# P8106 Midterm Final Report

Ila Kanneboyina, Shayne Estill, Naomi Simon-Kumar (ns3782)

03/24/2025

## Contents

# 1. Introduction

## 1.1. Background and Study Objective

This study aims to build a prediction model for antibody responses to a newly authorized vaccine, as measured by log-transformed antibody levels from dried blood spot samples. Our objective is to develop an accurate model to improve understanding of vaccine effectiveness across different population segments and to support the monitoring of immune protection over time.

## 1.2. Data Source and Description

The primary dataset (**dat1.RData**) contains demographic and clinical information from participants in a vaccine response study. Variables include age, gender, race/ethnicity, body mass index (BMI), blood pressure, cholesterol levels, diabetes, hypertension, and time since vaccination (in days). The outcome of interest is the log-transformed antibody level, measured using dried blood spot samples. The dataset consists of records from 5000 participants with 13 variables, excluding the unique identifier variable. A second independent dataset (**dat2.RData**) with identical structure was collected several months later to assess model generalizability. The dataset consists of records from 1000 participants.

# 2. Exploratory Analysis

## 2.1. Dataset Overview

## 2.2. Summary Statistics

## 2.3. Correlation Analysis

## 2.4. Exploratory Plots

# 3. Model Training

## 3.1. Data Preparation

## 3.2. Cross-validation approach

For model training and evaluation, we implemented 10-fold cross-validation on the training dataset (80% of the original data) using the trainControl function This partitioned the training data into 10 equal subsets, where each model was trained on 9 folds and validated on the remaining fold, rotating through all folds.

### 3.3. Models

#### 3.3.1. Linear Regression

#### 3.3.2. Elastic Net

#### 3.3.3. MARS

#### 3.3.4. GAM

## 4. Results

### 4.1. Model Comparison and Selection

### 4.2. Final Model: xxx

- Detailed specification of the selected model (coefficients, terms, etc.)
- Visualisations if appropriate

#### 4.2.1. Evaluation of model performance

Test set performance metrics & interpretation

### 4.3. Interpretation

i.e., summary of model results

## 5. Conclusion

## 6. References

If relevant

## 7. Appendices

If relevant