

Proposal_11/11

Ze Li, Qianying Wu, Zheshu Jiang, Yimeng Cai

11/11/2023

Group Member:

Yimeng Cai (yc3577), Zheshu Jiang (zj2379), Ze Li (zl2746), Qianying Wu (qw2418)

Project Title:

Multiple Linear Regression Analysis on Socio-economic factors and Bioindicators of Type II Diabetes by using R

Project Motivation

Type II Diabetes is a global health concern that affects more than 37 million Americans and places a significant burden on healthcare systems. As a multifaceted disease, Type II diabetes was influenced by both socio-economic factors and biological indicators. By analyzing a comprehensive dataset from the Behavioral Risk Factor Surveillance System, we aim to uncover the correlations between diabetes prevalence and factors such as sex, education, income, lifestyle habits, and bioindicators like blood glucose and cholesterol levels. Through visualizations and statistical analysis in R, we aspire to highlight patterns that could inform better prevention and management strategies, thereby contributing to the broader dialogue on public health and socio-economic policies related to diabetes.

Data Resources:

<https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset/data>

- Data Description:

The dataset was collected by the Behavioral Risk Factor Surveillance System (BRFSS) in 2015. This original dataset contains responses from 441,455 individuals and has 330 features. For this project, we use the one with 253,680 responses and a three-class diabetes target variable indicating no diabetes, prediabetes, or diabetes. The other potential factors for diabetes are blood pressure, cholesterol, cholesterol check, BMI, stroke, heart disease, and habits of smoking, physical exercise, eating fruit.

Intended Final Products

- Demographics Visualization:

From scratch, to understand the basic demographics information around type II diabetes, we decide to perform data visualization of type II diabetes occurrences among population in different biological sex, educational level, and income level.

- Socio-economic Factors Analysis:

In order to study socio - economic factors related to type II diabetes occurrences, we aim to distinguish the following three factors - smoking status, alcohol consumption, and body mass index (BMI). By implementing data analysis, statistical tests, along with data visualizations, we are able to study the correlation between each single factor and frequencies of type II diabetes.

- Symptoms Association:

With the understanding of factors correlated with the disease, we plan to study the physiological symptoms associated with type II diabetes in order to decide which biological symptom is able to perform as a significant bioindicator or biomarker through statistical tests.

Planned Timeline

Timeline	Requirement	Predicted Timeline
11/13-17	Project review meeting	Nov 15th or 16th
12/9	Report	Nov 24
12/9	webpage and screencast	Dec 1-7
12/9	peer assessment	Dec 7-8
12/14	In class discussion	