1. A medical researcher is studying the relationship between age (x years) and volume of blood (y ml) pumped by each contraction of the heart. The researcher obtained the following data from a random sample of 8 patients.

Age (x)	20	25	30	45	55	60	65	70
Volume (y)	74	76	77	72	68	67	64	62

[You may use $\sum x = 370$, $S_{xx} = 2587.5$, $\sum y = 560$, $\sum y^2 = 39\,418$, $S_{xy} = -710$]

(a) Calculate S_{yy} **(2)**

(b) Calculate the product moment correlation coefficient for these data. **(2)**

(c) Interpret your value of the correlation coefficient. **(1)**

The researcher believes that a linear regression model may be appropriate to describe these data.

(d) State, giving a reason, whether or not your value of the correlation coefficient supports the researcher's belief. **(1)**

(e) Find the equation of the regression line of y on x, giving your answer in the form $y = a + bx$ **(4)**

Jack is a 40-year-old patient.

(f) (i) Use your regression line to estimate the volume of blood pumped by each contraction of Jack's heart.

(ii) Comment, giving a reason, on the reliability of your estimate. **(2)**

a) $S_{yy} = 218$ b) $PMCC = \dfrac{-710}{\sqrt{218 \times 2587.5}} = -0.945$

c) Strong evidence to suggest negative correlation exists. The older a person the lower the volume

d) yes as r is very close to -1

e) $b = \dfrac{Sxy}{Sxx} = -0.274$ $a = \bar{y} - b\bar{x} = 82.69$ $y = 82.7 - 0.274x$

f) i) $x = 40 \Rightarrow y = 71.7$ ii) reliable, interpolation r close to -1.

3. A biased four-sided die has faces marked 1, 3, 5 and 7. The random variable X represents the score on the die when it is rolled. The cumulative distribution function of X, $F(x)$, is given in the table below.

x	1	3	5	7
$F(x)$	0.2	0.5	0.9	1

(a) Find the probability distribution of X **(4)**

(b) Find $P(2 < X \leqslant 6)$ **(2)**

(c) Write down the value of $F(4)$ **(1)**

a)

x	1	3	5	7
p	0.2	0.3	0.4	0.1

b) $0.3 + 0.4 = 0.7$

c) $F(4) = P(x \leq 4) = 0.5$

4. The random variable $Y \sim N(\mu, \sigma^2)$

Given that $P(Y < 17) = 0.6$ find

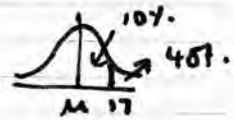
(a) $P(Y > 17)$ **(1)**

(b) $P(\mu < Y < 17)$ **(2)**

(c) $P(Y < \mu \mid Y < 17)$ **(2)**

a) $P(y > 17) = 40\% = 0.40$

b) $P\left(Z > \dfrac{17-\mu}{\sigma}\right) = 40\%$.



$P(\mu < Y < 17) = 0.10$

$P(y < \mu \mid y < 17) = \dfrac{P(y < \mu)}{P(y < 17)} = \dfrac{0.5}{0.6} = \dfrac{5}{6}$

2. The table below shows the distances (to the nearest km) travelled to work by the 50 employees in an office.

Distance (km)	Frequency (f)	Distance midpoint (x)
0 – 2	16	1.25
3 – 5	12	4
6 – 10	10	8
11 – 20	8	15.5
21 – 40	4	30.5

16, 28 → 25th, 38, 46, 50

[You may use $\sum fx = 394$, $\sum fx^2 = 6500$]

A histogram has been drawn to represent these data.

The bar representing the distance of 3 – 5 has a width of 1.5 cm and a height of 6 cm.

(a) Calculate the width and height of the bar representing the distance of 6 – 10 **(3)**

(b) Use linear interpolation to estimate the median distance travelled to work. **(2)**

(c) (i) Show that an estimate of the mean distance travelled to work is 7.88 km.

(ii) Estimate the standard deviation of the distances travelled to work. **(4)**

(d) Describe, giving a reason, the skewness of these data. **(2)**

Peng starts to work in this office as the 51st employee.

She travels a distance of 7.88 km to work.

(e) Without carrying out any further calculations, state, giving a reason, what effect Peng's addition to the workforce would have on your estimates of the

(i) mean,

(ii) median,

(iii) standard deviation

of the distances travelled to work. **(3)**

2.5 5.5 6cm

cw = 3 ÷2 = 1.5cm

5.5 10.5

cw = 5 ÷2 width = 2.5cm

A = 9cm²

F ≈ 12 ÷4 ×3

F = 10 ÷4×3

A = 7.5cm²

∴ h = 3cm

2.5 Q_2 5.5

16 25 28

$\frac{Q_2 - 2.5}{3} = \frac{9}{12}$ × + $Q_2 = 4.75$

c(i) $\bar{x} = \frac{\Sigma fx}{n} = \frac{394}{50} = 7.88$

ii) $S_{xx} = 6500 - 394^2 \div 50 = 3395.28$ $\sigma = \sqrt{\frac{S_{xx}}{n}} = 8.24$

d) Mean (7.88) > Median (4.75) ∴ positive skew

e) i) no change, 7.88 = mean.

ii) slight increase, $\frac{51}{2} = 25.5$th

iii) slight reduction $\Sigma(x - \bar{x})$ unchanged but n has increased by 1.

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

	6	4	5	-1	-7
x	-2	0	2	4	
$P(X = x)$	a	b	a	c	

where a, b and c are probabilities.

Given that $E(X) = 0.8$

(a) find the value of c. **(2)**

Given also that $E(X^2) = 5$ find

(b) the value of a and the value of b, **(4)**

(c) $Var(X)$ **(2)**

The random variable $Y = 5 - 3X$

Find

(d) $E(Y)$ **(1)**

(e) $Var(Y)$ **(2)**

(f) $P(Y \geqslant 0)$ **(4)**

a) $E(x) = -2a + 2a + 4c = 0.8 \quad \therefore c = 0.2$

b) $E(x^2) = 4a + 4a + 16c = 5 \quad 8a = 1.8 \quad a = 0.225$

$\Sigma p = 1 \Rightarrow 2a + b + c = 1 \quad \therefore b = 0.35$

c) $V(x) = E(x^2) - E(x)^2 = 5 - 0.8^2 = 4.36$

e) $V(5-3x) = 9V(x) = 39.24$ d) $E(5-3x) = 5 - 3(0.8) = 2.6$

f) $P(y \geq 0) = a + c = 0.425$

7. One event at *Pentor* sports day is throwing a tennis ball. The distance a child throws a tennis ball is modelled by a normal distribution with mean 32 m and standard deviation 12 m. Any child who throws the tennis ball more than 50 m is awarded a gold certificate.

(a) Show that, to 3 significant figures, 6.68% of children are awarded a gold certificate. **(3)**

A silver certificate is awarded to any child who throws the tennis ball more than d metres but less than 50 m.

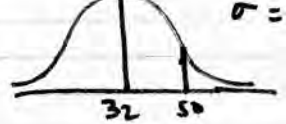
Given that 19.1% of the children are awarded a silver certificate,

(b) find the value of d. **(4)**

Three children are selected at random from those who take part in the throwing a tennis ball event.

(c) Find the probability that 1 is awarded a gold certificate and 2 are awarded silver certificates. Give your answer to 2 significant figures. **(3)**

a)



$\sigma = 12 \quad P(d > 50) = P\left(Z > \frac{50-32}{12}\right)$

$= P(Z > 1.5) = 1 - \Phi(1.5)$

$= 0.0668$

b) $P(D > d) = 0.191 + 0.0668 = 0.2578$

$\therefore \Phi(d) = 0.7422 \quad \therefore d = 0.65$

b) $P\left(Z > \frac{d-32}{12}\right) = 0.191 + 0.0668 = 0.2578$

$\therefore \Phi\left(\frac{d-32}{12}\right) = 0.7422 \quad \therefore \frac{d-32}{12} = 0.65$

$\therefore d = 39.8m$

c) $GSS \times 3 = 3 \times 0.0668 \times 0.191^2$

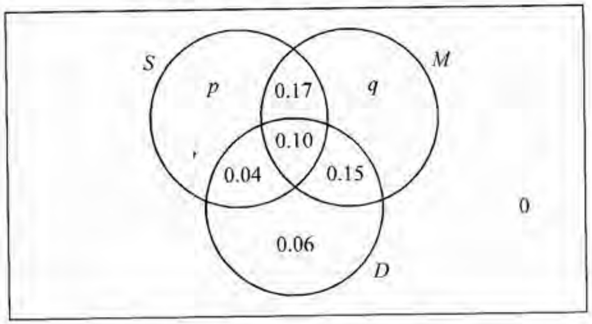
$= 0.00731$

6. The Venn diagram below shows the probabilities of customers having various combinations of a starter, main course or dessert at Polly's restaurant.

S = the event a customer has a starter.

M = the event a customer has a main course.

D = the event a customer has a dessert.



Given that the events S and D are statistically independent

(a) find the value of p. **(4)**

(b) Hence find the value of q. **(2)**

(c) Find

(i) $P(D \mid M \cap S)$

(ii) $P(D \mid M \cap S')$ **(4)**

One evening 63 customers are booked into Polly's restaurant for an office party. Polly has asked for their starter and main course orders before they arrive.

Of these 63 customers

27 ordered a main course and a starter,

36 ordered a main course without a starter.

(d) Estimate the number of desserts that these 63 customers will have. **(2)**

a) $P(S) \times P(D) = P(S \cap D)$ $\quad (0.31+p) \times 0.35 = 0.14$

$\therefore p = 0.09$

b) $q = 1 - \ldots = 0.39$

c) i) $P(D \mid M \cap S) = \dfrac{0.10}{0.27} = 0.37$

ii) $P(D \mid M \cap S') = \dfrac{0.15}{0.54} = 0.278$

d) 27×0.37
$+ 36 \times 0.278$

$= 20$