SOTSS: A SYSTEM FOR ONLINE TESTING OF STATISTICAL SKILLS

Derk Jan Kiewiet and Natacha Borgers
University of Groningen, the Netherlands
d.j.kiewiet@rug.nl

This paper introduces the system SOTSS *for online testing of statistical skills that involve software packages as* SPSS. *The emphasis of the system is on the interpretation of statistical output, and drawing conclusions based on these interpretations. Basic characteristics of the system are: 1) individual, 2) flexible, and 3) interactive. The system can examine statistical skills efficiently. The efficiency is achieved by the possibility to give exams at different time slots, without increasing the teacher's workload unacceptably. Besides, the system also avoids the possibility for students to pass on answers or questions to each other.*

INTRODUCTION

One of the key questions in teaching is: did the students actually acquire the knowledge and skills that the course was intended to convey. To answer this question, students are examined. The examination of students can be done at various levels (see e.g., Bloom, 1957). At the Faculty of Management, University of Groningen, the emphasis in statistics courses is on the interpretation and use of output of statistical software as *SPSS*. The calculation of the values of statistical parameters by using formulas is only marginally covered in the program. However, because of problems with space, time and computer facilities, we experience difficulties in assessing the skills of interpretation of statistical output. This paper describes a system that can be used to assess aforementioned skills for large groups of students.

THE PROBLEM

The Faculty of Management and Organization is part of the University of Groningen in the Netherlands. From Dutch perspective, it is a large faculty with about 3000 students. In the first year of the curriculum, students follow two introductory statistics courses. In these courses, attention is paid to descriptive as well as inferential statistical techniques. Although the faculty had some relapse in the number of students in the late 90s, the number of first years is rapidly growing. In 2005 there were 800 students who followed the statistics courses.

Central in the statistics program are: real life (business) situations, conceptual knowledge of statistics, interpretation of statistical output, and managerial implications. As a consequence, calculations get little or no attention; the reading and application of formulas is not trained. On the other hand, much time is devoted in learning how to use statistical software, in our case *SPSS*. The choices and consequences can be transformed into the following learning objectives: 1) the student is capable of producing the right *SPSS*-output in practical as well as in statistical sense, 2) the student can select the right results (numbers) from the output, 3) the student can interpret these results in the right statistical context, and 4) the student can draw sound conclusions from these interpretations. The learning objectives are denoted as the PSIC-model: Produce, Select, Interpret, and Conclude. The instruction of this model is done in small-scaled practicals. Because of the large number of students that attend the statistics courses, there are many groups in the practicals. Also, these groups are scattered in time.

As lecturers, we want information about the degree in which students actually acquired the PSIC-skills. However, examination of all the students at the same time is not possible because of lacking facilities (space and computers). Taking tests at different times will give problems because of fraud prevention (a copy of the exam is not possible because the students will tell each other the right answers, and different versions of the exam will increase the work load for teachers unacceptably). Therefore, a different solution is needed to overcome these problems.

Although the above description is specific for the Faculty of Management and Organization, our situation is not unique. Firstly, there has been a general trend towards emphasizing data and concepts instead of calculations (Moore, 1997; Moore and Cobb, 2000; Utts, 2002). The use of statistical software is not only widespread in applied statistics education in universities, it has also changed or influenced the focus of introductory statistics education.

Secondly, the group sizes of students participating in statistical courses are growing, partly due to synergy objectives of university managers. Moreover, the student population in general has increased and changed. As a result we teach a more diverse and less specialized student population (Moore, 1997). Even in many emergent and newly industrialized countries, the massification of university education is being replicated (Castle and Kelly, 2004). Therefore, many statistics teachers probably are familiar with the problems around testing skills as described above for the Faculty of Management and Organization in Groningen.

A POSSIBLE SOLUTION

The problem of giving exams simultaneously in large group statistics programs is difficult to solve due to infrastructural (space and time) as well as financial limitations (the number of computers available). The solution could be found in individualized online testing. By using a computer for this individualized process, the workload of teachers can be minimized. The computer will generate the individual exams, and the computer would also do the grading. This is in principle the same online testing process as we know for several years (for example, Bugbee Jr. and Bernt, 1990; Kinney, 2001; Vantaggiato, 1998). However the specific demands in statistics education, as a result of using statistical software, are more complicated. You also need individualized data sets, and as a result, students produce individualized answers to test questions. To overcome these difficulties, we have developed an online testing system for statistical purposes, using statistical software. This system is called *SOTSS* (System for Online Testing of Statistical Skills). The following operating procedure is applied (see also Figure 1):

1) The teacher creates a basic data set, including different variables and a large number of respondents or other research units (possibly automatically generated by the computer).
2) The teacher creates a test bank of exam questions, which refer to the variables of the basic data set and meet the learning objectives.
3) The student logs in for the exam.
4) The computer draws a random sample of respondents (research units) from the basic data set; the student receives this individual test data set.
5) The computer draws a stratified sample from the test bank of exam questions in order to create an individual exam for the student.
6) The student starts analyzing the individualized data set with *SPSS* to answer the questions of the individual exam.
7) The student enters the answers in the computer.
8) The computer conducts analyses for the individual test questions and the individual test data set.
9) The computer compares these answers with the produced and entered answers of the student.
10) The computer computes feedback (general or question specific).

This system contains three important basic characteristics: it is individual, flexible and interactive. Individual, because every student receives an individual exam as well as his or her own data set, and is individually graded. Flexible, because the test bank can be easily adapted by the teacher due to differences in educational contents. Moreover, different statistical techniques can be examined, as well different types of questions can be tested (depending on learning objectives). Besides, teachers can use different statistical software packages or different spreadsheet packages in their statistics classes. Interactive, because students actions and computer reactions are dependent of each other at different moments in the exam process.

At this moment we are working at the development of a prototype of the above-mentioned system. There is a lot of interest from other departments within the University of Groningen.

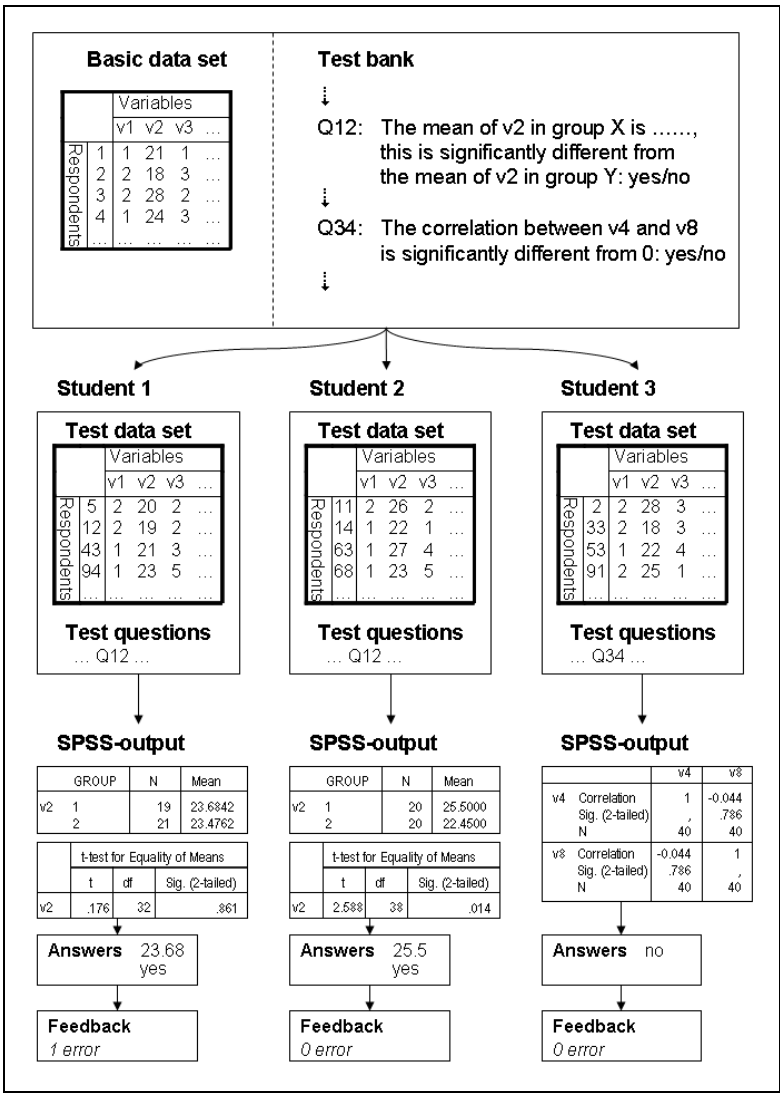

Figure 1

CONCLUSION

Examining statistical skills for large groups using statistical software is problematic in situations with limited availability of means. The above-described system can, despite limited facilities, examine statistical skills efficiently. The efficiency will be achieved by the possibility to give exams at different time slots, without increasing the teacher's workload unacceptably. Besides, the system also avoids the possibility for students to pass on answers or questions to each other.

REFERENCES

Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain.* New York: David McKay Co Inc.

Bugbee, Jr., A. C. and Bernt, F. M. (1990). Testing by computer: Findings in six years of use 1982-1988. *Journal of Research on Computing in Education*, 23(1), 87-101.

Castle, R. and Kelly, D. (2004). International Education: Quality assurance and standards in offshore teaching. *Exemplars and Problems, Quality in Higher Educations*, 10(1), 51-57.

Kinney, N. E. (2001). A guide to design and testing in online psychology courses. *Psychology Learning and Teaching*, 1, 16-20.

Moore, D. S. and Cobb, G. W. (2000). Statistics and mathematics: Tension and cooperation. *American Mathematical Monthly*, 107(7), 615-630.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, 65, 123-165

Utts, J. (2002). What educated citizens should know about statistics and probability. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Vantaggiato, A. (1989). Automatic exams and course administration on the web. In *WebNet 98 World Conference of the WWW, Internet and Intranet Proceedings.*