BIOST 536: Homework 5

Department of Biostatistics @ University of Washington

8 November 2024

1.

Suppose we are interested in whether there is an association between education and risk of CHD within 10 years (TenYearCHD). Education is a 4-level ordinal variable, where 1 = less than high school; 2 = high school diploma/GED; 3 = some college/vocational school; and 4 = college degree (BA,BS) or more. Consider a logistic regression model that treats education as a categorical (indicator) exposure variable and adjusts for the confounders: sex, age, BMI, current smoking status, total cholesterol, and systolic blood pressure (male, age, BMI, currentSmoker, totChol, sysBP).

a. Fit the above described model and interpret the coefficients of the education variable.

b. State the null hypothesis for the association between education and outcome for this model. How do we state the alternative hypothesis?

c. Provide a statistical test that you could use to test the hypothesis in 1b and how many degrees of freedom does this test have? (Hint: there is more than one valid test to choose from)

d. Perform the statistical test in 1c and state your conclusions.

e. If you were only looking at the individual coefficients for the education levels in this model, would you have been able to make the same conclusion as in 1d? Why or why not?

f. Suppose you treated education as a grouped linear variable instead, write down the model for logit(p) using the same confounders in part 1a.

g. Fit the model in 1f and state your conclusions about whether there is a linear trend in the association between education and CHD outcome.

h. Can you do a likelihood ratio test to compare the model in 1f to the model with education as a categorial variable? If yes, how many degrees of freedom does this test have?

2.

Now suppose we are interested in the same outcome, but with BMI as the exposure, and considering the categorical education variable and all other covariates from the model fit in 1a as confounders.

a. Treating BMI as a continuous variable, test whether there is a linear association between BMI and the outcome.

b. Create a 4-level ordinal variable BMIcat that categorizes the level of obesity according to standard cutoffs from the NHANES, where BMIcat=1 if BMI<18; BMIcat =2 if 18 BMI <30; BMIcat=3 if 30 BMI < 35; and BMI=4 if BMI35. Fit a model that uses this grouped exposure variable for the association between BMI and outcome that adjusts for same covariates but assumes the least amount of structure between the association of BMI and outcome.

c. Write down the null hypothesis that there is no association between BMIcat and outcome. Perform a statistical test of this hypothesis and state your conclusions. Is it different from 2a?

d. Calculate the OR and 95% confidence interval for a 2-unit increase in BMI cat.

e. Refit the same model in 2b, but now using these centered continuous variables: age centered at 60, systolic blood pressure centered at 140, and total cholesterol centered at 200, and leaving remaining other variables the same. What is the interpretation of the intercept in this model compared to the one fit in 2b?

f. Make a plot of logit(p) versus BMIcat using the model in 2e, for a female non-smoker aged 60 years with less than high school education, systolic blood pressure =140 and total cholesterol = 200. Looking at this plot, is there a suggestion that there is a non-linear association between BMI and logit (p)?

g. Fit a model that treats BMI as continuous but fits a nonlinear association (your choice) between BMI and outcome, adjusting for same confounders as in 2a. In this model, perform a test for whether there is an association between BMI and outcome and state your conclusion.

h. Optional: can you plot the non-linear association you found in 2g.

End of report. Code appendix begins on the next page.

Code Appendix

```
# clear environment
rm(list=ls())

# setup options
knitr::opts_chunk$set(echo = FALSE, warning=FALSE, message=FALSE)
options(knitr.kable.NA = '-', digits = 2)
labs = knitr::all_labels()
labs = labs[!labs %in% c("setup", "allcode")]
# load relevant packages
library(dplyr)      # data frame manipulation
library(ggplot2)
library(gtsummary)  # table summaries
# library(dagR)       # DAG simulation
library(broom)        # model coefficient table
# library(knitr)      # pretty tables
library(rigr)         # regression

# load data
load("../data/Framingham.Rdata")
# names(esph)
# dim(esph)
```

End of document.