TEACHING INTRODUCTORY STATISTICS : MORE DATA, LESS LECTURING

Abdul Razak Saleh, Ph.D.
Email: razak@webmail.uum.edu.my
Azizah Ahmad
Email: azie@webmail.uum.edu.my
School of Information Technology
Universiti Utara Malaysia
06010 UUM Sintok
Malaysia

ABSTRACT

Study conducted at the Northern University of Malaysia (UUM) showed that the performance of the students in the Introductory Statistics course (TS 1013) were below expectation. All students in TS 1013 are required to take a 40-question Statistics Readiness Test (SRT) during the first week of class. The test consist of multiple choice question constructed to test the knowledge of basic algebra and mathematical reasoning. The mean score is 64.7%. About 40% of the students score less than 60%. It was found that poor background in mathematics is a barrier for students to score a good result in Introductory Statistics course. This study suggested that students with poor mathematics background should be exposed with more data but less lecture. This task was accomplished by using SAS (Statistical Analysis System) software in teaching the course. The SAS system is a software system for data analysis and report writing. The system is available on many computer system such as mainframes, minicomputers, and personal computers under different kind of operating systems. With base SAS software you can store data values and retrieve them, modify data, compute simple statistics, and create reports all in one SAS session. The Introductory Statistics course is redesigned by incorporating an appropriate SAS procedure in TS 1013.

Keywords: Teaching Introductory Statistics Course, Statistical Package, Introductory Statistics

INTRODUCTION

Universiti Utara Malaysia (UUM) was founded in 16 February 1984 and is the sixth university in Malaysia. The University was established for the pursuit and development of management education. Its academic programmes focus on such disciplines as management, accountancy, economics, information technology, public administration and human resource development. Currently UUM has 8 schools that offered 21 different degree programmes. So far UUM has produced 15,316 graduates and currently there are about 15,834 students undertaking their undergraduate studies.

All UUM students are required to successfully completed the university core courses which comprise of humanity, language, co-curriculum, quantitative, accounting and management. The quantitative courses however, specifically Introductory Statistics (course code : TS1013) is considered extremely difficult by most students. Almost every semester the rate of failure for this course is considered high as compared to other courses. There are several factors that can contribute to the poor performance of these students. One of them is due to the fact that most of these students have very little mathematical background. The size of the class also plays an important aspect for these problem where most of these classes are big. Student's workload is another factor that can contribute to this problem where students have to carry up to 23 credit hours per semester in order to complete their 3-year degree programme. Most of these students are taking at least 18 credit hours per semester (Saleh, 2000).

Courses registration at UUM is done through a random method system where students are supposed to do a pre-registration for the following semester one week before their final exam at current semester. During the pre-registration, students are only required to list down all the courses to be enrolled next semester. The time slots and lecturers are unknown to them at this time. A computer system which uses a random method time tabling will generate classes and lecturers accordingly. The TS1013 class normally consists of an average of 100 students. The dasses are conducted using conventional method of classroom teaching where assessment is divided into 3 parts including class assignments, two one-hour examinations and final examination. Final exam would be 50% of the overall assessment.

LITERATURE REVIEW

Is method of teaching elementary statistics courses now is the best approach that can be offered to students? Current method is more towards exposing students to heavy mathematical concept including statistical formulas and equations, which are very hard for the students to relate to the actual real world problems. According to some statisticians, this method of teaching has to undergo some changes from currently using more towards mathematical concept to data analysis (Hogg, 1991). Many of them have agreed to introduce a practical way of data gathering. Bradstreet (1996) introduced courses based on practical and Dietz (1993) assigned projects to his students to be completed in 3 weeks. Smith (1998) has made some modification to the Elementary Statistics course by introducing new technique and learning style using assignments and projects. Besides written reports, students are also required to present their works. With the new method, it is reported that students' achievements have increased tremendously and students seemed to give positive feedback on the new approach.

Steinhorst and Keeler (1995) suggested the method of teaching Elementary Statistics course has to include active learning of statistical concept by exposing students to real live examples related to the topics. It took them about 4 years to design and gather enough materials such as classes' contents, exercises and test items. Teaching method was conducted in several ways including classroom teaching, laboratories and group activities. The purpose of the method was to build a concept of understanding statistics. The study reveals that students' interest and performance were increased.

Giraud (1997) made a study on two groups of students undertaking Elementary Statistics course using classroom method and cooperative learning method. The first group, which consisted of 44 students, was introduced to the cooperative learning method and the other group of 51 students was having a conventional classroom method. There was no significance different between the two groups in terms of gender composition, CGPA, number of learning hour per week and Statistics Readiness Test (SRT) score. Giraud concluded that students' performances in the cooperative learning are better compared the other group of classroom method.

The SAS system is a software system for data analysis and report writing. The system is available on many computer systems such as mainframes, minicomputers, and personal computers under different kind of operating systems. SAS software has developed over the years and the current capabilities have gone beyond the original function of solving statistical problems. SAS users can store data values and retrieve them, modify data, compute simple statistics, and create reports all in one SAS session (Schlotzhauer & Littell, 1991). Other functions of SAS include data management, reports production in various formats such as text or graphics, project management, programming language, file management and forecasting. Users can choose either to use menu driven option or to do their own programming in order to use SAS. For more detail on SAS software, please refer to Dilorio and Hardy (1996), Elliott (1995) and SAS Institute Inc. (1992).

OBJECTIVES

An exploratory survey was carried out to have a general idea of problems faced by UUM students regarding their poor performance in Introductory Statistics course. The specific objectives are:

- To define a relationship between student's mathematical competency background and student's achievement in the Introductory Statistics course
- To propose the incorporation of statistical package in statistics courses

METHODOLOGY

Population and sample

A population for this study consists of all 2100 students undertaken TS1013 course for semester November Session 1999/2000. This study considered a sub-population of 600 students taken from one big class out of the whole population. This particular class is handled by 3 lecturers who took turns to deliver their lectures according to the topics. Samples of 190 students from this sub-population are taken using a systematic sampling method.

Instrument

During the first week of the semester, the respondents were asked to fill in their personal particulars in the biodata form such as race, gender, UUM's entry academic qualification, mathematics' achievement in Malaysian Certificate of Education (SPM), SPM's field of study, current Cumulative Grade Point Average (CGPA), number of accumulated credit hours and current undertaking credit hours. Respondents were also been tested on the 40 questions Statistics Readiness Test (SRT) during this one week period. The test consists of multiple-choice questions constructed to test knowledge of basic algebra and mathematical reasoning. They were also been asked to make a log consisting their preparation towards achieving a good grade in this particular course including date, period, topic and study methods.

Course assessments are based on class assignments (10%), 2 one-hour examinations (40%) and final examination (50%). However for the purpose of this paper, only the first one-hour examination (will be called Test 1), which was held on 27 June 2000, will be reported.

RESULTS AND DISCUSSIONS

The majority of students involved in the study are Malaysian Higher School Certificate (STPM) holders (76.3%) while the rest are from the matriculation programmes (22.5%) and diploma holders (1.2%). More than half of them are female students (69.8%).

The performance of TS1013 students in Statistics Readiness Test (SRT)

The mean score for the SRT performance is 64.5% with variance of 56.7%. As shown in Table 1, more than 40% of these students have grade C and below. It means that they have scored less than 60 points from a 100-points test which is considered not competence enough to enroll in any statistics courses (please refer to Appendix A for detail description of the grading system). Therefore, it clearly shows that their mathematical competency background is not sufficient.

Table 1: Descriptive Statistics for SRT performance result.

Grade	Frequency	Percentage
A	48	28.4
B	53	31.4
C	25	14.7
D	23	13.1
F	21	12.4

Relationship Between SRT Score and Test 1 Score/ Gender and Test 1 Score

The correlation coefficient between SRT score and Test 1 score is 0.5747. The test showed that the relationship between SRT and Test 1 is very significant (p-value = 0.0001). However there is no significance relationship between gender of students and Test1 performance (p-value = 0.704).

Students' Achievement Based On The First Test

Test 1 assessments are divided into 3 categories of questions, which would test the students' understanding on statistical theory, calculation as well as discussion. TS1013 students are supposed to learn and understand the basic theory of statistics in this course and they should be able to interpret the result accordingly.

Table 2: Descriptive Statistics for Test 1 achievement result.

Test 1 Category	Percentage (%)					
	A	B	C	D	F	Mean (%)
Theory	34.3	36.7	18.3	3.6	7.1	69.43
Calculation	30.8	24.9	8.3	11.2	24.9	61.39
Discussion	8.9	14.8	13.0	14.2	49.1	42.18

Table 2 shows that the mean score for the discussion is the lowest among the three categories. More than 40% of these students had an F grade and less than 10% of them had grade A. However mean scores for theory and calculation categories doesn't seem to have so much difference in value.

ANOVA test was done to test the null hypotheses that all means are equal. The result of the test was to reject the null hypotheses (p-value = 0.0001). A multiple comparison method such as Tukey's Studentized Range (HSD), Waller-Duncan K-ratio T-test and Bonferroni (Dunn) T-test was used to compare the means. All test concluded that these 3 means are significantly difference.

The test results revealed that these students were actually having problem in understanding and interpreting the statistics result. They need more time to explain the meaning of values or figures obtained from the formulas or the equations taught during the class. They should be concentrating more on interpreting results and spend less time on learning theories and calculations. The use of statistical packages would be very beneficial, as it would reduce time taken to do mass calculation especially involving big volume of data.

RECOMMENDATION

This study has identified the weaknesses in the concept of teaching Introductory Statistics for students with lower mathematical background. It revealed the method of exposing students with heavy mathematical calculation and theories doesn't help them to better understand the statistical concept. Therefore, it is suggested that a new approach of using more data would be used in the course. By analyzing more data will help students to better understand the meaning of statistical concepts covered in the syllabus with the help of software packages. Testing for the normality of data set for example, would require sufficient amount of data in order to produce a more presentable normal curve. Since it would normally involve a high volume of data, effort and time needed are also high. This task can only be simplified with the help of a good statistical package.

Analyzing and presenting output can be so much easier and faster using software packages available of the shelf in the market. A normality testing case mentioned above can only be accomplished in three simple lines of Statistical Analysis System (SAS) coding as shown in the example:

```
Proc univariate data=sasuser.chiangmai plot;
   Var uji1;
Run;
```

Appendix B shows the detail output as a result after executing the above source code. The system has calculated or plot the mean, variance, standard deviation, coefficient of variance (CV), quartiles, extreme

values, stem and leaf plot, box plot and normal probability plot of the given set of data in less than 10 seconds. These tasks are actually covering 21 % of the overall TS1013 course syllabus.

The usage of software packages is very helpful and can save a lot student's learning time. Therefore they can concentrate more on understanding the theories and interpreting the results instead. In conclusion, incorporating statistical packages in the teaching of statistical courses is highly recommended in addition to theories in TS1013 course.

REFERENCES

Bradstreet, T.E. (1996). "Teaching Introductory Statistics Course So That Nonstatisticians Experience Statistical Reasoning". The American Statistician, (50): 69-78.

Dietz, E.J. (1993). "A Cooperative Learning Activity on Methods of Selecting a Sample". The American Statistician, (47): 104-108.

Dilorio, F.C. and Kenneth A. Hardy (1996). *Quick Start to Data Anlysis with SAS,* Belmont: Wadsworth Publishing Company.

Elliott, R.J. (1995). *Learning SAS in The Computer Lab,* Belmont: Wadsworth Publishing Company.

Giraud, G. (1997). "Cooperative Learning and Statistics Instruction". Journal of Statistics Education, 5(3).

Hogg, R.V. (1991). "Statistical Education: Improvements Are Badly Needed". The American Statistician, (45): 342-343.

Saleh, A.R. (2000). "Teaching and Learning Statistics in the New Millenium : A Case Study at Universiti Utara Malaysia" Paper presented at National Seminar on General Study in the New Millenium, National University of Malaysia.

SAS Institute Inc. (1992). *Doing More With SAS/Assist Software, Version 6, First Edition, North Carolina.*

Schlotzhauer, S.D. and Ramon C. Littell (1987). *SAS System for Elementary Statistical Analysis*, North Carolina: SAS Institute Inc.

Steinhorst, R.K. and Keeler, C.M. (1995). "Developing Material for Introductory Statistics Courses from a Conceptual, Active Learning Viewpoint". Journal of Statistics Education, 3(3).

Smith, G. (1998). "Learning Statistics by Doing Statistics ". Journal of Statistics Education, 3(6).

APPENDIX A

Detail Description of The Grading System

Gred	Score (%)
A	100.00 – 79.45
B	59.45 – 79.44
C	49.45 – 59.44
D	39.45 49.44
F	0.00 – 39.44

APPENDIX B

Output From SAS Procedure

Univariate Procedure

Variable=UJI1P

Moments

N	169	Sum Wgts	169
Mean	60.5503	Sum	10233
Std Dev	19.84521	Variance	393.8323
Skewness	-0.17128	Kurtosis	-1.0297
USS	685775	CSS	66163.82
CV	32.77475	Std Mean	1.526554
T:Mean=0	39.66468	Pr>|T|	0.0001
Num ^= 0	169	Num > 0	169
M(Sign)	84.5	Pr>=|M|	0.0001
Sgn Rank	7182.5	Pr>=|S|	0.0001

Quantiles(Def=5)

100% Max	94	99%	94
75% Q3	80	95%	90
50% Med	61	90%	86
25% Q1	44	10%	32
0% Min	19	5%	27
		1%	20
Range	75		
Q3-Q1	36		
Mode	60		

Extremes

Lowest	Obs	Highest	Obs
19(	115)	90(	149)
20(	128)	92(	46)
22(	107)	92(	113)
23(	168)	94(	7)
24(	88)	94(	147)

Stem	Leaf	#	Box plot
9	000000002244	12	|
8	56668888	8	|
8	000000000122222222244444	24	+-----+
7	6788899	7	| |
7	000024444	9	| |
6	566667888888	12	| |
6	000000000111122234444 4	22	*--+--*
5	556666688889	12	| |
5	000000122234	12	| |
4	6778888	7	| |
4	00000012222244	14	+-----+
3	666667778889	12	|
3	002224	6	|
2	6667888	7	|
2	0234	4	|
1	9	1	|

----+----+----+----+----

Multiply Stem.Leaf by 10**+1

Univariate Procedure

Variable=UJI1P