Homework 4

BSTA 512/612

2024-02-22

Directions

- Please upload your homework to Sakai. **Upload both your .Rmd code file and the knitted .html file**.
- For each question, make sure to include all code and resulting output in the html file to support your answers.
- Show the work of your calculations using R code within a code chunk. Make sure that both your code and output are visible in the knitted html file.
- Write all answers in complete sentences as if communicating the results to a collaborator.
  - Points (usually 0.5-1) will be deducted for not including a sentence summarizing results in the context of the research study.
  - Questions not requiring a sentence are: *none - include a summary for all questions*

  Tip: It is a good idea to try kitting your document from time to time as you go along! Note that knitting automatically saves your Rmd file and knitting frequently helps you catch your errors more quickly.

Chapter 11 & 12 (questions NOT from book)

Question 1

Use the data from Chapter 12 Problem 3 to answer the questions below.

a)

a. How many dummy variable(s) do you need to create for the categorical variable Diet (protein-rich vs. protein-poor)? Create the dummy variable(s) with the reference cell coding approach{0,1}.

b)

b. At a level of significance alpha = .05, test whether if Age is significantly associated with Height. Would this association be modified depending on diet group (e.g., rich-protein or poor-protein)? In other words, is Diet an effect-modifier that changes the association between Height and Age? Justify your answer (e.g., perform a hypothesis test at a level of alpha =.1).

Note: recall that an effect modifier is an interaction.

c)

c. From the results obtained in part b, should we perform an assessment of a confounder for Diet? Justify your answer. Perform such an assessment if needed.

d)

d. Perform a regression analysis on the model obtained from the results obtained from parts a- c. Write down a general regression equation that is applicable to both groups—rich-protein vs.poor-protein. Write down regression lines for each specific groups—rich-protein or poor-protein.

Question 2

Use the data from Chapter 9 Problem 5 to answer the questions below.

a)

a. Use $\alpha = 0.05$, test whether the (crude) association between Y and X1 could be established.

b)

b. Use$\alpha = 0.1$, test whether X3 is an effect modifier of the association between Y and X1.

Note: To identify effect modifiers, we perform a hypothesis test of interaction term, e.g.,X1*X3. That is: The full model includes X1, X3, X1*X3. the reduced model includes X1 and X3

c)

c. From the result obtained in part b, do we need to perform an assessment of a confounder for X3? Justify your answer. Perform such an assessment if needed.

d)

d. Perform an assessment of a confounder for X2 which potentially changes the association between Y and X1.

e)

e. From the results in parts a-d, what is your final association model?

Chapter 15: Polynomial Regression (questions NOT from book)

Use the data from Chapter 5 Problem 18 to answer the questions below.

a)

a. Obtain scatter plot: Y vs. X. Does linear trend support the relationship between Y and X?

b)

b. At the level alpha =.05, test whether the linear relationship could be established between Y and X.

c)

c. At the level alpha =.05, test whether the quadratic term (X^2) should be included in the model to improve the prediction in Y, given the linear term (X) is already in the model.

d)

d. From the result obtained from part c, should we test if the linear term (X) is necessary to be included in the model, given the quadratic term is already in the model? Explain your answer.

e)

e. From the results you obtain from parts c-d, should we further examine whether cubic term (X^3) or fourth polynomial degree (i.e., X^4) to improve the prediction in Y? Explain your answer and report the result of such a test if needed.

f)

f. From parts a – e, what is the final model that you have obtained? Interpret the R-square result from this model in the context of the study. Plot fitted value curve vs. X overlaid with scatter plot. Comments about the fitting model.