1. The random variable G has the distribution $N(\mu, \sigma^2)$. One hundred observations of G are taken. The results are summarised in the following table.

Interval	$G < 40.0$	$40.0 \leqslant G < 60.0$	$G \geqslant 60.0$
Frequency	17	58	25

i. By considering $P(G < 40.0)$, write down an equation involving μ and σ.

[2]

ii. Find a second equation involving μ and σ. Hence calculate values for μ and σ.

[4]

iii. Explain why your answers are only estimates.

[1]

2. The random variable Y is normally distributed with mean μ and variance σ^2. It is found that $P(Y > 150.0) = 0.0228$ and $P(Y > 143.0) = 0.9332$. Find the values of μ and σ.

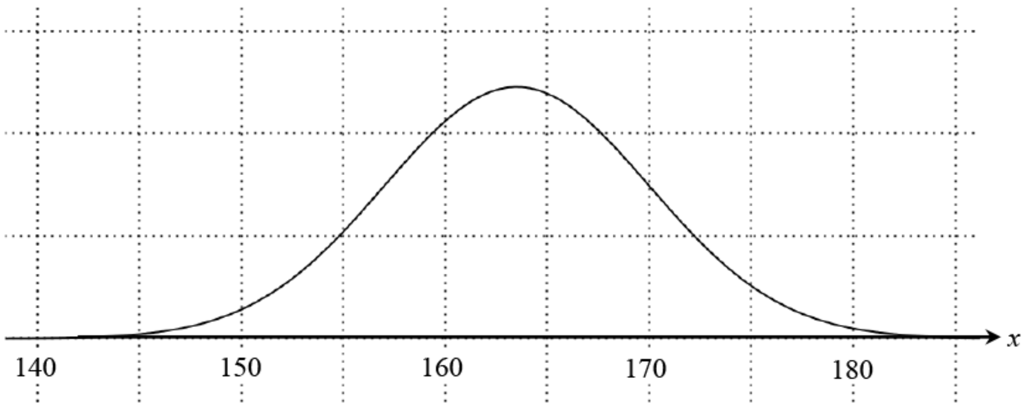
[6]

3. The mass, in kilograms, of a packet of flour is a normally distributed random variable with mean 1.03 and variance σ^2. Given that 5% of packets have mass less than 1.00 kg, find the percentage of packets with mass greater than 1.05 kg.

[6]

4. (a) The heights of English men aged 25 to 34 are normally distributed with mean 178 cm and standard deviation 8 cm. Three English men aged 25 to 34 are chosen at random. Find the probability that all three of them have a height less than 194 cm. [3]

(b) The diagram shows the distribution of heights of Scottish women aged 25 to 34.



It is given that the distribution is approximately normal. Use the diagram above to estimate the standard deviation of these heights, explaining your method. [3]

5. A market gardener records the masses of a random sample of 100 of this year's crop of plums. The table shows his results.

Mass, m grams	$m < 25$	$25 \le m < 35$	$35 \le m < 45$	$45 \le m < 55$	$55 \le m < 65$	$65 \le m < 75$	$m \ge 75$
Number of plums	0	3	29	36	30	2	0

(a) Explain why the normal distribution might be a reasonable model for this distribution. [1]

The market gardener models the distribution of masses by $N(47.5, 10^2)$.

(b) Find the number of plums in the sample that this model would predict to have masses in the range

(i) $35 \le m < 45$, [2]

(ii) $m < 25$. [2]

(c) Use your answers to parts **(b)(i)** and **(b)(ii)** to comment on the suitability of this model. [1]

(d) The market gardener plans to use this model to predict the distribution of the masses of next year's crop of plums. Comment on this plan. [1]

6. The variable X has the distribution $N(20, 4^2)$.

(a) Given that $P(X < a) = 0.1$, find a. [1]

(b) Given that $P(b < X < c) = 0.95$, find a possible pair of values of b and c. [2]

7. The heaviest 17% of rococo apples are classified as large, and the lightest 17% are classified as small. The remainder are classified as medium. The limits within which the masses of medium rococo apples lie are 96 g and 120 g. Stating a necessary assumption, estimate the mass of the heaviest rococo apple. [4]

8. The masses, X grams, of tomatoes are normally distributed. Half of the tomatoes have masses greater than 56.0 g and 70% of the tomatoes have masses greater than 53.0 g.

(a) Find the percentage of tomatoes with masses greater than 59.0 g. [2]

(b) Find the percentage of tomatoes with masses greater than 65.0 g. [4]

(c) Given that $P(a < X < 50) = 0.1$, find a. [3]

9. (a) The variable X has the distribution $N(20, 9)$.

(i) Find $P(X > 25)$. [1]

(ii) Given that $P(X > a) = 0.2$, find a. [1]

(iii) Find b such that $P(20 - b < X < 20 + b) = 0.5$. [3]

(b) The variable Y has the distribution $N\left(\mu, \frac{\mu^2}{9}\right)$. Find $P(Y > 1.5\mu)$. [3]

10. **In this question you must show detailed reasoning.**

The probability that Paul's train to work is late on any day is 0.15, independently of other days.

(a) The number of days on which Paul's train to work is late during a 450-day period is denoted by the random variable Y. Find a value of a such that $\mathrm{P}(Y > a) \approx \frac{1}{6}$. [3]

In the expansion of $(0.15 + 0.85)^{50}$, the terms involving 0.15^r and 0.15^{r+1} are denoted by T_r and T_{r+1} respectively.

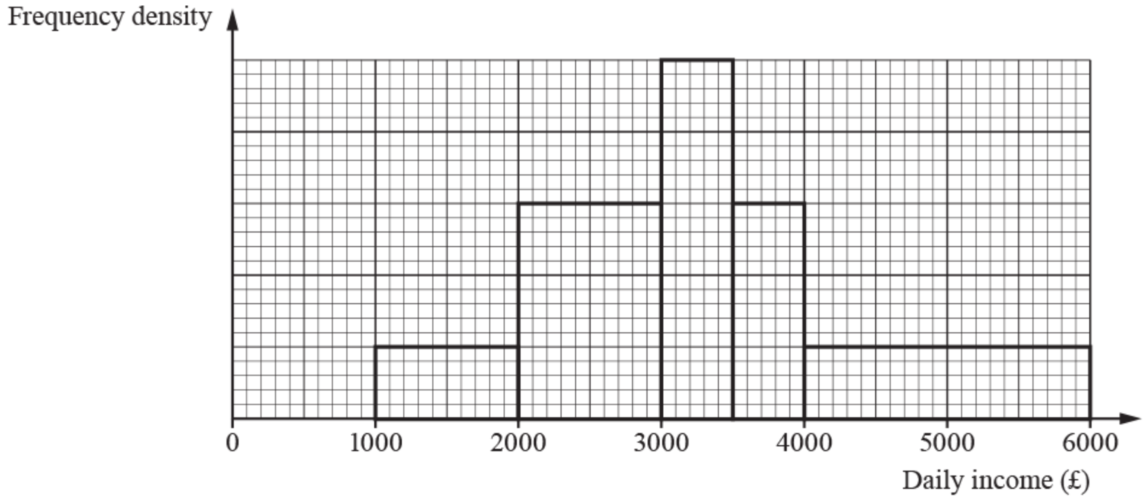
(b) Show that $\dfrac{T_r}{T_{r+1}} = \dfrac{17(r+1)}{3(50-r)}$. [3]

(c) The number of days on which Paul's train to work is late during a 50-day period is modelled by the random variable X.

(i) Find the values of r for which $\mathrm{P}(X = r) \le \mathrm{P}(X = r + 1)$. [4]

(ii) Hence find the most likely number of days on which the train will be late during a 50-day period. [2]

11. The finance department of a retail firm recorded the daily income each day for 300 days. The results are summarised in the histogram.



(a) Find the number of days on which the daily income was between £4000 and £6000. [3]

(b) Calculate an estimate of the number of days on which the daily income was between £2700 and £3600. [3]

(c) Use the midpoints of the classes to show that an estimate of the mean daily income is £3275. [2]

An estimate of the standard deviation of the daily income is £1060. The finance department uses the distribution $N(3275, 1060^2)$ to model the daily income, in pounds.

(d) Calculate the number of days on which, according to this model, the daily income would be between £4000 and £6000. [2]

(e) It is given that approximately 95% of values of the distribution $N(\mu, \sigma^2)$ lie within the range $\mu, \pm 2\sigma$. Without further calculation, use this fact to comment briefly on whether the proposed model is a good fit to the data illustrated in the histogram. [2]

12. A fair dice is thrown 1000 times and the number, X, of throws on which the score is 6 is noted.

(a) (i) State the distribution of X. [1]

(ii) Explain why a normal distribution would be an appropriate approximation to the distribution of X. [1]

(b) Use a normal distribution to find two positive integer values, a and b, such that $P(a < X < b) \approx 0.4$. [5]

(c) For your two values of a and b, use the distribution of part **(a)(i)** to find the value of $P(a < X < b)$, correct to 3 significant figures. [2]

END OF QUESTION paper

Mark scheme

Question			Answer/Indicative content	Marks	Part marks and guidance	
1		i	$\frac{\mu - 40}{\sigma} = 0.9544$	M1	Standardise with μ and σ and equate to Φ^{-1}, allow σ^2 but not $\sqrt{n}$, allow 1−, cc, wrong signs. P(…): M0 here. But can recover both marks from part (ii).	
		i		B1	[0.954, 0.955] seen **Examiner's Comments** Placing a question about finding normal parameters in the context of a frequency distribution did not worry most candidates, though some inevitably attempted to use a factor of √100 in the standard deviation. Signs were generally well handled, and the correct answers were often seen. Only a few used 0.17 or 0.25 in their equations, but continuity corrections in this context continue to be seen quite often. If the distribution is stated to be normal then a continuity correction must not be used.	
		ii	$\frac{60 - \mu}{\sigma} = 0.674(5)$	M1	Standardise as in (i) but do not give if "1 –" or wrong signs in *either* equation	
		ii		B1	[0.674, 0.675] seen. (Other errors lead to loss of A marks.)	
		ii	Solve to get σ = 12.3 [12.278]	A1	σ, a.r.t. 12.3, cwo	
		ii	μ = 51.7(18)	A1	μ, a.r.t. 51.7, cwo [NB: ***CARE!*** either or both can be obtained from wrong equns.] {note for scoris zoning – (i) to be visible in marking (ii)} **Examiner's Comments** Placing a question about finding normal parameters in the context of a frequency distribution did not worry most candidates, though some inevitably attempted to use a factor of √100 in the standard deviation. Signs were generally well handled, and the correct answers were often seen. Only a few used 0.17 or 0.25 in their equations, but continuity corrections in this context continue to be seen quite often. If the distribution is stated to be normal then a continuity correction must not be used.	

		iii	Based on a sample / small sample, etc	B1	Any similar comment, e.g. “frequencies not probabilities” (but not *just* “*n* is small”) *and* no wrong comments. Not “because data is grouped”. No scattergun. **Examiner's Comments** Many candidates realised that the probabilities were based on only a sample rather than on the whole population. However, there were also many who attempted to use a familiar answer to a different question, namely the routine answer to S1 questions about why calculations of sample mean and variance were not exact: “you don't know the exact data values, only the ranges”. Others said “it's only approximately a normal”, even though it was clearly stated in the question that the distribution *was* normal.	The Normal Distribution
			Total	7		
2			$\frac{150-\mu}{\sigma}=2.00$	M1	Standardise with σ, μ at least once, ignore cc, √ errors, equate to z	z not used, e.g. equated to 0.0228 and 0.9332 or 0.5092 and 0.8246: max M0M1
			$\frac{143-\mu}{\sigma}=-1.5$	A1	Both LHS and signs of RHS correct	One z, one not: M1A0B0
				B1	Both z-values correct to 3 SF	
				M1	Correct method for solution	Withhold if elimination done wrongly
			Solve to get	A1	μ ∈ [145.95, 146.05) www	√ σ or σ^2: can get M1A0B1M1A1A0
			μ = 146, σ = 2	A1	μ ∈ [1.995, 2.005) or $\mu^2 = 4$ www **Examiner's Comments** A very confident start to the paper by many. Fully correct answers were common, though inevitably there were some who made sign errors or who failed to use the tables in reverse.	cc: M1A0B1M1A0A0
			Total	6		

3			$\frac{1.03-1.00}{\sigma}=1.645$	M1 dep*	Standardise and equate to Φ^{-1}, allow wrong sign, σ^2, 1−, cc etc	The Normal Distribution
				A1	All correct apart possibly from value of Φ^{-1}	
				B1	1.645 seen anywhere, allow −1.645, can be implied	
			[σ = 0.0182… ≈ $^{6/329}$]	*M1	Solve to find σ, or eliminate σ, dependent on first M1	
			$1-\Phi\left(\frac{1.05-1.03}{\sigma}\right)=1-\Phi(1.0966)$	M1	Standardise with μ = 1.03, use Φ, answer < 0.5, allow $\sqrt{}$ errors	
					Final answer in range [0.1355, 0.137] or [13.55%, 13.7%], must be from positive σ, not from σ^2 0.1333 from σ = 0.018 is 5+A0	
			= 1 − 0.8635 = **0.1365** or 13.6(5)%	A1	**Examiner's Comments** There were many fully correct answer to this question, with only a few making the usual mistakes such as sign errors, use of σ^2 instead of σ, or 0.05 instead of 1.645. 13.6(5)%.	
			Total	6		
4		a	N (178, 8^2) and X< 194 oe	M1(AO1.1)	Soi	
			P(X<194) = 0.977(249868...)	A1(AO1.1)	BC	
			$0.977249868...^3$ = 0.933 (3 s.f.)	A1(AO1.1) [3]		
		b	E.g. |inflection −mean| E.g. $\frac{1}{2}$(97.5th percentile – mean)	M1(AO1.1a)	E.g. 170 – 163 E.g. $\frac{1}{2}$(176 – 163) — Figures are illustrative only	

			E.g. $\frac{1}{6}$(99.7th percentile – 0.3th percentile) = 6 to 7 E.g. Point of inflection is 1 sd from mean E.g. 95% of values within (approx) 2 sds of mean E.g. Almost all within (approx) 3 sds of mean	A1(AO1.1) E1(AO2.4) [3]	E.g. $\frac{1}{6}$(183 – 145) Statement matching method used	The Normal Distribution
			Total	6		
5		a	Symmetrical, high in middle, tails off at ends	B1(AO2.4) [1]	Any two of these | Not just bell shaped	
		b	a. P(35 < m < 45) = 0.296 Predicted no. = 30	M1(AO3.4) A1(AO1.1) [2]	Correct probability attempted Allow 29.6 or '29 or 30'	
		c	b. P(m < 25) = 0.0122 Predicted no. = 1	M1(AO3.4) A1(AO1.1) [2]	Correct probability attempted Allow 1.2 or '1 or 2'	
		d	29.6 close to 29 and 1.2 close to 0 Hence model (could be) suitable	B1(AO3.5a) [1]	Both needed | **OR B1** Model predicts some masses below 25 g, hence not suitable	
		e	E.g. Weather may cause different distribution	B1(AO3.5b) [1]	Any sensible reason why next year may be different	
			Total	7		

6		a	14.9 (3 sf)	B1(AO 1.1) [1]	BC	The Normal Distribution
		b	0.975 seen or implied $c = 27.8$, $b = 12.2$ (3 sf)	M1(AO 1.1a) A1(AO 1.1) [2]	Other solutions are possible	
			Total	3		
7			Assume masses normally distr. 66% of masses lie approximately within $\mu \pm \sigma$ and greatest mass $\approx \mu \pm 3\sigma$ sd = 0.5(120 – 96) (= 12) Greatest mass = 120 + 2×12 = 144 (g)	B1(AO 1.2) M1(AO 3.1b) M1(AO 1.1) A1(AO 1.1) [4]	or similar both stated or implied or greatest mass $= \frac{96+120}{2} + 3\times12 = 144$ Allow 144 to 145	
			Total	4		
8		a	$\mu = 56$ Percentage with masses > 59 g = 30%	B1(AO 1.1a) B1(AO 1.1) [2]	or 0.3	

		b	$\Phi\left(\frac{53-56}{\sigma}\right)=0.3,\ \frac{53-56}{\sigma}=-0.5244$ $\sigma = 5.721$ $X \sim N(56, '5.721'^2)$ soi $P(X > 65) = 0.0578$ or 5.78% (3 sf)	M1(AO 2.1) A1(AO 1.1) M1(AO 2.4) A1(AO 1.1) [4]	or $P(X > 65) = P(z > \frac{65-56}{5.721})$ $=P(z > 1.573)$ Or **BC** — ft their σ	The Normal Distribution
		c	$P(X < 50) = 0.1471$ $P(X < a) = 0.0471)$ $a = 46.4$ (3 sf)	M1(AO 1.1a) A1(AO 2.1) A1(AO 1.1) [3]		
			Total	9		
9		a	(i) 0.0478 or 0.048 (2 sf) (ii) 22.5 or 23 (2 sf)	B1 (AO 1.1) [1] B1	BC Examiner's Comments Most candidates answered this question correctly. A few used a standard deviation of 9.	

				(AO 1.1) [1]	BC **Examiner's Comments** Some candidates found $\Phi^{-1}(0.2) = 17.5$. A few used a standard deviation of 9	The Normal Distribution
			(iii) $P(X < 20 + b) = 0.75$ or $P(X > 20 + b) = 0.25$ $20 + b = 22.02..$ or 22.0 or 22 $b = 2.02$ or 2.0 (2 sf) Allow $b = 2$	M1 (AO 1.1a) A1 (AO 1.1) A1 (AO 1.1) [3]	$P(X < 20 - b) = 0.25$ $20 - b = 17.98$ or 18 $b = 22(.02)$ M1A1A0 T & I method: Try 2 values, one ≈ 2 M1 Correct probs for two values in [2, 2.1] A1 Correct probs for two values in [2, 2.05] & ans 2.0 or 2 A1 (0.495 & 0.516) **Examiner's Comments** Many candidates could not make the first step, which is to move from the given probability of 0.5 to a probability of either 0.25 or 0.75	

		b	$\frac{1.5\mu-\mu}{\mu/3}$ $=\frac{3}{2}$ P(X> 1.5μ) = 0.0668 or 0.67 (2 sf)	M1 (AO 1.1a) A1 (AO 1.1) A1 (AO 1.1) [3]	$\frac{4.5\sigma-3\sigma}{\sigma}$ | SC (eg) Let μ = 1; N(1, $\frac{1}{9}$) M1 | $X=\frac{3}{2}$ A0 | $P(X>\frac{3}{2})=0.067$ A1 __Examiner's Comments__ Many candidates chose a value for μ, without justification. These could score two out of the three marks. Many candidates used a standard deviation of $\frac{\mu^2}{9}$ and so found that μ did not cancel out. Some could not handle the simple algebra involved in substituting X = 1.5μ into $\frac{X-\mu}{\frac{\mu}{3}}$.	The Normal Distribution
			Total	8		
10		a	DR N(450×0.15, 450×0.15×0.85) or N(67.5, 57.375) oe $P(Y>\mu+\sigma)\approx\frac{1}{6}$ or $\Phi^{-1}(\frac{1}{6})=0.9674$	M1 (AO 3.1b)	seen or implied $P(Y<a)=\frac{5}{6}$ | B(450, 0.15) __with__ T & I method using ≥ one of 74, 75, 76, 61, 60, 59 P(X> 74) = 0.177	

			$'67.5' + \sqrt{'57.375'}$ or $'67.5' + 0.9674\times\sqrt{'57.375'}$ = 74 or 75 or 76	M1 (AO 1.2) A1 (AO 1.1) [3]	<table><tr><td>or 74.83 seen; ft their μ & σ for M1 only Integer. No ft Dep M1M1 Correct ans, inadequate wking: M0M0A0 NB 450/6 = 75 M0M0A0</td><td>$P(X > 75) = 0.145$ both a = 74 or 75 or 76</td></tr></table> <u>Examiner's Comments</u> Because this question required "detailed reasoning", correct answers did not necessarily score full marks. Thus, for example, some candidates gave the following working: $X \sim B(450. 0.15)$; $P(X < a) = \frac{5}{6}$; $X = 75$. These scored only one mark. Trial and improvement methods only scored marks if they were very clearly explained, with the distribution fully described <u>and</u> with at least two values close to 75 being tried, with the relevant probabilities actually seen. The better method was to use the normal approximation to the binomial and the fact that approximately $\frac{2}{3}$ of values lie within the range $\mu - \sigma < X < \mu + \sigma$.	The Normal Distribution

		b	$\dfrac{\frac{50!}{r!(50-r)!}\times 0.15^r \times 0.85^{50-r}}{\frac{50!}{(r+1)!(50-(r+1))!}\times 0.15^{r+1}\times 0.85^{50-(r+1)}}$ oe eg $\dfrac{\frac{1}{50-r}\times 0.85}{\frac{1}{r+1}\times 0.15}$ or $\dfrac{0.85}{50-r}\times\dfrac{r+1}{0.15}$ oe $=\dfrac{17(r+1)}{3(50-r)}$ AG	M1 (AO 1.1a) A1 (AO 2.1) A1 (AO 1.1) [3]	$\dfrac{^{50}C_r \times 0.15^r \times 0.85^{50-r}}{^{50}C_{r+1}\times 0.15^{r+1}\times 0.85^{50-(r+1)}}$ — Fully correct Any correct simplification without factorials **OR** without indices — **or** $\dfrac{17}{20}\times\dfrac{20}{3}\times\dfrac{r+1}{50-r}$ Any correct simplification without factorials **AND** without indices and correctly obtain result __Examiner's Comments__ Some candidates appeared not to understand the definition of T_r and T_{r+1}. Many candidates omitted the relevant powers of 0.85 in their expressions for T_r and T_{r+1}. Some wrote the correct expressions for both T_r and T_{r+1}, but were unable to find any common factors to cancel. Many attempted to expand the given binomial expression, but did not include a general term. Some candidates "fudged" the answers giving, for example, T_r = 0.85(r + 1) and T_{r+1} = 0.15(50 – r) etc.	The Normal Distribution
		c	(i) $\dfrac{17(r+1)}{3(50-r)}\le 1$ oe $17r + 17 \le 150 - 3r$ $20r \le 133$ oe	M1 (AO 3.1b) A1 (AO 1.1)	$\dfrac{1}{50-r}\times 0.85 \le \dfrac{1}{r+1}\times 0.15$ oe M1 — No factorials or indices $0.85(r + 1) \le 0.15(50 - r)$ $r \le 50\times0.15 - 0.85$ A1 — Correct, in form $ar \le b$ or	

			$r \leq 6.65$ r is an integer so $r \leq 6$	A1 (AO 1.1) A1 (AO 1.1) [4]	SC: P(X=6)=0.142, P(X=7)=0.157, P(X=8)=0.149 B1 (must be these three) hence r ≤ 6 B1dep | r < correct expr'n No wking B0B0 **Examiner's Comments** Many candidates did not see the connection with part (b). These went back to the binomial distribution but very few succeeded. Some candidates carried out a correct method but stopped after obtaining $r \leq 6.65$. Others found $r \leq 6.65$ but then gave the answer $r = 6$. A trial and improvement method could score a maximum of 2 marks in this question.	The Normal Distribution

			(ii) $P(X = r) \leq P(X = r + 1)$ for $r \leq 6$ Hence most likely value is r is 6 or 7 $\frac{P(X=6)}{P(X=7)} = \frac{17(6+1)}{3(50-6)} = 0.902 < 1$ Most likely value is 7	B1 (AO 2.1) B1 (AO 3.2a) [2]	or $P(X = 6) = 0.142$ & $P(X = 7) = 0.157$ — NOT 6.65 rounds to 7 B0B0 indep, but dep on some reasonable explanation — No expl'n: B0B0 **Examiner's Comments** Almost no candidates gave a correct solution based on their answer to part (c)(i). Some used trial and improvement, but did not consider enough values of X. (At least $X = 6$ and 7 were required). Some rounded their figure of 6.65 from part (c)(i) to 7. This did not score any marks.	The Normal Distribution
			Total	12		
11		a	Total area = 500 small squares $\frac{100}{500} \times 300$ = 60 (days)	B1 (AO1.2) M1 (AO1.1a) A1 (AO1.1) [3]	or 20 cm^2 or other units or $\frac{4}{20} \times 300$ or equivalent — May be implied	
		b	$\frac{3\times15+5\times25+1\times15}{500} \times 300$ = 111	M1 (AO2.1) M1 (AO1.1) A1 (AO1.1) [3]	M1 for denom & one term in num M1 M1 for correct × 300 — oe in other units	

		c	Frequencies: 30, 90, 75, 45, 60 $\frac{\Sigma fx}{300} = \frac{982500}{300}$ (= 3275 **AG**)	M1 (AO3.1a) A1 (AO1.1) [2]	Allow multiples of these correctly obtain 3275	The Normal Distribution
		d	P(4000 < x < 6000) × 300 = 0.2419 × 300 = 72.57 so 73 days	M1 (AO3.4) A1 (AO1.1) [2]	Attempted, using N(3275, 1060^2) BC accept truncation to 72 days	
		e	3270 + 2 × 1060 = 5395 In histogram, well over 2.5% of values are above 5395, so model not a good fit	B1 (AO3.4) E1 (AO2.2b) [2]		
			Total	12		
12		a	(i) X is binomial (ii) Large n	B1 (AO 3.3) [1] B1 (AO 3.3) [1]		
		b	$X \sim \text{N}(\frac{500}{3}, \frac{1250}{9})$ e.g. P(X < b) = 0.7 b = 173 or 174 $a = \frac{500}{3} - ((173 \text{ or } 174) - \frac{500}{3})$ = 160 or 159	B1 (AO 1.2) M1 (AO 3.4) A1 (AO 1.1) M1 (AO 3.4) A1 (AO 1.1) [5]	soi; allow N(167, 139) **BC** or P(X < a) = 0.3 Other correct methods score similarly eg $\Phi^{-1}(0.9)$ = 181 $\Phi^{-1}(0.5)$ = 166	

		c	$X \sim \text{Bin}(1000, \frac{1}{6})$ <table><tr><td>eg P(160 < X < 173)</td><td>or P(159 < X < 174)</td></tr><tr><td>= 0.69218 – 0.30280</td><td>or 0.72108 – 0.27355</td></tr><tr><td>= 0.389 (3 sf)</td><td>or 0.448 (3 sf)</td></tr></table>	M1 (AO 3.4) A1 (AO 1.1) [2]	<table><tr><td>BC NB ft their a and b</td><td>P(166 < X < 181) 0.87910 – 0.49812 = 0.381 (3 sf)</td></tr></table>	The Normal Distribution
			Total	9		