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

ORDINARY CERTIFICATE IN STATISTICS, 2007

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 6 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. (a) Briefly explain the difference between discrete and continuous data, giving one example of each type.

(4)

(b) Briefly explain the difference between inter- and intra-subject variation. Give an example of data where both types of variation might be found.

(4)

2. For 31 consecutive days, the numbers of vehicles passing a checkpoint in the period 10:00 to 11:00 were as shown.

Day 1	*Day 2*	*Day 3*	*Day 4*	*Day 5*	*Day 6*	*Day 7*	*Day 8*	*Day 9*	*Day 10*
355	432	516	467	394	218	175	428	523	419

Day 11	*Day 12*	*Day 13*	*Day 14*	*Day 15*	*Day 16*	*Day 17*	*Day 18*	*Day 19*	*Day 20*
384	439	224	156	518	556	494	369	388	251

Day 21	*Day 22*	*Day 23*	*Day 24*	*Day 25*	*Day 26*	*Day 27*	*Day 28*	*Day 29*	*Day 30*
208	476	381	464	508	529	253	184	394	475

Day 31
508

(i) Draw a stem-and-leaf diagram of the data with the stem being the hundreds value and the leaf the tens value (and the last digit not used).

(3)

(ii) Calculate the median and quartiles of the data, referring back to the original data to obtain values as accurate as possible.

(3)

(iii) Comment on and give a possible explanation for the main feature apparent in the data. What would be a more appropriate analysis of the data?

(3)

3. A sample of 100 shells from the Common Mussel (*Mytilus edulis*) has been collected from a beach in Scotland. The table shows the distribution of lengths, correct to the nearest millimetre.

Length (mm)	20 – 29	30 – 39	40 – 49	50 – 59	60 – 69
Frequency	2	7	11	15	18

Length (mm)	70 – 79	80 – 89	90 – 99	100 – 109	110 – 119
Frequency	17	12	9	5	4

(i) Draw up a cumulative frequency table showing, for each class, the maximum possible length and the cumulative frequency.
(3)

(ii) Plot on graph paper a cumulative frequency curve for the data.
(4)

(iii) Use this curve to estimate the median and the inter-quartile range of the data.
(4)

4. Water quality can be assessed by several different measures: higher quality is indicated by lower levels of suspended solids and chemicals, and by higher levels of dissolved oxygen. The table shows the quality measurements taken at various times during winter 2004/5 on a stream situated close to a major UK airport.

Date	*9 Nov*	*8 Dec*	*5 Feb*	*21 Feb*	*29 Mar*	*25 Apr*
Suspended solids (mg/l)	4	6	7	10	8	55
Ammonia (mg/l)	0.533	0.0006	0.74	0.0001	0.431	0.278
Nitrate (mg/l)	2.8	2.36	2.59	2.42	2.55	1.61
Orthophosphate (mg/l)	0.192	0.333	0.103	0.097	0.157	0.175
Dissolved oxygen (%)	81	83	87	89	104	84

(i) Copy and complete the following table, giving results correct to 3 significant figures.

Quality measure	*Mean*	*Standard deviation*	*Coefficient of variation*
Suspended solids	15.0 mg/l	19.7 mg/l	
Ammonia			
Nitrate	2.39 mg/l	0.411 mg/l	
Orthophosphate	0.176 mg/l	0.0858 mg/l	
Dissolved oxygen			

(11)

(ii) Comment on any unusual features in the original data and on any shown by your calculations. State with reasons whether your results suggest that the quality of water in the stream gives cause for concern.
(5)

5. Every Saturday I do my weekly shopping at the local village shop, the town centre supermarket or the out-of-town hypermarket. Where I go one week is influenced by where I went the week before. If I go to the hypermarket one week, the probability of going to the hypermarket the next week is 0.1 and the probability of going to the supermarket is 0.3. If I go to the supermarket, those probabilities are both 0.3. If I go to the village shop those probabilities are both 0.4.

(i) I go to the hypermarket in week 1. Draw a tree diagram for the situation in weeks 2 and 3. Write down the probabilities that I go to the hypermarket in (a) week 2, (b) week 3. Find the probability that I went to the supermarket in week 2 given that I went to the hypermarket in week 3. (9)

(ii) I now go to the hypermarket in week 1 with probability 20/71 and to the supermarket with probability 24/71. Draw a second tree diagram showing the situation in weeks 1 and 2. Show that the probability of going to the hypermarket will still be 20/71 in week 2. (6)

6. Data were collected on the times taken by ten skiers to reach an intermediate checkpoint, and the finish, in a downhill race. The table shows the times x, at the checkpoint, expressed as the number of seconds after 1 minute 28 seconds; and the times y, at the finish, expressed as the number of seconds after 1 minute 53 seconds.

x	1.2	2.5	1.5	0.9	0.8	1.5	2.1	1.8	1.2	1.9
y	1.0	2.4	1.2	0.6	0.3	1.3	1.7	1.6	0.7	1.4

$$\sum x = 15.4 \qquad \sum x^2 = 26.34 \qquad \sum y = 12.2 \qquad \sum y^2 = 18.24 \qquad \sum xy = 21.68$$

(i) Draw a carefully labelled scatter plot of the data. (5)

(ii) Calculate $\bar{x}$, $\bar{y}$, $\Sigma(x-\bar{x})^2$, $\Sigma(y-\bar{y})^2$, $\Sigma(x-\bar{x})(y-\bar{y})$ and the product-moment correlation coefficient. (5)

(iii) Comment on both the sign and the magnitude of the coefficient. (2)

(iv) Calculate the equation of the regression line of y on x and hence estimate the average finishing time of skiers who reach the checkpoint in a time of 1 minute 30 seconds. (5)

(v) Give the coordinates of two (x, y) points which lie on your line and plot the line on the scatter plot. (3)

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

(a) Trend.

(b) Seasonal variation.

(c) Additive model.

(3)

(ii) The values of quarterly sales in a furniture store over a 3½-year period are shown in the table, along with the 4-quarterly trend, calculated by the method of moving averages and centred.

Year	*Qtr*	*Sales(£000)*	*Trend(£000)*
2003	4	45	
2004	1	216	
	2	245	153.000
	3	67	175.000
	4	123	198.750
2005	1	314	226.375
	2	337	255.875
	3	196	282.125
	4	230	308.250
2006	1	417	333.000
	2	443	354.875
	3	288	377.875
	4	313	
2007	1	518	

Calculate the average seasonal variations for the four quarters, using an additive model. (There is no need to copy the table, but you should show your calculations clearly.)

(7)

(iii) Comment on what your calculations tell you about the sales in this store.

(1)

(iv) Say, with a reason, why a multiplicative model, rather than an additive model, might be appropriate for these data.

(1)

8. In an old recipe book belonging to my mother, I found that she had put the price of each item in 1975 next to the ingredients of a Christmas pudding recipe. A selection of the ingredients is shown in the table, along with the prices of these items in 1975 and in 2005.

Ingredient	*Price (£) in 1975*	*Price (£) in 2005*
Sugar (per kg)	0.16	0.72
Eggs (per dozen)	0.42	1.60
Raisins (per kg)	0.57	1.46
Ground almonds (per kg)	2.46	4.25
Brandy (70cl bottle)	2.20	9.50

(i) Calculate, correct to 1 decimal place, a simple mean of price relatives index number for 2005 using 1975 as base.

(3)

(ii) Comment on what this index tells you about the change in the cost of living over this 30-year period.

(2)

(iii) The recipe uses 200g of sugar, 4 eggs, 450g of raisins, 100g of ground almonds and 7cl of brandy. Calculate, correct to 1 decimal place, a base-weighted price relative index number for 2005 using 1975 as base and comment on its value.

(4)