

STAT 509 2017 Summer HW16

Instructor: Shiwen Shen

Lecture Day: June 8

1. Using the `gala` dataset, fit a simple linear regression model with **Species** as response and **Elevation** as explanatory variable. Show me the output.
 - (a) Draw the residual plot, as well as a solid line indicating 0 residual. Make comments.
 - (b) You may realize that we need a good transformation (for Y) to get better result. In the BoxCox framework, find the best transformation. This is equivalent to find the best λ .
 - (c) Refit the model with transformed Y , and draw the residual plot again. Compare two residual plot, what is your conclusion?
 - (d) Compare the Coefficient of Determination, is the transformed model better?
2. A researcher is interested in how variables, such as GRE (Graduate Record Exam scores) and GPA (grade point average) effect admission into graduate school. The response variable, admit (1) / don't admit (0), is a binary variable. *Download the `edu` dataset from the course webpage, save it in D drive, and run the following R code to load:*

```
edu <- read.table("D:/edudata.txt", sep="\t")
```

- (a) Fit a logistic regression model with **admit** as response and **gre** as explanatory variable. Show me the R output.
- (b) Is variable **gre** significant in the model? Why?
- (c) What is the function do we use to make estimate the probability that a student can be admitted with his/her GRE score?
- (d) If some student's GRE score is 700, what is the probability this student will be admitted?
- (e) Calculate the odds-ratio of **gre**, and make a solid interpretation.