7 Quanitifying uncertainty: proability, binomial and normal distributions

Skills check

1 There are 12 numbers in total.

a 2,5,11,17 are prime, so the probability that the number is prime $= \frac{4}{12} = \frac{1}{3}$.

b 1,5,9,11,17,25,27 are odd, so the probability that the number is odd $= \frac{7}{12}$.

c 1,4,9,16,25 are square, so the probability that the number is square $= \frac{5}{12}$.

2 Total number of people = 116.

a Total number of females is 57, so the probability is $\frac{57}{116}$.

b $\frac{12}{116} = \frac{3}{29}$.

c Total number of non-smokers is 98, so the probability is $\frac{98}{116} = \frac{49}{58}$.

3 Mean $\frac{9\times1+7\times2+3\times3+2\times6+1\times11}{9+7+3+2+1} = \frac{55}{22} = 2.5$.

Exercise 7A

1 $p = \frac{1}{3}$. (Each letter from RANDOM has probability of $\frac{1}{6}$ of being picked. 2 of these letters are also in MATHS. Hence, $p = \frac{1}{6}\times 2$).

2 a $p = \frac{17}{20}$. All numbers except for 1,2,3 can be hit.

b $p = \frac{14}{20} = \frac{7}{10}$. There are 14 numbers above 6.

c p = 1. The range of numbers is 1–20.

d $p = \frac{14}{20} = \frac{7}{10}$. All numbers 1 – 14 can be hit.

e $p = \frac{8}{20} = \frac{2}{5}$. 2,3,5,7,11,13,17,19 are prime.

f $p = \frac{4}{20} = \frac{1}{5}$. 1,4,9,16 are square.

g p = 0. There are no solutions to this equation in this set of positive integers.

3 a $p = \frac{6}{11}$.

b $p = \frac{3}{11}$ (1,4,9).

c $p = \frac{5}{11}$ (2,3,5,7,11).

d $p = \frac{2}{11}$ (1,9 are square and odd).

e p = 0 (no square numbers are prime).

f $p = \frac{4}{11}$ (2 is not odd).

g $p = \frac{1}{11}$ (2 is prime and even).

.

4 There are 10000 possible PINs because each digit can take 10 values.

a $p = \frac{1}{10000}$.

b $p = \frac{99}{10000}$.

c $p = \frac{1}{10}$ (last digit has to be 0 and other digits can take any value, assuming 0000 is considered to be divisible by 10).

d $p = \frac{9987}{10000}$ (13 combinations do not fall in this range.).

Exercise 7B

1 The total of 239 shoppers were surveyed.

a $p = \frac{73}{239}$. **b** $p = \frac{37}{136}$. **c** $1300 \times \frac{100}{239} \approx 544$ shoppers.

2 a $p = \frac{3}{8}$ (there are $4 \times 4 = 16$ possible pairs 3,3; 4,4; 5,5; 6,3; and 4,2 give a natural number).

b $320 \times \frac{3}{16} = 60$ (only three possibilities of positive difference: 4 − 3, 5 − 3 and 5 − 4.

3 a $154 \times \frac{6}{12} = 77$ (1,2,3,4,6,12 are factors of 12).

b $154 \times \frac{5}{12} \approx 64$ (2,3,5,7,11 are prime).

c $154 \times \frac{2}{12} \approx 26$ (2,3 are prime factors of 12).

4 $207 \times (0.15 + 0.25 + 0.12) \approx 108$.

5 $531 \times \frac{222}{347} \approx 340$.

6 $79 \times \frac{5}{11} \approx 36$.

7 a $573 \times (0.005 + 0.012) \approx 10$.

b Assume the distribution of the cars is the same every month of the year.

8 $67 \times 0.0137 + 313 \times 0.0041 \approx 2$.

9 $31 \times \frac{6}{24} \approx 8$.

10 Expected values of each dice: A $\frac{16}{6}$, B $\frac{18}{6}$, C $\frac{20}{6}$, D $\frac{18}{6}$, so dice C is most likely to win.

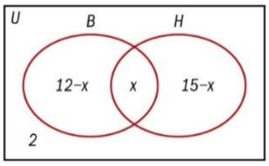
Exercise 7C

1 a Find x such that $127 = 81 + 70 - 29 + x$, so $x = 5$

b $p = \frac{70-29}{127} = \frac{41}{127}$.

c $10000 \times \frac{81-29}{127} \approx 4094$.

2 Create a Venn diagram.

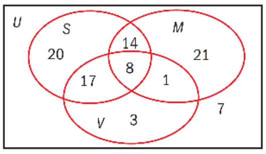


a $p = \frac{9}{20}$. Find x: 20 = (12 − x) + (15 − x) + x + 2, x = 9.

b $p = \frac{9}{12} = \frac{3}{4}$. Divide the number of students studying both subjects by the total number of students studying biology.

c $60 \times \frac{9}{20}$ = 27.

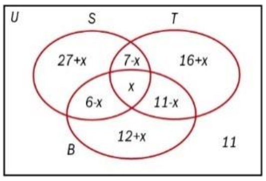
3 a



b $p = \frac{21+20+3}{91} = \frac{44}{91}$

c $p = \frac{14+1+17}{91} = \frac{32}{91}$

4 a



b 94 = (27 + x) + (7 − x) + (16 + x) + (6 − x) + x + (11 − x) + (12 + x) + 11,
94 = 90 + x, x = 4.

c $p \frac{7-x+6-x+11-x+x}{94} = \frac{16}{94} = \frac{8}{47}$.

Exercise 7D

1 Draw a sample space diagram.

	1	**4**	**9**	**16**
2	1	2	7	14
3	2	1	6	13
5	4	1	4	11
7	6	3	2	9
11	10	7	2	5
13	12	9	4	3

a $p = \frac{10}{24} = \frac{5}{12}$ (orange outcomes).

b $p = \frac{8}{24} = \frac{1}{3}$ (green outcomes).

2 a Draw a sample space diagram, then $p = \frac{16}{36} = \frac{4}{9}$ (green + orange areas).

	1	**2**	**3**	**4**	**5**	**6**
1	1	2	3	4	5	6

2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

b $945 \times \frac{27}{36} \approx 709$ (orange + brown areas).

3 Draw a sample space diagram.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	2	3	4	5	6	7	8
2	2	4	6	8	10	12	14	16
3	3	6	9	12	15	18	21	24
4	4	8	12	16	20	24	28	32

a $p = \frac{8}{32} = \frac{1}{4}$.

b $p = \frac{6}{32} = \frac{3}{16}$.

c $p = \frac{4}{32} = \frac{1}{8}$.

Draw a sample space diagram for Bethany's case.

	1	**2**	**3**	**4**	**5**	**6**
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36

d $p = \frac{9}{36} = \frac{1}{4}$.

e $p = \frac{13}{36}$.

f $p = \frac{20}{36} = \frac{5}{9}$ p = 20/36 = 5/9 (orange area).

g E.g. M and N are even.

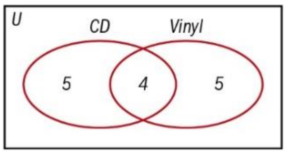
4 a

		Chromosome inherited from mother	
		X	X
Chromosome inherited from father	X	XX	XX
	Y	XY	YX

b There are 4 outcomes in total, 2 of which result in XX pair. Hence, $p = \frac{2}{4} = 0.5$.

Exercise 7E

1 Draw a Venn diagram. There are 10 artists with no choice of formats of their albums. Hence, $p = \frac{10}{14} = \frac{5}{7}$.



2 Draw a sample space diagram. Then, add probability of each factor: 1, 343, 1679616 and 5764801. This gives $p=\frac{1}{3}\times 1+\frac{1}{3}\times\frac{2}{3}+\frac{1}{3}\times\frac{2}{3}=\frac{7}{9}$.

	0	**3**	**8**
6	1	216	1679616
7	1	343	5764801

3 Total number of outcomes is $2^4=16$. Favourable outcomes can be written out as: *MMFF, MFMF, MFFM, FMMF, FMFM, FFMM,* i.e. there are 6 favourable outcomes. Hence, the probability $p=\frac{6}{16}=\frac{3}{8}$.

4 Probability that the number chosen is a multiple of 6 (30 from set X or 30, 60, 90 from set Y) is $p=\frac{1}{2}\times\frac{1}{7}+\frac{1}{2}\times\frac{3}{7}=\frac{2}{7}$. Hence, the expected number of points is $54\times\frac{2}{7}=15.4$.

5 Draw the sample space diagrams. P(R = 5) = $\frac{4}{36}=\frac{1}{9}$ and P(T = 5) = $\frac{3}{20}$ so therefore T = 5 is the more likely event.

R	**1**	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

T	**1**	**2**	**4**	**8**
1	2	3	5	9
2	3	4	6	10
3	4	5	7	11
4	5	6	8	12
5	6	7	9	13

6 A vs B: the winner table shows that the probability of dice B winning is P(B) = $\frac{12}{36}=\frac{1}{3}$ and the probability of dice A winning is P(A) = $\frac{24}{36}=\frac{2}{3}$.

C vs D: the winner table shows that that the probability of dice D winning is P(D) = $\frac{12}{36}=\frac{1}{3}$ and the probability of dice C winning is P(C) = $\frac{24}{36}=\frac{2}{3}$.

B vs A	**0**	**0**	**4**	**4**	**4**	**4**	**D vs C**	**2**	**2**	**2**	**2**	**6**	**6**
3	B	B	A	A	A	A	**1**	C	C	C	C	C	C
3	B	B	A	A	A	A	**1**	C	C	C	C	C	C
3	B	B	A	A	A	A	**1**	C	C	C	C	C	C

3	B	B	A	A	A	A	**5**	D	D	D	D	C	C
3	B	B	A	A	A	A	**5**	D	D	D	D	C	C
3	B	B	A	A	A	A	**5**	D	D	D	D	C	C

Exercise 7F

1 a $P(A) = \frac{3}{10}$, $P(B) = \frac{4}{10} = \frac{2}{5}$, $P(A \cap B) = \frac{1}{10}$ and $P(A \cup B) = \frac{6}{10} = \frac{3}{5}$.

b $P(A) + P(B) - P(A \cap B) = \frac{3}{10} + \frac{4}{10} - \frac{1}{10} = \frac{6}{10} = \frac{3}{5} = P(A \cup B)$.

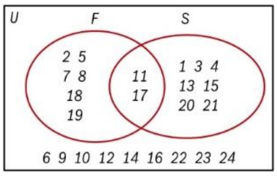
c Events A and B are not mutually exclusive because $P(A \cap B)$ is not 0.

2 a $P(C) = \frac{7}{12}$, $P(D) = \frac{5}{12}$, $P(C \cap D) = 0$ and $P(C \cup D) = \frac{12}{12} = 1$.

b $P(C) + P(D) = \frac{7}{12} + \frac{5}{12} = \frac{12}{12} = 1 = P(C \cup D)$.

c Events C and D are mutually exclusive because $P(C \cap D) = 0$.

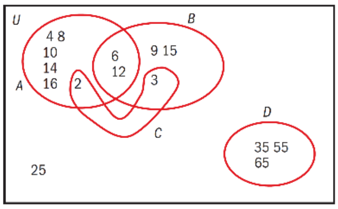
3 a Draw a Venn diagram.



b The events are *not* mutually exclusive because the intersection of the sets S and F (see Venn diagram) is not empty, i.e. both of the events can occur at the same time.

c $p = \frac{15}{24} = \frac{5}{8}$.

4 a Draw the diagram



b A & D, B & D and C & D form mutually exclusive pairs of events.

Exercise 7G

1 a Independent.

b Neither.

c Neither.

d Independent.

e Mutually exclusive.

f Neither.

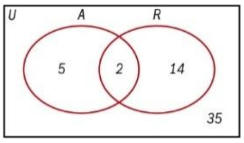
g Independent because $P(T|S) = \frac{7}{10} = P(T)$.

2 $P(A \cup V) = P(A) + P(V) - P(A \cap V)$, $P(A \cap V) = P(A)\,P(V)$ because the events are independent. Hence, $P(A \cup V) = 0.07 + 0.61 - 0.07 \times 0.61 = 0.6373$. $P(A \cup V)$ represents the probability of either the event A, or B, or both A and B happening.

3 a $P(S) \times P(M) = \frac{28+14}{51} \times \frac{14+3}{51} = 0.275$, $P(S \cap M) = \frac{14}{51} = 0.275$. Since $P(S \cap M) = P(M) \times P(S|M)$, $P(S|M)$ must be equal to $P(S)$, hence the events are independent.

b $P(S) = \frac{15+45}{99} = 0.61$ and $P(S|M) = \frac{15}{24} = 0.63$. They are not equal hence the events are not independent.

4 a Draw a Venn diagram. Since n(U) is given, n(A) = 7, n(R) = 16. The events are independent, so P(A∩R) $= \frac{1}{8} \times \frac{2}{7} = \frac{1}{28}$.



b $p = \frac{35+5+14}{56} = 0.964$.

5 a i P(A) = $\frac{2}{11}$

ii P(M|A) = P(M) = $\frac{2}{11}$ (independent events)

iii P(A ∩ M) = P(A) × P(M) = $\frac{4}{121}$

b i P(A) = $\frac{2}{11}$

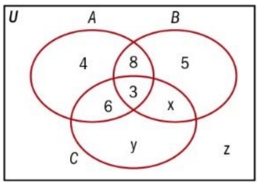
ii P(M|A) = $\frac{2}{10} = \frac{1}{5}$ (now a card is drawn from a set of 10 cards),

iii P(A∩M) = P(A) × P(M|A) = $\frac{2}{55}$

6 a P(B|A) = $\frac{8+3}{4+8+3+6}$ = 0.52.

b P(C|A) = $\frac{6+3}{4+8+3+6}$ = 0.43.

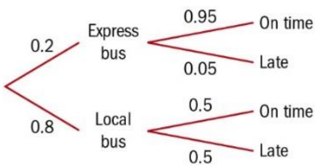
c P(C|B) = $\frac{x+3}{x+3+5+8}$, so x = 7. P(A|C) = $\frac{9}{x+3+6+y}$, so y = 11. Hence, z = 6.



d P(A) = $\frac{4+8+3+6}{50}$ = 0.42, so both A and B, and A and C are pairs of dependent events.

Exercise 7H

1 a



b p = 0.2 × 0.05 + 0.8 × 0.5 = 0.41

2 $p = \frac{13}{35} \times \frac{10+12}{35} + \frac{10}{35} \times \frac{13+12}{35} + \frac{12}{35} \times \frac{10+13}{35} = 0.66$

3 Best way to think about this problem is to consider the probability of *no* 6 occurring in the four throws. That is, $(5/6)^4$. Then, the probability to obtain at least one 6 is $1 - (5/6)^4$. Hence, the best presentation is the last choice. Even if drawing a tree diagram with 1296 branches might take a while, the worst representation is the second one, as it is an incorrect probability calculation.

4 **a** P(QCI) = 0.7 × 0.9 + 0.3 × 0.95 = 0.915.

b P(D|QCI) $= \frac{P(D \cap QCI)}{P(QCI)} = P(QCI \mid D) \times \frac{P(D)}{P(QCI)} = 0.95 \times \frac{0.3}{0.915} = 0.31$

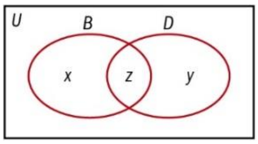
c 2000 × (1 − 0.915) = 170.

d Solve for x in (1 − x) × 0.9 + x × 0.95 = 0.93, x = 0.6 i.e. 60%.

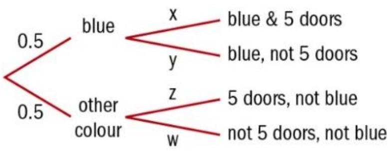
5 Probability of throwing a double six in one throw is $\frac{1}{36}$. Probability of not throwing a double six in one throw is $\frac{35}{36}$. Probability of not throwing a single double six in 24 throws is $\left(\frac{35}{36}\right)^{24}$.

Hence, Probability of throwing at least one double six in 24 throws is $1 - \left(\frac{35}{36}\right)^{24} \approx 0.491$.

6 Draw a Venn diagram labelling unknown quantities x, y, z. Then, construct the following simultaneous equations: x + z = 0.5, z + y = 0.3, x + y + z = 0.6. Solve the equations to obtain x = 0.3, y = 0.1, z = 0.2. Then, the probability that the car is not a blue car with five doors is 1 – z – y – x = 0.4.



Draw a tree diagram.



Then, construct the following simultaneous equations: 0.5(x + z) = 0.3, 0.5(1 + z) = 0.6, z = 0.2, x = 0.4, y = 0.6, w = 0.8, so the probability that the car hasn't got 5 doors and is not blue is 0.5w = 0.4 as before.

7 Probability that no boy is selected is: $\frac{15}{24} \times \frac{14}{23} \times \frac{13}{22} \times \frac{12}{21}$ = 0.128 because four choices are made from a decreasing in size set of girls while the total number of people also decreases. Hence, probability that at least one boy is selected is 1 – 0.13 = 0.872.

Exercise 7I

1 Table **a** does not represent a discrete probability distribution because the probabilities of all possibilities don't add up to 1. Table **b** does not represent a discrete probability distribution because P(B = 2) = −0.2 is negative. Table **c**, however, could represent a discrete probability distribution because all values 0 ≤ P(C = c) ≤ 1, and the probabilities add up to 1.

2 Since $\frac{1+2+3+4+5+6}{21} = 1$, $f(t) = \frac{t-4}{21}$ defines a discrete probability distribution on a given domain.

t	5	6	7	8	9	10
f(t)	$\frac{1}{21}$	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{6}{21}$

3 Find *k* such that $\frac{1+5+7+k}{19} = 1$, i.e. *k* = 6.

4 **a** Sample space *U* = {MMM, MMF, MFM, FMM, MFF, FMF, FFM, FFF}.

b To obtain P(*F* = *f*), divide the number of sequences which correspond to the particular outcome by the total number of sequences.

f	0	1	2	3
P(F = f)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

5 **a** P(A = 12) = 1 − 0.5 − 0.05 − 0.04 − 0.1 − 0.2 = 0.11.

b P(8 < A ≤ 10) = P(A = 9) + P(A = 10) = 0.14.

c P(A is no more than 9) = P(A = 5) + P(A = 8) + P(A = 9) = 0.59.

d P(A is at least 10) = P(A = 10) + P(A = 11) + P(A = 12) = 0.41.

e P(A > 8|A ≤ 11) = $\frac{P(A>8\cap A\le 11)}{P(A\le 11)}=\frac{0.04+0.1+0.2}{0.89}$ = 0.38.

6 **a** Probabilities don't add up to 1.

b Find p such that 0.28 + 0.2 + p + 3p = 1, i.e. p = 0.13.

7 Let P(T = 5) = p. Then, 0.2 + 0.15 + 0.1 + 4p + p = 1, so p = 0.11.

t	1	2	3	4	5
P(T = t)	0.2	0.15	0.1	0.44	0.11

Exercise 7J

1 To find k, note that P(B = b) sums to 1: k((4 − 0) + (4 − 1) + (4 − 2) + (4 − 3)) = 1, k = 0.1.Then, E(B) = 0 × 0.4 + 1 × 0.3 + 2 × 0.2 + 3 × 0.1 = 1.

2 The same probability distribution table applies for M = m. Then, E(M) = $0\times\frac{1}{8}+1\times\frac{3}{8}+2\times\frac{3}{8}+3\times\frac{1}{8}=\frac{12}{8}=1.5.$ Expected number of male and female births in a set of triplets is expected to be equal, and E(M) + E(F) = 3.

3 Construct a probability distribution table of a discrete variable K defined as the number of keys taken out of the handbag.

k	0	1	2
P(K = k)	$\frac{7}{10}\times\frac{6}{9}=\frac{7}{15}$	$\frac{3}{10}\times\frac{7}{9}+\frac{7}{10}\times\frac{3}{9}=\frac{7}{15}$	$\frac{3}{10}\times\frac{2}{9}=\frac{1}{15}$

Hence, the expected number of keys is E(K) $=0\times\frac{7}{15}+1\times\frac{7}{15}+2\times\frac{1}{15}=\frac{3}{5}$.

4 Note, in this problem coins and keys can be treated as the same object making the calculations easier. Construct a probability distribution table of a discrete variable M defined as the number of mints taken out of the handbag.

m	0	1	2
P(M = m)	$\frac{9}{17}\times\frac{8}{16}=\frac{9}{34}$	$\frac{8}{17}\times\frac{9}{16}+\frac{9}{17}\times\frac{8}{16}=\frac{18}{34}$	$\frac{8}{17}\times\frac{7}{16}=\frac{7}{34}$

Hence, the expected number of mints is E(M) $=0\times\frac{9}{34}+1\times\frac{18}{34}+2\times\frac{7}{34}=\frac{32}{34}=\frac{16}{17}$.

5 Expected prize E = US\$3 × $\frac{1}{4}$ + US\$7 × $\frac{1}{8}$ + US\$5 × $\frac{1}{8}$ + US\$2 × $\frac{1}{2}$ = US\$3.25. The expected prize is not US\$5, hence the game is not fair.

6 **a** P(US\$5000) = 0.0001, P(US\$1000) = 0.0005, P(US\$200) = 0.001.

b Since the price of the ticket is US\$10, E = − US\$10 + (US\$5000 × 0.0001 + US\$1000 × 0.0005 + US\$200 × 0.001) = − US\$10 + US\$1.20 = − US\$8.80.

c Expected value should be 0, so the price of a ticket should be US\$1.20.

7 **a** There are 16 outcomes in total, and the probability distribution table is:

d	1	2	3	4	6	8	9	12	16
P(D = d)	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

b P(D is a square number|D < 8) = $\dfrac{P(\text{is a square number} \cap D < 8)}{P(8 < 8)} = \dfrac{\frac{1}{16} + \frac{3}{16}}{\frac{10}{16}} = \dfrac{2}{5}$.

c To make the game fair, the price of the ticket should be equal to the expected value of the prize. E(D) = US\$12 × $\dfrac{1+2+1}{16}$ + US\$6 × $\dfrac{2+3+2+2+2+1}{16}$ = US\$7.5.

8 **a** P(B = 1) = 0.0001, P(B = 2) = 0.0001 × 0.9999, P(B = 3) = 0.0001 × 0.9999^2

b P(B = n) = 0.0001 × 0.9999^{n-1}, for n integer, because $n - 1$ bags of crisps don't contain the golden ticket and n^{th} bag does, $f(b)$ = P(B = b) = $0.0001(0.9999)^{b-1}$.

c $f(b)$ is defined for b positive integers, $b \geq 1$.

d p = P(B = 1) + P(B = 2) + … + P(B = 10) = 0.0001(1 + 0.9999 + … + 0.9999^9) = 0.0009996.

Exercise 7K

1 **a** $X \sim B(7, \frac{1}{2})$

b The 'success' probability is not constant because the die is not replaced.

c $X \sim B(4, \frac{1}{2})$

d $X \sim B(4, 0.3)$

e No 'success' probability.

2 Model this as $X \sim B(6, \frac{1}{2})$ and use technology to find the following probabilities.

a P(X = 3) = 0.313

b P(X ≤ 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 0.891

c P(3 ≤ X < 6) = P(X = 3) + P(X = 4) + P(X = 5) = 0.641

3 Model this as $X \sim B(30, 0.005)$

a P(X = 1) = 0.130

b P(X = 0) = 0.860

c P(X > 3) = 1 − P(X ≤ 3) = 0.0000154.

4 **a** Model this as $X \sim B(5, 0.17)$, P(X > 3) = 1 − P(X ≤ 3) = 0.00361.

b This is $(P(X > 3))^2$ = 0.0000130

5 **a** Model this as $X \sim B(10, 0.085)$ and find the probability of less than 5 panels failing P(X ≤ 4) = 0.999.

b $(P(X \leq 4))^6$ = 0.995.

6 Model this as $X \sim B\left(8, \frac{3}{8}\right)$

a P(X = 5) = 0.101.

b P(X < 5) = 0.863.

c P(X ≤ 5) = 0.964.

7 Model this as $X \sim B(5, 0.964)$, P(X = 4) = 0.15545 while $Y \sim B(5, 0.5)$, P(Y = 4) = 0.15625.

Exercise 7L

1 **a** P(X = 4) = 0.0535.

b P(X ≤ 4) = 0.991.

c P(1 ≤ X < 4) = P(X ≤ 3) − P(X ≤ 1) = 0.809.

d P(X ≥ 2) = 1 − P(X ≤ 1) = 0.558.

e $P(X \le 4|X \ge 2) = \frac{P(X \le 4 \cap X \ge 2)}{P(X \ge 2)} = \frac{P(X \le 4) - P(X \le 1)}{1 - P(X \le 1)} = 0.983.$

f Events are dependent because $P(X \le 4) \ne P(X \le 4|X \ge 2)$.

g $E(X) = np = 1.74$.

h Variance of $X = np(1 - p) = 1.24$.

2 Model this as $Q \sim B\left(7, \frac{1}{2}\right)$ because 2,3,5,7 are prime.

a $P(Q \ge 3) = 1 - P(Q \le 2) = 0.773$.

b $E(Q) = np = 3.5$.

c Variance of $Q = np(1 - p) = 1.75$.

3 **a** Model this as $R \sim B(10,0.78)$

i This is equivalent to 7 reds being thrown, $P(R = 7) = 0.224$

ii $P(3 < R < 7) = P(R \le 6) - P(R \le 3) = 0.157$.

b $P(A) = P(R > 7) = 1 - P(R \le 7) = 0.617$,

$P(B) = P(R < 3) = P(R \le 2) = 0.000160$,

$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0$ because $P(A \cap B) = 0$.

c Events A and B are *not* independent because $P(A|B) \ne P(A)$.

d Events A and B are mutually exclusive because they cannot occur together.

4 Model this as $R \sim B(10,0.1)$, assuming equal probability of the ball falling through the holes.

a $P(R \ge 5) = 1 - P(R \le 4) = 0.00163$

b First, find the probability David scores no points in one game: $P(R = 0) = 0.349$. Next, model this as $G \sim B(6,0.349)$, and find $P(G \ge 2) = 1 - P(G \le 1) = 0.679$.

5 **a** $R \sim B(5,0.964)$ and $B \sim B(5,0.5)$, so $E(R) = 5 \times 0.964 = 4.82$, $E(B) = 2.50$. On average, R scores higher than B.

b Variance of R is $np(1 - p) = 0.174$, variance of B is $np(1 - p) = 1.25$. The results for R are less well spread.

6 **a** To model the random variable A, use binomial distribution: $A \sim B(25,0.2)$.

b $P(A \le 5) = 0.617$.

c $P(A \ge 7) = 1 - P(A \le 6) = 0.220$.

d $P(A \le 3) = 0.234$.

e $E(A) = np = 25 \times 0.2 = 5$, on average, Alex can expect to get 5 answers right by randomly guessing.

f $P(A > 5) = 1 - P(A \le 5) = 1 - 0.62 = 0.383$.

g Alex is expected to score $5 \times 4 - 20 = 0$ points.

h For one student, the probability of answering at least 7 questions correctly is 0.220, so model this as $X \sim B(4,0.220)$. Then, $P(X \ge 2) = 1 - P(X \le 1) = 0.212$.

7 **a** Binomial distribution $T \sim B(538,0.91)$. Assume that whether an individual passenger turns up on time is independent of any other passenger.

b $P(T = 538) = 9.21 \times 10^{-23}$ – it is close to impossible for everyone to turn up on time.

c $P(T \ge 510) = 1 - P(T \le 509) = 0.000672$

d Increase n and check $P(T \ge 510)$, for example $n = 551$ gives $P(T \ge 510) = 0.11$ but $n = 550$ gives $P(T \ge 510) = 0.09$, so $n = 551$.

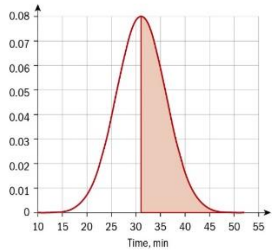
e $E(T) = np$, so choose $n = \frac{538}{0.91} = 591$.

f Using $n = 591$, $P(T = 538) = 0.0573$ and $P(T > 538) = 0.468$, so it is quite likely that more people than there are seats would show up and not very likely that exactly as many people as there are seats would show up.

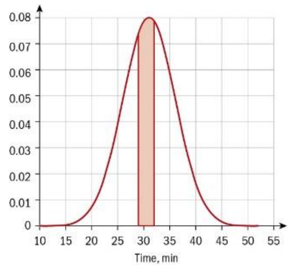
8 $f(p) = np(1-p) = n(p-p^2)$. The function has its maximum when $f'(p) = n(1-2p) = 0$, i.e. when $p = 0.5$ (check it's indeed a maximum and not a minimum by substituting e.g. $f(1) = 0 <$ $f(0.5) = 0.25n$).

Exercise 7M

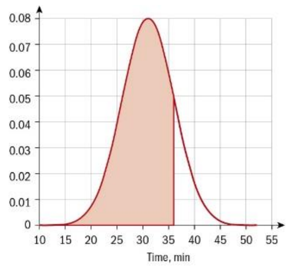
1 a



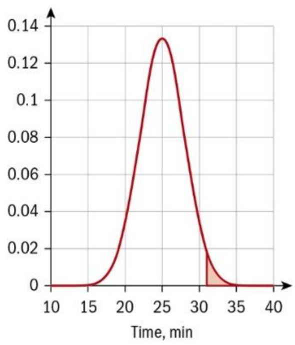
b



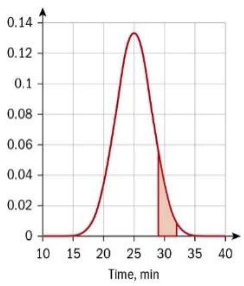
c



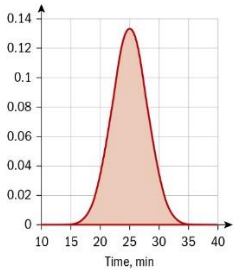
2 a



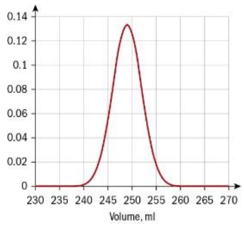
b



c



3 **a**

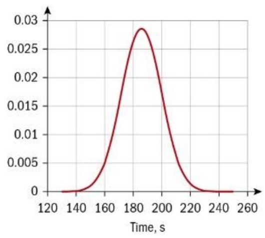


b Since approximately 68% of normal distribution data points are within a standard deviation from the mean and because $\mu - \sigma$ = 249ml − 3ml = 246ml, we expect 0.5 × (100 − 68)% = 16% of the shampoo bottles to contain less than 246ml.

c P(S < 246) = 0.159.

d In one bottle of shampoo, P(S ≥ 250) = 1 − P(S < 250) = 0.37. Model the sample of 200 shampoo bottles using binomial distribution. Let X be the number of bottles that will contain at least 250ml, so X ~ B(200, 0.37). Expected number of X E(X) = 200 × 0.37 = 74.

4 **a**



b Since approximately 95.5% of normal distribution data points are within two standard deviations from the mean and because $\mu + 2\sigma$ = 186s + 28s = 214s, we expect 0.5 × (100 − 95.5)% = 2.25% of the commuter trains to take at least 214s to board all the passengers.

c P(T ≥ 214) = 1 − P(T < 214) = 0.02275 ≈ 2.28%.

d For one commuter train, P(T > 200) = 1 − P(T ≤ 200) = 0.16. Model the sample of 176 commuter trains using binomial distribution. Let X be the number of trains that will take longer than 200 seconds to be fully boarded, so X ~ B(176,0.16). Expected number of X E(X) = 176 × 0.16 = 28.

5 **a** P(T < 17.1) = 0.5 (half of the data below the mean).

b P(T < 14) = 0.16 (68% of the data within standard deviation from the mean).

c P(T > 20.2) = 0.16 (68% of the data within standard deviation from the mean).

d First, P(T ≥ 23.3) = 0.0225 (95.5% of the data within two standard deviations from the mean), then P(14 ≤ T < 23.3) = P(T < 23.3) − P(T ≤ 14) = 1 − 0.0225 − 0.16 = 0.8175

e P(T < 7.8) = 0.0015 (99.7% of the data within three standard deviations from the mean).

f P(T < 23.3|T > 20.2) = $\frac{P(20.2 < T < 23.3)}{P(T < 20.2)} = \frac{1 - 0.0225 - (1 - 0.16)}{0.16} = 0.859$.

6 **a** P(Q < 4) = 0.483

b P(Q < 3.4) = 0.184

c P(Q > 5) = 1 − P(Q ≤ 5) = 0.0829

d P(3.5 ≤ Q < 4.5) = P(Q < 4.5) − P(Q < 3.5) = 0.525

e $P(Q < 4.9|Q > 2.9) = \frac{P(2.9 < Q < 4.9)}{P(Q > 2.9)} = 0.887$

7 **a** A2 (mean is in the middle and data is reasonably spread out), B4 (mean is in the middle and data is of similar density on both sides), C5 (mean is in the middle and most of the data is located at the centre), D1 (mean is in the middle but most data located at the edges), E3 (mean is more towards the left; data to the left of the mean is more dense than to the right)

b Histogram C follows the normal distribution (can approximate with a bell-shaped curve).

c p is true because symmetric histogram has the mean in the middle and quartiles and range is located symmetrically to both sides of the mean; q is true because the normal distribution has got a symmetric histogram; r is not true – histograms A and B provide a perfect counterexample.

Exercise 7N

1 Use inverse normal function to find r = 990g.

2 Use inverse normal function noting that 83% of the packs weigh less than t g: t = 384 g.

3 **a** s: false, t: true, u: true.

b Use inverse normal function and statement u to find Q_3 = 22331.3... g. Then, IQR = $(Q_3 - Q_2) \times 2$ = 405 g.

4 **a** Let $S \sim N(115.7, 10^2)$, then find $P(110 < S < 120) = P(S < 120) - P(S < 110) = 0.382$.

b Model this using the binomial distribution $X \sim B(8, 0.38)$, then the $E(X) = 8 \times 0.382 = 3.06$.

c Find the probability $P(X > 5) = 1 - P(X \le 5) = 0.0400$, assuming that the speeds of the cars are mutually independent.

5 **a** Let $T \sim N(182, 10^2)$, then $P(T > 190) = 1 - P(T \le 190) = 0.212$.

b Model this using the binomial distribution $X \sim B(7, 0.21)$, then $P(X \le 3) = 0.959$.

c Find $P(T < 165) = 0.0446$.

d Model this using the binomial distribution $Y \sim B(10000, 0.04)$, then $E(Y) = 10000 \times 0.0446 = 446$.

6 **a** Let $D \sim N(16, 5^2)$, then $P(13 < D < 15.3) = P(D < 15.3) - P(D < 13) = 0.170$.

b Use inverse normal function noting that 87% of employees travel at most x km to find that x = 21.6 km.

c First, find how many employees travel further than 14km to work: $P(D > 14) = 1 - P(D \le 14) = 0.66$, so there are $23109 \times 0.66 = 15252$ employees living further away than 14km. Hence, $0.91 \times 15252 = 13783 \approx 13800$ employees will fail to get to work on a snow day.

7 **a** Route A takes a shorter time on average, although has a larger deviation while Route B takes a longer time on average but has a very small standard deviation, so is more reliable.

b Let $A \sim N(42, 8^2)$ and $B \sim N(50, 3^2)$, find $P(A \le 45) = 0.646$ and $P(B \le 45) = 0.0478$, so choose route A.

c Model this using the binomial distribution $X \sim B(5, 0.646)$

i $P(X = 5) = 0.113$

ii $P(X \ge 3) = 1 - P(X \le 2) = 0.759$

iii $P(X = 3) = 0.338$, but we are only interested in 3 *consecutive* days. There are three ways in total to choose three consecutive days out of five (starting day 1, starting day 2 and starting day 3), but there are $\frac{5 \times 4}{2}$ ways to choose three days out of five in total. Hence, probability to arrive by 9am on exactly three consecutive days is $\frac{3}{10} \times 0.338 = 0.101$.

8 **a** Use inverse normal function noting that the standard deviation is 5 to find Q_3 = 73.4.

b Half the length of the box is the difference between Q_3 and the mean, 73 − 70 = 3.4, so the length of the box is less than 10 years.

Chapter review

1 **a** There are 12 square numbers between 1 and 150 (12^2 = 144) so the probability is $p = \frac{12}{150} = \frac{2}{25}$

b There are 51 numbers which are at least 100 and at most 150 so the probability is $p = \frac{51}{150}$

c There are 30 numbers divisible by 5 (all the numbers with last digit 0 or 5) between 1 and 150 so the probability is $p = \frac{30}{150} = \frac{1}{5}$

d There 11 numbers which are at least 1 and at most 11 so the probability is $p = \frac{11}{150}$

2 To find β, note that the probabilities for all possible values of k have to add up to 1. Hence, $\beta\left(3^2+2^2+1^2+0^2+1^2\right)=1$, $15\beta=1$, $\beta=\frac{1}{15}$. Then, E(K) = $0\times\frac{9}{15}+1\times\frac{4}{15}+2\times\frac{1}{15}+3\times\frac{0}{15}+4\times\frac{1}{5}=\frac{10}{15}=\frac{2}{3}$

3 **a** Construct simultaneous equations using the fact that probabilities add up to 1 and the expression for E(D): 0.3 + p + q + 0.15 + p − q + p + 2q = 1, i.e. 3p + 2q = 0.55 and 0 × 0.3 + 1 × (p + q) + 2 × 0.15 + 3 × (p − q) + 4 × (p + 2q) = 1.7, i.e. 4p + 3q = 0.7. Solve the equations to obtain $p=0.25, q=-0.1$.

b P(D = 3|D ≥ 1) = $\frac{P(D=3\cap D\geq 1)}{P(D\geq 1)}=\frac{P(D=3)}{1-P(D=0)}$ = $\frac{0.35}{0.7}$ = 0.5.

4 **a** Probability that the seed grows: p = 0.65 × 0.85 + 0.35 × 0.74 = 0.8115.

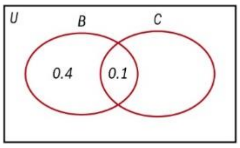
b Conditional probability: P(Green & Grows) = P(Green) × P(Grows|Green) = 0.65 × 0.85 = 0.5525.

c P(Red or Grows) = P(Red) + P(Green & Grows) = 0.35 + 0.5525 = 0.9025. Alternatively, P(Red or Grows) = 1 − P(Green & Doesn't grow) = 1 − 0.65 × 0.15 = 0.9025 leads to the same answer.

5 **a** In one throw, there is a $\frac{5}{6}$ chance of throwing no sixes at all. Hence, when the die is thrown n times, P(Throw at least one six in n throws) = 1 − P(Throw no sixes in n throws) = 1 − $\left(\frac{5}{6}\right)^n$.

b 1 − $\left(\frac{5}{6}\right)^n > 0.995$ corresponds to $\left(\frac{5}{6}\right)^n < 0.005$, $n > \frac{\log 0.005}{\log\frac{5}{6}}$ = 29.06... so take n = 30.

6 Draw a Venn diagram to visualise the situation. P(B′∪C) = P(B′) + P(B ∩ C) = 1 − P(B) + P(B∩C) = 1 − 0.4 − 0.1 + 0.1 = 0.6.



7 **a** Construct the table noting there are 36 outcomes in total and counting ways to obtain each of the value of t.

t	4	5	6	7	8	9	10
P(T = t)	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{8}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{4}{36}$	$\frac{1}{36}$

b Prime numbers: 5,7, square numbers: 4,9. For game to be fair, E(T) = 0. This gives the following equation: $\frac{5\times 4}{36}+\frac{7\times 10}{36}-x\left(\frac{1}{36}+\frac{4}{36}\right)$ = 0, i.e. 90 − 5x = 0, x = 18.

8 Let $W \sim N(65,11^2)$. Then:

a P(W > 70) = 0.325.

b Use inverse normal function to find UQ = 72.42kg and LQ = 57.58kg, so IQR = 14.8kg.

c Use inverse normal function for 92.7% to find 81.0kg.

d Use binomial distribution to model this: $X \sim$ B(8,0.325), then P($X \leq$ 3) = 0.758.

e P(W < 60) = 0.3247, so 1000 $\times$ 0.3247 $\approx$ 325.

9 a There are at most 5 turns before a green balls is definitely picked. P(Judith wins) = P(Judith wins on her first go) + P(Judith wins on her second go) + P(Judith wins on her third go).

P(Judith wins on her first go) = $\frac{3}{7}$, P(Judith wins on her second go) $= \frac{4}{7}\times\frac{3}{6}\times\frac{3}{5} = \frac{6}{35}$,

P(Judith wins on her third go) $= \frac{4}{7}\times\frac{3}{6}\times\frac{2}{5}\times\frac{1}{4} = \frac{4}{175}$. Hence, P(Judith wins) $= \frac{109}{175} =$ 0.629.

b Now that the ball chosen is replaced after each turn, it might take infinitely many turns until the green ball is picked. To find the new probability of Judith winning p, note that after Judith and Gilles both had an unsuccessful turn each, the probability of Judith winning from that point resets to the original value p and the following equation can be constructed:

$p = \frac{3}{7} + \frac{4}{7}\times\frac{4p}{7}$, $p = \frac{21}{33} =$ 0.636. Hence, Judith is more likely to win in this set up of the game.

10 a $\left(\frac{3}{16}\times\frac{13}{15}\right)+\left(\frac{13}{16}\times\frac{3}{15}\right)$ M1A1

$= \frac{39}{240}+\frac{39}{240} = \frac{78}{240}\left(=\frac{13}{40}\right)$ A1

b $1-\left(\frac{13}{16}\times\frac{12}{15}\right)$ M1A1

$1-\frac{156}{240} = \frac{84}{240}\left(=\frac{7}{20}\right)$ A1

11 a $0.7\times0.4\times0.8 = 0.224$ ~ M1A1

b $(0.7\times0.6\times0.2)+(0.3\times0.4\times0.2)+(0.3\times0.6\times0.8)$ M1A1

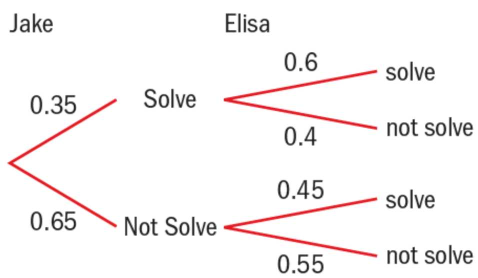
$= 0.252$ A1

c $0.4\times0.8 = 0.32$ M1A1

d $1-(0.3\times0.6\times0.2)$ M1A1

$= 0.964$ A1

12 a



M1A1A1

b $1-(0.65\times0.55) = 0.6425$ M1A1

c $\frac{P(\text{Jake and Elisa solve})}{P(\text{Elisa solve})} = \frac{0.35\times0.6}{(0.35\times0.6)+(0.65\times0.45)}$ M1A1A1

$= 0.418$ A1

13 a 0.4 M1A1

b 0.6 M1A1

c 0.75 M1A1

14 a $32+25-48 = 9$ M1A1

b $\frac{32-9}{48}$ M1A1

$=\frac{23}{48}$ A1

c $P(E|U)=\frac{9}{32}$ and $P(E)=\frac{32}{48}=\frac{2}{3}$ A1A1

$P(E|U)\neq P(E)$, so not independent. R1

15 a $\frac{43}{50}$ M1A1

b $\frac{7}{25}$ M1A1

c $\frac{5}{11}$ M1A1

d 0 M1A1

e $\frac{5}{34}$ M1A1

16 a $P(A\cap B)=P(A)P(B)=0.3\times 0.15=0.045$ M1A1

b $P(A\cup B)=P(A)+P(B)-P(A\cap B)=0.3+0.5-0.045=0.755$ M1A1

c $P(B'\cap A)=P(A)-P(A\cap B)=0.3-0.15=0.15$ M1A1

d $P(B|A')=\frac{P(B\cap A')}{P(A')}$ M1

$=\frac{P(B)-P(A\cap B)}{P(A')}=\frac{0.5-0.045}{0.7}$ M1

$=0.65$ A1

17 a $1-P(\text{no scoring})=1-0.72^4$ M1A1

$=0.731$ A1

b $4\times 0.28\times 0.76^3$ M1

$=0.492$ A1

c $1-P(\text{no goals})-P(\text{exactly one goal})$ M1

$=1-0.72^4-0.492$ A1

$=0.24$ A1

18 a Let X be the discrete random variable 'number of boys'.

So $X\sim B(10,0.512)$

$P(X=6)=\binom{10}{6}0.512^6(1-0.512)^4$ M1A1

$=0.215$ A1

b $P(X=0)=\binom{10}{0}0.512^0(1-0.512)^{10}=0.000766$ M1A1

c $P(X\le 4)=0.348$ M1A1

19 a $\frac{k}{2}+k+k^2+2k^2+\frac{k}{2}=1$ M1

$3k^2+2k-1=0$ A1

$(3k-1)(k+1)=0$ M1

$\Rightarrow k=\frac{1}{3}$ A1

b $E(X) = 0 \times \frac{k}{2} + 0.5 \times k + 1 \times k^2 + 1.5 \times 2k^2 + 2 \times \frac{k}{2}$ M1

$E(X) = \frac{k}{2} + k^2 + 3k^2 + k = 4k^2 + \frac{3k}{2}$ A1

$= 4\left(\frac{1}{3}\right)^2 + \frac{3}{2}\left(\frac{1}{3}\right) = \frac{4}{9} + \frac{1}{2} = \frac{17}{18}$ M1A1

c $P(X \geq 1.25) = 2k^2 + \frac{k}{2}$ M1

$= 2\left(\frac{1}{3}\right)^2 + \frac{1}{6} = \frac{2}{9} + \frac{1}{6} = \frac{7}{18}$ M1A1

20 a Let X be the discrete random variable 'time taken for Blossom to walk to her cafe'.

So $X \sim N(35, 3.4^2)$

$P(X > 37) = 0.278$ M1A1

b $P(X < 36.5) - P(X < 34)$ A1

$= 0.670 - 0.384$ A1

$= 0.286$ A1

c $P(X < 30) = 0.071$ M1

$0.071 \times 25 = 1.77$ M1

So approximately two occasions. A1

21 a Let X be the discrete random variable 'mass of a can of baked beans'.

Then $X \sim N(415, 12^2)$

Using GDC

$P(X > m) = 0.65$ M1A1

$\Rightarrow m = 410.4$ A1

b You require $P(X > 422.5 \mid X > 420)$ M1

$P(X > 422.5 \mid X > 420) = \frac{P(X > 422.5)}{P(X > 420)}$ M1

$= \frac{0.266}{0.338}$ A1

$= 0.787$ A1

c Using GDC

$P(X < 413.5) = 0.450$ M1A1

Now using $Y \sim B(144, 0.450)$ M1

$P(Y \geq 75) = 0.0524$ A1

8 Testing for validity: Spearman's, hypothesis testing and χ^2 test for independence

Skills check

1 $P(S) = \frac{2}{6} = \frac{1}{3}$ while $P(S|E) = \frac{1}{3}$. $P(S) = P(S|E)$ so the events are independent.

2 $D \sim N(35, 3^2)$, $P(D < 36) = 0.631$

3

x	0	1	2	3	4
P(X = x)	0.0016	0.0256	0.1536	0.4096	0.4096

Exercise 8A

1 **a** 1 (data monotonically increasing) **b** 1 (data monotonically increasing)

c − 1(data monotonically decreasing)

d 0(data is not consistently increasing or decreasing)

2 Put the ranked data into a GDC and obtain PMCC $r_s = 0.2$ so there is only weak positive correlation between the taste and value for money.

3 **a** The ranks are (note that when more than one piece of data have the same value, the average of the rank given is used):

x	7	6	5	4	3	2	1
y	1	2	3	4	6	6	6

Use GDC to find the PMCC for the ranked data: $r_s = -0.964$.

b The ranks are:

x	3	2	4	6	7	1	5
y	2	3	5	7	6	1	4

Use GDC to find the PMCC for the ranked data: $r_s = 0.893$.

4 **a** PMCC is used for linear relationships and the scatter plot shows the relationships is not linear.

b The ranks of A – L are:

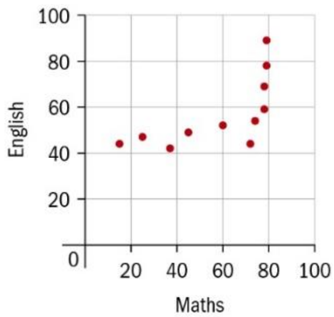
v	11	10	9	8	7	6	5	4	3	1	12	2
F	1	2	4	5	6.5	6.5	8	11	9	10	3	12

Use GDC to find the PMCC for the ranked data: $r_s = -0.942$.

c There is a strong negative correlation between velocity and force, as can be expected from the scatter plot. Since the force does not change significantly at high values of the velocity, the value of r_s could be affected significantly by small changes in data.

5 **a** Use GDC to find the PMCC for the data: $r_s = 0.670$.

b Scatter plot:



PMCC indicates that there is a positive correlation of medium strength between the English and Maths scores. The scatter plot shows strong but non-linear positive correlation between the scores..

c The ranks are:

Maths	11	10	9	8	7	6	5	3.5	3.5	1.5	1.5
English	9.5	8	11	7	6	9.5	5	4	3	2	1

Use GDC to find the PMCC for the ranked data: r_s = 0.883. This indicates a strong positive correlation between the scores which is a more realistic result given the scatter plot.

d Spearman's rank correlation because the data points are not linear.

6 **a** Because instead of quantitative data the ranks of the taste are given.

b The ranks are:

Taste rank	1	2	3	4	5	6
Cost rank	1	3	2	5	4	6

Use GDC to find the PMCC for the ranked data: r_s = 0.886, so there is a strong positive correlation between the price and taste of coffee.

7 **a** Data table:

x	0.82	1.28	1.78	1.46	2.46	2.48	2.02	3.02	2.98	7.46
y	0.86	1.56	1.22	0.62	0.84	1.76	1.82	1.42	0.62	4.98

Use GDC to find the PMCC

i with the outlier J: r = 0.874 **ii** without the outlier J: r = 0.0776

b **i** The ranked data table with the outlier J:

x	10	9	7	8	5	4	6	2	3	1
y	7	4	6	9.5	8	3	2	5	9.5	1

Use GDC to find the PMCC for the ranked data with the outlier J: r_s = 0.304.

ii The ranked data table without the outlier J:

x	9	8	6	7	4	3	5	1	2
y	6	3	5	8.5	7	2	1	4	8.5

Use GDC to find the PMCC for the ranked data without the outlier J: r_s = 0.0418.

c Even though both of the measures are affected by the outlier, Spearman's rank correlation coefficient is affected less.

Exercise 8B

1 **a** Probability that a person chosen at random likes black cars best is $\frac{22}{80}$. Probability that a person chosen at random is male is $\frac{38}{80}$. If the two events are independent, the expected number of males who prefer black cars is $80 \times \frac{22}{80} \times \frac{38}{80} = 10.45$.

b Probability that a person chosen at random likes white cars best is $\frac{18}{80}$. Probability that a person chosen at random is male is $\frac{42}{80}$. If the two events are independent, the expected number of males who prefer black cars is $80 \times \frac{18}{80} \times \frac{42}{80} = 9.45$.

c Use GDC to find the value $\chi^2 = 4.69$.

2 **a** Probability that a person chosen at random buys small coffee is $\frac{34}{110}$. Probability that a person chosen at random is male is $\frac{54}{110}$. If the two events are independent, the expected number of males who prefer black cars is $110 \times \frac{34}{110} \times \frac{54}{110} = 16.7$.

b Probability that a person chosen at random buys large coffee is $\frac{46}{110}$. Probability that a person chosen at random is female is $\frac{56}{110}$. If the two events are independent, the expected number of males who prefer black cars is $110 \times \frac{46}{110} \times \frac{56}{110} = 23.4$.

c Use GDC to find the value $\chi^2 = 5.21$.

3 Find totals:

Pet	**Rabbits**	**Guinea pigs**	**Hamsters**	**Totals**
Lettuce	16	16	28	60
Carrots	34	18	18	70
Totals	50	34	46	130

a Probability that a pet chosen at random eats carrots is $\frac{70}{130}$. Probability that a pet chosen at random is a rabbit is $\frac{50}{130}$. If the two events are independent, the expected number of rabbits who eat carrots is $130 \times \frac{70}{130} \times \frac{50}{130} = 26.9$.

b Probability that a pet chosen at random eats lettuce is $\frac{60}{130}$. Probability that a pet chosen at random is a hamster is $\frac{46}{130}$. If the two events are independent, the expected number of rabbits who eat carrots is $130 \times \frac{60}{130} \times \frac{46}{130} = 21.2$.

c Use GDC to find the value $\chi^2 = 8.05$.

4 Find totals:

Transport	**Car**	**Bus**	**Bicycle**	**Walk**	**Totals**
Male	12	12	28	8	60
Female	21	13	15	11	60
Totals	33	25	43	19	120

a Probability that a person chosen at random comes by bicycle is $\frac{43}{120}$. Probability that a person chosen at random is a male is $\frac{60}{120}$. If the two events are independent, the expected number of males who come by bicycle is $120 \times \frac{43}{120} \times \frac{60}{120} = 21.5$.

b Probability that a person chosen at random comes by car is $\frac{33}{120}$. Probability that a person chosen at random is a female is $\frac{60}{120}$. If the two events are independent, the expected number of females who come by car is $120 \times \frac{33}{120} \times \frac{60}{120} = 16.5$.

c Use GDC to find the value $\chi^2 = 6.90$.

Exercise 8C

1 a Contingency table:

Sport	Cycling	Basketball	Football	Totals
Males	7	10	6	23
Females	9	8	10	27
Totals	16	18	16	50

b H_0: favourite sport is independent of gender. H_1: favourite sport is not independent of gender.

c v = (rows − 1)(columns − 1) = $(2-1)\times(3-1) = 2$.

d Use GDC to find the values $\chi^2 = 1.16$ and $p = 0.560$.

e Expected values calculated by E(F/M liking C/B/F) = P(F/M) × P(likes C/B/F) × total:

Sport	Cycling	Basketball	Football
Males	7.36	8.28	7.36
Females	8.64	9.72	8.64

All expected values are greater than 5.

f $\chi^2 = 1.16 < 4.605$ so H_0 is accepted.

g Find p value for $\chi^2{}_c = 4.605$: $p_c = 0.10$, $p = 0.560 > 0.10$ so H_0 is accepted and p value supports the conclusion.

2 a H_0: favourite bread is independent of gender. H_1: favourite bread is not independent of gender.

b E(Female liking white bread) = P(Female) × P(likes white bread) × total = $\frac{41}{80}\times\frac{31}{80}\times 80$

$= 15.8875 \approx 15.9$

c v = (rows − 1)(columns − 1) = $(2-1)\times(4-1) = 3$.

d Use GDC to find the values $\chi^2 = 2.12$ and $p = 0.548$.

e $\chi^2 < \chi^2{}_c$ so H_0 is accepted.

3 a H_0: favourite genre of film is independent of age. H_1: favourite genre of film is not independent of age.

b E(20– 50 year-olds prefer horror films)

= P(20– 50 year-olds) × P(prefers horror films) × total = $\frac{130}{300}\times\frac{77}{300}\times 300 = 33.367 \approx 33.4$.

c v = (rows − 1)(columns − 1) = $(3-1)\times(3-1) = 4$.

d Use GDC to find the values $\chi^2 = 45.2$ and $p = 3.6\times10^{-9}$.

e $p < 0.10$ (for 10% test) so H_0 is rejected.

4 a H_0: favourite flavour of dog food is independent of breed. H_1: favourite flavour of dog food is not independent of breed.

b Expected values calculated by E(F/M liking C/B/F) = P(F/M) × P(likes C/B/F) × total:

Flavour	Beef	Chicken	Lamb
Boxer	11.7	7.00	9.33
Labrador	15.8	9.50	12.7
Poodle	14.6	8.75	11.7
Collie	7.92	4.75	6.33

Not all expected values are greater than 5.

c New contingency table:

Flavour	Beef	Chicken	Lamb	Totals
Boxer	14	6	8	28
Labrador	17	11	10	38
Poodle/Collie	19	13	22	54
Totals	50	30	40	120

d Use GDC to find the values $\chi^2 = 3.14$ and $p = 0.535$.

e $p = 0.535 > 0.05$ so H_0 is accepted.

5 a H_0: favourite flavour of chocolate is independent of gender. H_1: favourite flavour of chocolate is not independent of gender.

b v = (rows − 1)(columns − 1) = $(2-1)\times(3-1) = 2$.

c Use GDC to find the values $\chi^2 = 9.52$ and $p = 0.00856$.

d $\chi^2 = 9.52 > 9.210$so H_0 is rejected.

6 a H_0: GPA is independent of number of hours spent on social media. H_1: GPA is not independent of number of hours spent on social media.

b E(0− 9 hours and high GPA) = P(0− 9 hours) × P(high GPA) × total = $\frac{85}{270}\times\frac{99}{270}\times 270$

= $31.167 \approx 31.2$.

c v = (rows − 1)(columns − 1) = $(3-1)\times(3-1) = 4$.

d Use GDC to find the values $\chi^2 = 78.5$ and $p = 3.6\times10^{-16}$.

e $\chi^2 = 78.5 > 7.779$ so H_0 is rejected.

7 State the null hypothesis and the alternative hypothesis: H_0: the number of people walking their dog is independent of the time of the day. H_1: the number of people walking their dog is not independent of the time of the day.

Noting that the number of degrees of freedom is $v = 4$, use GDC to find the values $\chi^2 = 5.30$ and $p = 0.257$.

Since $0.257 > 0.05$ and $5.30 < 9.488$, H_0 is accepted.

8 a H_0: the number of bottles of water sold is independent of temperature. H_1: the number of bottles of water sold is not independent of temperature.

b v = (rows − 1)(columns − 1) = $(3-1)\times(3-1) = 4$.

c Use GDC to find the values $\chi^2 = 3.30$ and $p = 0.509$.

d Since $0.509 > 0.01$ and $3.30 < 13.277$, H_0 is accepted.

9 a H_0: annual salary is independent of the type of degree. H_1: annual salary is not independent of the type of degree.

b v = (rows − 1)(columns − 1) = $(3-1)\times(3-1) = 4$.

c Use GDC to find the values $\chi^2 = 24.4$ and $p = 6.53\times10^{-5}$.

d Since $6.53\times10^{-5} < 0.05$ and $24.4 > 9.488$, H_0 is rejected.

Exercise 8D

1 a Expected frequencies:

Colour	Frequency
Yellow	120
Orange	120
Red	120
Purple	120
Green	120

b $v = (n-1) = 5-1 = 4$

c H_0: The data satisfies a uniform distribution.

H_1: The data does not satisfy a uniform distribution.

Use GDC to find the values $\chi^2 = 10.45$ and $p = 0.0335$.

d Since $0.0335 < 0.05$ and $10.45 > 9.488$, H_0 is rejected.

2 **a** Expected frequencies:

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Freq	5	5	5	5	5	5	5	5	5	5	5	5

b $v = (n - 1) = 12 - 1 = 11$.

c H_0: The data satisfies a uniform distribution.

H_1: The data does not satisfy a uniform distribution.

Use GDC to find the values $\chi^2 = 6$ and $p = 0.873$.

d Since $0.873 > 0.10$and $6 < 17.275$, H_0 is accepted.

3 **a** E(number of calls) = $\dfrac{840}{7} = 120$.

b $v = (n - 1) = 7 - 1 = 6$.

c H_0: The data satisfies a uniform distribution.

H_1: The data does not satisfy a uniform distribution.

Use GDC to find the values $\chi^2 = 86.1$ and $p = 1.97 \times 10^{-16}$

d Since $1.97 \times 10^{-16} < 0.05$ and $86.1 > 12.592$, H_0 is rejected.

4 **a** Expected frequencies

Last digit	0	1	2	3	4	5	6	7	8	9
Frequency	49	49	49	49	49	49	49	49	49	49

b $v = (n - 1) = 10 - 1 = 9$.

c H_0: The data satisfies a uniform distribution.

H_1: The data does not satisfy a uniform distribution.

Use GDC to find the values $\chi^2 = 9.06$ and $p = 0.432$.

d Since $0.432 > 0.10$ and $9.06 < 14.684$, H_0 is accepted.

Exercise 8E

1 **a** H_0: The lengths are normally distributed with mean of 19 cm and standard deviation of 3 cm.

H_1: The lengths are not normally distributed with mean of 19 cm and standard deviation of 3 cm.

b Let $L \sim N(19, 3^2)$, then use GDC to find $P(9 < L < 12) = 0.00939$

c Expect $250 \times 0.00939 = 2.35$ fish.

d Repeat the same procedure to obtain the expected frequency table:

Length of fish, x cm	**Probability**	**Expected frequency**
$9 \le x < 12$	0.009386	2.35
$12 \le x < 15$	0.081396	20.35
$15 \le x < 18$	0.278230	69.56
$18 \le x < 21$	0.378066	94.52
$21 \le x < 24$	0.204702	51.18
$24 \le x < 27$	0.043960	10.99
$27 \le x < 30$	0.003708	0.927

e Combine the rows with expected frequencies less than five with the rows next to them, i.e. the top row with the second row and the last row with the second to last row.

f Updated table:

Length of fish, x cm	**Frequency**	**Expected Frequency**
$9 \le x < 15$	27	22.70
$15 \le x < 18$	71	69.56
$18 \le x < 21$	88	94.52
$21 \le x < 24$	52	51.18
$24 \le x < 30$	12	11.92

g $v = (n - 1) = 5 - 1 = 4$.

h Use GDC to find the values $\chi^2 = 1.31$ and $p = 0.860$.

i Since $0.860 > 0.05$ and $1.31 < 9.488$, H_0 is accepted.

2 Let $W \sim N(52, 3^2)$, and calculate the probabilities using GDC. Multiply them by 200 (total number of girls).

a Table with the expected frequencies:

Weight, *w* kg	$w < 45$	$45 \le w < 50$	$50 \le w < 55$	$55 \le w < 60$	$w \ge 60$
Expected frequency	1.96	48.54	117.77	30.96	0.77

b Merge the first and the last columns with their neighbouring columns to obtain an updates table:

Weight, *w* kg	$w < 50$	$50 \le w < 55$	$55 \le w$
Observed frequency	56	82	62
Expected frequency	50.50	117.77	31.73

c $v = (n - 1) = 3 - 1 = 2$.

d H_0: The weights are normally distributed with mean of 52 kg and standard deviation of 3 kg.

H_1: The weights are not normally distributed with mean of 52 kg and standard deviation of 3 kg.

Use GDC to find the values $\chi^2 = 40.3$ and $p = 1.74 \times 10^{-9}$.

e Since $1.74 \times 10^{-9} < 0.05$ and $40.3 > 5.991$, H_0 is rejected.

3 Let $X \sim N(65, 7.5^2)$, and calculate the probabilities using GDC. Multiply them by 300 (total number of students).

a Table with expected frequencies:

Grade, x%	$x < 50$	$50 \le x < 60$	$60 \le x < 70$	$70 \le x < 80$	$80 \le x$
Expected frequency	6.83	68.92	148.50	68.92	6.83

b $v = (n - 1) = 5 - 1 = 4$.

c H_0: The grades are normally distributed with mean of 65% and standard deviation of 7.5%.

H_1: The grades are not normally distributed with mean of 65% and standard deviation of 7.5%.

Use GDC to find the values $\chi^2 = 0.705$ and $p = 0.951$.

d Since $0.951 > 0.1$ and $0.705 < 7.779$, H_0 is accepted.

4 Let $H \sim N(250, 11^2)$, and calculate the probabilities using GDC. Multiply them by 250 (total number of elephants).

a Table with expected frequencies.

Height, *h* cm	$h < 235$	$235 \le h < 245$	$245 \le h < 255$	$255 \le h < 265$	$265 \le h$
Expected frequency	21.59	59.59	87.64	59.59	21.59

b $v = (n - 1) = 5 - 1 = 4$.

c H_0: The heights are normally distributed with mean of 250cm and standard deviation of 11 cm.

H_1: The heights are not normally distributed with mean of 250cm and standard deviation of 11 cm.

Use GDC to find the values $\chi^2 = 8.02$ and p = 0.0910.

d Since 0.0910 > 0.05 and 8.02 < 9.488, so H_0 is accepted.

5 Let $H \sim N(1200, 100^2)$, and calculate the probabilities using GDC. Multiply them by 400 (total number of light bubs).

a Table of expected frequencies.

Lifespan, *h* hours	$h <$ 1000	$1000 \le h <$ 1100	$1100 \le h <$ 1200	$1200 \le h <$ 1300	$1300 \le h <$ 1400	$1400 \le$ h
Freq	9.1	54.36	136.54	136.54	54.36	9.1

b $v = (n - 1) = 6 - 1 = 5$.

c H_0: The lifespan is normally distributed with mean of 1200 hours and standard deviation of 100 hours.

H_1: The heights are not normally distributed with mean of 1200 hours and standard deviation of 100 hours.

Use GDC to find the values $\chi^2 = 78.7$ and p = 1.5×10^{-15} .

d Since 1.5×10^{-15} < 0.05 and 78.7 > 11.070, so H_0 is rejected.

Exercise 8F

1 a Expected probabilities for $S \sim B(3, 0.75)$: $P(S = 0) = \binom{3}{0}0.75^0 0.25^3 = 0.015625$, $P(S = 1) = \binom{3}{1}0.75^1 0.25^2 = 0.140625$, $P(S = 2) = \binom{3}{2}0.75^2 0.25^1 = 0.421875$, $P(S = 3) = \binom{3}{3}0.75^3 0.25^0 = 0.421875$.

b Table of expected frequencies: multiply the probabilities by total number of seeds 50.

Number of seeds germinating	0	1	2	3
Expected frequency	0.78	7.03	21.09	21.09

c Expected value of no seeds germinating is less than 5. Hence, combine this with the data of 1 seed germinating:

Number of seeds germinating	0,1	2	3
Observed frequency	15	15	20
Expected frequency	7.81	21.09	21.09

d $v = (n - 1) = 3 - 1 = 2$.

e H_0: The number of germinating seeds follows a binomial distribution.

H_1: The number of germinating seeds does not follow a binomial distribution.

Use GDC to find the values $\chi^2 = 8.43$ and p = 0.0147.

f Since 0.0147 < 0.05 and 8.43 > 5.991, H_0 is rejected.

2 a Expected probabilities for $S \sim B(3, 0.5)$: $P(S = 0) = \binom{3}{0}0.5^0 0.5^3 = 0.125$, $P(S = 1) = \binom{3}{1}0.5^1 0.5^2 = 0.375$, $P(S = 2) = \binom{3}{2}0.5^2 0.5^1 = 0.375$, $P(S = 3) = \binom{3}{3}0.5^3 0.5^0 = 0.125$.

Table of expected frequencies: multiply the probabilities by total number of families: 100.

Number of boys	0	1	2	3
Expected frequency	12.5	37.5	37.5	12.5

b No expected frequency values below 5.

c $v = (n - 1) = 4 - 1 = 3$.

d H_0: The number of boys follows a binomial distribution.

H_1: The number of boys does not follow a binomial distribution.

Use GDC to find the values $\chi^2 = 10.77$ and $p = 0.0130$.

e Since 0.0130> 0.01 and 10.77 < 11.345, H_0 is accepted.

3 a Expected probabilities for $S \sim B(2, \frac{1}{6})$: $P(S = 0) = \binom{2}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^2 = 0.694$, $P(S = 1) =$

$\binom{2}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^1 = 0.278$, $P(S = 2) = \binom{2}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^0 = 0.0278$.

Table of expected frequencies: multiply the probabilities by total number of tosses: 250.

Number of 6s	0	1	2
Expected frequency	173.61	69.44	6.94

b No expected frequency values below 5.

c $v = (n - 1) = 3 - 1 = 2$.

d H_0: The number of 6s follows a binomial distribution.

H_1: The number of 6s does not follow a binomial distribution.

Use GDC to find the values $\chi^2 = 28.1$ and $p = 7.7 \times 10^{-7}$.

e Since 7.7 < 0.05 and 28.1 > 5.991, H_0 is rejected.

4 a Since there are four answers, P(getting a question right) = 0.25.

b,c Model this using the binomial distribution: $X \sim B(5,0.25)$.

Number correct	0	1	2	3	4	5
Probability	0.2373	0.3955	0.2637	0.0879	0.0146	0.0010
Expected frequency	118.65	197.75	131.84	43.95	7.30	0.50

d Expected frequency of 5 correct answers is less than 5, hence need to combine the data for 4 and 5 correct answers. The new table is:

Number correct	0	1	2	3	4,5
Observed frequency	38	66	177	132	87
Expected frequency	118.65	197.75	131.84	43.95	7.8

e $v = (n - 1) = 5 - 1 = 4$.

f H_0: The number of correct answers follows a binomial distribution.

H_1: The number of correct answers does not follow a binomial distribution.

Use GDC to find that $\chi^2 = 1139$ and $p = 0$.

g H_0 is rejected as 0 < 0.05 and 1138 > 9.488.

Exercise 8G

1 a H_0: $\bar{x}_1 = \bar{x}_2$ (there is no difference in the weights)

H_1: $\bar{x}_1 < \bar{x}_2$ (there is a difference in the weights of the apples: apples from the shade weigh less).

b This is a one-tailed test as Petra is trying to find out if the trees in the shade weigh less.

c Use GDC to find $t = -0.687$ and $p = 0.251$.

d Since 0.251 > 0.10, H_0 is accepted.

2 a H_0: $\bar{x}_1 = \bar{x}_2$ (there is no difference between the weights of town and country babies),

H_1: $\bar{x}_1 > \bar{x}_2$ (babies born in the country weigh more than babies born in the town).

b This is a one-tailed test as Fergus is trying to find out if babies born in the country weigh more than babies born in the town.

c Use GDC to find $t = -0.1913$ and $p = 0.575$.

d Since 0.575 > 0.10, H_0 is accepted. Comment: author used t-test at the 5% significance level, not sure why the answers are different.

3 **a** H_0: $\bar{x}_1 = \bar{x}_2$ (there is no difference between the lengths of the beans),
H_1: $\bar{x}_1 \neq \bar{x}_2$ (there is a difference between the lengths of the beans).
b This is a two-tailed test as Jocasta is interested in finding out if the lengths are different.
c Use GDC to find $t = -3.126$ and $p = 0.00584$.
d Since $0.00584 < 0.05$, H_0 is rejected.

4 **a** H_0: $\bar{x}_1 = \bar{x}_2$ (there is no difference between the lifetimes of the bulbs),
H_1: $\bar{x}_1 \neq \bar{x}_2$ (there is a difference between the lifetimes of the bulbs).
b This is a two-tailed test.
c Use GDC to find $t = 0.3$ and $p = 0.769$
d Since $0.769 > 0.05$, H_0 is accepted.

5 **a** H_0: $\bar{x}_1 = \bar{x}_2$ (there is no difference between the weights of the girls and the boys),
H_1: $\bar{x}_1 < \bar{x}_2$ (the boys weigh less than the girls).
b This is a one-tailed test (testing if the boys weighed less than the girls).
c Use GDC to find $t = -2.45$ and $p = 0.015$.
d Since $0.015 < 0.05$, H_0 is rejected.

6 **a** H_0: $\bar{x}_1 = \bar{x}_2$ (there is no difference between the weight loss with and without the remedy),
H_1: $\bar{x}_1 > \bar{x}_2$ (people lose more weight with the remedy than without it).
b This is a one-tailed test (testing if the weight loss is higher with the remedy).
c Use GDC to find $t = 2.84$ and $p = 0.00539$.
d Since $0.00539 < 0.01$, H_0 is rejected.

7 **a** H_0: $\bar{x}_1 = \bar{x}_2$ (there is no difference in the lengths of the sweetcorn cobs),
H_1: $\bar{x}_1 \neq \bar{x}_2$ (there is a difference in the lengths of the sweetcorn cobs).
b This is a two-tailed test.
c Use GDC to find $t = 0.535$ and $p = 0.600$.
d Since $0.600 > 0.10$, H_0 is accepted.

Chapter review

1 **a** The ranks are:

Height, cm	12	10.5	10.5	9	8	7	6	4.5	4.5	3	2	1
Time, s	2	1	3	4	5	8.5	6	7	8.5	10	11.5	11.5

Use GDC to find the PMCC for the ranked data: $r_s = -0.953$.
b It indicates strong and negative correlation between the height and the time it took to run, i.e. the taller the person, the faster they are.

2 **a** Instead of the quantitative data, the ranks of tennis players are given.
b The ranks are:

Tennis ranks	1	2	3	4	5	6	7	8
Aces	1	2	6	3	4.5	4.5	7	8

Use GDC to find the PMCC for the ranked data: $r_s = 0.850$.
c There is a strong and positive correlation between the tennis rank of the player and the number of aces they hit, i.e. the higher the rank, the more aces they are likely to hit.

3 **a** H_0: colours of the eggs laid are independent of the type of the hen. H_1: colours of the eggs laid are not independent of the type of hen.

b Probability that a hen chosen at random is Leghorn is $\frac{30}{90}$. Probability that an egg chosen at random is white is $\frac{42}{90}$. If the two events are independent, the expected number of white eggs laid by Leghorn hens is $90 \times \frac{30}{90} \times \frac{42}{90} = 14$.

c v = (rows − 1)(columns − 1) = $(2-1)\times(3-1) = 2$.

d Use GDC to find the values $\chi^2 = 21.7$ and $p = 1.94\times10^{-5}$.

e Since 21.7 > 5.991, H_0 is rejected.

4 **a** H_0: favourite colour of car is independent of gender; H_1: favourite colour of car is not independent of gender.

b Probability that a person chosen at random likes white cars is $\frac{20}{100}$. Probability that a person chosen at random is a male is $\frac{48}{100}$. If the two events are independent, the expected number of males who like white cars is $100\times\frac{20}{100}\times\frac{48}{100} = 9.6$.

c v = (rows − 1)(columns − 1) = $(2-1)\times(4-1) = 3$.

d Use GDC to find the values $\chi^2 = 9.43$ and $p = 0.0241$.

e Since 9.43 > 6.251, H_0 is rejected.

5 **a** The table of expected values:

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Expected frequencies	20	20	20	20	20	20	20

b $v = (n - 1) = 7 - 1 = 6$.

c H_0: The data satisfies a uniform distribution.

H_1: The data does not satisfy a uniform distribution.

Use GDC to find the values $\chi^2 = 10$and $p = 0.1247$, since 10 < 12.592, H_0 is accepted.

6 **a** The table of expected values:

Number on die	1	2	3	4	5	6
Expected frequencies	15	15	15	15	15	15

b $v = (n - 1) = 6 - 1 = 5$.

c H_0: The data satisfies a uniform distribution.

H_1: The data does not satisfy a uniform distribution.

Use GDC to find the values $\chi^2 = 0.67$ and $p = 0.98$, since 0.67 < 15.086, H_0 is accepted.

7 **a** Let $S \sim B(2, \frac{1}{2})$, then $P(S = 0) = 0.25$ and expected value is $0.25 \times 60 = 15$.

b Use binomial distribution defined above to find the expected frequencies:

Number of tails	0	1	2
Expected frequency	15	30	15

c All expected frequencies are higher than 5.

d $v = (n - 1) = 3 - 1 = 2$.

e H_0: The number of tails follows a binomial distribution.

H_1: The number of tails does not follow a binomial distribution.

Use GDC to find the values $\chi^2 = 1.2$ and $p = 0.549$.

f Since 1.2 < 5.991, H_0 is accepted.

8 Let $X \sim N(158, 4^2)$, and calculate the probabilities using GDC. Multiply them by 500(total number of girls).

a Table with expected frequencies:

Height, *x* cm	x < 152	152 ≤ x < 156	156 ≤ x < 160	160≤ x < 164	164 ≤ x
Expected frequency	33.40	120.87	191.46	120.87	33.4

b There are no expected values less than 5.

c $v = (n - 1) = 5 - 1 = 4$.

d H_0: The heights are normally distributed with mean of 158 cm and standard deviation of 4 cm.

H_1: The heights are not normally distributed with mean of 158 cm and standard deviation of 4 cm.

Use GDC to find the values $\chi^2 = 20.60$and $p = 0.00038$.

e Since 0.00038 < 0.10and 20.6 > 7.770, H_0 is rejected.

9 **a** H_0: $\bar{x}_1 = \bar{x}_2$ (there is no difference in the test scores)

H_1: $\bar{x}_1 \neq \bar{x}_2$ (there is a difference in the test scores).

b This is a two-tailed test.

c Use GDC to find $t = -0.421$ and $p = 0.678$.

d Since 0.678 < 0.05, H_0 is accpted.

10 **a** $H_0 : \mu_1 = \mu_2$ A1

$H_1 : \mu_1 < \mu_2$ A1

b One-tailed test A1

c $t-$ value $= -0.706$ M1A1

$p-$ value $= 0.244$ A1

d $0.244 > 0.05$ so accept H_0 M1R1

i.e. there is no significant difference in the results of the two schools.

11 **a** $H_0 : \mu_1 = \mu_2$ A1

$H_1 : \mu_1 > \mu_2$ A1

b $t-$ value $= 1.735$ M1A1

$p-$ value $= 0.0499$ A1

c $0.0499 < 0.05$ so reject H_0 . M1R1

i.e. there is significant evidence to suggest that older students read fewer books

12 **a** $H_0 : \mu_1 = \mu_2$ A1

$H_1 : \mu_1 \neq \mu_2$ A1

b $t-$ value $= 1.942$ M1A1

$p-$ value $= 0.0725$ A1

c $0.0725 < 0.1$ so reject H_0 . M1R1

i.e. there is significant evidence to suggest that there is a difference in average battery length

13 **a**

Age rank	5	4	8	1	6	7	2	3
Reaction rank	6	2.5	8	5	4	7	1	2.5

M1A1A1

b $r_s = 0.707$ M1A1

c r_s is positive and reasonably close to 1 , R1

indicating a fairly strong positive correlation between a person's age and their reaction time. R1

d Include a greater number of participants R1

Ensure the participants were spread equally throughout the age range R1

14 a

Age rank	1	9.5	3	12	11	2	7	9.5	6	8	5	4
Hours rank	6.5	3	12	1	9	11	5	2	4	8	10	6.5

M1A1A1

b $r_s = -0.596$ M1A1

c Neeve is incorrect. A1

A value of $r_s = -0.596$ indicates a small but significant negative correlation between a person's age and the hours per week they watch TV. R1

However, you cannot say this is *causal*. R1

i.e. You cannot conclude that your age *affects* the amount of TV you watch

15 A2 A1

B1 A1

C3 A1

D6 A1

E4 A1

F5 A1

16 a H_0 : Type of burger favoured is independent of age A1

H_1 : Type of burger favoured is not independent of age A1

b $(3-1)\times(4-1)=6$ A1

c $\chi^2_{calc} = 12.314$ M1A1

$p-$ value $= 0.0553$ A1

d $\chi^2_{calc} = 12.314 < 12.59$, therefore we accept H_0. A1R1

i.e. the type of burger favoured is independent of age.

17 a

	Smoker	Non-smoker	Totals
16-25 years	21.6	68.4	90
26-60 years	28.4	89.6	118
Totals	50	158	208

M1A1A1

b H_0 : Smoking is independent of age A1

H_1 : Smoking is not independent of age A1

$\chi^2_{calc} = 9.408$ M1A1

$p-$ value $= 0.00216$ A1

$\chi^2_{calc} = 9.408 > 6.64$, therefore we reject H_0 and accept H_1 A1R1

i.e. smoking is not independent of age

18 a H_0 : Movie preference is independent of gender A1

H_1 : Movie preference is not independent of gender A1

b $(5-1)\times(2-1)=4$ A1

c $\chi^2_{calc} = 11.111$ M1A1

$p-$ value $= 0.0253$ A1

d $\chi^2_{calc} = 11.111 > 9.49$, therefore we reject H_0 and accept H_1 A1R1

i.e. Movie preference is not independent of age.

19 a

Number of heads	0	1	2	3	4
Expected frequency	5	20	30	20	5

M1A1

b H_0 : The data satisfies a binomial distribution A1

H_1 : The data does not satisfy a binomial distribution A1

$\chi^2_{calc} = 7.583$ M1A1

$p-$ value $= 0.108$ A1

$\chi^2_{calc} = 7.583 < 9.49$, therefore we accept H_0 A1R1

i.e. The observed data fits a binomial distribution

20 a

Mark obtained	$0 \le x < 20$	$20 \le x < 40$	$40 \le x < 60$	$60 \le x < 80$	$80 \le x < 100$
Probability	0.004279	0.080233	0.365696	0.419444	0.121520
Expected frequency	2	22	97	111	32

M1A1A1

b Re-writing:

Mark obtained	$0 \le x < 40$	$40 \le x < 60$	$60 \le x < 80$	$80 \le x < 100$
Probability	0.0845	0.3657	0.4194	0.1215
Expected frequency	24	97	111	32

M1A1

Degrees of freedom $= 3$ A1

H_0 : The data satisfies a normal distribution A1

H_1 : The data does not satisfy a normal distribution A1

The critical value is $\chi^2_{(5\%)}(3) = 7.82$ A1

$\chi^2_{calc} = 10.47$ M1A1

$p-$value $= 0.015$ A1

$\chi^2_{calc} = 10.47 > 7.82$, therefore we reject H_0 and accept H_1 A1R1

i.e. The observed data does not fit a normal distribution with mean 62 and standard deviation 16.