Featuring the Use of Tobacco and E-cigarettes and Relevant Diseases

Abstract

- Most health care providers recommend electronic cigarette as a substitute to cigarette. This study investigates the potential advantage of e-cigarettes over cigarettes by analyzing the correlations between the use of cigarette and e-cigarette with related cancers, stroke, heart disease, and high blood pressure.

Abstract

- We analyze the 2017 BRFSS data from Centers for Disease Control and Prevention (CDC) using a logistic regression model. Our result shows that the use of cigarettes has strong correlations with cancers, stroke, and heart disease. Its correlation with high blood pressure is not pronounced. The use of e-cigarette does not show a statistically significant correlation with the above diseases.

Data Preparation

- We randomly select 22379 individual sample from the dataset and extract a subset of variables indicating the frequency an individual smokes cigarettes or e-cigarettes, and their health record relevant to cancers, stroke, heart diseases, and high blood pressure.
- We discarded the rows with multiple blank entries. We fill in the missing data in the remaining dataset with random values generated from the same distributions.

Logistic Regression Model

- We ran a series of logistic regression models trying to predict heart attack, stroke, cancers related to smoking, and coronary heart disease.
- By doing this, we split our data into training and test data (80:20 ratio), ran our regression model on our training data, and tested with our test data.

Result

- We found statistically significant relationships in the logistic regression model for smoking-related cancers with age, sex, and the use of cigarettes and e-cigarette. The misclassification errors in this model is high, due to the sparsity of the relevant data.
- For stroke, heart attack, and other heart diseases, we found their significant relationships with age, sex, and the use of cigarettes, but their relationships with the use of e-cigarettes is not significant. The misclassification errors are 24.6%, 24.1%, and 26.3%.

Next Step

- The use of electronic cigarettes became prevalent in the U.S. in 2010. Seven years is too short for some negative effects to be showed. A longer term observation to e-cigarettes smokers may reveal the hidden health risks that are not pronounced in this study.