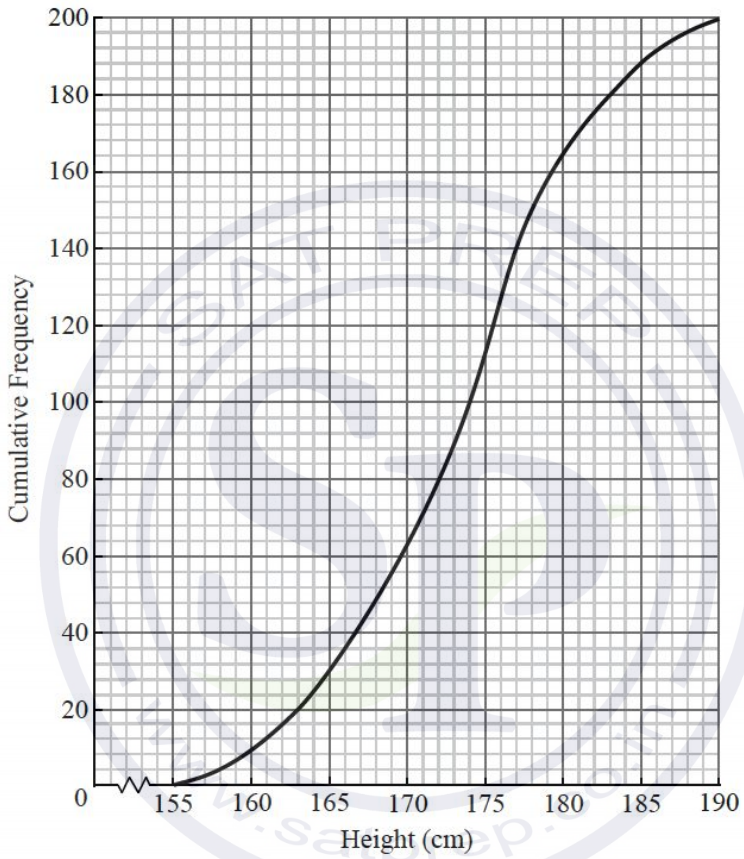
Subject – Math (Standard Level)
Topic - Statistics and Probability
Year - Nov 2011 – Nov 2019
Paper -2

Question 1

[Maximum mark: 6]

The cumulative frequency curve below represents the heights of 200 sixteen-year-old boys.



Use the graph to answer the following.

(a) Write down the median value. *[1 mark]*

(b) A boy is chosen at random. Find the probability that he is shorter than 161 cm. *[2 marks]*

(c) Given that 82 % of the boys are taller than h cm, find h. *[3 marks]*

Question 2

[Maximum mark: 16]

A company produces a large number of water containers. Each container has two parts, a bottle and a cap. The bottles and caps are tested to check that they are not defective.

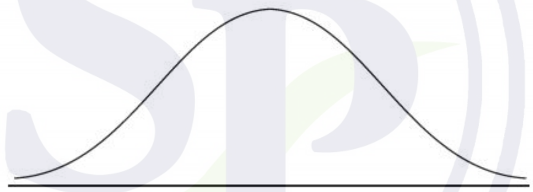
A cap has a probability of 0.012 of being defective. A random sample of 10 caps is selected for inspection.

(a) Find the probability that exactly one cap in the sample will be defective. *[2 marks]*

(b) The sample of caps passes inspection if at most one cap is defective. Find the probability that the sample passes inspection. *[2 marks]*

The heights of the bottles are normally distributed with a mean of 22 cm and a standard deviation of 0.3 cm.

(c) (i) **Copy** and complete the following diagram, shading the region representing where the heights are less than 22.63 cm.



(ii) Find the probability that the height of a bottle is less than 22.63 cm. *[5 marks]*

(d) (i) A bottle is accepted if its height lies between 21.37 cm and 22.63 cm. Find the probability that a bottle selected at random is accepted.

(ii) A sample of 10 bottles passes inspection if all of the bottles in the sample are accepted. Find the probability that the sample passes inspection. *[5 marks]*

(e) The bottles and caps are manufactured separately. A sample of 10 bottles and a sample of 10 caps are randomly selected for testing. Find the probability that both samples pass inspection. *[2 marks]*

Question 3

[Maximum mark: 7]

The probability of obtaining "tails" when a biased coin is tossed is 0.57. The coin is tossed ten times. Find the probability of obtaining

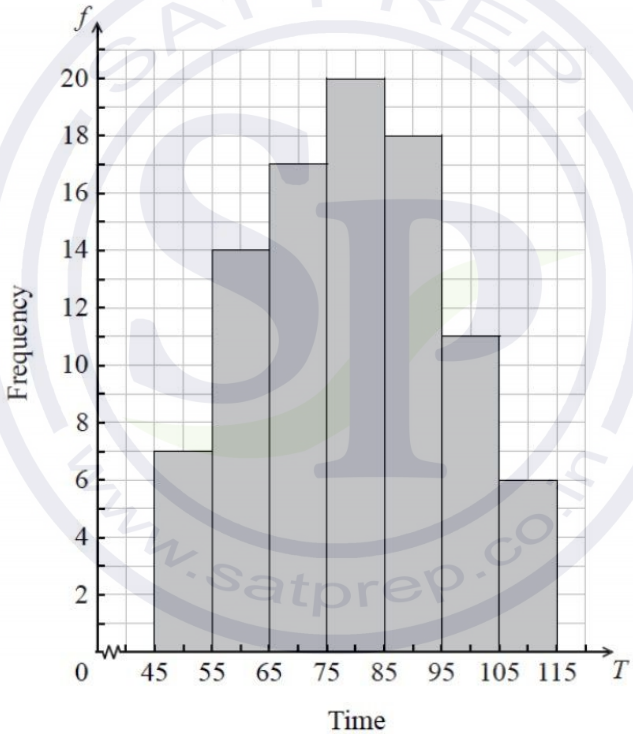
(a) **at least** four tails; *[4 marks]*

(b) the fourth tail on the tenth toss. *[3 marks]*

Question 4

[Maximum mark: 13]

The histogram below shows the time T seconds taken by 93 children to solve a puzzle.



The following is the frequency distribution for T.

Time	$45 \le T < 55$	$55 \le T < 65$	$65 \le T < 75$	$75 \le T < 85$	$85 \le T < 95$	$95 \le T < 105$	$105 \le T < 115$
Frequency	7	14	p	20	18	q	6

(a) (i) Write down the value of p and of q.

(ii) Write down the median class. *[3 marks]*

(b) A child is selected at random. Find the probability that the child takes less than 95 seconds to solve the puzzle. *[2 marks]*

Consider the class interval $45 \leq T < 55$.

(c) (i) Write down the interval width.

(ii) Write down the mid-interval value. *[2 marks]*

(d) Hence find an estimate for the

(i) mean;

(ii) standard deviation. *[4 marks]*

John assumes that T is normally distributed and uses this to estimate the probability that a child takes less than 95 seconds to solve the puzzle.

(e) Find John's estimate. *[2 marks]*

Question 5

[Maximum mark: 6]

The heights of a group of seven-year-old children are normally distributed with mean 117 cm and standard deviation 5 cm. A child is chosen at random from the group.

(a) Find the probability that this child is taller than 122.5 cm. *[3 marks]*

(b) The probability that this child is shorter than k cm is 0.65. Find the value of k. *[3 marks]*

Question 6

[Maximum mark: 8]

A factory makes lamps. The probability that a lamp is defective is 0.05. A random sample of 30 lamps is tested.

(a) Find the probability that there is at least one defective lamp in the sample. *[4 marks]*

(b) Given that there is at least one defective lamp in the sample, find the probability that there are at most two defective lamps. *[4 marks]*

Question 7

[Maximum mark: 8]

In a large city, the time taken to travel to work is normally distributed with mean μ and standard deviation σ. It is found that 4 % of the population take less than 5 minutes to get to work, and 70 % take less than 25 minutes.

Find the value of μ and of σ.

Question 8

[Maximum mark: 15]

At a large school, students are required to learn at least one language, Spanish or French. It is known that 75 % of the students learn Spanish, and 40 % learn French.

(a) Find the percentage of students who learn **both** Spanish and French. *[2 marks]*

(b) Find the percentage of students who learn Spanish, but not French. *[2 marks]*

At this school, 52 % of the students are girls, and 85 % of the girls learn Spanish.

(c) A student is chosen at random. Let G be the event that the student is a girl, and let S be the event that the student learns Spanish.

(i) Find $\mathrm{P}(G \cap S)$.

(ii) Show that G and S are **not** independent. *[5 marks]*

(d) A boy is chosen at random. Find the probability that he learns Spanish. *[6 marks]*

Question 9

[Maximum mark: 6]

Consider the following cumulative frequency table.

x	**Frequency**	**Cumulative frequency**
5	2	2
15	10	12
25	14	26
35	p	35
45	6	41

(a) Find the value of p. *[2 marks]*

(b) Find

(i) the mean;

(ii) the variance. *[4 marks]*

Question 10

[Maximum mark: 7]

A random variable X is normally distributed with $\mu = 150$ and $\sigma = 10$.

Find the interquartile range of X.

Question 11

[Maximum mark: 6]

The random variable X is normally distributed with mean 20 and standard deviation 5.

(a) Find $P(X \le 22.9)$. *[3 marks]*

(b) Given that $P(X < k) = 0.55$, find the value of k. *[3 marks]*

Question 12

[Maximum mark: 15]

A bag contains four gold balls and six silver balls.

(a) Two balls are drawn at random from the bag, with replacement. Let X be the number of gold balls drawn from the bag.

(i) Find $\mathrm{P}(X = 0)$.

(ii) Find $\mathrm{P}(X = 1)$.

(iii) Hence, find $\mathrm{E}(X)$. *[8 marks]*

Fourteen balls are drawn from the bag, with replacement.

(b) Find the probability that exactly five of the balls are gold. *[2 marks]*

(c) Find the probability that at most five of the balls are gold. *[2 marks]*

(d) Given that at most five of the balls are gold, find the probability that exactly five of the balls are gold. Give the answer correct to two decimal places. *[3 marks]*

Question 13

[Maximum mark: 6]

Two events A and B are such that $\mathrm{P}(A) = 0.2$ and $\mathrm{P}(A \cup B) = 0.5$.

(a) Given that A and B are mutually exclusive, find $\mathrm{P}(B)$. *[2]*

(b) Given that A and B are independent, find $\mathrm{P}(B)$. *[4]*

Question 14

[Maximum mark: 7]

The time taken for a student to complete a task is normally distributed with a mean of 20 minutes and a standard deviation of 1.25 minutes.

(a) A student is selected at random. Find the probability that the student completes the task in less than 21.8 minutes. *[2]*

(b) The probability that a student takes between k and 21.8 minutes is 0.3. Find the value of k. *[5]*

Question 15

[Maximum mark: 14]

Samantha goes to school five days a week. When it rains, the probability that she goes to school by bus is 0.5. When it does not rain, the probability that she goes to school by bus is 0.3. The probability that it rains on any given day is 0.2.

(a) On a randomly selected school day, find the probability that Samantha goes to school by bus. *[4]*

(b) Given that Samantha went to school by bus on Monday, find the probability that it was raining. *[3]*

(c) In a randomly chosen school week, find the probability that Samantha goes to school by bus on exactly three days. *[2]*

(d) After n school days, the probability that Samantha goes to school by bus at least once is greater than 0.95. Find the smallest value of n. *[5]*

Question 16

[Maximum mark: 7]

The following table shows the average weights (y kg) for given heights (x cm) in a population of men.

Heights (x cm)	165	170	175	180	185
Weights (y kg)	67.8	70.0	72.7	75.5	77.2

(a) The relationship between the variables is modelled by the regression equation $y = ax + b$.

(i) Write down the value of a and of b.

(ii) Hence, estimate the weight of a man whose height is 172 cm. *[4]*

(b) (i) Write down the correlation coefficient.

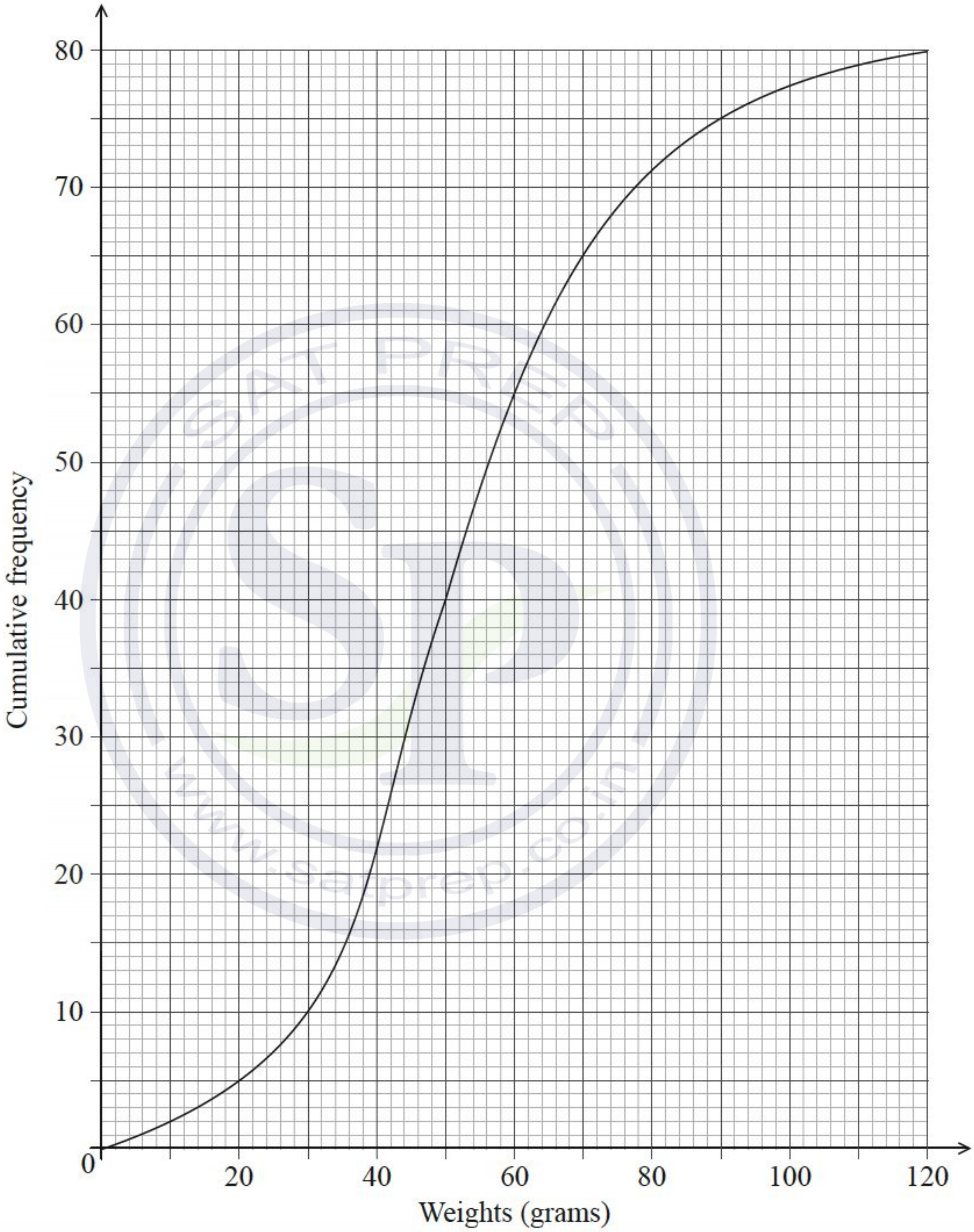
(ii) State which **two** of the following describe the correlation between the variables. *[3]*

strong	zero	positive
negative	no correlation	weak

Question 17

[Maximum mark: 16]

The weights in grams of 80 rats are shown in the following cumulative frequency diagram.



(Question 8 continued)

(a) (i) Write down the median weight of the rats.

(ii) Find the percentage of rats that weigh 70 grams or less. *[4]*

The same data is presented in the following table.

Weights w grams	$0 \le w \le 30$	$30 < w \le 60$	$60 < w \le 90$	$90 < w \le 120$
Frequency	p	45	q	5

(b) (i) Write down the value of p.

(ii) Find the value of q. *[4]*

(c) Use the values from the table to estimate the mean and standard deviation of the weights. *[3]*

Assume that the weights of these rats are normally distributed with the mean and standard deviation estimated in part (c).

(d) Find the percentage of rats that weigh 70 grams or less. *[2]*

(e) A sample of five rats is chosen at random. Find the probability that at most three rats weigh 70 grams or less. *[3]*

Question 18

[Maximum mark: 5]

The following table shows the amount of fuel (y litres) used by a car to travel certain distances (x km).

Distance (x km)	40	75	120	150	195
Amount of fuel (y litres)	3.6	6.5	9.9	13.1	16.2

This data can be modelled by the regression line with equation $y = ax + b$.

(a) (i) Write down the value of a and of b.

(ii) Explain what the gradient a represents. *[3]*

(b) Use the model to estimate the amount of fuel the car would use if it is driven 110 km. *[2]*

Question 19

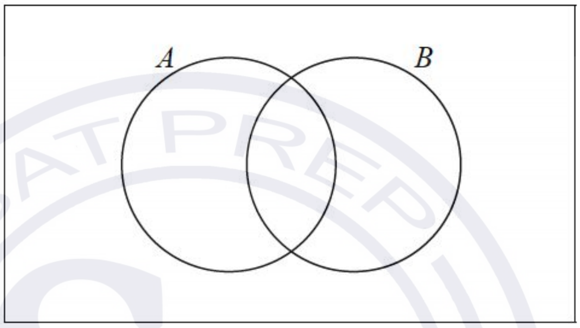
[Maximum mark: 7]

Let A and B be independent events, where $\mathrm{P}(A)=0.3$ and $\mathrm{P}(B)=0.6$.

(a) Find $\mathrm{P}(A \cap B)$. *[2]*

(b) Find $\mathrm{P}(A \cup B)$. *[2]*

(c) (i) On the following Venn diagram, shade the region that represents $A \cap B'$.



(ii) Find $\mathrm{P}(A \cap B')$. *[3]*

Question 20

[Maximum mark: 14]

A forest has a large number of tall trees. The heights of the trees are normally distributed with a mean of 53 metres and a standard deviation of 8 metres. Trees are classified as giant trees if they are more than 60 metres tall.

(a) A tree is selected at random from the forest.

(i) Find the probability that this tree is a giant.

(ii) Given that this tree is a giant, find the probability that it is taller than 70 metres. *[6]*

(b) Two trees are selected at random. Find the probability that they are both giants. *[2]*

(c) 100 trees are selected at random.

(i) Find the expected number of these trees that are giants.

(ii) Find the probability that at least 25 of these trees are giants. *[6]*

Question 21

[Maximum mark: 6]

The following table shows the Diploma score x and university entrance mark y for seven IB Diploma students.

Diploma score (x)	28	30	27	31	32	25	27
University entrance mark (y)	73.9	78.1	70.2	82.2	85.5	62.7	69.4

(a) Find the correlation coefficient. *[2]*

The relationship can be modelled by the regression line with equation $y = ax + b$.

(b) Write down the value of a and of b. *[2]*

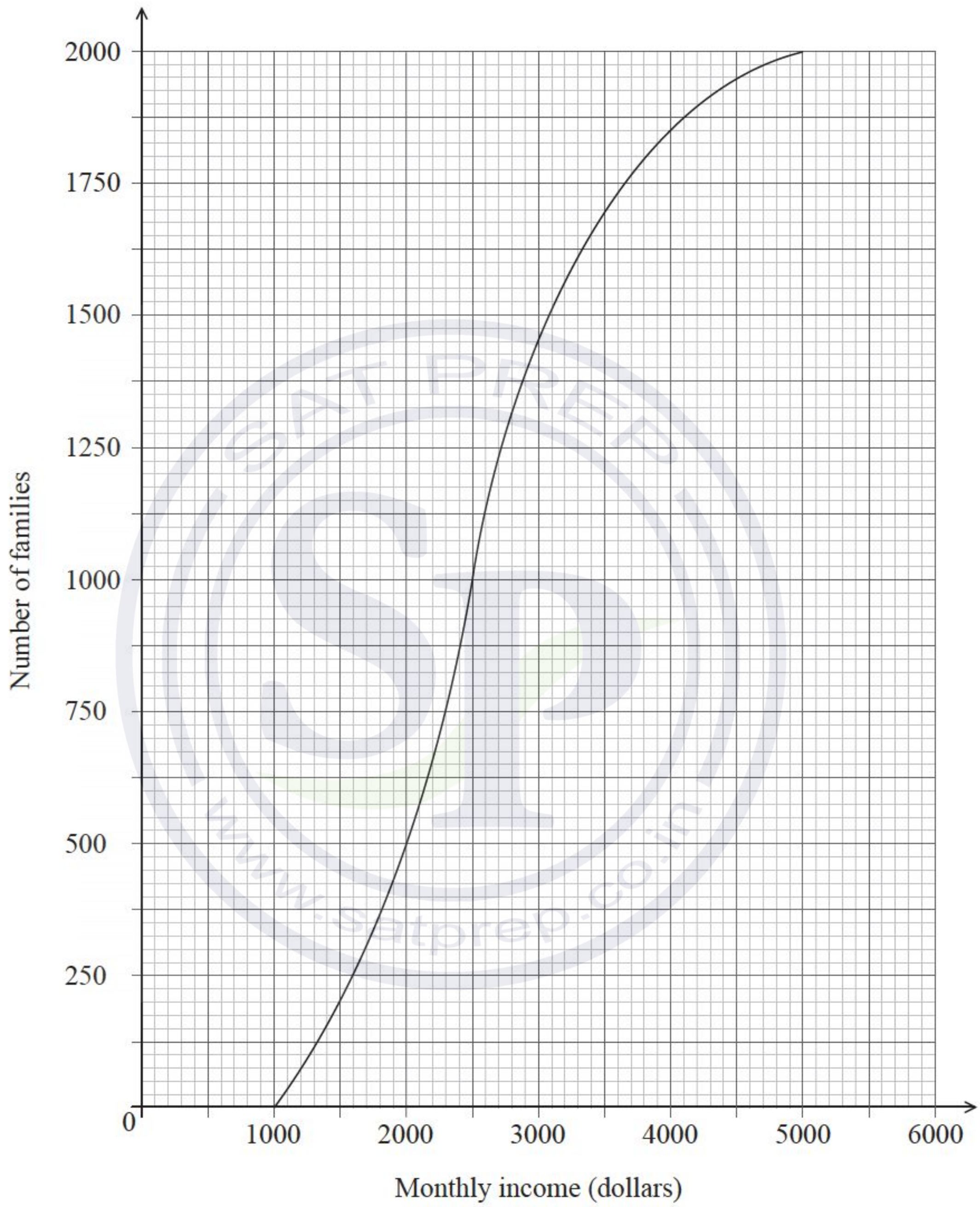
Rita scored a total of 26 in her IB Diploma.

(c) Use your regression line to estimate Rita's university entrance mark. *[2]*

Question 22

[Maximum mark: 15]

The following cumulative frequency graph shows the monthly income, I dollars, of 2000 families.



(a) Find the median monthly income. *[2]*

(b) (i) Write down the number of families who have a monthly income of 2000 dollars or less.

(ii) Find the number of families who have a monthly income of more than 4000 dollars. *[4]*

The 2000 families live in two different types of housing. The following table gives information about the number of families living in each type of housing and their monthly income I.

	$1000 < I \leq 2000$	$2000 < I \leq 4000$	$4000 < I \leq 5000$
Apartment	436	765	28
Villa	64	p	122

(c) Find the value of p. *[2]*

(d) A family is chosen at random.

(i) Find the probability that this family lives in an apartment.

(ii) Find the probability that this family lives in an apartment, given that its monthly income is greater than 4000 dollars. *[4]*

(e) Estimate the mean monthly income for families living in a villa. *[3]*

Question 23

[Maximum mark: 16]

The weights of fish in a lake are normally distributed with a mean of 760 g and standard deviation σ. It is known that 78.87% of the fish have weights between 705 g and 815 g.

(a) (i) Write down the probability that a fish weighs more than 760 g.

(ii) Find the probability that a fish weighs less than 815 g. *[4]*

(b) (i) Write down the standardized value for 815 g.

(ii) Hence or otherwise, find σ. *[4]*

A fishing contest takes place in the lake. Small fish, called tiddlers, are thrown back into the lake. The maximum weight of a tiddler is 1.5 standard deviations below the mean.

(c) Find the maximum weight of a tiddler. *[2]*

(d) A fish is caught at random. Find the probability that it is a tiddler. *[2]*

(e) 25% of the fish in the lake are salmon. 10% of the salmon are tiddlers. Given that a fish caught at random is a tiddler, find the probability that it is a salmon. *[4]*

Question 24

[Maximum mark: 7]

The following table shows the average number of hours per day spent watching television by seven mothers and each mother's youngest child.

Hours per day that a mother watches television (x)	2.5	3.0	3.2	3.3	4.0	4.5	5.8
Hours per day that her child watches television (y)	1.8	2.2	2.6	2.5	3.0	3.2	3.5

The relationship can be modelled by the regression line with equation $y = ax + b$.

(a) (i) Find the correlation coefficient.

(ii) Write down the value of a and of b. [4]

Elizabeth watches television for an average of 3.7 hours per day.

(b) Use your regression line to predict the average number of hours of television watched per day by Elizabeth's youngest child. Give your answer correct to one decimal place. [3]

Question 25

[Maximum mark: 16]

A company makes containers of yogurt. The volume of yogurt in the containers is normally distributed with a mean of $260\,\text{ml}$ and standard deviation of $6\,\text{ml}$.

A container which contains less than $250\,\text{ml}$ of yogurt is **underfilled**.

(a) A container is chosen at random. Find the probability that it is underfilled. [2]

The company decides that the probability of a container being underfilled should be reduced to 0.02. It decreases the standard deviation to σ and leaves the mean unchanged.

(b) Find σ. [4]

The company changes to the new standard deviation, σ, and leaves the mean unchanged. A container is chosen at random for inspection. It passes inspection if its volume of yogurt is between 250 and $271\,\text{ml}$.

(c) (i) Find the probability that it passes inspection.

(ii) Given that the container is **not** underfilled, find the probability that it passes inspection. [6]

(d) A sample of 50 containers is chosen at random. Find the probability that 48 or more of the containers pass inspection. [4]

Question 26

[Maximum mark: 6]

The following table shows the sales, y millions of dollars, of a company, x years after it opened.

Time after opening (x years)	2	4	6	8	10
Sales (y millions of dollars)	12	20	30	36	52

The relationship between the variables is modelled by the regression line with equation $y = ax + b$.

(a) (i) Find the value of a and of b.

(ii) Write down the value of r. [4]

(b) Hence estimate the sales in millions of dollars after seven years. [2]

Question 27

[Maximum mark: 16]

A machine manufactures a large number of nails. The length, L mm, of a nail is normally distributed, where $L \sim \mathrm{N}\left(50, \sigma^2\right)$.

(a) Find $\mathrm{P}(50-\sigma < L < 50+2\sigma)$. [3]

(b) The probability that the length of a nail is less than $53.92\,\mathrm{mm}$ is 0.975.
Show that $\sigma = 2.00$ (correct to three significant figures). [2]

All nails with length at least t mm are classified as large nails.

(c) A nail is chosen at random. The probability that it is a large nail is 0.75.
Find the value of t. [3]

(d) (i) A nail is chosen at random from the large nails. Find the probability that the length of this nail is less than $50.1\,\mathrm{mm}$.

(ii) Ten nails are chosen at random from the large nails. Find the probability that at least two nails have a length that is less than $50.1\,\mathrm{mm}$. [8]

Question 28

[Maximum mark: 5]

The following table shows the probability distribution of a discrete random variable X.

x	0	1	2	3
$\mathrm{P}(X=x)$	0.15	k	0.1	$2k$

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

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

Question 29

[Maximum mark: 7]

Let C and D be independent events, with $\mathrm{P}(C)=2k$ and $\mathrm{P}(D)=3k^2$, where $0<k<0.5$.

(a) Write down an expression for $\mathrm{P}(C\cap D)$ in terms of k. [2]

(b) Given that $\mathrm{P}(C\cap D)=0.162$, find k. [2]

(c) Find $\mathrm{P}(C'|D)$. [3]

Question 30

[Maximum mark: 16]

An environmental group records the numbers of coyotes and foxes in a wildlife reserve after t years, starting on 1 January 1995.

Let c be the number of coyotes in the reserve after t years. The following table shows the number of coyotes after t years.

number of years (t)	0	2	10	15	19
number of coyotes (c)	115	197	265	320	406

The relationship between the variables can be modelled by the regression equation $c = at + b$.

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

(b) Use the regression equation to estimate the number of coyotes in the reserve when $t = 7$. [3]

Let f be the number of foxes in the reserve after t years. The number of foxes can be modelled by the equation $f = \dfrac{2000}{1+99\mathrm{e}^{-kt}}$, where k is a constant.

(c) Find the number of foxes in the reserve on 1 January 1995. [3]

(d) After five years, there were 64 foxes in the reserve. Find k. [3]

(e) During which year were the number of coyotes the same as the number of foxes? [4]

Question 31

[Maximum mark: 14]

The masses of watermelons grown on a farm are normally distributed with a mean of $10\,\mathrm{kg}$. The watermelons are classified as small, medium or large.

A watermelon is small if its mass is less than $4\,\mathrm{kg}$. Five percent of the watermelons are classified as small.

(a) Find the standard deviation of the masses of the watermelons. [4]

The following table shows the percentages of small, medium and large watermelons grown on the farm.

small	medium	large
5%	57%	38%

A watermelon is large if its mass is greater than $w\,\mathrm{kg}$.

(b) Find the value of w. [2]

All the medium and large watermelons are delivered to a grocer.

(c) The grocer selects a watermelon at random from **this** delivery. Find the probability that it is medium. [3]

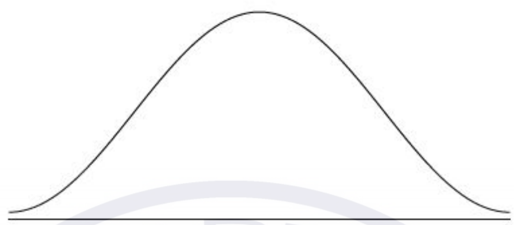
(d) The grocer sells all the medium watermelons for $\$1.75$ each, and all the large watermelons for $\$3.00$ each. His costs on this delivery are $\$300$, and his total profit is $\$150$. Find the number of watermelons in the delivery. [5]

Question 32

[Maximum mark: 6]

A random variable X is distributed normally with a mean of 20 and standard deviation of 4.

(a) On the following diagram, shade the region representing $\mathrm{P}(X \leq 25)$. [2]



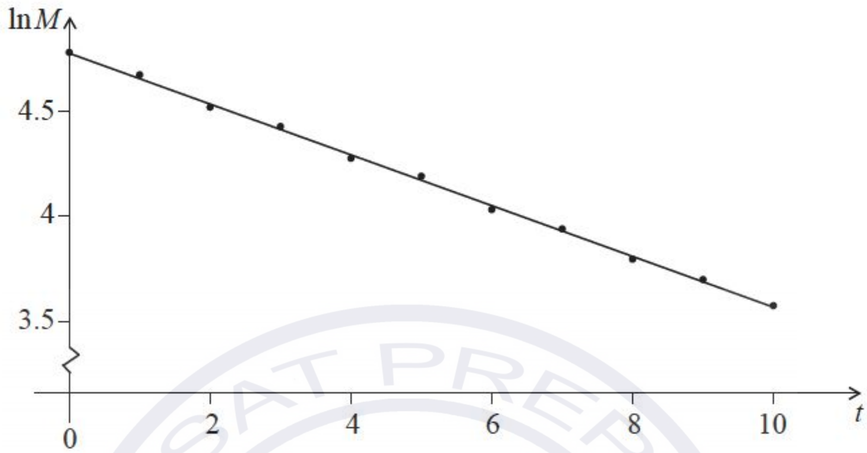
(b) Write down $\mathrm{P}(X \leq 25)$, correct to two decimal places. [2]

(c) Let $\mathrm{P}(X \leq c) = 0.7$. Write down the value of c. [2]

Question 33

[Maximum mark: 6]

The mass M of a decaying substance is measured at one minute intervals. The points $(t, \ln M)$ are plotted for $0 \leq t \leq 10$, where t is in minutes. The line of best fit is drawn. This is shown in the following diagram.



The correlation coefficient for this linear model is $r = -0.998$.

(a) State **two** words that describe the linear correlation between $\ln M$ and t. [2]

(b) The equation of the line of best fit is $\ln M = -0.12t + 4.67$. Given that $M = a \times b^t$, find the value of b. [4]

Question 34

[Maximum mark: 15]

A factory has two machines, A and B. The number of breakdowns of each machine is independent from day to day.

Let A be the number of breakdowns of Machine A on any given day. The probability distribution for A can be modelled by the following table.

a	0	1	2	3
$P(A = a)$	0.55	0.3	0.1	k

(a) Find k. [2]

(b) (i) A day is chosen at random. Write down the probability that Machine A has no breakdowns.

(ii) Five days are chosen at random. Find the probability that Machine A has no breakdowns on exactly four of these days. [3]

Let B be the number of breakdowns of Machine B on any given day. The probability distribution for B can be modelled by the following table.

b	0	1	2	3
$P(B = b)$	0.7	0.2	0.08	0.02

(c) Find $E(B)$. [2]

On Tuesday, the factory uses both Machine A and Machine B. The variables A and B are independent.

(d) (i) Find the probability that there are exactly two breakdowns on Tuesday.

(ii) Given that there are exactly two breakdowns on Tuesday, find the probability that both breakdowns are of Machine A. [8]

Question 35

[Maximum mark: 6]

A competition consists of two independent events, shooting at 100 targets and running for one hour.

The number of targets a contestant hits is the S score. The S scores are normally distributed with mean 65 and standard deviation 10.

(a) A contestant is chosen at random. Find the probability that their S score is less than 50. [2]

The distance in km that a contestant runs in one hour is the R score. The R scores are normally distributed with mean 12 and standard deviation 2.5. The R score is independent of the S score.

Contestants are disqualified if their S score is less than 50 **and** their R score is less than x km.

(b) Given that 1% of the contestants are disqualified, find the value of x. [4]

Question 36

[Maximum mark: 15]

The price of a used car depends partly on the distance it has travelled. The following table shows the distance and the price for seven cars on 1 January 2010.

Distance, x km	11 500	7500	13 600	10 800	9500	12 200	10 400
Price, y dollars	15 000	21 500	12 000	16 000	19 000	14 500	17 000

The relationship between x and y can be modelled by the regression equation $y = ax + b$.

(a) (i) Find the correlation coefficient.

(ii) Write down the value of a and of b. [4]

On 1 January 2010, Lina buys a car which has travelled $11\,000$ km.

(b) Use the regression equation to estimate the price of Lina's car, giving your answer to the nearest 100 dollars. [3]

The price of a car decreases by 5% each year.

(c) Calculate the price of Lina's car after 6 years. [4]

Lina will sell her car when its price reaches $10\,000$ dollars.

(d) Find the year when Lina sells her car. [4]

Question 37

[Maximum mark: 6]

The weights, W, of newborn babies in Australia are normally distributed with a mean 3.41 kg and standard deviation 0.57 kg. A newborn baby has a low birth weight if it weighs less than $w\text{ kg}$.

(a) Given that 5.3% of newborn babies have a low birth weight, find w. [3]

(b) A newborn baby has a low birth weight.
Find the probability that the baby weighs at least 2.15 kg. [3]

Question 38

[Maximum mark: 6]

A jar contains 5 red discs, 10 blue discs and m green discs. A disc is selected at random and replaced. This process is performed four times.

(a) Write down the probability that the first disc selected is red. [1]

(b) Let X be the number of red discs selected. Find the smallest value of m for which $\text{Var}(X) < 0.6$. [5]

Question 39

[Maximum mark: 16]

Ten students were surveyed about the number of hours, x, they spent browsing the Internet during week 1 of the school year. The results of the survey are given below.

$$\sum_{i=1}^{10} x_i = 252\,,\ \ \sigma = 5 \text{ and median} = 27\,.$$

(a) Find the mean number of hours spent browsing the Internet. [2]

(b) During week 2, the students worked on a major project and they each spent an additional five hours browsing the Internet. For week 2, write down

(i) the mean;

(ii) the standard deviation. [2]

(c) During week 3 each student spent 5% less time browsing the Internet than during week 1. For week 3, find

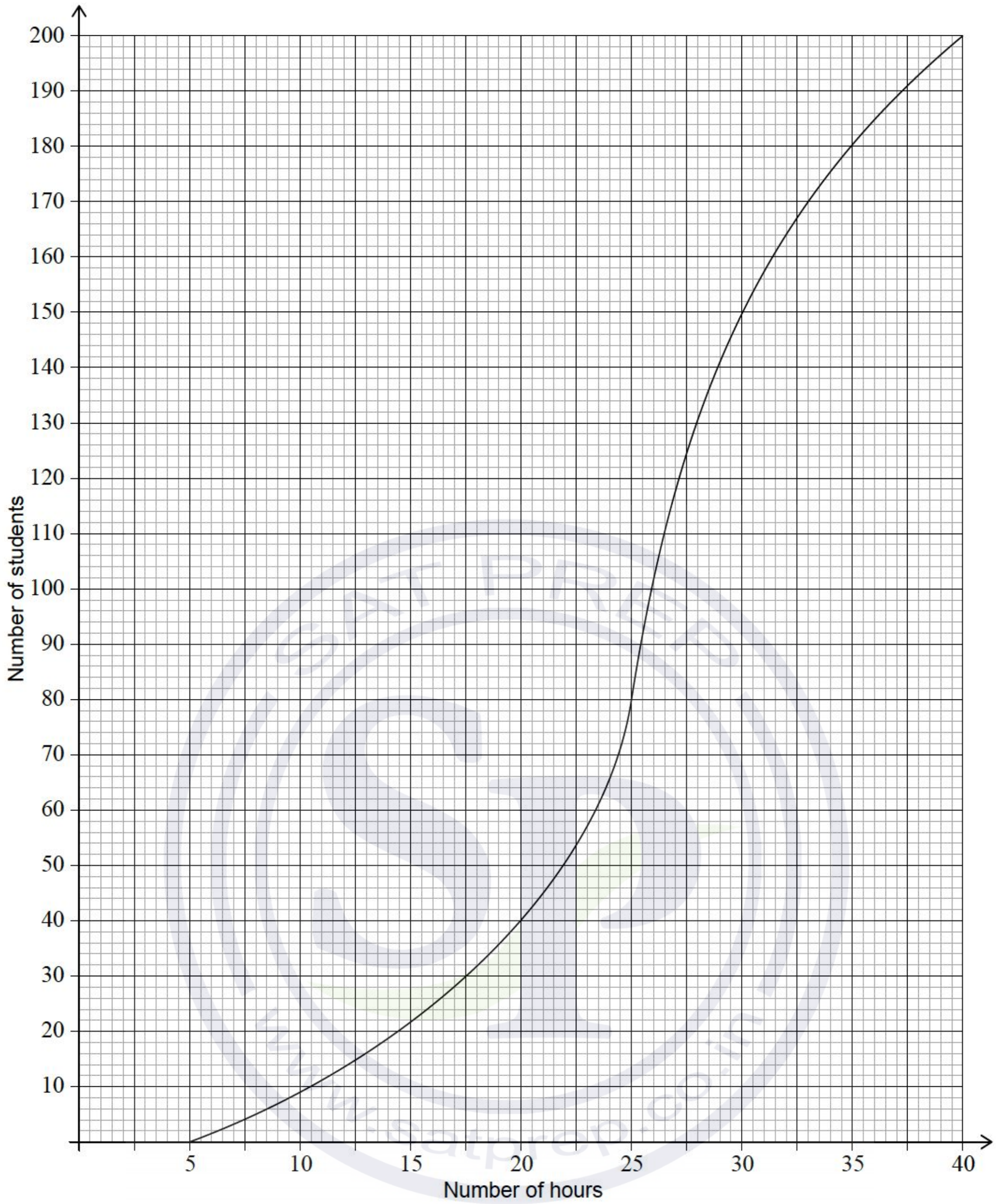
(i) the median;

(ii) the variance. [6]

(d) During week 4, the survey was extended to all 200 students in the school. The results are shown in the cumulative frequency graph on the following page.

(i) Find the number of students who spent between 25 and 30 hours browsing the Internet.

(ii) Given that 10% of the students spent more than k hours browsing the Internet, find the maximum value of k. [6]


Number of students
Number of hours
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
5
10
15
20
25
30
35
40
www.satpren.co.in

Question 40

[Maximum mark: 7]

Consider the following frequency table.

x	Frequency
2	8
4	15
7	21
10	28
11	3

(a) (i) Write down the mode.

(ii) Find the value of the range. [3]

(b) (i) Find the mean.

(ii) Find the variance. [4]

Question 41

[Maximum mark: 6]

In a large university the probability that a student is left handed is 0.08. A sample of 150 students is randomly selected from the university. Let k be the expected number of left-handed students in this sample.

(a) Find k. [2]

(b) Hence, find the probability that

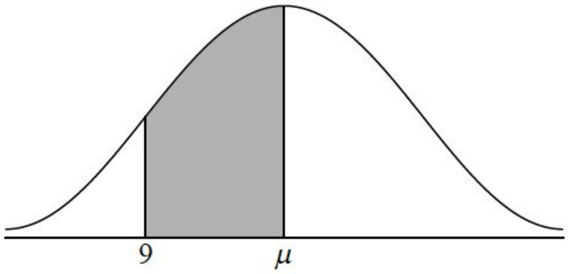
(i) exactly k students are left handed;

(ii) fewer than k students are left handed. [4]

Question 42

[Maximum mark: 15]

A random variable X is normally distributed with mean, μ. In the following diagram, the shaded region between 9 and μ represents 30% of the distribution.



(a) Find $P(X < 9)$. [2]

The standard deviation of X is 2.1.

(b) Find the value of μ. [3]

The random variable Y is normally distributed with mean λ and standard deviation 3.5.
The events $X > 9$ and $Y > 9$ are independent, and $P\big((X > 9) \cap (Y > 9)\big) = 0.4$.

(c) Find λ. [5]

(d) Given that $Y > 9$, find $P(Y < 13)$. [5]

Question 43

[Maximum mark: 7]

The maximum temperature T, in degrees Celsius, in a park on six randomly selected days is shown in the following table. The table also shows the number of visitors, N, to the park on each of those six days.

Maximum temperature (T)	4	5	17	31	29	11
Number of visitors (N)	24	26	36	38	46	28

The relationship between the variables can be modelled by the regression equation $N = aT + b$.

(a) (i) Find the value of a and of b.

(ii) Write down the value of r. [4]

(b) Use the regression equation to estimate the number of visitors on a day when the maximum temperature is $15\,^{\circ}\mathrm{C}$. [3]

Question 44

[Maximum mark: 15]

The following table shows a probability distribution for the random variable X, where $\mathrm{E}(X) = 1.2$.

x	0	1	2	3
$\mathrm{P}(X = x)$	p	$\frac{1}{2}$	$\frac{3}{10}$	q

(a) (i) Find q.

(ii) Find p. [4]

A bag contains white and blue marbles, with at least three of each colour. Three marbles are drawn from the bag, without replacement. The number of blue marbles drawn is given by the random variable X.

(b) (i) Write down the probability of drawing three blue marbles.

(ii) Explain why the probability of drawing three white marbles is $\frac{1}{6}$.

(iii) The bag contains a total of ten marbles of which w are white. Find w. [5]

A game is played in which three marbles are drawn from the bag of ten marbles, without replacement. A player wins a prize if three white marbles are drawn.

(c) Jill plays the game nine times. Find the probability that she wins exactly two prizes. [2]

(d) Grant plays the game until he wins two prizes. Find the probability that he wins his second prize on his eighth attempt. [4]

Question 45

[Maximum mark: 8]

A discrete random variable X has the following probability distribution.

X	0	1	2	3
$\mathrm{P}(X=x)$	0.475	$2k^2$	$\frac{k}{10}$	$6k^2$

(a) Find the value of k. [4]

(b) Write down $\mathrm{P}(X=2)$. [1]

(c) Find $\mathrm{P}(X=2 \mid X>0)$. [3]

Question 46

[Maximum mark: 7]

The heights of adult males in a country are normally distributed with a mean of $180\,\mathrm{cm}$ and a standard deviation of $\sigma\,\mathrm{cm}$. 17% of these men are shorter than $168\,\mathrm{cm}$. 80% of them have heights between $(192-h)\,\mathrm{cm}$ and $192\,\mathrm{cm}$.

Find the value of h.

Question 47

[Maximum mark: 14]

Adam is a beekeeper who collected data about monthly honey production in his bee hives. The data for six of his hives is shown in the following table.

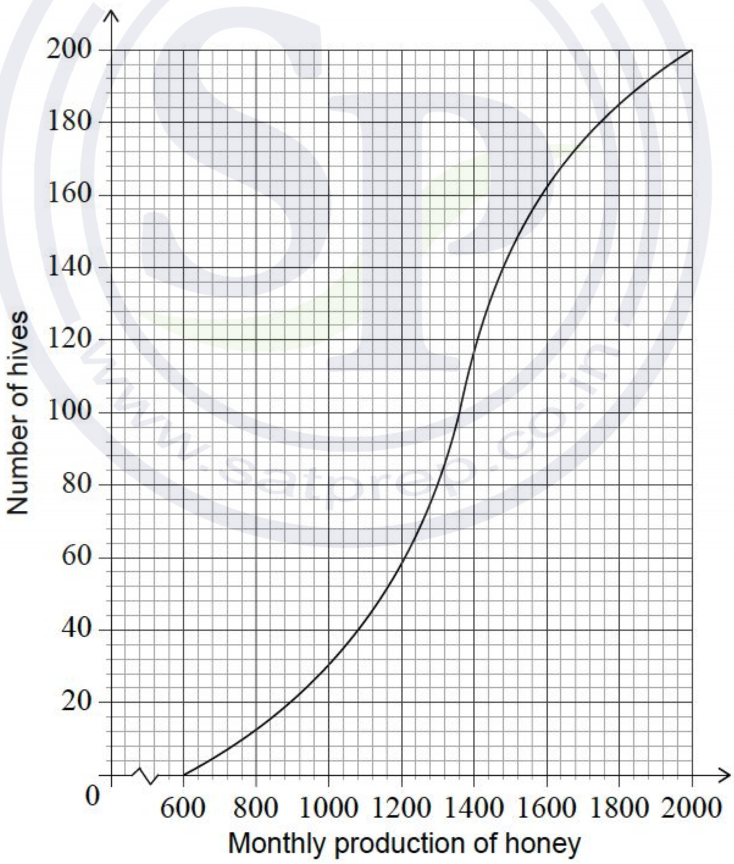
Number of bees (N)	190	220	250	285	305	320
Monthly honey production in grams (P)	900	1100	1200	1500	1700	1800

The relationship between the variables is modelled by the regression line with equation $P = aN + b$.

(a) Write down the value of a and of b. [3]

(b) Use this regression line to estimate the monthly honey production from a hive that has 270 bees. [2]

Adam has 200 hives in total. He collects data on the monthly honey production of all the hives. This data is shown in the following cumulative frequency graph.



Adam's hives are labelled as low, regular or high production, as defined in the following table.

Type of hive	low	regular	high
Monthly honey production in grams (P)	$P \le 1080$	$1080 < P \le k$	$P > k$

(c) Write down the number of low production hives. [1]

Adam knows that 128 of his hives have a regular production.

(d) Find

(i) the value of k;

(ii) the number of hives that have a high production. [5]

(e) Adam decides to increase the number of bees in each low production hive. Research suggests that there is a probability of 0.75 that a low production hive becomes a regular production hive. Calculate the probability that 30 low production hives become regular production hives. [3]

Question 48

[Maximum mark: 17]

The mass M of apples in grams is normally distributed with mean μ. The following table shows probabilities for values of M.

Values of M	$M < 93$	$93 \le M \le 119$	$M > 119$
$P(X)$	k	0.98	0.01

(a) (i) Write down the value of k.

(ii) Show that $\mu = 106$. [4]

(b) Find $P(M < 95)$. [5]

The apples are packed in bags of ten.

Any apples with a mass less than 95 g are classified as small.

(c) Find the probability that a bag of apples selected at random contains at most one small apple. [3]

(d) A crate contains 50 bags of apples. A crate is selected at random.

(i) Find the expected number of bags in this crate that contain at most one small apple.

(ii) Find the probability that at least 48 bags in this crate contain at most one small apple. [5]

Question 49

[Maximum mark: 6]

The following table shows the mean weight, y kg, of children who are x years old.

Age (x years)	1.25	2.25	3.5	4.4	5.85
Weight (y kg)	10	13	14	17	19

The relationship between the variables is modelled by the regression line with equation $y = ax + b$.

(a) (i) Find the value of a and of b.

(ii) Write down the correlation coefficient. [4]

(b) Use your equation to estimate the mean weight of a child that is 1.95 years old. [2]

Question 50

[Maximum mark: 17]

The weights, in grams, of oranges grown in an orchard, are normally distributed with a mean of $297\,\text{g}$. It is known that $79\,\%$ of the oranges weigh more than $289\,\text{g}$ and $9.5\,\%$ of the oranges weigh more than $310\,\text{g}$.

(a) Find the probability that an orange weighs between $289\,\text{g}$ and $310\,\text{g}$. [2]

The weights of the oranges have a standard deviation of σ.

(b) (i) Find the standardized value for $289\,\text{g}$.

(ii) Hence, find the value of σ. [5]

The grocer at a local grocery store will buy the oranges whose weights exceed the 35th percentile.

(c) To the nearest gram, find the minimum weight of an orange that the grocer will buy. [3]

The orchard packs oranges in boxes of 36.

(d) Find the probability that the grocer buys more than half the oranges in a box selected at random. [5]

The grocer selects two boxes at random.

(e) Find the probability that the grocer buys more than half the oranges in each box. [2]

Question 51

[Maximum mark: 13]

The following table shows values of $\ln x$ and $\ln y$.

$\ln x$	1.10	2.08	4.30	6.03
$\ln y$	5.63	5.22	4.18	3.41

The relationship between $\ln x$ and $\ln y$ can be modelled by the regression equation $\ln y = a \ln x + b$.

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

(b) Use the regression equation to estimate the value of y when $x = 3.57$. [3]

The relationship between x and y can be modelled using the formula $y = kx^n$, where $k \neq 0$, $n \neq 0$, $n \neq 1$.

(c) By expressing $\ln y$ in terms of $\ln x$, find the value of n and of k. [7]

Question 52

[Maximum mark: 6]

Two events A and B are such that $\mathrm{P}(A) = 0.62$ and $\mathrm{P}(A \cap B) = 0.18$.

(a) Find $\mathrm{P}(A \cap B')$. [2]

(b) Given that $\mathrm{P}\left((A \cup B)'\right) = 0.19$, find $\mathrm{P}(A \mid B')$. [4]

Question 53

[Maximum mark: 6]

A biased four-sided die is rolled. The following table gives the probability of each score.

Score	1	2	3	4
Probability	0.28	k	0.15	0.3

(a) Find the value of k. [2]

(b) Calculate the expected value of the score. [2]

(c) The die is rolled 80 times. On how many rolls would you expect to obtain a three? [2]

Question 54

[Maximum mark: 15]

A nationwide study on reaction time is conducted on participants in two age groups. The participants in Group X are less than 40 years old. Their reaction times are normally distributed with mean 0.489 seconds and standard deviation 0.07 seconds.

(a) A person is selected at random from Group X. Find the probability that their reaction time is greater than 0.65 seconds. [2]

The participants in Group Y are 40 years or older. Their reaction times are normally distributed with mean 0.592 seconds and standard deviation σ seconds.

(b) The probability that the reaction time of a person in Group Y is greater than 0.65 seconds is 0.396. Find the value of σ. [4]

In the study, 38% of the participants are in Group X.

(c) A randomly selected participant has a reaction time greater than 0.65 seconds. Find the probability that the participant is in Group X. [6]

(d) Ten of the participants with reaction times greater than 0.65 are selected at random. Find the probability that at least two of them are in Group X. [3]

Question 55

[Maximum mark: 6]

The following table shows the hand lengths and the heights of five athletes on a sports team.

Hand length (x cm)	21.0	21.9	21.0	20.3	20.8
Height (y cm)	178.3	185.0	177.1	169.0	174.6

The relationship between x and y can be modelled by the regression line with equation $y = ax + b$.

(a) (i) Find the value of a and of b.

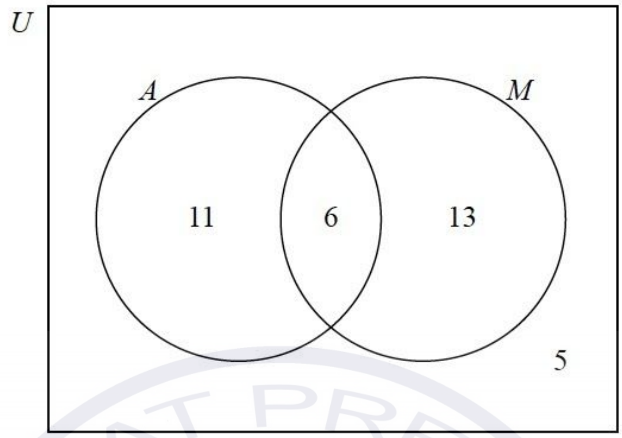
(ii) Write down the correlation coefficient. [4]

(b) Another athlete on this sports team has a hand length of 21.5 cm. Use the regression equation to estimate the height of this athlete. [2]

Question 56

[Maximum mark: 6]

In a group of 35 students, some take art class (A) and some take music class (M). 5 of these students do not take either class. This information is shown in the following Venn diagram.



(a) Write down the number of students in the group who take art class. [2]

(b) One student from the group is chosen at random. Find the probability that

(i) the student does not take art class;

(ii) the student takes either art class or music class, but not both. [4]

Question 57

[Maximum mark: 14]

At Penna Airport the probability, $\mathrm{P}(A)$, that all passengers arrive on time for a flight is 0.70. The probability, $\mathrm{P}(D)$, that a flight departs on time is 0.85. The probability that all passengers arrive on time for a flight and it departs on time is 0.65.

(a) Show that event A and event D are **not** independent. [2]

(b) (i) Find $\mathrm{P}(A \cap D')$.

(ii) Given that all passengers for a flight arrive on time, find the probability that the flight does **not** depart on time. [5]

The number of hours that pilots fly per week is normally distributed with a mean of 25 hours and a standard deviation σ. 90% of pilots fly less than 28 hours in a week.

(c) Find the value of σ. [3]

(d) All flights have two pilots. Find the percentage of flights where **both** pilots flew more than 30 hours last week. [4]

Question 58

[Maximum mark: 6]

A group of 7 adult men wanted to see if there was a relationship between their Body Mass Index (BMI) and their waist size. Their waist sizes, in centimetres, were recorded and their BMI calculated. The following table shows the results.

Waist (x cm)	58	63	75	82	93	98	105
BMI (y)	19	20	22	23	25	24	26

The relationship between x and y can be modelled by the regression equation $y = ax + b$.

(a) (i) Write down the value of a and of b.

(ii) Find the correlation coefficient. [4]

(b) Use the regression equation to estimate the BMI of an adult man whose waist size is $95\,\text{cm}$. [2]

Question 59

[Maximum mark: 16]

There are three fair six-sided dice. Each die has two green faces, two yellow faces and two red faces.

All three dice are rolled.

(a) (i) Find the probability of rolling exactly one red face.

(ii) Find the probability of rolling two or more red faces. [5]

Ted plays a game using these dice. The rules are:

- Having a turn means to roll all three dice.
- He wins $\$10$ for each green face rolled and adds this to his winnings.
- After a turn Ted can either:
  - end the game (and keep his winnings), or
  - have another turn (and try to increase his winnings).
- If two or more red faces are rolled in a turn, all winnings are lost and the game ends.

(b) Show that, after a turn, the probability that Ted adds exactly $\$10$ to his winnings is $\frac{1}{3}$. [5]

The random variable D ($\$$) represents how much is added to his winnings after a turn.

The following table shows the distribution for D, where $\$w$ represents his winnings in the game so far.

D ($\$$)	$-w$	0	10	20	30
$\mathrm{P}(D = d)$	x	y	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{27}$

(c) (i) Write down the value of x.

(ii) Hence, find the value of y. [3]

Ted will always have another turn if he expects an increase to his winnings.

(d) Find the least value of w for which Ted should end the game instead of having another turn. [3]

Question 60

[Maximum mark: 6]

A jigsaw puzzle consists of many differently shaped pieces that fit together to form a picture.

Jill is doing a 1000-piece jigsaw puzzle. She started by sorting the edge pieces from the interior pieces. Six times she stopped and counted how many of each type she had found. The following table indicates this information.

Edge pieces (x)	16	31	39	55	84	115
Interior pieces (y)	89	239	297	402	580	802

Jill models the relationship between these variables using the regression equation $y = ax + b$.

(a) Write down the value of a and of b. [3]

(b) Use the model to predict how many edge pieces she had found when she had sorted a **total** of 750 pieces. [3]

Question 61

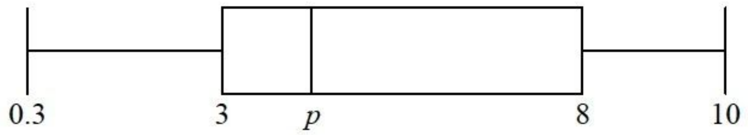
[Maximum mark: 5]

Ten students were asked for the distance, in km, from their home to school. Their responses are recorded below.

$$0.3 \quad 0.4 \quad 3 \quad 3 \quad 3.5 \quad 5 \quad 7 \quad 8 \quad 8 \quad 10$$

(a) For these data, find the mean distance from a student's home to school. [2]

The following box-and-whisker plot represents this data.



(b) Find the value of p. [1]

(c) Find the interquartile range. [2]

Question 62

[Maximum mark: 15]

SpeedWay airline flies from city A to city B. The flight time is normally distributed with a mean of 260 minutes and a standard deviation of 15 minutes.

A flight is considered late if it takes longer than 275 minutes.

(a) Calculate the probability a flight is **not** late. [2]

The flight is considered to be **on time** if it takes between m and 275 minutes. The probability that a flight is on time is 0.830.

(b) Find the value of m. [3]

During a week, SpeedWay has 12 flights from city A to city B. The time taken for any flight is independent of the time taken by any other flight.

(c) (i) Calculate the probability that at least 7 of these flights are **on time**.

(ii) Given that at least 7 of these flights are on time, find the probability that exactly 10 flights are on time. [7]

SpeedWay increases the number of flights from city A to city B to 20 flights each week, and improves their efficiency so that more flights are on time. The probability that at least 19 flights are on time is 0.788.

(d) A flight is chosen at random. Calculate the probability that it is on time. [3]

Question 63

[Maximum mark: 7]

The following table shows the probability distribution of a discrete random variable X, where $a \geq 0$ and $b \geq 0$.

x	1	4	a	$a+b-0.5$
$\mathrm{P}(X=x)$	0.2	0.5	b	a

(a) Show that $b = 0.3 - a$. [1]

(b) Find the difference between the greatest possible expected value and the least possible expected value. [6]

Question 64

[Maximum mark: 6]

The number of messages, M, that six randomly selected teenagers sent during the month of October is shown in the following table. The table also shows the time, T, that they spent talking on their phone during the same month.

Time spent talking on their phone (T minutes)	50	55	105	128	155	200
Number of messages (M)	358	340	740	731	800	992

The relationship between the variables can be modelled by the regression equation $M = aT + b$.

(a) Write down the value of a and of b. [3]

(b) Use your regression equation to predict the number of messages sent by a teenager that spent 154 minutes talking on their phone in October. [3]