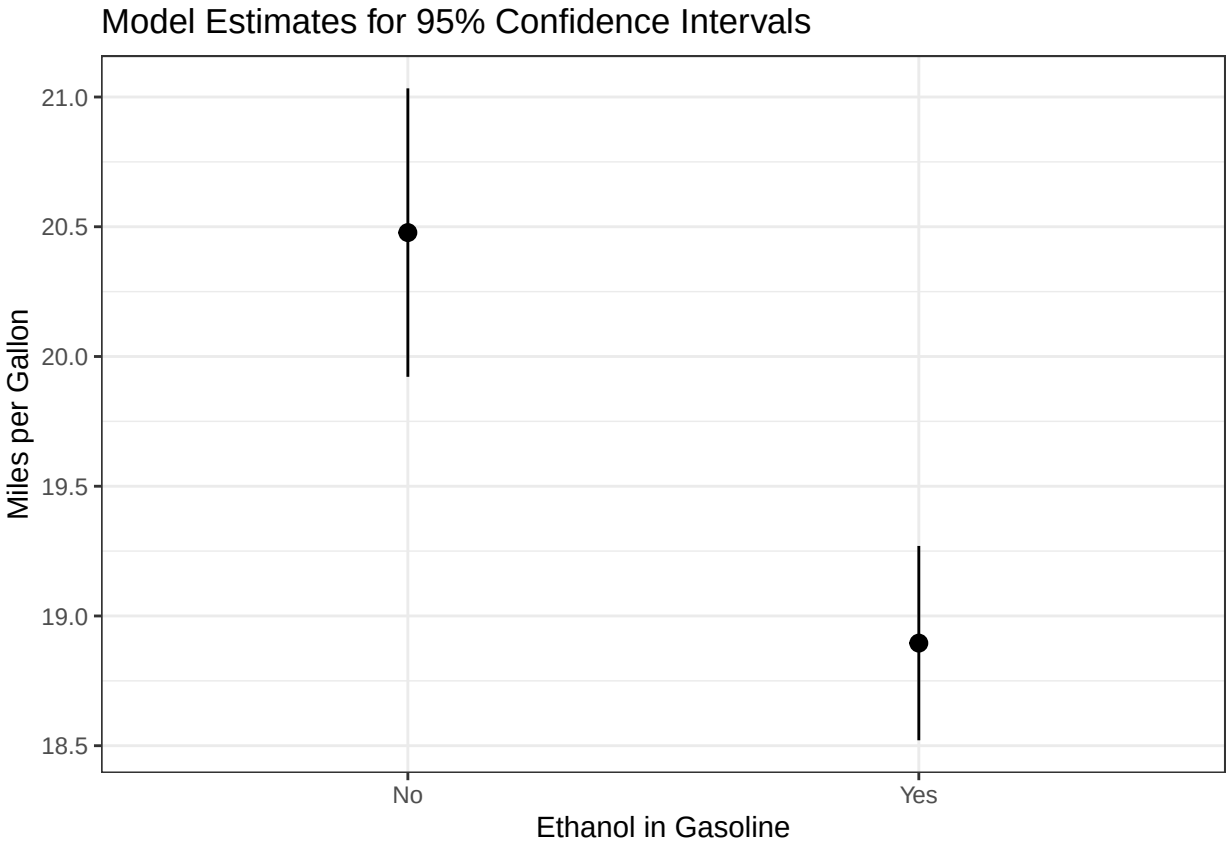
Here, β_0 represents the mean mpg for vehicles using gasoline with no ethanol, and β_1 represents the mean difference in mpg between the ethanol and non-ethanol groups.

This model was chosen because it allows us to estimate the average treatment effect of ethanol content on gas mileage, assuming linearity and normality of residuals. Since ethanol is a binary explanatory variable, this is equivalent to a two-sample comparison of means. An advantage of using a linear model compared to a more complicated model, such as a tree-based approach, is that it is highly explainable

From the fitted model, we estimate that the average gas mileage for vehicles using ethanol-blended gasoline is -1.58 mpg lower than that of vehicles using gasoline without ethanol. A 95% confidence interval for this difference is from -2.24 to -0.93 mpg.

This confidence interval does not include zero, indicating that there is a statistically significant difference in fuel efficiency between the two fuel types at the 5% level. Specifically, it suggests that ethanol-blended gasoline results in lower gas mileage on average.

The following figure shows the estimated mean mpg for each fuel type, along with 95% confidence intervals around the means:



The y-axis displays miles per gallon, and the x-axis distinguishes between gasoline with and without ethanol. Each point represents the sample mean mpg for that group, and the vertical bars represent 95% confidence intervals. The plot clearly shows that average mpg is lower when ethanol is present in the fuel, consistent with the regression results.

Discussion

The data used in this analysis were collected over a ten year period from a 2011 Toyota Sienna. At each gas fill-up, the number of miles on the trip meter was recorded and then reset, along with the amount of gas added, the ethanol content, and the fuel's octane rating. The miles per gallon (mpg) for each tank was then calculated from these entries. When ethanol options were available, one was chosen and noted. For the purposes of analysis, the ethanol variable was converted to a binary categorical variable with two levels: "Yes" for gasoline containing 10% ethanol and "No" for ethanol-free gasoline. Observations missing values for mpg or ethanol were removed to ensure accurate and interpretable results.

To address the primary research question of whether gas mileage is affected by the presence of ethanol, we used a simple linear regression model comparing the mean mpg between the ethanol and non-ethanol groups. This approach is equivalent to a two-sample t-test and is appropriate given the binary explanatory variable and continuous response.

The results suggest that gasoline containing ethanol results in lower average fuel efficiency. Specifically, vehicles using ethanol-blended gasoline had an average mpg that was 1.58 miles per gallon lower than those using ethanol-free gasoline. The 95% confidence interval for this difference, from -2.24 to -0.93 mpg, does not include zero, providing evidence of a statistically significant difference in mileage between the two fuel types.

In terms of assumptions, the model presumes that individual observations are independent, which is reason-

able given that each fill-up is treated as a separate event. While the dataset originates from a single vehicle, the fill-ups span various times and conditions. A review of the residuals indicates no serious violations of the assumptions of normality or constant variance, though residual diagnostics could be further explored.

Since the dataset is limited to a single make and model (a 2011 Toyota Sienna), the results may not generalize to all vehicles. However, they do provide suggestive evidence that ethanol-blended fuels reduce fuel efficiency in this context. Given that this is an observational study without randomized treatment assignment, the results should not be interpreted as establishing a causal relationship. Unmeasured factors, such as driving habits or load weight, could also influence fuel economy. Regardless, the consistency and size of the effect support the practical relevance of ethanol content in evaluating fuel efficiency.