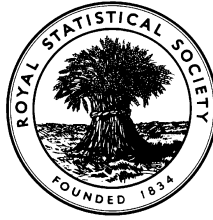# EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2011

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 $\begin{pmatrix} n \\ r \end{pmatrix}$ *is the same as* $^{n}C_{r}$*.*

1.   (a)   A simple random sample of 45 households was drawn from an area containing 14 848 households. The numbers y of persons per household in the sample were as follows.

> 5 6 3 3 2 3 3 3 4 4 3 2 7 4 3
> 5 4 4 3 3 4 3 3 1 2 4 3 4 2 4
> 6 3 3 3 4 2 5 4 4 2 4 3 1 4 4

(i)   Form a frequency distribution of y.

(3)

(ii)   Estimate the mean number of persons per household in the area, and obtain the estimated variance of your estimator.

(6)

(iii)   Estimate the proportion of households in the area that contain three or four persons, and obtain the estimated variance of your estimator.

(5)

(b)   A simple random sample of 1000 households is taken from a population of households.

In the sample there are 72 two-person households. State with reasons whether or not the 72 two-person households constitute a simple random sample of all two-person households in this population.

306 of the 1000 contain exactly one woman aged over 30 years. State with reasons whether or not the 306 women constitute a simple random sample of all women aged over 30 years in this population.

(6)

Turn over

2. The manager of a restaurant wishes to find out customers' opinions of the service provided, and she aims to do so by leaving comment cards on tables for customers to fill in.

    (i)   State three advantages and three drawbacks of this method of obtaining information.

    (6)

    (ii) To increase participation, the manager proposes to offer respondents the chance to enter a prize draw. What potential problems are there in using this incentive, and why?

    (2)

    (iii) Provide a draft comment card which asks for some basic information about customers such as

        • number in their party, sexes and age-groups

        • time of arrival

        • vegetarian or not

        • any allergies

    and which contains suitable questions to find out what the customers thought of

        • the speed of service

        • the friendliness and helpfulness of the staff

        • the quality of the food

        • the quality of the drinks

        • the cleanliness of tables.

    (12)

3. (a) (i) Explain the difference between a target population and a study population.

(3)

(ii) Explain what is meant by a sampling frame.

(3)

(iii) Discuss in general terms the problems of coverage that might arise when attempting to draw up a sampling frame for a study population.

(6)

(b) The table below is derived from the New Earnings Survey, which is an annual survey of employees in Great Britain based on a 1% random sample of the potential employed population. The information is obtained from employers. An employee's "standard working hours (per week)" is the number of hours for which that individual is contracted to work at normal (non-overtime) rates of pay.

Write a short report that highlights the main points shown in this table, explaining any statistical terms used.

(8)

	Men		Women	
	1975	1995	1975	1995
Average age (years)	39.5	38.7	40.0	38.5
Standard working hours (per week)				
Average	38.8	38.1	32.4	31.0
10th percentile	36.0	35.0	20.0	16.5
50th percentile	40.0	37.5	36.0	35.0
90th percentile	40.0	40.0	40.0	39.0

Turn over

4.	The display below shows the times to the nearest minute taken to complete a car treasure hunt in which 500 people took part.

Time to the nearest minute	Number of persons
40-45	56
45-50	61
50-55	65
55-60	69
60-65	102
65-70	83
70-75	64
	500

(i)	Explain why the classes as given are unsatisfactory. Propose a better set of boundary points and state what time values are included in each interval.

(4)

(ii)	Using your proposed boundary points,

(a)	represent these data by a frequency polygon,

(7)

(b)	obtain an estimate of the median time to complete the treasure hunt,

(4)

(c)	obtain estimates of the first and third quartiles and the inter-quartile range.

(5)