

# Statistics I, Fall 2012

## Project 1

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

**Note.** As this is a project instead of homework, the expectation on your performance is higher than that for homework. As you will see, the difficulty of the tasks in this project is not very high. So what we really care about is (1) how you design your methods and (2) how you present your ideas and results. Besides of getting correct results, try your best to make your report good, which means precise, concise, pretty, and easy to read.

## 1 Teams

Please find your team number (a capital letter from A to U) and your team mates in the sheet names “Teams” of the MS Excel file “StatFa12\_project1.xlsx”.

## 2 Tasks

1. In the MS Excel file “StatFa12\_project1.xlsx”, there is a sheet named “Task 1”. Use the data contained in the column corresponding to your team to perform the following tasks. For example, team A should use the data in column A.
  - (a) The 10000 values given to you is a set of population data, You know these values are generated from two well-known distributions introduced in Chapters 5 and 6, and your task is to identify them. An answer should include two distribution names as well as required parameters. For example, one possible answer is “an exponential distribution with rate 5 and a normal distribution with mean 15 and standard deviation 3.” Design your own way to identify the two underlying distributions, Verify your guess, and justify your method.
  - (b) In this part, visualize the sampling distribution of the sample mean with a sample size 100. You need to repeatedly sample and calculate sample means so that you may visualize the distribution. Does your result follow the central limit theorem? Explain it.
  - (c) In this part, with a sample size of 100, visualize the sampling distribution of the sample proportion of values being odd in its whole number part. For example, 15 is an odd number; 15.6 is also treated as “odd” as its whole number part (15) is odd. Does your result follow the central limit theorem? Explain it.
  - (d) In this part, with a sample size of 20 (NOT 100!), visualize the sampling distribution of the statistic

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2},$$

where  $S^2$  is the sample variance,  $\sigma^2$  is the population variance, and  $n = 20$  is the sample size. Does your result suggest that  $\chi^2$  follows a chi-square distribution? Explain it.

2. In the MS Excel file “StatFa12\_project1.xlsx”, there is a sheet named “Task 2”. Use the data contained in the column corresponding to your team to perform the following tasks.
  - (a) The 100 values given to you is a set of sample data, whose underlying population is uniformly distributed with unknown lower and upper bounds. Do you think your sample is also uniformly distributed? Why do you think so?
  - (b) Estimate the lower and upper bounds of the underlying uniform distribution of the population.  
**Note.** No one has taught you how to do this and you need to design your own way. This is where you may show your creativity.

### 3 Requirements

1. Limit your report to **six pages**. This means six single-sided sheets or three double-sided sheets.
2. Submit a **hard copy** of your report into the instructor's mailbox by **1pm, December 5th, 2012**.

### 4 Grading policy

1. All team members will get the same grades (but for Project 2 you will need to do peer review).
2. Grades breakdown:

Item	Percentage
Correctness of the results	40%
Whether the methods applied make sense	30%
Presentation and organization of the report	30%
Novelty and creativity (as a bonus)	10%

3. If your raw score is higher than 100, it will be treated as 100.