**Using R in Financial Statistics (spring 2019)  
Assignment 3**

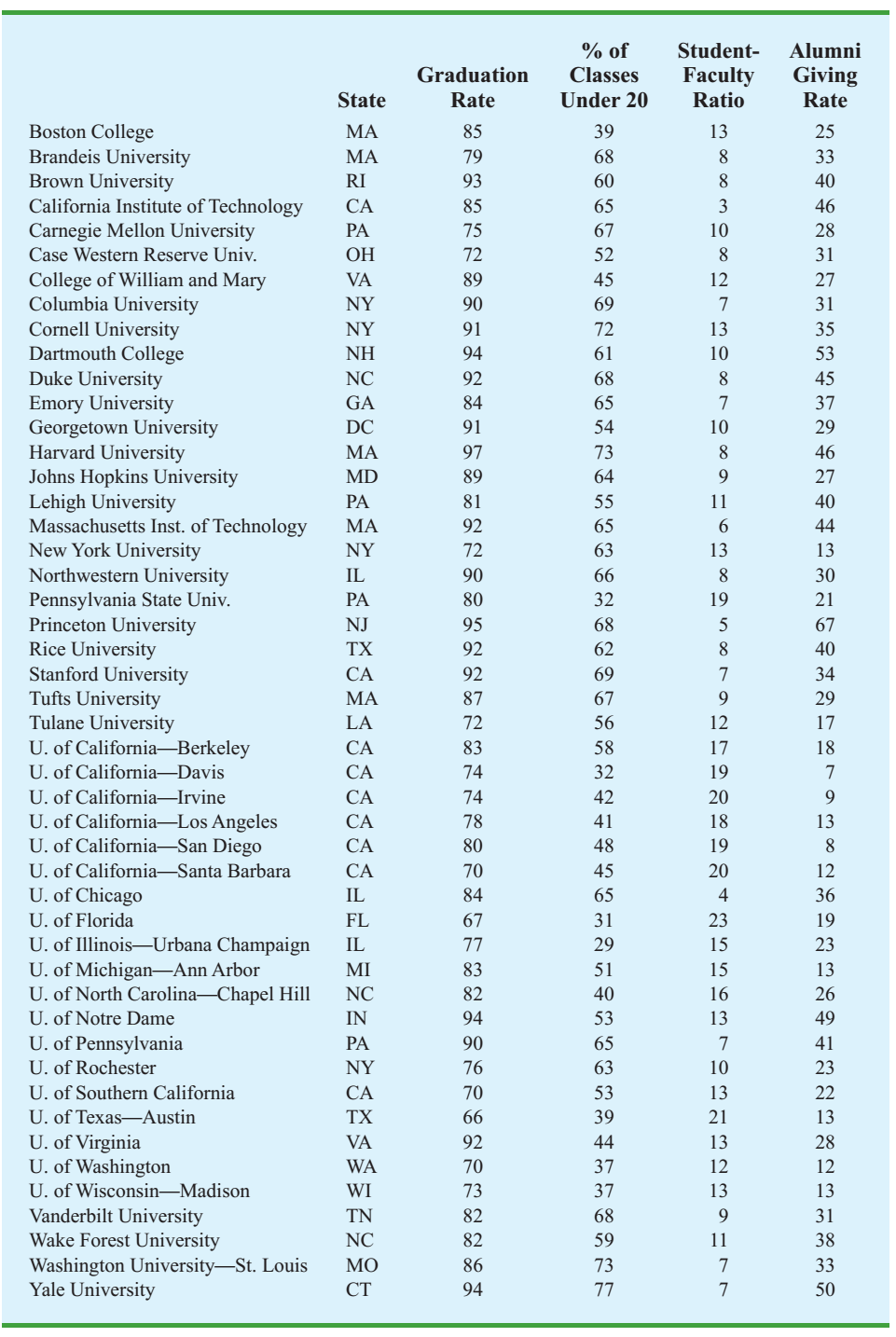
Due time: June 9 (Sunday), 24:00

**Part 1**

Alumni donations are an important source of revenue for colleges and universities. If administrators could determine the factors that could lead to increases in the percentage of alumni who make a donation, they might be able to implement policies that could lead to increased revenues. Research shows that students who are more satisfied with their contact with teachers are more likely to graduate. As a result, one might suspect that smaller class sizes and lower student-faculty ratios might lead to a higher percentage of satisfied graduates, which in turn might lead to increases in the percentage of alumni who make a donation. Table 1 shows the data for 48 US national universities (America’s Best Colleges, Year 2000 ed.). The column labeled Graduation Rate is the percentage of students who initially enrolled at the university and graduated. The column labeled % of Classes Under 20 shows the percentage of classes offered with fewer than 20 students. The column labeled Student-Faculty Ratio is the number of students enrolled divided by the total number of faculty. Finally, the column labeled Alumni Giving Rate is the percentage of alumni who made a donation to the university.

The data is stored in the file “alumni\_giving\_rate.csv”. Import the data into R and conduct the following analysis.

1. Find descriptive statistics of the data and summarize them into a table.
2. Use graphical analysis (such as scatterplot) to investigate the relationship between Alumni Giving Rate and each of the other variables.
3. Develop a multiple linear regression model that could be used to predict the Alumni Giving Rate using the data provided. This may include model specification and estimation. Summarize your findings with evidence and reasoning (possibly from the previous question).
4. Check the model assumptions.

Table 1. DATA FOR 48 US NATIONAL UNIVERSITIES  


**Part 2**

Attach the “swiss” data, which is about the standardized fertility measure and socio-economic indicators for each of 47 French-speaking provinces of Switzerland at about 1888. It is a data frame with 47 observations on 6 variables, each of which is measured as a percentage. The definitions of variables are given below. All variables but Fertility give proportions of the population.

Fertility common standardized fertility measure  
Agriculture % of males involved in agriculture as occupation  
Examination % draftees receiving highest mark on army examination  
Education % education beyond primary school for draftees.  
Catholic % ‘catholic’ (as opposed to ‘protestant’).  
Infant.Mortality live births who live less than 1 year.

Conduct the following analysis.

1. Calculate the mean of Fertility, and then partition the provinces into two groups, with group 1 including the provinces having above average Fertility measure, and group 2 including the remaining provinces. Use variable y to denote this group information.
2. Set the group 2 as the baseline group, and use logistic regression to show the relationship between y and the other variables except Fertility. Interpret the regression results.
3. Choose a model selection criterion, for instance, AIC, BIC, adjusted R square or Cp, and use it to select a reasonable model.

**Submission**

Save the source code as **assign3.R**

Summarize your results into a short report and save it as **assign3.pdf**Note: Do NOT just copy your running results. Use your own words to explain your reasonings and conclusions with supporting information (graphs, tables, etc.).

**[Important]** Send the above two files together to email: hjpszu@163.com, with email title  
RFS\_Assignment3\_studentID\_name  
E.g.: RFS\_Assignment3\_2018123456\_张三

You will receive an auto-reply only if your email is with a correct title.