

## Framingham Heart Study

The Framingham Heart Study is a long-term prospective study of the etiology of cardiovascular disease among a population of free-living subjects in the community of Framingham, Massachusetts. The Framingham Heart Study was a landmark study in epidemiology in that it was the first prospective study of cardiovascular disease and identified the concept of risk factors and their joint effects FHS Longitudinal Data Document

The dataset is a rather small subset of possible FHS datasets, having 4240 observations and 16 variables. The variables are as follows:

1. sex: the gender of the observations. The variable is a binary named "male" in the dataset.
2. age: Age at the time of medical examination in years.
3. education: A categorical variable of the participants education, with the levels: Some high school (1), high school/GED (2), some college/vocational school (3), college (4)
4. currentSmoker: Current cigarette smoking at the time of examinations
5. cigsPerDay: Number of cigarettes smoked each day
6. BPmeds: Use of Anti-hypertensive medication at exam
7. prevalentStroke: Prevalent Stroke (0 = free of disease)
8. prevalentHyp: Prevalent Hypertensive. Subject was defined as hypertensive if treated
9. diabetes: Diabetic according to criteria of first exam treated
10. totChol: Total cholesterol (mg/dL)
11. sysBP: Systolic Blood Pressure (mmHg)
12. diaBP: Diastolic blood pressure (mmHg)
13. BMI: Body Mass Index,  $\text{weight (kg)/height (m)}^2$
14. heartRate: Heart rate (beats/minute)
15. glucose: Blood glucose level (mg/dL)
16. TenYearCHD(response variable): The 10 year risk of coronary heart disease(CHD).

The goal is to make some predictive models on an FHS dataset and review some exploratory and modeling techniques.