To: Lex Luther, Managing Partner, Janzen Consulting Group

From: Data Analysis Team

Date: October 8, 2024

Re: Analysis of Electric Vehicle Population Data in Washington State

Mr. Luther, we have conducted an analysis of electric vehicle (EV) population data currently registered through the Washington State Department of Licensing (DOL) to inform our consulting strategy for clients in the automotive and energy sectors. This analysis is crucial for understanding EV adoption trends, which have significant implications for urban planning, energy infrastructure, and environmental policies in Washington State. Our analysis focuses on how Model Year impacts the Electric Range of EVs registered in Washington State. The target population for this study is all electric vehicles registered in Washington State as of October 2024. **Our hypothesis is that newer model year electric vehicles will have a greater electric range compared to older models**.
Spoiler: Contrary to our hypothesis, we found a statistically significant but very **weak negative relationship** between Model Year and Electric Range. However, **this relationship explains less than 0.1% of the variance in Electric Range, suggesting other factors may be more viable in determining an EV's range.**

We used the Electric Vehicle Population Data from the Washington State Department of Licensing, focusing on two key variables: Model Year (independent variable) and Electric Range (dependent variable).

Variable Analysis: **Model Year** encapsulates the year the electric vehicle was manufactured with numerical values representing years. The data spans from 1997 to 2024, with a mean of 2018 and median of 2018. The distribution is slightly left-skewed **(skewness = -0.25)**, indicating a **higher concentration of newer models**.

Summary Statistics for Model Year

N	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum	SD	Skewness
83283	1997	2017	2018	2018	2020	2024	3	-0.25

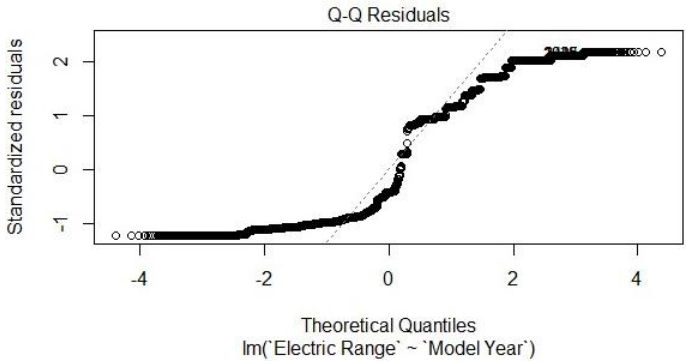
Electric Range encapsulates the distance an electric vehicle can travel on a single charge with the numerical values representing the distance of miles. The electric range varies widely from 6 to 337 miles, with a mean of 123.19 miles and median of 84 miles. The distribution is right-skewed (skewness = 0.43), **suggesting a concentration of vehicles with lower ranges and fewer high-range vehicles.** There were no missing values reported for either variable.

Summary Statistics for Electric Range in Miles

N	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum	SD	Skewness
83283	6	32	84	123.2	215	337	99.1	0.43

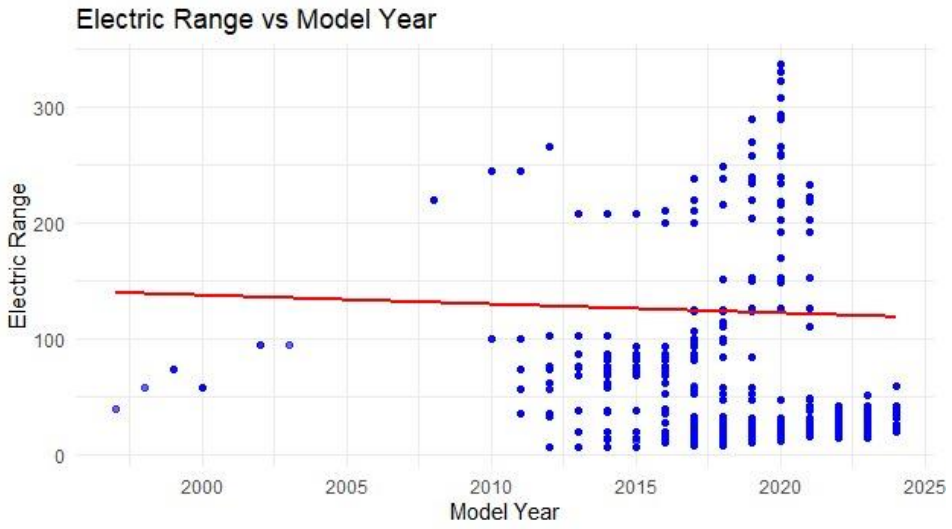
Regression Analysis: Our simple linear regression model yielded the following results utilizing **an alpha level of 0.05 (Ref. Table 1).** The model is statistically significant (**F-statistic: 45.42, p-value: 1.6e-11**). The coefficient for Model Year is -0.771 (p-value: 1.60e-11), **indicating a slight decrease in Electric Range for newer models.** The R^2 value is 0.0005451, meaning **the model explains only 0.05451% of the variance in Electric Range**. Regression model equation: **Electric Range = 1679.6944 - 0.7712(Model Year) + ε**

Summary of Assumptions: Linearity: Potential issue due to the pattern in the Residuals vs Fitted plot **(Table 4). Homoscedasticity**: Possible violation based on the clustering of residuals **(Table 3). Normality**: Violated, as seen from the Q-Q plot. **Influence/Outliers**: No significant violations observed in the Cook's distance and leverage plots **(Table 5).**



Statistical Significance: The findings are statistically significant at the conventional alpha level of 0.05. The entire model is statistically significant (F-statistic: 45.42, p-value: 1.6e-11). The independent variable (Model Year) is also statistically significant (p-value: 1.60e-11). We can reject the null hypothesis (H0) that there is no relationship between Model Year and Electric Range. Given the large sample size (N = 83,283) and statistical significance, we can generalize these findings to the target population of electric vehicles.

Substantive Significance:



- The IV regression coefficient is -0.771, indicating that for each year increase in Model Year, the Electric Range decreases by 0.771 miles.

- While statistically significant, this coefficient may not be substantively meaningful. A decrease of less than 1 mile per year is relatively small compared to the average Electric Range of 123.19 miles. The R^2 value is extremely low (0.0005451 or 0.05451%), indicating that Model Year explains only a tiny fraction of the variance in Electric Range. This suggests that other factors not included in the model are likely more important in determining Electric Range.

- For power analysis, the f^2 statistic is 0.0005454 (calculated as $R^2 / (1 - R^2)$). Given the large sample size, the power of this analysis is likely to be high, despite the small effect size. However, the practical significance of this high power (0.9999991) is limited due to the very small effect size.

Potential Weaknesses and Gaps: The model explains very little of the variance in Electric Range, suggesting important variables are missing. The negative relationship between Model Year and Electric Range is counterintuitive and may indicate confounding factors or a non-linear relationship. The model assumes a linear relationship, which may not be appropriate for this data. The analysis doesn't account for different makes or models of electric vehicles, which could significantly impact Electric Range.

Major Takeaways: While there is a statistically significant relationship between Model Year and Electric Range, the practical significance of this relationship is minimal. The model explains less than 0.1% of the variance in Electric Range, indicating that other factors are far more important in determining an electric vehicle's range. Surprisingly, the model suggests a slight decrease in range for newer models, which contradicts the general expectation of technological improvement over time. This analysis highlights the importance of considering both statistical and practical significance. A large sample size can lead to statistical significance even for very small effects. It's crucial to consider non-linear relationships and potential confounding variables in future analyses.

Conclusion: Given the weak relationship found and the counterintuitive direction of the effect, we recommend not acting on these findings alone. Instead, we suggest conducting a more comprehensive analysis that includes additional variables such as battery capacity, vehicle weight, and manufacturer. This would provide a more accurate and useful model for understanding factors influencing Electric Range. Additionally, exploring non-linear relationships and potential interactions between variables could yield more insightful results for decision-making regarding electric vehicle technology and marketing strategies.

Tables:

Table 1

Stargazer Output:

```
Regression Results
================================================
                          Dependent variable:
                      ---------------------------
                              Electric Range
------------------------------------------------
Model Year                      -0.771***
                                 (0.114)

Constant                      1,679.694***
                                (230.954)

------------------------------------------------
Observations                     83,283
R2                                0.001
Adjusted R2                       0.001
Residual Std. Error       99.095 (df = 83281)
F Statistic           45.420*** (df = 1; 83281)
================================================
Note:                *p<0.1; **p<0.05; ***p<0.01
```

Table 2

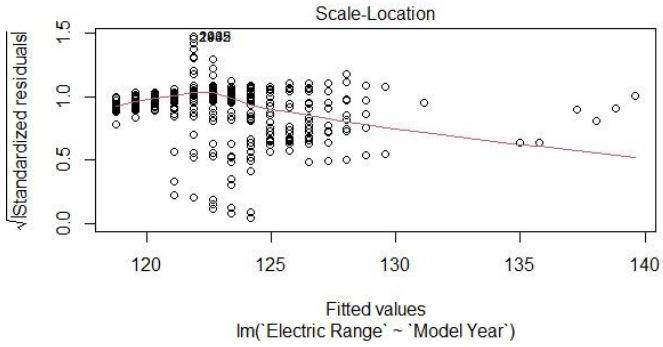


Table 3

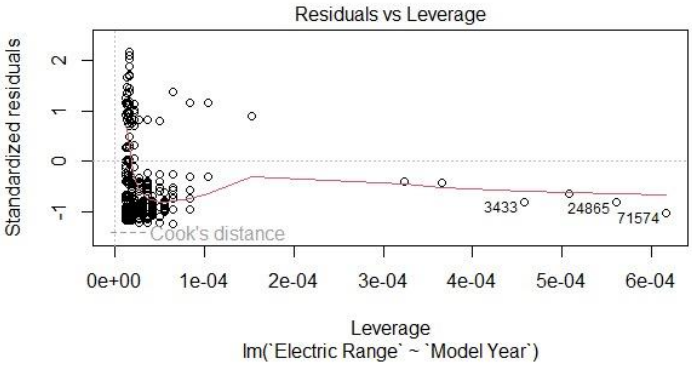


Table 4

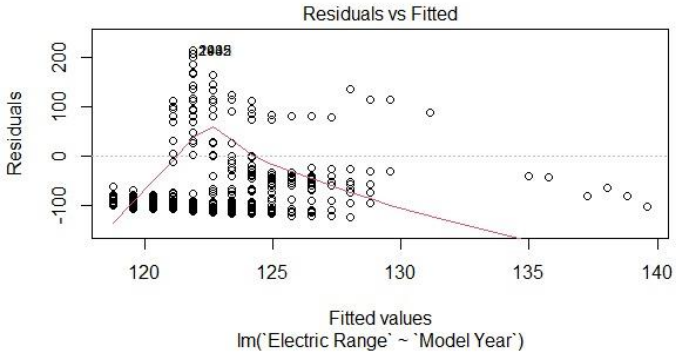
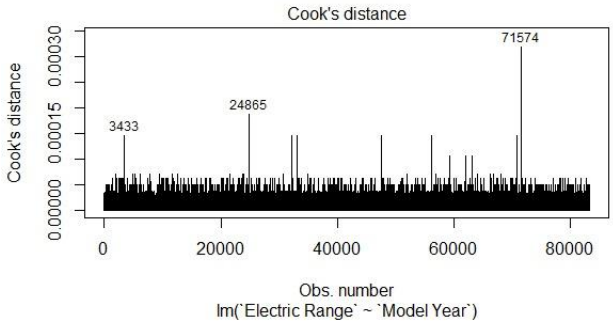


Table 5



References:

Publisher data.wa.gov. (2024, September 20). *State of Washington - Electric Vehicle Population Data*. Catalog. https://catalog.data.gov/dataset/electric-vehicle-population-data