Seoul Air Pollution 2017 Codebook

MinJung Kang

2018-12-12

### 1. Overview of Dataset

A <- read.csv("KMJ037\_(SeoulAir)c.csv")

* **Contents**: Air pollutants measured every hour for one week (2017.11.17 23:00 to 2017.11.24.23:00) in 25 districts of Seoul
* **Source**: Kaggle
* **URL**: <https://www.kaggle.com/jihyeseo/seoulairreport>

head(A)

## Date Time District NO2 O3 CO SO2 PM10 PM2.5 Y  
## 1 2017-11-24 23 Gangnam 0.038 0.004 0.4 0.005 16 10 AQG  
## 2 2017-11-24 22 Gangnam 0.031 0.008 0.4 0.005 17 9 AQG  
## 3 2017-11-24 21 Gangnam 0.025 0.012 0.4 0.005 18 11 AQG  
## 4 2017-11-24 20 Gangnam 0.033 0.007 0.4 0.005 21 12 AQG  
## 5 2017-11-24 19 Gangnam 0.033 0.008 0.4 0.005 20 10 AQG  
## 6 2017-11-24 18 Gangnam 0.026 0.011 0.4 0.005 21 10 AQG

* **Dimensions**

dim(A)

## [1] 4095 10

* **Column Names**

names(A)

## [1] "Date" "Time" "District" "NO2" "O3" "CO"   
## [7] "SO2" "PM10" "PM2.5" "Y"

* **Data Structure**

str(A)

## 'data.frame': 4095 obs. of 10 variables:  
## $ Date : Factor w/ 7 levels "2017-11-18","2017-11-19",..: 7 7 7 7 7 7 7 7 7 7 ...  
## $ Time : int 23 22 21 20 19 18 17 16 15 14 ...  
## $ District: Factor w/ 25 levels "Dobong","Dongdaemun",..: 7 7 7 7 7 7 7 7 7 7 ...  
## $ NO2 : num 0.038 0.031 0.025 0.033 0.033 0.026 0.021 0.017 0.015 0.015 ...  
## $ O3 : num 0.004 0.008 0.012 0.007 0.008 0.011 0.015 0.018 0.018 0.018 ...  
## $ CO : num 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.3 0.3 0.3 ...  
## $ SO2 : num 0.005 0.005 0.005 0.005 0.005 0.005 0.005 0.005 0.004 0.005 ...  
## $ PM10 : int 16 17 18 21 20 21 21 19 21 21 ...  
## $ PM2.5 : int 10 9 11 12 10 10 13 11 10 10 ...  
## $ Y : Factor w/ 5 levels "AQG","IT1","IT2",..: 1 1 1 1 1 1 1 1 1 1 ...

### 2. Description of variables

* **Codebook**

|  |  |  |
| --- | --- | --- |
| Variable | Description | Value |
| Date | Date of Measurement | YYYY-MM-DD |
| Time | Time of Measurement | HHMM |
| District | District Name of Measurement Station | Gangnam |
| NO2 | Nitrogen Dioxide (ppm) | 0.038 |
| O3 | Ozone (ppm) | 0.004 |
| CO | Carbon Monoxide (ppm) | 0.4 |
| SO2 | Sulfur Dioxide (ppm) | 0.005 |
| PM10 | PM10 Fine Dust (microgram/m3) | 16 |
| PM2.5 | PM2.5 Fine Dust (microgram/m3) | 10 |
| Y | PM2.5 Classification | AQG |

### 3. Subject of Analysis

* **Target Predictive Data**: “PM2.5” based on “Time” and “District”
* **Classification**: Based on WHO Air Quality Guidelines (for 24hour mean of PM2.5)

|  |  |  |
| --- | --- | --- |
| PM2.5 | Class | Description |
| 25 | AQG | Threshold of cardiopulmonary and lung cancer mortality increase in response to long-term exposure to PM2.5 |
| 37.5 | IT3 | Mortality risk 3% higher than AQG |
| 50 | IT2 | Mortality risk 9% higher than AQG |
| 75 | IT1 | Mortality risk 15% higher than AQG |
| 75+ | Off.T | Extremely high mortality risk |