**Final Project**

**EDA – CO2 emission by vehicles Data**

Madhavi Ghanta

Master of Data Science,

Bellevue University

DSC 530 T301 2237

Aug 8, 2023

**Statistical Question**

I have considered the dataset of Canada Co2 emission with regards to vehicles based in Canada. I have done the EDA over all lower end to high-end Car makes.

The question I want to address is: what vehicle mechanical demographic variables are most correlated with an increase in CO2 emissions?

I also had a few secondary questions I wanted to answer:

* if low end priced cars and high-end premium cars got a variation towards fuel consumption,
* if CVTs are more or less likely to be found on larger engines

**Outcome of EDA**

In my EDA I made my conclusion that the biggest factors in increasing a vehicle’s emissions are its variant, engine, and fuel consumption. The cylinder number goes hand-in-hand with displacement, so that is included as well. What surprised me was that transmission and fuel types did not affect emissions as much as said variables. Based on this, I would recommend purchasing a compact, midsize sedan, or station wagon, equipped with a 4-cylinder engine running regular gasoline, to reduce emissions. Such cars are usually very inexpensive, so it’s a win-win situation!

I then plotted a PMF of the fuel consumption of gas and premium gas cars. There was indeed a difference between premium Vehicles that had higher fuel consumption vehicles, it made me to run the hypothesis test to test the difference in means and the output yielded a p-value significantly smaller than zero, confirming that regular gasoline does, in fact, consume less fuel. It is also much less expensive, so the only reason one should purchase premium gas is if their car requires it.

The box plot of emissions vs car manufacturer. I found that (no surprise) all sports car and most luxury car brands are guilty of emitting more CO2 than other cars. Based on the plot, the following brands have the lowest emissions and are the brands I would recommend purchasing Fiat, Honda, Hyundai, Mazda, Mini, Smart, and Volkswagen. If one is adamant about purchasing a luxury car, one should consider Audi or Acura; these brands are in the middle of the emissions spectrum.

After calculating the percentage of CVT transmissions found for all cars with a certain number of cylinders and plotting this against the number of cylinders. The observation I made here is the occurrence of CVTs dropped with increasing engine size. Agreed, as CVTs are already very expensive and have issues with higher loads.

**Assumptions:**

Due to a lack of variables, the dataset did not contain information on vehicle year, engine torque/RPM, or weight, which I feel would certainly impact emissions. Variable “Year” would have also revealed how emissions changed over a subsequent amount of time. A few assumptions I made were that the trends seen in Canada apply to the UK, and that fuel octane ratings match that of the UK.

**Challenges faced:**

The main challenge I faced in understanding PMF and CDFs –that the reference books and the team’s discussion helped me to get on with it.

File resource:

<https://www.kaggle.com/debajyotipodder/co2-emission-by-vehicles>

**References:**

Downey, Allen B. (2014). *Think Stats.* O’Reilly Media.