OCR
Oxford Cambridge and RSA

A 56
B 50
C 45
D 40
E 35

Wednesday 14 June 2017 – Morning

A2 GCE MATHEMATICS (MEI)

4767/01 Statistics 2

QUESTION PAPER

Candidates answer on the Printed Answer Book.

OCR supplied materials:
- Printed Answer Book 4767/01
- MEI Examination Formulae and Tables (MF2)

Other materials required:
- Scientific or graphical calculator

Duration: 1 hour 30 minutes

6710721633

INSTRUCTIONS TO CANDIDATES

These instructions are the same on the Printed Answer Book and the Question Paper.

- The Question Paper will be found inside the Printed Answer Book.
- Write your name, centre number and candidate number in the spaces provided on the Printed Answer Book. Please write clearly and in capital letters.
- **Write your answer to each question in the space provided in the Printed Answer Book**. If additional space is required, you should use the lined page(s) at the end of this booklet. The question number(s) must be clearly shown.
- Use black ink. HB pencil may be used for graphs and diagrams only.
- Read each question carefully. Make sure you know what you have to do before starting your answer.
- Answer **all** the questions.
- Do **not** write in the barcodes.
- 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 FOR CANDIDATES

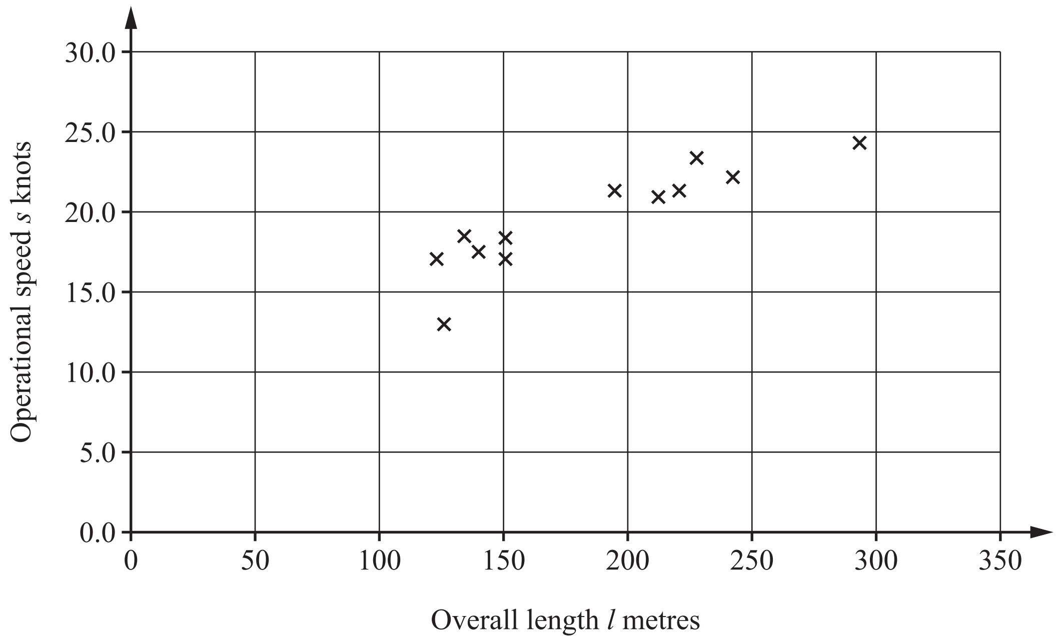
This information is the same on the Printed Answer Book and the Question Paper.

- The number of marks is given in brackets **[ ]** at the end of each question or part question on the Question Paper.
- 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 being used.
- The total number of marks for this paper is **72**.
- The Printed Answer Book consists of **12** pages. The Question Paper consists of **8** pages. Any blank pages are indicated.

INSTRUCTION TO EXAMS OFFICER/INVIGILATOR

- Do not send this Question Paper for marking; it should be retained in the centre or recycled. Please contact OCR Copyright should you wish to re-use this document.

1 The scatter diagram below illustrates the overall lengths l metres and the typical operational speeds s knots (nautical miles per hour) of 12 container ships. The length of a ship is one of the factors which determines its typical operational speed.



Summary statistics for these data are as follows:

$n = 12$, $\sum l = 2219$, $\sum s = 234.6$, $\sum l^2 = 443\,867$, $\sum s^2 = 4700.56$, $\sum ls = 45\,149.0$.

(i) State which of the two variables l and s is the independent variable and which is the dependent variable. Briefly explain your answers. **[2]**

(ii) Calculate the equation of the regression line of s on l. **[5]**

(iii) Interpret the coefficient of l in terms of the relationship between overall length and speed in the equation of the regression line found in part **(ii)**. **[2]**

(iv) Calculate the value of the residual for the data point where $l = 126$ and $s = 13.0$. **[3]**

(v) Use the equation of the regression line in part **(ii)** to calculate an estimate of the operational speed of a ship with overall length of 100 metres. Comment on the reliability of this estimate. **[2]**

If the data point in part **(iv)** is removed from the data set, the equation of the new regression line is

$$s = 0.0453l + 11.5.$$

(vi) Recalculate the estimate for an overall length of 100 metres using this new equation. Discuss which of these two estimates you think is likely to be more reliable. **[3]**

2 At a small hospital maternity department, there is an average of 1.3 births per day. Throughout this question, you should assume that births occur independently, at random times, and at a uniform average rate.

(i) Briefly explain the meaning of each of the three terms 'independently', 'random' and 'uniform', in the context of births at the maternity department. **[3]**

X represents the number of births at the hospital on a randomly chosen day.

(ii) State the distribution of X and also the variance of X. **[2]**

(iii) Find $\mathrm{P}(X > 3)$. **[2]**

(iv) Find the probability that there are exactly 3 births in a period of 3 days. **[2]**

There is an average of 0.4 home births per **week** in the area served by the hospital.

(v) Find the probability that the total number of births in a week (at home and in hospital) is at least 10. **[2]**

(vi) Use a suitable approximating distribution to find the probability that there is a total of at least 50 births in a period of 4 weeks. **[5]**

(vii) How realistic do you think the assumption of independence made at the start of this question is? **[2]**

3 The random variable X represents the weight, in grams, of a particular type of chocolate bar. It is known that X is Normally distributed with mean 50.7 and variance 0.72. On the wrapper it states that the bar weighs 50 grams.

(i) Find the proportion of these chocolate bars that actually weigh at least 50 grams. **[3]**

(ii) A quality control manager wishes to increase this proportion to 95%.

(A) Find the required value of the mean if the variance remains unchanged. **[4]**

(B) Find the required value of the variance if the mean remains unchanged. **[2]**

The weights of another type of chocolate bar are also Normally distributed. On the wrapper it states that the bar weighs 25 grams. It is known that 99% of these bars weigh at least 25.0 grams and 75% of them weigh at least 25.4 grams.

(iii) Find the probability that one of these bars weighs at least 26.0 grams. **[6]**

(iv) One bar of the first type (with the original mean and standard deviation) and 2 bars of the second type are selected at random. Find the probability that at least one of the bars has a weight less than that stated on its wrapper. **[2]**

4 **(a)** In an investigation into dietary supplements, a random sample of 200 adults was selected. Each of them was asked whether or not they regularly take dietary supplements. The 200 adults were categorised as 'Male under 50', 'Male 50 or older', 'Female under 50', 'Female 50 or older'. Their answers to the question are summarised in the table below.

	Yes	No
Male under 50	13	33
Male 50 or older	18	31
Female under 50	24	37
Female 50 or older	24	20

(i) Write down null and alternative hypotheses for a test to examine whether there is any association between category of adult and whether or not they regularly take dietary supplements. **[1]**

The expected frequencies under the null hypothesis for the usual χ^2 test are shown in the table below.

Expected frequency	Yes	No
Male under 50	18.17	27.83
Male 50 or older	19.36	29.65
Female under 50	24.10	36.91
Female 50 or older	17.38	26.62

(ii) Verify the expected frequency of 18.17 for Male under 50 answering Yes. **[2]**

The contributions to the test statistic are shown in the table below.

Contribution	Yes	No
Male under 50	1.4710	0.9604
Male 50 or older	0.0949	0.0619
Female under 50	0.0004	0.0002
Female 50 or older	2.5215	1.6463

(iii) Verify the contribution to the test statistic of 1.4710 for Male under 50 answering Yes. **[2]**

(iv) Given that the total of the contributions is 6.757, correct to 3 decimal places, complete the test at the 10% significance level. **[4]**

(v) For each category of adult, comment briefly on how their taking of dietary supplements compares with what would be expected if there were no association. **[3]**

(b) The breaking strength of a particular type of rope is known to be Normally distributed with mean 562 kg and a standard deviation of 27.4 kg when the rope is dry. A researcher believes that the breaking strength may be different when the rope is wet. He selects a random sample of 12 pieces of wet rope and finds that the mean of their breaking strengths is 547 kg. Carry out a test at the 10% significance level to investigate the researcher's belief. You may assume that the breaking strengths of wet ropes of this type are still Normally distributed with standard deviation 27.4 kg. **[8]**

END OF QUESTION PAPER

Question			Answer	Marks	Guidance
1	(i)		l independent & s dependent **l is independent** since the values of l are not subject to random variation, **or** values of l are controlled/pre-determined/set/chosen (by the manufacturer/researcher) **or s is dependent** since the values of s are subject to random variation.	B1 B1 **[2]**	 relevant comment regarding l **or** s do not accept "….changed"
1	(ii)		$\bar{s}$ = 234.6/12 (=19.55), $\bar{l}$ = 2219/12 (=184.917) $b = \frac{S_{ls}}{S_{ll}} = \frac{45149.0-(2219\times234.6/12)}{443867-2219^2/12} = \frac{1767.55}{33536.9} = 0.05270$ OR $b = \frac{45149.0/12-(19.55\times184.917)}{443867/12-184.917^2} = \frac{147.296}{2794.74} = 0.05270$ hence least squares regression line is: $s - \bar{s} = b(l - \bar{l})$ $\Rightarrow s - 19.55 = 0.05270\ (l - 184.917)$ $\Rightarrow s = 0.0527\ l + 9.80$ (accept $s = 0.05270\ l + 9.804$, $s = 0.05270\ l + 9.805$)	B1 M1 A1 M1 A1 **[5]**	for $\bar{s}$ and $\bar{l}$ seen (or can be implied by correct value of b) for attempt at gradient (b) with correct structure. See additional notes on 'structure'. for 0.0527. Allow 0.053 for equation of line with their b>0, $\bar{l}$ & $\bar{s}$ FT for complete equation in terms of s and l. Accept equation in terms of x and y only if defined as length and speed respectively. Allow $s = 0.053\ l + 9.8$ www See additional note RE over-specification
	(iii)		The coefficient of l is the additional speed resulting from an increase of 1 metre in length	B1 B1 **[2]**	for connecting **increase** in l increase in s for relating to **unit** increase in length.
	(iv)		$l = 126 \Rightarrow$ predicted speed = 0.0527 × 126 + 9.80 (= 16.4) Residual = 13.0 – 16.4 = – 3.4 (or – 3.44 or – 3.45)	 M1 M1 A1 **[3]**	 for prediction FT their equation for a subtraction involving 13.0 and their prediction, either way round SOI. FT only 13.0 – their prediction. See additional note RE over-specification

Question			Answer	Marks	Guidance
	(v)		0.0527 × 100 + 9.80 = 15.1 Might not be reliable as extrapolation	B1 B1 **[2]**	FT their equation. See additional note RE over-specification not reliable and extrapolation oe seen
	(vi)		0.0453 × 100 + 11.5 = 16.0 3sf The point where l = 126 and s = 13.0 may be an error and as such it might be better to use the second line which does not involve it. On the other hand the first model may be better as it uses all the available data.	B1 B1 B1 **[3]**	Allow 16 or 16.03 Allow "outlier" or equivalent for "error" or this point might suggest that a curve might be a better model in which case the first model would be better.
2	(i)		'Independently' means that the occurrence of one birth does not affect the **probability** of another birth occurring. 'Random' means that births occur with no particular **pattern**. 'uniform' means that the **average rate** of births is **constant** or the **average over any given time period** is **constant.**	B1 B1 B1 **[3]**	must be in context and include 'probability' or 'chance' but do not allow "the probability of a birth does not affect the probability of another" must be in context. Allow 'not predictable' do not allow "no particular order" must be in context
2	(ii)		X ~ Poisson (1.3) **Variance** = 1.3	B1 B1 **[2]**	allow X ~ Po(1.3) and X ~ P(1.3) and Poisson with λ = 1.3 must include 1.3 but do **not** allow Po(1.3,1.3) for variance = 1.3, allow σ^2 = 1.3, do not allow λ = 1.3
2	(iii)		From tables P(X > 3) = 1 – P($X \le 3$) = 1 – 0.9569 = 0.0431	 M1 A1 **[2]**	 Attempting 1 – P($X \le 3$) e.g. for 1 – 0.9463 (using λ = 1.4) or 1 – 0.6248 (using λ = 3.1) See additional note RE over-specification

Question			Answer	Marks	Guidance
2	(iv)		$\lambda = 3 \times 1.3 = 3.9$ P(3 births) $= \dfrac{e^{-3.9}3.9^3}{3!} = 0.2001$ Or from tables P(3 births) $= 0.4532 - 0.2531 = 0.2001$	B1 B1 **[2]**	for mean For 0.2001 Allow 0.200, 0.20, 0.2 www See additional note RE over-specification
2	(v)		$\lambda = 7 \times 1.3 + 0.4 = 9.5$ From tables $P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.5218 = 0.4782$	B1 B1 **[2]**	for mean for 0.4782 or 0.478www or 0.48www See additional note RE over-specification
2	(vi)		Normal approx. to the Poisson, $X \sim N(38, 38)$ $P(X \geq 50) = P\left(Z \geq \dfrac{49.5-38}{\sqrt{38}}\right)$ $= 1 - 0.9690$ $= 0.0310$	B1 B1 B1 M1 A1 **[5]**	for Normal approximation (SOI) for correct parameters (SOI) continuity correction i.e. 49.5 for correct structure of Normal probability calculation cao **(**Do not FT wrong or omitted CC) (answer from calculator = 0.031052 so accept 0.0310 or 0.0311) Allow 0.031www See additional note RE over-specification
2	(vii)		This assumption is not fully valid as there will be some multiple births but the proportion of multiple births is fairly small so it is not totally unreasonable	B1 B1 **[2]**	e.g. twins, triplets, …
3	(i)		$P(X \geq 50) = P\left(Z > \dfrac{50-50.7}{\sqrt{0.72}}\right) = P(Z > -0.825)$ $= \Phi(0.825)$ $= 0.7953$	M1 M1 A1 **[3]**	For standardizing. M0 for using "continuity corrections" e.g. 49.5, 49, 51, and/or σ = 0.72 used. Condone numerator reversed. For correct tail Cao allow 0.795 www See additional note RE over-specification

Question			Answer	Marks	Guidance
3	**(ii)**	**(*A*)**	P(weight > 50) = 0.95 so P($Z > \frac{50-\mu}{\sqrt{0.72}}$) = 0.95 $\Phi^{-1}(0.95) = -1.645$ $\frac{50-\mu}{\sqrt{0.72}} = -1.645$ $\mu = 50 + 1.645 \times \sqrt{0.72} = 51.395...$ = 51.4	B1 M1* M1dep* A1 **[4]**	For ±1.645 For equation for μ as seen or equivalent with their negative z-value. See additional note. Allow M1* if "continuity correction" and/or σ = 0.72 used and penalised in part (i). NOTE σ = 0.8485 (allow 0.85 or better) for rearranging to find μ cao allow 51.40 See additional note RE over-specification
3	**(ii)**	**(*B*)**	P(weight > 50) = 0.95 so P($Z > \frac{50-50.7}{\sigma}$) = 0.95 $\frac{50-50.7}{\sigma} = -1.645$ $\sigma = \frac{50-50.7}{-1.645} = 0.4255...$ Var = 0.4255^2 = 0.181	M1 A1 **[2]**	for equation as seen or equivalent allow M1 if "continuity correction" has been used and already penalised in part (i) or part(ii) for 0.181 or 0.1811 or 0.18www NOTE M0 A0 for 0.181 from $(-0.4255)^2$ See additional note RE over-specification

3	(iii)		$\mathrm{P}(Y > 25) = 0.99 \Rightarrow \mathrm{P}\left(Z > \frac{25-\mu}{\sigma}\right) = 0.99$ $\Rightarrow \frac{25-\mu}{\sigma} = \Phi^{-1}(0.99) = -2.326 \qquad \Rightarrow 25 = \mu - 2.326\sigma$ $\mathrm{P}(Y > 25.4) = 0.75 \Rightarrow \mathrm{P}\left(Z > \frac{25.4-\mu}{\sigma}\right) = 0.75$ $\Rightarrow \frac{25.4-\mu}{\sigma} = \Phi^{-1}(0.75) = -0.6745 \qquad \Rightarrow 25.4 = \mu - 0.6745\sigma$ $1.6515\sigma = 0.4 \qquad \sigma = 0.2422\ldots$ $\mu = 25 + 2.326 \times 0.2422\ldots \qquad \mu = 25.563\ldots$ $\mathrm{P}(\text{Weighs} > 26.0) = P\left(Z > \frac{26.0-25.563}{0.2422}\right) = 1 - \Phi(1.804) = 1 - 0.9644 = 0.0356$	B1 M1 A1 A1 A1 A1 **[6]**	for ±2.326 or ±0.6745 seen For obtaining two equations in terms of mean, standard deviation and their *z*- values (but not *z* = 0.99 or *z* = 0.75 or e.g.1 – 2.326) in any form equivalent to these. for at least one equation correct cao for answers in the range 0.0345 to 0.036 See additional note RE over-specification
3	(iv)		$1 - (0.7953 \times 0.99^2) = 1 - 0.7795$ $= 0.2205$	M1 A1 **[2]**	or equivalent FT their 3(i) allow 0.221www or 0.22www
4	(a)	(i)	H_0: no association between category of adult and taking dietary supplements. H_1: some association between category of adult and taking dietary supplements.	B1 **[1]**	Hypotheses must refer to ‘association’ and be in context. Allow hypotheses appropriately worded in terms of independence.
		(ii)	Expected frequency = (46 × 79)/200 = 18.17 AG	M1 A1 **[2]**	attempt at row total × column total/grand total 46, 79 and 200 used correctly **and 18.17 seen** **NB Answer given**

		(iii)	Contribution = $(13 - 18.17)^2 / 18.17$ = 1.4710 AG	M1 A1 **[2]**	for valid attempt at $(O\text{-}E)^2/E$ 13 and 18.17 used correctly **and 1.4710** or better **seen** **NB Answer given**
		(iv)	Refer to χ_3^2 Critical value at 10% level = 6.251 (6.757 > 6.251 so result is) significant There is sufficient evidence to **suggest/support** association between category of adult and taking dietary supplements NB if H_0 H_1 reversed do not award first B1 or final A1	B1 B1 M1 A1 **[4]**	for 3 degrees of freedom seen (e.g. in subscript) for 6.251 **- No further marks from here if wrong or omitted** for 'significant' or 'Accept H_1' or 'Reject H_0' seen For **non-assertive** conclusion in context. Do not allow 'relationship' or 'correlation' in place of 'association'
		(v)	**Large** contribution for **males under 50** suggests that there are fewer than expected saying **yes**. **Large** contribution for **females 50 or older** suggests that there are more than expected saying **yes**. **Small** contributions for the **other two groups** show that numbers are much as expected.	B1 B1 B1 **[3]**	or **large** contribution for **males under 50** suggests that there are more than expected saying **no**. NB if both comments are provided they must both be correct for B1 or **large** contribution for **females 50 or older** suggests that there are fewer than expected saying **no**. NB if both comments are provided they must both be correct for B1 Do not accept e.g. a few less/more Special Case – if sizes of contributions are not mentioned but comments are otherwise correct award SC1 Comments about what should have been observed (e.g. there should have been more males under 50 saying yes) get 0/3..

4	(b)		H_0: $\mu = 562$ H_1: $\mu \neq 562$	B1	For both correct **NB** for **H_1: μ < 562** award maxB0B1M1*A1B1(for – 1.282)depM0*A0A0
			Where μ denotes the **mean** breaking strength of **wet rope** of this type (in the population)	B1	For definition of μ in context. Do not allow any other symbols unless clearly defined as population mean.
			Test statistic $= \frac{547-562}{27.4/\sqrt{12}}$	M1*	must include √12 with numerator as seen
			$= \frac{-15}{7.910} = -1.896$	A1	cao for – 1.896
			Lower 5% level 2 tailed critical value of z = –1.645	B1	For 1.645 or – 1.645. No further marks from here if B0 awarded
			–1.896 < –1.645 so the result is significant.	M1dep*	For sensible comparison leading to a conclusion. Must be –1.645 unless it is clear that absolute values are being used.
			There is sufficient evidence to reject H0	A1	for correct conclusion
			There is sufficient evidence to suggest that the **mean** breaking strength of **wet** rope of this type is different (not equal to 562)	A1	for **non-assertive** conclusion in words in context. A0 for "…mean of wet rope has **changed**" FT candidate's test statistic only if both M marks earned
				[8]	See additional notes regarding alternative methods and sensible comparisons.

Additional notes Re Q4(b)

Critical Value Method c.v. = 562 – 1.645×27.4/√12 gets M1* B1
= 548.99 or 549.0 or 549 gets A1 (replacing A1 for -1.896)
547 < 548.99 with a conclusion gets M1dep* then final A1 A1 still available
NB if H_1: μ < 562 award maxB0B1M1*A1(for 551.9)B1(for - 1.282 used correctly) depM0*A0A0

Probability Method P(Z < –1.896) = 0.0289 or 0.029 gets B1 (replacing B1 for ±1.645)
0.0289 < 0.05 with conclusion gets M1dep* then final A1 A1 still available
NB if H_1: μ < 562 used award maxB0B1M1*A1B1(for 0.029)depM0*A0A0

Additional Note RE Over-specification

A0 or B0 for final answers given correct to 5sf or more potentially in Q1ii (final A1), Q1iv (final A1), Q1v (first B1), Q2iii(final A1), Q2iv (final B1), Q2v (final B1), Q2vi (final A1), Q3i (A1), Q3iiA (A1), Q3iiB (A1), Q3iii final A1), Q3iv (A1).
NOTE do not penalise over-specification more than twice in any single question or more than 4 times in a paper.

Additional Notes on Correct Structure in Q1(ii)

Equivalent calculations for finding b are allowed. For example use of $12S_{ls}/12S_{ll}$ is allowed. However, where these are mixed we award M0. e.g. use of $12S_{ls}/S_{ll}$ would earn M0. For M1 to be awarded, the structure of the calculation must be numerically equivalent to the one provided – NOTE if it is believed that the candidate has made an error in transcription of a number (for example using 2119 instead of 2219) we can allow M1 BOD if the structure is otherwise correct.

Additional Notes for Q3ii*A*

M1* is for forming a suitable equation using their z-value but it must be reasonably clear that the value used is a z-value – for example we do not allow 0.05 or 0.95 to be treated as z-values here. The M1dep* can be awarded if the candidate correctly rearranges their equation to find μ. Hence, use of an incorrect z-value could earn max B0M1*M1dep*A0.
If z = +1.645 is used then award B1 only to give 1/4 unless the numerator of the equation is reversed in which case the remaining marks are available.

Additional Notes on Sensible Comparisons

In Q4 (b) Neither – 1.896 < 0.05 nor 0.0289 < 1.645 are considered sensible as each compares a z-value with a probability.
Inequality sign reversed, e.g. – 1.896 > – 1.645, gets M0A0A0.
Comparing a negative with a positive z-value, e.g. – 1.896 < 1.645, gets M0A0A0.

<u>Additional Notes on Conclusions to Hypothesis Tests</u>

The following are examples of conclusions which are considered too assertive.
There is sufficient evidence to reject H_0 and **conclude** that…
"<u>there is</u> a positive association between…" or
"there seems to be <u>evidence that there is</u> a positive association between…" or
"the mean nicotine content <u>is</u> greater …."
"there doesn't appear to be association between…"
Also note that final conclusions **must refer to H_1 in context** for the final mark to be given.
e.g. In Q4a iv a conclusion simply stating that "the evidence suggests that there is association" gets A0 as this does not refer to the context.