CIE A/AS Level Topicals

9231 PAPER 4

FURTHER STATISTICS

EXAM-READY SERIES

TOPICAL PAPERS

- Past-exam questions
- Solutions
- Sorted by topic
- Examiner's comments

2008 to 2022

Visit: waltermaths.com for more learning resources

Disclaimer

The copyright and intellectual property owner of the material contained in this booklet is Cambridge Assessment International Education (CAIE). Our role is to curate these questions into organized topics, with the intention of benefiting students.
Walter Maths is not affiliated with and neither is it the official website for any of the stated organization. Therefore,for official announcements/news/updates regarding these qualifications, please visit: https://www.cambridgeinternational.org/

Contents

1. Continuous random variables 3

   Continuous random variables 30

2. Inference using normal and t-distributions 31

   Inference using normal and t-distributions 102

3. Chi-squared tests 103

   Chi-squared tests 148

4. Non-parametric tests 149

   Non-parametric tests 161

5. Probability generating functions 162

   Probability generating functions.....180

1. Continuous Random Variables

Q1.

A continuous random variable X has cumulative distribution function F given by

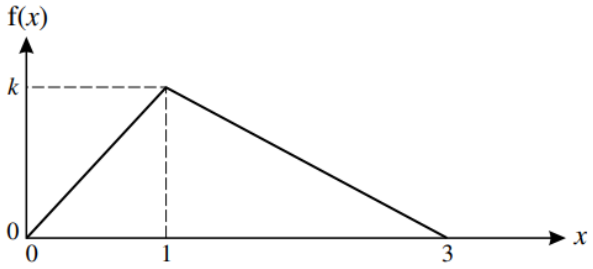
$$\mathrm{F}(x) = \begin{cases} 0 & x < -1, \\ \frac{1}{2}(x^3 + 1) & -1 \leqslant x \leqslant 1, \\ 1 & x > 1. \end{cases}$$

Find $\mathrm{P}\left(X \geqslant \frac{3}{4}\right)$, and state what can be deduced about the upper quartile of X. [3]

Obtain the cumulative distribution function of Y, where $Y = X^2$. [5]

9231/02/O/N/09 Q7

Q2.



The continuous random variable X takes values in the interval $0 \leqslant x \leqslant 3$ only. For $0 \leqslant x \leqslant 3$ the graph of its probability density function f consists of two straight line segments meeting at the point $(1, k)$, as shown in the diagram. Find k and hence show that the distribution function F is given by

$$\mathrm{F}(x) = \begin{cases} 0 & x \leqslant 0, \\ \frac{1}{3}x^2 & 0 < x \leqslant 1, \\ x - \frac{1}{2} - \frac{1}{6}x^2 & 1 < x \leqslant 3, \\ 1 & x > 3. \end{cases}$$ [6]

The random variable Y is given by $Y = X^2$. Find

(i) the probability density function of Y, [4]

(ii) the median value of Y. [4]

9231/21/M/J/11 Q11

1. Continuous Random Variables

Q3.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} 0 & x < 1, \\ \frac{1}{2} & 1 \leqslant x \leqslant 3, \\ 0 & x > 3. \end{cases}$$

Find the distribution function of X. [2]

The random variable Y is defined by $Y = X^3$. Find

(i) the probability density function of Y, [3]

(ii) the expected value and variance of Y. [3]

9231/21/O/N/11 Q6

Q4.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \dfrac{1}{2a} & -a \leqslant x \leqslant a, \\ 0 & \text{otherwise}, \end{cases}$$

where a is a positive constant. Find the distribution function of X. [3]

9231/23/M/J/12 Q9

Q5.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{2}{15}x & 1 \leqslant x \leqslant 4, \\ 0 & \text{otherwise}. \end{cases}$$

The random variable Y is defined by $Y = X^3$. Show that the distribution function G of Y is given by

$$G(y) = \begin{cases} 0 & y < 1, \\ \frac{1}{15}\left(y^{\frac{2}{3}} - 1\right) & 1 \leqslant y \leqslant 64, \\ 1 & y > 64. \end{cases}$$ [4]

Find

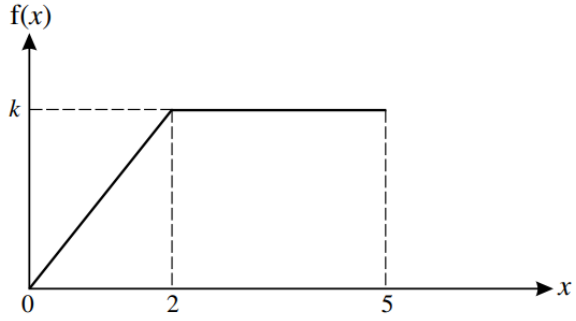
(i) the median value of Y, [3]

(ii) $E(Y)$. [4]

9231/21/O/N/12 Q7

1. Continuous Random Variables

Q6.



The continuous random variable X takes values in the interval $0 \leqslant x \leqslant 5$ only. For $0 \leqslant x \leqslant 5$ the graph of its probability density function f consists of two straight line segments, as shown in the diagram. Find k and show that f is given by

$$f(x) = \begin{cases} \frac{1}{8}x & 0 \leqslant x \leqslant 2, \\ \frac{1}{4} & 2 < x \leqslant 5, \\ 0 & \text{otherwise.} \end{cases}$$ [3]

The random variable Y is given by $Y = X^2$.

(i) Find the probability density function of Y. [6]

(ii) Show that $E(Y) = 10.25$. [3]

(iii) Show that the median of Y is the square of the median of X. [2]

9231/22/O/N/12 Q11

Q7.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{6}x & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

The random variable Y is defined by $Y = X^3$. Show that Y has probability density function g given by

$$g(y) = \begin{cases} \frac{1}{18}y^{-\frac{1}{3}} & 8 \leqslant y \leqslant 64, \\ 0 & \text{otherwise.} \end{cases}$$ [6]

Find $E(Y)$. [3]

9231/21/M/J/13 Q8

1. Continuous Random Variables

Q8.

Guided tours of a museum begin every 60 minutes. A randomly chosen tourist arrives X minutes after the start of a tour. The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \dfrac{(x-20)^2}{24\,000} & 0 < x < 60, \\ 0 & \text{otherwise.} \end{cases}$$

The random variable T is the time that the tourist has to wait for the next tour to begin. Show that the distribution function G of T is given by

$$G(t) = \begin{cases} 0 & t \leqslant 0, \\ \dfrac{8}{9} - \dfrac{(40-t)^3}{72\,000} & 0 < t < 60, \\ 1 & t \geqslant 60. \end{cases}$$ [5]

Find the median and the mean of T. [9]

9231/23/O/N/13 Q11

Q9.

The continuous random variable X has distribution function F given by

$$F(x) = \begin{cases} 0 & x < 2, \\ \frac{1}{8}x - \frac{1}{4} & 2 \leqslant x \leqslant 10, \\ 1 & x > 10. \end{cases}$$

Find the value of k for which $P(X \geqslant k) = 0.6$. [3]

The random variable Y is defined by $Y = 2 \ln X$. Find the distribution function of Y. [3]

Find the probability density function of Y and sketch its graph. [4]

9231/23/M/J/14 Q9

Q10.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{2} & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

The random variable Y is defined by $Y = X^3$. Find the distribution function of Y. [5]

Sketch the graph of the probability density function of Y. [3]

Find the probability that Y lies between its median value and its mean value. [4]

9231/21/O/N/14 Q10

1. Continuous Random Variables

Q11.

The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{21}x^2 & 1 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

The random variable Y is defined by $Y = X^2$. Show that Y has probability density function given by

$$g(y) = \begin{cases} \frac{1}{42}y^{\frac{1}{2}} & 1 \leqslant y \leqslant 16, \\ 0 & \text{otherwise.} \end{cases}$$ [5]

Find

(i) the median value of Y, [2]

(ii) the expected value of Y. [2]

9231/21/O/N/15 Q7

Q12.

The random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{6}x & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [3]

The random variable Y is defined by $Y = X^3$. Find

(ii) the probability density function of Y, [2]

(iii) the value of k for which $P(Y \geqslant k) = \frac{7}{12}$. [3]

9231/21/O/N/16 Q7

Q13.

8 The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{4}(x-1) & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [3]

The random variable Y is defined by $Y = (X-1)^3$.

(ii) Find the probability density function of Y. [4]

(iii) Find the median value of Y. [3]

9231/21/M/J/17 Q8

1. Continuous Random Variables

Q14.

The continuous random variable X has probability density function given by

$$f(x) = \begin{cases} \frac{1}{20}\left(3 - \frac{1}{\sqrt{x}}\right) & 1 \leqslant x \leqslant 9, \\ 0 & \text{otherwise.} \end{cases}$$

The random variable Y is defined by $Y = \sqrt{X}$.

(i) Show that the probability density function of Y is given by

$$g(y) = \begin{cases} \frac{1}{10}(3y - 1) & 1 \leqslant y \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$ [7]

(ii) Find the mean value of Y. [2]

9231/23/M/J/18 Q9

Q15.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{80}\left(3\sqrt{x} - \frac{8}{\sqrt{x}}\right) & 4 \leqslant x \leqslant 16, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [3]

The random variable Y is defined by $Y = \sqrt{X}$.

(ii) Find the probability density function of Y. [3]

9231/21/O/N/18 Q6

Q16.

The continuous random variable X has distribution function given by

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{90}(x^2 + x^4) & 0 \leqslant x \leqslant 3, \\ 1 & x > 3. \end{cases}$$

The random variable Y is defined by $Y = X^2$.

(i) Find the probability density function of Y. [4]

(ii) Find the mean value of Y. [2]

9231/22/O/N/18 Q7

1. Continuous Random Variables

Q17.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{3}{4x^2} + \frac{1}{4} & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [3]

(ii) Find the exact value of the interquartile range of X. [5]

9231/21/M/J/19 Q7

Q18.

The random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{30}\left(\frac{8}{x^2} + 3x^2 - 14\right) & 2 \leqslant x \leqslant 4, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of X. [3]

The random variable Y is defined by $Y = X^2$.

(ii) Find the probability density function of Y. [4]

(iii) Find the value of y such that $P(Y < y) = 0.8$. [3]

9231/21/O/N/19 Q10

Q19.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{3}{16}\left(2 - \sqrt{x}\right) & 0 \leqslant x < 1, \\ \frac{3}{16\sqrt{x}} & 1 \leqslant x \leqslant 9, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find $E(X)$. [3]

The random variable Y is such that $Y = \sqrt{X}$.

(b) Find the probability density function of Y. [5]

9231/41/M/J/20 Q3

1. Continuous Random Variables

Q20.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{1}{5}x & 0 \leqslant x < 2, \\ \frac{2}{15}(5-x) & 2 \leqslant x \leqslant 5, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function of X. [3]

(b) Find the median value of X. [2]

(c) Find $E(X^2)$. [2]

(d) Find $P(1 \leqslant X \leqslant 3)$. [2]

9231/43/M/J/20 Q3

Q21.

The continuous random variable X has cumulative distribution function F given by

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{60}(16x - x^2) & 0 \leqslant x \leqslant 6, \\ 1 & x > 6. \end{cases}$$

(a) Find the interquartile range of X. [4]

(b) Find $E(X^3)$. [4]

The random variable Y is such that $Y = \sqrt{X}$.

(c) Find the probability density function of Y. [3]

9231/41/O/N/20 Q6

1. Continuous Random Variables

Q22.

The continuous random variable X has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 2, \\ \frac{1}{60}x^2 - \frac{1}{15} & 2 \leqslant x \leqslant 8, \\ 1 & x > 8. \end{cases}$$

(a) Find $\mathrm{P}(3 \leqslant X \leqslant 6)$. [1]

(b) Find $\mathrm{E}\left(\sqrt{X}\right)$. [3]

(c) Find $\mathrm{Var}\left(\sqrt{X}\right)$. [2]

(d) The random variable Y is defined by $Y = X^3$. Find the probability density function of Y. [3]

9231/42/O/N/20 Q4

Q23.

The continuous random variable X has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < 0, \\ \frac{1}{81}x^2 & 0 \leqslant x \leqslant 9, \\ 1 & x > 9. \end{cases}$$

(a) Find $\mathrm{E}\left(\sqrt{X}\right)$. [3]

(b) Find $\mathrm{Var}\left(\sqrt{X}\right)$. [2]

(c) The random variable Y is given by $Y^3 = X$. Find the probability density function of Y. [3]

9231/41/M/J/21 Q3

Q24.

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} \frac{1}{8} & 0 \leqslant x < 1, \\ \frac{1}{28}(8-x) & 1 \leqslant x \leqslant 8, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the cumulative distribution function of X. [3]

(b) Find the value of the constant a such that $\mathrm{P}(X \leqslant a) = \frac{5}{7}$. [3]

The random variable Y is given by $Y = \sqrt[3]{X}$.

(c) Find the probability density function of Y. [5]

9231/43/M/J/21 Q6

1. Continuous Random Variables

Q25.

The continuous random variable X has cumulative distribution function F given by

$$\mathrm{F}(x) = \begin{cases} 0 & x < -1, \\ \frac{1}{2}(1+x)^2 & -1 \leqslant x \leqslant 0, \\ 1-\frac{1}{2}(1-x)^2 & 0 < x \leqslant 1, \\ 1 & x > 1. \end{cases}$$

(a) Find the probability density function of X. [2]

(b) Find $\mathrm{P}\left(-\frac{1}{2} \leqslant X \leqslant \frac{1}{2}\right)$. [2]

(c) Find $\mathrm{E}(X^2)$. [2]

(d) Find $\mathrm{Var}(X^2)$. [2]

9231/41/O/N/21 Q2

Q26.

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} a+\frac{1}{5}x & 0 \leqslant x < 1, \\ 2a-\frac{1}{5}x & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise}, \end{cases}$$

where a is a constant.

(a) Find the value of a. [3]

(b) Find $\mathrm{E}(X^2)$. [2]

(c) Find the cumulative distribution function of X. [3]

9231/42/O/N/21 Q3

Q27.

The continuous random variable X has probability density function f given by

$$\mathrm{f}(x) = \begin{cases} kx(4-x) & 0 \leqslant x < 2, \\ k(6-x) & 2 \leqslant x \leqslant 6, \\ 0 & \text{otherwise}, \end{cases}$$

where k is a constant.

(a) Show that $k = \frac{3}{40}$. [1]

(b) Given that $\mathrm{E}(X) = 2.5$, find $\mathrm{Var}(X)$. [3]

(c) Find the median value of X. [4]

9231/41/M/J/22 Q3

1. Continuous Random Variables

Q28.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} \frac{3}{8}\left(1+\frac{1}{x^2}\right) & 1 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find $E\left(\sqrt{X}\right)$. [3]

The random variable Y is given by $Y = X^2$.

(b) Find the probability density function of Y. [4]

(c) Find the 40th percentile of Y. [3]

9231/43/M/J/22 Q4

Q29.

The continuous random variable X has cumulative distribution function F given by

$$F(x) = \begin{cases} 0 & x < 0, \\ 1 - \frac{1}{144}(12-x)^2 & 0 \leqslant x \leqslant 12, \\ 1 & x > 12. \end{cases}$$

(a) Find the upper quartile of X. [2]

(b) Find $\text{Var}(X^2)$. [5]

The random variable Y is given by $Y = \sqrt{X}$.

(c) Find the probability density function of Y. [3]

9231/41/O/N/22 Q5

1. Continuous Random Variables

Q30.

The continuous random variable X has probability density function f given by

$$f(x) = \begin{cases} k & 0 \leqslant x < 1, \\ kx & 1 \leqslant x \leqslant 2, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a constant.

(a) Show that $k = \frac{2}{5}$. [1]

(b) Find the interquartile range of X. [5]

(c) Find $\text{Var}(X)$. [4]

9231/42/O/N/22 Q4

Solutions

Q1.

7	Find P($X \geq ¾$):	$1 - F(¾) = 1 - ½[(¾)^3 + 1]$	M1		
		$= 37/128$ *or* 0·289	*A1		
	State deduction about upper quartile Q_3 (dep *A1):	$Q_3 > ¾$	B1	(3)	
	Express cum. dist. fn. G of Y in terms of X:	$G(y) = P(Y \leqslant y) = P(-\sqrt{y} \leqslant X \leqslant \sqrt{y})$	M1 A1		
	Relate to F:	$= F(\sqrt{y}) - F(-\sqrt{y})$	M1		
	Simplify:	$= ½[(\sqrt{y})^3 + 1] - ½[(-\sqrt{y})^3 + 1]$			
		$= y^{3/2}$	A1		
	State G in full (√ on previous result):	0 $(y < 0)$, $y^{3/2}$ $(0 \leqslant y \leqslant 1)$, 1 $(y > 1)$	A1√	(5)	**[8]**

Q2.

(b)	Find k by equating area under graph to 1:	$½k + k = 1$, $k = ⅔$	M1 A1		
	Find f(x) for $0 < x \leq 1$ and $1 < x \leq 3$:	kx [$= ⅔x$], $½k(3 - x)$ [$= 1 - ⅓x$]	M1 A1		
	Integrate to find F(x) for $0 < x \leq 1$ and $1 < x \leq 3$:	$⅓x^2$, $x - ½ - x^2/6$ **A.G.**	M1 A1	6	
	(i) Relate dist. fn. G(y) of Y to X:	$G(y) = P(Y < y) = P(X^2 < y)$			
	(working may be omitted)	$= P(X < y^{1/2}) = F(y^{1/2})$			
		$= ⅓y$, $y^{1/2} - ½ - y/6$	M1 A1		
	Differentiate to find g(y):	$g(y) = ⅓$ $(0 < y \leq 1)$,			
		$½ y^{-1/2} - 1/6$ $(1 < y \leq 9)$,			
		[0 otherwise]	M1 A1	4	
	(ii) Use $G(m) = ½$ to find eqn for median m:	$m^{1/2} - ½ - m/6 = ½$	M1		
	Solve quadratic for $\sqrt{m}$:	$(\sqrt{m})^2 - 6\sqrt{m} + 6 = 0$, $\sqrt{m} = 3 \pm \sqrt{3}$	M1 A1		
	Select value of m in interval:	$m = (3 - \sqrt{3})^2 = 12 - 6\sqrt{3}$ *or* 1·61	B1	4	**[14]**

Q3.

6	Integrate to find F(x) for $1 \leq x \leq 3$:	$F(x) = ½(x - 1)$	B1		
	State F(x) for other intervals of x:	0 $(x < 1)$, 1 $(x > 3)$	B1	2	
	(i) Relate dist. fn. G(y) of Y to X:	$G(y) = P(Y < y) = P(X^3 < y)$			
	(working may be omitted)	$= P(X < y^{1/3}) = F(y^{1/3})$			
		$= ½(y^{1/3} - 1)$	M1 A1		
	Differentiate to find g(y):	$g(y) = y^{-2/3}/6$ $(1 \leq y \leq 27)$			
		[= 0 otherwise]	B1	3	
	(ii) Find expected value of Y (or X^3):	$E(Y) = \int_1^{27} y\,(y^{-2/3}/6)\,dy$			
		$= [y^{4/3}/8]_1^{27}$ *or* $[x^4/8]_1^3$			
		$= (81 - 1)/8 = 10$	B1		
	Find variance of Y:	$E(Y^2) = \int_1^{27} y^2\,(y^{-2/3}/6)\,dy$			
		$= [y^{7/3}/14]_1^{27}$ *or* $[x^7/14]_1^3$			
		$= (2187 - 1)/14 = 1093/7$			
		$\text{Var}(Y) = E(Y^2) - 10^2$			
		$= 393/7$ *or* 56·1[4]	M1 A1	3	**[8]**

Q4.

9	Integrate f(x) to find F(x) for $-a \le x \le a$ (A.E.F.): (Finding F(x) = $x/2a$ earns M1 A0)	F(x) = $x/2a + c$ = $(x + a)/2a$	M1 A1		
	State F(x) for other intervals:	F(x) = 0 ($x < -a$), F(x) = 1 ($x > a$)	B1	3	
	Relate dist. fn. G(y) of Y to X for central interval: (working may be omitted; ignore other intervals)	G(y) = P($Y < y$) = P($e^X < y$) = P($X < \ln y$) = F($\ln y$) = $(\ln y + a)/2a$; $(e^{-a} \le y \le e^a)$	M1 A1; A1	3	
	Find k from P($Y \ge k$) = ¼ with a = 4: (Using G(k) = ¼ can earn M1)	$1 - G(k) = ¼$, $(\ln k)/2a + ½ = ¾$	M1 A1		
		$\ln k = 2$, $k = e^2$ *or* 7·39	A1	3	**[9]**

Q5.

7	Integrate f(x) to find F(x) for $1 \le x \le 4$:	$F(x) = x^2/15 + c = (x^2 - 1)/15$	M1 A1		
	Relate dist. fn. G(y) of Y to X for $1 \le x \le 4$:	$G(y) = P(Y < y) = P(X^3 < y)$ $= P(X < y^{1/3}) = F(y^{1/3})$ $= (y^{2/3} - 1)/15$	**A.G** M1 A1	4	
(i)	Find relation for median m of Y:	$G(m) = ½$, $m^{2/3} = 17/2$	M1 A1		
	Evaluate m:	$m = 24{\cdot}8$	A1	3	
(ii)	EITHER Find g(y) and formulate E(Y):	$g(y) = 2y^{-1/3}/45$ $E(Y) = \int y g(y) dy = \int 2y^{2/3}/45 \, dy$	M1 A1		
	OR Formulate E(Y) in terms of X:	$E(Y) = E(X^3) = \int 2x^4/15 \, dx$	(M1 A1)		
	Integrate and apply limits:	$E(Y) = \left[\frac{2y^5}{75}\right]_1^{64}$ *or* $\left[\frac{2x^5}{75}\right]_1^4$ $= 2(1024 - 1)/75$ $= 682/25$ *or* 27·3	M1 A1	4	**[11]**

1. Continuous Random Variables

Q6.

	Find E(X) using $\int xf(x)dx$:	E(X) = $\int_2^4$ $(5x^2 - x^3 - 4x)/10$ dx M1 A1 = ½(4³–2³) – 3(4⁴–2⁴)/40 – 3(4²–2²)/5 = 28 – 18 – 7·2 = 2·8 *A1			
	Verify E(X) within 10% of 2·69 (A1 dep *A1):	(E(X) – 2·69)/2·69 = 0·041 < 0·1 *or* 1·1 × 2·69 = 2·96 > E(X) M1 A1		5	
	Show derivation of tabular entry:	60 $\int_2^{3.6}$ $(5x - x^2 - 4)/10$ dx M1 = 60[3($5x^2/2 - x^3/3 - 4x$)/10]$_2^{3.6}$ *or* [45$x^2 - 6x^3 - 72x$]$_2^{3.6}$ = 122·4 – 83·328 – 28·8 *or* 60 × 0·1712 = 10·272 **A.G** A1		2	
	State (at least) null hypothesis: Combine last 2 cells since exp. value < 5: Calculate χ^2 (to 2 d.p.): State or use consistent tabular value (to 2 d.p.): [or if no cells combined: Valid method for reaching conclusion: Conclusion (A.E.F., dep *A1, *B1):	H_0: f(x) fits data (A.E.F.) B1 *O*: . . . 8 *E*: . . . 14·208 B1 χ^2 = 0·8126 + 0·0584 + 0·2011 + 2·7135 = 3·78[47] M1 *A1 $\chi^2_{3\,0.9}$ = 6·25[1] $\chi^2_{4,\,0.9}$ = 7·78] *B1 Accept H_0 if χ^2 < tabular value M1 3·78 < 6·25 so f(x) does fit A1		7	**[14]**

Q7.

8	*EITHER* Integrate f(x) to find F(x) for $2 \le x \le 4$: Find G(y) for $8 \le y \le 64$: *OR* Find G(y) for $8 \le y \le 64$: State G(y) for other values of y: Differentiate to find g(y) for $8 \le y \le 64$: *EITHER:* Find E(Y) from $\int yg(y)$: *OR:* Find E(Y) from E(X^3): Evaluate:	$F(x) = \int_2^x f(x)\,dx$ M1 A1 G(y) = P($Y < y$) = P($X^3 < y$) = P($X < y^{1/3}$) = F($y^{1/3}$) = ($y^{2/3} - 4$)/12 M1 A1 G(y) = P($Y < y$) = P($X^3 < y$) = P($X < y^{1/3}$) = $\int_2^{y^{1/3}} f(x)\,dx$ (M1 A1) $= \left[x^2/12\right]_2^{y^{1/3}} = \left(y^{2/3} - 4\right)/12$ (M1 A1) G(y) = 0 ($y < 8$), 1 ($y > 64$) B1 g(y) = $y^{-1/3}/18$ **A.G.** B1 $(1/18)\int_8^{64} y^{2/3}\,dy = (1/30)\left[y^{5/3}\right]_8^{64}$ M1 A1 $(1/6)\int_2^4 x^4\,dx = (1/30)\left[x^5\right]_8^{64}$ (M1 A1) = ($4^5 - 2^5$)/30 = 992/30 = 496/15 *or* 33·1 A1	6 3	**9**

Q8.

11 (b)	Find F(x) for $0 < x < 60$ by integration:	F(x) = ⅓ $(x - 20)^3$/24000			
		+ ⅓ 20^3/24000	M1		
		= $(x - 20)^3$ / 72000 + 1/9	A1		
	State F(x) for other values of x:	F(x) = 0 ($x \leq 0$), 1 ($x \geq 60$)	B1		
	Find G(t) for $0 < t < 60$ from $X + T = 60$:	G(t) = P($T < t$)			
		= P($60 - X < t$)			
		= P($X > 60 - t$)			
		= 1 – F($60 - t$)			
		= $8/9 - (40 - t)^3$ / 72000 **A.G.**	M1 A1	5	
	Formulate eqn for median m of T:	$8/9 - (40 - m)^3$ / 72000 = ½	M1		
	Find value of m:	$(40 - m)^3 = (8/9 - ½)$ 72000			
		= 28000	M1		
		$m = 40 - 28000^{1/3} = 9{\cdot}63$	A1		
	Find g(t) for $0 < t < 60$:	g(t) = $(40 - t)^2$ / 24000	M1 A1		
	Find E(T) from $\int t\, g(t)\, dt$:	E(T) = $\int(40^2t - 80t^2 + t^3)\, dt$ /24000	M1		
		= $[½\, 40^2t^2 - ⅓\, 80t^3 + ¼t^4]_0^{60}$/24000	A1		
		= 120 – 240 + 135 = 15	A1	9	**14**

Q9.

9	Find k for which P($X \geqslant k$) = 0·6:	0·6 = 1 – F(k)	M1		
		= $1 - (k/8 - ¼)$	M1		
		k = 26/5 *or* 5·2	A1	3	
	Find G(y) from $Y = 2 \ln X$ for $2 \leqslant x \leqslant 10$:	G(y) = P($Y < y$) = P($2 \ln X < y$)			
	(allow < or ⩽ throughout)	= P($X < e^{y/2}$) = F($e^{y/2}$)			
	(result may be stated)	= $e^{y/2}/8 - ¼$ ($2 \ln 2 \leqslant y \leqslant 2 \ln 10$)			
		or ($\ln 4 \leqslant y \leqslant \ln 100$)			
		or ($1{\cdot}39 \leqslant y \leqslant 4{\cdot}61$)	M1 A1		
	State G(y) for other values of x:	0 ($y < 2 \ln 2$) *and* 1 ($y > 2 \ln 10$)	B1	3	
	Find g(y) for $2 \ln 2 \leqslant y \leqslant 2 \ln 10$:	g(y) = $e^{y/2}/16$	M1 A1		
	Sketch positive exponential for $2 \ln 2 \leqslant y \leqslant 2 \ln 10$		B1		
	Show g(y) = 0 on either side of this interval		B1	4	**10**

Q10.

10	Find F(x) for $1 \leqslant x \leqslant 3$:	$F(x) = \frac{1}{2}(x - 1)$	B1		
	Find G(y) from $Y = X^3$ for $1 \leqslant x \leqslant 3$:	$G(y) = P(Y < y) = P(X^3 < y)$ $= P(X < y^{\frac{1}{3}}) = F(y^{\frac{1}{3}})$			
	(result may be stated)	$= \frac{1}{2}(y^{\frac{1}{3}} - 1); (1 \leqslant y \leqslant 27)$	M1 A1; B1		
	State G(y) for other values of y:	0 ($y < 1$) *and* 1 ($y > 27$)	B1	5	
	Find g(y) for $1 \leqslant y \leqslant 27$ (√ on G(y)):	$g(y) = \frac{y^{-\frac{2}{3}}}{6}$ *or* $\frac{1}{6y^{\frac{2}{3}}}$	B1√		
	Sketch g(y) for $1 \leqslant y \leqslant 27$ with g(y) = 0 on either side of this interval		B1 B1	3	
	Find mean of Y:	$E(Y) = \int y\, g(y)\, dy = \int (\frac{y^{\frac{1}{3}}}{6})\, dy$			
	(no need to find median = 8)	$= \left[\frac{y^{\frac{4}{3}}}{8}\right]_1^{27} = 10$	M1 A1		
	Find probability Y lies between median and mean:	$G(10) - G(8)$ *or* $\left\lvert G(10) - \frac{1}{2}\right\rvert$ $= \frac{1}{2}\left(10^{\frac{1}{3}} - 8^{\frac{1}{3}}\right)$			
	(2 s.f. sufficient)	*or* $\left\lvert\frac{1}{2}\left(10^{\frac{1}{3}} - 1\right) - \frac{1}{2}\right\rvert = 0{\cdot}077$ [2]	M1 A1	4	**[12]**

Q11.

7	Find F(x) for $1 \leqslant x \leqslant 4$:	$F(x) = (x^3 - 1)/63$	**B1**		
	Find G(y) from $Y = X^2$ for $1 \leqslant x \leqslant 4$:	$G(y) = P(Y < y) = P(X^2 < y)$ $= P(X < y^{1/2}) = F(y^{1/2})$			
	(result may be stated)	$= (y^{3/2} - 1)/63$	**M1 A1**		
	Find g(y) for corresponding range of y:	$g(y) = y^{1/2}/42$ **A.G.**	**A1**		
	Find or state corresponding range of y :	$1 \leqslant y \leqslant 16$ **A.G.**	**B1**	5	
(i)	Find median value m of Y:	$(m^{3/2} - 1)/63 = ½$ $m = 32{\cdot}5^{2/3} = 10{\cdot}2$	**M1 A1**	2	
(ii)	Find E(Y) [or equivalently E(X^2)]:	$E(Y) = \int y\, g(y)\, dy = \int y^{3/2}\, dy\, /42$ $= [y^{5/2}]_1^{16}/105 = 1023/105$ $= 341/35$ *or* 9·74	**M1 A1**	2	**9**

1. Continuous Random Variables

Q12.

7 (i)	Find or state distribution function F(x) for $2 \leqslant x \leqslant 4$: Use F(2) = 0 or F(4) = 1 to find F(x):	$F(x) = \int f(x)\,dx = x^2/12 + c$ $F(x) = x^2/12 - \frac{1}{3}$ [$(2 \leqslant x \leqslant 4)$], 0 ($x < 2$), 1 ($x > 4$)	**M1 A1** **A1**	[3]
(ii)	Find or state G(y) from $Y = X^3$ for $2 \leqslant x \leqslant 4$: (allow < or $\leqslant$ throughout) (A0 if G(y) incorrect) Find g(y) by differentiation:	$G(y) = P(Y < y) = P(X^3 < y)$ $= P(X < y^{1/3}) = F(y^{1/3})$ $= y^{2/3}/12 - \frac{1}{3}$ $g(y) = (1 / 18)\, y^{-1/3}$ [for $8 \leqslant y \leqslant 64$, 0 otherwise]	**M1 A1**	[2]
(iii)	Formulate condition for k: (M0 for 7/12 = G(k)) Find k:	(AEF) $7/12 = 1 - G(k) = 1 - k^{2/3}/12 + \frac{1}{3}$ $k^{2/3} = 9$, $k = 27$	**M1 A1** **A1**	[3]

Q13.

8(i)	$F(x) = \int f(x)\,dx = x^2/8 - x/4$ [+ c]	**M1**	Find or state distribution function F(x) for $2 \leqslant x \leqslant 4$ using F(2) = 0 or F(4) = 1 to find c if necessary
	$= x^2/8 - x/4$ *or* $\{(x-1)^2 - 1\}/8$ (AEF)	**A1**	State F(x) for other values of x
	F(x) = 0 ($x < 2$), F(x) = 1 ($x > 4$)	**A1**	
	Total:	**3**	
8(ii)	*EITHER:* $G(y) = P(Y < y) = P((X-1)^3 < y)$ $= P(X < 1 + y^{1/3}) = F(1 + y^{1/3})$ $= (1 + y^{1/3})^2/8 - (1 + y^{1/3})/4$ *or* $(y^{2/3} - 1)/8$	**(M1 A1)**	Find or state G(y) for $2 \leqslant x \leqslant 4$ from $Y = (X-1)^3$ (allow < or $\leqslant$ throughout)
	OR: Use $x = 1 + y^{1/3}$ to find $f(x) = \frac{1}{4}\, y^{1/3}$ *and* $dx/dy = \frac{1}{3}\, y^{-2/3}$	**(M1 A1)**	Find f(x) *and* dx/dy for use in $g(y) = f(x) \times \lvert dx/dy \rvert$
	$g(y)$ [= $G'(y)$] = $(1/12)\, y^{-1/3}$ *or* $1 / (12\, y^{1/3})$	**A1**	Find g(y) in simplified form
	for $1 \leqslant y \leqslant 27$ [g(y) = 0 otherwise]	**A1**	State corresponding range of y for G(y) *or* g(y)
	Total:	**4**	
8(iii)	$(m^{2/3} - 1)/8 = \frac{1}{2}$	**M1**	Find median value m of Y from G(m) = ½
	$m^{2/3} = 5$, $m = \sqrt{125}$ *or* $5\sqrt{5}$ *or* 11·2	**M1 A1**	
	Total:	**3**	

1. Continuous Random Variables

Q14.

9(i)	*EITHER:* $F(x) = \int f(x)\,dx = (1/20)(3x - 2\sqrt{x}\,[+c])$ $c = -1$ so $F(x) = (1/20)(3x - 2\sqrt{x} - 1)$	**M1**	Find or state distribution function $F(x)$ for $1 \leqslant x \leqslant 9$ (may be implied by $G(y)$)
	or $(3/20)\,x - (1/10)\,\sqrt{x} - 1/20$ $G(y)\,[= P(Y < y) = P(\sqrt{X} < y) = P(X < y^2)\,]$	**A1**	Find or state $G(y)$ from $Y = \sqrt{X}$ for $1 \leqslant x \leqslant 9$ or $1 \leqslant y \leqslant 3$
	$= F(y^2)$ $= (1/20)(3y^2 - 2y - 1)$	**M1**	Allow A1√ as FT on expression found for $F(x)$
	or $(3/20)\,y^2 - (1/10)\,y - 1/20$	**A2**	Verify $g(y)$ (differentiation may be implied)
	$g(y) = G'(y) = (1/10)(3y - 1)$ **AG** [for $1 \leqslant y \leqslant 3$, $g(y) = 0$ otherwise]	**M1 A1**	**SC** Missing/incorrect c can earn M1 M1 A1√ M1 (max 4/7)
		7	
	OR: Use of $g(y) = f(x) \times \lvert dx / dy \rvert$	**(*M1)**	Reference to standard result required (not in syllabus)
	$f(x) = (1/20)(3 - 1/y)$ (dep *M1)	**(M1 A1)**	Find $f(x)$ using $x = y^2$
	$dx/dy = 2y$ (dep *M1)	**(M1 A1)**	Find dx/dy using $x = y^2$
	$g(y) = f(x) \times dx/dy = (1/10)(3y - 1)$ **AG** [for $1 \leqslant y \leqslant 3$, $g(y) = 0$ otherwise]	**(M1 A1)**	
		(7)	
9(ii)	$E(Y) = (1/10) \int (3y^2 - y)\,dy$	**M1**	Find mean of Y from $\int y\,g(y)\,dy$
	$= (1/10)\,[y^3 - \frac{1}{2}y^2]_1^3 = 11/5$ *or* 2·2	**A1**	
		2	

Q15.

6(i)	$F(x) = \int f(x)\,dx = (1/80)(2x^{3/2} - 16x^{1/2}\,[+c])$	**M1**	Find or state distribution function $F(x)$ for $4 \leqslant x \leqslant 16$ using $F(4) = 0$ or $F(16) = 1$ to find c if necessary
	$= (1/80)(2x^{3/2} - 16x^{1/2} + 16)$ *or* $(1/40)(x^{3/2} - 8x^{1/2} + 8)$ *or* $x^{3/2}/40 - x^{1/2}/5 + 1/5$ (AEF)	**A1**	State $F(x)$ for other values of x
	$F(x) = 0$ $(x < 4)$, $F(x) = 1$ $(x > 16)$	**A1**	
		3	
6(ii)	*EITHER:* $G(y)\,[= P(Y < y) = P(\sqrt{X} < y) = P(X < y^2)\,]$ $= F(y^2) = (1/40)(y^3 - 8y + 8)$ (AEF)	**M1A1**	Find or state $G(y)$ for $2 \leqslant y \leqslant 4$ from $Y = \sqrt{X}$ (allow < or ⩽ throughout; **FT** on constant term)
	OR: Use $x = y^2$ to find $f(x) = (1/80)(3y - 8/y)$ *and* $dx/dy = 2y$	**(M1A1)**	Find $f(x)$ *and* dx/dy for use in $g(y) = f(x) \times \lvert dx/dy \rvert$
	$g(y)\,[= G'(y)] = (1/40)(3y^2 - 8)$ *or* $(3/40)\,y^2 - 1/5$ [for $2 \leqslant y \leqslant 4$, $g(y) = 0$ otherwise]	**A1**	Find $g(y)$ in simplified form for $2 \leqslant y \leqslant 4$
		3	

1. Continuous Random Variables

Q16.

7(i)	*EITHER:* G(y) [= P($Y < y$) = P($X^2 < y$) = P($X < y^{1/2}$) = F($y^{1/2}$)] = (1/90) ($y + y^2$)	**M1A1**	Find or state G(y) for $0 \leqslant x \leqslant 3$ from $Y = X^2$ (allow < or $\leqslant$ throughout)
	OR: Use $x = y^{1/2}$ to find f(x) = (1/90) ($2x + 4x^3$) = (1/90) ($2y^{1/2} + 4y^{3/2}$) *and* dx/dy = 1 / $2y^{1/2}$	**(M1A1)**	Find f(x) *and* dx/dy for use in g(y) = f(x) × | dx/dy |
	g(y) [= G′(y)] = (1/90) (1 + 2y)	**A1**	Find g(y) in simplified form
	for $0 \leqslant y \leqslant 9$ [g(y) = 0 otherwise]	**A1**	State corresponding range of y at any stage
		4	
7(ii)	E(Y) = (1/90) $\int (y + 2y^2)$ dy	**M1**	Find mean of Y from $\int y\, g(y)\, dy$
	= (1/90) [½ y^2 + ⅔ y^3]$_0^9$ = 117/20 *or* 5.85	**A1**	
		2	

Q17.

7(i)	F(x) = $\int$ f(x) dx = – 3/(4x) + x/4 [+ c]	**M1**	Find or state distribution function F(x) for $1 \leqslant x \leqslant 3$
	= – 3/(4x) + x/4 + ½ *or* ¼ (– 3/x + x + 2)	**A1**	using F(1) = 0 or F(3) = 1 to find c if necessary
	F(x) = 0 ($x <$ *or* $\leqslant$ 1), F(x) = 1 ($x >$ *or* $\geqslant$ 3)	**A1**	State F(x) for other values of x
		3	
7(ii)	$\int_1^Q$ f(x) dx = – 3/4Q + Q/4 + ½ = ¼ [*or* ¾] (AEF)	**M1**	Formulate equation for either quartile value Q
	Q^2 + [*or* –] Q – 3 = 0	**A1**	
	Q_1 = ½ (– 1 + √13), Q_3 = ½ (1 + √13)	**A1 A1**	Find lower quartile Q_1 and upper quartile Q_3
	$Q_3 - Q_1$ = 1	**A1**	Find interquartile range
		5	

1. Continuous Random Variables

Q18.

10(i)	$F(x) = \int f(x)\,dx = (1/30)(-8/x + x^3 - 14x)$ [+ c]	**M1**	Find or state distribution function $F(x)$ for $2 \leqslant x \leqslant 4$
	$F(x) = (1/30)(-8/x + x^3 - 14x + 24)$	**M1**	Using $F(2) = 0$ or $F(4) = 1$ to find c if necessary. AEF
	$F(x) = 0$ ($x <$ *or* $\leqslant 2$), $F(x) = 1$ ($x >$ *or* $\geqslant 4$)	**A1**	State $F(x)$ for other values of x
		3	
10(ii)	$G(y) = P(Y < y) = P(X^2 < y)$ $G(y) = P(X < \sqrt{y}) = F(\sqrt{y})$ $G(y) = (1/30)\left(-8/y^{\frac{1}{2}} + y^{\frac{3}{2}} - 14y^{\frac{1}{2}} + 24\right)$	**M1** **A1**	Find or state $G(y)$ for $2 \leqslant x \leqslant 4$ from $Y = X^2$ (allow $<$ or $\leqslant$ throughout)
	Alternative method for question 10(ii)		
	Use $x = y^{\frac{1}{2}}$ to find $f(x) = (1/30)(8/y + 3y - 14), \frac{dx}{dy} = -\frac{1}{2}y^{-\frac{1}{2}}$	**(M1** **A1)**	Find $f(x)$ *and* $\frac{dx}{dy}$ for use in $g(y) = f(x) \times \left\lvert\frac{dy}{dx}\right\rvert$
	$g(y)[= G'(y)] = (1/30)\left(4/y^{\frac{3}{2}} + (3/2)y^{\frac{1}{2}} - 7/y^{\frac{1}{2}}\right)$ for $4 \leqslant y \leqslant 16$ [$g(y) = 0$ otherwise]	**A1** **A1**	Find $g(y)$. AEF State corresponding range of y for $G(y)$ *or* $g(y)$
		4	
10(iii)	$(1/30)\left(-8/y^{\frac{1}{2}} + y^{\frac{3}{2}} - 14y^{\frac{1}{2}} + 24\right) = 0.8$	**M1**	Set $G(y) = 0{\cdot}8$
	$-8 + y^2 - 14y = 0$, $y = 7 + \sqrt{57}$ *or* 14·5[5] [rejecting $7 - \sqrt{57}$; allow 14·6]	**M1 A1**	Rearrange to give quadratic in y and solve to find value of y
		3	

Q19.

3(a)	$E(X) = \int_0^1 \frac{3}{16}x\left(2 - \sqrt{x}\right)dx + \int_1^9 \frac{\frac{3}{16}x}{\sqrt{x}}dx$	**M1**
	$\frac{3}{16}\left[x^2 - \frac{2}{5}x^{\frac{5}{2}}\right] + \frac{3}{16}\left[\frac{2}{3}x^{\frac{3}{2}}\right]$	**A1**
	$\frac{269}{80}$	**A1**
		3
3(b)	$F(x) = \begin{cases} \frac{3}{8}\left(x - \frac{1}{3}x^{3/2}\right), & 0 \leqslant x < 1 \\ \frac{3}{8}x^{1/2} - \frac{1}{8}, & 1 \leqslant x \leqslant 9 \end{cases}$	**M1A1**
	$G(y) = \begin{cases} \frac{3}{8}\left(y^2 - \frac{1}{3}y^3\right), & 0 \leqslant y < 1 \\ \frac{3}{8}y - \frac{1}{8}, & 1 \leqslant y \leqslant 3 \end{cases}$	**M1A1**
	($G(y) = 0$ for $y \leqslant 0$ and $= 1$ for $y \geqslant 3$) $g(y) = \begin{cases} \frac{3}{8}\left(2y - y^2\right), & 0 \leqslant y < 1 \\ \frac{3}{8}, & 1 \leqslant y \leqslant 3 \end{cases}$ $= 0$ otherwise	**A1**
		5

Q20.

3(a)	$F(x)=\begin{cases}\frac{x^2}{10}, & 0\leqslant x<2\\ \frac{1}{15}(10x-x^2)-\frac{2}{3}, & 2\leqslant x\leqslant 5\end{cases}$	**M1A1**
	= 0 for $x < 0$ and = 1 for $x > 5$	**A1**
		3
3(b)	$F(m) = ½$ so $2m^2-20m+35=0$	**M1**
	$m=5-\frac{1}{2}\sqrt{30}=2.26$	**A1**
		2
3(c)	$E(X^2)=\int_0^2 x^2.\frac{x}{5}dx+\int_2^5\frac{2}{15}(5-x).x^2dx=\left[\frac{x^4}{20}\right]+\left[\frac{2}{15}\left(\frac{5}{3}x^3-\frac{x^4}{4}\right)\right]$	**M1**
	= 6.5	**A1**
		2

3(d)	F(3) – F(1)	**M1**
	$=\frac{11}{15}-\frac{1}{10}=\frac{19}{30}$	**A1**
	Alternative method for 3(d)	
	$\int_1^2\frac{x}{5}dx+\int_2^3\frac{2}{15}(5-x)dx=\left[\frac{x^2}{10}\right]+\left[\frac{2}{15}\left(5x-\frac{x^2}{2}\right)\right]$	**M1**
	$=\frac{3}{10}+\frac{1}{3}=\frac{19}{30}$	**A1**
		2

1. Continuous Random Variables

Q21.

6(a)	UQ: F(U) = 0.75: $u^2 - 16u + 45 = 0$	**M1**	Obtain quadratic equation for u or l
	$u = 8 - \sqrt{19}$ (= 3.64)	**A1**	Value of UQ
	LQ: $l^2 - 16l + 15 = 0$, $l = 1$	**A1**	Value of LQ
	IQR = UQ – LQ = $7 - \sqrt{19} = 2.64$	**A1 FT**	
		4	
6(b)	$f(x) = \begin{cases} \frac{1}{30}(8-x), & 0 \leqslant x \leqslant 6 \\ 0, & \text{otherwise} \end{cases}$	**B1**	May be implied, only need to see $f(x) = \frac{1}{30}(8-x)$
	Find E(X^3) by integration: $\frac{1}{30}\int_0^6 (8x^3 - x^4)\,dx$ $= \left[2x^4 - \frac{1}{5}x^5\right]_0^6$	**M1 A1**	Attempt integration (with PDF not CDF) limits not required Correct integrated expression, with limits
	= 34.56	**A1**	
		4	
6(c)	$G(y) = \frac{1}{60}(16y^2 - y^4)$	**M1**	CDF for Y
	$g(y) = \frac{1}{60}(32y - 4y^3)$ for $0 \leq y \leq \sqrt{6}$	**M1A1**	Differentiate to find PDF for Y Correct g(y) and correct range seen anywhere
		3	

Q22.

4(a)	F(6) – F(3) = $\frac{9}{20}$ or 0.45	**B1**	
		1	
4(b)	$f(x) = \begin{cases} \frac{1}{30}x, & 2 \leq x \leq 8 \\ 0, & \textit{otherwise} \end{cases}$	**B1**	May be implied. Only need to see $f(x) = \frac{1}{30}x$
	$E(\sqrt{X}) = \int_2^8 \frac{1}{30}x^{\frac{3}{2}}dx = \left[\frac{1}{75}x^{\frac{5}{2}}\right]$	**M1**	Integrated
	= 2.34	**A1**	2.338… Answer of 2.34 without $\frac{1}{75}x^{\frac{5}{2}}$ scores B1B1 (2/3)
		3	
4(c)	$\text{Var}(\sqrt{X}) = [\int_2^8 \frac{1}{30}x^2 dx = \left[\frac{1}{90}x^3\right]] - E(\sqrt{X})^2$ $\text{Var}(\sqrt{X}) = \left[\int_2^8 \frac{1}{30}x^2 dx = \left[\frac{1}{90}x^3\right]_2^8\right] - E(\sqrt{X})^2$	**M1**	
	= 5.6 – 2.338^2 = 0.133 or 0.134	**A1**	
		2	

Q23.

3(a)	$\int_0^9 \frac{2}{81}x^{1.5}\,dx$	**B1**	$f(x)=\frac{2}{81}x$, correct expression as integrand.
	$\frac{2}{81}\times\frac{2}{5}x^{2.5}$	**M1**	Integrate, FT only on their PDF expression.
	$\frac{12}{5}=2.4$	**A1**	
		3	
3(b)	$\int_0^9 \frac{2}{81}x^2\,dx-(their\ a)^2$	**M1**	
	$\frac{2}{81}\times\frac{9^3}{3}-2.4^2=\frac{6}{25}$	**A1**	
		2	
3(c)	$G(y)=\frac{1}{81}y^6$	**M1**	
	$g(y)=\frac{2}{27}y^5$	**M1**	
	Fully correct including 'for $0\leqslant y\leqslant\sqrt[3]{9}$ and 0 otherwise'	**A1**	Condone 2.08
		3	

Q24.

6(a)	$F(x)=\begin{cases}\frac{1}{8}x & 0\leqslant x<1,\\ \frac{1}{28}\left(8x-\frac{1}{2}x^2\right)-\frac{1}{7} & 1\leqslant x\leqslant 8.\end{cases}$	**M1** **A1**	Integrate both parts (all powers of x increase by 1), allow M1 if no constant term. Both parts correct, with domains. Any equivalent form, for example: $\frac{2}{7}x-\frac{1}{56}x^2-\frac{1}{7}$.
	$F(x)=0$ for $x<0$ and $F(x)=1$ for $x>8$.	**A1**	
	Alternative method for question 6(a)		
	$F(x)=\begin{cases}\frac{1}{8}x & 0\leqslant x<1,\\ 1-\frac{1}{56}(8-x)^2 & 1\leqslant x\leqslant 8.\end{cases}$	**M1** **A1**	Integrate both parts (all powers of x increase by 1), allow M1 if no constant term. Both parts correct, with domains. Any equivalent form, for example: $\frac{1}{56}(16x-x^2-8)$.
	$F(x)=0$ for $x<0$ and $F(x)=1$ for $x>8$.	**A1**	
		3	
6(b)	$F(a)=\frac{5}{7}$ leading to $\frac{1}{28}\left(8a-\frac{1}{2}a^2\right)-\frac{1}{7}=\frac{5}{7}$	***M1**	Correct method and attempt to simplify, ft their part **(a).**
	$a^2-16a+48=0$ $a=4\ [a=12]$	**DM1**	Obtain quadratic and solve.
	$a=4$	**A1**	Correct single answer, WWW.
		3	

6(c)	$G(y)=\begin{cases}\frac{1}{8}y^3 & 0 \leqslant y<1,\\ \frac{1}{56}\left(16y^3-y^6\right)-\frac{1}{7} & 1 \leqslant y \leqslant 2.\end{cases}$	**M1** **A1**	Substitute $x=y^3$ into both parts of their F(x) from part **(a)**. Correct functions in terms of y, any equivalent form.
	$0 \leqslant y<1$ and $1 \leqslant y \leqslant 2$seen in CDF or PDF	**A1**	Condone $\leqslant$ instead of $<$.
	$g(y)=\begin{cases}\frac{3}{8}y^2 & 0 \leqslant y<1,\\ \frac{3}{28}\left(8y^2-y^5\right) & 1 \leqslant y<2,\\ 0 & \text{otherwise.}\end{cases}$	**M1** **A1**	Differentiate their G(y), all powers of y decrease by 1. Must include 'otherwise'.
		5	

Q25.

2(a)	$f(x)=\begin{cases}1+x & -1\leqslant x\leqslant 0,\\ 1-x & 0<x\leqslant 1,\\ 0 & \text{otherwise.}\end{cases}$	**M1**	Differentiation attempted.
		A1	All correct, including 0 otherwise.
		2	
2(b)	$P\left(-\frac{1}{2}\leqslant X\leqslant\frac{1}{2}\right)=F\left(\frac{1}{2}\right)-F\left(-\frac{1}{2}\right)=\frac{7}{8}-\frac{1}{8}$	**M1**	Can do by integration with correct limits.
	$\frac{3}{4}$	**A1**	
		2	
2(c)	$\int_{-1}^{0}x^2(1+x)dx+\int_{0}^{1}x^2(1-x)dx$ $\left[\frac{1}{3}x^3+\frac{1}{4}x^4\right]+\left[\frac{1}{3}x^3-\frac{1}{4}x^4\right]$	**M1**	
	$\frac{1}{12}+\frac{1}{12}=\frac{1}{6}$	**A1**	
		2	

2(d)	$\int_{-1}^{0}x^4(1+x)dx+\int_{0}^{1}x^4(1-x)dx-\left(\frac{1}{6}\right)^2$ $\left[\frac{1}{5}x^5+\frac{1}{6}x^6\right]+\left[\frac{1}{5}x^5-\frac{1}{6}x^6\right]-\frac{1}{36}$	**M1**	Complete method with their mean squared explicit.
	$\frac{1}{15}-\frac{1}{36}=\frac{7}{180}$	**A1**	
		2	

1. Continuous Random Variables

Q26.

3(c)	$F(x)=\begin{cases}0 & x<0\\ \frac{1}{5}\left(2x+\frac{1}{2}x^2\right) & 0\leqslant x<1,\\ \frac{1}{5}\left(4x-\frac{1}{2}x^2-1\right) & 1\leqslant x\leqslant 2,\\ 1 & x>2.\end{cases}$	**M1**	Integration of their PDF.
		A1	Middle 2 parts correct.
	$=0\ \ (x<0)$ and $=1\ (x>2)$	**A1**	First and last parts correct and all domains correct, with no gaps.
		3	

Q27.

3(a)	$k2x^2-\frac{1}{3}x^3+k\left[6x-\frac{1}{2}x^2\right]=1$ $k\left[8-\frac{8}{3}+36-18-12+2\right]=\frac{40}{3}k=1$ $k=\frac{3}{40}$	**B1**	Integrate and equate to 1, no errors seen.
		1	
3(b)	$E(X^2)=k\int_0^2\left(4x^3-x^4\right)dx+k\int_2^6\left(6x^2-x^3\right)dx=$ $k\left[x^4-\frac{1}{5}x^5\right]+k\left[2x^3-\frac{1}{4}x^4\right]=k[9.6+96]=7.92$	**M1***	Attempt at $E(X^2)$, correct limits, integrated.
	$Var(X)=E(X^2)-E^2(X)=$ *their* $(7.92)-2.5^2$	**depM1**	Use of correct formula, with their $E(X^2)$.
	1.67	**A1**	
		3	

3(c)	$\int_0^2 kx(4-x)dx+\int_2^m k(6-x)=\frac{1}{2}$	**M1***	Correct expression and attempt to integrate.
	$\int_0^2 kx(4-x)dx=\frac{2}{5}$	**B1**	May be implied
	$\frac{2}{5}+k\left[6x-\frac{1}{2}x^2\right]=\frac{1}{2}$ $6m-\frac{1}{2}m^2-12+2=\frac{1}{10k}$ $m^2-12m+\frac{68}{3}=0$	**depM1**	First integral and $\frac{1}{2}$ may be seen combined as $\frac{1}{10}$. Obtain quadratic equation.
	$m=6\pm\frac{2}{3}\sqrt{30}$ $m=6-\frac{2}{3}\sqrt{30}\ (=2.35)$	**A1**	Single answer.
		4	

1. Continuous Random Variables

Q28.

4(a)	$\mathrm{E}(\sqrt{X}) = \int \sqrt{x}\frac{3}{8}\left(1+\frac{1}{x^2}\right)\mathrm{d}x \left(= \int_1^3 \frac{3}{8}\left(x^{0.5}+x^{-1.5}\right)\mathrm{d}x \right)$	**M1**	
	$\frac{3}{8}\left[\frac{2}{3}x^{1.5} - 2x^{-0.5}\right]$	**A1**	
	$\frac{3}{8}\left(2\sqrt{3} - \frac{2}{\sqrt{3}} - \frac{2}{3} + 2\right) = \frac{1}{2}\left(1+\sqrt{3}\right)$ $(= 1.37)$	**A1**	
		3	

Q29.

5(a)	$\mathrm{F}(U) = 0.75$, so $1 - \frac{1}{144}(12-U)^2 = \frac{3}{4}$	**M1**	
	$(12-U)^2 = 36, \quad 12-U = \pm 6,$ $U = 6$ only	**A1**	
		2	

5(b)	[f(x) =] $\quad \frac{1}{72}(12-x)$	**B1**	Differentiate given PDF.
	$\mathrm{E}(X^2) = \int_0^{12} \frac{1}{72}\left(12x^2 - x^3\right)dx = \frac{1}{72}\left(4x^3 - \frac{1}{4}x^4\right)$ [= 24]	**M1**	Integrate $x^2 f(x)$ with their $f(x)$.
	$E\left(X^4\right) = \int_0^{12} \frac{1}{72}\left(12x^4 - x^5\right)dx = \frac{1}{72}\left(\frac{12}{5}x^5 - \frac{1}{6}x^6\right)$ [= 1382.4]	**M1**	Integrate $x^4 f(x)$ with their $f(x)$.
	$\mathrm{Var}\left(X^2\right) = E\left(X^4\right) - E^2\left(X^2\right)$ $\quad [= 1382.4 - 24^2]$	**M1**	Using correct formula with numerical values possibly unsimplified.
	$1382.4 - 576 = 806(.4)$	**A1**	
		5	
5(c)	$\mathrm{G}(y) = \begin{cases} 0 & y < 0, \\ 1 - \frac{1}{144}\left(12-y^2\right)^2 & 0 \leqslant y \leqslant \sqrt{12}, \\ 1 & y > \sqrt{12}. \end{cases}$	**M1**	Change the variable: $1 - \frac{1}{144}\left(12-y^2\right)^2$ Allow domain unchanged at this stage.
	$\frac{1}{36}\left(12y - y^3\right)$	**M1**	Differentiate their $\mathrm{G}(y)$.
	$g(y) = \begin{cases} \frac{1}{36}\left(12y - y^3\right) & 0 \leqslant y \leqslant \sqrt{12} \\ 0 & \text{otherwise} \end{cases}$	**A1**	Fully correct.

Q30.

4(a)	$\int_0^1 k\,dx + \int_1^2 kx\,dx = [kx] + \left[\frac{k}{2}x^2\right], \quad k + \frac{3}{2}k = 1,\ k = \frac{2}{5}$	**B1**	Integration and $k + 3k/2 = 1$ oe seen.
		1	
4(b)	$F(x) = \begin{cases} kx & 0 \leqslant x < 1, \\ \frac{k}{2}x^2 + \frac{1}{5} & 1 \leqslant x \leqslant 2 \end{cases}$	**M1**	CDF or integration with $\frac{1}{5}x^2 + \frac{1}{5}$ seen.
	UQ: $\frac{k}{2}x^2 + \frac{1}{5} = \frac{3}{4}$,	**M1**	Method for UQ
	$\left[x^2 + 1 = 5 \times \frac{3}{4}, \quad x^2 = \frac{11}{4}\right] \quad x = \frac{1}{2}\sqrt{11}$	**A1**	UQ
	LQ: $kx = \frac{1}{4}, \quad x = \frac{5}{8}$	**B1**	LQ
	$\text{IQR} = \frac{1}{2}\sqrt{11} - \frac{5}{8} = 1.03(3)$	**A1**	CAO
		5	

4(c)	$E(X) = \int_0^1 kx\,dx + \int_1^2 kx^2\,dx = \left[\frac{k}{2}x^2\right] + \left[\frac{k}{3}x^3\right] = \frac{17}{15}$	**M1**	With correct limits.
	$E(X^2) = \int_0^1 kx^2\,dx + \int_1^2 kx^3\,dx = \left[\frac{k}{3}x^3\right] + \left[\frac{k}{4}x^4\right] = \frac{49}{30}$	**M1**	With correct limits.
	$\text{Var}(X) = \frac{49}{30} - \left(\frac{17}{15}\right)^2$	**M1**	Using correct formula with numerical values.
	$\frac{157}{450} \quad (= 0.349)$	**A1**	
		4	

2. Inference Using Normal and t-Distributions

Q1.

The time for a certain kind of plaster to dry has a normal distribution with mean μ minutes. In order to find a confidence interval for μ, the times for a random sample of 10 specimens to dry were measured. The results, x minutes, are summarised by

$$\Sigma x = 285.4, \quad \Sigma(x - \bar{x})^2 = 448.65.$$

Calculate a 95% confidence interval for μ. [5]

Estimate the width of a 95% confidence interval for μ based on a random sample of 20 specimens. [3]

9231/02/M/J/08 Q7

Q2.

A study was carried out to investigate the effect of moderate exercise on blood cholesterol levels. The levels were measured before and after a training programme administered to a random sample of 8 people. The results, in suitable units, are given in the following table.

Person	1	2	3	4	5	6	7	8
Before	195	209	184	215	295	198	310	191
After	197	201	192	204	266	183	272	168

Stating any assumption you need to make, test at the 10% significance level whether, after the training programme,

(i) there is a reduction in the population mean cholesterol level, [7]

(ii) there is a reduction of more than 5 units in the population mean cholesterol level. [3]

9231/02/M/J/08 Q9

Q3.

The independent random variables X and Y have normal distributions with the same variance σ^2. Samples of 5 observations of X and 10 observations of Y are made, and the results are summarised by $\Sigma x = 15$, $\Sigma x^2 = 128$, $\Sigma y = 36$ and $\Sigma y^2 = 980$. Find a pooled estimate of σ^2. [3]

9231/02/O/N/08 Q6

2. Inference Using Normal and t-Distributions

Q4.

The pulse rate of each member of a random sample of 25 adult UK males who exercise for a given period each week is measured in beats per minute. A 98% confidence interval for the mean pulse rate, μ beats per minute, for all such UK males was calculated as $61.21 < \mu < 64.39$, based on a t-distribution.

(i) Calculate the sample mean pulse rate and the standard deviation used in the calculation. [5]

(ii) State an assumption necessary for the validity of the confidence interval. [1]

(iii) The mean pulse rate for all UK males is 72 beats per minute. State, giving a reason, if it can be concluded that, on average, UK males who exercise have a reduced pulse rate. [2]

9231/02/O/N/08 Q7

Q5.

A perfume manufacturer had one bottle-filling machine, but because of increased sales a second machine was obtained. In order to compare the performance of the two machines, a random sample of 50 bottles filled by the first machine and a random sample of 60 bottles filled by the second machine were checked. The volumes of the contents from the first machine, x_1 ml, and from the second machine, x_2 ml, are summarised by

$$\Sigma x_1 = 1492.0, \quad \Sigma x_1^2 = 44\,529.52, \quad \Sigma x_2 = 1803.6, \quad \Sigma x_2^2 = 54\,220.84.$$

The volumes have distributions with means μ_1 ml and μ_2 ml for the first and second machines respectively. Test, at the 2% significance level, whether μ_2 is greater than μ_1. [10]

Find the set of values of α for which there would be evidence at the α% significance level that $\mu_2 - \mu_1 > 0.1$. [4]

9231/02/O/N/08 Q11

Q6.

Part of a research study of identical twins who had been separated at birth involved a random sample of 9 pairs, in which one twin had been raised by the natural parents and the other by adoptive parents. The IQ scores of these twins were measured, with the following results.

Twin pair	1	2	3	4	5	6	7	8	9
IQ of twin raised by natural parents	82	92	115	132	88	95	112	83	123
IQ of twin raised by adoptive parents	92	88	115	134	97	94	107	88	130

It may be assumed that the difference in IQ scores has a normal distribution. The mean IQ scores of separated twins raised by natural parents and by adoptive parents are denoted by μ_N and μ_A respectively. Obtain a 90% confidence interval for $\mu_N - \mu_A$. [6]

One of the researchers claimed that there was no evidence of a difference between the two population means. State, giving a reason, whether the confidence interval supports this claim. [2]

9231/02/M/J/09 Q8

2. Inference Using Normal and t-Distributions

Q7.

A study was made of the acidity levels in farmland on opposite sides of an island. The levels were measured at six randomly chosen points on the eastern side and at five randomly chosen points on the western side. The values obtained, in suitable units, are denoted by x_E and x_W respectively. The sample means $\bar{x}_E$ and $\bar{x}_W$, and unbiased estimates of the two population variances, s_E^2 and s_W^2, are as follows.

$$\bar{x}_E = 5.035,\ s_E^2 = 0.0231,\ \ \bar{x}_W = 4.782,\ s_W^2 = 0.0195.$$

The population means on the eastern and western sides are denoted by μ_E and μ_W respectively. State suitable hypotheses for a test for a difference between the mean acidity levels on the two sides of the island. [1]

Stating any required assumptions, obtain the rejection region for a test at the 5% significance level of whether the mean acidity levels differ on the two sides of the island. Give the conclusion of the test. [9]

Find the largest value of a for which the samples above provide evidence at the 5% significance level that $\mu_E - \mu_W > a$. [4]

9231/02/M/J/09 Q11

Q8.

A machine produces metal discs whose diameters have a normal distribution. The mean of this distribution is intended to be 10 cm. Accuracy is checked by measuring the diameters of a random sample of six discs. The diameters, in cm, are as follows.

10.03 10.02 9.98 10.06 10.08 10.01

Calculate a 99% confidence interval for the mean diameter of all discs produced by the machine. [5]

Deduce a 99% confidence interval for the mean circumference of all discs produced by the machine. [1]

9231/02/O/N/09 Q6

2. Inference Using Normal and t-Distributions

Q9.

150 sheep, chosen from a large flock of sheep, were divided into two groups of 75. Over a fixed period, one group had their grazing controlled and the other group grazed freely. The gains in weight, in kg, were recorded for each animal and the table below shows the sample means and the unbiased estimates of the population variances for the two samples.

	Sample mean	Unbiased estimate of population variance
Controlled grazing	19.14	20.54
Free grazing	15.36	9.84

It is required to test whether the population mean for sheep having their grazing controlled exceeds the population mean for sheep grazing freely by less than 5 kg. State, giving a reason, if it is necessary for the validity of the test to assume that the two population variances are equal. [1]

Stating any other assumption, carry out the test at the 5% significance level. [8]

9231/02/O/N/09 Q8

Q10.

An examination involved writing an essay. In order to compare the time taken to write the essay by students in two large colleges, a sample of 12 students from college A and a sample of 8 students from college B were randomly selected. The times, t_A and t_B, taken for these students to write the essay were measured, correct to the nearest minute, and are summarised by

$$n_A = 12, \quad \Sigma t_A = 257, \quad \Sigma t_A^2 = 5629, \quad n_B = 8, \quad \Sigma t_B = 206, \quad \Sigma t_B^2 = 5359.$$

Stating any required assumptions, calculate a 95% confidence interval for the difference in the population means. [8]

State, giving a reason, whether your confidence interval supports the statement that the population means, for the two colleges, are equal. [1]

9231/21/M/J/10 Q8

2. Inference Using Normal and t-Distributions

Q11.

Carpal Tunnel syndrome is a condition which affects a person's ability to grip with their hands. Researchers tested a treatment for this syndrome which was applied to 8 randomly chosen patients. A pre-treatment and a post-treatment test of grip was given to each patient, with the following results, measured in kg.

Patient	1	2	3	4	5	6	7	8
Pre-treatment grip	24.3	29.5	28.0	28.5	21.5	28.7	25.1	26.3
Post-treatment grip	28.3	34.6	30.3	31.6	21.5	29.8	26.0	27.5

Stating any required assumption, test, at the 1% significance level, whether the mean grip of people with the syndrome increases after undergoing the treatment. [9]

It is given that there is evidence at the 10% significance level that the mean grip increases by more than w kg. Find an inequality for w. [3]

9231/23/M/J/10 Q9

Q12.

The mean Intelligence Quotient (IQ) of a random sample of 15 pupils at School A is 109. The mean IQ of a random sample of 20 pupils at School B is 112. You may assume that the IQs for the populations from which these samples are taken are normally distributed, and that both distributions have standard deviation 15. Find a 90% confidence interval for $\mu_B - \mu_A$, where μ_A and μ_B are the population mean IQs. [6]

9231/02/O/N/10 Q6.

Q13.

A national athletics coach suspects that, on average, 200-metre runners' indoor times exceed their outdoor times by more than 0.1 seconds. In order to test this, the coach randomly selects eight 200-metre runners and records their indoor and outdoor times. The results, in seconds, are shown in the table.

Runner	A	B	C	D	E	F	G	H
Indoor time	21.5	21.8	20.9	21.2	21.4	21.4	21.2	21.0
Outdoor time	21.1	21.7	20.7	20.9	21.3	21.0	21.1	20.8

Stating suitable hypotheses and any necessary assumption that you make, test the coach's suspicion at the 2.5% level of significance. [10]

9231/02/O/N/10 Q9

2. Inference Using Normal and t-Distributions

Q14.

A greengrocer claims that his cabbages have a mean mass of more than 1.2 kg. In order to check his claim, he weighs 10 cabbages, chosen at random from his stock. The masses, in kg, are as follows.

1.26 1.24 1.17 1.23 1.18 1.25 1.19 1.20 1.21 1.18

Stating any assumption that you make, test at the 10% significance level whether the greengrocer's claim is supported by this evidence. [8]

9231/21/M/J/11 Q7

Q15.

In a crossword competition the times, x minutes, taken by a random sample of 6 entrants to complete a crossword are summarised as follows.

$$\Sigma x = 210.9 \qquad \Sigma(x - \bar{x})^2 = 151.2$$

The time to complete a crossword has a normal distribution with mean μ minutes. Calculate a 95% confidence interval for μ. [5]

Assume now that the standard deviation of the population is known to be 5.6 minutes. Find the smallest sample size that would lead to a 95% confidence interval for μ of width at most 5 minutes. [4]

9231/21/M/J/11 Q8

Q16.

Mr Lee asserts that boys are slower than girls at completing a particular mathematical puzzle. In order to test his assertion, a random sample of 40 boys and a random sample of 60 girls are selected from a large group of students who attempted the puzzle. The times taken by the boys, b minutes, and the times taken by the girls, g minutes, are summarised as follows.

$$\Sigma b = 92.0 \quad \Sigma b^2 = 216.5 \quad \Sigma g = 129.8 \quad \Sigma g^2 = 288.8$$

Test at the 2.5% significance level whether this evidence supports Mr Lee's assertion. [9]

9231/21/M/J/11 Q9

Q17.

The independent random variables X and Y have distributions with the same variance σ^2. Random samples of 5 observations of X and n observations of Y are made and the results are summarised by

$$\Sigma x = 5.5, \quad \Sigma x^2 = 15.05, \quad \Sigma y = 8.0, \quad \Sigma y^2 = 36.4.$$

Given that the pooled estimate of σ^2 is 3, find the value of n. [7]

9231/23/M/J/11 Q6

2. Inference Using Normal and t-Distributions

Q18.

A company decides that its employees should follow an exercise programme for 30 minutes each day, with the aim that they lose weight and increase productivity. The weights, in kg, of a random sample of 8 employees at the start of the programme and after following the programme for 6 weeks are shown in the table.

Employee	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Weight before (kg)	98.6	87.3	90.4	85.2	100.5	92.4	89.9	91.3
Weight after (kg)	93.5	85.2	88.2	84.6	95.4	89.3	86.0	87.6

Assuming that loss in weight is normally distributed, find a 95% confidence interval for the mean loss in weight of the company's employees. [6]

Test at the 5% significance level whether, after the exercise programme, there is a reduction of more than 2.5 kg in the population mean weight. [5]

9231/23/M/J/11 Q8

Q19.

A random sample of five metal rods produced by a machine is taken. Each rod is tested for hardness. The results, in suitable units, are as follows.

524 526 520 523 530

Assuming a normal distribution, calculate a 95% confidence interval for the population mean. [5]

Some adjustments are made to the machine. Assume that a normal distribution is still appropriate and that the population variance remains unchanged. A second random sample, this time of ten metal rods, is now taken. The results for hardness are as follows.

525 520 522 524 518 520 519 525 527 516

Stating suitable hypotheses, test at the 10% significance level whether there is any difference between the population means before and after the adjustments. [8]

9231/21/O/N/11 Q9

Q20.

A random sample of 8 observations of a normal random variable X gave the following summarised data, where $\bar{x}$ denotes the sample mean.

$$\Sigma x = 42.5 \qquad \Sigma(x - \bar{x})^2 = 15.519$$

Test, at the 5% significance level, whether the population mean of X is greater than 4.5. [7]

Calculate a 95% confidence interval for the population mean of X. [3]

9231/21/M/J/12 Q9

2. Inference Using Normal and t-Distributions

Q21.

Engineers are investigating the speed of the internet connection received by households in two towns P and Q. The speeds, in suitable units, in P and Q are denoted by x and y respectively. For a random sample of 50 houses in town P and a random sample of 40 houses in town Q the results are summarised as follows.

$$\Sigma x = 240 \qquad \Sigma x^2 = 1224 \qquad \Sigma y = 168 \qquad \Sigma y^2 = 754$$

Calculate a 95% confidence interval for $\mu_P - \mu_Q$, where μ_P and μ_Q are the population mean speeds for P and Q. [6]

Test, at the 1% significance level, whether μ_P is greater than μ_Q. [6]

9231/23/M/J/12 Q10

Q22.

The leaves from oak trees growing in two different areas A and B are being measured. The lengths, in cm, of a random sample of 7 oak leaves from area A are

$$6.2, \quad 8.3, \quad 7.8, \quad 9.3, \quad 10.2, \quad 8.4, \quad 7.2.$$

Assuming that the distribution is normal, find a 95% confidence interval for the mean length of oak leaves from area A. [5]

The lengths, in cm, of a random sample of 5 oak leaves from area B are

$$5.9, \quad 7.4, \quad 6.8, \quad 8.2, \quad 8.7.$$

Making suitable assumptions, which should be stated, test, at the 5% significance level, whether the mean length of oak leaves from area A is greater than the mean length of oak leaves from area B. [9]

9231/21/O/N/12 Q9

Q23.

7 The speed v at which a javelin is thrown by an athlete is measured in km h^{-1}. The results for 10 randomly chosen throws are summarised by

$$\Sigma v = 1110.8, \quad \Sigma(v - \bar{v})^2 = 333.9,$$

where $\bar{v}$ is the sample mean.

(i) Stating any necessary assumption, calculate a 99% confidence interval for the mean speed of a throw. [6]

The results for a further 5 randomly chosen throws are now combined with the above results. It is found that the sample variance is smaller than that used in part **(i)**.

(ii) State, with reasons, whether a 95% confidence interval calculated from the combined 15 results will be wider or less wide than that found in part **(i)**. [2]

9231/23/O/N/12 Q7

2. Inference Using Normal and t-Distributions

Q24.

Experiments are conducted to test the breaking strength of each of two types of rope, P and Q. A random sample of 50 ropes of type P and a random sample of 70 ropes of type Q are selected. The breaking strengths, p and q, measured in appropriate units, are summarised as follows.

$$\Sigma p = 321.2 \qquad \Sigma p^2 = 2120.0 \qquad \Sigma q = 475.3 \qquad \Sigma q^2 = 3310.0$$

Test, at the 10% significance level, whether the mean breaking strengths of type P and type Q ropes are the same. [10]

9231/23/O/N/12 Q9

Q25.

Each of a random sample of 6 cyclists from a cycling club is timed over two different 10 km courses. Their times, in minutes, are recorded in the following table.

Cyclist	A	B	C	D	E	F
Course 1	18.5	17.8	19.2	22.3	16.5	20.0
Course 2	20.2	20.4	18.1	20.6	18.5	20.5

Assuming that differences in time over the two courses are normally distributed, test at the 10% significance level whether the mean times over the two courses are different. [8]

9231/23/M/J/13 Q7

Q26.

A gardener P claims that a new type of fruit tree produces a higher annual mass of fruit than the type that he has previously grown. The old type of tree produced 5.2 kg of fruit per tree, on average. A random sample of 10 trees of the new type is chosen. The masses, x kg, of fruit produced are summarised as follows.

$$\Sigma x = 61.0 \qquad \Sigma x^2 = 384.0$$

Test, at the 5% significance level, whether gardener P's claim is justified, assuming a normal distribution. [6]

Another gardener Q has his own type of fruit tree. The masses, y kg, of fruit produced by a random sample of 10 trees grown by gardener Q are summarised as follows.

$$\Sigma y = 70.0 \qquad \Sigma y^2 = 500.6$$

Test, at the 5% significance level, whether the mean mass of fruit produced by gardener Q's trees is greater than the mean mass of fruit produced by gardener P's trees. You may assume that both distributions are normal and you should state any additional assumption. [8]

9231/21/M/J/13 Q9

2. Inference Using Normal and t-Distributions

Q27.

The number, x, of a certain type of sea shell was counted at 60 randomly chosen sites, each one metre square, along the coastline in country A. The number, y, of the same type of shell was counted at 50 randomly chosen sites, each one metre square, along the coastline in country B. The results are summarised as follows.

$$\Sigma x = 1752 \qquad \Sigma x^2 = 55\,500 \qquad \Sigma y = 1220 \qquad \Sigma y^2 = 33\,500$$

Find a 95% confidence interval for the difference between the mean number of sea shells, per square metre, on the coastlines in country A and in country B. [9]

9231/23/M/J/13 Q8

Q28.

A random sample of 10 observations of a normally distributed random variable X gave the following summarised data, where $\bar{x}$ denotes the sample mean.

$$\Sigma x = 70.4 \qquad \Sigma (x - \bar{x})^2 = 8.48$$

Test, at the 10% significance level, whether the population mean of X is less than 7.5. [7]

9231/21/O/N/13 Q7

Q29.

A factory produces bottles of spring water. The manager decides to assess the performance of the two machines that are used to fill the bottles with water. He selects a random sample of 60 bottles filled by the first machine X and a random sample of 80 bottles filled by the second machine Y. The volumes of water, x and y, measured in appropriate units, are summarised as follows.

$$\Sigma x = 58.2 \qquad \Sigma x^2 = 85.8 \qquad \Sigma y = 97.6 \qquad \Sigma y^2 = 188.6$$

A test at the α% significance level shows that the mean volume of water in bottles filled by machine X is less than the mean volume of water in bottles filled by machine Y. Find the set of possible values of α. [12]

9231/21/O/N/13 Q11

Q30.

Two independent random variables X and Y have distributions with the same variance σ^2. Random samples of n observations of X and $2n$ observations of Y are taken and the results are summarised by

$$\Sigma x = 10.0, \qquad \Sigma x^2 = 25.0, \qquad \Sigma y = 15.0, \qquad \Sigma y^2 = 43.5.$$

Given that the pooled estimate of σ^2 is 2, find the value of n. [7]

9231/23/O/N/13 Q7

2. Inference Using Normal and t-Distributions

Q31.

A random sample of 9 observations of a normally distributed random variable X gave the following summarised data.

$$\Sigma x = 94.5 \qquad \Sigma x^2 = 993.6$$

Test, at the 5% significance level, whether the population mean of X is 10.2. [7]

Calculate a 90% confidence interval for the population mean of X. [3]

9231/23/O/N/13 Q9

Q32.

Weekly expenses claimed by employees at two different branches, A and B, of a large company are being compared. Expenses claimed by an employee at branch A and by an employee at branch B are denoted by $\$x$ and $\$y$ respectively. A random sample of 60 employees from branch A and a random sample of 50 employees from branch B give the following summarised data.

$$\Sigma x = 6060 \qquad \Sigma x^2 = 626\,220 \qquad \Sigma y = 4750 \qquad \Sigma y^2 = 464\,500$$

Using a 2% significance level, test whether, on average, employees from branch A claim the same as employees from branch B. [9]

9231/21/M/J/14 Q8

Q33.

The time taken for a randomly chosen student at College P to complete a particular puzzle has a normal distribution with mean μ minutes. The times, x minutes, are recorded for a random sample of 8 students chosen from the college. The results are summarised as follows.

$$\Sigma x = 42.8 \qquad \Sigma x^2 = 236.0$$

Find a 95% confidence interval for μ. [4]

A test is carried out on this sample data, at the 10% significance level. The test supports the claim that $\mu > k$. Find the greatest possible value of k. [3]

A random sample, of size 12, is taken from the students at College Q. Their times to complete the puzzle give a sample mean of 4.60 minutes and an unbiased variance estimate of 1.962 minutes^2. Use a 2-sample test at the 10% significance level to test whether the mean time for students at College Q to complete the puzzle is less than the mean time for students at College P to complete the puzzle. You should state any assumptions necessary for the test to be valid. [7]

9231/21/M/J/14 Q11

2. Inference Using Normal and t-Distributions

Q34.

The lengths of a random sample of eight fish of a certain species are measured, in cm, as follows.

17.3 15.8 18.2 15.6 16.0 18.8 15.3 15.0

Assuming that lengths are normally distributed,

(i) test, at the 10% significance level, whether the population mean length of fish of this species is greater than 15.8 cm, [7]

(ii) calculate a 95% confidence interval for the population mean length of fish of this species. [4]

9231/23/M/J/14 Q10

Q35.

A random sample of 50 observations of a random variable X and a random sample of 60 observations of a random variable Y are taken. The results for the sample means, $\bar{x}$ and $\bar{y}$, and the unbiased estimates for the population variances, s_x^2 and s_y^2, respectively, are as follows.

$$\bar{x} = 25.4 \qquad \bar{y} = 23.6 \qquad s_x^2 = 23.2 \qquad s_y^2 = 27.8$$

A test, at the α% significance level, of the null hypothesis that the population means of X and Y are equal against the alternative hypothesis that they are not equal is carried out. Given that the null hypothesis is not rejected, find the set of possible values of α. [5]

9231/21/O/N/14 Q6

Q36.

Fish of a certain species live in two separate lakes, A and B. A zoologist claims that the mean length of fish in A is greater than the mean length of fish in B. To test his claim, he catches a random sample of 8 fish from A and a random sample of 6 fish from B. The lengths of the 8 fish from A, in appropriate units, are as follows.

15.3 12.0 15.1 11.2 14.4 13.8 12.4 11.8

Assuming a normal distribution, find a 95% confidence interval for the mean length of fish in A. [5]

The lengths of the 6 fish from B, in the same units, are as follows.

15.0 10.7 13.6 12.4 11.6 12.6

Stating any assumptions that you make, test at the 5% significance level whether the mean length of fish in A is greater than the mean length of fish in B. [7]

Calculate a 95% confidence interval for the difference in the mean lengths of fish from A and from B. [2]

9231/21/O/N/14 Q11

2. Inference Using Normal and t-Distributions

Q37.

The independent random variables X and Y have distributions with the same variance σ^2. Random samples of N observations of X and 10 observations of Y are taken, and the results are summarised by

$$\Sigma x = 5, \qquad \Sigma x^2 = 11, \qquad \Sigma y = 10, \qquad \Sigma y^2 = 160.$$

These data give a pooled estimate of 12 for σ^2. Find N. [4]

9231/21/M/J/15 Q6

Q38.

A random sample of 8 sunflower plants is taken from the large number grown by a gardener, and the heights of the plants are measured. A 95% confidence interval for the population mean, μ metres, is calculated from the sample data as $1.17 < \mu < 2.03$. Given that the height of a sunflower plant is denoted by x metres, find the values of Σx and Σx^2 for this sample of 8 plants. [7]

9231/21/M/J/15 Q7

Q39.

A large number of long jumpers are competing in a national long jump competition. The distances, in metres, jumped by a random sample of 7 competitors are as follows.

6.25 7.01 5.74 6.89 7.24 5.64 6.52

Assuming that distances are normally distributed, test, at the 5% significance level, whether the mean distance jumped by long jumpers in this competition is greater than 6.2 metres. [7]

The distances jumped by another random sample of 8 long jumpers in this competition are recorded. Using the data from this sample of 8 long jumpers, a 95% confidence interval for the population mean, μ metres, is calculated as $5.89 < \mu < 6.75$. Find the unbiased estimates for the population mean and population variance used in this calculation. [5]

9231/23/M/J/15 Q8

Q40.

The times taken, in hours, by cyclists from two different clubs, A and B, to complete a 50 km time trial are being compared. The times taken by a cyclist from club A and by a cyclist from club B are denoted by t_A and t_B respectively. A random sample of 50 cyclists from A and a random sample of 60 cyclists from B give the following summarised data.

$$\Sigma t_A = 102.0 \qquad \Sigma t_A^2 = 215.18 \qquad \Sigma t_B = 129.0 \qquad \Sigma t_B^2 = 282.3$$

Using a 5% significance level, test whether, on average, cyclists from club A take less time to complete the time trial than cyclists from club B. [10]

A test at the α% significance level shows that there is evidence that the population mean time for cyclists from club B exceeds the population mean time for cyclists from club A by more than 0.05 hours. Find the set of possible values of α. [4]

9231/23/M/J/15 Q10

2. Inference Using Normal and t-Distributions

Q41.

A random sample of 10 observations of a normal variable X gave the following summarised data, where $\bar{x}$ is the sample mean.

$$\Sigma x = 222.8 \qquad \Sigma(x - \bar{x})^2 = 4.12$$

Find a 95% confidence interval for the population mean. [5]

9231/21/O/N/15 Q5

Q42.

A random sample of 8 students is chosen from those sitting examinations in both Mathematics and French. Their marks in Mathematics, x, and in French, y, are summarised as follows.

$$\Sigma x = 472 \qquad \Sigma x^2 = 29\,950 \qquad \Sigma y = 400 \qquad \Sigma y^2 = 21\,226 \qquad \Sigma xy = 24\,879$$

Another student scored 72 marks in the Mathematics examination but was unable to sit the French examination. Estimate the mark that this student would have obtained in the French examination. [5]

Test, at the 5% significance level, whether there is non-zero correlation between marks in Mathematics and marks in French. [6]

9231/21/O/N/15 Q9

Q43.

A farmer A grows two types of potato plants, Royal and Majestic. A random sample of 10 Royal plants is taken and the potatoes from each plant are weighed. The total mass of potatoes on a plant is x kg. The data are summarised as follows.

$$\Sigma x = 42.0 \qquad \Sigma x^2 = 180.0$$

A random sample of 12 Majestic plants is taken. The total mass of potatoes on a plant is y kg. The data are summarised as follows.

$$\Sigma y = 57.6 \qquad \Sigma y^2 = 281.5$$

Test, at the 5% significance level, whether the population mean mass of potatoes from Royal plants is the same as the population mean mass of potatoes from Majestic plants. You may assume that both distributions are normal and you should state any additional assumption that you make. [9]

A neighbouring farmer B grows Crown potato plants. His plants produce 3.8 kg of potatoes per plant, on average. Farmer A claims that her Royal plants produce a higher mean mass of potatoes than Farmer B's Crown plants. Test, at the 5% significance level, whether Farmer A's claim is justified. [5]

9231/21/O/N/15 Q10

2. Inference Using Normal and t-Distributions

Q44.

A random sample of 9 observations of a normal variable X is taken. The results are summarised as follows.

$$\Sigma x = 24.6 \qquad \Sigma x^2 = 68.5$$

Test, at the 5% significance level, whether the population mean is greater than 2.5. [8]

9231/21/M/J/16 Q7

Q45.

Petra is studying a particular species of bird. She takes a random sample of 12 birds from nature reserve A and measures the wing span, x cm, for each bird. She then calculates a 95% confidence interval for the population mean wing span, μ cm, for birds of this species, assuming that wing spans are normally distributed. Later, she is not able to find the summary of the results for the sample, but she knows that the 95% confidence interval is $25.17 \leqslant \mu \leqslant 26.83$. Find the values of Σx and Σx^2 for this sample. [6]

Petra also measures the wing spans of a random sample of 7 birds from nature reserve B. Their wing spans, y cm, are as follows.

23.2 22.4 27.6 25.3 28.4 26.5 23.6

She believes that the mean wing span of birds found in nature reserve A is greater than the mean wing span of birds found in nature reserve B. Assuming that this second sample also comes from a normal distribution, with variance the same as the first distribution, test, at the 10% significance level, whether there is evidence to support Petra's belief. [8]

9231/21/M/J/16 Q11

Q46.

The annual salaries of workers at two factories, A and B, are to be compared. The salaries, in tens of thousands of dollars, at A and B, are denoted by x and y respectively. For a random sample of 40 workers in factory A and a random sample of 50 workers in factory B the results are as follows.

$$\Sigma x = 256.0 \qquad \Sigma x^2 = 1910.8 \qquad \Sigma y = 382.9 \qquad \Sigma y^2 = 3148.8$$

The population mean salaries for A and B are denoted by μ_A and μ_B respectively. The population variances for salaries at A and B cannot be assumed to be equal. Test, at the 1% significance level, whether μ_B is greater than μ_A. [10]

The width of an α% confidence interval for $\mu_B - \mu_A$ is found to be 1.82. Find the value of α. [4]

9231/23/M/J/16 Q11

2. Inference Using Normal and t-Distributions

Q47.

A random sample of 8 observations of a normal random variable X has mean $\bar{x}$, where

$$\bar{x} = 6.246 \quad \text{and} \quad \Sigma(x - \bar{x})^2 = 0.784.$$

Test, at the 5% significance level, whether the population mean of X is less than 6.44. [7]

9231/21/O/N/16 Q6

Q48.

The amounts spent on the weekly food shopping by families in the big city P and the small town Q are to be compared. The amounts spent, in dollars, in P and Q are denoted by x and y respectively. For a random sample of 60 families in P and a random sample of 50 families in Q, the amounts are summarised as follows.

$$\Sigma x = 9600 \qquad \Sigma x^2 = 1\,560\,000 \qquad \Sigma y = 7200 \qquad \Sigma y^2 = 1\,052\,500$$

Assuming a common population variance, find

(i) a pooled estimate for the population variance, [4]

(ii) a 95% confidence interval for the difference in the population means in P and Q. [5]

9231/21/O/N/16 Q8

Q49.

A farmer grows a particular type of fruit tree. On average, the mass of fruit produced per tree has been 6.2 kg. He has developed a new kind of soil and claims that the mean mass of fruit produced per tree when growing in this new soil has increased. A random sample of 10 trees grown in the new soil is chosen. The masses, x kg, of fruit produced are summarised as follows.

$$\Sigma x = 72.0 \qquad \Sigma x^2 = 542.0$$

Test at the 5% significance level whether the farmer's claim is justified, assuming a normal distribution. [7]

9231/21/M/J/17 Q7

Q50.

Two fish farmers X and Y produce a particular type of fish. Farmer X chooses a random sample of 8 of his fish and records the masses, x kg, as follows.

1.2 1.4 0.8 2.1 1.8 2.6 1.5 2.0

Farmer Y chooses a random sample of 10 of his fish and summarises the masses, y kg, as follows.

$$\Sigma y = 20.2 \qquad \Sigma y^2 = 44.6$$

You should assume that both distributions are normal with equal variances. Test at the 10% significance level whether the mean mass of fish produced by farmer X differs from the mean mass of fish produced by farmer Y. [10]

9231/21/M/J/17 Q9

2. Inference Using Normal and t-Distributions

Q51.

The independent variables X and Y have distributions with the same variance σ^2. Random samples of N observations of X and $2N$ observations of Y are taken, and the results are summarised by

$$\Sigma x = 4, \quad \Sigma x^2 = 10, \quad \Sigma y = 8, \quad \Sigma y^2 = 102.$$

These data give a pooled estimate of 10 for σ^2. Find N. [5]

9231/23/M/J/17 Q6

Q52.

The number, x, of beech trees was counted in each of 50 randomly chosen regions of equal size in beech forests in country A. The number, y, of beech trees was counted in each of 40 randomly chosen regions of the same equal size in beech forests in country B. The results are summarised as follows.

$$\Sigma x = 1416 \qquad \Sigma x^2 = 41\,100 \qquad \Sigma y = 888 \qquad \Sigma y^2 = 20\,140$$

Find a 95% confidence interval for the difference between the mean number of beech trees in regions of this size in country A and in country B. [9]

9231/23/M/J/17 Q8

Q53.

A factory produces bottles of an energy juice. Two different machines are used to fill empty bottles with the juice. The manager chooses a random sample of 50 bottles filled by machine X and a random sample of 60 bottles filled by machine Y. The volumes of juice, x and y respectively, measured in appropriate units, are summarised by

$$\Sigma x = 45.5, \quad \Sigma(x - \bar{x})^2 = 19.56, \quad \Sigma y = 72.3, \quad \Sigma(y - \bar{y})^2 = 30.25,$$

where $\bar{x}$ and $\bar{y}$ are the sample means of the volume of juice in the bottles filled by X and Y respectively.

(i) Find a 90% confidence interval for the difference between the mean volume of juice in bottles filled by machine X and the mean volume of juice in bottles filled by machine Y. [7]

A test at the α% significance level does not provide evidence that there is any difference in the means of the volume of juice in bottles filled by machine X and the volume of juice in bottles filled by machine Y.

(ii) Find the set of possible values of α. [6]

9231/21/O/N/17 Q10

Q54.

A large number of athletes are taking part in a competition. The masses, in kg, of a random sample of 7 athletes are as follows.

98.1 105.0 92.2 89.8 99.9 95.4 101.2

Assuming that masses are normally distributed, test, at the 10% significance level, whether the mean mass of athletes in this competition is equal to 94 kg. [7]

9231/21/M/J/18 Q7

2. Inference Using Normal and t-Distributions

Q55.

The times taken to run 400 metres by students at two large colleges P and Q are being compared. There is no evidence that the population variances are equal. The time taken by a student at college P and the time taken by a student at college Q are denoted by x seconds and y seconds respectively. A random sample of 50 students from college P and a random sample of 60 students from college Q give the following summarised data.

$$\Sigma x = 2620 \qquad \Sigma x^2 = 138\,200 \qquad \Sigma y = 3060 \qquad \Sigma y^2 = 157\,000$$

(i) Using a 10% significance level, test whether, on average, students from college P take longer to run 400 metres than students from college Q. [9]

(ii) Find a 90% confidence interval for the difference in the mean times taken to run 400 metres by students from colleges P and Q. [3]

9231/21/M/J/18 Q10

Q56.

During the summer months, all members of a large swimming club take part in intensive training. The times taken to swim 50 metres at the beginning of the summer and at the end of the summer are recorded for each member of the club. The time taken, in seconds, at the beginning of the summer is denoted by x and the time taken at the end of the summer is denoted by y. For a random sample of 9 members the results are shown in the following table.

Member	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
x	38.5	40.2	32.3	35.1	36.2	41.4	32.0	38.2	38.2
y	37.4	38.1	31.6	34.7	34.2	38.6	31.8	36.3	36.8

The swimming coach believes that, on average, the time taken by a swimmer to swim 50 metres will decrease by more than one second as a result of the intensive training.

(i) Stating suitable hypotheses and assuming a normal distribution, test the coach's belief at the 10% significance level. [8]

(ii) Find a 95% confidence interval for the population mean time taken to swim 50 metres after the intensive training, assuming a normal distribution. [4]

9231/23/M/J/18 Q10

2. Inference Using Normal and t-Distributions

Q57.

The weekly salaries of employees at two large electronics companies, A and B, are being compared. The weekly salaries of an employee from company A and an employee from company B are denoted by $\$x$ and $\$y$ respectively. A random sample of 50 employees from company A and a random sample of 40 employees from company B give the following summarised data.

$$\Sigma x = 5120 \qquad \Sigma x^2 = 531\,000 \qquad \Sigma y = 3760 \qquad \Sigma y^2 = 375\,135$$

(i) The population mean salaries of employees from companies A and B are denoted by $\$\mu_A$ and $\$\mu_B$ respectively. Using a 5% significance level, test the null hypothesis $\mu_A = \mu_B$ against the alternative hypothesis $\mu_A \neq \mu_B$. [8]

(ii) State, with a reason, whether any assumptions about the distributions of employees' salaries are needed for the test in part **(i)**. [1]

9231/21/O/N/18 Q8

Q58.

There are a large number of students at a particular college. The heights, in metres, of a random sample of 8 students are as follows.

1.75 1.72 1.62 1.70 1.82 1.75 1.68 1.84

You may assume that heights of students are normally distributed.

(i) Test, at the 5% significance level, whether the population mean height of students at this college is greater than 1.70 metres. [7]

(ii) Find a 95% confidence interval for the population mean height of students at this college. [3]

9231/21/O/N/18 Q9

Q59.

The heights, in metres, of a random sample of 8 trees of a particular type are as follows.

14.2 11.3 10.8 8.4 12.8 11.5 12.1 9.2

Assuming that heights of trees of this type are normally distributed, calculate a 95% confidence interval for the mean height of trees of this type. [6]

9231/22/O/N/18 Q6

2. Inference Using Normal and t-Distributions

Q60.

In a particular country, large numbers of ducks live on lakes A and B. The mass, in kg, of a duck on lake A is denoted by x and the mass, in kg, of a duck on lake B is denoted by y. A random sample of 8 ducks is taken from lake A and a random sample of 10 ducks is taken from lake B. Their masses are summarised as follows.

$$\Sigma x = 10.56 \qquad \Sigma x^2 = 14.1775 \qquad \Sigma y = 12.39 \qquad \Sigma y^2 = 15.894$$

A scientist claims that ducks on lake A are heavier on average than ducks on lake B.

(i) Test, at the 10% significance level, whether the scientist's claim is justified. You should assume that both distributions are normal and that their variances are equal. [9]

A second random sample of 8 ducks is taken from lake A and their masses are summarised as

$$\Sigma x = 10.24 \qquad \text{and} \qquad \Sigma(x - \bar{x})^2 = 0.294,$$

where $\bar{x}$ is the sample mean. The scientist now claims that the population mean mass of ducks on lake A is greater than p kg. A test of this claim is carried out at the 10% significance level, using only this second sample from lake A. This test supports the scientist's claim.

(ii) Find the greatest possible value of p. [5]

9231/22/O/N/18 Q11

Q61.

A large number of runners are attending a summer training camp. A random sample of 6 runners is chosen and their times to run 1500 m at the beginning of the camp and at the end of the camp are recorded. Their times, in minutes, are shown in the following table.

Runner	*A*	*B*	*C*	*D*	*E*	*F*
Time at beginning of camp	3.82	3.62	3.55	3.71	3.75	3.92
Time at end of camp	3.72	3.55	3.52	3.68	3.54	3.73

The organiser of the training camp claims that a runner's time will improve by more than 0.05 minutes between the beginning and end of the camp. Assuming that differences in time over the two runs are normally distributed, test at the 10% significance level whether the organiser's claim is justified. [8]

9231/21/M/J/19 Q8

2. Inference Using Normal and t-Distributions

Q62.

A farmer grows two different types of cherries, Type A and Type B. He assumes that the masses of each type are normally distributed. He chooses a random sample of 8 cherries of Type A. He finds that the sample mean mass is 15.1 g and that a 95% confidence interval for the population mean mass, μ g, is $13.5 \leqslant \mu \leqslant 16.7$.

(i) Find an unbiased estimate for the population variance of the masses of cherries of Type A. [3]

The farmer now chooses a random sample of 6 cherries of Type B and records their masses as follows.

12.2 13.3 16.4 14.0 13.9 15.4

(ii) Test at the 5% significance level whether the mean mass of cherries of Type B is less than the mean mass of cherries of Type A. You should assume that the population variances for the two types of cherry are equal. [9]

9231/21/M/J/19 Q11b

Q63.

A farmer grows large amounts of a certain crop. On average, the yield per plant has been 0.75 kg. The farmer has improved the soil in which the crop grows, and she claims that the yield per plant has increased. A random sample of 10 plants grown in the improved soil is chosen. The yields, x kg, are summarised as follows.

$$\Sigma x = 7.85 \qquad \Sigma x^2 = 6.19$$

(i) Test at the 5% significance level whether the farmer's claim is justified, assuming a normal distribution. [7]

(ii) Find a 95% confidence interval for the population mean yield for plants grown in the new soil. [3]

9231/23/M/J/19 Q9

Q64.

A company produces packets of sweets. Two different machines, A and B, are used to fill the packets. The manager decides to assess the performance of the two machines. He selects a random sample of 50 packets filled by machine A and a random sample of 60 packets filled by machine B. The masses of sweets, x kg, in packets filled by machine A and the masses of sweets, y kg, in packets filled by machine B are summarised as follows.

$$\Sigma x = 22.4 \qquad \Sigma x^2 = 10.1 \qquad \Sigma y = 28.8 \qquad \Sigma y^2 = 16.3$$

A test at the α% significance level provides evidence that the mean mass of sweets in packets filled by machine A is less than the mean mass of sweets in packets filled by machine B. Find the set of possible values of α. [12]

9231/23/M/J/19 Q11

2. Inference Using Normal and t-Distributions

Q65.

A random sample of 9 members is taken from the large number of members of a sports club, and their heights are measured. The heights of all the members of the club are assumed to be normally distributed. A 95% confidence interval for the population mean height, μ metres, is calculated from the data as $1.65 \leqslant \mu \leqslant 1.85$.

(i) Find an unbiased estimate for the population variance. [3]

(ii) Denoting the height of a member of the club by x metres, find Σx^2 for this sample of 9 members. [4]

9231/21/O/N/19 Q6

Q66.

A random sample of 8 elephants from region A is taken and their weights, x tonnes, are recorded. (1 tonne = 1000 kg.) The results are summarised as follows.

$$\Sigma x = 32.4 \qquad \Sigma x^2 = 131.82$$

A random sample of 10 elephants from region B is taken. Their weights give a sample mean of 3.78 tonnes and an unbiased variance estimate of 0.1555 tonnes^2. The distributions of the weights of elephants in regions A and B are both assumed to be normal with the same population variance. Test at the 10% significance level whether the mean weight of elephants in region A is the same as the mean weight of elephants in region B. [9]

9231/21/O/N/19 Q8

Q67.

A company has two different machines, X and Y, each of which fills empty cups with coffee. The manager is investigating the volumes of coffee, x and y, measured in appropriate units, in the cups filled by machines X and Y respectively. She chooses a random sample of 50 cups filled by machine X and a random sample of 40 cups filled by machine Y. The volumes are summarised as follows.

$$\Sigma x = 15.2 \quad \Sigma x^2 = 5.1 \quad \Sigma y = 13.4 \quad \Sigma y^2 = 4.8$$

The manager claims that there is no difference between the mean volume of coffee in cups filled by machine X and the mean volume of coffee in cups filled by machine Y.

Test the manager's claim at the 10% significance level. [9]

9231/41/M/J/20 Q4

2. Inference Using Normal and t-Distributions

Q68.

A large number of children are competing in a throwing competition. The distances, in metres, thrown by a random sample of 8 children are as follows.

19.8 22.1 24.4 21.5 20.8 26.3 23.7 25.0

(a) Assuming that distances are normally distributed, test, at the 5% significance level, whether the population mean distance thrown is more than 22.0 metres. [7]

(b) Find a 95% confidence interval for the population mean distance thrown. [3]

9231/41/M/J/20 Q5

Q69.

A random sample of 40 observations of a random variable X and a random sample of 50 observations of a random variable Y are taken. The resulting values for the sample means, $\bar{x}$ and $\bar{y}$, and the unbiased estimates, s_x^2 and s_y^2, for the population variances are as follows.

$$\bar{x} = 24.4 \quad \bar{y} = 17.2 \quad s_x^2 = 10.2 \quad s_y^2 = 11.1$$

Find a 90% confidence interval for the difference between the population means of X and Y. [5]

9231/43/M/J/20 Q2

Q70.

Students at two colleges, A and B, are competing in a computer games challenge.

(a) The time taken for a randomly chosen student from college A to complete the challenge has a normal distribution with mean μ minutes. The times taken, x minutes, are recorded for a random sample of 10 students chosen from college A. The results are summarised as follows.

$$\Sigma x = 828 \quad \Sigma x^2 = 68622$$

A test is carried out on the data at the 5% significance level and the result supports the claim that $\mu > k$.

Find the greatest possible value of k. [4]

(b) A random sample of 8 students is chosen from college B. Their times to complete the same challenge give a sample mean of 79.8 minutes and an unbiased variance estimate of 9.966 minutes^2.

Use a 2-sample test at the 5% significance level to test whether the mean time for students at college B to complete the challenge is the same as the mean time for students at college A to complete the challenge. You should assume that the two distributions are normal and have the same population variance. [7]

9231/43/M/J/20 Q5

2. Inference Using Normal and t-Distributions

Q71.

Kayla is investigating the lengths of the leaves of a certain type of tree found in two forests X and Y. She chooses a random sample of 40 leaves of this type from forest X and records their lengths, x cm. She also records the lengths, y cm, for a random sample of 60 leaves of this type from forest Y. Her results are summarised as follows.

$$\Sigma x = 242.0 \qquad \Sigma x^2 = 1587.0 \qquad \Sigma y = 373.2 \qquad \Sigma y^2 = 2532.6$$

Find a 90% confidence interval for the difference between the population mean lengths of leaves in forests X and Y. [7]

9231/41/O/N/20 Q1

Q72.

Members of the Sprints athletics club have been taking part in an intense training scheme, aimed at reducing their times taken to run 400 m. For a random sample of 9 athletes from the club, the times taken, in seconds, before and after the training scheme are given in the following table.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Time before	48.8	48.2	50.3	49.6	49.4	48.9	47.6	50.3	48.4
Time after	47.9	47.8	49.6	49.1	49.6	48.9	47.7	49.1	48.1

The organiser of the training scheme claims that on average an athlete's time will be reduced by at least 0.3 seconds.

Test at the 10% significance level whether the organiser's claim is justified, stating any assumption that you make. [8]

9231/41/O/N/20 Q4

Q73.

The heights of the members of a large sports club are normally distributed. A random sample of 11 members of the club is chosen and their heights, x cm, are measured. The results are summarised as follows, where $\bar{x}$ denotes the sample mean of x.

$$\bar{x} = 176.2 \qquad \Sigma(x-\bar{x})^2 = 313.1$$

Test, at the 5% significance level, the null hypothesis that the population mean height for members of this club is equal to 172.5 cm against the alternative hypothesis that the mean differs from 172.5 cm. [5]

9231/42/O/N/20 Q1

2. Inference Using Normal and t-Distributions

Q74.

Nassa is researching the lengths of a particular type of snake in two countries, A and B.

(a) He takes a random sample of 10 snakes of this type from country A and measures the length, x m, of each snake. He then calculates a 90% confidence interval for the population mean length, μ m, for snakes of this type, assuming that snake lengths have a normal distribution. This confidence interval is $3.36 \leqslant \mu \leqslant 4.22$.

Find the sample mean and an unbiased estimate for the population variance. [4]

(b) Nassa also measures the lengths, y m, of a random sample of 8 snakes of this type taken from country B. His results are summarised as follows.

$$\Sigma y = 27.86 \qquad \Sigma y^2 = 98.02$$

Nassa claims that the mean length of snakes of this type in country B is less than the mean length of snakes of this type in country A. Nassa assumes that his sample from country B also comes from a normal distribution, with the same variance as the distribution from country A.

Test at the 10% significance level whether there is evidence to support Nassa's claim. [8]

9231/42/O/N/20 Q6

Q75.

A random sample of 7 observations of a variable X are as follows.

8.26 7.78 7.92 8.04 8.27 7.95 8.34

The population mean of X is μ.

(a) Test, at the 10% significance level, the null hypothesis $\mu = 8.22$ against the alternative hypothesis $\mu < 8.22$. [6]

(b) State an assumption necessary for the test in part **(a)** to be valid. [1]

9231/41/M/J/21 Q1

Q76.

A scientist is investigating the lengths of the leaves of birch trees in different regions. He takes a random sample of 50 leaves from birch trees in region A and a random sample of 60 leaves from birch trees in region B. He records their lengths in cm, x and y, respectively. His results are summarised as follows.

$$\Sigma x = 282 \qquad \Sigma x^2 = 1596 \qquad \Sigma y = 328 \qquad \Sigma y^2 = 1808$$

The population mean lengths of leaves from birch trees in regions A and B are μ_A cm and μ_B cm respectively.

Carry out a test at the 5% significance level to test the null hypothesis $\mu_A = \mu_B$ against the alternative hypothesis $\mu_A \neq \mu_B$. [8]

9231/41/M/J/21 Q4

2. Inference Using Normal and t-Distributions

Q77.

Farmer A grows apples of a certain variety. Each tree produces 14.8 kg of apples, on average, per year. Farmer B grows apples of the same variety and claims that his apple trees produce a higher mass of apples per year than Farmer A's trees. The masses of apples from Farmer B's trees may be assumed to be normally distributed.

A random sample of 10 trees from Farmer B is chosen. The masses, x kg, of apples produced in a year are summarised as follows.

$$\Sigma x = 152.0 \qquad \Sigma x^2 = 2313.0$$

Test, at the 5% significance level, whether Farmer B's claim is justified. [6]

9231/43/M/J/21 Q1

Q78.

The heights, x m, of a random sample of 50 adult males from country A were recorded. The heights, y m, of a random sample of 40 adult males from country B were also recorded. The results are summarised as follows.

$$\Sigma x = 89.0 \qquad \Sigma x^2 = 159.4 \qquad \Sigma y = 67.2 \qquad \Sigma y^2 = 113.1$$

Find a 95% confidence interval for the difference between the mean heights of adult males from country A and adult males from country B. [8]

9231/43/M/J/21 Q3

Q79.

The times taken for students at a college to run 200 m have a normal distribution with mean μ s. The times, x s, are recorded for a random sample of 10 students from the college. The results are summarised as follows, where $\bar{x}$ is the sample mean.

$$\bar{x} = 25.6 \qquad \Sigma(x-\bar{x})^2 = 78.5$$

(a) Find a 90% confidence interval for μ. [4]

A test of the null hypothesis $\mu = k$ is carried out on this sample, using a 10% significance level. The test does not support the alternative hypothesis $\mu < k$.

(b) Find the greatest possible value of k. [3]

9231/41/O/N/21 Q1

2. Inference Using Normal and t-Distributions

Q80.

Manet has developed a new training course to help athletes improve their time taken to run 800 m. Manet claims that his course will decrease an athlete's time by more than 2 s on average. For a random sample of 10 athletes the times taken, in seconds, before and after the course are given in the following table.

Athlete	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Before	150	146	131	135	126	142	130	129	137	134
After	145	138	129	135	122	135	132	128	127	137

Use a *t*-test, at the 5% significance level, to test whether Manet's claim is justified, stating any assumption that you make. [8]

9231/41/O/N/21 Q4

Q81.

The number, x, of pine trees was counted in each of 40 randomly chosen regions of equal size in country A. The number, y, of pine trees was counted in each of 60 randomly chosen regions of the same equal size in country B. The results are summarised as follows.

$$\Sigma x = 752 \qquad \Sigma x^2 = 14320 \qquad \Sigma y = 1548 \qquad \Sigma y^2 = 40200$$

Find a 95% confidence interval for the difference between the mean number of pine trees in regions of this size in countries A and B. [7]

9231/42/O/N/21 Q1

Q82.

A scientist is investigating the masses of a particular type of fish found in lakes A and B. He chooses a random sample of 10 fish of this type from lake A and records their masses, x kg, as follows.

2.1 1.8 0.9 3.0 2.4 2.6 1.8 2.2 1.9 2.5

The scientist also chooses a random sample of 12 fish of this type from lake B, but he only has a summary of their masses, y kg, as follows.

$$\Sigma y = 24.48 \qquad \Sigma y^2 = 53.75$$

Test at the 10% significance level whether the mean mass of fish of this type in lake A is greater than the mean mass of fish of this type in lake B. You should state any assumptions that you need to make for the test to be valid. [10]

9231/42/O/N/21 Q6

2. Inference Using Normal and t-Distributions

Q83.

A manager is investigating the times taken by employees to complete a particular task as a result of the introduction of new technology. He claims that the mean time taken to complete the task is reduced by more than 0.4 minutes. He chooses a random sample of 10 employees. The times taken, in minutes, before and after the introduction of the new technology are recorded in the table.

Employee	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Time before new technology	10.2	9.8	12.4	11.6	10.8	11.2	14.6	10.6	12.3	11.0
Time after new technology	9.6	8.5	12.4	10.9	10.2	10.6	12.8	10.8	12.5	10.6

(a) Test at the 10% significance level whether the manager's claim is justified. [7]

(b) State an assumption that is necessary for this test to be valid. [1]

9231/41/M/J/22 Q1

Q84.

Raman is researching the heights of male giraffes in a particular region. Raman assumes that the heights of male giraffes in this region are normally distributed. He takes a random sample of 8 male giraffes from the region and measures the height, in metres, of each giraffe. These heights are as follows.

5.2 5.8 4.9 6.1 5.5 5.9 5.4 5.6

(a) Find a 90% confidence interval for the population mean height of male giraffes in this region. [5]

Raman claims that the population mean height of male giraffes in the region is less than 5.9 metres.

(b) Test at the 2.5% significance level whether this sample provides sufficient evidence to support Raman's claim. [4]

9231/41/M/J/22 Q5

Q85.

The times taken by members of a large quiz club to complete a challenge have a normal distribution with mean μ minutes. The times, x minutes, are recorded for a random sample of 8 members of the club. The results are summarised as follows, where $\bar{x}$ is the sample mean.

$$\bar{x} = 33.8 \qquad \Sigma(x-\bar{x})^2 = 94.5$$

Find a 95% confidence interval for μ. [4]

9231/43/M/J/22 Q1

2. Inference Using Normal and t-Distributions

Q86.

A company has two machines, A and B, which independently fill small bottles with a liquid. The volumes of liquid per bottle, in suitable units, filled by machines A and B are denoted by x and y respectively. A scientist at the company takes a random sample of 40 bottles filled by machine A and a random sample of 50 bottles filled by machine B. The results are summarised as follows.

$$\Sigma x = 1120 \qquad \Sigma x^2 = 31400 \qquad \Sigma y = 1370 \qquad \Sigma y^2 = 37600$$

The population means of the volumes of liquid in the bottles filled by machines A and B are denoted by μ_A and μ_B.

(a) Test at the 2% significance level whether there is any difference between μ_A and μ_B. [8]

(b) Find the set of values of α for which there would be evidence at the α% significance level that $\mu_A - \mu_B$ is greater than 0.25. [4]

9231/43/M/J/22 Q6

Q87.

Jasmine is researching the heights of pine trees in forests in two regions A and B. She chooses a random sample of 50 pine trees in region A and records their heights, x m. She also chooses a random sample of 60 pine trees in region B and records their heights, y m. Her results are summarised as follows.

$$\Sigma x = 1625 \qquad \Sigma x^2 = 53200 \qquad \Sigma y = 1854 \qquad \Sigma y^2 = 57900$$

Find a 95% confidence interval for the difference between the population mean heights of pine trees in regions A and B. [7]

9231/41/O/N/22 Q1

Q88.

A company manufactures copper pipes. The pipes are produced by two different machines, A and B. An inspector claims that the mean diameter of the pipes produced by machine A is greater than the mean diameter of the pipes produced by machine B. He takes a random sample of 12 pipes produced by machine A and measures their diameters, x cm. His results are summarised as follows.

$$\Sigma x = 6.24 \qquad \Sigma x^2 = 3.26$$

He also takes a random sample of 10 pipes produced by machine B and measures their diameters in cm. His results are as follows.

0.48 0.53 0.47 0.54 0.54 0.55 0.46 0.55 0.50 0.48

The diameters of the pipes produced by each machine are assumed to be normally distributed with equal population variances.

Test at the 2.5% significance level whether the data supports the inspector's claim. [9]

9231/41/O/N/22 Q6

2. Inference Using Normal and t-Distributions

Q89.

A basketball club has a large number of players. The heights, x m, of a random sample of 10 of these players are measured. A 90% confidence interval for the population mean height, μ m, of players in this club is calculated. It is assumed that heights are normally distributed. The confidence interval is $1.78 \leqslant \mu \leqslant 2.02$.

Find the values of $\sum x$ and $\sum x^2$ for this sample. [6]

9231/42/O/N/22 Q1

Q90.

A scientist is investigating the masses of birds of a certain species in country X and country Y. She takes a random sample of 50 birds of this species from country X and a random sample of 80 birds of this species from country Y. She records their masses in kg, x and y, respectively. Her results are summarised as follows.

$$\sum x = 75.5 \qquad \sum x^2 = 115.2 \qquad \sum y = 116.8 \qquad \sum y^2 = 172.6$$

The population mean masses of these birds in countries X and Y are μ_x kg and μ_y kg respectively.

Test, at the 5% significance level, the null hypothesis $\mu_x = \mu_y$ against the alternative hypothesis $\mu_x > \mu_y$. State your conclusion in the context of the question. [8]

9231/42/O/N/22 Q3

Solutions

Q1.

7	Use valid formula for C.I. with any t (or z) value:	$\bar{x} \pm t\sqrt{(s^2/n)}$ with $n = 10$ (*or* 9)	M1		
	Find mean of sample:	$\bar{x} = 28.54$	A1		
	Find population variance consistent with n (to 2 dp):	$s^2 = 49.85$ *or* 7.06^2 $(n = 10)$			
		or 44.865 *or* 6.698^2 $(n = 9)$	A1		
	Use correct tabular t value:	$t_{9,\,0.975} = 2.262$ (to 3 sf)	*A1		
	Evaluate C.I. (dep *A1):	28.54 ± 5.05 *or* [23.5, 33.6]	A1	5	
	Use valid formula for width (or CI) with any t value:	$2t\sqrt{((448.65 / 9) / 20)}$	M1		
	Use correct tabular t value:	$t_{19,\,0.975} = 2.093$ (to 3 sf)	*A1		
	Evaluate width of C.I. (dep *A1):	6.61	A1	3	**[8]**

Q2.

9	**(i)**	State hypotheses:	$H_0: \mu_B - \mu_A = 0$, $H_1: \mu_B - \mu_A > 0$	B1		
		State assumption (AEF):	Population of differences is normal	B1		
		Consider differences $B - A$ (*or* $A - B$):	–2 8 –8 11 29 15 38 23	M1		
		Estimate mean and population variance:	$\bar{x} = 114/8$ [= 14.25] *and*			
		(allow biased: 208.44 *or* 14.437^2)	$s^2 = (3292 - 8 \times 14.25^2) / 7$	M1		
			[= 238.21 *or* 15.434^2]			
		Calculate value of t:	$t = \bar{x} /(s/\sqrt{8}) = 2.61$[1]	*B1		
		Use correct tabular t value:	$t_{7,\,0.9} = 1.41$[5]	*B1		
		Correct conclusion (AEF, dep both *B1):	There is a reduction	B1	7	
	(ii)	Calculate t (M0 if inconsistent with s):	$t = (\bar{x} - 5) /(s/\sqrt{8}) = 1.69$[5]	M1*A1		
		Correct conclusion by comparing with				
		$t_{7,\,0.9} = 1.41$[5] (AEF, dep *A1):	There is a reduction > 5 units	B1	3	**[10]**

Q3.

6	Use standard formula for pooled estimate, e.g.:	$((128 - 15^2/5) + (980 - 36^2/10))/13$	M1		
	Award A1 for one term in numerator, e.g.:	$5 \times 16{\cdot}6$ *or* $10 \times 85{\cdot}04$ *or* 83 *or* 850·4			
		or $4 \times 20{\cdot}75$ *or* $9 \times 94{\cdot}5$	A1		
	Calculate value of pooled estimate:	71·8	A1	3	**[3]**

2. Inference Using Normal and t-Distributions

Q4.

7	**(i)** Find sample mean:	$\bar{x} = \frac{1}{2}(61{\cdot}21 + 64{\cdot}39) = 62{\cdot}8$	M1 A1		
	Use confidence interval formula:	$\bar{x} \pm ts/\sqrt{n}$ for any t	M1		
	Use correct tabular t:	$t_{24,099} = 2{\cdot}492$	A1		
	Calculate standard deviation:	$s = 1{\cdot}59 \times 5 / 2{\cdot}492 = 3{\cdot}19$	A1	5	
	(ii) State assumption (A.E.F.):	Population has normal distribution	B1	1	
	(iii) State valid reason (A.E.F.):	72 exceeds upper limit of interval	*B1		
	State conclusion (A.E.F., dep *B1):	Yes, it does reduce pulse rate	B1	2	**[8]**

Q5.

11 b	State hypotheses (A.E.F.):	H_0: $\mu_2 = \mu_1$, H_1: $\mu_2 > \mu_1$	B1		
	Calculate $\Sigma (x_i - \bar{x})^2$ (M1 for either)	8·24, 4·62[4]			
	or estimate variances:	0·168 *or* 0·165, 0·0784 *or* 0·0771	M1 A1 A1		
	Find s^2 (M0 if inconsistent denominators used):	$s^2 = 0{\cdot}168/50 + 0{\cdot}0784/60$			
		or $0{\cdot}165/49 + 0{\cdot}0771/59$			
		[= 0·00467]	*M1		
	Find value of z (dep *M1):	$z = (1803{\cdot}6/60 - 1492{\cdot}0/50) / s$	M1		
		$= (30{\cdot}06 - 29{\cdot}84)/0{\cdot}0683 = 3{\cdot}22$	*A1		
	S.R. Using pooled estimate of variance:	$s^2 = (8{\cdot}24 + 4{\cdot}62)/108 = 0{\cdot}119$	(M0)		
		$z = 0{\cdot}22 / s\sqrt{(1/50 + 1/60)} = 3{\cdot}33$	(B1)		
	Find tabular. value (to 2 dp):	$\Phi^{-1}(0{\cdot}98) = 2{\cdot}05$[4]	*B1		
	Compare values for conclusion (A1 dep *A1, *B1):	$\mu_2 > \mu_1$ (A.E.F., M1√ on values)	M1√ A1	10	
	Find limiting value of z (to 2 dp):	$z = (0{\cdot}22 - 0{\cdot}1)/s = 1{\cdot}756$	M1 A1		
	Find $\Phi(z)$ and hence values of α (to 1 dp):	$\Phi(z) = 0{\cdot}9604$, $\alpha \geq 3{\cdot}9$ *or* 4·0	M1 A1	4	
	$s^2 = 0{\cdot}119$ gives:	$z = 1{\cdot}816$, $\alpha \geq 3{\cdot}47$	(M1 M1)		**[14]**

2. Inference Using Normal and t-Distributions

Q6.

8	Consider differences $N - A$ (*or* $A - N$):	$-10\ \ 4\ \ 0\ \ -2\ \ -9\ \ 1\ \ 5\ \ -5\ \ -7$	M1		
	Estimate mean and population variance (to 3 sf):	$\bar{x} = -23/9$ [$= -2{\cdot}556$; allow +]			
		and $s^2 = (301 - 23^2/9)/8$			
		$= 30{\cdot}28$ *or* $5{\cdot}503^2$			
	(allow biased)	($26{\cdot}91$ *or* $5{\cdot}188^2$)	M1 A1		
	Use valid formula (M1 needs n consistent with s):	$\bar{x} \pm ts/\sqrt{n}$ for any t	M1		
	Use correct tabular t value:	$t_{8,\,0.95} = 1{\cdot}86$[0]	*B1		
	Evaluate confidence interval (dep *B1):	$-2{\cdot}56 \pm 3{\cdot}41$ *or* [$-5{\cdot}97$, $0{\cdot}86$]	*A1	6	
	Use valid argument consistent with interval:	0 lies in interval (A.E.F.)	M1		
	State conclusion (dep *A1 apart from rounding):	Yes, supports claim (A.E.F.)	A1	2	**[8]**

Q7.

11	State hypotheses:	H_0: $\mu_E = \mu_W$, H_1: $\mu_E \neq \mu_W$	B1	1	
OR	State assumption [A.E.F.]:	Two populations have Normal			
		distns. and common variance	B1		
	Estimate common variance:	$\sigma^2 = (5\times 0{\cdot}0231 + 4\times 0{\cdot}0195)/9$	M1		
		$= 0{\cdot}0215$ *or* $0{\cdot}1473^2$	A1		
	Use correct tabular value of t:	$t_{9,\,0.975} = 2{\cdot}26$[2]	B1		
	Formulate rejection region (with any t; allow >):	$\lvert \bar{x}_E - \bar{x}_W \rvert \geq t\sigma\sqrt{(1/6 + 1/5)}$	M1		
		$= 0{\cdot}201$	A1		
	Compare actual sample means with region:	$0{\cdot}253 > 0{\cdot}201$√			
	or compare calculated t with tabular t:	$2{\cdot}85 > 2{\cdot}26$[2]√	M1 A1		
	Consistent conclusion (A.E.F.; dep values above):	Mean acidity levels **do** differ	A1	9	
	State condition on a (with any t; allow > or =):	$\bar{x}_E - \bar{x}_W - a \geq t\sigma\sqrt{(1/6 + 1/5)}$	M1		
	Use correct tabular value of t:	$t_{9,\,0.95} = 1{\cdot}83$[3]	A1		
	Substitute to find largest value of a:	$a_{max} = 0{\cdot}253 - 0{\cdot}163 = 0{\cdot}09$ (2 dp)	M1 A1	4	**[14]**

2. Inference Using Normal and t-Distributions

Q8.

6	Find sample mean (2 dp required):	$\bar{x} = 60{\cdot}18 / 6 = 10{\cdot}03$	B1		
	Estimate population variance to 3 sf (allow biased here: 0·00107 *or* $0{\cdot}0327^2$):				
		$s^2 = (603{\cdot}6118 - 60{\cdot}18^2/6) / 5 = 0{\cdot}00128$ *or* $0{\cdot}0358^2$	B1		
	Use consistent formula for C.I. with any t value:	$\bar{x} \pm t\sqrt{(s^2/6)}$	M1√		
	Use of correct tabular value:	$t_{5,\,0.995} = 4{\cdot}03$[2]	*A1		
	C.I. correct to 2 dp in cm (dep *A1):	10·03 ± 0·06 *or* [9·97, 10·09]	A1	(5)	
	Find CI for mean circumference to 1 dp in cm:	$10{\cdot}03\pi \pm 0{\cdot}06\pi$			
		$= 31{\cdot}5 \pm 0{\cdot}2$ *or* [31·3, 31·7]	B1	(1)	**[6]**

Q9.

8	State if assumption necessary with valid reason:	No, since large samples (A.E.F.)	B1	(1)	
	State another valid assumption:	Samples must be random (A.E.F.)	B1		
	State hypotheses:	H_0: $\mu_c - \mu_f = 5$, H_1: $\mu_c - \mu_f < 5$	B1		
	Find z (A.E.F.):	$z = (19{\cdot}14 - 15{\cdot}36 - 5) /$			
		$\sqrt{(20{\cdot}54/75 + 9{\cdot}84/75)}$	M1 A1		
		$= -1{\cdot}22/\sqrt{(2\times15{\cdot}19/75)} = -1{\cdot}92$	*A1		
	Use correct tabular value of z:	$z_{0.95} = 1{\cdot}64$[5]	*B1		
	Consistent conclusion (A.E.F.; A1 dep *A1, *B1):	Does exceed by less than 5 kg.	M1√ A1	(8)	**[9]**

Q10.

8	State assumptions (A.E.F.):	Equal variances	B1		
		Normal populations	B1		
	Find difference in sample means to 2 dp, e.g.:	$\bar{x}_A - \bar{x}_B = 21{\cdot}417 - 25{\cdot}75 = -4{\cdot}33$	B1		
	Estimate common population variance:	$s^2 = (5629 - 257^2/12$			
		$+ 5359 - 206^2/8) / 18$			
		$= (124{\cdot}9 + 54{\cdot}5)/18$			
		$= 9{\cdot}968$ *or* $3{\cdot}157^2$ (3 sf)	M1 A1		
	Use of correct tabular t value:	$t_{18,\,0.975} = 2{\cdot}101$ (to 3 sf)	B1		
	Find confidence interval for e.g. $\mu_A - \mu_B$:	$\bar{x}_A - \bar{x}_B \pm ts\sqrt{(1/12 + 1/8)}$	M1		
	Evaluate:	$-4{\cdot}33 \pm 3{\cdot}03$ *or* [−7·36, −1·31]	A1*	8	
	State reason and conclusion (A.E.F.): (dep *A1 apart from rounding)	Interval does not include zero so statement not supported	B1	1	**[9]**

2. Inference Using Normal and t-Distributions

Q11.

10	State valid assumption (A.E.F.):	Popln. of diffs. has Normal distn.	B1		
	State hypotheses:	$H_0: \mu_2 - \mu_1 = 0$, $H_1: \mu_2 - \mu_1 > 0$	B1		
	Consider differences eg:	4·0 5·1 2·3 3·1 0 1·1 0·9 1·2	M1		
	Calculate sample mean:	$\bar{x} = 17{\cdot}7/8$ [= 2·2125]	M1		
	Estimate population variance:	$s^2 = (60{\cdot}37 - 17{\cdot}7^2/8) / 7$			
	(allow biased: 2·6511 *or* $1{\cdot}628^2$)	[= 3·0298 *or* $1{\cdot}741^2$]	M1		
	Calculate value of t (to 2 dp):	$t = \bar{x}/(s/\sqrt{8}) = 3{\cdot}60$ [*or* 3·59]	M1 *A1		
	Compare with correct tabular t value:	$t_{7,\,0.99} = 2{\cdot}998$ (to 2 dp)	*B1		
	Correct conclusion (AEF, dep *A1, *B1):	There is an increase	B1	9	
	Formulate inequality, with any tabular t value:	$(\bar{x} - w)/(s/\sqrt{8}) > t_{7,\,0.9}$	M1		
	Use correct tabular value (to 2 dp):	$t_{7,\,0.9} = 1{\cdot}415$	A1		
	Evaluate inequality for w:	$w < 1{\cdot}34$ [*or* ≤]	A1	3	
	S.R. Allow M1 A1 A0 if = or < used in inequality				**[12]**

Q12.

6	Use valid formula for C.I.:	$\bar{x}_B - \bar{x}_A \pm z\sigma\sqrt{(1/n_A + 1/n_B)}$	M2		
		$= 112 - 109 \pm z\ 15\ \sqrt{(1/15 + 1/20)}$	A1		
		$= 3 \pm 5{\cdot}123\,z$	A1		
	Use of correct tabular value:	$z_{0.995} = 1{\cdot}64$[5]	*A1		
	C.I. correct to 3 s.f. (dep *A1):	$3 \pm 8{\cdot}43$ *or* [-5·43, 11·4]	A1	6	**[6]**

Q13.

9	State both hypotheses (A.E.F.):	$H_0: \mu_1 - \mu_O = 0{\cdot}1$, $H_1: \mu_1 - \mu_O > 0{\cdot}1$	B1		
	State valid assumption for paired-sample test:	Popln. of diffs. has Normal distn.	B1		
	Consider differences eg:	0·4 0·1 0·2 0·3 0·1 0·4 0·1 0·2	M1		
	Calculate sample mean:	$\bar{d} = 1{\cdot}8 / 8$ [= 0·225]	M1		
	Estimate population variance:	$s^2 = (0{\cdot}52 - 1{\cdot}8^2/8) / 7$			
	(allow biased: 0·0144 *or* $0{\cdot}120^2$)	[= 0·0164 *or* $0{\cdot}128^2$]	M1		
	Calculate value of t (to 2 dp):	$t = (\bar{d} - 0{\cdot}1) / s\sqrt{(1/8)} = 2{\cdot}76$	M1*A1		
	Compare with correct tabular t value:	$t_{7,\,0.975} = 2{\cdot}36$[5]	*B1		
	Valid method for reaching conclusion:	Reject H_0 if χ^2 > tabular value	M1		
	Correct conclusion (AEF, dep *A1, *B1):	Coach's suspicion is correct	A1	10	**[10]**
	S.R.: State both hypotheses:	$H_0: \mu_1 - \mu_O = 0{\cdot}1$, $H_1: \mu_1 - \mu_O > 0{\cdot}1$	(B1)		
	State valid assumption for 2-sample test:	Both poplns. have Normal distns. *and* a common variance	(B1)		
	Calculate sample means:	170·4/8, 168·6/8 [= 21·3, 21·075]			
	and estimate population variance:	$s^2 = (3630{\cdot}1 - 170{\cdot}4^2/8 + 3553{\cdot}94 - 168{\cdot}6^2/8)/14$ [= 0·09107]	(M1)		
	Calculate value of t (to 2 dp):	$(0{\cdot}225 - 0{\cdot}1)/s\sqrt{(1/8 + 1/8)} = 0{\cdot}828$	(M1*A1)		
	Compare with correct tabular t value:	$t_{14,\,0.975} = 2{\cdot}14$[5]	(*B1)		
	Correct conclusion (AEF, dep *A1, *B1):	Coach's suspicion is not correct	(B1 max 7)		

Q14.

7	State valid assumption (A.E.F.):	Popln. of masses has normal distn.	B1		
	State both hypotheses:	$H_0: \mu = 1{\cdot}2$, $H_1: \mu > 1{\cdot}2$	B1		
	Calculate sample mean:	$\bar{x} = 12{\cdot}11 / 10$ [= 1·211]	B1		
	Estimate population variance:	$s^2 = (14{\cdot}6745 - 12{\cdot}11^2/10) / 9$			
	(allow biased: 0·000929 *or* $0{\cdot}0305^2$)	[= 0·00103 *or* $0{\cdot}0321^2$]	M1		
	Calculate value of t (to 2 dp):	$t = (\bar{x} - 1{\cdot}2) / (s/\sqrt{10}) = 1{\cdot}08$	M1 A1		
	Compare with correct tabular t value:	$t_{9,\,0.90} = 1{\cdot}38$[3]	B1		
	Conclusion consistent with values (A.E.F):	Claim is not correct	A1 √	8	**[8]**

Q15.

8	Calculate sample mean:	$\bar{d} = 35{\cdot}15$	B1		
	Estimate population variance (allow biased):	$s^2 = 30{\cdot}24$ [$5{\cdot}5^2$] (or 25·2 [$5{\cdot}02^2$])	B1		
	Find confidence interval (allow z in place of t) e.g.:	$35{\cdot}15 \pm t\sqrt{(30{\cdot}24/6)}$	M1		
	(inconsistent use of 5 or 6 loses M1)				
	Use of correct tabular value:	$t_{5,\,0.975} = 2{\cdot}571$ (2 d.p.)	A1		
	Evaluate C.I. correct to 3 s.f. (needs correct s, t):	$35{\cdot}15 \pm 5{\cdot}77$ *or* [29·4, 40·9]	A1	5	
	State inequality involving sample size n:	$1{\cdot}96 \times 5{\cdot}6 / \sqrt{n} \leq$ (*or* <) 2·5	M1 A1		
	(Equality or wrong critical value loses A1)				
	Solve for limiting value of n:	$4{\cdot}39^2 = 19{\cdot}3$	A1		
	State smallest sample size:	$n_{\min} = 20$	A1	4	**[9]**

Q16.

9	State hypotheses:	$H_0: \mu_b = \mu_g$, $H_1: \mu_b > \mu_g$	B1		
	Estimate population variance using boys' sample:	$s_b^2 = (216{\cdot}5 - 92{\cdot}0^2/40) / 39$			
	(allow use of biased: $\sigma_{b,40}^2 = 0{\cdot}1225$ *or* $0{\cdot}35^2$)	[= 0·126 *or* $0{\cdot}354^2$]	M1		
	Estimate population variance using girls' sample:	$s_g^2 = (288{\cdot}8 - 129{\cdot}8^2/60) / 59$			
	(allow use of biased: $\sigma_{g,60}^2 = 0{\cdot}133$ *or* $0{\cdot}365^2$)	[= 0·136 *or* $0{\cdot}368^2$]	M1		
	Estimate population variance for combined sample:	$s^2 = s_b^2/40 + s_g^2/60$			
		= 0·00541 *or* $0{\cdot}0736^2$			
	(allow use of $\sigma_{b,40}^2$, $\sigma_{g,60}^2$)	(*or* 0·00528 or $0{\cdot}0727^2$)	M1 A1		
	Calculate value of z (to 2 dp):	$z = (2{\cdot}3 - 2{\cdot}163)/s$			
		= 0·1367/0·0736 = 1·86 (*or* 1·88)	M1 A1		
	S.R. Allow (implicit) assumption of equal variances:				
	Find pooled estimate of common variance s^2:	$(40\sigma_{b,40}^2 + 60\sigma_{g,60}^2)/98 = 0{\cdot}132$ (M1A1)			
	Calculate value of z (to 2 dp):	$z = (2{\cdot}3 - 2{\cdot}163)/s\sqrt{(1/40+1/60)}$			
		= 1·85	(M1 A1)		
	Compare with correct tabular t value:	$z_{0.975} = 1{\cdot}96$	B1		
	Conclusion consistent with values (A.E.F):	Claim is not correct	A1 √	9	**[9]**

2. Inference Using Normal and t-Distributions

Q17.

6	Find pooled estimate:	$(15{\cdot}05 - 5{\cdot}5^2/5 + 36{\cdot}4 - 8^2/n)/(3+n)$	M1 A1		
	Equate to 3 and rearrange:	$45{\cdot}4 - 64/n = 9 + 3n$	M1 A1		
		$3n^2 - 36{\cdot}4n + 64 = 0$	A1		
	Solve for n:	$n = (36{\cdot}4 \pm 23{\cdot}6) / 6 = 10$	M1 A1	7	**7**

Q18.

8	Consider differences e.g.:	5·1 2·1 2·2 0·6 5·1 3·1 3·9 3·7	M1		
	Calculate sample mean:	$d = 25{\cdot}8 / 8 = 3{\cdot}225$	M1		
	Estimate population variance:	$s^2 = (100{\cdot}14 - 25{\cdot}8^2/8) / 7$			
	(allow biased here: $2{\cdot}117$ *or* $1{\cdot}455^2$)	$[= 2{\cdot}419$ *or* $1{\cdot}555^2]$	M1		
	Find confidence interval (allow z in place of t) e.g.:	$3{\cdot}225 \pm t\sqrt{(2{\cdot}419/8)}$	M1		
	(inconsistent use of 7 or 8 loses M1)				
	Use of correct tabular value:	$t_{7,\,0.975} = 2{\cdot}36[5]$	A1		
	Evaluate C.I. correct to 3 s.f. (in kg):	$3{\cdot}225 \pm 1{\cdot}301$ *or* $[1{\cdot}92, 4{\cdot}53]$	A1	6	
	State hypotheses:	$H_0: \mu_b - \mu_a = 2{\cdot}5$, $H_1: \mu_b - \mu_a > 2{\cdot}5$	B1		
	Calculate value of t (to 2 dp):	$t = (\bar{d} - 2{\cdot}5)/(s/\sqrt{8}) = 1{\cdot}32$	M1 *A1		
	Compare with correct tabular t value:	$t_{7,\,0.95} = 1{\cdot}89[5]$	*B1		
	Correct conclusion (AEF, dep *A1, *B1):	Reduction not more than 2·5	B1	5	**11**

Q19.

9	Calculate sample mean:	$\bar{d} = 2623/5 = 524{\cdot}6$	M1		
	Estimate population variance using 1st sample:	$s_1^2 = (1376081 - 2623^2/5)/4$			
	(allow biased here: $11{\cdot}04$ *or* $3{\cdot}323^2$)	$[= 13{\cdot}8$ *or* $3{\cdot}715^2]$	M1		
	Find confidence interval (allow z in place of t) e.g.:	$524{\cdot}6 \pm t\sqrt{(13{\cdot}8/5)}$	M1		
	(inconsistent use of 4 or 5 loses M1)				
	Use of correct tabular value:	$t_{4,\,0.975} = 2{\cdot}776$ (2 d.p.)	A1		
	Evaluate C.I. correct to 3 s.f. (needs correct s, t):	$524{\cdot}6 \pm 4{\cdot}6[1]$ *or* $[520{\cdot}0, 529{\cdot}2]$	A1	5	
	State hypotheses:	$H_0: \mu_b = \mu_a$, $H_1: \mu_b \neq \mu_a$	B1		
	Estimate population variance using 2nd sample:	$s_2^2 = (2720780 - 5216^2/10)/9$			
	(allow biased here: $11{\cdot}44$ *or* $3{\cdot}382^2$)	$[= 572/45$ *or* $12{\cdot}711$ *or* $3{\cdot}565^2]$	M1		
	Estimate population variance for combined sample:	$s^2 = (4\times13{\cdot}8 + 9\times12{\cdot}71)/13$			
		$= 848/65$ *or* $13{\cdot}05$	M1 A1		
	Calculate value of t (to 2 dp):	$t = (524{\cdot}6 - 521{\cdot}6)/(s\sqrt{(5^{-1} + 10^{-1})})$			
		$= 1{\cdot}52$	M1 *A1		
	Compare with correct tabular t value:	$t_{13,\,0.95} = 1{\cdot}77[1]$	*B1		
	Correct conclusion (AEF, dep *A1, *B1):	No difference in means	B1	8	**[13]**

2. Inference Using Normal and t-Distributions

Q20.

9	Calculate sample mean:	$\bar{d} = 42{\cdot}5 / 8 = 5{\cdot}3125$	M1		
	Estimate population variance:	$s^2 = 15{\cdot}519 / 7$			
	(allow biased here: $1{\cdot}940$ *or* $1{\cdot}393^2$)	$= 2{\cdot}217$ *or* $1{\cdot}489^2$	M1		
	State hypotheses (A.E.F.):	$H_0: \mu = 4{\cdot}5$, $H_1: \mu > 4{\cdot}5$	B1		
	Calculate value of t (to 3 sf):	$t = (\bar{d} - 4{\cdot}5)/(s/\sqrt{8}) = 1{\cdot}54$	M1 *A1		
	State or use correct tabular t value:	$t_{7,\,0.95} = 1{\cdot}89$[5]	*B1		
	(or can compare $\bar{d}$ with $4{\cdot}5 + 0{\cdot}998 = 5{\cdot}49$[8])				
	Correct conclusion (AEF, dep *A1, *B1):	Mean is not greater than 4·5	B1	7	
	Find confidence interval (allow z in place of t) e.g.:	$5{\cdot}3125 \pm t\sqrt{\{15{\cdot}519/(7 \times 8)\}}$	M1		
	Use of correct tabular value:	$t_{7,\,0.975} = 2{\cdot}36$[5]	A1		
	Evaluate C.I. correct to 3 s.f.:	$5{\cdot}31 \pm 1{\cdot}24$[5] *or* [4·07, 6·56]	A1	3	**[10]**

Q21.

10	Estimate population variance using P's sample:	$s_P^2 = (1224 - 240^2/50) / 49$			
	(allow use of biased: $\sigma_{P,50}^2 = 1{\cdot}44$ *or* $1{\cdot}2^2$)	$= 1{\cdot}469$ *or* $72/49$ *or* $1{\cdot}212^2$	M1 A1		
	Estimate population variance using Q's sample:	$s_Q^2 = (754 - 168^2/40) / 39$			
	(allow use of biased: $\sigma_{Q,40}^2 = 1{\cdot}21$ *or* $1{\cdot}1^2$)	$= 1{\cdot}241$ *or* $242/195$ *or* $1{\cdot}114^2$	A1		
	Estimate population variance for combined sample:	$s^2 = s_P^2/50 + s_Q^2/40$			
	(allow use of $\sigma_{P,50}^2$, $\sigma_{Q,40}^2$: $0{\cdot}05905$ *or* $0{\cdot}243^2$)	$= 0{\cdot}0604$ *or* $0{\cdot}246^2$	M1		
	Find confidence interval:	$(240/50 - 168/40) \pm zs$	M1		
	Use $z = 1{\cdot}96$ and evaluate correct to 3 d.p.:	$0{\cdot}6 \pm 0{\cdot}482$ *or* [0·118, 1·082]	A1	6	
	(allow use of $0{\cdot}05905$ instead of $0{\cdot}0604$:	$0{\cdot}6 \pm 0{\cdot}476$ *or* [0·124, 1·076])			
	State hypotheses:	$H_0: \mu_P = \mu_Q$, $H_1: \mu_P > \mu_Q$	B1		
	Calculate value of z (to 2 d.p.):	$z = (240/50 - 168/40)/s = 2{\cdot}44$	M1 *A1		
		(allow $0{\cdot}6/0{\cdot}243 = 2{\cdot}47$)			
	State or use correct tabular z value:	$z_{0.99} = 2{\cdot}326$ [*or* 2·33]	*B1		
	(or can compare 0·6 with 0·572 or 0·56[5])				
	Valid method for reaching conclusion:	Reject H_0 if $z >$ tabular value	M1		
	Correct conclusion (AEF, dep *A1, *B1):	μ_P is greater than μ_Q	A1	6	
	S.R. Allow (implicit) assumption of equal variances, but deduct A1 if not explicit:				
	Pooled estimate of common variance s^2:	$(50\sigma_{P,50}^2 + 40\sigma_{Q,40}^2)/88 = 1{\cdot}368$			
	Confidence interval:	$0{\cdot}6 \pm 1{\cdot}96s\sqrt{(1/50+1/40)}$			
		$= 0{\cdot}6 \pm 0{\cdot}486$ *or* [0·114, 1·086]			
	Value of z (to 2 dp):	$0{\cdot}6/s\sqrt{(1/50+1/40)} = 2{\cdot}42$			**[12]**

2. Inference Using Normal and t-Distributions

Q22.

9	Estimate population variance using A's sample:	$s_A^2 = (481{\cdot}1 - 57{\cdot}4^2/7) / 6$			
	(allow use of biased here: $1{\cdot}489$ *or* $1{\cdot}22^2$)	$= 521/300$ *or* $1{\cdot}737$ *or* $1{\cdot}318^2$	M1 A1		
	Find confidence interval:	$57{\cdot}4/7 \pm t\sqrt{(s_A^2/7)}$	M1		
	State or use correct tabular value of t:	$t_{6,0.975} = 2{\cdot}447$ [*or* $2{\cdot}45$]	A1		
	Evaluate C.I. correct to 3 s.f.:	$8{\cdot}2 \pm 1{\cdot}22$ *or* [6·98, 9·42]	A1	5	
	State suitable assumptions (A.E.F.):	Population of B is Normal			
		and has same variance as for A	B1		
	State hypotheses:	$H_0: \mu_A = \mu_B$, $H_1: \mu_A > \mu_B$	B1		
	Estimate population variance using B's sample:	$s_B^2 = (278{\cdot}74 - 37^2/5) / 4$			
	(allow use of biased here: $0{\cdot}988$ *or* $0{\cdot}994^2$)	$= 1{\cdot}235$ *or* $1{\cdot}111^2$	B1		
	Estimate population variance for combined sample:	$s^2 = (6s_A^2 + 4s_B^2) / 10$			
		$= 192/125$ *or* $1{\cdot}536$ *or* $1{\cdot}239^2$	M1 A1		
	Calculate value of t (to 2 d.p.):	$t = (57{\cdot}4/7 - 37/5)/ s\sqrt{(1/7+1/5)}$	M1		
		$= 0{\cdot}8/0{\cdot}726 = 1{\cdot}10$[2]	*A1		
	State or use correct tabular value	$t_{10,0.95} = 1{\cdot}812$ [*or* $1{\cdot}81$]	*B1		
	Correct conclusion (AEF, dep *A1, *B1):	μ_A is not greater than μ_B	B1		
	S.R.: Deduct only A1 if intermediate result to 3 s.f.				
	S.R.: Invalid method for calculating t (max 6/9):	$t = 0{\cdot}8/\sqrt{(s_A^2/7 + s_B^2/5)}$	(M1)		
		$= 0{\cdot}8/0{\cdot}704 = 1{\cdot}14$	(A1)	9	**[14]**

Q23.

7 (i)	State suitable assumption (A.E.F.):	Population is Normal	B1		
	Find confidence interval:	$1110{\cdot}8/10 \pm t\sqrt{(333{\cdot}9/90)}$	M1 A1		
		$= 111{\cdot}1 \pm t\sqrt{3{\cdot}71}$	A1		
	State or use correct tabular value of t:	$t_{9,0.995} = 3{\cdot}25$	A1		
	Evaluate C.I.:	111 ± 6 *or* [105, 117]	A1	6	
(ii)	Compare t, est. variance s and n:	t and s smaller, n larger	M1		
	Deduce effect on width of C.I. (A.E.F.):	Width is less than in **(i)**	A1	2	**[8]**
	S.R. B1 if valid apart from considering n				

Q24.

9	State hypotheses: $H_0: \mu_P = \mu_Q$, $H_1: \mu_P \neq \mu_Q$ B1			
	Estimate population variance using P's sample: $s_P^2 = (2120 - 321{\cdot}2^2/50) / 49$			
	(allow use of biased: $\sigma_{P,50}^2 = 1{\cdot}132$ *or* $1{\cdot}064^2$) [$= 1{\cdot}155$ *or* $1{\cdot}075^2$] M1			
	Estimate population variance using Q's sample: $s_Q^2 = (3310 - 475{\cdot}3^2/70) / 69$			
	(allow use of biased: $\sigma_{Q,70}^2 = 1{\cdot}182$ *or* $1{\cdot}087^2$) [$= 1{\cdot}199$ *or* $1{\cdot}095^2$] M1			
	Estimate population variance for combined sample: $s^2 = s_P^2/50 + s_Q^2/70$			
	$= 0{\cdot}04023$ *or* $0{\cdot}2006^2$			
	(allow use of $\sigma_{P,50}^2$, $\sigma_{Q,70}^2$) (*or* $0{\cdot}03949$ *or* $0{\cdot}1987^2$) M1 A1			
	Calculate value of z (to 2 d.p., either sign): $z = (6{\cdot}424 - 6{\cdot}79) / s$ M1 A1			
	$= -0{\cdot}366/0{\cdot}2006 = -1{\cdot}82[5]$			
	(*or* $-1{\cdot}84$) A1			
	S.R. Allow (implicit) assumption of equal variances,			
	but deduct A1 if not explicit:			
	Find pooled estimate of common variance s^2: $(50\sigma_{P,50}^2 + 70\sigma_{Q,70}^2)/118$			
	$= 1{\cdot}180$ *or* $1{\cdot}086^2$ (M1A1)			
	Calculate value of z (to 2 d.p.): $z = (6{\cdot}424 - 6{\cdot}79)/s\sqrt{(1/50+1/70)}$ (M1 A1)			
	$= -1{\cdot}82$ (A1)			
	State or use correct tabular z value: $z_{0.95} = 1{\cdot}645$ (to 2 d.p.) B1			
	Conclusion consistent with values (A.E.F): Breaking strengths not the same A1✓		10	**[10]**

Q25.

9	Find sample mean and estimate population variance: (allow biased here: $1{\cdot}19$ *or* $1{\cdot}091^2$) State hypotheses (B0 for $\bar{x}$ …): Calculate value of t (to 3 s.f.): State or use correct tabular t value: (or can compare $x = 6{\cdot}1$ with $5{\cdot}2 + 0{\cdot}667 = 5{\cdot}87$) Correct conclusion (A.E.F., ✓ on t, dep *B1): State hypotheses (B0 for $\bar{x}$ …), e.g.: State assumption (A.E.F.): Find sample mean [and estimate variance] for Y: *EITHER* Estimate (pooled) common varia *OR* (Equivalently) estimate common variance: Calculate value of t (to 3 s.f.): State or use correct tabular t value: (or can compare $\bar{y} - \bar{x} = 0{\cdot}9$ with $0{\cdot}867$) Correct conclusion (A.E.F., ✓ on t, dep *B1):	$\bar{x} = 6{\cdot}1$, $s_x^2 = (384 - 61^2/10)/9$ $= 1{\cdot}322$ *or* $119/90$ *or* $1{\cdot}15^2$ M1 $H_0: \mu = 5{\cdot}2$, $H_1: \mu > 5{\cdot}2$ B1 $t = (\bar{x} - 5{\cdot}2)/(s_x/\sqrt{10}) = 2{\cdot}47[5]$ M1 A1 $t_{9,\,0.95} = 1{\cdot}83[3]$ *B1 $t > 1{\cdot}83$ so mean is greater B1✓		6	
		$H_0: \mu_P = \mu_Q$, $H_1: \mu_P < \mu_Q$ B1 Distributions have equal variances B1 $y = 7{\cdot}0$, $[s_y^2 = (500{\cdot}6 - 70^2/10)/9$ $= 1{\cdot}178$ *or* $53/45$ *or* $1{\cdot}085^2]$ $s^2 = (9 \times 1{\cdot}322 + 9 \times 1{\cdot}178)/18$ *or* $(384 - 61^2/10 + 500{\cdot}6 - 70^2/10)/18$ $= 1.25$ *or* 1.18^2 M1 A1 $s^2 = (1{\cdot}322 + 1{\cdot}178)/10 = 0{\cdot}25$ (M1 A1) $t = (\bar{y} - \bar{x})/s = 1{\cdot}8$ (M1 A1) $t_{18,\,0.95} = 1{\cdot}73[4]$ *B1 $t > 1{\cdot}73$ so Q's mean is greater B1✓	8	**14**

2. Inference Using Normal and t-Distributions

Q26.

8	Find both sample means:	$\bar{x} = 29{\cdot}2, \quad \bar{y} = 24{\cdot}4$	B1		
	Estimate both popn. variances (to 4 s.f.)	$s_x^2 = (55500 - 1752^2/60) / 59$			
		and $s_y^2 = (33500 - 1220^2/50) / 49$	M1 A1		
	(allow biased here: 72·36 *and* 74·64)	$s_x^2 = 73{\cdot}59$ *and* $s_y^2 = 76{\cdot}16$	A1		
	EITHER: Estimate combined variance (3 s.f.):	$s^2 = s_x^2/60 + s_y^2/50$	M1		
		$= 2{\cdot}750$ *or* $1{\cdot}658^2$	A1		
	Use this s to find conf. interval:	$(\bar{x} - \bar{y}) \pm 1{\cdot}96\, s$	M1 A1		
	Evaluate:	$4{\cdot}8 \pm 3{\cdot}25$ *or* $[1{\cdot}55, 8{\cdot}05]$	A1		
	OR: Estimate common variance (to 3 s.f.):	$s^2 = (59 s_x^2 + 49 s_y^2) / 108$			
	(note s_x and s_y not needed explicitly)	*or* $(55500 - 1752^2/60 + 33500 - 1220^2/50) / 108$			
		$= 74{\cdot}8$ *or* $8{\cdot}65^2$ *or* 3364/45	(M1 A1)		
	Use this s to find conf. interval:	$(\bar{x} - \bar{y}) \pm 1{\cdot}96\, s\, \sqrt{(1/60 + 1/50)}$	(M1 A1)		
	Evaluate:	$4{\cdot}8 \pm 3{\cdot}24[5]$ *or* $[1{\cdot}55[5], 8{\cdot}04[5]]$	(A1)	9	**9**

Q27.

8	Find both sample means:	$\bar{x} = 29{\cdot}2, \quad \bar{y} = 24{\cdot}4$	B1		
	Estimate both popn. variances (to 4 s.f.)	$s_x^2 = (55500 - 1752^2/60) / 59$			
		and $s_y^2 = (33500 - 1220^2/50) / 49$	M1 A1		
	(allow biased here: 72·36 *and* 74·64)	$s_x^2 = 73{\cdot}59$ *and* $s_y^2 = 76{\cdot}16$	A1		
	EITHER: Estimate combined variance (3 s.f.):	$s^2 = s_x^2/60 + s_y^2/50$	M1		
		$= 2{\cdot}750$ *or* $1{\cdot}658^2$	A1		
	Use this s to find conf. interval:	$(\bar{x} - \bar{y}) \pm 1{\cdot}96\, s$	M1 A1		
	Evaluate:	$4{\cdot}8 \pm 3{\cdot}25$ *or* $[1{\cdot}55, 8{\cdot}05]$	A1		
	OR: Estimate common variance (to 3 s.f.):	$s^2 = (59 s_x^2 + 49 s_y^2) / 108$			
	(note s_x and s_y not needed explicitly)	*or* $(55500 - 1752^2/60 + 33500 - 1220^2/50) / 108$			
		$= 74{\cdot}8$ *or* $8{\cdot}65^2$ *or* 3364/45	(M1 A1)		
	Use this s to find conf. interval:	$(\bar{x} - \bar{y}) \pm 1{\cdot}96\, s\, \sqrt{(1/60 + 1/50)}$	(M1 A1)		
	Evaluate:	$4{\cdot}8 \pm 3{\cdot}24[5]$ *or* $[1{\cdot}55[5], 8{\cdot}04[5]]$	(A1)	9	**9**

2. Inference Using Normal and t-Distributions

Q28.

7	State hypotheses (A.E.F.):	$H_0: \mu = 7{\cdot}5$, $H_1: \mu < 7{\cdot}5$	B1		
	Calculate sample mean:	$\bar{x} = 70{\cdot}4 / 10 = 7{\cdot}04$	M1		
	Estimate population variance:	$s^2 = 8{\cdot}48 / 9 = 211/225$ *or*			
	(allow biased here: $0{\cdot}848$ *or* $0{\cdot}9209^2$)	$0{\cdot}9422$ *or* $0{\cdot}9707^2$	M1		
	Calculate value of t (to 3 sf):	$t = (\bar{x} - 7{\cdot}5)/(s/\sqrt{10}) = \pm 1{\cdot}49$[9]	M1 *A1		
	State or use correct tabular t value:	$t_{9,\,0.9} = 1{\cdot}38$[3]	*B1		
	(or can compare $\bar{x}$ with $7{\cdot}5 - 0{\cdot}425 = 7{\cdot}07$[5])				
	correct conclusion (AEF, dep *A1, *B1):	Mean is less than 7·5	B1	7	**[7]**

Q29.

11b	State hypotheses:	$H_0: \mu_X = \mu_Y$, $H_1: \mu_X < \mu_Y$	B1		
	Estimate population variance using X's sample:	$s_X^2 = (85{\cdot}8 - 58{\cdot}2^2/60) / 59$			
	allow use of biased: $\sigma_{X,60}^2 = 0{\cdot}4891$				
	or $0{\cdot}6994^2$)	[= $0{\cdot}4974$ *or* $0{\cdot}7053^2$]	M1 A1		
	Estimate population variance using Y's sample:	$s_Y^2 = (188{\cdot}6 - 97{\cdot}6^2/80) / 79$			
	(allow use of biased: $\sigma_{Y,80}^2 = 0{\cdot}8691$				
	or $0{\cdot}9323^2$)	[= $0{\cdot}8801$ *or* $0{\cdot}9381^2$]	M1 A1		
	Estimate population variance for combined sample:	$s^2 = s_X^2/60 + s_Y^2/80$			
		$= 0{\cdot}01929$ *or* $0{\cdot}1389^2$			
	(allow use of $\sigma_{X,60}^2$, $\sigma_{Y,80}^2$)	(*or* $0{\cdot}01901$ *or* $0{\cdot}1379^2$)	M1 A1		
	Calculate value of z (to 2 d.p., either sign):	$z = (0{\cdot}97 - 1{\cdot}22) / s$	M1		
		$= -0{\cdot}25/0{\cdot}1389 = -1{\cdot}80$			
		(*or* $-1{\cdot}813$)	A1		
	Find $\Phi(z)$:	$\Phi(z) = 0{\cdot}9641$ (*or* $0{\cdot}9651$)	M1		
	Find corresponding significance level:	$3{\cdot}59$ (*or* $3{\cdot}49$)	M1		
	Find set of possible values of α (to 1 d.p.):	$\alpha > 3{\cdot}6$	A1	12	
	S.R. Allow (implicit) assumption of equal variances as follows, but deduct A1 if not explicit:				
	Find pooled estimate of common variance s^2:	$(60\sigma_{X,60}^2 + 80\sigma_{Y,80}^2)/138$			
		$= 0{\cdot}7165$ *or* $0{\cdot}8465^2$			
	Calculate value of z (to 2 d.p.):	$z = (0{\cdot}97 - 1{\cdot}22)/s\sqrt{(1/60+1/80)}$			
		$= -1{\cdot}73$			
	Find set of possible values of α (to 1 d.p.):	$\Phi(z) = 0{\cdot}9582$, $\alpha > 4{\cdot}2$			**[12]**

Q30.

7	Find $\Sigma (x - \bar{x})^2$:	$\Sigma (x - \bar{x})^2 = 25 - 10^2/n$	B1		
	Find $\Sigma (y - \bar{y})^2$:	$\Sigma (y - \bar{y})^2 = 43{\cdot}5 - 15^2/2n$	B1		
	Equate pooled estimate of σ^2 to 2:	$(25 - 10^2/n + 43{\cdot}5 - 15^2/2n) /$			
		$(3n - 2) = 2$	M1 A1		
	Rearrange to give quadratic eqn for n:	$12n^2 - 145n + 425 = 0$	M1		
	Find n (integer value):	$n = (145 \pm 25)/24 = 5$ ($\neq 7{\cdot}08$)	M1 A1	7	**7**

2. Inference Using Normal and t-Distributions

Q31.

9	Calculate sample mean:	$\bar{x} = 94{\cdot}5 / 9 = 10{\cdot}5$	M1		
	Estimate population variance:	$s^2 = (993{\cdot}6 - 94{\cdot}5^2/9) / 8$			
	(allow biased here: 0·15 *or* $0{\cdot}3873^2$)	$= 0{\cdot}16875$ *or* $0{\cdot}4108^2$	M1		
	State hypotheses (A.E.F.):	$H_0: \mu = 10{\cdot}2$, $H_1: \mu \neq 10{\cdot}2$	B1		
	Calculate value of t (to 3 s.f.):	$t = (\bar{x} - 10{\cdot}2)/(s/\sqrt{9}) = 2{\cdot}19$	M1 *A1		
	State or use correct tabular t value (to 3 s.f.):	$t_{8,\,0.975} = 2{\cdot}306$	*B1		
	(or can compare $\bar{x}$ with $10{\cdot}2 + 0{\cdot}316 = 10{\cdot}52$)				
	Correct conclusion (AEF, dep *A1, *B1):	Population mean is 10·2	B1	7	
	Find confidence interval (allow z in place of t) e.g.:	$10{\cdot}5 \pm t\sqrt{\{1{\cdot}35/(8 \times 9)\}}$	M1		
	Use of correct tabular value:	$t_{8,\,0.95} = 1{\cdot}86[0]$	A1		
	Evaluate C.I. correct to 3 s.f.:	$10{\cdot}5 \pm 0{\cdot}255$ *or* [10·2, 10·8]	A1	3	**10**

Q32.

8	State hypotheses (B0 for $\bar{x}$ …):	$H_0: \mu_X = \mu_Y$, $H_1: \mu_X \neq \mu_Y$	B1		
	Estimate both popln. variances using two samples: (allow use of biased: $\sigma_{X,60}^2 = 236$ *or* $15{\cdot}36^2$)	$S_x^2 = (626\,220 - 6060^2/60) / 59$ [= 240 *or* $15{\cdot}49^2$]			
	(allow use of biased: $\sigma_{Y,50}^2 = 265$ *or* $16{\cdot}28^2$)	*And* $s_Y^2 = (464\,500 - 4750^2/50) / 49$ [= 270·4 *or* $16{\cdot}44^2$]	M1 A1		
	Estimate population variance for combined sample:	$s^2 = s_X^2/60 + s_Y^2/50$			
	(allow $\sigma_{X,60}^2/60 + \sigma_{Y,50}^2/50$: 9·233 *or* $3{\cdot}039^2$)	$= 9{\cdot}408$ *or* $3{\cdot}067^2$	M1 A1		
	Calculate value of z (to 2 d.p., either sign):	$z = (101 - 95) / s$	M1		
		$= 6/3{\cdot}067 = 1{\cdot}96$ (*or* 1·97)	A1		
	State or use correct tabular z – value (to 2 d.p.):	$z_{0.99} = 2{\cdot}326$ *or* 2·33 (allow 2·36)	B1		
	(or can compare 6 with e.g. $2{\cdot}326\,s = 7{\cdot}13$ *or* 7·07)				
	Correct conclusion (A.E.F, √on z – values):	[Accept H_0] Claims are the same	B1√		
	S.R. Assuming equal population variances: Find pooled estimate of common variance s^2:	Hypotheses; Explicit assumption : $s^2 = (626\,220 - 6060^2/60 +$	(B1; B1)		
		$464\,500 - 4750^2/50) / 108$	(M1 A1)		
	Calculate value of z (to 2 d.p., either sign):	$z = 6 / s\sqrt{(1/60+1/50)} = 1{\cdot}97$	(M1 A1)		
		$= 253{\cdot}8$ *or* $15{\cdot}93^2$	(A1)		
	Tabular value; conclusion	As above)	(B1; B1√)		**9**

2. Inference Using Normal and t-Distributions

Q33.

B	Estimate population variance:	$s_P^2 = (236{\cdot}0 - 42{\cdot}8^2/8)/7$			
	(allow biased here: 0·8775 or $0{\cdot}9367^2$)	$= 351/350$ or $1{\cdot}003$ or $1{\cdot}001^2$	M1		
	Find confidence interval (allow z in place of t) e.g.:	$42{\cdot}8/8 \pm t\sqrt{}(s_P^2/8)$	M1		
	Use correct tabular t-value:	$t_{7,\,0.975} = 2{\cdot}365$	A1		
	Evaluate C.I. correct to 2 d.p.:	$5{\cdot}35 \pm 0{\cdot}84$ or $[4{\cdot}51, 6{\cdot}19]$	A1	4	
	Formulate inequality for k (or equality for k_{max}):	$(5{\cdot}35 - k)/\sqrt{(s_P^2/8)} \geqslant$ [or $>$] t	M1		
	Use correct tabular t-value:	$t_{7,\,0.9} = 1{\cdot}415$	A1		
	Solve for k_{max} (A0 if = or ⩽ was used for k above):	$5{\cdot}35 - k \geqslant 0{\cdot}50$, $k_{max} = 4{\cdot}85$	A1	3	
	State hypotheses (B0 for $\bar{x}$ …), e.g.:	H_0: $\mu_P = \mu_Q$, H_1: $\mu_P > \mu_Q$	B1		
	State assumption (A.E.F.):	Normal distns. for [P and] Q			
		and equal variances	B1		
	Estimate (pooled) common variance:	$s^2 = (7 \times 1{\cdot}003 + 11 \times 1{\cdot}962)/18$			
		$= 1{\cdot}589$ or $1{\cdot}261^2$	M1 A1		
	Calculate value of t (to 3 s.f.):	$t = (5{\cdot}35 - 4{\cdot}60)/(s\sqrt{(1/8 + 1/12)})$			
		$= 1{\cdot}30$	M1 A1		
	Correct conclusion (A.E.F., $\sqrt{}$ on t):	$t < t_{18,\,0.9} = 1{\cdot}33$ so Q's mean is not less than P's	B1$\sqrt{}$	7	**14**

Q34.

10 **(i)**	Calculate sample mean:	$\bar{x} = 132/8 = 16{\cdot}5$	M1		
	Estimate population variance:	$s^2 = (2192{\cdot}06 - 132^2/8)/7$			
	(allow biased here: $1{\cdot}757$ *or* $1{\cdot}326^2$)	$= 703/350$ *or* $2{\cdot}009$ *or* $1{\cdot}417^2$	M1		
	State hypotheses (A.E.F.; B0 for $\bar{x}$):	$H_0: \mu = 15{\cdot}8$, $H_1: \mu > 15{\cdot}8$	B1		
	Calculate value of t (to 3 s.f.):	$t = (\bar{x} - 15{\cdot}8)/(s/\sqrt{8}) = 1{\cdot}4[0]$	M1 A1		
	State or use correct tabular t–value (to 3 s.f.):	$t_{7,\,0.9} = 1{\cdot}41[5]$	B1		
	(or can compare $\bar{x}$ with $15{\cdot}8 + 0{\cdot}709 = 16{\cdot}51$)				
	Consistent conclusion (A.E.F, $\sqrt{}$ on both t–values):	[Accept H_0:]			
		Popn. mean not greater than 15·8	B1$\sqrt{}$	7	
(ii)	Find confidence interval (allow z in place of t) e.g.:	$16{\cdot}5 \pm t\sqrt{(2{\cdot}009/8)}$	M1 A1		
	(Use of 15·8 does not lose M1)				
	Use of correct tabular value:	$t_{7,\,0.975} = 2{\cdot}36[5]$	A1		
	Evaluate C.I. correct to 3 s.f.:	$16{\cdot}5 \pm 1{\cdot}18[5]$ *or* $[15{\cdot}3, 17{\cdot}7$	A1	4	**11**

2. Inference Using Normal and t-Distributions

Q35.

6	Estimate population variance for combined sample: $s^2 = \frac{s_X^2}{50} + \frac{s_Y^2}{60}$			
	$= \frac{1391}{1500}$ *or* 0·9273 *or* 0.9630^2	M1		
	Calculate value of z (to 2 d.p., either sign): $z = \frac{1.8}{s} = 1.869$	M1 A1		
	Find $\Phi(z)$ and set of possible values of α (to 1 d.p.):			
	$\Phi(z) = 0.9692$ [*or* 96·92%]			
	(M1 A0 for $\alpha \leqslant 3.1$ *or* $\alpha > 93.8$) $\alpha \leqslant$ (*or* <) 6·2 (allow 6·1)	M1 A1	5	**[5]**

Q36.

11b	Estimate population variance using A's sample: $s_A^2 = \frac{\left(1422.34 - \frac{106^2}{8}\right)}{7}$			
	(allow use of biased: $\sigma_{A,8}^2 = 2.23$ *or* 1.493^2) $= \frac{446}{175}$ *or* 2·549 *or* 1.596^2	M1 A1		
	Find confidence interval: $\frac{106}{8} \pm t\sqrt{\left(\frac{s_X^2}{8}\right)}$	M1		
	State or use correct tabular value of t: $t_{7,0.975} = 2.36$ [5]	A1		
	Evaluate C.I. correct to 1 d.p.: 13·25 ± 1·335 *or* [11·9, 14·6]	A1	5	
	State suitable assumptions (A.E.F.): Distribution of B is Normal with same population variance	B1		
	State hypotheses (B0 for $\bar{a}$ …) e.g.: $H_0: \mu_A = \mu_B$, $H_1: \mu_A > \mu_B$	B1		

	Estimate (or imply) B's population variance:	$s_B^2 = \dfrac{\left(971.53 - \frac{75.9^2}{6}\right)}{5}$			
	(allow use of biased: $\sigma_{B,6}^2 = 1{\cdot}899$ *or* $1{\cdot}378^2$)	$= 2{\cdot}279$ *or* $1{\cdot}510^2$			
	and find pooled estimate of common variance s^2:	$s^2 = \dfrac{\left(7s_A^2 + 5s_B^2\right)}{12}$			
		$= \dfrac{(17.84 + 11.395)}{12}$			
		$= 2{\cdot}436$ *or* $1{\cdot}561^2$	M1		
	Calculate value of t (to 2 d.p., either sign):	$t = \dfrac{(13.25 - 12.65)}{s\sqrt{\left(8^{-1} + 6^{-1}\right)}}$			
		$= \dfrac{0.6}{0.8430} = 0{\cdot}712$	M1 A1		
	State or use correct tabular t-value (to 2 d.p.): (or compare 0·6 with $1{\cdot}782\ s\sqrt{(8^{-1} + 6^{-1})} = 1{\cdot}50$)	$t_{12,0.95} = 1{\cdot}782$	B1		
	Consistent conclusion (AEF, √ on two t values):	[Accept H_0]; mean lengths are the same	B1√	7	
	Find confidence interval for the difference:	$13{\cdot}25 - 12{\cdot}65 \pm t\,s\sqrt{(8^{-1} + 6^{-1})}$	M1		
	Evaluate C.I. with $t_{12,0.975} = 2{\cdot}179$, to 2 d.p.:	$0{\cdot}6 \pm 1{\cdot}84$ *or* $[-1{\cdot}24, 2{\cdot}44]$	A1	2	**[14]**

Q37.

6	Equate pooled estimate of σ^2 to 12:	$(11 - 5^2/N + 160 - 10^2/10)\ /$ $(N + 10 - 2) = 12$	M1 A1		
	Formulate and solve relevant quadratic eqn. for N:	$12N^2 - 65N + 25 = 0,\ N = 5$	M1 A1	4	**4**

Q38.

7	Find Σx via sample mean $\bar{x}$:	$\Sigma x = 8\bar{x} = 8 \times \frac{1}{2}(1.17 + 2.03)$ $= 8 \times 1.6 = 12.8$	M1 A1		
	Find estimate of population variance s^2:	$t\sqrt{(s^2/8)} = \frac{1}{2}(2.03 - 1.17)$ [= 0.43]	M1		
	Use of correct tabular value (1·96 leads to 23·2):	$t_{7,\,0.975} = 2.36$[5]	A1		
	(to 3 d.p.)	$s^2 = 0.2645$ *or* 32/121 *or* 0.5143^2	A1		
	Find Σx^2 from s^2:	$s^2 = \{\Sigma x^2 - (\Sigma x)^2/8\}/7$			
	(M0 for $s^2 = \{\ldots\}/8$)	$\Sigma x^2 = 7 \times 0.2645 + 12.8^2/8 = 22.3$	M1 A1	7	**7**

Q39.

8	Calculate sample mean:	$\bar{x} = 45.29 / 7 = 6.47$	M1		
	Estimate population variance:	$s_1^2 = (295.36 - 45.29^2/7) / 6$			
	(allow biased here: 0.3334 *or* 0.5774^2)	$= 0.389$ *or* 0.6236^2	M1		
	State hypotheses: (AEF; B0 for $\bar{x}$)	$H_0: \mu = 6{\cdot}2$, $H_1: \mu > 6.2$	B1		
	Calculate value of t (or $-t$; to 3 s.f.):	$t = (\bar{x} - 6.2)/(s_1/\sqrt{7}) = 1.15$	M1 A1		
	State or use correct tabular t–value (to 3 s.f.):	$t_{6,\,0.95} = 1.94[3]$	B1		
	(or can compare $\bar{x}$ with $6.2 + 0.458 = 6.66$)				
	Consistent conclusion (AEF, ft on both t–values):	[Accept H_0]			
		Popn. mean not greater than 6.2	B1✓		
	Find estimate of population mean:	$\frac{1}{2}(5.89 + 6.75) = 6.32$	M1 A1	7	
	Find estimate of population variance s_2^2:	$t\sqrt{(s_2^2/8)} = \frac{1}{2}(6.75 - 5.89)$ [= 0.43]	M1		
	SR Allow M1 for $t\sqrt{(s_2^2/7)}$, max 4/5				
	Use of correct tabular value:	$t_{7,\,0.975} = 2.36[5]$	A1		
	(Use of 2.36 may lose final A1)	[$s_2 = 0.5143$ *or* $4\sqrt{2}/11$]			
	(to 3 d.p.)	$s_2^2 = 0.2645$ *or* $32/121$	A1	5	**12**

Q40.

10B	State hypotheses (B0 for $\bar{t}_A$ …), e.g.:	$H_0: \mu_A = \mu_B$, $H_1: \mu_A < \mu_B$	B1		
	Estimate both popn. variances (to 3 d.p..):	$s_A^2 = (215.18 - 102^2/50) / 49$			
		and $s_B^2 = (282.3 - 129^2/60) / 59$	M1 A1		
	(allow biased here: 0.1420 *or* 0.3768^2	$s_A^2 = 0.1449$ *or* $71/490$			
		or 0.3807^2			
	and 0.0825 *or* 0.2872^2)	*and* $s_B^2 = 0.0839$ *or* $99/1180$			
		or 0.2897^2	A1		
	EITHER: Estimate combined variance:	$s^2 = s_A^2/50 + s_B^2/60$	M1		
	(to 3 s.f.)	$= 0.004296$ *or* 0.06555^2	A1		
	Calculate value of z (or $-z$):	$z = (\bar{t}_B - \bar{t}_A) / s$			
	(to 3 s.f.)	$= (2.15 - 2.04) / s = 1.68$	M1 A1		
	OR: Pooled estimate of common variance:	$s^2 = (49\,s_A^2 + 59\,s_B^2) / 108$			
	(note s_A^2 and s_B^2 not needed explicitly)	*or* $(215.18 - 102^2/50 +$			
		$282.3 - 129^2/60) / 108$			
	(to 3 s.f.)	$= 0.1116$ *or* 0.334^2	(M1 A1)		
	Calculate value of z (or $-z$):	$z = (\bar{t}_B - \bar{t}_A) / s\sqrt{(1/50 + 1/60)}$			
	(to 3 s.f.)	$= 0.11 / 0.06396 = 1{\cdot}72$	(M1 A1)		
	State or use correct tabular z value: (to 3 s.f.)	$z_{0.95} = 1.645$(allow 1.658)	*B1		
	(or can compare $\bar{t}_B - \bar{t}_A = 0.11$ with 0.105)				
	Correct conclusion (AEF, ft on z, dep *B1):	$z >$ tabular value			
		so club A takes less time	B1✓	10	
	Calculate value of z (either sign):	$z = (0.11 - 0.05)/s = 0.06/s$			
	(to 3 s.f.)	$= 0.9154$[or 0.9381]	M1 A1		
	Find $\Phi(z)$ and set of possible values of α:	$\Phi(z) = 0.820$[or 0.826]			
	(M0 for $\alpha \sim 82$ or 0.82) (to 1 d.p.)	$\alpha \geqslant$ (*or* >) 18.0[or 17.4]	M1 A1	4	
	SR Allow B1 (max 3/4) for $\alpha < 18.0$ [or 17.4]				**14**

2. Inference Using Normal and t-Distributions

Q41.

5	Find or use sample mean and estimate population variance: (allow biased here: 0·412 *or* $0{\cdot}642^2$)	$\bar{x} = 222{\cdot}8 / 10 = 22{\cdot}28$ $s^2 = 4{\cdot}12 / 9$ $= 0{\cdot}458$ *or* $103/225$ *or* $0{\cdot}677^2$	**M1**		
	Find confidence interval (allow z in place of t) e.g.:	$22{\cdot}28 \pm t\sqrt{(0{\cdot}458 / 10)}$	**M1 A1**		
	Use of correct tabular value:	$t_{9,\,0.975} = 2{\cdot}26[2]$	**A1**		
	Evaluate C.I. correct to 3 s.f.:	$22{\cdot}3 \pm 0{\cdot}48[4]$ *or* $[21{\cdot}8, 22{\cdot}8]$	**A1**	5	**5**

Q43.

10B	State hypotheses (B0 for $\bar{x}$ …), e.g.:	H_0: $\mu_X = \mu_Y$, H_1: $\mu_X \neq \mu_Y$	**B1**		
	State assumption (AEF):	Distributions have equal variances	**B1**		
	Find sample means and estimate popln. variances: (allow biased here: 0·36 *or* $0{\cdot}6^2$)	$\bar{x} = 4{\cdot}2$, $\bar{y} = 4{\cdot}8$ $s_X^2 = (180 - 42^2/10) / 9$ $= 0{\cdot}4$ *or* $0{\cdot}6325^2$ $s_Y^2 = (281{\cdot}5 - 57{\cdot}6^2/12) / 11$			
	(allow biased here: 0·4183 *or* $0{\cdot}6468^2$)	$= 0{\cdot}4564$ *or* $251/550$ *or* $0{\cdot}6755^2$	**M1**		
	Estimate (pooled) common variance: (note s_X^2 and s_Y^2 not needed explicitly)	$s^2 = (9\,s_X^2 + 11\,s_Y^2) / 20$ (AEF) *or* $(180 - 42^2/10 + 281{\cdot}5 - 57{\cdot}6^2/12) / 20$ $= 0{\cdot}431$ *or* $0{\cdot}6565^2$	**M1 A1**		
	Calculate value of t (to 3 s.f.):	$[-]\,t = (\bar{y} - \bar{x}) / s\sqrt{(1/10 + 1/12)}$ $= 2{\cdot}13$	**M1 A1**		
	State or use correct tabular t value: (or can compare $\bar{y} - \bar{x} = 0{\cdot}6$ with 0·586)	$t_{20,\,0.975} = 2{\cdot}086$ [allow 2·09]	**B1***		
	Correct conclusion (AEF, √ on t, dep *B1): **S.R.** Implicitly taking s_X^2, s_Y^2 as popln. variances: (may also earn first B1 B1 M1) Comparison with $z_{0.975}$ and conclusion: (can earn at most 5/9)	$t > 2{\cdot}09$ so mean masses not same $z = (\bar{y} - \bar{x}) / \sqrt{(s_X^2/10 + s_Y^2/12)}$ $= 0{\cdot}6 / \sqrt{(0{\cdot}078} = 2{\cdot}15$ $2{\cdot}15 > 1{\cdot}96$ so mean masses not same	**B1√** **(B1)**	9	
	State hypotheses (B0 for $\bar{x}$ …), e.g.:	H_0: $\mu_X = 3{\cdot}8$, H_1: $\mu_X > 3{\cdot}8$ *or* H_0: $\mu_X = \mu_Z$, H_1: $\mu_X > \mu_Z$	**B1**		
	Calculate value of t using s_X from above:	$t = (4{\cdot}2 - 3{\cdot}8) / (s_X/\sqrt{10}) = 2{\cdot}0$	**M1 A1**		
	State or use correct tabular t value: (or can compare 0·4 with 0·367)	$t_{9,\,0.95} = 1{\cdot}833$ [allow 1·83]	**B1***		
	Correct conclusion (A.E.F., √ on t, dep *B1):	$t > 1{\cdot}833$, so claim is justified *or* mean mass of Royals > mean mass of Crowns	**B1√**	5	**14**

2. Inference Using Normal and t-Distributions

Q44.

7	Calculate sample mean: Estimate population variance: (allow biased here: 0·14 *or* $0{\cdot}3742^2$) State hypotheses (AEF; B0 for $\bar{x}$ ...): Calculate value of t (to 3 s.f.): State or use correct tabular t-value (to 3 s.f.) : (or can compare $\bar{x}$ with 2·5 + 0·246 = 2·746) State or imply valid method for conclusion e.g.: Conclusion (AEF, requires both values correct):	$\bar{x} = 24{\cdot}6 / 9$ *or* 41/15 *or* 2·73[3] $s^2 = (68{\cdot}5 - 24{\cdot}6^2/9) / 8$ $= 63/400$ *or* 0·1575 *or* $0{\cdot}3969^2$ $H_0: \mu = 2{\cdot}5$, $H_1: \mu > 2{\cdot}5$ $t = (\bar{x} - 2{\cdot}5)/(s/\sqrt{9}) = 1{\cdot}76$ $t_{8,\,0.95} = 1{\cdot}86[0]$ Accept H_0 if $t <$ tabular value 1·76 [±0·02] < 1·86 so popln. mean not greater than 2·5	B1 M1 B1 M1 A1 B1 M1 A1	[8]

Q45.

11B	Find Σx via sample mean $\bar{x}$: 26·83) *EITHER:* Find estimated s.d. s_A, or s_A^2: with $t = t_{11,\,0.975} = 2{\cdot}201$ (to 3 s.f.) Find Σx^2 from s_A or s_A^2: (M0 for $s_A^2 = \{\ldots\}/12$) *OR:* Find estimated s.d. σ_A, or σ_A^2: with $t = t_{11,\,0.975} = 2{\cdot}201$ (to 3 s.f.) Find Σx^2 from σ_A or σ_A^2: (M0 for $\sigma_A^2 = \{\ldots\}/11$) State hypotheses (B0 for $\bar{x}_A$ ...), e.g.: Estimate B's popln. variance (to 3 d.p..): (allow biased here: 4·578) Find pooled estimate of common variance: (note s_B^2 may be implied, earning M1 above) Calculate value of t (or $-t$): (to 3 s.f.) State or use correct tabular t value: (to 3 s.f.) (or can compare $\bar{x}_A - \bar{x}_B = 0{\cdot}7143$ with 1·536) Consistent conclusion (AEF, √ on t, dep *B1): <

2. Inference Using Normal and t-Distributions

Q46.

11 (b)	State hypotheses (B0 for $\bar{x}_A$ …), e.g.:	H_0: $\mu_B = \mu_A$, H_1: $\mu_B > \mu_A$	**B1**	
	Estimate both popn. variances:	$s_A^2 = (1910{\cdot}8 - 256^2/40) / 39$		
	and	$s_B^2 = (3148{\cdot}8 - 382{\cdot}9^2/50) / 49$	**M1 A1**	
	(allow biased here: 6·81 *or* $2{\cdot}610^2$	$s_A^2 = 6{\cdot}985$ *or* $454/65$ *or* $2{\cdot}643^2$		
	and 4·331 *or* $2{\cdot}081^2$) *and*	$s_B^2 = 4{\cdot}419$ *or* $2{\cdot}102^2$	**A1**	
	Estimate combined variance (may be implied):	$s^2 = s_A^2/40 + s_B^2/50$	**M1**	
	(to 3 s.f)	$= 0{\cdot}263$ *or* $0{\cdot}513^2$	**A1**	
	Calculate value of z (or $-z$):	$z = (\bar{x}_B - \bar{x}_A) / s$		
		$= (7{\cdot}658 - 6{\cdot}4) / s$		
	(to 3 s.f.)	$= 1{\cdot}258/0{\cdot}5128 = 2{\cdot}45$	**M1 A1**	
	State or use correct tabular z value:(to 3 s.f.)	$z_{0.99} = 2{\cdot}326$(**SC**: allow 2·358)	***B1**	
	(or can compare $\bar{x}_B - \bar{x}_A = 1{\cdot}258$ with 1·19)			
	Correct conclusion (AEF, √ on z, dep *B1)	:$z >$ tabular value		
		so μ_B is greater than μ_A	**B1**√	
	Note: Basing test on equal variances can earn first 4 marks only			[10]
	Calculate value of z using width of interval:	$z = 1{\cdot}82 / 2s = 0{\cdot}91 / 0{\cdot}5128$		
	(allow M1 for $z = 1{\cdot}82 / s$)	$= 1{\cdot}77$[46]	**M1 A1**	
	Find $\Phi(z)$ and value of α :	$\Phi(z) = 0{\cdot}962 = 1 - ½(1 - \alpha/100)$		
	(overlooking % is misread;	$\alpha = 100\{1 - 2(1 - 0{\cdot}962)\}$		
	M0 for $\alpha = 96{\cdot}2$ or finding $100 - \alpha$)	$= 92{\cdot}4$	**M1 A1**	[4]

Q47.

6	Estimate population variance:	$s^2 = 0{\cdot}784 / 7$		
	(allow biased here: 0·098 *or* $0{\cdot}3130^2$)	$= 0{\cdot}112$ *or* $14/125$ *or* $0{\cdot}3347^2$	**M1**	
	State hypotheses (AEF; B0 for $\bar{x}$):	H_0: $\mu = 6{\cdot}44$, H_1: $\mu < 6{\cdot}44$	**B1**	
	Calculate value of t (either sign; to 3 s.f.):	$t = (6{\cdot}44 - \bar{x})/(s/\sqrt{8}) = 1{\cdot}64$	**M1 A1**	
	State or use correct tabular t-value (to 3 s.f.):	$t_{7,\,0.95} = 1{\cdot}89$[5]	**B1**	
	(or can compare $\bar{x}$ with 6·4 – 0·224 = 6·176)			
	State or imply valid method for conclusion e.g.:	Accept H_0 if $t <$ tabular value	**M1**	
	Conclusion (AEF, requires both values correct):	1·64 < 1·89 so		
		popln. mean not less than 6·44	**A1**	[7]

Q48.

8 (i)	*EITHER:* Estimate P's popln. variance (to 3 d.p.):	$s_P^2 = (1\,560\,000 - 9600^2/60) / 59$	**M1**	
	(allow biased here: 400)	[= 406·78]		
	Estimate Q's popln. variance (to 3 d.p.):	$s_Q^2 = (1\,052\,500 - 7200^2/50) / 49$	**M1**	
	(allow biased here: 314)	[= 320·41]		
	Find pooled estimate of common variance:	$s^2 = (59\,s_P^2 + 49\,s_Q^2) / 108$		
		$= 367{\cdot}6$ *or* 368 *or* $9925/27 s^2 = (1\,560\,000 - 9600^2/60$	**M1 A1**	
	OR: Find pooled estimate of common variance:	$s^2 = (1\,560\,000 - 9600^2/60$		
		$+ 1\,052\,500 - 7200^2/50) / 108$		
		$= 367{\cdot}6$ *or* 368 *or* $9925/27$	**(M3 A1)**	[4]
(ii)	Find confidence interval for the difference:	$9600/60 - 7200/50$ [= 160 – 144]		
		$\pm z\,s\sqrt{(60^{-1} + 50^{-1})}$	**M1 A1**	
	Use appropriate tabular value (to 2 d.p.):	$z_{0.975} = 1{\cdot}96$ *or* $t_{120,\,0.975} = 1{\cdot}98$	**A1**	
	Evaluate confidence interval (AEF, to 1 d.p.):	$16 \pm 7{\cdot}2$ *or* [8·8, 23·2]		
		or $16 \pm 7{\cdot}3$ *or* [8·7, 23·3]	**M1 A1**	[5]
	SR Using combined variance $s_P^2/60 + s_Q^2/50 = 13{\cdot}19$:			
	(i) M1, M1 as above; then M0 A0 (max 2/4)			
	(ii) M1 A0 for $9600/60 - 7200/50 \pm z\,s$			
	A1 for tabular value as above			
	M1 A0 for evaluating interval $16 \pm 7{\cdot}1$ (*or* 7·2)			
	(max 3/5)			

Q49.

7	$\bar{x} = 7{\cdot}2$	**B1**	Find sample mean
	$s^2 = (542 - 72^2/10) / 9$ $[= 118/45$ *or* $2{\cdot}622$ *or* $1{\cdot}619^2]$	**M1**	Estimate population variance (allow biased here: $2{\cdot}36$ *or* $1{\cdot}536^2$)
	H_0: $\mu = 6{\cdot}2$, H_1: $\mu > 6{\cdot}2$ (AEF)	**B1**	State hypotheses (**B0** for $\bar{x}$ ...)
	$t_{9,\,0.95} = 1{\cdot}83[3]$	**B1**	State or use correct tabular *t*-value
	$t = (\bar{x} - 6{\cdot}2) / (s/\sqrt{10}) = 1{\cdot}95$ [Accept H_1:]	**M1 A1**	Find value of *t* (or can compare $\bar{x}$ with $6{\cdot}2 + 0{\cdot}939 = 7{\cdot}14$) Consistent conclusion
	Claim (of mean mass increased) is justified (AEF)	**B1 FT**	(FT on both *t*-values)
	Total:	**7**	

Q50.

9	H_0: $\mu_X = \mu_Y$, H_1: $\mu_X \neq \mu_Y$ (AEF)	**B1**	State hypotheses (**B0** for $\bar{x}$ ...)
	$x = 13{\cdot}4/8$ *or* $1{\cdot}67[5]$, $\bar{y} = 2{\cdot}02$ (all to 3 s.f.)	**B1**	Find sample means (values to 3 s.f. throughout)
	$s_X^2 = (24{\cdot}7 - 13{\cdot}4^2/8) / 7$ $= 451/1400$ *or* $0{\cdot}3221$ *or* $0{\cdot}5678^2$ *and* $s_Y^2 = (44{\cdot}6 - 20{\cdot}2^2/10) / 9$ $= 949/2250$ *or* $0{\cdot}4218$ *or* $0{\cdot}6494^2$	**M1**	Estimate or imply popln. variances (allow biased here: $0{\cdot}2819$ *or* $0{\cdot}5309^2$) (allow biased here: $0{\cdot}3796$ *or* $0{\cdot}6161^2$)
	$s^2 = (7 s_X^2 + 9 s_Y^2) / 16$ (AEF) *or* $(24{\cdot}7 - 13{\cdot}4^2/8 + 44{\cdot}6 - 20{\cdot}2^2/10) / 16$	**M1 A1**	Estimate (pooled) common variance (note s_X^2 and s_Y^2 not needed explicitly)
	$= 6051/16\,000$ *or* $0{\cdot}3782$ *or* $0{\cdot}6150^2$	**A1**	
	$t_{16,\,0.95} = 1{\cdot}746$	***B1**	State or use correct tabular *t* value
	[–] $t = (\bar{y} - \bar{x}) / s\sqrt{(1/8 + 1/10)} = 1{\cdot}18$	**M1 A1**	Find value of *t* (or can compare $\bar{y} - \bar{x} = 0{\cdot}345$ with $0{\cdot}509$)
	$t < 1{\cdot}75$ so mean masses are the same (AEF)	**DB1 FT**	Correct conclusion (FT on *t*, dep ***B1**)
	SR: $Z = (\bar{y} - \bar{x}) / \sqrt{(s_X^2/8 + s_Y^2/10)}$ $= 0{\cdot}345 / \sqrt{(0{\cdot}078)} = 1{\cdot}20$	**(B1)**	SR: Implicitly taking s_X^2, s_Y^2 as unequal popln. variances (may also earn first **B1 B1 M1**)
	$Z < 1{\cdot}645$ so mean masses are the same (AEF)	**(B1FT)**	Comparison with $Z_{0.95}$ and conclusion (FT on *Z*) (can earn at most 5/10)
	Total:	**10**	

Q51.

6	$(10 - 4^2/N + 102 - 8^2/2N) / (N + 2N - 2)$ (AEF)	**M1 A1**	State or find expression for pooled estimate of σ^2 (confusing biased/unbiased estimates may still earn **M1**)
	$112 - 48/N = 10\,(3N - 2)$, $15N^2 - 66N + 24 = 0$	**M1 A1**	Equate to 10 and rearrange as quadratic
	$N = (66 \pm 54) / 30 = 4$	**A1**	Solve quadratic for *N*, rejecting root $0{\cdot}4$
	Total:	**5**	

Q52.

8	$\bar{x} = 28{\cdot}32$ *and* $\bar{y} = 22{\cdot}2$	**B1**	Find both sample means
	$s_A^2 = (41\ 100 - 1416^2/50) / 49$ *and* $s_B^2 = (20\ 140 - 888^2/40) / 39$	**M1**	Estimate both population variances
	$s_A^2 = 20{\cdot}39$ *and* $s_B^2 = 164/15$ *or* $10{\cdot}93$ (to 3 s.f.)	**A1**	(allow biased here: $19{\cdot}98$ *and* $10{\cdot}66$)
	EITHER: $s^2 = s_A^2/50 + s_B^2/40 = 0{\cdot}681$ *or* $0{\cdot}825^2$ ($19{\cdot}98/50 + 10{\cdot}66/40 = 0{\cdot}666$ is M1 A0, max 8/9)	**(M1 A1**	*EITHER:* Estimate combined variance (if biased values used wrongly here, giving $zs = 1{\cdot}60$, only this A1 is lost)
	$\bar{x} - \bar{y} \pm z\,s$	**M1**	Find confidence interval for difference
	$z_{0.975} = 1{\cdot}96$	**A1**	Use appropriate tabular value
	$z\,s = 1{\cdot}62$	**A1**	Evaluate semi-interval length
	$6{\cdot}12 \pm 1{\cdot}62$ *or* $[4{\cdot}50, 7{\cdot}74]$	**A1)**	State confidence interval (in either form)

	OR: Assume equal [population] variances $s^2 = (49\ s_A^2 + 39\ s_B^2) / 88$ *or* $(41\ 100 - 1416^2/50 + 20\ 140 - 888^2/40) / 88$	**(B1**	*OR:* State assumption Find pooled estimate of common variance (**M1 A1** for s_A^2 and s_B^2 may be implied here)
	$= 16{\cdot}2$ *or* $4{\cdot}02^2$	**B1**	
	$\bar{x} - \bar{y} \pm z\,s\,\sqrt{(1/50 + 1/40)}$	**M1**	Find confidence interval for difference
	$z_{0.975} = 1{\cdot}96$	**A1**	Use appropriate tabular z-value (or appropriate t-value from calculator or interpolation)
	($t_{88,\,0.975} = 1{\cdot}9873$ *or* $1{\cdot}99$) $1{\cdot}67$ ($1{\cdot}70$)	**A1**	Evaluate semi-interval length
	$6{\cdot}12 \pm 1{\cdot}67$ *or* $[4{\cdot}45, 7{\cdot}79]$ ($6{\cdot}12 \pm 1{\cdot}70$ *or* $[4{\cdot}42, 7{\cdot}82]$)	**A1)**	Evaluate confidence interval (in either form)
	Total:	**9**	

Q53.

10(i)	$\bar{x} = 0{\cdot}91$, $\bar{y} = 1{\cdot}205$ $s_X^2 = 19{\cdot}56 / 49$ [= 489/1225 *or* 0·3992] *and* $s_Y^2 = 30{\cdot}25 / 59$ [= 121/236 *or* 0·5127]	**B1** **M1**	Find both sample means Estimate both population variances (allow biased here: 0·3912 *and* 0·5042)
	EITHER: $s^2 = s_X^2/50 + s_Y^2/60$	**(M1**	Estimate or imply combined variance
	= 0·01653 *or* $0{\cdot}1286^2$ (to 3 s.f. throughout)	**A1**	
	[±] $(\bar{y} - \bar{x}) \pm z\,s$	**M1**	Find confidence interval for difference $Y - X$ *or* $X - Y$
	$z_{0.95} = 1{\cdot}645$ [±] 0·295 ± 0·211 (allow 0·212)	**A1**	Use appropriate tabular value Evaluate confidence interval (either form)
	or [±] [0·084, 0·506] (allow [±] [0·083, 0·507])	**A1)**	
	OR: Assume equal [population] variances $s^2 = (49\,s_X^2 + 59\,s_Y^2) / 108$ *or* (19·56 + 30·25) / 108	**(B1**	State assumption Find or imply pooled estimate of common variance (note s_X^2 and s_Y^2 not needed explicitly so first M1 may be implied by result)
	= 4981/10800 *or* 0·461 *or* $0{\cdot}679^2$	**B1**	
	[±] $(\bar{y} - \bar{x}) \pm z\,s\,\sqrt{(1/50 + 1/60)}$	**M1**	Find confidence interval for difference $Y - X$ *or* $X - Y$
	$z_{0.95} = 1{\cdot}645$	**A1**	Use appropriate tabular z-value (or t-value from calculator)
	[±] 0·295 ± 0·214 *or* [±] [0·081, 0·509]	**A1)**	Evaluate confidence interval (either form)
		7	

10(ii)	$z = (1{\cdot}205 - 0{\cdot}91) / s = 0{\cdot}295 / s$ = 2·29[4] [or 2·26[9]] (to 3 s.f.)	**M1 A1**	Find value of z (either sign)
	(z) = 0·989[1] [or 0·988[4]]	**A1**	Find $\Phi(z)$
	100 × (1 – 0·989) × 2 = 2·2 [or 2·3] (to 1 d.p.)	**M1 A1**	Find limiting value for α, based on two-tail test (**M0** for basing on one-tail test)
	$\alpha <$ (*or* ⩽) 2·2 [or 2·3]	**A1**	Find set of possible values of α (Treat α instead of α% as misread)
		6	

Q54.

7	$\bar{x}$ = 681·6/7 = 97·37 (allow 97·4 for this B1)	**B1**	Find sample mean
	$s^2 = (66\,536{\cdot}1 - 681{\cdot}6^2/7) / 6$ [= 27·96 *or* $5{\cdot}287^2$]	**M1**	Estimate population variance (allow biased here: 23·96 *or* $4{\cdot}895^2$)
	H_0: $\mu = 94$, H_1: $\mu \neq 94$ (AEF)	**B1**	State hypotheses (B0 for $\bar{x}$ ...)
	$t = (\bar{x} - 94)/(s/\sqrt{7}) = 1{\cdot}69$	**M1 A1**	Find value of t
	$t_{7,\,0.95} = 1{\cdot}94$[3]	**B1**	State or use correct tabular t-value (or can compare $\bar{x}$ with 94 + 3·88 = 97·88)
	[Accept H_0:] Mean mass is equal to 94 kg (AEF)	**B1√**	Consistent conclusion (FT on both t-values)
		7	

2. Inference Using Normal and t-Distributions

Q55.

10(i)	H_0: $\mu_P = \mu_Q$, H_1: $\mu_P > \mu_Q$	**B1**	State hypotheses (B0 for $\bar{x}$ …)
	$s_P^2 = (138\,200 - 2620^2/50) / 49 = 18{\cdot}61$ *or* 912/49 *and* $s_Q^2 = (157\,000 - 3060^2/60) / 59 = 15{\cdot}93$ *or* 940/59	**M1 A1**	Estimate both popn. variances (to 3 s.f.) (allow biased here: 18·24 *and* 15·67)
	$s^2 = s_P^2/50 + s_Q^2/60 = 0{\cdot}6378$ *or* $0{\cdot}7986^2$	**M1 A1**	Estimate combined variance (to 3 s.f.; may be implicit)
	$z_{0.9} = 1{\cdot}28$[2]	***B1**	State or use correct tabular z value
	$z = (\bar{y} - \bar{x}) / s = (52{\cdot}4 - 51) / s = 1{\cdot}75$ $z >$ tabular value so [accept H_1 and]	**M1 A1**	Calculate value of z (or $-z$) (or can compare $\bar{y} - \bar{x} = 1{\cdot}02$ with 1·4)
	college *P* students take more time (AEF)	**B1√**	Correctly stated conclusion (FT on z, dep *B1)
	[**SC**: $s^2 = (912 + 940)/108 = 17{\cdot}15$ *or* $4{\cdot}141^2$ $z = 1{\cdot}4 / s\sqrt{(1/50 + 1/60)} = 1{\cdot}77$]		**SC:** Using pooled estimate of common variance can earn B1 M1 A1 (may be implied) M0 *B1 M0 B1√ (max 5/9)
		9	
10(ii)	$\bar{x} - \bar{y} \pm z\,s$ (*or* $\bar{y} - \bar{x} \pm z\,s$)	**M1**	Find confidence interval for difference
	$z_{0.95} = 1{\cdot}64$[5]	**A1**	Use appropriate tabular value
	$1{\cdot}4 \pm 1{\cdot}31$ *or* [0·09, 2·71] *or* $-1{\cdot}4 \pm 1{\cdot}31$ *or* [–2·71, –0·09] [**SC**: e.g. $\bar{x} - \bar{y} \pm z\,s\sqrt{(1/50 + 1/60)} = 1{\cdot}4 \pm 1{\cdot}30$]	**A1**	Evaluate confidence interval (either form) **SC**: Allow M1 A1 if common variance used (max 2/3)
		3	

Q56.

10(i)	H_0: $\mu_x - \mu_y = 1$, H_1: $\mu_x - \mu_y > 1$ (AEF)	**B1**	State both hypotheses (B0 for $\bar{x}$ …)
	d_i: 1·1 2·1 0·7 0·4 2·0 2·8 0·2 1·9 1·4	**M1**	Consider differences d_i, e.g. $x_i - y_i$
	$\bar{d} = 12{\cdot}6 / 9 = 1{\cdot}4$	**B1**	Find sample mean
	$s^2 = (23{\cdot}72 - 12{\cdot}6^2/9) / 8$ [$= 0{\cdot}76$ *or* $0{\cdot}8718^2$]	**M1**	Estimate population variance (allow biased here: [$0{\cdot}6756$ *or* $0{\cdot}8219^2$])
	$t_{8,\,0.9} = 1{\cdot}397$ *or* 1·40	**B1**	State or use correct tabular t-value
	$t = (\bar{d} - 1) / (s/\sqrt{9}) = 1{\cdot}38$ [Accept H_0:] No evidence for coach's belief	**M1 A1**	Find t (or compare $\bar{d} - 1 = 0{\cdot}4$ with $t_{8,\,0.9}\,s/\sqrt{9} = 0{\cdot}406$) Consistent conclusion (FT on both t-values)
	or time will not decrease by more than 1 s (AEF)	**B1√**	
			SC Wrong (hypothesis) test can earn only B1 for hypotheses B1√ for conclusion (max 2/8)
		8	
10(ii)	$s_y^2 = (11\,395{\cdot}79 - 319{\cdot}5^2/9) / 8$ $= 6{\cdot}693$ *or* $2{\cdot}587^2$ (to 3 sf)	**B1**	Estimate population variance using y's sample (allow use of biased here: 5·950 *or* $2{\cdot}439^2$)
	$319{\cdot}5/9 \pm t\sqrt{(s_y^2/9)}$	**M1**	Find confidence interval (M0 if s^2 not s_y^2)
	$t_{8,\,0.975} = 2{\cdot}306$ *or* 2·31	**A1**	State or use correct tabular value of t
	$35{\cdot}5 \pm 1{\cdot}99$ *or* [33·5, 37·5] (allow $35{\cdot}5 \pm 2{\cdot}0$)	**A1**	Evaluate C.I. (either form) **SC** B1 (max 1/4) for $1{\cdot}4 \pm 0{\cdot}67$ *or* [0·73, 2·07]
		4	

Q57.

8(i)	$s_A^2 = (531\,000 - 5120^2/50) / 49$ *and* $s_B^2 = (375\,135 - 3760^2/40) / 39$ [= 6712/49] [= 21695/39]	**M1**	Estimate both population variances
	$s_A^2 = 136.98$ *and* $s_B^2 = 556.28$ (to 3 s.f. throughout)	**A1**	(allow biased here: 134.24 *and* 542.4)
	EITHER: $s^2 = s_A^2/50 + s_B^2/40 = 16.65$ *or* 4.08^2	**M1A1**	*EITHER:* Estimate combined variance
	$z_{0.975} = 1.96$	***B1**	State or use correct tabular z (or t) value
	$z = (\bar{y} - \bar{x}) / s = (102.4 - 94) / s = 2.06$ $z >$ tabular value [accept H_1]	**M1A1**	Calculate value of z (or $-z$) (or can compare $\bar{y} - \bar{x} = 8.4$ with 8.0)
	so $\mu_A \neq \mu_B$ *or* salaries do differ (AEF)	**B1**	Correct conclusion (**FT** on z, dep *B1)
	OR: Assume equal [population] variances $s^2 = (49\, s_A^2 + 39\, s_B^2) / 88$ *or* $(531\,000 - 5120^2/50 + 375\,135 - 3760^2/40) / 88$	**(B1**	*OR:* State assumption about variances (may be in part (ii)) Find pooled estimate of common variance (M1 A1 for s_A^2 and s_B^2 may be implied here)
	$= 28407/88$ *or* 322.8 *or* 17.97^2	**B1**	
	$z_{0.975} = 1.96$	***B1**	State or use correct tabular z (or t) value
	$z = 8.4 / s\sqrt{(1/50 + 1/40)} = 2.20$	**M1A1**	Calculate value of z (or $-z$) (or can compare $\bar{y} - \bar{x} = 8.4$ with 7.47)
	$z >$ tabular value [accept H_1] so $\mu_A \neq \mu_B$ *or* salaries do differ (AEF)	**B1)**	Correct conclusion (**FT** on z, dep *B1)
		8	

8(ii)	No assumption [of normality] needed since large samples *or* due to central limit theorem	**B1**	Valid comment on required assumptions (AEF) (assumption of equal variances may earn B1 above)
		1	

Q58.

9(i)	$\bar{x} = 13.88/8 = 1.735$ (allow 1.73 or 1.74 for this B1)	**B1**	Find sample mean
	$s^2 = (24.1182 - 13.88^2/8) / 7$ $[= 13/2500$ *or* 0.0052 *or* $0.07211^2]$	**M1**	Estimate population variance (allow biased here: 0.00455 *or* 0.06745^2)
	$H_0: \mu = 1.7$, $H_1: \mu > 1.7$ (AEF)	**B1**	State hypotheses (B0 for $\bar{x}$ ...)
	$t = (\bar{x} - 1.7) / (s/\sqrt{8}) = 1.37$	**M1A1**	Find value of t
	$t_{7,\,0.95} = 1.89$[5]	**B1**	State or use correct tabular t-value (or can compare $\bar{x}$ with $1.7 + 0.048 = 1.75$)
	[Accept H_0:] Mean ht. not greater than 1.7 m (AEF)	**B1**	Consistent conclusion (**FT** on both t-values)
		7	
9(ii)	$13.88/8 \pm t\sqrt{(s^2/8)}$	**M1**	Find confidence interval
	$t_{7,\,0.975} = 2.36$[5]	**A1**	State or use correct tabular value of t
	1.73[5] ± 0.06 *or* [1.67, 1.80]	**A1**	Evaluate confidence interval (either form)
		3	

Q59.

6	$\bar{x} = 90.3 / 8 = 11.2875$ (to 4 s.f.)	**B1**	Find sample mean
	$s^2 = (1043.67 - 90.3^2/8) / 7$	**M1**	Estimate population variance
	$= 19\,527 / 5600$ *or* 3.487 [*or* 1.867^2] (to 4 s.f.)	**A1**	(allow biased here: 3.051 *or* 1.747^2)
	$90.3 / 8 \pm t\sqrt{(s^2/8)}$	**M1**	Find confidence interval
	$t_{7,\,0.975} = 2.365$ (to 4 s.f.)	**A1**	State or use correct tabular value of t
	11.3 ± 1.6 *or* [9.7, 12.8[5]]	**A1**	Evaluate C.I. (either form)
		6	

Q60.

11B(i)	H_0: $\mu_x = \mu_y$, H_1: $\mu_x > \mu_y$ (or in terms of μ_A, μ_B)	**B1**	State hypotheses (B0 for $\bar{x}$ …)
	$s_x^2 = (14.1775 - 10.56^2/8) / 7 = 0.03404$ *and* $s_y^2 = (15.894 - 12.39^2/10) / 9 = 0.06031$ (to 3 s.f.)	**M1A1**	Estimate both population variances (allow biased here: 0.02979 *and* 0.05428)
	$s^2 = (7 s_x^2 + 9 s_y^2) / 16$ *or* $(14.1775 - 10.56^2/8 + 15.894 - 12.39^2/10) / 16$ $= 0.78\,109 / 16$ *or* 0.04882 *or* 0.22095^2 (to 3 s.f.)	**M1A1**	Find pooled estimate of common variance (M1 A1 for s_x^2 and s_y^2 may be implied here)
	$t_{16,\,0.9} = 1.337$ (to 3 s.f.)	***B1**	State or use correct tabular t value
	$t = (1.32 - 1.239) / s\sqrt{(1/10 + 1/8)} = 0.773$	**M1A1**	Calculate value of t (or $-t$) (or can compare $\bar{y} - \bar{x} = 0.081$ with 0.140)
	$t <$ tabular value so claim not justified *or* A's not heavier than B's (AEF)	**B1**	Correct conclusion (**FT** on t, dep *B1)
	SC: $z = (1.32 - 1.239) / \sqrt{(s_x^2/8 + s_y^2/10)}$ $= 0.081 / \sqrt{(0.078)} = 0.799$	**(B1**	**SC:** Implicitly taking s_x^2, s_y^2 as unequal popln. variances (may also earn first B1 M1 A1)
	$z < 1.282$ so claim is not justified (AEF)	**B1)**	Comparison with $z_{0.9}$ and conclusion (**FT** on z; max 5/9)
		9	
11B(ii)	$\bar{x} = 1.28$ *and* $s^2 = 0.294 / 7$ [$= 0.042$ (*or* 0.205^2)]	**M1**	Find sample mean & estimate popn. var (allow M1 if 0.294 / 8)
	$t = (1.28 - p) / \sqrt{(0.042/8)}$ (AEF)	**M1A1**	Find value of t
	$t > 1.415$ (< is A0), $p < 1.28 - 0.1025$, $p_{max} = 1.18$	**M1A1**	Find p_{max} by comparison with tabular value, here $t_{7,\,0.9}$
		5	

Q61.

8	H_0: $\mu_b - \mu_e = 0{\cdot}05$, H_1: $\mu_b - \mu_e > 0{\cdot}05$ (AEF)	**B1**	State both hypotheses (B0 for $\bar{x}$ …)
	d_i: 0·10 0·07 0·03 0·03 0·21 0·19 (or in sec)	**M1**	Consider differences d_i from e.g. $x_b - y_e$
	$\bar{d} = 0{\cdot}63 / 6 = 0{\cdot}105$ (or 6·3 sec)	**B1**	Find sample mean
	$s^2 = (0{\cdot}0969 - 0{\cdot}63^2/6) / 5$ [= 123/20 000 *or* 0·00615 *or* $0{\cdot}0784^2$](or 22·14)	**M1**	Estimate population variance (allow biased here: [41/8000 *or* 0·005125 *or* $0{\cdot}0716^2$])
	$t_{5,\,0.9} = 1{\cdot}476$ (to 3 sf)	**B1**	State or use correct tabular *t*-value
	$t = (\bar{d} - 0{\cdot}05) / (s/\sqrt{6}) = 1{\cdot}72$	**M1 A1**	Find value of *t* (or compare $\bar{d} - 0{\cdot}05 = 0{\cdot}055$ with $(t_{5,\,0.9})s/\sqrt{6} = 0{\cdot}0473$)
	[Reject H_0:] Evidence for organiser's belief *or* times improve by more than 0·05 min (AEF)	**B1**	**FT** on both *t*-values Consistent conclusion
			SC Wrong type of hypothesis test can earn only B1 for hypotheses B1FT for conclusion (max 2/8)
		8	

Q62.

11O(i)	$ts_A / \sqrt{8} = ½ (16{\cdot}7 - 13{\cdot}5)$ [= 1·6]	**M1**	Relate s_A to semi-width of confidence interval
	$t_{7,\,0.975} = 2{\cdot}365$ (to 3 s.f.)	**A1**	State or use correct tabular *t* value
	[$s_A = \sqrt{8} \times 1{\cdot}6 / 2{\cdot}365 = 1{\cdot}9135$], $s_A^2 = 3{\cdot}66$[16]	**A1**	Hence find unbiased estimate of *A*'s population variance
		3	
11O(ii)	H_0: $\mu_A = \mu_B$, H_1: $\mu_A > \mu_B$ (AEF)	**B1**	State hypotheses (B0 for $\bar{x}$ …)
	[$\bar{x}_A = 15.1$], $\bar{x}_B = 85.2 / 6 = 14.2$	**B1**	Find sample mean for *B*
	$s_B^2 = (1221{\cdot}06 - 85{\cdot}2^2/6) / 5$ = 561/250 *or* 2·244 *or* $1{\cdot}498^2$ (all to 3 s.f.)	**M1**	Estimate or imply population variance for *B* (allow biased here: 1·87 *or* $1{\cdot}367^2$)
	$s^2 = (7\,s_A^2 + 5\,s_B^2) / 12 = 3{\cdot}0709$ *or* $1{\cdot}752^2$	**M1 A1**	Estimate (pooled) common variance (s_B^2 not needed explicitly)
	$t_{12,\,0.95} = 1{\cdot}782$	**B1**	State or use correct tabular *t* value
	[-] $t = (\bar{x}_A - \bar{x}_B) / (s\sqrt{(1/8 + 1/6)}) = 0{\cdot}951$ $t < 1{\cdot}78$ so [accept H_0]	**M1 A1**	Find value of *t* (or can compare $\bar{x}_A - \bar{x}_B = 0{\cdot}9$ with 1·69) Correct conclusion
	mean mass of *B* not less than mean mass of *A* (AEF)	**B1**	
		9	
			SC1: Implicitly taking s_A^2, s_B^2 as unequal population variances (may also earn first B1 B1 M1) $z = (\bar{x}_A - \bar{x}_B) / \sqrt{(s_A^2/8 + s_B^2/6)}$ $= 0{\cdot}9 / \sqrt{(0{\cdot}8317)} = 0{\cdot}987$ $z < 1{\cdot}645$ so
			DepSC1: mean mass of *B* not less than mean mass of *A* (AEF) Comparison with $z_{0.95}$ and conclusion (FT on *z*) (can earn at most 5/9)

Q63.

9(i)	$\bar{x} = 0{\cdot}785$	**B1**	Find sample mean
	$s^2 = (6{\cdot}19 - 7{\cdot}85^2/10) / 9$ [= 37/12 000 *or* 0·003083 *or* $0{\cdot}05553^2$]	**M1**	Estimate population variance (allow biased here: 0·002775 *or* $0{\cdot}05268^2$)
	H_0: $\mu = 0{\cdot}75$, H_1: $\mu > 0{\cdot}75$ (AEF)	**B1**	State hypotheses (B0 for $\bar{x}$ ...)
	$t_{9,\,0.95} = 1{\cdot}83$[3]	**B1**	State or use correct tabular *t*-value
	$t = (\bar{x} - 0{\cdot}75)/(s/\sqrt{10}) = 1{\cdot}99$ [Reject H_0 and accept H_1]	**M1 A1**	Find value of *t* (or compare $\bar{x}$ with 0·75 + 0·032 = 0·782) Consistent conclusion
	Claim (of yield per plant increased) is justified (AEF)	**B1**	**FT** on both *t*-values (must be *t* value)
		7	
9(ii)	$\bar{x} \pm t\sqrt{(s^2/10)}$	**M1**	Find confidence interval (must be a *t* value)
	$t_{9,\,0.975} = 2{\cdot}26$[2]	**B1**	State or use correct tabular value of *t*
	0·785 ± 0·04 *or* [0·745, 0·825]	**A1**	Evaluate confidence interval (either form)
		3	

Q64.

11O	$\bar{x} - \bar{y} = 0{\cdot}448 - 0{\cdot}48 = [-]\,0{\cdot}032$	**B1**	Find difference in sample means (either sign; may be implied)
	$s_X^2 = (10{\cdot}1 - 22{\cdot}4^2/50) / 49 = 0{\cdot}0648 / 49$ [= 0·001322]	**M1 A1**	Estimate both population variances (may be implied)
	$s_Y^2 = (16{\cdot}3 - 28{\cdot}8^2/60) / 59 = 2{\cdot}476 / 59$ [= 0·04197]	**A1**	(allow biased here: 0·001296 *and* 0·04127)
	$s_C^2 = s_X^2/50 + s_Y^2/60$	**M1**	Estimate combined variance
	= 0·0007259 *or* $0{\cdot}02694^2$ (to 3 s.f. throughout)	**A1**	
	$z = 0{\cdot}032 / s_C$ [or $0{\cdot}032 / s_P\sqrt{(1/50 + 1/60)}$] = 1·188 [or 1·0895]	**M1 A1**	Find value of *z* (either sign)
	$\Phi(z) = 0{\cdot}8826$ [or 0·8620]	**A1**	Find $\Phi(z)$
	$100 \times (1 - \Phi(z)) = 11{\cdot}7$ [or 13·8]	**M1 A1**	Find limiting value for α, based on one-tail test (M0 for basing on two-tail test)
	$\alpha >$ (*or* $\geqslant$) 11·7 [or 13·8]	**A1**	Find set of possible values of α Allow 11.8 if 11.74 seen (Misreading α% as α loses only last A1)

11O	**Alternative method for question 11O**		
	$\bar{x} - \bar{y} = 0{\cdot}448 - 0{\cdot}48 = [-]\,0{\cdot}032$	**B1**	Find difference in sample means (either sign; may be implied)
	$s_X^2 = (10{\cdot}1 - 22{\cdot}4^2/50) / 49 = 0{\cdot}0648 / 49$ [= 0·001322]	**M1 A1**	Estimate both population variances (may be implied)
	$s_Y^2 = (16{\cdot}3 - 28{\cdot}8^2/60) / 59 = 2{\cdot}476 / 59$ [= 0·04197]	**A1**	(allow biased here: 0·001296 *and* 0·04127)
	Assume equal [population] variances	**B1**	State assumption
	$s_P^2 = (49\, s_X^2 + 59\, s_Y^2) / 108$ *or* $(0{\cdot}0648 + 2{\cdot}476) / 108$ $= 0{\cdot}02353$ *or* $0{\cdot}1534^2$	**B1**	(Find pooled estimate of common variance s_X^2 and s_Y^2 not needed explicitly so may be implied by result)
	$z = 0{\cdot}032 / s_C$ [or $0{\cdot}032 / s_P\sqrt{(1/50 + 1/60)}$] $= 1{\cdot}188$ [or 1·0895]	**M1 A1**	Find value of z (either sign)
	$\Phi(z) = 0{\cdot}8826$ [or 0·8620]	**A1**	Find $\Phi(z)$
	$100 \times (1 - \Phi(z)) = 11{\cdot}7$ [or 13·8]	**M1 A1**	Find limiting value for α, based on one-tail test (M0 for basing on two-tail test)
	$\alpha >$ (*or* $\geqslant$) 11·7 [or 13·8]	**A1**	Find set of possible values of α Allow 11.8 if 11.74 seen (Misreading α% as α loses only last A1)
		12	

Q65.

6(i)	$t\sqrt{(s^2/9)} = \frac{1}{2}(1{\cdot}85 - 1{\cdot}65)$ [= 0·1]	**M1**	Find estimate s^2 of population variance (must be t)
	$t_{8,\,0.975} = 2{\cdot}306$ (to 3 s.f.)	**A1**	Use of correct tabular t-value
	$s^2 = 9 \times 0{\cdot}04337^2 = 0{\cdot}0169$ *or* $0{\cdot}130[1]^2$	**A1**	
		3	
6(ii)	$\bar{x} = \frac{1}{2}(1{\cdot}65 + 1{\cdot}85) = 1{\cdot}75$ or $\Sigma x = 9 \times 1{\cdot}75 = 15{\cdot}75$	**M1 A1**	Find sample mean $\bar{x}$
	$s^2 = (\Sigma x^2 - 9 \times \bar{x}^2) / 8$ or $\{\Sigma x^2 - (\Sigma x)^2 / 9\} / 8$	**M1**	*or* Σx
	$\Sigma x^2 = 8 \times 0{\cdot}0169 + 15{\cdot}75^2/9 = 27{\cdot}7$	**A1**	Find Σx^2 from s^2
		4	

Q66.

8	$\bar{x}_A = 32{\cdot}4 / 8 = 4{\cdot}05$	**B1**	Find sample mean for A
	$s_A^2 = (131{\cdot}82 - 32{\cdot}4^2/8) / 7$ $s_A^2 = 3/35$ (or $0{\cdot}08571$ or $0{\cdot}2928^2$ both to 3 s.f.)	**M1**	Estimate or imply popln. variance for A
	H_0: $\mu_A = \mu_B$, H_1: $\mu_A \neq \mu_B$	**B1**	State hypotheses. AEF
	$s^2 = (7\, s_A^2 + 9\, s_B^2) / 16 = 0{\cdot}12497$ or $0{\cdot}3535^2$ or $\frac{3999}{32000}$	**M1 A1**	Estimate (pooled) common variance
	$t_{16,\,0.95} = 1{\cdot}746$	**B1***	State or use correct tabular t value
	$[-]\, t = (x_A - \bar{x}_B) / s\sqrt{(1/8 + 1/10)}$	**M1**	
	$= 0{\cdot}27 / 0{\cdot}1677 = 1{\cdot}61$	**A1**	Find value of t (or can compare $\bar{x}_A - \bar{x}_B = 0{\cdot}27$ with 0·293)
	$t < 1{\cdot}75$ so [accept H_0] mean masses are the same	**DB1**	Correct conclusion (FT on t, dep B1*). AEF
		9	

2. Inference Using Normal and t-Distributions

Q67.

4	$H_0: \mu_x = \mu_y$ $H_1: \mu_x \neq \mu_y$	**B1**
	$s_x^2 = \frac{1}{49}\left(5.1 - \frac{15.2^2}{50}\right) = 0.0097796$: $s_y^2 = \frac{1}{39}\left(4.8 - \frac{13.4^2}{40}\right) = 0.007974$	**M1A1**
	$s^2 = \frac{0.00977959}{50} + \frac{0.007974}{40} = 0.0003949$	**M1A1**
	$z = \frac{0.304 - 0.335}{\sqrt{0.0003949}} = (-)1.56$	**M1A1**
	Compare with 1.645	**M1**
	Accept H_0: insufficient evidence to reject manager's claim	**A1**
		9

Q68.

5(a)	$\Sigma x = 183.6$, $\Sigma x^2 = 4249.08$, $\bar{x} = 22.95$	**B1**
	$s^2 = \frac{1}{7}\left(4249.08 - \frac{183.6^2}{8}\right) = 5.066$	**M1**
	$H_0: \mu = 22.0$, $H_1: \mu > 22.0$	**B1**
	$t = \frac{22.95 - 22.0}{\sqrt{\frac{s^2}{8}}} = 1.194$	**M1A1**
	Compare t with correct tabular value 1.895	**M1**
	Accept H_0: mean distance thrown is not more than 22.0 m	**A1**
		7
5(b)	$22.95 \pm t\sqrt{\frac{s^2}{8}}$	**M1**
	With $t = 2.365$	**B1**
	[21.1, 24.8]	**A1**
		3

Q69.

2	$s^2 = \frac{10.2}{40} + \frac{11.1}{50} = 0.477$	**M1A1**
	$CI = (24.4 - 17.2) \pm zs$	**M1**
	$= (24.4 - 17.2) \pm 1.645\sqrt{0.477}$	**A1**
	$= [6.06, 8.34]$	**A1**
		5

2. Inference Using Normal and t-Distributions

Q70.

5(a)	$\bar{x} = \frac{828}{10} = 82.8 \quad s^2 = \frac{1}{9}\left(68622 - \frac{828^2}{10}\right) = 7.0667$	**B1**
	$\frac{82.8 - k}{\sqrt{\frac{s^2}{10}}} \geqslant t$ where $t = 1.833$ (**M1** if equality, **A1** for inequality and correct t value)	**M1 A1**
	$k \leqslant 81.259 \quad k \leqslant 81.3$	**A1**
		4
5(b)	H_0: $\mu_A = \mu_B$ H_1: $\mu_A \neq \mu_B$	**B1**
	Pooled variance $= \frac{9 \times 7.0667 + 7 \times 9.966}{10 + 8 - 2} = s_p^2$	**M1**
	$= 8.335$	**A1**
	$t = \frac{82.8 - 79.8}{s_p\sqrt{\frac{1}{10} + \frac{1}{8}}} = 2.19$	**M1A1**
	Compare with 2.12 ($t(16, 0.975)$) and reject H_0	**M1**
	Population means are not the same	**A1**
		7

Q71.

1	$s_x^2 = \frac{1}{39}\left(1587.0 - \frac{242.0^2}{40}\right) = 3.15128 \quad \left(= \frac{1229}{390}\right)$ $s_y^2 = \frac{1}{59}\left(2532.6 - \frac{373.2^2}{60}\right) = 3.58129 \quad \left(= \frac{26412}{7375}\right)$	**M1 A1**	Either unsimplified Both correct
	$s^2 = \frac{3.15128}{40} + \frac{3.58129}{60} = 0.13847$	**M1 A1**	Pooled variance is M0A0 or $s = 0.3721$
	$6.05 - 6.22 \pm zs$	**M1**	FT their s, must be a z value
	$= 0.17 \pm 1.645\sqrt{0.13847}$	**A1**	With 1.645
	$= [-0.442, 0.782]$ or $[-0.782, 0.442]$	**A1**	
		7	

2. Inference Using Normal and t-Distributions

Q72.

4	Assume (population) differences are normally distributed	**B1**	
	H_0: $\mu_X - \mu_Y = 0.3$ H_1: $\mu_X - \mu_Y > 0.3$	**B1**	
	Diff: 0.9 0.4 0.7 0.5–0.2 0–0.1 1.2 0.3	**M1**	Signed differences
	$\Sigma d = 3.7,\ \Sigma d^2 = 3.29$ $\bar{d} = 0.411,\ s^2 = \frac{1}{8}\left(3.29 - \frac{3.7^2}{9}\right) = 0.2211$	**M1**	$s = 0.4702$
	$t = \frac{0.411 - 0.3}{\sqrt{\frac{0.2211}{9}}}$ $= 0.708$ or 0.709	**M1 A1**	
	'0.708' < 1.397	**M1**	Compare *their* 0.708 with 1.397
	Accept H_0 Insufficient evidence to support claim	**A1 FT**	In context, except possibly hypotheses Level of uncertainty in language used. No contradictions
		8	

Q73.

1	$s^2 = \frac{313.1}{10} = 31.3$	**B1**	Can be implied
	$t = \frac{176.2 - 172.5}{s/\sqrt{11}}$	**M1**	
	$= 2.19$	**A1**	
	'2.19' < 2.228	**M1**	Comparison of their t value with 2.228
	accept H_0: insufficient evidence to reject mean height is 172.5 OR sufficient evidence to accept mean height is 172.5	**A1 FT**	Correct conclusion in context, level of uncertainty in language used. No contradictions. CWO Do not accept: 'mean height is 172.5' Do not accept use of symbols that are undefined e.g. insufficient evidence to reject $\mu = 172.5$
		5	

2. Inference Using Normal and t-Distributions

Q74.

6(a)	$\bar{x} = \frac{1}{2}(4.22 + 3.36) = 3.79$	**B1**	
	$4.22 - 3.36 = \frac{2ts}{\sqrt{10}}$	**M1**	Using z value implies M0
	With $t = 1.833$	**A1**	
	$(s = 0.7418)$ variance = 0.55(0)	**A1**	
		4	
6(b)	H_0: $\mu_A = \mu_B$ H_1: $\mu_A > \mu_B$	**B1**	Not $\bar{x}$
	$s^2 = \frac{1}{7}\left(98.02 - \frac{27.86^2}{8}\right) = 0.142(507)$	**M1**	$= \frac{1}{7}(98.02 - 97.02)$
	Pooled estimate $= \frac{9 \times 0.550316 + 7 \times 0.142507}{16}$	**M1**	
	$= 0.372$	**A1**	
	$t = \frac{\frac{27.86}{8} - 3.79}{\sqrt{0.372}\sqrt{\frac{1}{10} + \frac{1}{8}}} = (-)1.063$	**M1A1**	
	'1.063' < 1.337 oe	**M1**	Compare *their* value with 1.337
	Accept H_0 Nassa's claim is not supported	**A1ft**	Correct conclusion in context, level of uncertainty in language used. No contradictions.
		8	

Q75.

1(a)	$\bar{x} = \frac{56.56}{7} = 8.08$	**B1**	Accept unsimplified.
	$s^2 = \frac{1}{6}\left(457.275 - \frac{56.56^2}{7}\right) = 0.045033 \left[= \frac{1351}{30000}\right]$	**M1**	Accept unsimplified.
	$t = \frac{8.08 - 8.22}{\sqrt{\frac{s^2}{7}}} = -1.745$	**M1 A1**	1.74 – 1.75
	Tabular value 1.440 1.745 > 1.440 , so reject H_0	**M1**	Comparison with 1.440 and correct FT conclusion.
	There is insufficient evidence to support the hypothesis that the (population) mean is 8.22 OR sufficient evidence to suggest mean is less than 8.22 OR sufficient evidence to suggest $\mu < 8.22$ OE	**A1**	Correct conclusion, in context, following correct work. Level of uncertainty in language is used.
		6	
1(b)	Underlying distribution is normal/ population is normal	**B1**	X is normal distributed.
		1	

Q76.

4	$s_x^2 = \frac{1}{49}\left(1596 - \frac{282^2}{50}\right) = 0.11265 \left[= \frac{138}{1225}\right]$	**M1**	One correct, unsimplified.
	$s_y^2 = \frac{1}{59}\left(1808 - \frac{328^2}{60}\right) = 0.25311 \left[= \frac{224}{885}\right]$	**A1**	Both correct to 3sf.
	$s^2 = \frac{0.11265}{50} + \frac{0.25312}{60} = 0.006471\ldots$	**M1 A1**	May be implied.
	$z = \frac{\frac{282}{50} - \frac{328}{60}}{\text{their } s}$	**M1**	
	$[z] = 2.155$	**A1**	Tolerance 2.15 – 2.16
	Comparison with tabular value 1.96 OE $2.155 > 1.96$ Reject H_0	**M1**	Comparison with 1.96 OE and conclusion.
	Sufficient evidence to reject population means are equal OR sufficient evidence to accept population means are not equal OR sufficient evidence to suggest $\mu_A \neq \mu_B$ OE	**A1**	Correct conclusion following correct work. Level of uncertainty in language is used.
		8	

Q77.

1	$s^2 = \frac{1}{9}\left(2313 - \frac{152^2}{10}\right) = 0.28889$	**M1**	$\frac{13}{45}$, accept unsimplified.
	$H_0 : \mu = 14.8$ $H_1 : \mu > 14.8$	**B1**	If μ not used, 'population mean' required. Must see 14.8, must be = and >.
	$[t =] \frac{\frac{152}{10} - 14.8}{s/\sqrt{10}}$	**M1**	Using unbiased estimate. *p*-value is 0.0215.
	$[t =] 2.35$	**A1**	Rounds to 2.35.
	Tabular value = 1.833 '2.35' > 1.833 Reject H_0	**M1**	Comparison with 1.833 and correct FT conclusion.
	Sufficient evidence to accept Farmer B's claim oe	**A1**	Correct conclusion, in context, following correct work. Level of uncertainty in language is used. Allow ±1 difference in third significant figure of their *t* for this mark.
		6	

Q78.

3	$\bar{x}=\frac{89}{50}=1.78 \quad \bar{y}=\frac{67.2}{40}=1.68$	**B1**	Implied by 0.1 in the CI formula.
	$\sigma_x^2=\frac{1}{49}\left(159.4-\frac{89^2}{50}\right)=0.02 \quad \left(=\frac{1}{50}\right)$	**M1**	One variance correct unsimplified.
	$\sigma_y^2=\frac{1}{39}\left(113.1-\frac{67.2^2}{40}\right)=0.0052308 \quad \left(=\frac{17}{3250}\right)$	**A1**	Both correctly calculated.
	$\sigma^2=\frac{0.02}{50}+\frac{0.0052308}{40}=\left(\frac{69}{130000}\right)$	**M1**	Unsimplified expression must be seen, 40 and 50 in correct places.
	$[\sigma^2=]\ 0.0005308$	**A1**	May be implied by correct final answer.
	CI : $1.78-1.68\pm z\times\sqrt{0.0005308}$	**M1**	Correct form for CI with a z-value.
	with $z=1.96$	**A1**	
	0.10 ± 0.0452 or $[0.0548, 0.145]$	**A1**	Allow either form.
		8	SC Assuming equal variances award B1M0A0, M1A1 (for use of pooled variance formula, gives 0.0135), M1A1A0.

Q79.

1(a)	$s^2=\frac{78.5}{9}=8.7222$	**B1**	$\frac{157}{18}$
	CI: $25.6\pm t\sqrt{\frac{s^2}{10}}$	**M1**	Correct expression with a t value.
	With $t=1.833$	**A1**	With correct t value.
	$25.6\pm1.71(2)$ or [23.9, 27.3]	**A1**	Accept in either form, ISW. Accept inequality form.
		4	
1(b)	$\frac{25.6-k}{\sqrt{\frac{s^2}{10}}}\geqslant t$	**M1**	
	With $t=-1.383$	**A1**	Allow 1.383.
	26.9	**A1**	CAO
		3	

2. Inference Using Normal and t-Distributions

Q80.

4	$H_0 : \mu_B - \mu_A = 2$ and $H_1 : \mu_B - \mu_A > 2$	**B1**	Or use of μ_d.
	Differences: 5 8 2 0 4 7 −2 1 10 −3	**M1**	Allow one error.
	$\Sigma d = 32, \quad \Sigma d^2 = 272$ $\bar{d} = 3.2, \ s^2 = \frac{1}{9}\left(272 - \frac{32^2}{10}\right) = 18.84 \ \left(= \frac{848}{45}\right)$	**M1**	Sample mean and variance.
	$t = \frac{3.2 - 2}{\sqrt{\frac{s^2}{10}}} = 0.874$	**M1 A1**	
	Compare with tabular value 1.833: 0.874 < 1.833. Accept H_0.	**M1**	Compare their value with 1.833 and conclusion.
	Insufficient evidence to support claim.	**A1**	Correct conclusion, in context, following correct work. Level of uncertainty in language is used.
	Assumption: population differences are normally distributed	**B1**	
		8	

Q81.

1	$s_x^2 = \frac{1}{39}\left(14320 - \frac{752^2}{40}\right) = 4.677 \ = \frac{304}{65}$	**M1**	One correct unsimplified.
	$s_y^2 = \frac{1}{59}\left(40200 - \frac{1548^2}{60}\right) = 4.434 \ = \frac{1308}{295}$	**A1**	Both correct to at least 3sf.
	$s^2 = \frac{4.6769}{40} + \frac{4.4339}{60}$	**M1**	
	$0.1908 \ \left(= \frac{3659}{19175}\right)$	**A1**	
	$\text{CI} = \left(\frac{752}{40} - \frac{1548}{60}\right) \pm 1.96s$	**M1A1**	Correct form with a z-value with 1.96 s can be unsimplified.
	-7.0 ± 0.856 = [−7.86, −6.14]	**A1**	Accept in either form, ISW. Accept [6.14, 7.86]. Accept inequality form.
		7	

Q82.

6	$H_0 : \mu_A = \mu_B$ and $H_1 : \mu_A > \mu_B$	**B1**	
	$\Sigma x = 21.2 \quad \Sigma x^2 = 47.92$ $s_x^2 = \frac{1}{9}\left(47.92 - \frac{21.2^2}{10}\right) = 0.33067 = \frac{124}{375}$ $s_y^2 = \frac{1}{11}\left(53.75 - \frac{24.48^2}{12}\right) = 0.34644 = \frac{9527}{27500}$	**M1 A1**	
	$s^2 = \frac{9\times 0.33067 + 11\times 0.34644}{10+12-2} = 0.3393$	**M1 A1**	
	$t = \frac{2.12 - 2.04}{s\sqrt{\frac{1}{10}+\frac{1}{12}}} = 0.321$	**M1 A1**	Accept 0.321 – 0.322.
	Critical value: 1.325 $0.321 < 1.325$ accept H_0.	**M1**	Compare their value with 1.325 and correct FT conclusion.
	Insufficient evidence that mean of A is greater than mean of B	**A1**	Correct conclusion, in context, following correct work. Level of uncertainty in language is used.
	EITHER: Distributions are normal and equal variances OR: Distributions are normal	**B1**	Assumptions consistent with method used. Accept 'population is normal'. 'Distribution of difference of means is normal'.
	Alternative method for question 6		
	$H_0 : \mu_A = \mu_B$ and $H_1 : \mu_A > \mu_B$	**B1**	

	$\Sigma x = 21.2 \quad \Sigma x^2 = 47.92$ $s_x^2 = \frac{1}{9}\left(47.92 - \frac{21.2^2}{10}\right) = 0.33067 = \frac{124}{375}$ $s_y^2 = \frac{1}{11}\left(53.75 - \frac{24.48^2}{12}\right) = 0.34644 = \frac{9527}{27500}$	**M1** **A1**	For one correct unsimplified. For both correct.
	$s^2 = \frac{0.33067}{10} + \frac{0.34644}{12} = 0.061936$	**M1 A1**	
	$t = \frac{2.12 - 2.04}{s} = 0.321(5)$	**M1 A1**	Accept 0.321 – 0.322.
	Critical value: 1.325 $0.321 < 1.325$ accept H_0.	**M1**	Compare their value with 1.325 and correct FT conclusion.
	Insufficient evidence that mean of A is greater than mean of B.	**A1**	Correct conclusion, in context, following correct work. Level of uncertainty in language is used.
	EITHER: Distributions are normal and equal variances OR: Distributions are normal.	**B1**	Assumptions consistent with method used. Accept 'population is normal'. 'Distribution of difference of means is normal'.
		10	

2. Inference Using Normal and t-Distributions

Q83.

1(a)	Differences: 0.6 1.3 0.0 0.7 0.6 0.6 1.8 –0.2 –0.2 0.4	**M1**	Attempt at differences, allow one error.
	$\Sigma d = 5.6 \quad \Sigma d^2 = 6.74$ $s^2 = \frac{1}{9}\left(6.74 - \frac{5.6^2}{10}\right) = 0.400(4)$	**M1**	
	H_0: $\mu_B - \mu_A = 0.4$ and H_1: $\mu_B - \mu_A > 0.4$	**B1**	Allow use of μ_d
	$t = \frac{0.56 - 0.4}{\sqrt{\frac{0.4004}{10}}}$ (0.800)	**M1**	
	$t = 0.7996$ (0.800)	**A1**	
	Critical value = $t_{0.90}(9) = 1.383$ Compare: '0.7996' < 1.383 Accept H_0	**M1**	Compare calculated value with 1.383 and correct FT conclusion.
	Insufficient evidence to support manager's claim.	**A1**	Correct conclusion, in context. Level of uncertainty in language is used.
		7	
1(b)	Distribution of **population** differences is normal	**B1**	
		1	

Q84.

5(a)	$\Sigma x = 44.4, \quad \Sigma x^2 = 247.48$	**B1**	
	$s^2 = \frac{1}{7}\left(247.48 - \frac{44.4^2}{8}\right) = 0.1514\left(= \frac{53}{350}\right)$	**M1**	
	CI: $\frac{44.4}{8} \pm 1.895\sqrt{\frac{0.1514}{8}}$	**M1**	Correct formula with a t value.
	1.895	**B1**	1.895 used.
	$5.55 \pm 0.261 = [5.29, 5.81]$	**A1**	Either form.
		5	
5(b)	$H_0 : \mu = 5.9$ and $H_1 : \mu < 5.9$	**B1**	
	$t = \frac{5.55 - 5.9}{\sqrt{\frac{s^2}{8}}} = -2.544$	**M1**	
	Critical value $t_{0.975}(7) = 2.365$ Compare '– 2.544' < – 2.365, reject H_0.	**M1**	Compare calculated value with -2.365 and correct FT conclusion.
	Sufficient evidence to support Raman's claim.	**A1**	Correct conclusion, in context. Level of uncertainty in language is used.
		4	

Q85.

1	$s^2 = \frac{94.5}{7} (= 13.5)$	**B1**	
	CI: $33.8 \pm 2.365\sqrt{\frac{s^2}{8}}$	**M1**	Correct formula with a t value, must be 33.8.
		B1	2.365 used.
	$33.8 \pm 3.072 = [30.7, 36.9]$	**A1**	Either form.
		4	

Q86.

6(a)	$s_x^2 = \frac{1}{39}\left(31400 - \frac{1120^2}{40}\right) = 1.0256$ $s_y^2 = \frac{1}{49}\left(37600 - \frac{1370^2}{50}\right) = 1.2653$	**M1**	$\frac{40}{39}$ $\frac{62}{49}$ Both required
	$s^2 = \frac{1.0256}{40} + \frac{1.2653}{50}$	**M1**	$\frac{2434}{47775}$ Condone $s^2 = \frac{1.03}{40} + \frac{1.27}{50}$ or 0.05115.
	$s^2 = 0.050947$	**A1**	May be implied by 2.66 as z value.
	$H_0 : \mu_a - \mu_b = 0$ and $H_1 : \mu_a - \mu_b \neq 0$	**B1**	
	$z = \frac{\frac{1120}{40} - \frac{1370}{50}}{s} = \frac{0.6}{\sqrt{0.050947}}$	**M1**	
	$z = 2.66$	**A1**	
	Critical value of $z = 2.326$ $2.66 > 2.326$ Reject H_0	**M1**	Compare their z value with 2.326 and correct FT conclusion.
	Sufficient evidence to support difference in **population** means oe	**A1**	Correct conclusion, in context. Level of uncertainty in language is used.
		8	

6(b)	$z = \frac{0.6 - 0.25}{s}$	**M1**	Condone 1.54.
	$z = 1.55$	**A1**	
	0.9392 – 0.9395 or 0.0605 – 0.0608	**A1**	
	$\alpha \geqslant 6.05$	**A1**	Condone 6.05 – 6.08, accept >
		4	

2. Inference Using Normal and t-Distributions

Q87.

1	$\left[\bar{x}=\frac{1625}{50}=32.5\right]\quad s_x^2=\frac{1}{49}\left(53200-\frac{1625^2}{50}\right)=7.908$	**M1**	$\frac{775}{98}$
	$\left[\bar{y}=\frac{1854}{60}=30.9\right]\quad s_y^2=\frac{1}{59}\left(57900-\frac{1854^2}{60}\right)=10.3627$	**A1**	$\frac{3057}{295}$ Both correct
	$s^2=\frac{7.908}{50}+\frac{10.3627}{60}\ [=0.33087]$	**M1 A1**	
	CI: $(32.5-30.9)\pm1.96s$	**M1 A1**	Correct form with a z-value. Correct with 1.96.
	$1.6\pm1.1274=[0.473,2.73]$	**A1**	At least 3sf.
		7	Pooled variance used: M1A1 M0A0 M1A1A0 max 4/7

Q88.

6	$s_x^2=\frac{1}{11}\left(3.26-\frac{6.24^2}{12}\right)=0.001382$ $s_y^2=\frac{1}{9}\left(2.6124-\frac{5.1^2}{10}\right)=0.001267$	**M1**	$\frac{19}{13750}$ $\frac{19}{15000}$
		A1	Both correct.
	$s^2=\frac{11\times0.001382+9\times0.001267}{12+10-2}=0.001330$	**M1 A1**	Find pooled variance.
	$t=\frac{\frac{6.24}{12}-\frac{5.1}{10}}{s\sqrt{\frac{1}{12}+\frac{1}{10}}}=0.640$	**M1 A1**	Use correct formula for t.
	$H_0:\mu_a=\mu_b\quad H_1:\mu_a>\mu_b$	**B1**	
	Critical value is 2.086 ($t_{0.975}(20)$) 0.640 < 2.086 Accept H_0	**M1**	Compare their calculated value with 2.086 and conclusion.
	There is insufficient evidence to support inspector's claim	**A1**	All correct except possibly B1, in context, level of uncertainty in language. 'Prove' scores A0
		9	Pooled variance not used: M1A1 M0A0 M0A0 B1 M1A0 max 4/9

2. Inference Using Normal and t-Distributions

Q89.

1	$2.02=\bar{x}+1.833\sqrt{\frac{s^2}{10}}$ $1.78=\bar{x}-1.833\sqrt{\frac{s^2}{10}}$ Add, $\bar{x}=\frac{2.02+1.78}{2}=1.90$	**M1**	Add.
	$\Sigma x=19$	**A1**	
	Subtract: $2.02-1.78=2\times1.833\sqrt{\frac{s^2}{10}}$	**M1**	Allow 1.372, 1.383, 1.812 instead of 1.833.
	$s^2=0.042859$	**A1**	May be implied. $s=0.20702$
	But $s^2=\frac{1}{9}(\Sigma x^2-\frac{(\Sigma x)^2}{10})$	**M1**	
	$\Sigma x^2=9s^2+\frac{19^2}{10}=36.5$	**A1**	36.486
			Using 1.645: maximum M1A1 M0A0 M1A0 3/6
		6	

Q90.

3	$s_x^2=\frac{1}{49}\left(115.2-\frac{75.5^2}{50}\right)=0.02439$ $s_y^2=\frac{1}{79}\left(172.6-\frac{116.8^2}{80}\right)=0.02623$	**M1**	$\frac{239}{9800}$ $\frac{259}{9875}$
		A1	Both correct
	$s^2=\frac{0.02439}{50}+\frac{0.02623}{80}$ $[=0.00081565]$	**M1 A1**	May be implied by 1.75 for z
	$z=\frac{\frac{75.5}{50}-\frac{116.8}{80}}{s}$	**M1**	
	1.75	**A1**	
	Compare with 1.645: $1.75>1.645$ Reject null hypothesis	**M1**	Using areas, $0.04<0.05$.
	Sufficient evidence to suggest that population mean in country X is greater than population mean in country Y	**A1**	Correct conclusion in context, following correct work, level of uncertainty in language. 'Prove' gives A0.
		8	Pooled variance: $z=1.74$ M1A1 M0A0M0A0 M1A0 max 3/8

3. Chi-squared tests

Q1.

Each of 100 coins is tossed repeatedly until a head is obtained, and the number, x, of tosses needed is recorded. The results are summarised in the following table.

x	1	2	3	4	5	$\geqslant 6$
Frequency	56	20	11	10	3	0

Assume that, for each coin, the probability of obtaining a head has the same value p. State a distribution which could be used to fit the above data, and use the sample mean to obtain 0.5435, correct to 4 decimal places, as an estimate of p. [4]

Carry out a goodness of fit test at the $2\frac{1}{2}\%$ level. [8]

Without further calculation, state two changes which would be necessary if a similar method is to be used on the data above to test whether all the coins are fair. [2]

9231/02/M/J/08 Q11b

Q2.

A sample of 100 observations of the continuous random variable T was obtained and the values are summarised in the following table.

Interval	$1 \leqslant t < 1.5$	$1.5 \leqslant t < 2$	$2 \leqslant t < 2.5$	$2.5 \leqslant t < 3$
Frequency	64	17	16	3

It is required to test the goodness of fit of the distribution with probability density function given by

$$f(t) = \begin{cases} \dfrac{9}{4t^3} & 1 \leqslant t < 3, \\ 0 & \text{otherwise.} \end{cases}$$

The relevant expected values are as follows.

Interval	$1 \leqslant t < 1.5$	$1.5 \leqslant t < 2$	$2 \leqslant t < 2.5$	$2.5 \leqslant t < 3$
Expected frequency	62.5	21.875	10.125	5.5

Show how the expected value 10.125 is obtained. [3]

Carry out the test, at the 10% significance level. [7]

9231/02/O/N/08 Q9

3. Chi-squared tests

Q3.

The proportions of blood types A, B, AB and O in the Australian population are 38%, 10%, 3% and 49% respectively. In order to test whether the population in Sydney conforms to these figures, a random sample of 200 residents is selected. The table shows the observed frequencies of these types in the sample.

Blood Type	A	B	AB	O
Frequency	57	24	9	110

Carry out a suitable test at the 5% significance level. [7]

Find the smallest sample size that could be used for the test. [2]

9231/02/M/J/09 Q9

Q4.

It has been found that 60% of the computer chips produced in a factory are faulty. As part of quality control, 100 samples of 4 chips are selected at random, and each chip is tested. The number of faulty chips in each sample is recorded, with the results given in the following table.

Number of faulty chips	0	1	2	3	4
Number of samples	2	12	27	49	10

The expected values for a binomial distribution with parameters $n = 4$ and $p = 0.6$ are given in the following table.

Number of faulty chips	0	1	2	3	4
Expected value	2.56	15.36	34.56	34.56	12.96

Show how the expected value 34.56 corresponding to 2 faulty chips is obtained. [2]

Carry out a goodness of fit test at the 5% significance level, and state what can be deduced from the outcome of the test. [8]

9231/02/O/N/09 Q9

Q5.

Three new flu vaccines, A, B and C, were tested on 500 volunteers. The vaccines were assigned randomly to the volunteers and 178 received A, 149 received B and 173 received C. During the following year, 30 of the volunteers given A caught flu, 29 of the volunteers given B caught flu, and 16 of the volunteers given C caught flu. Carry out a suitable test for independence at the 5% significance level. [10]

Without using a statistical test, decide which of the vaccines appears to be most effective. [3]

9231/21/M/J/10 Q10

3. Chi-squared tests

Q6.

Benford's Law states that, in many tables containing large numbers of numerical values, the probability distribution of the leading non-zero digit D is given by

$$\mathrm{P}(D = d) = \log_{10}\left(\frac{d+1}{d}\right), \quad d = 1, 2, \ldots, 9.$$

The following table shows a summary of a random sample of 100 non-zero leading digits taken from a table of cumulative probabilities for the Poisson distribution.

Leading digit	1	2	3	4	5	⩾6
Frequency	22	21	13	11	11	22

Carry out a suitable goodness of fit test at the 10% significance level. [8]

9231/23/M/J/10 Q7

Q7.

The owner of three driving schools, A, B and C, wished to assess whether there was an association between passing the driving test and the school attended. He selected a random sample of learner drivers from each of his schools and recorded the numbers of passes and failures at each school. The results that he obtained are shown in the table below.

	Driving school attended		
	A	*B*	*C*
Passes	23	15	17
Failures	27	25	43

Using a χ^2-test and a 5% level of significance, test whether there is an association between passing or failing the driving test and the driving school attended. [7]

9231/02/O/N/10 Q8

Q8.

A random sample of residents in a town took part in a survey. They were asked whether they would prefer the local council to spend money on improving the local bus service or on improving the quality of road surfaces. The responses are shown in the following table, classified according to the area of the town in which the residents live.

	Area 1	Area 2	Area 3
Local bus service	73	36	30
Road surfaces	47	44	20

Using a 5% significance level, test whether there is an association between the area lived in and preference for improving the local bus service or improving the quality of road surfaces. [7]

9231/21/M/J/11 Q6

3. Chi-squared tests

Q9.

A family was asked to record the number of letters delivered to their house on each of 200 randomly chosen weekdays. The results are summarised in the following table.

Number of letters	0	1	2	3	4	5	⩾6
Number of days	57	60	53	25	4	1	0

It is suggested that the number of letters delivered each weekday has a Poisson distribution. By finding the mean and variance for this sample, comment on the appropriateness of this suggestion. [3]

The following table includes some of the expected values, correct to 3 decimal places, using a Poisson distribution with mean equal to the sample mean for the above data.

Number of letters	0	1	2	3	4	5	⩾6
Expected number of days	53.964	70.693	p	q	6.622	1.735	0.463

(i) Show that $p = 46.304$, correct to 3 decimal places, and find q. [2]

(ii) Carry out a goodness of fit test at the 10% significance level. [9]

9231/23/M/J/11 Q10

Q10.

A sample of 216 observations of the continuous random variable X was obtained and the results are summarised in the following table.

Interval	$0 \leqslant x < 1$	$1 \leqslant x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 4$	$4 \leqslant x < 5$	$5 \leqslant x < 6$
Observed frequency	1	3	15	31	59	107

It is suggested that these results are consistent with a distribution having probability density function f given by

$$f(x) = \begin{cases} kx^2 & 0 \leqslant x < 6, \\ 0 & \text{otherwise,} \end{cases}$$

where k is a positive constant. The relevant expected frequencies are given in the following table.

Interval	$0 \leqslant x < 1$	$1 \leqslant x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 4$	$4 \leqslant x < 5$	$5 \leqslant x < 6$
Expected frequency	1	7	a	b	c	91

(i) Show that $a = 19$ and find the values of b and c. [4]

(ii) Carry out a goodness of fit test at the 10% significance level. [7]

9231/21/O/N/11 Q8

3. Chi-squared tests

Q11.

Random samples of employees are taken from two companies, A and B. Each employee is asked which of three types of coffee (Cappuccino, Latte, Ground) they prefer. The results are shown in the following table.

	Cappuccino	Latte	Ground
Company A	60	52	32
Company B	35	40	31

Test, at the 5% significance level, whether coffee preferences of employees are independent of their company. [7]

Larger random samples, consisting of N times as many employees from each company, are taken. In each company, the proportions of employees preferring the three types of coffee remain unchanged. Find the least possible value of N that would lead to the conclusion, at the 1% significance level, that coffee preferences of employees are not independent of their company. [4]

9231/21/M/J/12 Q10

Q12.

Residents of three towns A, B and C were asked to grade the reliability of their digital television signal as good, satisfactory or poor. A random sample of responses from each town is taken and the numbers in each category are given in the following table.

	Good	Satisfactory	Poor
Town A	24	34	14
Town B	58	60	26
Town C	20	34	30

Test, at the 2.5% significance level, whether grade of reliability is independent of town. [7]

Identify which town makes the greatest contribution to the test statistic and relate your answer to the context of the question. [2]

9231/23/M/J/12 Q8

3. Chi-squared tests

Q13.

A continuous random variable X is believed to have the probability density function f given by

$$f(x) = \begin{cases} \frac{3}{10}(5x - x^2 - 4) & 2 \leqslant x < 4, \\ 0 & \text{otherwise.} \end{cases}$$

A random sample of 60 observations was taken and these values are summarised in the following grouped frequency table.

Interval	$2 \leqslant x < 2.4$	$2.4 \leqslant x < 2.8$	$2.8 \leqslant x < 3.2$	$3.2 \leqslant x < 3.6$	$3.6 \leqslant x < 4$
Observed frequency	19	17	16	8	0

The estimated mean, based on the grouped data in the table above, is 2.69, correct to 2 decimal places. It is decided that a goodness of fit test will only be conducted if the mean predicted from the probability density function is within 10% of the estimated mean. Show that this condition is satisfied. [5]

The relevant expected frequencies are as follows.

Interval	$2 \leqslant x < 2.4$	$2.4 \leqslant x < 2.8$	$2.8 \leqslant x < 3.2$	$3.2 \leqslant x < 3.6$	$3.6 \leqslant x < 4$
Expected frequency	15.456	16.032	14.304	10.272	3.936

Show how the expected frequency for the interval $3.2 \leqslant x < 3.6$ is obtained. [2]

Carry out the goodness of fit test at the 10% significance level. [7]

9231/21/O/N/12 Q10

Q14.

Drinking glasses are sold in packs of 4. The manufacturer conducts a survey to assess the quality of the glasses. The results from a sample of 50 randomly chosen packs are summarised in the following table.

Number of perfect glasses	0	1	2	3	4
Number of packs	1	3	10	17	19

Fit a binomial distribution to the data and carry out a goodness of fit test at the 10% significance level. [9]

9231/23/O/N/12 Q8

3. Chi-squared tests

Q15.

A random sample of 80 observations of the continuous random variable X was taken and the values are summarised in the following table.

Interval	$2 \leqslant x < 3$	$3 \leqslant x < 4$	$4 \leqslant x < 5$	$5 \leqslant x < 6$
Observed frequency	36	29	9	6

It is required to test the goodness of fit of the distribution having probability density function f given by

$$f(x) = \begin{cases} \frac{3}{x^2} & 2 \leqslant x < 6, \\ 0 & \text{otherwise.} \end{cases}$$

Show that the expected frequency for the interval $2 \leqslant x < 3$ is 40 and calculate the remaining expected frequencies. [4]

Carry out a goodness of fit test, at the 10% significance level. [5]

9231/21/M/J/13 Q7

Q16.

A researcher is investigating the relationship between the political allegiance of university students and their childhood environment. He chooses a random sample of 100 students and finds that 60 have political allegiance to the Alliance party. He also classifies their childhood environment as rural or urban, and finds that 45 had a rural childhood. The researcher carries out a test, at the 10% significance level, on this data and finds that political allegiance is independent of childhood environment. Given that A is the number of students in the sample who both support the Alliance party and have a rural childhood, find the greatest and least possible values of A. [10]

A second random sample of size $100N$, where N is an integer, is taken from the university student population. It is found that the proportions supporting the Alliance party from urban and rural childhoods are the same as in the first sample. Given that the value of A in the first sample was 29, find the greatest possible value of N that would lead to the same conclusion (that political allegiance is independent of childhood environment) from a test, at the 10% significance level, on this second set of data. [4]

9231/23/M/J/13 Q11

3. Chi-squared tests

Q17.

Customers were asked which of three brands of coffee, A, B and C, they prefer. For a random sample of 80 male customers and 60 female customers, the numbers preferring each brand are shown in the following table.

	A	B	C
Male	32	36	12
Female	18	30	12

Test, at the 5% significance level, whether there is a difference between coffee preferences of male and female customers. [8]

A larger random sample is now taken. It consists of $80n$ male customers and $60n$ female customers, where n is a positive integer. It is found that the proportions choosing each brand are identical to those in the smaller sample. Find the least value of n that would lead to a different conclusion for the 5% significance level hypothesis test. [4]

9231/21/O/N/13 Q10

Q18.

A factory produces china mugs. Random samples of size 6 are selected at regular intervals, and the mugs are inspected for defects. During one week, 100 samples are selected and the numbers of defective mugs found are summarised in the following table.

Number of defective mugs	0	1	2	3	4	5	6
Number of samples	11	43	35	8	2	1	0

Fit a binomial distribution to the data and carry out a goodness of fit test at the 5% significance level. [10]

9231/23/O/N/13 Q8

3. Chi-squared tests

Q19.

A random sample of 200 observations of the continuous random variable X was taken and the values are summarised in the following table.

Interval	$1 \leqslant x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 4$	$4 \leqslant x < 5$	$5 \leqslant x < 6$	$6 \leqslant x < 7$	$7 \leqslant x < 8$
Observed frequency	63	45	32	25	22	7	6

It is required to test the goodness of fit of the distribution with probability density function f given by

$$f(x) = \begin{cases} \dfrac{1}{x \ln 8} & 1 \leqslant x < 8, \\ 0 & \text{otherwise.} \end{cases}$$

The relevant expected frequencies, correct to 2 decimal places, are given in the following table.

Interval	$1 \leqslant x < 2$	$2 \leqslant x < 3$	$3 \leqslant x < 4$	$4 \leqslant x < 5$	$5 \leqslant x < 6$	$6 \leqslant x < 7$	$7 \leqslant x < 8$
Expected frequency	66.67	p	27.67	q	17.54	14.83	12.84

Show that $p = 39.00$, correct to 2 decimal places, and find the value of q. [4]

Carry out a goodness of fit test at the 5% significance level. [6]

9231/21/M/J/14 Q9

Q20.

A random sample of 200 is taken from the adult population of a town and classified by age-group and preferred type of car. The results are given in the following table.

	Hatchback	Estate	Convertible
Under 25 years	32	11	17
Between 25 and 50 years	45	24	6
Over 50 years	31	16	18

Test, at the 5% significance level, whether preferred type of car is independent of age-group. [8]

9231/23/M/J/14 Q8

Q21.

The numbers of a particular type of laptop computer sold by a store on each of 100 consecutive Saturdays are summarised in the following table.

Number sold	0	1	2	3	4	5	6	7	$\geqslant 8$
Number of Saturdays	7	20	39	16	14	2	1	1	0

Fit a Poisson distribution to the data and carry out a goodness of fit test at the 2.5% significance level. [9]

9231/21/O/N/14 Q8

3. Chi-squared tests

Q22.

Each of 200 identically biased dice is thrown repeatedly until an even number is obtained. The number of throws, x, needed is recorded and the results are summarised in the following table.

x	1	2	3	4	5	6	$\geqslant 7$
Frequency	126	43	22	3	5	1	0

State a type of distribution that could be used to fit the data given in the table above. [1]

Fit a distribution of this type in which the probability of throwing an even number for each die is 0.6 and carry out a goodness of fit test at the 5% significance level. [8]

For each of these dice, it is known that the probability of obtaining a 6 when it is thrown is 0.25. Ten of these dice are each thrown 5 times. Find the probability that at least one 6 is obtained on exactly 4 of the 10 dice. [5]

9231/21/M/J/15 Q11

Q23.

The reliability of the broadband connection received from two suppliers, A and B, is classified as good, fair or poor by a random sample of householders. The information collected is summarised in the following table.

		Reliability		
		Good	Fair	Poor
Supplier	A	65	63	33
	B	51	44	44

Test, at the 5% significance level, whether reliability is independent of supplier. [8]

9231/23/M/J/15 Q6

Q24.

The number of goals scored by a certain football team was recorded for each of 100 matches, and the results are summarised in the following table.

Number of goals	0	1	2	3	4	5	6 or more
Frequency	12	16	31	25	13	3	0

Fit a Poisson distribution to the data, and test its goodness of fit at the 5% significance level. [10]

9231/21/O/N/15 Q8

3. Chi-squared tests

Q25.

Applicants for a national teacher training course are required to pass a mathematics test. Each year, the applicants are tested in groups of 6 and the number of successful applicants in each group is recorded. The overall proportion of successful applicants has remained constant over the years and is equal to 60% of the applicants. The results from 150 randomly chosen groups are shown in the following table.

Number of successful applicants	0	1	2	3	4	5	6
Number of groups	1	3	25	51	38	30	2

Test, at the 5% significance level, the goodness of fit of the distribution B(6, 0.6) for the number of successful applicants in a group. [10]

9231/21/M/J/16 Q9

Q26.

A company produces four different flavours of ice cream: mint, strawberry, chocolate and fudge. Customers were asked which of the four flavours they preferred. The responses from a random sample of 80 male customers and 70 female customers are given in the following table.

	Mint	Strawberry	Chocolate	Fudge
Male	25	12	30	13
Female	16	22	25	7

Test, at the 10% significance level, whether there is a difference between ice cream flavour preferences of male and female customers. [9]

9231/23/M/J/16 Q9

3. Chi-squared tests

Q27.

The number of visitors arriving at an art exhibition is recorded for each 10-minute period of time during the ten hours that it is open on a particular day. The results are as follows.

Number of visitors in a 10-minute period	0	1	2	3	4	5	6	7	8	⩾9
Number of 10-minute periods	2	2	12	8	11	13	4	7	1	0

(i) Calculate the mean and variance for this sample and explain whether your answers support a suggestion that a Poisson distribution might be a suitable model for the number of visitors in a 10-minute period. [3]

(ii) Use an appropriate Poisson distribution to find the two expected frequencies missing from the following table. [2]

Number of visitors in a 10-minute period	0	1	2	3	4	5	6	7	8	⩾9
Expected number of 10-minute periods	1.10		8.79		11.72	9.38	6.25	3.57	1.79	1.28

(iii) Test, at the 10% significance level, the goodness of fit of this Poisson distribution to the data. [8]

9231/21/O/N/16 Q9.

Q28.

A shop is supplied with large quantities of plant pots in packs of six. These pots can be damaged easily if they are not packed carefully. The manager of the shop is a statistician and he believes that the number of damaged pots in a pack of six has a binomial distribution. He chooses a random sample of 250 packs and records the numbers of damaged pots per pack. His results are shown in the following table.

Number of damaged pots per pack (x)	0	1	2	3	4	5	6
Frequency	48	69	78	32	22	1	0

(i) Show that the mean number of damaged pots per pack in this sample is 1.656. [1]

The following table shows some of the expected frequencies, correct to 2 decimal places, using an appropriate binomial distribution.

Number of damaged pots per pack (x)	0	1	2	3	4	5	6
Expected frequency	36.01	82.36	a	39.89	b	1.74	0.11

(ii) Find the values of a and b, correct to 2 decimal places [5]

(iii) Use a goodness-of-fit test at the 1% significance level to determine whether the manager's belief is justified. [8]

9231/21/M/J/17 Q11

3. Chi-squared tests

Q29.

Roberto owns a small hotel and offers accommodation to guests. Over a period of 100 nights, the numbers of rooms, x, that are occupied each night at Roberto's hotel and the corresponding frequencies are shown in the following table.

Number of rooms occupied (x)	0	1	2	3	4	5	6	$\geqslant 7$
Number of nights	4	9	18	26	20	16	7	0

(i) Show that the mean number of rooms that are occupied each night is 3.25. [1]

The following table shows most of the corresponding expected frequencies, correct to 2 decimal places, using a Poisson distribution with mean 3.25.

Number of rooms occupied (x)	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	4	9	18	26	20	16	7	0
Expected frequency	3.88	12.60	20.48	22.18	18.02	11.72		

(ii) Show how the expected value of 22.18, for $x = 3$, is obtained and find the expected values for $x = 6$ and for $x \geqslant 7$. [4]

(iii) Use a goodness-of-fit test at the 5% significance level to determine whether the Poisson distribution is a suitable model for the number of rooms occupied each night at Roberto's hotel. [7]

9231/23/M/J/17 Q10

Q30.

Members of a Statistics club are voting to elect a new president of the club. Members must choose to vote either by post or by text or by email. The method of voting chosen by a random sample of 60 male members and 40 female members is given in the following table.

		Method of voting		
		Post	Text	Email
Gender	Male	10	12	38
	Female	5	21	14

Test, at the 1% significance level, whether there is an association between method of voting and gender. [8]

9231/21/O/N/17 Q8

3. Chi-squared tests

Q31.

A manufacturer produces three types of car: hatchbacks, saloons and estates. Each type of car is available in one of three colours: silver, blue and red. The manufacturer wants to know whether the popularity of the colour of the car is related to the type of car. A random sample of 300 cars chosen by customers gives the information summarised in the following table.

		Colour of car		
		Silver	Blue	Red
Type of car	Hatchback	53	36	41
	Saloon	29	40	31
	Estate	28	24	18

Test at the 10% significance level whether the colour of car chosen by customers is independent of the type of car. [8]

9231/21/M/J/18 Q8

Q32.

A scientist carries out an experiment to investigate the quantity X, which takes the values 0, 1, 2, 3, 4, 5 or 6. He believes that the values taken by X follow a binomial distribution. He conducts 250 trials. His results are summarised in the following table.

x	0	1	2	3	4	5	6
Observed frequency	22	83	72	53	17	3	0

(i) Show that unbiased estimates of the mean and variance for these results are 1.876 and 1.266 respectively, correct to 3 decimal places. By evaluating the mean and variance of the distribution B(6, 0.313), explain why X could have this distribution. [4]

The expected frequencies corresponding to the distribution B(6, 0.313) are shown in the following table.

x	0	1	2	3	4	5	6
Observed frequency	22	83	72	53	17	3	0
Expected frequency	26.3	71.9	81.8	49.7	17.0	3.1	0.2

(ii) Show how the expected frequency for $x = 4$ is calculated. [2]

(iii) Test at the 5% significance level whether the scientist's belief is correct. [8]

9231/23/M/J/18 Q11

3. Chi-squared tests

Q33.

A machine is used to produce metal rods. When the machine is working efficiently, the lengths, x cm, of the rods have a normal distribution with mean 150 cm and standard deviation 1.2 cm. The machine is checked regularly by taking random samples of 200 rods. The latest results are shown in the following table.

Interval	$146 \leqslant x < 147$	$147 \leqslant x < 148$	$148 \leqslant x < 149$	$149 \leqslant x < 150$
Observed frequency	1	2	23	52

$150 \leqslant x < 151$	$151 \leqslant x < 152$	$152 \leqslant x < 153$	$153 \leqslant x < 154$
69	36	15	2

As a first check, the sample is used to calculate an estimate for the mean.

(i) Show that an estimate for the mean from this sample is close to 150 cm. [2]

As a second check, the results are tested for goodness of fit of the normal distribution with mean 150 cm and standard deviation 1.2 cm. The relevant expected frequencies, found using the normal distribution function given in the List of Formulae (MF10), are shown in the following table.

Interval	$x < 147$	$147 \leqslant x < 148$	$148 \leqslant x < 149$	$149 \leqslant x < 150$
Observed frequency	1	2	23	52
Expected frequency	1.24	8.32	30.94	59.50

$150 \leqslant x < 151$	$151 \leqslant x < 152$	$152 \leqslant x < 153$	$153 \leqslant x$
69	36	15	2
59.50	30.94	8.32	1.24

(ii) Show how the expected frequency for $151 \leqslant x < 152$ is obtained. [3]

(iii) Test, at the 5% significance level, the goodness of fit of the normal distribution to the results. [7]

9231/21/O/N/18 Q11b

3. Chi-squared tests

Q34.

The number of accidents, x, that occur each day on a motorway are recorded over a period of 40 days. The results are shown in the following table.

Number of accidents	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	3	5	8	10	5	7	2	0

(i) Show that the mean number of accidents each day is 2.95 and calculate the variance for this sample. Explain why these values suggest that a Poisson distribution might fit the data. [3]

A Poisson distribution with mean 2.95, as found from the data, is used to calculate the expected frequencies, correct to 2 decimal places. The results are shown in the following table.

Number of accidents	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	3	5	8	10	5	7	2	0
Expected frequency	2.09	6.18	9.11	8.96	6.61	3.90	1.92	1.23

(ii) Show how the expected frequency of 6.61 for $x = 4$ is obtained. [2]

(iii) Test at the 5% significance level the goodness of fit of this Poisson distribution to the data. [7]

9231/22/O/N/18 Q10

Q35.

A random sample of 50 observations of the continuous random variable X was taken and the values are summarised in the following table.

Interval	$0 \leqslant x < 0.8$	$0.8 \leqslant x < 1.6$	$1.6 \leqslant x < 2.4$	$2.4 \leqslant x < 3.2$	$3.2 \leqslant x < 4$
Observed frequency	18	16	8	6	2

It is required to test the goodness of fit of the distribution with probability density function f given by

$$f(x) = \begin{cases} \frac{3}{16}(4-x)^{\frac{1}{2}} & 0 \leqslant x < 4, \\ 0 & \text{otherwise.} \end{cases}$$

The relevant expected frequencies, correct to 2 decimal places, are given in the following table.

Interval	$0 \leqslant x < 0.8$	$0.8 \leqslant x < 1.6$	$1.6 \leqslant x < 2.4$	$2.4 \leqslant x < 3.2$	$3.2 \leqslant x < 4$
Expected frequency	14.22	12.54	10.59	8.18	4.47

(i) Show how the expected frequency for $1.6 \leqslant x < 2.4$ is obtained. [3]

(ii) Carry out a goodness of fit test at the 5% significance level. [7]

9231/21/M/J/19 Q9

3. Chi-squared tests

Q36.

Two salesmen, A and B, work at a company that arranges different types of holidays: self-catering, hotel and cruise. The table shows, for a random sample of 150 holidays, the number of each type arranged by each salesman.

		Type of holiday		
		Self-catering	Hotel	Cruise
Salesman	A	25	38	21
	B	28	21	17

Test at the 10% significance level whether the type of holiday arranged is independent of the salesman. [8]

9231/23/M/J/19 Q8

Q37.

The number of puncture repairs carried out each week by a small repair shop is recorded over a period of 40 weeks. The results are shown in the following table.

Number of repairs in a week	0	1	2	3	4	5	$\geqslant 6$
Number of weeks	6	15	9	6	3	1	0

(i) Calculate the mean and variance for the number of repairs in a week and comment on the possible suitability of a Poisson distribution to model the data. [3]

Records over a longer period of time indicate that the mean number of repairs in a week is 1.6. The following table shows some of the expected frequencies, correct to 3 decimal places, for a period of 40 weeks using a Poisson distribution with mean 1.6.

Number of repairs in a week	0	1	2	3	4	5	$\geqslant 6$
Expected frequency	8.076	12.921	10.337	5.513	2.205	a	b

(ii) Show that $a = 0.706$ and find the value of the constant b. [3]

(iii) Carry out a goodness of fit test of a Poisson distribution with mean 1.6, using a 10% significance level. [8]

9231/21/O/N/19 Q11

3. Chi-squared tests

Q38.

Two randomly selected groups of students, with similar ranges of abilities, take the same examination in different rooms. One group of 140 students takes the examination with background music playing. The other group of 210 students takes the examination in silence. Each student is awarded a grade for their performance in the examination and the numbers from each group gaining each grade are shown in the following table.

	Grade awarded		
	A	B	C
Background music	49	51	40
Silence	93	68	49

Test at the 10% significance level whether grades awarded are independent of whether background music is playing during the examination. [6]

9231/41/M/J/20 Q1

Q39.

Young children are learning to read using two different reading schemes, A and B. The standards achieved are measured against the national average standard achieved and classified as above average, average or below average. For two randomly chosen groups of young children, the numbers in each category are shown in the table.

	Standard achieved		
	Above average	Average	Below average
Scheme A	31	35	22
Scheme B	19	50	43

Test at the 5% significance level whether standard achieved is independent of the reading scheme used. [6]

9231/43/M/J/20 Q1

Q40.

Apples are sold in bags of 5. Based on her previous experience, Freya claims that the probability of any apple weighing more than 100 grams is 0.35, independently of other apples in the bag.

The apples in a random sample of 150 bags are checked and the number, x, in each bag weighing more than 100 grams is recorded. The results are shown in the following table.

x	0	1	2	3	4	5
Frequency	12	39	46	37	12	4

Carry out a goodness of fit test at the 5% significance level and hence comment on Freya's claim. [7]

9231/41/O/N/20 Q3

3. Chi-squared tests

Q41.

A random sample of 200 observations of the continuous random variable X was taken and the values are summarised in the following table.

Interval	$0 \leqslant x < 0.5$	$0.5 \leqslant x < 1$	$1 \leqslant x < 1.5$	$1.5 \leqslant x < 2$	$2 \leqslant x < 2.5$	$2.5 \leqslant x < 3$
Observed frequency	5	23	40	41	46	45

It is required to test the goodness of fit of the distribution with probability density function f given by

$$f(x) = \begin{cases} \frac{1}{9}x(4-x) & 0 \leqslant x \leqslant 3, \\ 0 & \text{otherwise.} \end{cases}$$

Most of the relevant expected frequencies, correct to 2 decimal places, are given in the following table.

Interval	$0 \leqslant x < 0.5$	$0.5 \leqslant x < 1$	$1 \leqslant x < 1.5$	$1.5 \leqslant x < 2$	$2 \leqslant x < 2.5$	$2.5 \leqslant x < 3$
Expected frequency	p	q	37.96	43.52	43.52	37.96

(a) Show that $p = 10.19$ and find the value of q. [3]

(b) Carry out a goodness of fit test, at the 5% significance level, to test whether f is a satisfactory model for the data. [4]

9231/42/O/N/20 Q3

Q42.

A driving school employs four instructors to prepare people for their driving test. The allocation of people to instructors is random. For each of the instructors, the following table gives the number of people who passed and the number who failed their driving test last year.

	Instructor A	Instructor B	Instructor C	Instructor D	Total
Pass	72	42	52	68	234
Fail	33	34	41	58	166
Total	105	76	93	126	400

Test at the 10% significance level whether success in the driving test is independent of the instructor. [7]

9231/41/M/J/21 Q2

3. Chi-squared tests

Q43.

Chai packs china mugs into cardboard boxes. Chai's manager suspects that breakages occur at random times and that the number of breakages may follow a Poisson distribution. He takes a small sample of observations and finds that the number of breakages in a one-hour period has a mean of 2.4 and a standard deviation of 1.5.

(a) Explain how this information tends to support the manager's suspicion. [2]

The manager now takes a larger sample and claims that the numbers of breakages in a one-hour period follow a Poisson distribution. The numbers of breakages in a random sample of 180 one-hour periods are summarised in the following table.

Number of breakages	0	1	2	3	4	5	6	7 or more
Frequency	21	33	46	31	23	16	10	0

The mean number of breakages calculated from this sample is 2.5.

(b) Use the data from this larger sample to carry out a goodness of fit test, at the 10% significance level, to test the claim. [8]

9231/43/M/J/21 Q5

Q44.

A supermarket sells pears in packs of 8. Some of the pears in a pack may not be ripe, and the supermarket manager claims that the number of unripe pears in a pack can be modelled by the distribution B(8, 0.15).

A random sample of 150 packs was selected and the number of unripe pears in each pack was recorded. The following table shows the observed frequencies together with some of the expected frequencies using the manager's binomial distribution.

Number of unripe pears per pack	0	1	2	3	4	5	$\geqslant 6$
Observed frequency	35	48	43	15	6	3	0
Expected frequency	40.874	p	35.641	12.579	2.775	0.392	q

(a) Find the values of p and q. [2]

(b) Carry out a goodness of fit test, at the 5% significance level, to test whether the manager's claim is justified. [6]

9231/41/O/N/21 Q3

3. Chi-squared tests

Q45.

It is claimed that the heights of a particular age group of boys follow a normal distribution with mean 125 cm and standard deviation 12 cm. Observations for a randomly chosen group of 60 boys in this age group are summarised in the following table. The table also gives the expected frequencies, correct to 2 decimal places, based on the normal distribution with mean 125 cm and standard deviation 12 cm.

Height, x cm	$x < 100$	$100 \leqslant x < 110$	$110 \leqslant x < 120$	$120 \leqslant x < 130$	$130 \leqslant x < 140$	$x \geqslant 140$
Observed frequency	0	3	15	23	11	8
Expected frequency	1.12	5.22	13.97	19.38	13.97	6.34

(a) Show how the expected frequency for $130 \leqslant x < 140$ is obtained. [2]

(b) Carry out a goodness of fit test, at the 5% significance level, to determine whether the claim is supported by the data. [6]

9231/42/O/N/21 Q2

Q46.

A scientist is investigating the numbers of a particular type of butterfly in a certain region. He claims that the numbers of these butterflies found per square metre can be modelled by a Poisson distribution with mean 2.5. He takes a random sample of 120 areas, each of one square metre, and counts the number of these butterflies in each of these areas. The following table shows the observed frequencies together with some of the expected frequencies using the scientist's Poisson distribution.

Number per square metre	0	1	2	3	4	5	6	$\geqslant 7$
Observed frequency	12	20	36	32	13	6	1	0
Expected frequency	9.85	24.63	30.78	25.65	p	8.02	3.34	q

(a) Find the values of p and q, correct to 2 decimal places. [2]

(b) Carry out a goodness of fit test, at the 10% significance level, to test the scientist's claim. [6]

9231/41/M/J/22 Q4

3. Chi-squared tests

Q47.

A scientist is investigating the size of shells at various beach locations. She selects four beach locations and takes a random sample of shells from each of these beaches. She classifies each shell as large or small. Her results are summarised in the following table.

		Beach location				
		A	*B*	*C*	*D*	Total
Size of shell	Large	68	69	96	81	314
	Small	28	55	64	39	186
	Total	96	124	160	120	500

Test, at the 10% significance level, whether the size of shell is independent of the beach location. [7]

9231/43/M/J/22 Q2

Q48.

An organisation runs courses to train students to become engineers. These students are taught in groups of 8. The director of the organisation claims that on average 60% of the students in a group achieve a pass. A random sample of 150 groups of 8 students is chosen. The following table shows the observed frequencies together with some of the expected frequencies using the appropriate binomial distribution.

Number of passes per group	0	1	2	3	4	5	6	7	8
Observed frequency	0	0	8	24	45	36	26	10	1
Expected frequency	p	1.180	6.193	18.579	34.836	q	r	13.437	2.519

(a) Find the values of p, q and r giving your answers correct to 3 decimal places. [2]

(b) Carry out a goodness of fit test, at the 10% significance level, to test whether there is evidence to reject the director's claim. [6]

9231/41/O/N/22 Q2

3. Chi-squared tests

Q49.

In the colleges in three regions of a particular country, students are given individual targets to achieve. Their performance is measured against their individual target and graded as 'above target', 'on target' or 'below target'. For a random sample of students from each of the three regions, the observed frequencies are summarised in the following table.

		Region			
		A	*B*	*C*	Total
Performance	Above target	62	41	44	147
	On target	102	94	95	291
	Below target	56	45	61	162
	Total	220	180	200	600

Test, at the 10% significance level, whether performance is independent of region. [7]

9231/42/O/N/22 Q2

Solutions

Q1.

OR	State distribution:	Geometric	B1		
	Find sample mean:	$\bar{x} = 1.84$	B1		
	Estimate p from sample mean:	$1 / \bar{x} = 0.5435$ **A.G.**	M1 A1	4	
	Calculate expected values (correct to ± 0.01):	54.35 24.81 11.33 5.17 2.36 1.98	*M1 A1		
	Combine values as appropriate:	Last 3 to give 9.51 (dep *M1)	M1√		
	Calculate value of χ^2 :	$\chi^2 \approx 2.27$ [± 0.01]	M1 *A1		
	Compare with consistent tab. value (to 2 dp):	$\chi_{2,\,0.975}^{\,2} = 7.378$ [$\chi_4^2 = 11.14$]	*B1√		
	Consistent conclusion (A1 dep *A1, *B1):	Distribution fits data (A.E.F.)	M1√ A1	8	
	State 2 changes (A.E.F.):	Change estimated p to 0.5	B1		
		Increase degrees of freedom by 1	B1	2	**[14]**

Q2.

9	Integrate f(t) to give F(t):	$F(t) = -9/8t^2$	M1		
	Apply limits:	$F(2{\cdot}5) - F(2) = -(9/8)(2{\cdot}5^{-2} - 2^{-2})$	A1		
	Evaluate and multiply by 100:	10·125 **A.G.**	A1	3	
	State hypotheses (A.E.F.):	H_0: f(t) fits data, H_1: doesn't fit	B1		
	Find χ^2 (A1 if at least 3 terms correct):	$\chi^2 = 1{\cdot}5^2/62{\cdot}5 + 4{\cdot}875^2/21{\cdot}875$			
		$+ 5{\cdot}875^2/10{\cdot}125 + 2{\cdot}5^2/5{\cdot}5$	M1 A1		
	Evaluate χ^2:	$= 0{\cdot}036 + 1{\cdot}086 + 3{\cdot}409 + 1{\cdot}136$			
		$= 5{\cdot}67$ [± 0·01]	*A1		
	Compare with consistent tab. value (to 2 dp):	$\chi_{3,\,0.9}^{\,2} = 6{\cdot}251$	*B1√		
	($\chi_{2,\,0.9}^{\,2} = 4{\cdot}605$, $\chi_{1,\,0.9}^{\,2} = 2{\cdot}706$)				
	Consistent conclusion (A1 dep *A1, *B1):	Distribution fits data (A.E.F.)	M1√ A1	7	**[10]**

Q3.

9	State (at least) null hypothesis (A.E.F.):	H_0: Sydney popn. has same freq.			
		[0·38, 0·10, 0·03, 0·49]	B1		
	Calculate expected values:	76 20 6 98	M1 A1		
	Calculate value of χ^2 (M1 even if cells combined):	$\chi^2 \approx 8{\cdot}5$[2]	M1 *A1		
	Compare with consistent tab. value (to 2 dp):	$\chi_{3,\,0.95}^{\,2} = 7{\cdot}815$ [$\chi_2^2 = 5{\cdot}991$]	*B1√		
	Correct conclusion (dep *A1, *B1):	Sydney does **not** conform (AEF)	B1	7	
	Find smallest size n with expected values ≥ 5 [or >]:	$n \geq 5/0{\cdot}03 = 166{\cdot}7$ so n_{min} is 167	M1 A1	2	**[9]**

Q4.

9	Show how value is found (allow M1 if 100 omitted):	$100\ {}^4C_2\ 0{\cdot}4^2 \times 0{\cdot}6^2 = 34{\cdot}56$ **A.G.**	M1 A1	(2)	
	State (at least) null hypothesis:	H_0: B(4, 0·6) fits data (A.E.F.)	B1		
	Combine adjacent cells since exp. value < 5:	*O*: 14 27 49 10			
		E: 17·92 34·56 34·56 12·96	*M1		
	Calculate value of χ^2 (to 2 dp ; A1 dep *M1):	$\chi^2 = 9{\cdot}22$	M1 A1		
	Compare with consistent tabular value (to 2 dp):	$\chi^2_{3,\,0.95} = 7{\cdot}815$ (cells combined)			
		$\chi^2_{4,\,0.95} = 9{\cdot}488$ (not combined)	B1		
	Valid method for reaching conclusion:	Reject H_0 if χ^2 > tabular value	M1		
	Correct conclusion (A.E.F., requires correct values):	9·22 > 7·81[5] so distn. does not fit	*A1		
	State valid deduction (dep *A1; allow $p \neq 0{\cdot}6$):	Prob. of faulty chips ≠ 0·6 (A.E.F.)	B1	(8)	**[10]**

Q5.

10	Tabulate observed data with totals:	*A* *B* *C* *Flu* 30 29 16 75 *No flu* 148 120 157 425 178 149 173	M1		
	Find expected values: (lose A1 if 1 or 2 errors; lose A1 if rounded to integers)	*A* *B* *C* *Flu* 26·7 22·35 25·95 *No flu* 151·3 126·65 147·05	M1 A2		
	State (at least) null hypothesis (A.E.F.):	H_0: Catching flu indep of vaccine	B1		
	Calculate value of χ^2 (to 2 dp):	$\chi^2 = 7{\cdot}30$	M1 A1		
	S.R. If rounded to integers above allow (to 2 dp):	$\chi^2 = 7{\cdot}53$ (earns max 8/10)	(B1)		
	Compare with consistent tabular value (to 2 dp):	$\chi^2_{2,\,0.95} = 5{\cdot}991$	B1		
	Valid method for reaching conclusion:	Reject H_0 if χ^2 > tabular value	M1		
	Correct conclusion (A.E.F., requires correct values):	Catching flu depends on vaccine	A1	10	
	Find proportions (or complements) for *A*,*B*,*C*:	0·169, 0·195, 0·092 (to 2 dp)	M1 A1		
	Correct conclusion (A.E.F., requires correct values):	*C* appears most effective	A1	3	**[13]**

Q6.

7	State (at least) null hypothesis (A.E.F.):	H_0: Data conforms to Law	B1		
	Calculate expected values (to 1 dp):	30·10 17·61 12·49 9·69 7·92 22·18	M1 A1		
	Calculate value of χ^2:	$\chi^2 = 4{\cdot}23 \pm 0{\cdot}02$	M1 A1		
	Compare with consistent tabular value (to 2 dp):	$\chi^2_{5,\,0.9} = 9{\cdot}236$	B1		
	Valid method for reaching conclusion:	Reject H_0 if χ^2 > tabular value	M1		
	Correct conclusion (A.E.F., requires correct values):	Data conforms to Benford's Law	A1	8	**[8]**

3. Chi-squared tests

Q7.

8	Find expected values to (at least) 1 dp: (lose A1 if one or more errors or if rounded to integers) State (at least) null hypothesis (A.E.F.): Calculate value of χ^2 : **S.R.** If rounded to integers above allow: Compare with tabular value (to 2 dp): Valid method for reaching conclusion: Correct conclusion (A.E.F., requires correct values):	*A* *B* *C* *Passes* 18·33 14·67 22·00 *Failures* 31·67 25·33 38·00 M1 A1 H_0: Test result indep of school B1 $\chi^2 = 3{\cdot}7 \pm 0{\cdot}02$ B1 $\chi^2 = 3{\cdot}96$ *or* 4·0 (earns max 6/7) (B1) $\chi^2_{2,\,0.95} = 5{\cdot}99$ B1 Reject H_0 if $\chi^2 >$ tabular value M1 No association A1	7	**[7]**

Q8.

6	Find expected values (to 1 dp): (lose A1 if rounded to integers) State (at least) null hypothesis (A.E.F.): Calculate value of χ^2: Compare with consistent tabular value (to 2 dp): Conclusion consistent with values (A.E.F):	66·72 44·48 27·80 53·28 35·52 22·20 M1 A1 H_0: No association between them B1 $\chi^2 = 5{\cdot}36 \pm 0{\cdot}03$ M1 A1 $\chi^2_{2,\,0.95} = 5{\cdot}991$ B1 No association A1 √	7	**[7]**

Q9.

(b)	Find mean and variance of sample:	262/200 = 1·31 *and* $(586 - 262^2/200) / 200 = 1{\cdot}21[39]$ M1 A1			
	Valid comment (AEF, needs values approx correct):				
	Values close, so distn. appropriate		B1	3	
(i)	State and evaluate expression for p	**A.G.**: $p = 200\,(1{\cdot}31^2/2)e^{-1.31} = 46{\cdot}304$	B1		
	Find q (can use $\Sigma E_i = 200$):	$q = 200\,(1{\cdot}31^3/6)e^{-1.31} = 20{\cdot}2[19]$	B1	2	
(ii)	State (at least) null hypothesis:	H_0: Poisson fits data (A.E.F.)	B1		
	Combine last 3 cells since exp. value < 5:	*O*: . . . 5			
		E: . . . 8·82	*M1 A1		
	Calculate χ^2 (to 2 dp ; A1 dep *M1):	$\chi^2 = 5{\cdot}54$	M1 A1		
	Compare consistent tabular value (to 2 dp):	$\chi^2_{3,\,0.9} = 6{\cdot}251$	M1 A1		
	(A1 dep *M1)	$[\chi^2_{4,\,0.9} = 7{\cdot}779,\ \chi^2_{5,\,0.9} = 9{\cdot}236]$			
	Valid method for reaching conclusion:	Accept H_0 if $\chi^2 <$ tabular value	M1		
	Conclusion (A.E.F., needs correct values):	5·54 < 6·25 so Poisson does fit	A1	9	**14**

Q10.

8	(i)	Find value of k by integrating f(x):	$[\frac{1}{3}kx^3]_0^6 = 1$, $k = 3/6^3 = 1/72$	B1		
		State and evaluate expression for a: **A.G.**	$a = 216\ [\frac{1}{3}kx^3]_2^3 = 3^3 - 2^3 = 19$	B1		
		Find b and c:	$b = 216\ [\frac{1}{3}kx^3]_3^4 = 37$,			
		(**MR**: f(x) as distn. of table: max 3/4)	$c = 216\ [\frac{1}{3}kx^3]_4^5$ *or* 216–155 = 61	M1 A1	4	
	(ii)	State (at least) null hypothesis:	H_0: f(x) fits data (A.E.F.)	B1		
		Combine first 2 cells since exp. value < 5:	O: 4 ...			
			E: 8 ...	B1		
		Calculate χ^2 (to 2 dp):	$\chi^2 = 6{\cdot}69$[4]	M1 *A1		
		Compare consistent tabular value (to 2 dp):	$\chi^2_{4,\,0.9} = 7{\cdot}779$			
		[or if 3 or 0 cells combined:	$\chi^2_{3,\,0.9} = 6{\cdot}251$, $\chi^2_{5,\,0.9} = 9{\cdot}236$]	*B1		
		Valid method for reaching conclusion:	Accept H_0 if χ^2 < tabular value	M1		
		Conclusion (A.E.F., dep *A1, *B1):	6·69 < 7·78 so f(x) does fit	A1	7	**[11]**

Q11.

10	Find expected values (to 1 d.p.):	54·72 52·992 36·288			
	(lose A1 if rounded to integers)	40·28 39·008 26·712	M1 A1		
	State (at least) null hypothesis (A.E.F.):	H_0: Preferences are independent	B1		
	Calculate value of χ^2:	$\chi^2 = 0{\cdot}5095 + 0{\cdot}0186 + 0{\cdot}5067$			
		$+ 0{\cdot}6921 + 0{\cdot}0252 + 0{\cdot}6883$			
		$= 2{\cdot}44$ or $2{\cdot}45$	M1 A1		
	State or use correct tabular χ^2 value (to 3 sf):	$\chi^2_{2,\,0.95} = 5{\cdot}99$[1]	B1		
	Conclusion consistent with values (A.E.F):	Preferences are independent	A1 √	7	
	Calculate new value χ^2_{new} of χ^2:	$\chi^2_{new} = N \times \chi^2$	M1		
	State or use correct tabular χ^2 value:	$\chi^2_{2,\,0.99} = 9{\cdot}21$	B1		
	Find N_{min}:	$N > 9{\cdot}21/2{\cdot}45$, $N_{min} = 4$	M1 A1	4	**[11]**

Q12.

8	Find expected values (to 1 d.p.):	24·48 30·72 16·8			
	(lose A1 if rounded to integers)	48·96 61·44 33·6			
		28·56 35·84 19·6	M1 A1		
	State (at least) null hypothesis (A.E.F.):	H_0: Grade independent of town	B1		
	Calculate value of χ^2 to 1 d.p.:	$\chi^2 = 0{\cdot}009 + 0{\cdot}350 + 0{\cdot}467$			
		$+ 1{\cdot}669 + 0{\cdot}034 + 1{\cdot}719$			
		$+ 2{\cdot}567 + 0{\cdot}094 + 5{\cdot}518$			
		$= 12{\cdot}4$[3]	M1 A1		
	Compare with consistent tabular value (to 1 d.p.):	$\chi^2_{4,\,0.975} = 11{\cdot}1$[4]	B1		
	Conclusion consistent with values (A.E.F):	Grade not independent of town	A1 √	7	
	State town with max. contribn. (√ on exp. values):	Town C	B1 √		
	State valid comment (√ on exp. values), e.g.:	Signal poorer than expected in C	B1 √	2	**[9]**

3. Chi-squared tests

Q13.

(b)	Find E(X) using $\int xf(x)dx$:	E(X) = $\int_{2}^{4}$ $(5x^2 - x^3 - 4x)/10$ dx = ½(4³–2³) – 3(4⁴–2⁴)/40 – 3(4²–2²)/5 = 28 – 18 – 7·2 = 2·8	M1 A1 *A1	5	
	Verify E(X) within 10% of 2·69 (A1 dep *A1):	(E(X) – 2·69)/2·69 = 0·041 < 0·1 *or* 1·1 × 2·69 = 2·96 > E(X)	M1 A1		
	Show derivation of tabular entry:	60 $\int_{3.2}^{3.6}$ $(5x - x^2 - 4)/10$ dx = 60[3($5x^2/2 - x^3/3 - 4x$)/10]$_{3.2}^{3.6}$ *or* [$45x^2 - 6x^3 - 72x$]$_{3.2}^{3.6}$ = 122·4 – 83·328 – 28·8 *or* 60 × 0·1712 = 10·272	M1 **A.G** A1	2	
	State (at least) null hypothesis: Combine last 2 cells since exp. value < 5:	H_0: f(x) fits data (A.E.F.) *O*: . . . 8 *E*: . . . 14·208	B1 B1		
	Calculate χ^2 (to 2 d.p.):	χ^2 = 0·8126 + 0·0584 + 0·2011 + 2·7135 = 3·78[47]	M1 *A1		
	State or use consistent tabular value (to 2 d.p.): [or if no cells combined: Valid method for reaching conclusion: Conclusion (A.E.F., dep *A1, *B1):	$\chi^2_{3\,0.9}$ = 6·25[1] $\chi^2_{4,\,0.9}$ = 7·78] Accept H_0 if χ^2 < tabular value 3·78 < 6·25 so f(x) does fit	*B1 M1 A1	7	**[14]**

Q14.

8	Find value of p for binomial dist.:	mean = 150/50 = 3, p = ¾	M1 A1		
	Find expected binomial values (to 2 d.p.):	0·20 2·34 10·55 21·09 15·82	M1 A1		
	Combine adjacent cells since exp. value < 5:	*O*: 14 17 19 *E*: 13·09 21·09 15·82	 *M1		
	Calculate value of χ^2 (to 2 d.p. ; A1 dep *M1):	χ^2 = 1·50	M1 *A1		
	State or use consistent tabular value (to 2 d.p.):	$\chi^2_{1,\,0.9}$ = 2·706 (cells combined) [$\chi^2_{2,\,0.9}$ = 4·605, $\chi^2_{3,\,0.9}$ = 6·251]	*B1		
	Correct conclusion (A.E.F., dep *A1, *B1):	1·50 < 2·71 so distn. does fit	A1	9	**[9]**

3. Chi-squared tests

Q15.

7	Find 1st expected frequency E_2:	$E_2 = 80\int_2^3 3x^{-2}\,dx$			
	Find other expected frequencies:	$= 80\left[-3x^{-1}\right]_2^3$		4	
		$= 40$ **A.G.**	M1 A1		
		$E_3 = 20,\ E_4 = 12,\ E_5 = 8$	M1 A1		
	State (at least) null hypothesis:	H_0: f(x) fits data (A.E.F.)	B1		
	Calculate χ^2 (to 2 d.p.):	$\chi^2 = 0{\cdot}4 + 4{\cdot}05 + 0{\cdot}75 + 0{\cdot}5 = 5{\cdot}7$	M1 A1		
	State or use consistent tabular value:	$\chi^2_{3,\,0.9} = 6{\cdot}25$[1]	*B1		
	Conclusion (A.E.F., √ on χ^2, dep *B1):	$\chi^2 < 6{\cdot}25$ so f(x) does fit	B1√	5	**9**

Q16.

	Find positive value of μ when $\theta = 45°$:	$2\mu^2 + 4\mu - 1 = 0$			
		$\mu = \sqrt{(3/2)} - 1$ *or* 0·225	A.E.F M1 A1		
	Find positive value of μ when $\theta = 0°$:	$\mu = 1/\sqrt{2}$ *or* 0·707	A.E.F B1		
	State set of possible values of μ (A.E.F.):	$[\sqrt{(3/2)}-1, 1/\sqrt{2}]$ *or* [0·225, 0·707]	A1	4	**14**
(b)	Tabulate observed values:	*Rural* *Urban* *Support* A $60-A$ *Not support* $45-A$ $A-5$	M1 A1		
	Find corresponding expected values:	27 33 18 22	M1 A1		
	Calculate value of χ^2:]	$\chi^2 = (27-A)^2/27 + (A-27)^2/18$			
		$+ (A-27)^2/33 + (27-A)^2/22$	M1		
		$= (50/297)(A-27)^2$			
		or $0{\cdot}168$[35] $(A-27)^2$	A1		
	State or use correct tabular χ^2 value:	$\chi^2_{tab} = \chi^2_{1,\,0.9} = 2{\cdot}706$ (to 3 s.f.)	B1		
	Use conclusion of independence to find eqn. for A:	$(50/297)(A-27)^2 < \chi^2_{tab}$	M1		
	Find bounds for A (to 3 s.f.):	$(A-27)^2 < 16{\cdot}07$	A1		
	(integer valued reqd for this A1)	$A_{min} = 23$ *and* $A_{max} = 31$	A1	10	
	Relate new value χ^2_{new} to original χ^2:	$\chi^2_{new} = N \times \chi^2$	M1		
	Use conclusion of independence to find eqn. for A:	$0{\cdot}168$[35] $N(A-27)^2 < \chi^2_{tab}$	M1		
	Find N_{max} with $A = 29$:	$N < 2{\cdot}706/(4\times 0{\cdot}16835) = 4{\cdot}02$	A1		
	(integer value reqd for this A1)	$N_{max} = 4$	A1	4	**14**

3. Chi-squared tests

Q17.

10	State (at least) null hypothesis (A.E.F.):	H_0: No difference in preferences [or independent]	B1		
	Find expected values (to 1 d.p.): (lose A1 if rounded to integers)	28·57 37·71 13·71 21·43 28·29 10·29	M1 A1		
	Calculate value of χ^2:	$\chi^2 = 0{\cdot}411 + 0{\cdot}078 + 0{\cdot}214 + 0{\cdot}549 + 0{\cdot}104 + 0{\cdot}286$ $= 1{\cdot}64$ (*or* 1·61 if 1 d.p. used)	M1 A1		
	State or use correct tabular χ^2 value (to 3 s.f.):	$\chi^2_{2,\,0.95} = 5{\cdot}99$[1]	B1		
	Valid method for reaching conclusion:	Accept H_0 if $\chi^2 \leq$ tabular value	M1		
	Conclusion consistent with correct values (A.E.F):	No difference in preferences	A1	8	
	Calculate new value χ^2_{new} of χ^2:	$\chi^2_{new} = n \times \chi^2$	M1		
	State or use correct tabular χ^2 value:	$\chi^2_{2,\,0.95} = 5{\cdot}99$[1]	B1		
	Find n_{min}:	$n > 5{\cdot}99/1{\cdot}64$, $n_{min} = 4$	M1 A1	4	**[12]**

Q18.

8	Find value of p for binomial dist.:	mean = 150/100, $p = 1{\cdot}5/6 = ¼$	M1 A1		
	State (at least) null hypothesis:	H_0: B(6, p) fits data (A.E.F.)	B1		
	Find expected binomial values (to 1 d.p.):	17·80 35·60 29·66 13·18 3·30 0.44 0.02	M1 A1		
	Combine last four cells since exp. value < 5:	*O*: 11 43 35 11			
		E: 17·80 35·60 29·66 16·94	*M1		
	Calculate value of χ^2 (to 2 d.p.; A1 dep *M1):	$\chi^2 = 2{\cdot}60 + 1{\cdot}54 + 0{\cdot}96 + 2{\cdot}08$ $= 7{\cdot}18$ (*or* 7·14 if 1 d.p.)	M1 *A1		
	State or use consistent tabular value (to 2 d.p.):	$\chi^2_{2,\,0.95} = 5{\cdot}991$ (cells combined) [$\chi^2_{3,\,0.95} = 7{\cdot}815$, $\chi^2_{4,\,0.95} = 9{\cdot}488$]	*B1		
	Correct conclusion (A.E.F., dep *A1, *B1):	$\chi^2 > 5{\cdot}99$ so distn. does not fit	B1	10	**10**

Q19.

9	Find expected frequency p:	$p = 200 \int_2^3 (1 / x \ln 8)\,dx$ $= (200 / \ln 8)\,[\ln x]_2^3$ $= 200 \times 0{\cdot}1950 = 39{\cdot}00$ **A.G.**	M1A1		
	Find q by similar method *or* by using total of 200:	$q = 21{\cdot}46$ *or* 21·45	M1A1	4	
	State (at least) null hypothesis:	H_0: f(*x*) fits data (A.E.F.)	B1		
	Calculate χ^2 (to 3 s.f.):	$\chi^2 = 0{\cdot}202 + 0{\cdot}923 + 0{\cdot}678 + 0{\cdot}584 + 1{\cdot}134 + 4{\cdot}134 + 3{\cdot}644 = 11{\cdot}3$	M1A1		
	State or use correct tabular χ^2 value (to 3 s.f.)::	$\chi^2_{6,\,0.95} = 12{\cdot}59$	B1		
	Valid method for reaching conclusion:	Accept H_0 if $\chi^2 \leq$ tabular value	M1		
	Conclusion consistent with correct values (A.E.F):	Distribution fits observations	A1	6	**10**

3. Chi-squared tests

Q20.

8	Find expected values (to 1 d.p.):	32·4 15·3 12·3 40·5 19·125 15·375 35·1 16·575 13·325 M1 A1		
	State (at least) null hypothesis (A.E.F.):	H_0: Car type is independent B1		
	Calculate χ^2 (to 1 d.p.):	χ^2 = 0·005 + 1·208 + 1·796 + 0·500 + 1·243 + 5·716 + 0·479 + 0·020 + 1·640 = 12·6 (allow 12·7) M1 A1		
	State or use correct tabular χ^2 value (to 3 s.f.):	$\chi^2_{4,\,0.95}$ = 9·488 *OR* 9·49 B1		
	Valid method for reaching conclusion:	Reject H_0 if χ^2 > tabular value M1		
	Conclusion consistent with correct values (A.E.F):	Car type is dependent on age A1	8	**8**

Q21.

8	Find mean of sample data for use in Poisson distn.:	$\lambda = \frac{225}{100}$ = 2·25 B1		
	State (at least) null hypothesis (A.E.F.):	H_0: Poisson distn. fits data B1		
	Find expected values $\frac{100\lambda^r e^{-\lambda}}{r!}$ (to 1 d.p.): (ignore incorrect final value here for M1)	10·540 23·715 26·679 20·009 11·255 5·065 1·899 0·6105 0·2275 M1 A1		
	Combine last four cells so that exp. value ⩾ 5:	O_i: . . . 16 14 4 E_i: . . . 20·009 11·255 7·802 *M1		
	Calculate value of χ^2 (to 1 d.p.; A1 dep *M1):	χ^2 = 1·189 + 0·582 + 5·690 + 0·803 + 0·6695 + 1·853 = 10·8 (allow 10·7) M1 A1		
	State or use consistent tabular value (to 1 d.p.):	$\chi^2_{4,\,0.975}$ = 11·14 (if cells combined) B1 [$\chi^2_{7,\,0.975}$ = 16·01, $\chi^2_{5,\,0.975}$ = 12·83]		
	Consistent conclusion (A.E.F., √ on two χ^2 values):	χ^2 < 11·1 so Poisson distn. fits B1√	9	**[9]**

3. Chi-squared tests

Q22.

11B	State suitable distribution:	Geometric	B1	1	
	State (at least) null hypothesis: (AEF)	H_0: Distn. fits data *or* $p = 0.6$	B1		
	(B0 for "It is a good fit")				
	Find exp. values using $200pq^{x-1}$ with $p = 0.6$, $q = 0.4$:	120 48 19.2 7.68			
	(ignore incorrect final value here,	3.072 1.2288 0.8192	M1		
	e.g. 0.4915 which can earn max 5/8				
	Combine last 3 cells since exp. value < 5:	*O*: . . . 6			
		E: . . . 5.12	B1		
	Calculate χ^2 (result correct to 3 s.f.):	$\chi^2 = 0.3 + 0.5208 + 0.4083$			
		$+ 2.8519 + 0.15125 = 4.23$[2]	M1 A1		
	State or use consistent tabular value (to 3 s.f.):	5 cells: $\chi^2_{4,0.95} = 9.488$			
	[or if 2 or no cells combined:	6 cells: $\chi^2_{5,0.95} = 11.07$			
		7 cells: $\chi^2_{6,0.95} = 12.59$			
	or if 4 cells combined, as with 0·4915:	4 cells: $\chi^2_{3,0.95} = 7.815$]	B1√		
	Valid method for reaching conclusion:	Accept H_0 if χ^2 < tabular value	M1		
	Conclusion (AEF, requires both values correct):	4.23 < 9.49 so distn fits *or* $p = 0.6$	A1		
	(Allow A1 for "It is a good fit")				
	Find prob. p of at least one 6 on 5 throws of one die:	$p = 1 - 0.75^5 = 0.7627$ (4 s.f.)	M1 A1	8	
	Find prob. of at least one 6 on exactly 4 of 10 dice:	${}^{10}C_4 \times$; $p^4(1-p)^6$	M1; M1		
		$= 210 \times 6.043 \times 10^{-5}$			
		$= 0.0126$ *or* 0.0127	A1	5	**14**

Q23.

6	State (at least) null hypothesis (AEF):	H_0: Reliability is independent	B1		
	Find expected values (to 1 d.p.):	62.253 57.423 41.323			
	(lose A1 if rounded to integers)	53.747 49.577 35.677	M1 A1		
	Calculate value of χ^2:	$\chi^2 = 0.1212 + 0.5416 + 1.6764$			
		$+ 0.1404 + 0.6274 + 1.9417$			
	(allow 5.03 if 1 d.p. exp. values used)	$= 5{\cdot}05$	M1 A1		
	State or use correct tabular χ^2 value (to 3 s.f.):	$\chi^2_{2,0.95} = 5.99$[1]	B1		
	Valid method for reaching conclusion:	Accept H_0 if χ^2 < tabular value	M1		
	Correct conclusion, from correct values: (AEF)	Reliability is indep. of supplier	A1	8	**8**

3. Chi-squared tests

Q24.

8	Find mean of sample data [for use in Poisson distn.]:	$\lambda = 220/100 = 2{\cdot}2$	**B1**		
	State (at least) null hypothesis (AEF):	H_0: Poisson distn. fits data *or* $\lambda = 2{\cdot}2$	**B1**		
	Find expected values $100\lambda^r e^{-\lambda}/r!$ (to 1 d.p.): (ignore incorrect final value here for M1)	11·080 24·377 26·814 19·664 10·8151 4·759 2·491	**M1 A1**		
	Combine last two cells so that exp. value ⩾ 5:	O_i: 3 E_i: 7·25	**M1***		
	Calculate value of χ^2 (to 2 d.p.; A1 dep M1*): (allow 7·95 if 1 d.p. exp.values used)	$\chi^2 = 0{\cdot}076 + 2{\cdot}879 + 0{\cdot}653 + 1{\cdot}448 + 0{\cdot}441 + 2{\cdot}491$ $= 7{\cdot}99$	**M1 A1**		
	State or use consistent tabular value (to 3 s.f.):	5 cells: $\chi^2_{3,0.95} = 7{\cdot}815$ 6 cells: $\chi^2_{4,0.95} = 9{\cdot}488$ (correct) 7 cells: $\chi^2_{5,0.95} = 11{\cdot}07$	**B1**		
	State or imply valid method for conclusion e.g.:	Accept H_0 if χ^2 < tabular value	**M1**		
	Conclusion (AEF, requires both values correct): Not combining cells [so $\chi^2 = 8{\cdot}64$] can earn B1 B1 M1 A1 M0 M1 B1 M1 (max 7)	Distn fits *or* $\lambda = 2{\cdot}2$	**A1**	10	**10**

Q25.

9	State (at least) null hypothesis in full: (AEF)	H_0: Given distribution fits data	**B1**		
	Find exp. values using 150 ${}^6C_i\, q^{6-i}\, p^i$ with $p = 0{\cdot}6$, $q = 0{\cdot}4$ (to 3 s.f.): (allow A1 if only one error or if all values correct to 2 s.f.)	0·6144 5·5296 20·736 41·472 46·656 27·9936 6·9984	**M1 A2**		
	Combine 0·6144 & 5·5296 since 1st exp. value < 5:	O_i: 4 . . . E_i: 6·14[4] . . .			
	Calculate χ^2 (result correct to 3 s.f.):	$\chi^2 = 0{\cdot}748 + 0{\cdot}877 + 2{\cdot}189 + 1{\cdot}606 + 0{\cdot}144 + 3{\cdot}570 = 9{\cdot}13$	**B1**		
	State or use consistent tabular value (to 3 s.f.): [or if no or more cells combined:	6 cells: $\chi^2_{5,0.95} = 11{\cdot}07$ 7 cells: $\chi^2_{6,0.95} = 12{\cdot}59$ 5 cells: $\chi^2_{4,0.95} = 9{\cdot}488$ 4 cells: $\chi^2_{3,0.95} = 7{\cdot}815$ 3 cells: $\chi^2_{2,0.95} = 5{\cdot}991$]	**M1 A1** **B1√**		
	State or imply valid method for conclusion e.g.:	Accept H_0 if χ^2 < tabular value	**M1**		
	Conclusion (AEF, requires both values correct): [data] (Allow A1 here for e.g. "It is a good fit") **S.C.** 150 ${}^6C_i\, q^i\, p^{6-i}$ can earn B1 M1 B1 M1 B1 M1 (max 6/10)	9·13 [± 0·01] < 11·1 so distn. fits	**A1**		[10]

3. Chi-squared tests

Q26.

9	State (at least) null hypothesis (AEF): Find expected values (to 1 d.p.): (lose A1 if one error or rounded to integers) Calculate value of χ^2 (to 3 s.f.): State or use correct tabular χ^2 value (to 3 s.f.): State or imply valid method for conclusion e.g.: Correct conclusion, from correct values: (AEF)	H_0: No difference in preferences 21·86718·13329·33310·667 19·13315·86725·6679·333 $\chi^2 = 0·449 + 2·075 + 0·015 + 0·510$ $+ 0·513 + 2·371 + 0·017 + 0·583$ $= 6·53$ (allow 6·52) $\chi^2_{3,\,0.9} = 6·25$[1] Reject H_0 if $\chi^2 \geqslant$ tabular value Difference in preferences	**B1** **M1 A2** **M1 A1** **B1** **M1** **A1**	[9]

Q27.

9 (i)	Find mean and variance of sample data.: State valid reason why Poisson distn. suitable (AEF): (allow unsuitable since $4 \neq 3·57$)	$\bar{x} = 240/60 = 4$ $\sigma^2 = 1174/60 - 4^2 = 3·57$ $4 \approx 3·57$ (no √ on $\bar{x}$, σ^2)	**B1** **B1** **B1**	[3]
(ii)	Find expected values $60\lambda^r e^{-\lambda} / r!$ with $\lambda = 4$:	4·40; $11·7_{[21]}$ (to 1 d.p.)	**B1; B1**	[2]
(iii)	State (at least) null hypothesis (AEF): Combine cells so that all exp. value $\geqslant 5$: Calculate value of χ^2 (to 2 d.p.; A1 dep *M1): State or use consistent tabular value (to 2 d.p.): [or fewer or no cells combined: State or imply valid method for conclusion e.g.: Conclusion (AEF, requires both values correct):	H_0: [Poisson] distribution fits data O_i : 4 12 … 6 8 E_i : 5·50 8·79 … 6·25 6·64 $\chi^2 = 0·409 + 1·172 + 1·181$ $+ 0·044 + 1·397 + 0·81 + 0·279$ $= 5·29$ 7 cells: $\chi^2_{5,\,0.9} = 9·236$ 8 cells: $\chi^2_{6,\,0.9} = 10·64$ 9 cells: $\chi^2_{7,\,0.9} = 12·02$ 10 cells: $\chi^2_{8,\,0.9} = 13·36$] Accept H_0 if $\chi^2 <$ tabular value 5·29 < 9·24 so distn. fits	**B1** ***M1 A1** **M1 DA1** **B1√** **M1** **A1**	[8]

Q28.

11(b)(i)	$x = (1/250)\,\Sigma\, x\, f(x) = 414/250 = 1·656$ **AG**	**B1**	Verify given mean
	Total:	**1**	
11(b)(ii)	$p = x/6 = 0·276$, $q = 0·724$	**M1 A1**	Use 250 ${}^6C_i\ q^{6-i}\ p^i$ and find p and q
	$a = 250\ {}^6C_2\ q^4\ p^2 = 78·49 \pm 0·01$ (to 2 d.p.)	**A2**	Find either exp. value
	$b = 250\ {}^6C_4\ q^2\ p^4 = 11·41 \pm 0·01$ (to 2 d.p.)	**A1**	Find other exp. value (deduct single **A1** if either value given to only 1 d.p.)
	Total:	**5**	
11(b)(iii)	H_0: Distribution fits data *or* distribution is binomial (AEF)	**B1**	State (at least) null hypothesis in full Combine values consistent with all exp. values $\geqslant 5$
	O_i: 48 69 78 32 23 E_i: 36·01 82·36 78·49 39·89 13·26 (± 0·01)	**M1FT A1**	(FT for **M1** but not **A1** on values of a, b)
	$\chi^2 = 3·992 + 2·167 + 0·003 + 1·561 + 7·154$	**M1**	Find χ^2
	$= 14·9$	**A1**	
	No. n of cells: 7 6 5 4 3 $\chi^2_{n-2,\,0.99}$: 15·09 13·28 11·34 9·210 6·635	**B1FT**	State or use consistent tabular value $\chi^2_{n-2,\,0.99}$ (to 3 s.f.) [FT on number, n, of cells used to find χ^2]
	Accept H_1 if $\chi^2 >$ tabular value (AEF) 14·9 [± 0·1] > 11·34 so distn. doesn't fit [data]	**M1**	State or imply valid method for conclusion Conclusion (requires both values correct)
	or manager's belief not justified (AEF)	**A1**	
	Total:	**8**	

Q29.

10(i)	$\bar{x} = (1/100)\,\Sigma\, x\, f(x) = 325/100 = 3{\cdot}25$ AG	**B1**	Verify given mean (**B0** for $\bar{x} = 325/100 = 3{\cdot}25$)
	Total:	**1**	
10(ii)	$E_3 = 100\ \lambda^3\ e^{-\lambda}/3!$ with $\lambda = 3{\cdot}25$	**M1 A1**	State expression for reqd. expected value E_3 (**M1** for E_3 *or* E_6)
	$E_6 = 100\ \lambda^6\ e^{-\lambda}/6! = 6{\cdot}35$	**A1**	Find exp. value E_6
	$E_{\geqslant 7} = 100 - \Sigma_0^{\,6}\, E_i = 4{\cdot}77$	**A1**	Find exp. value $E_{\geqslant 7}$
	Total:	**4**	
10(iii)	H_0: Distribution fits data (AEF)	**B1**	State (at least) null hypothesis in full
	O_i: 13 18 26 20 16 7 E_i: 16·48 20·48 22·18 18·02 11·72 11·12	**M1FT**	Combine values consistent with all exp. values $\geqslant 5$ (FT on E_6 and $E_{\geqslant 7}$)
	$\chi^2 = 0{\cdot}735 + 0{\cdot}300 + 0{\cdot}658 + 0{\cdot}218 + 1{\cdot}563 + 1{\cdot}527$	**M1**	Find χ^2
	$= 5{\cdot}00$	**A1**	
	No. n of cells: 8 7 6 5 4 $\chi^2_{n-2,\,0.95}$: 12·59 11·07 9·488 7·815 5·991	**B1FT**	State or use consistent tabular value $\chi^2_{n-2,\,0.95}$ (to 3 s.f.) [FT on number, n, of cells used to find χ^2]
	Accept H_0 if χ^2 < tabular value (AEF) 5·00 [± 0·1] < 9·49 so distn. fits [data]	**M1**	State or imply valid method for conclusion Conclusion (requires both values correct)
	or distn. is a suitable model (AEF)	**A1**	
	Total:	**7**	

Q30.

8	H_0: No association *or* method is independent of gender (AEF)	**B1**	State (at least) null hypothesis
	E_i: 9·0 19·8 31·2 6·0 13·2 20·8	**M1 A1**	Find expected values E_i (**A0** if rounded to integers)
	$\chi^2 = 0{\cdot}111 + 3{\cdot}073 + 1{\cdot}482$ $+ 0{\cdot}167 + 4{\cdot}610 + 2{\cdot}223$ (to 2 d.p.) $= 11{\cdot}7$ (*or* 12·3)	**M1 A1**	Find value of χ^2 from $\Sigma\,(E_i - O_i)^2 / E_i$ [or $\Sigma\, O_i^2/E_i - n$] (allow 12·3 if integer values of E_i used)
	$\chi^2_{2,\,0.99} = 9{\cdot}21$	**B1**	State or use correct tabular χ^2 value
	Reject H_0 if χ^2 > tabular value (AEF)	**M1**	Valid method for reaching conclusion
	11·7 [± 0·1] > 9·21 so there is an association	**A1**	Correct conclusion, from correct values
		8	

3. Chi-squared tests

Q31.

8	H_0: Colour is independent of type *or* no association between colour and type (AEF)	**B1**	State (at least) null hypothesis in full
	E_i: 47·67 43·33 39 36·67 33·33 30 25·67 23·33 21 (to 3 s.f.)	**M1 A1**	Find expected values E_i (A0 if rounded to integers)
	$X^2 = 0·597 + 1·241 + 0·103$ $+ 1·603 + 1·333 + 0·033$ $+ 0·212 + 0·019 + 0·429 = 5·57$ (to 3 s.f.)	**M1 A1**	Find value of χ^2 from $\Sigma (E_i - O_i)^2 / E_i$ [or $\Sigma O_i^2 / E_i - n$] (allow 5·64 for this A1 if integer values of E_i used)
	$\chi_{4, 0.9}^2 = 7·779$ *or* 7·78	**B1**	State or use correct tabular χ^2 value
	Accept H_0 if $X^2 <$ tabular value (AEF)	**M1**	Valid method for reaching conclusion
	5·57 [± 0·01] < 7·78 so independent *or* no association (AEF)	**A1**	Correct (abbreviated) conclusion, from approx. correct values
		8	

Q32.

11B(i)	$\bar{x} = 469/250 = 1·876$ *and* **AG** $\sigma^2 = (1195 - 469^2/250) / 249 = 1·266$ **AG**	**B1**	Verify mean and unbiased variance of data, showing method B0 if biased variance, but allow other marks (max 3/4)
	$6 \times 0·313 = 1·878$ *and* $6 \times 0·313 \times 0·687 = 1·29$[0]	**M1 A1**	Find mean *np and* variance *npq* of binomial distribution
	Means and variances are similar for X and B(6, 0·313)	**A1**	Valid explanation (AEF; needs 1·878, 1·290 correct to 0·002)
		4	
11B(ii)	$250 \times {}^6C_4 \times 0·313^4 \times 0·687^2$ [$= 250 \times 0·06795 = 16·99$] Allow $49·7 \times ({}^6C_4 \times 0·313) / ({}^6C_3 \times 0·687)$	**M1 A1**	Show how expected frequency for $x = 4$ is found (AEF for ${}^6C_4 = 15$)
		2	
11B(iii)	H_0: [Binomial] distribution fits data (AEF)	**B1**	State (at least) null hypothesis
	O_i: 22 83 72 53 20 E_i: 26·3 71·9 81·8 49·7 20·3	***M1 A1**	Combine last 3 cells so that all exp. value ≥ 5
	$X^2 = 0·703 + 1·714 + 1·174 + 0·219 + 0·004$ $= 3·81$ (to 3 s.f.)	**M1 A1**	Find value of χ^2 from $\Sigma (E_i - O_i)^2 / E_i$ [or $\Sigma O_i^2/E_i - n$] (A1 dep *M1)
	No. n of cells: 7 6 5 $\chi_{n-2, 0.95}^2$: 11·07 9·488 7·815	**B1√**	State or use consistent tabular value $\chi_{n-2, 0.95}^2$ (to 3 s.f.) [FT on number, n, of cells used to find X^2]
	Accept H_0 if $X^2 <$ tabular value	**M1**	State or imply valid method for conclusion
	3·81 [± 0·01] < 7·81[5] so distribution fits *or* scientist's belief is correct	**A1**	Conclusion (AEF; requires both values approx. correct)
		8	

3. Chi-squared tests

Q33.

11B(i)	$\bar{x} = (1/200)\,\Sigma\, x\, f(x) = 30\,064 / 200 = 150.3$[2]	**M1A1**	Find sample mean to 4 s.f. using mid-interval values (B1 for 29 964 / 200 = 149.8[2] or 30 164 / 200 = 150.8[2])
		2	
11B(ii)	$P(151 \leqslant X < 152)$ $= P((151 - 150)/1.2 \leqslant Z < (152 - 150)/1.2)$ $= P(0.833 \leqslant Z < 1.667)$ $= 0.9522 - 0.7975$ [= 0.1547] (to 4 s.f.)	**M1A1**	Find $P(151 \leqslant X < 152)$
	$E_{151} = 190.44 - 159.5$ *or* $200 \times 0.1547 = 30.94$ AG	**A1**	and hence expected frequency for $151 \leqslant x < 152$
		3	
11B(iii)	H_0: [Normal] distribution fits data (AEF)	**B1**	State (at least) null hypothesis in full
	O_i: 3 23 52 69 36 17 E_i: 9.56 30.94 59.50 59.50 30.94 9.56	**M1**	Combine values consistent with all exp. values ⩾ 5
	$X^2 = 4.501 + 2.038 + 0.945 + 1.517 + 0.828 + 5.790$	**M1**	Find value of X^2 from $\Sigma\,(E_i - O_i)^2 / E_i$ [or $\Sigma\, O_i^2/E_i - n$]
	$= 15.6$ (to 3 s.f.)	**A1**	
	No. n of cells: 8 7 6 5 $\chi^2_{n-1,\,0.95}$: 14.07 12.59 11.07 9.488	**B1**	State or use consistent tabular value $\chi^2_{n-1,\,0.95}$ (to 3 s.f.) [**FT** on number, n, of cells used to find X^2]
	Accept H_1 if X^2 > tabular value (AEF)	**M1**	State or imply valid method for conclusion
	15.6 [± 0.1] > 11.1 so distn. doesn't fit [data] (AEF)	**A1**	Conclusion (requires both values correct)
		7	

Q34.

10(i)	$\bar{x} = \Sigma\, x\, f(x) = 118/40 = 2.95$ AG	**B1**	Verify mean of sample data (B0 for $\bar{x} = 118/40 = 2.95$)
	$\sigma_n^2 = (454 - 118^2/40) / 40 = 2.65$ *or* $\sigma_{n-1}^2 = (454 - 118^2/40) / 39 = 2.72$	**B1**	Find variance of sample data (accept either σ_n^2 or σ_{n-1}^2)
	$2.95 \approx 2.65$ *or* 2.72	**B1**	Valid comment on correct values
		3	
10(ii)	$E_4 = 40\,\lambda^4\, e^{-\lambda}/4!$ with $\lambda = 2.95$ [= 40 × 0.1652] (allow $(\lambda/4) \times E_3 = 0.7375 \times 8.96$)	**M1A1**	State expression for reqd. expected value E_4
		2	
10(iii)	H_0: [Poisson] distribution fits data (AEF)	**B1**	State (at least) null hypothesis in full
	O_i: 8 8 10 5 9 E_i: 8.27 9.11 8.96 6.61 7.05	**M1**	Combine values consistent with all exp. values ⩾ 5
	$X^2 = 0.0088 + 0.1352 + 0.1207 + 0.3921 + 0.5394$	**M1**	Find value of X^2 from $\Sigma\,(E_i - O_i)^2 / E_i$ [or $\Sigma\, O_i^2/E_i - n$]
	$= 1.20$ (to 3 s.f.)	**A1**	
	No. n of cells: 8 7 6 5 $\chi^2_{n-2,\,0.95}$: 12.59 11.07 9.488 7.815	**B1**	State or use consistent tabular value $\chi^2_{n-2,\,0.95}$ (to 3 s.f.) [**FT** on number, n, of cells used to find X^2]
	Accept H_0 if X^2 < tabular value (AEF)	**M1**	State or imply valid method for conclusion
	1.20 [± 0.1] < 7.81[5] so [Poisson] distn. fits [data] *or* distn. is a suitable model (AEF)	**A1**	Conclusion (requires both values correct)
		7	

Q35.

9 (i)	$E_3 = (3/16) \int_{1.6}^{2.4} (4-x)^{1/2}\, dx$ $= (3/16) [-(2/3)(4-x)^{3/2}]_{1.6}^{2.4}$	**M1**	State or imply expression for required expected value E_3 of X
	$= (2 \cdot 4^{3/2} - 1 \cdot 6^{3/2}) / 8 = 1 \cdot 694/8$ *or* $0 \cdot 2118$	**A1**	Find expected value E_3 (may be implied in finding 50 E_3) (M1 A1 requires adequate explicit working)
	$50\ E_3 = 10 \cdot 59$ AG	**A1**	Hence verify corresponding expected frequency
		3	
9(ii)	H_0: Distribution fits data (AEF)	**B1**	State (at least) null hypothesis in full
	O_i 18 16 8 8 E_i: 14·22 12·54 10·59 12·65 $X^2 = 1 \cdot 005 + 0 \cdot 955 + 0 \cdot 633 + 1 \cdot 709 = 4 \cdot 30$	**M1** **M1 A1**	Combine values consistent with all exp. values ≥ 5 Find value of X^2 from $\Sigma (E_i - O_i)^2 / E_i$ [or $\Sigma O_i^2/E_i - n$]
	No. n of cells: 5 4 3 $\chi^2_{n-1,\,0.95}$: 9·488 7·815 5·991 (to 3 s.f.)	**B1**	**FT** on number, n, of cells used to find X^2 State or use consistent tabular value $\chi^2_{n-1,\,0.95}$
	Accept H_0 if X^2 < tabular value (AEF)	**M1**	State or imply valid method for conclusion
	$4 \cdot 30$ [± 0·01] < 7·81[5] so distribution fits [data] *or* distribution is a suitable model (AEF)	**A1**	Conclusion (requires both values approx. correct)
		7	

Q36.

8	H_0: Holidays are independent of salesman *or* no association between holidays and salesman (AEF)	**B1**	State (at least) null hypothesis Find expected values (lose A1 if rounded to integers) Find value of X^2 from $\Sigma (E_i - O_i)^2 / E_i$ [or $\Sigma O_i^2/E_i - n$] State or use correct tabular χ^2 value
	E_i: 29·68 33·04 21·28 23·32 25·96 16·72 (to 1 d.p.)	**M1 A1**	
	X^2 = 0·7380 + 0·7446 + 0·0037 + 0·9392 + 0·9477 + 0·0047	**M1**	
	= 3·38 (to 3 s.f.)	**A1**	
	$\chi^2_{2,\,0.9}$ = 4·605 (to 3 s.f.)	**B1**	
	Accept H_0 if X^2 < tabular value	**M1**	Compare their calculated value with their X^2 value and appropriate conclusion
	Type of holidays is independent of salesman (AEF)	**A1**	Correct conclusion, from correct values $(3.37 - 3.39)$
		8	

3. Chi-squared tests

Q37.

11B(i)	$\bar{x} = (1/40)\,\Sigma\, x\ \mathrm{f}(x) = 68/40 = 1{\cdot}7$	**B1**	Find mean of sample
	$(1/40)\,\Sigma\, x^2\ \mathrm{f}(x) = 178/40 = 4{\cdot}45$, Var $= 4{\cdot}45 - 1{\cdot}7^2 = 1{\cdot}56$	**B1**	Find variance of sample
	Mean and variance are similar so Poisson may be suitable	**B1**	State valid comment
		3	
11B(ii)	$a = 40 \times 1{\cdot}6^5\, e^{-1.6}/5! = 40 \times 0{\cdot}01764$	**M1**	AG
	$a = 0{\cdot}706$	**A1**	Verify a from Poisson term
	$b = 40 - 39{\cdot}758 = 0{\cdot}242$	**B1***	Find b
		3	
11B(iii)	H_0: Distribution fits/models data	**B1**	State (at least) null hypothesis in full
	O_i: 6 15 9 10 E_i: 8·076 12·921 10·337 8·666	**DM1** **A1**	Combine values consistent with all exp. values $\geqslant 5$ (FT on b, dep B1*)
	$X^2 =$ 0·5337 + 0·3345 + 0·1729 + 0·2053 = 1·25	**M1**	Find value of X^2 from $\Sigma\,(E_i - O_i)^2 / E_i$ [or $\Sigma\, O_i^2/E_i - n$]
	$X^2 = 1{\cdot}25$	**A1**	
	No. n of cells: 7 6 5 4 3 $\chi^2_{n-1,\,0.9}$: 10·64 9·236 7·779 6·251 4·605	**DB1**	State or use consistent tabular value $\chi^2_{n-1,\,0.9}$ (to 3 s.f.) [FT on number, n, of cells used to find X^2]
	Accept H_0 if X^2 < tabular value (using their values)	**M1**	AEF
	1·25 [± 0·1] < 6·25 so distn. fits [data] or distn. is a suitable model	**A1**	Conclusion (requires both values approx. correct). AEF
		8	

Q38.

1	49 E = 56.8 | 51 E = 47.6 | 40 E = 35.6 93 E = 85.2 | 68 E = 71.4 | 49 E = 53.4	**M1A1**
	$\frac{(49-56.8)^2}{56.8} + \frac{(51-47.6)^2}{47.6} + \frac{(40-35.6)^2}{35.6} + \frac{(93-85.2)^2}{85.2} + \frac{(68-71.4)^2}{71.4} + \frac{(49-53.4)^2}{53.4}$	**M1**
	= 3.096 (3.10)	**A1**
	Use appropriate tabular value = 4.605	**M1**
	3.096 < 4.605 so the grades awarded are independent of the background	**A1**
		6

3. Chi-squared tests

Q39.

1	<table><tr><td>31 E = 22</td><td>35 E = 37.4</td><td>22 E = 28.6</td></tr><tr><td>19 E = 28</td><td>50 E = 47.6</td><td>43 E = 36.4</td></tr></table>	**M1A1**
	$\frac{(31-22)^2}{22}+\frac{(35-37.4)^2}{37.4}+\frac{(22-28.6)^2}{28.6}+\frac{(19-28)^2}{28}+\frac{(50-47.6)^2}{47.6}+\frac{(43-36.4)^2}{36.4}$	**M1**
	9.569 (9.57)	**A1**
	Tabular value = 5.991 and comparing 9.569 > 5.991	**M1**
	So standard achieved is dependent on reading scheme used	**A1**
		6

Q40.

3	<table><tr><td>x</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>Observed freq</td><td>12</td><td>39</td><td>46</td><td>37</td><td>12</td><td>4</td></tr><tr><td>Bin prob</td><td>0.11603</td><td>0.31236</td><td>0.33642</td><td>0.18115</td><td>0.04877</td><td>0.00525</td></tr><tr><td>Expected freq</td><td>17.404</td><td>46.858</td><td>50.462</td><td>27.172</td><td>7.316</td><td>0.7878</td></tr></table>	**M1 A1**	Attempt at E values (at least 4 correct) All correct, to 2 dp or better
	Add last two columns: 16, 8.104	**M1**	
	$\frac{(17.404-12)^2}{17.404}+\frac{(46.858-39)^2}{46.858}+\frac{(50.462-46)^2}{50.462}+$ $\frac{(27.172-37)^2}{27.172}+\frac{(8.104-16)^2}{8.104}$ = 14.64 or 14.65	**M1 A1**	Accept 14.6 – 14.7
	'14.64' > 9.49	**M1**	Compare *their* value with 9.49
	Freya's claim is not supported or Data does not fit the distribution	**A1 FT**	Correct conclusion in context, FT only *their* 14.64 SC for Poisson M0M1M1M0 max 2
		7	

3. Chi-squared tests

Q41.

<table>
<tr><td>3(a)</td><td>$\int_0^{0.5}\left(\frac{4}{9}x-\frac{1}{9}x^2\right)dx \quad \left[=\frac{1}{9}\left[2\times0.5^2-\frac{1}{3}0.5^3\right]\right]$

$(=0.0509(26)$ or $\frac{11}{216})$</td><td>M1</td><td>Statement of integration with correct limits</td></tr>
<tr><td></td><td>Freq = p = 0.0509(26) × 200 = 10.19 AG</td><td>A1</td><td>Requires sight of 0.0509 or $\frac{11}{216}$</td></tr>
<tr><td></td><td>By addition to 200, q = 26.85 $\left(\frac{725}{27}\right)$</td><td>B1</td><td>Or by integration</td></tr>
<tr><td></td><td></td><td>3</td><td></td></tr>
<tr><td>3(b)</td><td><table><tr><td>O</td><td>5</td><td>23</td><td>40</td><td>41</td><td>46</td><td>45</td></tr><tr><td>E</td><td>10.19</td><td>26.85</td><td>37.96</td><td>43.52</td><td>43.52</td><td>37.96</td></tr></table>Test statistic = 2.6433 + 0.5520 + 0.1096 + 0.1459 + 0.1413 + 1.3056</td><td>M1</td><td>At least 4 correct</td></tr>
<tr><td></td><td>4.89 or 4.90</td><td>A1</td><td></td></tr>
<tr><td></td><td>‘4.90’ < 11.07:</td><td>M1</td><td>Compare their value with 11.07</td></tr>
<tr><td></td><td>PDF is a satisfactory model for the data.</td><td>A1 FT</td><td>Correct conclusion in context, ft only their 4.90</td></tr>
<tr><td></td><td></td><td>4</td><td></td></tr>
</table>

Q42.

<table>
<tr><td>2</td><td><table><tr><td>72 (61.425)</td><td>42 (44.46)</td><td>52 (54.405)</td><td>68 (73.71)</td></tr><tr><td>33 (43.575)</td><td>34 (31.54)</td><td>41 (38.595)</td><td>58 (52.29)</td></tr></table></td><td>M1</td><td>At least 3 correct to 4sf.</td></tr>
<tr><td></td><td>Calculates E values (in brackets above)</td><td>A1</td><td>All correct to 4sf (if 6.037 not found).</td></tr>
<tr><td></td><td>Test value: $\Sigma\frac{(O-E)^2}{E}$ = 1.8206 + 0.1361 + 0.1063 +0.4423 + 2.5664 + 0.1919 + 0.1499 + 0.6235</td><td>M1</td><td>Correct formula used.</td></tr>
<tr><td></td><td>= 6.037</td><td>A1</td><td>Correct to 3 sf.</td></tr>
<tr><td></td><td>H_0: (driving) test success is independent of instructor</td><td>B1</td><td></td></tr>
<tr><td></td><td>Tabular value: 3 df 10% = 6.251
6.04 < 6.251
Accept H_0</td><td>M1</td><td>Compare with correct tabular value and conclusion.</td></tr>
<tr><td></td><td>Sufficient evidence to suggest that (driving) test success is independent of instructor</td><td>A1</td><td>Correct conclusion, in context, following correct work. Level of uncertainty in language is used.
Allow ±1 in the third significant figure of test value.</td></tr>
<tr><td></td><td></td><td>7</td><td></td></tr>
</table>

3. Chi-squared tests

Q43.

5(a)	(Mean of distribution = 2.4) Variance = 2.25	**M1**	Allow 1.5^2.
	Mean approx. equal to variance (so Poisson might be suitable)	**A1**	Must have 'approximately' oe, not equality. Must have 2.25.
		2	
5(b)	Po(2.5) leads to frequencies:	**M1**	At least 3 correct to 3sf.
	14.77[5], 36.938, 46.173, 38.477, 24.048, 12.024, 5.010, 2.554	**A1**	All correct to 4sf.
	Combine last two values: 7.564	**M1**	FT their table values if final figure is less than 5.
	$\Sigma\frac{(O-E)^2}{E} = 2.6227 + 0.4198 + 0.00065 + 1.4529 + 0.04570 + 1.3148 + 0.7845$	**M1**	Apply correct formula to their frequencies.
	6.64	**A1**	Correct to 3sf.
5(b)	H_0: distribution fits the data	**B1**	Must mention data and distribution .
	Tabular value, 5 degrees of freedom, is 9.236 '6.64' < 9.236, so accept H_0	**M1**	Compare with correct tabular value and conclusion.
	Insufficient evidence to show that data does not follow a Poisson distribution	**A1**	Correct conclusion, in context, following correct work. Level of uncertainty in language is used. Allow ±1 difference in third significant figure.
		8	

Q44.

3(a)	<table><tr><td>x</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6 or more</td></tr><tr><td>O</td><td>35</td><td>48</td><td>43</td><td>15</td><td>6</td><td>3</td><td>0</td></tr><tr><td>E</td><td>40.874</td><td>57.704</td><td>35.641</td><td>12.579</td><td>2.775</td><td>0.392</td><td>0.0352</td></tr></table> $p = 57.704$	**B1**	3 dp or better.
	$q = 0.035$ or 0.036	**B1**	Accept numbers which round to 0.035 or 0.036.
		2	

3(b)	Combine frequencies less than 5: last 4 columns give 24 / 15.7812	**M1**	Allow last 2 or 3 columns combined for this M1.
	$\Sigma\frac{(O-E)^2}{E} = 0.8444 + 1.6319 + 1.5199 + 4.2803$	**M1**	
	8.28	**A1**	Accept 8.27 – 8.28. **SC:** 25.60, if values not combined, scores M1A0. **SC:** 14.96 if last 3 combined, scores M1A0. **SC:** if last 2 combined, scores M1A0.
	H_0: B(8, 0.15) fits the data	**B1**	Must mention distribution and data.
	3 degrees of freedom, tabular value = 7.815. 8.28 > 7.815 Reject H_0.	**M1**	Compare their value with 7.815 and correct FT conclusion 3 columns combined compared with 9.488 and correct conclusion
	Insufficient evidence to support manager's claim.	**A1**	Correct conclusion, in context, following correct work. Level of uncertainty in language is used.
		6	

3. Chi-squared tests

Q45.

<table>
<tr><td>2(a)</td><td>$P(130 \leqslant x < 140) = P\left(\frac{130-125}{12} \leqslant Z < \frac{140-125}{12}\right)$
$P(0.4167 < Z < 1.25)$</td><td>M1</td><td>Must be seen.
Must see 0.4167 and 1.25, may be implied by 0.8944 and 0.6616
Accept 0.417.</td></tr>
<tr><td></td><td>$0.8944 - 0.6616 = 0.2328$
Multiply by 60 leads to 13.97</td><td>A1</td><td>AG
Accept 0.6615. Need to see 0.2328 or 0.2329 or 13.968 or $(0.8944 - 0.6616) \times 60$ ·</td></tr>
<tr><td></td><td></td><td>2</td><td></td></tr>
<tr><td>2(b)</td><td><table><tr><td>0</td><td>3</td><td>15</td><td>23</td><td>11</td><td>8</td></tr><tr><td>1.12</td><td>5.22</td><td>13.97</td><td>19.38</td><td>13.97</td><td>6.34</td></tr></table></td><td>M1</td><td>Combine first two values.</td></tr>
<tr><td></td><td>Test stat = 1.7588 + 0.0761 + 0.6743 + 0.6309 + 0.4352</td><td>M1</td><td></td></tr>
<tr><td></td><td>3.58</td><td>A1</td><td>Accept 3.57 – 3.58 .
SC: 3.88, if values not combined, scores M1A0.</td></tr>
<tr><td></td><td>H_0 : distribution fits data
N(125, 12^2) is a good model for the data oe</td><td>B1</td><td>Must mention distribution and data.</td></tr>
<tr><td></td><td>4 degrees of freedom, so tabular value = 9.488
3.58 < 9.488
Accept H_0 .</td><td>M1</td><td>Compare their value with 9.488 (9.49) and correct FT conclusion.
Or, if values not combined, compare their value with 11.07 and correct FT conclusion for M1A0.</td></tr>
<tr><td></td><td>There is sufficient evidence that the normal distribution fits the data
There is sufficient evidence that the claim is supported by the data</td><td>A1</td><td>Correct conclusion, in context, following correct work. Level of uncertainty in language is used.</td></tr>
<tr><td></td><td></td><td>6</td><td></td></tr>
</table>

Q46.

4(a)	$p = 16.03$	**B1**	
	$q = 1.70$	**B1**	Condone 1.7
		2	

4(b)	<table><tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>$\geqslant 7$</td></tr><tr><td>12</td><td>20</td><td>36</td><td>32</td><td>13</td><td>6</td><td>1</td><td>0</td></tr><tr><td>9.85</td><td>24.63</td><td>30.78</td><td>25.65</td><td>***16.03***</td><td>8.02</td><td>3.34</td><td>***1.70***</td></tr></table> Combine last two columns: O value = 1, E value = 5.04	**M1**	Add last 2 columns (or 3 columns: 13.06).
	Calculate values of $\frac{(O-E)^2}{E}$: 0.4693 0.8704 0.8853 1.5720 0.5727 0.5088 3.2384	**M1**	Calculate values.
	Test statistic = 8.12	**A1**	
	H_0: data fits a Poisson distribution with mean 2.5 H_1: data does not fit a Poisson distribution with mean 2.5	**B1**	Need data and distribution. e.g Data fits the given distribution. Number of butterflies per square metre fits Po(2.5).
	Critical value of chi-squared = 10.64 Compare: '8.12' < 10.64, accept H_0	**M1**	Compare their test statistic with 10.64. Note: allow 9.236 if 3 columns combined or 12.02 if none combined.
	There is sufficient evidence to support the scientist's claim / there is sufficient evidence to suggest that the data fits a Poisson distribution with mean 2.5	**A1**	Correct conclusion, in context. Level of uncertainty in language is used.
		6	

3. Chi-squared tests

Q47.

2	H_0 : the size of shell is independent of the location H_1 : the size of shell is not independent of the location	**B1**	
	<table><tr><td>O</td><td>68</td><td>69</td><td>96</td><td>81</td></tr><tr><td>E</td><td>60.29</td><td>77.87</td><td>100.48</td><td>75.36</td></tr></table> <table><tr><td>O</td><td>28</td><td>55</td><td>64</td><td>39</td></tr><tr><td>E</td><td>35.71</td><td>46.13</td><td>59.52</td><td>44.64</td></tr></table>	**M1 A1**	Attempt at calculation of E values.
	Calculate values of $\frac{(O-E)^2}{E}$: 0.9859 1.0104 0.1997 0.4221 1.6646 1.7055 0.3372 0.7126	**M1**	Attempt at Chi-squared contributions. Accuracy of ±0.1.
	Test statistic = 7.04	**A1**	Correct total to 3sf.
	Tabular value for 3df = 6.251 7.04 > 6.251 Reject H_0	**M1**	Compare with 6.251 and correct FT conclusion. May be implied by FT correct conclusion in words.
	Insufficient evidence to suggest size of shell is independent of location/Evidence suggests that size of shell is dependent on (beach) location	**A1**	Correct conclusion, in context. Level of uncertainty in language is used.
		7	

Q48.

2(a)	Using B(8, 0.6) $p = 0.098, \quad q = 41.804, \quad r = 31.353$	**B1**	One correct.
		B1	Other two correct. If B0B0 scored, SCB1 if all 3 correct but not rounded to 3dp.
		2	
2(b)	H_0: the data fits the binomial distribution B(8, 0.6) H_1: the data does not fit the binomial distribution B(8, 0.6)	**B1**	
	Add first 3 columns: O value 8, E value 7.471 AND last two columns: O value 11, E value 15.956	**M1**	Both.
	Chi-squared sum: 0.0374 + 1.5817 + 2.9655 + 0.8058 + 0.9139 + 1.5394	**M1**	Method must be seen.
	7.84	**A1**	If M0 awarded SCB1 for 7.84.
	Critical value, 5df = 9.236 7.84 < 9.236 Accept H_0	**M1**	Compare calculated value with 9.236 OR 13.36 if no columns added. 12.02 adding only last 2 columns. 10.64 adding only first 3 columns.
	There is insufficient evidence to reject the director's claim.	**A1**	Correct work only except possibly B1 for hypotheses, in context, level of uncertainty in language. 'prove' scores A0
		6	

3. Chi-squared tests

Q49.

<table>
<tr><td rowspan="8">2</td><td>H_0: performance is independent of region
H_1: performance is not independent of region</td><td>B1</td><td></td></tr>
<tr><td><table><tr><td>62 (53.9)</td><td>41 (44.1)</td><td>44 (49)</td></tr><tr><td>102 (106.7)</td><td>94 (87.3)</td><td>95 (97)</td></tr><tr><td>56 (59.4)</td><td>45 (48.6)</td><td>61 (54)</td></tr></table></td><td>M1 A1</td><td>Calculate expected frequencies, allow 1 error.</td></tr>
<tr><td>Test stat = 1.217 + 0.218 + 0.510 + 0.207 + 0.514 + 0.041 + 0.195 + 0.267 + 0.907</td><td>M1</td><td>Calculate test statistic.</td></tr>
<tr><td>4.08</td><td>A1</td><td>SCB2 for 4.08 if totally unsupported.
SCM1A1 B1 for 4.08 supported only by expected frequencies.</td></tr>
<tr><td>Tabular value, 4df = 7.779
1.08 < 7.779
Accept H_0</td><td>M1</td><td>Compare with 7.779 and conclusion.</td></tr>
<tr><td>Insufficient evidence that performance is not independent of region</td><td>A1</td><td>Correct conclusion in context, following correct work, level of uncertainty in language.
‘Prove’ gives A0.</td></tr>
<tr><td></td><td>7</td><td></td></tr>
</table>

4. Non-parametric tests

Q1.

(a) State briefly the circumstances under which a non-parametric test of significance should be used rather than a parametric test. [1]

The level of pollution in a river was measured at 12 randomly chosen locations. The results, in suitable units, are shown below, where higher values represent greater pollution.

5.62 5.73 6.55 6.81 6.10 5.75 5.87 6.47 5.86 6.26 6.99 5.91

(b) Use a Wilcoxon signed-rank test to test whether the average pollution level in the river is more than 6.00. Use a 5% significance level. [6]

9231/04/SP/20 Q1

Q2.

The times, in milliseconds, taken by a computer to perform a certain task were recorded on 10 randomly chosen occasions. The times were as follows.

6.44 6.16 5.62 5.82 6.51 6.62 6.19 6.42 6.34 6.28

It is claimed that the median time to complete the task is 6.4 milliseconds.

(a) Carry out a Wilcoxon signed-rank test at the 5% significance level to test this claim. [6]

(b) State an underlying assumption that is made when using a Wilcoxon signed-rank test. [1]

9231/41/M/J/20 Q2

Q3.

A biologist is studying the effect of nutrients on the heights to which plants grow. A random sample of 24 similar young plants is divided into two equal groups *A* and *B*. The plants in group *A* are fed with nutrients and water and the plants in group *B* are given only water. After four weeks, the height, in cm, of each plant is measured and the results are as follows.

Group *A*	12.3	11.8	12.1	13.2	11.1	10.6	13.8	12.0	12.2	12.4	13.5	13.9
Group *B*	11.7	10.8	10.9	11.3	11.2	12.6	11.0	10.5	11.9	12.5	10.7	11.6

The biologist decides to carry out a test at the 5% significance level to test whether the nutrients have resulted in an increase in growth.

(a) She carries out a Wilcoxon rank-sum test. Give a reason why this is an appropriate choice of test. [1]

(b) Carry out the Wilcoxon rank-sum test for these results. [10]

9231/43/M/J/20 Q6

4. Non-parametric tests

Q4.

Metal rods produced by a certain factory are claimed to have a median breaking strength of 200 tonnes. For a random sample of 9 rods, the breaking strengths, measured in tonnes, were as follows.

210 186 188 208 184 191 215 198 196

A scientist believes that the median breaking strength of metal rods produced by this factory is less than 200 tonnes.

(a) Use a Wilcoxon signed-rank test, at the 5% significance level, to test whether there is evidence to support the scientist's belief. [6]

(b) Give a reason why a Wilcoxon signed-rank test is preferable to a sign test, when both are valid. [1]

9231/41/O/N/20 Q2

Q5.

A large school is holding an essay competition and each student has submitted an essay. To ensure fairness, each essay is given a mark out of 100 by two different judges. The marks awarded to the essays submitted by a random sample of 12 students are shown in the following table.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Judge 1	62	74	52	48	68	55	56	64	37	70	81	59
Judge 2	65	70	47	49	76	74	67	54	50	77	72	75

(a) One of the students claims that Judge 2 is awarding higher marks than Judge 1.

Carry out a Wilcoxon matched-pairs signed-rank test at the 5% significance level to test whether the data supports the student's claim. [7]

It is discovered later that the marks awarded to student *A* have been entered incorrectly. In fact, Judge 1 awarded 65 marks and Judge 2 awarded 62 marks.

(b) By considering how this change affects the test statistic, explain why the conclusion of the test carried out in part **(a)** remains the same. [2]

9231/42/O/N/20 Q2

4. Non-parametric tests

Q6.

Georgio has designed two new uniforms X and Y for the employees of an airline company. A random sample of 11 employees are each asked to assess each of the two uniforms for practicality and appearance, and to give a total score out of 100. The scores are given in the table.

Employee	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
Uniform X	82	74	42	59	60	73	94	98	62	36	50
Uniform Y	78	75	63	56	67	82	99	90	72	48	61

(a) Give a reason why a Wilcoxon signed-rank test may be more appropriate than a t-test for investigating whether there is any evidence of a preference for one of the uniforms. [1]

(b) Carry out a Wilcoxon matched-pairs signed-rank test at the 10% significance level. [7]

9231/41/M/J/21 Q5

Q7.

A company is developing a new flavour of chocolate by varying the quantities of the ingredients. A random selection of 9 flavours of chocolate are judged by two tasters who each give marks out of 100 to each flavour of chocolate.

Chocolate	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*
Taster 1	72	86	75	92	98	79	87	60	62
Taster 2	84	72	74	95	85	87	82	75	68

Carry out a Wilcoxon matched-pairs signed-rank test at the 10% significance level to investigate whether, on average, there is a difference between marks awarded by the two tasters. [7]

9231/43/M/J/21 Q2

Q8.

The blood cholesterol levels, measured in suitable units, of a random sample of 11 women and a random sample of 12 men are shown below.

Women	51	55	242	167	152	256	75	137	98	238	235	
Men	311	262	170	302	175	320	220	260	72	351	86	333

Carry out a Wilcoxon rank-sum test, at the 5% significance level, to test whether, on average, there is a difference in cholesterol levels between women and men. [9]

9231/41/O/N/21 Q6

Q9.

Applicants for a particular college take a written test when they attend for interview. There are two different written tests, A and B, and each applicant takes one or the other. The interviewer wants to determine whether the medians of the distribution of marks obtained in the two tests are equal. The marks obtained by a random sample of 8 applicants who took test A and a random sample of 8 applicants who took test B are as follows.

Test A	46	32	29	12	33	18	25	40
Test B	36	28	49	37	48	35	41	31

(a) Carry out a Wilcoxon rank-sum test at the 5% significance level to determine whether there is a difference in the population median marks obtained in the two tests. [6]

The interviewer considers using the given information to carry out a paired sample t-test to determine whether there is a difference in the population means for the two tests.

(b) Give two reasons why it is not appropriate to use this test. [2]

9231/42/O/N/21 Q4

Q10.

A teacher at a large college gave a mathematical puzzle to all the students. The median time taken by a random sample of 24 students to complete the puzzle was 18.0 minutes. The students were then given practice in solving puzzles. Two weeks later, the students were given another mathematical puzzle of the same type as the first. The times, in minutes, taken by the random sample of 24 students to complete this puzzle are as follows.

18.2	17.5	16.4	15.1	20.5	26.5	19.2	23.2
17.9	18.8	25.8	19.9	17.7	16.2	17.3	16.6
17.1	20.1	20.3	12.6	16.0	21.4	22.7	18.4

The teacher claims that the practice has not made any difference to the average time taken to complete a puzzle of this type.

Carry out a Wilcoxon signed-rank test, at the 10% significance level, to test whether there is sufficient evidence to reject the teacher's claim. [10]

9231/42/M/J/22 Q6

4. Non-parametric tests

Q11.

A manager claims that the lengths of the rubber tubes that his company produces have a median of 5.50 cm. The lengths, in cm, of a random sample of 11 tubes produced by this company are as follows.

5.56 5.45 5.47 5.58 5.54 5.52 5.60 5.35 5.59 5.51 5.62

It is required to test at the 10% significance level the null hypothesis that the population median length is 5.50 cm against the alternative hypothesis that the population median length is not equal to 5.50 cm.

Show that both a sign test and a Wilcoxon signed-rank test give the same conclusion and state this conclusion. [9]

9231/43/M/J/22 Q5

Q12.

A large college is holding a piano competition. Each student has played a particular piece of music and two judges have each awarded a mark out of 80. The marks awarded to a random sample of 14 students are shown in the following table.

Student	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*
Judge 1	79	54	63	74	69	52	50	57	55	42	63	55	56	48
Judge 2	75	62	60	73	76	41	31	51	45	55	49	50	65	36

(a) One of the students claims that on average Judge 1 is awarding higher marks than Judge 2. Carry out a Wilcoxon matched-pairs signed-rank test at the 5% significance level to test whether the data supports the student's claim. [7]

(b) Give a reason why it is preferable to use a Wilcoxon matched-pairs signed-rank test in this situation rather than a paired sample *t*-test. [1]

9231/41/O/N/22 Q3

4. Non-parametric tests

Q13.

The manager of a technology company A claims that his employees earn more per year than the employees at technology company B. The amounts earned per year, in hundreds of dollars, by a random sample of 12 employees from company A and an independent random sample of 12 employees from company B are shown below.

Company A	461	482	374	512	415	452	502	427	398	545	612	359
Company B	454	506	491	384	361	443	401	472	414	342	355	437

(a) Carry out a Wilcoxon rank-sum test at the 5% significance level to test whether the manager's claim is supported by the data. [9]

(b) Explain whether a paired sample t-test would be appropriate to test the manager's claim if earnings are normally distributed. [1]

9231/42/O/N/22 Q6

Solutions

Q1.

1(a)	When the population cannot be assumed to be normally distributed	**1**	**B1**	
1(b)	H_0: population median is 6.00, H_1: population median is greater than 6.00	1	**B1**	Both hypotheses stated
	Calculate deviations and resulting signed ranks	1	**M1**	
	<table><tr><td>Devs</td><td>−0.38</td><td>−0.27</td><td>0.55</td><td>0.81</td><td>0.10</td><td>−0.25</td><td>−0.13</td><td>0.47</td><td>−0.14</td><td>0.26</td><td>0.99</td><td>−0.09</td></tr><tr><td>Rank</td><td>−8</td><td>−7</td><td>10</td><td>11</td><td>2</td><td>−5</td><td>−3</td><td>9</td><td>−4</td><td>6</td><td>12</td><td>−1</td></tr></table>	1	**A1**	
	Test statistic $T = 8 + 7 + 5 + 3 + 4 + 1 = 28$	1	**A1**	
	Compare with correct critical value 17	1	**M1**	
	Conclusion: accept H_0; insufficient evidence that the median is greater than 6.00	1	**A1FT**	Conclusion to be stated in context, not just 'not significant'; follow through their value of T
		6		

Q2.

2(a)	H_0: population median is 6.40, H_1: population median $\neq$ 6.40	**B1**
	Calculate differences and signed ranks <table><tr><td>0.04</td><td>–0.24</td><td>–0.78</td><td>–0.58</td><td>0.11</td><td>0.22</td><td>–0.21</td><td>0.02</td><td>–0.06</td><td>–0.12</td></tr><tr><td>2</td><td>–8</td><td>–10</td><td>–9</td><td>4</td><td>7</td><td>–6</td><td>1</td><td>–3</td><td>–5</td></tr></table>	**M1A1**
	Test statistic = 4 + 7 + 2 +1 =14	**A1**
	Compare with correct critical value 8	**M1**
	Accept H_0: insufficient evidence to reject claim (**FT** *their* test statistic)	**A1 FT**
		6
2(b)	Symmetrically distributed about the median	**B1**
		1

4. Non-parametric tests

Q3.

<table>
<tr><td>6(a)</td><td>Difference of location test for populations not known to be normal</td><td>B1</td></tr>
<tr><td></td><td></td><td>1</td></tr>
<tr><td>6(b)</td><td>
<table>
<tr><td>10.6</td><td>2</td><td>10.5</td><td>1</td></tr>
<tr><td>11.1</td><td>7</td><td>10.7</td><td>3</td></tr>
<tr><td>11.8</td><td>12</td><td>10.8</td><td>4</td></tr>
<tr><td>12.0</td><td>14</td><td>10.9</td><td>5</td></tr>
<tr><td>12.1</td><td>15</td><td>11.0</td><td>6</td></tr>
<tr><td>12.2</td><td>16</td><td>11.2</td><td>8</td></tr>
<tr><td>12.3</td><td>17</td><td>11.3</td><td>9</td></tr>
<tr><td>12.4</td><td>18</td><td>11.6</td><td>10</td></tr>
<tr><td>13.2</td><td>21</td><td>11.7</td><td>11</td></tr>
<tr><td>13.5</td><td>22</td><td>11.9</td><td>13</td></tr>
<tr><td>13.8</td><td>23</td><td>12.5</td><td>19</td></tr>
<tr><td>13.9</td><td>24</td><td>12.6</td><td>20</td></tr>
</table>
</td><td>M1</td></tr>
<tr><td></td><td>Total ranks: 109 or 191</td><td>A1</td></tr>
<tr><td></td><td>H_0: $m_x = m_y$ H_1: $m_x > m_y$</td><td>B1</td></tr>
<tr><td></td><td>Use normal approximation with attempts at mean and variance</td><td>M1</td></tr>
<tr><td></td><td>mean = 150, variance = 300</td><td>A1</td></tr>
</table>

Q4.

2(a)	H_0: $m = 200$ H_1: $m < 200$	**B1**	Allow 'median' in words, allow m not defined
	10 −14 −12 8 −16 −9 15 −2 −4	**M1**	Signed differences (at most 3 errors)
	Ranks: 5 −7 −6 3 −9 −4 8 −1 −2	**A1**	Award for correct rank order, ignore signs
	Sum ranks T = 16	**A1**	CWO
	Critical value 8 and compare '16' > 8	**M1**	Compare their T with 8
	Accept H_0 Insufficient evidence to support scientist's belief	**A1**	In context, all correct, except possibly hypotheses. Level of uncertainty in language used. No contradictions
		6	
2(b)	Magnitude of differences from median are taken into account	**B1**	Must mention magnitude and differences
		1	

Q5.

2(a)	H_0: difference in (population) medians = 0 H_1: difference in (population) medians < 0 (or >0)	**B1 B1**	Correct hypotheses (median or m) All notation identified e.g. m = population median if used
	Differences: −3 4 5 −1 −8 −19 −11 10 −13 −7 9 −16	**M1**	At most 3 errors
	Ranks: −2 3 4 −1 −6 −12 −9 8 −10 −5 7 −11	**A1**	Award for correct rank order, ignore signs
	Sum T = 3 + 4 + 8 + 7 = 22	**A1**	cwo
	Compare with critical value 17: '22' > 17	**M1**	Compare *their* T with 17
	Accept H_0 Data does not support student's claim	**A1**	In context, all correct, except possibly second B1 Level of uncertainty in language used. No contradictions.
		7	
2(b)	Rank for A becomes + 2,	**M1**	T = 24
	Changing sign of difference can only reduce evidence in favour of the claim.	**A1**	still > 17 and test result unchanged
		2	

Q6.

5(a)	Underlying distribution or population of differences is unknown Underlying distributions or population of scores for both X and Y are unknown	**B1**	Not known to be normal. Condone 'scores for X and Y cannot be assumed to be normally distributed'.
		1	
5(b)	H_0 : difference of population medians = 0	**B1**	Correct hypotheses, allow m but not μ or mean.
	H_1 : difference of population median ≠ 0	**B1**	'population' included.
	Diff: 4 −1 −21 3 −7 −9 −5 8 −10 −12 −11	**M1**	Differences, allow one error.
	Rank: 3 −1 −11 2 −5 −7 −4 6 −8 −10 −9	**M1**	Signed ranks, allow one sign error.
	$[Q=-55,]P=11$	**A1**	$P = 11$
	Critical tabular value = 13 '11' < 13 so reject H_0	**M1**	Comparison with 13 and correct ft conclusion.
	Sufficient evidence of a preference for one of the uniforms OE	**A1**	Correct conclusion, in context, following correct work, except possibly second B1. Level of uncertainty in language is used.
		7	

4. Non-parametric tests

Q7.

2	H_0: difference in population medians = 0 ($m_a = m_b$)	**B1**	Correct hypotheses, allow m_d . Do not allow μ or mean.
	H_1: difference in population medians ≠ 0 ($m_a \neq m_b$)	**B1**	'population' included.
	Diff: 12 −14 −1 3 −13 8 −5 15 6	**M1**	Allow one error.
	Rank: 6 −8 −1 2 −7 5 −3 9 4	**M1**	Allow one error.
	Smaller sum 19	**A1**	(Other sum is 26.)
	Critical tabular value = 8 '19' > 8 so accept H_0	**M1**	Comparison with 8 and correct FT conclusion.
	There is insufficient evidence that marks differ	**A1**	Correct conclusion, in context, following correct work, except possibly the hypotheses. Level of uncertainty in language is used.
		7	

Q8.

6		**M1**	Attempt at ranking.
	Total ranks: 95	**A1**	
	$H_0 : m_x = m_y$ and $H_1 : m_x \neq m_y$	**B1**	Allow words, must include 'population' median.
	Use normal approximation with attempts at mean and variance.	**M1**	
	Mean =132, variance = 264	**A1**	
	$\frac{95.5-132}{\sqrt{264}}$	**M1**	Allow no or wrong continuity correction for M1 only.

	51	1	72	3
	55	2	86	5
	75	4	170	10
	98	6	175	11
	137	7	220	12
	152	8	260	17
	167	9	262	18
	235	13	302	19
	238	14	311	20
	242	15	320	21
	256	16	333	22
			351	23

6	– 2.246	**A1**	CAO
	Critical value is – 1.96. – 2.246 < – 1.96 reject H_0 .	**M1**	Compare their value with – 1.96. Or area comparison 0.0123 or 0.0124 with 0.025 and FT conclusion.
	There is sufficient evidence of a difference in levels.	**A1**	Correct conclusion, in context, following correct work. Level of uncertainty in language is used.
		9	

Q9.

4(a)	12 1 28 4 18 2 31 6 25 3 35 9 29 5 36 10 32 7 37 11 33 8 41 13 40 12 48 15 46 14 49 16	**M1**	Attempt at ranking.
	Test statistic: 52	**A1**	
	$H_0 : m_x = m_y$ and $H_1 : m_x \neq m_y$	**B1**	Allow in words but 'population' must be included.
	Critical value for (8, 8) is 49.	***B1**	Allow 51 if clearly one-tail test in hypotheses.
	52 > 49 Accept H_0	**DM1**	Compare their calculated value with 49 and correct FT conclusion.
	Insufficient evidence of difference in medians.	**A1**	Correct conclusion, in context, following correct work. Level of uncertainty in language is used.
		6	
4(b)	Not a paired sample.	**B1**	
	Underlying distribution/population not (known to be) normal Underlying distribution/population unknown	**B1**	B0 for 'data is not normally distributed' B0 for 'marks are not normally distributed'
		2	

Q10.

6	Differences/rank: 0.2 −0.5 −1.6 −2.9 2.5 8.5 1.2 5.2 **2 −5 −11 −18 17 24 9 21** −0.1 0.8 7.8 1.9 −0.3 −1.8 −0.7 −1.4 **−1 7 23 13 −3 −12 −6 −10** −0.9 2.1 2.3 −5.4 −2.0 3.4 4.7 0.4 **−8 15 16 −22 −14 19 20 4**	**M1 A1**	Attempt at differences, allow 4 errors.
	Q = 110 P = 190	**A1**	Either sum correct.
	H_0: population median =18.0 and H_1: population median ≠ 18.0	**B1**	Must be 'population', allow *m*.
	Mean = $\frac{1}{4}n(n+1) = \frac{1}{4} \times 24 \times 25 = 150$	**B1**	Normal approximation: mean.
	Variance = $\frac{1}{24}n(n+1)(2n+1) = \frac{1}{24} \times 24 \times 25 \times 49 = 1225$	**B1**	
	$\frac{110.5 - 150}{\sqrt{1225}}$	**M1**	Attempt at test statistic, allow incorrect or missing cc, FT *their* 110.
	−1.129	**A1**	

6	10% 2-tail, critical value is 1.645 $-1.129 > -1.645$ or $0.1296 > 0.05$, accept H_0.	**M1**	Compare with -1.645, correct FT conclusion.
	Insufficient evidence to suggest that median is not 18.0/ Insufficient evidence to reject teacher's claim.	**A1**	Correct conclusion, in context. Level of uncertainty in language is used.
		10	

Q11.

5	Signs: + – – + + + + – + + + and use Binomial	**M1**	11 signs (values not needed) allow one error.
	Consider B(11, 0.5) with P($X \geqslant 8$) or P($X \leqslant 3$)	**M1**	Use of correct binomial.
	Compare 0.227 with 0.1 or 0.113 with 0.05	**A1**	
	Differences: 0.06 –0.05 –0.03 0.08 0.04 0.02 0.10 –0.15 0.09 0.01 0.12	**M1**	Allow one error.
	Signed Ranks: 6 -5 -3 7 4 2 9 -11 8 1 10	**M1**	Allow one swap.
	(P = 47) Q = 19	**A1**	
	Critical value from table = 13	**B1**	
	19 > 13 Accept H_0	**M1**	FT *their* (19) compared with 13 and correct FT conclusion
	Insufficient evidence to support median not being 5.50	**A1**	BOTH conclusions correct from correct working, in context. Level of uncertainty in language.
		9	

Q12.

3(a)	Differences: 4 -8 3 1 –7 11 19 6 10 –13 14 5 –9 12	**M1**	Attempt at signed differences, allow 2 errors.
	Signed Ranks: 3 –7 2 1 –6 10 14 5 9 –12 13 4 –8 11	**M1**	Ranks, allow 4 errors.
	[P = 72, Q = 33] T = 33	**A1**	33 clearly identified.
	H_0: difference in population medians = 0 H_1: difference in population medians > 0	**B1**	Correct, allow m for this mark.
	Critical tabular value = 25	**B1**	
	33 > 25 accept H_0	**M1**	Compares their value of sum of ranks with their 25 and conclusion.
	Insufficient evidence to support claim	**A1**	All correct except possibly first B1, in context, level of uncertainty in language. 'Prove' scores A0.
		7	
3(b)	Underlying/population distribution of differences is unknown (not known to be normal)	**B1**	
		1	

Q13.

6(a)		**M1**	Wrong test: max B1 for hypotheses, B1B1 for correct mean and variance Attempt at rankings (allow up to 4 errors)

	359	3	342	1
	374	5	355	2
	398	7	361	4
	415	10	384	6
	427	11	401	8
	452	14	414	9
	461	16	437	12
	482	18	443	13
	502	20	454	15
	512	22	472	17
	545	23	491	19
	612	24	506	21

Test statistic = 127	**A1**	Clearly identified
H_0: population medians are equal H_1: population median for X is greater than population median for Y	**B1**	
Mean = $\frac{1}{2}\times 12\times 25 = 150$	**B1**	
Variance = $\frac{1}{12}\times 12\times 12\times 25 = 300$	**B1**	

5. Probability Generating Functions

Q1.

Aisha has a bag containing 3 red balls and 3 white balls. She selects a ball at random, notes its colour and returns it to the bag; the same process is repeated twice more. The number of red balls selected by Aisha is denoted by X.

(a) Find the probability generating function $G_X(t)$ of X. [2]

Basant also has a bag containing 3 red balls and 3 white balls. He selects three balls at random, without replacement, from his bag. The number of red balls selected by Basant is denoted by Y.

(b) Find the probability generating function $G_Y(t)$ of Y. [3]

The random variable Z is the total number of red balls selected by Aisha and Basant.

(c) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(d) Use the probability generating function of Z to find E(Z) and Var(Z). [5]

9231/04/SP/20 Q6

Q2.

A bag contains 4 red balls and 6 blue balls. Rassa selects two balls at random, without replacement, from the bag. The number of red balls selected by Rassa is denoted by X.

(a) Find the probability generating function, $G_X(t)$, of X. [2]

Rassa also tosses two coins. One coin is biased so that the probability of a head is $\frac{2}{3}$. The other coin is biased so that the probability of a head is p. The probability generating function of Y, the number of heads obtained by Rassa, is $G_Y(t)$. The coefficient of t in $G_Y(t)$ is $\frac{7}{12}$.

(b) Find $G_Y(t)$. [3]

The random variable Z is the sum of the number of red balls selected and the number of heads obtained by Rassa.

(c) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(d) Use the probability generating function of Z to find E(Z). [2]

9231/41/M/J/20 Q6

Q3.

The discrete random variable X has probability generating function $G_X(t)$ given by

$$G_X(t) = 0.2t + 0.5t^2 + 0.3t^3.$$

The random variable Y is the sum of two independent observations of X.

(a) Find the probability generating function of Y, giving your answer as an expanded polynomial in t. [3]

(b) Use the probability generating function of Y to find $E(Y)$ and $Var(Y)$. [5]

9231/43/M/J/20 Q4

Q4.

Keira has two unbiased coins. She tosses both coins. The number of heads obtained by Keira is denoted by X.

(a) Find the probability generating function $G_X(t)$ of X. [1]

Hassan has three coins, two of which are biased so that the probability of obtaining a head when the coin is tossed is $\frac{1}{3}$. The corresponding probability for the third coin is $\frac{1}{4}$. The number of heads obtained by Hassan when he tosses these three coins is denoted by Y.

(b) Find the probability generating function $G_Y(t)$ of Y. [3]

The random variable Z is the total number of heads obtained by Keira and Hassan.

(c) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(d) Use the probability generating function of Z to find $E(Z)$. [2]

9231/41/O/N/20 Q5

Q5.

The random variable X has the binomial distribution $B(n, p)$.

(a) Write down an expression for $P(X = r)$ and hence show that the probability generating function of X is $(q + pt)^n$, where $q = 1 - p$. [3]

(b) Use the probability generating function of X to prove that $E(X) = np$ and $Var(X) = np(1 - p)$. [5]

9231/42/O/N/20 Q5

Q6.

Tanji has a bag containing 4 red balls and 2 blue balls. He selects 3 balls at random from the bag, without replacement. The number of red balls selected by Tanji is denoted by X.

(a) Find the probability generating function $G_X(t)$ of X. [2]

Tanji also has two coins, each biased so that the probability of obtaining a head when it is thrown is $\frac{1}{4}$. He throws the two coins at the same time. The number of heads obtained is denoted by Y.

(b) Find the probability generating function $G_Y(t)$ of Y. [2]

The random variable Z is the sum of the number of red balls selected by Tanji and the number of heads obtained.

(c) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(d) Use the probability generating function of Z to find $E(Z)$ and $\text{Var}(Z)$. [5]

9231/41/M/J/21 Q6

Q7.

X is a discrete random variable which takes the values 0, 2, 4, … . The probability generating function of X is given by

$$G_X(t) = \frac{1}{3-2t^2}.$$

(a) Find $E(X)$ and $\text{Var}(X)$. [5]

(b) Find $P(X = 4)$. [3]

9231/43/M/J/21 Q4

Q8.

Nine balls labelled 1, 2, 3, 4, 5, 6, 7, 8, 9 are placed in a bag. Kai selects three balls at random from the bag, without replacement. The random variable X is the number of balls selected by Kai that are labelled with a multiple of 3.

(a) Find the probability generating function $G_X(t)$ of X. [3]

The balls are replaced in the bag.

Jacob now selects two balls at random from the bag, without replacement. The random variable Y is the number of balls selected by Jacob that are labelled with an even number.

(b) Find the probability generating function $G_Y(t)$ of Y. [2]

The random variable Z is the sum of the number of balls that are labelled with a multiple of 3 selected by Kai and the number of balls that are labelled with an even number selected by Jacob.

(c) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(d) Use the probability generating function of Z to find $E(Z)$. [2]

9231/41/O/N/21 Q5

Q9.

The random variable X is such that $P(X = r) = kr^2$ for $r = 1,2,3,4$, where k is a constant.

(a) Find the value of k. [1]

(b) Find the probability generating function $G_X(t)$ of X. [2]

The random variable Y has probability generating function $G_Y(t) = \frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2$.

The random variable Z is the sum of X and Y.

(c) Assuming that X and Y are independent, find the probability generating function $G_Z(t)$ of Z as a polynomial in t. [3]

(d) Given that $E(Z) = \frac{13}{3}$, use $G_Z(t)$ to find $Var(Z)$. [3]

9231/42/O/N/21 Q5

Q10.

Andrew has five coins. Three of them are unbiased. The other two are biased such that the probability of obtaining a head when one of them is tossed is $\frac{3}{5}$.

Andrew tosses all five coins. It is given that the probability generating function of X, the number of heads obtained on the unbiased coins, is $G_X(t)$, where

$$G_X(t) = \tfrac{1}{8} + \tfrac{3}{8}t + \tfrac{3}{8}t^2 + \tfrac{1}{8}t^3.$$

(i) Find $G_Y(t)$, the probability generating function of Y, the number of heads on the biased coins. **[3]**

(ii) The random variable Z is the total number of heads obtained when Andrew tosses all five coins. Find the probability generating function of Z, giving your answer as a polynomial. **[3]**

(iii) Find $E(Z)$ and $Var(Z)$. **[6]**

(iv) Write down the value of $P(Z = 3)$. **[1]**

Q11.

The discrete random variable Y has probability generating function

$$G_Y(t) = 0.09t^2 + 0.24t^3 + 0.34t^4 + 0.24t^5 + 0.09t^6.$$

(i) Find the mean and variance of Y. **[5]**

Y is the sum of two independent observations of a random variable X.

(ii) Find the probability generating function of X, expressing your answer as a cubic polynomial in t. **[3]**

(iii) Write down the value of $P(X = 2)$. **[1]**

Q12.

The probability generating function of the random variable X is given by

$$G(t) = \frac{1 + at}{4 - t},$$

where a is a constant.

(i) Find the value of a. [2]

(ii) Find $P(X = 3)$. [4]

The sum of 3 independent observations of X is denoted by Y. The probability generating function of Y is denoted by $H(t)$.

(iii) Use $H(t)$ to find $E(Y)$. [5]

(iv) By considering $H(-1) + H(1)$, show that $P(Y \text{ is an even number}) = \frac{62}{125}$. [2]

Q13.

The probability generating function of the discrete random variable Y is given by

$$G_Y(t) = \frac{a + bt^3}{t},$$

where a and b are constants.

(i) Given that $E(Y) = -0.7$, find the values of a and b. **[4]**

(ii) Find $\text{Var}(Y)$. **[2]**

(iii) Find the probability that the sum of 10 random observations of Y is -7. **[4]**

Q14.

The probability generating function of the discrete random variable X is $\frac{e^{4t^2}}{e^4}$. Find

(i) $E(X)$, **[3]**

(ii) $P(X = 2)$. **[3]**

5. Probability Generating Functions

Q15.

The random variable X has the distribution $B(n, p)$.

(i) Show, from the definition, that the probability generating function of X is $(q + pt)^n$, where $q = 1 - p$. **[2]**

(ii) The independent random variable Y has the distribution $B(2n, p)$ and $T = X + Y$. Use probability generating functions to determine the distribution of T, giving its parameters. **[4]**

Q16.

The probability generating function, $G_Y(t)$, of the random variable Y is given by

$$G_Y(t) = 0.04 + 0.2t + 0.37t^2 + 0.3t^3 + 0.09t^4.$$

(a) Find Var(Y). [4]

The random variable Y is the sum of two independent observations of the random variable X.

(b) Find the probability generating function of X, giving your answer as a polynomial in t. [3]

9231/41/M/J/22 Q2

Q17.

George throws two coins, A and B, at the same time. Coin A is biased so that the probability of obtaining a head is a. Coin B is biased so that the probability of obtaining a head is b, where $b < a$. The probability generating function of X, the number of heads obtained by George, is $G_X(t)$. The coefficients of t and t^2 in $G_X(t)$ are $\frac{5}{12}$ and $\frac{1}{12}$ respectively.

(a) Find the value of a. [2]

The random variable Y is the sum of two independent observations of X.

(b) Find the probability generating function of Y, giving your answer as a polynomial in t. [3]

(c) Find Var(Y). [3]

9231/43/M/J/22 Q3

5. Probability Generating Functions

Q18.

Jason has three biased coins. For each coin the probability of obtaining a head when it is thrown is $\frac{2}{3}$. Jason throws all three coins. The number of heads obtained is denoted by X.

(a) Find the probability generating function $G_X(t)$ of X. [3]

Jason also has two unbiased coins. He throws all five coins. The number of heads obtained from the two unbiased coins is denoted by Y. It is given that $G_Y(t) = \frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2$. The random variable Z is the total number of heads obtained when Jason throws all five coins.

(b) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(c) Find $E(Z)$. [2]

9231/41/O/N/22 Q4

Q19.

A 6-sided dice, A, with faces numbered 1, 2, 3, 4, 5, 6 is biased so that the probability of throwing a 6 is $\frac{1}{4}$. The random variable X is the number of 6s obtained when dice A is thrown twice.

(a) Find the probability generating function of X. [2]

A second dice, B, with faces numbered 1, 2, 3, 4, 5, 6 is unbiased. The random variable Y is the number of 6s obtained when dice B is thrown twice.

The random variable Z is the total number of 6s obtained when both dice are thrown twice.

(b) Find the probability generating function of Z, expressing your answer as a polynomial. [3]

(c) Find $\text{Var}(Z)$. [3]

9231/42/O/N/22 Q5

Solutions

Q1.

6(a)	Probabilities $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$ for 0, 1, 2, 3 reds	1	**B1**	
	$G_X(t) = \frac{1}{8} + \frac{3}{8}t + \frac{3}{8}t^2 + \frac{1}{8}t^3$	1	**B1FT**	Follow through their probabilities so long as $\Sigma p = 1$
		2		
6(b)	Correct value $\frac{9}{20}$ or 0.45 for $P(Y = 1)$ or $P(Y = 2)$	1	**B1**	
	Probabilities $\frac{1}{20}, \frac{9}{20}, \frac{9}{20}, \frac{1}{20}$ for 0, 1, 2, 3 reds	1	**B1**	
	$G_Y(t) = \frac{1}{20} + \frac{9}{20}t + \frac{9}{20}t^2 + \frac{1}{20}t^3$	1	**B1FT**	Follow through their probabilities so long as $\Sigma p = 1$
		3		
6(c)	$G_Z(t) = G_X(t) \times G_Y(t)$	1	**B1**	Stated or implied
	Complete expansion of the product of their cubics	1	**M1**	
	$G_Z(t) = \frac{1}{160}(1 + 12t + 39t^2 + 56t^3 + 39t^4 + 12t^5 + t^6)$	1	**A1**	Or equivalent, e.g. with explicit fractional coefficients
		3		
6(d)	Attempt to differentiate $G_Z(t)$ and evaluate $G'_Z(1)$	1	**M1**	
	$E(Z) = 3$	1	**A1**	Correctly obtained from $G'_Z(1)$
	Attempt to find second derivative $G''_Z(t)$	1	**M1**	
	Use of $G''_Z(1) + G'_Z(1) - (G'_Z(1))^2$	1	**M1**	
	$\text{Var}(Z) = 1.2$	1	**A1**	Correct value correctly obtained
		5		

Q2.

6(a)	$P(BB) = \frac{6}{10} \times \frac{5}{9} = \frac{1}{3}$ $P(RB, BR) = \frac{6}{10} \times \frac{4}{9} = \frac{8}{15}$ $P(RR) = \frac{4}{10} \times \frac{3}{9} = \frac{2}{15}$	**M1**
	$\frac{1}{3} + \frac{8}{15}t + \frac{2}{15}t^2$ (**FT** *their* probabilities)	**A1 FT**
		2
6(b)	$P(1\text{ H}) = \frac{2}{3}(1-p) + \frac{1}{3}p = \frac{7}{12}$	**M1**
	$p = \frac{1}{4}$	**A1**
	PGF of $Y = \frac{1}{4} + \frac{7}{12}t + \frac{1}{6}t^2$	**A1**
		3
6(c)	PGF of $Z = \left(\frac{1}{3} + \frac{8}{15}t + \frac{2}{15}t^2\right)\left(\frac{1}{4} + \frac{7}{12}t + \frac{1}{6}t^2\right)$	**B1M1**
	$= \frac{1}{180}(15 + 59t + 72t^2 + 30t^3 + 4t^4)$ **AEF**	**A1**
		3
6(d)	Attempt to differentiate their $G_Z(t)$ and evaluate $G'_Z(1)$	**M1**
	$E(Z) = (59 + 144 + 90 + 16)/180 = \frac{309}{180} = 1.72$	**A1**
		2

5. Probability Generating Functions

Q3.

4(a)	$G_X(t) = 0.2t + 0.5t^2 + 0.3t^3$	**B1**
	$G_Y(t) = \left(0.2t + 0.5t^2 + 0.3t^3\right)^2$	**M1**
	$= 0.04t^2 + 0.2t^3 + 0.37t^4 + 0.3t^5 + 0.09t^6$	**A1**
		3
4(b)	$G'_Y(t) = 0.08t + 0.6t^2 + 1.48t^3 + 1.5t^4 + 0.54t^5$	**M1**
	$E(Y) = 4.2$	**A1**
	$G''_Y(t) = 0.08 + 1.2t + 4.44t^2 + 6t^3 + 2.7t^4$	**M1**
	Use $G''_Y(1) + G'_Y(1) - (G'_Y(1))^2$	**M1**
	$\text{Var}(Y) = 14.42 + 4.2 - 4.2^2 = 0.98$	**A1**
		5

Q4.

5(a)	$G_X(t) = \frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2$	**B1**	Accept $(0.5+0.5t)^2$
		1	
5(b)	$P(0H) = \frac{12}{36}$ $P(1H) = \frac{16}{36}$ $P(2H) = \frac{7}{36}$ $P(3H) = \frac{1}{36}$	**M1 A1**	Attempt at probs, at least 2 correct All correct
	$G_Y(t) = \frac{12}{36} + \frac{16}{36}t + \frac{7}{36}t^2 + \frac{1}{36}t^3$	**B1 FT**	FT *their* probabilities, must be cubic with 4 non-zero terms
		3	
5(c)	$G_Z(t) = \left(\frac{1}{4} + \frac{1}{2}t + \frac{1}{4}t^2\right)\left(\frac{12}{36} + \frac{16}{36}t + \frac{7}{36}t^2 + \frac{1}{36}t^3\right)$	**M1**	Attempt to multiply their two PGF
	$= \frac{1}{144}\left(12 + 40t + 51t^2 + 31t^3 + 9t^4 + t^5\right)$	**M1 A1**	Obtain quintic expression and collect terms
		3	
5(d)	$G'_Z(t) = \frac{1}{144}\left(40 + 102t + 93t^2 + 36t^3 + 5t^4\right)$	**M1**	Differentiate
	$E(Z) = G'_Z(1) = \frac{23}{12} = (= 1.92)$	**A1**	
		2	
5(e)	2	**B1 FT**	FT power of term with largest coefficient in *their* $G_Z(t)$
		1	

Q5.

5(a)	$P(X=r) = {}^nC_r p^r (1-p)^{n-r}$ or ${}^nC_r p^r q^{n-r}$	**B1**	
	$G_X(t) = \sum_0^n nCr\, p^r (1-p)^{n-r} t^r$	**M1**	Accept minimum of 4 terms including the last. Shown with specific value of n is M0
	$\sum_0^n nCr(pt)^r (1-p)^{n-r}$ $= (q+pt)^n$	**A1**	At least one intermediate step to be shown, with p and t grouped AG
		3	
5(b)	$G'_X(t) = n(q+pt)^{n-1} \times p$	**M1**	
	So $E(X) = G'_X(1) = np(q+p)^{n-1}$ and $q+p=1$ so $E(X) = np$	**A1**	
	$G''_X(t) = n(n-1)(q+pt)^{n-2} \times p \times p$	**M1**	
	$Var(X) = n(n-1)p^2 + np - (np)^2$	**M1**	
	$np(1-p)$	**A1**	
		5	

Q6.

6(a)	P(3R, 2R, 1R, 0R) $= \frac{1}{5}, \frac{3}{5}, \frac{1}{5}, 0$	**M1**	At least 2 correct probabilities used in a polynomial.
	$G_X(t) = \frac{1}{5}t + \frac{3}{5}t^2 + \frac{1}{5}t^3$	**A1**	
		2	
6(b)	P(2H, 1H, 0H) $= \frac{1}{16}, \frac{6}{16}, \frac{9}{16}$	**M1**	One correct probability and all three adding to 1 used in a polynomial.
	$G_Y(t) = \frac{9}{16} + \frac{6}{16}t + \frac{1}{16}t^2$	**A1**	
		2	
6(c)	Attempt to multiply results from part **(a)** and part **(b)**	**M1**	
	Obtain polynomial	**M1**	
	$\frac{9}{80}t + \frac{33}{80}t^2 + \frac{28}{80}t^3 + \frac{9}{80}t^4 + \frac{1}{80}t^5$	**A1**	Accept exact equivalent decimals.
		3	
6(d)	$E(Z) = G_Z'(1) = \frac{1}{80}(9+66+84+36+5)$	**M1**	Differentiate and put $t = 1$.
	$= \frac{200}{80} = 2.5$	**A1**	
	$G_Z''(1) = \frac{1}{80}(66+168+108+20) = 4.525 = \left[\frac{362}{80}\right]$	**M1**	

6(d)	$\text{Var}(Z) = G''(1) + G'(1) - (G'(1))^2$	**M1**	Use result.
	$\frac{362}{80} + \frac{200}{80} - \left(\frac{200}{80}\right)^2 = 0.775 = \left[\frac{31}{40}\right]$	**A1**	
		5	

Q7.

4(a)	$G'(t) = \frac{4t}{(3-2t^2)^2}$	**M1**	Differentiate to obtain $\frac{kt}{(3-2t^2)^2}$ OE.
	so $E(X) = G'(1) = 4$	**A1**	CAO WWW
	$G''(t) = \frac{12+24t^2}{(3-2t^2)^3}$ or $4(3-2t^2)^{-2} + 32t^2(3-2t^2)^{-3}$	**M1**	OE. Differentiate, allow only numerical or sign slips.
	$\text{Var}(X) = G''(1) + G'(1) - (G'(1))^2 = 36 + 4 - 16$	**M1**	Substitute their values into correct formula, dependent on attempt at $G''(t)$.
	$[\text{Var}(X)] = 24$	**A1**	CAO WWW
		5	
4(b)	$G_X(t) = \frac{1}{3-2t^2} = (3-2t^2)^{-1}$ $= \frac{1}{3}\left(1 + \frac{2}{3}t^2 + \frac{4}{9}t^4 + \ldots\right)$	**M1**	Expand given expression or give expression for the term in t^4.
	$P(X = 4)$ = their coefficient of t^4	**M1**	Found from legitimate method.
	$[P(X = 4) =] \frac{4}{27}$	**A1**	WWW
		3	

Q8.

5(a)	$P(3, 6, 9) = \frac{3}{9} \times \frac{2}{8} \times \frac{1}{7} = \frac{6}{504}$ $P(\text{Two of } 3, 6, 9) = \frac{3}{9} \times \frac{2}{8} \times \frac{6}{7} \times 3 = \frac{108}{504}$ $P(\text{one of } 3, 6, 9) = \frac{3}{9} \times \frac{6}{8} \times \frac{5}{7} \times 3 = \frac{270}{504}$ $P(\text{none of } 3, 6, 9) = \frac{6}{9} \times \frac{5}{8} \times \frac{4}{7} = \frac{120}{504}$	**B1**	At least 2 probabilities correct.
	$G_X(t) = \frac{20}{84} + \frac{45}{84}t + \frac{18}{84}t^2 + \frac{1}{84}t^3$	**M1** **A1**	Attempt with at least 3 probabilities in a polynomial, CAO.
		3	
5(b)	$P(\text{both even}) = \frac{12}{72}$ $P(\text{one even}) = \frac{40}{72}$ $P(\text{no even}) = \frac{20}{72}$	**M1**	
	$G_Y(t) = \frac{5}{18} + \frac{10}{18}t + \frac{3}{18}t^2$	**A1**	CAO
		2	

5(c)	$\frac{1}{1512}\left(\frac{20}{84}+\frac{45}{84}t+\frac{18}{84}t^2+\frac{1}{84}t^3\right)\left(\frac{5}{18}+\frac{10}{18}t+\frac{3}{18}t^2\right)$	**M1**	Method and attempt to multiply.
		M1	Multiplication to obtain single quintic polynomial.
	$\frac{1}{1512}(100+425t+600t^2+320t^3+64t^4+3t^5)$	**A1**	CAO
		3	
5(d)	$G'(1)=\frac{1}{1512}(425+1200+960+256+15)$	**M1**	
	$\frac{17}{9}$	**A1**	CWO
		2	

Q9.

5(a)	$(1+4+9+16)k=1$ so $k=\frac{1}{30}$	**B1**	
		1	
5(b)	$G_X(t)=\frac{1}{30}t+\frac{4}{30}t^2+\frac{9}{30}t^3+\frac{16}{30}t^4$	**M1A1**	Using their k in a polynomial, at least two terms correct for their k.
		2	
5(c)	$\left(\frac{1}{30}t+\frac{4}{30}t^2+\frac{9}{30}t^3+\frac{16}{30}t^4\right)\left(\frac{1}{4}+\frac{1}{2}t+\frac{1}{4}t^2\right)$	**M1**	Method and attempt to multiply.
	$\frac{1}{120}\left(t+6t^2+18t^3+38t^4+41t^5+16t^6\right)$	**M1A1**	Multiplication to obtain single polynomial of order 6.
		3	
5(d)	Given: G'(1) = 13/3 $G''(t)=\frac{1}{120}(12+108t+456t^2+820t^3+480t^4)$	**M1**	Differentiate twice.
	$\text{Var}(X)=G''(1)+\frac{13}{3}-\left(\frac{13}{3}\right)^2$	**M1**	Use correct formula.
	$\frac{1876}{120}+\frac{13}{3}-\frac{169}{9}=\frac{107}{90}$ or 1.19	**A1**	CAO
		3	

Q10.

6	(i)		$P(0) = 0.16, P(1) = 0.48, P(2) = 0.36$ $G_Y(t) = 0.16 + 0.48t + 0.36t^2$	B1 M1A1 **[3]**	
	(ii)		$G_X(t) \times G_Y(t)$ soi $0.02 + 0.12t + 0.285t^2 + 0.335t^3 + 0.195t^4 + 0.045t^5$	M1 A1A1 **[3]**	At least 4 terms correct; All correct.
	(iii)		$E(Z) = G_Z'(1)$ [$=0.12+0.57t+1.005t^2+0.78t^3+0.225t^4$] Sub $t = 1$ 2.7 Attempt 2nd derivative of G_Z Attempt $G''(1) + G'(1) - G'(1))^2$ $(G''(1) = 5.82)$ 1.23 Alternative methods. 3x0.5 or 2x0.6; added ; 2.7 M1M1A1 3x0.5x0.5 or 2x0.6x0.4; added; 1.23 M1M1A1 P(Z=0)=0.02 etc B1 ;E(Z)=Σzp=2.7 M1A1 $E(Z^2)=\Sigma z^2p=(8.52)$ M1;-2.7^2 M1 1.23A1	M1 M1dep A1 M1 M1dep A1 **[6]**	Differentiate. -ve var, M0
	(iv)		0.335	B1ft **[1]**	Coeff t^3 from (ii)

Q11.

4	**(i)**		G'Y(t)= $0.18t + 0.72t^2 + 1.36t^3 + 1.2t^4 + 0.54t^5$ Sub t=1,(0.18+0.72+1.36+1.2+0.54)	 M1	≥3 terms correct. M1 for diffn and sub t=1	y 2 3 4 5 6 p 0.09 0.24 0.34 0.24 0.09 B1 E(Y)=0.18+….=4 B1 E(Y2)=0.36+…..=17.2 M1 Var (Y)=17.2-42=1.2 M1A1
			4 G''Y(t)=$0.18+1.44t+4.08t^2+4.8t^3+2.7t^4$ Sub t=1 and use correct formula for Var (0.18+1.44+4.08+4.8+2.7+"4"-"16") 1.2	A1 M1 M1 A1 **[5]**	≥3 terms correct ft	
	(ii)		Attempt to factorise into $(0.3t+\ldots)(0.3t+\ldots)$ $0.3t+0.4t^2+0.3t^3$	M1M1 A1 **[3]**	Attempt to find √(GY(t)) seen or implied.	Or $0.3t^3$ twice.
	(iii)		0.4	B1ft	Allow non-numerical answer, but coeff of t2 is not enough.	

Q12.

7 (i)	$G(1) = 1$	M1		
	$a = 2$	A1	**2**	
(ii)	$(1+2t)/(4-t) = c\,(1+2t)(1-\frac{1}{4}t)^{-1}$	M1		$c = \frac{1}{4}$ or 4
	$= \frac{1}{4}(1+2t)(1+\frac{1}{4}t+(\frac{1}{4}t)^2+\ldots)$	A1		
	Coefficient of $t^3 = \frac{1}{4}[(\frac{1}{4})^3 + 2(\frac{1}{4})^2]$	M1√		With 2 terms from previous line
	$= {}^{9}/_{256}$			A1 **4**
(iii)	$H(t) = \left(\frac{1+2t}{4-t}\right)^3$	B1		
	$H'(t) = 3\left(\frac{1+2t}{4-t}\right)^2\left[\frac{2(4-t)+1+2t}{(4-t)^2}\right]$	M1A1		
	$E(Y) = H'(1)$	M1		
	$=3$	A1	**5**	
(iv)	$H(1) = p_0 + p_1 + p_2 + p_3 + p_4 + \ldots = 1$			
	$H(-1) = p_0 - p_1 + p_2 - p_3 + p_4 - \ldots = -{}^{1}/_{125}$			
	Add: $2(p_0+p_2+p_4+\ldots) = 1 - {}^{1}/_{125}$	M1		With sufficient detail
	${}^{1}/_{2}(1 - {}^{1}/_{125})$ AG	A1	**2 (13)**	

Q13.

4 (i)	$G(1)=1$ $[a+b=1]$ $G'(1) = -0.7$ $[-a+2b = -0.7]$ Solve to obtain $\boldsymbol{a = 0.9}$, $\boldsymbol{b = 0.1}$	M1 M1 M1 A1 **4**	
(ii)	$G''(t)$ $[=1.8/t^3 + 0.2]$ and $G''(1) + G'(1) - [G'(1)^2]$ used Var $= 2 - 0.7 - 0.7^2$ $= \mathbf{0.81}$	M1 A1 **2**	
(iii)	$[(0.9 + 0.1t^3)/t]^{10}$ Method to obtain coefficient of t^{-7} $10\times0.9^9\times0.1$ $= \mathbf{0.387}$ to 3SF	M1 M1 A1 ft A1 **4** **{10}**	$[(a + bt^3)/t]^{10}$ For both Use of MGF. $10a^9b$

Q14.

2(i)	$G'(t) = 8te^{4t^2}/e^4$ $E(X) = G'(1)$ $= 8$	M1A1 A1 **3**	M1 for ct^2/e^4
(ii)	EITHER: $G(t) = e^{-4}(1 + 4t^2+\ldots)$ $P(X=2)$ = coefficient of $t^2 = 4e^{-4}$ or $4/e^4$ or 0.0733 OR $G''(t) = (8+64t^2)e^{4t^2-4}$ $P(X=2) = \frac{1}{2}G''(0) = 4e^{-4}$ or $4/e^4$ or 0.0733	M1A1 A1 **3** M1A1 A1 **(6)**	Expand in powers of t M1 for reasonable attempt at M"(t)

Q15.

1 (i)	$\sum_{x=0}^{n}\binom{n}{x}p^x q^{n-x}t^x$	M1	From $E(t^X)$
	$=\sum_{x=0}^{n}\binom{n}{x}(pt)^x q^{n-x}$	A1 **2**	M1A0 $\sum$ without limits $G_X(t)$=q+pt M1 then argument A0
(ii)	$G_T(t)=(q+pt)^n(q+pt)^{2n}$ $=(q+pt)^{3n}$ So $T\sim B(3n,p)$	M1 A1 M1 A1 **4** **[6]**	Multiplying pgfs For B For parameters

Q16.

2(a)	$G_Y'(t)=0.2+0.74t+0.9t^2+0.36t^3$ $G_Y''(t)=0.74+1.8t+1.08t^2$	**M1**	PGF differentiated twice.
	$G_Y'(1)=2.2$ $G_Y''(1)=3.62$	**A1**	
	$\text{Var}(Y)=G_Y''(1)+G_Y'(1)-(G_Y'(1))^2$ $3.62+2.2-2.2^2$	**M1**	Correct formula used and attempt at Var(Y).
	0.98	**A1**	
		4	
2(b)	$G_Y(t)=\left(a+bt+ct^2\right)^2$	**M1***	Correct method.
	So $a^2=0.04,\quad a=0.2$ $c^2=0.09,\quad c=0.3$ $2ab=0.2,\quad b=\frac{0.1}{0.2}=0.5$	**depM1**	Attempt to find all coefficients.
	$G_X(t)=0.2+0.5t+0.3t^2$	**A1**	*Note:* $G_Y=\frac{1}{100}(1+t)^2(2+3t)^2$ M1, accept without $\frac{1}{100}$. So $G_X=\frac{1}{10}(1+t)(2+3t)$ dep M1, must have $\frac{1}{10}$. Final answer A1.
		3	

5. Probability Generating Functions

Q17.

3(a)	$G_X(t) = \frac{6}{12} + \frac{5}{12}t + \frac{1}{12}t^2$ P(0 heads) = $(1-a)(1-b) = \frac{6}{12}$ P(1 head) = $a(1-b) + (1-a)b = \frac{5}{12}$ P(2 heads) = $ab = \frac{1}{12}$ $\left(a+b = \frac{7}{12},\ ab = \frac{1}{12}\right)$	**M1**	Obtain two correct equations and attempt to solve.
	Solve to give $a = \frac{1}{3}$ $\left(b = \frac{1}{4}\right)$	**A1**	Correct value for a.
		2	
3(b)	$G_Y(t) = \left(\frac{6}{12} + \frac{5}{12}t + \frac{1}{12}t^2\right)^2$	**M1**	Square their $G_X(t)$.
	$\frac{1}{144}\left(36 + 60t + 37t^2 + 10t^3 + t^4\right)$	**M1**	Obtain quartic polynomial from 3-term $G_X(t)$.
		A1	
		3	

3(c)	$G_Y'(t) = \frac{1}{144}\left(60 + 74t + 30t^2 + 4t^3\right)$ $G_Y''(t) = \frac{1}{144}\left(74 + 60t + 12t^2\right)$	**M1**	Differentiate their $G_Y(t)$ twice.
	$\text{Var}(Y) = G_Y''(1) + G_Y'(1) - (G_Y'(1))^2$ $\frac{1}{144}(146) + \frac{7}{6} - \frac{49}{36} = \frac{59}{72} = 0.819$	**M1**	Use correct formula and attempt at Var(Y).
	0.819	**A1**	
		3	

Q18.

4(a)	$P(0\ H)=\frac{1}{3}\times\frac{1}{3}\times\frac{1}{3}=\frac{1}{27}$ $P(1\ H)=\frac{2}{3}\times\frac{1}{3}\times\frac{1}{3}\times3=\frac{6}{27}$ $P(2\ H)=\frac{1}{3}\times\frac{2}{3}\times\frac{2}{3}\times3=\frac{12}{27}$ $P(3\ H)=\frac{2}{3}\times\frac{2}{3}\times\frac{2}{3}=\frac{8}{27}$	**M1**	At least 2 probabilities correct.
	$G_X(t)=\frac{1}{27}+\frac{6}{27}t+\frac{12}{27}t^2+\frac{8}{27}t^3$	**M1A1**	Correct form, ft their probabilities.
	OR: PGF for one coin is $\frac{1}{3}+\frac{2}{3}t$	**M1**	
	For 3 coins, PGF is $\left(\frac{1}{3}+\frac{2}{3}t\right)^3$	**M1A1**	
		3	
4(b)	$G_Z(t)=\frac{1}{108}\left(1+2t+t^2\right)\left(1+6t+12t^2+8t^3\right)$	**M1**	Correct method.
	$\frac{1}{108}\left(1+8t+25t^2+38t^3+28t^4+8t^5\right)$	**M1**	Obtain quintic polynomial.
	$\frac{1}{108}+\frac{2}{27}t+\frac{25}{108}t^2+\frac{19}{54}t^3+\frac{7}{27}t^4+\frac{2}{27}t^5$	**A1**	
		3	

4(c)	$G_Z'(t)=\frac{1}{108}\left(8+50t+114t^2+112t^3+40t^4\right)$ $E(Z)=\frac{1}{108}\left(8+50+114+112+40\right)$	**M1**	Differentiate and evaluate at $t=1$
	$\frac{324}{108}=3$	**A1**	
	Alternative method for questions 4(c)		
	$\frac{2}{3}+\frac{2}{3}+\frac{2}{3}+\frac{1}{2}+\frac{1}{2}$	**M1**	Summing expected values for each coin
	3	**A1**	
		2	

Q19.

5(a)	$G_X(t)=\frac{9}{16}+\frac{6}{16}t+\frac{1}{16}t^2$	**M1 A1**	2 probabilities correct, in a polynomial
		2	

5(b)	$G_Z(t) = \left(\frac{9}{16} + \frac{6}{16}t + \frac{1}{16}t^2\right)\left(\frac{25}{36} + \frac{10}{36}t + \frac{1}{36}t^2\right)$	**M1**	Second PGF correct and multiplied by part (a).
	$\frac{1}{576}\left(225 + 240t + 94t^2 + 16t^3 + t^4\right)$ Or $\frac{25}{64} + \frac{5}{12}t + \frac{47}{288}t^2 + \frac{1}{36}t^3 + \frac{1}{576}t^4$	**M1** **A1**	Obtains a quartic polynomial. Note: $\frac{1}{576}(t+3)^2(t+5)^2$ scores M1M1A0.
		3	
5(c)	$G_Z'(t) = \frac{1}{576}\left(240 + 188t + 48t^2 + 4t^3\right)$ $G_Z''(t) = \frac{1}{576}\left(188 + 96t + 12t^3\right)$	**M1**	Differentiate twice.
	$\text{Var}(Z) = \frac{1}{576}(188 + 96 + 12) + \frac{5}{6} - \frac{25}{36}$	**M1**	Use correct formula.
	$\frac{47}{72}$	**A1**	
	OR: $\text{Var}(Z) = 2\times\frac{1}{4}\times\frac{3}{4} + 2\times\frac{1}{6}\times\frac{5}{6}$	**M1** **M1**	One term correct. Two terms present and added.
	$\frac{47}{72}$	**A1**	
		3	
5(d)	1	**B1 FT**	FT their power with greatest coefficient in part (b).
		1	