National Qualifications 2024

2024 Statistics

Advanced Higher – Paper 2

Question Paper Finalised Marking Instructions

© Scottish Qualifications Authority 2024

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

The information in this document may be reproduced in support of SQA qualifications only on a non-commercial basis. If it is reproduced, SQA must be clearly acknowledged as the source. If it is to be reproduced for any other purpose, written permission must be obtained from permissions@sqa.org.uk.

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

Vertical: $\bullet^5$ $x = 2$ and $y = 5$ $\quad$ $\bullet^6$ $x = -4$ 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 $1\frac{1}{4}$ $\qquad$ $\frac{43}{1}$ must be simplified to 43

$\frac{15}{0{\cdot}3}$ must be simplified to 50 $\qquad$ $\frac{4\frac{}{}5}{3}$ 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

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

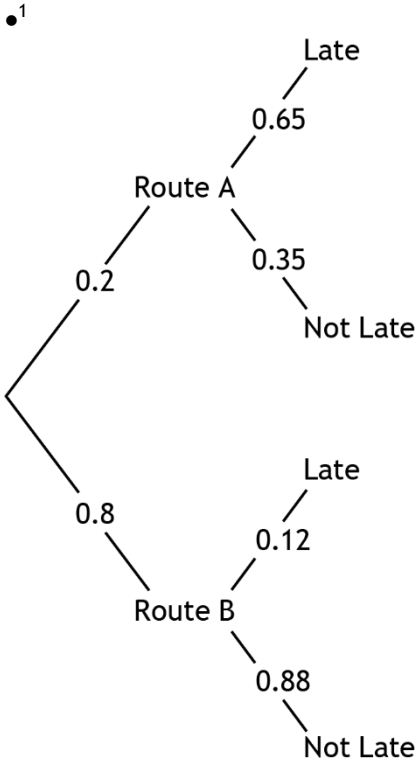
For example:

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.

Marking instructions for each question

Question			Generic scheme	Illustrative scheme	Max mark
1.			$\bullet^1$ calculate upper fence $\bullet^2$ calculate lower fence $\bullet^3$ appropriate comment	$\bullet^1$ $90 + 1.5 \times 27 = 130.5$ $\bullet^2$ $63 - 1.5 \times 27 = 22.5$ $\bullet^3$ $11 < 22.5$ so it is a possible outlier	**3**
Notes:					
Commonly Observed Responses: **Candidate A** calculates interquartile range incorrectly. Mark $\bullet^1$ not available Marks $\bullet^2$ and $\bullet^3$ available, as follow through errors.					

Question			Generic scheme	Illustrative scheme	Max mark
2.	(a)		$\bullet^1$ appropriate strategy	$\bullet^1$ 	**2**
			$\bullet^2$ calculate probability	$\bullet^2$ 0.226	
Notes: 1. Mark $\bullet^1$ can be implied by mark $\bullet^2$					
Commonly Observed Responses:					
	(b)		$\bullet^3$ appropriate strategy	$\bullet^3$ $P(B \mid L) = \frac{P(B \cap L)}{P(L)}$	**3**
			$\bullet^4$ correct substitution	$\bullet^4$ $\frac{0.096}{0.226}$	
			$\bullet^5$ calculate probability	$\bullet^5$ 0.4248	
Notes: 1. Mark $\bullet^3$ can be implied from mark $\bullet^4$. 2. For $\bullet^4$, the substitution must be consistent with the communicated strategy. 3. For $\bullet^5$, the probability must be given to at least 3 significant figures. 4. For $\bullet^5$, the probability must be a valid value, $0 \leq p \leq 1$					
Commonly Observed Responses:					

Question			Generic scheme	Illustrative scheme	Max mark
3.	(a)		$\bullet^1$ state distribution $\bullet^2$ calculate probability	$\bullet^1$ $X \sim \text{B}(12, 0.88)$ $\bullet^2$ $\text{P}(X = 9) = 0.1203$	**2**
Notes: 1. For $\bullet^1$, do not accept 'B(n,p)' unless both parameter values are stated.					
Commonly Observed Responses:					
	(b)		$\bullet^3$ correct approximation $\bullet^4$ correct continuity correction $\bullet^5$ calculate z-value $\bullet^6$ calculate probability	$\bullet^3$ $X \approx \text{N}(42.24,\ 5.0688)$ $\bullet^4$ $\text{P}\left(Z > \dfrac{36.5 - 42.24}{\sqrt{5.0688}}\right)$ $\bullet^5$ $\text{P}(Z > -2.55)$ $\bullet^6$ 0.9946	**4**
Notes: 1. For $\bullet^3$, do not penalise omission of double tilde. 2. Mark $\bullet^3$ can be implied by $\bullet^4$.					
Commonly Observed Responses: **Candidate A** calculates exact probability, without using a normal approximation. Marks $\bullet^3$ $\bullet^4$ $\bullet^5$ not available Mark $\bullet^6$ available for probability value of 0.9903 **Candidate B** $\text{P}(X > 36) = \text{P}\left(Z > \dfrac{35.5 - 42.24}{\sqrt{5.0688}}\right) = \text{P}(Z > -2.99) = 0.9986$ award 3/4 ✓×✓$_1$✓$_1$ **Candidate C** $\text{P}(X > 36) = \text{P}\left(Z > \dfrac{36 - 42.24}{\sqrt{5.0688}}\right) = \text{P}(Z > -2.77) = 0.9972$ award 3/4 ✓×✓$_1$✓$_1$ **Candidate D** $\text{P}(X \geq 36) = \text{P}\left(Z > \dfrac{36.5 - 42.24}{\sqrt{5.0688}}\right) = \text{P}(Z > -2.55) = 0.9946$ award 3/4 ✓×✓$_1$✓$_1$					

Question			Generic scheme	Illustrative scheme	Max mark
4.			$\bullet^1$ state hypotheses	$\bullet^1$ H_0: The grades fit a uniform distribution U(5) H_1: The grades do not fit a uniform distribution U(5)	**6**
			$\bullet^2$ correct expected frequency	$\bullet^2$ 174	
			$\bullet^3$ calculate test statistic	$\bullet^3$ $X^2 = 6.5977$	
			$\bullet^4$ correct critical value	$\bullet^4$ $X^2_{4,0.90} = 7.779$	
			$\bullet^5$ deal with H_0	$\bullet^5$ 6.5977 < 7.779 so we do not reject H_0 at the 10% level of significance	
			$\bullet^6$ appropriate conclusion	$\bullet^6$ and conclude that there is no evidence to suggest that the grade frequencies do not fit a U(5) uniform distribution	

Notes:

1. For $\bullet^1$, do not penalise omission of U(5), but mention of 'uniform' is required.
2. For $\bullet^1$, the context of 'grades' must be included.
3. For $\bullet^4$, also accept p-value = 0.1587.
4. For $\bullet^5$, do not accept 'accept H_0'
5. For $\bullet^6$, context of grade frequencies must be included.
6. For $\bullet^6$, do not accept conclusions that are too definite. Phrasing must include 'evidence to conclude…', or 'evidence to suggest…', or similar.

Commonly Observed Responses:

Question			Generic scheme	Illustrative scheme	Max mark
5.	(a)		$\bullet^1$ appropriate comment $\bullet^2$ appropriate comment	$\bullet^1$ non-linear and … $\bullet^2$ …negatively associated	**2**

Notes:

1. For $\bullet^2$, also accept:
   'inversely proportional'
   'inverse relationship'
   'as attempts increase, errors decrease'
2. For $\bullet^2$, do not accept:
   'negatively correlated' (as it is a non-linear relationship)
   'negative exponential'
   'negative logarithmic'
   'decreasing curve'

Commonly Observed Responses:

	(b)	(i)	$\bullet^3$ calculate 'S_{xy}' $= S_{x\frac{1}{y}}$ $\bullet^4$ calculate b $\bullet^5$ calculate a $\bullet^6$ state equation	$\bullet^3$ $S_{x\frac{1}{y}} = 34.03 - \frac{55 \times 3.45}{9}$ $= 12.9467$ $\bullet^4$ $b = \frac{12.9467}{158.89} = 0.0815$ $\bullet^5$ $a = \frac{3.45}{9} - 0.0815 \times \frac{55}{9}$ $= -0.115$ $\bullet^6$ $\frac{1}{y} = -0.115 + 0.0815x$	**4**

Notes:

1. For $\bullet^3$, $\bullet^4$ and $\bullet^5$, do not penalise the notational use of y instead of $\frac{1}{y}$.
2. For $\bullet^3$, $\bullet^4$, $\bullet^5$ and $\bullet^6$, values must be given to at least 2 decimal places.

Commonly Observed Responses:

Question			Generic scheme	Illustrative scheme	Max mark
5.	(b)	(ii)			**5**
			•7 correct strategy	•7 $\frac{1}{\hat{Y}_i} \pm t_{7,0.975}\, s\sqrt{1+\frac{1}{n}+\frac{(x-\bar{x})^2}{S_{xx}}}$	
			•8 calculate fitted value	•8 −0.115 + 0.0815 × 7 = 0.4555	
			•9 appropriate substitution	•9 $0.4555 \pm 2.365 \times 0.078\sqrt{1+\frac{1}{9}+\frac{\left(7-\frac{55}{9}\right)^2}{158.89}}$	
			•10 calculate limits for $\frac{1}{y}$	•10 (0.2606, 0.6504)	
			•11 calculate limits for y	•11 (1.538, 3.837)	

Notes:

1. For •7, do not penalise the notational use of $\hat{Y}$ instead of $\frac{1}{\hat{Y}}$
2. For •8, fitted value must be given to at least 2 decimal places.
3. For •10, limits must match illustrative scheme values, rounded to 2 decimal places
4. For •10, also accept:
   (0.2609, 0.6506) if calculated exactly from the summary values.
   (0.2620, 0.6495) if calculated exactly from the raw data.
5. For •11, also accept:
   (1.5370, 3.832) if calculated exactly from the summary values.
   (1.5396, 3.8162) if calculated exactly from the raw data.
   '2 or 3 errors'

Commonly Observed Responses:

Candidate A calculates confidence interval, instead of a prediction interval.
Mark •7 not available.
Mark •8 available, as above.
Mark •9 available for $0.4555 \pm 2.365 \times 0.078\sqrt{\frac{1}{9}+\frac{\left(7-\frac{55}{9}\right)^2}{158.89}}$
Mark •10 available for (0.3926, 0.5184).
Mark •11 available for (1.929, 2.547).

Question			Generic scheme	Illustrative scheme	Max mark
6.	(a)		$\bullet^1$ appropriate assumption	$\bullet^1$ each cracker's success at working properly is independent of all the other crackers	**5**
			$\bullet^2$ appropriate assumption	$\bullet^2$ the probability of success (ie working properly) of the sampled crackers is constant	
			$\bullet^3$ appropriate strategy	$\bullet^3$ $\hat{p} \pm z_{0.995}\sqrt{\frac{\hat{p}\hat{q}}{n}}$	
			$\bullet^4$ correct substitution	$\bullet^4$ $\frac{14}{20} \pm 2.58\sqrt{\frac{\frac{14}{20}\cdot\frac{6}{20}}{20}}$	
			$\bullet^5$ calculate interval	$\bullet^5$ (0.436,0.964)	

Notes:

1. For $\bullet^1$ and $\bullet^2$, context must refer to the event 'cracker successfully works'
2. For $\bullet^3$, do not penalise omission of hats.
3. Mark $\bullet^3$ can be implied from $\bullet^4$, if done correctly.
4. For $\bullet^4$, if not done correctly, look to $\bullet^3$ to judge if the substitution is correct for the stated subscript.
5. For $\bullet^5$, interval values must be given to at least 2 decimal places.

Question			Generic scheme	Illustrative scheme	Max mark
6.	**(a)**		**(continued)**		

Commonly Observed Responses:

Candidate A provides more than two assumptions for marks $\bullet^1$ and $\bullet^2$.
Treat each comment as using the same strategy, mark each possible combination of pairs of comments and then award the lowest mark from these combinations:

Number of comments	Number correct	Number incorrect	Marks Awarded
3	0	3	0
3	1	2	0
3	2	1	1
3	3	0	2
4	0	4	0
4	1	3	0
4	2	2	0
4	3	1	1
4	4	0	2

Candidate B $\frac{14}{20} \pm 2.58\sqrt{\frac{0.75\times 0.25}{20}} = (0.4502, 0.9498)$

Mark $\bullet^3$ available.
Mark $\bullet^4$ not available.
Mark $\bullet^5$ available, as a follow through error.

Candidate C uses a t distribution, instead of a z distribution

Mark $\bullet^3$ not available.

Mark $\bullet^4$ is available for $\frac{14}{20} \pm 2.861\sqrt{\frac{\frac{14}{20}\cdot\frac{6}{20}}{20}}$ where $t_{19,0.995} = 2.861$

Mark $\bullet^5$ is available for $(0.4068, 0.9932)$

	(b)		$\bullet^6$ appropriate comment $\bullet^7$ appropriate decision	$\bullet^6$ the CI contains 75% $\bullet^7$ so the belief is supported	**2**

Notes:

1. For $\bullet^7$, do not accept decisions involving a probabilistic interpretation, such as '99% chance of being correct'.
2. For $\bullet^7$, if the candidate claims the belief is supported and then proceeds to contradict their decision, then do not award $\bullet^7$.

Commonly Observed Responses:

Question			Generic scheme	Illustrative scheme	Max mark
7.			$\bullet^1$ correct E(X) $\bullet^2$ correct V(X) $\bullet^3$ calculate E(Y) $\bullet^4$ calculate V(Y) $\bullet^5$ calculate SD(Y)	$\bullet^1$ $\frac{9}{2}=4.5$ $\bullet^2$ $\frac{63}{12}=5.25$ $\bullet^3$ $\begin{cases}=3\mathrm{E}(X)-2\\=3\times4.5-2\\=11.5\end{cases}$ $\bullet^4$ $\begin{cases}\mathrm{V}(3X)\\=3^2\times\frac{63}{12}\\=47.25\end{cases}$ $\bullet^5$ $\mathrm{SD}(Y)=6.87$	**5**
Notes:					
Commonly Observed Responses:					

Question	Generic scheme	Illustrative scheme	Max mark
8.	•[1] appropriate hypotheses	•[1] $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 > \mu_2$	**9**
	•[2] correct test statistic	•[2] $t_{n_1+n_2-2} = \dfrac{\overline{X}_1 - \overline{X}_2}{s\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$	
	•[3] correct s^2	•[3] $s^2 = 12.0429$	
	•[4] calculate t	•[4] t = 1.154	
	•[5] state critical value	•[5] 5% cv is $t_{18,0.95} = 1.734$	
	•[6] deal with H_0	•[6] $1.154 < 1.734$ so we cannot reject H_0 at the 5% significance level	
	•[7] appropriate conclusion	•[7] and conclude that there is insufficient evidence to suggest that there is a greater (population) mean height in Area 1	
	•[8] appropriate assumption	•[8] the test assumes that plant heights are distributed normally…	
	•[9] appropriate assumption	•[9] with equal variances for the two areas	

Notes:

1. For •[1], do not accept H_0: mean$_1$ = mean$_2$, etc, as must refer to population mean.
2. For •[5], also accept p-value = 0.1318.
3. For •[5], also accept other significance levels that are clearly stated, with appropriate follow through
4. For •[6], do not accept 'accept H_0'
5. For •[7], conclusion must be phrased in terms of H_1
6. For •[7], do not accept conclusions that are too definite. Phrasing must include 'evidence to conclude…', or 'evidence to suggest…', or similar.
7. For •[7], •[8] and •[9], context must clearly refer to 'plant height', or 'plant growth'.
8. For •[8] or •[9], also accept 'plant heights are independent'

Commonly Observed Responses:

Candidate A uses a z-test, instead of a t-test

Marks •[2] and •[3] not available.

Mark •[4] available for z = 1.157

Mark •[5] available for $z_{0.95} = 1.64$ or a p-value of 0.1236

Marks •[6] and •[7] available for consistent 'deal with H_0' and 'appropriate conclusion'

Marks •[8] and •[9] available for assumptions about normality and independence only. Do not accept the assumption of 'equal variances for the two areas'

Question			Generic scheme	Illustrative scheme	Max mark
9.	(a)		$\bullet^1$ identifies distribution parameter $\bullet^2$ appropriate assumption	$\bullet^1$ $\lambda = 2.5$ $\bullet^2$ component failures are independent of each other	**2**

Notes:

1. For $\bullet^1$, also accept $X \sim \text{Po}(2.5)$.
2. For $\bullet^2$, also accept:
   'component failure occurs at a constant **mean** rate'
   'component failure occur at random points in time'
   'component failures cannot occur simultaneously'
3. For $\bullet^2$, do not accept '… constant average rate', as 'average' does not specify the mean.

Commonly Observed Responses:

Candidate A provides more than one assumption for mark $\bullet^2$
If 2 assumptions given, at least one must be valid and correct to gain $\bullet^2$
If 3 or 4 assumptions given, at least two must be valid and correct to gain $\bullet^2$

	(b)		$\bullet^3$ appropriate probability $\bullet^4$ calculate probability	$\bullet^3$ P(X = 0) = $\bullet^4$ 0.0821	**2**

Notes:

1. Mark $\bullet^3$ can be implied from mark $\bullet^4$.
2. For $\bullet^4$, must give probability to at least 3 decimal places.

Commonly Observed Responses:

Question			Generic scheme	Illustrative scheme	Max mark
9.	(c)		•^{5} correct strategy •^{6} appropriate conclusion	•^{5} $P(X \leq n) \geq 0.90$ •^{6} 5 components are required to ensure a 90% chance of completing the rally	**2**

Notes:

1. For •^{5}, also accept:
   $P(X \leq n) > 0.90$
   $P(X > n) < 0.10$
   $P(X > n) \leq 0.10$
   strategy involving 'trial and improvement'

Commonly Observed Responses:

Candidate A attempts to use a normal approximation
Mark •^{5} not available, as $\lambda < 10$, so the approximation is not appropriate.
Mark •^{6} only available if the resulting calculated value for n is substituted back into a Poisson distribution calculation to check that it meets the criteria.

Candidate B does not use cumulative probabilities, so gains no marks.
$X \sim \text{Po}(2.5)$
$P(X = 4) = 0.1336$
$P(X = 5) = 0.0668$

Question	Generic scheme	Illustrative scheme	Max mark
10.	$\bullet^1$ appropriate strategy	$\bullet^1$ $\bar{x} \pm z_{0.95} \frac{\sigma}{\sqrt{n}}$	**5**
	$\bullet^2$ correct z	$\bullet^2$ $\bar{x} \pm 1.64 \frac{\sigma}{\sqrt{n}}$	
	$\bullet^3$ appropriate strategy	$\bullet^3$ $\frac{2 \times 1.64 \times 2.9}{\sqrt{n}} < 1.4$	
	$\bullet^4$ solve for n	$\bullet^4$ $n > 46.1623$	
	$\bullet^5$ appropriate value of n	$\bullet^5$ 47 should be sampled	

Notes:

1. For $\bullet^1$, also accept $z_{0.95} \frac{\sigma}{\sqrt{n}}$.
2. For $\bullet^3$, also accept $\frac{1.64 \times 2.9}{\sqrt{n}} < 0.7$
3. For $\bullet^3$, also accept statement that uses equality

Commonly Observed Responses:

Candidate A does not half the interval width

Mark $\bullet^3$ not available for $\frac{1.64 \times 2.9}{\sqrt{n}} < 1.4$

Mark $\bullet^4$ available for $n > 11.54$ as a follow through error.
Mark $\bullet^5$ available for $n = 12$ as a follow through error.

Question			Generic scheme	Illustrative scheme	Max mark
11.	(a)	(i)	$\bullet^1$ correct variables	$\bullet^1$ age and pulse rate	**1**
Notes: 1. For $\bullet^1$, candidates must only state these two variables to gain the mark.					
Commonly Observed Responses:					
		(ii)	$\bullet^2$ correct assumption	$\bullet^2$ (heights of people in) each group would have a same shape and variability	**1**
Notes:					
Commonly Observed Responses:					

Question			Generic scheme	Illustrative scheme	Max mark
11.	(b)		$\bullet^3$ correct data type $\bullet^4$ correct variables	$\bullet^3$ categorical data $\bullet^4$ activity level and smoker	**2**

Notes:

1. For $\bullet^3$, also accept 'qualitative data'
2. For $\bullet^4$, candidates must state both the variables to gain the mark (see below)

Commonly Observed Responses:

Candidate A response of 'categorical data' or 'qualitative data'
Mark $\bullet^3$ is awarded
Mark $\bullet^4$ available for at least two from age, pulse rate, activity level and smoker, with no other variables mentioned.
(The inclusion of age and pulse rate as additional acceptable responses to those given in the illustrative scheme acknowledges, in this context, that these variables might be used to form a contingency table)

Candidate B response of 'numerical data' or 'quantitive data'
Mark $\bullet^3$ not available
Mark $\bullet^4$ available for at least two from age, height, mass, BMI and pulse rate, with no other variables mentioned.

Candidate C response of 'continuous data'
Mark $\bullet^3$ not available
Mark $\bullet^4$ available for at least two from age, height, mass, BMI, with no other variables mentioned.

Candidate D response of 'discrete data'
Mark $\bullet^3$ not available
Mark $\bullet^4$ available for age and pulse rate, with no other variables mentioned.
or
Mark $\bullet^4$ available for writing the same (incorrect) response as already given in part (a)(i).

	(c)		$\bullet^5$ appropriate statistic	$\bullet^5$ r, the (sample) product moment correlation coefficient	**1**

Notes:

1. For $\bullet^5$, also accept:
   'correlation coefficient'
   'pmcc'
   'r'
2. For $\bullet^5$, do not accept:
   'coefficient of determination', as it refers to a model, not a relationship.
   'rho' or 'ρ', as it is a parameter, and not a statistic.

Commonly Observed Responses:

Question			Generic scheme	Illustrative scheme	Max mark
12.	(a)		$\bullet^1$ appropriate assumption	$\bullet^1$ Assuming that the masses of all jars and masses of all honey are all independent of each other	**6**
			$\bullet^2$ combine random variables	$\bullet^2$ $T = (J_1 + J_2 + \ldots + J_{48}) + (H_1 + H_2 + \ldots + H_{48})$	
			$\bullet^3$ calculate μ	$\bullet^3$ 25056	
			$\bullet^4$ calculate σ^2	$\bullet^4$ 1056	
			$\bullet^5$ appropriate strategy	$\bullet^5$ $\begin{cases} \mathrm{P}(T > 25000) \\ = \mathrm{P}\left(Z > \dfrac{25000 - 25056}{\sqrt{1056}}\right) \\ = \mathrm{P}(Z > -1.723) \end{cases}$	
			$\bullet^6$ calculate probability	$\bullet^6$ 0.9573	

Notes:

1. For $\bullet^1$ and $\bullet^2$, there should be evidence that 96 distinct random variables are being combined and they all need to be independent from each other (to support the valid use of the law of variance for the addition of several random variables)
2. For $\bullet^1$, do not accept assumptions that refer to just honey, or just jars.
3. For $\bullet^2$, do not accept $T = 48J + 48H$
4. Mark $\bullet^2$ is only available when the strategy used has been clearly communicated.
5. For $\bullet^6$, also accept 0.9576 (obtained by exact calculation).

Commonly Observed Responses:

Candidate A $\mathrm{V}(48J + 48H) = 48\mathrm{V}(J) + 48\mathrm{V}(H) = 1056$

Mark $\bullet^2$ not available
Mark $\bullet^4$ not available

Candidate B $\mathrm{V}(48J + 48H) = 48^2\mathrm{V}(J) + 48^2\mathrm{V}(H) = 50688$

Mark $\bullet^1$ available for an assumption that refers to the mass of a single jar being independent from a single honey (as that is consistent with their workings)
Mark $\bullet^2$ not available
Mark $\bullet^4$ available for 50688
Mark $\bullet^5$ available for $\mathrm{P}\left(Z > \dfrac{25000 - 25056}{\sqrt{50688}}\right) = \mathrm{P}(Z > -0.25)$
Mark $\bullet^6$ available for 0.5987

Question			Generic scheme	Illustrative scheme	Max mark
12.	(b)		•[7] state hypotheses	•[7] $H_0 : \mu = 522$ $H_1 : \mu \neq 522$	**6**
			•[8] calculate z value	•[8] $z = \dfrac{527.5 - 522}{5/\sqrt{10}} = 3.47851$	
			•[9] correct critical value	•[9] $z_{0.995} = 2.58$	
			•[10] deal with H_0	•[10] 3.48 >2.58 so we reject H_0 at the 1% level of significance	
			•[11] appropriate conclusion	•[11] conclude that there is evidence to suggest that the (population) mean mass does not equal 522g	
			•[12] appropriate assumption	•[12] the standard deviation is unchanged	

Notes:

1. For •[9] , also accept p-value = 2 × 0.000252 = 0.000504.
2. For •[10], do not accept 'accept H_1'
3. For •[11], conclusion must be phrased in terms of H_1
4. For •[11], do not accept conclusions that are too definite. Phrasing must include 'evidence to conclude…', or 'evidence to suggest…', or similar.

Commonly Observed Responses:

Candidate A uses a t-test, instead of a z-test

Mark •[7] available.

Mark •[8] not available.

Mark •[9] available for $t_{9,0.95} = 3.250$ or p-value = 2 × 0.0003477 = 0.0006955.

Marks •[10] and •[11] are available for consistent 'deal with H_0' and 'appropriate conclusion'

Marks •[12] not available

Question			Generic scheme	Illustrative scheme	Max mark
13.	(a)		$\bullet^1$ correct substitution $\bullet^2$ calculate limits	$\bullet^1$ 2σ limits are $102 \pm 2 \times \frac{0.13}{\sqrt{5}}$ $\bullet^2$ 101.88, 102.12	**2**
Notes: 1. For $\bullet^2$, treat (101.88, 102.12) as bad form, with no penalty					
Commonly Observed Responses: **Candidate A** $102 \pm 2 \times 0.13$ =(101.74, 102.26) award 1/2 $\times\checkmark_1$ **Candidate B** $102 \pm 2 \times \frac{0.13}{5}$ =(101.948, 102.052) award 1/2 $\times\checkmark_1$					
	(b)		$\bullet^3$ strategy to ensure it is in control $\bullet^4$ appropriate bound $\bullet^5$ appropriate bound	$\bullet^3$ $\ldots \leq \overline{x} \leq \ldots$ $\bullet^4$ 101.88 $\bullet^5$ 102.17	**3**
Notes: 1. For $\bullet^3$, also accept: $\ldots < \overline{x} < \ldots$ (…, …) 2. For $\bullet^3$, also accept an appropriate sentence, such as 'more than … and less than …' 'between … and …' 'above … and below …' 3. Treat (102.17, 101.88) as bad form, with no penalty.					
Commonly Observed Responses:					

Question			Generic scheme	Illustrative scheme	Max mark
14.	(a)	(i)	•[1] appropriate statement	•[1] distribution is (approximately) normal	**1**
Notes: 1. For •[1], the use of statistical notation gains 0 marks.					
Commonly Observed Responses:					
		(ii)	•[2] appropriate statement •[3] appropriate statement	•[2] distribution's mean is equal to the **population mean** •[3] distribution's variance is equal to the **population variance** divided by the sample size.	**2**
Notes: 1. For •[2] and •[3], the use of statistical notation gains 0 marks. 2. For •[2], only accept reference to population mean, and do not accept sample mean. 3. For •[3], only accept reference to population variance, and do not accept sample variance. 4. For •[3], also accept reference to population standard deviation divided by the square root of the sample size					
Commonly Observed Responses:					
	(b)	(i)	•[4] appropriate reason	•[4] the distribution of weights is already known to be normally distributed	**1**
Notes: 1. For •[4], also accept responses referring to how the CLT relates to the distribution and parameters of the sample mean, whilst the study seeks parameters of the parent distribution.					
Commonly Observed Responses:					
		(ii)	•[5] appropriate reason	•[5] non-random (convenience) sampling has been used	**1**
Notes: 1. For •[5], also accept responses referring to whether birth weights are all independent by giving examples of multiple births, such as twins. 2. For •[5], also accept responses referring to other appropriate geographical or cluster effects, such as: high altitude regions poor nutritional health low levels of antenatal care babies may have been born premature					
Commonly Observed Responses:					

[END OF MARKING INSTRUCTIONS]