

Final Review

Questions

May 7, 2018

```
r ## numbers >= 10^5 will be denoted in scientific notation, ## and rounded to 2
digits options(scipen = 3, digits = 4) suppressMessages( require(car))
```

The Data

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. The study began in 1948 and 5,209 subjects were initially enrolled in the study. Participants have been examined biennially since the inception of the study and all subjects are continuously followed through regular surveillance for cardiovascular outcomes. Clinic examination data has included cardiovascular disease risk factors and markers of disease such as blood pressure, blood chemistry, lung function, smoking history, health behaviors, ECG tracings, Echocardiography, and medication use. Through regular surveillance of area hospitals, participant contact, and death certificates, the Framingham Heart Study reviews and adjudicates events for the occurrence of Angina Pectoris, Myocardial Infarction, Heart Failure, and Cerebrovascular disease.

The dataset is a subset of the data collected as part of the Framingham study and includes laboratory, clinic, questionnaire, and adjudicated event data on 4,434 participants. Participant clinic data was collected during three examination periods, approximately 6 years apart, from roughly 1956 to 1968. Each participant was followed for a total of 24 years for the outcome of the following events: Angina Pectoris, Myocardial Infarction, Atherothrombotic Infarction or Cerebral Hemorrhage (Stroke) or death.

Logistic Model

Our goal will be to build a model to explain hypertension. We will do this cross-sectionally and consider what covariates at baseline explain whether or not a subject will have hypertension. Your goal with this will be to go through the output given here and pick the best model:

1. Discuss the outcome variable (what is the outcome, how many subjects have this outcome? ...)
2. Make a table of the values for the simple logistic regressions. Comment on what you see here.
3. Walk through the process of multivariate model building.
 - a. Look over the models, do the coefficients look the same as the simple logistic regression?
 - b. What model would you choose?
 - Make sure to show tests comparing some of the models and why you chose the model you chose.
 - Remember some tests are shown that may not be done appropriately.

Survival Model

We will now build a model to explain hypertension with survival data. Your goal with this will be to go through the output given here and pick the best model:

1. Discuss the outcome variable (what is the outcome, how many subjects have this outcome? ...)
2. Consider the 4 models.
 - Do the models contain significant coefficients?
 - Do the models meet the proportional hazards assumptions?
3. Can we pick a final Model?
 - Does this have the same variables as the logistic?