![OCR — Oxford Cambridge and RSA]

GCE

Mathematics (MEI)

Unit **4767**: Statistics 2

Advanced GCE

Mark Scheme for June 2017

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
U1	Mark for correct units
G1	Mark for a correct feature on a graph
M1 dep*	Method 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
ito	In terms of

Subject-specific Marking Instructions for GCE Mathematics (MEI) Statistics strand

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, award marks according to the spirit of the basic scheme; if you are in any doubt whatsoever (especially if several marks or candidates are involved) 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, eg 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, eg 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, exactly what is acceptable will be detailed in the mark scheme rationale. If this is not the case please consult your Team Leader.

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   Wrong or missing units in an answer should not lead to the loss of a mark unless the scheme specifically indicates otherwise.

Candidates are expected to give numerical answers to an appropriate degree of accuracy. 3 significant figures may often be the norm for this, but this always needs to be considered in the context of the problem in hand. For example, in quoting probabilities from Normal tables, we generally expect *some* evidence of interpolation and so quotation to 4 decimal places will often be appropriate. But even this does not always apply – quotations of the standard critical points for significance tests such as 1.96, 1.645, 2.576 (maybe even 2.58 – but not 2.57) will commonly suffice, especially if the calculated value of a test statistic is nowhere near any of these values. Sensible discretion *must* be exercised in such cases.

Discretion must also be exercised in the case of small variations in the degree of accuracy to which an answer is given. For example, if 3 significant figures are expected (either because of an explicit instruction or because the general context of a problem demands it) but only 2 are given, loss of an accuracy ("A") mark is likely to be appropriate; but if 4 significant figures

are given, this should not normally be penalised. Likewise, answers which are slightly deviant from what is expected in a very minor manner (for example a Normal probability given, after an attempt at interpolation, as 0.6418 whereas 0.6417 was expected) should not be penalised. However, answers which are *grossly* over- or under-specified should normally result in the loss of a mark. This includes cases such as, for example, insistence that the value of a test statistic is (say) 2.128888446667 merely because that is the value that happened to come off the candidate's calculator. Note that this applies to answers that are given as final stages of calculations; intermediate working should usually be carried out, and quoted, to a greater degree of accuracy to avoid the danger of premature approximation.

The situation regarding any particular cases where the accuracy of the answer may be a marking issue should be detailed in the mark scheme rationale. If in doubt, contact your Team Leader.

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   Genuine misreading (of numbers or symbols, occasionally even of text) occurs. If this results in the object and/or difficulty of the question being considerably changed, it is likely that all the marks for that question, or section of the question, will be lost. However, misreads are often such that the object and/or difficulty remain substantially unaltered; these cases are considered below.

The simple rule is that *all* method ("M") marks [and of course all independent ("B") marks] remain accessible but at least some accuracy ("A") marks do not. It is difficult to legislate in an overall sense beyond this global statement because misreads, even when the object and/or difficulty remains unchanged, can vary greatly in their effects. For example, a misread of 1.02 as 10.2 (perhaps as a quoted value of a sample mean) may well be catastrophic; whereas a misread of 1.6748 as 1.6746 may have so slight an effect as to be almost unnoticeable in the candidate's work.

A misread should normally attract *some* penalty, though this would often be only 1 mark and should rarely if ever be more than 2. Commonly in sections of questions where there is a numerical answer either at the end of the section or to be obtained and commented on (eg the value of a test statistic), this answer will have an "A" mark that may actually be designated as "cao" [correct answer only]. This should be interpreted *strictly* – if the misread has led to failure to obtain this value, then this "A" mark must be withheld even if all method marks have been earned. It will also often be the case that such a mark is implicitly "cao"

even if not explicitly designated as such.

On the other hand, we commonly allow "fresh starts" within a question or part of question. For example, a follow-through of the candidate's value of a test statistic is generally allowed (and often explicitly stated as such within the marking scheme), so that the candidate may exhibit knowledge of how to compare it with a critical value and draw conclusions. Such "fresh starts" are not affected by any earlier misreads.

A misread may be of a symbol rather than a number – for example, an algebraic symbol in a mathematical expression. Such misreads are more likely to bring about a considerable change in the object and/or difficulty of the question; but, if they do not, they should be treated as far as possible in the same way as numerical misreads, *mutatis mutandis*. This also applied to misreads of text, which are fairly rare but can cause major problems in fair marking.

The situation regarding any particular cases that arise while you are marking for which you feel you need detailed guidance should be discussed with your Team Leader.

Note that a miscopy of the candidate's own working is not a misread but an accuracy error.

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

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

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

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

3	(iii)		$P(Y > 25) = 0.99 \Rightarrow P\left(Z > \dfrac{25-\mu}{\sigma}\right) = 0.99$ $\Rightarrow \dfrac{25-\mu}{\sigma} = \Phi^{-1}(0.99) = -2.326 \qquad \Rightarrow 25 = \mu - 2.326\sigma$ $P(Y > 25.4) = 0.75 \Rightarrow P\left(Z > \dfrac{25.4-\mu}{\sigma}\right) = 0.75$ $\Rightarrow \dfrac{25.4-\mu}{\sigma} = \Phi^{-1}(0.75) = -0.6745 \qquad \Rightarrow 25.4 = \mu - 0.6745\sigma$	B1 M1 A1	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
			$1.6515\sigma = 0.4 \qquad\qquad \sigma = 0.2422\ldots$	A1	
			$\mu = 25 + 2.326 \times 0.2422\ldots \qquad \mu = 25.563\ldots$	A1	
			$P(\text{Weighs} > 26.0) = P\left(Z > \dfrac{26.0-25.563}{0.2422}\right) = 1 - \Phi(1.804) = 1 - 0.9644 = 0.0356$	A1 **[6]**	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 $= (46 \times 79)/200$ $= 18.17$ AG	M1 A1 **[2]**	attempt at row total $\times$ 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 $(O-E)^2/E$ 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. **[3]**	B1 B1 B1	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: $\mu < 562$** 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 $= \dfrac{547-562}{27.4/\sqrt{12}}$	M1*	must include $\sqrt{12}$ with numerator as seen
			$= \dfrac{-15}{7.910} = -1.896$	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: $\mu < 562$ 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: $\mu < 562$ 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.

Additional Notes on Conclusions to Hypothesis Tests

The following are examples of conclusions which are considered too assertive.
There is sufficient evidence to reject H_0 and **conclude** that…
"there is a positive association between…" or
"there seems to be evidence that there is a positive association between…" or
"the mean nicotine content is 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.

PART OF THE
CAMBRIDGE ASSESSMENT
GROUP

UKAS
MANAGEMENT
SYSTEMS
001