# Multivariate Analysis
## 2019 Spring Final Exam

The data provided is related with direct marketing campaigns (phone calls) of a local retail bank. The goal is to build up your models, using R and Stan together with the exam questions, to predict the success rate of calls.

**Data Description:**
Input variables
# bank client data
- age (numeric)
- job: type of job (categorical: 'admin.', 'blue-collar', 'entrepreneur', 'housemaid', 'management', 'retired', 'self-employed', 'services', 'student', 'technician', 'unemployed', 'unknown')
- marital: marital status (categorical: 'divorced', 'married', 'single', 'unknown'; note: 'divorced' means divorced or widowed)
- education (categorical: 'basic.4y', 'basic.6y', 'basic.9y', 'high.school', 'illiterate', 'professional.course', 'university.degree', 'unknown')
- default: has credit in default? (categorical: 'no', 'yes', 'unknown')
- housing: has housing loan? (categorical: 'no', 'yes', 'unknown')
- loan: has personal loan? (categorical: 'no', 'yes', 'unknown')

# related with the last contact of the current campaign
- contact: contact communication type (categorical: 'cellular', 'telephone')
- month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
- day_of_week: last contact day of the week (categorical: 'mon', 'tue', 'wed', 'thu', 'fri')
- duration: last contact duration, in seconds (numeric).
  ***Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.***

# other attributes
- campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
- pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
- previous: number of contacts performed before this campaign and for this client (numeric)
- poutcome: outcome of the previous marketing campaign (categorical: 'failure', 'nonexistent', 'success')

# social and economic context attributes
- emp.var.rate: employment variation rate - quarterly indicator (numeric)
- cons.price.idx: consumer price index - monthly indicator (numeric)
- cons.conf.idx: consumer confidence index - monthly indicator (numeric)
- taiwan3m: taiwan three-month interbank offered rate indicator (numeric)
- nr.employed: number of employees - quarterly indicator (numeric)

Output variable (desired target):
- y - has the client subscribed a term deposit? (binary: 'yes', 'no')

## Question

1) Part 1 (20%). Build up a single-level statistical model using the below data set to predict the success rate with data: final 1.csv

2) Part 2 (20%). Build up an advanced statistical model (option: monster, mixture, or multilevel), with the following data set: final 1.csv

3) Part 3 (20%). Model comparison: calculate WAIC of the two models developed.

4) Part 4 (20%). Make an ensemble model by combining the two models developed.

5) Part 5 (20%). After answering the questions above, check the estimation accuracy using the following data set: final 2.csv. Specifically, use the three models developed to examine whether they make acceptable prediction performance via calculating the correct rate of these three models, respectively.