ROYAL STATISTICAL SOCIETY

EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2012

MODULE 1 : Data collection and interpretation

Time allowed: One and a half hours

*Candidates should answer **THREE** questions.*

Each question carries 20 marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

*Candidates may use calculators in accordance with the regulations published in
the Society's "Guide to Examinations" (document Ex1).*

The notation log *denotes logarithm to base **e**.*
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ *is the same as* ^nC_r.

This examination paper consists of 4 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1.   The table below summarises the ages at the last birthday of patients seen by a doctor in his surgery during one particular working day.

Age in years	Number of patients
Under 15	13
15–29	10
30–44	12
45–59	15
60–74	21
75–89	9
Total	80

(i)   Estimates of the mean and standard deviation of the ages of these patients are required.

Peter calculates these by taking the mid-points of the intervals as 7.5, 22.5, 37.5, 52.5, 67.5 and 82.5.

Paul uses mid-points of 7, 22, 37, 52, 67 and 82.

Mary uses the same mid-points as Paul but then adds 0.5 to both her answers.

Select the method you think is best and obtain the mean and standard deviation of the ages of the patients. You should show details of your working.

State briefly why you did not choose either of the other methods.

(10)

(ii)   Calculate an estimate of the median age of these patients.

(4)

(iii)   It is found that two of the patients categorised in the 75–89 age group were in fact over 90 years old. Explain how your answers for the mean, standard deviation and median might be affected by this information.

Write a short comment on the ages of patients seen by the doctor on this day.

(6)

2.   One of the courses available from a provider of distance learning education has run for five years. Students on the course are allocated to one of fifty regions. In twenty-five of these regions, students are allocated to one of three tutors; in seventeen regions, to one of two tutors; in the remaining eight regions, there is a single tutor. There are between 12 and 18 students in each tutor's tutorial group. The provider is planning a survey of the students to find out their opinions of the course.

(a)   (i)   Discuss the relative merits of stratified and cluster sampling methods. Suggest ways of selecting a sample of regions using each of these methods.

(6)

(ii)   A sample of students is to be chosen by simple random sampling from selected regions. Explain in detail how this sample might be achieved using random number tables.

(4)

(iii)   Two methods of asking the tutors to choose samples of students from their group and to distribute questionnaires to the selected students are under consideration. One is a simple random sample of the tutor's students registered on the course to whom questionnaires would be posted. The other is a quota sample of students attending the final class. Discuss the relative merits of these two methods.

(4)

(b)   There are intakes to the course every six months. In one tutor's group the numbers of students in the last four years have been as follows.

12   14   13   16   12   15   17   18

(i)   Plot these figures.

(3)

(ii)   Comment on the pattern of intake to the course.

(3)

3. (a) A university plans to survey a sample of those who have graduated within the last two years. Construct questions on the following topics that would be suitable for inclusion in a mail survey of the graduates.

   - Month, year and age when secondary education was completed.
   - Details of subjects studied in the last year in secondary education with any qualifications gained.
   - Year of recent graduation from the university, degree title and class of degree obtained.
   - Details of any qualifications gained between secondary education and the most recent degree from this university.
   - <u>Main</u> occupation in each of the last two years.
   - For those who are working – their job title, current annual salary, the name of the employer and location of the employment.

   (14)

   (b) What are the main differences between questionnaires designed for mail surveys and those for use with an interviewer?

   (6)

4. *Non-response bias* and *response error* are both sources of error in social surveys.

   (i) Explain what these two sources are and what might cause them.

   (8)

   (ii) Suggest, with examples, ways of reducing them.

   (9)

   (iii) Comment on the idea that decrease of non-response bias could increase response error.

   (3)