

# Homework 4, Due October 27 at 11:45am

*Zach White*

**Provide clearly documented code in an R markdown document where appropriate. Whenever possible, provide statistical evidence to support your answers.**

## 1. Heart Failure Severity

Using the Duke Databank for Cardiovascular Diseases data in `cathhw4.Rdata`, consider the outcome of heart failure severity (0=no heart failure, 4=most severe) as a function of age group, gender, and smoking history. Note the variable definitions:

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))

GENDER (1=female, 0=male)

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= $\geq 80$ )

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

### 1a. Descriptive Statistics

Explore relationships among the four variables using graphical displays and describe your findings.

### 1b. Baseline Category Logistic Model

Fit a baseline category logistic model to the heart failure severity data, with the baseline category defined as CHFSEV=0, and main effects of the predictors age group, gender, and smoking history. Use the functional form of age group that provides the best fit to the data and assess whether any interaction terms are needed. Provide a clearly-labeled table that gives the OR's and 95% CI's relating the predictors to heart failure severity and interpret all estimates that are statistically significant at the  $\alpha = 0.05$  level.

### 1c. Prediction in Baseline Category Logistic Model

Using your preferred linear predictor from 1b, find and report in a table the predicted probabilities of each level of heart failure severity from the baseline category logistic model for (i) a 25 year-old woman who has never smoked, (ii) a 25 year-old woman who smokes, (iii) a 25 year-old man who has never smoked, (iv) a 25 year-old man who smokes, (v) an 80 year-old woman who has never smoked, (vi) an 80 year-old woman who smokes, (vii) an 80 year-old man who has never smoked, (viii) an 80 year-old man who smokes.

### 1d. Proportional Odds Model

Evaluate whether the proportional odds assumption is satisfied for any of the predictors, using the same functional forms as in 1b. Provide test results to support your conclusions.

### **1e. Interpretation of Estimates in Proportional Odds Model**

Regardless of your conclusions in part 1c, use the same functional forms as in 1b and provide a clearly-labeled table that gives the OR's and 95% CI's relating the predictors to heart failure severity in the (fully) proportional odds model. Interpret all estimates that are statistically significant at the  $\alpha = 0.05$  level.

### **1f. Prediction in Proportional Odds Model**

Using your preferred linear predictor from 1b, find and report in a table the predicted probabilities of each level of heart failure severity from the proportional odds model for (i) a 25 year-old woman who has never smoked, (ii) a 25 year-old woman who smokes, (iii) a 25 year-old man who has never smoked, (iv) a 25 year-old man who smokes, (v) an 80 year-old woman who has never smoked, (vi) an 80 year-old woman who smokes, (vii) an 80 year-old man who has never smoked, (viii) an 80 year-old man who smokes.

## **2. Agresti 8.34 (Separate Logistic Regressions)**