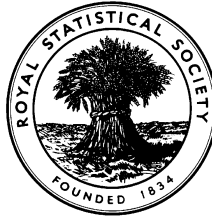**EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY**
(formerly the Examinations of the Institute of Statisticians)

ORDINARY CERTIFICATE IN STATISTICS, 2003

Paper II

Time Allowed: Three Hours

*Candidates may attempt **all** the questions.*

The number of marks allotted to each question or part-question is shown in brackets.

The total for the whole paper is 100.

A pass may be obtained by scoring at least 50 marks.

Graph paper and Official tables are provided.

Candidates may use silent, cordless, non-programmable electronic calculators.

*Where a calculator is used the **method** of calculation should be stated in full.*

1

1. The following is an extract from a letter to *The Times* (a newspaper) on 21 January 2002.

> *I became aware that most of the Letters to the Editor in The Times were from men. I decided to count the letters in 100 consecutive issues of the paper. Between September 18, 2001 and January 12, 2002, you published 1942 letters on this page; 244 were from women.*
>
> *I classified by subject 100 letters from ladies which appeared between October 3 and November 21, 2001. The top three topics were: home and family, 18 letters; terrorism and war, 17; education, 12. Next, with five letters each, came social questions and the arts.*

    (i)   Construct a table showing the distribution of Letters to the Editor by sex in the period September 18, 2001 to January 12, 2002.

            Draw a bar chart to illustrate the table.

                      (4)

    (ii)   Construct a table showing the distribution of the 100 ladies' letters by subject.

            Draw a bar chart to illustrate the table.

                      (5)

    (iii)   State, with reasons, whether you would consider it preferable to use pie charts rather than bar charts to show these tables.

                      (2)

2. In each of the following quotations, discuss the error in the section in italics and correct the error.

    (i)   "Migraines affect approximately 14% of women and 6% of men, *that is one fifth of the population,* assuming that there are equal numbers of men and women in the population."

    (ii)   "Luxurious sofa – price slashed from £999 to £333, *a reduction of 200%.*"

    (iii)   "The insurance premium is reduced by 70% for five years without claims, a further reduction of 10% of the reduced premium is given for buying on the internet, *that's a total reduction of 80%.*"

                      (6)

Turn over

3. There are 25 contestants labelled *A, B, …, Y* who take part in a quiz show. In the first round they have to answer two general knowledge questions. The labels of those getting the questions **incorrect** are given below.

Question 1 incorrect   *A C E G H P T V*

Question 2 incorrect   *A D F G I P Q S V Y*

(i) Copy and complete the following contingency table.

(2)

	Question 2 incorrect	*Question 2 correct*	*Total*
Question 1 incorrect			
Question 1 correct			
Total			

(ii) From the contingency table find

    (a)   the proportion getting Question 1 correct,

    (b)   the proportion getting Question 2 correct,

    (c)   the proportion getting both questions correct,

    (d)   the proportion who get at least one question correct,

    (e)   the proportion of those who get Question 1 correct who also get Question 2 correct,

    (f)   the proportion of those who get Question 2 incorrect who also get Question 1 incorrect.

(6)

(iii) Draw up a frequency table showing the numbers of contestants who answered 0, 1, 2 questions correctly. Find the mean and the standard deviation of the number of questions answered correctly.

(4)

Turn over

4. The EFG Bank employs 72,000 people in its UK operations. The Human Resources Director is eager to compare the distribution of salaries within the bank with those in other banks operating in the UK.

Distribution of salaries in EFG Bank, year ending 31 December 2002

Salary	Numbers of Employees
Under £10,000	6200
£10,000 but under £15,000	12000
£15,000 but under £20,000	15600
£20,000 but under £25,000	14200
£25,000 but under £30,000	11900
£30,000 but under £40,000	7000
£40,000 but under £50,000	3500
£50,000 but under £100,000	1500
£100,000 or more	100

(i) Estimate, by calculation, the median, the upper and lower quartiles and the ninety-fifth percentile of this distribution. Give your answers correct to the nearest £100. In which salary group does the fifth percentile lie?

(10)

(ii) The corresponding values for the UK banking industry as a whole are as follows.

Median: £23,200

Upper Quartile: £32,300

Lower Quartile: £15,700

Ninety-fifth percentile: £48,600

Fifth percentile: £12,100

Comment briefly on the similarities and differences between the data for the UK as a whole and your calculations for EFG Bank. Suggest how they might have arisen.

(3)

(iii) Explain why it may be preferable to use the median, quartiles and percentiles for describing the location and dispersion of this distribution, rather than the mean and the standard deviation.

(2)

4

5. (i) Explain briefly what the product-moment correlation coefficient, r, of two variables measures.

(2)

(ii) Draw a sketch of a possible scatter plot of two variables x and y when

(a) $r = +1$,

(b) $r = -1$,

(c) $r = 0$.

(3)

(iii) Give one similarity and one difference between the product-moment correlation coefficient and Spearman's rank correlation coefficient, r_s.

(2)

(iv) Sketch a scatter plot for which r_s is equal to $+1$ and r is not equal to $+1$.

(2)

(v) Data for eight randomly chosen finishers in a men's ski race in the 2002 Winter Olympics are given below. The data give the finishing time for each competitor and his speed at a particular check point on the course.

Competitor	Time (seconds)	Speed (km/h)
A	100.30	106.1
B	101.24	102.2
C	102.54	97.5
D	99.13	104.3
E	100.81	101.1
F	99.78	105.5
G	101.66	105.8
H	100.58	104.0

Calculate Spearman's rank correlation coefficient for the data and comment on the result.

(5)

(vi) Suppose that, in (v), the speed data had been given in miles per hour. What effect would this have on r_s?

(2)

5

Turn over

6.   A family carefully monitors its expenditure on food and has noted that the amount spent on fresh fruit, in particular, has risen from September 1999 to September 2002. Details of prices and weekly consumption are given in the table.

Fruit price and consumption data

	September 1999		September 2002	
Fruit	*Price*	*Consumption*	*Price*	*Consumption*
Apples	60p per kg	1.5kg	75p per kg	2kg
Oranges	12p each	15	15p each	20
Bananas	80p per kg	1.5kg	75p per kg	1kg
Grapefruit	25p each	8	35p each	10

They have asked a neighbour, who is studying statistics, to estimate by how much the price of fruit has risen over the three-year period and the neighbour decides to calculate a weighted index of prices, based on September 1999.

(i)   Explain why a weighted index number is preferable to an unweighted index number in this situation.

(1)

(ii)   State whether you would calculate a Laspeyres aggregate index of fruit prices or a Paasche aggregate index of fruit prices, giving a reason for your choice.

(2)

(iii)   Calculate your chosen index number, based on September 1999. Interpret your result for the family.

(7)

7. (i) Explain what you understand by the following terms in relation to a time series.

   (a) Trend.

   (b) Seasonal component.

   (c) Additive model.

   (d) Multiplicative model.

   Explain when it would be appropriate to use a multiplicative model rather than an additive model for time series analysis.

   (5)

   (ii) The numbers of passengers travelling through a regional airport have been recorded quarterly as shown in the table.

   Airport passengers (×100,000)

		Quarter			
		1	*2*	*3*	*4*
	1999	12	15	21	13
	2000	15	19	27	17
Year	*2001*	21	28	36	16
	2002	22	32	44	25
	2003	25			

   Estimate the trend of passenger numbers by calculating centred four-quarterly moving averages. Comment on any irregularity in the trend.

   (7)

   (iii) Use a multiplicative model to estimate average seasonal indices for each quarter. Describe, in non-technical language, the fluctuation between quarters.

   (7)

Turn over

8.   Three coins are labelled *A, B* and *C*. Coin *A* is unbiased. For *B* it is estimated that the probability of a head is 1/3. For *C* it is estimated that the probability of a head is 1/4.

(i)   Explain briefly how these estimates for *B* and *C* might have been obtained.

(2)

In the rest of this question, assume these estimates are correct.

(ii)   The three coins are each tossed once. By drawing a tree diagram, or otherwise, find the probability of every possible outcome.

(5)

(iii)   Hence write down the probabilities of obtaining 3 heads, 2 heads, 1 head and 0 heads.

(2)

(iv)   If 2 heads were obtained in total, what is the probability that one of the heads was on coin *A*?

(2)