

**STAT 2131:**  
**Applied Statistical Methods I**  
**HW #4**  
**Due Tuesday, Nov 2nd**

1. You are working on a study whose goal is to determine how the region of the country (“region”) a hospital is in affects a patient in that hospital’s chance of developing a hospital-acquired infection (“risk”). The data for this study is in senicdata.txt (from KNNL book appendix C.1). You want to look at this relationship while controlling for the number of beds (“beds”) in the hospital, the number of nurses in the hospital (“nnurse”), and the percent of available facilities and services at the hospital (“facilities”).
  - (a) Fit a first order regression model which address the main question.
  - (b) Do graphical checks to see if you should add any second order terms ( $X_1^2$ ,  $X_1X_2$  etc). What second order terms do you think you will need in your model?
  - (c) Fit the full second order model. What is the hypothesis that no interaction terms are needed? Do this test and report its p-value.
  - (d) Are there any other terms you think you might not put in your report? Fit the model you would like to report.
  - (e) Write a short paragraph describing your findings. Be sure to report the results of a test that address the primary question.
2. You are working on a study that uses the same data set that you used in Question 1. The epidemiologist you are working with suspects that older people feel energized when they are around young people and that the relationship between age and risk of infection might be different in teaching hospitals (“school” = 1) where patients interact with young people on a daily basis and non-teaching hospitals (“school” = 2). She wants you to test this hypothesis using a linear regression model while controlling for the number of nurses working at the hospital. She also informs you that it is well accepted in the field that the number of nurses in a hospital has a non-linear association with infection.

Find an appropriate regression model, fit it, and test if there is a difference in the relationship between age and risk of infection in teaching and non-teaching hospitals. Write a short summary to the epidemiologist. In this summary, be sure to directly address the epidemiologist’s question, quantify interesting relationships, and describe why you chose the model that you did.
3. Refer to the same SENIC data in question 1 and 2. Consider the regression relation of average length of stay (“stay”) to infection risk. Here the length of stay is the response.

- (a) Obtain Bonferroni joint confidence intervals for  $\beta_0$  and  $\beta_1$ , with a 90 percent family coverage.
- (b) A researcher suggest that  $\beta_0$  should be approximately 7 and  $\beta_1$  should be approximately 1. Do the joint intervals in part (a) support this expectation?
- (c) It is desired to estimate the expected hospital stay for persons with infection risks  $X = 2, 3, 4, 5$  with family coverage 95 percent. Which procedure, the Working-Hotelling or the Bonferroni is more efficient here?