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

	State	Graduation Rate	% of Classes Under 20	Student-Faculty Ratio	Alumni Giving Rate
Boston College	MA	85	39	13	25
Brandeis University	MA	79	68	8	33
Brown University	RI	93	60	8	40
California Institute of Technology	CA	85	65	3	46
Carnegie Mellon University	PA	75	67	10	28
Case Western Reserve Univ.	OH	72	52	8	31
College of William and Mary	VA	89	45	12	27
Columbia University	NY	90	69	7	31
Cornell University	NY	91	72	13	35
Dartmouth College	NH	94	61	10	53
Duke University	NC	92	68	8	45
Emory University	GA	84	65	7	37
Georgetown University	DC	91	54	10	29
Harvard University	MA	97	73	8	46
Johns Hopkins University	MD	89	64	9	27
Lehigh University	PA	81	55	11	40
Massachusetts Inst. of Technology	MA	92	65	6	44
New York University	NY	72	63	13	13
Northwestern University	IL	90	66	8	30
Pennsylvania State Univ.	PA	80	32	19	21
Princeton University	NJ	95	68	5	67
Rice University	TX	92	62	8	40
Stanford University	CA	92	69	7	34
Tufts University	MA	87	67	9	29
Tulane University	LA	72	56	12	17
U. of California—Berkeley	CA	83	58	17	18
U. of California—Davis	CA	74	32	19	7
U. of California—Irvine	CA	74	42	20	9
U. of California—Los Angeles	CA	78	41	18	13
U. of California—San Diego	CA	80	48	19	8
U. of California—Santa Barbara	CA	70	45	20	12
U. of Chicago	IL	84	65	4	36
U. of Florida	FL	67	31	23	19
U. of Illinois—Urbana Champaign	IL	77	29	15	23
U. of Michigan—Ann Arbor	MI	83	51	15	13
U. of North Carolina—Chapel Hill	NC	82	40	16	26
U. of Notre Dame	IN	94	53	13	49
U. of Pennsylvania	PA	90	65	7	41
U. of Rochester	NY	76	63	10	23
U. of Southern California	CA	70	53	13	22
U. of Texas—Austin	TX	66	39	21	13
U. of Virginia	VA	92	44	13	28
U. of Washington	WA	70	37	12	12
U. of Wisconsin—Madison	WI	73	37	13	13
Vanderbilt University	TN	82	68	9	31
Wake Forest University	NC	82	59	11	38
Washington University—St. Louis	MO	86	73	7	33
Yale University	CT	94	77	7	50

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.