ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

THE ROYAL STATISTICAL SOCIETY

2015 EXAMINATIONS − SOLUTIONS

ORDINARY CERTIFICATE – MODULE 2

The Society is providing these solutions to assist candidates preparing for the examinations in 2017.

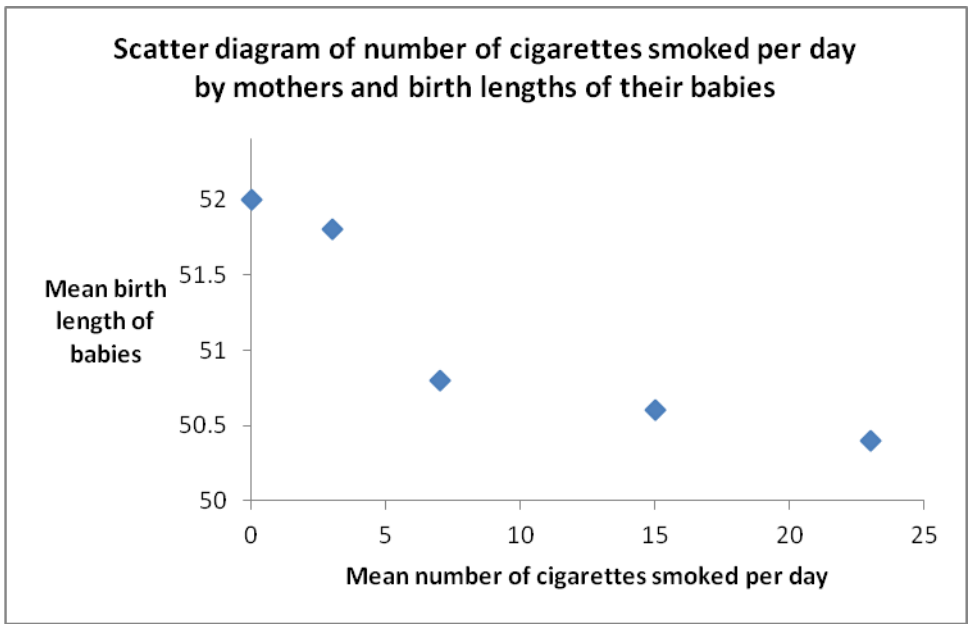
The solutions are intended as learning aids and should not be seen as "model answers".

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.

1(i)



(4)

(Title and axis labels (1) scales (1) points (1) presentation (1))

(ii) Birth length decreases with increasing cigarette consumption. (1)
Distinct difference *between* figures for low consumption and figures for higher consumption. (1)
Much less variation within figures for low consumption and figures for higher consumption. (1)

(iii) A regression line may be appropriate as a representation of the clear trend for birth length to decrease with increasing smoking.
However, there is some evidence of non-linearity. (NB: the data points are means for groups of size 250.)
It is difficult to say for sure when there are only 5 data points. (2)

(Award 1 mark for each of two of the above comments.)

Total (9)

2 (i)

	Northeast	Midwest	South	West	Totals
Decrease	30	***32***	51	56	169
Keep same	***143***	165	189	***166***	***663***
Increase	93	***158***	229	***74***	554
Totals	266	355	***469***	296	***1386***

(3)

(Deduct one mark per error down to zero.)

(ii)

	Northeast	Midwest	South	West
Decrease	11.3%	9.0%	**10.9%**	18.9%
Keep same	53.8%	46.5%	**40.3%**	56.1%
Increase	35.0%	44.5%	**48.8%**	25.0%
Totals	100.0%	100.0%	100.0%	100.0%

(2)

(Deduct one mark per error down to zero.)

(iii)(a) Support for decreasing expenditure broadly similar. (1)
Support for increasing expenditure, rather than keeping it the same, stronger in the Midwest than the Northeast. (1)

(iii)(b) Support for decreasing and for keeping it the same higher in West than South. (1)
Support for increasing very much stronger in the South than the West. (1)

(Total 9)

3 (i)	The broad trend is for the number of deaths to decrease.	(1)
	Both time periods show decreases in almost all cases,	(1)
	but the rate of decrease is greater in 2005-2010.	(1)
(ii)	Increases in deaths, or no decrease, could be caused by rapidly increasing rates of car ownership. E.g.: Hungary 2000-2005, Romania 2000-2010.	(1)(1)
	Another reason eg road safety campaigns; increasing affluence	(1)(1)
	(Accept other plausible explanations and examples.)	
(iii)	Greatest rate of decrease 2000-2010.	(1)
	Best: Estonia. Worst: Romania.	(1)(1)
	(Accept other similar statements)	
(iv)	Deaths per million of population; deaths per road mileage	(1)(1)
	(Accept other reasonable suggestions)	
		(Total 12)

4 (i) Mean: 13.5 (1)
Variance: 68.286 (2)
(Variance with divisor n is 59.75)

(ii) Overall reductions in mean systolic and mean diastolic. (1)
Systolic reductions greater than diastolic. (1)
Greater variation in systolic reduction than in diastolic. (1)
For the last mark, a comment on greater coefficient of variation for diastolic reduction than for systolic receives 1 mark.

(iii) $\sum wx = 2718$ (1)
Covariance of w and x: 1.5 (divisor $n - 1$) or 1.3125 (divisor n) (1)
pmcc is 0.0595 (0.059461) (1)
Award all 3 marks for a correct answer irrespective of what working is shown.

(iv) Moderate positive correlation between u and v indicates that reductions in systolic pressure are greatest for those with high initial systolic. (1)
Moderate negative correlation between v and w indicates that systolic reduction is greater for those with lower initial diastolic. (1)
Approximately zero correlation between x and w indicates that there is no association between diastolic reduction and initial diastolic. (1)

(Total 12)

5 (i) P(all 4 have myopia) = $0.264 \times 0.145 \times 0.183 \times 0.207 = 0.00145$ (2)
Correct product (1), answer (1)

(ii) P(none have myopia) = $(1 - 0.264) \times (1 - 0.145) \times (1 - 0.183) \times (1 - 0.207) =$ 0.4077 (3)
1 – … (1), product (1), answer (1)

(iii) P(at least one has myopia) = 1 – P(none have myopia) = $1 - (1 - 0.239)^3 =$ 0.5593 (3)
1 – P(…) (1), $1 - (1 - 0.239)^3$ (1), answer (1)

(iv) Require smallest n such that $(1 - 0.239)^n < 0.05$. Smallest n is 11. (3)
Correct formulation (1), evidence of trial and error or taking logs (1), answer (1)

(Total 11)

6 (i) Simple price relatives:

Wheat	2.580	or	258.0	if base year = 100 convention is used
Barley	2.602		260.2	
Corn	3.441		344.1	(3)

Correct method (1), any correct answer (1), other two correct (1)

(ii) Average annual rates of inflation:

Wheat	2.579551^(1/10) – 1 = 0.099395 hence 9.9%	
Barley	10.0% (10.0345%)	
Corn	13.2% (13.1533%)	(3)

Correct method (1), any correct answer (1), other two correct (1)

(iii) $\sum P_{2002} Q_{2002} = 290141.8$ (1)

$\sum P_{2012} Q_{2012} = 454114.8$ (1)

Hence index is 454114.8 / 290141.8 = 1.565148 (1.565 or 156.5) (1)

(iv) $\sum P_{2002} Q_{2012} = 175707.1$ (1)

Hence Paasche index is 454114.8 / 175707.1 = 2.584499 (2.584 or 258.4) (1)

The Paasche index uses current year quantities to compare price levels. (1)

(Total 12)

7 (i) Clear rising trend. (1)
Seasonal variation across the year. (1)
Variation is not constant. (1)

(ii) The variation appears to increase steadily over time as the trend line rises. An additive model would give constant variation. A multiplicative model would give greater variation for higher trend values. (2)
Allow (1) for a partial explanation along these lines.

(iii) $285.75 = (0.5 \times 242 + (233 + \ldots + 278) + 0.5 \times 284) / 12$ (1)
$a = (0.5 \times 233 + (267 + \ldots + 284) + 0.5 \times 277) / 12 = 289.33$ (2)
Correct expression (1), answer (1)

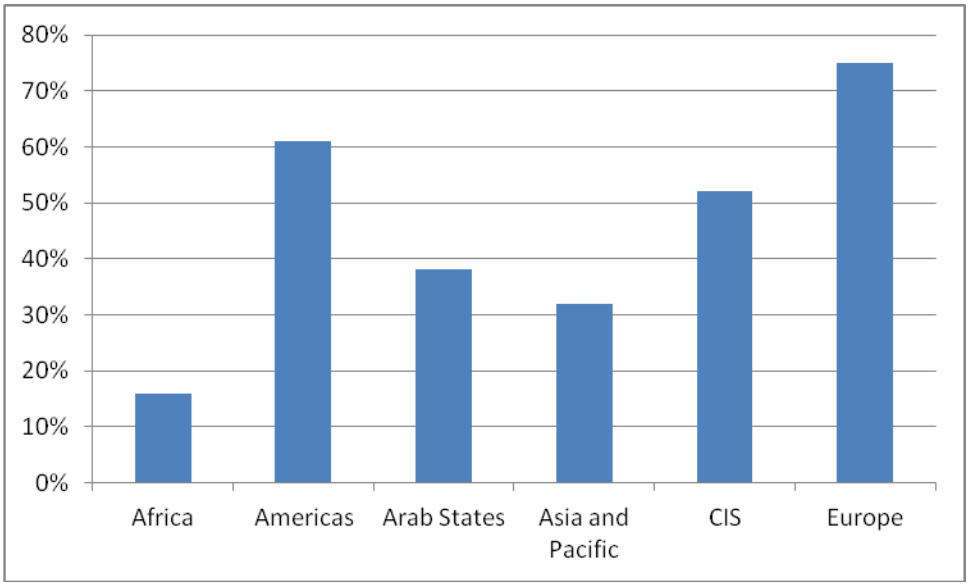
(iv) The additive differences from trend are: 78.25, 83.46, 95.54, 109.17, 117.29 (1)
The multiplicative differences are: 1.274, 1.253, 1.259, 1.286, 1.272 (1)
The additive differences are clearly increasing by a substantial amount. (1)
The multiplicative differences are nearly constant, (1)

(Total 12)

8 (i) $7.1 \times 0.39 = 2.8$ (billion) (1)

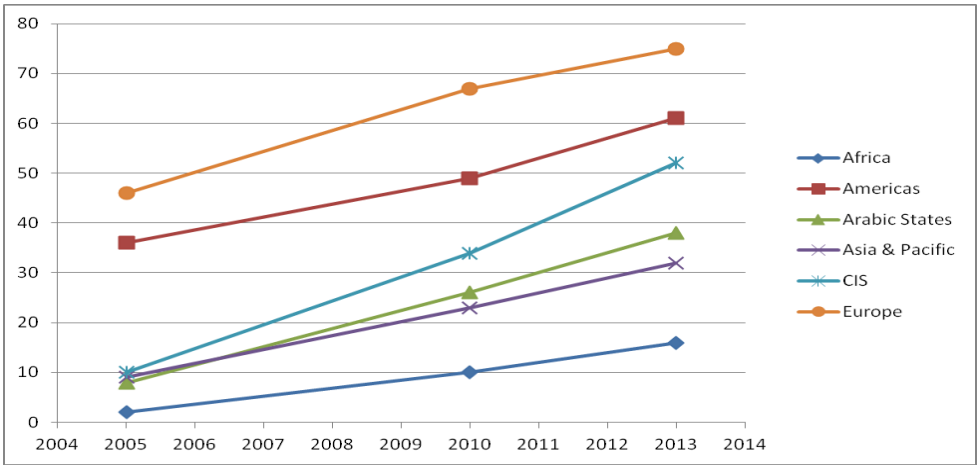
(ii) Developing world: $7.1 \times 0.82 = 5.8$ billion people
Of whom, $5.8 \times 0.31 = 1.8$ billion use the internet (1)
Developed world: 1.3 billion people
Of whom, $1.3 \times 0.77 = 1.0$ billion use the internet (1)
The total of 2.8 billion is consistent with part (i) (1)

(iii) Bar chart showing proportions of users of the internet by region of the world



Correct heights (1), correct labels (1), title (1)

(iv) Line graphs showing proportions of population using the internet in different regions of the world for the years 2005, 2010 and 2013



Correct graphs with labels (2), title (1)

(v) Regional groupings do not seem appropriate: e.g. Americas covers the highly developed North and the developing South; or Europe includes the economically advanced West and the less advanced East.

(2)
Answer (1), explanation (1)

(Total 12)

9 (i) The outlier is 128. It is a long way below all the other data points. (No formal calculation required.) (2)

(ii) An outlier may be a genuine item of data, in which case it may be appropriate to keep the outlier in the data set when analyzing the data. (1)
An outlier may be an error (or its status may be unclear), in which case it may be appropriate to remove it from the data set when analyzing the data. (1)

(iii) In this case, 128 might be an error: perhaps it should have been 182. Or it could be correct: 1.28m is a plausible height for someone with dwarfism. (2)

(iv) The mean would increase. (1)
The quartiles (1) and the median (1) would not decrease (might increase).
Any increase in the median is likely to be less than that in the mean. (1)
The standard deviation would decrease. (1)

(Total 11)