1 20% of packets of a certain kind of cereal contain a free gift. Jane buys one packet a week for 8 weeks. The number of free gifts that Jane receives is denoted by X. Assuming that Jane's 8 packets can be regarded as a random sample, find

(i) $\mathrm{P}(X = 3)$, **[3]**

(ii) $\mathrm{P}(X \geqslant 3)$, **[2]**

(iii) $\mathrm{E}(X)$. **[2]**

1(i) $X \sim B(8, 0.2)$

$P(X=3) = \binom{8}{3} \times 0.2^3 \times 0.8^5$

$= 0.14680064$

$= \underline{\underline{0.147}}$ ✓ (3sf)

(ii) $P(X \geqslant 3) = 1 - P(X \leq 2)$

$P(X \leq 2)$ (from tables) $= 0.7969$

$P(X \geqslant 3) = 1 - 0.7969$

$= 0.2031$

$= \underline{\underline{0.203}}$ ✓ (3sf)

1(iii) $E(X) = np$ ← for a BINOMIAL distribution

$= 8 \times 0.2$

$= \underline{\underline{1.6}}$ ✓

2 Two judges placed 7 dancers in rank order. Both judges placed dancers A and B in the first two places, but in opposite orders. The judges agreed about the ranks for all the other 5 dancers. Calculate the value of Spearman's rank correlation coefficient. **[4]**

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

DANCER	Judge1	Judge2	d	d^2
A	1	2	-1	1
B	2	1	1	1
C	3	3	0	0
D	4	4	0	0
E	5	5	0	0
F	6	6	0	0
G	7	7	0	0
				2

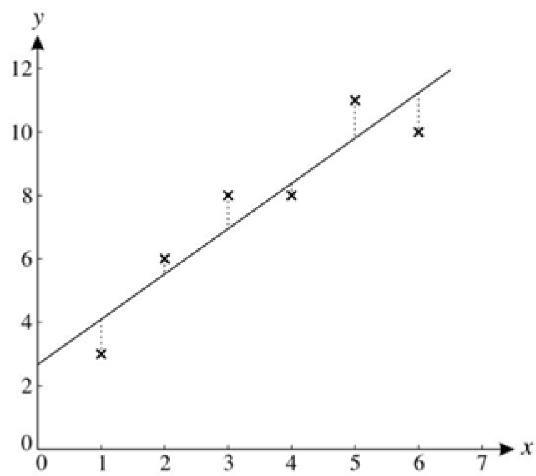
$$r_s = 1 - \frac{(6\times 2)}{(7\times 48)} = \frac{27}{28} = 0.964$$

(3sf)

3 In an agricultural experiment, the relationship between the amount of water supplied, x units, and the yield, y units, was investigated. Six values of x were chosen and for each value of x the corresponding value of y was measured. The results are shown in the table.

x	1	2	3	4	5	6
y	3	6	8	8	11	10

These results, together with the regression line of y on x, are plotted on the graph.



(i) Give a reason why the regression line of x on y is not suitable in this context. **[1]**

(ii) Explain the significance, for the regression line of y on x, of the distances shown by the vertical dotted lines in the diagram. **[2]**

(iii) Calculate the value of the product moment correlation coefficient, r. **[3]**

(iv) Comment on your value of r in relation to the diagram. **[2]**

3(i) Because the yield is dependent upon the amount of water used.

(ii) The sum of the squares of the length of these vertical lines has been minimised and that is what determines the equation of the regression line.

3(iii)

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

$$S_{xy} = \sum xy - \frac{(\sum x \sum y)}{n}$$

$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

x	y	xy	x^2	y^2
1	3	3	1	9
2	6	12	4	36
3	8	24	9	64
4	8	32	16	64
5	11	55	25	121
6	10	60	36	100
21	46	176	91	394

$$S_{xy} = 186 - \frac{(21 \times 46)}{6}$$

$$= 25$$

$$S_{xx} = 91 - \frac{(21)^2}{6} = 17.5$$

$$S_{yy} = 394 - \frac{(46)^2}{6} = 41.\dot{3}$$

$$r = \frac{25}{\sqrt{(17.5 \times 41.\dot{3})}}$$

$$= 0.929545749$$

$$= \underline{\underline{0.930}} \text{ (3sf)}$$ ✓

3(iv) The calculated value of r is 0.93, this indicates a very strong correlation between yield and the amount of water used. This strong relationship is shown by the regression sloping upwards to the right and the points on the scatter plot lying close to the line.

4 30% of people own a Talk-2 phone. People are selected at random, one at a time, and asked whether they own a Talk-2 phone. The number of people questioned, up to and including the first person who owns a Talk-2 phone, is denoted by X. Find

(i) $P(X = 4)$, **[3]**

(ii) $P(X > 4)$, **[2]**

(iii) $P(X < 6)$. **[3]**

4 (i) $X \sim Geo(0.3)$

$P(X=4) = 0.7^3 \times 0.3 = 0.1029$

$= \underline{0.103}$ (3sf)

(ii) $P(X>4) = 0.7^4 = 0.2401$

$= \underline{0.240}$ (3sf)

(iii) $P(X<6) = 1 - P(X>5)$

$= 1 - (0.7^5)$

$= 0.83193$

$= \underline{0.832}$ (3sf)

5 The diameters of 100 pebbles were measured. The measurements rounded to the nearest millimetre, x, are summarised in the table.

x	$10 \leqslant x \leqslant 19$	$20 \leqslant x \leqslant 24$	$25 \leqslant x \leqslant 29$	$30 \leqslant x \leqslant 49$
Number of stones	25	22	29	24

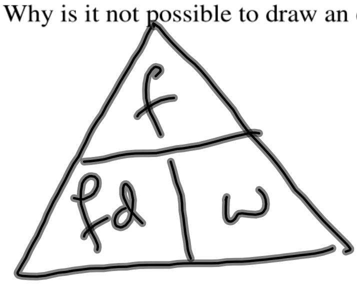
These data are to be presented on a statistical diagram.

(i) For a histogram, find the frequency density of the $10 \leqslant x \leqslant 19$ class. **[2]**

(ii) For a cumulative frequency graph, state the coordinates of the first two points that should be plotted. **[2]**

(iii) Why is it not possible to draw an exact box-and-whisker plot to illustrate the data? **[1]**

5(i)



Actual width of $10 \leq x \leq 19$ is $19.5 - 9.5 = 10$

$$fd = \frac{25}{10} = 2.5 \checkmark$$

(ii) $(19.5, 25) \checkmark$ however MS expects

$(24.5, 47) \checkmark$ $(9.5, 0)$ as 1st coordinate!

(iii) Reasonable estimates of the median and upper and lower quartiles can be obtained but we aren't given exact minimum and maximum values.

✓ acceptable.

6 Last year Eleanor played 11 rounds of golf. Her scores were as follows:

79, 71, 80, 67, 67, 74, 66, 65, 71, 66, 64.

(i) Calculate the mean of these scores and show that the standard deviation is 5.31, correct to 3 significant figures. **[4]**

(ii) Find the median and interquartile range of the scores. **[4]**

This year, Eleanor also played 11 rounds of golf. The standard deviation of her scores was 4.23, correct to 3 significant figures, and the interquartile range was the same as last year.

(iii) Give a possible reason why the standard deviation of her scores was lower than last year although her interquartile range was unchanged. **[1]**

In golf, smaller scores mean a better standard of play than larger scores. Ken suggests that since the standard deviation was smaller this year, Eleanor's overall standard has improved.

(iv) Explain why Ken is wrong. **[1]**

(v) State what the smaller standard deviation does show about Eleanor's play. **[1]**

$$\bar{x} = \frac{\sum x}{n} \qquad \sigma^2 = \frac{\sum x^2}{n} - \bar{x}^2$$

$$\sigma = \sqrt{\sigma^2}$$

x	x^2
79	6241
71	5041
80	6400
67	4489
67	4489
74	5476
66	4356
65	4225
71	5041
66	4356
64	4096
770	54210

$\sum x = 770$

$\sum x^2 = 54210$

$\bar{x} = \frac{770}{11} = \underline{\underline{70}}$ ✓

$\sigma^2 = \frac{54210}{11} - 70^2$

$= 28.1\dot{8}$

$\sigma = \sqrt{28.1\dot{8}}$

$= \underline{\underline{5.31}}$ ✓ (3sf)

6(ii) 64, 65, 66, 66, 67, 67, 71, 71, 74, 79, 80

MEDIAN = 67

LQ = 66

UQ = 74

IQR = 74 − 66 = 8

(iii) The standard deviation is affected by the spread of all of the values whereas the IQR only by the middle 50%.
If her lowest scores below the LQ were closer to the LQ or her scores higher than the UQ were closer to the UQ either of these situations would reduce the standard deviation.

6(iv) Ken is wrong because standard deviation is a measure of spread not a measure of location.

The mean score would be a better indicator of improving or worsening performance.

(v) The smaller s.d. indicates the Eleanors scores are becoming more consistent.

7 Three letters are selected at random from the 8 letters of the word COMPUTER, without regard to order.

(i) Find the number of possible selections of 3 letters. **[2]**

(ii) Find the probability that the letter P is included in the selection. **[3]**

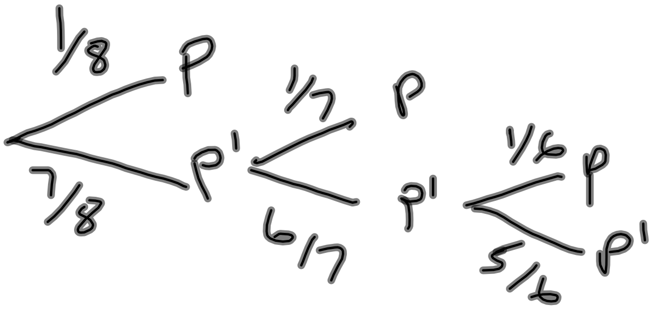
Three letters are now selected at random, one at a time, from the 8 letters of the word COMPUTER, and are placed in order in a line.

(iii) Find the probability that the 3 letters form the word TOP. **[3]**

7(i) C O M P U T E R

$\binom{8}{3}$ $\quad \underline{\underline{56}}$ ✓

(ii)



$$\frac{1}{8} + \left(\frac{7}{8} \times \frac{1}{7}\right) + \left(\frac{7}{8} \times \frac{6}{7} \times \frac{1}{6}\right)$$

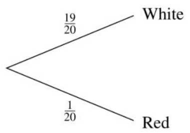
$$= \underline{\underline{\frac{3}{8}}}$$ ✓

(iii) $\frac{1}{8} \times \frac{1}{7} \times \frac{1}{6} = \frac{1}{336}$ ✓

8 A game at a charity event uses a bag containing 19 white counters and 1 red counter. To play the game once a player takes counters at random from the bag, one at a time, without replacement. If the red counter is taken, the player wins a prize and the game ends. If not, the game ends when 3 white counters have been taken. Niko plays the game once.

(i) (a) Copy and complete the tree diagram showing the probabilities for Niko. **[4]**

First counter



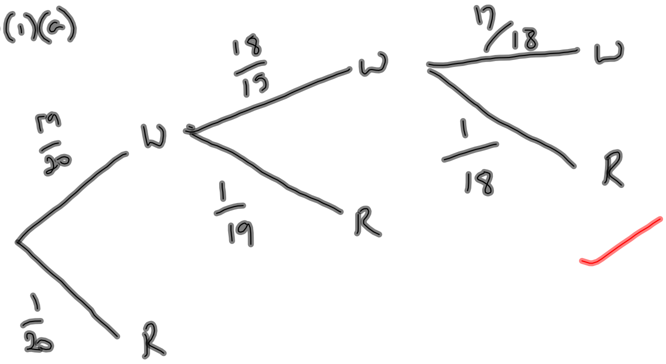
(b) Find the probability that Niko will win a prize. **[3]**

(ii) The number of counters that Niko takes is denoted by X.

(a) Find $P(X = 3)$. **[2]**

(b) Find $E(X)$. **[4]**

8(i)(a)



(b) $P(\text{no Red}) = \frac{19}{20} \times \frac{18}{19} \times \frac{17}{18}$

$= \frac{17}{20}$

$P(\text{win}) = 1 - \frac{17}{20} = \frac{3}{20}$

(ii)

x	1	2	3	
p	$1/20$	$\frac{1}{20}$	$\frac{9}{10}$	
px	$1/20$	$2/20$	$54/20$	$\sum px = \frac{57}{20}$

$P(2\text{ goes}) = \frac{19}{20} \times \frac{1}{19} = \frac{1}{20}$

$P(3\text{ goes}) = \frac{19}{20} \times \frac{18}{19} = \frac{9}{10}$

$P(x=3) = \frac{9}{10}$

(b) $E(X) = \frac{57}{20} = 2.85$

9 Repeated independent trials of a certain experiment are carried out. On each trial the probability of success is 0.12.

(i) Find the smallest value of n such that the probability of at least one success in n trials is more than 0.95. **[3]**

(ii) Find the probability that the 3rd success occurs on the 7th trial. **[5]**

9i) $p = 0.12$ $\quad X \sim B(n, 0.12)$

$P(X \geq 1) = 1 - P(X=0)$

$P(X \geq 1) > 0.95$

$1 - P(X=0) > 0.95$

so $\quad P(X=0) < 0.05$

$\binom{n}{0} \times 0.12^0 \times 0.88^n < 0.05$

$0.88^n < 0.05$

$0.88^{10} = 0.279$

$0.88^{20} = 0.078$

$0.88^{21} = 0.068$

$0.88^{22} = 0.060$

$0.88^{23} = 0.053$

$0.88^{24} = 0.047$

$\underline{\underline{n = 24}}$ ✓

9(ii) Must end 6th trial with 2 successes

$\binom{6}{2} \times 0.12^2 \times 0.88^4 = 0.1295341978$

Must succeed on 7th trial

$0.130 \times 0.12 \doteqdot 0.0155410373$

$= \underline{\underline{0.0155}}$ (3sf)