Cardiovascular Disease Research at Duke: Case Study 1

Zach White

Note: you are not to communicate about the case study with anyone in any format, written, oral, or otherwise, with the exception of Prof. Herring. Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity. Any communication about the case study except with Prof. Herring will be treated as a violation of Duke's honor code.

Assignment

Using data from the Duke Databank for Cardiovascular Diseases on patients undergoing a cardiac catheterization procedure for suspected disease from 1985-2013, we will explore factors predictive of a subsequent myocardial infarction (MI). These factors of interest include patient demographics, medical history, and test results. This dataset, cs1.RData, has been de-identified in order to remove individually identifiable protected health information.

Your job is to use a generalized linear model or other closely related method to examine the relationship between predictors in the dataset and the outcome of myocardial infarction (MI, or heart attack), to evaluate strengths and limitations of analysis models (including predictive ability), and to describe the associations between the predictors and MI in language suitable for a scientific journal. Informative graphics should be used whenever possible. The results should be summarized in an R markdown document not exceeding 6 pages (using standard font) after compilation, containing sections for methods and results.

Assessment

Your report should be submitted on Sakai as an R markdown file (no page limit), and an accompanying .pdf (6 page limit) should be **turned in to me in class on Oct 13 at 11:45am**. All components should be completely reproducible from the markdown file (that is, all I should have to do is point to the right location of the original data set on my computer and run your code; you may wish to use a computing lab machine to test reproducibility). The data should be kept secure (a password-protected machine is fine) and not shared with others. Grading will be based in large part on the following aspects of your report.

1. Methods (description and validity of analysis plan, including model structure and procedures for determining how the predictors will be used to address the scientific question)
2. Results

- Logical presentation of findings supported by analysis
- Complete explanation of data and results (e.g., point and interval estimates provided, direction and interpretation of associations clearly described)
- Validity of model results clearly addressed
- Figures/tables with clear labels and captions, referenced and described in text (no obvious exclusions in selection of figures and tables)
- Discussion of limitations of findings: be honest; don't overstate quality of model or confidence in results! Usually a paper in the field just provides support for or against some hypotheses, rather than completely solving the problem of what causes MI, so do not overreach.

3. Reproducibility (analysis easily reproduced using code provided)
4. General

- References where appropriate

- Careful editing of manuscript for items including spelling, punctuation, and grammar
- Proper formatting of analysis results (publication format, as tables in a journal article, is strongly preferred to raw R output)
- Document nicely formatted (easy to read text, logical ordering, text/captions clear and not in too small font)

Data Description

The outcome, MI, is coded MI=1=subsequent MI, 0=no subsequent MI.

The predictors of interest include the following.

AGE_G (categorized as 1=18-24; 2=25-29; 3=30-34; 4=35-39; 5=40-44; 6=45-49; 7=50-54; 8=55-59; 9=60-64; 10=65-69; 11=70-74; 12=75-79; 13=>=80)

RACE_G (1=Caucasian, 2=African American, 3=Other)

HXSMOKE (1=history of smoking, 0=no history of smoking)

CHFSEV (severity of congestive heart failure, 0=None, 1-4 correspond to New York Heart Association classifications I (cardiac disease does not limit normal activity though symptoms may present with exertion), II (slight limitation of ordinary activity though could have symptoms with light activity such as walking more than 2 blocks or climbing more than one flight of stairs), III (patient comfortable at rest but marked limitations to physical activity), and IV (patient has symptoms even at rest that increase with any physical activity))

HXANGINA (1=history of anginal pain, 0=no history of anginal pain (excludes pain from any prior MI))

HXCEREB (1=history of cerebrovascular disease, 0=no history of cerebrovascular disease)

HXCHF (1=history of congestive heart failure not due to acute MI, 0=no such history)

HXCOPD (1=history of chronic obstructive pulmonary disease, 0=no such history)

HXDIAB (1=history of diabetes, 0=no history of diabetes)

HXHTN (1=history of clinically significant hypertension, 0=no such history)

HXHYL (1=history of hyperlipidemia, 0=no such history)

HXMI (1=history of prior MI, 0=no prior MI)

LVEF_R (left ventricular ejection fraction (%) as measured during the catheterization)

NUMDZV (number of significantly diseased vessels, measured during the catheterization)