National
Qualifications
2017

2017 Statistics

Advanced Higher

Finalised Marking Instructions

SQA

©

General marking principles for Advanced Higher Statistics

This information is provided to help you understand the general principles you must apply when marking candidate responses to questions in this Paper. These principles must be read in conjunction with the detailed marking instructions, which identify the key features required in candidate responses.

For each question the marking instructions are generally in two sections, namely Illustrative Scheme and Generic Scheme. The Illustrative Scheme covers methods which are commonly seen throughout the marking. The Generic Scheme indicates the rationale for which each mark is awarded. In general, markers should use the Illustrative Scheme and only use the Generic Scheme where a candidate has used a method not covered in the Illustrative Scheme.

(a) Marks for each candidate response must always be assigned in line with these general marking principles and the detailed marking instructions for this assessment.

(b) Marking should always be positive. This means that, for each candidate response, marks are accumulated for the demonstration of relevant skills, knowledge and understanding: they are not deducted from a maximum on the basis of errors or omissions.

(c) If a specific candidate response does not seem to be covered by either the principles or detailed Marking Instructions, and you are uncertain how to assess it, you must seek guidance from your Team Leader.

(d) Credit must be assigned in accordance with the specific assessment guidelines.

(e) One mark is available for each •. There are no half marks.

(f) Working subsequent to an error must be **followed through**, with possible credit for the subsequent working, provided that the level of difficulty involved is approximately similar. Where, subsequent to an error, the working for a follow through mark has been eased, the follow through mark cannot be awarded.

(g) As indicated on the front of the question paper, full credit should only be given where the solution contains appropriate working. Unless specifically mentioned in the marking instructions, a correct answer with no working receives no credit.

(h) Candidates may use any mathematically correct method to answer questions except in cases where a particular method is specified or excluded.

(i) As a consequence of an error perceived to be trivial, casual or insignificant, eg $6 \times 6 = 12$ candidates lose the opportunity of gaining a mark. However, note the second example in comment **(j)**.

(j) Where a transcription error (paper to script or within script) occurs, the candidate should normally lose the opportunity to be awarded the next process mark, eg

$$x^2+5x+7=9x+4$$
$$x-4x+3=0$$
$$x=1$$

This is a transcription error and so the mark is not awarded. (refers to $x-4x+3=0$)

Eased as no longer a solution of a quadratic equation so mark is not awarded. (refers to $x=1$)

$$x^2+5x+7=9x+4$$
$$x-4x+3=0$$
$$(x-3)(x-1)=0$$
$$x=1 \text{ or } 3$$

Exceptionally this error is not treated as a transcription error as the candidate deals with the intended quadratic equation. The candidate has been given the benefit of the doubt and all marks awarded. (refers to $x-4x+3=0$)

(k) **Horizontal/vertical marking**
Where a question results in two pairs of solutions, this technique should be applied, but only if indicated in the detailed marking instructions for the question.

Example:

	$\bullet^5$	$\bullet^6$
$\bullet^5$	*x* = 2	*x* = −4
$\bullet^6$	*y* = 5	*y* = −7

Horizontal: $\bullet^5$ *x* = 2 and *x* = −4
$\bullet^6$ *y* = 5 and *y* = −7

Vertical: $\bullet^5$ *x* = 2 and *y* = 5
$\bullet^6$ *x* = −4 and *y* = −7

Markers should choose whichever method benefits the candidate, but **not** a combination of both.

(l) In final answers, unless specifically mentioned in the detailed marking instructions, numerical values should be simplified as far as possible, eg:

$\frac{15}{12}$ must be simplified to $\frac{5}{4}$ or $1\frac{1}{4}$

$\frac{43}{1}$ must be simplified to 43

$\frac{15}{0\cdot3}$ must be simplified to 50

$\frac{4/5}{3}$ must be simplified to $\frac{4}{15}$

$\sqrt{64}$ must be simplified to 8*

*The square root of perfect squares up to and including 100 must be known.

(m) Commonly Observed Responses (COR) are shown in the marking instructions to help mark common and/or non-routine solutions. CORs may also be used as a guide when marking similar non-routine candidate responses.

(n) Unless specifically mentioned in the marking instructions, the following should not be penalised:

- Working subsequent to a correct answer
- Correct working in the wrong part of a question
- Legitimate variations in numerical answers/algebraic expressions, eg angles in degrees rounded to nearest degree
- Omission of units
- Bad form (bad form only becomes bad form if subsequent working is correct), eg $(x^3+2x^2+3x+2)(2x+1)$ written as $(x^3+2x^2+3x+2)\times 2x+1$

  $2x^4+4x^3+6x^2+4x+x^3+2x^2+3x+2$ written as $2x^4+5x^3+8x^2+7x+2$ gains full credit

- Repeated error within a question, but not between questions or papers

(o) In any 'Show that…' question, where the candidate has to arrive at a required result, the last mark of that part is not available as a follow-through from a previous error unless specified in the detailed marking instructions.

(p) All working should be carefully checked, even where a fundamental misunderstanding is apparent early in the candidate's response. Marks may still be available later in the question so reference must be made continually to the marking instructions. The appearance of the correct answer does not necessarily indicate that the candidate has gained all the available marks.

(q) Scored-out working which has not been replaced should be marked where still legible. However, if the scored out working has been replaced, only the work which has not been scored out should be marked.

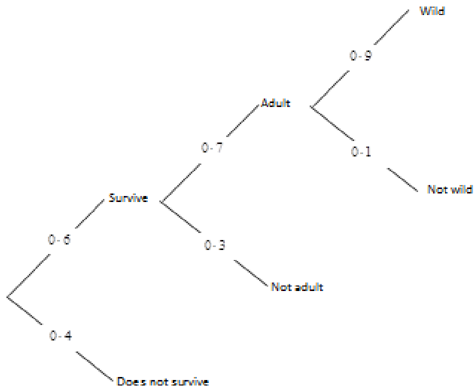
(r) Where a candidate has made multiple attempts using the same strategy and not identified their final answer, mark all attempts and award the lowest mark.

Where a candidate has tried different valid strategies, apply the above ruling to attempts within each strategy and then award the highest resultant mark.

For example:

Strategy 1 attempt 1 is worth 3 marks.	Strategy 2 attempt 1 is worth 1 mark.
Strategy 1 attempt 2 is worth 4 marks.	Strategy 2 attempt 2 is worth 5 marks.
From the att empts using strategy 1, the resultant mark would be 3.	From the attempts using strategy 2, the resultant mark would be 1.

In this case, award 3 marks.

Question			Generic scheme	Illustrative scheme	Max mark
1	(a)	(i)	•[1] appropriate strategy	•[1]  **OR** $0\cdot6\times0\cdot7\times0\cdot9=$	**3**
			•[2] correct probability	•[2] 0·378	
		(ii)	•[3] correct probability	•[3] $0\cdot6\times0\cdot7\times0\cdot1=0\cdot042$	
Notes:					
	(b)		•[4] correct probability	•[4] $0\cdot4+0\cdot6\times0\cdot3=0\cdot58$	**2**
			•[5] calculate proportion	•[5] 0·4/0·58 = 0·69	
Notes:					

<table>
<tr><th colspan="3">Question</th><th>Generic scheme</th><th>Illustrative scheme</th><th>Max mark</th></tr>
<tr><td>2</td><td>(a)</td><td></td><td>•1 correct probability distribution</td><td>•1
<table><tr><td>x</td><td>20</td><td>30</td><td>40</td><td>50</td><td>60</td></tr><tr><td>p(x)</td><td>$\frac{1}{5}$</td><td>$\frac{1}{5}$</td><td>$\frac{1}{5}$</td><td>$\frac{1}{5}$</td><td>$\frac{1}{5}$</td></tr></table></td><td>1</td></tr>
<tr><td colspan="6">Notes:</td></tr>
<tr><td></td><td>(b)</td><td>(i)</td><td>•2 calculate E(X)

•3 calculate V(X)</td><td>•2 E(X) = 40

•3 V(X) = 200</td><td>3</td></tr>
<tr><td></td><td>(ii)</td><td></td><td>•4 calculate probability</td><td>•4 $P(X > \mu) = \frac{2}{5}$</td><td></td></tr>
<tr><td colspan="6">Notes:</td></tr>
</table>

Question			Generic scheme	Illustrative scheme	Max mark
3	(a)	(i)	•[1] appropriate reason	•[1] the Board may not be willing to release such confidential information	**4**
		(ii)	•[2] appropriate sampling method •[3] appropriate strategy •[4] continue method	•[2] cluster sampling •[3] chose, say, 3 centres at random •[4] use all the data from the 3 centres	
Notes:					
	(b)		•[5] appropriate hypotheses •[6] correct s^2 •[7] calculate t •[8] correct cv •[9] deal with H_0 •[10] appropriate conclusion	•[5] $H_o: \mu_p = \mu_v \quad H_1: \mu_p < \mu_v$ •[6] $s^2 = \dfrac{(n_p - 1)s_p^2 + (n_v - 1)s_v^2}{(n_p + n_v - 2)} = 83\cdot 29$ •[7] $t_{n_p+n_v-2} = \dfrac{\bar{x}_p - \bar{x}_v}{s\sqrt{\dfrac{1}{n_p} + \dfrac{1}{n_v}}} = -1\cdot 497$ •[8] $t_{28,0\cdot950} = -\ 1\cdot701$ •[9] 1·497 < 1·701 so we cannot reject Ho at the 5% level of significance •[10] and conclude that there is no evidence of pianists being awarded lower marks	**6**
Notes: the alternative p-value approach (hereafter the PvA) would record that $P(t_{28,0\cdot95} \le 1\cdot 497) = 0\cdot 0728 > 0\cdot 05 \ldots$					

Question			Generic scheme	Illustrative scheme	Max mark
4	(a)		•[1] correct first fence •[2] calculate second fence •[3] appropriate comment	•[1] 29 − 1·5 x 3 = 24·5 •[2] 32 + 1·5 x 3 = 36·5 •[3] 22 is an outlier	3
Notes:					
	(b)	(i)	•[4] appropriate comment	•[4] not at all useful	3
		(ii)	•[5] appropriate diagram •[6] appropriate analysis	•[5] scatterplot •[6] least squares regression analysis	
Notes:					

Question			Generic scheme	Illustrative scheme	Max mark
5	(a)	(i)	•[1] correct limits	•[1] $-6\cdot1\pm6\times\frac{9\cdot0561}{\sqrt{5}}=[-30\cdot40,\ 18\cdot20]$	**2**
		(ii)	•[2] correct week	•[2] week 25	
Notes:					
	(b)	(i)	•[3] correct probability	•[3] $p = 2\times0\cdot5^9 = 0\cdot0039$	**3**
		(ii)	•[4] correct week •[5] correct week	•[4] ninth consecutive point above target value occurs in week 22 •[5] second out of three values above 2 sigma occurs in week 16	
Notes:					

Question			Generic scheme	Illustrative scheme	Max mark
6			•[1] appropriate assumption •[2] appropriate strategy •[3] correct μ •[4] correct σ^2 •[5] appropriate strategy •[6] calculate probability	•[1] assuming that the weights of skiers and equipment are independent •[2] $T = (S1 + S2 + \ldots + S130) + (E1 + E2 + \ldots + E130)$ •[3] 11050 •[4] 2340 •[5] $P(T>11200)$ $= P\left(z > \frac{11200 - 11050}{\sqrt{2340}}\right)$ •[6] = 0.001	**6**
Notes:					

Question			Generic scheme	Illustrative scheme	Max mark
7	(a)		•[1] correct strategy •[2] correct t value •[3] calculate interval	•[1] 95% CI is given by $\bar{x} \pm t\frac{s}{\sqrt{n}}$ •[2] $t_{17,0\cdot975}$ =2·110 •[3] (6·73, 8·19)	**3**
Notes:					
	(b)	(i)	•[4] appropriate observation •[5] appropriate comment	•[4] 5·87 is outwith the CI •[5] so there is evidence that the number of large trees may have declined	**3**
		(ii)	•[6] appropriate comment	•[6] no because no other factors have been considered	
Notes:					

Question			Generic scheme	Illustrative scheme	Max mark
8	(a)		$\bullet^1$ correct parameters $\bullet^2$ appropriate strategy $\bullet^2$ calculate probability	$\bullet^1$ $\mu = \sigma^2 = 4$ $\bullet^2$ $\text{P}(2 \le W \le 6)$ $\bullet^3$ = 0·7977	**3**
Notes:					
	(b)		$\bullet^4$ appropriate strategy $\bullet^5$ calculate probability	$\bullet^4$ $\text{P}(-1 \le Z \le 1)$ $\bullet^5$ = 0·6827	**2**
Notes:					
	(c)		$\bullet^6$ correct □ $\bullet^7$ appropriate strategy $\bullet^8$ calculate probability	$\bullet^6$ $\sigma = \frac{b-a}{2\sqrt{3}} = \frac{2}{\sqrt{3}}$ $\bullet^7$ $\text{P}(8 - \frac{2}{\sqrt{3}} \le Y \le 8 + \frac{2}{\sqrt{3}})$ $\bullet^8$ $= \frac{1}{4}.\frac{4}{\sqrt{3}}$ = 0·5774	**3**
Notes:					

Question			Generic scheme	Illustrative scheme	Max mark
9	(a)		•[1] identify problem •[2] appropriate solution	•[1] A building, path or concrete area might be at the centre ie no soil •[2] Sample as close to the centre as possible	**2**
Notes: other acceptable responses are possible eg sample a random point in the grid					
	(b)		•[3] correct distribution •[4] correct parameters •[5] appropriate comment	•[3] $\bar{X}$ is approx N •[4] $\mathrm{N}\left(165\cdot6, \frac{23\cdot1^2}{25}\right)$ •[5] the parent distribution is skewed but n = 25 so the CLT is invoked	**3**
Notes:					
	(c)		•[6] appropriate hypotheses •[7] correct z-value •[8] correct critical value •[9] deal with H_0 •[10] appropriate conclusion	•[6] $H_0 : \mu = 165.6$ $H_1 : \mu > 165.6$ •[7] $z = \frac{174\cdot5 - 165\cdot6}{23\cdot1/5} = 1\cdot93$ •[8] $cv = 1\cdot64$ •[9] 1·93 > 1·64 so reject H_0 at the 5% level of significance •[10] and conclude that there is evidence that the soil lead concentration in village A is greater than 165·6 mg/kg	**5**
Notes: the PvA approach would record that $\mathrm{P}(Z \geq 1\cdot93) = 0\cdot0268 < 0\cdot05$					

Question			Generic scheme	Illustrative scheme	Max mark
10	(a)		•[1] correct distribution •[2] calculate probability	•[1] $X \sim B(20, 0{\cdot}2)$ •[2] $P(X = 3) = 0.2053$	2
Notes:					
	(b)		•[3] correct distribution •[4] correct continuity correction •[5] calculate z- values •[6] calculate probability	•[3] $X \approx N(9, 7{\cdot}2)$ •[4] $P(4{\cdot}5 < X < 10{\cdot}5)$ •[5] $P(-1{\cdot}68 < Z < 0{\cdot}56)$ •[6] 0·6658	4
Notes:					
	(c)		•[7] appropriate hypothesis •[8] correct E_i •[9] calculate value of x^2 •[10] correct critical value •[11] deal with H_0 •[12] appropriate conclusion	•[7] H_0: there is no association between sex and recapture status H_1: there is an association •[8] expected frequencies are 62·49 | 46·51 250·51 | 186·49 •[9] $x^2 = 0{\cdot}9427$ •[10] $x^2_{1,0\cdot950} = 3{\cdot}841$ •[11] $3{\cdot}841 > 0{\cdot}9427$ so we cannot reject H_0 at the 5% level of significance •[12] and conclude that there is no evidence of an association between sex and the likelihood of recapture	6
Notes: the PvA approach would record that $P(\chi^2_{1,0\cdot95} \geq 0\cdot9427) = 0\cdot3316 > 0\cdot05 \ldots$. Alternatively, using a z-test for proportions yields a pooled proportion of $0\cdot1996$ and a z-score of $0\cdot9831 \ . \ . \ .$					

Question			Generic scheme	Illustrative scheme	Max mark
11	(a)		•[1] correct strategy •[2] correct substitution •[3] calculate interval •[4] appropriate comment	•[1] 99% CI for p is $\hat{p} \pm z_{0.995}\sqrt{\frac{\hat{p}\hat{q}}{n}}$ •[2] $= 0 \cdot 61 \pm 2 \cdot 58\sqrt{\frac{0 \cdot 61.0 \cdot 39}{100}}$ •[3] = (0·484, 0·736) •[4] the calculation involves a normal approximation to the binomial distribution	**4**
Notes:					
	(b)		•[5] appropriate strategy •[6] correct substitution •[7] calculate *n*	•[5] lower limit > 50% •[6] $0 \cdot 58 - 2 \cdot 58\sqrt{\frac{0 \cdot 58.0 \cdot 42}{n}} > 0 \cdot 50$ •[7] $n \geq 254$	**3**
Notes:					

Question			Generic scheme	Illustrative scheme	Max mark
12	(a)		•1 identify lack of normality •2 identify same shape	•1 the distribution of both data sets is skewed •2 both data sets appear to have similar shaped distributions	**2**
Notes: other acceptable responses are possible eg the parent distribution is unknown					
	(b)	(i)	•3 correct hypotheses •4 correct sum of ranks •5 correct critical value •6 deal with H_0 •7 appropriate conclusion	•3 $H_0 : \eta_1 = \eta_2$ $H_1 : \eta_1 \neq \eta_2$ •4 SoR for Strip 2 is 340 •5 cv = 348 •6 340<348 so we reject H_0 at the 10% level of significance •7 and conclude that there is evidence that the median bluebell coverage in strips 1 and 2 is different	**8**
		(ii)	•8 deal with 5% level •9 appropriate comment	•8 340>337 so we cannot reject H_0. There is no evidence, at the 5% level of significance, that there is a difference between median bluebell coverage in strips 1 and 2 •9 the sensitivity of the result to the significance level indicates that no firm conclusion can be drawn regarding whether or not bluebell coverage in the two strips is different	
		(iii)	•10 appropriate suggestion	•10 sample more quadrats within each section	
Notes: other acceptable suggestions are possible					

Question		Generic scheme	Illustrative scheme	Max mark
13	(a)	•[1] appropriate comment •[2] appropriate comment	•[1] There appears to be no linear relationship between the two variables and •[2] the data for Florida seems to be an outlier	**2**
Notes: Acceptable alternatives might be that there appears to be a linear relationship apart from the outlier OR because of the outlier it is not possible to comment on a linear relationship OR both Florida & South Florida appear to the outliers.				
	(b)	•[3] correct value of R^2 •[4] appropriate comment	•[3] $R^2 = \frac{7\cdot5234^2}{26\cdot2676 \times 2\cdot6395} = 0\cdot8164$ •[4] 82% of the variation in the y variable, $\log_{10}$ (Number of species), can be attributed to variation in $\log_{10}$ Area	**2**
Notes: •[4] can also be awarded for "82% of the variation in y can be attributed to the linear model"				
	(c)	•[5] correct strategy •[6] correct fitted value •[7] correct SSR •[8] calculate s •[9] appropriate substitution •[10] calculate limits •[11] calculate species limits	•[5] $\hat{Y}_i \pm t_{11,\,0\cdot950}\, s\sqrt{1+\frac{1}{n}+\frac{(x-\overline{x})^2}{S_{xx}}}$ •[6] $\hat{y} = 0\cdot9202 + 0\cdot2864 \times 2\cdot5441 = 1\cdot6488$ •[7] $SSR = 2\cdot6395 - \frac{7\cdot5234^2}{26\cdot2676} = 0\cdot4847$ •[8] $s = \sqrt{\frac{SSR}{n-2}} = \sqrt{\frac{0\cdot4847}{11}} = 0\cdot2099$ •[9] $1\cdot6488 \pm 1\cdot796 \times 0\cdot2099\sqrt{1+\frac{1}{13}+\frac{(2\cdot5441-1\cdot3365)^2}{26\cdot2676}}$ •[10] (1·248, 2·050) •[11] Minimum no. of species = $10^{1\cdot248}$ =17·62, so 18 Maximum no. of species = $10^{2\cdot050}$ =112·2, so 112	**7**
Notes:				

[END OF MARKING INSTRUCTIONS]