Statistics Year 2 exam questions Edexcel

NOTE:
Please be aware that in the Year 13 collections, you will miss questions from the Year 13 papers. However, these questions are intentionally included because they align with Year 12 content and topics.

Arancha Ruiz

Ch. 1: Regression, Correlation and Hypothesis Testing

June 2024 Question 2

2. Amar is studying the flight of a bird from its nest.

He measures the bird's height above the ground, h metres, at time t seconds for 10 values of t
Amar finds the equation of the regression line for the data to be $h = 38.6 - 1.28t$

(a) Interpret the gradient of this line.

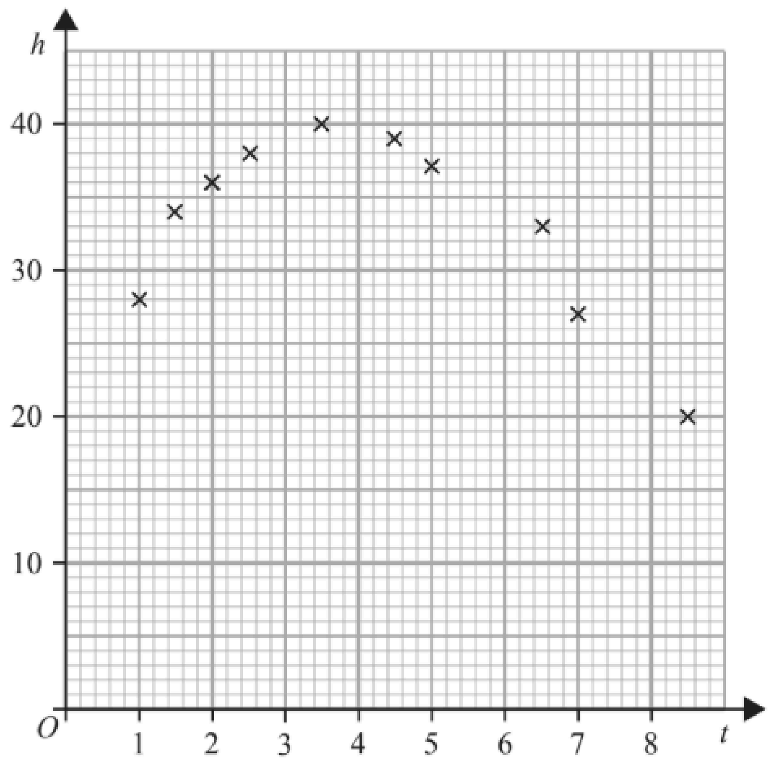
(1)

The product moment correlation coefficient between h and t is -0.510

(b) Test whether or not there is evidence of a negative correlation between the height above the ground and the time during the flight.
You should
- state your hypotheses clearly
- use a 5% level of significance
- state the critical value used

(3)

Jane draws the following scatter diagram for Amar's data.



(c) With reference to the scatter diagram, state, giving a reason, whether or not the regression line $h = 38.6 - 1.28t$ is an appropriate model for these data.

(1)

Jane suggests an improved model using the variable $u = (t - k)^2$ where k is a constant.

She obtains the equation $h = 38.1 - 0.78u$

(d) Choose a suitable value for k to write Jane's improved model for h in terms of t only.

(1)

ANSWER

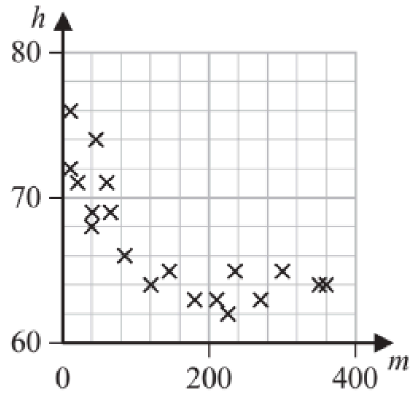
Qu 2	Scheme	Marks	AO
(a)	e.g. The height (h) decreases by about 1.28 m for each second of the flight	B1 **(1)**	3.4
(b)	$H_0 : \rho = 0 \quad H_1 : \rho < 0$	B1	2.5
	[5% 1-tail cv =] ($\pm$) 0.5494	M1	1.1a
	[$r = -0.510$ not sig] there is insufficient (o.e.) evidence of a negative correlation between height (or h) and time (or t)	A1 **(3)**	2.2b
(c)	No – since points seem to follow a curve/quadratic (rather than a line) or since points are "non-linear" but regression line/ model is linear or e.g. between ($t = 5$ and 7)height drops by much more than 2.56 m or e.g. gradient is positive up to $t = 3.5$ (line gradient < 0) or e.g. gradient is positive initially (line gradient < 0) or e.g. gradient is positive and then negative	B1 **(1)**	2.4
(d)	[$h = 38.1 - 0.78\,(t - k)^2$ with] a suitable k i.e. in the range 3~4.5	B1 **(1)**	3.3
		(6 marks)	

June 2022 Question 6

6. Anna is investigating the relationship between exercise and resting heart rate. She takes a random sample of 19 people in her year at school and records for each person

- their resting heart rate, h beats per minute
- the number of minutes, m, spent exercising each week

Her results are shown on the scatter diagram.



(a) Interpret the nature of the relationship between h and m

(1)

Anna codes the data using the formulae

$$x = \log_{10} m$$
$$y = \log_{10} h$$

The product moment correlation coefficient between x and y is -0.897

(b) Test whether or not there is significant evidence of a negative correlation between x and y
You should

- state your hypotheses clearly
- use a 5% level of significance
- state the critical value used

(3)

The equation of the line of best fit of y on x is

$$y = -0.05x + 1.92$$

(c) Use the equation of the line of best fit of y on x to find a model for h on m in the form

$$h = am^k$$

where a and k are constants to be found.

(5)

ANSWER

Question	Scheme		Marks	AOs
6(a)	eg As the number of minutes <u>exercise</u> (m) increases the resting <u>heart rate</u> (h) decreases **or** the gradient of the curve is becoming flatter with increasing m: diminishing effect of each <u>additional minute</u> of <u>exercise</u>		B1	2.4
			(1)	
(b)	$\mathrm{H}_0: \rho = 0 \;\; \mathrm{H}_1: \rho < 0$		B1	2.5
	Critical value – 0.3887 (Allow $\pm$)		M1	1.1b
	There is evidence that the product moment **<u>correlation</u>** is **<u>less than 0</u>**/ **<u>there is a negative correlation</u>**		A1	2.2b
			(3)	
(c)	$\log_{10} h = -0.05\log_{10} m + 1.92$	$h = am^k \rightarrow \log_{10} h = \log_{10} am^k$	M1	1.1b
	$\log_{10} h = -\log_{10} m^{0.05} + 1.92$ or $\log_{10} h = \log_{10} m^{-0.05} + 1.92$ or $h = 10^{1.92 - 0.05\log_{10} m}$ oe	$\log_{10} h = \log_{10} a + \log_{10} m^k$ or $\log_{10} a = 1.92$	M1	2.1
	$\log_{10} hm^{0.05} = 1.92$ or $\log_{10}\left(\dfrac{h}{m^{-0.05}}\right) = 1.92$ or $h = 10^{1.92} \times 10^{-0.05\log_{10} m}$ oe	$\log_{10} h = \log_{10} a + k\log_{10} m$	M1	1.1b
	$hm^{0.05} = 10^{1.92}$ or $\dfrac{h}{m^{-0.05}} = 10^{1.92}$ or $h = 10^{1.92} \times 10^{\log_{10} m^{-0.05}}$	$\log_{10} a = 1.92$ and $k = -0.05$	M1	1.1b
	$h = 10^{1.92} m^{-0.05}$ or $h = 83.17\ldots m^{-0.05}$ or $a =$ awrt 83.17 **and** $k = -0.05$		A1	1.1b
			(5)	
	Notes:			**(9 marks)**

Video solution:

https://youtu.be/4rFzrJdw2tQ

November 2021 question 2

2. Marc took a random sample of 16 students from a school and for each student recorded

- the number of letters, x, in their last name
- the number of letters, y, in their first name

His results are shown in the scatter diagram on the next page.

(a) Describe the correlation between x and y.

(1)

Marc suggests that parents with long last names tend to give their children shorter first names.

(b) Using the scatter diagram comment on Marc's suggestion, giving a reason for your answer.

(1)

The results from Marc's random sample of 16 observations are given in the table below.

x	3	6	8	7	5	3	11	3	4	5	4	9	7	10	6	6
y	7	7	4	4	6	8	5	5	8	4	7	4	5	5	6	3

(c) Use your calculator to find the product moment correlation coefficient between x and y for these data.

(1)

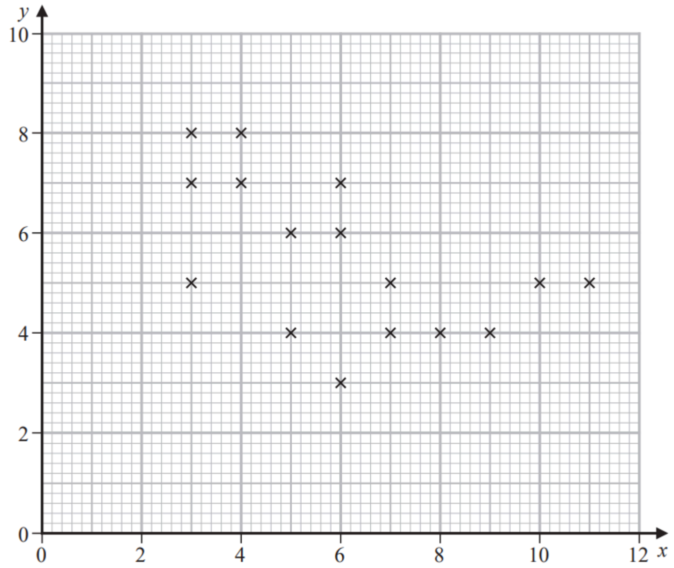
(d) Test whether or not there is evidence of a negative correlation between the number of letters in the last name and the number of letters in the first name.

You should

- state your hypotheses clearly
- use a 5% level of significance

(3)

Continues in next page



ANSWER

Qu 2	Scheme	Marks	AO
(a)	Negative	B1	1.2
		(1)	
(b)	Marc's suggestion is compatible because it's negative correlation	B1	2.4
		(1)	
(c)	$(r =)\ -0.54458266\ldots$ awrt $\mathbf{-0.545}$	B1	1.1b
		(1)	
(d)	$H_0 : \rho = 0 \qquad H_1 : \rho < 0$	B1	2.5
	[5% 1-tail cv =] $(\pm)$ 0.4259	M1	1.1a
	(significant result / reject H_0)		
	There is evidence of negative correlation between the number of letters in (or length of) a student's last name and their first name	A1	2.2b
		(3)	
		(6 marks)	

October 2020 question 2

2. A random sample of 15 days is taken from the large data set for Perth in June and July 1987.

The scatter diagram in Figure 1 displays the values of two of the variables for these 15 days.

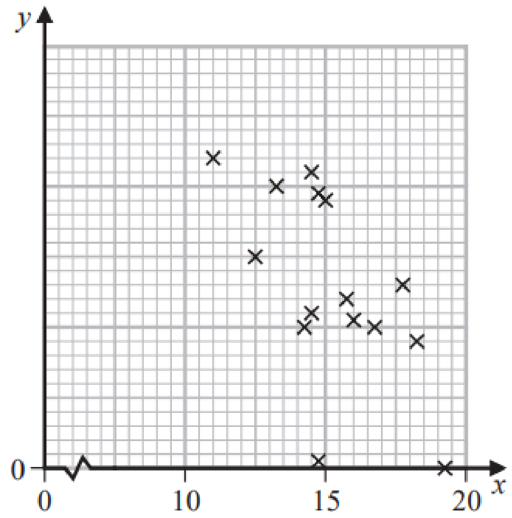


Figure 1

(a) Describe the correlation.

(1)

The variable on the x-axis is Daily Mean Temperature measured in °C.

(b) Using your knowledge of the large data set,

(i) suggest which variable is on the y-axis,

(ii) state the units that are used in the large data set for this variable.

(2)

Stav believes that there is a correlation between Daily Total Sunshine and Daily Maximum Relative Humidity at Heathrow.

He calculates the product moment correlation coefficient between these two variables for a random sample of 30 days and obtains $r = -0.377$

(c) Carry out a suitable test to investigate Stav's belief at a 5% level of significance. State clearly

- your hypotheses
- your critical value

(3)

On a random day at Heathrow the Daily Maximum Relative Humidity was 97%

(d) Comment on the number of hours of sunshine you would expect on that day, giving a reason for your answer.

(1)

ANSWER

Qu 2	Scheme	Marks	AO
(a)	Negative	B1	1.2
		(1)	
(b)(i)	Rainfall | <u>or</u> | Pressure	B1	2.2b
(ii)	mm | | hPa or Pascals or hectopascals or mb or millibars	B1ft	1.1b
		(2)	
(c)	$\mathrm{H}_0 : \rho = 0 \qquad \mathrm{H}_1 : \rho \neq 0$	B1	2.5
	Critical value: $-0.361(0)$	M1	1.1b
	$r < -0.3610$ so significant result and there is evidence of a correlation between Daily Total <u>Sunshine</u> and Daily Maximum Relative <u>Humidity</u>	A1	2.2b
		(3)	
(d)	Humidity is high and there is evidence of correlation and $r < 0$ So expect amount of sunshine to be <u>lower</u> than the <u>average</u> for Heathrow(oe)	B1	2.2b
		(1)	
		(7 marks)	

June 2019 question 3

3. Barbara is investigating the relationship between average income (GDP per capita), x US dollars, and average annual carbon dioxide (CO_2) emissions, y tonnes, for different countries.

She takes a random sample of 24 countries and finds the product moment correlation coefficient between average annual CO_2 emissions and average income to be 0.446

(a) Stating your hypotheses clearly, test, at the 5% level of significance, whether or not the product moment correlation coefficient for all countries is greater than zero.

(3)

Barbara believes that a non-linear model would be a better fit to the data.
She codes the data using the coding $m = \log_{10} x$ and $c = \log_{10} y$ and obtains the model $c = -1.82 + 0.89m$

The product moment correlation coefficient between c and m is found to be 0.882

(b) Explain how this value supports Barbara's belief.

(1)

(c) Show that the relationship between y and x can be written in the form $y = ax^n$ where a and n are constants to be found.

(5)

ANSWER

Question	Scheme		Marks	AOs
3(a)	$H_0 : \rho = 0$ $H_1 : \rho > 0$		B1	2.5
	Critical value 0.3438		M1	1.1a
	(0.446 > 0.3438) so there is evidence that the product moment correlation coefficient (pmcc) is greater than 0/there is positive correlation		A1	2.2b
			(3)	
(b)	The value is close(r) to 1 **or** there is strong(er) (positive) correlation		B1	2.4
			(1)	
(c)	$\log_{10} y = -1.82 + 0.89(\log_{10} x)$	$y = ax^n \rightarrow$ $\log_{10} y = \log_{10}(ax^n)$	M1	1.1b
	$y = 10^{-1.82+0.89(\log_{10} x)}$	$\log_{10} y = \log_{10} a + \log_{10} x^n$	M1	2.1
	$y = 10^{-1.82} \times 10^{0.89(\log_{10} x)}$ $[= 10^{-1.82} \times 10^{(\log_{10} x)^{0.89}}]$	$\log_{10} y = \log_{10} a + n\log_{10} x$ $[\log_{10} a = -1.82, n = 0.89]$	M1	1.1b
	$y = 0.015x^{0.89}$	$y = 0.015x^{0.89}$	A1A1	1.1b 1.1b
			(5)	
				(9 marks)

June 2018 Question 2

2. Tessa owns a small clothes shop in a seaside town. She records the weekly sales figures, £w, and the average weekly temperature, t°C, for 8 weeks during the summer.
The product moment correlation coefficient for these data is −0.915

(a) Stating your hypotheses clearly and using a 5% level of significance, test whether or not the correlation between sales figures and average weekly temperature is negative.
(3)

(b) Suggest a possible reason for this correlation.
(1)

Tessa suggests that a linear regression model could be used to model these data.

(c) State, giving a reason, whether or not the correlation coefficient is consistent with Tessa's suggestion.
(1)

(d) State, giving a reason, which variable would be the explanatory variable.
(1)

Tessa calculated the linear regression equation as $w = 10\,755 - 171t$

(e) Give an interpretation of the gradient of this regression equation.
(1)

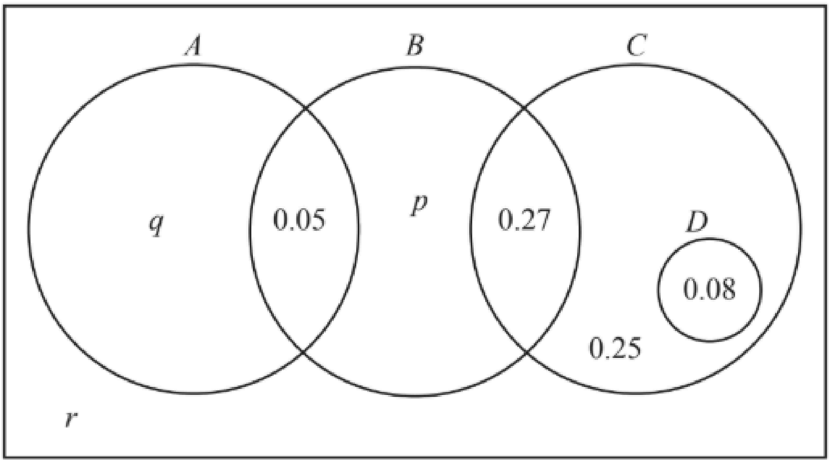
ANSWER

Qu 2	Scheme	Marks	AO
(a)	$H_0: \rho = 0 \quad H_1: \rho < 0$	B1	2.5
	Critical value: -0.6215 (Allow any cv in range $0.5 < \lvert cv \rvert < 0.75$)	M1	1.1a
	$r < -0.6215$ so significant result and there is evidence of a negative correlation between w and t	A1	2.2b
		(3)	
(b)	e.g. As temperature increases people spend more time on the beach and less time shopping (o.e.)	B1	2.4
		(1)	
(c)	Since r is close to -1 , it **is** consistent with the suggestion	B1	2.4
		(1)	
(d)	t will be the explanatory variable since sales are likely to depend on the temperature	B1	2.4
		(1)	
(e)	Every degree rise in temperature leads to a drop in weekly earnings of £171	B1	3.4
		(1)	
		(7 marks)	

Ch. 2: Conditional Probability

June 2024 Question 6

6. The Venn diagram, where p, q and r are probabilities, shows the events A, B, C and D and associated probabilities.



(a) State any pair of mutually exclusive events from A, B, C and D **(1)**

The events B and C are independent.

(b) Find the value of p **(2)**

(c) Find the greatest possible value of $\mathrm{P}(A \mid B')$ **(3)**

Given that $\mathrm{P}(B \mid A') = 0.5$

(d) find the value of q and the value of r **(3)**

(e) Find $\mathrm{P}\left(\left[A \cup B\right]' \cap C\right)$ **(1)**

(f) Use set notation to write an expression for the event with probability p **(1)**

ANSWER

Qu 6	Scheme	Marks	AO
(a)	A, C or A, D or B, D [Allow things like $A \cap D$]	B1	1.2
		(1)	
(b)	$\text{P}(C) = 0.6$ and $\text{P}(B) = p + 0.32$ and $\text{P}(B \cap C) = 0.27$ or $(0.08 + 0.25 + 0.27) \times (0.27 + 0.05 + p) = 0.27$ or $0.27 + 0.05 + p = \frac{0.27}{0.6} = 0.45$	M1	1.1b
	$[p + 0.32 = 0.45 \text{ so}]$ $p = \underline{\mathbf{0.13}}$	A1	2.2a
		(2)	
(c)	$[\text{P}(A \mid B')] = \dfrac{q}{q + r + 0.25 + 0.08}$ or $\dfrac{q}{1 - (0.05 + \text{"}0.13\text{"} + 0.27)}$ or $\dfrac{q}{0.55}$	M1	2.1
	$q + r = 1 - 0.65 -$ "0.13" [= 0.22]	M1	1.1b
	Since $r \geqslant 0$ the greatest value of q is "0.22" so $\text{P}(A \mid B') \leqslant \underline{\mathbf{0.4}}$ or $\underline{\frac{2}{5}}$	A1	2.2a
		(3)	
(d)	$[\text{P}(B \mid A') =] \dfrac{0.27 + \text{"}0.13\text{"}}{0.6 + \text{"}0.13\text{"} + r} = 0.5$ or $\dfrac{0.27 + \text{"}0.13\text{"}}{1 - (q + 0.05)} = 0.5$	M1	1.1b
	$r = \underline{\mathbf{0.07}}$,	A1	1.1b
	$q = \underline{\mathbf{0.15}}$	A1ft	1.1b
		(3)	
(e)	$\left[\text{P}\left([A \cup B]' \cap C\right) = \right] [0.25 + 0.08] = \underline{\mathbf{0.33}}$	B1	1.1b
		(1)	
(f)	e.g. $B \cap [A \cup C]'$ or $B \cap A' \cap C'$ or $(B \cap A') \cap (B \cap C')$ o.e.	B1	1.1b
		(1)	
		(11 marks)	

June 2023 Question 5

5. Tisam is playing a game.
She uses a ball, a cup and a spinner.

The random variable X represents the number the spinner lands on when it is spun.
The probability distribution of X is given in the following table

x	20	50	80	100
$P(X = x)$	a	b	c	d

where a, b, c and d are probabilities.

To play the game

- the spinner is spun to obtain a value of x
- Tisam then stands x cm from the cup and tries to throw the ball into the cup

The event S represents the event that Tisam successfully throws the ball into the cup.

To model this game Tisam assumes that

- $P(S \mid \{X = x\}) = \dfrac{k}{x}$ where k is a constant
- $P(S \cap \{X = x\})$ should be the same whatever value of x is obtained from the spinner

Using Tisam's model,

(a) show that $c = \dfrac{8}{5}b$

(2)

(b) find the probability distribution of X

(5)

Nav tries, a large number of times, to throw the ball into the cup from a distance of 100 cm.
He successfully gets the ball in the cup 30% of the time.

(c) State, giving a reason, why Tisam's model of this game is not suitable to describe Nav playing the game for all values of X

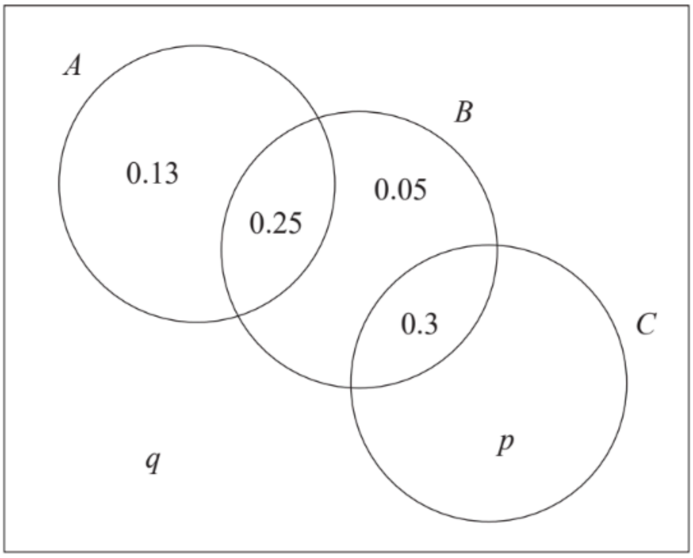
(1)

ANSWER

Qu 5	Scheme	Marks	AO
(a)	$P(S \cap \{X=50\}) = P(S \cap \{X=80\})[= \text{a constant, } V] \Rightarrow b \times \frac{k}{50} = c \times \frac{k}{80}$ May see: $\frac{k}{50} = \frac{V}{b}$ <u>and</u> $\frac{k}{80} = \frac{V}{c}$ (condone any <u>letter</u> for V even S)	M1	3.1a
	So $\underline{c = \frac{8}{5}b}$ *	A1cso*	1.1b
		(2)	
(b)	$d = 2b$ **or** $a = \frac{2}{5}b$ **or** $c = 4a$ **or** $d = 5a$ **or** $d = \frac{5}{4}c$	M1 A1	2.1 3.3
	$\frac{2}{5}b + b + \frac{8}{5}b + 2b = 1$	M1	2.1
	$\Rightarrow 5b = 1$ so $b = \frac{1}{5}$ (o.e.)	A1	1.1b
	$\underline{\underline{a = \frac{2}{25} \quad b = \frac{1}{5} \quad c = \frac{8}{25} \quad d = \frac{2}{5}}}$	A1	3.2a
		(5)	
(c)	[Experiment suggests for Nav] $P(S \mid \{X=100\}) = 0.3 \Rightarrow k = 30$ **or** $0.3 = \frac{V}{0.4} \Rightarrow V = 0.12$ So model won't work since $P(S \mid X=20) = \frac{30}{20}$ <u>or</u> $\frac{0.12}{0.08}$ and so would be greater than 1	B1	2.4
		(1)	
		(8 marks)	

June 2023 Question 1

1. The Venn diagram, where p and q are probabilities, shows the three events A, B and C and their associated probabilities.



(a) Find $P(A)$

(1)

The events B and C are independent.

(b) Find the value of p and the value of q

(3)

(c) Find $P(A|B')$

(2)

ANSWER

Qu 1	Scheme	Marks	AO
(a)	[0.13 + 0.25 =] $\underline{\mathbf{0.38}}$	B1	1.1b
		(1)	
(b)	Independence implies:		
	e.g. $\left[\mathrm{P}(B\cap C)=\mathrm{P}(B)\times\mathrm{P}(C)\quad\Rightarrow\right]\quad 0.3=(0.3+0.05+0.25)\times(0.3+p)$	M1	1.1b
	So $p=\underline{\mathbf{0.2}}$	A1	1.1b
	[Sum of probabilities = 1 gives] $q=\underline{\mathbf{0.07}}$	B1ft	1.1b
		(3)	
(c)	$[\mathrm{P}(A\mid B')=]\dfrac{\mathrm{P}(A\cap B')}{\mathrm{P}(B')}$ or $\dfrac{0.13}{(1-0.6)\text{ or }(0.13+"0.2"+"0.07")}$	M1	1.1b
	$=\dfrac{13}{40}$ or $\underline{\mathbf{0.325}}$	A1	1.1b
		(2)	
		(6 marks)	

June 2022 Question 5

5. A company has 1825 employees.
The employees are classified as professional, skilled or elementary.

The following table shows

- the number of employees in each classification
- the two areas, A or B, where the employees live

	A	B
Professional	740	380
Skilled	275	90
Elementary	260	80

An employee is chosen at random.

Find the probability that this employee

(a) is skilled,

(1)

(b) lives in area B and is not a professional.

(1)

Some classifications of employees are more likely to work from home.

- 65% of professional employees in both area A and area B work from home
- 40% of skilled employees in both area A and area B work from home
- 5% of elementary employees in both area A and area B work from home
- Event F is that the employee is a professional
- Event H is that the employee works from home
- Event R is that the employee is from area A

(c) Using this information, complete the Venn diagram on the opposite page.

(4)

(d) Find $P(R' \cap F)$

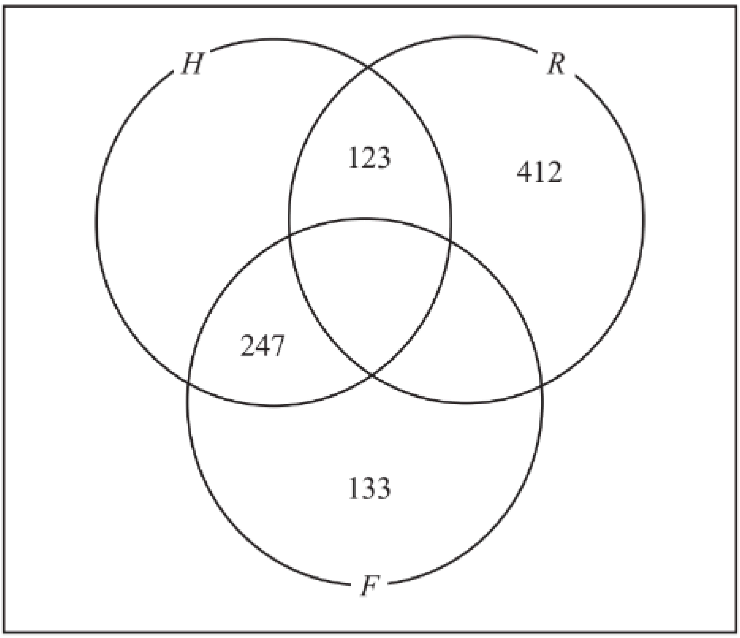
(1)

(e) Find $P([H \cup R]')$

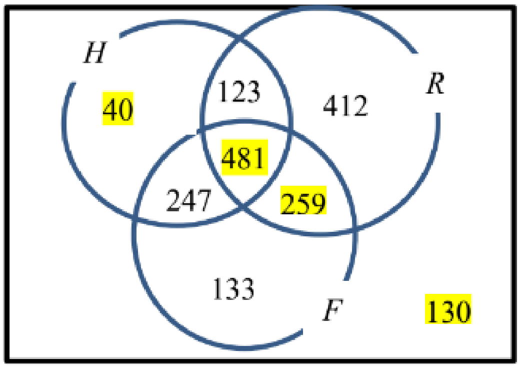
(1)

(f) Find $P(F \mid H)$

(2)



ANSWER

Question	Scheme	Marks	AOs
5(a)	$\dfrac{365}{1825}$ or $\dfrac{1}{5}$ or 0.2 oe	B1	1.1b
		(1)	
(b)	$\dfrac{170}{1825}$ or $\dfrac{34}{365}$ or awrt 0.093	B1	1.1b
		(1)	
(c)	$90\times0.4+80\times0.05[=40]$ or $90\times0.6+80\times0.95[=130]$ or $740\times0.65[=481]$ or $740\times0.35[=259]$ 	M1 B1 B1 A1	3.1b 1.1b 1.1b 1.1b
		(4)	
(d)	$\mathrm{P}(R'\cap F)=\dfrac{380}{1825}\left[=\dfrac{76}{365}=0.208...\right]$ oe awrt 0.208	B1	1.1b
		(1)	
(e)	$\left[\dfrac{133\ +\ "130"}{1825}=\right]\dfrac{"263"}{1825}$ awrt 0.144	B1ft	1.1b
		(1)	
(f)	$\dfrac{247+"481"}{247+"481"+123+"40"}$	M1	3.4
	$=\dfrac{728}{891}$ awrt 0.817	A1	1.1b
		(2)	
	Notes:	**(10 marks)**	

November 2021 question 1

1. (a) State one disadvantage of using quota sampling compared with simple random sampling.

(1)

In a university 8% of students are members of the university dance club.

A random sample of 36 students is taken from the university.

The random variable X represents the number of these students who are members of the dance club.

(b) Using a suitable model for X, find

(i) $P(X = 4)$

(ii) $P(X \geqslant 7)$

(3)

Only 40% of the university dance club members can dance the tango.

(c) Find the probability that a student is a member of the university dance club and can dance the tango.

(1)

A random sample of 50 students is taken from the university.

(d) Find the probability that fewer than 3 of these students are members of the university dance club and can dance the tango.

(2)

ANSWER

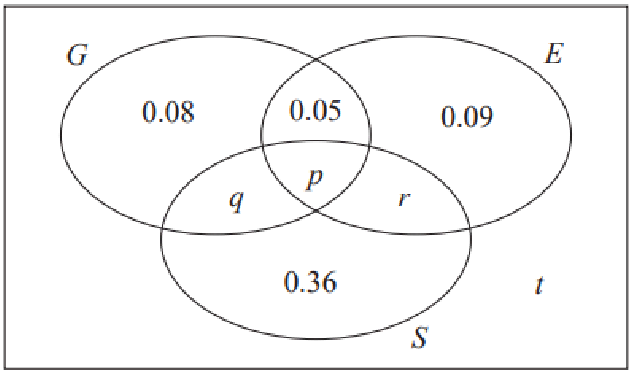
Qu 1	Scheme	Marks	AO
(a)	**Disadvantage:** e.g. Not random; cannot use (reliably) for inferences	B1	1.1b
		(1)	
(b)	[Sight or correct use of] $X \sim \mathrm{B}(36,\ 0.08)$	M1	3.3
(i)	$\mathrm{P}(X = 4) = 0.167387\ldots$ awrt $\underline{\mathbf{0.167}}$	A1	1.1b
(ii)	$[\mathrm{P}(X \geqslant 7) = 1 - \mathrm{P}(X \leqslant 6) =]$ $0.022233\ldots$ awrt $\underline{\mathbf{0.0222}}$	A1	1.1b
		(3)	
(c)	P(In dance club and dance tango) $= 0.4 \times 0.08 = \underline{\mathbf{0.032}}$ or $\underline{\underline{\dfrac{4}{125}}}$ or $\underline{\mathbf{3.2\%}}$	B1	1.1b
		(1)	
(d)	[Let T = those who can dance the Tango. Sight or use of] $T \sim \mathrm{B}(50,\ \text{“}0.032\text{”})$	M1	3.3
	$[\mathrm{P}(T < 3) = \mathrm{P}(T \leqslant 2) =]$ $0.7850815\ldots$ awrt $\underline{\mathbf{0.785}}$	A1	1.1b
		(2)	
		(7 marks)	

November 2021 question 4

4. A large college produces three magazines.
One magazine is about green issues, one is about equality and one is about sports.
A student at the college is selected at random and the events G, E and S are defined as follows

G is the event that the student reads the magazine about green issues
E is the event that the student reads the magazine about equality
S is the event that the student reads the magazine about sports

The Venn diagram, where p, q, r and t are probabilities, gives the probability for each subset.



(a) Find the proportion of students in the college who read exactly one of these magazines.
(1)

No students read all three magazines and $P(G) = 0.25$

(b) Find

(i) the value of p

(ii) the value of q
(3)

Given that $P(S \mid E) = \frac{5}{12}$

(c) find

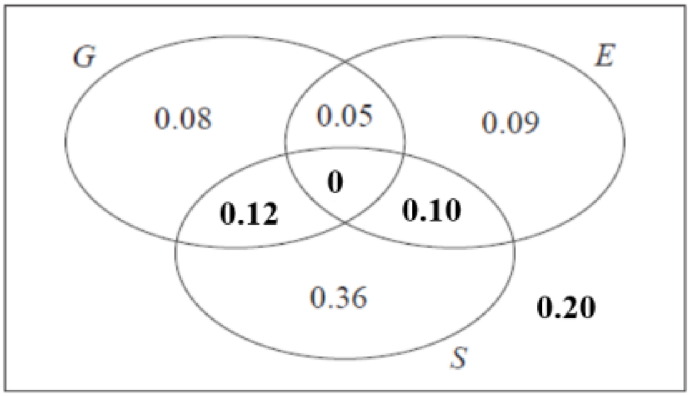
(i) the value of r

(ii) the value of t
(4)

(d) Determine whether or not the events $(S \cap E')$ and G are independent.
Show your working clearly.
(3)

ANSWER

Qu 4	Scheme	Marks	AO
(a)	$0.08+0.09+0.36=\underline{\mathbf{0.53}}$	B1	1.1b
		(1)	
(b)(i)	$\left[\mathrm{P}(G\cap E\cap S)=0\quad\Rightarrow\right]\quad \underline{p=0}$	B1	1.1b
(ii)	$[\mathrm{P}(G)=0.25\ \Rightarrow]\ 0.08+0.05+q+"p"=0.25$	M1	1.1b
	$\underline{q=0.12}$	A1	1.1b
		(3)	
(c)(i)	$\left[\mathrm{P}(S\mid E)=\frac{5}{12}\quad\Rightarrow\right]\frac{r+"p"}{r+"p"+0.09+0.05}=\frac{5}{12}$	M1 A1ft	3.1a 1.1b
	$[12r=5r+5\times0.14\quad\Rightarrow]\quad\underline{r=0.10}$	A1	1.1b
(ii)	$[0.08+0.05+"0.12"+"0"+0.09+"0.10"+0.36+t=1\quad\Rightarrow]\quad\underline{t=0.20}$	B1ft	1.1b
		(4)	
(d)	$\mathrm{P}(S\cap E')=0.36+\text{“}q\text{”}\ [=0.48]$	B1ft	1.1b
	$\mathrm{P}\left(\left[(S\cap E')\right]\cap G\right)="q"[=0.12]$ **and** $\mathrm{P}(G)=0.25$ **and** $\mathrm{P}(S\cap E')\times\mathrm{P}(G)="0.48"\times\frac{1}{4}$ or 0.12	M1	2.1
	$\mathrm{P}(S\cap E')\times\mathrm{P}(G)=0.12=\mathrm{P}\left(\left[(S\cap E')\right]\cap G\right)$ so **are independent**	A1	2.2a
		(3)	
		(11 marks)	



November 2021 question 6

6. The discrete random variable X has the following probability distribution

x	a	b	c
$P(X = x)$	$\log_{36} a$	$\log_{36} b$	$\log_{36} c$

where

- a, b and c are distinct integers $(a < b < c)$
- all the probabilities are greater than zero

(a) Find

(i) the value of a

(ii) the value of b

(iii) the value of c

Show your working clearly.

(5)

The independent random variables X_1 and X_2 each have the same distribution as X

(b) Find $P(X_1 = X_2)$

(2)

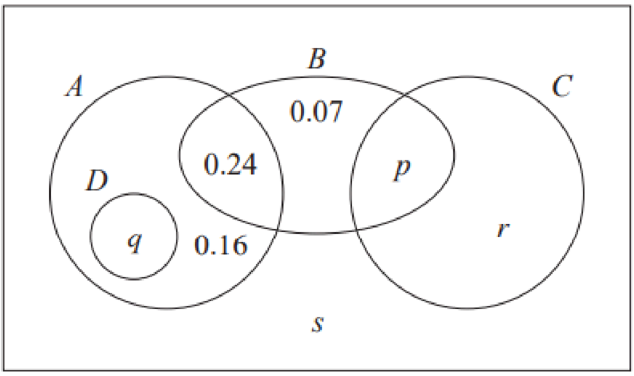
ANSWER

Qu 6	Scheme	Marks	AO
(a)	[Sum of probs = 1 implies] $\log_{36} a + \log_{36} b + \log_{36} c = 1$	M1	3.1a
	$\Rightarrow \log_{36}(abc) = 1$ so $abc = 36$	A1	3.4
	All probabilities greater than 0 implies each of a, b and c > 1	B1	2.2a
	$36 = 2^2 \times 3^2$ (or 3 numbers that multiply to give 36 e.g. 2, 2, 9 etc)	dM1	2.1
	Since a, b and c are distinct must be **2, 3, 6** **($a = 2, b = 3, c = 6$)**	A1	3.2a
		(5)	
(b)	$(\log_{36} a)^2 + (\log_{36} b)^2 + (\log_{36} c)^2$	M1	3.4
	[= 0.0374137…+ 0.09398737…+0.25]		
	$= 0.38140\ldots$ awrt **0.381**	A1	1.1b
		(2)	
		(7 marks)	

https://youtu.be/VHMoGF_qVCI

October 2020 question 1

1. The Venn diagram shows the probabilities associated with four events, A, B, C and D



(a) Write down any pair of mutually exclusive events from A, B, C and D

(1)

Given that $P(B) = 0.4$

(b) find the value of p

(1)

Given also that A and B are independent

(c) find the value of q

(2)

Given further that $P(B' \mid C) = 0.64$

(d) find

(i) the value of r

(ii) the value of s

(4)

ANSWER

Qu 1	Scheme	Marks	AO
(a)	A, C or D, B or D, C	B1 **(1)**	1.2
(b)	[$p = 0.4 - 0.07 - 0.24 =$] $\underline{\mathbf{0.09}}$	B1 **(1)**	1.1b
(c)	A and B independent implies		1.1b
	$\mathrm{P}(A)\times 0.4 = 0.24$ or $(q+0.16+0.24)\times 0.4 = 0.24$	M1	
	so $\mathrm{P}(A) = 0.6$ and $q = \underline{\mathbf{0.20}}$	A1cso **(2)**	1.1b
(d)(i)	$\mathrm{P}(B' \mid C) = 0.64$ gives $\dfrac{r}{r+p} = 0.64$ or $\dfrac{r}{r+"0.09"} = 0.64$	M1	3.1a
	$r = 0.64r + 0.64$ "p" so $0.36r = 0.0576$ so $r = \underline{\mathbf{0.16}}$	A1	1.1b
(ii)	Using sum of probabilities = 1 e.g. "0.6" + 0.07 + "0.25" + s =1	M1	1.1b
	so $s = \underline{\mathbf{0.08}}$	A1 **(4)**	1.1b
		(8 marks)	

October 2020 question 4

4. The discrete random variable D has the following probability distribution

d	10	20	30	40	50
$\mathrm{P}(D=d)$	$\frac{k}{10}$	$\frac{k}{20}$	$\frac{k}{30}$	$\frac{k}{40}$	$\frac{k}{50}$

where k is a constant.

(a) Show that the value of k is $\frac{600}{137}$

(2)

The random variables D_1 and D_2 are independent and each have the same distribution as D.

(b) Find $\mathrm{P}(D_1 + D_2 = 80)$
Give your answer to 3 significant figures.

(3)

A single observation of D is made.

The value obtained, d, is the common difference of an arithmetic sequence.

The first 4 terms of this arithmetic sequence are the angles, measured in degrees, of quadrilateral Q

(c) Find the exact probability that the smallest angle of Q is more than 50°

(5)

ANSWER

Qu 4	Scheme	Marks	AO
(a)	$\frac{k}{10}+\frac{k}{20}+\frac{k}{30}+\frac{k}{40}+\frac{k}{50}=1$ or $\frac{1}{600}(60k+30k+20k+15k+12k)=1$	M1	1.1b
	So $k=\frac{600}{137}$ (*)	A1cso	1.1b
		(2)	
(b)	(Cases are:) $D_1=30, D_2=50$ and $D_1=50, D_2=30$ and $D_1=40, D_2=40$	M1	2.1
	$\text{P}(D_1+D_2=80)=\frac{k}{50}\times\frac{k}{30}\times 2+\left(\frac{k}{40}\right)^2$	M1	3.4
	$=0.0375619\ldots$ awrt $\underline{\mathbf{0.0376}}$	A1	1.1b
		(3)	
(c)	Angles are: $a,\quad a+d,\quad a+2d,\quad a+3d$	M1	3.1a
	$S_4=a+(a+d)+(a+2d)+(a+3d)=360$	M1	2.1
	$2a+3d=180$ (o.e.)	A1	2.2a
	Smallest angle is $a>50$ consider cases: $d=10$ so $a=75$ <u>or</u> $d=20$ so $a=60$ [$d=30$ gives $a=45$ no good]	M1	3.1b
	$\text{P}(D=10\text{ or }20)=\frac{3k}{20}=\underline{\underline{\frac{90}{137}}}$	A1	1.1b
		(5)	
		(10 marks)	

June 2019 question 1

1. Three bags, A, B and C, each contain 1 red marble and some green marbles.

   Bag A contains 1 red marble and 9 green marbles only
   Bag B contains 1 red marble and 4 green marbles only
   Bag C contains 1 red marble and 2 green marbles only

   Sasha selects at random one marble from bag A.
   If he selects a red marble, he stops selecting.
   If the marble is green, he continues by selecting at random one marble from bag B.
   If he selects a red marble, he stops selecting.
   If the marble is green, he continues by selecting at random one marble from bag C.

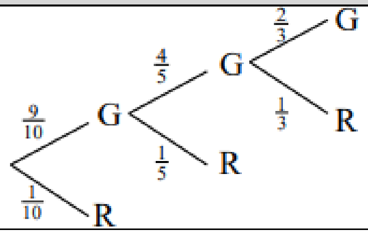
   (a) Draw a tree diagram to represent this information.
   (2)

   (b) Find the probability that Sasha selects 3 green marbles.
   (2)

   (c) Find the probability that Sasha selects at least 1 marble of each colour.
   (2)

   (d) Given that Sasha selects a red marble, find the probability that he selects it from bag B.
   (2)

ANSWER

Question	Scheme	Marks	AOs
1(a)		B1 dB1	1.1b 1.1b
		(2)	
(b)	$\frac{9}{10}\times\frac{4}{5}\times\frac{2}{3}$	M1	1.1b
	$=\frac{12}{25}(=0.48)$	A1	1.1b
		(2)	
(c)	$\frac{9}{10}\times\frac{1}{5}+\frac{9}{10}\times\frac{4}{5}\times\frac{1}{3}$ or $1-\left(\frac{1}{10}+\frac{9}{10}\times\frac{4}{5}\times\frac{2}{3}\right)$	M1	3.1b
	$=\frac{21}{50}$ $(=0.42)$	A1	1.1b
		(2)	
(d)	[P(Red from B|Red selected)] $=\frac{\frac{9}{10}\times\frac{1}{5}}{\frac{1}{10}+\frac{9}{10}\times\frac{1}{5}+\frac{9}{10}\times\frac{4}{5}\times\frac{1}{3}}\left[=\frac{\frac{9}{50}}{\frac{13}{25}}\right]$	M1	3.1b
	$=\frac{9}{26}$	A1	1.1b
		(2)	
		(8 marks)	

Ch. 3: The Normal Distribution

June 2024 Question 5

5. The records for a school athletics club show that the height, H metres, achieved by students in the high jump is normally distributed with mean 1.4 metres and standard deviation 0.15 metres.

(a) Find the proportion of these students achieving a height of more than 1.6 metres.

(1)

The records also show that the time, T seconds, to run 1500 metres is normally distributed with mean 330 seconds and standard deviation 26 seconds.

The school's Head would like to use these distributions to estimate the proportion of students from the school athletics club who can jump higher than 1.6 metres **and** can run 1500 metres in less than 5 minutes.

(b) State a necessary assumption about H and T for the Head to calculate an estimate of this proportion.

(1)

(c) Find the Head's estimate of this proportion.

(3)

Students in the school athletics club also throw the discus.

The random variable $D \sim \mathrm{N}\left(\mu, \sigma^2\right)$ represents the distance, in metres, that a student can throw the discus.

Given that $\mathrm{P}(D < 16.3) = 0.30$ and $\mathrm{P}(D > 29.0) = 0.10$

(d) calculate the value of μ and the value of σ

(5)

ANSWER

Qu 5	Scheme	Marks	AO
(a)	[P($H > 1.6$) =] 0.091211… = awrt **0.0912**	B1	1.1b
		(1)	
(b)	Need H and T to be independent <u>or</u> events $\{H > 1.6\}$ and $\{T < 300\}$ are independent	B1	2.4
		(1)	
(c)	[P($T < 300$) =] 0.124(2816…)	M1	3.4
	Prob both is: "0.0912..."×"0.124..."	M1	1.1b
	$= 0.011335…$ = awrt **0.0113**	A1	1.1b
		(3)	
(d)	$\frac{16.3-\mu}{\sigma} = -0.5244(0051...)$, $\frac{29-\mu}{\sigma} = 1.2816$ (calc: 1.28155156…)	M1M1	3.1a 1.1b
	e.g. $29 - 16.3 = \sigma(\text{"}1.2816\text{"} - \text{"}-0.5244\text{"})$	M1	1.1b
	$\sigma = 7.032115...$ = awrt **7.03**	A1	1.1b
	$\mu = 19.9876...$ = $\underline{\mathbf{19.95}} \leqslant \underline{\mu} \leqslant \underline{\mathbf{20.0}}$	A1	3.2a
		(5)	
		(10 marks)	

June 2024 Question 1

1. Xian rolls a fair die 10 times.

The random variable X represents the number of times the die lands on a six.

(a) Using a suitable distribution for X, find

(i) $P(X = 3)$

(ii) $P(X < 3)$

(3)

Xian repeats this experiment each day for 60 days and records the number of days when $X = 3$

(b) Find the probability that there were at least 12 days when $X = 3$

(3)

(c) Find an estimate for the total number of sixes that Xian will roll during these 60 days.

(1)

(d) Use a normal approximation to estimate the probability that Xian rolls a total of more than 95 sixes during these 60 days.

(4)

ANSWER

Qu 1	Scheme	Marks	AO
(a)	$X \sim B(10, \frac{1}{6})$ [Allow 0.167 or better for $\frac{1}{6}$]	M1	3.3
(i)	[$P(X = 3) =$] 0.155045… **awrt <u>0.155</u>**	A1	1.1b
(ii)	[$P(X < 3) = P(X \leqslant 2) =$] 0.775226… **awrt <u>0.775</u>**	A1	1.1b
		(3)	
(b)	[Let D = no. of days when $X = 3$] $D \sim B(60, \text{"}0.155\text{"})$	M1	3.3
	$P(D \geqslant 12) = 1 - P(D \leqslant 11)$ [Allow $1 - P(D < 12)$]	M1	3.4
	$= 1 - 0.78819\ldots$ **awrt <u>0.212</u>**	A1	1.1b
		(3)	
(c)	[$n = 600, p = \frac{1}{6}$] estimate = **<u>100</u>**	B1	3.4
		(1)	
(d)	[S = total no. of sixes over 60 days.] $S \approx T \sim N\left(\text{"100"}, \sqrt{\frac{5}{6} \times 100}^{\,2}\right)$	M1A1	3.3,1.1b
	$P(S > 95) \approx P([T >]95.5)$ <u>or</u> $P\left([Z >]\frac{95.5 - \text{"100"}}{\text{"9.128..."}}\right)$ <u>or</u> $P([Z >] -0.49..)$	M1	3.4
	$= 0.688976\ldots$ **awrt <u>0.689</u>**	A1	1.1b
		(4)	
		(11 marks)	

June 2023 Question 6

6. A medical researcher is studying the number of hours, T, a patient stays in hospital following a particular operation.

The histogram on the page opposite summarises the results for a random sample of 90 patients.

(a) Use the histogram to estimate $P(10 < T < 30)$

(2)

For these 90 patients the time spent in hospital following the operation had

- a mean of 14.9 hours
- a standard deviation of 9.3 hours

Tomas suggests that T can be modelled by $N(14.9, 9.3^2)$

(b) With reference to the histogram, state, giving a reason, whether or not Tomas' model could be suitable.

(1)

Xiang suggests that the frequency polygon based on this histogram could be modelled by a curve with equation

$$y = kxe^{-x} \quad 0 \leqslant x \leqslant 4$$

where

- x is measured in **tens of hours**
- k is a constant

(c) Use algebraic integration to show that

$$\int_0^n xe^{-x}\,dx = 1 - (n+1)e^{-n}$$

(4)

(d) Show that, for Xiang's model, $k = 99$ to the nearest integer.

(3)

(e) Estimate $P(10 < T < 30)$ using

(i) Tomas' model of $T \sim N(14.9, 9.3^2)$

(1)

(ii) Xiang's curve with equation $y = 99xe^{-x}$ and the answer to part (c)

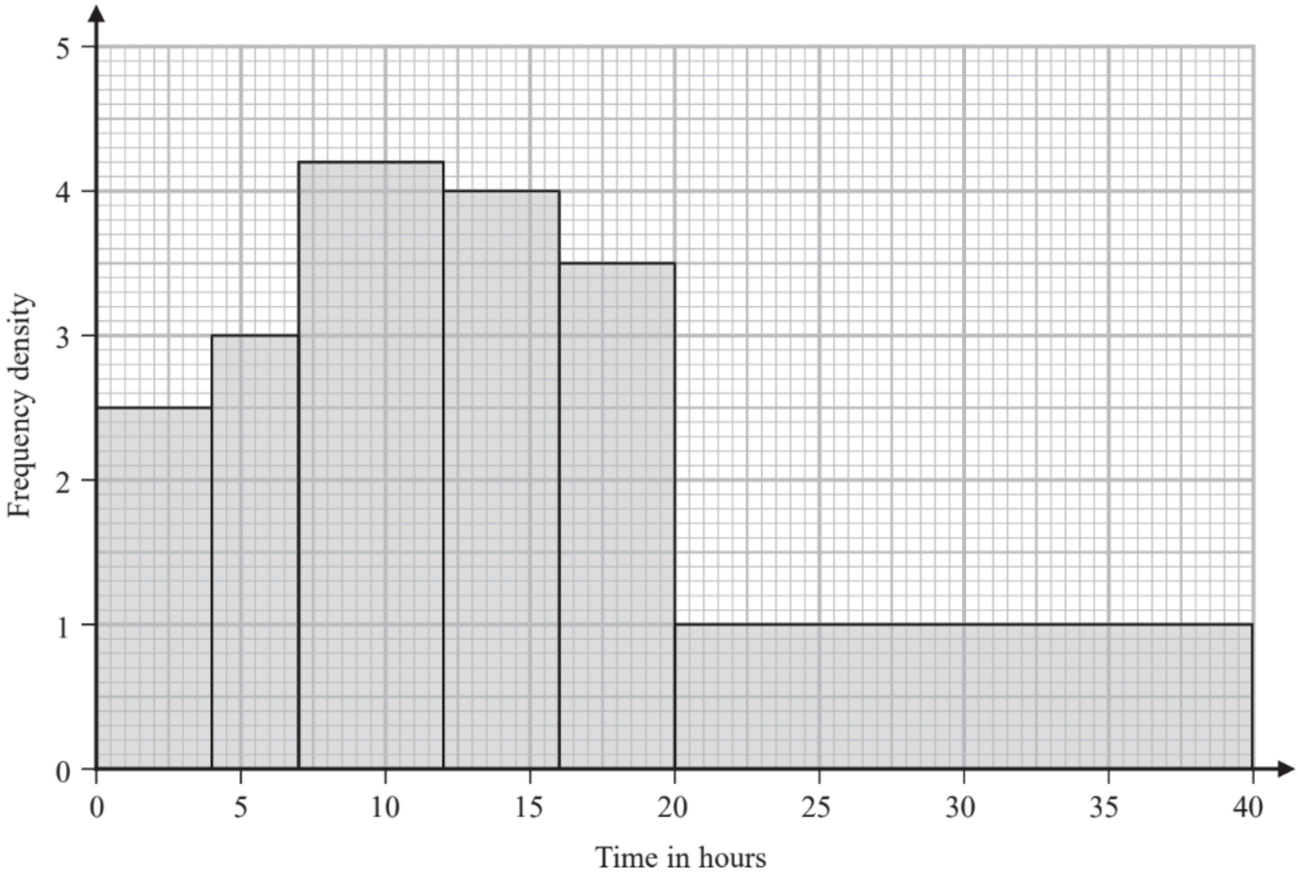
(2)

The researcher decides to use Xiang's curve to model $P(a < T < b)$

(f) State one limitation of Xiang's model.

(1)

Question 6 continued



ANSWER

Qu 6	Scheme	Marks	AO
(a)	$2\times4.2\,,4\times4\,,4\times3.5\,,10\times1$ $(=8.4+16+14+10=48.4)$	M1	1.1b
	[So P$(10<T<30)$ =] $\left[\dfrac{48.4}{90}\right]=\dfrac{121}{225}=0.53777...$ **<u>0.53~0.54</u>** (2sf OK)	A1	1.1b
		(2)	
(b)	(Not suitable as) data is not symmetric <u>or</u> is skew (normal is symmetric) ("Even" distribution or a diagram <u>on its own</u> is not enough so B0)	B1 **(1)**	2.4
(c)	$\int xe^{-x}\ (\mathrm{d}x)=\int x\mathrm{d}(-e^{-x})$	M1	2.1
	$=\left[-xe^{-x}\right]-\int\left(-e^{-x}\right)(\mathrm{d}x)\quad(+c)$	A1	1.1b
	$\int_0^n xe^{-x}\ (\mathrm{d}x)=\left[-xe^{-x}-e^{-x}\right]_0^n=\left(-ne^{-n}-e^{-n}\right)-\left[-(0)-1\right]$	dM1	1.1b
	$=\underline{1-(n+1)e^{-n}}$ (*)	A1cso*	1.1b
		(4)	
(d)	Require area = 90 i.e. $k\int_{(0)}^{(n)} xe^{-x}\ \mathrm{d}x=90$ (ignore limits)	M1	3.1a
	Using the result in part (c) with $n=4$ gives $k\left[1-5e^{-4}\right]=90$	M1	2.1
	(k =) **<u>99</u>**(.0729…) (*)	A1cso* **(3)**	1.1b
(e)(i)	[P$(10<T<30)$ =] 0.64863… awrt **<u>0.649</u>**	B1 **(1)**	1.1b
(ii)	[No. of patients =] $(99)\left[\left(1-4e^{-3}\right)-\left(1-2e^{-1}\right)\right]$ $(=53.1..)$	M1	3.4
	Prob $=\dfrac{0.5366...\times99}{90}=0.59027...$[or 0.5907...] = awrt **<u>0.590 or 0.591</u>**	A1 **(2)**	3.2a
(f)	eg Patients might stay longer than 40 hours (Can ignore other comments unless clearly contradictory.)	B1 **(1)**	3.5b
		(14 marks)	

June 2023 Question 4

4. A study was made of adult men from region A of a country.
It was found that their heights were normally distributed with a mean of 175.4 cm and standard deviation 6.8 cm.

(a) Find the proportion of these men that are taller than 180 cm.

(1)

A student claimed that the mean height of adult men from region B of this country was different from the mean height of adult men from region A.

A random sample of 52 adult men from region B had a mean height of 177.2 cm

The student assumed that the standard deviation of heights of adult men was 6.8 cm both for region A and region B.

(b) Use a suitable test to assess the student's claim.
You should

- state your hypotheses clearly
- use a 5% level of significance

(4)

(c) Find the p-value for the test in part (b)

(1)

ANSWER

Qu 4	Scheme	Marks	AO
(a)	[Let N = height from region A; P(N > 180) =] 0.24937… awrt **<u>0.249</u>**	B1 **(1)**	1.1b
(b)	$H_0 : \mu = 175.4 \quad H_1 : \mu \neq 175.4$	B1	2.5
	[S = height from region B] $\bar{S} \sim N\left(175.4, \frac{6.8^2}{52}\right)$ Allow σ^2 = awrt 0.889	M1	3.3
	[P($\bar{S}$ > 177.2)] = 0.02814…	A1	3.4
	[0.028… > 0.025, Not sig, do not reject H_0] <u>Insufficient</u> evidence to <u>support</u> student's <u>claim</u>	A1 **(4)**	2.2b
(c)	[p-value = $2 \times 0.02814...$ =] 0.05628… in range **<u>0.056~0.06</u>** or **<u>5.6(%)~6(%)</u>**	B1ft **(1)**	1.2
		(6 marks)	

June 2022 Question 1

1. George throws a ball at a target 15 times.
Each time George throws the ball, the probability of the ball hitting the target is 0.48

The random variable X represents the number of times George hits the target in 15 throws.

(a) Find

(i) $\mathrm{P}(X = 3)$

(ii) $\mathrm{P}(X \geqslant 5)$

(3)

George now throws the ball at the target 250 times.

(b) Use a normal approximation to calculate the probability that he will hit the target more than 110 times.

(3)

ANSWER

Question	Scheme		Marks	AOs
1(a)(i)	$X \sim B(15, 0.48)$		M1	3.3
	$P(X=3)=0.019668...$	awrt 0.0197	A1	3.4
(ii)	$\left[P(X \geqslant 5)=1-P(X \leqslant 4)\right]=0.92013...$	awrt 0.920	A1	1.1b
			(3)	
(b)	Y is the number of hits	M is the number of misses		
	$Y \sim N(120, 62.4)$	$M \sim N(130, 62.4)$	B1	3.3
	$P(X>110) \approx P(Y>110.5)$ $\left[=P\left(Z>\frac{110.5-"120"}{\sqrt{"62.4"}}\right)\right]$	$P(X>110) \approx P(M<139.5)$ $\left[=P\left(Z<\frac{139.5-"130"}{\sqrt{"62.4"}}\right)\right]$	M1	3.4
	$=0.88544\ldots$		A1	1.1b
			(3)	
				(6 marks)

Video solution:

https://youtu.be/pQZFNXX4JDU

June 2022 Question 2

2. A manufacturer uses a machine to make metal rods.

The length of a metal rod, L cm, is normally distributed with

- a mean of 8 cm
- a standard deviation of x cm

Given that the proportion of metal rods less than 7.902 cm in length is 2.5%

(a) show that $x = 0.05$ to 2 decimal places.

(2)

(b) Calculate the proportion of metal rods that are between 7.94 cm and 8.09 cm in length.

(1)

The **cost** of producing a single metal rod is 20p

A metal rod

- where $L < 7.94$ is **sold** for scrap for 5p
- where $7.94 \leqslant L \leqslant 8.09$ is **sold** for 50p
- where $L > 8.09$ is shortened for an extra **cost** of 10p and then **sold** for 50p

(c) Calculate the expected profit per 500 of the metal rods.
Give your answer to the nearest pound.

(5)

The same manufacturer makes metal hinges in large batches.

The hinges each have a probability of 0.015 of having a fault.

A random sample of 200 hinges is taken from each batch and the batch is accepted if fewer than 6 hinges are faulty.

The manufacturer's aim is for 95% of batches to be accepted.

(d) Explain whether the manufacturer is likely to achieve its aim.

(4)

ANSWER

Qu	Scheme		Marks	AOs
2(a)	$\left[P(L<7.902)=0.025\Rightarrow\right]\ \dfrac{7.902-8}{x}=-1.96$ oe		M1	3.4
	$[x=]0.05\,*$		A1cso*	1.1b
	SC B1(mark as M0A1) for $\dfrac{7.902-8}{0.05}=-1.96\Rightarrow 0.024998$			
			(2)	
(b)	$P(7.94\leqslant L\leqslant 8.09)=0.8490\ldots$ **awrt 0.849**		B1	1.1b
			(1)	
(c)	$[P(L<7.94)=]\ 0.115069\ldots$ (awrt 0.115) **or** $[P(L>8.09)=]\ 0.03593\ldots$ (awrt 0.036)		B1	1.1b
	$[P(L<7.94)=]\ 0.115069\ldots$ (awrt 0.115) **&** $[P(L>8.09)=]\ 0.03593\ldots$ (awrt 0.036)		B1	1.1b
	Expected income per 500 rods = $\sum(\text{Income}\times\text{probability}\times 500)$ $(500\times"0.849"\times 0.5)+(500\times"0.1150..."\times 0.05)+(500\times"0.03593..."\times 0.4)$ **or** Expected profit per rod = $\sum(\text{Profit}\times\text{probability})$ $0.30\times"0.849"+-0.15\times"0.1150..."+0.20\times"0.03593..."$ [= 0.2446..]		M1	3.4
	Expected profit per 500 rods $500\times\sum(\text{Profit}\times\text{probability})$ **or** $\sum(\text{Income}\times\text{probability}\times 500)-500\times 0.2$ $=500\times"0.2446..."$ **or** $=$ "222.3" -500×0.2		M1d	3.1b
	$=[£]122.3\ldots$ **awrt [£]122**		A1	1.1b
			(5)	
(d)	Let $X\sim B(200, 0.015)$		M1	3.3
	$P(X\leqslant 5)=$	$P(X\geqslant 6)=$	M1	1.1b
	0.9176…	0.0824	A1	1.1b
	Manufacturer is unlikely to achieve their aim since <u>$0.9176<0.95$</u>	Manufacturer is unlikely to achieve their aim since <u>$0.0824>0.05$</u>	A1ft	2.4
			(4)	
	Notes:		**(12 marks)**	

Video solution:

https://youtu.be/JbKRjbZgFj8

November 2021 question 5

5. The heights of females from a country are normally distributed with

- a mean of 166.5 cm
- a standard deviation of 6.1 cm

Given that 1% of females from this country are shorter than k cm,

(a) find the value of k

(2)

(b) Find the proportion of females from this country with heights between 150 cm and 175 cm

(1)

A female, from this country, is chosen at random from those with heights between 150 cm and 175 cm

(c) Find the probability that her height is more than 160 cm

(4)

The heights of females from a different country are normally distributed with a standard deviation of 7.4 cm

Mia believes that the mean height of females from this country is less than 166.5 cm

Mia takes a random sample of 50 females from this country and finds the mean of her sample is 164.6 cm

(d) Carry out a suitable test to assess Mia's belief.
You should

- state your hypotheses clearly
- use a 5% level of significance

(4)

ANSWER

Qu 5	Scheme	Marks	AO
(a)	$\left[\text{Let } F \sim \mathrm{N}\left(166.5, 6.1^2\right)\right]$ $\mathrm{P}(F < k) = 0.01 \Rightarrow \dfrac{k-166.5}{6.1} = -2.3263$	M1	3.4
	$k = 152.309\ldots$ **152** or awrt **152.3**	A1	1.1b
		(2)	
(b)	$[\ \mathrm{P}(150 < F < 175) =]\ 0.914840\ldots$ awrt **0.915**	B1	1.1b
		(1)	
(c)	$\mathrm{P}(F > 160 \mid 150 < F < 175)$	M1	3.1b
	$= \dfrac{\mathrm{P}(160 < F < 175)}{\mathrm{P}(150 < F < 175)}$ or $\dfrac{\mathrm{P}(160 < F < 175)}{\text{"(b)"}}$	M1	1.1b
	$= \dfrac{0.7749487...}{\text{"0.91484..."}}$	A1ft	1.1b
	$= 0.84708\ldots$ awrt **0.847**	A1	1.1b
		(4)	
(d)	$\mathrm{H}_0 : \mu = 166.5 \quad \mathrm{H}_1 : \mu < 166.5$	B1	2.5
	[Let X = height of female from 2nd country] $\bar{X} \sim \mathrm{N}\left(166.5, \left(\dfrac{7.4}{\sqrt{50}}\right)^2\right)$	M1	3.3
	$\mathrm{P}\left(\bar{X} < 164.6\right) = 0.03472...$	A1	3.4
	[0.0347… < 0.05 so significant or reject H0] There is evidence to support Mia's belief	dA1	2.2b
		(4)	
		(11 marks)	

October 2020 question 5

5. A health centre claims that the time a doctor spends with a patient can be modelled by a normal distribution with a mean of 10 minutes and a standard deviation of 4 minutes.

(a) Using this model, find the probability that the time spent with a randomly selected patient is more than 15 minutes.

(1)

Some patients complain that the mean time the doctor spends with a patient is more than 10 minutes.

The receptionist takes a random sample of 20 patients and finds that the mean time the doctor spends with a patient is 11.5 minutes.

(b) Stating your hypotheses clearly and using a 5% significance level, test whether or not there is evidence to support the patients' complaint.

(4)

The health centre also claims that the time a dentist spends with a patient during a routine appointment, T minutes, can be modelled by the normal distribution where $T \sim N(5, 3.5^2)$

(c) Using this model,

(i) find the probability that a routine appointment with the dentist takes less than 2 minutes

(1)

(ii) find $P(T < 2 \mid T > 0)$

(3)

(iii) hence explain why this normal distribution may not be a good model for T.

(1)

The dentist believes that she cannot complete a routine appointment in less than 2 minutes.

She suggests that the health centre should use a refined model only including values of $T > 2$

(d) Find the median time for a routine appointment using this new model, giving your answer correct to one decimal place.

(5)

ANSWER

Qu 5	Scheme	Marks	AO
(a)	{Let X = time spent, $\mathrm{P}(X > 15) =$ } 0.105649… awrt **<u>0.106</u>**	B1 **(1)**	1.1b
(b)	$\mathrm{H}_0 : \mu = 10 \quad \mathrm{H}_1 : \mu > 10$	B1	2.5
	$\bar{X} \sim \mathrm{N}\left(10, \left(\frac{4}{\sqrt{20}}\right)^2\right)$; $\mathrm{P}\left(\bar{X} > 11.5\right) = 0.046766\ldots$ [Condone 0.9532…]	M1;A1	3.3;3.4
	[This is significant ($< 5\%$) so] there is evidence to support the complaint	A1 **(4)**	2.2b
(c)(i)	[$\mathrm{P}(T < 2) =$] 0.1956… awrt **<u>0.196</u>**	B1 **(1)**	1.1b
(ii)	Require $\frac{\mathrm{P}(0 < T < 2)}{\mathrm{P}(T > 0)} = \frac{0.119119...}{0.923436...}$; $= 0.1289955\ldots$ awrt **<u>0.129</u>**	M1 A1;A1 **(3)**	3.4 1.1bx2
(iii)	The current model suggests **non-negligible** probability of T values < 0 which is impossible	B1 **(1)**	3.5b
(d)	Require t such that $\mathrm{P}(T > t \mid T > 2) = 0.5$ <u>or</u> $\mathrm{P}(T < t \mid T > 2) = 0.5$	M1	3.1b
	e.g. $\frac{\mathrm{P}(T > t)}{\mathrm{P}(T > 2)} = 0.5$; so $\mathrm{P}(T > t) = 0.5 \times [1 - \text{(c)(i)}]$ or $\mathrm{P}(T > t) = 0.5 \times 0.8043..$	M1; A1ft	1.1b 3.4
	[i.e. $\mathrm{P}(T > t) = 0.40\ldots$ implies] $\frac{t - 5}{3.5} = 0.2533$ <u>or</u> $\mathrm{P}(T < t) =$ "0.5978.."	M1	1.1b
	$t = 5.886\ldots$ <u>or</u> from calculator 5.867… so awrt **<u>5.9</u>**	A1 **(5)**	1.1b
		(15 marks)	

June 2019 question 2

2.

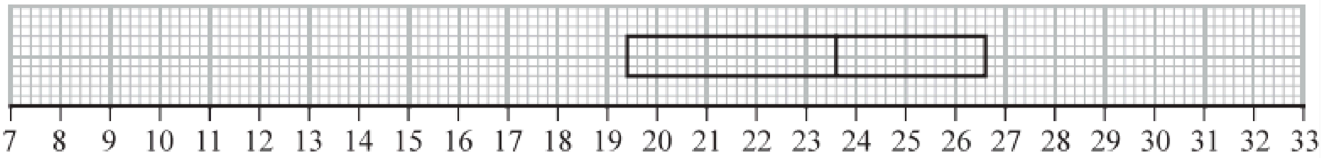


Figure 1

The partially completed box plot in Figure 1 shows the distribution of daily mean air temperatures using the data from the large data set for Beijing in 2015

An outlier is defined as a value
more than 1.5 × IQR below Q_1 or
more than 1.5 × IQR above Q_3

The three lowest air temperatures in the data set are 7.6 °C, 8.1 °C and 9.1 °C
The highest air temperature in the data set is 32.5 °C

(a) Complete the box plot in Figure 1 showing clearly any outliers.
(4)

(b) Using your knowledge of the large data set, suggest from which month the two outliers are likely to have come.
(1)

Using the data from the large data set, Simon produced the following summary statistics for the daily mean air temperature, x °C, for Beijing in 2015

$$n = 184 \qquad \sum x = 4153.6 \qquad S_{xx} = 4952.906$$

(c) Show that, to 3 significant figures, the standard deviation is 5.19 °C
(1)

Simon decides to model the air temperatures with the random variable

$$T \sim \mathrm{N}(22.6, 5.19^2)$$

(d) Using Simon's model, calculate the 10th to 90th interpercentile range.
(3)

Simon wants to model another variable from the large data set for Beijing using a normal distribution.

(e) State two variables from the large data set for Beijing that are **not** suitable to be modelled by a normal distribution. Give a reason for each answer.
(2)

ANSWER

Question	Scheme	Marks	AOs
2(a)	IQR = 26.6 – 19.4 [= 7.2]	B1	2.1
	19.4 – 1.5 × ‘7.2’ [= 8.6] or 26.6 + 1.5 × ‘7.2’ [= 37.4]	M1	1.1b
	Plotting one upper whisker to 32.5 **and** one lower whisker to 8.6 or 9.1	A1	1.1b
	Plotting 7.6 and 8.1 as the only two outliers	A1	1.1b
		(4)	
(b)	October (since it is the month with the coldest temperatures between May and October in Beijing)	B1	2.4
		(1)	
(c)	$[\sigma =]\sqrt{\frac{4952.906}{184}}$ or e.g. $[\sigma =]\sqrt{\frac{S_{xx}}{n}} = 5.188...$ [=5.19*]	B1cso*	1.1b
		(1)	
(d)	$z = (\pm)\ 1.28(16)$; $[P_{90} =]29.251...$ or $[P_{10} =]15.948...$	B1	3.1b
	$2 \times 1.2816 \times 5.19$; ‘29.251…’ – ‘15.948…’	M1	1.1b
	= awrt **13.3**	A1	1.1b
		(3)	
(e)	Daily mean wind speed/Beaufort conversion since it is qualitative Rainfall since it is not symmetric/lots of days with 0 rainfall	B1 B1	2.4 2.4
		(2)	
		(11 marks)	

June 2019 question 5

5. A machine puts liquid into bottles of perfume. The amount of liquid put into each bottle, D ml, follows a normal distribution with mean 25 ml

Given that 15% of bottles contain less than 24.63 ml

(a) find, to 2 decimal places, the value of k such that $P(24.63 < D < k) = 0.45$

(5)

A random sample of 200 bottles is taken.

(b) Using a normal approximation, find the probability that fewer than half of these bottles contain between 24.63 ml and k ml

(3)

The machine is adjusted so that the standard deviation of the liquid put in the bottles is now 0.16 ml

Following the adjustments, Hannah believes that the mean amount of liquid put in each bottle is less than 25 ml

She takes a random sample of 20 bottles and finds the mean amount of liquid to be 24.94 ml

(c) Test Hannah's belief at the 5% level of significance.
You should state your hypotheses clearly.

(5)

ANSWER

Question	Scheme	Marks	AOs
5(a)	$\frac{24.63-25}{'\sigma'}=-1.0364$	M1	3.1b
	$[\sigma=]0.357$ (must come from compatible signs)	A1	1.1b
	$\mathrm{P}(D>k)=0.4$ or $\mathrm{P}(D<k)=0.6$	B1	1.1b
	$\frac{k-25}{'0.357'}=0.2533$	M1	3.4
	$k=$ awrt **25.09**	A1	1.1b
		(5)	
(b)	$[Y\sim \mathrm{B}(200, 0.45)\rightarrow]\ W\sim \mathrm{N}(90, 49.5)$	B1	3.3
	$\mathrm{P}(Y<100)\approx \mathrm{P}(W<99.5)\left[=\mathrm{P}\left(Z<\frac{99.5-90}{\sqrt{49.5}}\right)\right]$	M1	3.4
	$=0.9115\ldots$ awrt **0.912**	A1	1.1b
		(3)	
(c)	$\mathrm{H}_0: \mu=25 \qquad \mathrm{H}_1: \mu<25$	B1	2.5
	$[\bar{D}\sim]\mathrm{N}\left(25, \frac{0.16^2}{20}\right)$	M1	3.3
	$\mathrm{P}(\bar{D}<24.94)[=\mathrm{P}(Z<-1.677...)]=0.046766...$	A1	3.4
	$p=0.047<0.05$ or $z=-1.677\ldots<-1.6449$ or $24.94<24.94115\ldots$ or reject H_0/in the critical region/significant	M1	1.1b
	There is sufficient evidence to support Hannah's belief.	A1	2.2b
		(5)	
			(13 marks)

June 2018 Question 5

5. The lifetime, L hours, of a battery has a normal distribution with mean 18 hours and standard deviation 4 hours.

Alice's calculator requires 4 batteries and will stop working when any one battery reaches the end of its lifetime.

(a) Find the probability that a randomly selected battery will last for longer than 16 hours.

(1)

At the start of her exams Alice put 4 new batteries in her calculator.
She has used her calculator for 16 hours, but has another 4 hours of exams to sit.

(b) Find the probability that her calculator will not stop working for Alice's remaining exams.

(5)

Alice only has 2 new batteries so, after the first 16 hours of her exams, although her calculator is still working, she randomly selects 2 of the batteries from her calculator and replaces these with the 2 new batteries.

(c) Show that the probability that her calculator will not stop working for the remainder of her exams is 0.199 to 3 significant figures.

(3)

After her exams, Alice believed that the lifetime of the batteries was more than 18 hours. She took a random sample of 20 of these batteries and found that their mean lifetime was 19.2 hours.

(d) Stating your hypotheses clearly and using a 5% level of significance, test Alice's belief.

(5)

ANSWER

Qu 5	Scheme	Marks	AO
(a)	$P(L>16)=0.69146\ldots$ **awrt 0.691**	B1 (1)	1.1b
(b)	$P(L>20\mid L>16)=\dfrac{P(L>20)}{P(L>16)}$	M1	3.1b
	$=\dfrac{0.308537...}{(a)}$ <u>or</u> $\dfrac{1-(a)}{(a)}, =0.44621\ldots$	A1ft, A1	1.1b 1.1b
	For calc to work require $(0.44621...)^4=0.03964\ldots$ **awrt <u>0.0396</u>**	dM1 A1 (5)	2.1 1.1b
(c)	Require: $[P(L>4)]^2\times[P(L>20\mid L>16)]^2$	M1	1.1a
	$=(0.99976...)^2\times(\text{"}0.44621...\text{"})^2$	A1ft	1.1b
	$=0.19901\ldots$ **awrt <u>0.199</u> (*)**	A1cso* (3)	1.1b
(d)	$H_0:\mu=18\quad H_1:\mu>18$	B1	2.5
	$\bar{L}\sim N\left(18,\left(\dfrac{4}{\sqrt{20}}\right)^2\right)$	M1	3.3
	$P(\bar{L}>19.2)=P(Z>1.3416\ldots)=0.089856\ldots$	A1	3.4
	$(0.0899>5\%)$ <u>or</u> $(19.2<19.5)$ <u>or</u> $1.34<1.6449$ so not significant	A1	1.1b
	Insufficient evidence to support Alice's claim (or belief)	A1 (5)	3.5a
		(14 marks)	