Statistics and Data Analysis, Fall 2015
Sample Final Exam

Instructor: Ling-Chieh Kung
Department of Information Management
National Taiwan University

1. The 20 numbers contained in the sheet "P1" form a random sample. The population is known to be normally distributed. The population size is 20000.

   (a) Construct a 95% confidence interval for population mean from the sample data.

2. Two groups of students took the same course taught by two different instructors. Class 1 had 12 students while class 2 had 20 students. After they took an exam at the same time, their scores were recorded in the sheet "P2" in tow columns. We consider these scores as sample data (from two populations) in evaluating the teaching effectiveness (in only one aspect, of course) of the two instructors. Let μ_i be the population mean of students' grades of this exam by taking this course from instructor i, $i = 1, 2$. Both populations are assumed to be normally distributed. The population standard deviations for the two classes are known to be 16 and 19, respectively.

   (a) For each instructor, we wonder whether the average scores of all her/his students (not limited in the sampled class) are higher than 70. Consider the following hypothesis testing at a 5% significance level:

   $$\begin{aligned} &\mathrm{H}_0 : \mu_1 = 70 \\ &\mathrm{H}_\mathrm{a} : \mu_i > 70, \end{aligned}$$

   where $i = 1, 2$. What are the p-values for classes 1 and 2?

3. The sheet "P3" contains the same data set as "P2" in a different format. Each row is the information of one student, including her/his class and scores.

   (a) Construct a regression model to use *class* to explain *scores*. Write down the regression formula, R^2, and the p-value(s) of independent variable(s). Comment on the significance of independent variable(s).

4. The sheet "P4" contains the scores of three classes of students taking the same exam. For each student, her/his gender and number of credits in the same semester are also recorded.

   Note. For the following problem, if you define your own variables, clearly define them.

   (a) Construct a regression model to use *class* to explain *scores*. Let class 1 be the reference level. Write down the regression formula, R^2, and the p-value(s) of independent variable(s). Interpret the coefficients and p-value(s) of independent variable(s).

5. Answer the following true/false questions. Please do not provide any explanation.

   (a) To construct a confidence interval, it is always required for the population to be normally distributed.