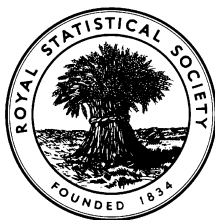**EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY**
(formerly the Examinations of the Institute of Statisticians)

ORDINARY CERTIFICATE IN STATISTICS, 2002

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.   A large organisation has been accused of being "ageist", i.e. tending not to employ older people. In response, it publishes the following table showing the age distribution of its current employees.

Age last birthday (years)	Number of employees
15 – 19	240
20 – 24	340
25 – 29	360
30 – 39	420
40 – 49	380
50 – 64	240

(i)   Draw a histogram on graph paper to show the data.

(6)

(ii)   State, with reasons, whether the data suggest that the organisation is ageist.

(2)

(iii)   Explain why the way the data have been presented in the table may be misleading to the casual observer.

(2)

3

Turn over

2. Explain what is meant by the dispersion of a set of data.

(1)

Define three different measures of dispersion for a set of data and give one advantage and one disadvantage of each measure.

(9)

Which of your measures would you recommend in calculating the dispersion of a set of data giving the wages of **all** employees in a company? Give brief reasons for your choice.

(3)

3. The table shows 100 three-digit numbers x that have been generated using the random number function on an electronic calculator.

Values of x

838	114	017	839	129	298	136	886	058	368
250	877	314	554	200	032	114	415	244	479
522	290	983	522	492	715	160	545	636	642
867	260	202	351	322	134	209	164	025	027
984	319	593	161	035	359	999	243	502	993
830	723	314	574	126	426	601	558	692	867
488	102	834	029	750	425	427	465	681	978
558	385	393	398	592	926	337	683	792	659
078	105	957	150	927	789	904	188	102	299
616	610	877	377	737	610	067	878	472	344

You are given that $\sum x = 47\,118$ and $\sum x^2 = 30\,710\,404$.

(i) Calculate the mean and the standard deviation of this sample of random numbers.

(2)

(ii) Group the data into a frequency distribution using classes 000 – 199, 200 – 399, …, 800 – 999.

(4)

(iii) Calculate the mean and the standard deviation of this grouped frequency distribution.

(6)

(iv) Comment on your results in parts (i) and (iii) and explain why they are not identical.

(4)

Turn over

4.   In a survey of shopping habits, participants were asked to indicate with a tick any of three transport factors that they considered to be important to their choice of shopping venue. The following results were obtained.

		% who ticked the factor and no other	*% who did not tick the factor*
Factor	*Distance*	12	38
	Convenience of parking	6	42
	Fuel availability	8	44

You are also told that 22% ticked all three factors.

(i)   Draw a suitable Venn diagram that divides the space into 8 regions, and mark the appropriate percentages in each region.

(9)

(ii)   Write down in a table the probabilities that a randomly chosen participant ticked 0, 1, 2, 3 factors.

(2)

(iii)   Write down in a table the probabilities that, given that a randomly chosen participant ticked at least one factor, the participant ticked 1, 2, 3 factors.

(3)

5.   The expenditure (£M, or millions of pounds) on buildings and equipment in one region of the country is given in the table.

Expenditure (£M) on buildings and equipment 1986 – 2000

Year	Expenditure	Year	Expenditure	Year	Expenditure
1986	200	1992	257	1998	426
1987	203	1993	273	1999	449
1988	207	1994	305	2000	452
1989	220	1995	341		
1990	242	1996	379		
1991	256	1997	395		

(i)   Draw a line graph of the data.

(4)

(ii)   Calculate a three-year moving average and plot it on your graph.

(5)

(iii) Explain why moving averages are not satisfactory for predicting the trend of expenditure during the period 2001 – 2005.

(3)

Turn over

6.   Consider the following three diagrams and in every case

      (i)   state why the diagram may be misleading,

                                                                    (6)

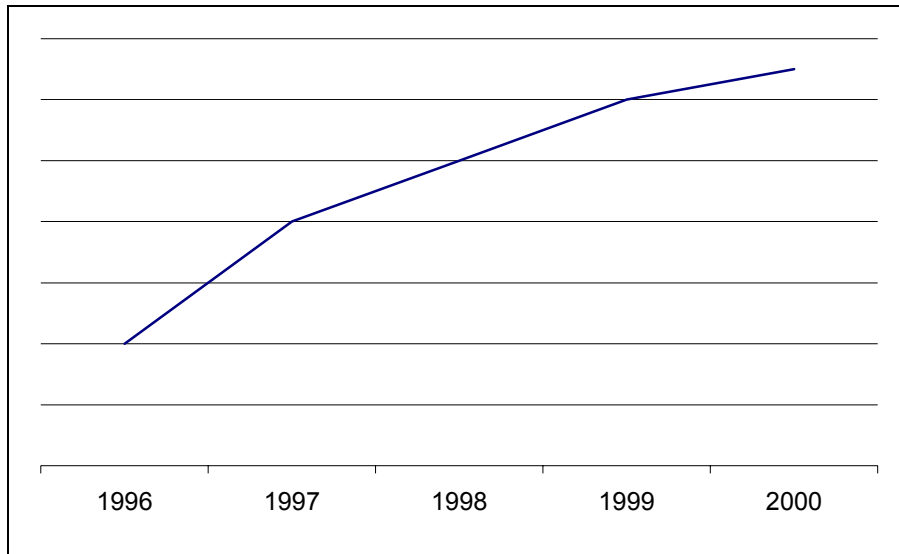   (ii)   interpret the diagram as far as possible.
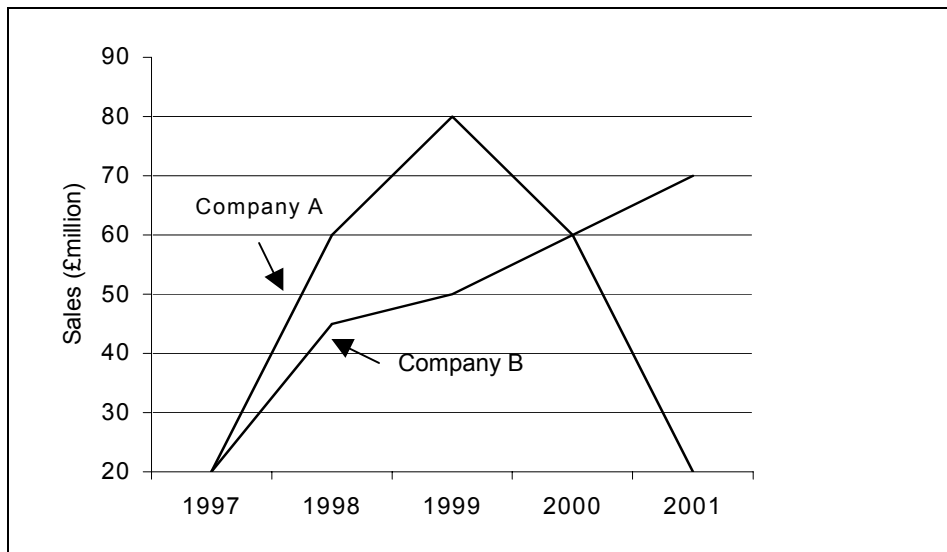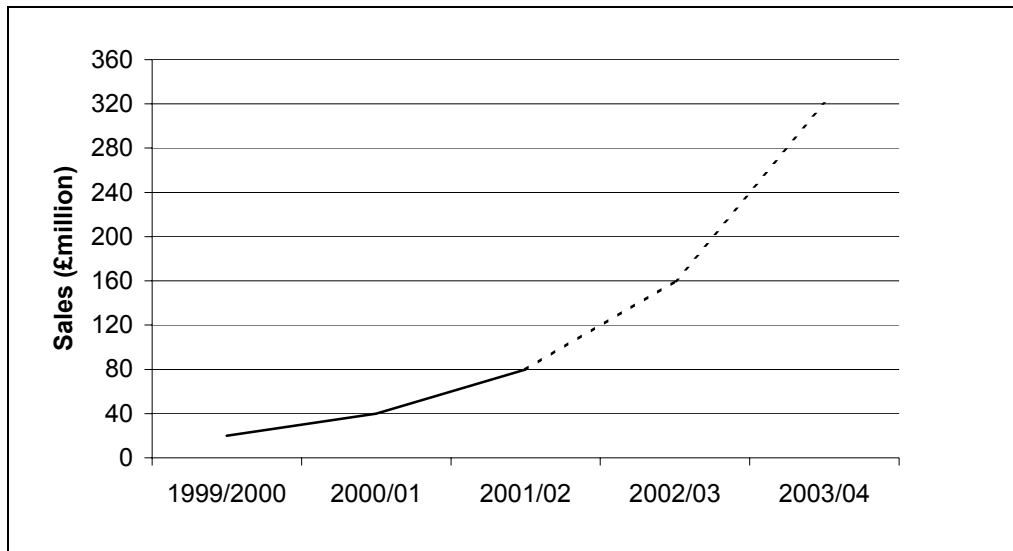
                                                                    (3)

Diagram 1 – Sales up!

Diagram 2 – Disastrous results for Company A

Diagram 3 is on the next page

Diagram 3 – Internet company to prosper in 2003/04

7. Andrew is a member of a Trade Union and is keen to monitor the rates at which his earnings have increased over the past five years. His annual earnings before tax have been as follows.

Year	1997	1998	1999	2000	2001
Earnings (£)	14 590	15 203	15 735	16 191	16 596

(i) Using 1997 (=100) as base year, calculate an index number of earnings for each year from 1998 to 2001. Give your answers correct to 1 decimal place.

(2)

(ii) Use the chain-base method to calculate an index number of earnings for each year from 1998 to 2001.

(2)

(iii) Interpret the results for Andrew.

(4)

Turn over

8.   Maccal Ferries runs ferries to the Scottish Islands from the mainland. The data below give the prices (in £) for a return ticket for a driver and for a car on each of 10 routes.

Route	A	B	C	D	E	F	G	H	I	J
Driver cost (x)	20	23	27	33	28	42	38	23	22	19
Car cost (y)	92	107	124	165	105	163	143	85	100	83

You are given that

$$\sum x^2 = 8\ 113, \qquad \sum y^2 = 144\ 671, \qquad \sum xy = 34\ 046.$$

(i)   Plot a scatter diagram of the data, marking the letters near your points.

(4)

(ii)   Find the correlation coefficient between x and y and comment on its value.

(6)

(iii)   Find the regression line that predicts car cost for a given driver cost.

(4)

(iv)   Plot the line on your scatter diagram. Which route gives the cheapest actual car cost compared to predicted cost and which the most expensive?

(4)