An Integrated Software for Teaching Statistics, S-LINK

Jung Jin Lee
Soong Sil University, Department of Statistics
1-1 Sangdo-dong, Dongjak-ku
Seoul, Korea 156-743
jjlee@stat.soongsil.ac.kr

Gun Seog Kang
Soong Sil University, Department of Statistics
1-1 Sangdo-dong, Dongjak-ku
Seoul, Korea 156-743
gskang@stat.soongsil.ac.kr

1. Introduction

Statistics is widely used in many research areas and the number of students who want to learn statistics has been increased rapidly. However, learning statistics is not easy for novices, especially whose background are social sciences or liberal arts rather than natural sciences and engineering. 'How to teach the complicated statistical theories efficiently' by using computer technologies has been a long issue. The traditional statistical softwares such as SAS and SPSS are designed mainly for data processing and therefore they are not so helpful to teach statistical theories.

This paper discusses an integrated software for teaching statistics, called Statistics – Learning Information Network Knowledge system(S-LINK). The S-LINK uses recent network technologies and multimedia components such as interactive visualization, animation, movies, etc. The S-LINK consists of five components; e-book, dynamic motion and lecture movie, modules on Computer Aided Teaching for Statistics(CATS), evaluation component, and data processing component for interactive experiments (Figure 1).

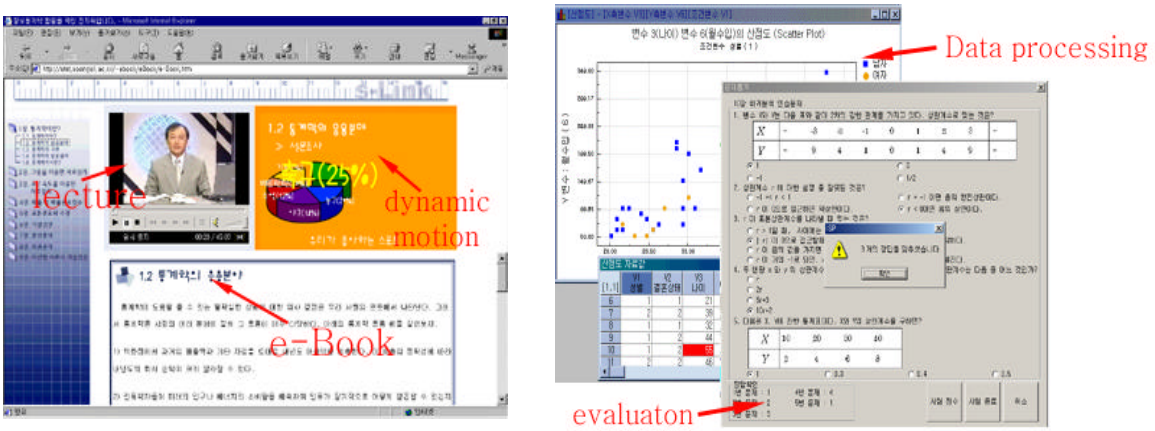

Figure 1. An integrated software system for teaching statistics, the S-LINK

The e-book component is not only a hyper-linked electronic textbook, but also has the ability to link relevant web sites of related subject. The movie component allows students to take lectures on any subject whenever they want and to watch related

dynamic motions. The CATS component includes many modules designed for teaching statistical theories by using computer which are not easy to teach in a lecture. In section 2, we focus on some of these modules. The evaluation component includes on-line test banks for self study. The data processing component includes its own data processing module like a statistical package which enables an interactive experiment what students learned from the e-book and lecture. The data processing component has its own engine that can manipulate data by spreadsheet and uses GUI environments. The S-LINK system has been developed by using Visual C++, Component Object Model(COM), FLASH, JAVA script, and Active Server Page(ASP).

2. Modules on Computer Aided Teaching for Statistics

The CATS component of the S-LINK system consists of many modules designed for teaching statistical theories that are not easy in a classroom lecture. In this section we focus on modules for teaching Probability Distributions, Sampling Distributions, Confidence Interval, Testing Hypothesis, and Regression Analysis.

2.1 Probability Distributions

Any books on statistics allocate many pages of tables for probabilities and percentile points of probability distributions. Students should spend much time to learn how to use these tables which is not easy for a novice. Also, a statistics book usually provides limited number of tables because of page limitation, for example, only 95, 97.5, 99 and 99.5 percentiles of F distribution. Therefore, it is not easy for a student to understand the shape of F distribution, to calculate 5 or 1 percentiles of F which need a formulae, and sometimes impossible to calculate p-value of an F test.

The CATS provides modules on distributions of Normal, t, Chi-square, F, Binomial, Poisson, Hypergeometric. In case of continuous distributions, given parameters of the distribution, the system provides both probabilities and percentiles of three types such as P(X < a), P(a < X < b) and P(X > b). In case of discrete distributions, the system provides three probabilities such as P(X <= a), P(X = a) and P(X >=a) simultaneously (Figure 2).

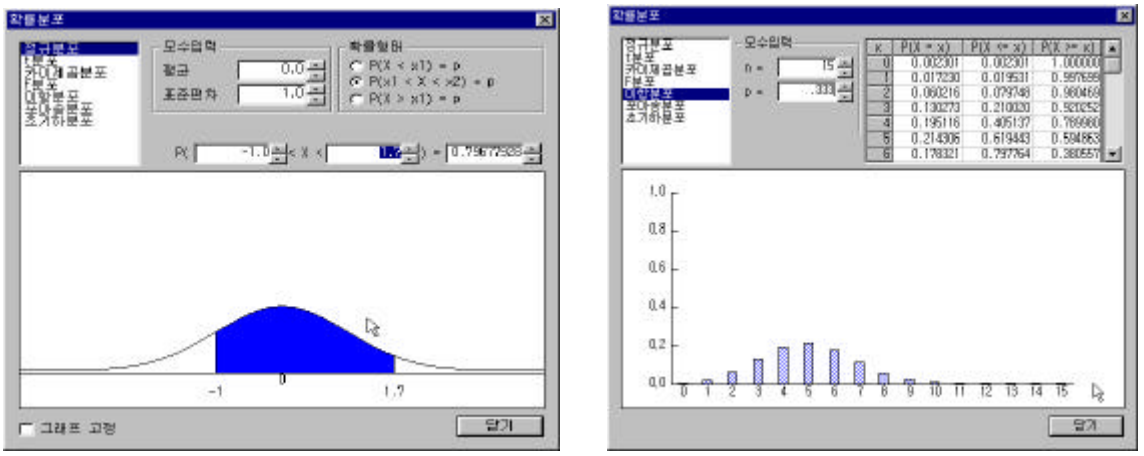

Figure 2: Calculating probabilities and percentiles of Normal and Binomial distributions

2.2 Sampling Distributions

One of the most difficult theories in statistics for a novice would be the sampling distributions including the Central Limit Theorem. The CATS provides a tool for experimenting the sampling distributions of sample means, sample variances, and sample proportions. Students are able to check how the sampling distribution of sample means changes depending on population distributions and sample sizes. If a student selects a population distribution in the system, provides the parameters of the distribution, population size, three different sample sizes, and the number of repetitions of sampling, then the system shows sampling distributions for each of three sample sizes (Figure 3).

2.3 Confidence Interval

Students often confuse the meaning of confidence interval for a population mean. In order to understand the meaning of 95% confidence interval, a student provides parameters of a normal population with population size, sample size, and number of repetitions, and then the CATS shows a simulation experiment of the repeated confidence intervals. Student can check how many of the confidence intervals include actual population mean (Figure 3).

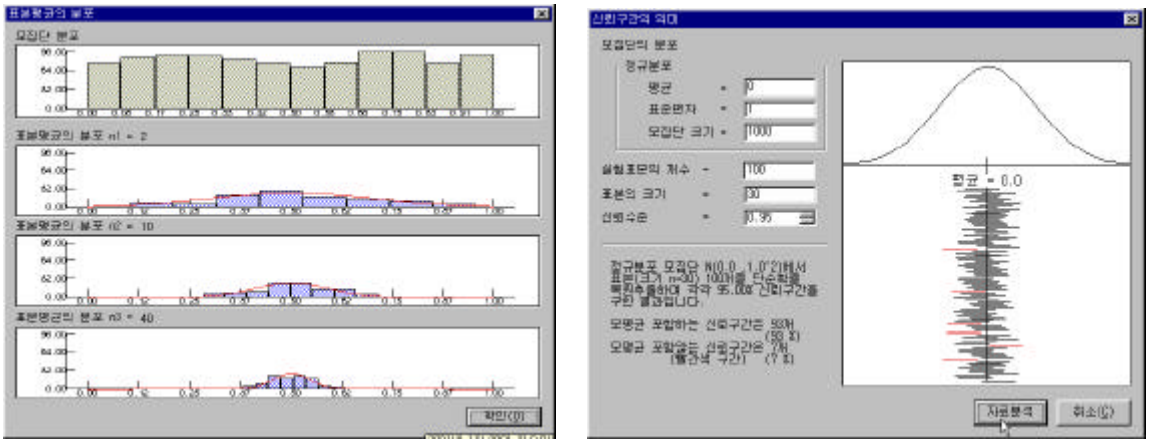

Figure 3. Experiment for sampling distribution and confidence interval

2.4 Testing Hypothesis

Testing hypothesis is also one of the most difficult theories in statistics for a novice student. In order to understand the meaning of testing hypothesis, the CATS draws sampling distributions under each hypothesis and shows location of sample statistics on the sampling distribution. Therefore a student can easily understand the meaning of significance level, p-value, type I error, and type II error of a test (Figure 4).

2.5. Regression Analysis

We can teach theories of regression analysis in a classroom, but it is not easy to show on a scatter plot how the estimated regression line looks like and how an outlier can influence on the result of regression analysis, for example, R^2 value. The CATS system can change any point in a scatter plot by dragging mouse on the point, and a student can easily check the effect of single outlier in a regression analysis (Figure 4).

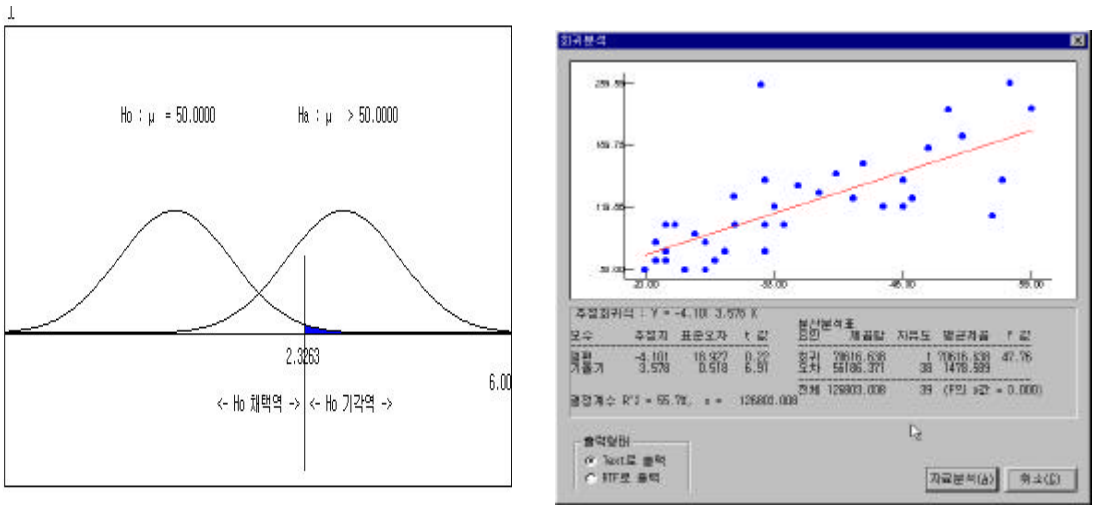

Figure 4. Experiments for testing hypothesis and regression analysis

3. Discussion

This paper introduced the S-LINK system which includes several ideas to teach and to practice the complicated statistical theories for a novice. The S-LINK also includes a hyper-linked textbook, multimedia components, and interactive experiment as traditional statistical software in order to help teaching/learning statistics. Developing a new statistical software that uses all advanced information technologies is not an easy task that might need international cooperation.

REFERENCE

Danan, Brianean, and Matthewco. (2000): *Asp 3.0 Programmer's Reference*, Wrox Pr Inc.

Davison, Kirsty. (1997): statwise. Multimedia Statistics Courseware. *Maths & Stats* 8, <http://www.stats.gla.ac.uk/cti/activities/reiviews/97_08/statwise.html>.

Lee, J. and Kang, G. (2000): *Statistical Information Analysis by Using ISP*, Freedom Academy Inc.

Mittag, H. (2000): Learning and Teaching Statistics in a Networked World. Proceedings of the 68th Conference of Japan Statistical Society, p269.

West, R., Ogden, R., and Rossini, A. (1998): Statistical Tools on the World Wide Web, *The American Statistician*, **3**, 257-262

RESUME

Jung Jin Lee and Gun Seog Kang are professors of statistics at the Soong Sil University in Seoul, Korea. Lee received his Ph.D. in operations research from the Case Western Reserve University and was a professor at the University of Toledo and Mississippi State University in USA. Lee worked on Tantalus Inc and was a visiting scholar of IBM Watson Research Center, Rutgers University, and the Institute of Statistical Mathematics in Japan. Kang received his Ph.D. in Statistics from the University of Wisconsin-Madison and was a researcher at the North Carolina State University.