

EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

ORDINARY CERTIFICATE IN STATISTICS, 2017

MODULE 2 : Analysis and presentation of data

Time allowed: Three hours

*Candidates may attempt **all** the questions.*

The number of marks allotted to each question or part-question is shown in brackets.

The total for the whole paper is 100.

A pass may be obtained by scoring at least 50 marks.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in the Society's "Guide to Examinations" (document Ex1).

This examination paper consists of 8 printed pages.

This front cover is page 1.

Question 1 starts on page 2.

There are 10 questions altogether in the paper.

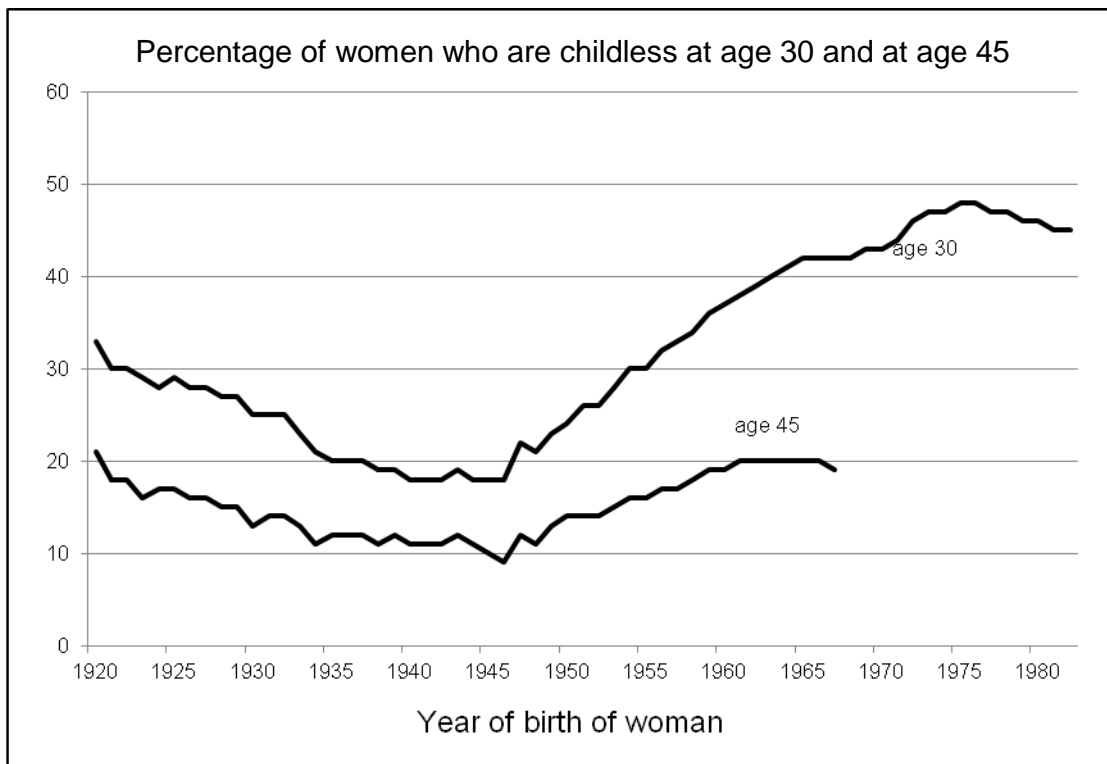
1. According to standard genetic theory, eye colour is independent of sex.

In a survey carried out in the United States, 2026 students (roughly half of them female and half male) completed a questionnaire in which they were asked the colour of their eyes; 33.4% of the females said that they had blue eyes, and 39.1% of the males said that they had blue eyes.

These data appear to contradict genetic theory. Give three distinct reasons why this might have happened. In each case discuss briefly how plausible the reason is.

(6)

2. The chart below relates to women in the UK. It shows the variation over time in the proportions who were childless at age 30 and age 45. (Source: Office for National Statistics, 2013)



- (i) Explain briefly what the two graphs show. Include in your answer an interpretation of the variation in the vertical distance between the two graphs.

(4)

- (ii) A newspaper article interpreted this chart as an indication that family size is decreasing. State whether or not that interpretation is correct, giving a reason for your answer.

(2)

3. The table below shows the proportions of citizens living in poverty in 18 countries in the European Union (EU) in 2010. The percentages are shown for the total population and by age band. Poverty is defined as living in a household with less than 60% of the median household income in the country of residence.

Poverty according to age (60% of median income) Units: %						
	Total	Less than 16 years	From 16 to 24 years	From 25 to 49 years	From 50 to 64 years	65 years or more
Czech Republic	9.0	13.6	12.9	8.0	6.8	6.8
Netherlands	10.3	13.5	18.6	9.3	7.6	5.9
Austria	12.1	14.7	13.1	10.6	9.4	15.2
Hungary	12.3	20.1	17.9	12.7	8.6	4.1
Sweden	12.9	12.4	26.9	10.6	5.9	15.5
Finland	13.1	11.2	23.3	9.9	10.3	18.3
Denmark	13.3	10.7	32.2	11.6	5.7	17.7
France	13.5	18.4	23.1	12.3	8.4	9.7
Belgium	14.6	18.5	14.8	11.4	12.3	19.4
Germany	15.6	17.2	19.1	14.1	17.0	14.1
United Kingdom	17.1	20.0	21.2	13.7	14.4	21.4
Poland	17.6	22.1	21.7	16.1	16.3	14.2
Portugal	17.9	20.9	21.9	14.9	16.1	21.0
Italy	18.2	24.3	24.2	17.7	13.1	16.6
Greece	20.1	22.3	27.8	18.1	17.3	21.3
Bulgaria	20.7	26.4	20.3	15.8	15.6	32.2
Spain	20.7	25.3	25.3	18.6	18.1	21.7
Romania	21.1	31.3	25.3	20.3	14.3	16.7
European Union	16.4	20.2	21.6	14.8	13.5	15.9

Source: Eurostat. Year of data: 2010

- (i) Compare and contrast the figures for poverty in Hungary and Sweden. (4)
- (ii) Another table (not shown here) gives the poverty thresholds in Austria and Romania as 958 and 176 euro per month respectively. Find the difference in the median incomes in these two countries. (2)
- (iii) An article based on these data states that 82.5 million citizens in the EU live in poverty. Use this figure to calculate the population of the EU. (2)

4. The table below is based on data collected in the 2011 census in the UK. The data shown are for the residents of one town in the North-West of England who were in work.

Respondents were asked to state their home postcode and the postcode of their place of work. This information was used to calculate the distance they travelled to work. Respondents were able to specify that they worked mainly at or from home, or that their work pattern did not involve regular travel to a fixed place of work. (The last group are indicated in the table as 'Other'.)

<i>Distance travelled to work</i>	<i>All persons</i>	<i>Males</i>	<i>Females</i>
All categories	137 978	71 329	66 649
Less than 10 km	80 753	34 806	45 947
10 km to less than 30 km	26 946	15 554	11 392
30 km and over	6 524	4 637	1 887
Work mainly at or from home	13 641	8 438	5 203
Other	10 114	7 894	2 220

- (i) Calculate, for males and females separately, the percentages working mainly at or from home, travelling less than 10 km, travelling between 10 km and 30 km, and travelling 30 km and over.

Present these figures in a suitable chart for comparing male and female travelling distances.

(6)

- (ii) Suppose that you were presenting a report about these data on a radio programme. Describe four features of the data as you would present them in your report.

(4)

5. A certain professional examination is taken by large numbers of candidates. Those who fail at the first attempt may make a second attempt if they wish. Those who fail at the second attempt are not permitted to take the examination again.

Past records show that 70% of candidates pass at the first attempt.

- (i) Find the probability that, in a randomly chosen group of three candidates,
(a) all three pass at the first attempt,
(b) exactly one passes at the first attempt. (4)

Of those who do not pass at the first attempt, two-thirds decide to resit the examination. Records show that 60% of resit candidates pass the examination.

- (ii) Find the probability that a randomly chosen candidate passes the examination at either the first or the second attempt. (3)

Distinctions are awarded to 30% of candidates who pass at the first attempt and to 20% of candidates who pass at the second attempt.

- (iii) Given that a randomly chosen candidate is awarded a distinction, find the probability that this candidate passed at the first attempt. (5)

6. The table shows the numbers of siblings (i.e. brothers and sisters) among a group of 450 university students, each from a different family.

Number of siblings	0	1	2	3	4	5	6	7
Number of responses	66	130	140	74	19	15	4	2

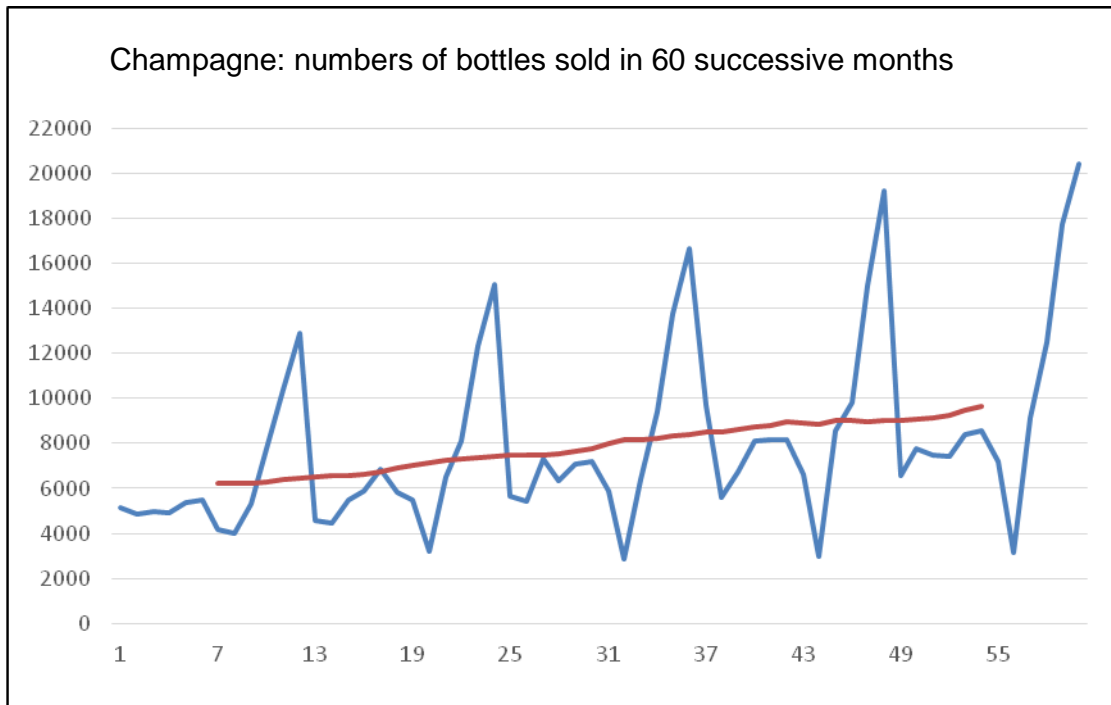
- (i) Find the mean and standard deviation of the number of siblings. (6)
- (ii) Deduce the mean and standard deviation of the number of children in the 450 families. (2)
- (iii) Give two distinct reasons why the mean obtained in part (ii) would not be a good estimate of the number of children per family in the population. (2)

7. The table shows price and circulation data for 10 national daily newspapers in the UK in June 2004 and June 2014. The price is in pence; the circulation, given in thousands, is the average number of copies sold per day in June. The newspapers are classified as 'Quality' or 'Popular' titles.

<i>Title</i>	<i>Type</i>	<i>Price 2004</i>	<i>Price 2014</i>	<i>Circulation (000s) 2004</i>	<i>Circulation (000s) 2014</i>
The Guardian	Quality	55	160	368	185
The Financial Times	Quality	100	250	390	197
The Times	Quality	50	120	614	373
The Daily Telegraph	Quality	60	140	916	515
The Independent	Quality	60	140	193	47
The Daily Mirror	Popular	35	55	1789	919
The Daily Mail	Popular	40	60	2222	1589
The Daily Express	Popular	40	55	926	480
The Sun	Popular	30	40	3254	2034
The Daily Star	Popular	35	40	884	467

- (i) Find the largest and smallest price relatives for the 10 newspapers in 2014, taking 2004 as the base year. (2)
- (ii) Find the average annual rate of price increase for The Financial Times from 2004 to 2014. (3)
- (iii) Calculate the percentage reduction in total circulation from 2004 to 2014 for
 (a) Quality titles,
 (b) Popular titles. (2)
- (iv) For each newspaper, calculate the percentage price increase from 2004 to 2014. Use these to calculate the simple (that is, unweighted) average percentage price increase from 2004 to 2014 for
 (a) Quality titles,
 (b) Popular titles.
 Comment on your results. (4)
- (v) Calculate the Laspeyres price index for the whole dataset.
 Explain briefly the underlying principle of the Laspeyres index and hence say how useful or otherwise it is in this case. (5)

8. The graph below shows the numbers of bottles of champagne sold by a large wine merchant in 60 successive months. It also shows a centred moving average.



- (i) If you had not been told that these data were monthly, what features of the graph would lead you to deduce that they were indeed monthly data? (2)
- (ii) Describe the trend and variation shown in the data. (3)
- (iii) Explain by reference to the graph whether an additive or multiplicative model would appear to be a better fit to the data.

Investigate whether or not your answer is correct by reference to the following extracts from the table of data used to draw the graph.

<i>Month</i>	<i>Actual sales</i>	<i>Moving average</i>
12	12 868	6451
24	15 058	7421
36	16 698	8414
48	19 218	9020

- (iv) Let the monthly sales figures be denoted by m_1, m_2, \dots, m_{60} . Give a formula for the first and a formula for the last of the centred moving averages. State clearly where each of these moving averages should be plotted. (3)

9. A random sample of 50 adult patients at a clinic had their percentage of body fat, x , and their body mass index, y , measured. The results are summarised as follows.

$$n = 50, \Sigma x = 894.1, \Sigma y = 1280.9, \Sigma x^2 = 20734.71, \Sigma y^2 = 33471.43, \Sigma xy = 24308.09.$$

- (i) Calculate the product moment correlation coefficient for the data. (3)
- (ii) Calculate the equation of the linear regression line of y on x . (3)
- (iii) Use the equation found in part (ii) to estimate the body mass index for a patient with 10% body fat. Discuss briefly whether it would be appropriate to use the equation to estimate the percentage of body fat for a patient with a body mass index of 25. (3)

10. Girls at a nationally representative sample of 100 schools in the United States were monitored for sports injuries during a school year. The sports monitored were soccer, volleyball, basketball and softball.

The table shows the number of sessions in which girls participated in the various sports, subdivided to show practice sessions and competitive sessions separately. The numbers of injuries sustained are subdivided similarly.

	Number of sessions		Number of injuries	
	Practice	Competitive	Practice	Competitive
Soccer	98 166	43 415	108	226
Volleyball	75 544	43 691	112	84
Basketball	132 836	53 325	182	192
Softball	88 362	46 727	70	83

- (i) In the following questions, you should support your answers with appropriate calculations.

Find the sport which has

- (a) the highest ratio of competitive to practice sessions, (2)
- (b) the lowest ratio of practice to competitive injuries, (2)
- (c) the highest injury rate in practice sessions, (2)
- (d) the lowest overall injury rate. (2)
- (ii) Discuss briefly the additional information you would wish to have before coming to a judgement about which of these sports is the most dangerous. (2)