Particle Emissions from Cars at an Intersection

Dean Gladish

Math 245, Professor Loy. April 17, 2018.

Introduction

Particulate matter, or PM (a complex mixture of nitrates, sulfates, and other organic particles), is a topic of interest to many environmental health scientists due to its propensity to cause bronchitis and other potentially fatal diseases. Thus, we seek to find out if it is possible to predict PM levels based on human activity; using 500 observations about particulate matter, we want to find if there is any correlation.

Data

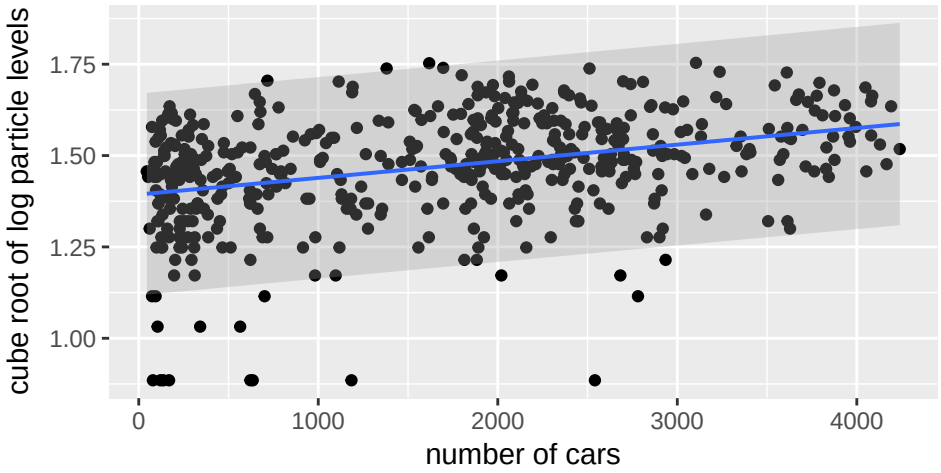
Our data consists of 500 observations about particulate matter (PM) numbers, measured in parts per million, that were taken in conjunction with data on the number of cars passing through an intersection. There are no missing data and they are given in no particular order. Our analysis includes 21 outliers for the sake of accurate representation.

Table 1: Table 1. Summary Statistics for PM and Number of Cars

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
Particles (ppm)	2	15	27	37.88	48	220	35
Number of Cars	45	523.5	1851.5	1683.2	2509.5	4239	1154.5

Exploratory data analysis reveals skewness, high variability, and outliers; to account for this non-linearity, we perform analysis in which the cubic root of the log-transformed Particle concentration is used as our response variable.

Figure 1. Scatterplot of transformed PM levels fitted with 95% prediction intervals



Results

We use the simple linear regression model

$$\widehat{\mu}\{\sqrt[3]{log(Particles)}|Cars\} = \beta_0 + \beta_1 Cars$$

where Particles denotes particle pollution in parts per million and Cars denotes the number of cars at an intersection. The following table gives our estimates of the parameters for this model.

Table 2: Table 2. Coefficients of our Model

	Estimate	Standard error	t value	P-value
intercept	1.393	0.01109	125.69	2e-16
Cars	4.548e-05	5.433e-06	8.37	5.85e-16

Our model indicates that increasing the number of cars by a factor of two is associated with a 3.706% increase in the median PM level at the time. The 95% confidence interval for this increase is (3.705%, 3.708%). This model explains 12.3% of the variability in the cubic root of the log of particle concentrations.

For this intersection given that there are 2000 cars present, we predict that there should be 1865 pollutants per million particles. Our 95% prediction interval for this estimate is 58 to 282980.

Discussion

Our model predicts with 95% certainty that the particulate level will generally remain within 50.6 and 208343 parts per million with an average around 50.6 ppm. As a policy prescription we can only recommend that health scientists suggest limiting the number of cars to below one thousand as this reduces the incidence of extreme outliers.

Furthermore, due to the observational nature of this study we can not make any causal inferences related to human auto activity and the measure of PM. The general trend is indicative that the number of cars on this intersection is slightly positively associated with particulate matter levels in a rather exponential manner. We recommend further data collection under different weather conditions and a spatially broader means of PM assessment. Our equal variances and normally distributed error assumptions being met for the most part, our simple linear model provides a reliable yet sometimes inaccurate prediction of the number of particles in the air based on the number of cars in place.