

BST 210 HOMEWORK #5

Due 11:59pm, Tuesday, October 29, 2019

***Please be sure to submit your assignment by 11:55ish pm (or before) to prevent any glitches in the upload from precluding your timely submission.**

***Please work well in advance, getting help during office hours and labs, as there will be no extensions given for this assignment, outside of extreme, extenuating circumstances which must be communicated in advance to the primary instructor.**

There is 1 problem with 6 parts (a-f) in this homework assignment. Please double check that you have provided a response for each part of the problem, before you submit.

BST 210 Problem set policies:

- *We encourage you to discuss homework with your fellow students (or with the instructor or the TAs), but you must write your own final answers, in your own words.*
- *Please include the appropriate computer output in your solution if that helps you to answer a question, but be sure to interpret your findings in words – submitting only output is not sufficient for full credit.*
- *Homework assignments will not be accepted late (other than for extreme emergency, but the primary instructor must be reached in advance).*
- *Be complete in your responses; not verbose, to get full scores.*
- *All homework must be submitted online via Canvas by 11:59pm on Tuesday.*

Consider the Framingham Heart Study data set that we used previously in a lab session. Here we focus on predicting “death from any cause” (mortality) over the 24-year period of follow-up, and focus on continuous BMI (body mass index), participant sex, and age at exam (or age category) as independent variables. The dataset and a help file are available in the HW5 folder on the course website.

Problem 1

a) Use logistic regression to assess the effects of (continuous) BMI on mortality. Briefly interpret your model. What are your conclusions? Also estimate an odds ratio and a 95% confidence interval for the effect of a 5-unit change in BMI.

b) One way to assess possible nonlinear effects of BMI (on the logit scale) is to run a logistic regression model including (linear) BMI and (quadratic) BMI² in the same model. Generate a BMI² term, run models containing only the linear term and then including both the linear and quadratic terms, and determine if the quadratic term is needed or not. What happens to the linear effect when the quadratic term is included in the model? Also, graph the fitted probabilities from these two models overlaid on the same plot and (briefly) compare.

c) For the model including both linear and quadratic BMI, estimate the odds ratio for a 5-unit increase in BMI (comparing 25 to 20) and for a 5-unit increase in BMI (comparing 35 to 30). (Because we have a quadratic BMI term in the model, these two odds ratio estimates should differ, because BMI is “interacting with itself”.)

d) Go back to using only the linear BMI term. Perform some descriptive statistics or graphical display to assess the association between BMI and participant sex. Then perform an appropriate set of logistic regression analyses to determine whether or not sex is a confounder or an effect modifier of the effect of (continuous) BMI on mortality. What are your conclusions (in words) about the effect of BMI on mortality, considering the additional effects of sex? (Hint: It may be helpful to create a 0/1 indicator variable for sex, e.g., “Female = 1 for females, Female = 0 for males”.)

e) Now considering age and age category alone (not BMI or sex), compare models using (continuous) age, (ordinal) age category (i.e., age category used as a continuous covariate), and (categorical) age category. Which approach do you feel best models the effect of age on mortality? Justify your response. (It may be helpful to look at or plot fitted probabilities or run a hypothesis test.)

f) Perform an appropriate set of logistic regression analyses to determine whether or not age category is a confounder or an effect modifier of the possible effect of (continuous) BMI on mortality. What are your conclusions about the effect of BMI on mortality, considering the additional effects of age category?