Regression, correlation and hypothesis testing 1A

1 a As noted at the beginning of Section 1.1, the equation $Y=1.2+0.4X$ can be rewritten as $\log y=1.2+0.4\log x$, which is of the form $\log y=\log a+n\log x$ and so $y=ax^n$.

b $Y=1.2+0.4X$

$\Rightarrow \log y=1.2+0.4\log x$

$\Rightarrow y=10^{1.2+0.4\log x}=10^{1.2}\times 10^{0.4\log x}$

$\Rightarrow y=10^{1.2}\times 10^{\log x^{0.4}}=10^{1.2}\times x^{0.4}$

Therefore $a=10^{1.2}\approx 15.8$ (3 s.f.) and $n=0.4$

2 a As noted at the beginning of Section 1.1, the equation $Y=0.4+1.6X$ can be rewritten as $\log y=0.4+1.6x$, which is of the form $\log y=\log k+x\log b$ and so $y=kb^x$.

b $Y=0.4+1.6X$

$\Rightarrow \log y=0.4+1.6x$

$\Rightarrow y=10^{0.4+1.6x}=10^{0.4}\times 10^{1.6x}$

$\Rightarrow y=10^{0.4}\times\left(10^{1.6}\right)^x$

Therefore $k=10^{0.4}\approx 2.51$ (3 s.f.) and $b=10^{1.6}\approx 39.8$.

3 In the linear model $Y=mX+c$, where m and c are constants,
$Y=\log y$ and $X=\log x$, so $\log y=m\log x+c$

Therefore $c=\log a$

The point (0, 172) lies on the line, so $c=172$ and $\log a=172\Rightarrow a=10^{172}$

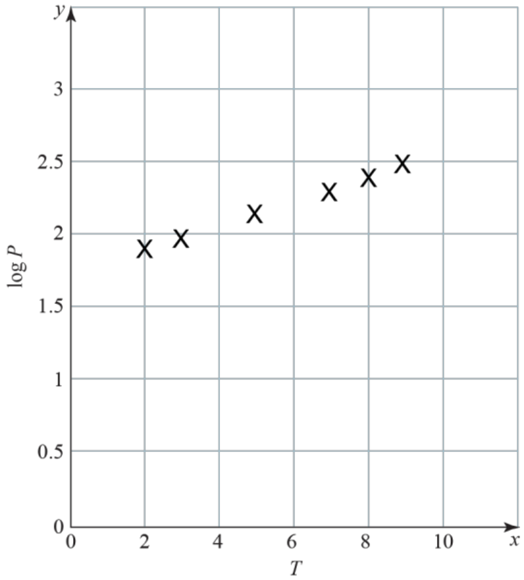
(23, 109) lies on $Y=mX+172$:
$109=23m+172$

$\Rightarrow 23m=109-172$

$\Rightarrow m=\dfrac{-63}{23}\approx -2.739$ (3 d.p.).

4 a

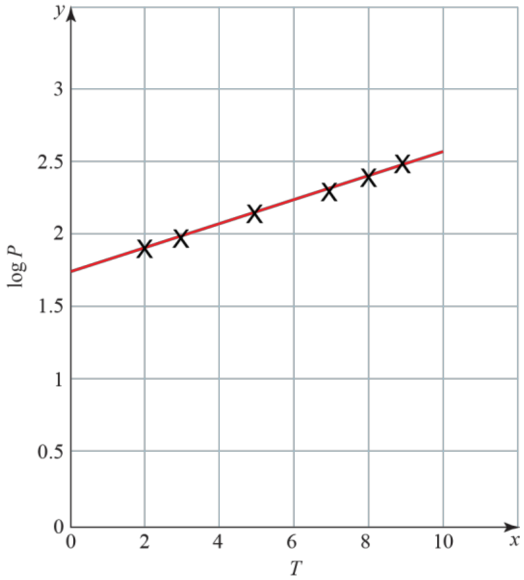
T	2	3	5	7	8	9
log P	1.86	1.93	2.10	2.25	2.33	2.41



b The points seem to lie on a straight line with a positive gradient, which suggests a strong positive correlation.

c Yes – the variables show a linear relationship when log P is plotted against T.

d



If log $P = mT + c$ then $P = 10^c(10^m)^T$. Measuring the gradient and intercept from the line of best fit with computer provides $c = 1.69927$ and $m = 0.07901$. These then give $a = 50.0345502089$ and $b = 1.19949930315$. Allow c between 1.65 and 1.8 so that a can be between 44.7 and 63.1. Since the gradient is small, it is better found using the original data points. Allow m between 0.077 and 0.082 so that b can be between 1.19 and 1.21.

4 e The approximate model is $P = 50.1 \times 1.2^T$ and so increasing T by 1 gives

$50.1 \times 1.2^{T+1} = (50.1 \times 1.2^T) \times 1.2$

which means increasing T by 1 corresponds to an increase of the population by 20%. Note that T is recorded in months, and so for every month that passes, the population of moles increases by 20%.

5 a The equation $t = a + bn$ is the equation of a straight line, but the data on the scatter diagram are not close to a straight line.

b $y = -0.301 + 0.6x$

$\Rightarrow \log t = -0.301 + 0.6 \log n$

$\Rightarrow t = 10^{-0.301 + 0.6 \log n} = 10^{-0.301} \times 10^{0.6 \log n}$

$\Rightarrow t = 10^{-0.301} \times 10^{\log n^{0.6}}$

$\Rightarrow t = 10^{-0.301} \times n^{0.6}$

Therefore $a = 10^{-0.301} \approx 0.5$ (3 s.f.) and $k = 0.6$.

6 $y = 1.31x - 0.41$

$\Rightarrow \log r = 1.31 \log c - 0.41$

$\Rightarrow r = 10^{1.31 \log c - 0.41} = 10^{1.31 \log c} \times 10^{-0.41}$

$\Rightarrow r = 10^{\log c^{1.31}} \times 10^{-0.41} = c^{1.31} \times 10^{-0.41}$

Therefore $r = 0.389 \times c^{1.31}$ (3 s.f.).

7 $y = 0.0023 + 1.8x$

$\Rightarrow \log m = 0.0023 + 1.8 \log h$

$\Rightarrow m = 10^{0.0023 + 1.8 \log h} = 10^{0.0023} \times 10^{1.8 \log h}$

$\Rightarrow m = 10^{0.0023} \times 10^{\log h^{1.8}} = 10^{0.0023} \times h^{1.8}$

Therefore $a = 10^{0.0023} \approx 1.0$ (3 s.f.) and $n = 1.8$.

8 a $y = 0.09 + 0.05x$

$\Rightarrow \log g = 0.09 + 0.05t$

$\Rightarrow g = 10^{0.09 + 0.05t} = 10^{0.09} \times 10^{0.05t}$

$\Rightarrow g = 10^{0.09} \times \left(10^{0.05}\right)^t$

Therefore $a = 10^{0.09} \approx 1.23$ and $b \approx 1.12$ (3 s.f.)

b If you increase the temperature by 1 °C, b is the increase in the growth rate g, i.e. b is the rate of change of g per degree.

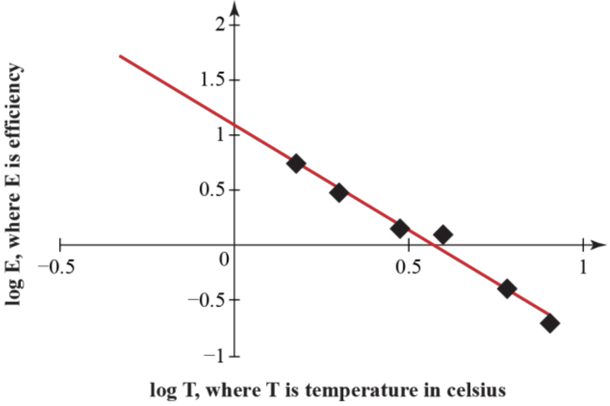
c 35 °C is outside of the range of data (extrapolation).

Challenge

a Construct a table of values for log$T$ and log$E$:

log *T*	0.0792	0.176	0.301	0.477	0.602	0.778	0.903
log *E*	0.954	0.74	0.477	0.146	0.0969	−0.398	−0.699

Plot the scatter diagram and draw a line of best fit:



The fact that the data are fitted by a straight line shows the validity of the relationship.

b The y-intercept of the line of best fit is 1.1 (to 2 s.f.).
So $\log a = 1.1$ (approximately)
$a = 10^{1.1} = 12.58925\ldots = 12.6$
From the graph, the gradient of the line of best fit is approximately −1.90 (to 3 s.f.), so $b = -1.90$.

c The model is of the form $\log E = \log a + b\log T$, but the expression $\log a + b\log T$ is not defined when $T = 0$ since $\log(0)$ is undefined (it approaches $-\infty$ as $T \to 0^+$).

Regression, correlation and hypothesis testing 1B

1 a $r = 0.9$ is a good approximation, since the points lie roughly, but not exactly, on a straight line. Remember that the value of r tells you how ‘close’ the data is to having a perfect positive or negative linear relationship.

b Clearly r is negative, and the data is not as close to being linear as in part **a**. $r = -0.7$ is therefore a good approximation.

c The data seems to have some negative correlation, but is rather ‘random’. Because so many points would lie far away from a line of best fit, $r = -0.3$ is a good approximation.

2 a The product moment correlation coefficient gives the type (positive or negative) and strength of linear correlation between v and m.

b By inputting the (ordered) data into your calculator, $r = 0.870$ (to 3 s.f.).

3 a $r = -0.854$ (to 3 s.f.)

b There is a negative correlation. The relatively older young people took less time to reach the required level.

4 a The completed table should read:

Time, *t*	1	2	4	5	7
Atoms, *n*	231	41	17	7	2
log *n*	2.36	1.61	1.23	0.845	0.301

b $r = -0.980$ (to 3 s.f.)

c There is an almost perfect negative correlation with the data in the form log n against t, which suggests an exponential decay curve. (This uses knowledge from the previous section.)

d $y = 2.487 - 0.320x$

$\Rightarrow \log n = 2.487 - 0.320t$

$\Rightarrow n = 10^{2.487-0.320t} = 10^{2.487} \times 10^{-0.320t}$

$\Rightarrow n = 10^{2.487} \times \left(10^{-0.320}\right)^{t}$

Therefore $a = 10^{2.487} = 307$ (3 s.f.) and $b = 10^{-0.320} = 0.479$ (3 s.f.).

5 a

Width, w	3	4	6	8	11
Mass, m	23	40	80	147	265
log w	0.4771	0.6021	0.7782	0.9031	1.041
log m	1.362	1.602	1.903	2.167	2.423

b $r = 0.9996$

c A graph of log w against log m is close to a straight line as the value of r is close to 1, therefore $m = kw^n$ is a good model for this data.

d $y = 0.464 + 1.88x$

$\Rightarrow \log m = 0.464 + 1.88 \log w$

$\Rightarrow m = 10^{(0.464+1.88\log w)}$

$\Rightarrow m = 10^{0.464} \times w^{1.88}$

Therefore $k = 10^{0.464} = 2.91$ (3 s.f.) and $n = 1.88$ (3 s.f.).

6 a $r = -0.833$ (3 s.f.)

b -0.833 is close to -1 so the data values show a strong to moderate negative correlation. A linear regression model is suitable for these data.

7 a 'tr' should be interpreted as a trace, which means a small amount.

b $r = -0.473$ (3 s.f.), treating 'tr' values as zero.

c The data show a weak negative correlation so a linear model may not be best; there may be other variables affecting the relationship or a different model might be a better fit.

Challenge

Take logs of the data in order to compute all of the required relationships:

x	3.1	5.6	7.1	8.6	9.4	10.7
y	3.2	4.8	5.7	6.5	6.9	7.6
log x	0.491	0.748	0.851	0.934	0.973	1.03
log y	0.505	0.681	0.756	0.813	0.839	0.881

Compute the PMCC for x and log y: $r = 0.985$ (3 s.f.).
Compute the PMCC for log x and log y: $r = 1.00$ (3 s.f.).
Therefore the data indicate that log x and log y have a strong positive linear relationship. From the previous section, the data indicate a relationship of the form $y = kx^n$.

Regression, correlation and hypothesis testing 1C

1 **a** H_0: $\rho = 0$, H_1: $\rho \neq 0$, critical value = ±0.3120. Reject H_0: there is reason to believe at the 5% level of significance that there is a correlation between the scores.

b H_0: $\rho = 0$, H_1: $\rho \neq 0$, critical value = ±0.3665. Accept H_0: there is no evidence of correlation between the two scores at the 2% level of significance.

2 **a** $r = -0.960$ (3 s.f.)

b H_0: $\rho = 0$, H_1: $\rho \neq 0$, critical value = ±0.8745. Reject H_0: there is reason to believe at the 1% level of significance that there is a correlation between the scores.

3 **a** The product moment correlation coefficient measures the type and strength of linear correlation between two variables.

b $r = 0.935$ (Get this value directly from your calculator.)

c

$$\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{array}\right\} \text{ 1-tail } \alpha = 0.05$$

test statistic $= 0.935$

critical values $= 0.4973$

t.s. > c.v., so reject H_0.

Conclude there is positive correlation between theoretical Biology and practical Biology marks – this implies that students who do well in theoretical Biology tests also tend to do well in practical Biology tests.

d There is a probability of 0.05 that the null hypothesis is true.

4 **a** $r = 0.68556...$

so $r = 0.686$ (3 s.f.)

(NB. In the exam get this directly from your calculator. If you set up a table of results you are likely to run out of time.)

b H_0: $\rho = 0$, H_1: $\rho > 0$, critical value = 0.6215. Reject H_0: there is reason to believe that there is a linear correlation between the English and Mathematics marks.

5 $r = 0.793$

(NB. In the exam get this directly from your calculator. If you set up a table of results you are likely to run out of time.)

$$\left.\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{array}\right\} \text{ 1-tail } \alpha = 0.01$$

test statistic $= 0.793$

critical values $= 0.8822$

t.s. < c.v. so accept H_0.

Conclude there is insufficient evidence at the 1% significance level to support the company's belief.

6 H_0: $\rho = 0$, H_1: $\rho < 0$, critical value = −0.4409. Accept H_0. There is evidence that the researcher is incorrect to believe that there is negative correlation between the amount of solvent and the rate of the reaction.

7 The safari ranger's test.
Type: 1-tailed
$H_0 : \rho = 0$
$H_1 : \rho > 0$
Sample size: 10
$r = 0.66$
He has sufficient evidence to reject H_0. The corresponding part of the table reads:

0.10	0.05	0.025	0.01	0.005	Sample size
0.4428	494	0.6319	0.715	0.7646	10

Therefore the least possible significance level for the ranger's test is 2.5%.

8 The information from the question is as follows:
Type: 1-tailed
$H_0 : \rho = 0$
$H_1 : \rho > 0$
Sample size: unknown
$r = 0.715$.
He has sufficient evidence to reject H_0. Part of the corresponding column of the table reads:

PMCC at 0.025 level of significance	Sample size
0.9500	4
0.8783	5
0.8114	6
0.7545	7
0.7067	8

Therefore the smallest possible sample size is 8.

9 a $r = -0.846$ (3 s.f.)

b H_0: $\rho = 0$, H_1: $\rho < 0$, critical value = −0.8822. Accept H_0. There is evidence that the employee is incorrect to believe that there is a negative correlation between humidity and visibility.

10 a This is a two-tailed test, so the scientist would need to halve the significance level.

b ± 0.4438

Regression, correlation and hypothesis testing Mixed exercise 1

1 a

Base area, x (cm^2)	1.1	1.3	1.9	2.2	2.5	3.7
Time, t (hours)	0.7	0.9	1.5	1.8	2.2	3.8
log x	0.0414	0.114	0.279	0.342	0.398	0.568
log t	−0.155	−0.0458	0.176	0.255	0.342	0.580

Calculating the PMCC for log x and log t: $r = 0.9998$.

b r is close to 1, so a graph of log t against log x shows a straight line, suggesting that the relationship is in the form $t = ax^n$.

c $\log t = -0.215 + 1.38 \log x$

$\Rightarrow t = 10^{-0.215+1.38\log x} = 10^{-0.215} \times 10^{1.38\log x}$

$\Rightarrow t = 10^{-0.215} \times 10^{\log x^{1.38}} = 10^{-0.215} \times x^{1.38}$

Therefore $a = 10^{-0.215} \approx 0.617$ (3 s.f.) and $n = 1.38$

2 a

Temperature, t (°C)	38	51	72	83	89	94
Dry residue, d (grams)	4.3	11.7	58.6	136.7	217.0	318.8
y = log d	0.633	1.07	1.77	2.14	2.34	2.50

$y = -0.635 + 0.0334x$

$\Rightarrow \log d = -0.635 + 0.0334t$

$\Rightarrow d = 10^{(-0.635+0.0334)^t} = 10^{0.635} \times 10^{0.0334t}$

$\Rightarrow d = 10^{-0.635} \times \left(10^{0.0334}\right)^t$

Therefore $a = 10^{-0.635} = 0.232$ (3 s.f.) and $b = 10^{0.0334} = 1.08$ (3 s.f.)

b 151 °C is outside the range of the data (extrapolation).

3 As a person's age increases their score on a memory test decreases.

4 a Each cow should be given 7 units. The yield levels off at this point. This can be seen even more clearly by drawing a scatter plot.

b $r = 0.952$ (3 s.f.)

c It would be less than 0.952. The yield of the last three cows is no greater than that of the seventh cow.

5 **a** $r = -0.972$ (to 3 s.f.)

b There is strong negative correlation. As c increases, f decreases.

6 **a** $r = 0.340$ (3 s.f.)

b $H_0: \rho = 0$
$H_1: \rho \neq 0$

Sample size = 10

Significance level in each tail = 0.025

From the table, critical values of r for a 2.5% significance level with a sample size of 10 are $r = \pm 0.6319$

So the critical region is $r < -0.6319$ and $r > 0.6319$

$0.340 < 0.6319$ so do not reject H_0.

There is not sufficient evidence, at the 5% level of significance, of correlation between age and salary. This means that an older person in this profession does not necessarily earn more than a younger person.

7 **a** $r = 0.937$ (3 s.f.)

b H_0: $\rho = 0$, H_1: $\rho \neq 0$, critical value = ±0.6319. Reject H_0. There is evidence that there is a correlation between the age of a machine and its maintenance costs.

8

$$\left.\begin{aligned} H_0 &: \rho = 0 \\ H_1 &: \rho < 0 \end{aligned}\right\} \text{ 1-tail } \alpha = 0.05$$

Test statistic $= -0.975$

$n = 9$, critical value $= -0.5822$

Lower tail test, t.s. < c.v. since $-0.975 < -0.5822$ reject H_0.

Conclude there is evidence of negative correlation. There is evidence that the greater the height above sea level, the lower the temperature at 7.00 a.m. is likely to be.

9

$$\left.\begin{aligned} H_0 &: \rho = 0 \\ H_1 &: \rho > 0 \end{aligned}\right\} \text{ 1-tail } \alpha = 0.05$$

Test statistic $= r = 0.972$

$n = 9$, critical value $= 0.5822$

Upper tail test, t.s. > c.v. since $0.972 > 0.5822$ so reject H_0.

Conclude there is evidence of a positive association between age and weight. This means the older a baby is, the heavier it is likely to be.

10 a $r = 0.940$ (to 3 s.f.)

b H_0: $\rho = 0$, H_1: $\rho > 0$, critical value 0.7293. Reject H_0. There is evidence that sunshine hours and ice cream sales are positively correlated.

11 $r = 0.843$ (3 s.f.), H_0: $\rho = 0$, H_1: $\rho > 0$, critical value 0.8054. Reject H_0. There is evidence that mean windspeed and daily maximum gust are positively correlated.

12 $r = -0.793$ (3 s.f.), H_0: $\rho = 0$, H_1: $\rho < 0$, critical value −0.7545. Reject H_0. There is evidence that temperature and pressure are negatively correlated.

Conditional probability 2A

1 a This is the set of anything not in set *B* but in set *A*. So the shaded region consists of the part of *A* which does not intersect with *B*, i.e. $A \cap B'$.

b The shaded region includes all of B and the region outside of A and B, i.e. $B \cup A'$.

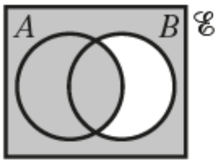
c There are two regions to describe. The first is the intersection of *A* and *B*, i.e. $A \cap B$ and the second is everything that is not in either *A* or *B*, i.e. $A' \cap B'$. Therefore the shaded region is $(A \cap B) \cup (A' \cap B')$.

d The shaded region is anything that is in *A* and *B* and *C*, i.e. $A \cap B \cap C$.

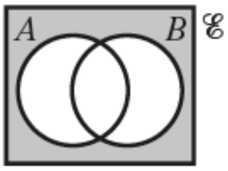
e The shaded region is anything that is either in *A* or *B* or *C*, i.e. $A \cup B \cup C$.

f The shaded region is anything that is either in *A* or *B* but is not in *C*. So the shaded region consists of the part of $A \cup B$ which does not intersect with *C*, i.e. $(A \cup B) \cap C'$.

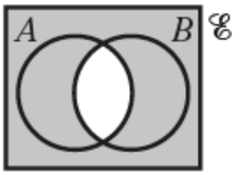
2 a Shade set *A*. The set B' consists of the region outside of *A* and *B* and the region inside *A* that does not intersect B. Therefore $A \cup B'$ is the region consisting of both these regions.



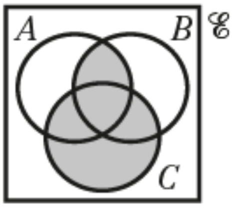
b Since this is an intersection, the region must satisfy both conditions. The first is to be in A'. This consists of two regions: one inside *B* and not in $A \cap B$; and one outside of *A* and *B*. The second condition is to be in B'. Again, this consists of two regions: one inside *A* and not in $A \cap B$; and one outside of *A* and *B*. Therefore $A' \cap B'$ is the region outside of *A* and *B* (since this region was in both A' and B'). One way to help picture this is to shade the regions A' and B' differently (either with different colours or using a different pattern for each). The intersection is then the region that includes both colours or patterns.



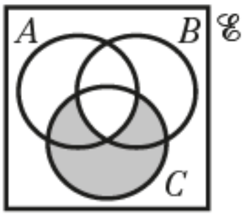
c In order to describe $(A \cap B)'$ it is sensible to first describe $A \cap B$. This is the single region which is included in both *A* and *B*. The complement is then everything *except* this region.



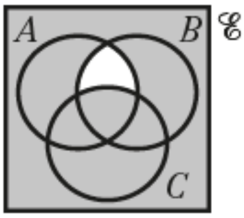
3 a The set $(A \cap B) \cup C$ is the union of the sets $A \cap B$ and C. On the blank diagram, the set $A \cap B$ consists of the two regions that are both contained within A and B. The remaining regions within set C can then be shaded in.



b First describe $A' \cup B'$. The set $A' \cup B'$ is everything apart from $A \cap B$. So the intersection of $A' \cup B'$ and C is everything in C apart from that part of C that intersects $A \cap B$.



3 c First describe $A \cap B \cap C'$. Brackets have not been included since for any sets X, Y and Z $(X \cap Y) \cap Z = X \cap (Y \cap Z)$. The intersection of $A \cap B$ and C' is the region within $A \cap B$ that does not intersect C. Therefore $(A \cap B \cap C')'$ is everything *except* this region.



4 a K is the event ‘the card chosen is a king’.

$$P(K) = \frac{4}{52} = \frac{1}{13}$$

b C is the event ‘the card chosen is a club’.

$$P(C) = \frac{1}{4}$$

c $C \cap K$ is the event ‘the card chosen is the king of clubs’.

$$P(C \cap K) = \frac{1}{52}$$

d $C \cup K$ is the event ‘the card chosen is a club or a king or both’.

$$P(C \cup K) = \frac{16}{52} = \frac{4}{13}$$

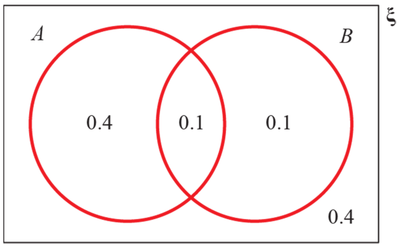
4 e C' is the event 'the card chosen is a not a club'.

$$\mathrm{P}(C') = \frac{3}{4}$$

f $K' \cap C$ is the event 'the card chosen is not a king and is a club'.

$$\mathrm{P}(K' \cap C) = \frac{12}{52} = \frac{3}{13}$$

5 Use the information in the question to draw a Venn diagram that will help in answering each part.



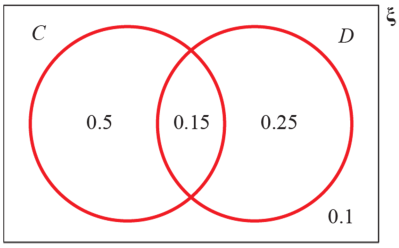
a $A \cup B$ is the region contained by sets A and B. So $\mathrm{P}(A \cup B) = 0.4 + 0.1 + 0.1 = 0.6$

b B' is the region that is not in set B. $\mathrm{P}(B') = 0.8$

c $A \cap B'$ is the region inside set A but outside set B. $\mathrm{P}(A \cap B') = 0.4$

d $A \cup B'$ is the region inside set A and the region outside set B, i.e. everything but the region inside set B that is not also in set A. $\mathrm{P}(A \cup B') = 0.4 + 0.1 + 0.4 = 0.9$

6 Use the information in the question to draw a Venn diagram that will help in answering each part.



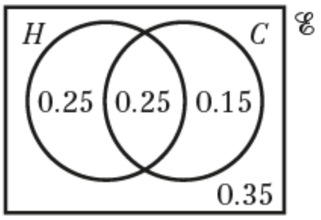
a $C' \cap D$ is the region inside set D but outside set C. $\mathrm{P}(C' \cap D) = 0.25$

b $C \cap D'$ is the region inside set C but outside set D. $\mathrm{P}(C \cap D') = 0.5$

c $\mathrm{P}(C) = 0.65$

d $C' \cup D'$ is the region outside set C and the region outside set D, i.e. everything but the region that is in both sets C and D. $\mathrm{P}(C' \cup D') = 0.85$

7 a

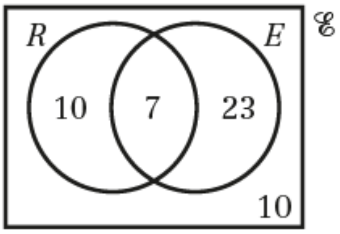


b i $\text{P}(H \cup C)$ means that either one of $H \cap C'$, $H \cap C$ or $H' \cap C$ occurs. Alternatively,
$\text{P}(H \cup C) = \text{P}(H) + \text{P}(C) - \text{P}(H \cap C) = 0.5 + 0.4 - 0.25 = 0.65$

ii $H' \cap C$ is the region inside set *C* but outside set *H*. $\text{P}(H' \cap C) = 0.15$

iii $H \cup C'$ is the region inside set *H* and the region outside set *C*, i.e. everything but the region inside set *C* that is not also in set *H*. $\text{P}(H \cup C') = 0.25 + 0.25 + 0.35 = 0.85$

8 a Only the possible outcomes of the two events need to considered, and so the Venn diagram should consist of two circles, one labelled 'R' for red and one labelled 'E' for even. They should intersect.



b i Note that $n(R \cup E) = n(R) + n(E) - n(R \cap E)$

$n(R \cap E) = n(R) + n(E) - n(R \cup E)$

$\Rightarrow n(R \cap E) = 17 + 30 - 40 = 7$

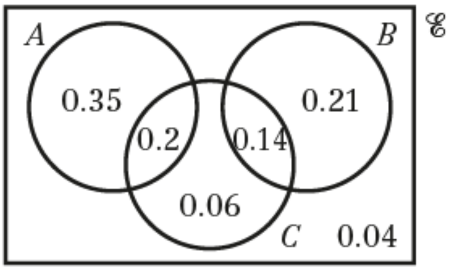
ii The region $R' \cap E'$ lies outside of both *R* and *E*.
Since there are 50 counters, $n(R' \cap E') = 50 - n(R \cup E) = 50 - 40 = 10$

So $\text{P}(R' \cap E') = \frac{10}{50} = \frac{1}{5} = 0.2$

iii From part **b i** $n(R \cap E) = 7$, so $n(R \cap E)' = 50 - 7 = 43$

So $\text{P}((R \cap E)') = \frac{43}{50} = 0.86$

9 a Since *A* and *B* are mutually exclusive, $\mathrm{P}(A \cap B) = 0$ and they need no intersection on the Venn diagram. From the question, $\mathrm{P}(A \cap C) = 0.2$ and so this can immediately be added to the diagram. Since *B* and *C* are independent, $\mathrm{P}(B \cap C) = \mathrm{P}(B) \times \mathrm{P}(C) = 0.35 \times 0.4 = 0.14$ and this can also be added to the diagram. The remaining region in *B* must be $\mathrm{P}(B) - \mathrm{P}(B \cap C) = 0.35 - 0.14 = 0.21$, the remaining region for *A* must be $\mathrm{P}(A) - \mathrm{P}(A \cap C) = 0.55 - 0.2 = 0.35$ and the remaining region for *C* must be $\mathrm{P}(C) - \mathrm{P}(A \cap C) - \mathrm{P}(B \cap C) = 0.4 - 0.2 - 0.14 = 0.06$. This means that the region outside of *A*, *B* and *C* must be 1 – 0.35 – 0.2 – 0.21 – 0.14 – 0.06 = 0.04.



b i The set $A' \cap B'$ must be outside of *A* and outside of *B*. These two regions are labelled 0.06 and 0.04. Therefore $\mathrm{P}(A' \cap B') = 0.06 + 0.04 = 0.1$

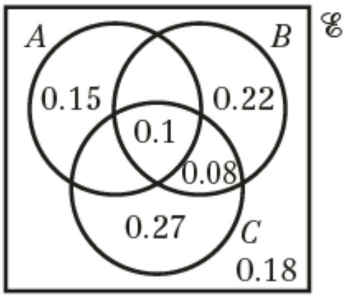
ii The region $B \cap C'$ is the region inside set *B* but outside set *C,* it is labelled 0.21 on the Venn diagram and is disjoint from *A*. Therefore $\mathrm{P}(A \cup (B \cap C')) = P(A) + 0.21 = 0.55 + 0.21 = 0.76$

iii Since $A \cap C$ consists of a single region, $(A \cap C)'$ consists of everything in the diagram except for that region. But B' includes the region $A \cap C$ and so $(A \cap C)' \cup B$ includes everything in the diagram, and so $\mathrm{P}((A \cap C)' \cup B') = 1$

10 a Start with a Venn diagram with all possible intersections. Then find the region $A \cap B \cap C$, which is at the centre of the diagram, and label it 0.1.

Now, since *A* and *B* are independent, $\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B) = 0.25 \times 0.4 = 0.1$, and as *B* and *C* are independent $\mathrm{P}(B \cap C) = \mathrm{P}(B) \times \mathrm{P}(C) = 0.4 \times 0.45 = 0.18$. Use these results to find values for the other intersections. $\mathrm{P}(A \cap B \cap C') = \mathrm{P}(A \cap B) - \mathrm{P}(A \cap B \cap C) = 0.1 - 0.1 = 0$;
$\mathrm{P}(B \cap C \cap A') = \mathrm{P}(B \cap C) - \mathrm{P}(A \cap B \cap C) = 0.18 - 0.1 = 0.08$; and $\mathrm{P}(A \cap C \cap B') = 0$ is given in the question.

Now find values for the remaining parts of the diagram. For example,
$\mathrm{P}(A \cap B' \cap C') = \mathrm{P}(A) - \mathrm{P}(A \cap B \cap C') - \mathrm{P}(A \cap C \cap B') - \mathrm{P}(A \cap B \cap C) = 0.25 - 0 - 0 - 0.1 = 0.15$
Similarly, $\mathrm{P}(B \cap A' \cap C') = 0.4 - 0.1 - 0.08 = 0.22$ and $\mathrm{P}(C \cap A' \cap B') = 0.45 - 0.1 - 0.08 = 0.27$
Finally calculate the region outside sets *A*, *B* and *C*,
$\mathrm{P}(A \cup B \cup C)' = 1 - 0.15 - 0.1 - 0.22 - 0.08 - 0.27 = 0.18$

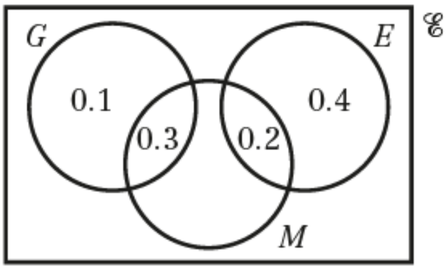


10 b i There are several ways to work out the regions that comprise the set $A' \cap (B' \cup C)$. One way is to determine, for each region, whether it lies in A' and $B' \cup C$. Alternatively, find the regions within A' (there are four) and then note that only one of these does not lie in $B' \cup C$. Summing the three remaining probabilities yields $\mathrm{P}(A' \cap (B' \cup C)) = 0.27 + 0.08 + 0.18 = 0.53$

ii The required region must be contained within C. Three of the four regions in C also lie in $A \cup B$, summing the probabilities yields $\mathrm{P}((A \cup B) \cap C) = 0 + 0.1 + 0.08 = 0.18$

c $\mathrm{P}(A') = 1 - P(A) = 0.75$, P($C$)= 0.45 and, from the Venn diagram, $\mathrm{P}(A' \cap C) = 0.08 + 0.27 = 0.35$. Since $\mathrm{P}(A') \times \mathrm{P}(C) = 0.75 \times 0.45 = 0.3375 \neq 0.35$, the events A' and C are not independent.

11 a Since $\mathrm{P}(G \cap E) = 0$, it follows that $\mathrm{P}(M \cap G \cap E) = 0$. So $\mathrm{P}(M \cap G \cap E') = \mathrm{P}(M \cap G) = 0.3$ and $\mathrm{P}(G \cap M') = \mathrm{P}(G) - \mathrm{P}(G \cap M) = 0.4 - 0.3 = 0.1$. This only accounts for 40% of the book club, 60% is unaccounted for, but P(E) = 0.6, so this 60% read epic fiction. So all the remaining members who read murder mysteries must also read epic fiction. Therefore $\mathrm{P}(M \cap E' \cap G') = 0$, $\mathrm{P}(M \cap E \cap G') = \mathrm{P}(M) - \mathrm{P}(M \cap G) = 0.5 - 0.3 = 0.2$, and $\mathrm{P}(E \cap M' \cap G') = 0.6 - 0.2 = 0.4$.



b i $\mathrm{P}(M \cup G) = \mathrm{P}(M \cup G \cup E) - \mathrm{P}(E \cap M' \cap G') = 1 - 0.4 = 0.6$

ii In this case $\mathrm{P}((M \cap G) \cup (M \cap E)) = \mathrm{P}((M \cap G \cap E') \cup (M \cap G' \cap E))$ and so the required probability is $\mathrm{P}(M \cap G \cap E') + \mathrm{P}(M \cap G' \cap E) = 0.3 + 0.2 = 0.5$

c $\mathrm{P}(G') = 0.6$, $\mathrm{P}(M) = 0.5$ and so $\mathrm{P}(G') \times \mathrm{P}(M) = 0.6 \times 0.5 = 0.3$. Since $\mathrm{P}(G' \cap M) = 0.2$, the events are not independent.

12 a Since A and B are independent, $\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B) = x \times y = xy$

b $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) = x + y - xy$

c $\mathrm{P}(A \cup B') = \mathrm{P}(A) + \mathrm{P}(A' \cap B')$ and since
$\mathrm{P}(A' \cap B') = 1 - \mathrm{P}(A \cup B) = 1 - (x + y - xy) = 1 - x - y + xy$ this means
$\mathrm{P}(A \cup B') = \mathrm{P}(A) + 1 - x - y + xy = x + 1 - x - y + xy = 1 - y + xy$

Challenge

a Use that the events are independent.

$$\begin{aligned}\text{P}(A\cap B\cap C) &= \text{P}((A\cap B)\cap C)\\ &= \text{P}(A\cap B)\times\text{P}(C)\\ &= \text{P}(A)\times\text{P}(B)\times\text{P}(C)\\ &= xyz\end{aligned}$$

b Using similar logic to the identity $\text{P}(A\cup B)=\text{P}(A)+\text{P}(B)-\text{P}(A\cap B)$, build up to the correct expression. First, x represents one circle and its intersections with the other two circles being shaded. Then $x + y - xy$ represents two circles and their intersections with the third being shaded. Finally $x + y - xy + z - xz - yz$ represents all three circles shaded except for where all three intersect. From part **a**, the final expression is therefore $x + y - xy + z - xz - yz + xyz$.

An alternative approach is to start by considering $A\cup B$
$\text{P}(A\cup B)=\text{P}(A)+\text{P}(B)-\text{P}(A\cap B)=x+y-xy$

Now find the union of $A\cup B$ and C
$\text{P}(A\cup B\cup C)=\text{P}(A\cup B)+\text{P}(C)-\text{P}((A\cup B)\cap C)=x+y+z-xy-\text{P}((A\cup B)\cap C)$ **(1)**

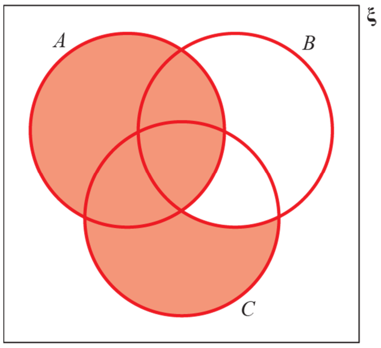
$(A\cup B)\cap C$ consists of the intersections of C with just A, with just B and with both A and B
So $(A\cup B)\cap C=(C\cap A\cap B')+(C\cap B\cap A')+(A\cap B\cap C)$

Consider the probabilities of each of these three regions in turn
$\text{P}(A\cap B\cap C)=xyz$ from part **a**
$\text{P}(C\cap A\cap B')=\text{P}(C\cap A)-\text{P}(A\cap B\cap C)=xz-xyz$
$\text{P}(C\cap B\cap A')=\text{P}(C\cap B)-\text{P}(A\cap B\cap C)=yz-xyz$
So $\text{P}(A\cup B)\cap C=xz-xyz+yz-xyz+xyz=xz+yz-xyz$ **(2)**

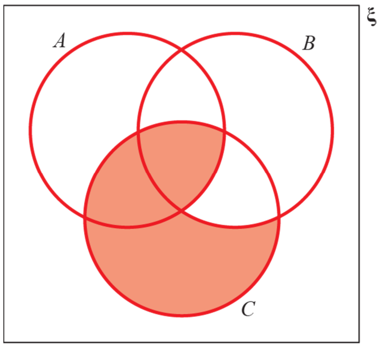
Now substitute the result for $\text{P}(A\cup B)\cap C$ from equation (**2**) into equation (**1**). This gives
$\text{P}(A\cup B\cup C)=x+y+z-xy-xz-yz+xyz$

Challenge

c First understand the region on a Venn diagram. The set $A \cup B'$ corresponds to the shaded regions:



Therefore the set $(A \cup B') \cap C$ corresponds to the shaded regions:



The unshaded part of C is the region $C \cap B \cap A'$

$\mathrm{P}(C \cap B \cap A') = P(C \cap B) - P(A \cap B \cap C) = yz - xyz$

So $\mathrm{P}((A \cup B') \cap C) = \mathrm{P}(C) - \mathrm{P}(C \cap B \cap A') = z - yz + xyz$

Conditional probability 2B

1 **a** There are 29 male students out of a total of 60 students.

$$\text{P(Male)} = \frac{29}{60}$$

b Restrict the sample space to the 29 male students; 18 of these prefer curry.

$$\text{P(Curry|Male)} = \frac{18}{29}$$

c Restrict the sample space to the 35 students that prefer curry; 18 of these are male.

$$\text{P(Male|Curry)} = \frac{18}{35}$$

d Restrict the sample space to the 31 female students; 14 of these prefer pizza.

$$\text{P(Pizza|Female)} = \frac{14}{31}$$

2 **a** By simple subtraction, there are 43 male members of the club (75 – 32 = 43). Of these 21 play badminton (43 – 22 = 21).

	Badminton	**Squash**	**Total**
Male	21	22	43
Female	15	17	32
Total	36	39	75

b **i** Restrict the sample space to the 39 members that play squash; 22 of these are male.

$$\text{P(Male|Squash)} = \frac{22}{39}$$

ii Restrict the sample space to the 36 members that play badminton; 15 of these are female.

$$\text{P(Female|Badminton)} = \frac{15}{36} = \frac{5}{12}$$

iii Restrict the sample space to the 32 members that are female; 17 of these play squash.

$$\text{P(Squash|Female)} = \frac{17}{32}$$

3 a There are 35 boys (80 – 45 = 35), of which 10 like chocolate (35 – 2 – 23 = 10).
Of the girls, 20 like strawberry (45 – 13 – 12 = 20).

	Girls	**Boys**	**Total**
Vanilla	13	2	15
Chocolate	12	10	22
Strawberry	20	23	43
Total	45	35	80

b i Restrict the sample space to the 43 children that like strawberry; 23 of these are boys.

$$\text{P(Boy|Strawberry)} = \frac{23}{43}$$

ii Restrict the sample space to the 15 children that like vanilla; 13 of these are girls.

$$\text{P(Girl|Vanilla)} = \frac{13}{15}$$

iii Restrict the sample space to the 35 boys; 10 of these like chocolate.

$$\text{P(Chocolate|Boy)} = \frac{10}{35} = \frac{2}{7}$$

4 a

Blue spinner

		1	**2**	**3**	**4**
Red spinner	**1**	2	3	4	5
	2	3	4	5	6
	3	4	5	6	7
	4	5	6	7	8

b i There 4 outcomes where $X = 5$, and 16 possible outcomes in total.

$$\text{P}(X = 5) = \frac{4}{16} = \frac{1}{4}$$

ii There are 4 equally likely outcomes where the red spinner is 2; and for one of these $X = 3$.

$$\text{P}(X = 3|\text{Red spinner is 2}) = \frac{1}{4}$$

iii There are 4 equally likely outcomes where $X = 5$, and for one of these the blue spinner is 3.

$$\text{P(Blue spinner is 3}|X = 5) = \frac{1}{4}$$

5 **a**

Dice 1

Dice 2

	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

b There are 6 outcomes where Dice 1 shows 5, and for one of these the product is 20.

$$\text{P(Product is 20|Dice 1 shows a 5)} = \frac{1}{6}$$

c There are 4 outcomes where the product is 2, and for one of these Dice 2 shows a 6.

$$\text{P(Dice 2 shows a 6|Product is 12)} = \frac{1}{4}$$

d All outcomes are equally likely.

6 $$\text{P(Ace|Diamond)} = \frac{\text{P(Ace of diamonds)}}{\text{P(Diamond)}} = \frac{\frac{1}{52}}{\frac{13}{52}} = \frac{1}{13}$$

7 Drawing a sample space diagram can be helpful in answering this question.

Coin 1

Coin 2

	H	**T**
H	HH	TH
T	HT	TT

a Note there are three outcomes where at least one coin lands on a head.

$$\text{P(HH|H)} = \frac{\text{P(Head and Head)}}{\text{P(Head)}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

7 **b** $$\text{P(Head and Tail|Head)} = \frac{\text{P(Head and Tail)}}{\text{P(Head)}} = \frac{\frac{2}{4}}{\frac{3}{4}} = \frac{2}{3}$$

c Assume that the coins are not biased.

8 **a** 64 students do not watch sport (120 – 56 = 64).
43 students do not watch drama (120 – 77 = 43).

Use the fact that of those who watch drama, 18 also watch sport to complete the table.
For example, this means that 38 students who watch sport do not watch drama (56 – 18 = 38), and 59 students who watch drama do not watch sport (77 – 18 = 59).

Given that 43 students do not watch drama, but 38 students who do not watch drama watch sport, this means 5 students do not watch drama or sport (43 – 38 = 5).

	Watches drama (D)	**Does not watch drama (D′)**	**Total**
Watches sport (S)	18	38	56
Does not watch sport (S′)	59	5	64
Total	77	43	120

b **i** The probability that the student does not watch drama.

$$\text{P}(D') = \frac{43}{120}$$

ii The probability that the student does not watch sport ot drama.

$$\text{P}(S' \cap D') = \frac{5}{120} = \frac{1}{24}$$

iii The probability that the student also watches sport if they watch drama.

$$\text{P}(S|D) = \frac{18}{77}$$

iv The probability that the student does not watch drama if they watch sport.

$$\text{P}(D'|S) = \frac{38}{56} = \frac{19}{28}$$

9 a

	Women	Men	Total
Stick	26	18	44
No stick	37	29	66
Total	63	47	110

b i $\text{P(Uses a stick)} = \frac{44}{110} = \frac{2}{5}$

ii Restrict the sample space to the 63 women; 26 of these use a stick.

$\text{P(Uses a stick|Female)} = \frac{26}{63}$

iii Restrict the sample space to those who use a stick; 18 of these are men.

$\text{P(Male|Uses a stick)} = \frac{18}{44} = \frac{9}{22}$

10 Build up a table to show the options as follows. First note that as there are 450 female owners, so there are 300 male owners (750 – 450 = 300). Consider those who own cats. 320 owners in total own a cat. Since no one owns more than one type of pet, this means that 430 owners do not own a cat (750 – 320 = 430).

175 female owners have a cat. Since there are 450 female owners, this means that 275 female owners do not own a cat (450 – 175 = 275). 145 male owners own a cat (320 – 175 = 145) and so 155 male owners do not own a cat (300 – 145 = 155). This gives this table:

	Owns a cat	Does not own a cat	Total
Female	175	275	450
Male	145	155	300
Total	320	430	750

Of the 430 owners who do not own a cat, 250 of them own a dog. Therefore 180 of the owners own another type of pet (430 – 250 = 180). Since 25 males own another type of pet, this means that 155 women own another type of pet (180 – 25 = 155).

10 Finally, of the 450 women, 175 own a cat and 155 own something other than a cat or a dog. Therefore 120 women own a dog (450 – 175 –155 = 120) and 130 men own a dog (300 – 145 – 25 = 130). This information is summarised in this table:

	Owns a cat	**Owns a dog**	**Owns another type of pet**	**Total**
Female	175	120	155	450
Male	145	130	25	300
Total	320	250	180	750

a The probability that the owner does not own a dog or a cat.

$$P(D' \cap C') = \frac{180}{750} = \frac{6}{25}$$

b The probability that a male owner (i.e. not female) owns a dog.

$$P(D|F') = \frac{130}{300} = \frac{13}{30}$$

c The probability that a cat owner is male (i.e. not female).

$$P(F'|C) = \frac{145}{320} = \frac{29}{64}$$

d The probability that a female owner does not own a dog or a cat.

$$P((D' \cap C')|F) = \frac{155}{450} = \frac{31}{90}$$

Conditional probability 2C

1 **a** The probability $A \cup B$ includes all cases where either event A or event B occurs. So sum the probabilities for these three regions $A \cap B'$, $A \cap B$ and $B \cap A'$.

This gives $\mathrm{P}(A \cup B) = 0.3 + 0.12 + 0.28 = 0.7$

b The probability that A occurs given that B occurs means that we are only selecting from those situations where B occurs. So the sample space is restricted to just circle B. The denominator of the fraction is 0.12 + 0.28 = 0.4. The numerator is when A also occurs i.e. when both A and B occur, which is the region $A \cap B$.

Therefore $\mathrm{P}(A|B) = \dfrac{0.12}{0.4} = 0.3$

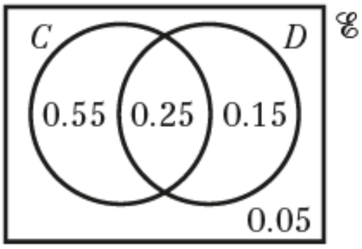
c The sample space is restricted to those instances where A has not occurred i.e. the regions $B \cap A'$ or $B' \cap A'$. This means the denominator will be 0.28 + 0.3 = 0.58. The numerator will consist of the cases where B has occurred i.e. $B \cap A'$.

Therefore $\mathrm{P}(B|A') = \dfrac{0.28}{0.58} = 0.483$ (3 s.f.)

d The sample space is restricted to those instances where A or B has occurred i.e. the region $A \cup B$. From part **a** this has probability 0.7. The numerator will consist of the cases where B has occurred i.e. either $B \cap A'$ or $B \cap A$.

Therefore $\mathrm{P}(B|A \cup B) = \dfrac{0.28 + 0.12}{0.7} = \dfrac{0.4}{0.7} = 0.571$ (3 s.f.)

2 **a** Fill in $\mathrm{P}(C \cap D) = 0.25$ on the Venn diagram, and then calculate $\mathrm{P}(C \cap D') = 0.8 - 0.25 = 0.55$, $\mathrm{P}(D \cap C') = 0.4 - 0.25 = 0.15$ and $\mathrm{P}(C \cup D)' = 1 - 0.25 - 0.55 - 0.15 = 0.05$



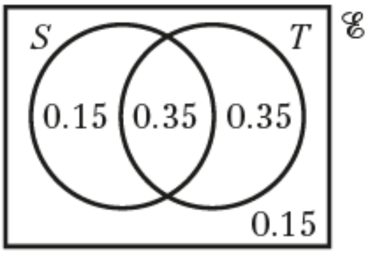
b **i** $\mathrm{P}(C \cup D) = 0.55 + 0.25 + 0.15 = 0.95$

ii $\mathrm{P}(C|D) = \dfrac{\mathrm{P}(C \cap D)}{\mathrm{P}(D)} = \dfrac{0.25}{0.4} = 0.625$

iii $\mathrm{P}(D|C) = \dfrac{\mathrm{P}(D \cap C)}{\mathrm{P}(C)} = \dfrac{0.25}{0.8} = 0.3125$

iv $\mathrm{P}(D'|C') = \dfrac{\mathrm{P}(D' \cap C')}{\mathrm{P}(C')} = \dfrac{0.05}{0.15 + 0.05} = 0.25$

3 **a** Since S and T are independent, $\text{P}(S \cap T) = \text{P}(S) \times \text{P}(T) = 0.5 \times 0.7 = 0.35$, and use this result to fill in the Venn diagram.



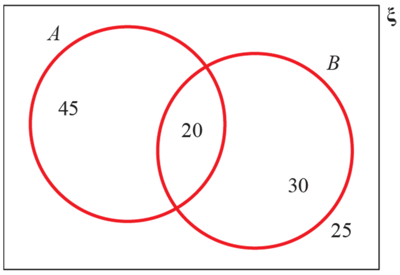
b **i** This is calculated to complete the Venn diagram in part **a**, $\text{P}(S \cap T) = 0.35$

ii $$\text{P}(S|T) = \frac{\text{P}(S \cap T)}{\text{P}(T)} = \frac{0.35}{0.7} = 0.5$$

iii $$\text{P}(T|S') = \frac{\text{P}(T \cap S')}{\text{P}(S')} = \frac{0.35}{0.5} = 0.7$$

iv $$\text{P}(S|(S' \cup T')) = \frac{\text{P}(S \cap (S' \cup T'))}{\text{P}(S' \cup T')} = \frac{\text{P}(S \cap T')}{0.15 + 0.35 + 0.15} = \frac{0.15}{0.65} = 0.231 \text{ (3 s.f.)}$$

4 **a** First produce a Venn diagram with the numbers of people in each region.



The Venn diagram can now be used to find the required probabilities. From the diagram,

$$\text{P}(A \cap B') = \frac{45}{120} = \frac{3}{8} = 0.375$$

b $$\text{P}(A|B) = \frac{\text{P}(A \cap B)}{\text{P}(B)} = \frac{\frac{20}{120}}{\frac{50}{120}} = \frac{20}{50} = \frac{2}{5} = 0.4$$

c $$\text{P}(B|A') = \frac{\text{P}(B \cap A')}{\text{P}(A')} = \frac{30}{55} = \frac{6}{11} = 0.545 \text{ (3 s.f.)}$$

d $$\text{P}(A|(A \cup B)) = \frac{\text{P}(A \cap (A \cup B))}{\text{P}(A \cup B)} = \frac{\text{P}(A)}{\text{P}(A \cup B)} = \frac{65}{95} = \frac{13}{19} = 0.684 \text{ (3 s.f.)}$$

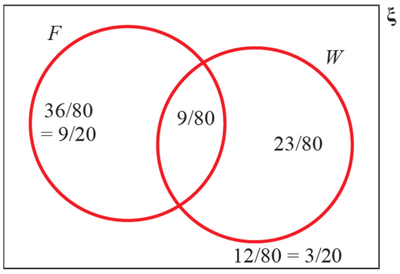
5 **a** Note that 12 cats like neither brand of food. So 80 – 12 = 68 cats like Feskers or Whilix or both. Use this and the other information in the question to calculate $\mathrm{P}(F \cap W)$ as follows:

$$\mathrm{P}(F \cup W) = \mathrm{P}(F) + \mathrm{P}(W) - \mathrm{P}(F \cap W)$$

$$\Rightarrow \mathrm{P}(F \cap W) = \mathrm{P}(F) + \mathrm{P}(W) - \mathrm{P}(F \cup W)$$

$$\text{So } \mathrm{P}(F \cap W) = \frac{45}{80} + \frac{32}{80} - \frac{68}{80} = \frac{9}{80}$$

This is a Venn diagram showing the result:



b $$\mathrm{P}(F|W) = \frac{\mathrm{P}(F \cap W)}{\mathrm{P}(W)} = \frac{\frac{9}{80}}{\frac{32}{80}} = \frac{9}{32} = 0.281 \text{ (3 s.f.)}$$

c $$\mathrm{P}(W|F) = \frac{\mathrm{P}(F \cap W)}{\mathrm{P}(F)} = \frac{\frac{9}{80}}{\frac{45}{80}} = \frac{9}{45} = \frac{1}{5} = 0.2$$

d $$\mathrm{P}(W'|F') = \frac{\mathrm{P}(F' \cap W')}{\mathrm{P}(F')} = \frac{\frac{12}{80}}{\frac{23+12}{80}} = \frac{12}{35} = 0.343 \text{ (3 s.f.)}$$

6 **a** $$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{0.2+0.1}{0.2+0.1+0.08+0.12} = \frac{0.3}{0.5} = 0.6$$

b $$\mathrm{P}(C|A') = \frac{\mathrm{P}(C \cap A')}{\mathrm{P}(A')} = \frac{0.1+0.08}{0.1+0.08+0.12+0.15} = \frac{0.18}{0.45} = 0.4$$

c $$\mathrm{P}((A \cap B)|C') = \frac{\mathrm{P}(A \cap B \cap C')}{\mathrm{P}(C')} = \frac{0.2}{0.2+0.2+0.12+0.15} = \frac{0.2}{0.67} = 0.299 \text{ (3 s.f.)}$$

d $$\mathrm{P}(C|(A' \cup B')) = \frac{\mathrm{P}(C \cap (A' \cup B'))}{\mathrm{P}(A' \cup B')} = \frac{0.05+0.08+0.1}{0.2+0.05+0.08+0.12+0.1+0.15} = \frac{0.23}{0.7} = 0.329 \text{ (3 s.f.)}$$

7 **a** The fact that the student must watch at least one of the TV programmes means that the student is selected from a region contained in $A \cup B \cup C$. Therefore this question should be interpreted as:

$$\mathrm{P}(C|(A \cup B \cup C)) = \frac{\mathrm{P}(C \cap (A \cup B \cup C))}{\mathrm{P}(A \cup B \cup C)} = \frac{\mathrm{P}(C)}{\left(\frac{23}{29}\right)} = \frac{\left(\frac{9}{29}\right)}{\left(\frac{23}{29}\right)} = \frac{9}{23} = 0.391 \text{ (3 s.f.)}$$

The other way to do this is to note that only 23 students watch at least one of the TV programmes, and of these 9 watch programme *C*.

b The standard method is as follows:

$$\mathrm{P}(\text{exactly two programmes}|A \cup B \cup C) = \frac{\mathrm{P}(A \cap B) + \mathrm{P}(A \cap C) + \mathrm{P}(B \cap C) - 3\mathrm{P}(A \cap B \cap C)}{\mathrm{P}(A \cup B \cup C)}$$

$$= \frac{\left(\frac{2}{29} + \frac{0}{29} + \frac{1}{29} - 3\frac{0}{29}\right)}{\left(\frac{23}{29}\right)} = \frac{3}{23} = 0.130 \text{ (3 s.f.)}$$

An alternative method is to note that 2 + 1 = 3 students watch exactly two of the programmes (they watch *A* and *B*, and *B* and *C*, respectively) and so 3 out of the 23 students that watch at least one of the TV programmes watch exactly two of the programmes.

c $$\mathrm{P}(B) = \frac{2+7+1}{29} = \frac{10}{29} = 0.345 \text{ (3 s.f.)}$$

$$\mathrm{P}(B|C) = \frac{\mathrm{P}(B \cap C)}{\mathrm{P}(C)} = \frac{\frac{1}{29}}{\frac{9}{29}} = \frac{1}{9} = 0.111 \text{ (3 s.f.)}$$

So $\mathrm{P}(B) \neq \mathrm{P}(B\,|\,C)$ and the events are not independent.

8 a $P(A \cap B) = 0$ since A and B are mutually exclusive.

$P(B \cap C) = P(B) \times P(C) = 0.6 \times 0.5 = 0.3$ since B and C are independent.

$P(B \cap C') = P(B) - P(B \cap C) = 0.6 - 0.3 = 0.3$

As $P(A \cap B) = 0$, $P(A \cup C) = 1 - P(B \cap C') - P(A' \cup B' \cup C') = 1 - 0.3 - 0.1 = 0.6$

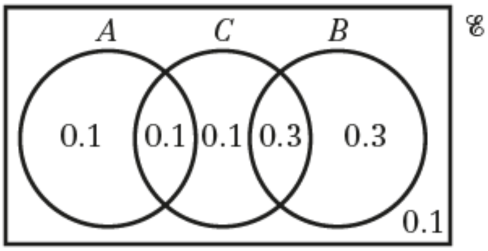
$P(A \cup C) = P(A) + P(C) - P(A \cap C)$

$\Rightarrow P(A \cap C) = P(A) + P(C) - P(A \cup C) = 0.2 + 0.5 - 0.6 = 0.1$

Now it is straightforward to work out remaining regions for the Venn diagram

$P(A \cap C') = 0.2 - P(A \cap C) = 0.2 - 0.1 = 0.1$

$P(C \cap A' \cap B') = 0.5 - P(A \cap C) - P(B \cap C) = 0.5 - 0.1 - 0.3 = 0.1$

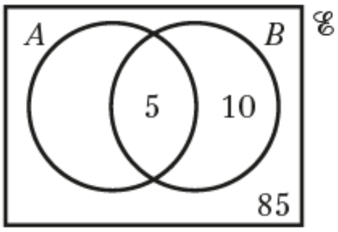


b i $$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{0.1}{0.5} = 0.2$$

ii $$P(B|C') = \frac{P(B \cap C')}{P(C')} = \frac{0.3}{0.1 + 0.3 + 0.1} = \frac{0.3}{0.5} = 0.6$$

iii $$P(C|(A \cup B)) = \frac{P(C \cap (A \cup B))}{P(A \cup B)} = \frac{0.1 + 0.3}{0.1 + 0.1 + 0.3 + 0.3} = \frac{0.4}{0.8} = 0.5$$

9 a All of the people who have the disease test positive, which means that there are no people in A who are not in $A \cap B$. There are also 10 people who test positive but do not have the disease. These people lie in B but do not lie in A, i.e. they lie in $B \cap A'$. There are 100 – 10 – 5 = 85 people who do not have the disease and do not test positive, so they lie in $A' \cap B'$. Therefore the Venn diagram should show:



b $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.05}{0.15} = \frac{1}{3} = 0.333 \text{ (3 s.f.)}$$

c The test would allow the doctor to find all of the people who have the disease, but only one third of those who tested positive would actually have the disease. This means that two thirds of the people who were told they had the disease would actually not have it.

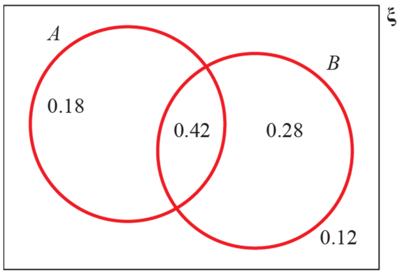
10 a Since $\mathrm{P}(A' \cap B') = 0.12$, this means that $\mathrm{P}(A \cup B) = 1 - 0.12 = 0.88$

Now find $\mathrm{P}(A \cap B)$

$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$

$\Rightarrow \mathrm{P}(A \cap B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cup B) = 0.6 + 0.7 - 0.88 = 0.42$

This allows a Venn diagram of the probabilties of the two events to be produced, which can be used to answer each part of the question.



$$\mathrm{P}(B|A') = \frac{\mathrm{P}(B \cap A')}{\mathrm{P}(A')} = \frac{0.28}{0.28 + 0.12} = \frac{0.28}{0.4} = 0.7$$

b $\mathrm{P}(B|A) = \dfrac{\mathrm{P}(B \cap A)}{\mathrm{P}(A)} = \dfrac{0.42}{0.6} = 0.7$

c Since $\mathrm{P}(B|A) = \mathrm{P}(B|A') = \mathrm{P}(B)$, the events A and B are independent.

11 $\mathrm{P}(A|B) = \mathrm{P}(B')$

$$\Rightarrow \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = P(B') = 0.2 + 0.1 = 0.3$$

$\Rightarrow \mathrm{P}(A \cap B) = 0.3\mathrm{P}(B)$

$\Rightarrow x = 0.3(x + y)$

The probabilities must sum to 1, so $0.2 + x + y + 0.1 = 1 \Rightarrow x + y = 0.7$

Substituting for $x + y$ gives

$x = 0.3(0.7) = 0.21$

$y = 0.7 - x = 0.7 - 0.21 = 0.49$

12 $\mathrm{P}(A|B) = \mathrm{P}(A')$

$$\Rightarrow \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \mathrm{P}(A')$$

$$\Rightarrow \frac{c}{c+d} = d + 0.2 \qquad \textbf{(1)}$$

The probabilities must sum to 1, so

$$0.3 + c + d + 0.2 = 1 \Rightarrow c + d = 0.5 \qquad \textbf{(2)}$$

Substituting for $c + d$ in the equation (**1**) gives

$$\frac{c}{c+d} = d + 0.2 \Rightarrow c = 0.5d + 0.1 \qquad \textbf{(3)}$$

Substituting this equation for c in equation (**2**) gives

$$0.5d + 0.1 + d = 0.5 \Rightarrow 1.5d = 0.4 \Rightarrow d = \frac{4}{15}$$

Finally, using equation (**3**) gives

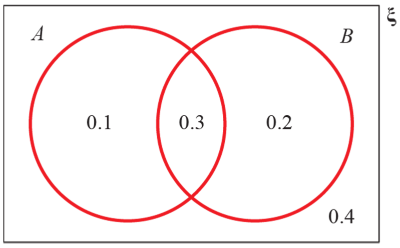
$$c = 0.5 \times \frac{4}{15} + 0.1 = \frac{4}{30} + \frac{1}{10} = \frac{4}{30} + \frac{3}{30} = \frac{7}{30}$$

Conditional probability 2D

1 a Rewrite the addition formula to obtain

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.4 + 0.5 - 0.6 = 0.3$$

Use this result to complete a Venn diagram to help answer the remaining parts of the question.



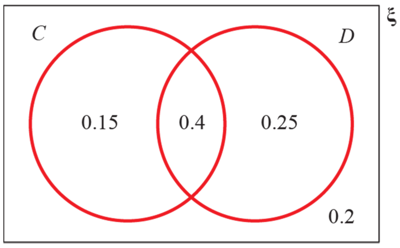
b $P(A') = 0.2 + 0.4 = 0.6$

c $P(A \cup B') = 0.4 + 0.4 = 0.8$

d $P(A' \cup B) = 0.5 + 0.4 = 0.9$

2 a $P(C \cup D) = P(C) + P(D) - P(C \cap D) = 0.55 + 0.65 - 0.4 = 0.8$

b



i The required region is the part ‘outside’ of *C* and *D*, which can be found since all of the probabilities must sum to 1.

$$P(C' \cap D') = 1 - P(C \cup D) = 1 - 0.8 = 0.2$$

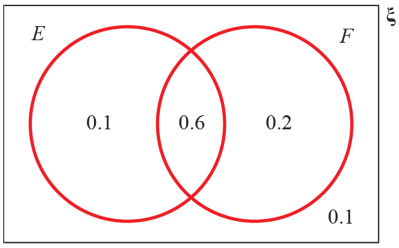
ii $P(C|D) = \dfrac{P(C \cap D)}{P(D)} = \dfrac{0.4}{0.65} = 0.615 \text{ (3 s.f.)}$

iii $P(C|D') = \dfrac{P(C \cap D')}{P(D')} = \dfrac{0.15}{0.35} = \dfrac{3}{7} = 0.429 \text{ (3 s.f.)}$

c From part **bii**, it is known that $P(C|D) \neq P(C)$ so the two events are not independent.
Alternatively, show that $P(C) \times P(D) \neq P(C \cap D)$.

3 **a** $P(E \cup F) = P(E) + P(F) - P(E \cap F) = 0.7 + 0.8 - 0.6 = 0.9$

b



i The required region is within E as well as everything outside F. It includes three of the four regions in the Venn diagram.
$P(E \cup F') = 0.1 + 0.6 + 0.1 = 0.8$

ii The required region is that part of F that does not intersect E.
$P(E \cap F') = 0.2$

iii $P(E|F') = \dfrac{P(E \cap F')}{P(F')} = \dfrac{0.1}{0.1 + 0.1} = \dfrac{1}{2} = 0.5$

4 **a**

$$\begin{aligned} P(T \cup Q) &= P(T) + P(Q) - P(T \cap Q) \\ 0.75 &= 3P(T \cap Q) + 3P(T \cap Q) - P(T \cap Q) \\ 5P(T \cap Q) &= 0.75 \\ P(T \cap Q) &= 0.15 \end{aligned}$$

b As $P(T) = P(Q)$, using $P(T \cup Q) = P(T) + P(Q) - P(T \cap Q)$ gives

$0.75 = 2P(T) - 0.15$

$\Rightarrow 2P(T) = 0.9$

$\Rightarrow P(T) = 0.45$

c $P(Q') = 1 - P(Q) = 1 - P(T) = 1 - 0.45 = 0.55$

d $P(T' \cap Q') = 1 - P(T \cup Q) = 1 - 0.75 = 0.25$

e $P(T \cap Q') = P(T) - P(T \cap Q) = 0.45 - 0.15 = 0.3$

5 Let F be the event has a freezer and D be the event has a dishwasher. The question requires finding $P(F \cap D)$. Use the addition formula
$P(F \cap D) = P(F) + P(D) - P(F \cup D) = 0.7 + 0.2 - 0.8 = 0.1$

6 **a** Use the multiplication formula for conditional probability to find $\text{P}(A \cap B)$

$$\text{P}(A \cap B) = \text{P}(A \mid B) \times \text{P}(B) = 0.4 \times 0.5 = 0.2$$

Now use the multiplication formula again to find $\text{P}(B|A)$

$$\text{P}(B|A) = \frac{\text{P}(B \cap A)}{\text{P}(A)} = \frac{0.2}{0.4} = \frac{1}{2} = 0.5$$

b Use the addition formula to find $\text{P}(A \cup B)$

$$\text{P}(A \cup B) = \text{P}(A) + \text{P}(B) - \text{P}(A \cap B) = 0.4 + 0.5 - 0.2 = 0.7$$

Now $\text{P}(A' \cap B')$ can be found as it is the region outside $\text{P}(A \cup B)$

$$\text{P}(A' \cap B') = 1 - \text{P}(A \cup B) = 1 - 0.7 = 0.3$$

c $\text{P}(A' \cap B) = \text{P}(B) - \text{P}(A \cup B) = 0.5 - 0.2 = 0.3$

7 **a** First use the addition formula to find $\text{P}(A \cap B)$

$$\text{P}(A \cap B) = \text{P}(A) + \text{P}(B) - \text{P}(A \cup B) = \frac{1}{4} + \frac{1}{2} - \frac{3}{5} = \frac{3}{20}$$

Now use the multiplication formula to

$$\text{P}(A|B) = \frac{\text{P}(A \cap B)}{\text{P}(B)} = \frac{\frac{3}{20}}{\frac{1}{2}} = \frac{3}{10} = 0.3$$

b $\text{P}(A' \cap B) = \text{P}(B) - \text{P}(A \cap B) = \frac{1}{2} - \frac{3}{20} = \frac{7}{20} = 0.35$

c $\text{P}(A' \cap B') = 1 - P(A \cup B) = 1 - \frac{3}{5} = \frac{2}{5} = 0.4$

8 **a** $\text{P}(C \cap D) = \text{P}(C|\text{D}) \times \text{P}(D) = \frac{1}{3} \times \frac{1}{4} = \frac{1}{12} = 0.0833$ (3 s.f.)

b $\text{P}(C \cap D') = \text{P}(C|\text{D}') \times \text{P}(D') = \frac{1}{5} \times \left(1 - \frac{1}{4}\right) = \frac{1}{5} \times \frac{3}{4} = \frac{3}{20} = 0.15$

c $\text{P}(C) = \text{P}(C \cap D') + \text{P}(C \cap D) = \frac{3}{20} + \frac{1}{12} = \frac{9}{60} + \frac{5}{60} = \frac{14}{60} = \frac{7}{30} = 0.233$ (3 s.f.)

d $\text{P}(D|C) = \frac{\text{P}(D \cap C)}{\text{P}(C)} = \frac{\frac{1}{12}}{\frac{7}{30}} = \frac{30}{84} = \frac{5}{14} = 0.357$ (3 s.f.)

e $\text{P}(D'|C) = 1 - \text{P}(D|C) = 1 - \frac{5}{14} = \frac{9}{14} = 0.643$ (3 s.f.)

8 f $\mathrm{P}(D'|C') = \dfrac{\mathrm{P}(C' \cap D')}{\mathrm{P}(C')} = \dfrac{1-\mathrm{P}(C \cup D)}{\mathrm{P}(C')}$

However, $\mathrm{P}(C') = 1 - \mathrm{P}(C) = \dfrac{7}{30} = \dfrac{23}{30}$

And $\mathrm{P}(C \cup D) = \mathrm{P}(C) + \mathrm{P}(D) - \mathrm{P}(C \cap D) = \dfrac{7}{30} + \dfrac{1}{4} - \dfrac{1}{12} = \dfrac{24}{60} = \dfrac{2}{5}$

So $\mathrm{P}(D'|C') = \dfrac{\mathrm{P}(C' \cap D')}{\mathrm{P}(C')} = \dfrac{1-\mathrm{P}(C \cup D)}{\mathrm{P}(C')} = \dfrac{1-\frac{2}{5}}{\frac{23}{30}} = \dfrac{3}{5} \times \dfrac{30}{23} = \dfrac{18}{23} = 0.783 \text{ (3 s.f.)}$

9 a $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) = 0.42 + 0.37 - 0.12 = 0.67$

b $\mathrm{P}(A|B') = \dfrac{\mathrm{P}(A \cap B')}{\mathrm{P}(B')} = \dfrac{\mathrm{P}(A) - \mathrm{P}(A \cap B)}{1 - \mathrm{P}(B)} = \dfrac{0.42 - 0.12}{1 - 0.37} = \dfrac{0.3}{0.63} = 0.476 \text{ (3 s.f.)}$

c Since the events A and C are independent, $\mathrm{P}(A \cap C) = \mathrm{P}(A) \times \mathrm{P}(C) = 0.42 \times 0.3 = 0.126$

d Since B and C are mutually exclusive, there is no need to have an intersection between B and C on the diagram. Work out the probabilities associated with each region as follows:

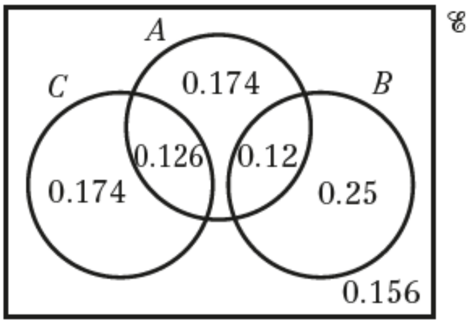
$\mathrm{P}(C \cap A') = \mathrm{P}(C) - \mathrm{P}(A \cap C) = 0.3 - 0.126 = 0.174$

$\mathrm{P}(B \cap A') = \mathrm{P}(B) - \mathrm{P}(A \cap B) = 0.37 - 0.12 = 0.25$

$\mathrm{P}(A \cap B' \cap C') = \mathrm{P}(A) - \mathrm{P}(A \cap B) - \mathrm{P}(A \cap C) = 0.42 - 0.12 - 0.126 = 0.174$

$\mathrm{P}(A \cup B \cup C) = 0.174 + 0.126 + 0.174 + 0.12 + 0.25 = 0.844$

$\mathrm{P}(A' \cap B' \cap C') = 1 - \mathrm{P}(A \cup B \cup C) = 1 - 0.844 = 0.156$



e $\mathrm{P}((A' \cup C)') = 1 - \mathrm{P}(A' \cup C)$

Use the Venn diagram to find $\mathrm{P}(A' \cup C) = 0.174 + 0.126 + 0.25 + 0.156 = 0.706$

So $\mathrm{P}((A' \cup C)') = 1 - 0.706 = 0.294$

10 a B and C are independent: $\text{P}(B \cap C) = \text{P}(B) \times \text{P}(C) = 0.7 \times 0.4 = 0.28$

b Using part **a**, $\text{P}(B|C) = \dfrac{\text{P}(B \cap C)}{\text{P}(C)} = \dfrac{0.28}{0.4} = \dfrac{7}{10} = 0.7$

c $\text{P}(A|B') = \dfrac{\text{P}(A \cap B')}{\text{P}(B')} = \dfrac{\text{P}(A) - \text{P}(A \cap B)}{1 - \text{P}(B)} = \dfrac{0.4 - 0.3}{1 - 0.7} = \dfrac{0.1}{0.3} = 0.333 \text{ (3 s.f.)}$

d $\text{P}((B \cap C)|A') = \dfrac{\text{P}((B \cap C) \cap A')}{\text{P}(A')} = \dfrac{\text{P}(B \cap C) - P(A \cap B \cap C)}{1 - \text{P}(A)}$

As A and C are mutually exclusive, $\text{P}(A \cap B \cap C) = 0$

So $\text{P}((B \cap C)|A') = \dfrac{\text{P}(B \cap C)}{1 - \text{P}(A)} = \dfrac{0.28}{1 - 0.4} = \dfrac{0.28}{0.6} = 0.467 \text{ (3 s.f.)}$

11 a This requires finding $\text{P}(A \cap B)$
First find $\text{P}(A \cup B)$
$\text{P}(A \cup B) = 0.9$ as $\text{P}(A \cup B) + \text{P}(A' \cap B') = 1$
Using the addition rule gives
$\text{P}(A \cap B) = \text{P}(A) + \text{P}(B) - \text{P}(A \cup B) = 0.3 + 0.7 - 0.9 = 0.1$

b This requires finding $\text{P}(A|B)$
$\text{P}(A|B) = \dfrac{\text{P}(A \cap B)}{\text{P}(B)} = \dfrac{0.1}{0.7} = 0.143 \text{ (3 s.f.)}$

c Test whether the events are independent
$\text{P}(A) \times \text{P}(B) = 0.3 \times 0.7 = 0.21,\ \text{P}(A \cap B) = 0.1$
So the events are not independent. If Anna is late, Bella is *less* likely to be late and vice versa.

12 a The probability that both John and Kayleigh win their matches is $\text{P}(J \cap K)$
$\text{P}(J \cap K) = \text{P}(J) + \text{P}(K) - \text{P}(J \cup K) = 0.6 + 0.7 - 0.8 = 0.5$

b $\text{P}(J|K') = \dfrac{\text{P}(J \cap K')}{\text{P}(K')} = \dfrac{\text{P}(J) - \text{P}(J \cap K)}{1 - \text{P}(K)} = \dfrac{0.6 - 0.5}{1 - 0.7} = \dfrac{0.1}{0.3} = 0.333 \text{ (3 s.f.)}$

c $\text{P}(K|J) = \dfrac{\text{P}(J \cap K)}{\text{P}(J)} = \dfrac{0.5}{0.6} = 0.833 \text{ (3 s.f.)}$

d $\text{P}(K|J) = 0.833 \text{ (3 s.f.)},\ \text{P}(K) = 0.7,$ so $\text{P}(K|J) \neq \text{P}(K)\text{P}(K|J) = 0.833\ldots \neq \text{P}(K) = 0.7$. So J and K are not independent.

Challenge

a The probability function must sum to 1. Therefore $k+2k+3k+4k+5k=1 \Rightarrow 15k=1 \Rightarrow k=\frac{1}{15}$

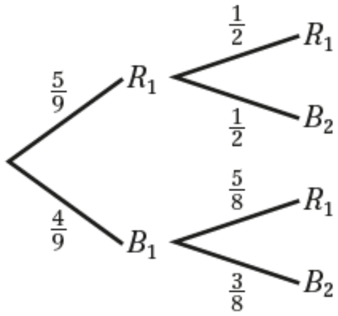
b $\mathrm{P}(X=5|X>2)=\frac{\mathrm{P}(X=5)}{\mathrm{P}(X>2)}=\frac{\frac{5}{15}}{1-\mathrm{P}(X=1 \cup X=2)}=\frac{\frac{5}{15}}{1-\frac{3}{15}}=\frac{\frac{5}{15}}{\frac{12}{15}}=\frac{5}{12}$

c

$$\mathrm{P}(X \text{ is odd}|X \text{ is prime})=\frac{\mathrm{P}(X \text{ is odd and prime})}{\mathrm{P}(X \text{ is prime})}=\frac{\mathrm{P}(X=3 \cup X=5)}{\mathrm{P}(X=2 \cup X=3 \cup X=5)}=\frac{\frac{3+5}{15}}{\frac{2+3+5}{15}}=\frac{8}{10}=\frac{4}{5}$$

Conditional probability 2E

1 **a** When the first token removed is red, there are 8 tokens remaining in the bag, 4 red and 4 blue. When the first token removed is blue, there are 8 tokens remaining in the bag, 5 red and 3 blue.



b The answer can be read off from the tree diagram, following the lower branch (first blue) and then the red branch.

So $\text{P(second red}|\text{first blue)} = \frac{5}{8}$

c $$\text{P(first red}|\text{second blue)} = \frac{\text{P(first red and second blue)}}{\text{P(second blue)}} = \frac{\frac{5}{9}\times\frac{1}{2}}{\left(\frac{5}{9}\times\frac{1}{2}\right)+\left(\frac{4}{9}\times\frac{3}{8}\right)} = \frac{\frac{5}{18}}{\frac{32}{72}} = \frac{20}{32} = \frac{5}{8}$$

d $$\text{P(first blue}|\text{tokens different colours)} = \frac{\text{P(first blue and second red)}}{\text{P(first blue and second red)} + \text{P(first red and second blue)}}$$

$$= \frac{\frac{4}{9}\times\frac{5}{8}}{\left(\frac{4}{9}\times\frac{5}{8}\right)+\left(\frac{5}{9}\times\frac{1}{2}\right)} = \frac{\frac{20}{72}}{\frac{40}{72}} = \frac{20}{40} = \frac{1}{2}$$

e $$\text{P(tokens same colour}|\text{second token red)} = \frac{\text{P(first red and second red)}}{\text{P(first red and second red)} + \text{P(first blue and second red)}}$$

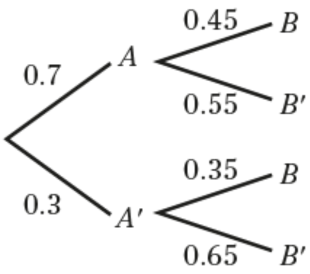
$$= \frac{\frac{5}{9}\times\frac{1}{2}}{\left(\frac{5}{9}\times\frac{1}{2}\right)+\left(\frac{4}{9}\times\frac{5}{8}\right)} = \frac{\frac{5}{18}}{\frac{40}{72}} = \frac{\frac{20}{72}}{\frac{40}{72}} = \frac{20}{40} = \frac{1}{2}$$

2 **a** $\mathrm{P}(A) = 0.7 \Rightarrow \mathrm{P}(A') = 1 - 0.7 = 0.3$

$\mathrm{P}(B|A) = 0.45 \Rightarrow \mathrm{P}(B'|A) = 1 - 0.45 = 0.55$

$\mathrm{P}(B|A') = 0.35 \Rightarrow \mathrm{P}(B'|A') = 1 - 0.35 = 0.65$

Therefore the completed tree diagram should be:



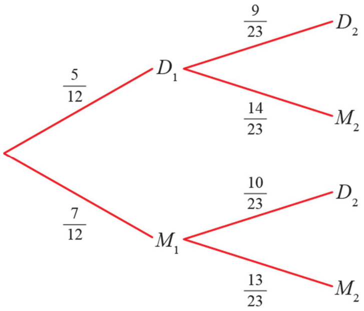
b **i** $\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B|A) = 0.7 \times 0.45 = 0.315$

ii $\mathrm{P}(A' \cap B') = \mathrm{P}(A') \times \mathrm{P}(B'|A') = 0.3 \times 0.65 = 0.195$

iii $\mathrm{P}(A|B) = \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \dfrac{0.315}{\mathrm{P}(A \cap B) + \mathrm{P}(A' \cap B)} = \dfrac{0.315}{0.315 + 0.3 \times 0.35} = \dfrac{0.315}{0.42} = 0.75$

3 **a** There are 10 dark chocolates in a box of 24, meaning the probability of choosing a dark chocolate is $\frac{10}{24} = \frac{5}{12}$. Similarly there are 14 milk chocolates out of the 24, and so the probability of choosing a dark chocolate is $\frac{14}{24} = \frac{7}{12}$.

Once Linda has eaten one chocolate, there are 23 chocolates left in the box. If the first chocolate she ate was a dark one, the probability of choosing another dark chocolate is $\frac{9}{23}$, and the probability of choosing a milk chocolate is $\frac{14}{23}$. If the first chocolate she ate was a milk one, the probability of a dark chocolate is $\frac{10}{23}$, and the probability of choosing another milk chocolate is $\frac{13}{23}$.



b $\mathrm{P}(\text{dark and dark}) = \dfrac{5}{12} \times \dfrac{9}{23} = \dfrac{15}{92} = 0.163$ (3 s.f.)

3 **c** $\mathrm{P}(\text{one dark and one milk}) = \mathrm{P}(D_1 \cap M_2) + \mathrm{P}(M_1 \cap D_2)$

$$= \frac{5}{12} \times \frac{14}{23} + \frac{7}{12} \times \frac{10}{23}$$

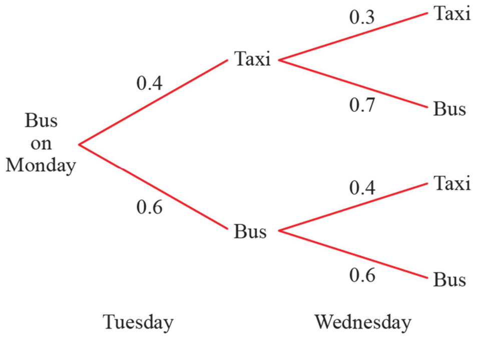
$$= \frac{70}{276} + \frac{70}{276} = \frac{140}{276} = \frac{35}{69} = 0.507 \text{ (3 s.f.)}$$

d $\text{P(dark and dark|at least one dark)} = \frac{\text{P(dark and dark)}}{\text{P(at least one dark)}}$

$$= \frac{\text{P}(D_1 \cap D_2)}{\text{P}(D_1 \cap D_2) + \text{P}(D_1 \cap M_2) + \text{P}(M_1 \cap D_2)}$$

$$= \frac{\frac{5}{12} \times \frac{9}{23}}{\frac{5}{12} \times \frac{9}{23} + \frac{5}{12} \times \frac{14}{23} + \frac{7}{12} \times \frac{10}{23}} = \frac{\frac{45}{276}}{\frac{185}{276}} = \frac{45}{185} = \frac{9}{37} = 0.243 \text{ (3 s.f.)}$$

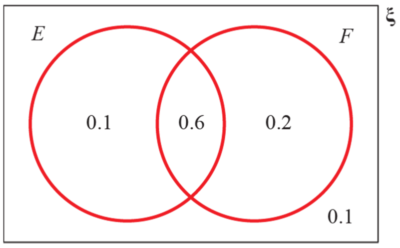
4 Use the information in the question to produce a tree diagram covering Jean's possible travel arrangements on Tuesday and Wednesday as follows:



Now sum the probabilities of Jean taking a taxi to work on Wednesday

$$\begin{aligned}\text{P(taxi on Wednesday)} &= 0.4 \times 0.3 + 0.6 \times 0.4 \\ &= 0.12 + 0.24 \\ &= 0.36\end{aligned}$$

5 Represent the information as a tree diagram. The coins are chosen at random, so there is a probability of $\frac{1}{2}$ of choosing each coin.

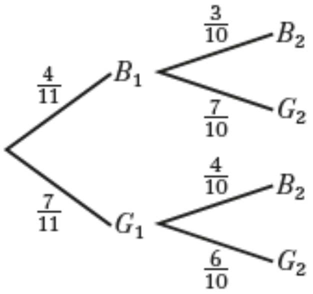


a $\text{P(head)} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} = 0.25$

5 b $\text{P(Fair|tail)} = \frac{\text{P(Fair and Tail)}}{\text{P(Tail)}}$

$$= \frac{\frac{1}{2} \times \frac{1}{2}}{\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 1} = \frac{1}{3} = 0.333 \text{ (3 s.f.)}$$

6 a Since the first ball selected is not replaced, there are 10 balls in the bag when the second ball is selected.



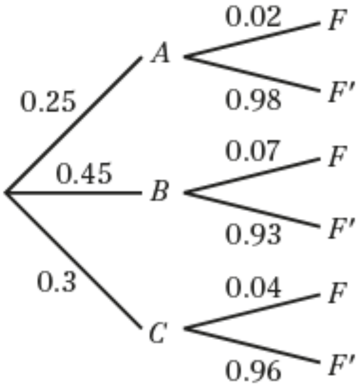
b $\text{P(second ball is green)} = \text{P}(B_1 \cap G_2) + \text{P}(G_1 \cap G_2)$

$$= \left(\frac{4}{11} \times \frac{7}{10}\right) + \left(\frac{7}{11} \times \frac{6}{10}\right) = \frac{28+42}{110} = \frac{7}{11} = 0.636 \text{ (3 s.f.)}$$

c $\text{P(both balls are green|second ball is green)} = \frac{\text{P(both balls are green)}}{\text{P(second ball is green)}}$

$$= \frac{\frac{7}{11} \times \frac{6}{10}}{\frac{70}{110}} = \frac{\frac{42}{110}}{\frac{70}{110}} = \frac{42}{70} = \frac{3}{5} = 0.6$$

7 a The probability of the sheet coming from *A, B* or *C* is given in the question. In each case, the probability that a sheet is flawed immediately provides the probability that it is not flawed since the two probabilities must sum to 1. Therefore the completed tree diagram is:



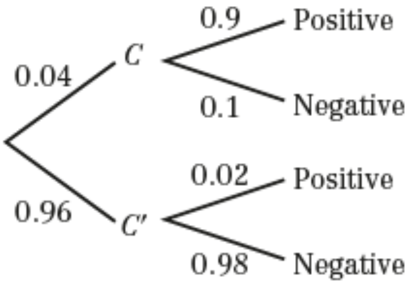
b i $\text{P(produced by } B \cap \text{flawed)} = 0.45 \times 0.07 = 0.0315$

7 b ii P(flawed) can be found by summing $\text{P(produced by } A \cap \text{flawed)}$, $\text{P(produced by } B \cap \text{flawed)}$ and $\text{P(produced by } C \cap \text{flawed)}$. Therefore

$$\text{P(flawed)} = 0.25 \times 0.02 + 0.0315 + 0.3 \times 0.04 = 0.0485$$

c $\text{P(produced by } A|\text{flawed)} = \dfrac{\text{P(produced by } A \cap \text{flawed)}}{\text{P(flawed)}} = \dfrac{0.25 \times 0.02}{0.0485} = 0.103 \text{ (3 s.f.)}$

8 **a** The reliability of the test depends on whether the person has the condition (C) or not (C').



b $\text{P(tests negative)} = \text{P}(C \cap \text{tests negative}) + \text{P}(C' \cap \text{tests negative})$

$= 0.04 \times 0.1 + 0.96 \times 0.98 = 0.9448 = 0.945 \text{ (3 s.f.)}$

c $\text{P(has condition|tests negative)} = \dfrac{\text{P}(C \cap \text{tests negative})}{\text{P(tests negative)}}$

$= \dfrac{0.04 \times 0.1}{0.9448} = 0.00423 \text{ (3 s.f.)}$

d From the data in the question, the test fails to find 10% of the people with the condition (since it has a 0.1 chance of producing a negative result when a person has the condition).

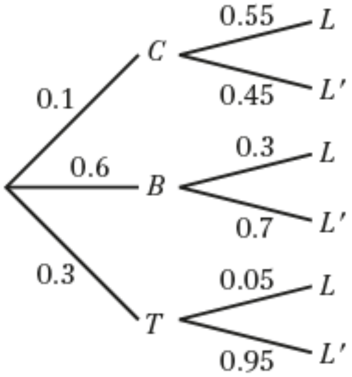
Consider also false positives, the case of a person who does not have the condition returning a positive result.

$\text{P(does not have the condition|tests positive)} = \dfrac{\text{P}(C' \cap \text{tests positive})}{\text{P(tests positive)}} = \dfrac{0.96 \times 0.02}{1 - 0.9448} = 0.348 \text{ (3 s.f.)}$

So over one third of the positive tests are false positives.

This means that if the test was used on the entire population, 10% of the people with the condition would not be identified and over one third of the people with a positive result would actually not have the condition.

9 a Since the probabilities of being late are given, the probabilities for being on time (i.e. not late) for each type of transport are known, sine the probabilities must sum to 1. Therefore the completed tree diagram should be as follows:



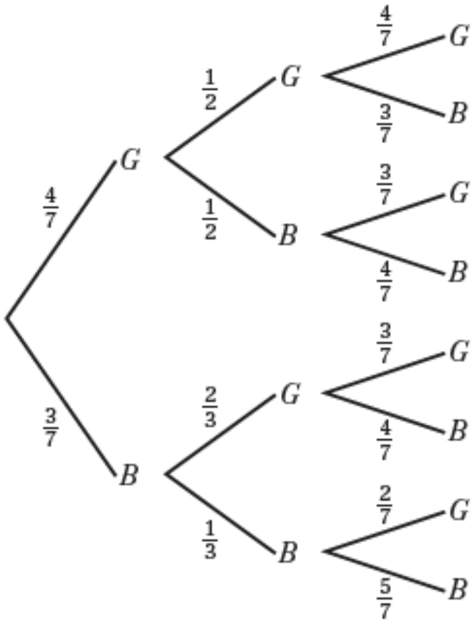
b **i** $\text{P(Bill travels by train and is late)} = 0.3 \times 0.05 = 0.015$

ii To find P(Bill is late), sum P(Bill travels by car and is late), P(Bill travels by bus and is late) and P(Bill travels by train and is late).

$\text{P(Bill is late)} = 0.1 \times 0.55 + 0.6 \times 0.3 + 0.3 \times 0.05 = 0.25$

c $\text{P(Bill travels by bus or train|Bill is late)} = \dfrac{0.6 \times 0.3 + 0.3 \times 0.05}{0.25} = 0.78$

10 a The two counters being drawn from box *A* can be modelled using a tree diagram. In each case, the number of counters of each colour in box *B* is then known, and so the third set of branches can be labelled to represent the drawing of the counter from box *B*. Therefore the completed tree diagram should be:



b $\text{P}(C) = \text{P}(GG) + \text{P}(BB) = \left(\dfrac{4}{7} \times \dfrac{1}{2}\right) + \left(\dfrac{3}{7} \times \dfrac{1}{3}\right) = \dfrac{2}{7} + \dfrac{1}{7} = \dfrac{3}{7}$

10 c $\mathrm{P}(D)=\mathrm{P}(GGB)+\mathrm{P}(GBB)+\mathrm{P}(BGB)+\mathrm{P}(BBB)$

$$=\left(\frac{4}{7}\times\frac{1}{2}\times\frac{3}{7}\right)+\left(\frac{4}{7}\times\frac{1}{2}\times\frac{4}{7}\right)+\left(\frac{3}{7}\times\frac{2}{3}\times\frac{4}{7}\right)+\left(\frac{3}{7}\times\frac{1}{3}\times\frac{5}{7}\right)$$

$$=\frac{6}{49}+\frac{8}{49}+\frac{8}{49}+\frac{5}{49}=\frac{27}{49}$$

d The calculation will be similar to that for P(D), but with the first and second counters being the same colour.

$$\mathrm{P}(C\cap D)=\mathrm{P}(GGB)+\mathrm{P}(BBB)=\left(\frac{4}{7}\times\frac{1}{2}\times\frac{3}{7}\right)+\left(\frac{3}{7}\times\frac{1}{3}\times\frac{5}{7}\right)=\frac{6}{49}+\frac{5}{49}=\frac{11}{49}$$

e Use the addition formula

$$\mathrm{P}(C\cup D)=\mathrm{P}(C)+\mathrm{P}(D)-\mathrm{P}(C\cap D)=\frac{3}{7}+\frac{27}{49}-\frac{11}{49}=\frac{21+27-11}{49}=\frac{37}{49}$$

f The required probability is:

$$\frac{\mathrm{P}(GGG)}{\mathrm{P}(GGG)+\mathrm{P}(BBB)}=\frac{\frac{4}{7}\times\frac{1}{2}\times\frac{4}{7}}{\left(\frac{4}{7}\times\frac{1}{2}\times\frac{4}{7}\right)+\left(\frac{3}{7}\times\frac{1}{3}\times\frac{5}{7}\right)}=\frac{\frac{8}{49}}{\frac{8}{49}+\frac{5}{49}}=\frac{8}{13}=0.615\text{ (3 s.f.)}$$

11 She has not taken into account the fact that after the first jelly bean is selected, there are only 9 jelly beans left in the box. So if the first jelly bean selected is sweet, the probability that the second bean is sweet is $\frac{6}{9}$ not $\frac{7}{10}$.

This is the correct solution.

$$\mathrm{P}(\text{both jelly beans are sweet})=\frac{7}{10}\times\frac{6}{9}=\frac{7}{15}$$

$$\mathrm{P}(\text{at least one jelly bean is sweet})=1-\mathrm{P}(\text{neither jelly bean is sweet})=1-\left(\frac{3}{10}\times\frac{2}{9}\right)=\frac{14}{15}$$

$$\mathrm{P}(\text{both are sweet given at least one is sweet})=\frac{\frac{7}{15}}{\frac{14}{15}}=\frac{7}{14}=0.5$$

The correct answer is therefore 0.5.

Conditional probability Mixed exercise 2

1 **a** $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B) = 0.4 + 0.35 - 0.2 = 0.55$

b $\mathrm{P}(A' \cap B') = 1 - \mathrm{P}(A \cup B) = 1 - 0.55 = 0.45$

c $\mathrm{P}(B|A) = \dfrac{\mathrm{P}(B \cap A)}{\mathrm{P}(A)} = \dfrac{0.2}{0.4} = 0.5$

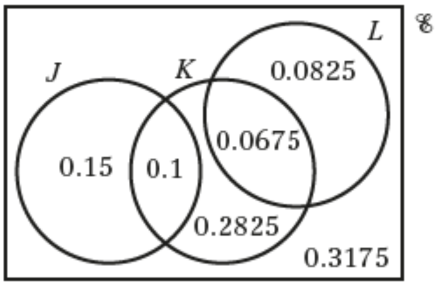
d $\mathrm{P}(A'|B) = \dfrac{\mathrm{P}(A' \cap B)}{\mathrm{P}(B)} = \dfrac{\mathrm{P}(B) - \mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \dfrac{0.15}{0.35} = 0.429$ (3 s.f.)

2 **a** Work out each region of the Venn diagram from the information provided in the question.

First, as J and L are mutually exclusive, $\mathrm{P}(J \cap L) = \varnothing$
So $\mathrm{P}(J \cap K' \cap L') = \mathrm{P}(J) - \mathrm{P}(J \cap K) = 0.25 - 0.1 = 0.15$

As K and L are independent $\mathrm{P}(K \cap L) = \mathrm{P}(K) \times \mathrm{P}(L) = 0.45 \times 0.15 = 0.0675$
So $\mathrm{P}(L \cap K') = \mathrm{P}(L) - \mathrm{P}(L \cap K) = 0.15 - 0.0625 = 0.0825$
And $\mathrm{P}(K \cap J' \cap L') = \mathrm{P}(K) - \mathrm{P}(J \cap K) - \mathrm{P}(K \cap L) = 0.45 - 0.1 - 0.0675 = 0.2825$

Find the outer region by subtracting the sum of all the other regions from 1
$\mathrm{P}(J' \cap K' \cap L') = 1 - 0.15 - 0.1 - 0.2825 - 0.0675 - 0.0825 = 0.3175$



b **i** $\mathrm{P}(J \cup K) = 0.15 + 0.1 + 0.2825 + 0.0675 = 0.6$

ii $\mathrm{P}(J' \cap L') = 0.2825 + 0.3175 = 0.6$

iii $\mathrm{P}(J|K) = \dfrac{\mathrm{P}(J \cap K)}{\mathrm{P}(K)} = \dfrac{0.1}{0.45} = 0.222$ (3 s.f.)

iv $\mathrm{P}(K|J' \cap L') = \dfrac{\mathrm{P}(K \cap (J' \cap L'))}{\mathrm{P}(J' \cap L')} = \dfrac{0.2825}{0.6} = 0.471$ (3 s.f.)

3 a $\mathrm{P}(F \cap S') + \mathrm{P}(S \cap F') = \mathrm{P}(F) - \mathrm{P}(F \cap S) + \mathrm{P}(F) - \mathrm{P}(F \cap S)$

$$= \frac{35 - 27 + 45 - 27}{60} = \frac{26}{60} = 0.433 \text{ (3 s.f.)}$$

b $\mathrm{P}(F|S) = \dfrac{\mathrm{P}(F \cap S)}{\mathrm{P}(S)} = \dfrac{27}{45} = 0.6$

c $\mathrm{P}(S|F') = \dfrac{\mathrm{P}(S \cap F')}{\mathrm{P}(F')} = \dfrac{45 - 27}{60 - 35} = \dfrac{18}{25} = 0.72$

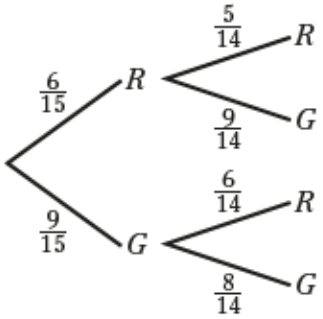
d There are 6 students that study just French and wear glasses ($8 \times 0.75 = 6$) and 9 students that study just Spanish and wear glasses ($18 \times 0.5 = 9$), so the required probability is

$$\mathrm{P}(\text{studies one language and wears glasses}) = \frac{6+9}{60} = \frac{15}{60} = 0.25$$

e There are 26 students studying one language (from part **a**). Of these, 15 wear glasses (from part **d**).

$$\mathrm{P}(\text{wears glasses}|\text{studies one language}) = \frac{15}{26} = 0.577 \text{ (3 s.f.)}$$

4 a



b i $\mathrm{P}(GG) = \dfrac{9}{15} \times \dfrac{8}{14} = \dfrac{3}{5} \times \dfrac{4}{7} = \dfrac{12}{35} = 0.343 \text{ (3 s.f.)}$

ii There are two different ways to obtain balls that are different colours:

$$\mathrm{P}(RG) + \mathrm{P}(GR) = \left(\frac{6}{15} \times \frac{9}{14}\right) + \left(\frac{9}{15} \times \frac{6}{14}\right) = \frac{2 \times 9}{5 \times 7} = \frac{18}{35} = 0.514 \text{ (3 s.f.)}$$

c There are 4 different outcomes:

$\mathrm{P}(RRR) + \mathrm{P}(RGR) + \mathrm{P}(GRR) + \mathrm{P}(GGR)$

$$= \left(\frac{6}{15} \times \frac{5}{14} \times \frac{4}{13}\right) + \left(\frac{6}{15} \times \frac{9}{14} \times \frac{5}{13}\right) + \left(\frac{9}{15} \times \frac{6}{14} \times \frac{5}{13}\right) + \left(\frac{9}{15} \times \frac{8}{14} \times \frac{6}{13}\right)$$

$$= \frac{120 + 270 + 270 + 432}{2730} = \frac{1092}{2730} = 0.4$$

4 d The only way for this to occur is to draw a green ball each time. The corresponding probability is:

$$\text{P}(GGGG) = \frac{9}{15} \times \frac{8}{14} \times \frac{7}{13} \times \frac{6}{12} = \frac{3 \times 2}{5 \times 13} = \frac{6}{65} = 0.0923 \text{ (3 s.f.)}$$

5 a Either Colin or Anne must win both sets. Therefore the required probability is:

$$\text{P(match over in two sets)} = (0.7 \times 0.8) + (0.3 \times 0.6) = 0.56 + 0.18 = 0.74$$

b $\text{P(Anne wins|match over in two sets)} = \dfrac{0.7 \times 0.8}{0.74} = \dfrac{0.56}{0.74} = 0.757 \text{ (3 s.f.)}$

c The three ways that Anne can win the match are: wins first set, wins second set; wins first set, loses second set, wins tiebreaker; loses first set, wins second set, wins tiebreaker.

$$\begin{aligned}\text{P(Anne wins match)} &= (0.7 \times 0.8) + (0.7 \times 0.2 \times 0.55) + (0.3 \times 0.4 \times 0.55) \\ &= 0.56 + 0.077 + 0.066 = 0.703\end{aligned}$$

6 a There are 20 kittens with neither black nor white paws (75 – 26 – 14 – 15 = 20).

$$\text{P(neither white or black paws)} = \frac{20}{75} = \frac{4}{15} = 0.267 \text{ (3 s.f.)}$$

b There are 41 kittens with some black paws (26 + 15 = 41).

$$\text{P(black and white paws|some black paws)} = \frac{15}{41} = 0.366 \text{ (3 s.f.)}$$

c This is selection without replacement (since the first kitten chosen is not put back).

$$\text{P(both kittens have all black paws)} = \frac{26}{75} \times \frac{25}{74} = \frac{13}{3 \times 37} = \frac{13}{111} = 0.117 \text{ (3 s.f.)}$$

d There are 29 kittens with some white paws (14 + 15 = 29).

$$\text{P(both kittens have some white paws)} = \frac{29}{75} \times \frac{28}{74} = \frac{812}{5550} = 0.146 \text{ (3 s.f.)}$$

7 a Using the fact that A and B are independent: $\text{P}(A) \times \text{P}(B) = \text{P}(A \cap B) \Rightarrow \text{P}(B) = \dfrac{0.12}{0.4} = 0.3$

b Use the addition formula to find $\text{P}(A \cup B)$

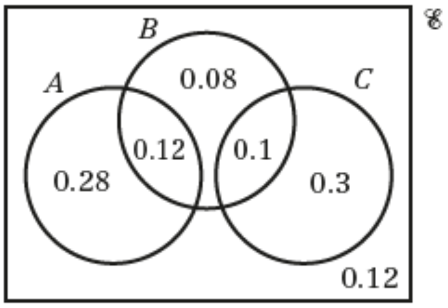
$\text{P}(A \cup B) = \text{P}(A) + \text{P}(B) - \text{P}(A \cap B) = 0.4 + 0.3 - 0.12 = 0.58$

$\text{P}(A' \cap B') = 1 - \text{P}(A \cup B) = 1 - 0.58 = 0.42$

7 c As A and C are mutually exclusive

$\mathrm{P}(A \cap B' \cap C') = P(A) - \mathrm{P}(A \cap B) = 0.4 - 0.12 = 0.28$
$\mathrm{P}(C \cap A' \cap B') = P(C) - \mathrm{P}(B \cap C) = 0.4 - 0.1 = 0.3$
$\mathrm{P}(B \cap A' \cap C') = P(B) - \mathrm{P}(A \cap B) - P(B \cap C) = 0.3 - 0.12 - 0.1 = 0.08$

Find the outer region by subtracting the sum of all the other regions from 1
$\mathrm{P}(A' \cap B' \cap C') = 1 - 0.28 - 0.12 - 0.08 - 0.1 - 0.3 = 0.12$



d i $\mathrm{P}(B|C) = \dfrac{\mathrm{P}(B \cap C)}{\mathrm{P}(C)} = \dfrac{0.1}{0.4} = 0.25$

ii The required region must be contained within A, and not include B (the condition on C is irrelevant since A and C are mutually exclusive). Therefore, $\mathrm{P}(A \cap (B' \cup C)) = 0.28$

8 a It may be that neither team scores in the match, and it is a 0–0 draw.

b $\mathrm{P}(\text{team } A \text{ scores first}) = \mathrm{P}(\text{team } A \text{ scores first and wins}) + \mathrm{P}(\text{team } A \text{ scores first and does not win})$
So $\mathrm{P}(\text{team } A \text{ scores first and does not win}) = 0.6 - 0.48 = 0.12$

c From the question $\mathrm{P}(A \text{ wins}|B \text{ scores first}) = 0.3$. Using the multiplication formula gives

$$\mathrm{P}(A \text{ wins}|B \text{ scores first}) = \frac{\mathrm{P}(A \text{ wins} \cap B \text{ scores first})}{\mathrm{P}(B \text{ scores first})} = 0.3$$

$\Rightarrow \mathrm{P}(A \text{ wins} \cap B \text{ scores first}) = 0.3 \times 0.35 = 0.105$

Now find the required probability

$$\mathrm{P}(B \text{ scores first}|A \text{ wins}) = \frac{\mathrm{P}(A \text{ wins} \cap B \text{ scores first})}{\mathrm{P}(A \text{ wins})} = \frac{0.105}{0.48 + 0.105} = \frac{0.105}{0.585} = 0.179 \text{ (3 s.f.)}$$

Challenge

a Let $\mathrm{P}(A \cap B) = k$
As $\mathrm{P}(A \cap B) \leqslant P(B) \Rightarrow k \leqslant 0.2$
A and B could be mutually exclusive, meaning $\mathrm{P}(A \cap B) = 0$, so $0 \leqslant k \leqslant 2$

Now, $\mathrm{P}(A \cap B') = \mathrm{P}(A) - \mathrm{P}(A \cap B)$, so $p = 0.6 - k \Rightarrow 0.4 \leqslant p \leqslant 0.6$

Challenge

b Use the fact that $\mathrm{P}(A \cap C) = \mathrm{P}(A \cap B \cap C) + \mathrm{P}(A \cap B' \cap C)$
So $\mathrm{P}(A \cap B' \cap C) = \mathrm{P}(A \cap C) - \mathrm{P}(A \cap B \cap C) = \mathrm{P}(A \cap C) - 0.1$

Consider the range of $\mathrm{P}(A \cap C)$
$\mathrm{P}(A \cap C) \leqslant \mathrm{P}(A) \Rightarrow \mathrm{P}(A \cap C) \leqslant 0.6$

By the multiplication formula $\mathrm{P}(A \cup C) = \mathrm{P}(A) + \mathrm{P}(C) - \mathrm{P}(A \cap C)$
So $\mathrm{P}(A \cap C) = \mathrm{P}(A) + \mathrm{P}(C) - \mathrm{P}(A \cup C) = 1.3 - \mathrm{P}(A \cup C)$
As $\mathrm{P}(A \cup C) \leqslant 1 \Rightarrow \mathrm{P}(A \cap C) \geqslant 0.3$

So $0.3 \leqslant \mathrm{P}(A \cap C) \leqslant 0.6$ and as $\mathrm{P}(A \cap B' \cap C) = \mathrm{P}(A \cap C) - 0.1$ this gives the result that $0.3 - 0.1 \leqslant \mathrm{P}(A \cap B' \cap C) \leqslant 0.6 - 0.1$, so $0.2 \leqslant q \leqslant 0.5$

The normal distribution 3A

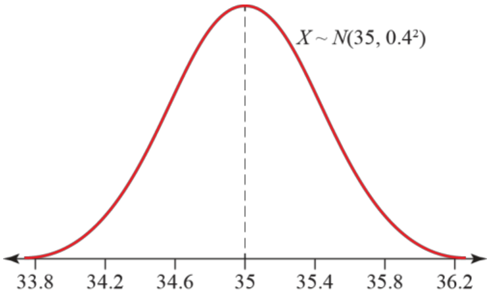
1 **a** Continuous – lengths can take any value.

b Discrete – scores can only take certain values.

c Continuous – masses can take any value.

d Discrete – shoe sizes can only take certain values.

2 Since the mean is 35 mm, the distribution should be symmetrical about this value. Since the standard deviation is 0.4 mm, 68% of the data should lie in the range 34.6 mm to 35.4 mm and 95% of the data should lie in the range 34.2 mm to 35.8 mm.



3 One of the key features of normal distributions is that they are symmetrical about the mean (which is equal to the mode and the median). This curve shows that the bank employees' incomes are not equally distributed either side of its peak (the modal income), so the normal distribution is not a suitable model.

4 **a** Since $\sigma^2 = 16$, the standard deviation is $\sqrt{16} = 4$ cm and the mean is 120 cm.
68% of the pupils are expected to have an armspan within one standard deviation of the mean, i.e. within the interval 116 cm to 124 cm.

b The given interval, 112 cm to 128 cm, includes all of the pupils whose armspans are up to two standard deviations either side of the mean. Therefore 95% of the pupils can be expected to have an armspan within this range.

5 If 68% of the adders have a length between 93 cm (100 cm − 7 cm) and 107 cm (100 cm + 7 cm), then the standard deviation is 7 cm. Therefore the variance, σ^2, is $7^2 = 49$.

6 Since 95% of the data should lie within two standard deviations of the mean, 2.5% of the data should be two standard deviations or more below the mean and 2.5% of the data should be two standard deviations or more above the mean. Since 2.5% of the dormice weigh more than 70 grams, this means that 70 grams is two standard deviations above the mean. The standard deviation is 5 grams (the square root of the variance) so the mean is 60 grams.

7 In a normal distribution, 68% of the data lies within one standard deviation, σ, of the mean, μ, and 32% lies outside of this range. Therefore 16% of the data lies below $\sigma - \mu$, and 16% lies above $\sigma + \mu$. Also, 95% of the data lies between $\mu - 2\sigma$ and $\mu + 2\sigma$. Therefore 2.5% lies below $\mu - 2\sigma$ and 2.5% lies above $\mu + 2\sigma$.

The question states that 84% of the pigs weigh more than 52 kg, so 16% weigh less than 52 kg. Hence $52 = \mu - \sigma$. Also, 97.5% of the pigs weigh more than 47.5 kg, so 2.5% weigh less than 47.5 kg and $47.5 = \mu - 2\sigma$. From these two equations, $\sigma = 52 - 47.5 = 4.5$ kg and so $\sigma^2 = (4.5)^2 = 20.25$.
Hence $\mu = 52 + \sigma = 52 + 4.5 = 56.5$ kg.

8 a Since the normal distribution is symmetrical and the mean is equal to the median and the mode, half the scores should be above 45 and half should be below. Therefore $\mathrm{P}(S > 45) = 0.5$.

b Since the data within one standard deviation of the mean should be 68% of the sample, $\mathrm{P}(30 < S < 60) = 0.68$.

c Since the data within two standard deviations of the mean should be 95% of the sample, $\mathrm{P}(15 < S < 75) = 0.95$.

d Alexia is incorrect: although $\mathrm{P}(X > 100) > 0,$ the value is very small as 100 as more than three standard deviations from the mean, so the model as a whole is still reasonable.

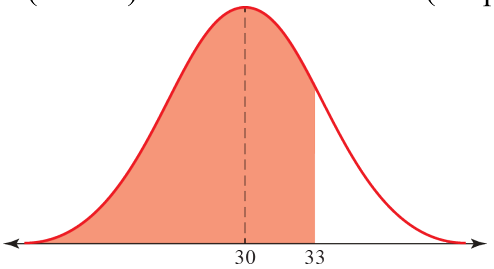
9 a The mean of the normal distribution is where the highest point on the curve appears. From the sketch, this is around 36 cm.

b The points of inflection of the normal occur at $\mu + \sigma$ and $\mu - \sigma$. From the sketch, these points lie somewhere in the intervals [33, 34] and [38, 39]. Since the mean is around 36 cm, this means that the standard deviation should be somewhere between 2 cm and 3 cm.

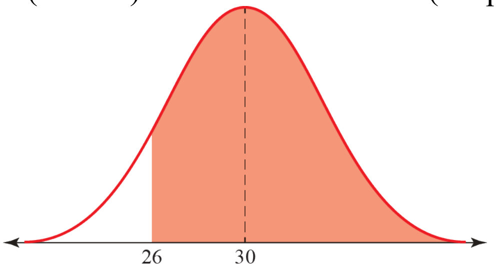
The normal distribution 3B

1 Use the Normal CD function on your calculator, with $\mu = 30$ and $\sigma = 2$.

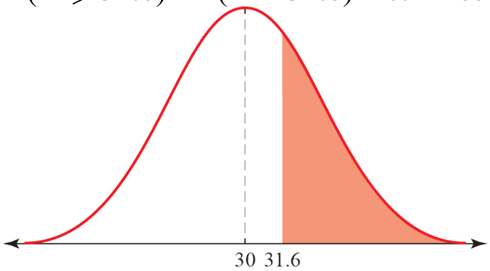
a Set a small value for the lower limit, e.g. 0.
$P(X < 33) = 0.93319... = 0.9332$ (4 d.p.)



b Set a large value for the upper limit, e.g. 1000.
$P(X > 26) = 0.97724... = 0.9772$ (4 d.p.)

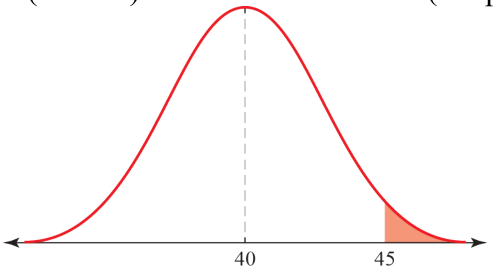


c Set a large value for the upper limit, e.g. 1000.
$P(X \geqslant 31.6) = P(X > 31.6) = 0.21185... = 0.2119$ (4 d.p.)

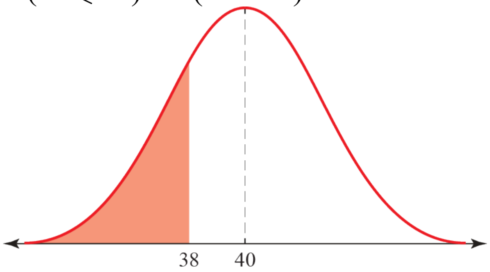


2 Use the Normal CD function on your calculator, with $\mu = 40$ and $\sigma = \sqrt{9} = 3$.

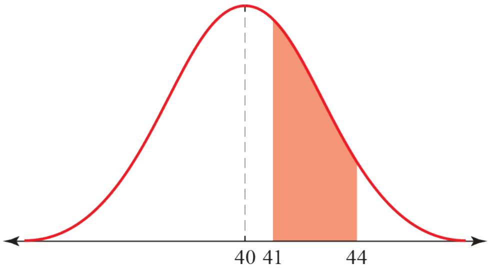
a Set a large value for the upper limit, e.g. 1000.
$P(X > 45) = 0.04779... = 0.0478$ (4 d.p.)



2 b Set a small value for the lower limit, e.g. 0.
$P(X \leqslant 38) = P(X < 38) = 0.25249... = 0.2525$ (4 d.p.)

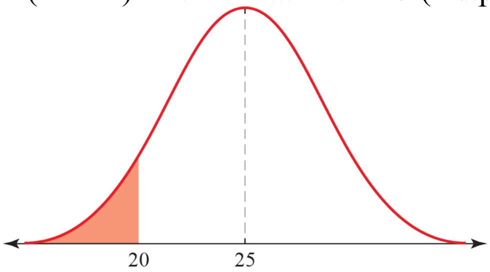


c $P(41 \leqslant X \leqslant 44) = P(41 < X < 44) = 0.27823... = 0.2782$ (4 d.p.)

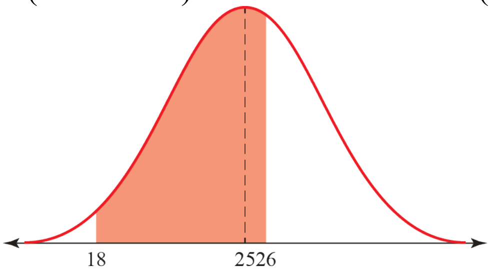


3 Use the Normal CD function on your calculator, with $\mu = 25$ and $\sigma = \sqrt{25} = 5$.

a Set a small value for the lower limit, e.g. 0.
$P(Y < 20) = 0.15865... = 0.1587$ (4 d.p.)



b $P(18 < Y < 26) = 0.49850... = 0.4985$ (4 d.p.)



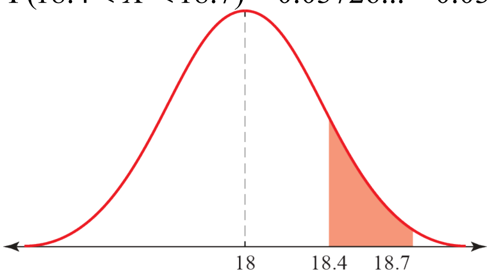
c Set a large value for the upper limit, e.g. 1000.
$P(Y > 23.8) = 0.59483... = 0.5948$ (4 d.p.)

4 Use the Normal CD function on your calculator, with $\mu = 18$ and $\sigma = \sqrt{10}$.

a Set a large value for the upper limit, e.g. 1000.
$P(X \geqslant 20) = P(X > 20) = 0.26354... = 0.2635$ (4 d.p.)

b Set a small value for the lower limit, e.g. 0.
$P(X < 15) = 0.17139... = 0.1714$ (4 d.p.)

4 c $P(18.4 < X < 18.7) = 0.03726... = 0.0373$ (4 d.p.)



5 Use the Normal CD function on your calculator, with $\mu = 15$ and $\sigma = 1.5$.

a i Set a large value for the upper limit, e.g. 1000.
$P(M > 14) = 0.74750... = 0.7474$ (4 d.p.)

ii Set a small value for the lower limit, e.g. 0.
$P(M < 14) = 0.25249... = 0.2525$ (4 d.p.)

b $P(M > 14) + P(M < 14) = 0.7475 + 0.2525 = 1$
The sum is 1, as the combined probabilities include every possible value.

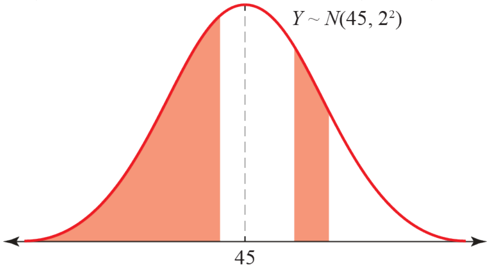
6 a Use the Normal CD function on your calculator, with $\mu = 4.5, \sigma = \sqrt{0.4}$ and a small value for the lower limit.
$P(T < 4.2) = 0.31762... = 0.3176$ (4 d.p.)

b $P(T > 4.2) = 1 - P(T < 4.2) = 1 - 0.3176 = 0.6824$ (4 d.p.)

7 Use the Normal CD function on your calculator, with $\mu = 45$ and $\sigma = 2$.

a $P(Y < 41 \text{ or } Y > 47) = 1 - P(41 < Y < 47)$
Using your calculator, $P(41 < Y < 47) = 0.81859...$
So $P(Y < 41 \text{ or } Y > 47) = 1 - 0.81859... = 0.1814$ (4 d.p.)

b $P(Y < 44 \text{ or } 46.5 < Y < 47.5) = P(Y < 44) + P(46.5 < Y < 47.5)$



Using your calculator, $P(Y < 44) = 0.30853...$ and $P(46.5 < Y < 47.5) = 0.12097...$
so $P(Y < 44 \text{ or } 46.5 < Y < 47.5) = 0.30853... + 0.12097... = 0.4295$ (4 d.p.)

8 Use the Normal CD function on your calculator, with $\mu = 6$ and $\sigma = 0.8$.

a i A suitable upper limit is 10, giving $P(X < 7) = 0.10564... = 0.1056$ (4 d.p.)

ii A suitable lower limit is 2, giving $P(X < 5) = 0.10564... = 0.1056$ (4 d.p.)

b Since these are independent events, the probability is $P(X < 5)^3$, i.e.
$(0.10564...)^3 = 0.00117... = 0.0012$ (4 d.p.)

9 Use the Normal CD function on your calculator, with $\mu = 500$ and $\sigma = 14$.

a **i** A suitable upper limit is 570, giving $\text{P}(W > 505) = 0.36049... = 0.3605$ (4 d.p.)

ii A suitable lower limit is 430, giving $\text{P}(W < 490) = 0.23752... = 0.2375$ (4 d.p.)

b Since these are independent events, the probability is $\text{P}(W > 490)^4$.
$\text{P}(W > 490) = 1 - \text{P}(W < 490) = 1 - 0.23752... = 0.76248...$
So the probability is $(0.76248\ldots)^4 = 0.33799\ldots = 0.3380$ (4 d.p.).

10 Use the Normal CD function on your calculator, with $\mu = 165$ and $\sigma = 3.5$.

a A suitable lower limit is 10, giving $\text{P}(h < 160) = 0.07656... = 0.0766$ (4 d.p.)

b $\text{P}(168 < h < 174) = 0.19061... = 0.1906$ (4 d.p.)

c Use the binomial distribution $X \sim \text{B}(20, 0.1906)$.
Using the binomial CD function on your calculator:
$\text{P}(X \geqslant 5) = 1 - \text{P}(X \leqslant 4) = 1 - 0.67035... = 0.3296$ (4 d.p.)

11 Use the Normal CD function on your calculator, with $\mu = 13$ and $\sigma = 0.1$.

a A suitable lower limit is 12.5, giving $\text{P}(D < 12.8) = 0.02274... = 0.0227$ (4 d.p.)

b P('perfect') = $\text{P}(12.9 < D < 13.1) = 0.68268... = 0.6827$ (4 d.p.)
Use the binomial distribution $X \sim \text{B}(40, 0.6827)$.
Using the binomial CD function on your calculator:
$\text{P}(X > 25) = 1 - \text{P}(X \leqslant 25) = 1 - 0.26549... = 0.7345$ (4 d.p.)

12 Use the Normal CD function on your calculator, with $\mu = 480$ and $\sigma = 40$.

a A suitable upper limit is 680, giving $\text{P}(X > 490) = 0.40129... = 0.4013$ (4 d.p.)

b $\text{P}(470 < X < 490) = 0.19741... = 0.1974$ (4 d.p.)
Use the binomial distribution $Y \sim \text{B}(30, 0.1974)$.
Using the binomial CD function on your calculator:
$\text{P}(X \geqslant 15) = 1 - \text{P}(X \leqslant 14) = 1 - 0.99980141... = 0.0001986$ (4 s.f.)

The normal distribution 3C

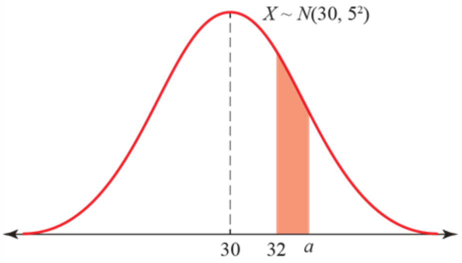
1 Use the inverse normal distribution function on your calculator, with $\mu = 30$ and $\sigma = 5$.

a $\text{P}(X < a) = 0.3 \Rightarrow a = 27.377... = 27.38$ (2 d.p.).

b $\text{P}(X < a) = 0.75 \Rightarrow a = 33.372... = 33.37$ (2 d.p.).

c $\text{P}(X > a) = 0.4 \Rightarrow \text{P}(X < a) = 0.6 \Rightarrow a = 31.266... = 31.27$ (2 d.p.).

d Since $P(32 < X < a) = 0.2$, it must be that $a > 32$.



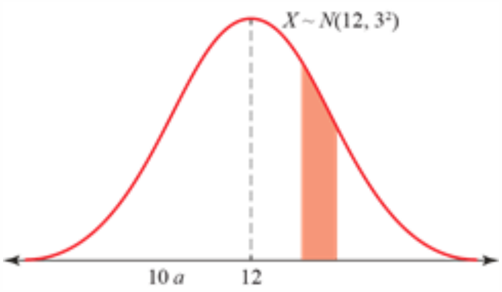
$$\begin{aligned}&\text{P}(32 < X < a) = 0.2\\&\Rightarrow \text{P}(X < a) - \text{P}(X < 32) = 0.2\\&\Rightarrow \text{P}(X < a) = 0.2 + \text{P}(X > 32) = 0.2 + 0.6554 = 0.8554\\&\Rightarrow a = 35.299... = 35.30 \text{ (2 d.p.)}\end{aligned}$$

2 Use the inverse normal distribution function on your calculator, with $\mu = 12$ and $\sigma = 3$.

a $\text{P}(X < a) = 0.1 \Rightarrow a = 8.155... = 8.16$ (2 d.p.).

b $\text{P}(X > a) = 0.65 \Rightarrow \text{P}(X < a) = 0.35 \Rightarrow a = 10.844... = 10.84$ (2 d.p.).

c



$$\begin{aligned}&\text{P}(10 \leqslant X \leqslant a) = \text{P}(X \leqslant a) - \text{P}(X \leqslant 10) = 0.25\\&\Rightarrow \text{P}(X < a) = 0.25 + \text{P}(X < 10) = 0.25 + 0.2525 = 0.5025\\&\Rightarrow a = 12.018... = 12.02 \text{ (2 d.p.)}\end{aligned}$$

d $\text{P}(a < X < 14) = \text{P}(X < 14) - \text{P}(X < a) = 0.32$

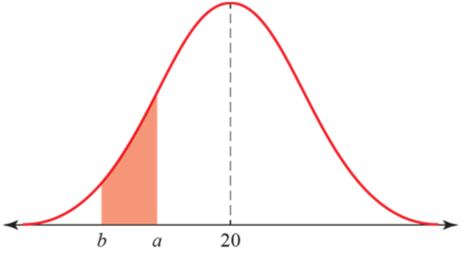
$$\begin{aligned}&\Rightarrow \text{P}(X < a) = \text{P}(X < 14) - 0.32 = 0.7475 - 0.32 = 0.4275\\&\Rightarrow a = 11.451... = 11.45 \text{ (2 d.p.)}\end{aligned}$$

3 Use the inverse normal distribution function on your calculator, with $\mu = 20$ and $\sigma = \sqrt{12}$.

a **i** $\text{P}(X < a) = 0.40 \Rightarrow a = 19.12... = 19.1$ (1 d.p.)

ii $\text{P}(X > b) = 0.6915 \Rightarrow \text{P}(X < b) = 0.3085 \Rightarrow b = 18.26... = 18.3$ (1 d.p.)

3 b $P(b < X < a) = P(X < a) - P(X < b) = 0.40 - (1 - 0.6915) = 0.0915$

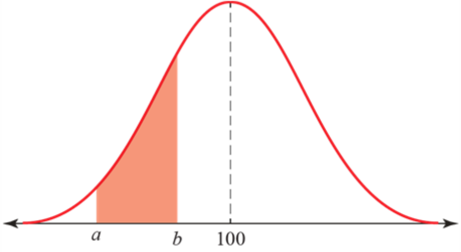


4 Use the inverse normal distribution function on your calculator, with $\mu = 100$ and $\sigma = 15$.

a i $P(Y > a) = 0.975 \Rightarrow P(Y < a) = 0.025 \Rightarrow a = 70.60... = 70.6$ (1 d.p.)

ii $P(Y < b) = 0.10 \Rightarrow b = 80.77... = 80.8$ (1 d.p.)

b $P(a < Y < b) = P(Y < b) - P(Y < a) = 0.10 - (1 - 0.975) = 0.10 - 0.025 = 0.075$

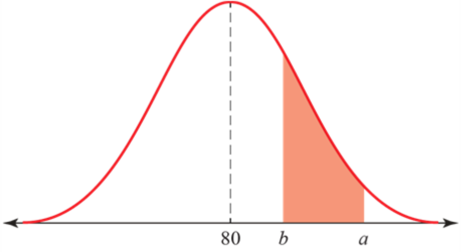


5 Use the inverse normal distribution function on your calculator, with $\mu = 80$ and $\sigma = \sqrt{16} = 4$.

a i $P(X > a) = 0.40 \Rightarrow P(X < a) = 0.60 \Rightarrow a = 81.01... = 81.0$ (1 d.p.)

ii $P(X < b) = 0.5636 \Rightarrow b = 80.64... = 80.6$ (1 d.p.)

b $P(b < X < a) = P(X < a) - P(X < b) = (1 - 0.40) - 0.5636 = 0.60 - 0.5636 = 0.0364$ (4 d.p.)



6 Use the inverse normal distribution function on your calculator, with $\mu = 4.5$ and $\sigma = 0.6$.

a By definition, the lower quartile is the point Q_1 such that $P(M < Q_1) = 0.25$.
$P(M < Q_1) = 0.25 \Rightarrow Q_1 = 4.0953... = 4.095$ kg (3 d.p.).

b Let a be the 80th percentile, so that $P(M < a) = 0.8$.
$P(M < a) = 0.8 \Rightarrow a = 5.0049... = 5.005$ kg (3 d.p.)

c The mean is 4.5, and since the data is normally distributed, this means that 50% of the badgers will have a mass less than 4.5 kg, i.e. $Q_2 = 4.5$ kg.

7 Use the inverse normal distribution function on your calculator, with $\mu = 72$ and $\sigma = 6$.

a $P(X < a) = 0.6 \Rightarrow a = 73.520... = 73.52$ (2 d.p.).

7 b $P(X < Q_1) = 0.25 \Rightarrow Q_1 = 67.953...$ and $P(X < Q_3) = 0.75 \Rightarrow Q_3 = 76.046...$
So the interquartile range = $Q_3 - Q_1 = 76.046 - 67.953 = 8.093 = 8.09$ (2 d.p.)

8 Use the inverse normal distribution function on your calculator, with $\mu = 60$ and $\sigma = 2$.

a $\text{P}(Y > y) = 0.2 \Rightarrow \text{P}(Y < y) = 0.8 \Rightarrow y = 61.683... = 61.68$ (2 d.p.)

b $\text{P}(X < a) = 0.1 \Rightarrow a = 57.436...$ and $\text{P}(X < b) = 0.9 \Rightarrow b = 62.563...$
So the 10% to 90% interpercentile range of masses is $b - a = 5.127 = 5.13$ grams (2 d.p.).

c Tom is correct: the data is assumed to be normally distributed, so the median is equal to the mean.

9 a The short coat should be suitable for the shortest 30% of the men.
Since $\text{P}(H < a) = 0.3 \Rightarrow a = 165,$ this means that the short coat should be suitable for the men who are up to 165 cm tall.
The long coat should be suitable for the tallest 20% of the men.
Since $\text{P}(H > b) = 0.2 \Rightarrow \text{P}(H < b) = 0.8 \Rightarrow b = 178,$ this means that the long coat should be suitable for the men who are more than 178 cm tall.
The regular coat is then suitable for those in between, i.e. those men who are between 165 cm and 178 cm tall.

b There are many assumptions made by the model. It is assumed, for example, that the men likely to shop at the stores selling these frock coats have the same distribution of heights as those in the ‘large group’; that the men’s measurements from their necks to near the floor are normally distributed; that arm lengths are also normally distributed; that people are ‘in proportion’ (that, generally, taller men have longer arm lengths and higher shoulders than shorter men); and that the population follows the normal distribution over the whole range of values, i.e. that there are no extreme outliers.

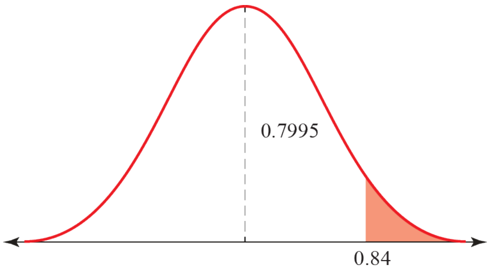
The normal distribution 3D

1 Use the Normal CD function on your calculator, with $\mu = 0, \sigma = 1$ and a small value for the lower limit, e.g. −10.

a $P(Z < 2.12) = 0.98299... = 0.9830$ (4 d.p.)

b $P(Z < 1.36) = 0.91308... = 0.9131$ (4 d.p.)

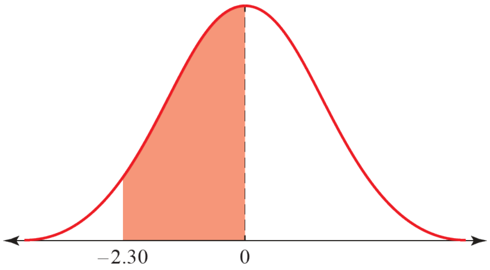
c



$P(Z > 0.84)$
$= 1 - P(Z < 0.84)$
$= 1 - 0.79954...$
$= 0.20045... = 0.2005$ (4 d.p.)

d $P(Z < -0.38) = 0.35197... = 0.3520$ (4 d.p.)

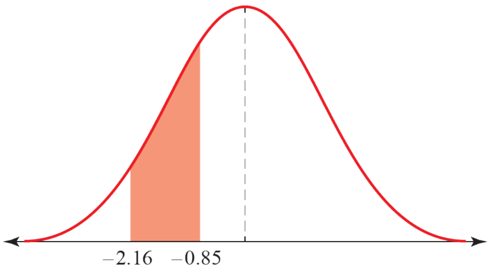
e



$P(-2.30 < Z < 0)$
$= 0.5 - P(Z < -2.30)$
$= 0.5 - 0.1072...$
$= 0.48929... = 0.4893$ (4 d.p.)

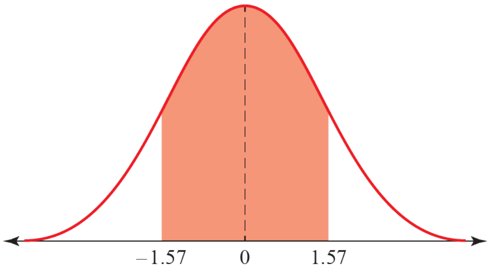
f $P(Z < -1.63) = 0.05155... = 0.0516$ (4 d.p.)

1 **g**



$P(-2.16 < Z < -0.85)$
$= P(Z < -0.85) - P(Z < -2.16)$
$= 0.19766... - 0.01538...$
$= 0.18227... = 0.1823$ (4 d.p.)

h



$P(-1.57 < Z < 1.57)$
$= 2 \times (0.5 - P(Z < -1.57)$
$= 2 \times (0.5 - 0.05820...)$
$= 2 \times 0.44179...$
$= 0.88358... = 0.8836$ (4 d.p.)

2 Use the inverse normal distribution function on your calculator, with $\mu = 0$ and $\sigma = 1$.

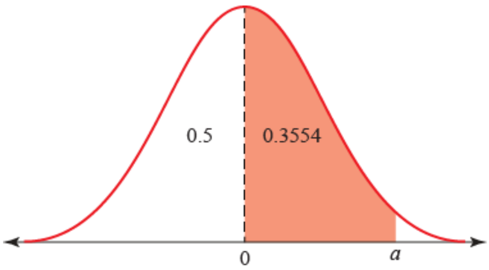
a $P(Z < a) = 0.9082 \Rightarrow a = 1.32975... = 1.3298$ (4 d.p.)

b $P(Z > a) = 0.0314$
$\Rightarrow P(Z < a) = 0.9686$
$\Rightarrow a = 1.86060... = 1.8606$ (4 d.p.)

c $P(Z > a) = 0.15$
$\Rightarrow P(Z < a) = 0.85$
$\Rightarrow a = 1.03643... = 1.0364$ (4 d.p.)
(Alternatively, use the table of percentage points with $p = 0.15 \Rightarrow a = 1.0364$)

d $P(Z > a) = 0.95$
$\Rightarrow P(Z < a) = 0.05$
$\Rightarrow a = -1.64485... = -1.6449$ (4 d.p.)
(Alternatively, use the table of percentage points with $p = 0.05 \Rightarrow -a = 1.6449 \Rightarrow a = -1.6449$)

2 **e**

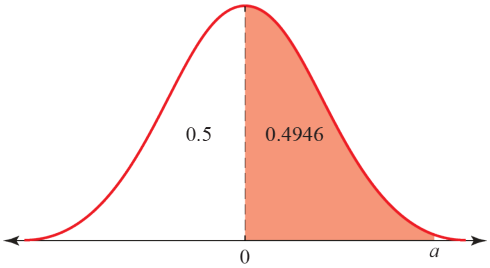


$P(0 < Z < a) = 0.3554$

$\Rightarrow P(Z < a) = 0.8554$

$\Rightarrow a = 1.05987... = 1.0599$ (4 d.p.)

f

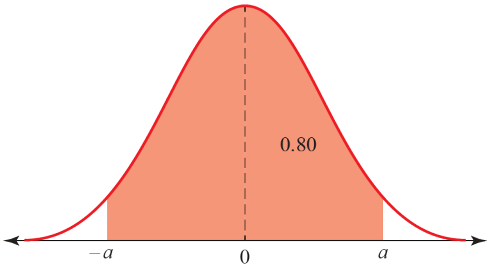


$P(0 < Z < a) = 0.4946$

$\Rightarrow P(Z < a) = 0.9946$

$\Rightarrow a = 2.54910... = 2.5491$ (4 d.p.)

g



$P(-a < Z < a) = 0.80$

$\Rightarrow P(-a < Z < 0) = 0.40$

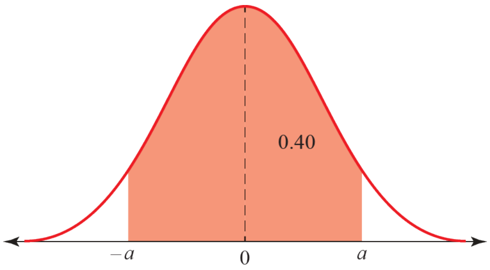
$\Rightarrow P(-a < Z) = 0.10$

$\Rightarrow -a = -1.28155...$

$\Rightarrow a = 1.2816$ (4 d.p.)

(Alternatively, use the table of percentage points with $p = 0.10 \Rightarrow a = 1.2816$)

2 h



$\mathrm{P}(-a < Z < a) = 0.40$

$\Rightarrow \mathrm{P}(-a < Z < 0) = 0.20$

$\Rightarrow \mathrm{P}(-a < Z) = 0.30$

$\Rightarrow -a = -0.52440...$

$\Rightarrow a = 0.5244$ (4 d.p.)

(Alternatively, use the table of percentage points with $p = 0.30 \Rightarrow a = 0.5244$)

3 a $x = 0.8 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{0.8-0.8}{0.05} = 0$

b $x = 0.792 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{0.792-0.8}{0.05} = -0.16$

c $x = 0.81 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{0.81-0.8}{0.05} = 0.2$

d $x = 0.837 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{0.837-0.8}{0.05} = 0.74$

4 a $x = 154 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{154-154}{12} = 0 \Rightarrow \mathrm{P}(X < 154) = \Phi(0)$

b $x = 160 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{160-154}{12} = 0.5 \Rightarrow \mathrm{P}(X < 160) = \Phi(0.5)$

c $x = 151 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{151-154}{12} = -0.25 \Rightarrow \mathrm{P}(X > 151) = 1 - \mathrm{P}(X < 151) = 1 - \Phi(-0.25)$

d $x = 140 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{140-154}{12} = -\dfrac{7}{6}$

$x = 155 \Rightarrow z = \dfrac{x-\mu}{\sigma} = \dfrac{155-154}{12} = \dfrac{1}{12}$

$\Rightarrow \mathrm{P}(140 < X < 155) = \mathrm{P}(X < 155) - \mathrm{P}(X < 140) = \Phi\left(\dfrac{1}{12}\right) - \Phi\left(-\dfrac{7}{6}\right)$

5 a $\mathrm{P}(Z > z) = 0.025 \Rightarrow p = 0.025$

Using the percentage points table, $p = 0.025 \Rightarrow z = 1.96$

5 b Using the formula $z = \dfrac{x-\mu}{\sigma}$:

$$\begin{aligned}1.96 &= \frac{x-80}{4}\\ x-80 &= 4\times 1.96\\ x &= 80+7.84\\ &= 87.84\end{aligned}$$

A score of 87.8 (3 s.f.) is needed to get on the programme.

6 a From the percentage points table, $p = 0.15 \Rightarrow z = 1.0364$

Therefore $\text{P}(Z > 1.0364) = 0.15$, hence $\text{P}(Z < -1.0364) = 0.15$, so $z = -1.0364$

b Using the formula $z = \dfrac{x-\mu}{\sigma}$:

$$\begin{aligned}-1.0364 &= \frac{x-57}{2}\\ x-57 &= 2\times(-1.0364)\\ x &= 57-2.0728\\ &= 54.9272\end{aligned}$$

The size of a ‘petite’ hat is 54.9 cm (3 s.f.).

7 a The 90th percentile corresponds to $p = 0.1$.

From the percentage points table, $p = 0.10 \Rightarrow z = 1.2816$

By the symmetry of the normal distribution, the 10th percentile is at $z = -1.2816$

So the 10% to 90% interpercentile range corresponds to $-1.2816 < z < 1.2816$

b A ‘standard’ light bulb should have a range of life within the above range, but for $\text{N}(1175, 56)$.

Using the formula $z = \dfrac{x-\mu}{\sigma}$ with $z = -1.2816$:

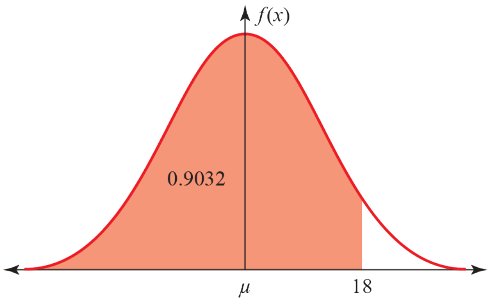
$$\begin{aligned}-1.2816 &= \frac{x-1175}{56}\\ x-1175 &= 56\times(-1.2816)\\ x &= 1175-71.7696\\ &= 1103.2304\end{aligned}$$

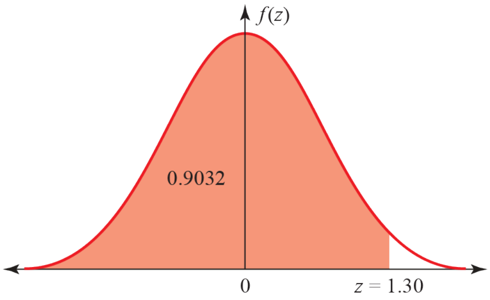
Similarly, for $z = 1.2816$, $x = 1175 + 71.7696 = 1246.7696$.

So the range of life for a ‘standard’ bulb is 1103 to 1247 hours.

The normal distribution 3E

1 $P(X < 18) = 0.9032 \Rightarrow P\left(Z < \frac{18-\mu}{5}\right) = 0.9032$



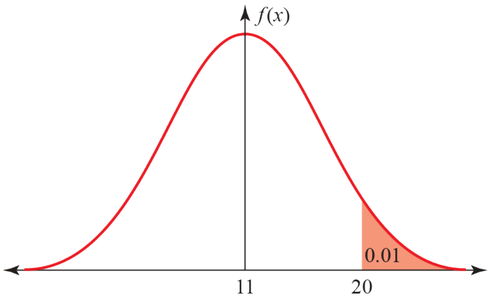


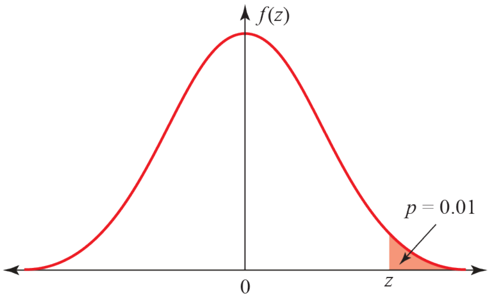
Using the inverse normal function, $z = 1.30$

$$\text{so } 1.30 = \frac{18-\mu}{5}$$

$$\mu = 18 - 5 \times 1.30 = 11.5$$

2 $\mathrm{P}(X > 20) = 0.01 \Rightarrow \mathrm{P}\left(Z < \frac{20-11}{\sigma}\right) = 0.01$





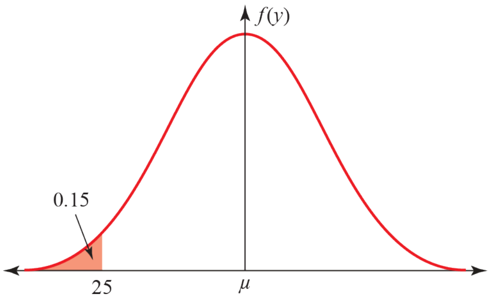
Using the inverse normal function, $z = 2.3263...$

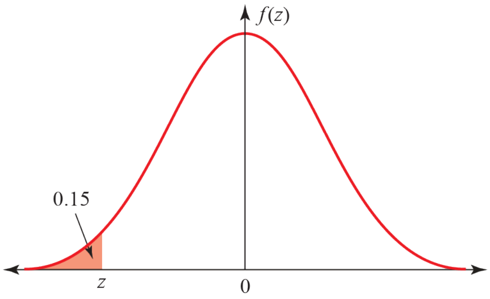
$$\text{so } 2.3263... = \frac{20-11}{\sigma}$$

$$\sigma = \frac{9}{2.3263...}$$

$$= 3.8687... = 3.87 \text{ (3 s.f.)}$$

3 $\mathrm{P}(Y < 25) = 0.15 \Rightarrow \mathrm{P}\left(Z < \dfrac{25-\mu}{\sqrt{40}}\right) = 0.15$





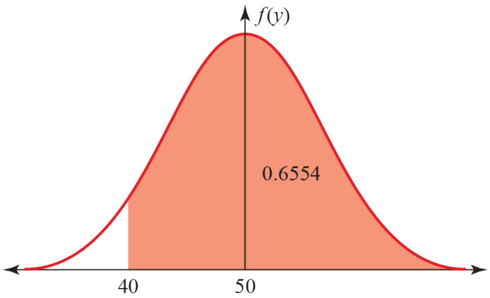
Using the inverse normal function, $z = -1.0364...$

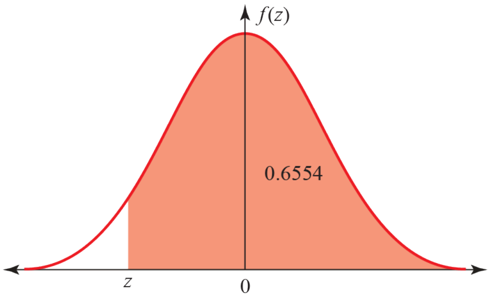
$$\text{so } -1.0364... = \frac{25-\mu}{\sqrt{40}}$$

$$\mu = \sqrt{40} \times (-1.0364...)$$

$$= 31.554... = 31.6 \text{ (3 s.f.)}$$

4 $\mathrm{P}(Y > 40) = 0.6554 \Rightarrow \mathrm{P}\left(Z > \dfrac{40-50}{\sigma}\right) = 0.6554$





Using the inverse normal function, $z = -0.3999...$

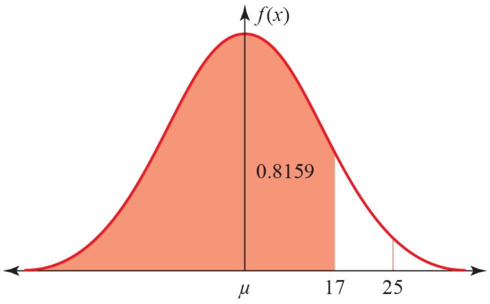
$$\text{so } -0.3999... = \frac{40-50}{\sigma}$$

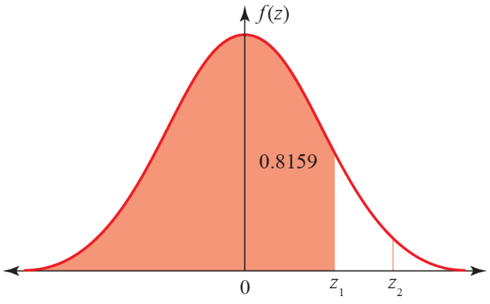
$$\sigma = \frac{10}{0.3999...} = 25.0 \text{ (3 s.f.)}$$

5 Using the inverse normal function,

$$\mathrm{P}(X < 17) = 0.8159 \Rightarrow \mathrm{P}\left(Z < \frac{17-\mu}{\sigma}\right) = 0.8159 \Rightarrow z_1 = 0.8998...$$

$$\mathrm{P}(X < 25) = 0.9970 \Rightarrow \mathrm{P}\left(Z < \frac{25-\mu}{\sigma}\right) = 0.9970 \Rightarrow z_2 = 2.7477...$$





So $0.8998\sigma = 17 - \mu$ **(1)**

and $2.7477\sigma = 25 - \mu$ **(2)**

(2) – **(1)**: $1.8479\sigma = 8$

$\sigma = 4.329...$

Substituting into (**2**):

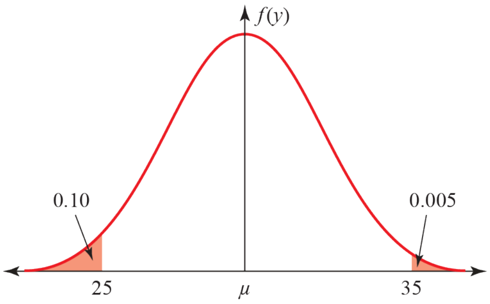
$\mu = 17 - 0.8998 \times 4.329... = 13.104...$

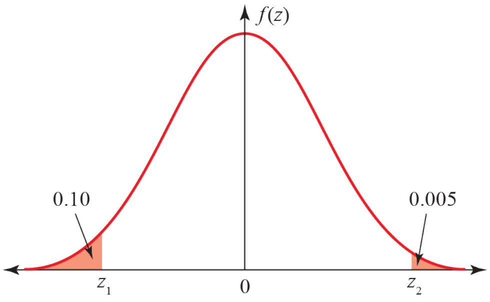
So $\mu = 13.1$ and $\sigma = 4.33$ (3 s.f.)

6 Using the inverse normal function (or the percentage points table),

$$\mathrm{P}(Y < 25) = 0.10 \Rightarrow \mathrm{P}\left(Z < \frac{25-\mu}{\sigma}\right) = 0.10 \Rightarrow z_1 = -1.28155...$$

$$\mathrm{P}(Y > 35) = 0.005 \Rightarrow \mathrm{P}\left(Z > \frac{35-\mu}{\sigma}\right) = 0.005 \Rightarrow z_2 = 2.57582...$$





So $-1.2816\sigma = 25 - \mu$ **(1)**

and $2.5758\sigma = 35 - \mu$ **(2)**

(2) – **(1)**: $3.8574\sigma = 10$

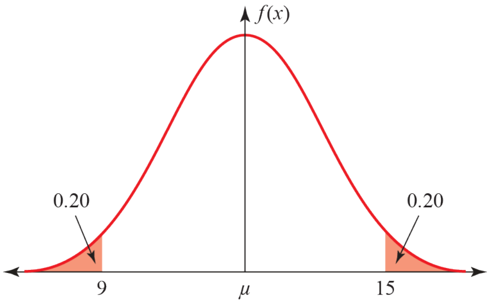
$\sigma = 2.5924...$

Substituting into (**2**):

$\mu = 35 - 2.5758 \times 2.5924... = 28.322...$

So $\mu = 28.3$ and $\sigma = 2.59$ (3 s.f.)

7



By symmetry, $\mu = \frac{1}{2}(9 + 15) = 12$

$$\mathrm{P}(X > 15) = 0.20 \Rightarrow \mathrm{P}\left(Z > \frac{15 - 12}{\sigma}\right) = 0.20$$

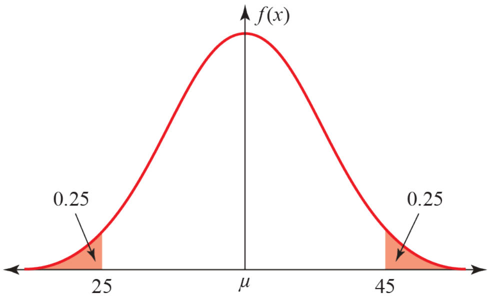
Using the inverse normal function (or the percentage points table), $z = 0.8416...$

$$\text{so } 0.8416 = \frac{3}{\sigma}$$

$$\sigma = \frac{3}{0.8416} = 3.564...$$

So $\mu = 12$ and $\sigma = 3.56$ (3 s.f.)

8



By symmetry, $\mu = \frac{1}{2}(25 + 45) = 35$

$$\mathrm{P}(X > 45) = 0.25 \Rightarrow \mathrm{P}\left(Z > \frac{45 - 35}{\sigma}\right) = 0.25$$

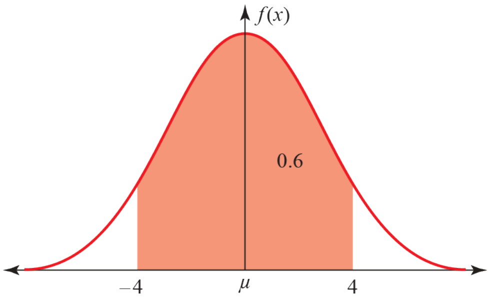
Using the inverse normal function, $z = 0.6744...$

$$\text{so } 0.6744 = \frac{10}{\sigma}$$

$$\sigma = \frac{10}{0.6744} = 14.82...$$

So $\mu = 35$ and $\sigma = 14.8$ (3 s.f.)

9 $\mu = 0$ (given) so $\mathrm{P}(X > 4) = 0.2$ and $\mathrm{P}\left(Z > \frac{4-0}{\sigma}\right) = 0.2$



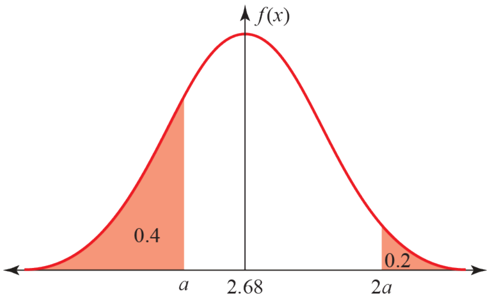
Using the inverse normal function (or the percentage points table), $z = 0.8416...$

so $0.8416 = \frac{4}{\sigma}$

$$\sigma = \frac{4}{0.8416} = 4.752...$$

So $\sigma = 4.75$ (3 s.f.)

10



Using the inverse normal function (or the percentage points table),

$\mathrm{P}(X > 2a) = 0.2 \Rightarrow \mathrm{P}\left(Z < \frac{2a-2.68}{\sigma}\right) = 0.2 \Rightarrow z_1 = 0.8416...$

$\mathrm{P}(X < a) = 0.4 \Rightarrow \mathrm{P}\left(Z < \frac{a-2.68}{\sigma}\right) = 0.4 \Rightarrow z_2 = -0.2533...$

So $0.8416\sigma = 2a - 2.68$ **(1)**

and $-0.2533\sigma = a - 2.68$ **(2)**

(2) $\times 2$: $-0.5066\sigma = 2a - 5.36$ **(3)**

(1) − **(3)**: $1.3482\sigma = 2.68$

$$\sigma = 1.9878...$$

Substituting into (**2**):

$a = 2.68 - 0.2533 \times 1.9878... = 2.176...$

So $\sigma = 1.99$ and $a = 2.18$ (3 s.f.)

11 a The distribution is $D \sim \text{N}(\mu, 5^2)$.

$$\text{P}(D > 200) = 0.75 \Rightarrow \text{P}(D < 200) = 0.25 \Rightarrow \text{P}\left(Z < \frac{200-\mu}{5}\right) = 0.25$$

Using the inverse normal function, $z = -0.6744...$

$$\text{so } -0.6744... = \frac{200-\mu}{5}$$

$$\mu = 200 + 5 \times 0.6744...$$

$$= 203.37... = 203\,\text{mm (3 s.f.)}$$

b $\text{P}(204 < D < 206) = \text{P}(D < 206) - \text{P}(D < 204)$

$$= \text{P}\left(Z < \frac{206 - 203.37...}{5}\right) - \text{P}\left(Z < \frac{204 - 203.37...}{5}\right)$$

$$= \text{P}(Z < 0.5256) - \text{P}(Z < 0.1256)$$

$$= 0.70041... - 0.54997...$$

$$= 0.15045 = 0.1504 \text{ (4 d.p.)}$$

c $\text{P}(D > 205) = \text{P}(Z > 0.3256) = 1 - 0.62763... = 0.37237...$
So the probability that all three bowls are greater than 205 mm in diameter is $\text{P}(D > 205)^3$,
i.e. $(0.37237...)^3 = 0.05163... = 0.0516$ (3 s.f.).

12 a The distribution is $T \sim \text{N}(2.5, \sigma)$.

$$\text{P}(T < 2.55) = 0.65 \Rightarrow \text{P}\left(Z < \frac{2.55 - 2.5}{\sigma}\right) = 0.65$$

Using the inverse normal function, $z = 0.38532...$

$$\text{so } \frac{2.55 - 2.5}{\sigma} = 0.38532...$$

$$0.05 = \sigma \times 0.38532...$$

$$\sigma = 0.12976... = 0.1298 \text{ (4 d.p.)}$$

b $\text{P}(2.4 < T < 2.6) = \text{P}(T < 2.6) - \text{P}(T < 2.4)$ (4 s.f.).

$$= \text{P}\left(Z < \frac{2.6 - 2.5}{0.12976...}\right) - \text{P}\left(Z < \frac{2.4 - 2.5}{0.12976...}\right)$$

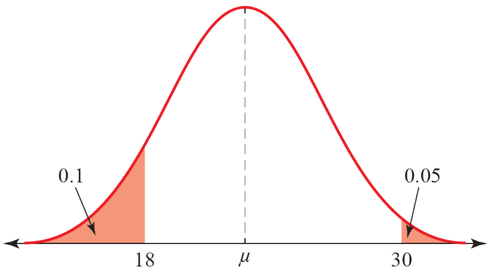
$$= \text{P}(Z < 0.77065...) - \text{P}(Z < -0.77065...)$$

$$= 0.77954... - 0.22045...$$

$$= 0.55908... = 0.5591 \text{ (4 d.p.)}$$

c Use the binomial distribution $X \sim \text{B}(20, 0.5591)$.
Using the binomial CD function,
$\text{P}(X \geqslant 15) = 1 - \text{P}(X \leqslant 14) = 1 - 93501... = 0.06497...$
So the probability that at least 15 table cloths can be sold is 0.0650 (4 d.p.)

13 a



b $P(M < 18) = 0.10 \Rightarrow P\left(Z < \frac{18-\mu}{\sigma}\right) = 0.10 \Rightarrow z_1 = -1.28155...$

$P(M > 30) = 0.05 \Rightarrow P\left(Z > \frac{30-\mu}{\sigma}\right) = 0.05 \Rightarrow z_2 = 1.64485...$

So $-1.28155... \times \sigma = 18 - \mu$ **(1)**

and $1.64485... \times \sigma = 30 - \mu$ **(2)**

(2) − **(1)**: $2.92640... \times \sigma = 12$

$\sigma = 4.10059...$

Substituting into **(2)**:

$\mu = 30 - 1.64485... \times 4.10059... = 23.25512...$

So $\mu = 23.26$ and $\sigma = 4.101$ (4 s.f.)

c $P(M > 25) = 1 - P\left(Z > \frac{25 - 23.25512...}{4.10059...}\right) = 0.33522...$

Use the binomial distribution $X \sim B(10, 0.33522...)$

Using the binomial CD function,

$P(X \geqslant 4) = 1 - P(X \leqslant 3) = 1 - 0.55408... = 0.44591...$

So the probability that at least 4 have a mass greater than 25 kg is 0.4459 (4 d.p.)

14 a $P(L < 16) = 0.20 \Rightarrow P\left(Z < \frac{16-\mu}{\sigma}\right) = 0.20 \Rightarrow z_1 = -0.84162...$

$P(L > 18) = 0.10 \Rightarrow P\left(Z > \frac{18-\mu}{\sigma}\right) = 0.10 \Rightarrow z_2 = 1.28155...$

So $-0.84162... \times \sigma = 16 - \mu$ **(1)**

and $1.28155... \times \sigma = 18 - \mu$ **(2)**

(2) − **(1)**: $2.12317... \times \sigma = 2$

$\sigma = 0.94198...$

Substituting into **(2)**:

$\mu = 18 - 1.28155... \times 0.94198... = 16.79279...$

So $\mu = 16.79$ and $\sigma = 0.9420$ (4 s.f.)

b $P(L < Q_1) = 0.25 \Rightarrow z_1 = -0.67448... \Rightarrow \frac{Q_1 - 16.79279...}{0.94198...} = -0.67448... \Rightarrow Q_1 = 16.15743...$

$P(L < Q_3) = 0.75 \Rightarrow z_2 = 0.67448... \Rightarrow \frac{Q_3 - 16.79279...}{0.94198...} = 0.67448... \Rightarrow Q_3 = 17.42814...$

The interquartile range is $Q_3 - Q_1 = 17.42814... - 16.15743... = 1.2701... = 1.27$ (2 d.p.)

Challenge

a Let the quartiles be $Q_3 = \mu + z\sigma$ and $Q_1 = \mu + z\sigma$.
Then the interquartile range is $Q_3 - Q_1 = q = (\mu + z\sigma) - (\mu - z\sigma) = 2z\sigma$
z is such that $\Phi(z) = 0.75$ so, using the inverse normal function, $z = 0.67448...$
So $q = 2 \times 0.67448... \times \sigma = 1.34987... \times \sigma$
and hence $\sigma = 0.74130... \times p = 0.741p$ (3 s.f.)

b Since $q = (\mu + z\sigma) - (\mu - z\sigma) = 2z\sigma$ (i.e. the μ s cancel), q is not dependent on μ.
So it is not possible to write an equation for q in terms of μ, and vice versa.

The normal distribution 3F

1 **a** **i** Yes, since $n = 120$ is large (> 50) and $p = 0.6$ is close to 0.5.

ii $\mu = np = 120 \times 0.6 = 72$ and $\sigma = \sqrt{np(1-p)} = \sqrt{72 \times 0.4} = \sqrt{28.8} = 5.37$ (3 s.f.)
$X \sim \mathrm{B}(72, 5.37^2)$

b **i** No, $n = 6$ is not large enough (< 50).

c **i** Yes, since $n = 250$ is large (> 50) and $p = 0.52$ is close to 0.5.

ii $\mu = np = 250 \times 0.52 = 130$ and $\sigma = \sqrt{np(1-p)} = \sqrt{130 \times 0.48} = \sqrt{62.4} = 7.90$ (3 s.f.)
$X \sim \mathrm{B}(130, 7.90^2)$

d **i** No, $p = 0.98$ is too far from 0.5.

e **i** Yes, since $n = 400$ is large (> 50) and $p = 0.48$ is close to 0.5.

ii $\mu = np = 400 \times 0.48 = 192$ and $\sigma = \sqrt{np(1-p)} = \sqrt{192 \times 0.52} = \sqrt{99.84} = 9.99$ (3 s.f.)
$X \sim \mathrm{B}(192, 9.99^2)$

f **i** Yes, since $n = 1000$ is large (> 50) and $p = 0.58$ is close to 0.5.

ii $\mu = np = 1000 \times 0.58 = 580$ and $\sigma = \sqrt{np(1-p)} = \sqrt{580 \times 0.42} = \sqrt{243.6} = 15.6$ (3 s.f.)
$X \sim \mathrm{B}(580, 15.6^2)$

2 A normal approximation is valid since $n = 150$ is large (> 50) and $p = 0.45$ is close to 0.5.
$\mu = np = 150 \times 0.45 = 67.5$ and $\sigma = \sqrt{np(1-p)} = \sqrt{67.5 \times 0.55} = \sqrt{37.125} = 6.093$ (4 s.f.)

a $\mathrm{P}(X \leqslant 60) \approx \mathrm{P}(Y < 60.5) = 0.1253$ (4 s.f.)

b $\mathrm{P}(X > 75) \approx \mathrm{P}(Y > 75.5) = 0.0946$ (4 s.f.)

c $\mathrm{P}(65 \leqslant X \leqslant 80) \approx \mathrm{P}(64.5 < Y < 80.5) = 0.6723$ (4 s.f.)

3 A normal approximation is valid since $n = 200$ is large (> 50) and $p = 0.53$ is close to 0.5.
$\mu = np = 200 \times 0.53 = 106$ and $\sigma = \sqrt{np(1-p)} = \sqrt{106 \times 0.47} = \sqrt{49.82} = 7.058$ (4 s.f.)

a $\mathrm{P}(X < 90) \approx \mathrm{P}(Y < 89.5) = 0.0097$ (4 s.f.)

b $\mathrm{P}(100 \leqslant X < 110) \approx \mathrm{P}(99.5 < Y < 109.5) = 0.5115$ (4 s.f.)

c $\mathrm{P}(X = 105) \approx \mathrm{P}(104.5 < Y < 105.5) = 0.0559$ (4 s.f.)

4 A normal approximation is valid since $n = 100$ is large (> 50) and $p = 0.6$ is close to 0.5.
$\mu = np = 100 \times 0.6 = 60$ and $\sigma = \sqrt{np(1-p)} = \sqrt{60 \times 0.4} = \sqrt{24} = 4.899$ (4 s.f.)

a $\mathrm{P}(X > 58) \approx \mathrm{P}(Y > 58.5) = 0.6203$ (4 s.f.)

b $\mathrm{P}(60 < X \leqslant 72) \approx \mathrm{P}(60.5 < Y < 72.5) = 0.4540$ (4 s.f.)

c $\mathrm{P}(X = 70) \approx \mathrm{P}(69.5 < Y < 70.5) = 0.0102$ (4 s.f.)

5 Let X = number of heads in 70 tosses of a fair coin, so $X \sim \mathrm{B}(70, 0.5)$.
Since $p = 0.5$ and 70 is large, X can be approximated by the normal distribution $Y \sim \mathrm{N}(\mu,\sigma^2)$,
where $\mu = 70 \times 0.5 = 35$ and $\sigma = \sqrt{70 \times 0.5 \times 0.5} = \sqrt{17.5}$
So $Y \sim \mathrm{N}(35,17.5)$
$\mathrm{P}(X > 45) \approx \mathrm{P}(Y \geqslant 45.5) = 0.0060$

6 A normal approximation is valid since $n = 1200$ is large and p is close to 0.5.

$$\mu = np = 1200 \times \frac{50}{101} = 594.059$$

and $\sigma = \sqrt{np(1-p)} = \sqrt{594.059 \times \frac{51}{101}} = \sqrt{299.97059\ldots} = 17.32$ (4 s.f.)
So $Y \sim \mathrm{N}(594.059, 299.97\ldots)$
$\mathrm{P}(X \geqslant 600) \approx \mathrm{P}(Y > 599.5) = 0.3767$ (4 s.f.)

7 a The number of trials, n, must be large (> 50), and the success probability, p, must be close to 0.5.

b Using the binomial distribution, $\mathrm{P}(X = 10) = \binom{20}{10} \times 0.45^{10} \times 0.55^{10} = 0.1593$ (4 s.f.)

c A normal approximation is valid since $n = 240$ is large and $p = 0.45$ is close to 0.5.
$\mu = np = 240 \times 0.45 = 108$ and $\sigma = \sqrt{np(1-p)} = \sqrt{108 \times 0.55} = \sqrt{59.4} = 7.707$ (4 s.f.)
So $Y \sim \mathrm{N}(108, 59.4)$
$\mathrm{P}(X < 110) \approx \mathrm{P}(Y < 109.5) = 0.5772$ (4 s.f.)

d $\mathrm{P}(X \geqslant q) = 0.2 \Rightarrow \mathrm{P}(Y > (q - 0.5)) = 0.2$
Using the inverse normal function,
$\mathrm{P}(Y > (q - 0.5)) = 0.2 \Rightarrow q - 0.5 = 114.485 \Rightarrow q = 114.985$
So $q = 115$

8 a Using the cumulative binomial function with $N = 30$ and $p = 0.52$,
$\mathrm{P}(X < 17) = \mathrm{P}(X \leqslant 16) = 0.6277$ (4 s.f.)

b A normal approximation is valid since $n = 600$ is large and $p = 0.52$ is close to 0.5.
$\mu = np = 600 \times 0.52 = 312$ and $\sigma = \sqrt{np(1-p)} = \sqrt{312 \times 0.48} = \sqrt{149.76} = 12.24$ (4 s.f.)
So $Y \sim \mathrm{N}(312, 149.76)$
$\mathrm{P}(300 \leqslant X \leqslant 350) \approx \mathrm{P}(299.5 < Y < 350.5) = 0.8456$ (4 s.f.)

9 **a** Using the binomial distribution, $\mathrm{P}(X=55)=\binom{100}{55}\times 0.56^{55}\times 0.44^{45}=0.07838\ (4\text{ s.f.})$

b A normal approximation is valid since $n = 100$ is large and $p = 0.56$ is close to 0.5.

$\mu = np = 100\times 0.56 = 56$ and $\sigma=\sqrt{np(1-p)}=\sqrt{56\times 0.44}=\sqrt{24.64}=4.964\ (4\text{ s.f.})$

So $Y \sim \mathrm{B}(56, 24.64)$

$\mathrm{P}(X=55)\approx \mathrm{P}(54.5<Y<55.5)=0.07863\ (4\text{ s.f.})$

Percentage error $=\dfrac{0.07838...-0.07863...}{0.07838...}\times 100=-0.31\%\ (2\text{ d.p.})$

The normal distribution 3G

1 **a** $H_0 : \mu = 21,\ \ H_1 : \mu \neq 21$, two-tailed test with 2.5% in each tail.

Assume H_0, so that $X \sim N(21, 1.5^2)$ and $\overline{X} \sim \left(21, \frac{1.5^2}{20}\right)$ or $\overline{X} \sim (21, 0.0.3354^2)$

Using the cumulative normal function, $P(\overline{X} > 21.2) = 0.2755$ (4 d.p.)
$0.2755 > 0.025$ so not significant. Do not reject H_0.

b $H_0 : \mu = 100,\ \ H_1 : \mu < 100$, one-tailed test at the 5% level.

Assume H_0, so that $X \sim N(100, 5^2)$ and $\overline{X} \sim \left(100, \frac{5^2}{36}\right)$ or $\overline{X} \sim (100, 0.8333^2)$

$P(\overline{X} < 98.5) = 0.0359$ (4 d.p.)
$0.0359 < 0.05$ so significant. Do not reject H_0.

c $H_0 : \mu = 5,\ \ H_1 : \mu \neq 5$, two-tailed test with 2.5% in each tail.

Assume H_0, so that $X \sim N(5, 3^2)$ and $\overline{X} \sim \left(5, \frac{3^2}{25}\right)$ or $\overline{X} \sim (5, 0.6^2)$

$P(\overline{X} > 6.1) = 0.0334$ (4 d.p.)
$0.0334 > 0.025$ so not significant. Do not reject H_0.

d $H_0 : \mu = 15,\ \ H_1 : \mu > 15$, one-tailed test at the 1% level.

Assume H_0, so that $X \sim N(15, 3.5^2)$ and $\overline{X} \sim \left(15, \frac{3.5^2}{40}\right)$ or $\overline{X} \sim (5, 0.5534^2)$

$P(\overline{X} > 16.5) = 0.0034$ (4 d.p.)
$0.0034 < 0.01$ so significant. Reject H_0.

e $H_0 : \mu = 50,\ \ H_1 : \mu \neq 50$, two-tailed test with 0.5% in each tail.

Assume H_0, so that $X \sim N(50, 4^2)$ and $\overline{X} \sim \left(50, \frac{4^2}{60}\right)$ or $\overline{X} \sim (50, 0.5163^2)$

$P(\overline{X} < 48.9) = 0.0166$ (4 d.p.)
$0.0166 > 0.005$ so not significant. Do not reject H_0.

2 **a** $H_0 : \mu = 120,\ \ H_1 : \mu < 120$, one-tailed test at the 5% level.

Assume H_0, so that $X \sim N(120, 2^2)$ and $\overline{X} \sim \left(120, \frac{2^2}{30}\right)$

Let $Z = \dfrac{\overline{X} - 120}{\frac{2}{\sqrt{30}}}$

Using the inverse normal function, $P(Z < z) = 0.05 \Rightarrow z = -1.6449$

$$-1.6449 = \frac{\overline{X} - 120}{\frac{2}{\sqrt{30}}} \Rightarrow \overline{X} = 120 - 1.6449 \times \frac{2}{\sqrt{30}} = 119.39...$$

So the critical region is $\overline{X} < 119.39...$ or 119 (3 s.f.).

2 b $H_0 : \mu = 12.5,\ \ H_1 : \mu > 12.5$, one-tailed test at the 1% level.

Assume H_0, so that $X \sim N(12.5, 1.5^2)$ and $\overline{X} \sim \left(12.5, \frac{1.5^2}{25}\right)$

Let $Z = \dfrac{\overline{X} - 12.5}{\frac{1.5}{\sqrt{25}}} = \dfrac{\overline{X} - 12.5}{\frac{1.5}{5}}$

$P(Z > z) = 0.01 \Rightarrow z = 2.3263$

$2.3263 = \dfrac{\overline{X} - 12.5}{\frac{1.5}{5}} \Rightarrow \overline{X} = 12.5 + 2.3263 \times \dfrac{3}{10} = 13.197...$

So the critical region is $\overline{X} > 13.197...$ or 13.2 (3 s.f.) (3 s.f.).

c $H_0 : \mu = 85,\ \ H_1 : \mu > 85$, one-tailed test at the 10% level.

Assume H_0, so that $X \sim N(85, 4^2)$ and $\overline{X} \sim \left(85, \frac{4^2}{50}\right)$

Let $Z = \dfrac{\overline{X} - 85}{\frac{4}{\sqrt{50}}}$

$P(Z < z) = 0.10 \Rightarrow z = -1.2816$

$-1.2816 = \dfrac{\overline{X} - 85}{\frac{4}{\sqrt{50}}} \Rightarrow \overline{X} = 85 - 1.2816 \times \dfrac{4}{\sqrt{50}} = 84.275...$

So the critical region is $\overline{X} < 84.275...$ or 84.3 (3 s.f.).

d $H_0 : \mu = 0,\ \ H_1 : \mu \neq 0$, two-tailed test with 2.5% in each tail.

Assume H_0, so that $X \sim N(0, 3^2)$ and $\overline{X} \sim \left(0, \frac{3^2}{45}\right)$

Let $Z = \dfrac{\overline{X} - 0}{\frac{3}{\sqrt{45}}}$

$P(Z < z) = 0.025 \Rightarrow z = -1.9600 \Rightarrow \overline{X} = -1.96 \times \dfrac{3}{\sqrt{45}} = -0.8765...$

$P(Z > z) = 0.025 \Rightarrow z = 1.9600 \Rightarrow \overline{X} = 1.96 \times \dfrac{3}{\sqrt{45}} = 0.8765...$

So the critical region is $\overline{X} < -0.877$ or $\overline{X} > 0.877$ (3 s.f.).

e $H_0 : \mu = -8,\ \ H_1 : \mu \neq -8$, two-tailed test with 0.5% in each tail.

Assume H_0, so that $X \sim N(-8, 1.2^2)$ and $\overline{X} \sim \left(-8, \frac{1.2^2}{20}\right)$

Let $Z = \dfrac{\overline{X} - (-8)}{\frac{1.2}{\sqrt{20}}}$

$P(Z < z) = 0.005 \Rightarrow z = -2.5758 \Rightarrow \overline{X} = -8 - 2.5758 \times \dfrac{1.2}{\sqrt{20}} = -8.6911...$

$P(Z > z) = 0.005 \Rightarrow z = 2.5758 \Rightarrow \overline{X} = -8 + 2.5758 \times \dfrac{1.2}{\sqrt{20}} = -7.3088...$

So the critical region is $\overline{X} < -8.69$ or $\overline{X} > -7.31$ (3 s.f.).

3 $\sigma = 15,\ n = 25,\ \overline{x} = 179$

$\mathrm{H}_0 : \mu = 185$ (no improvement), $\mathrm{H}_1 : \mu < 185$ (shorter time)**,** one-tailed test at the 5% level.

Assume H_0, so that $X \sim \mathrm{N}(185, 15^2)$ and $\overline{X} \sim \left(15, \dfrac{15^2}{25}\right)$ or $\overline{X} \sim \left(15, 3^2\right)$

Using the cumulative normal function, $\mathrm{P}(\overline{X} < 179) = 0.0227$ (4 d.p.)

$0.0227 < 0.05$ so significant, so reject H_0.

There is evidence that the new formula is an improvement.

4 a The psychologist wishes to test whether the score has increased (not just changed). Therefore $\mathrm{H}_0 : \mu = 100,\ \ \mathrm{H}_1 : \mu > 100$, one-tailed test at the 2.5% level.

Assume H_0, so that $X \sim \mathrm{N}(100, 15^2)$ and $\overline{X} \sim \left(100, \dfrac{15^2}{80}\right)$ or $\overline{X} \sim (100, 1.6771^2)$

Using the inverse normal function, $\mathrm{P}(\overline{X} > \overline{x}) = 0.025 \Rightarrow \overline{x} = 103.287\ldots$

So the critical region is $\overline{X} > 103.287\ldots$ or 103.29 (3 s.f.).

b $102.5 < 103.29$ so there is not sufficient evidence to reject H_0, i.e. there is not sufficient evidence to say, at the 2.5% level, that eating chocolate before taking an IQ test improves the result.

5 a $\sigma = 0.15,\ n = 30,\ \overline{x} = 8.95$

$\mathrm{H}_0 : \mu = 9$ (no change), $\mathrm{H}_1 : \mu \neq 9$ (change in mean diameter)

Two-tailed test with 2.5% in each tail

Assume H_0, so that $X \sim \mathrm{N}(9.0, 0.15^2)$ and $\overline{X} \sim \left(9.0, \dfrac{0.15^2}{30}\right)$ or $\overline{X} \sim \left(9.0, 0.0274^2\right)$

Using the cumulative normal function, $\mathrm{P}(\overline{X} < 8.95) = 0.033944\ldots = 0.0340$ (4 d.p.)

$0.0340 > 0.025$ so not significant, so do not reject H_0.

There is not enough evidence to conclude that there has been a change in the mean diameter.

b Two-tailed test so double to probability to find the p-value

p-value $= 0.033944\ldots \times 2 = \ 0.0678$ (4 d.p.)

6 a First find the mean of the distribution.

$\mathrm{P}(D > 5.62) = 0.05$

Using the inverse normal function (or the percentage points table), $p = 0.05 \Rightarrow z = 1.6449$

Using the formula $z = \dfrac{x - \mu}{\sigma}$, $\dfrac{5.62 - \mu}{0.1} = 1.6449 \Rightarrow 5.62 - \mu = 0.16449 \Rightarrow \mu = 5.4555$

The probability that a randomly chosen bolt can be sold is $\mathrm{P}(5.1 \leqslant D \leqslant 5.6)$

Using the cumulative normal function, $\mathrm{P}(5.1 \leqslant D \leqslant 5.6) = 0.92558\ldots$

So the probability that a randomly chosen bolt can be sold is 0.9256 (4 d.p.).

b Use the binomial distribution $N \sim \mathrm{B}(12, 1 - 0.9256)$ or $N \sim \mathrm{B}(12, 1 - 0.0744)$.

Using the cumulative binomial function, $\mathrm{P}(N < 3) = \mathrm{P}(N \leqslant 2) = 0.94549\ldots$

So the probability that fewer than three cannot be sold is 0.9455 (4 d.p.).

6 c Test to determine whether the mean diameter is less than 5.7 mm. Therefore $H_0 : \mu = 5.7$, $H_1 : \mu < 5.7$, one-tailed test at the 2.5% level.

Assume H_0, so that $Y \sim N(5.7, 0.08^2)$ and $\overline{Y} \sim \left(5.7, \frac{0.08^2}{10}\right)$ or $\overline{Y} \sim (5.7, 0.025298^2)$

Using the cumulative normal function, $P(\overline{Y} < 5.65) = 0.02405... < 0.025$ (one-tailed) so reject H_0.

There is sufficient evidence to suggest that the mean diameter is less than 5.7 mm.

7 a $P(M > 160) = 0.025$

Using the inverse normal function (or the percentage points table), $p = 0.025 \Rightarrow z = 1.9599$

Using the formula $z = \frac{x - \mu}{\sigma}$, $\frac{160 - \mu}{12} = 1.9599 \Rightarrow 160 - \mu = 23.52 \Rightarrow \mu = 136.48$

So the mean mass of a European water vole is 136.48 g (2 d.p.).

b Using the cumulative normal function, $P(M > 150) = 0.1299$

Use the binomial distribution $N \sim B(8, 0.1299)$.

Using the cumulative binomial function, $P(N \geqslant 4) = 1 - P(N \leqslant 3) = 1 - 0.98708... = 0.01291...$

The probability that at least 4 voles have a mass greater than 150 g is 0.0129 (4 d.p.).

7 c Test to determine whether the mean mass is different from 860 grams. Therefore $H_0 : \mu = 860$, $H_1 : \mu \neq 860$, two-tailed test with 5% in each tail.

Assume H_0, so that $N \sim N(860, 85^2)$ and $\overline{N} \sim \left(860, \frac{85^2}{15}\right)$ or $\overline{N} \sim (860, 21.946^2)$

Using the cumulative normal function, $P(\overline{N} > 875) = 0.24715...$

$0.24715 > 0.05$ so not significant, do not reject H_0.

There is insufficient evidence to suggest that the mean mass of all water rats is different from 860 g.

8 Test to determine whether the daily mean windspeed is greater than 9.5 knots. Therefore $H_0 : \mu = 9.5$, $H_1 : \mu > 9.5$, one-tailed test at the 2.5% level.

Assume H_0, so that $X \sim N(9.5, 3.1^2)$ and $\overline{X} \sim \left(9.5, \frac{3.1^2}{25}\right)$ or $\overline{X} \sim (9.5, 0.62^2)$

Using the inverse normal function, $P(\overline{X} > \overline{x}) = 0.025 \Rightarrow \overline{x} = 10.715...$

So the critical region is $\overline{X} \geqslant 10.715...$

$12.2 > 10.715$ so there is sufficient evidence to reject H_0, i.e. there is sufficient evidence to say, at the 2.5% level, that the daily mean windspeed is greater than 9.5 knots.

The normal distribution Mixed exercise 3

1 $H \sim \mathrm{N}(178, 4^2)$

a Using the normal CD function, $\mathrm{P}(H > 185) = 0.04059... = 0.0401$ (4 d.p.)

b Using the normal CD function, $\mathrm{P}(H < 180) = 0.69146...$
The probability that three men, selected at random, all satisfy this criterion is
$\mathrm{P}(H < 180)^3 = 0.33060... = 0.3306$ (4 d.p.).

c Using the inverse normal function, $\mathrm{P}(H > h) = 0.005 \Rightarrow h = 188.03...$
To the nearest centimetre, the height of a door frame needs to be at least 188 cm.

2 $W \sim \mathrm{N}(32.5,\ 2.2^2)$

a Using the normal CD function, $\mathrm{P}(W < 30) = 0.12790...$
The percentage of sheets weighing less than 30kg is 12.8% (3 s.f.).

b Using the normal CD function, $\mathrm{P}(31.6 < W < 34.8) = 0.51085...$
So 51.1% of sheets satisfy Bob's requirements.

3 $T \sim \mathrm{N}(48,\ 8^2)$

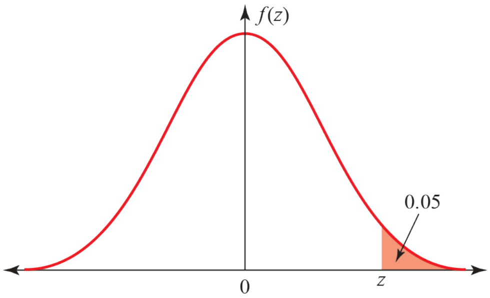
a Using the normal CD function, $\mathrm{P}(T > 60) = 0.06680...$
The probability that a battery will last for more than 60 hours is 0.0668 (4 d.p.).

b Using the normal CD function, $\mathrm{P}(T < 35) = 0.05208...$
The probability that a battery will last for less than 35 hours is 0.0521 (4 d.p.).

c Use the binomial distribution $X \sim \mathrm{B}(30, 0.05208\ldots)$
Using the binomial CD function, $\mathrm{P}(X \leqslant 3) = 0.93145...$
The probability that three or fewer last less than 35 hours is 0.9315 (4 d.p.).

4 $X \sim \mathrm{N}(24, \sigma^2)$

a $\mathrm{P}(X > 30) = 0.05 \Rightarrow \mathrm{P}\left(Z > \frac{30-\mu}{\sigma}\right) = 0.05$



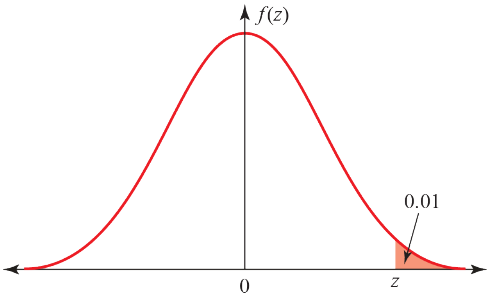
Using the inverse normal function, $z = -1.64485...$

$$\text{so } 1.64485... = \frac{30-24}{\sigma}$$

$$\sigma = \frac{6}{1.64485...} = 3.647... = 3.65 \text{ (3 s.f.)}$$

b Using the normal CD function, $\mathrm{P}(X < 20) = 0.13636... = 0.136$ (3 d.p.)

c $\mathrm{P}(X > d) = 0.01 \Rightarrow \mathrm{P}\left(Z > \frac{d-\mu}{\sigma}\right) = 0.01$



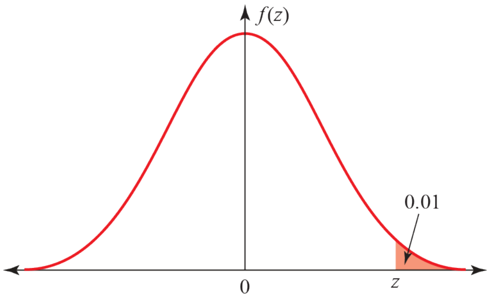
Using the inverse normal function, $z = 2.32634...$

$$\text{so } 2.32634... = \frac{d-24}{3.647...}$$

$$d = 32.485... = 32.5 \text{ (3 s.f.)}$$

5 $L \sim \mathrm{N}(120,\ \sigma^2)$

a $\mathrm{P}(L > 140) = 0.01 \Rightarrow \mathrm{P}\left(Z > \frac{140 - \mu}{\sigma}\right) = 0.01$



Using the inverse normal function, $z = 2.32634...$

$$\text{so } 2.32634... = \frac{140 - 120}{\sigma}$$

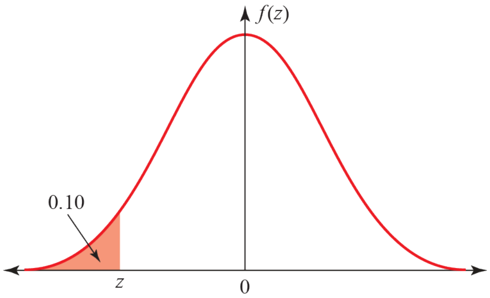
$$\sigma = \frac{20}{2.32634...} = 8.59716...$$

So the standard deviation of the volume dispensed is 8.60 ml (3 s.f.).

b Using the normal CD function, $\mathrm{P}(L < 110) = 0.12237...$

The probability that the machine dispenses less than 110ml is 0.122 (3 s.f.).

c $\mathrm{P}(L < c) = 0.10 \Rightarrow \mathrm{P}\left(Z < \frac{c - \mu}{\sigma}\right) = 0.10$



Using the inverse normal function, $z = -1.28155...$

$$\text{so } -1.2816 = \frac{c - 120}{8.59716...}$$

$$c = 108.982...$$

To the nearest millilitre, the largest volume leading to a refund is 109 ml.

6 **a** $\mathrm{P}(X < 20) = 0.25$ and $\mathrm{P}(X < 40) = 0.75$

Using the inverse normal function (or the percentage points table),

$$\mathrm{P}(X < 20) = 0.25 \Rightarrow \mathrm{P}\left(Z < \frac{20-\mu}{\sigma}\right) = 0.25 \Rightarrow z_1 = -0.67448...$$

$$\mathrm{P}(X < 40) = 0.75 \Rightarrow \mathrm{P}\left(Z < \frac{40-\mu}{\sigma}\right) = 0.75 \Rightarrow z_2 = 0.67448...$$

So $-0.6745\sigma = 20 - \mu$ **(1)**

and $0.6745\sigma = 40 - \mu$ **(2)**

(2) − **(1)**: $1.3489\sigma = 20$

$\sigma = 14.826...$

Substituting into **(2)**:

$\mu = 40 - 0.6745 \times 14.826... = 29.99...$

So $\mu = 30$ and $\sigma = 14.8$ (3 s.f.)

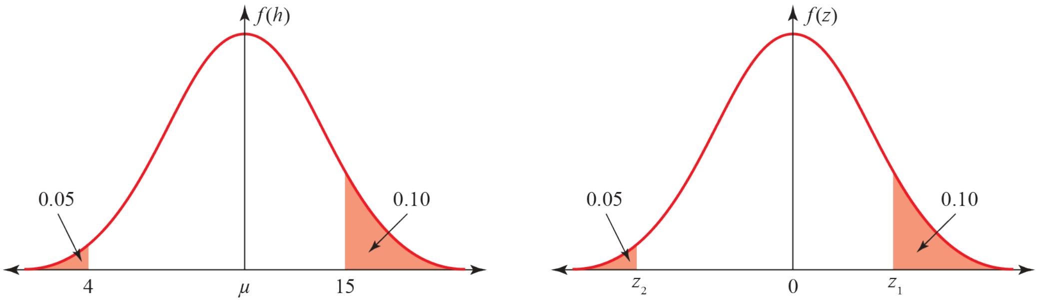
b Using the inverse normal CD function with $\mu = 30$ and $\sigma = 14.826...$,

$\mathrm{P}(X < a) = 0.1 \Rightarrow a = 10.999...$ and $\mathrm{P}(X < b) = 0.9 \Rightarrow b = 49.000...$

So the 10% to 90% interpercentile range is $49.0 - 11.0 = 38.0$

7 $$\mathrm{P}(H > 15) = 0.10 \Rightarrow \mathrm{P}\left(Z > \frac{15-\mu}{\sigma}\right) = 0.10 \Rightarrow z_1 = 1.28155...$$

$$\mathrm{P}(H < 4) = 0.05 \Rightarrow \mathrm{P}\left(Z > \frac{4-\mu}{\sigma}\right) = 0.05 \Rightarrow z_2 = -1.64485...$$



So $-1.6449\sigma = 4 - \mu$

$1.2816\sigma = 15 - \mu$

Subtract $2.9265\sigma = 11$

$\sigma = 3.7587... = 3.76\,\text{cm}$ (3 s.f.)

$\mu = 15 - 1.2816\sigma = 10.2\,\text{cm}$

8 **a** $T \sim \mathrm{N}(80,\ 10^2)$

Using the normal CD function, $\mathrm{P}(T > 85) = 0.30853... = 0.3085$ (4 d.p.)

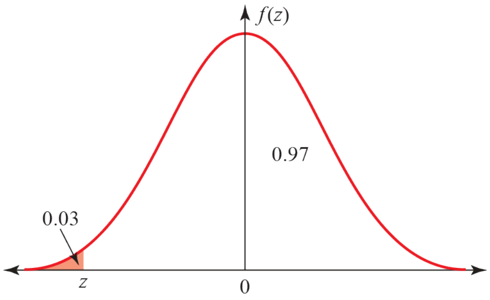
b $S \sim \mathrm{N}(100,\ 15^2)$

Using the normal CD function, $\mathrm{P}(S > 105) = 0.36944... = 0.3694$ (4 d.p.)

c The student's score on the first test was better, since fewer of the students got this score or higher.

9 $J \sim \text{N}(108,\ \sigma^2)$

a $\text{P}(J < 100) = 0.03 \Rightarrow \text{P}\left(Z < \dfrac{100-\mu}{\sigma}\right) = 0.03$



Using the inverse normal function, $z = -1.88079...$

so $-1.88079... = \dfrac{100-108}{\sigma}$

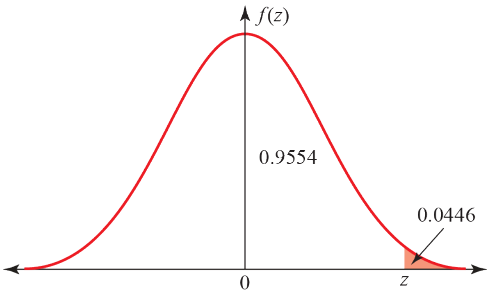
$$\sigma = 4.2535... = 4.25\text{ g } (3\text{ s.f.}).$$

The standard deviation is 4.25 g (3 s.f.).

b Using the normal CD function, $\text{P}(J > 115) = 0.0499... = 0.050$ (3 d.p.)

c Use the binomial distribution $X \sim \text{B}(25, 0.05)$
Using the binomial CD function,
$\text{P}(X \leqslant 2) = 0.87289... = 0.8729$ (4 d.p.)

10 $T \sim \text{N}(\mu, 3.8^2)$ and $\text{P}(T > 15) = 0.0446$



a $\text{P}(T > 15) = 0.0446 \Rightarrow \text{P}\left(Z > \dfrac{X-\mu}{\sigma}\right) = 0.0446 \Rightarrow z = 1.70$

so $1.70 = \dfrac{15-\mu}{3.8}$

$$\mu = 15 - 3.8 \times 1.70$$
$$= 8.54 \text{ minutes } (3 \text{ s.f.})$$

b $\text{P}(T < 5) = \text{P}\left(Z < \dfrac{5-8.54}{3.8}\right)$

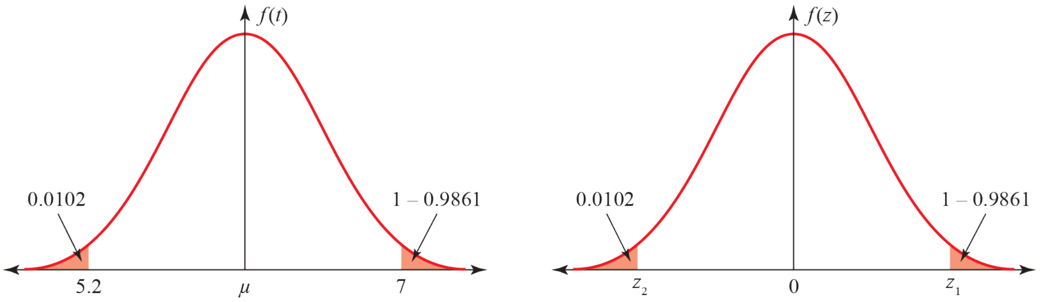
$$= \text{P}(Z < -0.93...)$$
$$= 0.17577... = 0.1758 \text{ (4 d.p.)}$$

11 $T \sim \mathrm{N}(\mu, \sigma^2)$

Using the inverse normal function,

$$\mathrm{P}(T < 7) = 0.9861 \Rightarrow \mathrm{P}\left(Z < \frac{7-\mu}{\sigma}\right) = 0.9861 \Rightarrow z_1 = 2.20009...$$

$$\mathrm{P}(T < 5.2) = 0.0102 \Rightarrow \mathrm{P}\left(Z < \frac{5.2-\mu}{\sigma}\right) = 0.0102 \Rightarrow z_2 = -2.31890...$$



So $2.2001\sigma = 7 - \mu$ **(1)**

and $-2.3189\sigma = 5.2 - \mu$ **(2)**

(1) − **(2)**: $4.5190\sigma = 1.8$

$\sigma = 0.3983...$

Substituting into **(1)**:

$\mu = 7 - 2.2001 \times 0.3983... = 6.123...$

So the mean thickness of the shelving is 6.12 mm and the standard deviation is 0.398 mm (3 s.f.).

12 Let X = number of heads in 60 tosses of a fair coin, so $X \sim \mathrm{B}(60, 0.5)$.

Since $p = 0.5$ and 60 is large, X can be approximated by the normal distribution $Y \sim \mathrm{N}(\mu, \sigma^2)$, where $\mu = 60 \times 0.5 = 30$ and $\sigma = \sqrt{60 \times 0.5 \times 0.5} = \sqrt{15}$

So $Y \sim \mathrm{N}(30, 15)$

$\mathrm{P}(X < 25) \approx \mathrm{P}(Y < 24.5) = 0.07779... = 0.0778$ (3 s.f.)

13 a The distribution is binomial, $\mathrm{B}(100, 0.40)$.

The binomial distribution can be approximated by the normal distribution when n is large (> 50) and p is close to 0.5. Here $n = 100$ and $p = 0.4$ so both of these conditions are satisfied.

b $\mu = np = 100 \times 0.4 = 40$ and $\sigma = \sqrt{np(1-p)} = \sqrt{40 \times 0.6} = \sqrt{24} = 4.899$ (4 s.f.)

c $\mathrm{P}(X \geqslant 50) \approx \mathrm{P}(Y \geqslant 49.5) = 0.02623... = 0.0262$ (3 s.f.)

14 a $\mathrm{P}(X = 65) = \binom{120}{65} \times 0.46^{65} \times 0.54^{55} = 0.01467$ (4 s.f.) or 0.0147 (4 d.p.)

b The distribution satisfies the two necessary conditions for an approximation by a normal distribution to be valid: $n = 120$ is large (> 50) and $p = 0.46$ is close to 0.5.

$\mu = np = 120 \times 0.46 = 55.2$ and $\sigma = \sqrt{np(1-p)} = \sqrt{55.2 \times 0.54} = \sqrt{29.808} = 5.460$ (4 s.f.)

14 c $Y \sim N(55.2, 5.460^2)$

Using the normal CD function, $P(X = 65) \approx P(64.5 < Y < 65.5) = 0.01463...$

Percentage error $= \frac{0.01467 - 0.01463}{0.01467} \times 100 = \frac{0.00004}{0.01467} \times 100 = 0.27\%$

15 a The distribution satisfies the two necessary conditions for an approximation by a normal distribution to be valid: $n = 300$ is large (> 50) and $p = 0.6$ is close to 0.5.

b $\mu = np = 300 \times 0.6 = 180$ and $\sigma = \sqrt{np(1-p)} = \sqrt{180 \times 0.4} = \sqrt{72} = 8.485$ (4 s.f.)

So $Y \sim N(180, 8.485^2)$

$P(150 < Y \leqslant 180) \approx P(150.5 < N < 180.5) = 0.52324... = 0.5232$ (4 s.f.)

c Using the inverse normal distribution, $P(N < a) = 0.05 \Rightarrow a = 166.04$

So $P(N < 166.5) > 0.05$ and $P(N < 165.5) < 0.05$

So $165.5 < y < 166.5$, i.e. the smallest value of y such that $P(Y < y) < 0.05$ is $y = 166$.

16 The distribution satisfies the two necessary conditions for an approximation by a normal distribution to be valid: $n = 80$ is large (> 50) and $p = 0.4$ is close to 0.5.

$\mu = np = 80 \times 0.4 = 32$ and $\sigma = \sqrt{np(1-p)} = \sqrt{32 \times 0.6} = \sqrt{19.2} = 4.382$ (4 s.f.)

So $Y \sim N(32, 4.382^2)$

$P(X > 30) \approx P(Y > 30.5) = 0.63394... = 0.6339$ (4 s.f.)

17 a Use the binomial distribution $X \sim B(20, 0.55)$

Using the binomial CD function, $P(X > 10) = 1 - P(X \leqslant 10) = 1 - 0.40863... = 0.5914$ (4 s.f.)

b A normal approximation is valid since $n = 200$ is large (> 50) and $p = 0.55$ is close to 0.5.

$\mu = np = 200 \times 0.55 = 110$ and $\sigma = \sqrt{np(1-p)} = \sqrt{110 \times 0.45} = \sqrt{49.5} = 7.036$ (4 s.f.)

So $Y \sim N(110, 7.036^2)$

$P(X \leqslant 95) \approx P(Y < 95.5) = 0.01965... = 0.0197$ (3 s.f.)

c It seems unlikely that the company's claim is correct: if it were true, the chance of only 95 (or fewer) seedlings producing apples from a sample of 200 seedlings would be less than 2%.

18 a Use the binomial distribution $X \sim B(25, 0.52)$

Using the binomial CD function, $P(X > 12) = 1 - P(X \leqslant 12) = 1 - 0.41992... = 0.5801$ (4 s.f.)

b A normal approximation is valid since $n = 300$ is large (> 50) and $p = 0.52$ is close to 0.5.

$\mu = np = 300 \times 0.52 = 156$ and $\sigma = \sqrt{np(1-p)} = \sqrt{156 \times 0.48} = \sqrt{74.88} = 8.653$ (4 s.f.)

So $Y \sim N(156, 8.653^2)$

$P(X \geqslant 170) \approx P(Y > 169.5) = 0.05936... = 0.0594$ (4 s.f.)

c There is a greater than 5% chance that 170 people out of 300 would be cured, therefore there is insufficient evidence for the herbalist's claim that the new remedy is more effective than the original remedy.

19 $X \sim \mathrm{N}(\mu, 2^2)$

$\mathrm{H}_0 : \mu = 7, \ \mathrm{H}_1 : \mu > 7$, one-tailed test at the 5% level.

Assume H_0, so that $X \sim \mathrm{N}(7, 2^2)$ and $\overline{X} \sim \left(7, \frac{2^2}{25}\right)$

Let $Z = \dfrac{\overline{X} - 7}{\frac{2}{\sqrt{25}}}$

Using the inverse normal function, $\mathrm{P}(Z > z) = 0.05 \Rightarrow z = 1.6449$

$1.6449 = \dfrac{\overline{X} - 7}{\frac{2}{5}} \Rightarrow \overline{X} = 7 + 1.6449 \times \dfrac{2}{5} = 7.6579...$

So the critical region is $\overline{X} > 7.6579...$ or 7.66 cm (3 s.f.).

20 Let B represent the amount of water in a bottle, so $B \sim \mathrm{N}(\mu, 2^2)$.

$\mathrm{H}_0 : \mu = 125, \mathrm{H}_1 : \mu < 125$, one-tailed test at the 5% level.

Assume H_0, so that $B \sim \mathrm{N}(125, 2^2)$ and $\overline{B} \sim \left(125, \frac{2^2}{15}\right)$

Using the normal CD function, $\mathrm{P}(\overline{B} < 124.2) = 0.06066... = 0.0607$ (3 s.f.)

$0.0607 > 0.05$ so not significant, so accept H_0.

There is insufficient evidence to conclude that the mean content of a bottle is lower than the manufacturer's claim.

21 Let B represent the breaking strength, so $B \sim \mathrm{N}(170.2, 10.5^2)$.

a Using the normal CD function, $\mathrm{P}(174.5 < B < 175.5) = 0.03421... = 0.0342$ (3 s.f.)

b $n = 50$ so $\overline{B} \sim \mathrm{N}\left(170.2, \dfrac{10.5^2}{50}\right)$

Using the normal CD function, $\mathrm{P}(\overline{B} > 172.4) = 0.06922... = 0.0692$ (3 s.f.)

c $\mathrm{H}_0 : \mu = 170.2, \mathrm{H}_1 : \mu > 170.2$ one-tailed test at the 5% level.

Assume H_0, so that $B \sim \mathrm{N}(170.2, 10.5^2)$ and $\overline{B} \sim \left(170.2, \dfrac{10.5^2}{50}\right)$ (as before).

Using the normal CD function, $\mathrm{P}(\overline{B} > 172.4) = 0.06922... = 0.0692$ (3 s.f.)

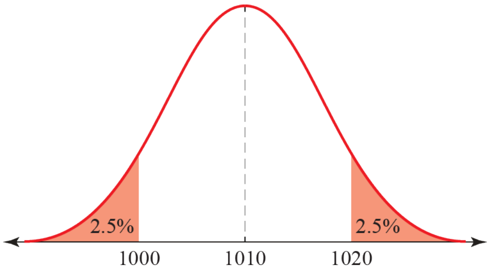
This is the p-value for the hypothesis test.

$0.0692 > 0.05$ so not significant, so accept H_0.

There is insufficient evidence to conclude that the mean breaking strength is increased.

22 Let W represent the weight of sugar in a packet, so $W \sim \mathrm{N}(1010, \sigma^2)$.

a $\mathrm{P}\left(1000 < W < 1020\right) = 0.95 \Rightarrow \mathrm{P}(W < 1000) = 0.025 \Rightarrow \mathrm{P}\left(Z < \dfrac{1000 - 1010}{\sigma}\right) = 0.025$



Using the inverse normal function, $z = -1.95996...$

$$\text{so} \quad -1.95996... = \frac{1000 - 1010}{\sigma}$$

$$\sigma = \frac{-10}{-1.95996..} = 5.1021...$$

$$\sigma^2 = 26.031... = 26.03 \text{ (2 d.p.)}$$

b $n = 8$ and $\sum x = 8109.1$, so $\bar{x} = 1013.6375$

$\mathrm{H}_0 : \mu = 1010$, $\mathrm{H}_1 : \mu \neq 1010$, two-tailed test with 1% in each tail.

Assume H_0, so that $W \sim \mathrm{N}(1010, 26.03)$ and $\overline{W} \sim \left(1010, \dfrac{26.03}{8}\right)$

Using the normal CD function, $\mathrm{P}(\overline{W} > 1013.6375) = 0.02187... = 0.0219$ (3 s.f.)

$0.0219 > 0.01$ so not significant, so accept H_0.

There is insufficient evidence of a deviation in the mean from 1010, so we can assume that condition **i** is being met.

23 Let D represent the diameter of a little-gull egg, so $D \sim \mathrm{N}(4.11, 0.19^2)$.

a Using the normal CD function, $\mathrm{P}(3.9 < D < 4.5) = 0.84542... = 0.8454$ (4 s.f.)

b $\sigma = 0.19$, $n = 8$, $\sum d = 34.5$, $\bar{d} = 4.3125$

$\mathrm{H}_0 : \mu = 4.11$, $\mathrm{H}_1 : \mu \neq 4.11$, two-tailed test with 0.5% in each tail.

Assume H_0, so that $D \sim \mathrm{N}(4.11, 0.19^2)$ and $\overline{D} \sim \left(4.11, \dfrac{0.19^2}{8}\right)$

Using the normal CD function, $\mathrm{P}(\overline{D} > 4.3125) = 0.00128... = 0.0013$ (2 s.f.)

p-value $= 2 \times 0.00128\ldots = 0.00258 < 0.01$ so significant, so reject H_0.

There is evidence that the mean diameter of eggs from this island is different from elsewhere.

24 **a** $X \sim \mathrm{N}\left(\mu, \sigma^2\right)$

$$\overline{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

24 b $\mathrm{P}\left(\left|\overline{X}-\mu\right|<15\right)=\mathrm{P}\left(|Z|<\frac{15}{\left(\frac{\sigma}{\sqrt{40}}\right)}\right)$

Require $\mathrm{P}\left(|Z|<\frac{15\sqrt{n}}{40}\right)>0.95 \Rightarrow \mathrm{P}\left(Z<\frac{15\sqrt{n}}{40}\right)<0.025$ (by symmetry)

Using the inverse normal function, $z=-1.95996...$

so $\frac{15\sqrt{n}}{40}>1.95996...$

$$\sqrt{n}>\frac{40\times 1.95996...}{15}=5.2266...$$

$$n>27.317...$$

So a sample of at least 28 is needed.

Challenge

a Use the binomial distribution $X \sim \mathrm{B}(15,0.48)$.
Using the binomial CD function, $\mathrm{P}(X>8)=1-\mathrm{P}(X\leqslant 8)=1-0.74903...=0.2510$ (4 s.f.)

b The distribution satisfies the two necessary conditions for an approximation by a normal distribution to be valid: $n=250$ is large (> 50) and $p=0.48$ is close to 0.5.

$\mu=np=250\times 0.48=120$ and $\sigma=\sqrt{np(1-p)}=\sqrt{120\times 0.52}=\sqrt{62.4}=7.90$ (3 s.f.)

Test to determine whether the mean is different from 120 (the expected number of supporters based on the manager's claim):

$\mathrm{H}_0:\mu=120,\ \mathrm{H}_1:\mu\neq 120$, two-tailed test with 2.5% in each tail.

Assume H_0, so that $X\sim \mathrm{N}(120,7.9^2)$ and $\overline{X}\sim\left(120,\frac{7.9^2}{250}\right)$

Using the inverse normal function,

$\mathrm{P}(\overline{X}<\overline{x})=0.025\Rightarrow \overline{x}=104.516$ and $\mathrm{P}(\overline{X}>\overline{x})=0.025\Rightarrow \overline{x}=135.484$

So the critical region approximated for the binomial distribution is $\overline{X}\leqslant 105$ or $\overline{X}\geqslant 135$.

(Note: 105 lies in the critical region because, as part of the normal distribution, it includes the region between 104.5 and 105.5 and 104.513 is where the critical region starts. Similarly 135 lies in the critical region because, as part of the normal distribution, it includes the region between 134.5 and 135.5 and 135.487 is where the critical region starts.)

c Since $102<105$, there is sufficient evidence to reject H_0, i.e. there is sufficient evidence to say, at the 5% level, that the level of support for the manager is different from 48%.

Review exercise 1

1 a Produce a table for the values of log *s* and log *t*:

log *s*	0.3010	0.6532	0.7924	0.8633	0.9590
log *t*	−0.4815	0.0607	0.2455	0.3324	0.4698

which produces $r = 0.9992$

b Since r is very close to 1, this indicates that log *s* by log *t* is very close to being linear, which means that *s* and *t* are related by an equation of the form $t = as^n$ (beginning of Section 1.1).

c Rearranging the equation:

$\log t = -0.9051 + \log s^{1.4437}$

$\Rightarrow t = 10^{-0.9051+\log s^{1.4437}} = 10^{-0.9051} \times 10^{\log s^{1.4437}}$

$\Rightarrow t = 10^{-0.9051} \times s^{1.4437}$

and so $a = 10^{-0.9051} = 0.1244$ (4 s.f.) and $n = 1.4437$

2 a Rearranging the equation:

$y = -0.2139 + 0.0172x$

$\Rightarrow \log t = -0.2139 + 0.0172P$

$\Rightarrow t = 10^{-0.2139+0.0172P} = 10^{-0.2139} \times 10^{0.0172P}$

$\Rightarrow t = 10^{-0.2139} \times \left(10^{0.0172}\right)^P$

Therefore $a = 10^{-0.2139} = 0.611$ (3 s.f.) and $b = 10^{0.0172} = 1.04$ (3 s.f.).

b Not in the range of data (extrapolation).

3 a $r = \dfrac{59.524}{\sqrt{152.444 \times 26.589}}$

$= 0.93494$ (the formulae for this is under S1 in the formula book).

b Make sure your hypotheses are clearly written using the parameter ρ:

$H_0 : \rho = 0, \quad H_1 : \rho > 0$

Test statistic: r = 0.935

Critical value at 1% = 0.7155

(Look up the value under 0.01 in the table for product moment coefficient; quote the figure in full.)

$0.935 > 0.7155$

Draw a conclusion in the context of the question:

So reject H_0 : levels of serum and disease are positively correlated.

4 $r = -0.4063$, critical value for $n = 6$ is -0.6084, so no evidence.

5 **a** $H_0 : \rho = 0$

$H_1 : \rho < 0$

From the data, $r = -0.9313$. Since the critical value for $n = 5$ is -0.8783, there is sufficient evidence to reject H_0, i.e. at the 2.5% level of significance, there is sufficient evidence to say that there is negative correlation between the number of miles done by a one-year-old car and its value.

b If a 1% level of significance was used, then the critical value for $n = 5$ is -0.9343 and so there would not be sufficient evidence to reject H_0.

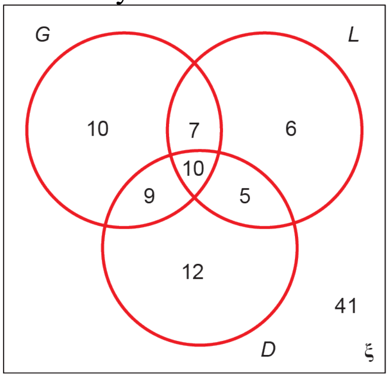
6 **a**
$$\begin{aligned}\mathrm{P}(\text{tourism}) &= \frac{50}{148}\\ &= \frac{25}{74}\\ &= 0.338 \text{ (3 s.f.)}\end{aligned}$$

b The words 'given that' in the question tell you to use conditional probability:
$$\begin{aligned}\mathrm{P}(\text{no glasses} \mid \text{tourism}) &= \frac{\mathrm{P}(G' \cap T)}{\mathrm{P}(T)}\\ &= \frac{\frac{23}{148}}{\frac{50}{148}}\\ &= \frac{23}{50}\\ &= 0.46\end{aligned}$$

c It often helps to write down which combinations you want:
$$\begin{aligned}\mathrm{P}(\text{right-handed}) &= \mathrm{P}(E \cap RH) + \mathrm{P}(T \cap RH) + \mathrm{P}(C \cap RH)\\ &= \frac{30}{148} \times 0.8 + \frac{50}{148} \times 0.7 + \frac{68}{148} \times 0.75\\ &= \frac{55}{74}\\ &= 0.743 \text{ (3 s.f.)}\end{aligned}$$

d The words 'given that' in the question tell you to use conditional probability:
$$\begin{aligned}\mathrm{P}(\text{engineering} \mid \text{right-handed}) &= \frac{\mathrm{P}(E \cap RH)}{\mathrm{P}(RH)}\\ &= \frac{\frac{30}{148} \times 0.8}{\frac{55}{74}}\\ &= \frac{12}{55}\\ &= 0.218 \text{ (3 s.f.)}\end{aligned}$$

7 **a** Start in the middle of the Venn diagram and work outwards. Remember the rectangle and those not in any of the circles. Your numbers should total 100.



b $$\mathrm{P}(G, L', D') = \frac{10}{100}$$
$$= \frac{1}{10} = 0.1$$

c $$\mathrm{P}(G', L', D') = \frac{41}{100} = 0.41$$

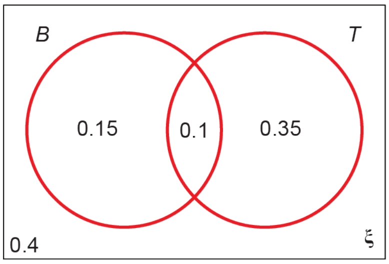
d $$\mathrm{P}(\text{only two attributes}) = \frac{9+7+5}{100}$$
$$= \frac{21}{100} = 0.21$$

e The word ‘given’ in the question tells you to use conditional probability:

$$\mathrm{P}(G \mid L \cap D) = \frac{\mathrm{P}(G \mid L \cap D)}{\mathrm{P}(\mathrm{L} \mid D)}$$
$$= \frac{\frac{10}{100}}{\frac{15}{100}}$$
$$= \frac{10}{15}$$
$$= \frac{2}{3} = 0.667 \text{ (3 s.f.)}$$

8 **a** $$\mathrm{P}(B \cup T) = \mathrm{P}(B) + \mathrm{P}(T) - \mathrm{P}(B \cap T)$$
$$0.6 = 0.25 + 0.45 - \mathrm{P}(B \cap T)$$
$$\mathrm{P}(B \cap T) = 0.1$$

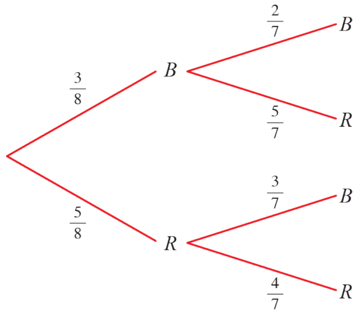
b When drawing the Venn diagram remember to draw a rectangle around the circles and add the probability 0.4.
Remember the total in circle $B = 0.25$ and the total in circle $T = 0.45$.



8 c The words ‘given that’ in the question tell you to use conditional probability:

$$\begin{aligned}\mathrm{P}\left(B\cap T'\mid B\cup T\right)&=\frac{0.15}{0.6}\\&=\frac{1}{4}\\&=0.25\end{aligned}$$

9 a

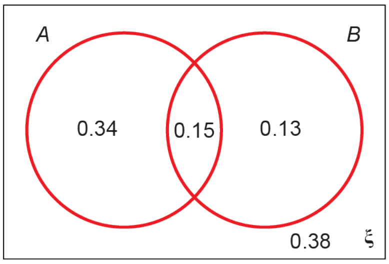


b i There are two different situations where the second counter drawn is blue. These are BB and RB. Therefore the probability is: $\left(\frac{3}{8}\times\frac{2}{7}\right)+\left(\frac{5}{8}\times\frac{3}{7}\right)=\frac{6+15}{56}=\frac{21}{56}=\frac{3}{8}=0.375$.

ii $\mathrm{P(both\ blue\mid 2nd\ blue)}=\dfrac{\mathrm{P(both\ blue\ and\ 2nd\ blue)}}{\mathrm{P(2nd\ blue)}}=\dfrac{\mathrm{P(both\ blue)}}{\mathrm{P(2nd\ blue)}}=\dfrac{\left(\frac{3}{8}\times\frac{2}{7}\right)}{\left(\frac{3}{8}\right)}=\dfrac{2}{7}$

10 a The first two probabilities allow two spaces in the Venn diagram to be filled in. $\mathrm{P}(A\cup B)=\mathrm{P}(A\cap B')+\mathrm{P}(A'\cap B)+\mathrm{P}(A\cap B)$, and this can be rearranged to see that $P(A\cap B)=0.15$

Finally, $\mathrm{P}(A\cup B)=0.62\Rightarrow \mathrm{P}((A\cup B)')=0.38$. The completed Venn diagram is therefore:



b $\mathrm{P}(A)=0.34+0.15=0.49$ and $\mathrm{P}(B)=0.13+0.15=0.28$

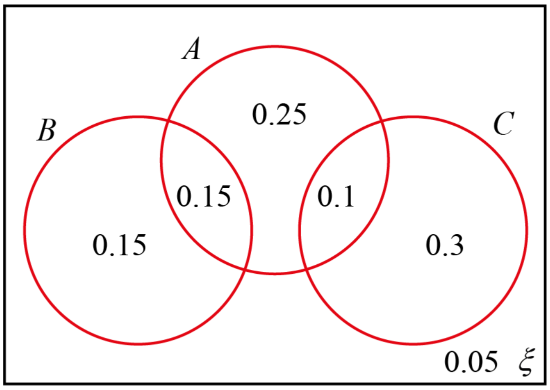
c $\mathrm{P}(A\mid B')=\dfrac{\mathrm{P}(A\cap B')}{\mathrm{P}(B')}=\dfrac{0.34}{1-\mathrm{P}(B)}=\dfrac{0.34}{0.72}=0.472$ (3 d.p.).

d If A and B are independent, then $\mathrm{P}(A)=\mathrm{P}(A\mid B)=\mathrm{P}(A\mid B')$. From parts **b** and **c**, this is not the case. Therefore they are not independent.

11 a $\mathrm{P}(A\cap B)=\mathrm{P}(A)\times\mathrm{P}(B)\Rightarrow \mathrm{P}(A)=\mathrm{P}(A\cap B)\div\mathrm{P}(B)=0.15\div0.3=0.5$

11 b $\mathrm{P}(A\cup B)=\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A\cap B)=0.5+0.3-0.15=0.65\Rightarrow \mathrm{P}(A'\cap B')=1-0.65=0.35$

c Since B and C are mutually exclusive, they do not intersect.
The intersection of A and C should be 0.1 but $\mathrm{P}(A)=0.5$, allowing $\mathrm{P}(A\cap B'\cap C')$ to be calculated. The filled-in probabilities sum to 0.95, and so $\mathrm{P}(A'\cap B'\cap C')=0.05$.
Therefore, the filled-in Venn diagram should look like:



d i $\mathrm{P}(A\,|\,C)=\dfrac{\mathrm{P}(A\cap C)}{\mathrm{P}(C)}=\dfrac{0.1}{0.4}=0.25$

ii The set $A\cap(B\cup C')$ must be contained within A. First find the set $B\cup C'$: this is made up from four distinct regions on the above Venn diagram, with labels 0.15, 0.15, 0.25 and 0.05. Restricting to those regions that are also contained within A leaves those labelled 0.15 and 0.25. Therefore, $\mathrm{P}(A\cap(B\cup C'))=0.15+0.25=0.4$

iii From part **ii**, $\mathrm{P}(B\cup C')=0.15+0.15+0.25+0.05=0.6$. Therefore

$$\mathrm{P}(A\,|\,(B\cup C'))=\frac{\mathrm{P}(A\cap(B\cup C'))}{\mathrm{P}(B\cup C')}=\frac{0.4}{0.6}=\frac{2}{3}$$

12 a There are two different events going on: ‘Joanna oversleeps’ (O) and ‘Joanna is late for college’ (L). From the context, we cannot assume that these are independent events.
Drawing a Venn diagram, none of the regions can immediately be filled in. We are told that $\mathrm{P}(O)=0.15$ and so $\mathrm{P}(J$ does not oversleep$)=\mathrm{P}(O')=0.85$. The other two statements can be interpreted as $\dfrac{\mathrm{P}(L\cap O)}{\mathrm{P}(O)}=0.75$ and $\dfrac{\mathrm{P}(L\cap O')}{\mathrm{P}(O')}=0.1$

Filling in the first one:

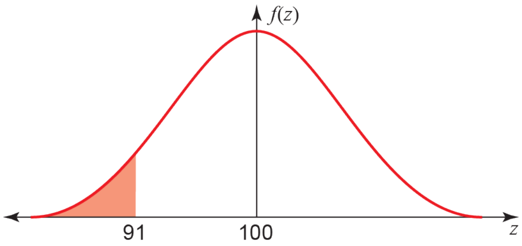
$$\frac{\mathrm{P}(L\cap O)}{\mathrm{P}(O)}=0.75\Rightarrow\frac{\mathrm{P}(L\cap O)}{0.15}=0.75\Rightarrow \mathrm{P}(L\cap O)=0.1125$$

Also, $\dfrac{\mathrm{P}(L\cap O')}{0.85}=0.1\Rightarrow \mathrm{P}(L\cap O')=0.085$

Therefore, $\mathrm{P}(L)=\mathrm{P}(L\cap O)+\mathrm{P}(L\cap O')=0.1125+0.085=0.1975$

b $\mathrm{P}(L\mid O)=\dfrac{\mathrm{P}(L\cap O)}{\mathrm{P}(O)}=\dfrac{0.1125}{0.1975}=\dfrac{45}{79}=0.5696$ (4 s.f.).

13 a Drawing a diagram will help you to work out the correct area:

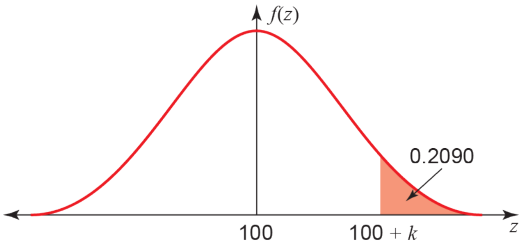


Using $z = \frac{x-\mu}{\sigma}$. As 91 is to the left of 100, your z value should be negative.

$$\begin{aligned}P(X<91) &= P\left(Z<\frac{91-100}{15}\right)\\ &= P(Z<-0.6)\\ &= 1-0.7257\\ &= 0.2743\end{aligned}$$

(The tables give $P(Z<0.6) = P(Z>-0.6)$, so you want $1-$ this probability.)

b

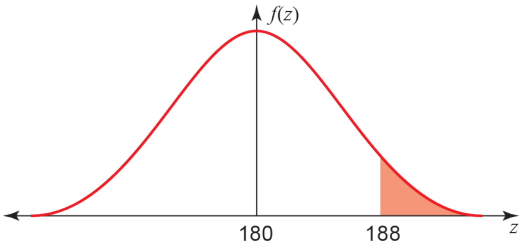


As 0.2090 is not in the table of percentage points you must work out the larger area:
$1-0.2090=0.7910$
Use the first table or calculator to find the z value. It is positive as $100+k$ is to the right of 100.
$P(X>100+k)=0.2090$ or $P(X<100+k)=0.791$

$$\frac{100+k-100}{15}=0.81$$
$$k=12$$

14 a Let H be the random variable ~ height of athletes, so $H \sim \mathrm{N}\left(180,\ 5.2^2\right)$

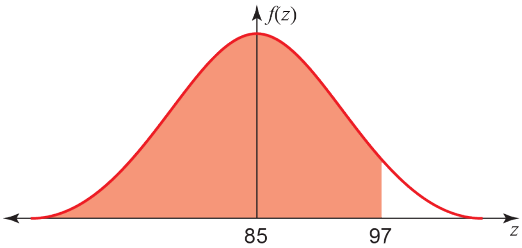
Drawing a diagram will help you to work out the correct area:



Using $z = \dfrac{x-\mu}{\sigma}$. As 188 is to the right of 180 your z value should be positive. The tables give $\mathrm{P}(Z < 1.54)$ so you want $1 -$ this probability:

$$\begin{aligned}\mathrm{P}(H > 188) &= \mathrm{P}\left(Z > \frac{188-100}{5.2}\right)\\ &= \mathrm{P}(Z > 1.54)\\ &= 1 - 0.9382\\ &= 0.0618\end{aligned}$$

b Let W be the random variable ~ weight of athletes, so $W \sim \mathrm{N}\left(85,\ 7.1^2\right)$



Using $z = \dfrac{x-\mu}{\sigma}$. As 97 is to the right of 85, your z value should be positive.

$$\begin{aligned}\mathrm{P}(W < 97) &= \mathrm{P}\left(Z < \frac{97-85}{7.1}\right)\\ &= \mathrm{P}(Z < 1.69)\\ &= 0.9545\end{aligned}$$

c $\mathrm{P}(W > 97) = 1 - \mathrm{P}(W < 97)$, so

$$\begin{aligned}\mathrm{P}(H > 188 \,\&\, W > 97) &= 0.618(1 - 0.9545)\\ &= 0.00281\end{aligned}$$

d Use the context of the question when you are commenting:
The evidence suggests that height and weight are positively correlated/linked, so assumption of independence is not sensible.

15 a Use the table of percentage points or calculator to find z. You must use at least the four decimal places given in the table.

$$\mathrm{P}(Z > a) = 0.2$$
$$a = 0.8416$$
$$\mathrm{P}(Z < b) = 0.3$$
$$b = -0.5244$$

0.5244 is negative since 1.65 is to the left of the centre. 0.8416 is positive as 1.78 is to the right of the centre.

Using $z = \dfrac{x-\mu}{\sigma}$:

$$\frac{1.78-\mu}{\sigma} = 0.8416 \Rightarrow 1.78-\mu = 0.8416\sigma \quad (\mathbf{1})$$
$$\frac{1.65-\mu}{\sigma} = -0.5244 \Rightarrow 1.65-\mu = 0.5244\sigma \quad (\mathbf{2})$$

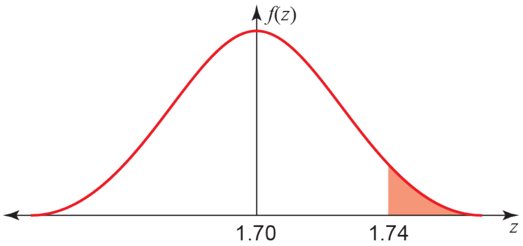
Solving simultaneously, $(\mathbf{1})-(\mathbf{2})$:

$$0.13 = 1.366\sigma$$
$$\sigma = 0.095 \text{ m}$$

Substitute in (**1**): $1.78 - \mu = 0.8416 \times 0.095$

$$\mu = 1.70\,\text{m}$$

b



Using $z = \dfrac{x-\mu}{\sigma}$:

$$\mathrm{P}(\text{height} > 1.74) = \mathrm{P}\left(z > \frac{1.74-1.70}{0.095}\right)$$
$$= \mathrm{P}(z > 0.42) \text{ (the tables give P}(Z < 0.42)\text{ so you need } 1 - \text{this probability)}$$
$$= 1 - 0.6628$$
$$= 0.3372 \text{ (calculator gives 0.3369)}$$

16 a $\mathrm{P}(D < 21.5) = 0.32$ and $\mathrm{P}(Z < a) = 0.32 \Rightarrow a = -0.467$. Therefore

$$\frac{21.5-\mu}{\sigma} = -0.467 \Rightarrow 21.5 - 22 = -0.467\sigma \Rightarrow \sigma = \frac{0.5}{0.467} = 1.071 \text{ (4 s.f.).}$$

b $\mathrm{P}(21 < D < 22.5) = \mathrm{P}(D < 22.5) - \mathrm{P}(D < 21) = 0.5045$ (4 s.f.).

c $\mathrm{P}(B \geq 10) = 1 - \mathrm{P}(B \leq 9) = 1 - 0.01899 = 0.98101$ (using 4 s.f. for the value given by the binomial distribution) or 0.981 (4 s.f.).

17 a Let W be the random variable 'the number of white plants'. Then $W \sim \mathrm{B}(12,\ 0.45)$ ('batches of 12': $n = 12$; '45% have white flowers': $p = 0.45$).

$$\mathrm{P}(W=5)=\binom{12}{5}0.45^5\,0.55^7 \text{ (you can also use tables: } P(W \le 5)-P(W \le 4))$$
$$=0.2225$$

b Batches of 12, so: 7 white, 5 coloured; 8 white, 4 coloured; etc.

$$\mathrm{P}(W \ge 7)=1-\mathrm{P}(W \le 6)$$
$$=1-0.7393$$
$$=0.2607$$

c Use your answer to part **b**: $p = 0.2607$, $n = 10$:

$$\mathrm{P}(\text{exactly } 3)=\binom{10}{3}(0.2607)^3\,(1-0.2607)^7$$
$$=0.2567$$

d A normal approximation is valid, since n is large (> 50) and p is close to 0.5. Therefore $\mu = np = 150\times0.45 = 67.5$ and $\sigma=\sqrt{np(1-p)}=\sqrt{67.5\times0.55}=\sqrt{37.125}=6.093$ (4 s.f.). Now $\mathrm{P}(X>75)\approx\mathrm{P}(N>75.5)=0.0946$ (3 s.f.).

18 a Using the binomial distribution, $\mathrm{P}(B=35)=\binom{80}{35}\times0.48^{35}\times0.52^{45}=0.06703$.

b A normal approximation is valid, since n is large (> 50) and p is close to 0.5. Therefore $\mu = np = 80\times0.48 = 38.4$ and $\sigma=\sqrt{np(1-p)}=\sqrt{38.4\times0.52}=\sqrt{19.968}=4.469$ (4 s.f.). Now $\mathrm{P}(B=35)\approx\mathrm{P}(34.5<N<35.5)=0.0668$ (3 s.f.).

Percentage error is $\dfrac{0.06703-0.0668}{0.06703}=0.34\%$.

19 Remember to identify which is H_0 and which is H_1. This is a one-tail test since we are only interested in whether the time taken to solve the puzzle has reduced. You must use the correct parameter (μ):

$\mathrm{H}_0:\mu=18 \quad \mathrm{H}_1:\mu<18$

Using $z=\dfrac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$, $z=\dfrac{(16.5-18)}{\left(\frac{3}{\sqrt{15}}\right)}=-1.9364...$

Using the percentage point table and quoting the figure in full:

5% one tail c.v. is $z=-1.6449$

$-1.9364<-1.6449$, so

significant *or* reject H_0 *or* in critical region.

State your conclusion in the context of the question:

There is evidence that the (mean) time to complete the puzzles has reduced.

Or Robert is getting faster (at doing the puzzles).

20 a $\text{P}(Z < a) = 0.05 \Rightarrow -1.645$. Using that $\text{P}(L < 1.7) = 0.05$ means that

$$\frac{1.7-\mu}{0.4} = -1.645 \Rightarrow 1.7 - \mu = -0.658 \Rightarrow \mu = 2.358$$

b $\text{P}(L > 2.3) = 0.5576$ (4 s.f.) and so, using the binomial distribution,
$\text{P}(B \geq 6) = 1 - \text{P}(B \leq 5) = 1 - 0.4758 = 0.5242$ (4 s.f.).

c It is thought that the mean length of the female rattlesnakes is 1.9 m, and a hypothesis test is needed to conclude whether the mean length is not equal to 1.9 m. Therefore,

$H_0 : \mu = 1.9$

$H_1 : \mu \neq 1.9$

Sample size: 20. Therefore, the sample population is initially thought to have distribution $\bar{M} \sim N\left(1.9, \frac{0.3^2}{20}\right)$. By using the inverse normal distribution, $\text{P}(\bar{M} < 1.768) = 0.025$ and $\text{P}(\bar{M} > 2.032) = 0.025$, meaning that the critical region is below 1.768 and above 2.032

d There is sufficient evidence to reject H_0, since 2.09 > 2.032; i.e. there is sufficient evidence to say, at the 5% level, that the mean length of the female rattlesnakes is not equal to 1.9 metres.

21 It is thought that the daily mean temperature in Hurn is less than 12 °C, and so a hypothesis test is needed to conclude whether, at the 5% level of significance, the mean temperature is less than 12 °C. Therefore,

$H_0 : \mu = 12$

$H_1 : \mu < 12$

Sample size: 20. Therefore, the sample population is initially thought to have distribution $\bar{T} \sim N\left(12, \frac{2.3^2}{20}\right)$. By using the inverse normal distribution, $\text{P}(\bar{T} < 11.154) = 0.05$, meaning that the critical region consists of all values below 11.154. Since 11.1 < 11.154, there is sufficient evidence to reject H_0; i.e. there is sufficient evidence to say, at the 5% level, that the mean daily temperature in Hurn is less than 12 °C.

Challenge

1 a Since A and B could be mutually exclusive, $\text{P}(A \cap B) \geq 0$. Since $\text{P}(A \cap B) \leq \text{P}(B) = 0.3$, we have that $0 \leq \text{P}(A \cap B) \leq 0.3$ and so $q = \text{P}(A \cap B') = \text{P}(A) - \text{P}(A \cap B)$. Therefore $0.4 \leq p \leq 0.7$

b First, $\text{P}(B \cap C) \leq \text{P}(B) = 0.3$ and so $q \leq \text{P}(B \cap C) - \text{P}(A \cap B \cap C) \leq 0.25$. Moreover, it is possible to draw a Venn diagram where $q = 0$, and so $0 \leq q \leq 0.25$

Challenge

2 **a** We wish to use a hypothesis test to determine (at the 10% significance level) whether the support for the politician is 53%. A normal distribution is suitable, and we use the model given by $\mu = np = 300 \times 0.53 = 159$ and $\sigma = \sqrt{np(1-p)} = \sqrt{159 \times 0.47} = \sqrt{74.73} = 8.645$ (4 s.f.).

Therefore,

$H_0 : \mu = 159$

$H_1 : \mu \neq 159$

By using the inverse normal distribution, $\mathrm{P}(\overline{X} < 144.78) = 0.05$ and $\mathrm{P}(\overline{X} > 173.22) = 0.05$ (2 d.p.) and so the critical region consists of the values below 144.78 and above 173.22

b Since 173 is not within the critical region, there is not sufficient evidence to reject H_0 at the 10% significance level; i.e. there is not sufficient evidence to say, at the 10% level, that the politician's claim that they have support from 53% of the constituents is false.