I analyzed a dataset of 7,385 vehicles over the 2013-2020 model years in the Canadian market. My goal was to answer the question: what engine variables are most correlated with an increase with carbon dioxide (CO2) emissions? I also had a few secondary questions I wanted to answer: 1) if gas cars and premium gas cars got different fuel consumption ratings, 2) if CVTs are more or less likely to be found on larger engines, and 3) which car manufacturers one should avoid to be environmentally conscious.

I concluded that the biggest factors in increasing a vehicle’s emissions are its shape, engine displacement, and fuel consumption. 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 consumptions of gas and premium gas cars. There was indeed a difference (premium gas had higher fuel consumption), so I ran a hypothesis test to test the difference in means. This 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.

I then calculated the percentage of CVT transmissions found for all cars with a certain number of cylinders and plotted this against the number of cylinders. I saw that the occurrence of CVTs dropped with increasing engine size. This makes sense, as CVTs are already very expensive and have issues with higher loads.

I then constructed a 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 on purchasing a luxury car, they should consider Audi or Acura; these brands are in the middle of the emissions spectrum.

I feel that a lot of variables were missed. My dataset did not contain information on vehicle year, engine torque/RPM, or weight, which I feel would certainly impact emissions. Year would have also revealed how emissions changed over time. A few assumptions I made were that the trends seen in Canada apply to the US, and that fuel octane ratings match that of the US. One challenge I faced was running PMFs/CDFs using Python libraries rather than ThinkStats2, which I still could not resolve. In the future, I would like to fully understand how to run code using only libraries rather than relying on external files.

Word count: 490

References:

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