ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2015

MODULE 8 : Survey sampling and estimation

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* ${}^{n}C_{r}$.

This examination paper consists of 8 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

1. Wildlife managers want to estimate the total number of caribou in the Nelchina herd located in south central Alaska. The density of caribou differs dramatically in different types of habitat. A preliminary aerial investigation has identified the area used by the herd, and divided it into six strata based on habitat type.

For the main survey, the organiser decides to divide the area into sub-areas called quadrats, each of size 4 km^2. The survey is conducted by selecting a simple random sample of quadrats from each stratum, and for each quadrat the area is searched by aircraft to locate and then photograph the animals; the number of caribou, *y*, in each quadrat is counted in the photographs.

The sample means and standard deviations of the measurements, *y*, in each stratum based on a sample of 211 quadrats are as follows.

Stratum (h)	Map Quadrats N_h	Sample Quadrats n_h	Sample Mean $\bar{y}_h$	Sample Standard Deviation s_h
1	400	98	24.1	74.7
2	40	10	25.6	63.7
3	100	37	267.6	589.5
4	40	6	179.0	151.0
5	70	39	293.7	351.5
6	120	21	33.2	99.0
Total	770	211		

(i) The sampling frame for this survey is a land map. Discuss briefly what problems are likely to be associated with this type of sample.

(4)

(ii) The formula for the variance of the estimator of a population total based on a stratified (random) sample is

$$V = \sum_{h=1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}.$$

Define the terms N_h, S_h and n_h in the formula above.

(2)

(iii) Using the data above, estimate the total number of caribou in the herd and obtain an approximate 95% confidence interval for this total.

(7)

(iv) For this survey discuss briefly the merits of using stratified sampling rather than simple random sampling.

(4)

(v) Given that stratified sampling is used for this survey, discuss briefly the merits of using optimal rather than proportional allocation.

(3)

2. A Government agency is becoming increasingly concerned with the problem of obesity among the adult population. To explore the extent of the problem, the agency is to fund a national survey of 5000 adults. The survey will cover social and demographic variables as well as those that relate to lifestyle and consumption.

Participating households will be visited and interviewed, and individual members of the household will be asked to keep a diary of the types and quantities of food and drink consumed, and when, over a specified time period. After this period, there will be a further short interview to check on the completeness of the recorded entries and to collect information on any unusual circumstances or illness during the period which might have affected eating behaviour.

(i) Outline the advantages and disadvantages of collecting information for this survey using a diary instead of relying on a questionnaire.

(6)

(ii) Questions on lifestyle and consumption can be sensitive. Why might participants who fail to answer such interview questions or fail to complete the dietary diary give rise to bias in the results of this survey? How could you use the questions about a participant's background to investigate the bias?

Discuss other sources of errors that could arise in surveys such as the one described.

(7)

(iii) One topic that the interview will cover is participation in physical activity. Participants will be asked the following question about their usual exercise behaviour:

Do you *never* exercise, *occasionally* exercise, or *frequently* exercise?

Comment on the weaknesses of this question. Suggest questions to find out about the frequency and duration of participation in physical activity.

(7)

3. A household survey to collect information about people's housing circumstances was conducted in a community of 23 000 households. A researcher wishes to use the data to study the extent of overcrowding and under-occupation in this community. A simple random sample of 1500 households yielded the following information on size of household and number of bedrooms per household.

	Number of Households
All households in sample	1500
Number of persons per household	
one	434
two	525
three	247
four	200
five	65
six	20
seven	5
eight	4
Number of bedrooms per household	
one	158
two	406
three	645
four	228
five +	63

(i) Obtain a point estimate and an approximate 95% confidence interval for

(a) the mean number of persons per household in this community (a measure of overcrowding),

(b) the proportion of single-person households in this community (a measure of under-occupation).

(12)

(ii) The researcher has heard of an alternative measure of overcrowding to that given in part (i), based on number of persons per bedroom. Explain how you would estimate this quantity from the information above. Discuss the properties of this estimator.

(3)

(iii) Discuss any practical difficulties that might arise in this survey

(a) in defining "households",

(b) in defining "overcrowding",

(c) in including information from single-person households.

(5)

4. (a) A survey is being planned to find out whether people have been victims of crime and to obtain details of any crimes involved. Discuss whether it is better to collect information by means of a *telephone interview* or by a *self-completion* questionnaire.

(7)

(b) An orange grower is to sell a truckload of oranges. The oranges are packed into 140 crates containing 120 oranges each. Before agreeing to buy the oranges, the potential buyer wants to estimate the quantity of juice in the oranges, and proposes to inspect a sample of oranges.

Convenience sampling chooses the items which are most accessible while sampling is in progress. Suggest reasons why *cluster sampling* might be preferred to convenience sampling for this sampling inspection. How do *one-stage* and *two-stage* cluster sampling differ in the context of this example?

(7)

(c) A survey exploring finances of full-time undergraduate students at a university is being planned. One objective is to estimate the mean weekly earnings of such students in paid work during term time. The university has 11 500 full-time undergraduate students.

The organiser of the survey wants to estimate the mean weekly earnings of such students in paid work during term time to within ±£3. If the organiser were to use simple random sampling, show that this aim would be achieved with a sample size of about 510. The population standard deviation may be assumed to be £35.38.

Discuss briefly the practical difficulties that might arise in planning this survey.

(6)

BLANK PAGE

BLANK PAGE

BLANK PAGE