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

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

0.0478

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

22.5

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

$P(20-b < X) = 0.25$

$\frac{20-b-20}{3} = 0.674 \quad b = 2.023$

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

$P\left(\frac{1.5\mu-\mu}{\frac{\mu}{3}}\right) = P(1.5)$ $\quad P(x > 1.5\mu)$

$= 0.067$

2 Briony suspects that a particular 6-sided dice is biased in favour of 2. She plans to throw the dice 35 times and note the number of times that it shows a 2. She will then carry out a test at the 4% significance level. Find the rejection region for the test. **[7]**

$X \sim B(35, \frac{1}{6})$

$P(n \geq x) \leq 0.04$ using trial and improvement

$n = 11$

$1 - P(x \leq n) \leq 0.04$

$P(x \leq n) \geq 0.96$

3

A certain forest contains only trees of a particular species. Dipak wished to take a random sample of 5 trees from the forest. He numbered the trees from 1 to 784. Then, using his calculator, he generated the random digits 14781049. Using these digits, Dipak formed 5 three-digit numbers. He took the first, second and third digits, followed by the second, third and fourth digits and so on. In this way he obtained the following list of numbers for his sample.

147 478 781 104 49

(a) Explain why Dipak omitted the number 810 from his list. **[1]**

No tree has this number

(b) Explain why Dipak's sample is not random. **[1]**

Each tree is not equally likely to be chosen

(c) Carry out the test at the 2% significance level. [7]

$H_0: \mu = 4.2$ $\bar{X} \sim N(4.2, \frac{0.8^2}{58})$

$H_1: \mu < 4.2$ $P(\bar{X} < 4)$

$= 0.0385$

$0.0385 > 0.02$

$\therefore$ don't reject H_0 as there is insufficient evidence to suggest the mean is different at a 2% level

4 Christa used Pearson's product-moment correlation coefficient, r, to compare the use of public transport with the use of private vehicles for travel to work in the UK.

(a) Using the pre-release data set for all 348 UK Local Authorities, she considered the following four variables.

Number of employees using public transport	x
Number of employees using private vehicles	y
Proportion of employees using public transport	a
Proportion of employees using private vehicles	b

(i)Explain, in context, why you would expect strong, positive correlation between x and y. [1]

As the number of employees for a company increases there will be more people going to work $\therefore$ more of both methods used

strong negative as decreasing a increases b eg due to higher paid jobs $\therefore$ people can afford b

(ii)Suggest what effect this outlier is likely to have on the value of r for the whole country. [1]

decrease

(iii) What can you deduce about the area of the London Borough represented by the outlier? Explain your answer. [1]

No effect

a greater proportion use public transport compared to the expected proportion driving $\therefore$ has very good public transport links

5 The discrete random variable X takes values 1, 2, 3, 4 and 5, and its probability distribution is defined as follows.

$$\mathrm{P}(X=x)=\begin{cases} a & x=1, \\ \frac{1}{2}\mathrm{P}(X=x-1) & x=2,3,4,5, \\ 0 & \text{otherwise,} \end{cases}$$

where a is a constant.

(a) Show that $a=\frac{16}{31}$. **[2]**

$a+\frac{a}{2}+\frac{a}{4}+\frac{a}{8}+\frac{a}{16}=1 \quad \frac{31a}{16}=1 \quad a=\frac{16}{31}$

The discrete probability distribution for X is given in the table.

x	1	2	3	4	5
$\mathrm{P}(X=x)$	$\frac{16}{31}$	$\frac{8}{31}$	$\frac{4}{31}$	$\frac{2}{31}$	$\frac{1}{31}$

(b) Find the probability that X is odd. **[1]**

$P(X=1)+P(X=3)+P(X=5)=\frac{21}{31}$

Two independent values of X are chosen, and their sum S is found.

(c) Find the probability that S is odd. **[2]**

2 × odd then even

$=2\times\frac{21}{31}\times\frac{10}{31}=0.437$

(d) Find the probability that S is greater than 8, given that S is odd. **[3]**

$\rightarrow 4,5$ or $5,4$ $\quad \frac{2\times\frac{1}{31}\times\frac{2}{31}}{0.437}=\frac{1}{105}$

Sheila sometimes needs several attempts to start her car in the morning. She models the number of attempts she needs by the discrete random variable Y defined as follows.

$$\mathrm{P}(Y=y+1)=\frac{1}{2}\mathrm{P}(Y=y) \quad \text{for all positive integers } y.$$

(e) Find $\mathrm{P}(Y=1)$. **[2]**

$y+\frac{y}{2}+\frac{y}{4}+\frac{y}{8}\ldots$ $\quad r=0.5$ $\quad a=P(Y=1)$ $\quad \frac{a}{1-r}=0.5$ $\quad a=0.25=P(Y=1)$

(f) Give a reason why one of the variables, X or Y, might be more appropriate as a model for the number of attempts that Sheila needs to start her car. **[1]**

Y as no max attempts

6 **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 $P(Y > a) \approx \frac{1}{6}$. **[3]**

$Y \sim N(450 \times 0.15, 450 \times 0.15 \times 0.85$

$\sim N(67.5, 57.375),$

$P(Y < a) = \frac{5}{6}$

$\therefore a = 75$

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 $\frac{T_r}{T_{r+1}} = \frac{17(r+1)}{3(50-r)}$. **[3]**

$$\frac{\frac{50!}{r!(50-r)!} \times 0.15^r \times 0.85^{50-r}}{\frac{50!}{(r+1)!(50-(r+1))!} \times 0.15^{r+1} \times 0.85^{50-(r+1)}}$$

(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 $P(X = r) \leqslant P(X = r+1)$. **[4]**

$\frac{17(r+1)}{3(50-r)} \leq 1$ so $20r \leq 133$

$r \leq 6.65$

so $\underline{r \leq 6}$

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

Most likely is 6 or 7 and

$P(X=6) = 0.142$, $P(X=7) = 0.15750$

7 is most likey.

Total Marks for Question Set 2: 53 Marks