**STROKE PREDICTION**

J. Spencer Shaw

**1. Introduction**

According to the CDC, stroke is a leading cause of death in the United States, and approximately 795,000 suffer from one each year. Strokes are preventable, and there are many factors that help to prevent or lower the chances of one. This report will be concerned with predicting whether someone will have a stroke given 10 parameters (p = 10) about a patient. The dataset used contains information on 4,909 individuals (n = 4,909) each identified by a unique identification number. This report will aim to use single- and multiple-logistic regression techniques in attempt to answer the overarching question: Is it possible to determine whether a given patient will have a stroke given the 10 predictors utilized in the dataset?

**2. Exploratory Data Analysis**

Initial analysis of the dataset reveals a sample size of 4,909 unique patients and 12 variables for each. 2 of these variables are a unique identification number of the patient and whether they have had a stroke. This left 10 predictor variables to analyze.

**3. Analysis**

**4. Conclusions and Interpretations**

**5. Shortcomings**

**6. Appendix**