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

ORDINARY CERTIFICATE IN STATISTICS, 2005

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 calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

This examination paper consists of 7 pages, **each printed on one side only**.
This front cover is page 1.
Question 1 starts on page 2.

There are 8 questions altogether in the paper.

1. In the year 2000, the National Center for Health Statistics estimated that the numbers (in millions) of older people in the USA who suffered from Alzheimer's Disease, by age group, were

Age 65–74	Age 75–84	Age 85+
0.3	2.4	1.8

The Center's projections for 2050 (in millions) are

Age 65–74	Age 75–84	Age 85+
0.4	4.8	8.0

(i) Draw separate bar charts, using the same scale, to represent the data. (6)

(ii) Comment on what these charts show. (2)

(iii) Give one advantage and one disadvantage of pie charts as opposed to bar charts in representing data. (2)

2. In the 1881 British Census, ages are recorded as age at last birthday up to and including Census Day (3 April).

(i) A man is recorded as age 26. Write down his minimum and maximum possible age, in years and days, explaining carefully how you obtained your values. (2)

(ii) A husband is recorded as age 36 and his wife as age 34. Again working in years and days, write down

(a) the minimum and maximum period by which the husband is older than the wife, (1)

(b) the mid-point of these two values. (1)

(iii) The table below gives the ages of 10 husbands and their wives. Working in complete years only, find the mean and standard deviation of

(a) the difference in age between the older and the younger of each couple, (4)

(b) the difference in age between the husband and his wife. (4)

Husband's age	26	32	29	51	60	45	59	77	40	65
Wife's age	24	27	40	48	64	45	62	64	37	66

3. The charge in a town centre car park depends on how long a car is parked, rounded up to the next hour. Penalty charges are levied for stays of more than six hours. For a random sample of 60 cars, the parking times were recorded to the nearest minute. A stem-and-leaf diagram of these times is shown below.

Times (minutes) in car park

hundreds	tens
0	1124
0	5555556
1	0123344
1	5566667777799
2	022233334
2	556799
3	00123
3	5555666
4	
4	9
5	2

(i) Explain clearly what the entries in the diagram signify. By reference to the final entry 5 | 2, identify the "stem" and the "leaf".

(3)

(ii) Comment on the distribution of times. What does this distribution suggest about the pricing policy at the car park?

(3)

(iii) Assuming that each of the times recorded had a units digit of 5, find the median and quartiles of the distribution of times.

(3)

(iv) Draw a box and whisker plot of the distribution. Comment on the skewness of the distribution.

(3)

4. The manager of a busy supermarket carried out a survey and recorded the lengths of time that a sample of 520 customers had to wait at the checkout before being served. The results are shown in the table.

Waiting time (minutes)	Number of customers
Less than 0.5	16
0.5 but less than 1	22
1 but less than 2	74
2 but less than 3	76
3 but less than 4	68
4 but less than 5	60
5 but less than 7.5	56
7.5 but less than 10	54
10 but less than 15	42
15 but less than 20	36
20 or more	16

(i) Draw a cumulative frequency polygon (ogive) of the data.

(6)

(ii) Use your graph to estimate the median and quartiles of the distribution.

(4)

(iii) Estimate, by calculation using linear interpolation, the median of the distribution.

(3)

(iv) What aspects of these waiting times seem satisfactory and what should give that manager cause for concern?

(2)

5. A restaurant in a certain area currently serves quiches that contain four fresh eggs. Research in the area has shown that one in four locally produced eggs contains the salmonella bacterium. A health inspector has therefore suggested that in order to avoid food poisoning with salmonella the restaurant should use no more than three eggs in its quiches.

(i) Comment on the logic behind the inspector's suggestion.

(2)

Assuming that any egg has, independently, the same chance as any other of containing salmonella bacterium, find

(ii) the probability that a three-egg quiche is salmonella free,

(1)

(iii) the probability that a four-egg quiche is salmonella free,

(1)

(iv) for what values of n an n-egg quiche would have a less than 10% probability of being salmonella free.

(3)

(v) Comment on the assumption of independence between the eggs in a given quiche.

(1)

6. The table shows the quarterly number of reported major road accidents within the boundaries of a large town from 2002 to 2004.

Year	Quarter	Accidents (y)
2002	1	42
	2	38
	3	36
	4	46
2003	1	48
	2	43
	3	42
	4	47
2004	1	52
	2	47
	3	45
	4	49

(i) Draw a time chart of the data.

(5)

(ii) Explain why it is appropriate to calculate a four-quarterly centred moving average to estimate the trend of accidents.

(1)

(iii) Estimate the trend of accidents in this way and plot it on your chart.

(4)

(iv) Comment on the trend.

(1)

(v) Numbering the quarters from 2002 Qtr 1 to 2004 Qtr 4 in order from 1 to 12 and using these as x values, calculate the least squares regression line of Accidents (y) on Quarter number.

(8)

(vi) Plot this line on your chart and compare it with the moving average trend.

(4)

Turn over

7. In the preliminary rounds of the 2003 Rugby Union World Cup, the points scored for and against each of the eight teams who reached the quarter-finals are given in the table below.

Team	Points scored	Points conceded
Australia	273	32
Ireland	141	56
France	204	70
Scotland	102	97
England	255	47
South Africa	184	60
New Zealand	282	57
Wales	132	98

Rank the teams according to their success in (i) scoring points, and (ii) not conceding points. Calculate Spearman's rank correlation coefficient between the two rankings. Comment on the answers you have found.

(8)

8. A car manufacturing company has produced four models of car for a number of years. The costs of production of the models are shown in the table together with the numbers produced each week in 2000 and 2004.

Model	**Number produced each week**		**Cost of production, in thousands of pounds per car**	
	2000	*2004*	*2000*	*2004*
A	160	200	7.5	8.4
B	240	240	10.2	12.0
C	500	750	9.8	9.6
D	200	140	12.4	13.2

(i) Using 2000 as base year, calculate production cost relatives for each model for 2004, correct to 1 decimal place.

(6)

(ii) Comment on what the relatives tell you about the changes in cost from 2000 to 2004.

(2)

(iii) Using the numbers produced in 2004 as weights, calculate a weighted mean of production cost relatives index for 2004, with 2000 as base.

(4)