National
Qualifications
SPECIMEN ONLY

SQ26/AH/01 Statistics

Date — Not applicable

Duration — 3 hours

Total marks — 100

Attempt ALL questions.

You may use a calculator.

Full credit will be given only to solutions which contain appropriate working.

State the units for your answer where appropriate.

Write your answers clearly in the answer booklet provided. In the answer booklet you must clearly identify the question number you are attempting.

Use **blue** or **black** ink.

Before leaving the examination room you must give your answer booklet to the Invigilator; if you do not, you may lose all the marks for this paper.

A booklet of Statistical Formulae and Tables is supplied for all candidates.

MARKS

Total marks — 100

Attempt ALL questions

1. The waiting times, to the nearest minute, for a sample of patients attending a hospital clinic were recorded as follows:

2 2 3 6 7 7 9 10 13 14 29

State the values of the median and quartiles.

Hence determine the upper fence for this data.

Comment on the waiting time of 29 minutes. **3**

2. A competition is run at a local bridge club where, each week, players compete in pairs against other pairs. Scores are given as a percentage for each pair with the highest percentage being the winning score.

A random sample of nine scores obtained by a particular pair, during the first few months of this year's competition, yielded the following percentages:

53·7 58·9 56·4 57·1 62·0 55·3 59·5 61·1 57·6

with $\sum x = 521{\cdot}6$ and $\sum x^2 = 30287{\cdot}58$

(a) Stating one assumption that needs to be made, calculate a 95% confidence interval for the population mean percentage score for this pair. **6**

(b) The mean percentage score for this pair in last year's competition was 55·4.

State, with a reason, whether or not you would agree with the opinion that their current scores yield evidence of an improvement on last year. **2**

3. In an examination, the probabilities of three candidates, Ahmed, Ben and Christine solving a certain problem are $\frac{5}{6}$, $\frac{3}{4}$ and $\frac{4}{5}$ respectively.

(a) Stating any assumption made, calculate the probability that the examiner will receive at least two correct solutions for this problem from these candidates. **4**

(b) Given that the examiner receives exactly one correct solution, determine the probability that this solution is provided by Christine. **5**

MARKS

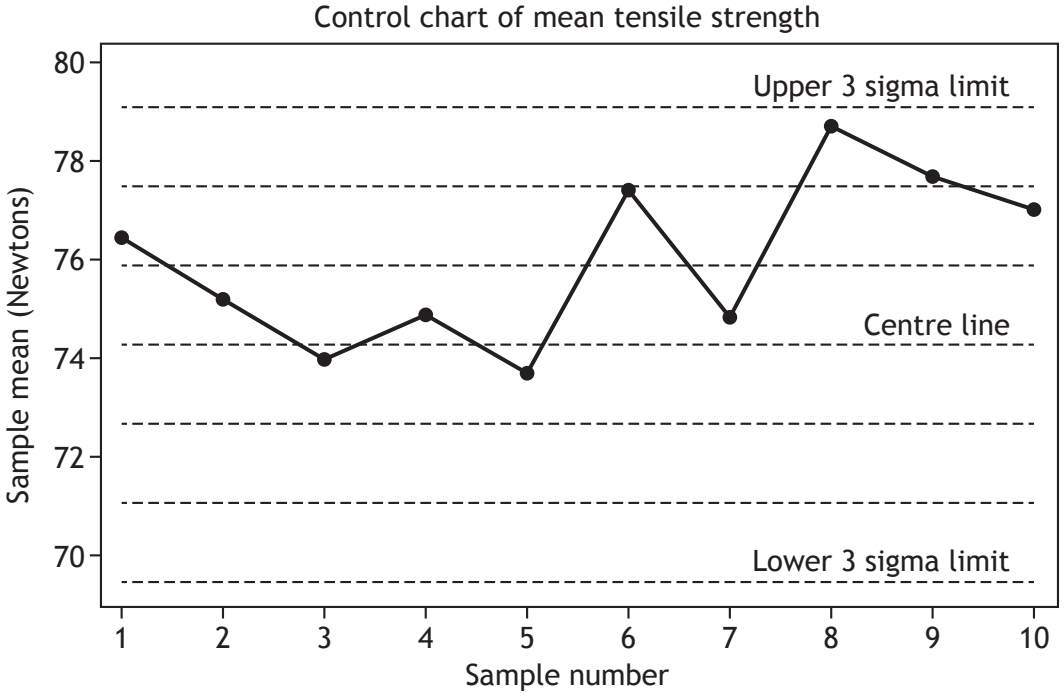
4. Widgets produced in Work Cell A are known to have tensile strengths with mean 74·3 N and standard deviation 3·2 N when the manufacturing process is operating in a stable manner. In an attempt to increase tensile strength, modifications were made to the process used in the Cell. In order to monitor the modified process, random samples of four widgets were taken from the widgets produced each hour and their tensile strengths determined.

(a) Calculate three sigma limits for a control chart of the mean tensile strengths of the samples using the given historical parameters. **2**

The first ten consecutive hourly samples yielded the following means:

76·5 75·2 74·0 74·8 73·7 77·4 74·8 78·7 77·7 77·0

These are plotted on the control chart below with both one and two sigma limits shown in addition to the centre line and three sigma limits.



(b) Identify the feature of this control chart which indicates that the process is out of statistical control. State, giving a reason, whether or not you think the team on Work Cell A would be pleased with this lack of statistical control. **2**

(c) As a result of this analysis, suggest what Work Cell A could do to continue to monitor the tensile strength of the widgets produced. **2**

MARKS

5. A random sample of eight sixth-year pupils was taken. Each of the pupils threw three darts at a dartboard with the objective of obtaining as high a total score as possible.

They then had a few days to practise and the experiment was repeated to see if their scores had improved.

Pupil	A	B	C	D	E	F	G	H
Score before	63	40	31	40	18	40	89	38
Score after	62	45	23	58	38	60	80	43

(a) Test the claim that practice improves the performance of sixth-year pupils, using a Wilcoxon test. **6**

(b) Using a normal approximation, calculate an appropriate p-value for this test. Comment on the use of the normal approximation for this experiment. **4**

6. An unbiased regular octahedral (eight-faced) die is thrown. Let the random variable X be the score shown.

(a) Calculate $\mathrm{E}(X)$ and $\mathrm{V}(X)$. **2**

An unbiased cubical die is also thrown. Let the random variable Y be the score shown on this die.

(b) Calculate $\mathrm{E}(3X - 2Y)$ and $\mathrm{V}(3X - 2Y)$. **4**

7. In order to meet current acoustic standards in the UK for separating floors between flats, the impact sound transmission (IST) should not exceed 62 dB. The IST results (dB) for a certain type of flooring can be adequately modelled by the N(53·4, $4{\cdot}7^2$) distribution.

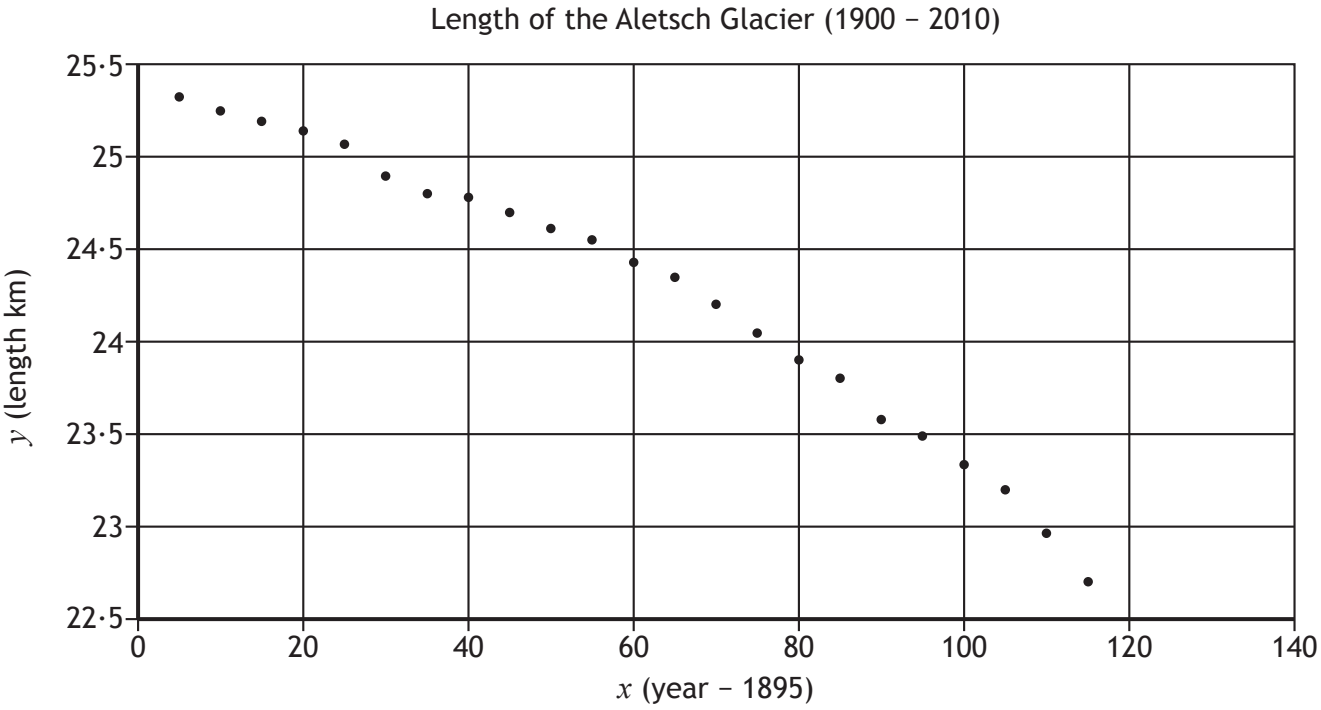
(a) Calculate the proportion of such floors that would fail to meet the 62 dB standard. **3**

(b) Stating any assumption required, calculate the value to which the mean IST for this type of floor would have to be reduced in order to ensure that at most 1% of floors would fail to meet the 62 dB standard. **4**

MARKS

8. The Swiss Glacier Monitoring Network has records of the length y (km) of the Aletsch Glacier from 1900 until the present day.

The x and y data is shown on a scatterplot (where x = year − 1895).



A straight line can be used to model this data giving a regression equation of $y = -0{\cdot}0228x + 25{\cdot}653$ with a coefficient of determination $R^2 = 0{\cdot}971$.

(a) (i) Interpret the slope parameter in terms of the length of the glacier. **1**

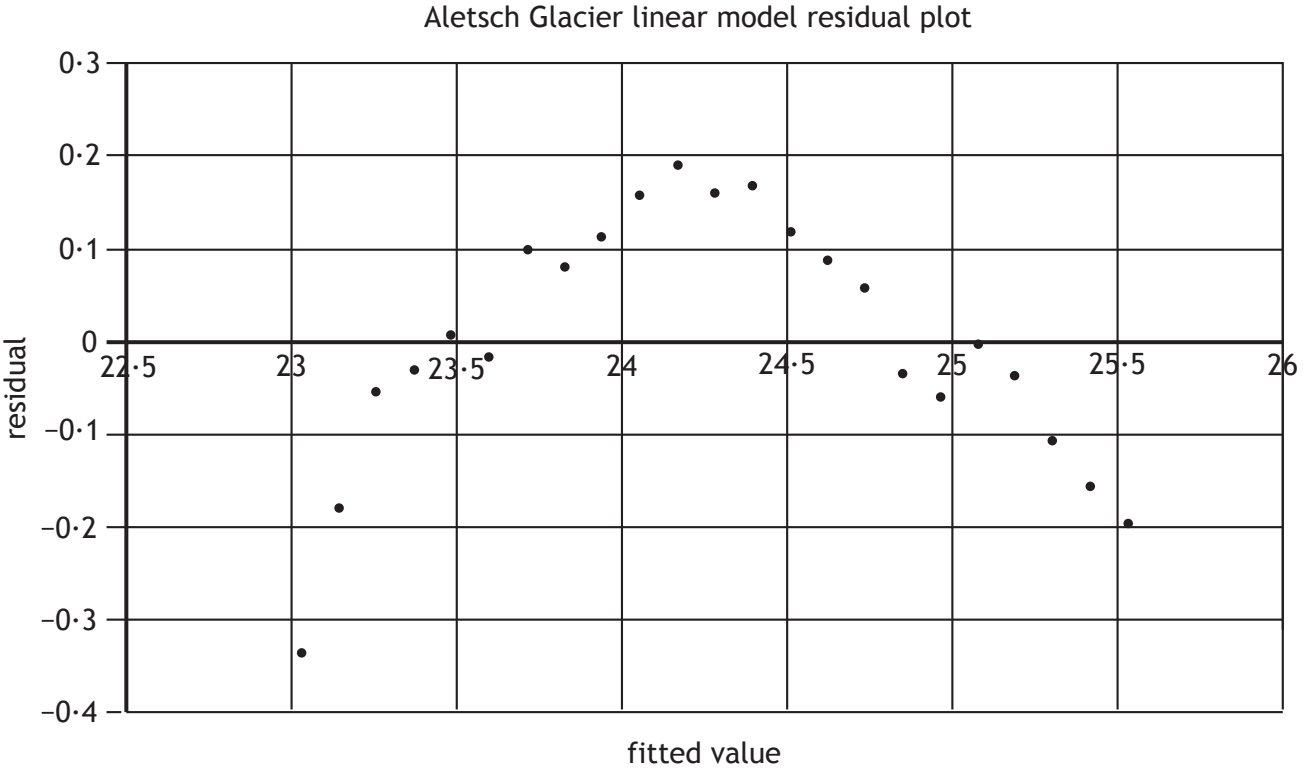
(ii) Determine the product moment correlation coefficient and comment on what it measures. **2**

The following results are obtained:

$n = 23 \quad \bar{x} = 60 \quad S_{xx} = 25300 \quad SSR = 0{\cdot}3923$

8. (continued)

Using the linear model the following residual plot is obtained:



(b) (i) Using the linear model, calculate a 90% prediction interval for the length of the glacier for the year 2030. **4**

(ii) Interpret this interval in the context of the question. **1**

(iii) Give **two** reasons why this prediction interval may be unreliable. **2**

9. Researchers investigated whether or not subjects could distinguish between two brands of a drink by using the triangle test.

The test involves giving subjects three blind samples, chosen at random, two of which contain the same drink and one of which contains a different drink. Subjects taste all three samples and are asked to pick out the one that is different.

(a) State the probability of identifying the different drink by chance alone. **1**

(b) In a group of eight subjects, calculate the probability that exactly five correctly identify the different drink by chance. State an assumption that you have made in calculating your answer. **3**

In a larger group of 42 subjects, 19 identify the different drink.

(c) Using an approximate distribution, perform a hypothesis test at the 5% level to determine whether or not this provides evidence that members of this group are more likely to identify the different drink than by chance. **7**

MARKS

10. In order to find out if there is any evidence of gender difference in performance in a national examination, random samples of marks are taken.

Summary statistics are as follows:

Boys: $n_b = 50,\ \overline{x}_b = 78,\ s_b = 7$

Girls: $n_g = 40,\ \overline{x}_g = 74,\ s_g = 8$

(a) Choose an appropriate test for samples of this size and state a condition under which it may be applied. **2**

(b) (i) Perform the test and comment on any differences in conclusion between using a 5% and a 1% level of significance. **7**

(ii) If the original hypothesis had been that the boys performed better than the girls, what conclusion would have been reached at the 1% level? **2**

(c) Give **two** recommendations about the proper use of hypothesis tests suggested by the results in part (b). **2**

11. In an investigation of whether taste preference changes with age, it was found that 56% of a random sample of 300 people under 40 years of age preferred the new brand of a low fat spread to the old one, whereas 48% of a random sample of 200 people over 40 showed the same preference.

Test the hypothesis that there is a difference in preference between the age groups. **7**

12. The random variables F and G are the ages of the trees in two independent sustainable forests:

Population F has mean of 10 and standard deviation of 2.
Population G has mean of 8 and standard deviation of 1.

A random sample of size 100 is drawn from each forest and the sample means denoted by $\overline{F}$ and $\overline{G}$.

(a) Calculate $\mathrm{E}\left(\overline{F} - \overline{G}\right)$ and $\mathrm{SD}\left(\overline{F} - \overline{G}\right)$.

Hence, find the approximate probability that the absolute difference between the two sample means is less than 1·5. **7**

(b) Explain why the probability above is only approximate. Justify the use of this approximation in the context of this question. **3**

[END OF SPECIMEN QUESTION PAPER]

National
Qualifications
SPECIMEN ONLY

SQ26/AH/01 Statistics

Marking Instructions

These Marking Instructions have been provided to show how SQA would mark this Specimen Question Paper.

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.

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

(d) 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 is easier, candidates lose the opportunity to gain credit.

(e) Where transcription errors occur, candidates would normally lose the opportunity to gain a processing mark.

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

(g) Unless specifically mentioned in the Detailed Marking Instructions, do not penalise:

- working subsequent to a correct answer
- correct working in the wrong part of a question
- legitimate variations in solutions
- repeated errors within a question

Definitions of Mathematics-specific command words used in this Specimen Question Paper

Comment (on): use statistics to describe and/or explain a statement or result.

Determine: determine an answer from given facts, figures, or information.

Find: obtain an answer showing relevant stages of working.

Hence: use the previous answer to proceed.

Hence, or otherwise: use the previous answer to proceed; however, another method may alternatively be used.

Perform: carry out, conduct.

Test: use a statistical test to assess whether or not there is evidence in favour of a given claim or statement.

Detailed Marking Instructions for each question

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
1			Ans: it is an outlier	**3**	
			•[1] correct values •[2] calculate upper fence •[3] appropriate comment		•[1] 3, 7 and 13 •[2] upper fence = 13 + 1·5 × 10 = 28 •[3] 29 is an outlier
Notes:					
2	**a**		Ans: (55·89, 60·03)	**6**	
			•[1] correct strategy •[2] correct mean •[3] correct standard deviation •[4] correct *t* •[5] calculate interval •[6] appropriate assumption		•[1] an approximate 95% CI is given by $\bar{x} \pm t_{n-1,0\cdot975}\frac{s}{\sqrt{n}}$ •[2] $\bar{x} = 57\cdot96$ •[3] $s = 2\cdot692$ •[4] $t_{8,0\cdot975} = 2\cdot306$ •[5] 95% CI is $(55\cdot89, 60\cdot03)$ •[6] scores are normally distributed
2	**b**		Ans: evidence of an improvement in scores	**2**	
			•[7] appropriate reason •[8] appropriate comment		•[7] the whole CI is above 55·4% and so •[8] there is evidence of an improvement in scores
Notes:					
3	**a**		Ans: $\frac{107}{120}$	**4**	
			•[1] appropriate assumption •[2] correct strategy		•[1] the solutions provided by the candidates in the examination are independent •[2] P(≥ 2 correct solutions) $= \begin{matrix} \mathrm{P}(\mathrm{A}\cap\mathrm{B}\cap\mathrm{C}) + \mathrm{P}(\mathrm{A}\cap\mathrm{B}\cap\bar{\mathrm{C}}) \\ +\mathrm{P}(\mathrm{A}\cap\bar{\mathrm{B}}\cap\mathrm{C}) + \mathrm{P}(\bar{\mathrm{A}}\cap\mathrm{B}\cap\mathrm{C}) \end{matrix}$

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
			•[3] appropriate substitution		•[3] $= \frac{5}{6}\times\frac{3}{4}\times\frac{4}{5} + \frac{5}{6}\times\frac{3}{4}\times\frac{1}{5} + \frac{5}{6}\times\frac{1}{4}\times\frac{4}{5} + \frac{1}{6}\times\frac{3}{4}\times\frac{4}{5}$
			•[4] calculate probability		•[4] $= \frac{107}{120}$
3	**b**		Ans: $= \frac{1}{3}$	**5**	
			•[5] correct strategy		•[5] $\mathrm{P}(1 \text{ correct solution})$ $= \mathrm{P}(\mathrm{A}\cap\bar{\mathrm{B}}\cap\bar{\mathrm{C}}) + \mathrm{P}(\bar{\mathrm{A}}\cap\mathrm{B}\cap\bar{\mathrm{C}}) + \mathrm{P}(\bar{\mathrm{A}}\cap\bar{\mathrm{B}}\cap\mathrm{C})$
			•[6] calculate probability		•[6] $= \frac{5}{6}\times\frac{1}{4}\times\frac{1}{5} + \frac{1}{6}\times\frac{3}{4}\times\frac{1}{5} + \frac{1}{6}\times\frac{1}{4}\times\frac{4}{5} = \frac{1}{10}$
			•[7] correct strategy		•[7] $\mathrm{P}(\mathrm{C} \mid 1 \text{ correct solution}) = \frac{\mathrm{P}(\bar{\mathrm{A}}\cap\bar{\mathrm{B}}\cap\mathrm{C})}{\mathrm{P}(1 \text{ correct solution})}$
			•[8] appropriate substitution		•[8] $= \frac{\frac{1}{6}\times\frac{1}{4}\times\frac{3}{4}}{\frac{1}{10}}$
			•[9] calculate probability		•[9] $= \frac{1}{3}$
Note: Other methods, eg tree diagrams and Venn diagrams, are acceptable.					
4	**a**		Ans: (69·5, 79·1), normally distributed	**2**	
			•[1] correct strategy		•[1] three sigma limits are given by $\mu \pm 3\frac{\sigma}{\sqrt{n}}$
			•[2] calculate both limits		•[2] 69·5 and 79·1
4	**b**		Ans: WECO rule and pleased	**2**	
			•[3] appropriate WECO rule		•[3] samples 8 and 9 are above 2-sigma (WECO: 2 out of 3 consecutive points above 2-sigma)

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
			•[4] appropriate comment		•[4] pleased with very high tensile strength
4	**c**		Ans: start a new chart with higher limits	**2**	
			•[5] appropriate idea •[6] appropriate change		•[5] start a new control chart with . . . •[6] the limits set a bit higher
Notes:					
5	**a**		Ans: no evidence of improvement	**6**	
			•[1] appropriate hypotheses •[2] correct ranking •[3] correct statistics •[4] correct critical value •[5] deal with H_0 •[6] appropriate conclusion		•[1] H_0: practice has made no difference H_1: practice has improved performance •[2] Diff 1 −5 −5 8 9 −18 −20 −20 Rank 1 −2·5 −2·5 4 5 −6 −7·5 −7·5 •[3] $n = 8 \quad W_+ = 10 \quad W_- = 26$ •[4] 5% cv is 5 •[5] 10 > 5 so we cannot reject H_0 at the 5% level of significance •[6] and conclude that there is no evidence of an improvement
5	**b**		Ans: 0·1468	**4**	
			•[7] correct parameters •[8] use of continuity correction •[9] calculate p-value •[10] appropriate comment		•[7] $E(V) = 18 \quad V(V) = 51$ •[8] $z = \frac{10 - 18 + 0{\cdot}5}{\sqrt{51}}$ •[9] p = 0·1468 •[10] the sample size is too small for a NA
Notes:					
6	**a**		Ans: $\frac{9}{2}$ and $\frac{21}{4}$	**2**	
			•[1] correct $E(X)$		•[1] $X \sim U(8)$ gives $E(X) = \frac{8+1}{2} = \frac{9}{2}$

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
			•[2] correct $\mathrm{V}(X)$		•[2] $\mathrm{V}(X)=\frac{8^2-1}{12}=\frac{21}{4}$
6	**b**		Ans: $\frac{707}{12}$	**4**	
			•[3] correct $\mathrm{E}(Y)$ and $\mathrm{V}(Y)$ •[4] calculate $\mathrm{E}(3X-2Y)$ •[5] knowing to square and add •[6] calculate $\mathrm{V}(3X-2Y)$		•[3] $\mathrm{E}(Y)=\frac{k+1}{2}=\frac{7}{2}$, $\mathrm{V}(Y)=\frac{k^2-1}{12}=\frac{35}{12}$ •[4] $\mathrm{E}(3X-2Y)=3\times\frac{9}{2}-2\times\frac{7}{2}=\frac{13}{2}$ •[5] $\mathrm{V}(3X-2Y)=3^2\times\frac{21}{4}+2^2\times\frac{35}{12}$ •[6] $=\frac{707}{12}$
Notes:					
7	**a**		Ans: 0·0336	**3**	
			•[1] appropriate strategy •[2] appropriate strategy •[3] calculate probability		•[1] $X\sim N(53{\cdot}4,\ 4{\cdot}7^2)\ \Rightarrow \mathrm{P}(X>62)$ •[2] $\mathrm{P}\left(Z>\frac{62-53{\cdot}4}{4{\cdot}7}\right)$ •[3] $\mathrm{P}(Z>1{\cdot}83)=0{\cdot}0336$
7	**b**		Ans: $\mu=51{\cdot}05$	**4**	
			•[4] appropriate assumption •[5] correct z-value for 1% •[6] correct strategy •[7] calculate μ		•[4] the variability remains unchanged •[5] z = 2·33 •[6] $\frac{Z-\mu}{\sigma}=\frac{62-\mu}{\sigma}=2{\cdot}33$ •[7] $\mu\leq 51{\cdot}04$
Notes:					
8	**a**	i	Ans: interpretation	**1**	
			•[1] appropriate interpretation		•[1] the glacier loses on average 22·8 m in length p.a.
		ii	Ans: −0·985 and comment	**2**	

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
			•[2] correct r •[3] appropriate comment		•[2] $r = -0{\cdot}985$ •[3] which indicates a strong linear association
8	**b**	i	Ans: $(22{\cdot}310, 22{\cdot}840)$	**4**	
			•[4] correct x and $\hat{Y}$ •[5] correct s •[6] correct t •[7] calculate limits		•[4] $x = 2030 - 1895 = 135$ $\hat{Y} = -0{\cdot}0228 \times 135 + 25{\cdot}653 = 22{\cdot}575$ •[5] $s = \sqrt{\frac{0{\cdot}3923}{21}} = 0{\cdot}1367$ •[6] $t_{21,0{\cdot}950} = 1{\cdot}721$ •[7] (22·310, 22·840)
		ii	Ans: interpretation	**1**	
			•[8] appropriate interpretation		•[8] we are 90% sure that in 2030 the length of the glacier will lie in the interval (22·1, 22·84) km
		iii	Ans: extrapolation and quadratic	**2**	
			•[9] appropriate reason •[10] appropriate reason		•[9] 2030 requires extrapolation •[10] the residual plot indicates that the relationship is not linear
Notes:					
9	**a**		Ans: $\frac{1}{3}$	**1**	
			•[1] correct probability		•[1] $\frac{1}{3}$
9	**b**		Ans: 0·068, subjects' answers are independent	**3**	
			•[2] use binomial distribution •[3] calculate probability •[4] appropriate assumption		•[2] $\mathrm{P}(X = 5) = \binom{8}{5}\left(\frac{1}{3}\right)^5\left(\frac{2}{3}\right)^3$ •[3] 0·068 •[4] the answers of subjects are independent of one another
9	**c**		Ans: no significant evidence	**7**	

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
			•[5] correct hypotheses		•[5] $H_0 : p = \frac{1}{3}$ $H_1 : p > \frac{1}{3}$
			•[6] identify approximate normal distribution		•[6] $X \sim B(42, \frac{1}{3}) \approx N(14, \frac{28}{3})$
			•[7] subtract continuity correction		•[7] 18·5
			•[8] calculate z		•[8] $P(X \geq 19) = P(Z \geq \frac{18 \cdot 5 - 14}{\sqrt{\frac{28}{3}}}) = P(Z \geq 1 \cdot 47)$
			•[9] correct critical value		•[9] 5% cv for Z is 1·645
			•[10] deal with H_0		•[10] 1·47 < 1·645 so we cannot reject H_0 at the 5% level of significance
			•[11] context		•[11] and conclude that there is no evidence that this group can identify the different drink
Note: The alternative and acceptable p-value approach would record that $P(Z > 1 \cdot 47) = 0 \cdot 0708 > 0 \cdot 05$ etc.					
10	**a**		Ans: test and assumption	**2**	
			•[1] appropriate test		•[1] two-sample z-test
			•[2] appropriate assumption		•[2] marks are normally distributed (or both sample sizes exceed 20)
10	**b**	i	Ans: evidence of a difference	**7**	
			•[3] appropriate hypotheses		•[3] $H_0: \mu_1 = \mu_2$ $H_1: \mu_1 \neq \mu_2$
			•[4] correct test statistic		•[4] $Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
			•[5] appropriate substitution		•[5] $= \frac{74 - 78 - 0}{\sqrt{\frac{64}{40} + \frac{49}{50}}}$
			•[6] calculate z		•[6] $= -2 \cdot 49$
			•[7] correct critical values and inequalities		•[7] $-2 \cdot 58 < -2 \cdot 49 < -1 \cdot 96$

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
			•[8] deal with H_0		•[8] so we can reject H_0 at the 5% level of significance but not at the 1% level
			•[9] appropriate conclusion		•[9] and conclude that there is some evidence of a difference in population means, the boys appearing to perform slightly better than the girls
10	**b**	ii	Ans: evidence of boys performing better	**2**	
			•[10] correct new critical value and inequality		•[10] 1% cv is now −2·33 > −2·49
			•[11] appropriate conclusion		•[11] we have evidence at the 1% level that the boys performed better
10	**c**		Ans: recommendations	**2**	
			•[12] appropriate comment		•[12] both the significance level and the alternative hypothesis
			•[13] appropriate comment		•[13] should be specified before the test is performed
Note: •[4] to •[6] can only be awarded if the correct test is chosen. The p-value approach would record that $2\mathrm{P}(Z < -2{\cdot}49) = 0{\cdot}0128$ and that $0{\cdot}01 < 0{\cdot}0128 < 0{\cdot}05$ etc					
11			Ans: no evidence of a difference	**7**	
			•[1] appropriate hypotheses		•[1] $\mathrm{H}_0: p_1 = p_2 \quad \mathrm{H}_1: p_1 \neq p_2$
			•[2] correct test statistic		•[2] $Z = \dfrac{(p_1 - p_2)}{\sqrt{pq(\frac{1}{n_1} + \frac{1}{n_2})}}$
			•[3] correct p		•[3] $p = \dfrac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0{\cdot}528$
			•[4] appropriate substitution		•[4] $z = \dfrac{0{\cdot}56 - 0{\cdot}48}{\sqrt{(0{\cdot}528)(0{\cdot}472)(\frac{1}{300} + \frac{1}{200})}}$
			•[5] calculate z		•[5] $= 1{\cdot}76$

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
			•[6] deal with H_0		•[6] 1·76<1·96 so we cannot reject H_0 at the 5% level of significance
			•[7] appropriate conclusion		•[7] and conclude that there is no evidence of a difference in preference between the two age groups
Note: •[2] to •[4] can only be awarded if the correct test is chosen. The p-value approach would record that $2\mathrm{P}(Z > 1{\cdot}76) = 0{\cdot}0784 > 0{\cdot}05$ etc					
12	**a**		Ans: 2, −2, 0·2236, 0·0127	**7**	
			•[1] correct expectation		•[1] 2
			•[2] use $\frac{\sigma^2}{n}$		•[2] $\frac{4}{100}+\ldots$
			•[3] correct standard error		•[3] $\sqrt{\frac{4}{100}+\frac{1}{100}} = 0{\cdot}2236$
			•[4] appreciate there are two possible differences		•[4] $\mathrm{P}(-1{\cdot}5 < \bar{X}_p - \bar{X}_q < 1{\cdot}5)$
			•[5] invoke the CLT		•[5] $= \mathrm{P}(\ldots\ldots\ Z\ \ldots\ldots)$
			•[6] correct strategy		•[6] $= \mathrm{P}\left(\frac{-1{\cdot}5-2}{0{\cdot}2236} < Z < \frac{1{\cdot}5-2}{0{\cdot}2236}\right)$
			•[7] calculate probability		•[7] = 0·0127
Note: If the possibility of two differences is not appreciated then •[4] and •[6] cannot be awarded.					
12	**b**		Ans: details of the CLT	**3**	
			•[8] state CLT		•[8] the CLT states that the distribution of sample means is only approximately normal if the parent distributions are not themselves normal
			•[9] appropriate justification		•[9] the samples are large
			•[10] appropriate justification		•[10] so the parent population distribution does not matter

[END OF SPECIMEN MARKING INSTRUCTIONS]

Published: September 2016

Change since last published:

Clarification made to questions and marking instructions.

Inclusion of marking instructions for candidates who are using a *p*-value approach on a graphic calculator.