National Qualifications 2025

2025 Statistics

Advanced Higher Paper 1

Question Paper Finalised Marking Instructions

These marking instructions have been prepared by examination teams for use by SQA appointed markers when marking external course assessments.

General marking principles for Advanced Higher Statistics

Always apply these general principles. Use them in conjunction with the detailed marking instructions, which identify the key features required in candidates' responses.

The marking instructions for each question are generally in two sections:

- *generic scheme — this indicates why each mark is awarded*
- *illustrative scheme — this covers methods which are commonly seen throughout the marking*

In general, you should use the illustrative scheme. Only use the generic scheme where a candidate has used a method not covered in the illustrative scheme.

(a) Always use positive marking. This means candidates accumulate marks for the demonstration of relevant skills, knowledge and understanding; marks are not deducted for errors or omissions.

(b) If you are uncertain how to assess a specific candidate response because it is not covered by the general marking principles or the detailed marking instructions, you must seek guidance from your team leader.

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

(d) If a candidate's response contains an error, all working subsequent to this error must still be marked. Only award marks if the level of difficulty in their working is similar to the level of difficulty in the illustrative scheme.

(e) Only award full marks where the solution contains appropriate working. A correct answer with no working receives no mark, unless specifically mentioned in the marking instructions.

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

(g) If an error is trivial, casual or insignificant, for example 6 × 6 = 12, candidates lose the opportunity to gain a mark, except for instances such as the second example in point (h) below.

(h) If a candidate makes a transcription error (question paper to script or within script), they lose the opportunity to gain the next process mark, for example

This is a transcription error and so the mark is not awarded.

This is no longer a solution of a quadratic equation, so the mark is not awarded.

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

The following example is an exception to the above

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.

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

(i) Horizontal/vertical marking

If a question results in two pairs of solutions, apply the following technique, 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 $x = -4$ Vertical: $\bullet^5$ $x = 2$ and $y = 5$

$\bullet^6$ $y = 5$ and $y = -7$ $\bullet^6$ $x = -4$ and $y = -7$

You must choose whichever method benefits the candidate, **not** a combination of both.

(j) In final answers, candidates should simplify numerical values as far as possible unless specifically mentioned in the detailed marking instruction. For example

$\frac{15}{12}$ must be simplified to $\frac{5}{4}$ or $1\frac{1}{4}$ $\frac{43}{1}$ must be simplified to 43

$\frac{15}{0{\cdot}3}$ 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 144 must be known.

(k) Do not penalise candidates for any of the following, unless specifically mentioned in the detailed marking instructions:

- working subsequent to a correct answer
- correct working in the wrong part of a question
- legitimate variations in numerical answers/algebraic expressions, for example angles in degrees rounded to nearest degree
- omission of units
- bad form (bad form only becomes bad form if subsequent working is correct), for example

$\left(x^3 + 2x^2 + 3x + 2\right)\left(2x + 1\right)$ *written as*

$\left(x^3 + 2x^2 + 3x + 2\right) \times 2x + 1$

$= 2x^4 + 5x^3 + 8x^2 + 7x + 2$

gains full credit

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

(l) In any 'Show that…' question, where candidates have to arrive at a required result, the last mark is not awarded as a follow-through from a previous error, unless specified in the detailed marking instructions.

(m) You must check all working carefully, even where a fundamental misunderstanding is apparent early in a candidate's response. You may still be able to award marks later in the question so you must refer continually to the marking instructions. The appearance of the correct answer does not necessarily indicate that you can award all the available marks to a candidate.

(n) You should mark legible scored-out working that has not been replaced. However, if the scored-out working has been replaced, you must only mark the replacement working.

(o) If candidates make multiple attempts using the same strategy and do not identify their final answer, mark all attempts and award the lowest mark. If candidates try different valid strategies, apply the above rule to attempts within each strategy and then award the highest 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 attempts 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.

Note: Marking from Image (MFI) annotation change from 2025

A double cross-tick is used to indicate correct working which is irrelevant or insufficient to score any marks. In MFI marking instructions prior to 2025 this was shown as **ü₂** or ü2.

From 2025, the double cross-tick will no longer be used in MFI. A cross or omission symbol will be used instead.

Marking Instructions for each question

Question			Generic scheme	Illustrative scheme	Max mark
1.	(a)		$\bullet^1$ appropriate assumption $\bullet^2$ appropriate assumption $\bullet^3$ appropriate assumption	$\bullet^1$ independence of occurrence of filler sounds $\bullet^2$ no two filler sounds happen at the same time $\bullet^3$ the mean rate of use of filler sounds is constant (over time)	**3**
Notes: 1. For $\bullet^1$, $\bullet^2$, $\bullet^3$, the context of 'filler sounds' needs to be mentioned in at least one response. 2. For $\bullet^1$, also accept: 'independence of use of filler sounds'. 'independence of timing of filler sounds'. 3. For $\bullet^1$, do not accept 'independence of (choice of) filler sounds'. 4. For $\bullet^3$, do not accept 'the average rate….'.					
Commonly Observed Responses:					
	(b)		$\bullet^4$ appropriate description	$\bullet^4$ rescale both axes to be the same across all three graphs	**1**
Notes: 1. For $\bullet^4$, also accept 'rescale the vertical axes to be the same across all three graphs'. 2. For $\bullet^4$, do not accept 'rescale x-axis scale only', or similar.					
Commonly Observed Responses:					
	(c)		$\bullet^5$ correct value	$\bullet^5$ $\dfrac{97}{28} = 3.4643$	**1**
Notes: 1. For $\bullet^5$, do not accept 3.5 with no further working. 2. For $\bullet^5$, the value can be obtained from reading the report's values of 97 and 14 via $\dfrac{97}{2\times 14}$.					
Commonly Observed Responses:					
	(d)		$\bullet^6$ calculate cumulative probability $\bullet^7$ calculate expected frequency	$\bullet^6$ $\mathrm{P}(X \geq 5) = 1 - \mathrm{P}(X \leq 4) = 0.2746$ $\bullet^7$ … × 28 = 7.69	**2**
Notes: 1. For $\bullet^6$ and $\bullet^7$, also accept $28 - (0.85 + 2.96 + 5.18 + 6.04 + 5.29) = 7.68$ 2. Evidence for mark $\bullet^6$ may be found in the response for mark $\bullet^7$.					
Commonly Observed Responses:					

Question			Generic scheme	Illustrative scheme	Max mark
1.	(e)		$\bullet^8$ correct degrees of freedom $\bullet^9$ appropriate justification	$\bullet^8$ $6 - 2 = 4$ $\bullet^9$ we are estimating the distribution parameter from the observed data	**2**

Notes:

1. For $\bullet^9$, also accept inclusion of information about both constraints (sum of expected frequencies being fixed; mean of expected frequencies also fixed).
2. If $6 - 1 = 5$ is given for $\bullet^8$, then $\bullet^9$ is available for 'we have one constraint of the total sample size'.

Commonly Observed Responses:

Question			Generic scheme	Illustrative scheme	Max mark
1.	(f)	(i)	•[10] appropriate explanation	•[10] none of expected frequencies should be less than one	**1**

Notes:

1. For •[10], reference must explicitly be made to expected frequencies.
2. For •[10], also accept 'at present only 4 out of 6 of the expected frequencies are greater than or equal to 5, (which is only 66.7%), when it should be at least 80%'.

Commonly Observed Responses:

		(ii)	•[11] correct combined frequencies •[12] calculate test statistic	•[11] O_i = 3 and E_i = 3.81 •[12] $X^2 = 0.331$	**2**

Notes:

1. Mark •[11] can be implied from a correct response to mark •[12].
2. For •[12], allow carry through from part (d) – see accepted commonly observed responses, below.

Commonly Observed Responses:

Candidate A

$$X^2 = \frac{(3-3.81)^2}{3.81} + \ldots + \frac{(8-6.94)^2}{6.94} = 0.1722 + \ldots + 0.1619 = 0.4801$$

Candidate B

$$X^2 = \frac{(3-3.81)^2}{3.81} + \ldots + \frac{(8-7.68)^2}{7.68} = 0.1722 + \ldots + 0.0133 = 0.3315$$

Candidate C

$$X^2 = \frac{(3-3.81)^2}{3.81} + \ldots + \frac{(8-7.69)^2}{7.69} = 0.1722 + \ldots + 0.0125 = 0.3307$$

Candidate D

$$X^2 = \frac{(3-3.8052)^2}{3.8052} + \ldots + \frac{(8-6.94)^2}{6.94} = 0.1704 + \ldots + 0.1619 = 0.4783$$

Candidate E

$$X^2 = \frac{(3-3.8052)^2}{3.8052} + \ldots + \frac{(8-7.68)^2}{7.68} = 0.1704 + \ldots + 0.0133 = 0.3287$$

Candidate F

$$X^2 = \frac{(3-3.8052)^2}{3.8052} + \ldots + \frac{(8-7.69)^2}{7.69} = 0.1704 + \ldots + 0.0125 = 0.3289$$

Question			Generic scheme	Illustrative scheme	Max mark
1.	(f)	(iii)	•[13] state hypotheses	•[13] H_0: number of filler sounds (in 30 seconds) ~ Po(3.5) H_1: number of filler sounds (in 30 seconds) is **not** ~ Po(3.5)	**4**
			•[14] state critical value	•[14] $\chi^2_{3,0.90} = 6.251$	
			•[15] deal with H_0	•[15] as $0.331 < 6.251$ do not reject H_0 (at the 10% level)	
			•[16] appropriate conclusion	•[16] conclude we have insufficient evidence to suggest that the number of filler sounds (in 30 seconds) is not a Po(3.5) distribution	

Notes:

1. For •[13], also accept 'H_0: X distributed as Po(3.5) and H_1: X not distributed as Po(3.5)'.
2. For •[14], also accept p-value = 0.9545
3. For •[14], if an incorrect number of degrees of freedom were stated in (e) from only subtracting 1, then mark •[14] can be awarded for $\chi^2_{4,0.90} = 7.779$ as a follow through error.
4. For •[15], also accept '$0.9545 > 0.10$ so do not reject H_0'.
5. For •[15], also accept use of the test statistic value calculated in part (f)(ii).
6. For •[15], do not accept 'accept H_0'.
7. For •[16], do not accept conclusions phrased in terms of H_0 (eg. 'we have evidence to suggest that the number of filler sounds in 30 seconds is a Po(3.5) distribution').
8. For •[16], do not accept conclusions that are too definite. Phrasing must include 'evidence to conclude…', or 'evidence to suggest…', or similar.

Commonly Observed Responses:

	(g)		•[17] appropriate assumption	•[17] the (population) distributions (of the number of filler sounds in 30 seconds) were normally distributed	**1**

Notes:

1. For •[17], do not accept responses that mention 'sample'.

Commonly Observed Responses:

Question			Generic scheme	Illustrative scheme	Max mark
2.	(a)		$\bullet^1$ correct value	$\bullet^1$ $293 + 2 = 295$	**1**
Notes:					
Commonly Observed Responses:					
	(b)		$\bullet^2$ deal with H_0 $\bullet^3$ appropriate conclusion	$\bullet^2$ p-value $< 0.0001 < 0.05$, so we reject H_0 $\bullet^3$ and conclude that there is evidence that the (linear) correlation between sprint times and hurdle times is non-zero	**2**
Notes: 1. For $\bullet^2$, also accept: '$13.876 > 1.96$, so reject H_0'. 'p-value is very small, so reject H_0'. 2. For $\bullet^2$, do not accept: 'accept H_1'. 'p-value < 0.025' (if a significance level is mentioned, only 0.05 is accepted). 'p-value < 0.0001, so reject H_0'. 3. For $\bullet^3$, also accept: '…linear association between sprint times and hurdle times'. '…linear relationship between sprint times and hurdle times'. '…positive correlation between sprint times and hurdle times'. 4. For $\bullet^3$, do not accept 'sprint times and hurdle times have a correlation'. 5. For $\bullet^3$, do not accept conclusions that are too definite. Phrasing must include 'evidence to conclude…', or 'evidence to suggest…', or similar.					
Commonly Observed Responses:					
	(c)		$\bullet^4$ appropriate comment $\bullet^5$ appropriate comment	$\bullet^4$ $E(\varepsilon_i) = 0$ for all x_i, appears to be valid since the residuals are randomly scattered and centred around zero $\bullet^5$ $V(\varepsilon_i)$ = constant for all x_i, appears to be valid since the distribution of residuals does not depend upon the fitted value	**2**
Notes: 1. For $\bullet^4$ and $\bullet^5$, do not penalise mention of normality. 2. For $\bullet^4$ and $\bullet^5$, do not accept any mention of 'independence'. 3. For $\bullet^4$ and $\bullet^5$, do not accept 'the assumptions are valid' with no other comments.					
Commonly Observed Responses:					

Question			Generic scheme	Illustrative scheme	Max mark
2.	(d)		•[6] correct value •[7] correct value	•[6] $\begin{cases} 24.1366 - 0.9665 \times 13.09 \\ = 11.4851 \end{cases}$ •[7] $\begin{cases} 24.1366 + (24.1366 - 22.5224) \\ = 25.7508 \end{cases}$	**2**
Notes:					
Commonly Observed Responses:					
	(e)		•[8] appropriate assumption	•[8] (residuals are) normally distributed	**1**
Notes:					
Commonly Observed Responses:					
	(f)		•[9] correct interpretation	•[9] we would expect her (individual) sprint time (in the same competition) to be between 22.5224 and 25.7508 seconds, 99% of the time	**1**
Notes: 1. For •[9], also accept upper bound value calculated in mark •[7] of part (d). 2. For •[9], also accept '… 99 times out of 100'. 3. For •[9], do not accept any of the following phrases: '99% confident'. '99% chance'. '99% certain'. 'with a 99% likelihood'.					
Commonly Observed Responses:					
	(g)		•[10] appropriate explanation •[11] appropriate suggestion	•[10] the model has fitted sprint times (y) on hurdle times (x) •[11] a new model fitting hurdle times (x) on sprint times (y) needs to be used	**2**
Notes:					
Commonly Observed Responses:					

Question			Generic Scheme	Illustrative Scheme	Max Mark
2.	(h)		•[12] correct explanation •[13] appropriate reason	•[12] a confidence interval is where we'd expect the **mean** sprint time to be captured, (from repeated competitions where the hurdles time was 13.09 seconds) •[13] Katerina Johnson-Thompson's individual sprint time was not a mean sprint time	**2**
Notes:					
Commonly Observed Responses:					

[END OF MARKING INSTRUCTIONS]