THE ROYAL STATISTICAL SOCIETY

2003 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.

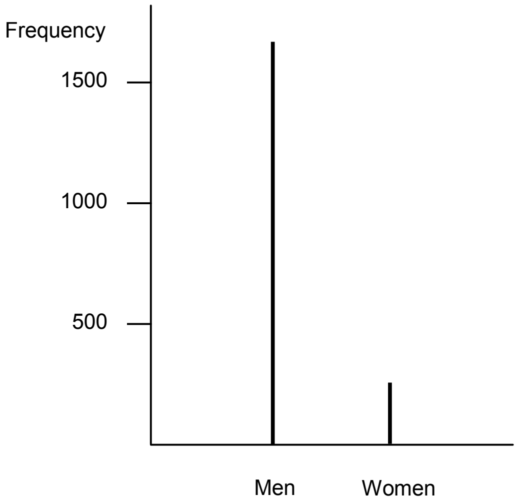
The Society will not enter into any correspondence in respect of these solutions.

Ordinary Certificate, Paper II, 2003. Question 1

(i) Letters to the Editor, 18 September 2001 – 12 January 2002.

Sex	Frequency
Men	1698
Women	244
TOTAL	1942

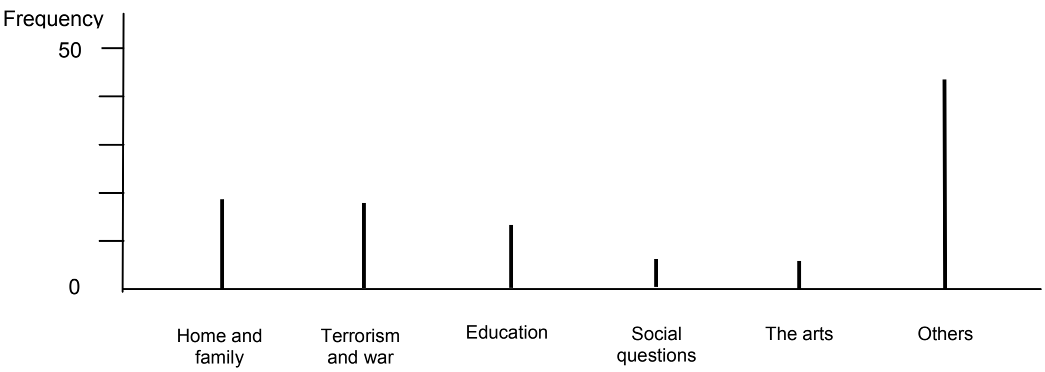
Source: "The Times", 21 January 2002.



(ii) Letters to the Editor from ladies, 18 September 2001 – 12 January 2002.

Subject	Frequency
Home and family	18
Terrorism and war	17
Education	12
Social questions	5
The arts	5
Others	43
TOTAL	100

Source: "The Times", 21 January 2002.



(iii) Pie charts are less easy to draw than bar charts, but they show proportions of the total rather than actual numbers of letters. In (i) this would be more useful. A bar chart would be better in (ii) to make exact comparisons between the subjects, although a pie chart would help to emphasise the proportion in the "others" category.

Ordinary Certificate, Paper II, 2003. Question 2

(i) One-fifth is 20%. The quotation has simply added 14% and 6%, instead of averaging them, weighted according to the numbers of each sex in the population.

If there were equal numbers, the weights would be ½. The percentage affected in the whole population would therefore be (½ × 14) + (½ × 6) = 10.

(ii) The reduction is actually $\frac{999-333}{999}=\frac{666}{999}=66.7\%$.

(The quotation had divided by 333, not 999.)

(iii) The further 10% reduction applies to the reduced premium, not the original, i.e. it is 10% of the 30% no-claims premium.

So the total reduction is $70+\frac{1}{10}(30)=73\%$, not 80%.

Ordinary Certificate, Paper II, 2003. Question 3

(i)

	Question 2 incorrect	*Question 2 correct*	*Total*
Question 1 incorrect	4	4	8
Question 1 correct	6	11	17
Total	10	15	25

(ii) (a) $\frac{17}{25} = 0.68$

(b) $\frac{15}{25} = 0.60$

(c) $\frac{11}{25} = 0.44$

(d) $\frac{21}{25} = 0.84$

(e) $\frac{11}{17} = 0.65$

(f) $\frac{4}{10} = 0.4$

(iii)

Number correct	0	1	2	*Total*
Frequency	4	10	11	25

$$\text{Mean} = \frac{(4\times0)+(10\times1)+(11\times2)}{25} = \frac{32}{25} = 1.28.$$

$$\text{Variance} = \frac{1}{24}\left(\sum fx^2 - \frac{\left(\sum fx\right)^2}{25}\right) = \frac{1}{24}\left(\left[(4\times0)+(10\times1)+(11\times4)\right]-\frac{32^2}{25}\right)$$

$$= \frac{1}{24}(54-40.96) = 0.5433, \text{ so standard deviation} = 0.737$$

[standard deviation = 0.72 if divisor 25 is used]

Ordinary Certificate, Paper II, 2003. Question 4

Salaries for EFG Bank:-

Salary (£)	*Frequency*	*Cumulative frequency*
Under £10,000	6200	6200
£10,000 but under £15,000	12000	18200
£15,000 but under £20,000	15600	33800
£20,000 but under £25,000	14200	48000
£25,000 but under £30,000	11900	59900
£30,000 but under £40,000	7000	66900
£40,000 but under £50,000	3500	70400
£50,000 but under £100,000	1500	71900
£100,000 or more	100	72000

(i) Median is at ½(m_1 + m_2) where m_1 and m_2 are the values of the 36000th and 36001th observations. There are 33800 up to £19999.5 [£19999.99 could be argued for, and similarly in the rest of the question, but this possibility is ignored; it would make hardly any difference to the answers]. 36000.5 – 33800 = 2200.5, and the next interval is £5000 wide with frequency 14200. So we need to go 2200.5/14200 of the way through it to locate the median, which will therefore be 19999.5 + (2200.5/14200)×5000 = £20774, or £20800 to the nearest £100.

For the upper quartile, ¾(72000) = 54000 [very slightly different definitions of percentiles are similarly ignored – they would make hardly any difference; the 6000/11900 below is similarly an approximation]. There are 48000 up to £24999.5. The next interval is £5000 wide with frequency 11900. So we need to go (approximately) 6000/11900 of the way through it, to Q_3 = 24999.5 + (6000/11900)×5000 = £27521, or £27500 to the nearest £100.

For the lower quartile, ¼(72000) = 18000 and thus we similarly get Q_1 = 9999.5 + (11800/12000)×5000 = £14916, or £14900 to the nearest £100.

For the 95th percentile, 0.95×72000 = 68400 and we similarly get that the 95th percentile is at 39999.5 + (1500/3500)×10000 = £44285, or £44300 to the nearest £100.

5% of 72000 is at 3600, which is in the "under £10000" group.

(ii)

	5th percentile	Q_1	Median	Q_3	95th percentile
UK	12100	15700	23200	32300	48600
EFG	< 10000	14900	20800	27500	44300

EFG's statistics are all lower than the corresponding UK ones, differences tending to increase further up the scale.

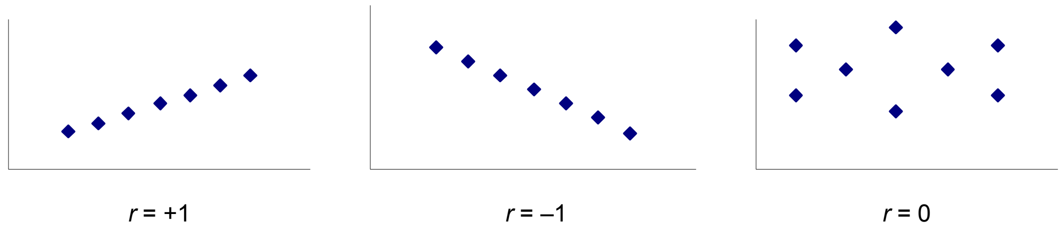
Possibly EFG has a different pattern of workforce compared with banks in general, with more younger workers and/or more part-timers, or more employees in call-centres.

(iii) The mean and standard deviation would be inflated by the salary figures in the (open-ended) top salary range, and we would need to make assumptions about the limits of the uppermost interval (and, less importantly, the lowest interval) in order to do the calculations. The percentiles allow more detailed comparison of the differences between UK and EFG than could be made using only the mean and standard deviation.

Ordinary Certificate, Paper II, 2003. Question 5

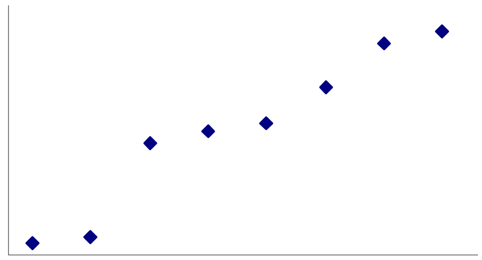
(i) r measures the degree of linear relationship between the two variables.

(ii)



(iii) The product-moment correlation coefficient uses the actual recorded values of x and y; Spearman's rank correlation coefficient ranks the values in order and uses only the rankings. Both coefficients are calculated from the same basic formula (although Spearman's is usually expressed in a different way).

(iv)



(v)

	A	B	C	D	E	F	G	H
Time rank	3	6	8	1	5	2	7	4
Speed rank	8	3	1	5	2	6	7	4
Difference (d)	–5	3	7	–4	3	–4	0	0

$\Sigma d^2 = 25 + 9 + 49 + 16 + 9 + 16 + 0 + 0 = 124$

$$r_s = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)} = 1 - \frac{6\times124}{8\times63} = -0.476\,.$$

Faster speed is associated with shorter time, so the coefficient is negative. But the association is not very strong and so r_s is not near to –1.

(vi) No effect, because the actual values are not used, only their rankings, which would not be altered.

Ordinary Certificate, Paper II, 2003. Question 6

(i) A weighted index number takes account of the amounts consumed, giving greater weight to the prices of fruit with higher consumption.

(ii) A Paasche index number is based on the current consumption pattern and so is more up to date.

[NOTE. If it is intended to go on monitoring price changes using an index number, a Laspeyres index, which is base-weighted, would be useful.]

(iii) The Paasche index is $\frac{\sum p_1 q_1}{\sum p_0 q_1}$, where p_0 and p_1 are 1999 and 2002 prices respectively and q_1 is 2002 consumption.

$$\sum p_1 q_1 = (2\times 0.75)+(20\times 0.15)+(1\times 0.75)+(10\times 0.35) = 8.75.$$

$$\sum p_0 q_1 = (2\times 0.60)+(20\times 0.12)+(1\times 0.80)+(10\times 0.25) = 6.90.$$

Therefore the Paasche index is $\frac{8.75}{6.90}\times 100 = 126.8$.

Fruit prices rose by 26.8%, using this index, from September 1999 to September 2002.

Ordinary Certificate, Paper II, 2003. Question 7

(i) (a) Trend is the underlying long-term movement of the series.

(b) The seasonal component is short-term regular variation about the trend, and can be daily or weekly as well as quarterly – even in some cases time of day.

(c) In an additive model, observations are expressed as the sum

trend + seasonal component + random residual "noise" .

(d) In a multiplicative model, these three items are multiplied together instead of being added.

An additive model is appropriate for a slowly changing trend component, while a multiplicative model is better when trend is changing rapidly. In a multiplicative model, it is assumed that the ratio of seasonal variation to trend is constant in each season over time.

(ii)

Year	Quarter	Passengers (x 100,000)	4-quarter mv avge	Trend	Actual/Trend (%) [for part (iii)]
1999	1	12			
	2	15	15.25		
	3	21	16.00	15.625	134.40
	4	13	17.00	16.500	78.79
2000	1	15	18.50	17.750	84.51
	2	19	19.50	19.000	100.00
	3	27	21.00	20.250	133.33
	4	17	23.25	22.125	76.84
2001	1	21	25.50	24.375	86.15
	2	28	25.25	25.375	110.34
	3	36	25.50	25.375	141.87
	4	16	26.50	26.000	61.54
2002	1	22	28.50	27.500	80.00
	2	32	30.75	29.625	108.02
	3	44	31.50	31.125	141.37
	4	25			
2003	1	25			

The trend rises until the second quarter of 2001, and then it rises again after a short period of being constant.

(iii) See the table above for calculations of (actual ÷ trend). Average seasonal variations (%) are calculated as follows:-

	Q1	Q2	Q3	Q4	
1999			134.40	78.79	
2000	84.51	100.00	133.33	76.84	
2001	86.15	110.34	141.87	61.54	
2002	80.00	108.02	141.37		
Mean	83.55	106.12	137.74	72.39	Sum = 399.8
x 400/399.8	**83.59**	**106.17**	**137.81**	**72.43**	Sum = 400 ✓

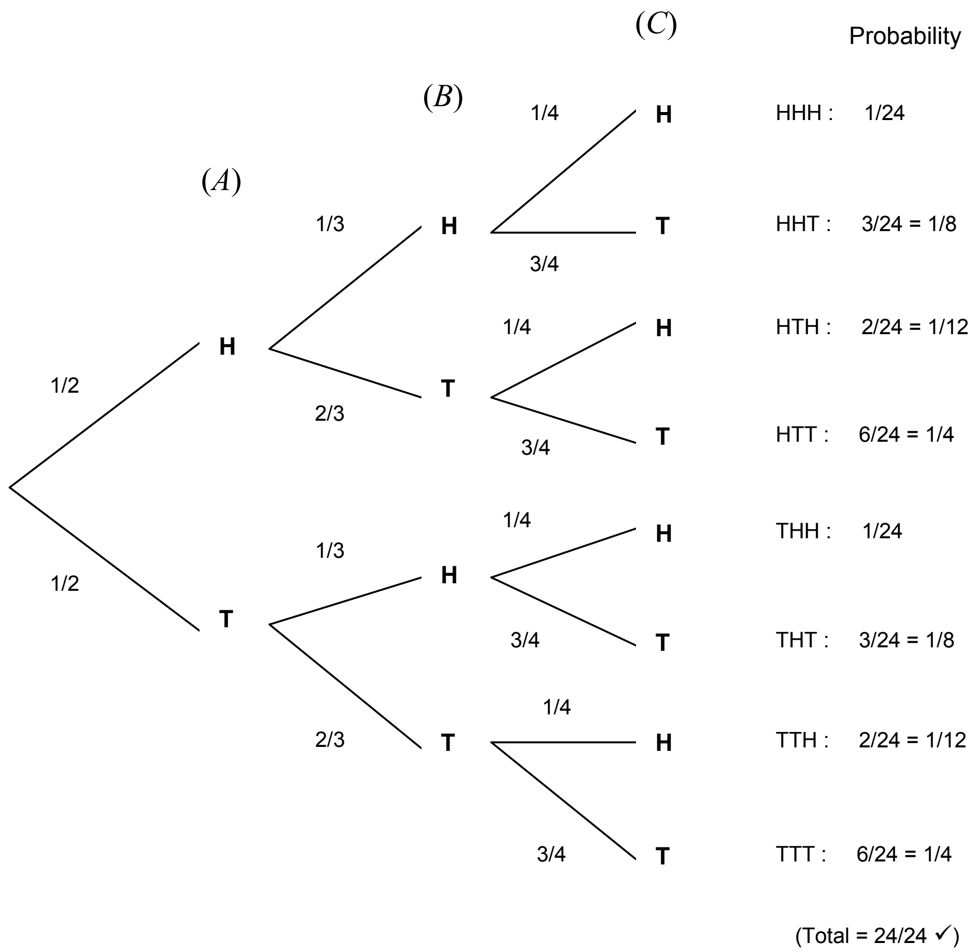
In quarters 1 and 4, actual numbers are below trend, namely about 84% and 72% of trend respectively (i.e. 16% and 28% below). In quarters 2 and 3, actual numbers are above trend, namely about 106% and 138% of trend respectively (i.e. 6% and 38% above).

Ordinary Certificate, Paper II, 2003. Question 8

(i) Coins *B* and *C* could have been tossed a large number of times, and the proportions of heads obtained. The estimates could then have been expressed to the nearest simple fraction.

[Coin *A* could have been dealt with in the same way, or if it was thought to be "fair" (e.g. a new undamaged coin) a smaller number of tosses could have been used and the result tested against the hypothesis that $P(\text{head}) = \frac{1}{2}$, using this as the "true" value of the probability if the hypothesis is not rejected.]

(ii)



(iii) $P(3 \text{ heads}) = 1/24$, $P(2 \text{ heads}) = 1/4$, $P(1 \text{ head}) = 11/24$, $P(0 \text{ heads}) = 1/4$.

(iv) $P(2 \text{ heads}) = 1/4$. For the outcomes where one head is on coin *A*, we have $P(\text{HHT}) + P(\text{HTH}) = 5/24$. Hence the required probability is $\frac{5}{24} \div \frac{1}{4} = \frac{5}{6}$.