"Knowing the path is good but not enough, walking the path with determination leads to destiny"

AS Edexcel Statistics (WST01) Classified Questions

2019 – 2023

1.1 Representation and analysing the data:

Reference Notes:

Reference Notes:

Q1.

Jim records the length, l mm, of 81 salmon. The data are coded using $x = l - 600$ and the following summary statistics are obtained.

$$n = 81 \qquad \sum x = 3711 \qquad \sum x^2 = 475\,181$$

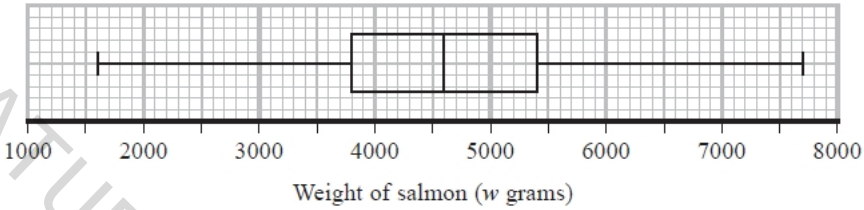
(a) Find the mean length of these salmon.

(3)

(b) Find the variance of the lengths of these salmon.

(2)

The weight, w grams, of each of the 81 salmon is recorded to the nearest gram.
The recorded results for the 81 salmon are summarised in the box plot below.



(c) Find the maximum number of salmon that have weights in the interval

$$4600 < w \leq 7700$$

(1)

Raj says that the box plot is incorrect as Jim has not included outliers.

For these data an outlier is defined as a value that is more than

1.5 × IQR above the upper quartile or 1.5 × IQR below the lower quartile

(d) Show that there are no outliers.

(3)

(Total for question = 9 marks)

(**Q03 WST01/01, June 2023**)

Extra space for working:

Q2.

Morgan is investigating the body length, b centimetres, of squirrels.

A random sample of 8 squirrels is taken and the data for each squirrel is coded using

$$x = \frac{b - 21}{2}$$

The results for the coded data are summarised below

$$\sum x = -1.2 \qquad \sum x^2 = 5.1$$

(a) Find the mean of b

(3)

(b) Find the standard deviation of b

(3)

A 9th squirrel is added to the sample. Given that for all 9 squirrels $\sum x = 0$

(c) find

(i) the body length of the 9th squirrel,

(2)

(ii) the standard deviation of x for all 9 squirrels.

(2)

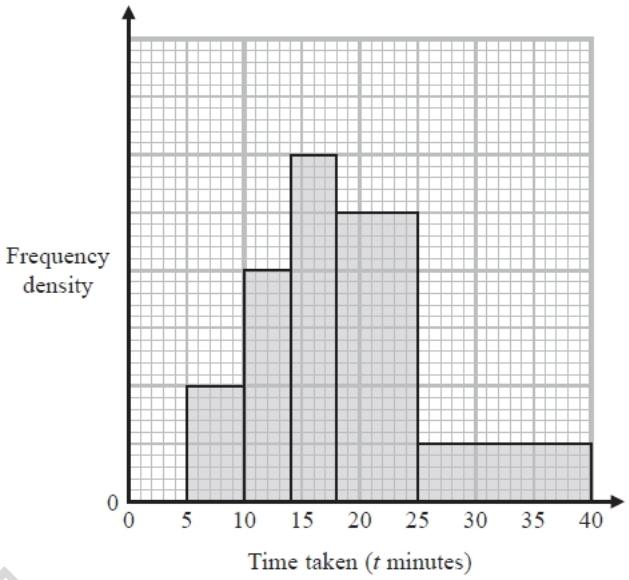
(Total for question = 10 marks)

(**Q03 WST01/01, Oct 2022**)

Extra space for working:

Q3.

The histogram shows the times taken, t minutes, by each of 100 people to swim 500 metres.



(a) Use the histogram to complete the frequency table for the times taken by the 100 people to swim 500 metres.

Time taken (t minutes)	5 – 10	10 – 14	14 – 18	18 – 25	25 – 40
Frequency (f)	10	16	24		

(1)

(b) Estimate the number of people who took less than 16 minutes to swim 500 metres.

(2)

(c) Find an estimate for the mean time taken to swim 500 metres.

(2)

Given that $\sum ft^2 = 41\,033$

(d) find an estimate for the standard deviation of the times taken to swim 500 metres.

(2)

Given that $Q_3 = 23$

(e) use linear interpolation to estimate the interquartile range of the times taken to swim 500 metres.

(3)

(Total for question = 10 marks)

(**Q01 WST01/01, Jan 2023**)

Extra space for working:

Q4.

The stem lengths of a sample of 120 tulips are recorded in the grouped frequency table below.

Stem length (cm)	Frequency
$40 \leqslant x < 42$	12
$42 \leqslant x < 45$	18
$45 \leqslant x < 50$	23
$50 \leqslant x < 55$	35
$55 \leqslant x < 58$	24
$58 \leqslant x < 60$	8

A histogram is drawn to represent these data.

The area of the bar representing the $40 \leqslant x < 42$ class is 16.5 cm^2

(a) Calculate the exact area of the bar representing the $40 \leqslant x < 42$ class.

(2)

The height of the tallest bar in the histogram is 10 cm.

(b) Find the exact height of the second tallest bar.

(3)

Q_1 for these data is 45 cm.

(c) Use linear interpolation to find an estimate for

(i) Q_2
(ii) the interquartile range.

(4)

One measure of skewness is given by

$$\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

(d) By calculating this measure, describe the skewness of these data.

(2)

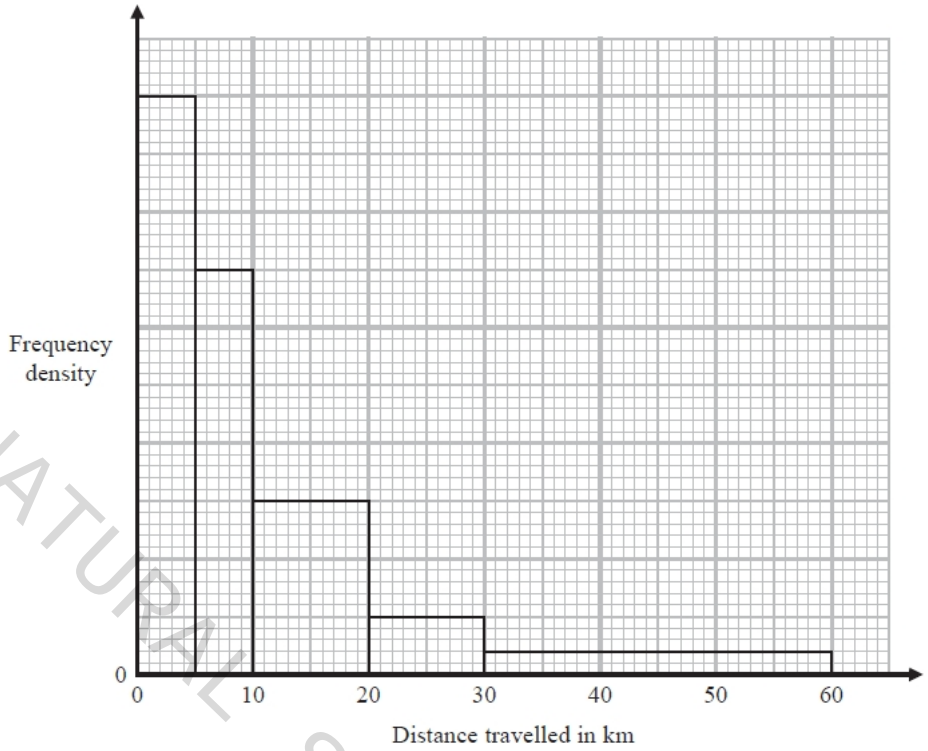
(Total for question = 11 marks)

(**Q01 WST01/01, Oct 2022**)

Extra space for working:

Q5.

The histogram shows the distances, in km, that 274 people travel to work.



Given that 60 of these people travel between 10 km and 20 km to work, estimate

(a) the number of people who travel between 22 km and 45 km to work,

(3)

(b) the median distance travelled to work by these 274 people,

(2)

(c) the mean distance travelled to work by these 274 people.

(3)

(Total for question = 8 marks)

(**Q01 WST01/01, June 2023**)

Extra space for working:

Q6.

A disc of radius 1 cm is rolled onto a horizontal grid of rectangles so that the disc is equally likely to land anywhere on the grid. Each rectangle is 5 cm long and 3 cm wide. There are no gaps between the rectangles and the grid is sufficiently large so that no discs roll off the grid.

If the disc lands inside a rectangle without covering any part of the edges of the rectangle then a prize is won.

By considering the possible positions for the centre of the disc,

(a) show that the probability of winning a prize on any particular roll is $\frac{1}{5}$

(3)

A group of 15 students each roll the disc onto the grid twenty times and record the number of times, x, that each student wins a prize. Their results are summarised as follows

$$\sum x = 61 \qquad \sum x^2 = 295$$

(b) Find the standard deviation of the number of prizes won per student.

(2)

A second group of 12 students each roll the disc onto the grid twenty times and the mean number of prizes won per student is 3.5 with a standard deviation of 2

(c) Find the mean and standard deviation of the number of prizes won per student for the whole group of 27 students.

(7)

The 27 students also recorded the number of times that the disc covered a corner of a rectangle and estimated the probability to be 0.2216 (to 4 decimal places).

(d) Explain how this probability could be used to find an estimate for the value of π and state the value of your estimate.

(3)

(Total for question = 15 marks)

(**Q06 WST01/01, Jan 2021**)

Extra space for working:

Q7.

The time taken to complete a puzzle, in minutes, is recorded for each person in a club. The times are summarised in a grouped frequency distribution and represented by a histogram.

One of the class intervals has a frequency of 20 and is shown by a bar of width 1.5 cm and height 12 cm on the histogram. The total area under the histogram is 94.5 cm^2

Find the number of people in the club.

(Total for question = 3 marks)

(**Q02 WST11/01, Specimen papers**)

Q8.

A researcher recorded the time, t minutes, spent using a mobile phone during a particular afternoon, for each child in a club.

The researcher coded the data using $v = \frac{t-5}{10}$ and the results are summarised in the table below.

Coded Time (v)	Frequency (f)	Coded Time Midpoint (m)
$0 \leqslant v < 5$	20	2.5
$5 \leqslant v < 10$	24	a
$10 \leqslant v < 15$	16	12.5
$15 \leqslant v < 20$	14	17.5
$20 \leqslant v < 30$	6	b

(You may use $\sum fm = 825$ and $\sum fm^2 = 12\,012.5$)

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

(1)

(b) Calculate an estimate of the mean of v.

(1)

(c) Calculate an estimate of the standard deviation of v.

(2)

(d) Use linear interpolation to estimate the median of v.

(2)

(e) Hence describe the skewness of the distribution. Give a reason for your answer.

(2)

(f) Calculate estimates of the mean and the standard deviation of the time spent using a mobile phone during the afternoon by the children in this club.

(4)

(Total for question = 12 marks)

(**Q04 WST11/01, Specimen papers**)

Extra space for working:

Q9.

A random sample of 100 carrots is taken from a farm and their lengths, L cm, recorded.
The data are summarised in the following table.

Length, L cm	Frequency, f	Class mid point, x cm
$5 \leqslant L < 8$	5	6.5
$8 \leqslant L < 10$	13	9
$10 \leqslant L < 12$	16	11
$12 \leqslant L < 15$	25	13.5
$15 \leqslant L < 20$	30	17.5
$20 \leqslant L < 28$	11	24

A histogram is drawn to represent these data.
The bar representing the class $5 \leq L < 8$ is 1.5 cm wide and 1 cm high.

(a) Find the width and height of the bar representing the class $15 \leq L < 20$

(3)

(b) Use linear interpolation to estimate the median length of these carrots.

(2)

(c) Estimate

(i) the mean length of these carrots,

(2)

(ii) the standard deviation of the lengths of these carrots.

(3)

A supermarket will only buy carrots with length between 9 cm and 22 cm.

(d) Estimate the proportion of carrots from the farm that the supermarket will buy.

(2)

Any carrots that the supermarket does not buy are sold as animal feed.

The farm makes a profit of 2.2 pence on each carrot sold to the supermarket, a profit of 0.8 pence on each carrot longer than 22 cm and a loss of 1.2 pence on each carrot shorter than 9 cm.

(e) Find an estimate of the mean profit per carrot made by the farm.

(2)

(Total for question = 14 marks)

(**Q03 WST01/01, June 2021**)

Extra space for working:

Q10.

The company *Seafield* requires contractors to record the number of hours they work each week. A random sample of 38 weeks is taken and the number of hours worked per week by contractor Kiana is summarised in the stem and leaf diagram below.

Stem	Leaf										
1	4	4	4	5	5	5	6	6	9	9	9 (11)
2	1	2	2	3	3	4	4	4	w	9	(10)
3	2	3	4	4	5	6	7	7	7	9	(10)
4	1	1	2	3							(4)
5	1	9									(2)
6	4										(1)

Key : 3|2 means 32

The quartiles for this distribution are summarised in the table below.

Q_1	Q_2	Q_3
x	26	y

(a) Find the values of w, x and y

(3)

Kiana is looking for outliers in the data. She decides to classify as outliers any observations greater than

$$Q_3 + 1.0 \times (Q_3 - Q_1)$$

(b) Showing your working clearly, identify any outliers that Kiana finds.

(2)

(c) Draw a box plot for these data in the space provided on the grid opposite.

(3)

(d) Use the formula

$$\text{skewness} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

to find the skewness of these data. Give your answer to 2 significant figures.

(2)

Kiana's new employer, *Landacre*, wishes to know the average number of hours per week she worked during her employment at *Seafield* to help calculate the cost of employing her.

(e) Explain why *Landacre* might prefer to know Kiana's mean, rather than median, number of hours worked per week.

(1)

(Total for question = 11 marks)

(Q01 WST01/01, June 2022)

Extra space for working:

Q11.

Gill buys a bag of logs to use in her stove. The lengths, l cm, of the 88 logs in the bag are summarised in the table below.

Length (l)	Frequency (f)
$15 < l \leqslant 20$	19
$20 < l \leqslant 25$	35
$25 < l \leqslant 27$	16
$27 < l \leqslant 30$	15
$30 < l \leqslant 40$	3

A histogram is drawn to represent these data.

The bar representing logs with length $27 < l \leqslant 30$ has a width of 1.5 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing log lengths of $20 < l \leqslant 25$

(3)

(b) Use linear interpolation to estimate the median of l

(2)

The maximum length of log Gill can use in her stove is 26 cm.

Gill estimates, using linear interpolation, that x logs from the bag will fit into her stove.

(c) Show that $x = 62$

(1)

Gill randomly selects 4 logs from the bag.

(d) Using $x = 62$, find the probability that all 4 logs will fit into her stove.

(2)

The weights, W grams, of the logs in the bag are coded using $y = 0.5w - 255$ and summarised by

$$n = 88 \quad \sum y = 924 \quad \sum y^2 = 12\,862$$

(e) Calculate

(i) the mean of W

(3)

(ii) the variance of W

(3)

(Total for question = 14 marks)

(Q03 WST01/01, June 2022)

Q12.

The production cost, £c million, of a film and the total ticket sales, £t million, earned by the film are recorded for a sample of 40 films.

Some summary statistics are given below.

$$\sum c = 1634 \quad \sum t = 1361 \quad \sum t^2 = 82\,873 \quad \sum ct = 83\,634 \quad S_{cc} = 28\,732.1$$

(a) Find the exact value of S_{tt} and the exact value of S_{ct}

(3)

(b) Calculate the value of the product moment correlation coefficient for these data.

(2)

(c) Give an interpretation of your answer to part (b)

(1)

(d) Show that the equation of the linear regression line of t on c can be written as

$$t = -5.84 + 0.976c$$

where the values of the intercept and gradient are given to 3 significant figures.

(3)

(e) Find the expected total ticket sales for a film with a production cost of £90 million.

(2)

Using the regression line in part (d)

(f) find the range of values of the production cost of a film for which the total ticket sales are less than 80% of its production cost.

(2)

(Total for question = 13 marks)

(**Q02 WST01/01, Oct 2022**)

Extra space for working:

Q13.

The stem and leaf diagram below shows the ages (in years) of the residents in a care home.

Age												Key: 4 | 3 is an age of 43
4	3											(1)
5	4											(1)
6	2	3	5	6	8	8	8	9	9			(9)
7	1	1	3	4	4	6	6	6	8	8	9	(11)
8	0	0	2	7	8	8	9					(7)
9	3	7										(2)

(a) Find the median age of the residents.

(1)

(b) Find the interquartile range (IQR) of the ages of the residents.

(2)

An outlier is defined as a value that is either

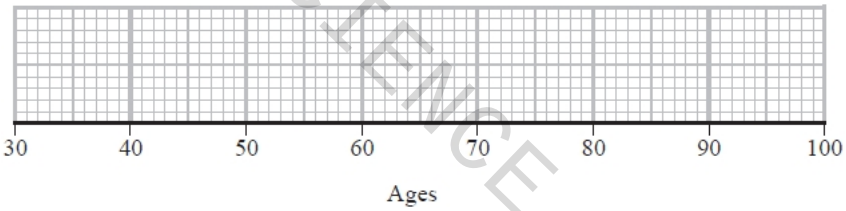
more than 1.5 × (IQR) below the lower quartile or
more than 1.5 × (IQR) above the upper quartile.

(c) Determine any outliers in these data. Show clearly any calculations that you use.

(3)

(d) On the grid below, draw a box plot to summarise these data.

(3)



(Total for question = 9 marks)

(**Q02 WST01/01, Jan 2021**)

Extra space for working:

Q14.

The weights, to the nearest kilogram, of a sample of 33 red kangaroos taken in December are summarised in the stem and leaf diagram below.

	Weight (kg)	Totals
1	6	(1)
2	3 6	(2)
3	2 4 6	(3)
4	2 5 5 6 6 7 8	(7)
5	3 4 7 7 7 8 9 9	(8)
6	0 2 2 3 3 7 8	(7)
7	2 8	(2)
8	2 6	(2)
9	4	(1)

Key: 3 | 2 represents 32 kg

(a) Find

(i) the value of the median
(ii) the value of Q_1 and the value of Q_3
for the weights of these red kangaroos.

(3)

For these data an outlier is defined as a value that is

greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$
or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$

(b) Show that there are 2 outliers for these data.

(3)

Figure 1 below shows a box plot for the weights of the same 33 red kangaroos taken in February, earlier in the year.

(c) In the space on Figure 1, draw a box plot to represent the weights of these red kangaroos in December.

(4)

(d) Compare the distribution of the weights of red kangaroos taken in February with the distribution of the weights of red kangaroos taken in December of the same year. You should interpret your comparisons in the context of the question.

(3)

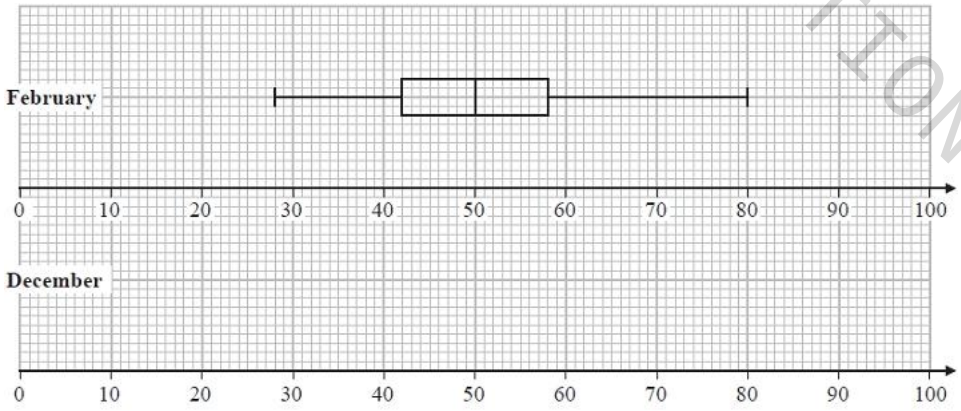


Figure 1

(Total for question = 13 marks)

(Q02 WST01/01, Oct 2023)

Extra space for working:

Extra space for working:

Q15.

The stem and leaf diagram shows the ages of the 35 male passengers on a cruise.

Age **Key:** 1 | 3 represents an age of 13 years

1	3	(1)
2	7 9	(2)
3	1 2 8 8	(4)
4	5 5 6 7 8 8 9	(7)
5	2 2 3 3 4 4 5 6 6 8	(10)
6	0 1 1 4 4 4 7	(7)
7	3 6	(2)
8	7 8	(2)

(a) Find the median age of the male passengers.

(1)

(b) Show that the interquartile range (IQR) of these ages is 16

(2)

An outlier is defined as a value that is more than

1.5 × IQR above the upper quartile

or

1.5 × IQR below the lower quartile

(c) Show that there are 3 outliers amongst these ages.

(3)

(d) On the grid in Figure 1 below, draw a box plot for the ages of the male passengers on the cruise.

(4)

Figure 1 below also shows a box plot for the ages of the female passengers on the cruise.

(e) Comment on any difference in the distributions of ages of male and female passengers on the cruise.

State the values of any statistics you have used to support your comment.

(1)

Anja, along with her 2 daughters and a granddaughter, now join the cruise.
Anja's granddaughter is younger than both of Anja's daughters.
Anja had her 23rd birthday on the day her eldest daughter was born.
When their 4 ages are included with the other female passengers on the cruise, the box plot does not change.

(f) State, giving reasons, what you can say about

(i) the granddaughter's age
(ii) Anja's age.

(3)

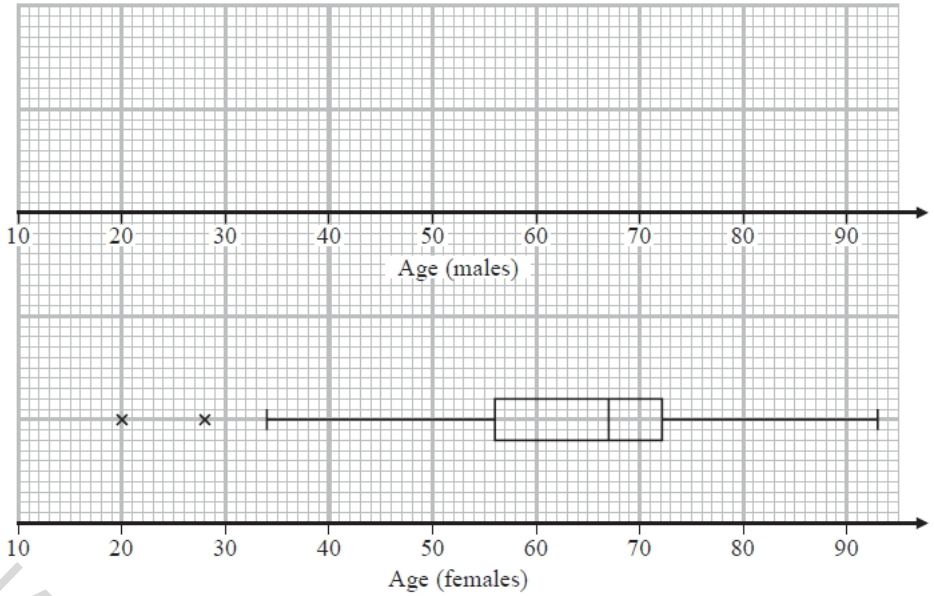


Figure 1

(Total for question = 14 marks)

(**Q03 WST01/01, Oct 2021**)

Extra space for working:

Extra space for working:

Q16.

The stem and leaf diagram shows the number of deliveries made by Pat each day for 24 days

Key: 10 | 8 represents 108 deliveries

10	8	9										(2)
11	0	3	6	6	6	8	8	9	9	9	9	(11)
12	4	5	5	5	5	5	5	8				(8)
13	a	b	c									(3)

where a, b and c are positive integers with $a < b < c$

An outlier is defined as any value greater than 1.5 × interquartile range above the upper quartile.

Given that there is only one outlier for these data,

(a) show that $c = 9$

(3)

The number of deliveries made by Pat each day is represented by d

The data in the stem and leaf diagram are coded using

$$x = d - 125$$

and the following summary statistics are obtained

$$\sum x = -96 \quad \text{and} \quad \sum (x - \bar{x})^2 = 1306$$

(b) Find the mean number of deliveries.

(3)

(c) Find the standard deviation of the number of deliveries.

(2)

One of these 24 days is selected at random. The random variable D represents the number of deliveries made by Pat on this day.

The random variable $X = D - 125$

(d) Find $P(D > 118 \mid X < 0)$

(2)

(Total for question = 10 marks)

(Q03 WST01/01, Jan 2022)

Extra space for working:

Extra space for working:

1.2 Probability:

Reference Notes:

Q1.

Sally plays a game in which she can either win or lose.

A turn consists of up to 3 games. On each turn Sally plays the game up to 3 times. If she wins the first 2 games or loses the first 2 games, then she will not play the 3rd game.

- The probability that Sally wins the first game in a turn is 0.7
- If Sally wins a game the probability that she wins the next game is 0.6
- If Sally loses a game the probability that she wins the next game is 0.2

(a) Use this information to complete the tree diagram below.

(3)

(b) Find the probability that Sally wins the first 2 games in a turn.

(2)

(c) Find the probability that Sally wins exactly 2 games in a turn.

(2)

Given that Sally wins 2 games in a turn,

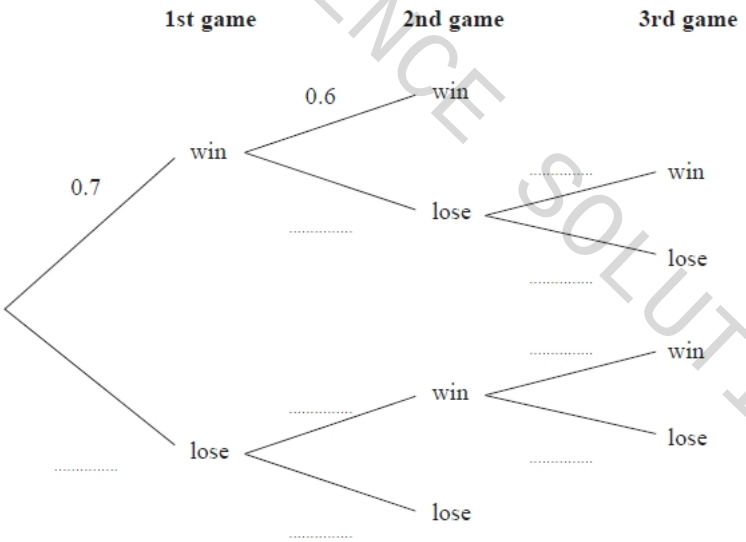
(d) find the probability that she won the first 2 games.

(2)

Given that Sally won the first game in a turn,

(e) find the probability that she won 2 games.

(2)



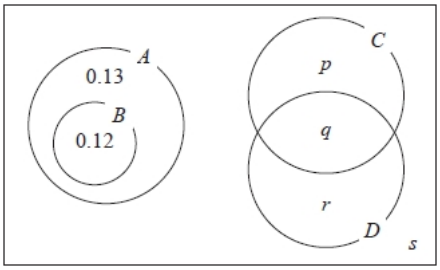
(Total for question = 11 marks)

(Q01 WST01/01, Oct 2023)

Extra space for working:

Q2.

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



(a) Write down the value of

(i) P(A)
(ii) P($A|B$)
(iii) P($A|C$)

(3)

Given that $P(B' \cap D') = \frac{7}{10}$ and $P(C|D) = \frac{3}{5}$

(b) find the exact value of q and the exact value of r

(6)

Given also that $P(B \cup C') = \frac{5}{8}$

(c) find the exact value of s

(2)

(Total for question = 11 marks)

(Q06 WST01/01, Oct 2022)

Extra space for working:

Q3.

Three events A, B and C are such that

$$P(A) = 0.1 \qquad P(B|A) = 0.3 \qquad P(A \cup B) = 0.25 \qquad P(C) = 0.5$$

Given that A and C are mutually exclusive

(a) find $P(A \cup C)$

(1)

(b) Show that $P(B) = 0.18$

(3)

Given also that B and C are independent,

(c) draw a Venn diagram to represent the events A, B and C and the probabilities associated with each region.

(5)

(Total for question = 9 marks)

(**Q06 WST01/01, June 2023**)

Q4.

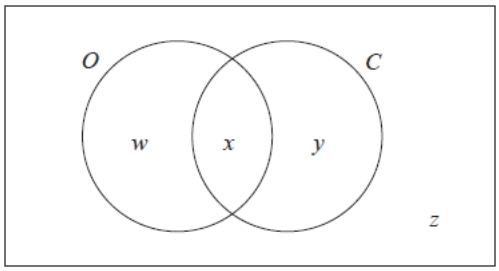
(i) Bob shops at a market each week. The event that

Bob buys carrots is denoted by C
Bob buys onions is denoted by O

At each visit, Bob may buy neither, or one, or both of these items. The probability that

Bob buys carrots is 0.65
Bob does **not** buy onions is 0.3
Bob buys onions but not carrots is 0.15

The Venn diagram below represents the events C and O



where w, x, y and z are probabilities.

(a) Find the value of w, the value of x, the value of y and the value of z

(4)

For one visit to the market,

(b) find the probability that Bob buys either carrots or onions but not both.

(1)

(c) Show that the events C and O are **not** independent.

(2)

(ii) F, G and H are 3 events. F and H are mutually exclusive. F and G are independent.

Given that

$$P(F) = \frac{2}{7} \qquad P(H) = \frac{1}{4} \qquad P(F \cup G) = \frac{5}{8}$$

(a) find $P(F \cup H)$

(1)

(b) find $P(G)$

(3)

(c) find $P(F \cap G)$

(1)

(Total for question = 12 marks)

(Q03 WST01/01, Oct 2023)

Extra space for working:

Q5.

Three events A, B and C are such that

$$P(A) = \frac{2}{5} \qquad P(C) = \frac{1}{2} \qquad P(A \cup B) = \frac{5}{8}$$

Given that A and C are mutually exclusive find

(a) $P(A \cup C)$

(1)

Given that A and B are independent

(b) show that $P(B) = \frac{3}{8}$

(4)

(c) Find $P(A \mid B)$

(1)

Given that $P(C' \cap B') = 0.3$

(d) draw a Venn diagram to represent the events A, B and C

(5)

(Total for question = 11 marks)

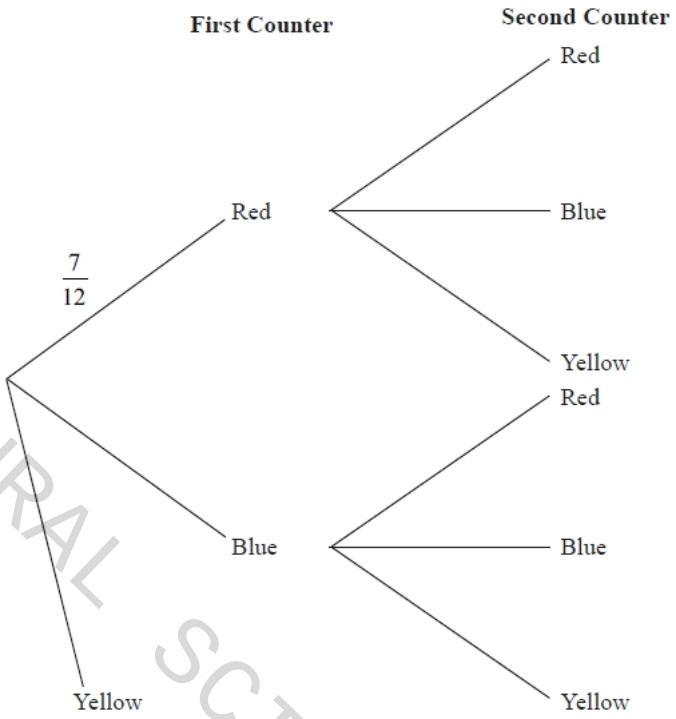
(**Q06 WST11/01, Specimen papers**)

Extra space for working:

Q6.

There are 7 red counters, 3 blue counters and 2 yellow counters in a bag. Gina selects a counter at random from the bag and keeps it. If the counter is yellow she does not select any more counters. If the counter is not yellow she randomly selects a second counter from the bag.

(a) Complete the tree diagram.



(2)

Given that Gina has selected a yellow counter,

(b) find the probability that she has 2 counters.

(3)

(Total for question = 5 marks)

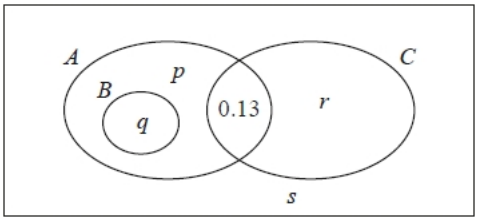
(Q01 WST01/01, June 2021)

Extra space for working:

Q7.

In the Venn diagram below, A, B and C are events and p, q, r and s are probabilities.

The events A and C are independent and $P(A) = 0.65$



(a) State which two of the events A, B and C are mutually exclusive.

(1)

(b) Find the value of r and the value of s.

(5)

The events $(A \cap C')$ and $(B \cup C)$ are also independent.

(c) Find the exact value of p and the exact value of q. Give your answers as fractions.

(6)

(Total for question = 12 marks)

(Q02 WST01/01, June 2021)

Extra space for working:

Q8.

The events *H* and *W* are such that

$$P(H) = \frac{3}{8} \qquad P(H \cup W) = \frac{3}{4}$$

Given that *H* and *W* are independent,

(a) show that P(W) = $\frac{3}{5}$

(4)

The event *N* is such that

$$P(N) = \frac{1}{15} \qquad P(H \cap N) = P(N)$$

(b) Find $P(N' | H)$

(2)

Given that *W* and *N* are mutually exclusive,

(c) draw a Venn diagram to represent the events *H*, *W* and *N* giving the exact probabilities of each region in the Venn diagram.

(5)

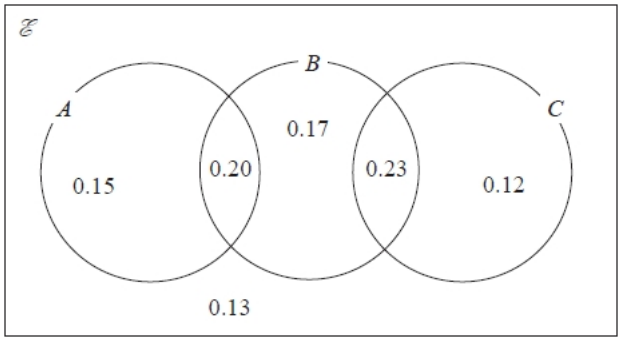
(Total for question = 11 marks)

(**Q04 WST01/01, June 2022**)

Extra space for working:

Q9.

The Venn diagram shows the events A, B and C and their associated probabilities.



Find

(a) P(B')

(1)

(b) P($A \cup C$)

(2)

(c) P($A \mid B'$)

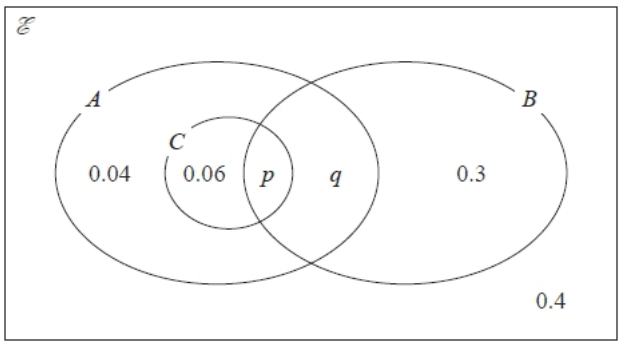
(2)

(Total for question = 5 marks)

(**Q01 WST01/01, Jan 2021**)

Q10.

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



(a) Find P(B)

(1)

(b) Determine whether or not A and B are independent.

(2)

Given that P(C | B) = P(C)

(c) find the value of p and the value of q

(3)

The event D is such that

- A and D are mutually exclusive
- P($B \cap D$) > 0

(d) On the Venn diagram show a possible position for the event D

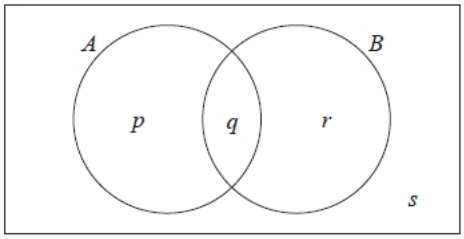
(1)

(Total for question = 7 marks)

(**Q01 WST01/01, Oct 2021**)

Q11.

(i) In the Venn diagram below, A and B represent events and p, q, r and s are probabilities.



$$P(A) = \frac{7}{25} \qquad P(B) = \frac{1}{5} \qquad P\left[(A \cap B') \cup (A' \cap B)\right] = \frac{8}{25}$$

(a) Use algebra to show that $2p + 2q + 2r = \frac{4}{5}$

(4)

(b) Find the value of p, the value of q, the value of r and the value of s

(5)

(ii) Two events, C and D, are such that

$$P(C) = \frac{x}{x+5} \qquad P(D) = \frac{5}{x}$$

where x is a positive constant.
By considering $P(C) + P(D)$ show that C and D**cannot** be mutually exclusive.

(4)

(Total for question = 13 marks)

(**Q04 WST01/01, Jan 2023**)

Q12.

A bag contains a large number of coloured counters. Each counter is labelled A, B or C

30% of the counters are labelled A
45% of the counters are labelled B
The rest of the counters are labelled C

It is known that

2% of the counters labelled A are red
4% of the counters labelled B are red
6% of the counters labelled C are red

One counter is selected at random from the bag.

(a) Complete the tree diagram below to illustrate this information.

(2)

(b) Calculate the probability that the counter is labelled A and is not red.

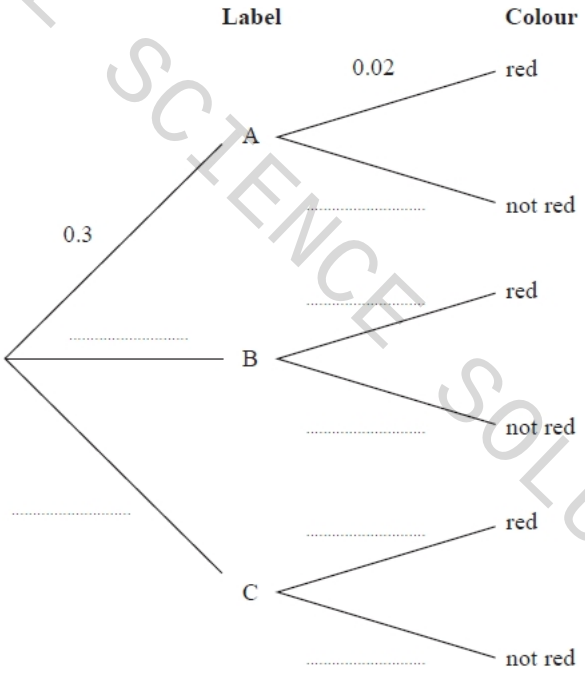
(2)

(c) Calculate the probability that the counter is red.

(2)

(d) Given that the counter is red, find the probability that it is labelled C

(3)



(Total for question = 9 marks)

(Q04 WST01/01, June 2023)

Extra space for working:

Q13.

Three bags **A**, **B** and **C** each contain coloured balls.

Bag **A** contains 4 red balls and 2 yellow balls only.
Bag **B** contains 4 red balls and 1 yellow ball only.
Bag **C** contains 6 red balls only.

In a game
Mike takes a ball at random from bag **A**, records the colour and places it in bag **C**.
He then takes a ball at random from bag **B**, records the colour and places it in bag **C**.
Finally, Mike takes a ball at random from bag **C** and records the colour.

(a) Complete the tree diagram below, to illustrate the game by adding the remaining branches and all probabilities.

(3)

(b) Show that the probability that Mike records a yellow ball exactly twice is $\frac{1}{10}$

(3)

Given that Mike records exactly 2 yellow balls,

(c) find the probability that the ball drawn from bag **A** is red.

(2)

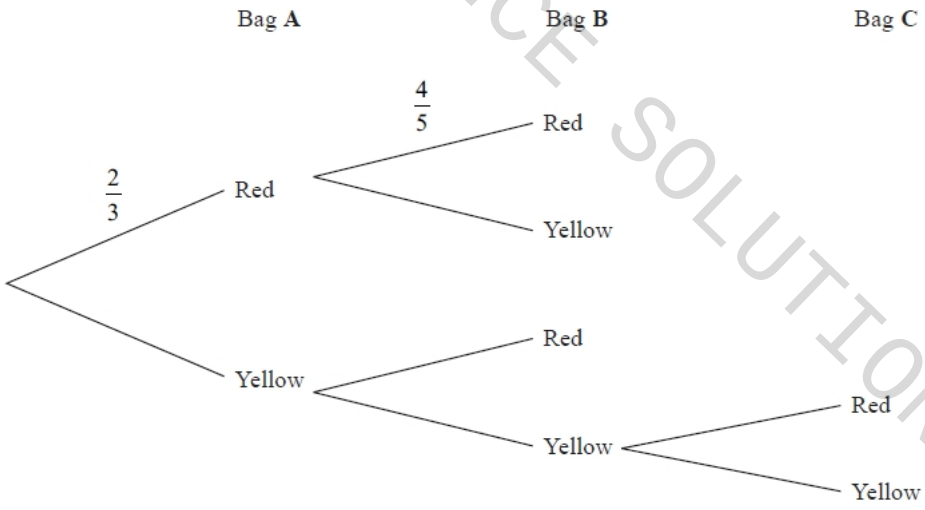
Mike plays this game a large number of times, each time starting with the bags containing balls as described above. The random variable X represents the number of yellow balls recorded in a single game.

(d) Find the probability distribution of X

(3)

(e) Find E(X)

(2)



(Total for question = 13 marks)

(**Q04 WST01/01, Oct 2021**)

Extra space for working:

Q14.

Two bags, ***X*** and ***Y***, each contain green marbles (G) and blue marbles (B) only.

- Bag ***X*** contains 5 green marbles and 4 blue marbles
- Bag ***Y*** contains 6 green marbles and 5 blue marbles

A marble is selected at random from bag ***X*** and placed in bag ***Y***
A second marble is selected at random from bag ***X*** and placed in bag ***Y***
A third marble is then selected, this time from bag ***Y***

(a) Use this information to complete the tree diagram shown below.

(3)

(b) Find the probability that the 2 marbles selected from bag ***X*** are of different colours.

(2)

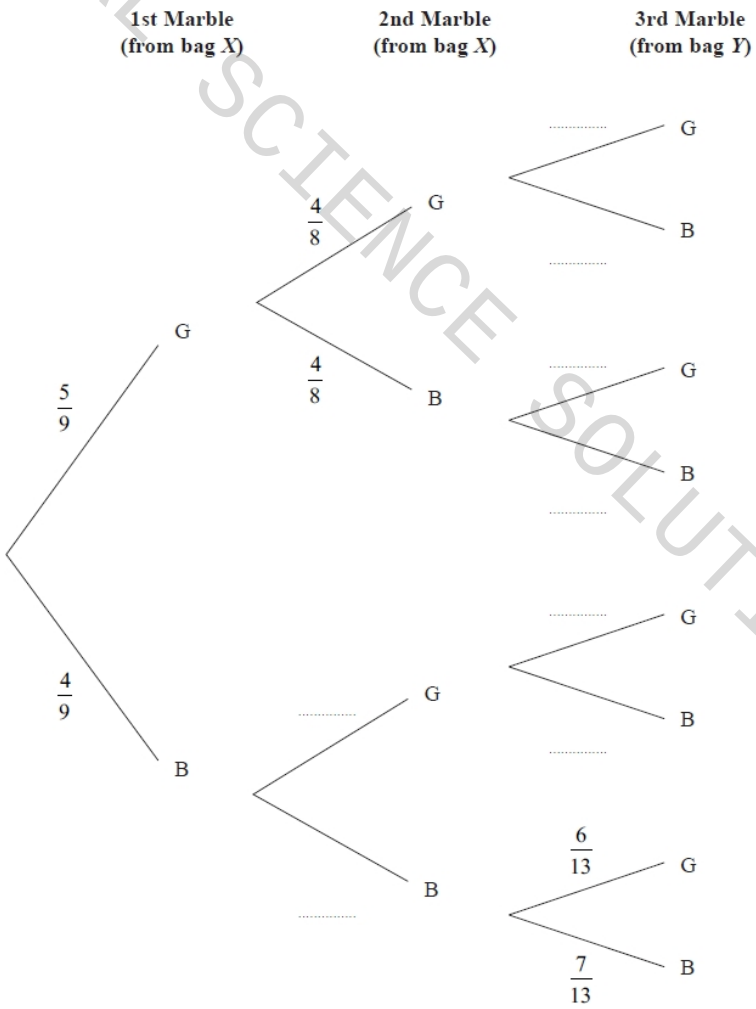
(c) Find the probability that all 3 marbles selected are the same colour.

(2)

Given that all three marbles selected are the same colour,

(d) find the probability that they are all green.

(3)



(Total for question = 10 marks)

(Q02 WST01/01, Jan 2023)

Extra space for working:

Q15.

The lengths, L mm, of housefly wings are normally distributed with $L \sim N(4.5, 0.4^2)$

(a) Find the probability that a randomly selected housefly has a wing length of less than 3.86 mm.

(3)

(b) Find

(i) the upper quartile (Q_3) of L
(ii) the lower quartile (Q_1) of L

(4)

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(c) Find these two outlier limits.

(3)

A housefly is selected at random.

(d) Using standardisation, show that the probability that this housefly is **not** an outlier is 0.993 to 3 decimal places.

(3)

Given that this housefly is **not** an outlier,

(e) showing your working, find the probability that the wing length of this housefly is greater than 5 mm.

(4)

(Total for question = 17 marks)

(**Q05 WST01/01, Jan 2023**)

Extra space for working:

Q16.

A disc of radius 1 cm is rolled onto a horizontal grid of rectangles so that the disc is equally likely to land anywhere on the grid. Each rectangle is 5 cm long and 3 cm wide. There are no gaps between the rectangles and the grid is sufficiently large so that no discs roll off the grid.

If the disc lands inside a rectangle without covering any part of the edges of the rectangle then a prize is won.

By considering the possible positions for the centre of the disc,

(a) show that the probability of winning a prize on any particular roll is $\frac{1}{5}$

(3)

A group of 15 students each roll the disc onto the grid twenty times and record the number of times, x, that each student wins a prize. Their results are summarised as follows

$$\sum x = 61 \qquad \sum x^2 = 295$$

(b) Find the standard deviation of the number of prizes won per student.

(2)

A second group of 12 students each roll the disc onto the grid twenty times and the mean number of prizes won per student is 3.5 with a standard deviation of 2

(c) Find the mean and standard deviation of the number of prizes won per student for the whole group of 27 students.

(7)

The 27 students also recorded the number of times that the disc covered a corner of a rectangle and estimated the probability to be 0.2216 (to 4 decimal places).

(d) Explain how this probability could be used to find an estimate for the value of π and state the value of your estimate.

(3)

(Total for question = 15 marks)

(**Q06 WST01/01, Jan 2021**)

Extra space for working:

Q17.

Xiang is designing shelves for a bookshop. The height, H cm, of books is modelled by the normal distribution with mean 25.1 cm and standard deviation 5.5 cm

(a) Show that P(H > 30.8) = 0.15

(3)

Xiang decided that the smallest 5% of books and books taller than 30.8 cm would not be placed on the shelves. All the other books will be placed on the shelves.

(b) Find the range of heights of books that will be placed on the shelves.

(3)

The books that will be placed on the shelves have heights classified as small, medium or large.
The numbers of small, medium and large books are in the ratios 2 : 3 : 3

(c) The medium books have heights x cm where $m < x < d$

(i) Show that d = 25.8 to 1 decimal place.

(3)

(ii) Find the value of m

(4)

Xiang wants 2 shelves for small books, 3 shelves for medium books and 3 shelves for large books.
These shelves will be placed one above another and made of wood that is 1 cm thick.

(d) Work out the minimum total height needed.

(2)

(Total for question = 15 marks)

(**Q06 WST01/01, Oct 2021**)

Extra space for working:

Q18.

A biased tetrahedral die has faces numbered 0, 1, 2 and 3. The die is rolled and the number face down on the die, X, is recorded. The probability distribution of X is

x	0	1	2	3
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$

If $X = 3$ then the final score is 3
If $X \neq 3$ then the die is rolled again and the final score is the sum of the two numbers.

The random variable T is the final score.

(a) Find $P(T = 2)$

(2)

(b) Find $P(T = 3)$

(3)

(c) Given that the die is rolled twice, find the probability that the final score is 3

(3)

(Total for question = 8 marks)

(**Q05 WST11/01, Specimen papers**)

Q19.

Kris works in the mailroom of a large company and is responsible for all the letters sent by the company. The weights of letters sent by the company, W grams, have a normal distribution with mean 165 g and standard deviation 35 g.

(a) Estimate the proportion of letters sent by the company that weigh less than 120 g.

(3)

Kris splits the letters to be sent into 3 categories: heavy, medium and light, with $\frac{1}{3}$ of the letters in each category.

(b) Find the weight limits that determine medium letters.

(4)

A heavy letter is chosen at random.

(c) Find the probability that this letter weighs less than 200 g.

(3)

Kris chooses a random sample of 3 letters from those in the mailroom one day.

(d) Find the probability that there is one letter in each of the 3 categories.

(3)

(Total for question = 13 marks)

(**Q04 WST01/01, June 2021**)

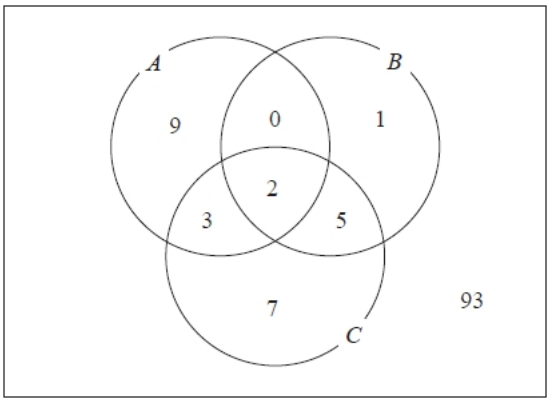
Extra space for working:

Q20.

A factory produces shoes.

A quality control inspector at the factory checks a sample of 120 shoes for each of three types of defect.
The Venn diagram represents the inspector's results.

A represents the event that a shoe has defective stitching
B represents the event that a shoe has defective colouring
C represents the event that a shoe has defective soles



One of the shoes in the sample is selected at random.

(a) Find the probability that it does **not** have defective soles.

(1)

(b) Find P($A \cap B \cap C'$)

(1)

(c) Find P($A \cup B \cup C'$)

(2)

(d) Find the probability that the shoe has at most one type of defect.

(2)

(e) Given the selected shoe has at most one type of defect, find the probability it has defective stitching.

(2)

The random variable X is the number of the events A, B, C that occur for a randomly selected shoe.

(f) Find E(X)

(3)

(Total for question = 11 marks)

(Q01 WST01/01, Jan 2022)

Extra space for working:

Q21.

Gill buys a bag of logs to use in her stove. The lengths, l cm, of the 88 logs in the bag are summarised in the table below.

Length (l)	Frequency (f)
$15 < l \leqslant 20$	19
$20 < l \leqslant 25$	35
$25 < l \leqslant 27$	16
$27 < l \leqslant 30$	15
$30 < l \leqslant 40$	3

A histogram is drawn to represent these data.

The bar representing logs with length $27 < l \leqslant 30$ has a width of 1.5 cm and a height of 4 cm.

(a) Calculate the width and height of the bar representing log lengths of $20 < l \leqslant 25$

(3)

(b) Use linear interpolation to estimate the median of l

(2)

The maximum length of log Gill can use in her stove is 26 cm.

Gill estimates, using linear interpolation, that x logs from the bag will fit into her stove.

(c) Show that $x = 62$

(1)

Gill randomly selects 4 logs from the bag.

(d) Using $x = 62$, find the probability that all 4 logs will fit into her stove.

(2)

The weights, W grams, of the logs in the bag are coded using $y = 0.5w - 255$ and summarised by

$$n = 88 \quad \sum y = 924 \quad \sum y^2 = 12\,862$$

(e) Calculate

(i) the mean of W

(3)

(ii) the variance of W

(3)

(Total for question = 14 marks)

(Q03 WST01/01, June 2022)

Extra space for working:

Q22.

The stem and leaf diagram shows the number of deliveries made by Pat each day for 24 days

Key: 10 | 8 represents 108 deliveries

10	8	9										(2)
11	0	3	6	6	6	8	8	9	9	9	9	(11)
12	4	5	5	5	5	5	5	8				(8)
13	a	b	c									(3)

where a, b and c are positive integers with $a < b < c$

An outlier is defined as any value greater than 1.5 × interquartile range above the upper quartile.

Given that there is only one outlier for these data,

(a) show that $c = 9$

(3)

The number of deliveries made by Pat each day is represented by d

The data in the stem and leaf diagram are coded using

$$x = d - 125$$

and the following summary statistics are obtained

$$\sum x = -96 \quad \text{and} \quad \sum (x - \bar{x})^2 = 1306$$

(b) Find the mean number of deliveries.

(3)

(c) Find the standard deviation of the number of deliveries.

(2)

One of these 24 days is selected at random. The random variable D represents the number of deliveries made by Pat on this day.

The random variable $X = D - 125$

(d) Find $P(D > 118 \mid X < 0)$

(2)

(Total for question = 10 marks)

(Q03 WST01/01, Jan 2022)

Extra space for working:

Q23.

A red spinner is designed so that the score R is given by the following probability distribution.

r	2	3	4	5	6
$P(R = r)$	0.25	0.3	0.15	0.1	0.2

(a) Show that $E(R^2) = 15.8$

(1)

Given also that $E(R) = 3.7$

(b) find the standard deviation of R, giving your answer to 2 decimal places.

(2)

A yellow spinner is designed so that the score Y is given by the probability distribution in the table below. The cumulative distribution function $F(y)$ is also given.

y	2	3	4	5	6
$P(Y = y)$	0.1	0.2	0.1	a	b
$F(y)$	0.1	0.3	0.4	c	d

(c) Write down the value of d

(1)

Given that $E(Y) = 4.55$

(d) find the value of c

(5)

Pabel and Jessie play a game with these two spinners.

Pabel uses the red spinner.

Jessie uses the yellow spinner.

They take turns to spin their spinner.

The winner is the first person whose spinner lands on the number 2 and the game ends.

Jessie spins her spinner first.

(e) Find the probability that Jessie wins on her second spin.

(2)

(f) Calculate the probability that, in a game, the score on Pabel's first spin is the same as the score on Jessie's first spin.

(3)

(Total for question = 14 marks)

(Q05 WST01/01, June 2022)

Extra space for working:

1.3 Correlation and Regression:

Reference Notes:

Q1.

The production cost, £c million, of a film and the total ticket sales, £t million, earned by the film are recorded for a sample of 40 films.

Some summary statistics are given below.

$$\sum c = 1634 \quad \sum t = 1361 \quad \sum t^2 = 82\,873 \quad \sum ct = 83\,634 \quad S_{cc} = 28\,732.1$$

(a) Find the exact value of S_{tt} and the exact value of S_{ct}

(3)

(b) Calculate the value of the product moment correlation coefficient for these data.

(2)

(c) Give an interpretation of your answer to part (b)

(1)

(d) Show that the equation of the linear regression line of t on c can be written as

$$t = -5.84 + 0.976c$$

where the values of the intercept and gradient are given to 3 significant figures.

(3)

(e) Find the expected total ticket sales for a film with a production cost of £90 million.

(2)

Using the regression line in part (d)

(f) find the range of values of the production cost of a film for which the total ticket sales are less than 80% of its production cost.

(2)

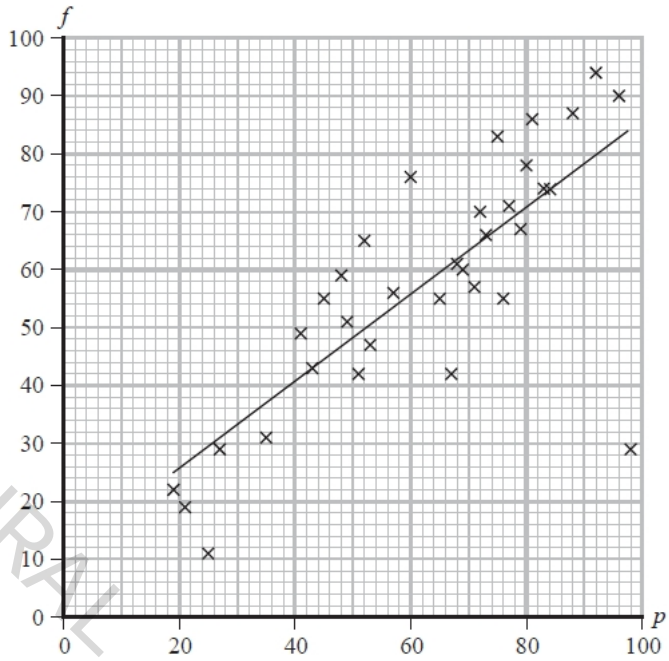
(Total for question = 13 marks)

(**Q02 WST01/01, Oct 2022**)

Extra space for working:

Q2.

Students on a psychology course were given a pre-test at the start of the course and a final exam at the end of the course. The teacher recorded the number of marks achieved on the pre-test, p, and the number of marks achieved on the final exam, f, for 34 students and displayed them on the scatter diagram.



The equation of the least squares regression line for these data is found to be

$$f = 10.8 + 0.748\,p$$

For these students, the mean number of marks on the pre-test is 62.4

(a) Use the regression model to find the mean number of marks on the final exam.

(2)

(b) Give an interpretation of the gradient of the regression line.

(1)

Considering the equation of the regression line, Priya says that she would expect someone who scored 0 marks on the pre-test to score 10.8 marks on the final exam.

(c) Comment on the reliability of Priya's statement.

(1)

(d) Write down the number of marks achieved on the final exam for the student who exceeded the expectation of the regression model by the largest number of marks.

(1)

(e) Find the range of values of p for which this regression model, $f = 10.8 + 0.748\,p$, predicts a greater number of marks on the final exam than on the pre-test.

(3)

Later the teacher discovers an error in the recorded data. The student who achieved a score of 98 on the pre-test, scored 92 not 29 on the final exam.

The summary statistics used for the model $f = 10.8 + 0.748p$ are corrected to include this information and a new least squares regression line is found.

Given the **original** summary statistics were,

$$n = 34 \quad \sum p = 2120 \quad \sum pf = 133\,486 \quad S_{pp} = 15\,573.76 \quad S_{pf} = 11\,648.35$$

(f) calculate the gradient of the new regression line. Show your working clearly.

(5)

(Total for question = 13 marks)

(**Q06 WST01/01, Jan 2022**)

Extra space for working:

Q3.

Two economics students, Andi and Behrouz, are studying some data relating to unemployment, x %, and increase in wages, y %, for a European country. The least squares regression line of y on x has equation

$$y = 3.684 - 0.3242x$$

and $\sum y = 23.7 \quad \sum y^2 = 42.63 \quad \sum x^2 = 756.81 \quad n = 16$

(a) Show that $S_{yy} = 7.524\,375$

(1)

(b) Find S_{xx}

(4)

(c) Find the product moment correlation coefficient between x and y.

(3)

Behrouz claims that, assuming the model is valid, the data show that when unemployment is 2% wages increase at over 3%

(d) Explain how Behrouz could have come to this conclusion.

(1)

Andi uses the formula

$$\text{range} = \text{mean} \pm 3 \times \text{standard deviation}$$

to estimate the range of values for x.

(e) Find estimates of the minimum value and the maximum value of x in these data using Andi's formula.

(3)

(f) Comment, giving a reason, on the reliability of Behrouz's claim.

(2)

Andi suggests using the regression line with equation $y = 3.684 - 0.3242x$ to estimate unemployment when wages are increasing at 2%

(g) Comment, giving a reason, on Andi's suggestion.

(2)

(Total for question = 16 marks)

(**Q06 WST01/01, June 2021**)

Extra space for working:

Q4.

A large company is analysing how much money it spends on paper in its offices each year. The number of employees in the office, x, and the amount spent on paper in a year, p ($ hundreds), in each of 12 randomly selected offices were recorded.

The results are summarised in the following statistics.

$$\sum x = 93 \qquad S_{xx} = 148.25 \qquad \sum p = 273 \qquad \sum p^2 = 6602.72 \qquad \sum xp = 2347$$

(a) Show that $S_{xp} = 231.25$

(1)

(b) Find the product moment correlation coefficient for these data.

(3)

(c) Find the equation of the regression line of p on x in the form $p = a + bx$

(4)

(d) Give an interpretation of the gradient of your regression line.

(1)

The director of the company wants to reduce the amount spent on paper each year.

He wants each office to aim for a model of the form $p = \frac{4}{5}a + \frac{1}{2}bx$ where a and b are the values found in part (c).

Using the data for the 93 employees from the 12 offices,

(e) estimate the percentage saving in the amount spent on paper each year by the company using the director's model.

(3)

(Total for question = 12 marks)

(**Q02 WST01/01, Oct 2021**)

Extra space for working:

Q5.

Stuart is investigating the relationship between Gross Domestic Product (GDP) and the size of the population for a particular country. He takes a random sample of 9 years and records the size of the population, t millions, and the GDP, g billion dollars for each of these years.

The data are summarised as

$$n = 9 \sum t = 7.87 \sum g = 144.84 \sum g^2 = 3624.41 \; S_{tt} = 1.29 \; S_{tg} = 40.25$$

(a) Calculate the product moment correlation coefficient between t and g

(3)

(b) Give an interpretation of your product moment correlation coefficient.

(1)

(c) Find the equation of the least squares regression line of g on t in the form $g = a + bt$

(4)

(d) Give an interpretation of the value of b in your regression line.

(1)

(e) (i) Use the regression line from part (c) to estimate the GDP, in billions of dollars, for a population of 7 000 000

(2)

(ii) Comment on the reliability of your answer in part (i). Give a reason, in context, for your answer.

(1)

Using the regression line from part (c), Stuart estimates that for a population increase of x million there will be an increase of 0.1 billion dollars in GDP.

(f) Find the value of x

(2)

(Total for question = 14 marks)

(**Q02 WST01/01, June 2022**)

Extra space for working:

Q6.

A company director wants to introduce a performance-related pay structure for her managers. A random sample of 15 managers is taken and the annual salary, y in £1000, was recorded for each manager. The director then calculated a performance score, x, for each of these managers.
The results are shown on the scatter diagram in Figure 1 on the next page.

(a) Describe the correlation between performance score and annual salary.

(1)

The results are also summarised in the following statistics.

$$\sum x = 465 \quad \sum y = 562 \quad S_{xx} = 2492 \quad \sum y^2 = 23140 \quad \sum xy = 19428$$

(b) (i) Show that $S_{xy} = 2006$

(1)

(ii) Find S_{yy}

(2)

(c) Find the product moment correlation coefficient between performance score and annual salary.

(2)

The director believes that there is a linear relationship between performance score and annual salary.

(d) State, giving a reason, whether or not these data are consistent with the director's belief.

(1)

(e) Calculate the equation of the regression line of y on x, in the form $y = a + bx$
Give the value of a and the value of b to 3 significant figures.

(4)

(f) Give an interpretation of the value of b.

(1)

(g) Plot your regression line on the scatter diagram in Figure 1

(2)

The director hears that one of the managers in the sample seems to be underperforming.

(h) On the scatter diagram, circle the point that best identifies this manager.

(1)

The director decides to use this regression line for the new performance related pay structure.

(i) Estimate, to 3 significant figures, the new salary of a manager with a performance score of 30

(2)

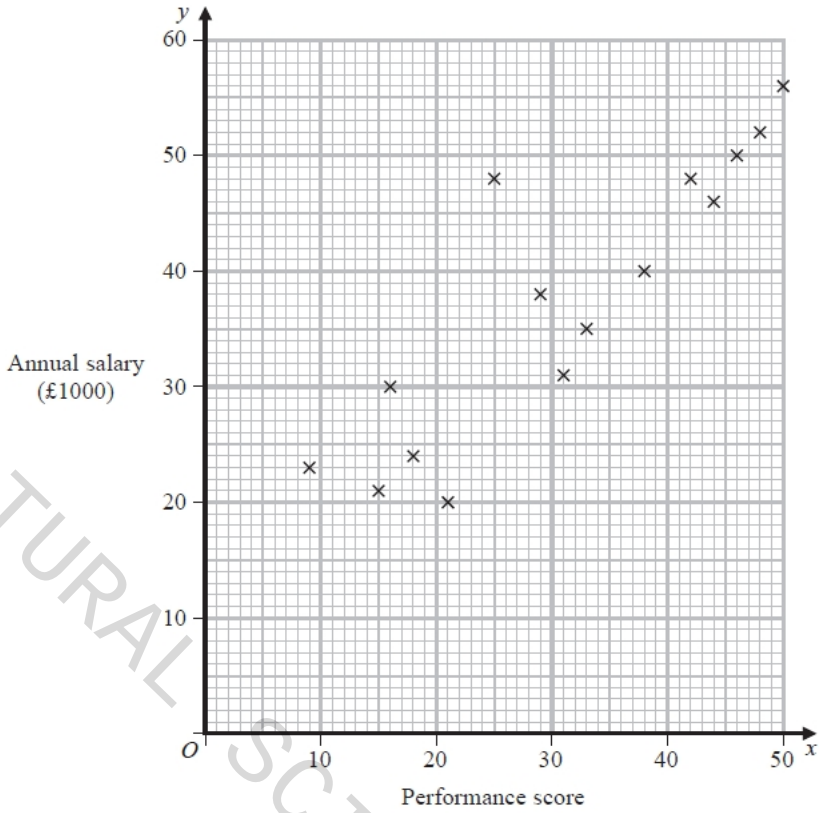


Figure 1

(Total for question = 17 marks)

(Q05 WST01/01, Jan 2021)

Extra space for working:

Q7.

A research student is investigating the maximum weight, y grams, of sugar that will dissolve in 100 grams of water at various temperatures, x °C, where $10 \leq x \leq 80$

The research student calculated the regression line of y on x and found it to be

$$y = 151.2 + 2.72x$$

(a) Give an interpretation of the gradient of the regression line.

(1)

(b) Use the regression line to estimate the maximum weight of sugar that will dissolve in 100 grams of water when the temperature is 90 °C.

(2)

(c) Comment on the reliability of your estimate, giving a reason for your answer.

(2)

Using the regression line of y on x and the following summary statistics

$$\sum y = 3119 \qquad \sum y^2 = 851\,093 \qquad \sum x^2 = 24\,500 \qquad n = 12$$

(d) show that the product moment correlation coefficient for these data is 0.988 to 3 decimal places.

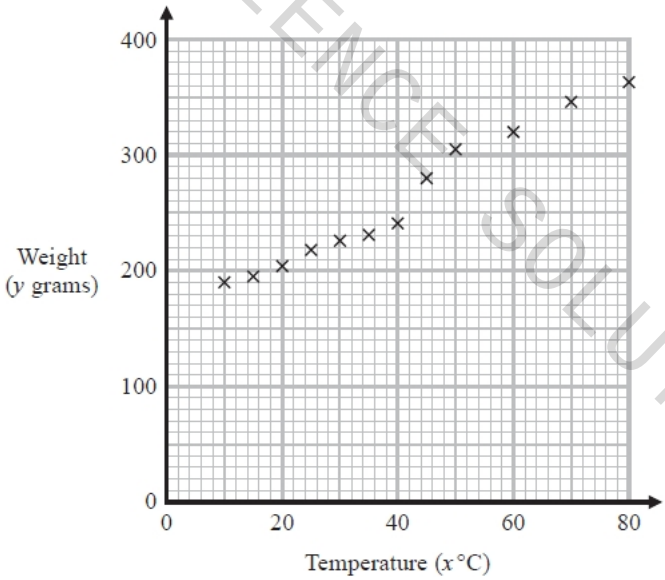
(7)

The research student's supervisor plotted the original data on a scatter diagram, below.

With reference to both the scatter diagram and the correlation coefficient,

(e) discuss the suitability of a linear regression model to describe the relationship between x and y.

(2)



(Total for question = 14 marks)

(Q06 WST01/01, Jan 2023)

Extra space for working:

Q8.

The variables x and y have the following regression equations based on the same 12 observations.

	Regression equation
y on x	$y = 1.4x + 1.5$
x on y	$x = 1.2 + 0.2y$

(a) (i) Find the point of intersection of these lines.

(ii) Hence show that $\sum x = 25$

(4)

Given that

$$\sum xy = \frac{6961}{60}$$

(b) Find S_{xy}

(4)

(c) Find the product moment correlation coefficient between x and y

(4)

(Total for question = 12 marks)

(**Q06 WST01/01, Oct 2023**)

Q9.

Two students, Olive and Shan, collect data on the weight, w grams, and the tail length, t cm, of 15 mice.

Olive summarised the data as follows

$$S_{tt} = 5.3173 \qquad \sum w^2 = 6089.12 \qquad \sum tw = 2304.53 \qquad \sum w = 297.8 \qquad \sum t = 114.8$$

(a) Calculate the value of S_{tw} and the value of S_{ww}

(3)

(b) Calculate the value of the product moment correlation coefficient between w and t

(2)

(c) Show that the equation of the regression line of w on t can be written as

$$w = -16.7 + 4.77t$$

(3)

(d) Give an interpretation of the gradient of the regression line.

(1)

(e) Explain why it would not be appropriate to use the regression line in part (c) to estimate the weight of a mouse with a tail length of 2 cm.

(2)

Shan decided to code the data using $x = t - 6$ and $y = \frac{w}{2} - 5$

(f) Write down the value of the product moment correlation coefficient between x and y

(1)

(g) Write down an equation of the regression line of y on x

You do not need to simplify your equation.

(1)

(Total for question = 13 marks)

(Q02 WST01/01, June 2023)

Extra space for working:

Q10.

Tom's car holds 50 litres of petrol when the fuel tank is full.

For each of 10 journeys, each starting with 50 litres of petrol in the fuel tank, Tom records the distance travelled, d kilometres, and the amount of petrol used, p litres.

The summary statistics for the 10 journeys are given below.

$$\sum d = 1029 \quad \sum p = 50.8 \quad \sum dp = 5240.8 \quad S_{dd} = 344.9 \quad S_{pp} = 0.576$$

(a) Calculate the product moment correlation coefficient between d and p

(3)

The amount of petrol remaining in the fuel tank for each journey, w litres, is recorded.

(b) (i) Write down an equation for w in terms of p

(ii) Hence, write down the value of the product moment correlation coefficient between w and p

(2)

(c) Write down the value of the product moment correlation coefficient between d and w

(1)

(Total for question = 6 marks)

(**Q02 WST01/01, Jan 2022**)

Extra space for working:

Q11.

The percentage oil content, p, and the weight, w milligrams, of each of 10 randomly selected sunflower seeds were recorded. These data are summarised below.

$$\sum w^2 = 41252 \qquad \sum wp = 27557.8 \qquad \sum w = 640 \qquad \sum p = 431 \qquad S_{pp} = 2.72$$

(a) Find the value of S_{ww} and the value of S_{wp}

(3)

(b) Calculate the product moment correlation coefficient between p and w

(2)

(c) Give an interpretation of your product moment correlation coefficient.

(1)

The equation of the regression line of p on w is given in the form $p = a + bw$

(d) Find the equation of the regression line of p on w

(4)

(e) Hence estimate the percentage oil content of a sunflower seed which weighs 60 milligrams.

(2)

(Total for question = 12 marks)

(**Q01 WST11/01, Specimen papers**)

1.4 Discrete Random Variable:

Reference Notes:

Q1.

The discrete random variable X has probability distribution

$$P(X = x) = \frac{1}{5} \qquad x = 1, 2, 3, 4, 5$$

(a) Write down the name given to this distribution.

(1)

Find

(b) P(X = 4)

(1)

(c) F(3)

(1)

(d) P($3X - 3 > X + 4$)

(2)

(e) Write down E(X)

(1)

(f) Find E(X^2)

(2)

(g) Hence find Var (X)

(2)

Given that E($aX - 3$) = 11.4

(h) find Var ($aX - 3$)

(4)

(Total for question = 14 marks)

(Q03 WST11/01, Specimen papers)

Extra space for working:

Q2.

The discrete random variable X has the following probability distribution

x	-2	-1	0	1	4
$P(X = x)$	a	b	c	b	a

Given that $E(X) = 0.5$

(a) find the value of a.

(2)

Given also that $Var(X) = 5.01$

(b) find the value of b and the value of c.

(5)

The random variable $Y = 5 - 8X$

(c) Find (i) $E(Y)$

(ii) $Var(Y)$

(3)

(d) Find $P(4X^2 > Y)$

(5)

(Total for question = 15 marks)

(**Q05 WST01/01, June 2021**)

Extra space for working:

Q3.

A red spinner is designed so that the score R is given by the following probability distribution.

r	2	3	4	5	6
$P(R = r)$	0.25	0.3	0.15	0.1	0.2

(a) Show that $E(R^2) = 15.8$

(1)

Given also that $E(R) = 3.7$

(b) find the standard deviation of R, giving your answer to 2 decimal places.

(2)

A yellow spinner is designed so that the score Y is given by the probability distribution in the table below. The cumulative distribution function $F(y)$ is also given.

y	2	3	4	5	6
$P(Y = y)$	0.1	0.2	0.1	a	b
$F(y)$	0.1	0.3	0.4	c	d

(c) Write down the value of d

(1)

Given that $E(Y) = 4.55$

(d) find the value of c

(5)

Pabel and Jessie play a game with these two spinners.

Pabel uses the red spinner.

Jessie uses the yellow spinner.

They take turns to spin their spinner.

The winner is the first person whose spinner lands on the number 2 and the game ends.

Jessie spins her spinner first.

(e) Find the probability that Jessie wins on her second spin.

(2)

(f) Calculate the probability that, in a game, the score on Pabel's first spin is the same as the score on Jessie's first spin.

(3)

(Total for question = 14 marks)

(Q05 WST01/01, June 2022)

Extra space for working:

Q4.

The random variable W has a discrete uniform distribution where

$$P(W = w) = \frac{1}{5} \quad \text{for } w = 1, 2, 3, 4, 5$$

(a) Find $P(2 \leq W < 3.5)$

(1)

The discrete random variable $X = 5 - 2W$

(b) Find $E(X)$

(3)

(c) Find $P(X < W)$

(2)

The discrete random variable $Y = \frac{1}{W}$

(d) Find

(i) the probability distribution of Y
(ii) $\text{Var}(Y)$, showing your working.

(5)

(e) Find $\text{Var}(2 - 3Y)$

(2)

(Total for question = 13 marks)

(**Q04 WST01/01, Jan 2022**)

Extra space for working:

Q5.

A spinner can land on the numbers 10, 12, 14 and 16 only and the probability of the spinner landing on each number is the same.

The random variable X represents the number that the spinner lands on when it is spun once.

(a) State the name of the probability distribution of X.

(1)

(b) (i) Write down the value of E(X)

(1)

(ii) Find Var(X)

(2)

A second spinner can land on the numbers 1, 2, 3, 4 and 5 only.
The random variable Y represents the number that this spinner lands on when it is spun once. The probability distribution of Y is given in the table below

y	1	2	3	4	5
$P(Y = y)$	$\frac{4}{30}$	$\frac{9}{30}$	$\frac{6}{30}$	$\frac{5}{30}$	$\frac{6}{30}$

(c) Find (i) E(Y)

(2)

(ii) Var(Y)

(3)

The random variable $W = aX + b$, where a and b are constants and $a > 0$
Given that E(W) = E(Y) and Var(W) = Var(Y)

(d) find the value of a and the value of b.

(5)

Each of the two spinners is spun once.

(e) Find P($W = Y$)

(2)

(Total for question = 16 marks)

(Q04 WST01/01, Jan 2021)

Extra space for working:

Q6.

The discrete random variable Y has the following probability distribution

y	-9	-5	0	5	9
$P(Y = y)$	q	r	u	r	q

where q, r and u are probabilities.

(a) Write down the value of E(Y)

(1)

The cumulative distribution function of Y is F(y)

Given that F(0) = $\frac{19}{30}$

(b) show that the value of u is $\frac{4}{15}$

(3)

Given also that Var (Y) = 37

(c) find the value of q and the value of r

(4)

The coordinates of a point P are (12, Y)

The random variable D represents the length of OP

(d) Find the probability distribution of D

(6)

(Total for question = 14 marks)

(**Q05 WST01/01, Oct 2021**)

Q7.

The cumulative distribution function of the discrete random variable W, which takes only the values 6, 7 and 8, is given by

$$F(W) = \frac{(w+3)(w-1)}{77} \quad \text{for } w = 6, 7, 8$$

Find E(W)

(4)

(Total for question = 4 marks)

(**Q04 WST01/01, Oct 2022**)

Q8.

The discrete random variable X has the following probability distribution.

x	1	2	3	4
$P(X = x)$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{2}{5}$

(a) Show that $E\left(\frac{1}{X}\right) = \frac{2}{5}$

(1)

(b) Find $\text{Var}\left(\frac{1}{X}\right)$

(3)

The random variable $Y = \frac{30}{X}$

(c) Find

(i) E(Y)
(ii) Var(Y)

(3)

(d) Find P($X < 3 | Y < 20$)

(5)

(Total for question = 12 marks)

(**Q04 WST01/01, Oct 2023**)

Extra space for working:

Q9.

A discrete random variable Y has probability function

$$P(Y = y) = \begin{cases} k(3-y) & y = 1, 2 \\ k(y^2 - 8) & y = 3, 4, 5 \\ k & y = 6 \\ 0 & \text{otherwise} \end{cases}$$

where k is a constant.

(a) Show that $k = \frac{1}{30}$

(2)

Find the exact value of

(b) $P(1 < Y \leq 4)$

(2)

(c) $E(Y)$

(2)

The random variable $X = 15 - 2Y$

(d) Calculate $P(Y \geq X)$

(3)

(e) Calculate Var (X)

(4)

(Total for question = 13 marks)

(Q05 WST01/01, June 2023)

Extra space for working:

Q10.

A biased tetrahedral die has faces numbered 0, 1, 2 and 3. The die is rolled and the number face down on the die, X, is recorded. The probability distribution of X is

x	0	1	2	3
$P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{2}$

If $X = 3$ then the final score is 3
If $X \neq 3$ then the die is rolled again and the final score is the sum of the two numbers.

The random variable T is the final score.

(a) Find $P(T = 2)$

(2)

(b) Find $P(T = 3)$

(3)

(c) Given that the die is rolled twice, find the probability that the final score is 3

(3)

(Total for question = 8 marks)

(**Q05 WST11/01, Specimen papers**)

Extra space for working:

Q11.

Adana selects one number at random from the distribution of X which has the following probability distribution.

x	0	5	10
$P(X = x)$	0.1	0.2	0.7

(a) Given that the number selected by Adana is not 5, write down the probability it is 0

(1)

(b) Show that $E(X^2) = 75$

(1)

(c) Find $Var(X)$

(3)

(d) Find $Var(4 - 3X)$

(2)

Bruno and Charlie each independently select one number at random from the distribution of X

(e) Find the probability that the number Bruno selects is greater than the number Charlie selects.

(3)

Devika multiplies Bruno's number by Charlie's number to obtain a product, D

(f) Determine the probability distribution of D

(4)

(Total for question = 14 marks)

(**Q07 WST01/01, Oct 2022**)

Extra space for working:

Q12.

The probability distribution of the discrete random variable X is given by

x	2	3	4
$P(X = x)$	a	0.4	$0.6 - a$

where a is a constant.

(a) Find, in terms of a, E(X)

(2)

(b) Find the range of the possible values of E(X)

(3)

Given that Var(X) = 0.56

(c) find the possible values of a

(6)

(Total for question = 11 marks)

(**Q03 WST01/01, Jan 2023**)

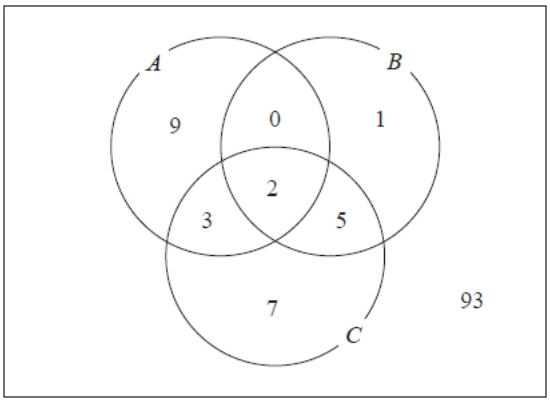
Extra space for working:

Q13.

A factory produces shoes.

A quality control inspector at the factory checks a sample of 120 shoes for each of three types of defect. The Venn diagram represents the inspector's results.

A represents the event that a shoe has defective stitching
B represents the event that a shoe has defective colouring
C represents the event that a shoe has defective soles



One of the shoes in the sample is selected at random.

(a) Find the probability that it does **not** have defective soles.

(1)

(b) Find P($A \cap B \cap C'$)

(1)

(c) Find P($A \cup B \cup C'$)

(2)

(d) Find the probability that the shoe has at most one type of defect.

(2)

(e) Given the selected shoe has at most one type of defect, find the probability it has defective stitching.

(2)

The random variable X is the number of the events A, B, C that occur for a randomly selected shoe.

(f) Find E(X)

(3)

(Total for question = 11 marks)

(**Q01 WST01/01, Jan 2022**)

1.5 The Normal Distribution:

Reference Notes:

Q1.

A machine fills bottles with water. The volume of water delivered by the machine to a bottle is X ml where $X \sim N(\mu, \sigma^2)$

One of these bottles of water is selected at random.

Given that $\mu = 503$ and $\sigma = 1.6$

(a) find

(i) $P(X > 505)$
(ii) $P(501 < X < 505)$

(5)

(b) Find w such that $P(1006 - w < X < w) = 0.9426$

(3)

Following adjustments to the machine, the volume of water delivered by the machine to a bottle is such that $\mu = 503$ and $\sigma = q$

Given that $P(X < r) = 0.01$ and $P(X > r + 6) = 0.05$

(c) find the value of r and the value of q

(7)

(Total for question = 15 marks)

(**Q07 WST11/01, Specimen papers**)

Extra space for working:

Q2.

The weights of packages that arrive at a factory are normally distributed with a mean of 18 kg and a standard deviation of 5.4 kg

(a) Find the probability that a randomly selected package weighs less than 10 kg

(3)

The heaviest 15% of packages are moved around the factory by Jemima using a forklift truck.

(b) Find the weight, in kg, of the lightest of these packages that Jemima will move.

(3)

One of the packages **not** moved by Jemima is selected at random.

(c) Find the probability that it weighs more than 18 kg

(4)

A delivery of 4 packages is made to the factory.
The weights of the packages are independent.

(d) Find the probability that exactly 2 of them will be moved by Jemima.

(3)

(Total for question = 13 marks)

(**Q03 WST01/01, Jan 2021**)

Extra space for working:

Q3.

A manufacturer fills bottles with oil.

The volume of oil in a bottle, Vml, is normally distributed with $V \sim N(100, 2.5^2)$

(a) Find $P(V > 104.9)$

(3)

(b) In a pack of 150 bottles, find the expected number of bottles containing more than 104.9 ml

(2)

(c) Find the value of v, to 2 decimal places, such that $P(V > v \mid V < 104.9) = 0.2801$

(6)

(Total for question = 11 marks)

(**Q06 WST01/01, June 2022**)

Extra space for working:

Q4.

The weights, *W* grams, of kiwi fruit grown on a farm are normally distributed with mean 80 grams and standard deviation 8 grams.

The table shows the classifications of the kiwi fruit by their weight, where *k* is a positive constant.

Small		Large		
Tiny	Petite	Extra	Jumbo	Mega
$w < 66$	$66 \leqslant w < 70$	$70 \leqslant w < 80$	$80 \leqslant w < k$	$w \geqslant k$

One kiwi fruit is selected at random from those grown on the farm.

(a) Find the probability that this kiwi fruit is Large.

(3)

35% of the kiwi fruit are Jumbo.

(b) Find the value of *k* to one decimal place.

(4)

75% of Tiny kiwi fruit weigh more than *y* grams.

(c) Find the value of *y* giving your answer to one decimal place.

(5)

(Total for question = 12 marks)

(**Q05 WST01/01, Oct 2022**)

Extra space for working:

Q5.

The weights, X grams, of a particular variety of fruit are normally distributed with

$$X \sim \mathrm{N}(210, 25^2)$$

A fruit of this variety is selected at random.

(a) Show that the probability that the weight of this fruit is less than 240 grams is 0.8849

(2)

(b) Find the probability that the weight of this fruit is between 190 grams and 240 grams.

(2)

(c) Find the value of k such that $\mathrm{P}(210 - k < X < 210 + k) = 0.95$

(3)

A wholesaler buys large numbers of this variety of fruit and classifies the lightest 15% as small.

(d) Find the maximum weight of a fruit that is classified as small.

You must show your working clearly.

(3)

The weights, Y grams, of a second variety of this fruit are normally distributed with

$$Y \sim \mathrm{N}(\mu, \sigma^2)$$

Given that 5% of these fruit weigh less than 152 grams and 40% weigh more than 180 grams,

(e) calculate the mean and standard deviation of the weights of this variety of fruit.

(5)

(Total for question = 15 marks)

(**Q05 WST01/01, Oct 2023**)

Extra space for working:

Q6.

Kris works in the mailroom of a large company and is responsible for all the letters sent by the company. The weights of letters sent by the company, W grams, have a normal distribution with mean 165 g and standard deviation 35 g.

(a) Estimate the proportion of letters sent by the company that weigh less than 120 g.

(3)

Kris splits the letters to be sent into 3 categories: heavy, medium and light, with $\frac{1}{3}$ of the letters in each category.

(b) Find the weight limits that determine medium letters.

(4)

A heavy letter is chosen at random.

(c) Find the probability that this letter weighs less than 200 g.

(3)

Kris chooses a random sample of 3 letters from those in the mailroom one day.

(d) Find the probability that there is one letter in each of the 3 categories.

(3)

(Total for question = 13 marks)

(**Q04 WST01/01, June 2021**)

Extra space for working:

Q7.

Xiang is designing shelves for a bookshop. The height, H cm, of books is modelled by the normal distribution with mean 25.1 cm and standard deviation 5.5 cm

(a) Show that P(H > 30.8) = 0.15

(3)

Xiang decided that the smallest 5% of books and books taller than 30.8 cm would not be placed on the shelves. All the other books will be placed on the shelves.

(b) Find the range of heights of books that will be placed on the shelves.

(3)

The books that will be placed on the shelves have heights classified as small, medium or large.
The numbers of small, medium and large books are in the ratios 2 : 3 : 3

(c) The medium books have heights x cm where $m < x < d$

(i) Show that d = 25.8 to 1 decimal place.

(3)

(ii) Find the value of m

(4)

Xiang wants 2 shelves for small books, 3 shelves for medium books and 3 shelves for large books.
These shelves will be placed one above another and made of wood that is 1 cm thick.

(d) Work out the minimum total height needed.

(2)

(Total for question = 15 marks)

(**Q06 WST01/01, Oct 2021**)

Extra space for working:

Q8.

A machine squeezes apples to extract their juice. The volume of juice, J ml, extracted from 1 kg of apples is modelled by a normal distribution with mean μ and standard deviation σ

Given that $\mu = 500$ and $\sigma = 25$ use standardisation to

(a) (i) show that $P(J > 510) = 0.3446$

(2)

(ii) calculate the value of d such that $P(J > d) = 0.9192$

(3)

Zen randomly selects 5 bags each containing 1 kg of apples and records the volume of juice extracted from each bag of apples.

(b) Calculate the probability that each of the 5 bags of apples produce less than 510 ml of juice.

(2)

Following adjustments to the machine, the volume of juice, R ml, extracted from 1 kg of apples is such that $\mu = 520$ and $\sigma = k$

Given that $P(R < r) = 0.15$ and $P(R > 3r - 800) = 0.005$

(c) find the value of r and the value of k

(7)

(Total for question = 14 marks)

(**Q07 WST01/01, June 2023**)

Extra space for working:

Q9.

Jia writes a computer program that randomly generates values from a normal distribution.

He sets the mean as 40 and the standard deviation as 2.4

(a) Find the probability that a particular value generated by the computer program is less than 37

(3)

Jia changes the mean to m but leaves the standard deviation as 2.4

The computer program then randomly generates 2 independent values from this normal distribution.

The probability that both of these values are greater than 32 is 0.16

(b) Find the value of m, giving your answer to 2 decimal places.

(4)

Jia now changes the mean to 4 and the standard deviation to 8

The computer program then randomly generates 5 independent values from this normal distribution.

(c) Find the probability that at least one of these values is negative.

(4)

(Total for question = 11 marks)

(**Q05 WST01/01, Jan 2022**)

Extra space for working:

Q10.

The lengths, L mm, of housefly wings are normally distributed with $L \sim N(4.5, 0.4^2)$

(a) Find the probability that a randomly selected housefly has a wing length of less than 3.86 mm.

(3)

(b) Find

(i) the upper quartile (Q_3) of L
(ii) the lower quartile (Q_1) of L

(4)

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(c) Find these two outlier limits.

(3)

A housefly is selected at random.

(d) Using standardisation, show that the probability that this housefly is **not** an outlier is 0.993 to 3 decimal places.

(3)

Given that this housefly is **not** an outlier,

(e) showing your working, find the probability that the wing length of this housefly is greater than 5 mm.

(4)

(Total for question = 17 marks)

(**Q05 WST01/01, Jan 2023**)

Extra space for working:

Q11.

The random variable W has a discrete uniform distribution where

$$P(W = w) = \frac{1}{5} \quad \text{for } w = 1, 2, 3, 4, 5$$

(a) Find $P(2 \leq W < 3.5)$

(1)

The discrete random variable $X = 5 - 2W$

(b) Find $E(X)$

(3)

(c) Find $P(X < W)$

(2)

The discrete random variable $Y = \frac{1}{W}$

(d) Find

(i) the probability distribution of Y
(ii) $\text{Var}(Y)$, showing your working.

(5)

(e) Find $\text{Var}(2 - 3Y)$

(2)

(Total for question = 13 marks)

(**Q04 WST01/01, Jan 2022**)

1.6

The Normal Distribution Function

The function tabulated below is $\Phi(z)$, defined as $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} dt$

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0.00	0.5000	0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772
0.01	0.5040	0.51	0.6950	1.01	0.8438	1.51	0.9345	2.02	0.9783
0.02	0.5080	0.52	0.6985	1.02	0.8461	1.52	0.9357	2.04	0.9793
0.03	0.5120	0.53	0.7019	1.03	0.8485	1.53	0.9370	2.06	0.9803
0.04	0.5160	0.54	0.7054	1.04	0.8508	1.54	0.9382	2.08	0.9812
0.05	0.5199	0.55	0.7088	1.05	0.8531	1.55	0.9394	2.10	0.9821
0.06	0.5239	0.56	0.7123	1.06	0.8554	1.56	0.9406	2.12	0.9830
0.07	0.5279	0.57	0.7157	1.07	0.8577	1.57	0.9418	2.14	0.9838
0.08	0.5319	0.58	0.7190	1.08	0.8599	1.58	0.9429	2.16	0.9846
0.09	0.5359	0.59	0.7224	1.09	0.8621	1.59	0.9441	2.18	0.9854
0.10	0.5398	0.60	0.7257	1.10	0.8643	1.60	0.9452	2.20	0.9861
0.11	0.5438	0.61	0.7291	1.11	0.8665	1.61	0.9463	2.22	0.9868
0.12	0.5478	0.62	0.7324	1.12	0.8686	1.62	0.9474	2.24	0.9875
0.13	0.5517	0.63	0.7357	1.13	0.8708	1.63	0.9484	2.26	0.9881
0.14	0.5557	0.64	0.7389	1.14	0.8729	1.64	0.9495	2.28	0.9887
0.15	0.5596	0.65	0.7422	1.15	0.8749	1.65	0.9505	2.30	0.9893
0.16	0.5636	0.66	0.7454	1.16	0.8770	1.66	0.9515	2.32	0.9898
0.17	0.5675	0.67	0.7486	1.17	0.8790	1.67	0.9525	2.34	0.9904
0.18	0.5714	0.68	0.7517	1.18	0.8810	1.68	0.9535	2.36	0.9909
0.19	0.5753	0.69	0.7549	1.19	0.8830	1.69	0.9545	2.38	0.9913
0.20	0.5793	0.70	0.7580	1.20	0.8849	1.70	0.9554	2.40	0.9918
0.21	0.5832	0.71	0.7611	1.21	0.8869	1.71	0.9564	2.42	0.9922
0.22	0.5871	0.72	0.7642	1.22	0.8888	1.72	0.9573	2.44	0.9927
0.23	0.5910	0.73	0.7673	1.23	0.8907	1.73	0.9582	2.46	0.9931
0.24	0.5948	0.74	0.7704	1.24	0.8925	1.74	0.9591	2.48	0.9934
0.25	0.5987	0.75	0.7734	1.25	0.8944	1.75	0.9599	2.50	0.9938
0.26	0.6026	0.76	0.7764	1.26	0.8962	1.76	0.9608	2.55	0.9946
0.27	0.6064	0.77	0.7794	1.27	0.8980	1.77	0.9616	2.60	0.9953
0.28	0.6103	0.78	0.7823	1.28	0.8997	1.78	0.9625	2.65	0.9960
0.29	0.6141	0.79	0.7852	1.29	0.9015	1.79	0.9633	2.70	0.9965
0.30	0.6179	0.80	0.7881	1.30	0.9032	1.80	0.9641	2.75	0.9970
0.31	0.6217	0.81	0.7910	1.31	0.9049	1.81	0.9649	2.80	0.9974
0.32	0.6255	0.82	0.7939	1.32	0.9066	1.82	0.9656	2.85	0.9978
0.33	0.6293	0.83	0.7967	1.33	0.9082	1.83	0.9664	2.90	0.9981
0.34	0.6331	0.84	0.7995	1.34	0.9099	1.84	0.9671	2.95	0.9984
0.35	0.6368	0.85	0.8023	1.35	0.9115	1.85	0.9678	3.00	0.9987
0.36	0.6406	0.86	0.8051	1.36	0.9131	1.86	0.9686	3.05	0.9989
0.37	0.6443	0.87	0.8078	1.37	0.9147	1.87	0.9693	3.10	0.9990
0.38	0.6480	0.88	0.8106	1.38	0.9162	1.88	0.9699	3.15	0.9992
0.39	0.6517	0.89	0.8133	1.39	0.9177	1.89	0.9706	3.20	0.9993
0.40	0.6554	0.90	0.8159	1.40	0.9192	1.90	0.9713	3.25	0.9994
0.41	0.6591	0.91	0.8186	1.41	0.9207	1.91	0.9719	3.30	0.9995
0.42	0.6628	0.92	0.8212	1.42	0.9222	1.92	0.9726	3.35	0.9996
0.43	0.6664	0.93	0.8238	1.43	0.9236	1.93	0.9732	3.40	0.9997
0.44	0.6700	0.94	0.8264	1.44	0.9251	1.94	0.9738	3.50	0.9998
0.45	0.6736	0.95	0.8289	1.45	0.9265	1.95	0.9744	3.60	0.9998
0.46	0.6772	0.96	0.8315	1.46	0.9279	1.96	0.9750	3.70	0.9999
0.47	0.6808	0.97	0.8340	1.47	0.9292	1.97	0.9756	3.80	0.9999
0.48	0.6844	0.98	0.8365	1.48	0.9306	1.98	0.9761	3.90	1.0000
0.49	0.6879	0.99	0.8389	1.49	0.9319	1.99	0.9767	4.00	1.0000
0.50	0.6915	1.00	0.8413	1.50	0.9332	2.00	0.9772		

Percentage Points Of The Normal Distribution

The values z in the table are those which a random variable $Z \sim N(0, 1)$ exceeds with probability p; that is, $P(Z > z) = 1 - \Phi(z) = p$.

p	z	p	z
0.5000	0.0000	0.0500	1.6449
0.4000	0.2533	0.0250	1.9600
0.3000	0.5244	0.0100	2.3263
0.2000	0.8416	0.0050	2.5758
0.1500	1.0364	0.0010	3.0902
0.1000	1.2816	0.0005	3.2905