Mortality and Major Factors

Jiasheng Xiao

Abstract

1. We choose mortality to be our response variable
2. Our purpose is to find which factors will have influenced on the Mortality. Does air pollution affect the Mortality?
3. We will use regression model to approach our result and will also use the relative weights to determine which factors have the most effect Mortality

Data set

Our data set

1. contains 17 variables
2. have 60 observations
3. from 60 Standard Metropolitan Statistical Areas in USA

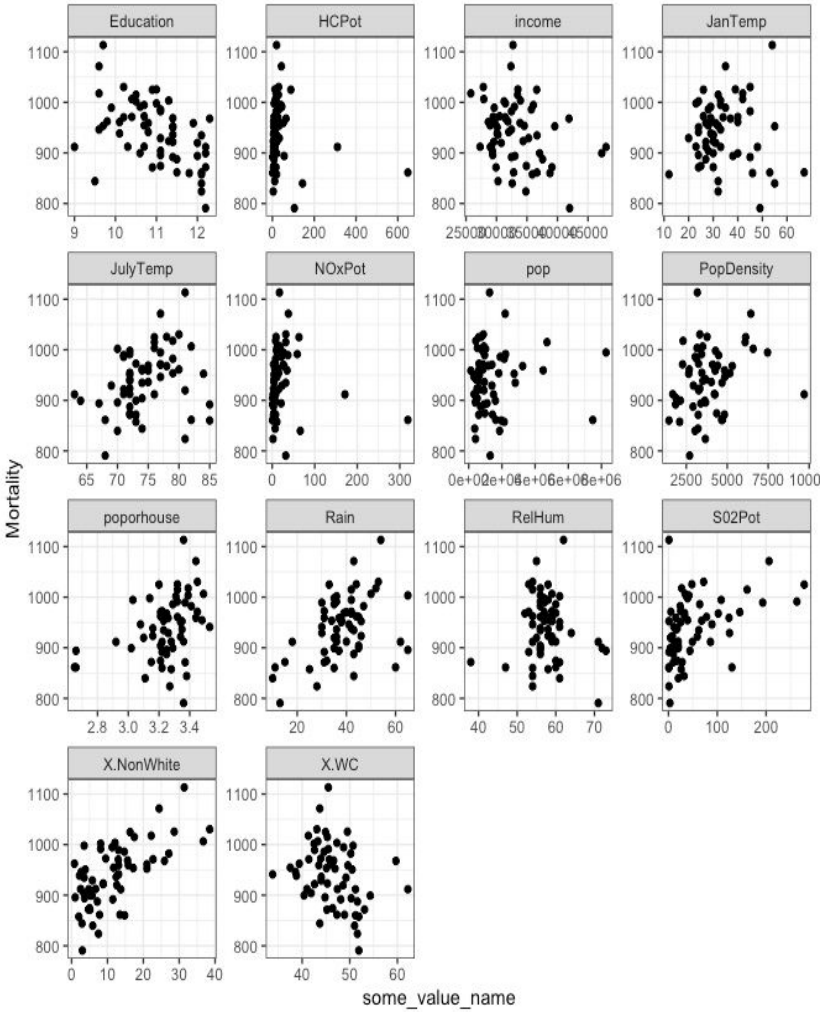
Replace Missing values

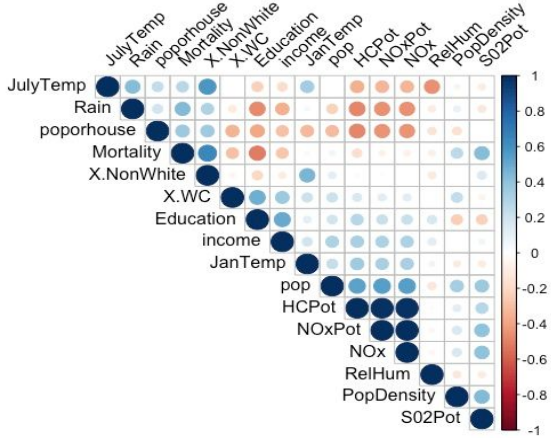
1. There are two missing values in our data set.
2. We use the median(914427 and 32452) of these variables (population and income)to replace the missing values (in FortWorth).

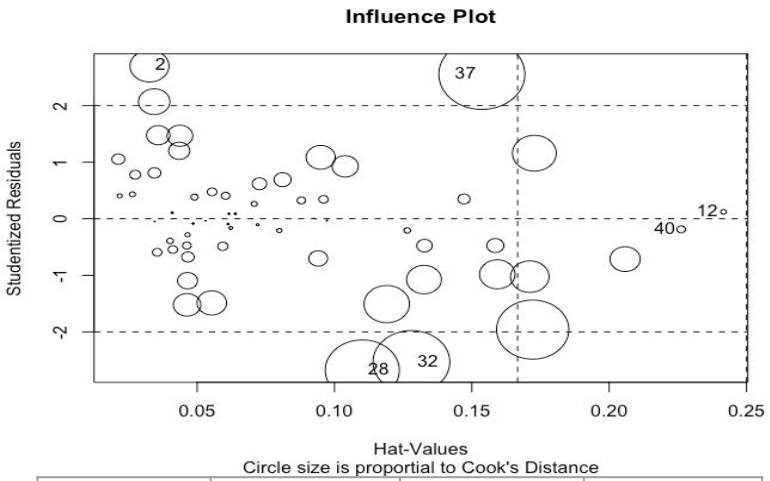
Predictor and outliers

Find predictors

Based on the scatter plot and correlation between mortality and other variables, we choose Rain, %Nowhite, SO2Pot, Education for our predictors .





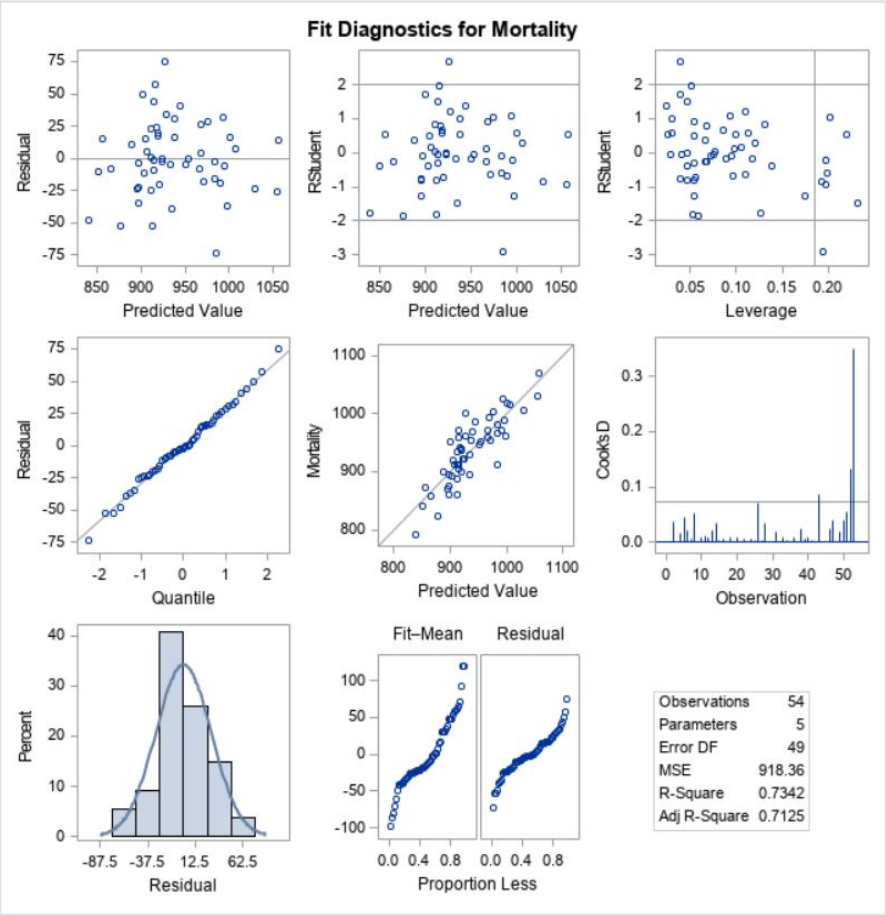


Obs	StudRes	Hat	CookD
2	2.69	0.03	0.04
12	0.12	0.24	0.01
28	-2.67	0.11	0.15
32	-2.53	0.12	0.17
37	2.55	0.15	0.21
40	-0.18	0.22	0.01

Find outliers

1. We use influence plot to look for the outliers
2. There are six outliers in our data set
3. We do not have to remove the outliers

OLS Assumption Test



Durbin Waston Test

DW Value	P-Value<DW
1.8	0.19

Breusch-Pagan Test

BP Value	P-Value>BP
6.95	0.14

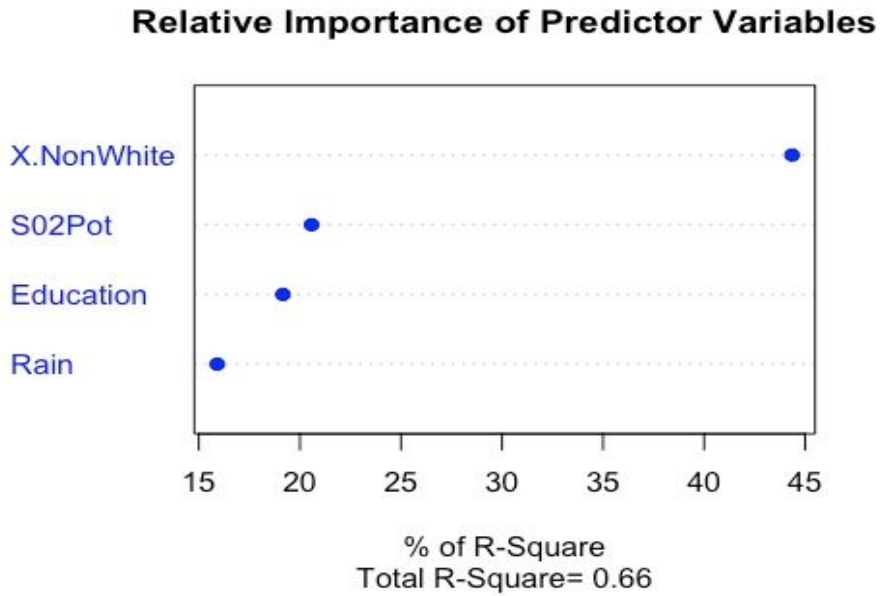
1. No Heteroscedasticity
2. Normal Distribution
3. No Autocorrelation

Variable	Parameter Estimate	Pr > |t|
Intercept	1022.85	<0.0001
rain	1.2	0.02
education	-16.89	0.018
nonwhite	3.32	<.0001
S02Pot	0.31	0.0004

R-Square	Adj R-Square
0.66	0.63

Variable	VIF
rain	1.53
education	1.57
nonwhite	1.17
SO2Pot	1.27

$$\hat{y} = 1022.85 + 1.2\hat{R} - 16.89\widehat{Ed} + 3.32\widehat{NW} + 0.31\hat{S}$$



Predictors	Weights
Education	19.16
SO2 Pot	20.57
Rain	15.91
X.NonWhite	44.35

Relative weights

Relative importance in this context means the proportion of the variance in y accounted for by xj.

It helps you figure out what variables contribute the most to r-squared.

X.NonWhite has the most significant effect on our response variable.

Conclusion

1. SO2Pot is an important air pollution indicator, and it is a significant predictor in our model. We could say that air pollution has effect on mortality.
2. Nonwhite and education reflect the social factor. So, mortality is also impacted by our social factor (eg. poverty and economy).
3. Rain is the natural indicator. Based on some researches, mortality was significantly associated with daily ambient temperatures and rainfall, after controlling for seasonality and long-term time trends, especially in rural area.

Q&A