Homework #4. Analyze real-world emissions data

Use the example data files provided in Learning Suite on the schedule for November 2

Feat.class.example.csv

1. Graph histograms of the fuel-based emission rates (g/kg-fuel) for HC, CO, and NOx for each date of measurements.

10 pts

Chart

Description automatically generated with low confidence

* 1. How much does the maximum emission rate at each location compare to the median emission rate? (maximum divided by the median emission rate)

10 pts

The maximum CO emission rate compared to the median, is between 52 to 504.

For NO it ranges between 53 to 171 (if you ignore the negative median 7/14/2022.)

For HC, the max to median is not as pronounced, ranging from 11 to 23.

Table

Description automatically generated

* 1. Which pollutant has the most skewed distribution?

6 pts

There are multiple ways to measure skew. Visually with the graphs above, and different formulas.

If we were to use the max/median as our measurement, then NO has the highest skew for all dates, except for 8/2/2022. That date the CO distribution is skewed by the very large outlier which has an outlier of 800 g/kg-fuel, which is ~500 times higher than the median CO measurement for that day.

Using the Pearson's second skewness coefficient (median skewness), the NO has a higher skew than CO, and CO has a higher skew than HC.

1. Do gasoline vehicles with high emissions of CO also tend to have high emissions of HC and NOx?

6 pts

There is a rather weak positive relationship between CO, HC, and NOx. (See graphs and R2 below). If a vehicle is high emitting at one of the three pollutants, then there is a slightly higher probability that the vehicle will also be high emitting at the other pollutants. However, due to the weak relationship among the pollutants, the probability only increases slightly, and just because a vehicle that is high emitter for one of the three pollutants evaluated, does not mean it is a high pollutant for the other two pollutants.

* 1. Compare the data graphically

10 pts

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

* 1. What’s the R2 correlation between the different pollutants?

CO and HC: R2 = 0.11

NO and CO: R2 = 0.08

NO and HC: R2 = 0.01

6 pts

1. Vehicle emission Inspection and Maintenance (I/M) Programs are designed to prevent high emitting vehicles from operating on the road, either by requiring the owner to fix their vehicle, or to remove it from the vehicle fleet.

Researchers have not found meaningful impacts of I/M programs on reducing vehicle emissions and improving air quality[1], [2]. In addition, I/M programs are shown to impact lower income households more than higher income households[3].

* 1. How do you think we should address vehicle emissions from high-emitting vehicles?

10 pts

My own thoughts on the issue:

Use money that goes for emissions testing to fund in-use surveillance (rather than just electronic scan of engine as an emissions tests)

Increase funding public transportation

Increase tolling (costs of driving)?

Increased penalty for purposely tampering with your vehicle. At the city or state level, make it a criminal offence for citizens to tamper with their emission control systems, and have it enforced by policy officers with penalty to lose their driver’s license

1. Calculate the **90%** confidence intervals of the mean CO emission rate for gasoline vehicles for each day. Are the gasoline vehicle rates from each site significantly different from one another?

Note: You can edit my R code for this problem to re-sample the measurements.

If you do not have familiarity with R, you can approximate using the t-distribution.

10 pts (distributions), + 6 pts (answering the question)

Using the t-distribution:

All locations are not significantly different from one another, except for Timp Hwy East and University Avenue on 7/14/2022, which are significantly different than one another.

Note: that the mean CO emission rate from Timp Hwy West is not significantly different than 0.

A screenshot of a computer

Description automatically generated with low confidence

Using resampling:

The uncertainty bounds for the first three days are quite similar. The lower bound and upper bound is slightly higher using the resampling methods.

One significant difference is the Timp Hwy West is now significantly larger than 0. (the confidence intervals don’t include 0).

Note: that the mean CO emission rate from Timp Hwy West is not significantly different than 0.

Otherwise, the inferences we make similar inferences about differences in the means between the four sampling days (e.g. Univ Ave on 7/12/22 and Timp Hwy East are not significantly different using both data sets).

Table

Description automatically generated with medium confidence

4.1 What are advantages and disadvantages of using the re-sampling method?

10 pts

Re-sampling Advantages:

Great for non-normal data.

It appropriately accounts for the impact of outliers

We don’t need to assume our data comes from a certain statistical distribution—we use the distribution of our own observations.

Disadvantages:

Could give misleading results if our sample does not represent our population of interest. For example, if we sampled more vehicles, perhaps we would see super-high emitters that we did not observe in our observation of 100 vehicles.

If the distribution of the population is known, it can be advantages

Unfamiliarity. Most people are used to confidence intervals based on t-distributions, but are not used to it based on re-sampling distributions.

Computations. Re-sampling requires a computer to re-sample your data set. For some problems, this could become computationally intensive.

1. Table MF-2 from Highway Statistics provides total sales of gasoline and special (diesel) fuel in 2020 by state in 1000’s of gallons[4].

We measured 50 diesel vehicles at 10 am across the four locations. The average diesel vehicle NO emission rate (g/kg) for the 10 am time period is:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| pollutant | mean | median | sd | n | min | max | lower.95 | upper.95 |
| CO | 13.59425 | 6.759731 | 30.53268 | 50 | -18.2548 | 190.3729 | 4.828864 | 22.35964 |
| HC | 5.590014 | 4.538387 | 21.32577 | 50 | -31.9126 | 113.5856 | -0.53223 | 11.71226 |
| NH3 | 0.133595 | 0.070125 | 0.434446 | 46 | -0.41951 | 2.41061 | 0.003155 | 0.264035 |
| NO | 4.631564 | 0.164322 | 10.75692 | 50 | -1.56201 | 46.8546 | 1.543446 | 7.719683 |
| NO2 | 0.536723 | 0.423853 | 1.270595 | 29 | -2.16289 | 3.984007 | 0.04486 | 1.028586 |

* 1. Use the road-side emission factors (g/kg-fuel) to estimate the total amount of NOx (NO + NO2) emitted from gasoline and diesel in Utah.

You can use the energy density provided in the MOVES fuel report. The density of gasoline is 2.839 kg/gallon, and the density diesel is 3.167 kg/gallon [5]

16 pts

Text

Description automatically generated

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | NOx emission rate | Fuel density | Utah Fuel Sales (Annual) | | NOx emissions | |
|  | (g/kg) | (kg/gallon) | gallons | kg | metric tonnes | % contribution |
| Gasoline | 1.75 | 2.839 | 1.16E+09 | 3.29E+09 | 5,767 | 40.2% |
| Diesel | 5.17 | 3.167 | 5.25E+08 | 1.66E+09 | 8,588 | 59.8% |
| Total |  |  |  |  | 14,355 | 100.0% |

* 1. Calculate the % contribution of vehicle NOx from gasoline and diesel vehicles.

6 pts above

* 1. How does the % contribution from gasoline and diesel vehicles compare to the NOx emission distribution estimated for Salt Lake City using MOVES in Homework #3?

6 pts

MOVES Results

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NOx | | | | | VOC | | | | |
| Fuel | 2022 | | 2030 | | % Change | 2022 | | 2030 | | % Change |
|  | kg | % | kg | % | % | kg | % | kg | % | % |
| Gasoline | 414.8 | 41% | 135.9 | 24% | -67% | 166.7 | 84% | 85.54 | 84% | -49% |
| Diesel Fuel | 600.2 | 59% | 438.3 | 76% | -27% | 30.6 | 15% | 14.23 | 14% | -54% |
| CNG | 2.6 | 0% | 2.4 | 0% | -8% | 1.3 | 1% | 1.84 | 2% | 37% |
| E85 | 0.3 | 0% | 0.2 | 0% | -38% | 0.2 | 0% | 0.10 | 0% | -33% |
| Total | 1017.9 | 100% | 576.7 | 100% | -43% | 198.8 | 100% | 101.71 | 100% | -49% |

The relative split between gasoline and diesel compares very well. (I’m surprised actually). Using the one-hour MOVES run, the gas/diesel split in 2022 was estimated to be 41%/59%. Using estimates from fuel sales, the gas/diesel split is 40%/60%.

* 1. Do you expect the two estimates to differ from one another? Why?

6 pts

Yes, I expect moderate differences, for several reasons, including:

* The HW#3 MOVES runs modeling emissions from additional vehicle types that were not measured at the roadside (particularly heavy-duty vehicles)
* The HW#3 MOVES SLC run is just looking at emissions from July 2022 at 5 pm (not emissions for the entire year).
* MOVES HW#3 is modeling activity on all the roadtypes in SLC at different average speeds, it is also including NOx emissions from extended idling and starts which were not measured at the roadside
* The relative split of gasoline/diesel estimate of emissions and fuel consumption should be different between just SLC (MOVES runs), and from the fuel sales (for the entire state of Utah)
* The roadside measurements are only based on the vehicle operation at three different locations for less than one hour
* MOVES emission rates are based to be representative of national conditions, and rely on default activity inputs.
* MOVES is based on default miles driven per year, which may not account for COVID in the defaults. The other method uses actual fuel sales.
  1. What answer do you think is more representative of the real-world emissions in Utah?

6 pts

Reasons for MOVES:

* Includes all the vehicle activity (all driving on freeways, different types of roads, and NOx emissions that occur not on the roads (start and extended idle)
* Includes all vehicle types (including heavy-duty vehicles)

Reasons for Road-side

* Fuel sales are more linked to NOx emissions, then VMT.
* We are using actual measurements of near-road vehicles, that accounts for the observed prevalence of high-emitters which is difficult to estimate using nation-wide data
* Does not rely on data on activity (or assumptions), regarding the vehicle age distribution, the vehicle speeds, the operating mode distribution, the vehicle class distribution. Each of these inputs (and others), leads to uncertainty in estimating real-world emissions.

1. Ideally, we would have more measurement locations and data. Use the hypothetical data in the CO.site.average.date.2023.csv, calculate the average CO gasoline vehicle emission rate for Utah County with 95% CI.

10 pts

(Correct) Use the mean of the location means (n=8 locations). My population of interest is the average gasoline emission rate for all of Utah County.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| mean | sd | n | min | max | tcrit | bound | lower.95 | upper.95 |
| 16.75 | 8.77 | 8 | 8.16 | 31.80 | 2.36 | 7.84 | 8.91 | 24.58 |

(Incorrect), use the daily mean observations (n=15 daily observations). However, the means conducted at the same location are correlated with one another. Due to the lower standard deviation, and the higher number of observation, the uncertainty bound of the mean is signicantly smaller than using the locations mean using the correct method.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| mean | sd | n | min | max | tcrit | bound | lower.95 | upper.95 |
| 15.95 | 7.37 | 15 | 6.71 | 31.80 | 2.14 | 4.22 | 11.72 | 20.17 |

* 1. What are you using as your random, independent variable?

10 pts.

The mean CO gasoline emission rate at each location is my random, independent variable.

N = 8 using location means, instead of N = 15 for daily means. However, the mean CO emission rates can be assumed to be random, independent variables, whereas, the daily mean observations are not independent, random variables, because the daily means sampled at the same location are correlated to one another (not random or independent).

[1] G. A. Bishop and D. H. Stedman, “A Decade of On-road Emissions Measurements,” *Environ. Sci. Technol.*, vol. 42, no. 5, pp. 1651–1656, Mar. 2008, doi: 10.1021/es702413b.

[2] N. J. Sanders and R. Sandler, “Technology and the Effectiveness of Regulatory Programs over Time: Vehicle Emissions and Smog Checks with a Changing Fleet,” *Journal of the Association of Environmental and Resource Economists*, vol. 7, no. 3, pp. 587–618, May 2020, doi: 10.1086/707954.

[3] R. J. Wessel, “Policing the poor: The impact of vehicle emissions inspection programs across income,” *Transportation Research Part D: Transport and Environment*, vol. 78, p. 102207, Jan. 2020, doi: 10.1016/j.trd.2019.102207.

[4] “Table MF-2 / Highway Statistics 2020 - Policy | Federal Highway Administration.” https://www.fhwa.dot.gov/policyinformation/statistics/2020/mf2.cfm (accessed Nov. 05, 2022).

[5] USEPA, “Greenhouse Gas and Energy Consumption Rates for Onroad Vehicles in MOVES3,” Office of Transportation and Air Quality, U.S. Environmental Protection Agency, Ann Arbor, MI, USA, EPA-420-R-20-015, Nov. 2020. [Online]. Available: https://www.epa.gov/moves/moves-onroad-technical-reports