THE ROYAL STATISTICAL SOCIETY

2004 EXAMINATIONS – SOLUTIONS

ORDINARY CERTIFICATE

PAPER II

The Society provides these solutions to assist candidates preparing for the examinations in future years and for the information of any other persons using the examinations.

The solutions should NOT be seen as "model answers". Rather, they have been written out in considerable detail and are intended as learning aids.

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

Ordinary Certificate, Paper II, 2004. Question 1

A quantitative variable is one measured on a numerical scale, while a qualitative variable is not numerical but categorical, each item being assigned to one of a set of categories.

Qualitative variables can be nominal or ordinal.

The categories of a nominal variable cannot be put into an order; they may for example be colours of objects (e.g. vehicles), source of origin (e.g. food products from different parts of the world) or ethnic group.

Ordinal variables can be arranged in a logical order, such as for example in a 1-to-5 scoring scale for an opinion (e.g. disagree up to strongly agree), sizes of motor cars (small, medium, large), vigour of plants (weak, average, good, very good).

Discrete variables are usually counted as integers, for example the number of vehicles passing along a road or the number of insects found on the leaves of a plant.

Continuous variables are precise measured variables, such as lengths, heights, times, which can take any value within a range.

Nominal : a bar chart with categories in no special order

Ordinal : a bar chart in the order given by the categories

Discrete : a bar diagram in numerical order of the variable

Continuous : a histogram with intervals in increasing order of the value of the variable.

Ordinary Certificate, Paper II, 2004. Question 2

(i) Stem and leaf diagrams, using £10 units as stems and £1 as leaves, for the two categories are as follows, after ordering the leaves for each stem.

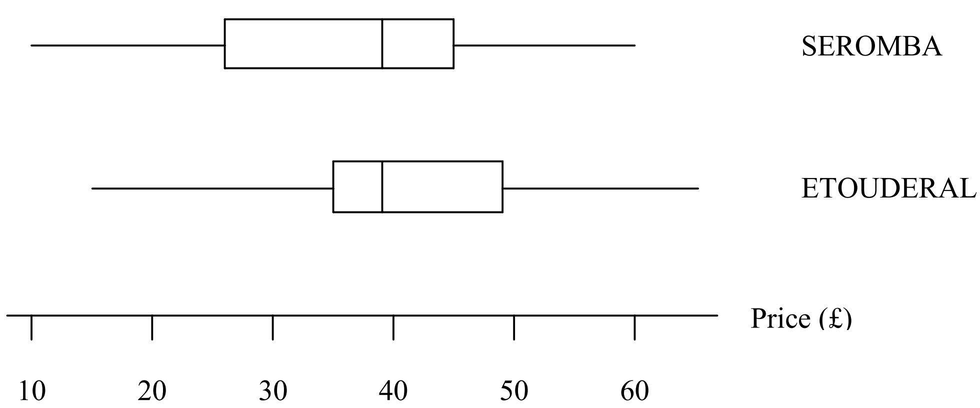
SEROMBA		Frequency	Cumulative frequency
1	0 6 6 8	4	4
2	2 3 4 4 5 5 5 6 6	9	13
3	0 0 0 0 0 0 5 5 5 7 8 8	12	25
4	0 0 0 0 0 0 0 0 5 5 5 5 5 5 7	15	40
5	0 0 0 0 2 5	6	46
6	0 0 0 0	4	50

ETOUDERAL		Frequency	Cumulative frequency
1	5 9	2	2
2	3 5 5 6 6 9 9 9 9	9	11
3	5 5 5 5 5 5 5 5 5 5 5 9 9 9 9 9 9 9 9 9	20	31
4	5 5 5 5 5 5 9 9 9	9	40
5	5 5 5 5 9 9 9	7	47
6	5 5 5	3	50

(ii) There are 50 prices in each list. The median is therefore midway between the 25th and 26th, and so is £39 for each [Seromba: ½(38+40); Etouderal: ½(39+39)].

The upper quartile is the [¾(50+1)]th value in the list, i.e. the 38¼th item; this is £45 for Seromba and £49 for Etouderal. The lower quartile is similarly the 12¾th item, which is £26 for Seromba and £35 for Etouderal. [Note. Alternative definitions of quartiles for discrete data sets are sometimes used. In the present cases, they would make no difference due to the repeated values in the lists.]

The maxima and minima are also used in the boxplots.



(iii) Both have a range of £50, Seromba from £10 to £60 and Etouderal from £15 to £65. Both have a median of £39. Seromba has more at the lower end of the price range. Etouderal has a large number at £35 – £39.

(iv) Almost all Etouderal products have prices ending in 5 or 9, the 9 being no doubt intended to give an impression that they are not all that expensive. Seromba has several ending in zeros, and none with 9s.

Ordinary Certificate, Paper II, 2004. Question 3

(i)

	Value	Minimum possible	Maximum possible
	2.7	2.65	2.75
	3.8	3.75	3.85
	3.0	2.95	3.05
	4.4	4.35	4.45
Total	13.9	13.7	14.1

(ii) Minimum for mean = 13.7/4 = 3.425.

Maximum for mean = 14.1/4 = 3.525.

(iii) Minimum possible SD = 0.7182, maximum = 0.8261 (these are given in the question).

The coefficient of variation is $\left(100\,\dfrac{\text{SD}}{\text{Mean}}\right)\%$.

$\dfrac{\text{Min SD}}{\text{Max mean}} = \dfrac{0.7182}{3.525} = 0.2037$. $\dfrac{\text{Max SD}}{\text{Min mean}} = \dfrac{0.8261}{3.425} = 0.2412$.

Hence the minimum possible coefficient of variation is 20.4% and the maximum is 24.1%.

Ordinary Certificate, Paper II, 2004. Question 4

(i) The proportion with no amenities is $1 - 0.84 = 0.16$.

The proportion with 3 is given as 0.16 and the proportion with 2 or more as 0.36. Hence the proportion with 2 is $0.36 - 0.16 = 0.20$.

This leaves a proportion of 0.48 with 1 amenity.

Summary table:

Number of amenities	0	1	2	3
Proportion	0.16	0.48	0.20	0.16

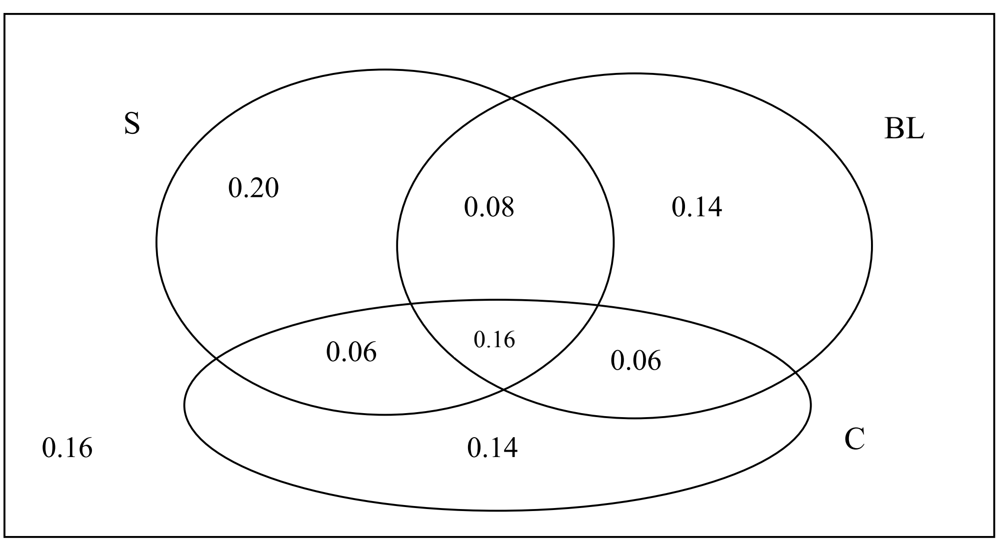
Mean $= (0 \times 0.16) + (1 \times 0.48) + (2 \times 0.20) + (3 \times 0.16) = 1.36$.

(ii) Put x = proportion having just C = proportion having just BL. The proportion having just S is then $x + 0.06$.

So, for one amenity, we have $0.48 = x + x + x + 0.06 = 3x + 0.06$, giving $3x = 0.42$ so that $x = 0.14$.

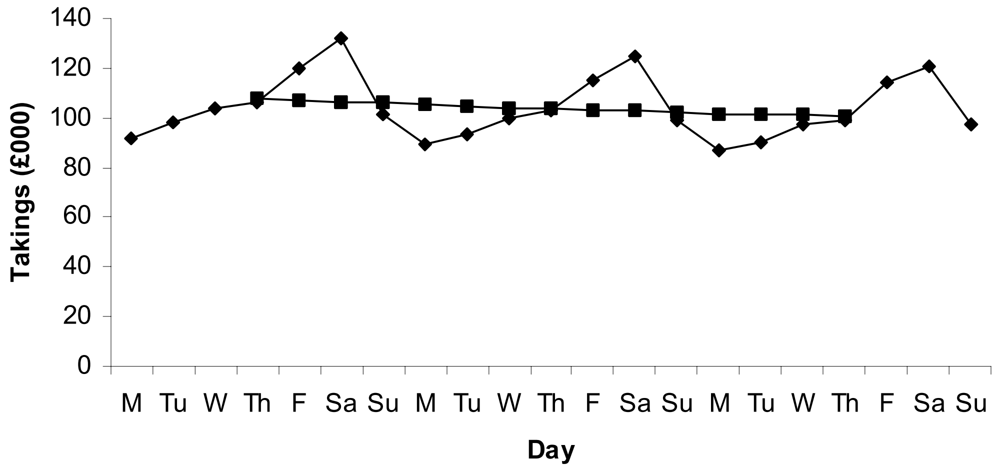
Now put y = proportion with (C+BL) = proportion with (C+S). Then the proportion having (S+BL) is $y + 0.02$. Thus for two amenities we have $3y + 0.02 = 0.20$, so $y = 0.06$.

The diagram can now be completed.



Ordinary Certificate, Paper II, 2004. Question 5

(i) The chart shows actual takings (diamond symbols) and the 7-day moving average calculated in part (iii) (square symbols).



(ii) Seasonal variation is a regular short-term variation from the trend. Here there is a 7-day variation, with more sales on Friday and Saturday, dropping on Sunday and Monday, and then increasing again steadily through the week.

(iii) A 7-day moving average will be needed.

	Takings (£000)	7-day total	7-day moving average
M	92		
Tu	98		
W	104		
Th	106	753	107.57
F	120	750	107.14
Sa	132	745	106.43
Su	101	741	105.86
M	89	738	105.43
Tu	93	733	104.71
W	100	726	103.71
Th	103	724	103.43
F	115	722	103.14
Sa	125	719	102.71
Su	99	716	102.29
M	87	712	101.71
Tu	90	711	101.57
W	97	707	101.00
Th	99	705	100.71
F	114		
Sa	121		
Su	97		

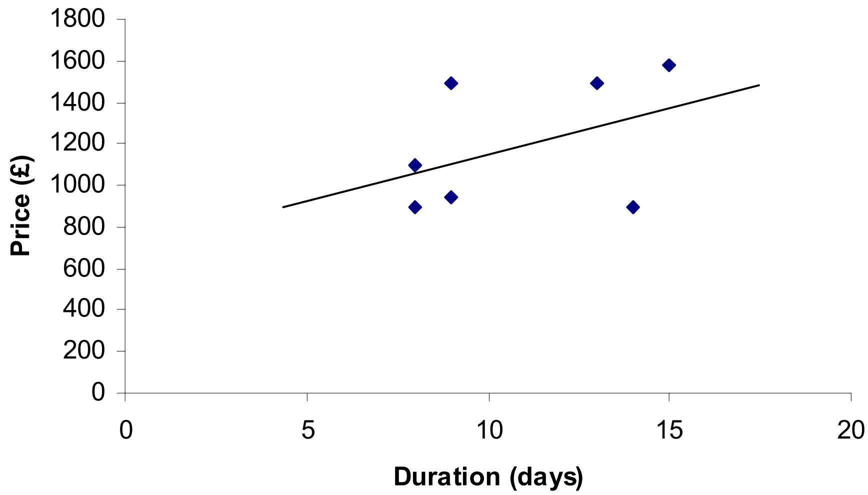
(iv) The trend is slowly downwards, almost linear.

<u>Ordinary Certificate, Paper II, 2004. Question 6</u>

(i) There are $n = 8$ destinations. Sample mean $\bar{x} = 84/8 = 10.5$ days.

Sample standard deviation $= \sqrt{\frac{1}{7}\left(944 - \frac{84^2}{8}\right)} = \sqrt{\frac{62}{7}} = 2.976.$

(ii)



There is very little suggestion of a linear relationship. It is possible to find holidays of the same length at very different prices.

(iii) $\bar{y} = 9490/8 = 1186.25.$

The estimate of the slope parameter is $\dfrac{102230 - (84)(9490)/8}{944 - 84^2/8} = \dfrac{2585}{62} = 41.69.$

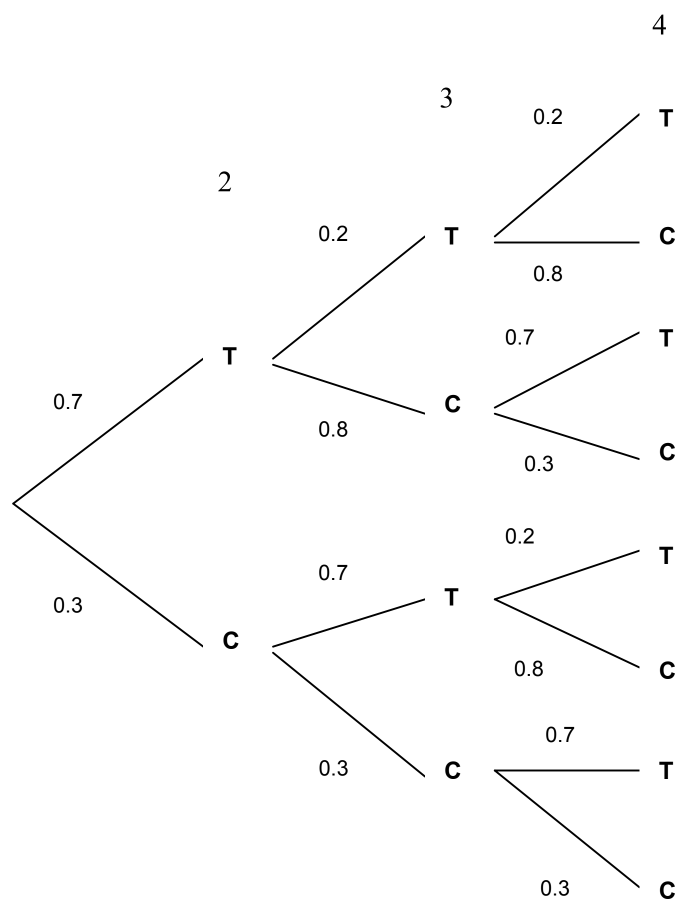
So the line is $y - 1186.25 = 41.69(x - 10.5) = 41.69x - 437.78$, i.e. $y = 748.5 + 41.7x$.

[This is plotted on the diagram above. Note for plotting that, for example, the line passes through $(\bar{x}, \bar{y})$ and (15, 1374).]

(iv) With $x = 21$, we get $y = 1624(.2)$. However, the data are very variable and predictions are therefore unreliable. Also there are no data above $x = 15$, so to extend the line to $x = 21$ is extremely hazardous.

Ordinary Certificate, Paper II, 2004. Question 7

(i) In the tree diagram, C represents cursory check and T represents thorough check. The tree "starts" with C at week 1; weeks 2, 3 and 4 are as shown.



(ii) From the tree diagram, the probabilities are (a) 0.7, (b) $(0.7\times0.2) + (0.3\times0.7) = 0.35$, (c) $(0.7\times0.2\times0.2) + (0.7\times0.8\times0.7) + (0.3\times0.7\times0.2) + (0.3\times0.3\times0.7) = 0.525$.

(iii) For $P(2|4)$, we have $P(2|4) = \dfrac{P(2 \text{ and } 4)}{P(4)}$ and, following the appropriate routes through the tree to obtain the numerator, we have

$$\frac{P(2 \text{ and } 4)}{P(4)} = \frac{(0.7\times0.2\times0.2)+(0.7\times0.8\times0.7)}{0.525} = \frac{0.42}{0.525} = 0.8 \ .$$

Ordinary Certificate, Paper II, 2004. Question 8

The table could be rewritten in the order of SMRs (increasing or decreasing equally suitable), with the statement "SMR for whole of South Trafford = 87" placed beneath it. (Possibly the statement "National SMR = 100" would also be useful.)

The number of deaths in South Trafford is 13% below the national average, but the electoral wards vary substantially. Sale Moor and St Martin's are over 20% higher than the national average, while Bowdon, Village, Mersey St Mary's, Hale and Timperley are all more than 25% less than the national average. Priory and Broadheath are very close to the national average, Altrincham is 15% below it and Broadlands 18% below it.

Why should Sale Moor and St Martin's be so high? Some study should be made of possible reasons, such as population profiles (age etc) and local conditions. Comparison with those at the opposite extreme of the table would be useful.