Type 2 Diabetes Prediction

Authors: Andre Niacaris, Ashina Jayan, Nicolas Thomas
University of Central Florida

Abstract

This research looks at the independent role of Body Mass Index (BMI) in predicting the likelihood and severity of Type 2 diabetes. We hypothesize that higher BMI is significantly associated with an increased risk and severity of Type 2 diabetes. Using a comprehensive medical dataset, we applied Exploratory Data Analysis (EDA) to visualize patterns and relationships between predictors. Our research aims to clarify the impact of body mass indexes in relation to Type 2 diabetes.

Introduction

More than 38 million Americans have diabetes, and about 90% to 95% have Type 2 diabetes. Type 2 diabetes is caused by multi-organ insulin resistance, and a decline in beta-cell insulin secretory function. To put it simply, the pancreas does not produce enough insulin to regulate the glucose (sugar) levels, and the cells respond poorly to insulin (resistance) so the cells don't take enough sugar. While BMI has been recognized as a key risk factor, there are some studies that suggest that other factors such as ethnicity and family history are more significant. We can understand and compare the predictive power of other key variables to BMI.

Data

Our data is from Kaggle. It is compiled from various reliable sources such as clinical studies, public health databases, and hospital records. The dataset has 13 classifications of diabetes as the main target variable. For our research, we are focusing on the Type 2 diabetes classification. There are 34 variables in total: one dependent variable (target) and 33 independent variables used for predicting the target variable. The independent variables are a mix of categorical and numerical data. We created data visualizations and conducted statistical analysis in R Studio. Specifically we applied logistic regression, chi-square test, odds ratio, logistic regression with interaction terms, and roc curve for model assessment.

Research Question and Hypothesis

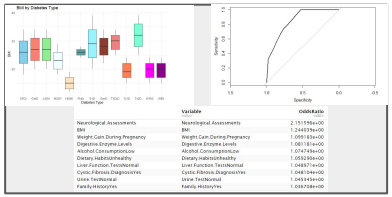
Research Question:
To what extent does BMI independently predict the likelihood and severity of Type 2 diabetes, and is a higher BMI associated with an increased risk of developing Type 2 diabetes compared to other factors?

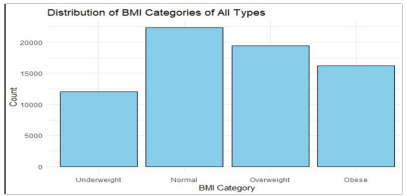
Hypothesis:
We hypothesize that a higher BMI independently predicts a greater likelihood and severity of Type 2 diabetes compared to other risk factors.

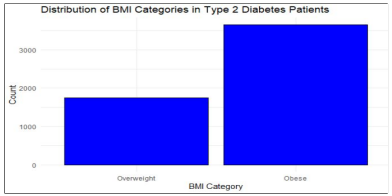
Analysis

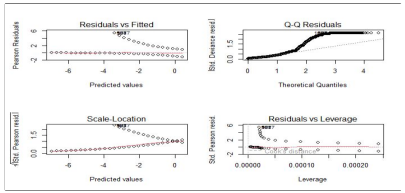
23.2% of the overall population is obese, and 0.0771 of the population have type 2 diabetes. Out of people with Type 2 Diabetes, 32.4% are not obese and 67.6% are obese.

After setting Type 2 Diabetes to 1 and all other Diabetes types to 0, we applied logistic regression on Target to BMI, finding an intercept of -10.206 and BMI had an estimate of 0.272. The odds ratio is 1.313, so every one-unit increase in BMI equates to the likelihood of the Target being Type 2 increasing by 31.3% The Odds Ratio for BMI was also the second highest, displaying how valuable of a predictor it is. The low p-value also indicated high significance between BMI and the Target.









Conclusion

The study confirms that BMI is a significant independent predictor of Type 2 diabetes. BMI is a critical predictor, but it's predictive power is surpassed by one other factor at least. The high prevalence of obesity (67.6%) among those with Type 2 diabetes highlights the importance of preventative strategies like weight loss. Future research should incorporate additional predictors along with BMI, as well as more advanced modeling techniques to assist in the improvement and comprehension of more preventative efforts.

Limitations

The dataset used is limited to individuals who are either prediabetic or diabetic. This prevents analysis of BMI's predictive abilities as it relates to non-diabetic populations. Including non-diabetic samples would enable a more comprehensive understanding of BMI's role across more population spectrums..

With there only being 33 predictors, there is a high likelihood that key confounding variables, such as activity level and genetics, are not accounted for. These unmeasured factors may better help to explain the relationship between Type 2 diabetes and BMI.

References

Batra, A. (2024, September 1). *Diabetes dataset*. Kaggle. https://www.kaggle.com/datasets/ankitbatra1210/diabetes-dataset/data

Centers for Disease Control and Prevention. (n.d.). Type 2 diabetes. Centers for Disease Control and Prevention. https://www.cdc.gov/diabetes/about/about-type-2-diabetes.html#:~:text=More%20than%2038%20million%20Americans,adults%20are%20also%20developing%20it.

Mayo Foundation for Medical Education and Research. (2023, March 14). Type 2 diabetes. Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/type-2-diabetes/symptoms-causes/syc-20351193

Klein , S. (2022, January 4). Why does obesity cause diabetes?. Cell Metabolism. https://www.sciencedirect.com/science/article/pii/S1550413121006318