National
Qualifications
EXEMPLAR PAPER ONLY

EP26/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. (a) With the aid of an example of each, define both discrete and continuous data. **4**

(b) Explain what is meant by stratified sampling. **1**

(c) State one advantage and one disadvantage of stratified sampling. **2**

2. A nurse who applied for a promoted post asked the charge nurse on his ward for a reference.

He expected to be given a reference that was positive, neutral or negative, with respective probabilities 0·6, 0·3 and 0·1.

He estimated that, with a positive reference from the charge nurse, he had a probability of 0·55 of securing the post; with a neutral reference he had a probability of 0·25; and that with a negative reference he had a probability of only 0·05 of securing the post.

Using his estimates:

(a) calculate the probability that he secured the post; **2**

(b) calculate the probability that he was given a positive reference by the charge nurse, given that he failed to secure the post. **4**

3. A commuter travels to work by car using one of two routes, A and B, and his wife claims that route A is generally faster than route B.

For three weeks the commuter selected the route to take by tossing a coin, using route A if the outcome was a head and route B otherwise.

Journey times (in minutes) were:

Route A	46	37	53	50	42	43	49	37
Route B	54	63	44	49	47	55	47	

(a) Display the data and give a brief interpretation. **2**

(b) Perform a Mann-Whitney test to investigate his wife's claim, stating the underlying assumption for the test. **7**

MARKS

4. A telephone poll of 1013 members of the general public, carried out for television, revealed that 539 people disagreed with a recent decision made by the Scottish Government.

(a) State a possible disadvantage of using this type of sample, and suggest under what circumstances this would not be a disadvantage. **2**

(b) Stating any assumption required, calculate an approximate 95% confidence interval for the proportion of the general public that disagreed with the decision.

Comment on the result in relation to the statement made by the organisation carrying out the poll that over half of Scots disagreed with the decision. **5**

5. The standard score for a random variable X is defined as $Z = \dfrac{X - \mu}{\sigma}$ where $\mu = \mathrm{E}(X)$ and $\sigma^2 = \mathrm{V}(X)$.

(a) Use the laws of expectation and variance to determine the mean and standard deviation of Z. **3**

(b) In national examinations in Mathematics and Music the mean marks were 50 and 65 respectively and the standard deviations 10 and 15 respectively.

A student scored 60 in Mathematics and 80 in Music.

Calculate this student's standard scores and comment. **2**

6. A helicopter with a payload of 900 kg is used to transport groups of eight climbers and their equipment into a wilderness area.

The weights (kg) of climbers are N(80, 16) and the weights (kg) of their equipment packs are N(30, 4).

Calculate the probability that the payload is exceeded, stating an assumption you have made. **6**

7. A country hosting a major international soccer tournament produced a special commemorative coin to be used by match referees and sold as a souvenir. Concern was expressed prior to the tournament that the coins were biased, yielding heads from 55% of tosses.

It was decided to select one of the coins at random and toss it 200 times. It was agreed to assess the coin biased, as suspected, if 110 or more heads were obtained and to assess the coin fair otherwise.

(a) If the coin is actually fair, use an appropriate approximation to estimate the probability that it is assessed biased in the way suspected. **5**

(b) Similarly, estimate the probability that a biased coin is assessed to be fair. **4**

(c) State a step that could be taken to reduce the risk of incorrect assessment of a coin as biased. **1**

MARKS

8. At the Vienna General Hospital in 1843 there were 5799 births in the two maternity clinics. Maternal mortality due to puerperal fever was of major concern. Classification of the births yielded the following contingency table.

		Clinic in which birth took place	
		No. 1	No. 2
Maternal outcome	Death from puerperal fever	274	164
	No death from puerperal fever	2786	2575

(a) Carry out a formal test of association between maternal outcome and the clinic in which the birth took place. **5**

(b) Summarise your findings, including numerical information, in a concise form that could be understood by someone with no knowledge of statistics. **2**

9. It is known from experience that a machine used to fill soft drinks bottles of various sizes delivers "fill volumes" that are normally distributed with standard deviation 0·1 ml.

In order to meet statutory requirements for fill volumes, it is desirable that the mean fill volume for 500 ml size bottles should be 500·30 ml. Shortly after the start of a production run of 500 ml bottles, a random sample of 10 bottles was taken and the fill volumes were:

500·40 500·22 500·31 500·13 500·16 500·18 500·32 500·31 500·18 500·44

(a) Use a statistical test to assess whether there is any evidence that the mean fill volume differs from 500·30 ml. **5**

(b) Under what circumstance would a *t*-test of the same hypotheses be required and how might this affect the conclusion? **2**

10. In a particular supermarket, the daily sales over a random sample of 12 days of one brand of butter have mean 124·5 packs and variance 129·96.

Following a nationwide television advertising campaign for the butter, the mean daily sales over another random sample of 12 days rose to 132·5 packs with variance 112·36.

Assess the evidence, at the 5% level of significance, that the mean number of daily sales has increased, stating any assumptions that you have made. **9**

MARKS

11. In a paper published in the journal, Environmental Pollution in 2005, Laureysens et al reported concentrations of aluminium in 13 poplar clones growing in a polluted area, measured once in August 2002 and once in November 2002. Concentrations of aluminium (micrograms of Al per gram of wood) were determined and are shown below.

Clone	August	November
Balsam Spire	8·1	11·2
Beaupre	10·0	16·3
Columbia River	18·3	12·7
Fritzi Pauley	13·3	11·1
Gaver	8·1	20·4
Gibecq	8·9	14·2
Hazendans	16·5	15·3
Hoogvorst	13·6	15·6
Primo	12·6	12·7
Raspalje	9·5	10·5
Trichobel	7·9	19·9
Unal	8·3	15·5
Wolterson	13·4	36·8

Perform a Wilcoxon signed-rank test to determine if there is any evidence that the median change in aluminium concentration from August to November differs significantly from zero.

Comment on any concern that you might have about the valid application of this test. **7**

MARKS

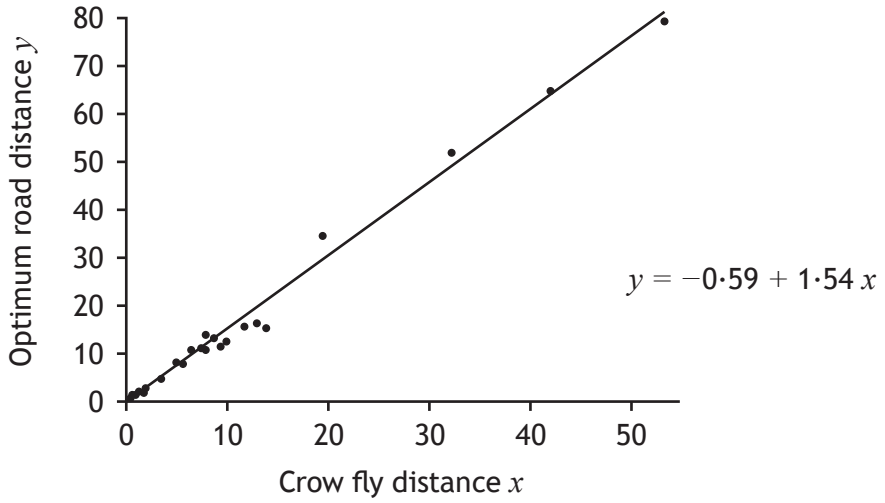
12. During the run-up to the 2012 Olympic Games, the Royal Statistical Society Centre for Statistical Education created the teaching and learning resource *Getting to the Point in 2012*.

The resource may be used to obtain both the direct (crow fly) distance (x) and the optimum road distance (y) between locations.

It was used to determine both distances from each of a random sample of 25 Scottish schools to the nearest location at which the passing of the Olympic Torch could be viewed.

For example, for Crieff High School the crow fly distance to view the torch at Kinross was 19·4 miles whereas the optimum distance by road was 34·7 miles.

A scatter plot of the data together with the least squares regression line of y on x is displayed below.



The data provide very strong evidence that the slope parameter differs significantly from zero.

(a) Give an interpretation of the slope parameter estimate of 1·54 in terms of optimum road distance per additional mile of crow fly distance. **1**

(b) Suggest a reason why a similar analysis of data for 25 schools in an urban area yielded a smaller value of the slope parameter. **1**

(c) The Scottish data gave $\bar{x} = 10 \cdot 55$, $S_{xx} = 4251 \cdot 662$ and $s^2 = 4 \cdot 669$.

(i) Calculate a 95% prediction interval for the actual optimum road distance for a Scottish school with a crow fly distance of 40 miles. **4**

(ii) Calculate a 95% confidence interval for the mean optimum road distance for a Scottish school with a crow fly distance of 40 miles. **1**

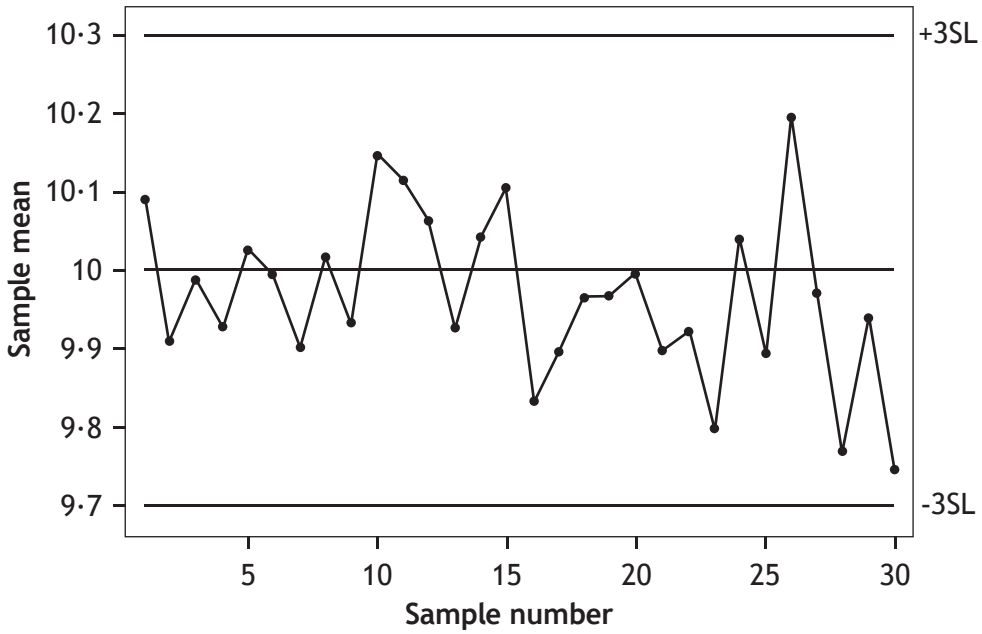
(iii) Give an interpretation for each interval. **2**

MARKS

13. A manufacturer of automotive parts monitors a critical dimension, X (mm), of a turbo compressor housing using a control chart for the means of samples of size 4, taken from the production line every 30 minutes.

When the manufacturing process is unaffected by special cause variation, ie in a state of statistical control, the random variable X has mean 10 and standard deviation 0·2.

A chart, with 3-sigma limits, of 30 consecutive samples is shown.

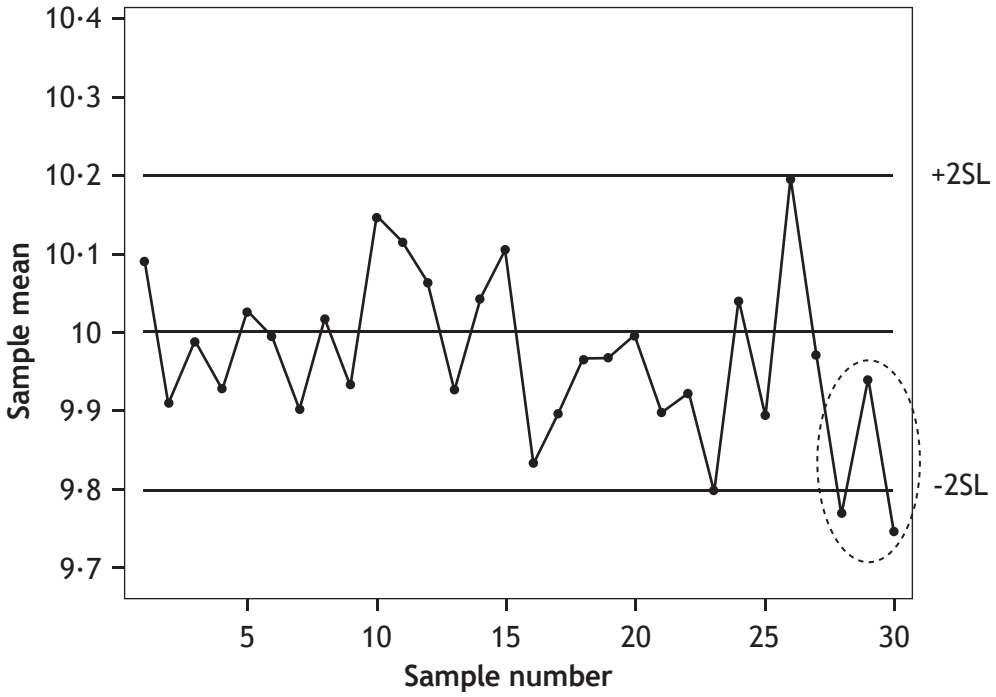


(a) Confirm the values of the limits shown and, on the assumptions that X is normally distributed and in a state of statistical control, demonstrate that the tables provided give probability 0·0026 that a sample mean will plot outwith them. **4**

13. (continued)

(b) In addition to taking the occurrence of a point outwith 3-sigma limits as evidence of the presence of special cause variation affecting the process, many control chart practitioners take the occurrence of two out of three consecutive points falling outwith 2-sigma limits on the same side of the centre line to provide such evidence.

The second chart shown indicates that this occurs with the final three samples for the 30 samples considered.



Explain why the probability of this type of signal occurring when the process is in a state of statistical control is given by:

$$2 \times \binom{3}{2} (0 \cdot 0228)^2 \, (0 \cdot 9772).$$

Show that it has a similar value to that of a point falling outwith 3-sigma limits when the process is in a state of statistical control. **4**

(c) Write down and evaluate an expression giving the probability of four out of five consecutive points lying outwith 1-sigma limits on the same side of the centre line, when the process is in statistical control. **3**

[END OF EXEMPLAR QUESTION PAPER]

Acknowledgement of Copyright

Question 11 Reprinted from *Environmental Pollution*, Vol 133, Issue 3, I. Laureysens, L. De Temmerman, T. Hatir, M. Van Gysel, R. Ceulemans, "Clonal variation in heavy metal accumulation and biomass production in poplar coppice culture. II. Vertical distribution and phytoextraction potential," Pages 541-551, © 2005, with permission from Elsevier.

Question 12 Reference to "Getting to the Point" (2012) by RSSCSE/ICSE from CensusAtSchool.org.uk. Reproduced by kind permission of International Centre for Statistical Education.

National
Qualifications
EXEMPLAR PAPER ONLY

EP26/AH/01 **Statistics**

Marking Instructions

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

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.

(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 Exemplar 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.

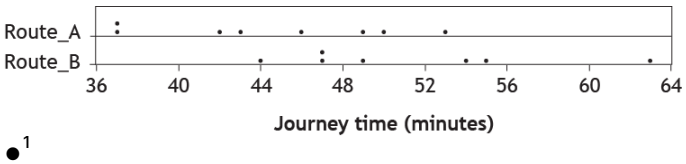
Hence, or otherwise: use the previous answer to proceed; however, another method may alternatively be used.

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	a			4	
			•[1] appropriate definition •[2] appropriate example •[3] appropriate definition •[4] appropriate example		•[1] discrete data usually involves counting •[2] eg the number of tomatoes on a plant •[3] continuous data usually involves measuring •[4] eg the heights of pupils in a class
1	b		•[5] appropriate statement	1	•[5] eg stratified sampling is used when the population can be divided into groups or strata
1	c		•[6] state one advantage	2	•[6] stratified sampling improves the precision of estimation OR stratified sampling can ensure a more representative sample
			•[7] state one disadvantage		•[7] stratified sampling may require more administrative effort and cost than simple random sampling
Notes:					
2	a		Ans: 0·41	2	
			•[1] appropriate strategy •[2] correct probability		•[1] $P(S) = P(+)P(S\|+) + P(n)P(S\|n) + P(-)P(S\|-)$ •[2] $= 0{\cdot}6\times\ 0{\cdot}55 + 0{\cdot}3\times 0{\cdot}25 + 0{\cdot}1\ \times 0{\cdot}05\ =\ 0{\cdot}41$
2	b		Ans: $\frac{27}{59}$	4	
			•[3] appropriate strategy •[4] appropriate continuation •[5] calculate probabilities •[6] calculate probability		•[3] $P(+\|\overline{S}) = \frac{P(+\cap\overline{S})}{P(\overline{S})}$ •[4] $= \frac{P(+)P(\overline{S}\|+)}{P(\overline{S})}$ •[5] $P(\overline{S}\|+) = 0{\cdot}45 \quad P(\overline{S}) = 0{\cdot}59$ •[6] $P(+\|\overline{S}) = \frac{0{\cdot}6\times 0{\cdot}45}{0{\cdot}59} = \frac{27}{59}$ or $0{\cdot}46$
Note: •[6] other methods, eg tree diagrams, are acceptable.					

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
3	**a**			**2**	
			•[1] correct diagram		 •[1]
			•[2] appropriate interpretation		•[2] the exploratory data analysis suggests that the claim may be justified
3	**b**		Ans: There is some evidence for the wife's claim	**7**	
			•[3] appropriate hypotheses		•[3] $\mathrm{H}_0 : \eta_a = \eta_b \quad \mathrm{H}_1 : \eta_a < \eta_b$
			•[4] correct rank sum		•[4] B Rank Sum: $5+7\cdot5+7\cdot5+9\cdot5+13+14+15=71\cdot5$
			•[5] recognise problem with W		•[5] W is the smaller of 71·5 and 40·5
			•[6] correct critical value		•[6] 5% cv is 41
			•[7] deal with H_0		•[7] $40\cdot5<41$ so we can reject H_0 at the 5% level of significance and
			•[8] appropriate conclusion		•[8] conclude that there is some evidence in favour of the wife's claim
			•[9] appropriate assumption		•[9] Mann-Whitney test assumes that the populations have the same shape and variability
Notes: The alternative of reverse ranking to obtain $W=1+2+3+6\cdot5+8+9+11=40\cdot5$ is also acceptable					
4	**a**			**2**	
			•[1] appropriate comment		•[1] telephone contact rules out certain members of the general public from inclusion in the sample
			•[2] appropriate comment		•[2] telephone ownership may be independent of view on this issue
4	**b**		Ans: the interval does not include 50% so, although marginal, the claim is probably reasonable	**5**	
			•[3] appropriate assumption		•[3] assuming the sample is random
			•[4] correct strategy		•[4] a 95% CI is given by $\hat{p} \pm 1{\cdot}96\sqrt{\frac{\hat{p}\hat{q}}{n}}$

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
			•[5] correct statistics		•[5] $\hat{p}=\frac{539}{1013}\quad \hat{q}=\frac{474}{1013}\quad n=1013$
			•[6] calculate CI		•[6] yielding an interval of (0·501, 0·563)
			•[7] appropriate comment		•[7] the interval does not include 50% so, although marginal, the claim is probably reasonable
Notes:					
5	**a**		Ans: proof	**3**	
			•[1] correct value of *n*		•[1] $\mathrm{E}(Z)=\mathrm{E}\left[\frac{1}{\sigma}(X-\mu)\right]=\frac{1}{\sigma}(\mathrm{E}(X)-\mathrm{E}(\mu)$ $=\frac{1}{\sigma}(\mu-\mu)=0$
			•[2] squaring the constant		•[2] $\mathrm{V}(Z)=\mathrm{V}\left[\frac{1}{\sigma}(X-\mu)\right]=\frac{1}{\sigma^2}\mathrm{V}(X-\mu)$
			•[3] acceptable completion		•[3] $=\frac{1}{\sigma^2}\mathrm{V}(X)=\frac{\sigma^2}{\sigma^2}=1$
	b		Ans: the student performed equally well in both subjects (in relation to the rest of the candidates)	**2**	
			•[4] correct scores		•[4] $Z_{Maths}=\frac{60-50}{10}=1$ $Z_{Music}=\frac{80-65}{15}=1$
			•[5] appropriate comment		•[5] the student performed equally well in both subjects (in relation to the rest of the candidates)
Notes:					
6			Ans: 0·0569	**6**	
			•[1] appropriate assumption		•[1] eg assuming that the weights are independent
			•[2] appropriate strategy		•[2] $T=\Sigma C+\Sigma P$
			•[3] correct E(*T*)		•[3] $\mathrm{E}(T)$ = 880
			•[4] correct variance		•[4] $\mathrm{V}(T)$ = 160

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
			•[5] appropriate strategy •[6] calculate probability		•[5] $\mathrm{P}(T > 900) = \mathrm{P}\left(Z > \frac{900-880}{\sqrt{160}}\right)$ •[6] = 0·0569
Notes:					
7	**a**		Ans: 0·0901	**5**	
			•[1] correct distribution •[2] appropriate approximation •[3] use of continuity correction •[4] appropriate strategy		•[1] $X \sim \mathrm{B}(200, 0{\cdot}5)$ •[2] $Y \sim \mathrm{N}(100, 50)$ •[3] $\mathrm{P}(X \geq 110) \approx \mathrm{P}(Y \geq 109{\cdot}5)$ •[4] $= \mathrm{P}\left(Z \geq \frac{109{\cdot}5-100}{\sqrt{50}}\right)$
			•[5] calculate probability		•[5] = 0·0901 (0·0896)
7	**b**		Ans: 0·4721	**4**	
			•[6] correct distribution •[7] appropriate approximation •[8] appropriate strategy •[9] correct probability		•[6] $\mathrm{P}(X < 110 \mid X \sim \mathrm{B}(200, 0{\cdot}55))$ •[7] $\approx \mathrm{P}(Y \leq 109{\cdot}5 \mid Y \sim \mathrm{N}(110, 49{\cdot}5))$ •[8] $= \mathrm{P}\left(Z \leq \frac{109{\cdot}5-110}{\sqrt{49{\cdot}5}}\right)$ •[9] = 0·4721 (0·4717)
7	**c**		Ans: the coin could be tossed more than 200 times •[10] appropriate suggestion	**1**	•[10] the coin could be tossed more than 200 times
Notes:					

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
8	**a**			**5**	
			•[1] appropriate hypotheses		•[1] H_0: there is no association between clinic and maternal outcome. H_1: there is an association between clinic and maternal outcome.
			•[2] correct value of chi-squared		•[2] $\chi^2 = 18{\cdot}22$
			•[3] appropriate cv		•[3] cv for 1df at 0·1% level is 10·827
			•[4] deal with H_0		•[4] $18{\cdot}22 > 10{\cdot}827$ so we reject H_0 at the 0·1% level of significance and
			•[5] appropriate conclusion		•[5] conclude that there is very strong evidence of an association between clinic and maternal outcome.
8	**b**			**2**	
			•[6] appropriate figures		•[6] Clinic 1, 9% death rate whereas Clinic 2, 6%
			•[7] appropriate comment		•[7] yielding evidence of a higher death rate in Clinic 1
Note: •[3] 5% level acceptable The alternative (and acceptable) p-value approach would record that $P\left(\chi_1^2 > 18{\cdot}22\right) = 0{\cdot}00002 < 0{\cdot}001$ etc. The 1% level of significance is acceptable.					
9	**a**			**5**	
			•[1] appropriate hypotheses		•[1] $H_0 : \mu = 500{\cdot}30$ $H_1 : \mu \neq 500{\cdot}30$
			•[2] correct value of mean		•[2] $\bar{x} = 500{\cdot}265$
			•[3] calculate z		•[3] $z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}} = \dfrac{500{\cdot}265 - 500{\cdot}3}{\dfrac{0{\cdot}1}{\sqrt{10}}} = -1{\cdot}11$
			•[4] deal with H_0		•[4] $-1{\cdot}11 > -1{\cdot}96$ so we cannot reject H_0 at the 5% level of significance and
			•[5] appropriate conclusion		•[5] conclude that there is no evidence of the mean fill volume differing from 500·30 ml
9	**b**			**2**	
			•[6] appropriate comment		•[6] if the population variance was unknown
			•[7] appropriate comment		•[7] it may lead to evidence that the mean volume differs from 500·30 ml
Notes: The p-value approach would record that $2\mathrm{P}(Z < -1{\cdot}11) = 0{\cdot}2670$ etc.					

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
10				**9**	
			•[1] appropriate hypotheses		•[1] $H_o: \mu_a = \mu_b \quad H_1: \mu_a > \mu_b$
			•[2] correct strategy		•[2] $s^2 = \frac{(n_a-1)s_a^2 + (n_b-1)s_b^2}{n_a+n_b-2}$
			•[3] correct value of s^2		•[3] = 121·16 $(s = 11\cdot007)$
			•[4] correct test statistic		•[4] $T_{n_a+n_b-2} = \frac{\overline{X}_a - \overline{X}_b}{s\sqrt{\frac{1}{n_a}+\frac{1}{n_b}}}$
			•[5] calculate t		•[5] = 1·780
			•[6] correct cv and inequality		•[6] $t_{22,0\cdot95} = 1\cdot717 < 1\cdot780$
			•[7] deal with H_0		•[7] so we reject H_0 at the 5% level of significance and conclude that there is evidence of an increase in the mean level of sales and there is evidence that the mean level has increased
			•[8] appropriate assumption		•[8] daily sales are normally distributed
			•[9] appropriate assumption		•[9] the population variances are equal
Notes: The p-value approach would record that $P(t > 1\cdot78) = 0\cdot0444 < 0.05$ etc.					
11				**7**	
			•[1] appropriate hypotheses		•[1] $H_0: \eta_n = \eta_a \quad H_1: \eta_n \neq \eta_a$
			•[2] correct ranks		•[2] 6 9 -8 -5 12 7 -3 4 1 2 11 10 13
			•[3] calculate statistics		•[3] $n = 13 \quad W_- = 16 \quad W_+ = 75$
			•[4] correct critical value		•[4] 5% cv is 17
			•[5] deal with H_0		•[5] 16 < 17 so we reject H_0 at the 5% level of significance
			•[6] appropriate conclusion		•[6] and conclude that there is evidence of a median change in aluminium concentration
			•[7] appropriate suggestion		•[7] the test assumes that the distributions are symmetric and this may not be the case
Notes:					

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
12	**a**			**1**	
			•[1] appropriate interpretation		•[1] for each additional crow fly mile one can expect to travel, on average, an additional 1·54 miles by road on the optimum route
12	**b**			**1**	
			•[2] appropriate reason		•[2] an urban area has almost certainly got more roads
12	**c**				
		(i)	•[3] correct $\hat{Y}_i$	4	•[3] $\hat{Y}_i = -0\cdot59 + 1\cdot54 \times 40 = 61\cdot01$
			•[4] correct strategy		•[4] PI is $\hat{Y}_i \pm t_{0\cdot975,n-2} s\sqrt{1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{S_{xx}}}$
			•[5] appropriate substitution		•[5] $61\cdot01 \pm 2\cdot069\sqrt{4\cdot669}\sqrt{1+\frac{1}{25}+\frac{(40-10\cdot55)^2}{4251\cdot662}}$
			•[6] calculate interval		•[6] PI is (56·0, 66·0)
		(ii)	•[7] calculate interval	1	•[7] CI is (58·8, 63·2)
		(iii)	•[8] appropriate interpretation	2	•[8] for a school with crow fly distance 40 miles, the PI indicates that one can be 95% confident that the optimum road distance will lie between 56·0 and 66·0 miles
			•[9] appropriate interpretation		•[9] the CI indicates that one can be 95% confident that mean optimum road distance for schools with crow fly distance 40 miles will lie between 58·8 and 63·2 miles
Notes:					
13	**a**		Ans: 0·0026	**4**	
			•[1] appropriate strategy		•[1] 3-sigmalimits are $\mu \pm 3\frac{\sigma}{\sqrt{n}}$
			•[2] correct limits		•[2] $= 10 \pm 3\frac{0\cdot2}{\sqrt{4}} = (9\cdot7, 10\cdot3)$
			•[3] overall comment		•[3] P(outwith 3σ limits) = P($Z \le -3$) + P($Z \ge 3$)
			•[4] correct probability		•[4] $0\cdot0013 + 0\cdot0013 = 0\cdot0026$

Question			Expected response (Give one mark for each •)	Max mark	Additional guidance (Illustration of evidence for awarding a mark at each •)
13	**b**			**4**	
			•[5] correct probability •[6] introduce binomial distribution •[7] explain probability •[8] appropriate conclusion		•[5] P(above 2σ limit)=P($Z \geq 2$)=$0 \cdot 0228$ •[6] assuming successive samples are independent, a binomial distribution may be used to obtain the probability of getting 2 out of 3 •[7] P(2/3)=$\binom{3}{2} p^2 q = \binom{3}{2}(0 \cdot 0228)^2(0 \cdot 9772) = 0 \cdot 0015$ •[8] and this should be doubled to allow for being below the lower 2-sigma limit to give 0·0030 which is close to 0·0026
13	**c**		Ans: 0·0053	**3**	
			•[9] correct probability		•[9] P(Z>1)=$0 \cdot 1587$
			•[10] use binomial distribution •[11] calculate probability		•[10] P(4/5)=$2\binom{5}{4}(0 \cdot 1587)^4(0 \cdot 8413)$ •[11] = 0·0053
Notes:					

[END OF EXEMPLAR 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.