The Cardiovascular Health Study (CHS) is a population-based, longitudinal study of coronaryheart disease and stroke in adults aged 65 years or older. The primary goal of the study is toidentify factors related to the onset and course of coronary heart disease and stroke in an elderly population. Data represents the 2690 (out of an original 5201) that have complete data. The goal of this study is to quantify the association between possible risk factors and coronary heart disease (CHD). The data set includes the following observations, recorded at a single time point: CHD (history of myocardial infarction, angina, or cardiac procedure, 1=yes and 0=no) Age (years) Hyperten (history of hypertension, 1=yes and 0=no) Diabetic (history of diabetes, 1=yes and 0=no) Gender (1=male and 0=female) HDL (high density lipoprotein level mg/dl) Smoke (1=never smoked, 2=former smoker, 3=current smoker) Height (cm) Weight (lbs) Packyrs (number of years smoking \* average number of packs smoked per day) Currsm (1= current smoker, 0=otherwise) Formsm(1=former smoker, 0=otherwise) Eversm(1=ever smoked, 0=never smoked)

* This problem will focus on smoking as a potential risk factor for CHD. This is anexercise, focused on logistic model interpretation, and is not necessarily the most appropriate analysis of the association between smoking and CHD.

1. Fit a logistic regression model to determine if smoking status is significantly associated with CHD. To do this, specify CHD as the dependent variable and smoke as the independent variable. Interpret the resulting coefficient estimates, including the intercept. (Comment about the p-values, ORs and CIs. Note: the intercept estimates the probability of CHD for never smokers here.)
2. Now, fit a logistic regression model to determine if pack-years is associated with CHD. First, describe the distribution of Packyears for the never, former, and current smokers with descriptive statistics. Then, describe the distribution of Packyears for those with and those without a history of CHD with descriptive statistics
3. Fit a logistic regression model to determine if smoking (as measured bypack-years) is significantly associated with CHD, where CHD is the dependent variable and Packyrs is the covariate. Interpret the resulting coefficient estimates, including the intercept. Compare the results from this approach to the modeling approach in part a.
4. What is the estimated odds ratio of CHD associated with a 5 unit increase in packyears, using the estimated model from part c?
5. What is the estimated probability of CHD for a subject with 0 packyears of smoking exposure? What is the estimated probability of CHD for a subject with 40packyears of smoking exposure? (Use the estimated model from part c)