A Trial of Statistical Education using Sports Data in Japan

Yamaguchi, Kazunori
Rikkyo University, College of Business
3-34-1 Nishi-Ikebukuro Toshima-ku Tokyo, 171-8501, Japan
E-mail: kyamagu@rikkyo.ac.jp

Sakaori, Fumitake
Chuo University, Department of Mathematics
1-13-27 Kasuga Bunkyo-ku Tokyo, 112-8551, Japan
E-mail: sakaori@math.chuo-u.ac.jp

Watanabe, Michiko
Toyo University, Faculty of Economics
5-28-20 Hakusan Bunkyo-ku Tokyo, 112-8606, Japan
E-mail: watanabe_michiko@nifty.com

1. Introduction

The establishment of computer networks has brought about an age of substantial information disclosure, in which various branches of knowledge and types of technologies that read and utilize statistical data are becoming increasingly important. In university education, regardless of the field—natural sciences or social sciences—the knowledge of statistics is essential in conducting experimental studies. For this reason, various faculties and departments in many universities adopt basic courses related to statistics.

In basic statistics education, where it is expected that positive analyses will be conducted, a practical type of education in which each student will be capable of using quantitative analytical skills is being increasingly regarded as highly desirable. A survey was conducted by Takeda (1995) and Senuma (2004) in Japan to determine what students were expected to study through mathematical studies at universities; this survey was conducted on all the companies listed in the Tokyo Stock Exchange. The results revealed that statistical education, which enables students to use data substantially, is regarded as highly desirable.

On the other hand, numerous statistics teachers in social science departments are of the opinion that students, in general, are hesitant to study the type of statistics that emphasizes mathematical aspects. Course materials utilizing the Internet and other multimedia resources have recently been developed and put to practical use in university education. Multimedia materials emphasize audio and visuals that can be interactively operated and verified. It is hoped that the use of multimedia will positively affect university education; however, no concrete lecture form that will create that positive effect has been standardized in the field of statistics. One of the possible reasons for this failure is that most of the syllabuses that are publicly available are developed in text form and are not based on Internet awareness or the course materials being converted into multimedia formats.

For the above reasons, we formed the “statistics” group from the social science-related fields within the Cyber Campus Consortium (CCC), which was developed at the Japan Universities Association for Computer Education. Further, we constructed a next-generation computerized syllabus system in order to initiate the establishment and use of databases for common development and so that multimedia course materials can be used in a concrete form for the purpose of standardizing the methodology of effective statistics education utilizing IT at universities (Watanabe and Yamaguchi 2006).

In a statistical education, it is needless to say that the motivation improvement to the student's study is important for the learning effect. It is also important that students have the analysis experiences by actual data in the study of statistics. It was not fictitious data, and the analysis practice that used actual data was done repeatedly by such a viewpoint in Rikkyo University. The lecture and the analysis practice has used

several real data sets including sports data in 2006, convenience store data collected though the point of sale, POS, as well as the social survey data.. This paper reports on educational trials using the sports data.

2. Sports data

Our group has used data sets which were provided by a company which provides data of several types of professional sports, to media and professional teams. Students could use real data sets from professional sports, baseball games and football matches in Japan. Almost all students were interested in such sports data. In Japan, baseball is one of the favorite sports and football is also very popular.

We introduce two examples we used in the basic statistics classes.

2.1 Baseball data: pitching data to study basic statistics and distributions

In the basic statistics course, one of the most important objectives is that students can understand the variations. Pitchers throw many types of balls. We can use two types of data. One is types of pitching balls, and another is speed of the balls, these are categorical and continuous data, respectively. Using these data sets, students learned how to summarize variations of data. Students used frequency tables, histograms, box plots and so on.

Figure 1 shows a distribution of speed of pitches by Daisuke Matsuzaka who is one of the most famous professional baseball players. He is playing for Boston Red Sox this year.

The distribution is multimodal, because the data includes six types of balls, for example, fastball, breaking ball, cut ball and so on. The student learnt the necessity for the classification. They considered what the overall mean represents in this case, and concluded that the overall mean did not have any role of presentation for the pitcher's skill.

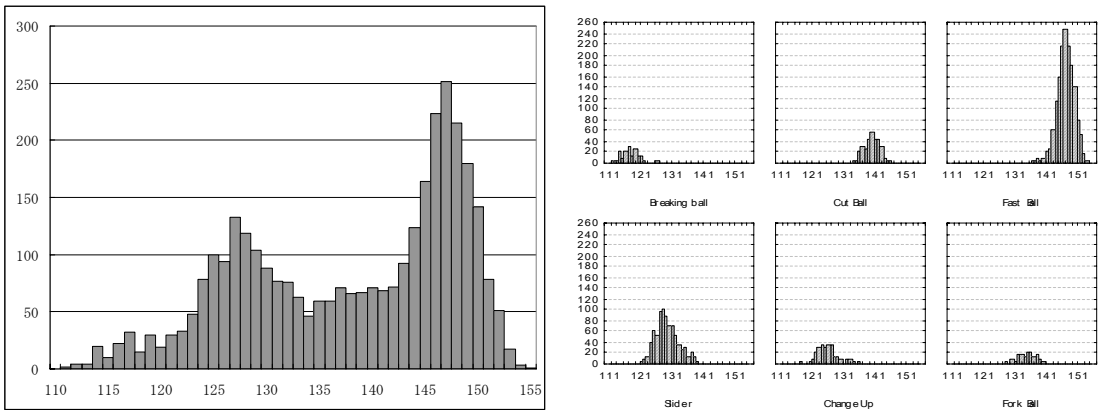

Figure 1 : A Histogram of the pitch speed (km/h) (left: all data, right: after classification)

Table 1: Mean and Standard deviation of each type of pitch

Type	N	mean	SD
Fastball	1560	147.1	2.8
Slider	814	129.1	3.7
Cutball	339	140.0	2.6
Breaking	172	117.7	3.1
Cange Up	281	126.0	3.8
Forkball	155	134.8	3.1
Total	3321	138.1	10.1

Table 1 shows the mean and standard deviation of each ball. Using this data set, students also studied the normal distribution, mean, standard deviation, and model fits. Mixture models were also introduced.

Figure 2 shows a result of fitting a mixture model of two normal distributions to the distribution of pitch speed of sliders.

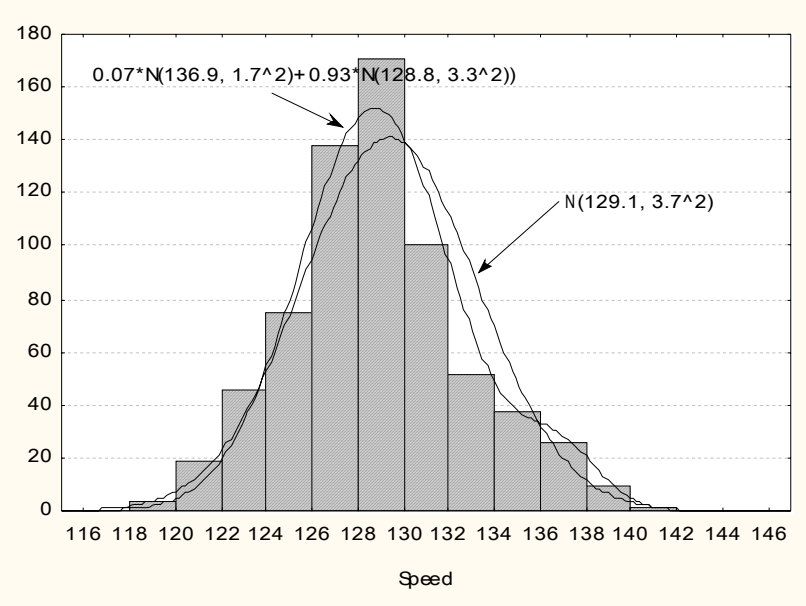

Figure 2: A Histogram of the pitch speed (km/h) of sliders and two fitted theoretical distributions($0.07N(136.9,\ 1.7^2) + 0.93N(128.8,\ 3.3^2)$ ***and*** $N(129.1,\ 3.7^2)$ ***).***

2.2 Baseball data: numbers of homerun and strike out to study correlation and causation

From baseball games, we could get several types of data. There are many good examples for the multivariate analysis. For example, the regression, principal component, and factor analysis, etc.

We usually observe positive correlation between the numbers of homerun and strikeouts (see the Figure 3.). Student can easily understand that correlation does not necessarily mean causation. Student can easily imagine the third variable in this case (Figure 4).

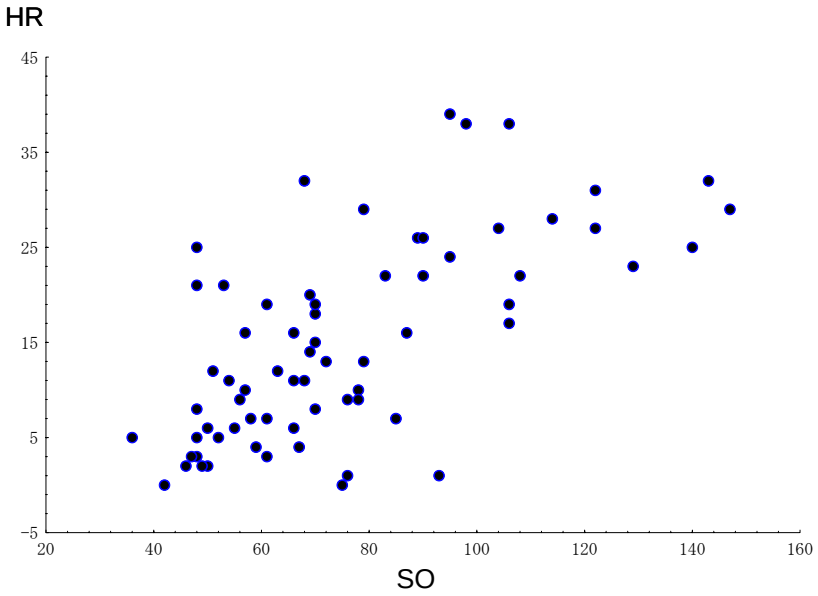

Figure 3: A scatter plot of the numbers of strike out and homerun

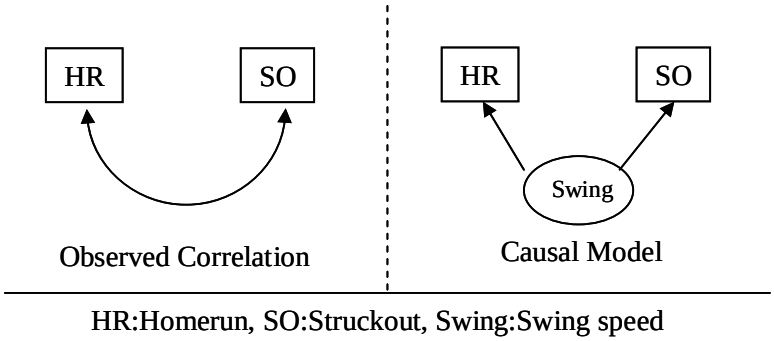

Figure 4: A path diagram for the numbers of strike out and homerun

3. Discussions

Seven important topics in basic statistics education have been introduced by Utts (2003), "Cause and effect", "Significance versus importance", "No effect versus low power", "Biases in surveys/questions", "Probable coincidences", "Confusion of the inverse", and "Average versus normal". The final example is ~~a~~ material for "Cause and effect". Students can understand "Average versus normal" by distributions of pitching speed. Our project group is investigating good materials from sports data ~~to~~ for teaching such topics. Data sets from football games were also used. For example, numbers of goals, yellow or red cards are useful for studying the Poisson distribution.

Other examples we have used will be shown in our presentation.

REFERENCES

Utts, J. (2003). What educated citizens should know about statistics and probability, The American Statistician, 57, 74-79.

Takeda. (1995) The Demand Level for Mathematical Education from Company Viewpoints, Journal of College and University Division of Japan Society of Mathematical Education, 2, 81–94. (in Japanese)

Senuma. (2004) Companies' Expectations for Arithmetic and Mathematical Education–Emphasis on Predictions Based on Data and Theoretical Thinking Power, and Improvement on Teaching Methods), Association of Science Education, 28, 34–42. (in Japanese)

Watanabe, M. and Yamaguchi, K. (2006) On demand statistics courses using new technologies in Japan, Proceedings of ICOT 7.