Secondary Analysis of carbon dioxide emissions and petroleum consumption

Mayel Espino

University of San Diego

# Author Note

# Abstract

.

*Keywords:*

# Carbon dioxide emissions and petroleum consumption

Global warming is ubiquitous in the news coverage, and in the political discourse. With the effects that Greenhouse gas (GHG) emissions on global warming becoming undeniable and considering that twenty eight percent of the total greenhouse emissions in the United States originate from transportation, according to the Environmental Protection Agency (EPA, web page) the need for a statistical model to study this in greater detail is clear. The objective of this study is to apply the statistical tools and methods we learned in class to formulate a statistical model. The statistical model will be tested and conclusions drawn about the hypotheses postulated on the model. The hypotheses will be used to determine, for example: What are the principal characteristics of the population that have the greatest effect on the emission of greenhouse gases. Further hypotheses should be formulated and conclusions drawn from this model given its relevance to the immediate future of humanity.

Best efforts will be made to ensure that the model created for this study and the hypotheses postulated in this study are unbiased and mathematically sound, so that that they can be trusted and proven useful for future use.

This is a secondary study, the original study was done by the EPA and the sample data was acquired from said study.

# Method

The data is comprised of a variety of metrics on attributes that define an automobile. We were provided a data sample obtained from FuelEconomy.gov web services(web page citation). The size is considerable, with over forty three thousand records. The sample has a mix of pure electric vehicles 0.5952% , Premium and Electricity 0.2131%, Premium Gas or Electricity 0.1228%, Regular Gas and Electricity 0.1320%, Regular Gas or Electricity 0.0093%, and internal combustion vehicles 98.93%. For all vehicles the same variables are measured and recoded, there are two categories of variables: key variables and control variables. The key variables are

primary fuel tailpipe carbon dioxide emissions in grams per mile, and annual primary-fuel petroleum consumption in barrels. The control variables are: Combined miles-per-gallon for the primary fuel type, make, engine displacement in litters, engine cylinders, combined luggage and passenger volume in cubic feet, vehicle type, transmission type, and primary fuel type.

The following table provides descriptive statistics for the key variables.

**Table 1**

*Discrete data distribution and location*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *co2TailpipeGpm* | *barrels08* | *comb08* | *displ* | *volume* |
| Minimum | 0.0 | 0.06 | 7.00 | 0.000 | 0.00 |
| 1st Quartile | 386.4 | 14.33 | 17.00 | 2.200 | 0.00 |
| Median | 444.4 | 16.48 | 20.00 | 3.000 | 86.00 |
| Mean | 462.8 | 17.15 | 20.85 | 3.287 | 66.93 |
| 3rd Quartile | 522.8 | 19.39 | 23.00 | 4.300 | 113.00 |
| Maximum | 1269.6 | 47.09 | 141.00 | 8.400 | 538.00 |

**Table 2**

Categorical data counts

|  |  |  |  |
| --- | --- | --- | --- |
| make | *count* | *mfrCode* | *count* |
| Chevrolet | 0.0 | blank | 30808 |
| Ford | 386.4 | GMX | 1786 |
| Dodge | 444.4 | BMX | 1378 |
| GMC | 462.8 | FMX | 1023 |
| Toyota | 522.8 | CRX | 964 |
| BMW | 1269.6 | TYX | 941 |
|  |  | Other | 6277 |

The following formula is the generalized lineal model (GLM), this formula helps us describe or model the relationships between all the variables involved. The GLM also accounts for an error ( alongside all the factors of our equation. The error is not a mistake but rather a deviation from our model. The is the that the rest of the variables B0, B1 … Bn do not account for.



B0 is the baseline value for our model, when representing the model on a graph, this is the location on the Y axis that our line intercepts.

The specific GLM we will use is the Linear Regression model. This model works well with continues variables, not categorical, so I am wondering why we have categorical values in our

MPG = B0 + B1(co2TailpipeGpm) +B2(barrels08) + B3(comb08) + B4(make) + B5(make\_id) + B6(mfrCode) + B7(displ) + B8(cylinders) + B9(volume) +B10(vehtype) + (2)



Table three and table four are the bivariate table which illustrates the sample proportions of two of our categorical variables. The variables being considered are vehicle type and transmission type.

**Table 3**

Bivariate table, part one

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | *Population*  *N(%)* | *Premium Gasoline*  *N(%)* | *Midgrade Gasoline*  *N(%)* | *Regular Gasoline*  *N(%)* |
| Vehicle Type |  |  |  |  |
| Unknown | 19,730(45.7%) | 3,491(27.3%) | 90(69.2%) | 15,346(53.4%) |
| Hatchback | 5,070(11.17%) | 1,313(10.3%) | 0(0%) | 3,535(12.3%) |
| 2 door | 6,394(14.8%) | 3,157(24.7%) | 12(9.2%) | 3,120(10.9%) |
| 4 door | 11,983(27.8%) | 4,840(37.8%) | 28(21.5%) | 6,732(23.4%) |
| Transmission Type |  |  |  |  |
| Automatic | 30,210(70.0%) | 9,411(73.5%) | 130(100%) | 19,588(68.2%) |
| Manual | 12,956(30.0%) | 3,390(26.5%) | 0(0%) | 9,143(31.8%) |

**Table 4**

Bivariate table, part two

|  |  |  |
| --- | --- | --- |
| Variable | *Diesel*  *N(%)* | *Electricity*  *N(%)* |
| Vehicle Type |  |  |
| Unknown | 34(56.7%) | 84(32.7%) |
| Hatchback | 2(3.3%) | 105(40.9%) |
| 2 door | 1(1.7%) | 1(0.4%) |
| 4 door | 23(38.3%) | 67(26.1%) |
| Transmission Type |  |  |
| Automatic | 60(100%) | 248(100%) |
| Manual | 0(0%) | 0(0%) |

From the tables we should note some key relationships between the proportions of the sample population and specific categories. One key relationship is the proportional amount of unknown vehicle type to the all the fuel types. From this I conclude that comparing fuel type to vehicle type is of very limited value. On the other hand, the comparison of transmission type to fuel type is very useful.

Table 5 shows a similar

**Table 5**

Association of Emissions Category by Fuel Type and Other Characteristics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | *Population*  *N(%)* | *Ultra-low emissions*  *N(%)* | *Very-low emissions*  *N(%)* | *Low emissions*  *N(%)* | *Standard*  *N(%)* |
| Primary Fuel Type |  |  |  |  |  |
| Premium Gasoline | 12,801 (29.6%) | 24 (7.5%) | 70(18.2%) | 1,169(21.0%) | 9,798(33.2%) |
| Midgrade Gasoline | 130(0.3%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 124 (0.4%) |
| Regular Gasoline | 28,733 (66.5%) | 40(12.5%) | 311(81.0%) | 4,066(73.2%) | 18,971(64.2%) |
| Diesel | 1,196 (2.8%) | 0(0.0%) | 0(0.0%) | 303(5.5%) | 629(2.1%) |
| Vehicle Type |  |  |  |  |  |
| Unknown | 19,730 (45.7%) | 91(28.3%) | 50(13.0%) | 739(13.3%) | 12,579(42.6%) |
| Hatchback | 5,070 (11.7%) | 122(38.0%) | 128(33.3%) | 1,820(32.8%) | 2,952(10.0%) |
| 2 door | 6,394 (14.8%) | 7(2.2%) | 11(2.9%) | 703(12.7%) | 5,193(17.6%) |
| 4 door | |  | | --- | | 11,983 (27.8%) | | 101(31.5%) | 195(50.8%) | 2,294(41.3%) | 8,819(29.9%) |
| Transmission Type |  |  |  |  |  |
| Automatic | 30,210 (70.0%) | 312(100.0) | 301(78.4%) | 3,202(57.6%) | 20,730(70.2%) |
| Manual | 12,956 (30.0%) | 0(0.0%) | 83(21.6%) | 2,354(42.4%) | 8,813(29.8) |

**Table 6**

Association of Emissions Category by Fuel Type and Other Characteristics

|  |  |  |
| --- | --- | --- |
| Variable | *Polluter*  *N(%)* | *Gross Polluter*  *N(%)* |
| Primary Fuel Type |  |  |
| Premium Gasoline | 1,262(21.4%) | 478(32.4%) |
| Midgrade Gasoline | 6(0.1%) | 0(0.0%) |
| Regular Gasoline | 4,358(73.9%) | 987(67.0%) |
| Diesel | 259(4.4%) | 5 (0.3%) |
| Natural Gas | 14(0.2) | 4 (0.3%) |
| Electricity | 0(0.0%) | 0(0.0%) |
| Vehicle Type |  |  |
| Unknown | 5,119(86.8%) | 1,152(78.2%) |
| Hatchback | 47(0.8%) | 1(0.1%) |
| 2 door | 339 (5.7%) | 141 (9.6%) |
| 4 door | 394 (6.7%) | 180 (12.2%) |
| Transmission Type |  |  |
| Automatic | 4,557(77.3%) | 1,108(75.2%) |
| Manual | 1,341(22.7%) | 365(24.8%) |

Correlation tables

**Table 7**

Pearson Correlation Coefficients (N=42,917)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | co2TailpipeGpm | barrels08 | comb08 | make\_id | displ |
| co2TailpipeGpm | 1.0000 | .9885 | (.9184) | (.2157) | .7954 |
| barrels08 | .9885 | 1.0000 | (.9050) | (.2117) | .7843 |
| comb08 | (.9184) | (.9050) | 1.0000 | .2072 | (.7327) |
| make\_id | (.2157) | (.2117) | .2072 | 1.0000 | (.2823) |
| displ | .7954 | .7843 | (.7327) | (.2823) | 1.0000 |
| cylinders | .7438 | .7337 | (.6863) | (.2670) | .9046 |
| volume | (.4323) | (.4266) | .4161 | .1165 | (.3628) |
| vehtype | (.3626) | (.3580) | .3313 | .0940 | (.2631) |
| emissionscat | .8894 | .8791 | (.8415) | (.1755) | .6703 |
| prifueltype | (.1128) | (.1084) | .1234 | .0710 | (.2149) |

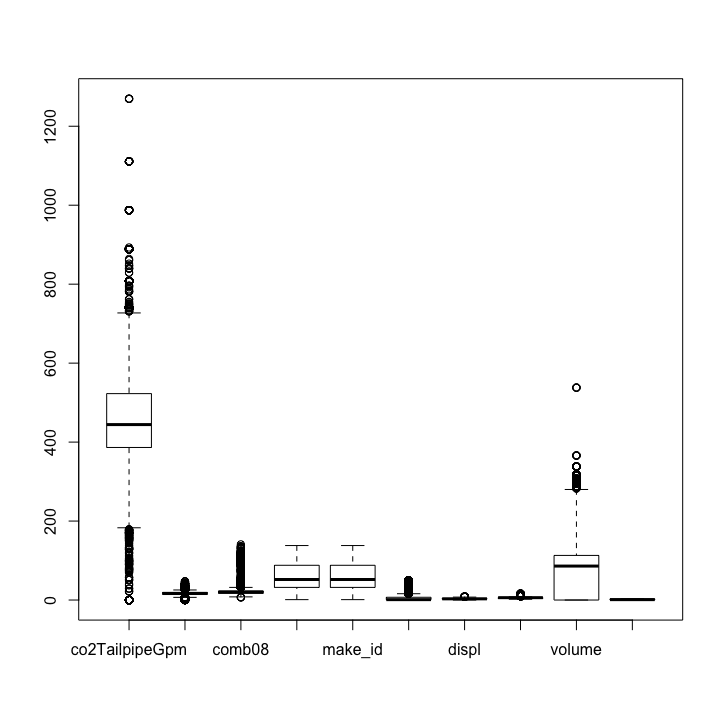
Table 8

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | cylinders | volume | vehtype | emissionscat | ??? |
| co2TailpipeGpm | .7438 | (.4323) | (.3626) | .8894 | (.1128) |
| barrels08 | .7337 | (.4266) | (.3580) | .8791 | (.1084) |
| comb08 | (.6863) | .4161 | .3313 | (.8415) | .1234 |
| make\_id | (.2670) | .1165 | .0940 | (.1755) | .0710 |
| displ | .9046 | (.3628) | (.2631) | .6703 | (.2149) |
| cylinders | 1.0000 | (.2648) | (.1524) | .6185 | (.2181) |
| volume | (.2648) | 1.0000 | .7418 | (.3627) | .0498 |
| vehtype | (.1524) | .7418 | 1.0000 | (.3054) | (.0340) |
| emissionscat | .6185 | (.3627) | (.3054) | 1.0000 | (.0874) |
| prifueltype | (.2181) | .0498 | (.0340) | (.0874) | 1.0000 |

# Appendix

The following box plot, figure 1, further illustrates the distribution

**Figure 1**



PDFco2 = f(x;µ = 465.538, (1)

References

<https://www.epa.gov/transportation-air-pollution-and-climate-change/carbon-pollution-transportation>

<https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions>