Laurence Kim
05/28/23
ELEN 520: Radhika Grover

Final Project Proposal

Topic:
Using a Multi Linear Regression and a logistic regression model to predict diabetes using various predictors such as pregnancies, glucose, blood pressure, skin thickness, insulin, bmi, diabetes pedigree function, and age.

Introduction:
Diabetes happens to run in both sides of my family and it has always been in the back of my mind. I always felt like I was eventually going to get it and it was just a matter of time. However, maybe just because people in my family tree have had it will not necessarily predict whether I or any of my other family members will. I use a multi linear regression model using different predictors to predict the response and find the correlation between predictors and the outcome.

Hypothesis Testing:
Response:
Diabetes outcome: To express the final result 1 is Yes and 0 is No

Predictors:

Pregnancies: To express the Number of pregnancies

Glucose: To express the Glucose level in blood

BloodPressure: To express the Blood pressure measurement

SkinThickness: To express the thickness of the skin

Insulin: To express the Insulin level in blood

BMI: To express the Body mass index

DiabetesPedigreeFunction: To express the Diabetes percentage

Age: To express the age

Null Hypothesis:
There is no relationship between the response variable and the 8 predictors.

Alternative Hypothesis:
There is a relationship between the response variable and the 8 predictors.

Dataset:
Dataset used is from Kaggle Website
https://www.kaggle.com/datasets/mathchi/diabetes-data-set