The association of access to clean water and infant mortality : Reproducing results from Aronow and Miller [2019, Sec. 4.4]

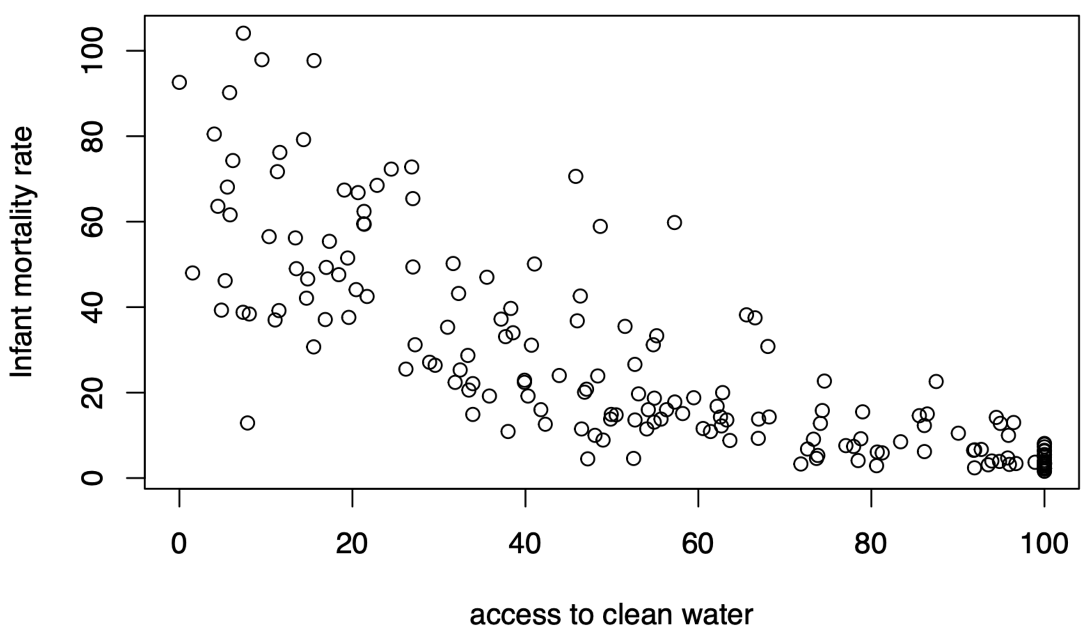
Anuradha Ramachandran

Introduction

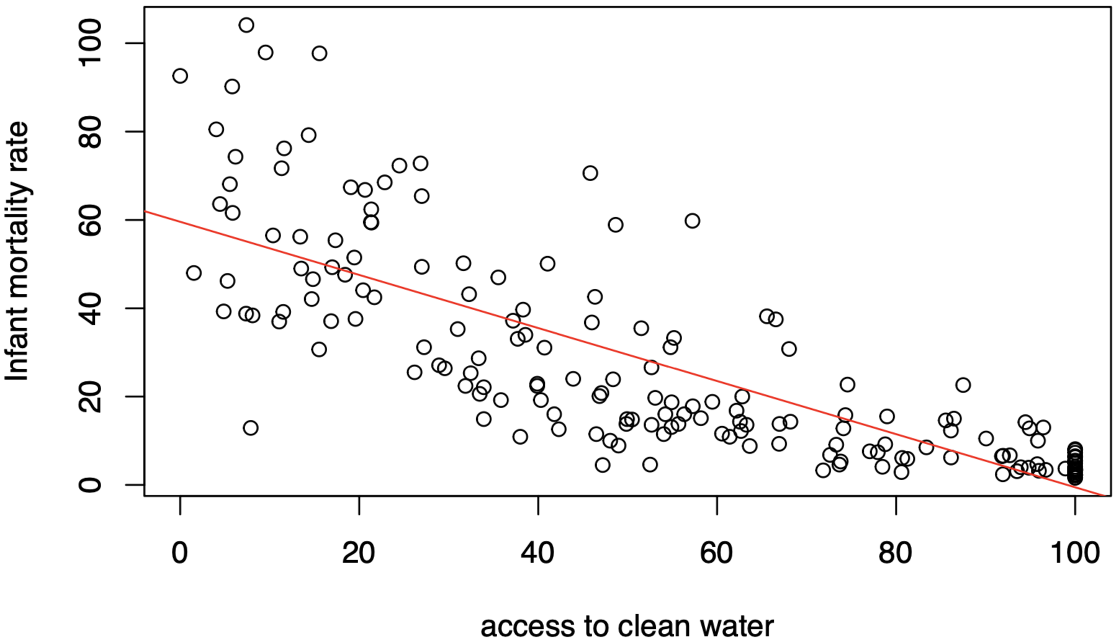
Quality of Governance dataset from January 2016 [Teorell et al., 2016, Aronow and Miller, 2019] is used to investigate the association between access to clean water and infant mortality. The dataset contains 184 rows(countries) and four columns:
• cname: country name
• wdi morinftot: infant mortality rate, as measured by the number of infants died before reach- ing one year of age per 1,000 live births, in a given year
• epi watsup: access to clean water, as measured by the percentage of the population with access to a source of clean driking water
• wdi accelectr: access to electricity, as measured by the percentage of the population with acceess to electricity
Let Y_i denote wdi_morinftot, X_i denote epi_watsup, and Z_i denote wdi_accelectr.

Linear regression model using only X_i to predict Y_i

From the scatter plot, there seems to be negative association between infant mortality rate and access to clean water, ie., as access to clean water increases, infant mortality rate decreases.



Fitting a simple linear regression model to predict the infant mortality using access to clean water, $Y_i = \hat{\beta}_{y1.x} + \hat{\beta}_{yx.1}X_i + \hat{e}_{y.1x,i}$
Estimates of the simple linear regression model:
Intercept: 59.5832: On average 60% infant mortality rate is observed when there is no access to clean water.
Slope: -0.6013: When the access to clean water increases by 1 unit, on average a 6% decrease in infant mortality rate is observed. 95% CI using standard error and non parametric bootstrap is (-0.66 ,-0.54) for $\hat{\beta}_{yx.1}$.



Linear regression model using both X_i and Z_i to predict Y_i

A regression model to predict infant mortality using both access to clean water and access to electricity, can be written as
$Y_i = \hat{\beta}_{y1.xz} + \hat{\beta}_{yx.1z}X_i + \hat{\beta}_{yz.1x}Z_i + \hat{e}_{y.1xz,i}$
The estimates of the coefficients are:
Intercept is 74.18 and coefficient of X is -0.35 and Z is -0.37
The magnitude of coefficient of access to clean water using the two covariate model (0.35) is lower than the univariate model(0.60). $\hat{\beta}_{yx.1z}$ is the coefficient in the linear model that explains only the effect of access to water on infant mortality rate without including the effects of electricity.

Quadractic Model

We are now going to make our model a bit more flexible, and use a quadratic specification.
$Y_i = \hat{\beta}_{y1.xx^2} + \hat{\beta}_{yx.1x^2}X_i + \hat{\beta}_{yx^2.1x}X_i^2 + \hat{e}_{y.1xx^2,i}$
$Infant\_mortalityrate = 71.7508 - 1.2555 * Access\_to\_cleanwater + 0.0059 * (Access\_to\_cleanwater)^2$

Estimated coefficients of this model are:
Intercept = 71.7508
Access to $cleanwater$ = -1.2555 Access to $cleanwater^2$ = 0.0059

The coefficient of X(Access to clean water) $\hat{\beta}_{yx.1.x^2}$ cannot be directly interpreted as impact of change in X_i

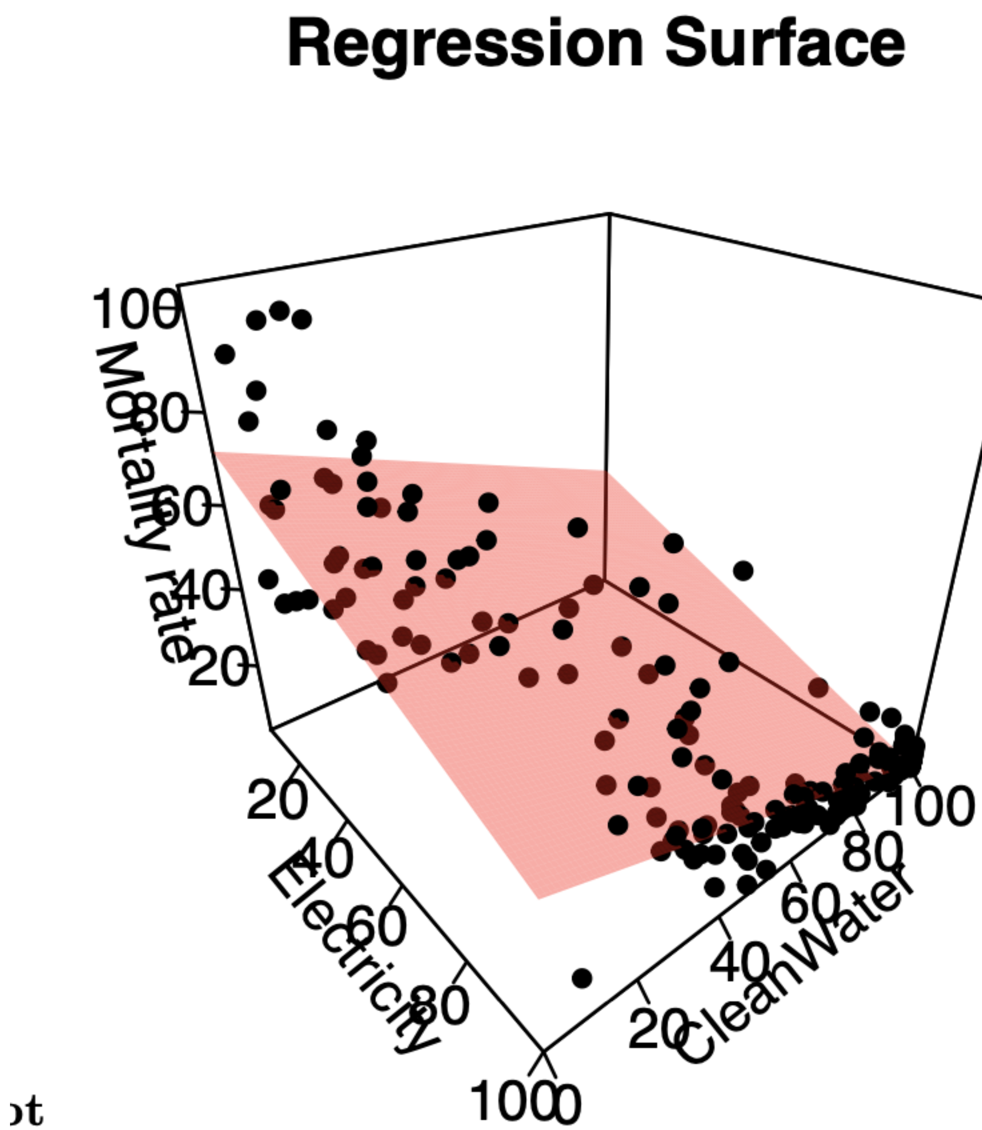
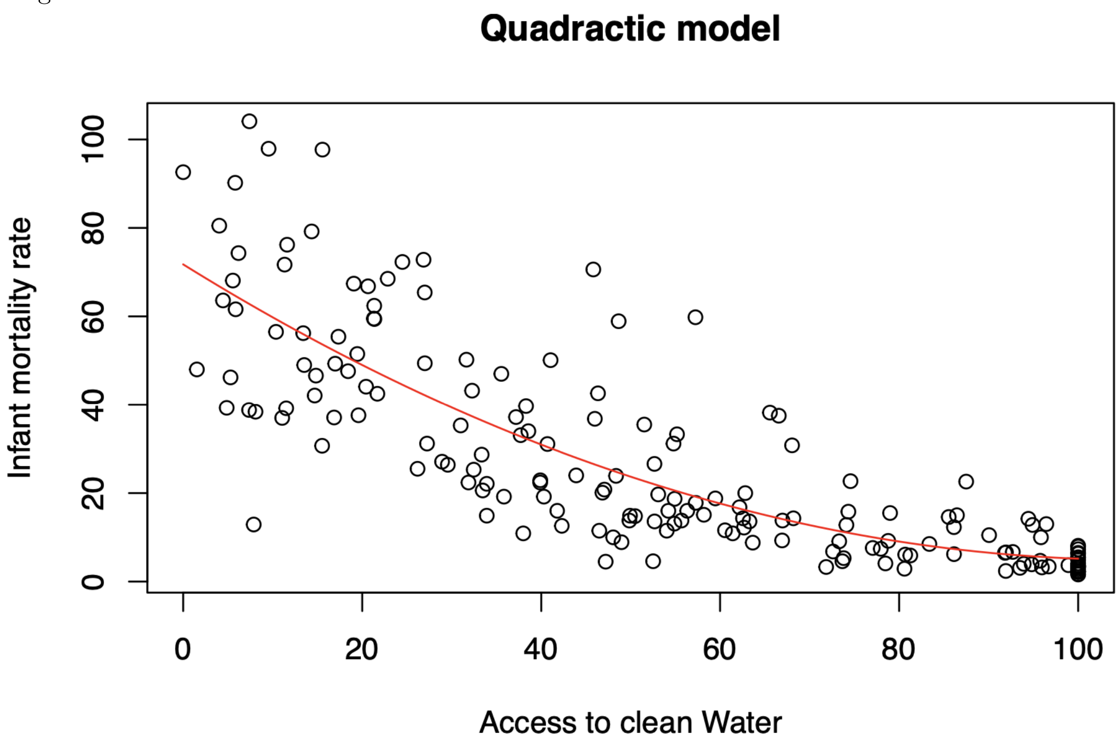

Regression Surface
100
80
60
40
20
Mortality rate
20
40
60
80
100
Electricity
0
20
40
60
80
100
CleanWater

Figure 1: 3D

on Y_i
The change of Y impacted by change in X $= -1.255 + 2 * 0.0059(Access\_to\_cleanwater)$.

The regression line is shown as:



The regression surface is shown as:

Regression Surface