Problem Set 7 (one page)

Statistics 24510-30040 (W19)

Due Tuesday, March 5, at the beginning of class.

Requirements Provide detailed derivations. Select only the relevant part of the output to be inserted. Attach your code or output as an appendix if necessary. Discussions allowed, the assignment should be devised and written by yourself completely.

Problem assignments (Relevant reading: Review Chapter 14.)

1. (*Inference on mean of a future observation*) Problem 13, Chapter 14, page 597 in Rice (3rd ed.).

2. (*Inference on a future observation*) Problem 14, Chapter 14, page 597-598.

3. (*Multiple linear model with indicator variable*)

   Data sets: (refer to data in Problem 44, chapter 14 of the text.)
   - https://www.stat.uchicago.edu/ meiwang/courses/w19stat2/asthma.txt
   - https://www.stat.uchicago.edu/ meiwang/courses/w19stat2/cystfibr.txt

   Let y be the response variable measuring respiratory `resistance`, x be the covariate `height` (cm),

   $$z = \begin{cases} 1, & \text{if the } i\text{th observation is made on a child with cystic fibrosis,} \\ 0, & \text{if the } i\text{th observation is made on a child with asthma.} \end{cases}$$

   Using indicator variable (*a.k.a.* "dummy variable") z, we can analyze both data sets simultaneously.

   Consider the model

   $$Y_i = \alpha_0 + \alpha_1 x_i + \beta_0 z_i + \beta_1 x_i z_i + \varepsilon_i, \qquad i = 1, \cdots, n, \tag{1}$$

   where $n = 42 + 24 = 46$ is the sample size combining both data sets.
   Assume that the errors ε_i are independent $N(0, \sigma^2)$ random variables.

   (a) Explain the meaning of the parameters $\alpha_0, \alpha_1, \beta_0, \beta_1$. In particular, explain, for all four parameters in turn, the consequence of setting the parameter to zero.

   (b) Use R to fit the model in (1) to the combined asthma/cystic-fibrosis data. Discuss the results of the four t-tests.

   (c) Use an F-test to test the hypothesis $H_0 : \beta_0 = \beta_1 = 0$. Explain the meaning of the hypothesis.

   (d) Use graphical displays to check the constant variance assumption made in the model (1).

4. (*Weighted Least Squares*) Problems 7, Chapter 14, page 593.

5. Redo parts (a), (b), and (d) of the previous question for the multiple general regression setting; i.e. $Y = X\beta + \varepsilon$, where the ε_i's (the components of ε) are indipendent with mean 0 and variance $var(\varepsilon_i) = \rho_i^2 \sigma^2$.

   As much as possible, write your solution using matrix notation. For example, if X is the original design matrix, then the transformed design matrix can be written as $R^{-1}X$, in which R is a diagonal matrix with as many row as X whose diagonal elements are the ρ_i's.