THE ROYAL STATISTICAL SOCIETY

2001 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, 2001. Question 1

(i) p for $z = 0.714$ is given by

$$p \text{ for } 0.70 + \frac{0.014}{0.02}\left(p \text{ for } 0.72 - p \text{ for } 0.70\right)$$

$$= 0.7580 + \frac{0.014}{0.02}\left(0.7642 - 0.7580\right)$$

$$= 0.7580 + 0.7 \times 0.0062$$

$$= 0.7623$$

This can be rewritten as $0.7580 + 0.7(0.7642 - 0.7580)$

or $(0.3 \times 0.7580) + (0.7 \times 0.7642)$

where the weights are 0.3 and 0.7.

(ii) When $z = 0.70$, $p_{min} = 0.75795$ and $p_{max} = 0.75805$.

When $z = 0.72$, $p_{min} = 0.76415$ and $p_{max} = 0.76425$.

The minimum for $z = 0.714$ is $(0.3 \times 0.75795) + (0.7 \times 0.76415)$
$= 0.76229$ i.e. 0.7623.

The maximum is $(0.3 \times 0.75805) + (0.7 \times 0.76425)$
$= 0.76239$ i.e. 0.7624.

[Note that the values calculated above are not necessarily accurate to 4 decimal places.]

Ordinary Certificate, Paper II, 2001. Question 2

For clarity, use the hundreds as stems, tens as leaves and discard the units. [An alternative would be first to round to the nearest 10.] Times in seconds.

Unordered, reading across rows:

0	7 6 7 8 8
1	7 0 3 1
2	1 5 9 7 8 9 5 1 8 7 9
3	1 2 2 2 3
4	2 1 5 4 8 9
5	0 5 2 2 8 8 5 5 0

Ordered:

0	6 7 7 8 8
1	0 1 3 7
2	1 1 5 5 7 7 8 8 9 9 9
3	1 2 2 2 3
4	1 2 4 5 8 9
5	0 0 2 2 5 5 5 8 8

The median and quartiles help to give information about the distribution of times. The median is about 300 (midway between the 20th and 21st entries in the ordered stem-and-leaf diagram). The lower quartile (between 10th and 11th) is about 210 and the upper quartile is about 485. So a quarter of the calls last less than about 210, half less than about 300, and a quarter are longer than about 485.

Inspection also shows that 14 (i.e. about one-third) of calls are concentrated in the period 250 – 330. Five of the 40 calls were 540 or longer (i.e. 9 minutes or more).

So the calls range in length from approximately 1 to 10 minutes, with a substantial number between 4 and 5½ minutes.

Ordinary Certificate, Paper II, 2001. Question 3

Cumulative frequencies and percentage frequencies are:

	< 1	< 2	< 3	< 6	< 12	< 24	< 36	TOTAL
A	2	5	12	32	65	85	95	100
%	2	5	12	32	65	85	95	100
B	40	60	78	100	120	140	160	200
%	20	30	39	50	60	70	80	100

(i) For the graphs, see the next page.

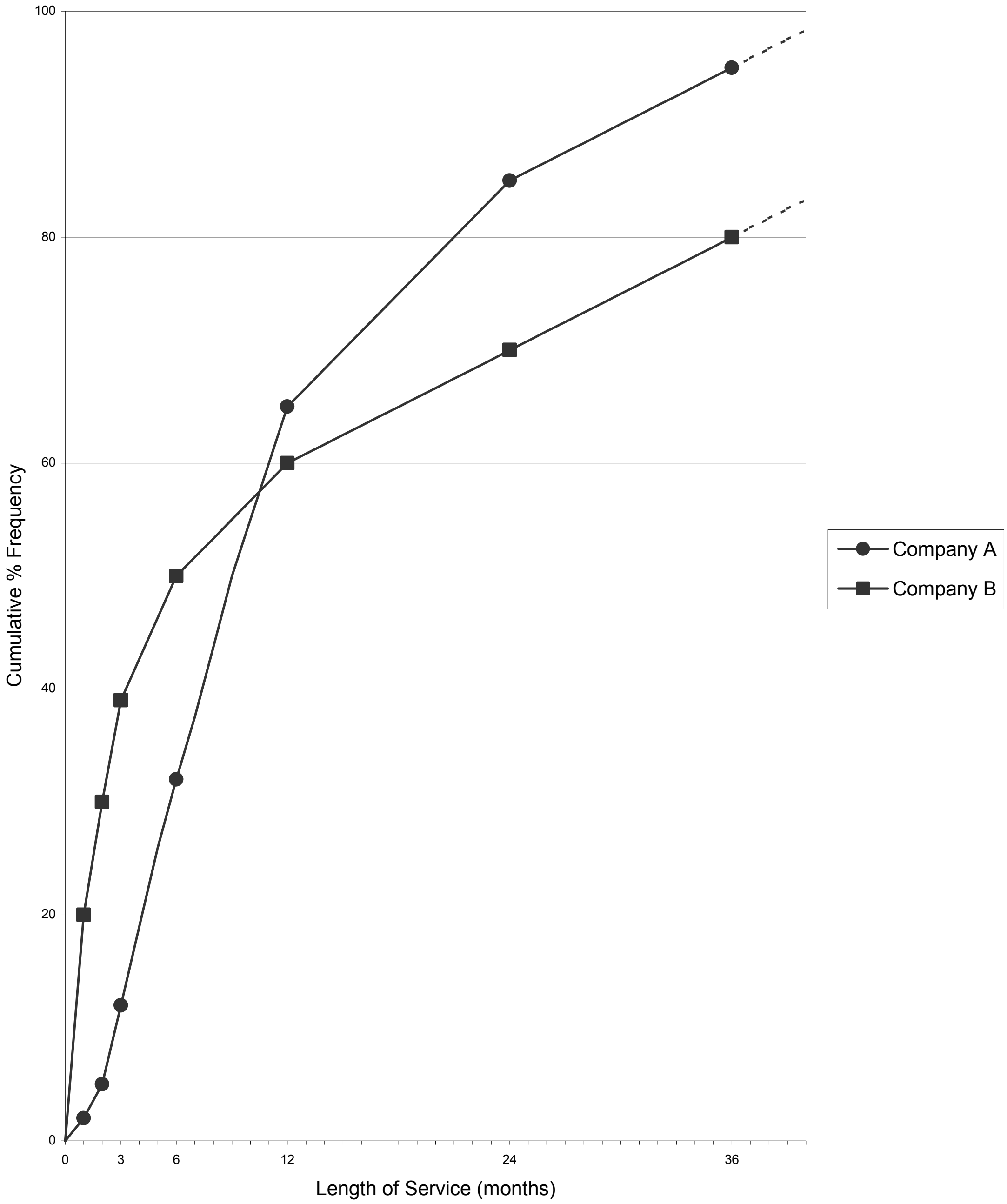
(ii) From the graphs, take the 50% points for (a) and (b), and use the 25% and 75% points in calculating (c) and (d).

(a) Median for *A* is 9 months.
(b) Median for *B* is 6 months.
(c) Quartiles for *A* are 5 months and 18 months (approximately) so the inter-quartile range is 13 months.
(d) Quartiles for *B* are 1½ months and 30 months (approximately) so the inter-quartile range is 28½ months.

(iii) Half of the employees at *A* leave within 9 months, at *B* within only 6 months. The inter-quartile ranges show that there is much more variability in the length of service in Company *B* than in *A*. A quarter of *B*'s employees stay more than 30 months, but for *A* the comparable figure is only 18 months. But a quarter of *B*'s employees leave within about 1½ months, whereas for *A* the comparable figure is about 5 months. So employees of *B* have a tendency either to leave almost at once or to stay a long time, whereas for *A* they do not begin leaving quite so soon but nearly all have left by the end of 3 years.

Continued on next page

Cumulative percentage frequency of length of service for companies A and B



Ordinary Certificate, Paper II, 2001. Question 4

Advantages: standard deviation is very useful in theoretical work and in statistical methods and inference.

Disadvantages: it is not easy to calculate, and its value is seriously influenced by extreme values in a set of data.

(i) For the Celsius temperatures, $\sum_{i=1}^{8} x_i = 146$, so $\bar{x} = 18.25$

$$\sum x_i^2 = 2738, \text{ so } s^2 = \frac{1}{7}\left(\sum x_i^2 - \frac{\{\sum x_i\}^2}{8}\right) = \frac{1}{7}(2738 - 2664.5)$$

$$= \frac{73.5}{7} = 10.5, \text{ so } s = 3.2404$$

(ii) If y = Fahrenheit temperature, we have $y = 32 + 1.8x$.

Hence $\bar{y} = 32 + 1.8\bar{x} = 64.85$

$s_y = 1.8 s_x = 1.8 \times 3.2404 = 5.8327$

(iii) The scaling $z = \dfrac{x - \bar{x}}{s_x}$ is to be applied.

Maximum and minimum values of x are 23 and 13, so for z they are

$\dfrac{23 - 18.25}{3.2404}$ and $\dfrac{13 - 18.25}{3.2404}$, i.e. 1.466 and –1.620.

The range is $1.466 - (-1.620) = 3.086$.

Ordinary Certificate, Paper II, 2001. Question 5

Probabilities are

		Male	Female	*Total*
Age	Under 25	0.15	0.12	0.27
	Over 25	0.45	0.28	0.73
	Total	0.6	0.4	1

using the information given.
(e.g. P(male and under 25) = 0.25 × 0.6 = 0.15).

(i) (a) 0.15; (b) 0.45; (c) 0.12; (d) 0.28.

(ii)

(a) P(under 25) = sum of entries in first row = 0.27.

(b) Males contribute 0.15 towards this 0.27, so required probability

$$=\frac{0.15}{0.27}=0.556.$$

(c) This is all except the class "female over 25", so required probability = 1 − 0.28 = 0.72.

(d) This is the complement of (c), so probability is 0.28.

(iii)

(a) The probabilities in this table must be multiplied by the corresponding probabilities in the table above, to obtain

$$(0.09\times 0.15)+(0.06\times 0.12)+(0.04\times 0.45)+(0.02\times 0.28)=0.0443.$$

(b) P(male and under 25|accident) =

$$\frac{P(\text{male and under 25 and accident})}{P(\text{accident})}=\frac{0.09\times 0.15}{0.0443}=0.305.$$

Ordinary Certificate, Paper II, 2001. Question 6

Pairs of data values may have been collected but not in the form of measurements. For example, two people may have tasted the same range of jams, each person ranking them from best to worst. Do these rankings agree, at least reasonably well, or are they very different? In other words, are the items ranked in approximately the same order by the two people? A rank correlation coefficient can be used to examine this.

The product-moment coefficient assumes measurements are bivariate Normally distributed. If this is not a reasonable assumption, the ranking of measurements rather than their actual values may be used to compare them. Data with extreme values in them can be handled in this way.

Spearman's rank coefficient uses rankings, and makes the same calculations on them as Pearson does on actual measurements.

A rank coefficient lies between -1 and $+1$.

(i)

Sample	(1)	(2)	(3)	(4)	(5)	(6)
Judge A	4	1	6	2	3	5
Judge B	4	1	6	2	3	5

(identical rankings)

(ii)

Sample	(1)	(2)	(3)	(4)	(5)	(6)
Judge A	3	4	1	5	6	2
Judge B	4	3	6	2	1	5

(diametrically opposite rankings)

(iii) See graph on next page. The last-but-one is the obvious outlier.

(iv) Rankings – smallest first.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
L	10	9	4	5	2	1	8	3	7	6
S	8	10	5	6	3	2	9	4	1	7
$d = L - S$	2	-1	-1	-1	-1	-1	-1	-1	6	-1

$\sum d^2 = 48.$ $\quad r_s = 1 - \dfrac{6\sum d^2}{n(n^2-1)} = 1 - \dfrac{6\times 48}{10\times 99} = 0.709$

Continued on next page

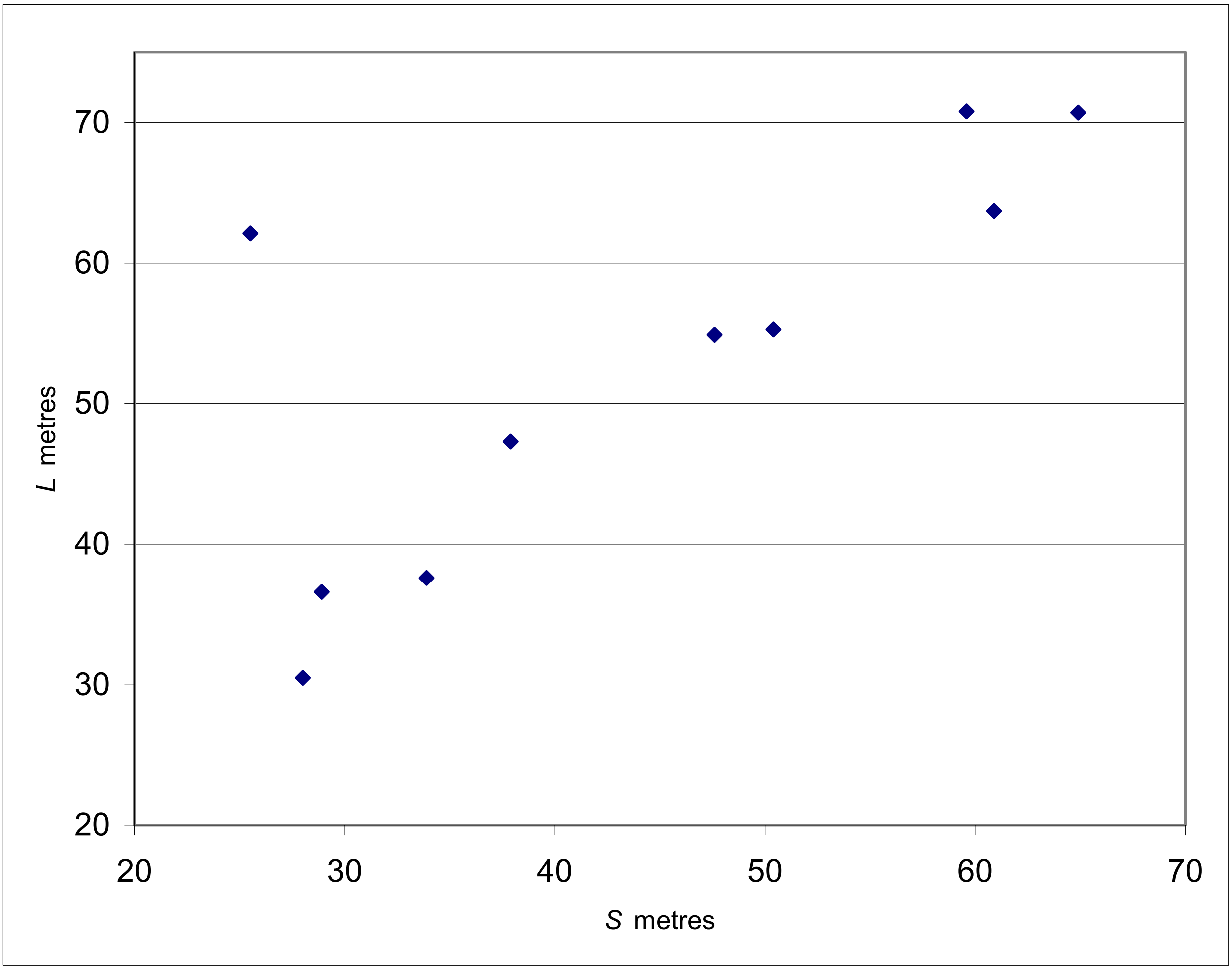
(v) Without the outlier ($L7$, $S1$), ranks become

L	9	8	4	5	2	1	7	3	6
S	7	9	4	5	2	1	8	3	6
d	2	−1	0	0	0	0	−1	0	0

$\sum d^2 = 6$. $\quad r_s = 1 - \frac{6 \times 6}{9 \times 80} = 0.95$

The rankings now agree almost perfectly. The outlier was clearly a different shape.

SCATTER DIAGRAM FOR PART (iii)



Ordinary Certificate, Paper II, 2001. Question 7

(i)

(a) The basic underlying long-term movement of the series.

(b) Short-term regular variation about the trend (e.g. seasonal, daily, time of day).

(c) A model in which the series value y is the sum of a trend component and a seasonal component and a random "residual" component.

(d) A model in which y is the product of these components.

If the trend is changing slowly, an additive model is likely to be suitable, but with rapid changes of trend the multiplicative version is preferable.

(ii) Taking seasonal variation as (column 2 – column 3):

Quarter	1	2	3	4	
1996			–34.125	15.000	
1997	6.000	19.125	–38.000	3.375	
1998	32.000	3.625	–37.375	8.125	
1999	16.875	9.625	–35.875	7.250	
(Unadjusted) average	18.292	10.792	–36.344	8.438	Total 1.178
					Adjustment 1.178/4 = 0.294
Adjusted average	17.998	10.498	–36.638	8.144	(0.002)
(to nearest 1000 units)	18	10	–37	8	

(iii) Predictions for 2001 (thousands of units) are

Quarter	1	2	3	4
	170 + 18 = 188	172 + 10 = 182	174 – 37 = 137	175 + 8 = 183

(Note. The predicted trend values for 2001 used in this calculation are given in the question.)

(iv) Unusual changes in the series, e.g. especially cold or warm winter causing particularly high or low usage, would make predictions inaccurate. Expansion of college activities would change the usage, and also make forecasts inaccurate. So would closure of a department using heavy electrical equipment.

Ordinary Certificate, Paper II, 2001. Question 8

(i) The correlation implies that as the age of men at marriage increases, so does that of women. They do not have to be equal, simply that younger men marry younger women and older men marry older women. If every man was 5 years older than his bride there would be correlation +1.

(ii) Index in 2000 compared with 1995 has increased $\frac{150.4}{124.2} \times 100\%$ i.e. 21.1%. This is the correct, direct comparison.

(iii) The statement would be true if we were given the median, not the mean. More than half will earn less than the mean, because the few highly paid specialists will make the distribution skew to the right.