OCR
Oxford Cambridge and RSA

Accredited

A Level Further Mathematics B (MEI)
Y422 Statistics Major
Sample Question Paper

Date – Morning/Afternoon
Time allowed: 2 hours 15 minutes

OCR supplied materials:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)

You must have:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)
- Scientific or graphical calculator

INSTRUCTIONS
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Complete the boxes provided on the Printed Answer Booklet with your name, centre number and candidate number.
- Answer **all** the questions.
- **Write your answer to each question in the space provided in the Printed Answer Booklet.**
- Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION
- The total number of marks for this paper is **120**.
- The marks for each question are shown in brackets **[ ]**.
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is used. You should communicate your method with correct reasoning.
- The Printed Answer Booklet consists of **20** pages. The Question Paper consists of **16** pages.

Section A (30 marks)

Answer **all** the questions.

1 In a promotion for a new type of cereal, a toy dinosaur is included in each pack. There are three different types of dinosaur to collect. They are distributed, with equal probability, randomly and independently in the packs. Sam is trying to collect all three of the dinosaurs.

(i) Find the probability that Sam has to open only 3 packs in order to collect all three dinosaurs. **[1]**

Sam continues to open packs until she has collected all three dinosaurs, but once she has opened 6 packs she gives up even if she has not found all three. The random variable X represents the number of packs which Sam opens.

(ii) Complete the table below, using the copy in the Printed Answer Booklet, to show the probability distribution of X.

r	3	4	5	6
$\mathrm{P}(X = r)$		$\frac{2}{9}$	$\frac{14}{81}$	

[1]

(iii) In this question you must show detailed reasoning.

Find

- $\mathrm{E}(X)$ and
- $\mathrm{Var}(X)$. **[5]**

2 The continuous random variable X takes values in the interval $-1 \le x \le 1$ and has probability density function

$$f(x) = \begin{cases} a & -1 \le x < 0 \\ a + x^2 & 0 \le x \le 1 \end{cases}$$

where a is a constant.

(i) (*A*) Sketch the probability density function. **[2]**

(*B*) Show that $a = \frac{1}{3}$. **[3]**

(ii) Find

(*A*) $P\left(X < \frac{1}{2}\right)$, **[2]**

(*B*) the mean of X. **[2]**

(iii) Show that the median of X satisfies the equation $2m^3 + 2m - 1 = 0$. **[3]**

3 A researcher is investigating factors that might affect how many hours per day different species of mammals spend asleep.

First she investigates human beings. She collects data on body mass index, x, and hours of sleep, y, for a random sample of people. A scatter diagram of the data is shown in Fig. 3.1 together with the regression line of y on x.

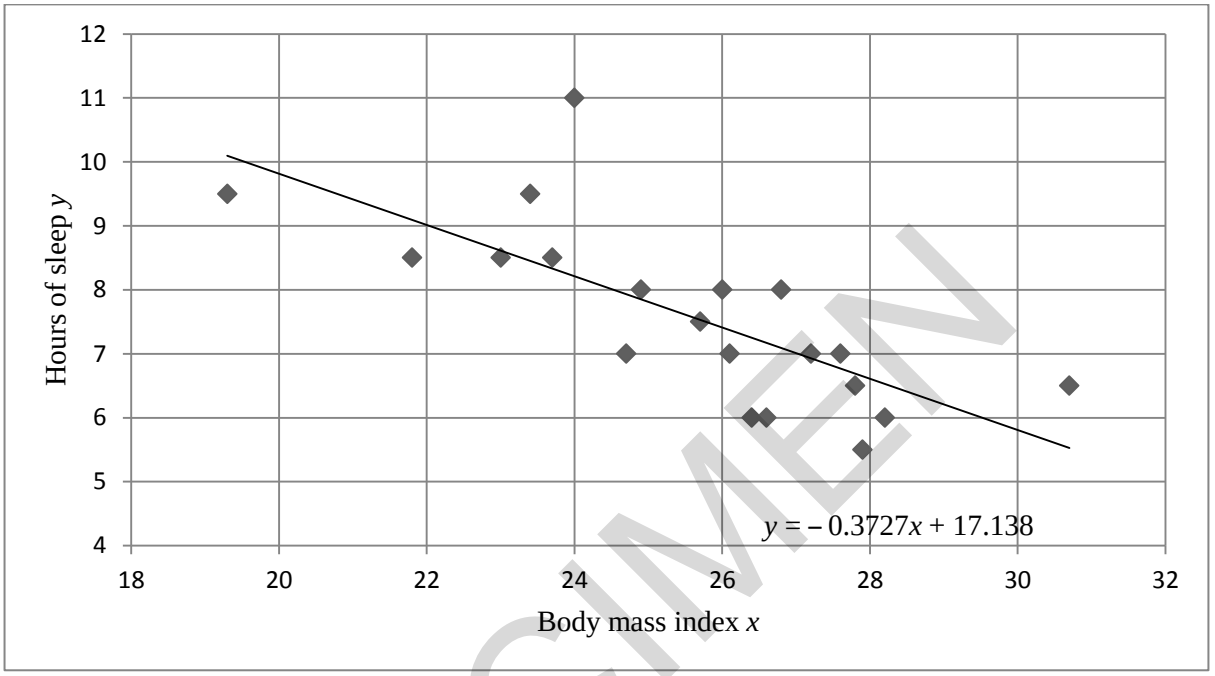


Fig. 3.1

(i) Calculate the residual for the data point which has the residual with the greatest magnitude. **[3]**

(ii) Use the equation of the regression line to estimate the mean number of hours spent asleep by a person with body mass index

(*A*) 26,

(*B*) 16,

commenting briefly on each of your predictions. **[4]**

The researcher then collects additional data for a large number of species of mammals and analyses different factors for effect size. Definitions of the variables measured for a typical animal of the species, the correlations between these variables, and guidelines often used when considering effect size are given in Fig. 3.2.

Variable	Definition
Body mass	Mass of animal in kg
Brain mass	Mass of brain in g
Hours of sleep/day	Number of hours per day spent asleep
Life span	How many years the animal lives
Danger	A measure of how dangerous the animal's situation is when asleep, taking into account predators and how protected the animal's den is: higher value indicates greater danger.

Correlations (pmcc)	Body Mass	Brain Mass	Hours of sleep/day	Life span	Danger
Body Mass	1.00				
Brain Mass	0.93	1.00			
Hours of sleep/day	−0.31	−0.36	1.00		
Life span	0.30	0.51	−0.41	1.00	
Danger	0.13	0.15	−0.59	0.06	1.00

Product moment correlation coefficient	Effect size
0.1	Small
0.3	Medium
0.5	Large

Fig. 3.2

(iii) State two conclusions the researcher might draw from these tables, relevant to her investigation into how many hours mammals spend asleep. **[2]**

One of the researcher's students notices the high correlation between body mass and brain mass and produces a scatter diagram for these two variables, shown in Fig. 3.3 below.

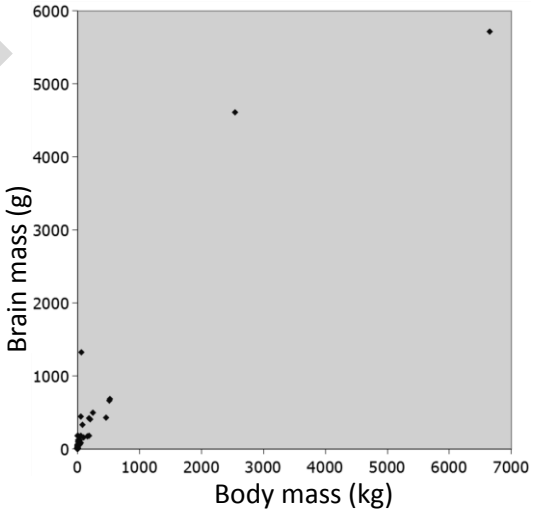


Fig. 3.3

(iv) Comment on the suitability of a linear model for these two variables. **[2]**

Section B (90 marks)

Answer **all** the questions.

4 A fair six-sided dice is rolled repeatedly. Find the probability of the following events.

(i) A five occurs for the first time on the fourth roll. **[1]**

(ii) A five occurs at least once in the first four rolls. **[2]**

(iii) A five occurs for the second time on the third roll. **[2]**

(iv) At least two fives occur in the first three rolls. **[2]**

The dice is rolled repeatedly until a five occurs for the second time.

(v) Find the expected number of rolls required for two fives to occur. Justify your answer. **[3]**

5 A particular brand of pasta is sold in bags of two different sizes. The mass of pasta in the large bags is advertised as being 1500 g; in fact it is Normally distributed with mean 1515 g and standard deviation 4.7 g. The mass of pasta in the small bags is advertised as being 500 g; in fact it is Normally distributed with mean 508 g and standard deviation 3.3 g.

(i) Find the probability that the total mass of pasta in 5 randomly selected small bags is less than 2550 g. **[3]**

(ii) Find the probability that the mass of pasta in a randomly selected large bag is greater than three times the mass of pasta in a randomly selected small bag. **[4]**

6 Fig. 6 shows the wages earned in the last 12 months by each of a random sample of American males aged between 16 and 65.

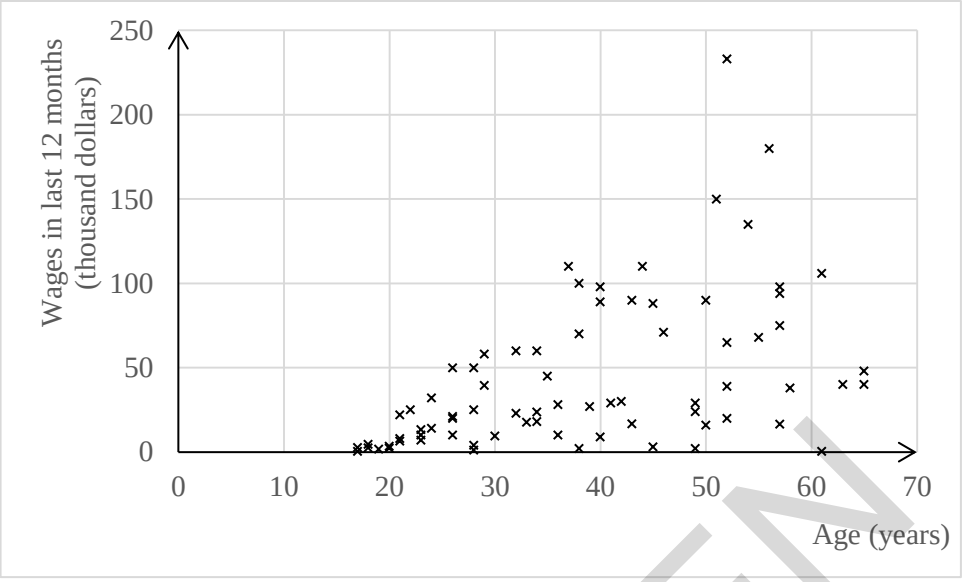


Fig.6

A researcher wishes to test whether the sample provides evidence of a tendency for higher wages to be earned by older men in the age range 16 to 65 in America.

(i) The researcher needs to decide whether to use a test based on Pearson's product moment correlation coefficient or Spearman's rank correlation coefficient. Use the information in Fig. 6 to decide which test is more appropriate. **[2]**

(ii) Should it be a one-tail or a two-tail test? Justify your answer. **[1]**

7 A newspaper reports that the average price of unleaded petrol in the UK is 110.2 p per litre.

The price, in pence, of a litre of unleaded petrol at a random sample of 15 petrol stations in Yorkshire is shown below together with some output from software used to analyse the data.

116.9	114.9	110.9	113.9	114.9
117.9	112.9	99.9	114.9	103.9
123.9	105.7	108.9	102.9	112.7

Statistics	
n	15
Mean	111.6733
σ	6.1877
s	6.4048
Σx	1675.1
Σx²	187638.31
Min	99.9
Q1	105.7
Median	112.9
Q3	114.9
Max	123.9

Fig. 7.1

n	15
Kolmogorov-Smirnov test	$p > 0.15$
Null hypothesis	The data can be modelled by a Normal distribution
Alternative hypothesis	The data cannot be modelled by a Normal distribution

Fig. 7.2

(i) Select a suitable hypothesis test to investigate whether there is any evidence that the average price of unleaded petrol in Yorkshire is different from 110.2 p. Justify your choice of test. **[3]**

(ii) Conduct the hypothesis test at the 5% level of significance. **[8]**

8 Natural background radiation consists of various particles, including neutrons. A detector is used to count the number of neutrons per second at a particular location.

(i) State the conditions required for a Poisson distribution to be a suitable model for the number of neutrons detected per second. **[2]**

The number of neutrons detected per second due to background radiation only is modelled by a Poisson distribution with mean 1.1.

(ii) Find the probability that the detector detects

(*A*) no neutrons in a randomly chosen second,

(*B*) at least 60 neutrons in a randomly chosen period of 1 minute. **[3]**

A neutron source is switched on. It emits neutrons which should all be contained in a protective casing. The detector is used to check whether any neutrons have not been contained; these are known as stray neutrons.

If the detector detects more than 8 neutrons in a period of 1 second, an alarm will be triggered in case this high reading is due to stray neutrons.

(iii) Suppose that there are no stray neutrons and so the neutrons detected are all due to the background radiation. Find the expected number of times the alarm is triggered in 1000 randomly chosen periods of 1 second. **[3]**

(iv) Suppose instead that stray neutrons are being produced at a rate of 3.4 per second in addition to the natural background radiation. Find the probability that at least one alarm will be triggered in 10 randomly chosen periods of 1 second. You should assume that all stray neutrons produced are detected. **[4]**

9 A random sample of adults in the UK were asked to state their primary source of news: television (**T**), internet (**I**), newspapers (**N**) or radio (**R**). The responses were classified by age group, and an analysis was carried out to see if there is any association between age group and primary source of news.

Fig. 9 is a screenshot showing part of the spreadsheet used to analyse the data. Some values in the spreadsheet have been deliberately omitted.

	A	B	C	D	E	F
1	**Source**	**Age group**				
2	**of news**	**18-32**	**33-47**	**48-64**	**65+**	
3	T	63	61	71	80	275
4	I	33	33	22	12	100
5	N	9	8	11	20	48
6	R	4	9	9	5	27
7		109	111	113	117	450
8						
9		**Expected frequencies**				
10		66.61	67.83	69.06	71.50	
11		24.22	24.67		26.00	
12		11.63	11.84	12.05	12.48	
13		6.54	6.66	6.78	7.02	
14						
15		**Contributions to the test statistic**				
16		0.20	0.69	0.05	1.01	
17		3.18	2.82		7.54	
18		0.59		0.09	4.53	
19		0.99	0.82	0.73	0.58	
20					**test statistic**	25.45

Fig. 9

(i) (*A*) State the sample size. **[1]**

(*B*) Give the name of the appropriate hypothesis test. **[1]**

(*C*) State the null and alternative hypotheses. **[1]**

(ii) Showing your calculations, find the missing values in cells

- D11,
- D17 and
- C18. **[4]**

(iii) Complete the appropriate hypothesis test at the 5% level of significance. **[4]**

(iv) Discuss briefly what the data suggest about primary source of news. You should make a comment for each age group. **[3]**

10 The label on a particular size of milk carton states that it contains 1.5 litres of milk. In an investigation at the packaging plant the contents, x litres, of each of 60 randomly selected cartons are measured. The data are summarised as follows.

$$\Sigma x = 89.758 \qquad \Sigma x^2 = 134.280$$

(i) Estimate the variance of the underlying population. **[2]**

(ii) Find a 95% confidence interval for the mean of the underlying population. **[4]**

(iii) What does the confidence interval which you have calculated suggest about the statement on the carton? **[1]**

Each day for 300 days a random sample of 60 cartons is selected and for each sample a 95% confidence interval is constructed.

(iv) Explain why the confidence intervals will not be identical. **[2]**

(v) What is the expected number of confidence intervals to contain the population mean? **[1]**

11 Two girls, Lili and Hui, play a game with a fair six-sided dice. The dice is thrown 10 times.

$X_1, X_2, \ldots, X_{10}$ represent the scores on the 1st, 2nd, … , 10th throws of the dice.
L denotes Lili's score and $L = 10X_1$.
H denotes Hui's score and $H = X_1 + X_2 + X_3 + ... + X_{10}$.

(i) Calculate

- $\mathrm{P}(L = 60)$ and
- $\mathrm{P}(H = 60)$. **[3]**

(ii) Without doing any further calculations, explain which girl's score has the greater standard deviation. **[1]**

(iii) Write down

- the name of the probability distribution of X_1,
- the value of $\mathrm{E}(X_1)$,
- the value of $\mathrm{Var}(X_1)$. **[3]**

(iv) Find

(*A*) $\mathrm{E}(L)$,

(*B*) $\mathrm{Var}(L)$,

(*C*) $\mathrm{E}(H)$,

(*D*) $\mathrm{Var}(H)$. **[5]**

The spreadsheet below shows a simulation of 25 plays of the game. The cell E3, highlighted, shows the score when the dice is thrown the fourth time in the first game.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Throw of dice											Lili's	Hui's
2		1	2	3	4	5	6	7	8	9	10		score	score
3	Game 1	3	5	2	1	1	3	1	1	1	4		30	22
4	Game 2	6	3	2	4	4	3	5	3	3	5		60	38
5	Game 3	6	4	2	6	5	2	1	5	2	3		60	36
6	Game 4	1	5	1	6	6	3	1	4	6	2		10	35
7	Game 5	4	4	3	1	6	4	4	1	6	2		40	35
8	Game 6	2	1	5	1	2	5	1	5	2	3		20	27
9	Game 7	1	1	3	4	4	5	6	3	4	2		10	33
10	Game 8	1	1	3	6	3	4	4	5	2	3		10	32
11	Game 9	2	2	2	4	3	2	1	5	5	6		20	32
12	Game 10	3	5	3	3	5	3	4	3	1	1		30	31
13	Game 11	5	3	6	5	5	4	2	1	1	5		50	37
14	Game 12	6	4	3	2	4	1	3	3	5	3		60	34
15	Game 13	2	3	2	1	2	2	2	2	2	1		20	19
16	Game 14	4	1	3	3	1	2	6	6	1	3		40	30
17	Game 15	5	1	2	6	3	4	6	3	6	4		50	40
18	Game 16	3	6	1	1	5	3	1	3	3	3		30	29
19	Game 17	5	2	5	2	4	5	2	2	3	4		50	34
20	Game 18	3	6	3	5	5	2	3	1	1	2		30	31
21	Game 19	6	6	3	1	5	6	3	4	1	6		60	41
22	Game 20	2	6	4	5	6	5	2	4	3	3		20	40
23	Game 21	5	3	5	4	5	3	3	6	6	1		50	41
24	Game 22	6	3	5	5	6	3	5	6	1	1		60	41
25	Game 23	5	4	5	5	6	4	2	1	3	6		50	41
26	Game 24	3	5	2	3	2	4	3	2	3	3		30	30
27	Game 25	5	2	4	2	4	5	2	2	5	2		50	33
28														
29											mean		37.60	33.68
30												sd	17.39	5.77

Fig. 11

(v) Use the simulation to estimate $\mathrm{P}(L > 40)$ and $\mathrm{P}(H > 40)$. **[2]**

(vi) (*A*) Calculate the exact value of $\mathrm{P}(L > 40)$. **[1]**

(*B*) Comment on how the exact value compares with your estimate of $\mathrm{P}(L > 40)$ in part **(v)**. **[1]**

Hui wonders whether it is appropriate to use the Central Limit Theorem to approximate the distribution of $X_1 + X_2 + X_3 + ... + X_{10}$.

(vii) (*A*) State what type of diagram Hui could draw, based on the output from the spreadsheet, to investigate this. **[1]**

(*B*) Explain how she should interpret the diagram. **[2]**

(viii) (*A*) Calculate an approximate value of $\mathrm{P}\left(X_1 + X_2 + X_3 + ... + X_{10} > 40\right)$ using the Central Limit Theorem. **[3]**

(*B*) Comment on how this value compares with your estimate of $\mathrm{P}(H > 40)$ in part **(v)**. **[1]**

END OF QUESTION PAPER

BLANK PAGE

BLANK PAGE

OCR
Oxford Cambridge and RSA

…day June 20XX – Morning/Afternoon

A Level Further Mathematics B (MEI)

Y422 Statistics Major

SAMPLE MARK SCHEME

Duration: 2 hours 15 minutes

MAXIMUM MARK 120

This document consists of 20 pages

Text Instructions

1. **Annotations and abbreviations**

Annotation in scoris	**Meaning**
✓and ✗	
BOD	Benefit of doubt
FT	Follow through
ISW	Ignore subsequent working
M0, M1	Method mark awarded 0, 1
A0, A1	Accuracy mark awarded 0, 1
B0, B1	Independent mark awarded 0, 1
SC	Special case
^	Omission sign
MR	Misread
Highlighting	
Other abbreviations in mark scheme	**Meaning**
E1	Mark for explaining a result or establishing a given result
dep*	Mark dependent on a previous mark, indicated by *
cao	Correct answer only
oe	Or equivalent
rot	Rounded or truncated
soi	Seen or implied
www	Without wrong working
AG	Answer given
awrt	Anything which rounds to
BC	By calculator
DR	This indicates that the instruction **In this question you must show detailed reasoning** appears in the question.

2. Subject-specific Marking Instructions for A Level Further Mathematics B (MEI)

a Annotations should be used whenever appropriate during your marking. The A, M and B annotations must be used on your standardisation scripts for responses that are not awarded either 0 or full marks. It is vital that you annotate standardisation scripts fully to show how the marks have been awarded. For subsequent marking you must make it clear how you have arrived at the mark you have awarded.

b An element of professional judgement is required in the marking of any written paper. Remember that the mark scheme is designed to assist in marking incorrect solutions. Correct solutions leading to correct answers are awarded full marks but work must not be judged on the answer alone, and answers that are given in the question, especially, must be validly obtained; key steps in the working must always be looked at and anything unfamiliar must be investigated thoroughly. Correct but unfamiliar or unexpected methods are often signalled by a correct result following an apparently incorrect method. Such work must be carefully assessed. When a candidate adopts a method which does not correspond to the mark scheme, escalate the question to your Team Leader who will decide on a course of action with the Principal Examiner.
If you are in any doubt whatsoever you should contact your Team Leader.

c The following types of marks are available.

M

A suitable method has been selected and *applied* in a manner which shows that the method is essentially understood. Method marks are not usually lost for numerical errors, algebraic slips or errors in units. However, it is not usually sufficient for a candidate just to indicate an intention of using some method or just to quote a formula; the formula or idea must be applied to the specific problem in hand, e.g. by substituting the relevant quantities into the formula. In some cases the nature of the errors allowed for the award of an M mark may be specified.

A

Accuracy mark, awarded for a correct answer or intermediate step correctly obtained. Accuracy marks cannot be given unless the associated Method mark is earned (or implied). Therefore M0 A1 cannot ever be awarded.

B

Mark for a correct result or statement independent of Method marks.

E

A given result is to be established or a result has to be explained. This usually requires more working or explanation than the establishment of an unknown result.

Unless otherwise indicated, marks once gained cannot subsequently be lost, e.g. wrong working following a correct form of answer is ignored. Sometimes this is reinforced in the mark scheme by the abbreviation isw. However, this would not apply to a case where a candidate passes through the correct answer as part of a wrong argument.

d When a part of a question has two or more 'method' steps, the M marks are in principle independent unless the scheme specifically says otherwise; and similarly where there are several B marks allocated. (The notation 'dep*' is used to indicate that a particular mark is dependent on an earlier, asterisked, mark in the scheme.) Of course, in practice it may happen that when a candidate has once gone wrong in a part of a question, the work from there on is worthless so that no more marks can sensibly be given. On the other hand, when two or more steps are successfully run together by the candidate, the earlier marks are implied and full credit must be given.

e The abbreviation FT implies that the A or B mark indicated is allowed for work correctly following on from previously incorrect results. Otherwise, A and B marks are given for correct work only – differences in notation are of course permitted. A (accuracy) marks are not given for answers obtained from incorrect working. When A or B marks are awarded for work at an intermediate stage of a solution, there may be various alternatives that are equally acceptable. In such cases, what is acceptable will be detailed in the mark scheme. If this is not the case please, escalate the question to your Team Leader who will decide on a course of action with the Principal Examiner.
Sometimes the answer to one part of a question is used in a later part of the same question. In this case, A marks will often be 'follow through'. In such cases you must ensure that you refer back to the answer of the previous part question even if this is not shown within the image zone. You may find it easier to mark follow through questions candidate-by-candidate rather than question-by-question.

f Unless units are specifically requested, there is no penalty for wrong or missing units as long as the answer is numerically correct and expressed either in SI or in the units of the question. (e.g. lengths will be assumed to be in metres unless in a particular question all the lengths are in km, when this would be assumed to be the unspecified unit.) We are usually quite flexible about the accuracy to which the final answer is expressed; over-specification is usually only penalised where the scheme explicitly says so. When a value is given in the paper only accept an answer correct to at least as many significant figures as the given value. This rule should be applied to each case. When a value is not given in the paper accept any answer that agrees with the correct value to 2 s.f. Follow through should be used so that only one mark is lost for each distinct accuracy error, except for errors due to premature approximation which should be penalised only once in the examination. There is no penalty for using a wrong value for *g*. E marks will be lost except when results agree to the accuracy required in the question.

g Rules for replaced work: if a candidate attempts a question more than once, and indicates which attempt he/she wishes to be marked, then examiners should do as the candidate requests; if there are two or more attempts at a question which have not been crossed out, examiners should mark what appears to be the last (complete) attempt and ignore the others. NB Follow these maths-specific instructions rather than those in the assessor handbook.

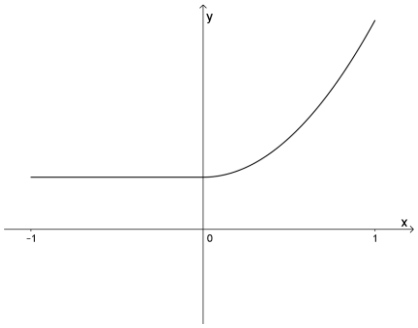
h For a genuine misreading (of numbers or symbols) which is such that the object and the difficulty of the question remain unaltered, mark according to the scheme but following through from the candidate's data. A penalty is then applied; 1 mark is generally appropriate, though this may differ for some papers. This is achieved by withholding one A mark in the question. Marks designated as cao may be awarded as long as there are no other errors. E marks are lost unless, by chance, the given results are established by equivalent working. 'Fresh starts' will not affect an earlier decision about a misread. Note that a miscopy of the candidate's own working is not a misread but an accuracy error.

i If a graphical calculator is used, some answers may be obtained with little or no working visible. Allow full marks for correct answers (provided, of course, that there is nothing in the wording of the question specifying that analytical methods are required). Where an answer is wrong but there is some evidence of method, allow appropriate method marks. Wrong answers with no supporting method score zero. If in doubt, consult your Team Leader.

j If in any case the scheme operates with considerable unfairness consult your Team Leader.

k Anything in the mark scheme which is in square brackets […] is not required for the mark to be earned on this occasion, but shows what a complete solution might look like

Question			Answer	Marks	AOs	Guidance	
1	(i)		$\frac{2}{9}$	**B1** **[1]**	**1.1**		
1	(ii)		r: 3, 4, 5, 6 $P(X = r)$: $\frac{2}{9}$, , , $\frac{31}{81}$	**B1** **[1]**	**1.1**		
1	(iii)		DR $E(X) = 3\times\frac{2}{9}+4\times\frac{2}{9}+5\times\frac{14}{81}+6\times\frac{31}{81}$ $=\frac{382}{81}=4.716$ $E(X^2) = 9\times\frac{2}{9}+16\times\frac{2}{9}+25\times\frac{14}{81}+36\times\frac{31}{81}$ $=\frac{1916}{81}=23.564$ $Var(X)=\frac{1916}{81}-\left(\frac{382}{81}\right)^2$ =1.413	**M1** **A1** **M1** **M1** **A1** **[5]**	**1.2** **1.1** **1.1** **1.1** **1.1**	For Σrp (at least 3 terms correct) FT their part **(ii)** cao For $\Sigma r^2 p$ (at least 3 terms correct) **M1**dep for – their $E(X)^2$ **A1** FT their $E(X)$ provided $Var(X) > 0$	

Question			Answer	Marks	AOs	Guidance	
2	(i)	(*A*)		**B1** **B1** **[2]**	**1.1** **1.1**	Shape of each part separately, domain correct All correct, including *y*-intercept, which may be labelled *a* or $\frac{1}{3}$	
2	(i)	(*B*)	Total area = 1 Area = $a+\int_0^1\left(a+x^2\right)\mathrm{d}x$ So $2a+\frac{1}{3}=1 \Rightarrow a=\frac{1}{3}$ AG	**B1** **M1** **A1** **[3]**	**1.2** **2.1** **2.2a**	Use of this principle somewhere in solution Attempt at two (or more) areas including a correct integral	
2	(ii)	(*A*)	$\frac{1}{3}+\int_0^{\frac{1}{2}}\left(\frac{1}{3}+x^2\right)\mathrm{d}x$ $=\frac{13}{24}=0.5417$	**M1** **A1** **[2]**	**1.1a** **1.1**	Attempt to find area from –1 to $\frac{1}{2}$ Must be seen BC As fraction, or given correct to 3 or 4 d.p.	
		(*B*)	$\mathrm{E}(X)=\int_{-1}^0 \frac{1}{3}x\mathrm{d}x+\int_0^1\left(\frac{1}{3}x+x^3\right)\mathrm{d}x$ $=\frac{1}{4}$	**M1** **A1** **[2]**	**1.1a** **1.1**	Must be seen BC	

Question			Answer	Marks	AOs	Guidance	
2	(iii)		area from –1 to 0 is $\frac{1}{3}$, so require				
			$\int_0^m \left(\frac{1}{3}+x^2\right)\mathrm{d}x = \frac{1}{6}$	E1	2.1		
			$\left[\frac{1}{3}x+\frac{1}{3}x^3\right]_0^m = \frac{1}{6}$	M1	1.1a		
			$\frac{1}{3}m+\frac{1}{3}m^3 = \frac{1}{6} \Rightarrow 2m^3+2m-1=0$ AG	A1	1.1		
				[3]			
3	(i)		At (24,11)	B1	1.1a		
			Residual	M1	1.1	Subtraction other way round scores **M1** only	
			$=11-(17.138-0.3727\times 24)=11-8.1932$				
			= 2.81	A1	1.1		
				[3]			
3	(ii)	(*A*)	x = 26 y = 7.45	B1	1.1		
			Interpolation and points lie fairly close to the line so probably a good estimate	E1	3.5a	Must mention both	
				[2]			
		(*B*)	x = 16 y = 11.17	B1	1.1		
			Extrapolation so probably not reliable	E1	3.5b		
				[2]			
3	(iii)		The only factor with a large effect size when correlated with hours of sleep is danger	E1	2.2b		
			It seems that the more dangerous the animal's situation, the less time it spends asleep	E1	2.2b	Or any other relevant comment, e.g. stating that the data do not demonstrate causality, or saying something relevant about the other factors	
				[2]			

Question			Answer	Marks	AOs	Guidance	
3	(iv)		There are outliers which affect the size of the pmcc … A linear model may well be suitable for the data with these outliers removed	E1 E1 [2]	3.5b 3.5c	Accept 'is suitable'. Or any other comment, e.g. redraw scatter diagram (or recalculate pmcc) without outliers	
4	(i)		$\left(\frac{5}{6}\right)^3 \times \frac{1}{6} = 0.0965$	B1 [1]	1.1		
4	(ii)		$1-\left(\frac{5}{6}\right)^4$ =0.518	M1 A1 [2]	1.1a 1.1		
4	(iii)		$\frac{5}{6}\times\left(\frac{1}{6}\right)^2 + \frac{1}{6}\times\frac{5}{6}\times\frac{1}{6}$ $= \frac{10}{216} = \frac{5}{108} = 0.0463$	M1 A1 [2]	3.1b 1.1		

Question			Answer	Marks	AOs	Guidance	
4	(iv)		$\frac{5}{108}+\left(\frac{1}{6}\right)^2$	M1	3.1b		
			$=\frac{8}{108}=\frac{2}{27}=0.0741$	A1	1.1		
			Alternative Method				
			$3\times\left(\frac{1}{6}\right)^2\times\frac{5}{6}+\left(\frac{1}{6}\right)^3$	M1			
			$=\frac{15}{216}+\frac{1}{216}=\frac{2}{27}=0.0741$	A1			
				[2]			
4	(v)		Expected value for one five = 6	E1	1.1	soi	
			Because geometric	E1	2.4		
			So for two fives expected value = 6 + 6 = 12	B1	2.1		
				[3]			
5	(i)		Mass of 5 small bags $\sim N(5\times508, 5\times3.3^2)$	B2	3.3	B1 For Normal and mean,	Distribution must be stated
			$\sim N(2540, 54.45)$		1.1	B1 For variance	
			$P(X < 2550) = 0.9123$	B1	3.4	BC FT their mean and variance	
				[3]			
5	(ii)		Mean of $L-3S=1515-3\times508=-9$	M1	3.3	Mean	
			Variance of $L-3S=4.7^2+9\times3.3^2$	M1	1.1	Method for variance	
			$L-3S \sim N(-9, 120.1)$	A1	1.1	Correct variance	
			$P(L-3S>0)=0.2058$	A1	3.4	BC	
				[4]			

Question			Answer	Marks	AOs	Guidance	
6	(i)		Shape of scatter diagram not approx. elliptical so no evidence of bivariate Normal required for test using pmcc to be valid	**B1** **B1** **[2]**	**3.5a** **2.4**	OR **B1** There does not appear to be a linear relationship OR **B1** so test using pmcc not appropriate [because that is for linear relationships]	
6	(ii)		One-tail because evidence of a positive relationship is sought	**B1** **[1]**	**2.4**	o.e.	
7	(i)		Kolmogorov-Smirnov *p*-value [greater than 0.15] indicates that the data could be from a Normal distribution Sample small with unknown population variance *t*-test	**E1** **E1** **B1** **[3]**	**3.5a** **2.4** **3.1b**		**SC2** For Wilcoxon test with • Cannot be sure the data are from a Normal distribution • Mean ≈ median indicates distribution is fairly symmetrical OR **SC1** For Wilcoxon with one of the above bullet points

Question			Answer	Marks	AOs	Guidance	
7	(ii)		H_0: $\mu = 110.2$ H_1: $\mu \neq 110.2$	**B1**	**1.1a**	Both hypotheses	
			Where μ is the population mean petrol price for Yorkshire	**B1**	**2.5**	Correct verbal definition	Allow hypotheses based on median if Wilcoxon chosen in **(i)**
			Test statistic is $\frac{111.6733 - 110.2}{6.4048/\sqrt{15}}$	**M1**	**3.3**		FT method for calculation of Wilcoxon statistic Price / -110.2 / rank 116.9 / 6.7 / 11 117.9 / 7.7 / 13 123.9 / 13.7 / 15 114.9 / 4.7 / 8 112.9 / 2.7 / 4 105.7 / -4.5 / 6 110.9 / 0.7 / 1 99.9 / -10.3 / 14 108.9 / -1.3 / 2 113.9 / 3.7 / 5 114.9 / 4.7 / 8 102.9 / -7.3 / 12 114.9 / 4.7 / 8 103.9 / -6.3 / 10 112.7 / 2.5 / 3
			= 0.8909	**A1**	**1.1**		FT test statistic = 44
			Use of t_{14}	**M1**	**3.4**		FT use of n = 15 row of Wilcoxon tables
			Critical value = 2.145	**A1**	**1.1**		FT Critical value = 25
			0.8089 < 2.145 so do not reject H_0	**M1**	**2.2b**		FT 44 > 25 so do not reject H_0
			There is insufficient evidence that the average price in Yorkshire is different from that in the UK	**A1** **[8]**	**3.5a**	Conclusion in context	

Question			Answer	Marks	AOs	Guidance	
8	(i)		The neutrons that are detected must occur randomly, independently and at a constant average rate.	**E1** **E1** **[2]**	**3.3** **3.3**	For randomly, independently For constant average rate or uniform rate but not constant rate	
8	(ii)	(*A*) (*B*)	P(0) = 0.333 $\lambda = 66$ P(at least 60) = 1 – 0.214 = 0.786	**B1** **M1** **A1** **[3]**	**1.1** **3.3** **3.4**	BC BC	
8	(iii)		P(more than 8 neutrons) = 1 – 0.999997573 = 0.000002427 Expected number = 1000 × 0.000002427 = 0.00243	 **B1** **M1** **A1** **[3]**	 **3.4** **1.1a** **1.1**	 BC	
8	(iv)		New $\lambda = 3.4 + 1.1 = 4.5$ P(No alarm triggered in 1 second) = 0.95974 P(At least one in 10 pds) = $1 - (0.95974)^{10}$ = 0.337	**B1** **B1** **M1** **A1** **[4]**	**3.1b** **3.4** **1.1a** **1.1**	 BC FT from here if $\lambda = 3.4$ used	

Question			Answer	Marks	AOs	Guidance	
9	(i)	(*A*)	Sample size = 450	**B1** **[1]**	**2.2a**		
9	(i)	(*B*)	Chi-squared test [for a contingency table]	**B1** **[1]**	**1.2**		
9	(i)	(*C*)	H_0: no association between age and news source H_1: some association between age and news source	**B1** **[1]**	**2.5**		
9	(ii)		D11= $\frac{113}{450}\times 100$ =25.11 C18 = $\frac{(8-11.84)^2}{11.84} = 1.25$ D17 = $\frac{(22-25.11)^2}{25.11} = 0.39$	**M1** **A1** **M1** **A1** **[4]**	**3.4** **1.1** **1.1** **1.1**	**M1** for $\frac{(O-E)^2}{E}$ applied at least once **A1** for both correct: accept 1.245, 0.385 (NB one can be calculated by subtraction)	
9	(iii)		Degrees of freedom = 9 Critical value = 16.92 Test statistic = 25.45 25.45 > 16.92 so reject H_0 There is sufficient evidence to suggest that there is some association between age and primary news source	**B1** **B1** **M1** **A1** **[4]**	**3.3** **1.1** **2.2b** **3.5a**		

Question			Answer	Marks	AOs	Guidance	
9	(iv)		For age group 18-32 and 33-47, the contributions of 3.18 and 2.82 show that more than expected have primary source the internet	E1	3.5a		
			For age group 65+, the contributions of 7.54 and 4.53 show that fewer than expected have primary source the internet and more than expected have primary source newspapers.	E1	3.5a		
			For age group 48 - 64 the contributions show that primary sources are as expected	E1 [3]	3.5a	Allow other suitable answers. Max 2 out of 3 if numerical values of contributions to test statistic not mentioned	

Question			Answer	Marks	AOs	Guidance	
10	(i)		Estimate of population variance = $\frac{134.280 - \frac{89.758^2}{60}}{59}$	M1	1.1		
			= 0.00008515	A1 [2]	1.1		
10	(ii)		1.49597	B1	1.1		
			$\pm 1.96 \times \sqrt{\frac{0.00008515}{60}}$	M1 M1	3.3 1.1		
			= 1.49597 ± 0.00233 or (1.4936,1.4983)	A1 [4]	3.4	Allow (1.494, 1.498)	
10	(iii)		It appears that the (population) mean content is not 1.5 litres as the calculated interval does not contain 1.5.	E1 [1]	3.5a		
10	(iv)		Each time a sample is taken it will be different, so e.g. will have a different mean hence different midpoint for confidence interval.	E1 E1 [2]	2.4 2.4	Samples vary …so confidence intervals vary	
10	(v)		$300 \times 0.95 = 285$	B1 [1]	1.1		

Question			Answer	Marks	AOs	Guidance	
11	(i)		$P(L = 60) = \frac{1}{6}$	B1	1.1		
			$P(H = 60) = \left(\frac{1}{6}\right)^{10}$	M1	1.1		
			$= 1.65 \times 10^{-8}$	A1	1.1		
				[3]			
11	(ii)		Lili's score because Hui's score uses more results so is likely to be closer to the mean	E1	2.2b	Any reasonable explanation e.g. Lili's score has greater standard deviation, as the extreme scores have higher probability	
				[1]			
11	(iii)		Discrete uniform	B1	3.3		
			$E(X_1) = 3.5$	B1	1.1		
			$Var(X_1) = \frac{35}{12} [= 2.917]$	B1	3.4		
				[3]			
11	(iv)	(*A*)	$E(L) = 35$			See (*C*)	
		(*B*)	$Var(L) = 10^2 \times \frac{35}{12}$	M1	1.2		
			$= \frac{875}{3} = 291.7$	A1	1.1		
		(*C*)	$E(H) = 35$	B1	1.1	Both expected values correct	
		(*D*)	$Var(H) = 10 \times \frac{35}{12}$	M1	2.4		
			$\frac{175}{6} = 29.17$	A1	1.1		
				[5]			

Question			Answer	Marks	AOs	Guidance	
11	(v)		Estimate of $P(L > 40) = \frac{11}{25}$ Estimate of $P(H > 40) = \frac{4}{25}$	**B1** **B1** **[2]**	**2.2b** **1.1**		
11	(vi)	(*A*)	$P(L > 40) = \frac{1}{3}$	**B1** **[1]**	**1.1**		
11	(vi)	(*B*)	Estimate 0.44, calculated value 0.33. Some way off but not totally unreasonable approximation with only 25 trials.	**E1** **[1]**	**3.2b**	Any sensible relevant comment Ft their **(v)**	
11	(vii)	(*A*)	Produce a normal probability plot of the 25 values of Hui's scores (or of the scores $\div 10$)	**B1** **[1]**	**1.2**		
11	(vii)	(*B*)	… if approximately a straight line then would appear to be from Normal distribution… … so Central Limit Theorem would seem to apply.	**E1** **E1** **[2]**	**2.4** **2.2b**		

Question			Answer	Marks	AOs	Guidance	
11	(viii)	(*A*)	Mean $\sim N\left(35, \frac{350}{12}\right)$	M1	1.2		
			P(Mean > 40) = P(Normal >40.5)	B1	3.4	Continuity correction – with value of 40.5 as border (may have 40.5 included)	
			So $P(H > 40) \approx 0.154$	A1	1.1	BC	
				[3]			
11	(viii)	(*B*)	Agrees well with $\frac{4}{25} = 0.16$	E1	3.2b	FT their **(v)**	
				[1]			

Question	AO1	AO2	AO3(PS)	AO3(M)	Totals
1i	1	0	0	0	1
1ii	1	0	0	0	1
1iii	5	0	0	0	5
2iA	2	0	0	0	2
2iB	1	2	0	0	3
2iiA	2	0	0	0	2
2iiB	2	0	0	0	2
2iii	2	1	0	0	3
3i	3	0	0	0	3
3iiA	1	0	0	1	2
3iiB	1	0	0	1	2
3iii	0	2	0	0	2
3iv	0	0	0	2	2
4i	1	0	0	0	1
4ii	2	0	0	0	2
4iii	1	0	1	0	2
4iv	1	0	1	0	2
4v	1	2	0	0	3
5i	1	0	0	2	3
5ii	2	0	0	2	4
6i	0	1	0	1	2
6ii	0	1	0	0	1
7i	0	1	1	1	3
7ii	3	2	0	3	8
8i	0	0	0	2	2
8iiA	1	0	0	0	1
8iiB	0	0	0	2	2
8iii	2	0	0	1	3
8iv	2	0	1	1	4
9iA	0	1	0	0	1
9iB	1	0	0	0	1
9iC	0	1	0	0	1
9ii	3	0	0	1	4
9iii	1	1	0	2	4
9iv	0	0	0	3	3
10i	2	0	0	0	2
10ii	2	0	0	2	4
10iii	0	0	0	1	1
10iv	0	2	0	0	2
10v	1	0	0	0	1

Question	AO1	AO2	AO3(PS)	AO3(M)	Totals
11i	3	0	0	0	3
11ii	0	1	0	0	1
11iii	1	0	0	2	3
11iv	4	1	0	0	5
11v	1	1	0	0	2
11viA	1	0	0	0	1
11viB	0	0	1	0	1
11viiA	1	0	0	0	1
11viiB	0	2	0	0	2
11viiiA	2	0	1	0	3
11viiiB	0	0	0	1	1
Totals	61	22	6	31	120

OCR
Oxford Cambridge and RSA

Accredited

A Level Further Mathematics B (MEI)

Y422 Statistics Major

Printed Answer Booklet

Date – Morning/Afternoon

Time allowed: 2 hours 15 minutes

OCR supplied materials:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)

You must have:
- Printed Answer Booklet
- Formulae Further Mathematics B (MEI)
- Scientific or graphical calculator

First name	
Last name	

Centre number						Candidate number				

INSTRUCTIONS
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Complete the boxes provided on the Printed Answer Booklet with your name, centre number and candidate number.
- Answer **all** the questions.
- **Write your answer to each question in the space provided in the Printed Answer Booklet.**
- Additional paper may be used if necessary but you must clearly show your candidate number, centre number and question number(s).
- Do **not** write in the bar codes.
- You are permitted to use a scientific or graphical calculator in this paper.
- Final answers should be given to a degree of accuracy appropriate to the context.

INFORMATION
- You are advised that an answer may receive **no marks** unless you show sufficient detail of the working to indicate that a correct method is used. You should communicate your method with correct reasoning.
- The Printed Answer Booklet consists of **20** pages. The Question Paper consists of **16** pages.

Section A (30 marks)

1 (i)

1 (ii)

r	3	4	5	6
$\mathrm{P}(X = r)$		$\frac{2}{9}$	$\frac{14}{81}$	

1 (iii)

2 (i) (*A*)

2 (i) (*B*)

2 (ii) (*A*)

2 (ii) (*B*)

2 (iii)

3 (i)	
3(ii) (*A*)	Estimate
	Comment
3(ii) (*B*)	Estimate
	Comment
3 (iii)	
3 (iv)	

Section B (90 marks)

4 (i)

4 (ii)

4 (iii)

4 (iv)

4 (v)	
5 (i)	
5 (ii)	

6(i)	
6(ii)	
7 (i)	

7 (ii)

8 (i)	
8(ii) (*A*)	
8(ii) (*B*)	
8 (iii)	

8 (iv)

9 (i) (*A*)

9 (i) (*B*)

9 (i) (*C*)

9 (ii)

9 (iii)

(answer space continued on next page)

9 (iii)	**(continued)**
9 (iv)	
10 (i)	

10 (ii)	
10 (iii)	
10 (iv)	
10 (v)	

11 (i)	
11 (ii)	
11 (iii)	

11 (iv) (*A*)

11 (iv) (*B*)

11 (iv) (*C*)

11 (iv) (*D*)

11 (v)	
11 (vi) (*A*)	
11 (vi) (*B*)	
11 (vii) (*A*)	
11 (vii) (*B*)	

11 (viii) (*A*)

11 (viii) (*B*)

DO NOT WRITE ON THIS PAGE

PLEASE DO NOT WRITE ON THIS PAGE