

## **Multiple-choice Examination Papers at a Basic University Statistics Course. Experience through 12 Years**

Jon Stene

*University of Copenhagen, Institute of Statistics,  
Studivestrate 6,  
DK-1455 Copenhagen K., Denmark  
e-mail: usijs@pc.ibt.dk*

Knut Conradsen

*Technical University of Denmark, Institute of Mathematical Modelling,  
Building 321, DTU,  
DK-2800 Lyngby, Denmark  
e-mail: kc@imm.dtu.dk*

Examination papers form an important part in the evaluation of the students' achievements of the educational goals of a course. Many basic courses with a large number of students run into the problem that all students are not evaluated in the same way. Those evaluated first may get a more thorough evaluation than those coming later. This is often the case with written examination papers where all students are confronted with the same set of questions and the examiners have to read through a large number of papers of 20 pages or more. Evaluation of examination papers in this way also make a heavy demand on manpower.

Traditional examination papers in statistics courses in Danish universities tend to consist of sets of two or three problems, each with a number of detailed questions. These problems are taken from different areas in the course. Other areas are not considered at all, but the chosen areas vary from set to set. This means that only limited areas of the course will be evaluated by each set. The general knowledge of most areas of the course will not be tested at all at the same time.

At the Technical University of Denmark the basic course in statistics runs for about 15 weeks and is held twice a year. Often more than 400 students take the course. The course has run for more than 25 years. The concerns mentioned above led in 1986 to the replacement of traditional sets by sets of multiple choice questions, which has now been used 25 times.

The structure of the exam paper has remained almost the same during this period. A set consists of a front page and of 30 questions starting at page 2 of the set. To each question 5 possible answers are given, only one being correct. In addition, there is a "don't know" option. The students are asked to mark one and only one of the 6 options and enter the answer into a box, one for each question, on the front page. Only the front page with the student's name and answers is handed in and forms the basis for the evaluation. The answers are typed into a computer.

The questions vary quite a lot. Together they cover most of the topics in the course. Some are theoretical ones where e.g. the expectation of a probability density has to be calculated, 5 options are given, only one being correct. Another example is different test statistics of a contingency table each with a numerical value, significance level and critical region. Only one among 5 options is correct. Some questions regard a recognition of the correct test for a given data problem. For ANOVA problems different sums of squares are given and the students are asked to choose the correct test statistic among 5 given ones. Some calculations are demanded, but long calculations, e.g. of sums of squares are avoided. Much effort is put into making the 5 options quite distinct and unambiguous.

Five points are given for the correct answer to a question, a wrong one gets -1, and no answer or "don't know" gets 0. All questions are weighted equally. Evaluation is based on the

sum of points for each student. This sum ranges theoretically from -30 to 150. The observed sum at an exam usually ranges from about nought to 150, which very few get. At the exam last autumn the median sum was 72 and the quartiles 47 and 99, respectively. These three quantities may vary considerably from term to term. One cause of variation is the person who constructs the set, who has also run the course that term. The difficulty of the various questions and of the whole set plays an important role. Young and less experienced teachers tend to construct more complicated questions than more experienced ones. The external examiner has to accept the whole set. He has been the same person for all these 25 sets and has played a modifying and homogenizing rôle. He does not formulate the questions himself, but suggests revisions. Often substantial revisions are made before a set is ready. A computer program calculates the sum for each student and a number of summary results.

A difficult problem is to translate the sums of points to the scale of marks and, in particular to determine the minimal number of points for passing the exam. In this connection one has to consider the possibility that a student by pure guessing can obtain a sufficiently large sum to pass the exam. As there are 5 options (except for the "don't know" one) the probability for guessing the right answer of a question is 0.2. By pure guessing the number of correct answers is binomially distributed with binomial parameter 0.2 and counting parameter equal to the number of questions considered, at most 30. If all 30 questions are answered, the probability for 9 correct ones, giving the sum 24, is 0.129 and for 12 correct ones with sum 42 the probability is 0.009.

The sets vary so much from term to term that the minimum number of points for passing cannot be kept constant. The same argument applies to the whole scale of marks. If the limits for the different marks had been kept constant, the mark a student got, would depend heavily on the term one entered for examination. Therefore, determination of these limits has been based on the cumulative frequency function of the sum of points for the actual term. Several ideas have been applied to determine the passing limit depending on how difficult the set has been. One regards the percentage of students allowed to pass, which might be put to 70%. If this idea is applied strictly and the set has been relatively simple some students, who have got a relatively large sum, and hence attained many of the goals of the course, would fail. Therefore, this principle is modified depending on the actual set. Similar arguments are applied for determining other limits.

The computer program also gives the percentage of answers to the six options for each question. This gives the teachers and the text book author a very good feedback with regard to problem areas in the course not fully understood and also with regard to the formulation of the different questions and options. With 30 questions large parts of the course material can be covered in a single set. In a set most of the questions should be so easy and straight forward that more than 40% of the students get the correct answer, some questions should be so easy that more than 70% of the answers are correct. Very few should be so difficult that less than 20% have correct answer. Then the probability of a correct guess is higher.

## **FRENCH RESUMÉ**

*À l'examen écrit du cours élémentaire en statistique à l'École Polytechnique danoise on pose une série de 30 questions à choix multiple. Seulement une possibilité est correcte. On fait 5 points pour une réponse correcte, -1 pour une réponse fautive et zéro pour aucune réponse. La notation est basée sur la somme de points des étudiants. Les sujets des 30 questions couvrent la matière du cours. Le type des questions varie beaucoup, quelques-unes sont théoriques, des autres traitent des applications et quelques-unes demandent des calculs numériques. La distribution de la somme varie d'un examen à un autre et dépend du constructeur de la série des questions. On discute des difficultés de la notation et de la composition de la série.*