

Hypothesis Testing Component

1. Difference in Mean Price Across Locations

Null Hypothesis (H0): There is no significant difference in the mean price of items across different locations.

Alternate Hypothesis (H1): There is a significant difference in the mean price of items across different locations.

Statistical Test: Kruskal-Wallis test

Why this test?

We chose the Kruskal-Wallis test for the following reasons:

- Our data was not normally distributed, which makes the use of a non-parametric test appropriate.
- We dealt with more than two independent samples.

Other Considered Tests

We considered using ANOVA as it is also used to compare means across multiple groups. However, it requires the data to be normally distributed, which was not the case with our data.

Metric for Success or Failure

We used the p-value to measure success and check for a statistically significant difference in the mean price across locations. A p-value close to zero (below our alpha level of 0.05) indicated that we should reject the null hypothesis.

Challenges with this test

We had to restructure the data due to inconsistencies in location names (e.g., "Ashland, MA" vs. "Ashland").

Results

Kruskal-Wallis H-test result: Stat=914.337, p=0.000

2. Difference in Mean Price Across Categories

Null Hypothesis (H0): There is no significant difference in the mean price of items across all categories.

Alternate Hypothesis (H1): There is a significant difference in the mean price of items across all categories.

Statistical Test: Kruskal-Wallis test

Why this test?

Again, we used the Kruskal-Wallis test due to:

- The lack of any particular distribution of the data.
- The use of multiple independent samples.

Other Considered Tests:

Although we considered ANOVA again, we decided against it due to the normal distribution requirement.

Metric for Success or Failure:

We compared the p-value to our alpha level (0.05), and with a p-value close to zero, we rejected the null hypothesis.

Challenges with this test

There were fewer challenges with data consistency across categories compared to locations.

Results

Kruskal-Wallis H-test result: Stat=1163.701, p=0.000

3. There is no Linear Relationship Between Mileage and Price in Vehicle Listings

Null Hypothesis (H0): There is no linear relationship between mileage and price in vehicle listings.

Alternate Hypothesis (H1): There is a linear relationship between mileage and price in vehicle listings.

Statistical Test: Simple linear regression

Why this test?

We chose linear regression to examine the continuous relationship between two variables, with mileage as the predictor.

Other Considered Tests

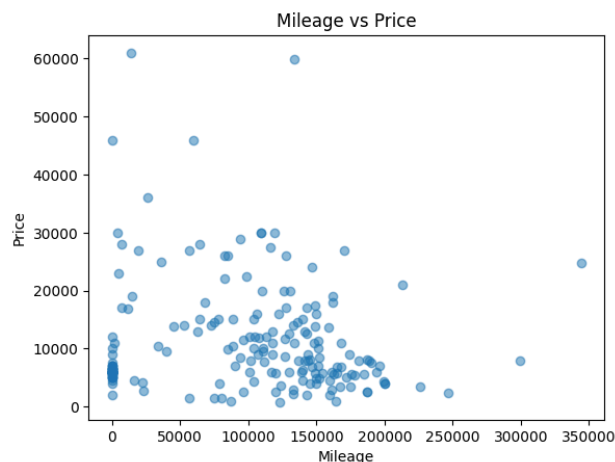
We considered using polynomial regression since the results of the linear regression test did not align with our data.

Metric for Success or Failure:

We used the correlation coefficient (r^2) to determine if there was a correlation between mileage and price in vehicle listings. Our linear regression test yielded an r^2 of -0.08, indicating that there is no relationship between the two variables.

Challenges with this test:

We used linear regression because we wanted to test if there was a linear relationship between the price and mileage. It turns out that linear regression is not the best way to capture the relationship between the mileage and price since the relationship might not be linear. As shown in the image below, we can observe that there is some correlation between the two variables. As expected, vehicles with lower mileage tend to have higher prices



Result

Mean Squared Error on Training Set: 33699795.680239655

Mean Squared Error on Testing Set: 25951609.13328298

The correlation coefficient on Test set: -0.08099521294104406

Summary

Difference in Mean Price Across Locations:

Using the Kruskal-Wallis test, we evaluated our null hypothesis that there is no significant difference in mean prices across locations. We got a p-value less than 0.05, leading us to reject the null hypothesis and conclude that there are significant differences in item prices among various locations.

Difference in Mean Price Across Categories:

We tested the null hypothesis using the Kruskal-Wallis test, which stated that there is no significant difference in the mean price of items across all categories. With a p-value less than 0.05, we rejected the null hypothesis, confirming significant price variations across categories.

Linear Relationship Between Mileage and Price In Vehicle Listings

We tested the Linear Relationship between mileage and price in vehicle listings using linear regression. Our null hypothesis was that “there is no linear relationship between mileage and price in vehicle listings” which we determined to be true (we did not reject the null hypothesis) based on the r^2 coefficient of -0.08.