STAT 420: Data Analysis Project (lumav2, Panda5)

Team Name

|  |  |
| --- | --- |
| Name | Net ID |
| Luma Vasiljevic | lumav2 |
| Sushanta Panda | Panda5 |

Title of the Project:

Forced Expiratory Volume (FEV) prediction for Childhood Respiratory Disease

Description of the dataset:

Forced Expiratory Volume (FEV) is an index of pulmonary function that measures the volume of the air expelled after one second of constant effort. The data contains the determinations of FEB on 654 children ages 6 – 22 who are seen in childhood respiratory disease study in 1980 in East Boston, Massachusetts. The data are part of a larger study to follow the change in pulmonary function over time in children.

Dataset available: <http://www.statsci.org/data/general/fev.html>

Below are the all variables in the dataset

|  |  |  |
| --- | --- | --- |
| Variable Name | Category | Description |
| ID | Numeric | Uniquely Identified Row |
| Age | Numeric | Age of the child |
| Height | Numeric | Height of the child in inches |
| Sex | Categorical | Sex of the child  Possible Values : Male / Female |
| Smoker | Categorical | Whether the child is a non-smoker or current smoker  Possible Values : Non / Current  Non = Non-Smoker  Current = Current Smoker |
| FEV | Numeric | FEV of the child in litres  **Response Variable** |

Background information of the dataset:

The data contains the determinations of FEB on 654 children ages 6 – 22 who are seen in childhood respiratory disease study in 1980 in East Boston, Massachusetts. The data are part of a larger study to follow the change in pulmonary function over time in children

*Note: No citation required for this source (*[*http://www.statsci.org/data/general/fev.html*](http://www.statsci.org/data/general/fev.html)*)*

Interest in dataset hope to explore:

This dataset has chosen by us because of personal interests, whether we can predict the child’s FEV with the help of the available predictor, rather going for a Pulmonary Function Test, which will identify any pulmonary disease of a child. And secondly, to do statistical analysis on what are the predictors responsible to increase / decrease the pulmonary function of the child and try to find answer on these lines.

Evidence that the data can be loaded into R:

Below is the evidence the data is loaded from the URL (<http://www.statsci.org/data/general/fev.txt>) into the R data frame, with data types and number of rows loaded into it

A screenshot of a social media post

Description automatically generated