THE ROYAL STATISTICAL SOCIETY

2005 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.

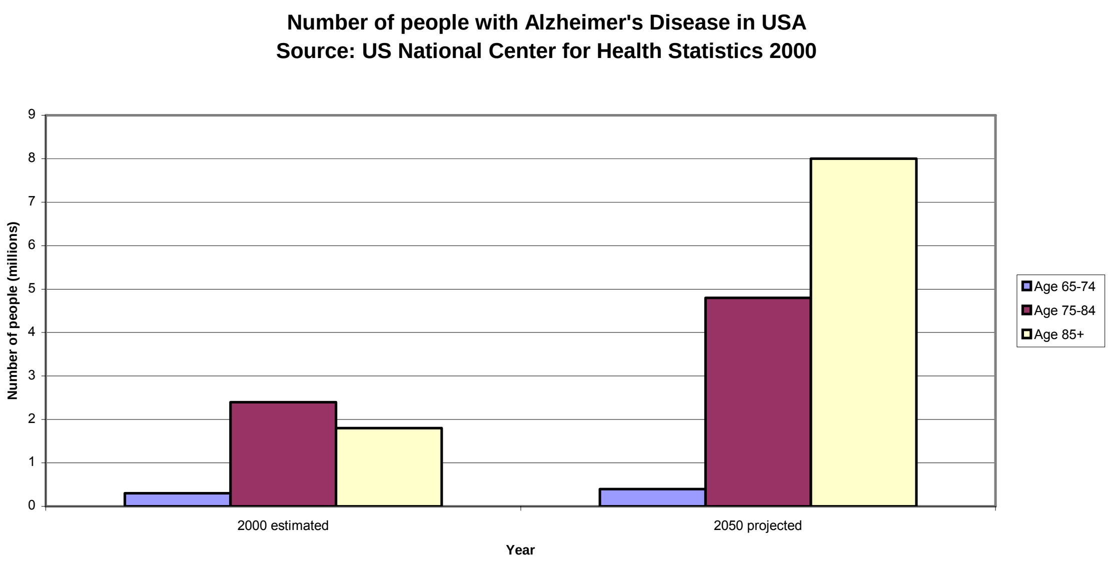
While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

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

Note. In accordance with the convention used in the Society's examination papers, the notation log denotes logarithm to base *e*. Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Ordinary Certificate, Paper II, 2005. Question 1

(i)



(ii) Total numbers in each age group increase in the projected figures, only slightly in the 65–74 group but substantially in 75–84 and even more in 85+. In the 2000 data, the number in the 85+ group is less than that in 75–84, but the prediction is that it will become much larger in the 85+ group by 2050. In 2000 there were relatively few in the 65–74 group, and this is predicted to continue to 2050.

(iii) Possible advantages of using a pie chart are

(*A*) it is more eye-catching, easier to interpret,
(*B*) it shows the proportions in each category.

Possible disadvantages are

(*A*) calculations (of angles) are needed before it can be drawn,
(*B*) there is no scale to read off actual numbers in each category,
(*C*) the absolute numbers with the disease are not shown.

<u>Ordinary Certificate, Paper II, 2005. Question 2</u>

(i) Maximum is 27 years minus 1 day; minimum 26 years exactly. If his birthday is on Census Day he will be recorded as having reached that age, and he will still be 26 until the actual day he reaches 27.

(ii) The minimum age difference is 1 year 1 day (husband is exactly 36 and wife will be 35 tomorrow). The maximum difference is 3 years minus 1 day (husband will be 37 tomorrow and wife is 34 today).

Mid-point of difference is 2 years.

(iii) (a) Differences (older – younger) are

2, 5, 11, 3, 4, 0, 3, 13, 3, 1. Mean $= \frac{45}{10} = 4.5$ years.

The sum of the squares of the differences is 4 + 25 + 121 + 9 + 16 + 0 + 9 + 169 + 9 + 1 = 363.

$$s^2 = \frac{1}{10}\left(363 - \frac{45^2}{10}\right) = 16.05$$ and therefore $s = 4.01$

regarding this as a <u>population</u> of couples, or

$$s^2 = \frac{1}{9}\left(363 - \frac{45^2}{10}\right) = 17.833$$ and therefore $s = 4.22$

regarding it as a <u>sample</u>.

(b) Differences (husband – wife) are

2, 5, –11, 3, –4, 0, –3, 13, 3, –1. Mean $= \frac{7}{10} = 0.7$ years.

$$s^2 = \frac{1}{10}\left(363 - \frac{7^2}{10}\right) = 35.81$$ and therefore $s = 5.98$ for a <u>population</u>, or

$$s^2 = \frac{1}{9}\left(363 - \frac{7^2}{10}\right) = 39.79$$ and therefore $s = 6.31$ for a <u>sample</u>.

Ordinary Certificate, Paper II, 2005. Question 3

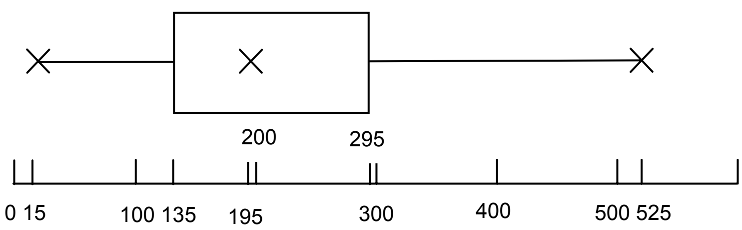
(i) These are the parking times to the nearest 10 minutes (below), i.e. 5 | 2 would be any time from 520 to 529 minutes. 5 is the "stem", which is in hundreds of minutes, and 2 is the "leaf", which is in tens of minutes. (Minutes less than 10 are ignored).

(ii) Six hours = 360 minutes. Seven vehicles stayed the full permitted time (or a *little* more), and only two were above (490 and 520). Six cars park for just below 60 minutes (1 hour), suggesting that there is one price for up to one hour and then an increase. There is a similar pattern leading up to 180 minutes (3 hours), pointing to a price increase at 3 hours. Also leading up to 240 minutes (4 hours) there is a concentration of leavers, suggesting another increase at 4 hours. There are no indications from these data of tariff change at 2 hours; there is a possibility at 5 hours.

(iii) The median is the $\frac{1}{2}(60+1)$th item in order in the list, i.e. between the 30th and 31st. Thus it is at 195 (with the assumptions given).

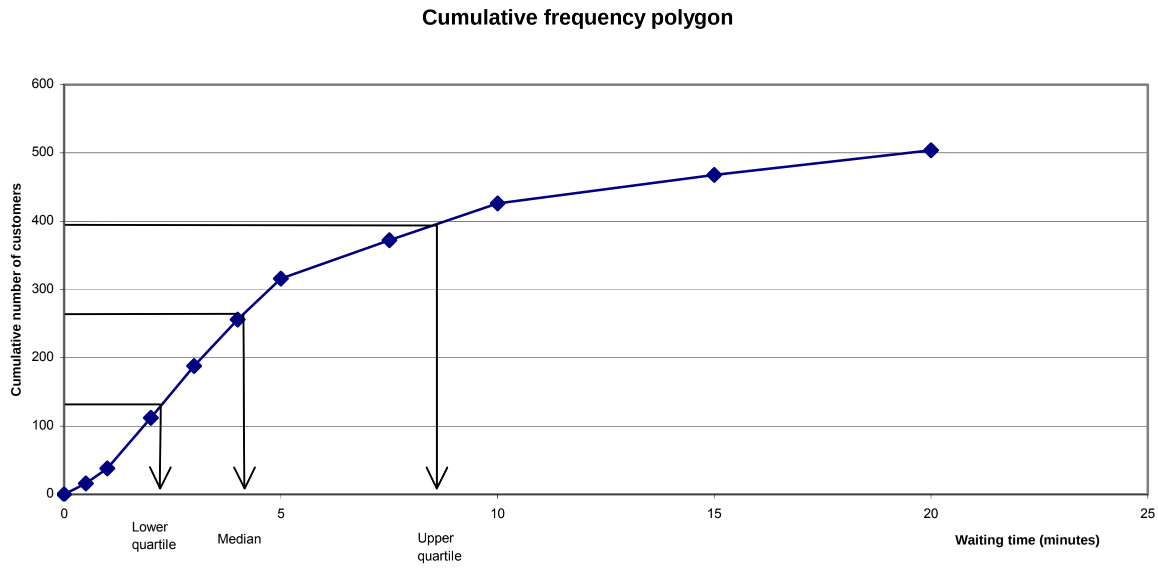
The lower and upper quartiles may be taken at the 15th and 45th observations in order (there are slightly differing conventions about this); these are at 135 and 295.

(iv) The first quartile is nearer to the median than the third quartile is, which indicates a distribution that is skew to the right (positively skew).



Ordinary Certificate, Paper II, 2005. Question 4

(i)



Cumulative frequencies are:

Up to	0.5	1	2	3	4	5	7.5	10	15	20	$\longrightarrow$
Cu freq	16	38	112	188	256	316	372	426	468	504	520

(ii) The median is at the $\left(\frac{520}{2}\right)$th item, i.e. it is found by drawing a horizontal line to the graph at height 260 and projecting down to the x-axis – this gives a waiting time of about 4.1 minutes. Similarly, using heights 130 and 390 in the y-direction, the quartiles are approximately 2.3 and 8.5 minutes.

(iii) The 260th customer in order of times is in the group 4 to 5. It is easier to calculate the median as $\frac{4}{60}$ of the way up from 4 (there are 256 up to 4 and 60 in the next interval, which is of width 1). Therefore $M = 4 + \frac{4}{60} = 4.07$ mins (i.e. 4 mins 4 seconds).

(iv) Half the customers wait less than 4.07 minutes BUT there is a very long tail to the right, with 10% waiting more than 15 minutes (and 25% more than 8.5 minutes (see (ii)).

Ordinary Certificate, Paper II, 2005. Question 5

(i) The inspector seems to believe that if only 3 eggs are used they will always be the 3 (out of 4) that do not contain salmonella! This is clearly not true, indeed totally lacking in logic. Unless we have tested them, we do not know; the bacterium is likely to be randomly distributed. We assume this in what follows.

(ii) Let S be the event "contains salmonella". We have $P(S) = 0.25$.

P(3-egg quiche salmonella free) = $(0.75)^3 = 0.422$.

(iii) Similarly, P(4-egg quiche salmonella free) = $(0.75)^4 = 0.316$.

(iv) The probability that an n-egg quiche is salmonella free is $(0.75)^n$. This is less than 0.1 if $n\log(0.75) < \log(0.1)$.

Hence the critical value of n is $\log(0.1)/\log(0.75) = \dfrac{-2.302585}{-0.287682} = 8.004$.

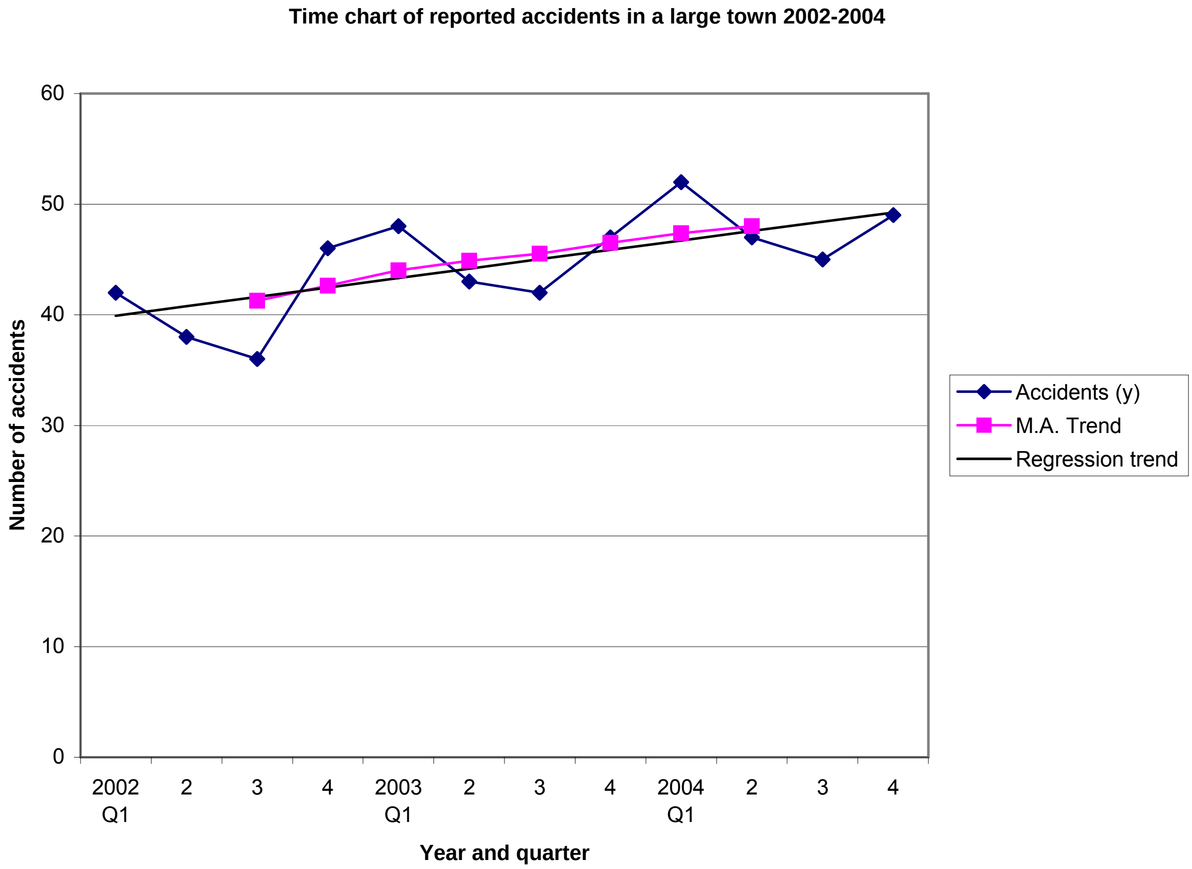
[Check: $(0.75)^8 = 0.1011$ and $(0.75)^9 = 0.07508$. (This result could be reached on a hand calculator very easily by successive multiplication.)]

We must take n at least 9, so the answer is $n \geq 9$.

(v) If locally produced eggs are used, independence is not likely since they are likely to have come from exactly the same source, so will all (or mostly) have the infection <u>or</u> all (or mostly) not. If however a large egg-packing company is used for supplies, independence is more likely since sources of supply will be numerous and mixing will occur.

<u>Ordinary Certificate, Paper II, 2005. Question 6</u>

(i)



(ii) The peaks and troughs recur at 4-quarterly intervals, so a 4-quarter moving average will smooth this out.

Solution continued on next page

(iii)

Year	Quarter	Quarter number (x)	Accidents (y)	4-quarter moving total	Sum of pairs of totals	Centred moving average trend
2002	1	1	42			
	2	2	38			
				162		
	3	3	36		330	41.250
				168		
	4	4	46		341	42.625
				173		
2003	1	5	48		352	44.000
				179		
	2	6	43		359	44.875
				180		
	3	7	42		364	45.500
				184		
	4	8	47		372	46.500
				188		
2004	1	9	52		379	47.375
				191		
	2	10	47		384	48.000
				193		
	3	11	45			
	4	12	49			
		$\Sigma x=78$	$\Sigma y=535$			

(iv) The moving average trend is very nearly linear.

Solution continued on next page

(v) The quarter numbers (x) have been added to table in part (iii).

Linear regression:

$\Sigma x^2 = 650$ $\Sigma y^2 = 24085$ $\Sigma xy = 3599$.

$\bar{x} = 6.5$, $\bar{y} = 44.583$.

$$\hat{b} = \frac{3599 - \frac{78 \times 535}{12}}{650 - \frac{78^2}{12}} = \frac{121.5}{143} = 0.8497$$

Line is $y - 44.583 = 0.8497(x - 6.5)$ or $y = 0.8497x + 39.060$.

(vi) The linear regression trend is very similar to the moving average, with the advantage of being estimated for <u>all</u> quarters (not omitting the end ones).

Ordinary Certificate, Paper II, 2005. Question 7

TEAM	A	I	F	S	E	SA	NZ	W
Scored	2	6	4	8	3	5	1	7
Conceded	1	3	6	7	2	5	4	8

(Highest scored ranked 1. Lowest conceded ranked 1.)

d (diff. in ranks)	1	3	–2	1	1	0	–3	–1

$\Sigma d^2 = 1 + 9 + 4 + 1 + 1 + 0 + 9 + 1 = 26.$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 26}{8\times 63} = 0.6905$$

The sample is too small for r_s to be worth testing for significance (in fact it is not significant) but the main contributions to Σd^2 are from Ireland and New Zealand. New Zealand scored most points but conceded a fair number; Ireland did not score very many.

<u>Ordinary Certificate, Paper II, 2005. Question 8</u>

(i) Cost relatives are (percentages):

$A \quad \dfrac{8.4}{7.5} = 1.120$ i.e. 112.0%;

$B \quad \dfrac{12.0}{10.2} = 1.176$ i.e. 117.6%;

$C \quad \dfrac{9.6}{9.8} = 0.980$ i.e. 98.0%;

$D \quad \dfrac{13.2}{12.4} = 1.065$ i.e. 106.5%.

(ii) Model C became 2% cheaper to produce, and the others were all more expensive: D by 6.5%, A by 12.0% , B by 17.6%.

(iii) $(200 \times 112.0) + (240 \times 117.6) + (750 \times 98.0) + (140 \times 106.5) = 139034$.

Total weekly in 2004 = 1330.

Therefore weighted index $= \dfrac{139034}{1330} = 104.5$.