STAT/Q SCI 403: Introduction to Resampling Method
Spring 2019
Homework 07

Instructions:

- You have to submit all your answers in a single PDF file generated by either LaTeX or *Rmarkdown*.
- You may use the LaTeXtemplate `HW_template.tex` to submit your answer.
- For questions using R, you have to attach your code in the PDF file. If the question ask you to plot something, you need to attach the plot in the PDF as well.
- If the question asks you to show a figure, the clarity of the figure will also be graded.
- The total score of this homework is 8 points.
- Questions with ♠ will be difficult questions.

Questions:

1. In this question, we will focus on the `iris` dataset. In particular, we treat the variable `Petal.Width` as the response variable and the variables `Sepal.Length`, `Sepal.Width`, and `Petal.Length` as covariates. We will use the three bootstrap approaches to analyze the uncertainty of the linear regression. When using the bootstrap, please use at least $B = 10,000$ bootstrap samples.

   (a) **(1 pt)** Fit a linear regression. What are the fitted coefficients?

   (b) **(2 pt)** Apply the empirical, residual, and wild bootstrap to find the variance of the four fitted coefficients (intercept, and slope of the three covariates). Use a matrix to compare the variance of each fitted coefficients under the three bootstrap approaches. You should report a 3×4 matrix or 4×3 matrix and you need to indicate what each column and row represent.

   (c) **(1 pt)** For the intercept, use a single plot to compare its distribution from the three bootstrap approaches. Hint: There are many ways to do it. Use your creativity.

2. In this question, we will use the dataset we used in Lab session 07, part 4 (from the file `binary.csv`). In particular, we will fit a logistic regression model with response $Y =$`admit` and two covariates `gre` and `gpa`. When using the bootstrap, please use at least $B = 10,000$ bootstrap samples.

(a) **(1 pt)** Use the parametric bootstrap to construct a 90% CI for the slope of `gpa`.

(b) **(1 pt)** Apply both the parametric and empirical bootstrap to estimate the standard error (standard deviation of the estimator) of the intercept, slope of `gre`, and slope of `gpa`. Use a single matrix to compare the standard errors of the three parameters obtained using the two bootstrap methods and the value from `summary()` function. You should report a 3×3 matrix and you need to indicate what each column and row represent.

(c) **(1 pt; ♠)** Assume that we are interested in the following quantity:

$$\lambda = P(\mathsf{admit} = 1 | \mathtt{gre} = 500, \mathtt{gpa} = 3.7).$$

Use a bootstrap method to compute a 90% confidence interval of λ.

(d) **(1 pt; ♠)** John and Sam are two undergraduates about to graduate. They both are interested in applying to a graduate school. John was lazy and did not pay much attention to the coursework so his GPA is 2.3. But later John realized that this will get him into trouble in applying to graduate school so he studied very hard in preparing the GRE test. He obtained a score of 700 in the GRE test. Sam, on the other hand, worked very hard in the coursework so his GPA is 3.9 and he obtained a score of 670 in the GRE test.

Under the logistic regression model and using the dataset, do they have distinct probabilities of receiving an admission? In other word, we want to test the null hypothesis

$$H_0 : P(\mathsf{admit} = 1 | \mathtt{gre} = 670, \mathtt{gpa} = 3.9) = P(\mathsf{admit} = 1 | \mathtt{gre} = 700, \mathtt{gpa} = 2.3).$$

Let the significance level $\alpha = 0.1$. Use a bootstrap approach to see if we can reject the null hypothesis.