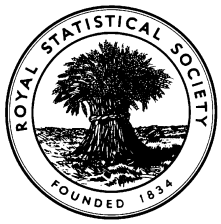
EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY
(formerly the Examinations of the Institute of Statisticians)



ORDINARY CERTIFICATE IN STATISTICS, 2001

Paper II

Time Allowed: Three Hours

There is no restriction on the number of questions that a candidate may attempt, nor on the order in which they are attempted. Candidates are not required to answer all the questions: they should answer as many as they can.

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.*

This examination paper consists of 10 printed pages. This front cover is page 1. The reverse of the front cover, which is intentionally left blank, is page 2. Question 1 starts on page 3.

There are 8 questions altogether in the paper.

1. A statistical table gives the following values for a probability, p, which depends on a value, z.

For $z = 0.70$, $p = 0.7580$.
For $z = 0.72$, $p = 0.7642$.

(i) Use linear interpolation to estimate p for $z = 0.714$. Show that this estimate can be written as a weighted average of the two given values of p, stating the weights clearly.

(4)

(ii) You are now told that the values of p given are correct to four decimal places. Write down the maximum and minimum possible values for both of the p values given.

Use the weighted average form to write down the maximum and minimum possible values of the estimate of p for $z = 0.714$.

(4)

2. A mail order company is considering whether or not to open a call centre so that it could take more orders by telephone and so reduce the costs of processing orders. In order to evaluate the length of time spent on the telephone when an order is placed, it has conducted a small pilot survey with a number of its long-established customers and asked them to use the telephone rather than the post when ordering. The durations of 40 such telephone calls are shown in the table.

Duration (seconds) of customer telephone order calls

422	317	503	325	76	217	411	69	255	299
172	274	325	552	522	526	72	86	325	286
299	255	459	586	581	556	555	103	215	139
444	114	501	489	85	286	270	497	333	299

Construct a stem and leaf diagram to display the data and comment on what it shows.

(6)

3. Two human resource managers in different companies want to compare the lengths of time that employees stay in their organisations before leaving. They consulted their records on 31 March and looked at all the employees who left the organisations in the previous three months. The data are shown in the table.

Number of employees

Length of service (months)	*Company A*	*Company B*
Less than 1	2	40
1 but less than 2	3	20
2 but less than 3	7	18
3 but less than 6	20	22
6 but less than 12	33	20
12 but less than 24	20	20
24 but less than 36	10	20
36 or more	5	40

(i) Draw, on a single graph, cumulative percentage frequency polygons for the two distributions.

(8)

(ii) From your graph, estimate

(a) the median length of service in Company *A*,

(b) the median length of service in Company *B*,

(c) the inter-quartile range of length of service in Company *A*,

(d) the inter-quartile range of length of service in Company *B*.

(4)

(iii) Write a brief report comparing the lengths of service at the two companies.

(4)

4. Give one advantage and one disadvantage of the standard deviation as a measure of the spread of a set of data values.

(2)

The maximum temperatures in degrees Celsius at 8 places in the British Isles on a July day last year were as follows:

18 13 17 20 23 19 15 21

(i) Calculate the mean and standard deviation of these temperatures.

(4)

(ii) Write down the mean and standard deviation of the temperatures in degrees Fahrenheit. (To convert degrees Celsius to degrees Fahrenheit, the temperature in Celsius should be multiplied by 1.8 and then have 32 added.)

(4)

(iii) The original data values need to be scaled (i.e. linearly transformed) to a set which have mean zero and standard deviation one. Find the scaled (transformed) values corresponding to the largest and smallest of the data values and write down the range of the transformed values.

(3)

5. An insurance company classifies drivers according to sex and to whether they are under 25 or 25 years and over. It finds that 60% of its drivers are male; 25% of the male drivers and 30% of the female drivers are under 25.

(i) Find the probabilities that a randomly chosen driver is in each of the four categories

(a) male and under 25,

(b) male and 25 or over,

(c) female and under 25,

(d) female and 25 or over.

(4)

(ii) Hence write down the probabilities of a driver being

(a) under 25,

(b) male given that the driver is under 25,

(c) male or under 25 (or both),

(d) neither male nor under 25.

(4)

(iii) The probability p of having at least one accident in a year is given in the table for the different classes of drivers whose distribution is as above.

Probability p of one or more accidents for different classes of driver

	under 25	*25 or over*
male	0.09	0.04
female	0.06	0.02

(a) Find the probability that a randomly chosen driver has at least one accident in a year.

(b) If a driver has at least one accident what is the probability that the driver is male and under 25?

(4)

6. Explain what is meant by a *rank correlation coefficient* and outline two circumstances in which a rank correlation coefficient might be used rather than an ordinary (i.e. product moment) correlation coefficient.

(4)

What are the maximum and minimum values taken by a rank correlation coefficient? For rankings of six quantities write down pairs of rankings which will give

(i) the maximum value,

(ii) the minimum value.

(4)

The table below gives the length, L, in metres and the wing-span, S, in metres of ten aeroplanes.

Values of length (L) and wing-span (S) for ten aeroplanes

L	70.8	70.7	47.3	54.9	36.6	30.5	63.7	37.6	62.1	55.3
S	59.6	64.9	37.9	47.6	28.9	28.0	60.9	33.9	25.5	50.4

(iii) Draw a scatter diagram of the data and identify the outlier.

(4)

(iv) Calculate Spearman's rank correlation coefficient for all ten aeroplanes.

(4)

(v) Recalculate the correlation coefficient excluding the outlier. Comment briefly on the two coefficients.

(4)

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 an additive model rather than a multiplicative model for time series analysis.

(5)

(ii) The table provides data about the use of electricity in a large college in the UK for each quarter from 1996 Qtr 1 to 2000 Qtr 2. The trend has been calculated by the method of moving averages. Calculate the average seasonal component for each type of quarter using an additive model.

Consumption of electricity 1996 - 2000

Year/Quarter	*Units used (000s)*	*Trend (000s of unit)*
1996/1	207	
/2	152	
/3	127	161.125
/4	176	161.000
1997/1	172	166.000
/2	186	166.875
/3	133	171.000
/4	177	173.625
1998/1	204	172.000
/2	175	171.375
/3	131	168.375
/4	174	165.875
1999/1	183	166.125
/2	176	166.375
/3	132	167.875
/4	175	167.750
2000/1	194	
/2	164	

(7)

Question Seven is continued on the next page.

(iii) The trend for each quarter of 2001 has been predicted as follows:

	Qtr 1	*Qtr 2*	*Qtr 3*	*Qtr 4*
Electricity consumption units (000s)	170	172	174	175

Using the seasonal components you have calculated in (ii), predict the actual consumption of electricity in each quarter of 2001.

(2)

NOTE: It is not necessary to draw a graph.

(iv) Suggest factors which could affect the accuracy of the forecasts.

(2)

8. Explain what is wrong with the conclusions in **bold type** in each of the following statements.

(i) In Town M, the correlation coefficient between the ages of men and women at marriage is 0.94. **Hence a man is likely to marry a woman of his own age.**

(ii) An index number of retail prices in April 1995 was 124.2 (Jan 1994 = 100). In April 2000, the value of the index was 150.4 (Jan 1994 = 100). **Hence prices had gone up by 26.2% between April 1995 and April 2000.**

(iii) Company N is a company employing 500 people, a few of whom are highly paid specialists. The mean annual salary of employees in the company is £15,355. **Hence half the employees earn less than £15,355 per year.**

(9)