Applications of *Excel* in teaching *Statistics*

Yang Shuqin
Department of Mathematics
Dalian Maritime University
Dalian
Liaoning Province 116026
People's Republic of China

This article was first published in 'The China Papers' Issue 5 and is reprinted here with permission from the author.

Abstract

The aim of this paper is to describe the teaching of *Statistics* at Dalian Maritime University. Problems associated with traditional teaching activities are identified and a new teaching strategy using *Excel* is proposed. Several examples are also introduced to illustrate the usefulness of *Excel* in *Statistics* teaching.

Introduction

For many years, I have been involved with the teaching of the course, *Probability Theory and Statistics* at the Dalian Maritime University. I have received a lot of feedback relating to the course from my students. Many of students think that the subject is important and useful for their future job, but it is difficult to learn and to understand. Some students totally lack interest in this subject. Even for excellent students, it is difficult for them to apply the knowledge they have learned in lectures to the practical problems. I often asked myself one question: 'why do so many students think that the *Probability Theory and Statistics* is difficult?' I have tried many methods to make the lectures easier and more effective, but the results have not been satisfactory. Finally I attributed this to the numbers of students in class and gave up.

For four months in 2005, I was part of the *Teaching Science in English* program held at the University of Sydney. It is a collaborative project between The University of Sydney and the China Scholarship Council. After attending the program, I have come to realise that the traditional methods used in the teaching activities and associated teaching materials like textbooks used for this course may be the reasons behind the lack of success in the teaching of Statistics in my university. To make statistics teaching more effective, some new teaching strategies have to be explored.

This paper is organised as follows. Firstly, I will outline the present situation regarding statistics teaching in our university, highlighting the problems existing in the traditional teaching activities. Then I will justify the need to introduce a new course, Introduction to Statistics for the first year students. Finally, I will explain how to use modern teaching strategies in the teaching of Statistics and how to use Microsoft Excel to make the subject less difficult and more appealing.

Present situations

At Dalian Maritime University, *Probability Theory and Statistics* is a compulsory course for second year students majoring in many diverse subjects such as Business, Computer Science and Technology, Environmental Science and Engineering, Information Technology, Transportation Management, etc. The course consists of 54 lectures (each lecture being 45 minutes) in one semester. Approximately 24 lectures are allocated for Probability theory, 18 hours for Statistics, 10 lectures for tutorials and 2 hours for the final examination. The method used in teaching is mainly teacher centred and lecture based: lecturers delivering knowledge while students watch, listen and take notes. If students require help, they may visit the mathematical teacher's office on any weekday afternoon.

Existing problems

Problems exist with this traditional teacher-centred teaching method. In each lecture, the teacher talks from beginning to end. There is no time for students to ask questions and to discuss the material with each other. Learning is based heavily on lectures and textbook learning with no laboratory work, teamwork, group discussions, or any complementary resources. The assessment is summative rather than formative. There is no feedback to or from students. The students don't know how much they have understood or what they have mastered. Also

teachers don't know how their students are progressing. On the student side, the students have many compulsory courses to learn, there is no time to review what they have studied in the lectures. To address these problems, we should change the teacher-centred learning into student-centred learning. Problem based learning (PBL) and case studies are important strategies in achieving student-centred learning.

Student-centred teaching

The major theoretical development in teaching methodology is a greater awareness of, and a move towards, the constructivist perspective and view of how people learn science, and student-centred approaches and student active science is compatible with the constructivist view of learning (Peat 2005). Learning should be active, integrated, cumulative, and connected. The teacher's role should be supportive, not directive. Teachers should act as facilitators, providing resources, guidance and instruction to learners (King 2005). The low efficiency in our teaching is the result of a lack of student-centred methodologies. Our teaching should be changed to improve the quality of teaching. The use of Microsoft Excel in Statistics teaching may be an appropriate method in achieving the desired outcomes mentioned above.

The statistics functions in *Excel* may be used in descriptive techniques, graphs, numerical summaries, estimation, hypothesis testing, and regression. These functions may also be used to explain most of the concepts in *Statistics*. More importantly, it should not be so difficulty for the first year students. After I return to China, I will suggest a new optional course *Introduction to Statistics* for first year students in any major. There are two purposes to suggesting this course. For the students majoring in engineering and science, it will be an introduction to *Probability Theory and Statistics*. Because many Chinese senior middle schools students do not have concepts about probability theory and statistics except limited ideas about arrays and combination, they often experience confusion when studying the probability theory and statistics. This course will give students a visual, intuitional interpretation of probability and statistics and help them to understand the later course *Probability Theory and Statistics*. For the students majoring in Business, Humanities and Social science, the course will provide them with the concepts of Statistics and techniques to calculate the statistical data and sketch them geometrically so that they can use Statistics as a tool in the future job.

Methodology

Problem based learning (PBL)

Census results can be used as a problem at the beginning of the course. There are many statistics, graphs and tables that result from a census and they have corresponding explanations in Probability theory and Statistics. Why does the census have this particular result? How is the result approached? What do the results mean? These questions can make student active learners in the course. By the end of the course, the students should try to obtain the results themselves. When they can calculate the numerals, make the graphs and tables, and use them to explain real world phenomena, then they will become more and more interested in learning the more difficult content and concepts. Successes can help students build up confidence in learning

Case studies

A great deal of material can be used in a case study. The following is an example for studying correlation and the regression line. A developmental psychologist believes that the age at which a normal child begins to speak words clearly is closely related to the age at which the child first begins to use complete sentences. A random sample of 33 normal children was taken. Let X be the age at which words are first clearly used, and Y be the age at which complete sentences are used. The data below give the values of X and Y in months. We can design some questions for the students to answer in the following lectures. (a) What is the correlation between X and Y? (b) Work out the true regression equation for predicting Y from X. (c) Suppose that the joint distribution of X and Y values is bivariate normal in form. Find the 95 percent confidence interval for the predicted Y age of a child by use of the regression equation given the X age of 13.0, 14.6, 12.2[3].

X	15.1	12.7	11.7	13.1	13.0	11.2	13.3	…..
Y	25.2	24.3	22.1	23.3	24.1	23.6	25.5	…..

After solving these problems the students will understand the concepts deeply.

Using *Excel* to teach the concepts

Excel is an appropriate and effective tool to assist first year students in learning Statistics. We can use Excel worksheet to remember and calculate the statistical numerals.

Figure 1 is an *Excel* worksheet. In fact, it has 21 columns but many columns were hidden to accommodate the size of the paper. I calculated the values for X, Y, XY, X^2, Y^2, the probability distributions of X and Y, the averages of X and Y, their standard deviations and their correlation etc. When a student clicks a cell with the mouse, the formula from which the value is calculated will appear on the function bar. If the student changes some values of X or Y in any cells, every cell will change and the graphs will reflect the changes of these cells. After some self learning and by following the instructions given by lecturers with different files, the students can manipulate the Excel worksheet and produce the correct answers to the problems they meet.

In the similar method, we can illustrate theorems. For example, when we teach the Central Limit Theorem we can use *Excel* to generate n variables X1, X2,……Xn with Bernuli distribution and same parameters. And we can get the value of the sum $\sum_{i=1}^{n} Xi$. By repeating the process, we can get the distribution of the sum. Comparing the distributions of the sum when n is different, we notice that when n approaches large numbers, the distribution of the sum will approach to the Normal distribution.

Microsoft Excel - STAT CONCEPTS

	A	B	F	G	I	J	K	L	M	N	Q	T	U
1	X	probability	Y	Probabilyty	Y^2	X^2	mean	39.12	mean	65.62	X*Y	Sxx	16492.99
2	35	0.1	7	0.06	49	1225	variance	756.7456	variance	5187.976	245	Sxy	22167.56
3	12	0.02	29	0.01	841	144	sd1	27.50901	sd2	72.0276	348	Syy	62600.73
4	2	0.05	1	0.1	1	4					2	COV(X,Y)	1563.207
5	69	0.3	46	0.2	2116	4761					3174	CORREL(X,Y)	0.776875
6	28	0.08	35	0.05	1225	784					980		
7	8	0.2	12	0.12	144	64					96		
8	62	0.03	178	0.2	31684	3844					11036		
9	30	0.1	34	0.08	1156	900					1020		
10	12	0.06	75	0.04	5625	144					900		
11	5	0	15	0.1	225	25					75		
12	86	0.06	240	0.04	57600	7396					20640		

Figure 1. The Probability Distribution and the Statistics Numerals

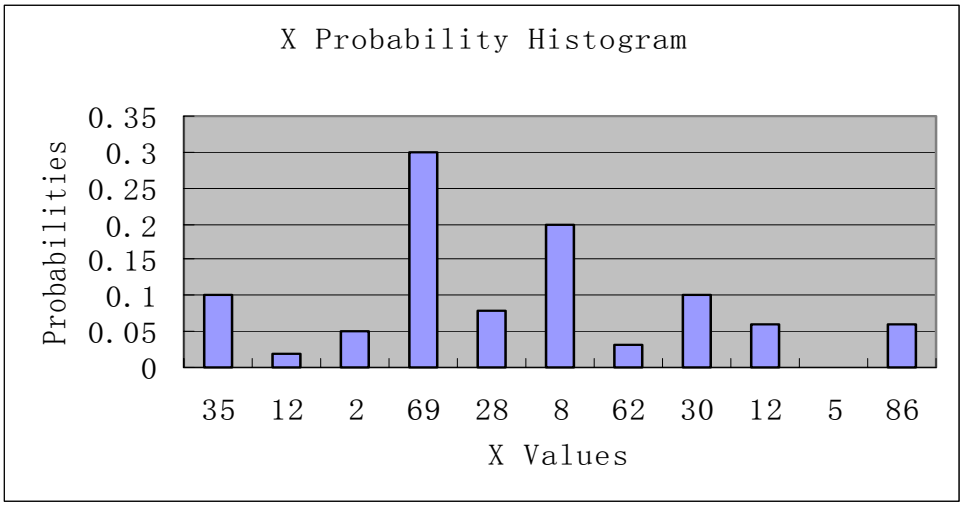


Figure 2. The Probability Distribution of X

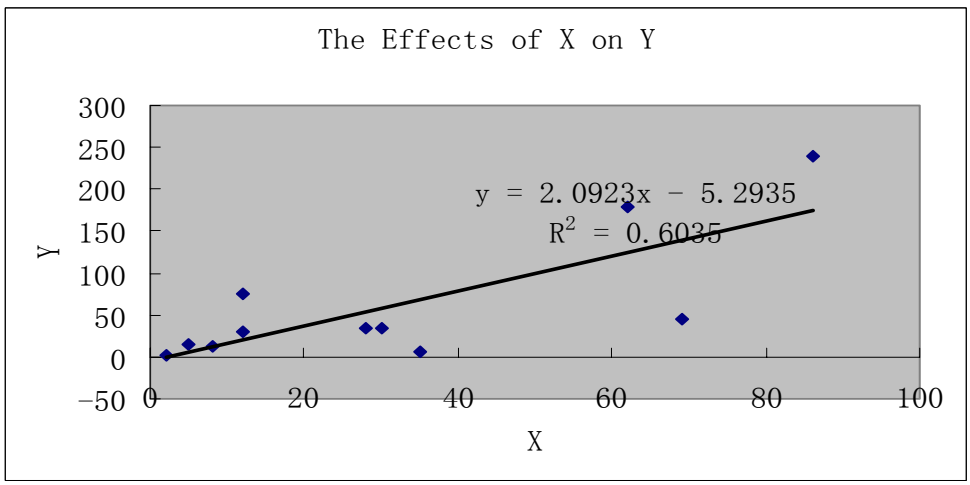
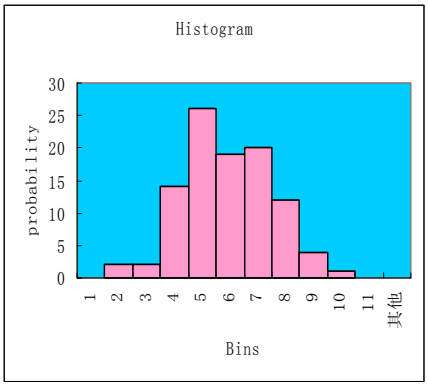
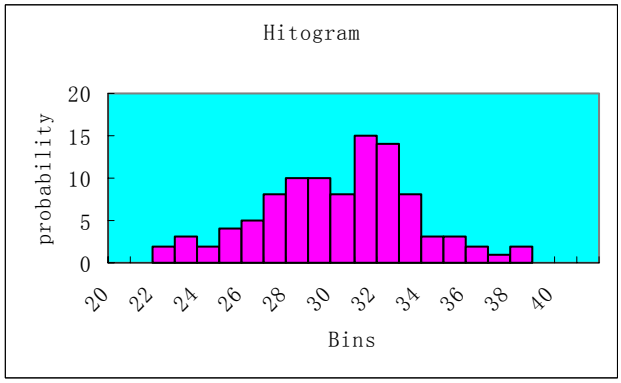


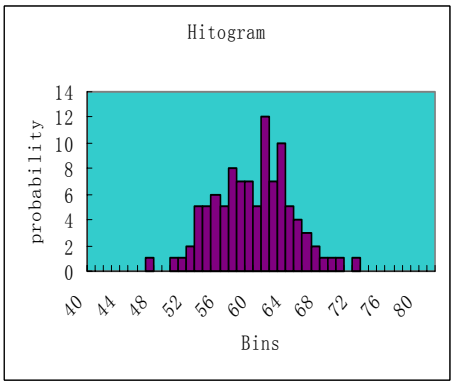
Figure 3. The Least Squares (Regression) Line of X and Y



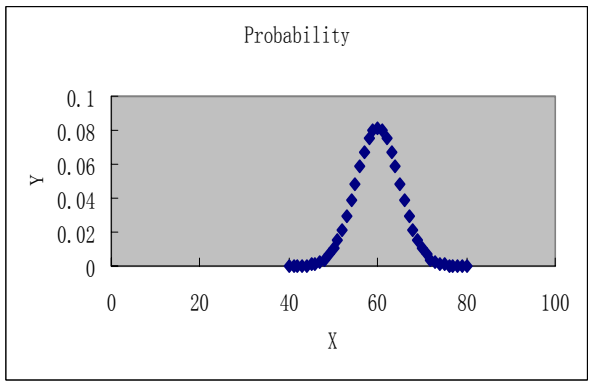
N=10



n=50



N=100



The Normal distribution N(60,24)

Conclusions

There are some difficulties in using *Excel* in teaching *Statistics*. Some *Excel* files used in the teaching should be prepared in advance to lead students to learning the method step by step. There may not be sufficient computers for this course due to financial problems and the numbers of students. But everything is hard to begin with. I will remember what I learned here and I will do my best to bring about reform in the teaching of Statistics in my university.

Acknowledgements

I wish to offer my thanks to the China Scholarship Council and University of Sydney, who provide us the opportunity to be part of this program. I would also like to thank Mike, Mary, Cecilia, Nicole, Kathy and other teachers for their excellent lectures. I am especially grateful to Dr.Howard D'Abrera. I was able to attend all of his lectures this semester. I would also like to thank Rubin Zhang, Qiying Wang and Lloyd for their direction in preparing this paper. I also express my thanks to my classmates for their help in my life and my career.

References

D'Abera, H. (2005) Excel Notes for LABS. The University of Sydney.

Hays, W. (1981) *Statistics* (3rd edition). Holt Rinehart and Winston.

King, M. (2005) Teaching Science in English, The University of Sydney.

Peat, M. (2005) Teaching Science in English, The University of Sydney.