MORTALITY ANALYSIS REPORT

Christina Thai
ID: 915508192

Statistics 108: Regression Analysis
University of Davis
Spring 2018

INTRODUCTION

This report will analyze the relation between pollution and mortality using data collected from 60 Standard Metropolitan Statistical Areas (SMSA) in the United States, obtained from the years 1959 – 1961 [Source: GC McDonald and JS Ayers, “Some applications of the ‘Chernoff Faces’: a technique for graphically representing multivariate data”, in Graphical Representation of Multivariate Data, Academic Press, 1978].

DATA COLLECTION

The data collected features environmental factors and demographic factors such as:

(1) [PRECIP] the mean annual precipitation (in inches),
(2) [EDUC] median number of school years completed by persons of age 25 or older
(3) [NONWHITE] percentage of population in 1960 that is non-white,
(4) [POOR] percentage of households with annual income under $3000 in 1960,
(5) [NOX] relative pollution of oxides of nitrogen (NO_X), and
(6) [SO2] relative pollution potential of Sulphur dioxide (SO_2).

*** NOTE: Relative pollution potential is the product of the tons emitted per day per square kilometer as a factor correcting for SMSA dimension and exposure.

The response variable [MORTALITY] is the age-adjusted mortality rate (deaths per 100,000 people in a population).

DATA QUALITY CHECK

To ensure that our data is “good”, we consider some graphic diagnostics for our predictors. We will use box plots to show the minimum and maximum value of each predictor, the first and third quartile, and median value. If the predictor variable values are skewed or have outliers, it is a good idea to transform the data, so it does not distort the fitted regression model.

OUTLIERS

Outliers of the predictor variables could influence the appropriateness of the fitted regression functions. Since the variables [NOX] and [SO2] are skewed and [NONWHITE] and [POOR] are skewed, we will transform them using the natural logarithm and cubic root, respectively.

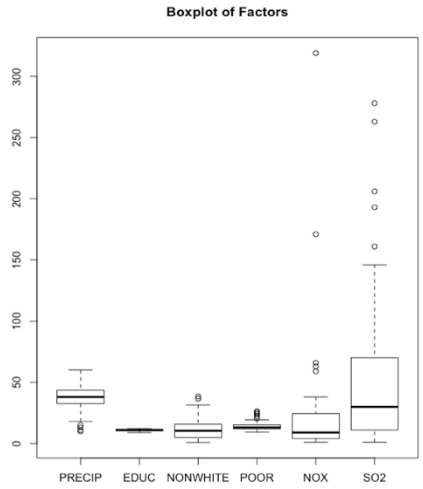


Figure 1: Boxplot of Explanatory Variables

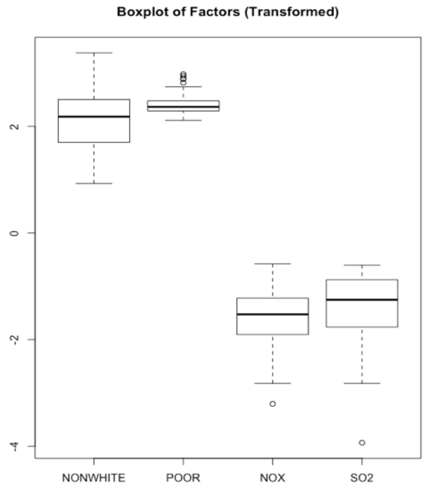


Figure 2: Boxplot of Transformed Predictors

Even after transforming the data, there are still outliers. Using the scatterplot matrix of the data we will examine in the distribution of data, located in the diagonal in the matrix, for each predictor.

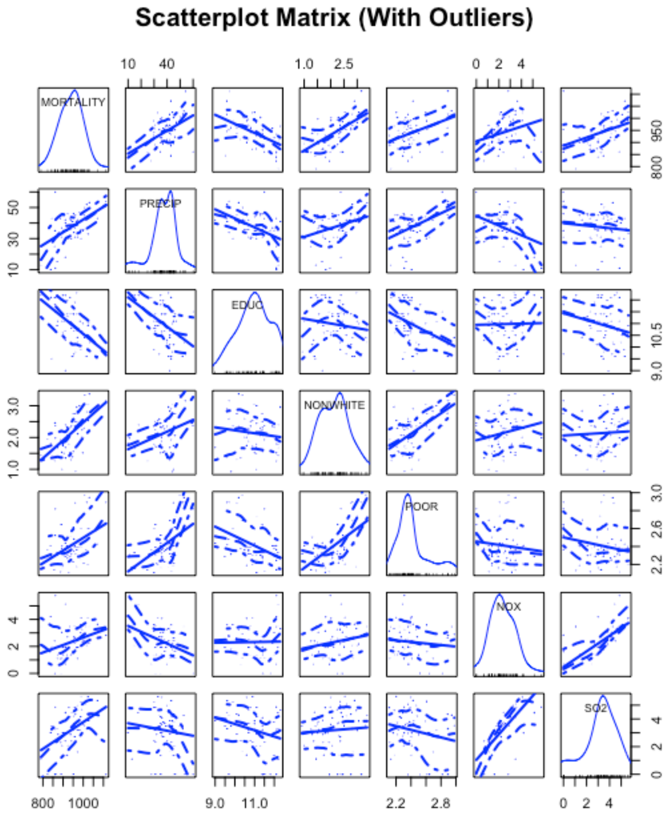


Figure 3: Scatterplot Matrix with Outliers

We will remove the outliers that appear to skew the distribution for each predictor. Referring to *Figure 3,* all the probability distributions of the predictors seem normal except for [POOR]. We will remove outliers from [POOR] until the distribution appears normal.

**NOTE: We removed nine outliers from [POOR] and there are 51 remaining values for the explanatory variable [POOR].

It is also important to note that removal of outliers may cause discrepancies in our analysis as the removed values could potentially hold significant information. Therefore, I will do analysis for both cases and conclude the difference.

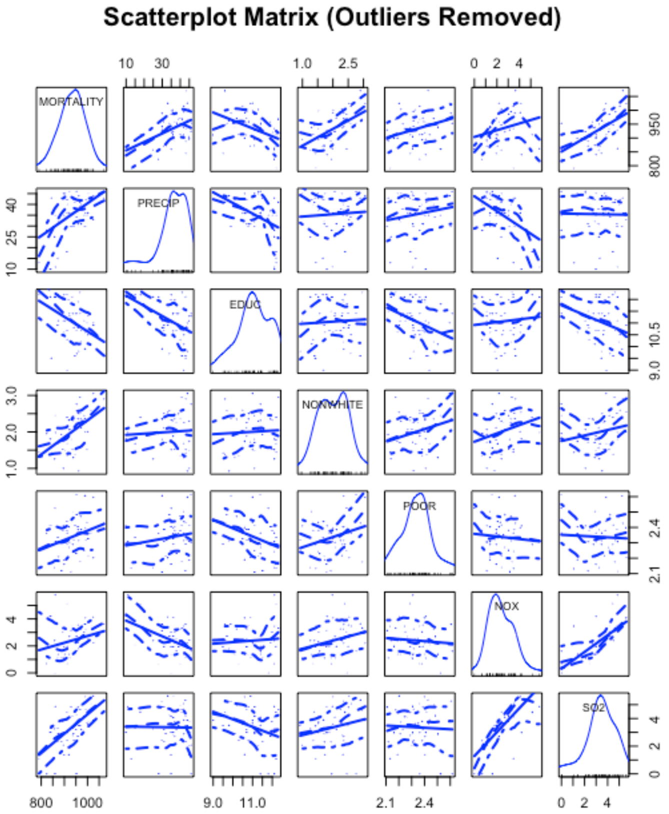
REGRESSION RELATION

To choose the functional form of the regression relation between each predictor variable and the response [MORTALITY], it is ideal to plot each explanatory variable against the response to determine if there is a linear relationship. Since there are many predictors, a scatterplot matrix is ideal since each box in the matrix shows the corresponding bivariate regression. Note that the diagonals in the matrix is the axis to determine which variable is being plotted.

For example, the scatter plot in row 2 and column 1 show a linear bivariate regression between [MORTALITY] and [PRESIP] and the scatter plot in row 5 and column 3 show the relation between [POOR] and [EDUC].

Analyzing the first column, or row, will provide a visual description of the regression relation between the response [MORTALITY] and each predictor variable. It appears that there are strong relationships between the response and all the predictors.

Figure 4: Scatterplot Matrix Without Outliers



A visual representation of the regression between [MORTALITY] and each predictor also provides a graphic description of the correlation for each predictor. As seen in row 1, there is a correlation with [MORTALITY], for all the predictors.

Additionally, we need to address the effects of multicollinearity. Scatter plots not in column or row 1 showing correlation can potentially imply a relationship between those two predictors and disrupt our resulting regression model. Figure 4 shows there is a relationship between:

[NOX] and [SO2]
[PRECIP] and [NOX]
[PRECIP] and [EDUC]
[EDUC] and [POOR].

We will not be considering interaction terms and instead, replace these terms with squares.

Below (Table 1.) is a Covariance-Variance matrix of the data that includes the outliers, which correspond to the relations we see in Figure 1.

	MORTALITY	PRECIP	EDUC	NONWHITE	POOR	NOX	SO2
MORTALITY	1	0.5094	-0.5109	0.6063	0.4099	0.2920	0.4031
PRECIP	0.5094	1	-0.4904	0.3193	0.4937	-0.3683	-0.1212
EDUC	-0.5109	-0.4904	1	-0.1359	-0.4167	0.017984	-0.2562
NONWHITE	0.6063	0.3193	-0.1359	1	0.6003	0.1977	0.05922
POOR	0.4099	0.4937	-0.4167	0.6003	1	-0.1041	-0.1955
NOX	0.2919	-0.3683	0.01798	0.1977	-0.1041	1	0.7328
SO2	0.4031	-0.1211	-0.2561	0.05921	-0.1955	0.7328	1

Table 1: ***Covariance Matrix of Data (with Outliers)***
The green cells denote moderate to strong correlations with the response, blue is moderate correlation between predictors, and yellow borderline strong correlation with potential of multicollinearity.

FITTING DATA IN MULTIPLE LINEAR REGRESSION

To estimate the regression function, we will be using the `summary()` command in R. This function also returns a five-number summary of the residuals, which will be significant during residual analysis. I will estimate four potential fits for linear regression. In all the summary outputs, the highlighted yellow implies a significance level of approximately 0 and blue for significance level of 0.001. The smaller the significance level, the smaller the probability of that predictor not having a estimated coefficient of zero (hypothesis testing for t-test).

(1) Data EXCLUDING Outliers:

```
Residuals:
    Min      1Q    Median     3Q     Max
-84.907 -16.704  -1.454    14.097  72.520

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 525.7308   161.2935   3.259 0.002157 **
PRECIP        2.5846     0.6654   3.884 0.000341 ***
EDUC          0.1517     7.4567   0.020 0.983864
NONWHITE     42.4116     9.8534   4.304 9.21e-05 ***
POOR         67.1742    42.6443   1.575 0.122369
NOX          -4.3365     7.6878  -0.564 0.575566
SO2          24.4736     6.2598   3.910 0.000316 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.37 on 44 degrees of freedom
  (9 observations deleted due to missingness)
Multiple R-squared:  0.7549, Adjusted R-squared:  0.7214
F-statistic: 22.58 on 6 and 44 DF,  p-value: 5.971e-12
```

The ‘Estimate’ column provides the estimated coefficients in our estimated fitted function:

$$\hat{Y} = (525.731) + (2.585)X_1 + (0.152)X_2 + (42.412)X_3 + (67.173)X_4 + (-4.337)X_5 + (24.474)X_6$$

Analyzing the coefficients of our estimated regression function shows that if all these factors are 0, the mortality rate would be 525.7308 (deaths per 100,000 people in a population).

The standard error is the standard deviation of the sampling distribution of the estimate of the coefficient. There seems to be a high standard error for the estimation coefficient of [EDUC], [POOR], [NOX], and [SO2].

The t-value is value of the test statistic for the hypothesis test of that estimated coefficient being 0. The p-value tells us the probability of the test statistic obtained is common if the null hypothesis (b = 0) is true. According to the summary's p-values, the predictors that play a significant role in our model are [PRECIP], [NONWHITE], and [SO2].

According to the multiple R-squared value generated in the summary, approximately 69.85% of the variation in [MORTALITY] can be explained by our model, which includes [PRECIP], [EDUC], [NONWHITE], [POOR], [NOX], and [SO2].

The obtained F-Statistic above is used measures the significance of the overall model, and not just one variable. The small p-value associated with the large F-statistic implies that the model is perhaps significant, which in this case is very good.

(2) Data INCLUDING Outliers

```
Residuals:
     Min       1Q   Median       3Q      Max
-104.554  -22.405    0.693   18.168   93.494

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 980.4750   141.9266   6.908 6.33e-09 ***
PRECIP        2.3748     0.6709   3.540 0.000844 ***
EDUC        -19.1004     7.6787  -2.487 0.016048 *
NONWHITE     49.9051    11.3256   4.406 5.15e-05 ***
POOR        -31.0975    34.5908  -0.899 0.372713
NOX          10.1044     7.1973   1.404 0.166178
SO2           8.0315     5.6263   1.427 0.159305
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 36.04 on 53 degrees of freedom
Multiple R-squared:  0.6985, Adjusted R-squared:  0.6644
F-statistic: 20.46 on 6 and 53 DF,  p-value: 3.139e-12
```

Using the data with outliers, the summary estimates the fitted regression line to be:

$$\hat{Y} = (980.475) + (2.375)X_1 + (-19.1)X_2 + (49.905)X_3 + (-31.098)X_4 + (10.104)X_5 + (8.032)X_6$$

Significant predictors are [PRECIP], [EDUC], and [NONWHITE] at approximately level 0.

(3) Squared Term of [NOX]

```
Residuals:
     Min       1Q   Median       3Q      Max
-102.734  -23.108   -2.776   20.972   83.585

Coefficients:
               Estimate Std. Error t value     Pr(>|t|)
(Intercept)    972.7274   141.8812   6.856     8.40e-09***
PRECIP           2.0584     0.6702   3.071     0.00339**
EDUC           -16.4570     7.5734  -2.173     0.03436*
NONWHITE        48.7069    11.0203   4.420     5.05e-05***
POOR           -18.3358    34.1909  -0.536     0.59405
nox            13.6910     7.2122    1.898     0.06321 .
nox2           -5.2629     2.5874   -2.034     0.04706 *
SO2             5.8666      5.5694   1.053     0.29705
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*'  0.05 '.'  0.1 ' ' 1

Residual standard error: 35.02 on 52 degrees of freedom
Multiple R-squared:  0.7207, Adjusted R-squared:  0.6831
F-statistic: 19.17 on 7 and 52 DF,  p-value: 2.207e-12
```

The fitted regression is

$$\hat{Y} = (972.728) + (2.058)x_1 + (-16.457)x_2 + (48.707)x_3 + (-18.336)x_4 + (13.691)x_5 + (-5.263)x_5^2 + (5.867)x_6 \quad \text{where } x_i = X_i - \text{mean}(X_i)\text{, and } i = 1, 2, \ldots, 6.$$

Significant predictors are [PRECIP], [EDUC], [NONWHITE], and [NOX]. Also note that this model shows the median of the residuals is the furthest from zero of all the models, which may be a sign that this fit has the most error and we should consider another model.

(4) Squared Terms of [EDUC], [NONWHITE], and [NOX]

```
Residuals:
    Min      1Q  Median      3Q     Max
-66.294 -17.029   0.238  13.569  54.653

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    662.2865  101.9451   6.497 8.51e-08 ***
PRECIP           1.8895    0.6452   2.928 0.005539 **
educ            -5.4026    6.8683  -0.787 0.436035
educ2          -17.9162    5.2990  -3.381 0.001597 **
nonwhite        43.9298   10.2384   4.291 0.000106 ***
nonwhite2       20.7940   14.7850   1.406 0.167132
POOR            67.5711   39.2712   1.721 0.092858 .
nox             4.7019    7.5857   0.620 0.538793
nox2            -3.9852    2.3395  -1.703 0.096050 .
SO2             19.0711    5.8233   3.275 0.002153 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '  1

Residual standard error: 27.03 on 41 degrees of freedom
  (9 observations deleted due to missingness)
Multiple R-squared:  0.8191, Adjusted R-squared:  0.7794
F-statistic: 20.63 on 9 and 41 DF,  p-value: 1.489e-12
```

The fitted regression function is

$$\hat{Y} = (662.287) + (1.89)x_1 + (-5.403)x_2 + (-17.912)x_2^2 + (43.93)x_3 + (20.79)x_3^2 + (67.571)x_4 + (4.702)x_5 + (-3.99)x_5^2 + (19.071)x_6$$

where $x_i = X_i - \text{mean}(X_i)$, and $i = 1, 2, \ldots, 6$.

Significant predictors are [PRECIP], [EDUC], [NONWHITE], [NOX], and $[\text{NOX}]^2$. In all of the summaries, the intercept is significant and is interpreted as the expected value of mortality rate when all other factors are 0.

Of the four models, (2) and (4) produce the smallest p-value associated with their respective large F-statistics. The F-statistic is a good indicator of whether there is a relationship between the variables and the model is a good fit for the data set. Additionally, those two models yielded the highest R_{adj}^2 values, which is a good measure of how much variability the model explains. Thus, only models (2) and (4) will be considered in the next section.

RESIDUAL ANALYSIS

Histograms and boxplots are useful for examining whether the error terms are reasonably normally distributed or identifying any outliers. Referring to the two histograms (Figure 5, 6) of the residuals of the two different models (one with linear terms and one with quadratic terms), the residuals of the polynomial model seem to be more normally distributed, which is preferred.

Figure 5: Residual of Linear Regression

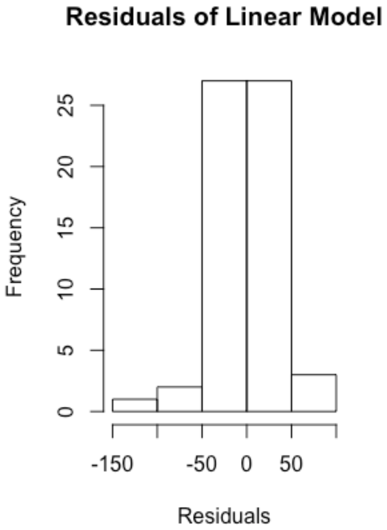
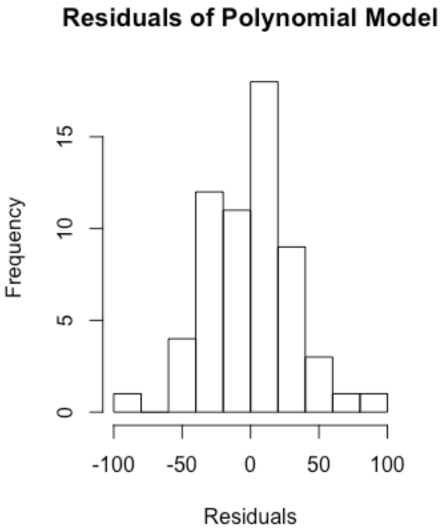


Figure 6: Histogram of Residuals in Polynomial Model



The boxplots on the right show the outliers of the residuals (Figures 7, 8). Both plots show outliers, but the outliers associated with the polynomial model are closer to 0, which is better because we want errors to be close to 0.

Figure 7: Boxplot of Residuals in Linear Regression

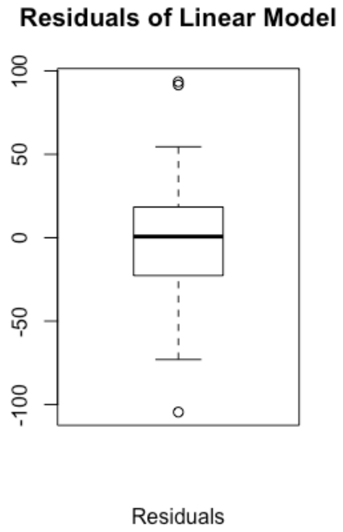
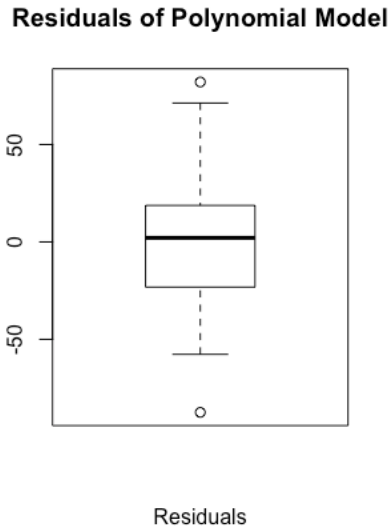


Figure 8: Boxplot of Residuals in Polynomial Model



A quantile-quantile plot shows that the residuals in the polynomial model are slightly more "normal" and are closer to 0 than the residuals in the regression with all linear terms.

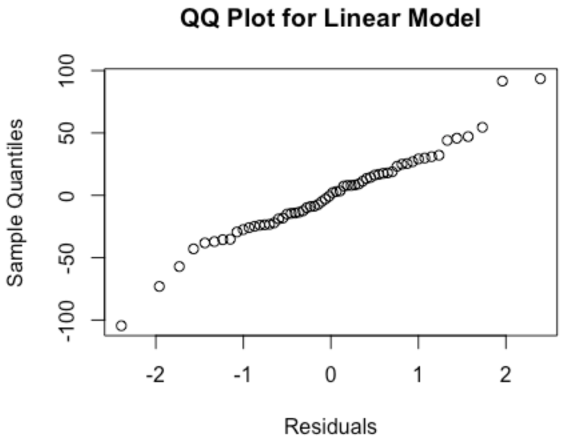


Figure 9: Q-Q Plot of Linear Model

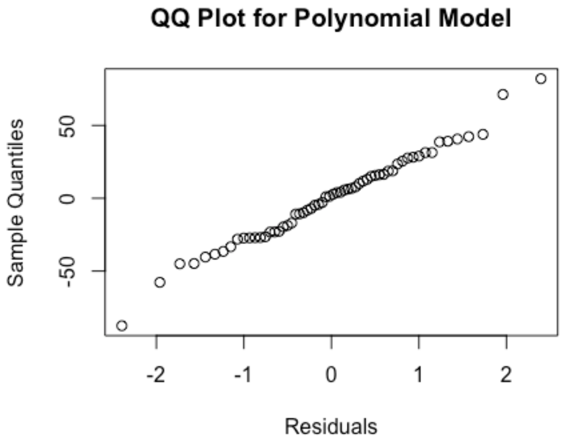
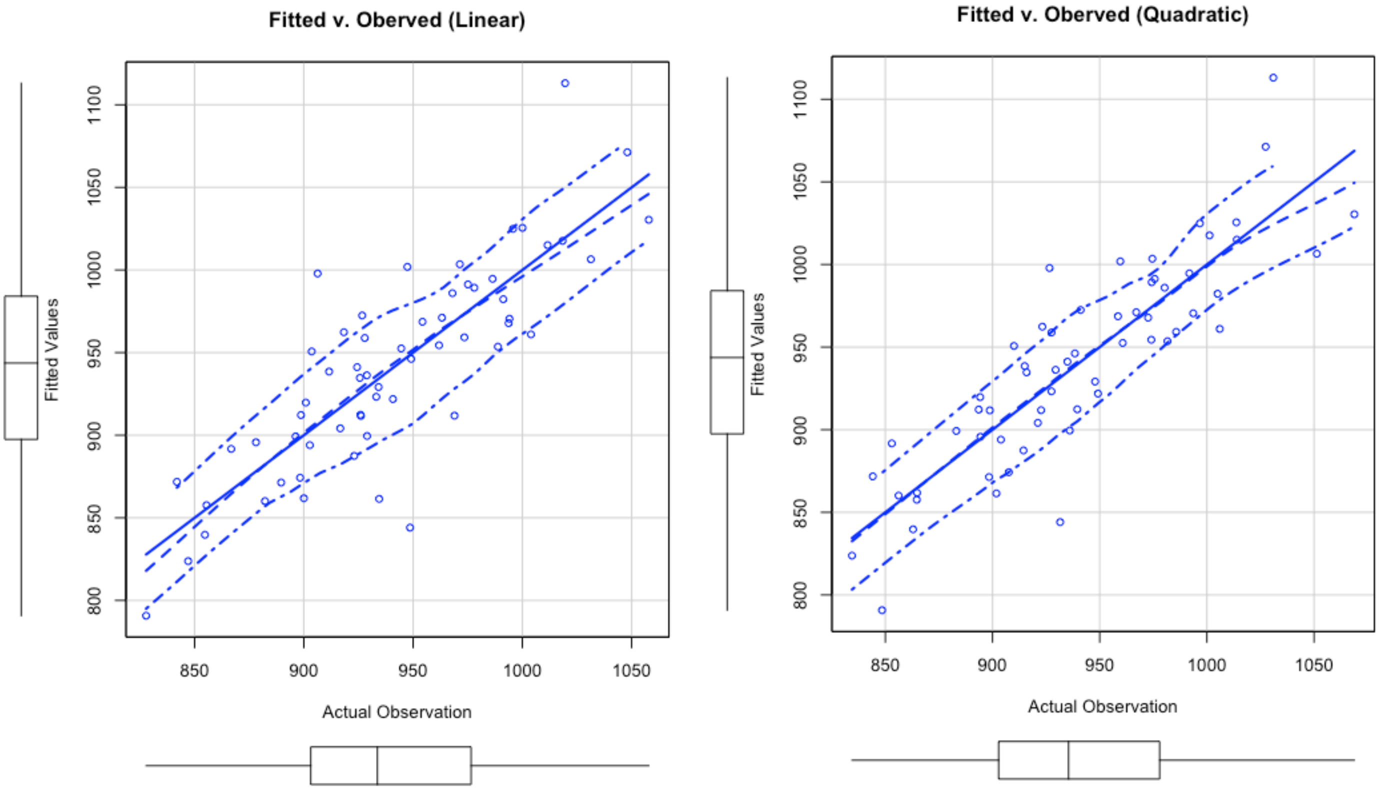


Figure 10: Q-Q Plot of Quadratic Model

A plot of the residuals against the fitted values is useful for assessing the appropriateness of the multiple regression function and the constancy of the variance of the error terms. Below is the fitted value for the two potential models chosen to be best fits. Even though we do not know which predictors cause the variability, it is obvious that Figure 11 has a smaller standard deviation and shows to be a better fit for the data. Moving forward, I will be conducting analysis only on the quadratic fitted regression model.

Figure 10-11: Fitted Values Against Observed Values of Mortality (Linear and Quadratic)



ANALYSIS OF VARIANCE (ANOVA)

```
Analysis of Variance Table

Response: MORTALITY
                  Df    Sum Sq      Mean Sq     F value     Pr(>F)
PRECIP            1      59256      59256       52.7021      2.357e-09 ***
educ              1      20492      20492       18.2261      8.747e-05 ***
educ2             1      7272       7272        6.4677       0.014126 *
nonwhite          1      48956      48956       43.5412      2.549e-08 ***
nonwhite2         1      2371       2371        2.1089       0.152691
POOR              1      10286      10286       9.1481       0.003924 **
nox               1      17593      17593       15.6473      0.000241 ***
nox2              1      5450       5450        4.8474       0.032330 *
SO2               1      380        380         0.3379       0.563677
Residuals         50     56218      1124
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '  1
```

The regression sum of squares is the square of the difference of the estimated value from the mean. A lower SSR does not imply that the fitted line yields a smaller difference from the mean. Rather, the p-value in the last column of the anova table relating to the significance levels is a better indicator. [PRECIP] has a significance level of close to 0, which implies that predictor has a significant influence on mortality. Highlighted above is the predictors that have a significant influence of at least 0.001 on mortality rates. This also provides insight on potential predictors that will best explain mortality, but we will have to perform more analysis to see how each predictor explains mortality within the model.

MODEL SELECTION AND DIAGNOSTICS

REMOVAL OF PREDICTORS

To confirm my analysis on significant predictors, using a stepwise multiple regression, I used the function `step(),` which takes the model including all of the predictors, then analyzes the model with all the predictors and eliminates insignificant predictor variables. If all of the variables were significant, then the final model would be the equation we fitted using the summary, but that was not the case.

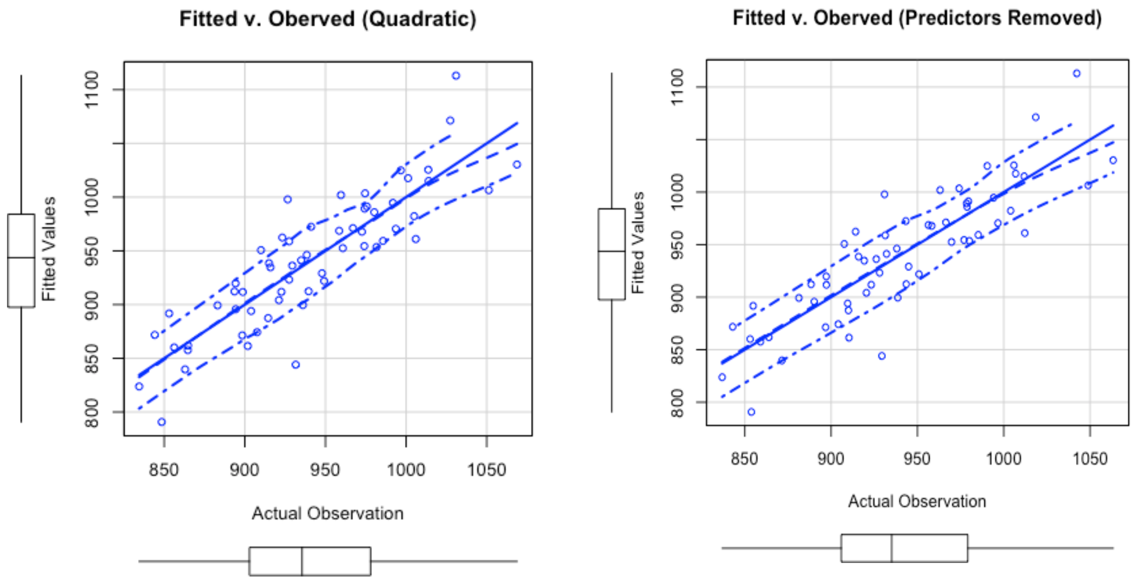
The "best" subset suggested our final model to be:

$$Y =(890.620) + (1.876)x_1 + (-21.621)x_2 + (-16.76)x_2^2 + (39.319)x_3 + (22.712)x_5 + (-6.233)x_5^2 +E$$

where $x_i = X_i - \text{mean}(X_i)$, and i = 1, 2, …, 6 and E is the error term.
The significant predictors shown in the model is [PRECIP], [EDUC], $[\text{EDUC}]^2$, [NONWHITE], [NOX], and $[\text{NOX}]^2$.

A side by side comparison of the model including all the predictors and with excluded predictors are very similar, but the graph on the right shows a much smoother regression.

Figure 12-11: Fitted Values Against Observed Values of Mortality (Predictor Influence)



The new fitted model after insignificant predictors were removed yielded a smaller p-value and larger F-statistic than that of the old model (3.116e-14 < 2.148e-12 and 26.2 < 17).

CONCLUSION

From the statistical analyses it can be concluded that the best model for determining mortality per 100,000 persons in a SMSA with the given factors of mean precipitation (in inches), median number of school years completed by person of age 25 or older, percentage of population in 1960 that is non-white, percentage of households with annual income under $3000 in 1960, and relative pollution of oxides of nitrogen, and relative pollution potential of Sulphur dioxide, is a polynomial regression model with the predictors of precipitation, education, non-white, and levels of nitrogen oxide. Even though correlation does not imply causation, with the given data, we can conclude that pollution has only a slight effect on mortality, but in addition to factors such as precipitation, education, and non-white.

With the absence of interaction terms, there could be confounding effects that were not accounted for in our final model. Additionally, there could be omitted predictors. For example, crime rate could have an immense impact on mortality. Potential factors that could improve our model would be increasing the sample size and collecting more modern data since the data collected was in 1960. This will allow us to produce a more accurate model for determining the effects of pollution on mortality. In regard to the step function used for predictor removal, the function stops once it finds a “good” model. There may be a better subset of the model that is not accounted for.

CODE APPENDIX

```
knitr::opts_chunk$set(echo = FALSE)
#Read data
dat_mortality = read.csv('mortality.csv', header = TRUE)
dat_m = dat_mortality[,c(7,1,2,3,4,5,6)]
#Transform NO, SO using  natural log
dat_m$NOX = log(dat_mortality$NOX)
dat_m$SO2 = log(dat_mortality$SO2)

#Transform NONWHITE, POOR using cubic root
dat_m$NONWHITE = (dat_mortality$NONWHITE)^(1/3)
dat_m$POOR = (dat_mortality$POOR)^(1/3)
#Visual representation of outliers
boxplot(dat_mortality[,1:6], main = "Boxplot of Factors")
boxplot(dat_m[,4:7], main = "Boxplot of Factors (Transformed)")
#reprint data to remove outliers
dat_mort = dat_m

#remove outliers
source("http://goo.gl/UUyEzD")
outlierKD(dat_mort, POOR)
library(car)
#Scatterplot Matrix of Outliers and Outliers Removed
scatterplotMatrix(dat_m, main = 'Scatterplot Matrix (With Outliers)', cex = 0
.01 )
scatterplotMatrix(dat_mort, main = 'Scatterplot Matrix (Outliers Removed)', c
ex = 0.01)

boxplot(dat_mort[,3:7], main = "Boxplot of Factors (Outliers Removed)")
#covariance matrix of data w outliers
cor(dat_m)
#[EDUC] quadratic fit
educ = dat_m$EDUC - mean(dat_m$EDUC)
educ2 = educ^2

fit_educ2 = lm(dat_m$MORTALITY ~ educ + educ2)
fit_educ = lm(dat_m$MORTALITY ~ dat_m$EDUC)

ed = qqnorm(fit_educ$res)
ed2 = qqnorm(fit_educ2$res)
plot(ed, xlab = "Theoretical Quantiles",
     ylab = "Standardized Residuals",
     main = "Normal QQ (Linear [EDUC])")
plot(ed2, xlab = "Theoretical Quantiles",
     ylab = "Standardized Residuals",
     main = "Normal QQ (Quadratic [EDUC])")
```

```
#[NONWHITE] quadratic fit
nonwhite = dat_m$NONWHITE - mean(dat_m$NONWHITE)
nonwhite2 = nonwhite^2

fit_nonwhite = lm(dat_m$MORTALITY ~ dat_m$NONWHITE)
fit_nonwhite2 = lm(dat_m$MORTALITY ~ nonwhite + nonwhite2)

non = qqnorm(fit_nonwhite$res)
non2 = qqnorm(fit_nonwhite2$res)
plot(non, xlab = "Theoretical Quantiles",
     ylab = "Standardized Residuals",
     main = "Normal QQ (Linear [NONWHITE])")
plot(non2, xlab = "Theoretical Quantiles",
     ylab = "Standardized Residuals",
     main = "Normal QQ (Quadratic [NONWHITE])")


#[NOX] quadratic fit
nox = dat_m$NOX - mean(dat_m$NOX)
nox2 = nox^2

fit_nox = lm(dat_m$MORTALITY ~ dat_mort$NOX)
fit_nox2 = lm(dat_m$MORTALITY ~ nox + nox2)

n = qqnorm(fit_nox$res)
n2 = qqnorm(fit_nox2$res)
plot(n, xlab = "Theoretical Quantiles",
     ylab = "Standardized Residuals",
     main = "Normal QQ (Linear [NOX])")
plot(n2, xlab = "Theoretical Quantiles",
     ylab = "Standardized Residuals",
     main = "Normal QQ (Quadratic [NOX])")
#model for data without outliers
full_l = lm(MORTALITY ~ ., data = dat_mort)

#model for data with outliers
full_out = lm(MORTALITY ~ ., data = dat_m)

#model for data with NOX^2
full_q2 = lm(formula = dat_m$MORTALITY ~
               dat_m$PRECIP + dat_m$EDUC + dat_m$NONWHITE + dat_m$POOR + (nox
+ nox2) + dat_m$SO2)


#model for data with three square terms
full_q3 = lm(formula = dat_m$MORTALITY ~
               dat_m$PRECIP + (educ + educ2) +
               (nonwhite + nonwhite2) +
               dat_m$POOR + (nox + nox2) + dat_m$SO2)
```

```
capture.output(summary(full_l), file = "summary_l.doc")
capture.output(summary(full_out), file = "summary_out.doc")
capture.output(summary(full_q2), file = "summary_q2.doc")
capture.output(summary(full_q3), file = "summary_q3.doc")
#Scatternplot of fitted values of linear and polynomial model
scatterplot(full_out$fitted.values, dat_m$MORTALITY,
            xlab = "Actual Observation", ylab = "Fitted Values",
            main = "Fitted v. Oberved (Linear)")

scatterplot(full_q3$fitted.values, dat_m$MORTALITY,
            xlab = "Actual Observation", ylab = "Fitted Values",
            main = "Fitted v. Oberved (Quadratic)")
#analysis of residuals for linear and polynomial model
boxplot(full_out$res, xlab = "Residuals",
        main = "Residuals of Linear Model")
hist(full_out$res, xlab = "Residuals",
     main = "Residuals of Linear Model")

boxplot(full_q3$res, xlab = "Residuals",
        main = "Residuals of Polynomial Model")
hist(full_q3$res, xlab = "Residuals",
     main = "Residuals of Polynomial Model")

qqnorm(full_out$res, xlab = "Residuals",
       main = "QQ Plot for Linear Model")

qqnorm(full_q3$res, xlab = "Residuals",
       main = "QQ Plot for Polynomial Model")
#Anova table for quadratic model
anova(full_q3)
capture.output(anova(full_q3), file = "anova_q3.doc")
#Removing predictors with stepwise
library('leaps')
library('MASS')
new_full_q3 = step(full_q3,
     scope = ~ dat_m$PRECIP + educ + educ2 +
       nonwhite + nonwhite2 + dat_m$POOR +
       nox + nox2 + dat_m$SO2,
     direction = 'backward')
capture.output(new_full_q3, file = 'step.doc')
#improved scatterplot matrix of fitted values and observed
scatterplot(new_full_q3$fitted.values, dat_m$MORTALITY,
            xlab = "Actual Observation", ylab = "Fitted Values",
            main = "Fitted v. Oberved (Predictors Removed)")
#reviewing models
summary(full_q3)
summary(new_full_q3)
```