OCR
Oxford Cambridge and RSA

GCE

Mathematics (MEI)

Unit **4767**: Statistics 2

Advanced GCE

Mark Scheme for June 2018

OCR (Oxford Cambridge and RSA) is a leading UK awarding body, providing a wide range of qualifications to meet the needs of candidates of all ages and abilities. OCR qualifications include AS/A Levels, Diplomas, GCSEs, Cambridge Nationals, Cambridge Technicals, Functional Skills, Key Skills, Entry Level qualifications, NVQs and vocational qualifications in areas such as IT, business, languages, teaching/training, administration and secretarial skills.

It is also responsible for developing new specifications to meet national requirements and the needs of students and teachers. OCR is a not-for-profit organisation; any surplus made is invested back into the establishment to help towards the development of qualifications and support, which keep pace with the changing needs of today's society.

This mark scheme is published as an aid to teachers and students, to indicate the requirements of the examination. It shows the basis on which marks were awarded by examiners. It does not indicate the details of the discussions which took place at an examiners' meeting before marking commenced.

All examiners are instructed that alternative correct answers and unexpected approaches in candidates' scripts must be given marks that fairly reflect the relevant knowledge and skills demonstrated.

Mark schemes should be read in conjunction with the published question papers and the report on the examination.

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

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)	**EITHER:** $S_{xy} = \Sigma xy - \frac{1}{n}\Sigma x \Sigma y = 739140 - \frac{1}{60} \times 10524 \times 4219$ (= -872.6) $S_{xx} = \Sigma x^2 - \frac{1}{n}(\Sigma x)^2 = 1849100 - \frac{1}{60} \times 10524^2$ (= 3190.4) $S_{yy} = \Sigma y^2 - \frac{1}{n}(\Sigma y)^2 = 303700 - \frac{1}{60} \times 4219^2$ (= 7033.9833…) $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-872.6}{\sqrt{3190.4 \times 7033.98}}$ = -0.1842 **OR:** cov $(x,y) = \frac{\sum xy}{n} - \overline{x}\overline{y} = 739140/60 - (10524/60 \times 4219/60)$ = -14.54333… rmsd(x) $= \sqrt{\frac{S_{xx}}{n}} = \sqrt{(3190.4/60)} = 7.29200…$ (msd(x) = 53.173...) rmsd(y) $= \sqrt{\frac{S_{yy}}{n}} = \sqrt{(7033.9833…/60)} = 10.8274…$ (msd(y) = 117.23...) $r = \frac{\text{cov(x,y)}}{rmsd(x)rmsd(y)} = \frac{-14.54333...}{7.29200... \times 10.8274...}$ = -0.1842	M1* M1* A1 M1dep* A1 **[5]** M1* M1* A1 M1dep* A1	For method for S_{xy} For method for at least one of S_{xx} or S_{yy} For at least one of S_{xy} S_{xx} or S_{yy} correct to 4sf. Can be implied by correct final answer. For structure of r For answers between – 0.1840 and – 0.1850. Allow – 0.184 www and – 0.185 www For method for cov (x,y) For method for at least one msd or rmsd For at least one of cov(x,y) one msd or rmsd correct to 4sf. Can be implied by correct final answer. For structure of r For answers between – 0.1840 and – 0.1850. Allow – 0.184 www and – 0.185 www Methods mixed – max 2/5	**Ao1 5**

Question			Answer	Marks	Guidance	
1	(ii)		H_0: $\rho = 0$ H_1: $\rho \neq 0$ (two-tailed test) where ρ is the (population) correlation coefficient between height and lifespan For $n = 60$, 5% critical value = ± **0.2542** Since -0.1842 > -0.2542 (or 0.1842 < 0.2542) the result is not significant. Thus we have insufficient evidence to reject H_0 There is insufficient evidence at the 5% level to suggest that there is correlation between height and lifespan.	B1 B1 B1 M1 A1 A1 **[6]**	For H_0, H_1 in symbols For defining ρ in context. Allow 'between x and y'. (must use ρ. Accept other variables, but not r, if defined as the **population** correlation coefficient between height and lifespan) For critical value No further marks from here if wrong. B0 for ±0.2545 For sensible comparison leading to any conclusion. FT their |r| <1 for final 3 marks For 'not significant' oe For non-assertive result in context	**Ao1 2** **Ao3 3** **Ao5 1**
1	(iii)		(Underlying population must have a) **bivariate Normal** distribution. **Yes**, since the points in the scatter diagram show a roughly **elliptical** shape.	B1 B1 **[2]**	Allow 'data comes from (a population with) a bivariate Normal distribution' but do not allow 'data has a bivariate Normal distribution' or 'both populations…' Allow **no,** since… …**not elliptical**	**Ao3 2**
1	(iv)		As the **population parameter is known,** There is no need to carry out a test.	E1* E1dep* **[2]**		**Ao2 2**
1	(v)		Correlation does not imply causation There could be another factor/other factors causing the relationship	B1 B1 **[2]**		**Ao3 2**
2	(i)		**Faults** occur randomly and independently There is a **uniform**/constant **mean rate** of occurrence	B1 B1 **[2]**	**Do not allow** 'events occur randomly…' Allow **uniform**/constant **average rate**…	**Ao3 2**
2	(ii)		Variance = 8.4	B1 **[1]**	Do not allow $\lambda = 8.4$ or Po(8.4)	**Ao3 1**

Question			Answer	Marks	Guidance	
2	(iii)		$P(X \geq 8) = 1 - 0.3987$ (from tables) $= 0.6013$	M1 A1 **[2]**		**Ao1 1 Ao5 1**
2	(iv)		Using Normal approx. to the Poisson, $X \sim N(42, 42)$	B1 B1	For Normal approximation **used**. For correct parameters (SOI)	**Ao2 2 Ao3 4**
			$P(X \geq 40) = P\left(Z \geq \frac{39.5-42}{\sqrt{42}}\right)$	B1	For 39.5 SOI **Beware use of 40, which gives z = - 0.3086 followed by incorrect use of tables – i.e. looking up -0.386 which leads to 0.6503.**	
				M1	For probability using correct tail	
			$= P(Z > -0.3858) = \Phi(0.3858) = 0.6502$ www	A1	CAO Allow 0.6503 www. Allow 0.650 www and 0.65www **(**Do not FT wrong or omitted CC)	
			Assumption: No additional assumption necessary since it is a Poisson process	E1 **[6]**		
2	(v)	(*A*)	$P(103 < X < 104) = P\left(\frac{103-103.2}{\sqrt{0.36}} \leq Z \leq \frac{104-103.2}{\sqrt{0.36}}\right)$	M1	For standardizing. M0 for using 'continuity corrections' or $\sigma = 0.36$	**Ao1 2 Ao5 1**
			$= P(-0.3333 < Z < 1.3333)$ $= \Phi(1.3333) - (1 - \Phi(0.3333))$ $= 0.9085 - (1 - 0.6305)$	M1	For correct structure	
			$= 0.5390$	A1 **[3]**	Accept awrt 0.539	

Question			Answer	Marks	Guidance	
2	(v)	(*B*)	$\Phi^{-1}(0.05) = -2.576$ $\frac{102-103.2}{\sigma} = -2.576$ $\sigma = \frac{102-103.2}{-2.576}$ (= 0.4658) So variance = $\sigma^2 = 0.2170$	B1 M1* M1dep* A1 **[4]**	For ±2.576 **1 – 2.576** gets B0M0*M0dep* if used later. For equation including σ, as seen or equivalent, with their z value which must lead to a positive value for σ (see additional note). M0 if c.c. used Allow other symbols for σ. Rearranging for σ CAO Allow 0.217 and allow 0.22 www	**Ao1 3** **Ao5 1**
3	(i)		P(X > 500) P(Z > -1.175) or 1 – P(Z < - 1.175) or P(Z < 1.175) = Φ(1.175) or 1 – Φ(– 1.175), leading to given answer of 0.8800(…)	 B1 B1 **[2]**	 For obtaining any of these probability statements involving Z correctly For either of these statements involving Φ leading to 0.8800(…) NOTE **AG**	**Ao1 1** **Ao4 1**
3	(ii)		$10 \times 0.88^9 \times 0.12 + 0.88^{10}$ = 0.6583	M1 A1 **[2]**	 Allow 0.658 www	**Ao2 2**
3	(iii)		Mean 100 × 0.88 = 88 , Var = 100 × 0.88 × 0.12 = 10.56 $X \sim$ N(88, 10.56) $P(X \geq 90) = P\left(Z \geq \frac{89.5-88}{\sqrt{10.56}}\right)$ = P(Z > 0.4616) = 1 – Φ(0.4616) = 1 – 0.6779 = 0.3221	 B1 B1 B1 M1 A1 **[5]**	 For using Normal approximation. For correct parameters (SOI) For continuity corr. For probability using correct tail CAO (Do not FT wrong or omitted CC) (Answer from calculator 0.3222) Accept 0.322 www	**Ao2 1** **Ao3 4**

Question			Answer	Marks	Guidance	
3	(iv)		$\bar{x}$ = 12580/25 = 503.2	B1		Ao1 2 Ao2 3 Ao3 2 Ao4 1 Ao5 1
			H_0: μ = 504.7;	B1	For hypotheses in symbols, both correct	
			H_1 : $\mu \neq$ 504.7	B1	For definition of μ in context. Accept other variables (apart from $\bar{x}$) if defined as the mean amount of active ingredient **in the population**	
			Where μ denotes the mean amount of active ingredient (in the population of pills)			
			Test statistic $= \frac{503.2-504.7}{4/\sqrt{25}} = \frac{-1.5}{0.8} = -1.875$	M1*	using their $\bar{x}$ and must include √25 M0* if numerator reversed.	
				A1	For – 1.875 or – 15/8	
			5% level 2 tailed critical value of z = ±1.96	B1	For (±) 1.96 but B0 if 1 ± 1.96 used later. No further marks from here if wrong. Allow B1 for -1.645 only with H_1: μ < 504.7 but no further marks in this case.	
			-1.875 > -1.96 (or 1.875 < 1.96)	M1dep*	FT candidate's test statistic only if negative for final 3 marks For sensible comparison as written leading to any conclusion.	
			The result is not significant. There is insufficient evidence to reject H_0	A1	For 'not significant', oe	
			There is insufficient evidence to suggest that the **mean amount** (of active ingredient) is not 504.7 milligrams.	A1	For conclusion in words in **context**	
				[9]		
3	(v)		The test has accepted H_0, which is a wrong result as the mean is no longer 504.7.	E1	E1 for recognising wrong result (Type II error)	Ao4 2
			If a larger sample had been selected, it would have been more likely that the change would have been detected.	E1	E1 for larger sample	
				[2]		

Question			Answer	Marks	Guidance	
4	(i)		H_0: no association between location and age category. H_1: some association between location and age category.	B1 [1]	Correct hypotheses in context. Allow hypotheses ito independence. B0 if correlation used in place of association. Ignore symbols if used.	Ao2 1
4	(ii)		Refer to χ^2_6 Critical value at 10% level = 10.64 8.752 < 10.6451 so result is **not significant** There is insufficient evidence to suggest that there is association between location and age category	B1 B1 M1 A1 [4]	for 6 deg of f for cv allow 10.645 No further marks from here if incorrect For sensible comparison leading to a correct conclusion. Allow '**Accept H_0**' or '**Reject H_1**'. NOTE **contradictory responses**, e.g. 'not significant so reject H_0', get M0A0. For non-assertive conclusion in context referring to H_1. A0 if correlation used in place of association.	Ao1 1 Ao2 2 Ao5 1
4	(iii)		Expected frequency = 88/200 × 43 = 18.92 Contribution = $(12 - 18.92)^2 / 18.92$ (= 2.5310)	M1 A1 A1 [3]	Method for f_e 18.92 seen For correct substitution into $(O-E)^2/E$ leading to answer of 2.5310. Condone numerator reversed. ***NB Answer given***	Ao1 3
4	(iv)		X^2 = 8.396 (or 8.3963) Critical value at 10% level = 7.779 8.396 > 7.779 so result is significant (oe) - there is sufficient evidence to suggest that there is association between location and age category	B1 B1 B1 [3]	 No further marks from here if incorrect. No FT if incorrect test statistic used.	Ao2 2 Ao5 1
4	(v)		Expected frequencies for housing estate for 2008 – 2012 = 16.13 (16.125) On the **housing estate,** there are two **high contributions** (or the contributions of **2.53 and 3.84**) that suggest that there are fewer cars in 2001-2007 group than expected and more cars in 2008-2012 group than expected. For the **other two locations and housing** estate in **2013 – 2017**, the **low contributions** suggest numbers are much **as expected**.	B1 M1 A1 E1	 Reference to either of these high contributions Correct interpretation for both of these groups E0 if (slightly) more/fewer mentioned. Note: comments referring to what would	Ao1 1 Ao4 3

Question			Answer	Marks	Guidance	
				[4]	be expected get A0 and/or E0	
4	(vi)		Advantage: Having more categories might reveal association which is not revealed by the present 3 categories. Disadvantage: Some of the expected frequencies might be less than 5 so categories would have to be combined.	E1 E1 [2]		Ao4 2

Additional notes Re Q3(iv)

Critical Value Method

c.v. = 504.7 – 1.96×4/√25 gets M1* B1

= 503.1(32) gets A1 (replacing A1 for -1.875)

503.2 > 503.1(32) with a conclusion gets M1dep* then final A1 A1 still available

NB if H_1: μ < 504.7 used award maxB1B0B1M1* B1(for -1.645) A1(for 503.4)depM0*A0A0

Probability Method

$P(Z < -1.875) = 0.0303$ gets B1 (for 0.0303 or 0.0304 replacing B1 for ±1.96)

0.0304 > 0.025 with conclusion gets M1dep* then final A1 A1 still available

NB if H_1: μ < 504.7 used award maxB1B0B1M1*A1B1(for 0.0303 or 0.0304)depM0*A0A0

Additional Note RE Over-specification

A0 or B0 for final answers given correct to 5sf or more. NOTE do not penalise over-specification more than twice in any single question or more than 4 times in a paper.

Additional Notes for Q2(v)*B*

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, e.g. – 1.645. Do not allow 0.005 or 0.995 to be treated as *z*-values here. The M1dep* can be awarded if the candidate correctly rearranges their equation to find σ. Hence, use of an incorrect *z*-value could earn max B0M1*M1dep*A0.

If $z = +2.576$ 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 Q1(ii) Writing – 0.2542 < – 0.1842 < 0.2542 or equivalent, leading to any conclusion, earns M1

In Q3 (iv) Neither – 1.875 < 0.05 nor 0.0304 < 1.96 are considered sensible as each compares a z-value with a probability.

Inequality sign reversed, e.g. – 1.875 < – 1.96, gets M0A0A0.

Comparing a negative with a positive *z*-value, e.g. – 1.875 < 1.96, gets M0A0A0, unless seen in an inequality which includes both tails, i.e. – 1.96 < – 1.875 < 1.96 which could lead M1A1A1.

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 an association between…" or
"there seems to be evidence that there is an association between…" or
"the mean amount of active ingredient has not changed …."
"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 Q4(ii) a conclusion simply stating that "the evidence does not suggest that there is an association" gets A0 as this does not refer to the context.

OCR (Oxford Cambridge and RSA Examinations)
The Triangle Building
Shaftesbury Road
Cambridge
CB2 8EA

OCR Customer Contact Centre

Education and Learning
Telephone: 01223 553998
Facsimile: 01223 552627
Email: general.qualifications@ocr.org.uk

www.ocr.org.uk

For staff training purposes and as part of our quality assurance programme your call may be recorded or monitored

Oxford Cambridge and RSA Examinations
is a Company Limited by Guarantee
Registered in England
Registered Office; The Triangle Building, Shaftesbury Road, Cambridge, CB2 8EA
Registered Company Number: 3484466
OCR is an exempt Charity

OCR (Oxford Cambridge and RSA Examinations)
Head office
Telephone: 01223 552552
Facsimile: 01223 552553