ISE790 Final Project Write-up

Yuan Zhang, Thomas Richardson
Operations Research
North Carolina State University

June 11, 2010

Introduction

There are about 23.6 million children and adults in the United States 7.8% of the population living with diabetes, and 1.6 million new cases of diabetes are diagnosed in people aged 20 years and older each year [**?**]. Complications of diabetes including heart disease, stroke, blindness, kidney disease and etc. Overall, the risk for death among people with diabetes is about twice that of people without diabetes of similar age. [**?**].

The U.S. is spending $174 billion for patients with diabetes, combining $116 billion for direct medical costs and $58 billion for indirect costs such as disability, work loss, and premature mortality [**?**]. Today diabetes prevention and early treatment are under-emphasized, if adults received their recommended diabetes screenings and early lifestyle intervention of medicine treatment, complications and disability could be avoided and billions of dollars can be saved [**?**].

The United States Health Expense Think Tank (USHETH) is a fictional group that wants to know the impact of preventive measures on diabetes. Furthermore, it would like to know how an increase or decrease in diabetes prevalence would affect the amount of money spent on health care. Specifically, USHETH wants to research on the following two questions:

1. USHETH wants to know diabetes influences a patient total health care expense. There are many estimates of the cost of diabetes care that have focused on diabetes-specific expenses, but it is known that diabetes might affect total health care expenses in many ways. There are many factors in the patient database that need to be taken into account so we propose to build a model to know the difference in health care costs for a person with diabetes from the costs for a person without the disease. It acknowledges that there are many factors that need to be taken into account.

2. USHETH needs to know how many people have diabetes by age and perhaps gender. This will be critical for the analysis. We know that BMI is established as one of the significant risk factors for diabetes [**?**], and studies show that lifestyle intervention can reduce the incidence of diabetes in persons at high risk [**?**]. USHETH wants to measure the impact of diabetes preventive measure. The measure focuses on people who have a BMI (body mass index) larger than 25 in adults. In children the threshold BMI varies by age linearly: at 5 years old, the threshold BMI is 17 and at 20 years old, the threshold BMI is the same for adults. These people will be enrolled in special programs to reduce their BMI by ten percent. USHETH then wants to know how many people would contract diabetes.

Proposed Approach

sort out important / most relevant parameters from list of 44

A large study was done in the United States to collect information on individuals, their health picture and how much money is spent on their behalf for health care. The sample provided is representative of the population and represents a snapshot of the country and its health care costs at a point in time.

In determining how effective the proposed measure will be, we must analyze the onset of diabetes in the stated BMI categories, and determine relevant correlations. Most open, as we also need to investigate how well other parameters predict onset of diabetes. This is difficult since

the data provides only a snapshot, and not a time series of data points. Therefore, we have little knowledge of how these parameters changed for individuals over time, though we can extrapolate the onset of particular things (diabetes, heart conditions, BMI changes) by trends in the data.

First, we propose to create a clustering algorithm that could accurately place data points into the correct diabetes category based on the parameters provided, and one with a more continuous range of placement tags, such as “very likely”,“likely”, “not very likely”, etc. The parameters from these models could likely be used to assess which parameters or sets of parameters are most correlated to an individual having diabetes. Then we propose to build a neural network as a underlying function between the various parameters (inputs) and the incidence of diabetes (output).

Implementation

Results

Performance Analysis

Conclusion and Discussion