ST. XAVIER'S COLLEGE (AUTONOMOUS), KOLKATA

DEPARTMENT OF STATISTICS

A STATISTICAL STUDY ON DIABETES USING LOGISTIC REGRESSION

B.SC. STATISTICS (HONOURS)

NAME: SOHINI MANDAL

ROLL NUMBER: 0481

SESSION: 2021-2024

SEMESTER: VI

NAME OF THE SUPERVISOR: PROF. MADHURA DASGUPTA

SYNOPSIS

Diabetes mellitus is a prevalent chronic metabolic disorder characterized by elevated blood sugar levels, posing significant health challenges worldwide. It occurs when the pancreas does not produce enough insulin or when the body is not able to effectively use the insulin that is produced. Timely identification of individuals at risk is essential for effective management and prevention of complications like cardiovascular diseases, kidney failure and so on.

This study employs Binomial Logistic Regression analysis to explore the relationship between various risk factors and the likelihood of diabetes onset, aiming to develop a predictive model for accurate detection. Before proceeding with regression, some descriptive studies are done, followed by some categorical data analysis, and checking for multicollinearity among the predictors. After fitting the regression model, a list of statistically significant predictor variables which influence the response is obtained. Confusion matrix is made to check the accuracy of this model.

Here, the dataset consists of several medical independent predictors which are continuous variables except one which is a categorical variable and one binary target variable, response. This dataset is taken from the "National Institute Of Diabetes And Kidney Diseases" repository. 768 patients were tested. Predictors include the number of pregnancies the patient has had, their BMI, insulin level, age, glucose level, diastolic blood pressure, skin thickness and diabetes pedigree function.

The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.