Review exercise 1

1 a $S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 8880 - \frac{130 \times 48}{8} = 8100$

$b = \frac{S_{xy}}{S_{xx}} = \frac{8100}{20487.4} = 0.395363... = 0.395$ (3 s.f.)

$a = \bar{y} - b\bar{x} = \frac{48}{8} - \frac{8100}{20487.4} \times \frac{130}{8} = -0.42468... = -0.425$ (3 s.f.)

So the equation of the regression line is: $y = -0.425 + 0.395x$

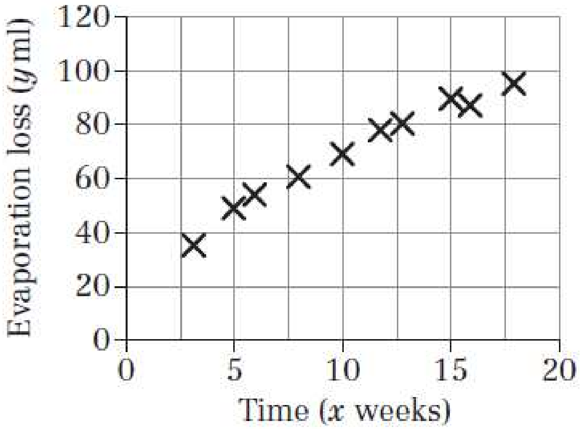
b $f - 100 = -0.42468... + 0.39536...(m - 250)$

$\Rightarrow f = 0.7353... + 0.3953...m$

$\Rightarrow f = 0.735 + 0.395m$ (giving the equation parameters to 3 s.f.)

c $f = 0.7353 + 0.3953 \times 235 = 93.6$ litres (3 s.f.)

2 a



b There appears to be a linear relationship between the variables as the points lie close to a straight line.

c The summary data for x and y are:

$\sum x = 106 \qquad \sum y = 704$

$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 8354 - \frac{106 \times 704}{10} = 891.6$

$S_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 1352 - \frac{106^2}{10} = 228.4$

$b = \frac{S_{xy}}{S_{xx}} = \frac{891.6}{228.4} = 3.90367... = 3.90$ (2 d.p.)

$a = \bar{y} - b\bar{x} = \frac{704}{10} - b\frac{106}{10} = 29.02$ (2 d.p.)

d For every extra week in storage, another 3.90 ml of chemical evaporates.

2 e i $y = 29.02 + 3.903 \times 19 = 103\text{ml}$ (3 s.f.)

ii $y = 29.02 + 3.903 \times 35 = 166\text{ml}$ (3 s.f.)

f i The value of 19 is close to the range of x, so the estimate should be reasonably reliable.

ii The value of 35 is well outside range of x, so the estimate is unreliable since there is no evidence that the model will continue to hold as x increases.

3 a Find the y values by subtracting 2460 form all the l values. The summary data for x and y are:

$\sum x = \sum t = 337.1 \qquad \sum y = 16.28$

$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 757.467 - \frac{337.1 \times 16.28}{8} = 71.4685$$

$$S_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 15965.01 - \frac{337.1^2}{8} = 1760.45875$$

b $b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{71.4685}{1760.45875} = 0.04059652 = 0.0406$ (3 s.f.)

$a = \bar{y} - b\bar{x} = \dfrac{16.28}{8} - 0.04059652 \times \dfrac{337.1}{8} = 0.324364 = 0.325$ (3 s.f.)

The equation of the regression line is: $y = 0.324 + 0.0406x$

c $t = 40$, therefore $x = 40$

$y = 0.3243 + 0.0406 \times 40 = 1.9483$

$l = 2460 + 1.9483 = 2461.95\text{ mm}$ (2 d.p.)

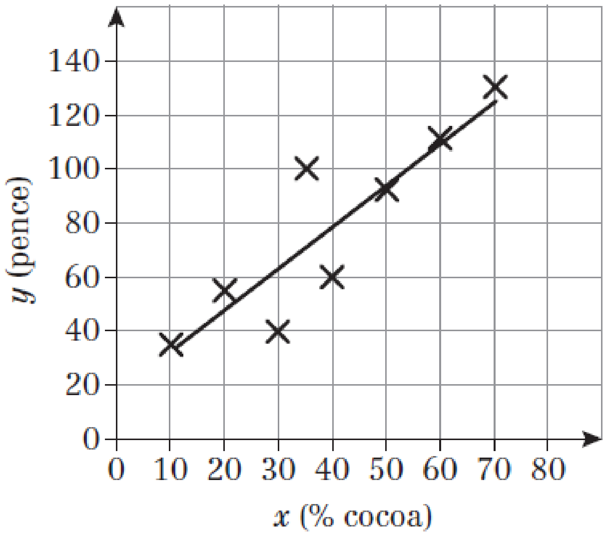
d $l - 2460 = 0.324 + 0.0406t$

$\Rightarrow l = 2460.324 + 0.0406t$

e At $t = 90$, $l = 2460.324 + 0.0406 \times 90 = 2463.98\text{mm}$ (2 d.p.)

f As 90 °C is well outside the range of data, the estimate is unlikely to be reliable.

4 a This is a scatter diagram of the data. (The diagram also shows the regression line, see part **d**.)



b $S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 28750 - \frac{315 \times 620}{8} = 4337.5$

$S_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 15225 - \frac{315^2}{8} = 2821.875$

c $b = \frac{S_{xy}}{S_{xx}} = 1.53709\ldots = 1.54 \text{ (3 s.f.)}$

$a = \overline{y} - b\overline{x} = \frac{620}{8} - b\frac{315}{8} = 16.976\ldots = 17.0 \text{ (3 s.f.)}$

d From part **c**, the equation of the regression line is: $y = 17.0 + 1.54x$
This line is shown on the scatter diagram (see answer for part **a**).

e i Brand D is overpriced, since this data point is a long way above the regression equation line.

ii Using the equation of the regression line: $y = 17 + 35 \times 1.54 = 71\text{p (2 s.f.)}$ pence

5 a This table sets out the residuals for each data point (x, y):

x	y	$y = 12.476 + 0.2311x$	ε
250	72	70.251	1.749
300	81	81.806	–0.806
340	90	91.05	–1.05
360	94	95.672	–1.672
385	102	101.4495	0.5505
400	106	104.916	1.084
450	115	116.471	–1.471
475	124	122.2485	1.7515

b Yes, a linear model is suitable as the residuals are randomly scattered about zero.

c $\text{RSS} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = 2090 - \frac{8980^2}{38850} = 14.3 \text{ (3 s.f.)}$

d The first mobile phone operator as the RSS is smaller.

6 a $S_{mm} = \sum m^2 - \frac{(\sum m)^2}{n} = 1768.47 - \frac{(101.9)^2}{6} = 37.868\ldots = 37.9 \text{ (3 s.f.)}$

$S_{md} = \sum md - \frac{\sum m \sum d}{n} = 868.06 - \frac{101.9 \times 49.7}{6} = 23.988\ldots = 24.0 \text{ (3 s.f.)}$

b $b = \frac{S_{md}}{S_{mm}} = \frac{23.988}{37.868} = 0.63346\ldots = 0.633 \text{ (3 s.f.)}$

$a = \bar{d} - b\bar{m} = \frac{49.7}{6} - 0.63346 \times \frac{101.9}{6} = -2.47504\ldots = -2.48 \text{ (3 s.f.)}$

Hence the equation of the regression line of d on m is: $d = -2.48 + 0.633m$

c $d = -2.475 + 0.6334 \times 15.5 = 7.34$ divorces per 1000 people (3 s.f.)

d $S_{dd} = 430.65 - \frac{(49.7)^2}{6} = 18.968\ldots$

$\text{RSS} = S_{dd} - \frac{(S_{md})^2}{S_{mm}} = 18.968 - \frac{(23.988)^2}{37.868} = 3.77 \text{ (3 s.f.)}$

e $\sum \varepsilon = 0 \Rightarrow -0.70365 - 0.3474 + 1.46885 + 0.64215 + x - 0.552 = 0 \Rightarrow x = -0.50765$

f A linear regression model is not suitable as the residuals are not randomly scattered about zero.

7 a $b = \frac{S_{xy}}{S_{xx}} = \frac{895.5714}{3.388571} = 264.29\ldots = 264 \text{ (3 s.f.)}$

$a = \bar{y} - b\bar{x} = \frac{4020}{7} - \frac{895.5714}{3.388571} \times \frac{141}{70} = 41.926\ldots = 41.9 \text{ (3 s.f.)}$

Hence the equation of the regression line of y on x is: $y = 41.9 + 264x$

b $\text{RSS} = S_{yy} - \frac{(S_{xy})^2}{S_{xx}} = 289771.4 - \frac{(895.5714)^2}{3.388571} = 53079.30\ldots = 53\,100 \text{ (3 s.f.)}$

c This table sets out the residuals for each data point (x, y):

x	y	$y = 41.9 + 264x$	ε
1.4	370	411.5	–41.5
1.5	440	437.9	2.1
2.5	660	701.9	–41.9
3.4	950	939.5	10.5
1.3	330	385.1	–55.1
2.2	550	622.7	–72.7
1.8	720	517.71	202.9

d The outlier is 1.8 since it has a much greater residual.

e This could be a legitimate data point since it could represent a company that thrives despite relatively low spending on advertising. So there may be an argument for retaining this data point.

f The new summary statistics are:

$\sum x = 12.3 \quad \sum x^2 = 28.55 \quad \sum y = 3300 \quad \sum xy = 7697$

$S_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 28.55 - \frac{12.3^2}{6} = 3.335$

$S_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 7697 - \frac{12.3 \times 3300}{6} = 932$

$b = \frac{S_{xy}}{S_{xx}} = \frac{932}{3.335} = 279.46\ldots = 279 \text{ (3 s.f.)}$

$a = \bar{y} - b\bar{x} = \frac{3300}{6} - \frac{932}{3.335} \times \frac{12.3}{6} = -22.893\ldots = -22.9 \text{ (3 s.f.)}$

Equation is: $y = -22.9 + 279x$

g $y = -22.89 + 279.46 \times 1.8 = 480 \text{ (3 s.f.)}$

So estimate monthly sales of £480 000

h The estimate is reliable since $x = 1.8$ is within the range of the data.

8 a $S_{ll} = \sum l^2 - \frac{\left(\sum l\right)^2}{n} = 15762.5 - \frac{(353.8)^2}{8} = 115.695$

$S_{lh} = \sum lh - \frac{\sum l \sum h}{n} = 58825.04 - \frac{353.8 \times 1325.1}{8} = 222.4925$

$b = \frac{S_{lh}}{S_{ll}} = \frac{222.4925}{115.695} = 1.92309\ldots = 1.92 \text{ (3 s.f.)}$

$\bar{l} = \frac{353.8}{8} = 44.225 \qquad \bar{h} = \frac{1325.1}{8} = 165.6375$

$a = 165.6375 - 1.92309 \times 44.225 = 80.588\ldots = 80.6 \text{ (3 s.f.)}$

Hence the equation of the regression line of h on l is: $h = 80.6 + 1.92l$

b $h = 80.6 + 1.92 \times 45.1 = 167 \text{ cm (3 s.f.)}$

c $\sum \varepsilon = 0 \Rightarrow 1.47138 + 0.03296 - 2.80543 - 2.94388 + 2.04074 + p - 1.24376 + 1.24089 = 0$

$\Rightarrow p = 2.2071$

d The model is suitable as the residuals are randomly scattered about zero.

e $S_{hh} = \sum h^2 - \frac{\left(\sum h\right)^2}{n} = 219944.9 - \frac{(1325.1)^2}{8} = 458.67875 \text{ (3 s.f.)}$

$\text{RSS} = S_{hh} - \frac{(S_{lh})^2}{S_{ll}} 458.67875 - \frac{(222.4925)^2}{115.695} = 30.8 \text{ (3 s.f.)}$

f The female sample is likely to have the best linear fit as the RSS is lower.

9 Diagram A: as x increases, y decreases. There is a negative correlation.
So this corresponds to $r = –0.79$.

Diagram B: There is no real pattern. There are several values of v for one value of u. There is very weak or no correlation. So this corresponds to $r = 0.08$.

Diagram C: As s increases, t increases. There is a positive correlation.
So this corresponds to $r = 0.68$.

10 a $r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-808.917}{\sqrt{113573 \times 8.657}} = -0.8157\ldots = -0.816 \text{ (3 s.f.)}$

b There is a negative correlation. The survey suggests that houses are cheaper the further they are from the railway station.

c To change miles to kilometres, multiply by 1.6. The coding is linear, so the product moment correlation coefficient will be unaffected by the coding. So the product moment correlation coefficient will still be –0.816.

11 a The shopper in the supermarket for 15 minutes spent $20 - 3 = £17$

b The summary statistics for *s* and *t* are:

$$\sum t = 212 \qquad \sum m = 61$$

$$S_{tm} = \sum tm - \frac{\sum t \sum m}{n} = 2485 - \frac{61 \times 212}{10} = 1191.8$$

$$S_{tt} = \sum t^2 - \frac{\left(\sum t\right)^2}{n} = 5478 - \frac{212^2}{10} = 983.6$$

$$S_{mm} = \sum m^2 - \frac{\left(\sum m\right)^2}{n} = 2101 - \frac{61^2}{10} = 1728.9$$

c $$r = \frac{S_{tm}}{\sqrt{S_{tt}S_{mm}}} = \frac{1191.8}{\sqrt{983.6 \times 1728.9}} = 0.91392\ldots = 0.914 \text{ (3 s.f.)}$$

d The coding is linear (m = *amount spent* – 20) so the product moment correlation coefficient will be unaffected by the coding. So the product moment correlation coefficient will still be 0.914.

e The product moment correlation coefficient of 0.914 suggests that the longer spent shopping the more money the customer spends. It would suggest a relationship between time spent shopping and money spent.

The product moment correlation coefficient of 0.178 suggests that there is no relationship between time spent shopping and money spent.

f The two sets of data might be very different because the data was collected at different times of the day – or on different days of the week – when shopping behaviour is not the same.

12 a The table shows the ranks and *d* and d^2 for each pair of ranks. Remember to rank the data. (Note, it does not matter whether the data is ranked from highest to lowest or vice versa as long as it is same for both judges. It is ranked from highest to lowest here.)

Display	***A***	***B***	***C***	***D***	***E***	***F***	***G***	***H***
Judge P	25	19	21	23	28	17	16	20
Judge Q	20	9	21	13	17	14	11	15
Rank, P	2	6	4	3	1	7	8	5
Rank, Q	2	8	1	6	3	5	7	4
d	0	–2	3	–3	–2	2	1	1
d^2	0	4	9	9	4	4	1	1

$$\sum d^2 = 32$$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \times 32}{8(8^2 - 1)} = 1 - \frac{8}{21} = \frac{13}{21} = 0.619 \text{ (3 s.f.)}$$

12 b $H_0: \rho = 0$ There is no correlation between the judges rankings
$H_1: \rho > 0$ There is a positive correlation between the judges rankings
Sample size = 8
Significance level = 0.05
The critical value for r_s for a 0.05 significance level with a sample size of 8 is $r_s = 0.6429$
As $0.619 < 0.6429$, accept H_0. There is insufficient evidence at the 5% significance level of a positive correlation between rankings of the judges – the competitor's claim is justified.

13 a The table shows the ranks and d and d^2 for each pair of ranks.

Shop	Distance	Price	$r_{distance}$	r_{price}	d	d^2
A	50	1.75	1	9	–8	64
B	175	1.20	2	7	–5	25
C	270	2.00	3	10	–7	49
D	375	1.05	4	6	–2	4
E	425	0.95	5	4	1	1
F	580	1.25	6	8	–2	4
G	710	0.80	7	2	5	25
H	790	0.75	8	1	7	49
I	890	1.00	9	5	4	16
J	980	0.85	10	3	7	49
					Total	286

$\sum d^2 = 286$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 286}{10(10^2-1)} = -0.733 \text{ (3 s.f.)}$$

b $H_0: \rho = 0$, $H_1: \rho < 0$
Sample size = 10
Significance level = 0.05
The critical value for r_s for a 0.05 significance level with a sample size of 10 is $r_s = -0.5636$.
As $-0.733 < -0.5636$, r_s lies within the critical region, so reject H_0. There is sufficient evidence at the 5% significance level that the price of an ice lolly and the distance from the pier are negatively correlated. The further from the pier you go, the less you are likely to pay of an ice lolly.

14 a There is no reason to assume that the variables are normally distributed. Therefore use Spearman's rank correlation coefficient.

14 b Let x be the deaths from pneumoconiosis and y the deaths from lung cancer.
The table shows the ranks and d and d^2 for each pair of ranks.

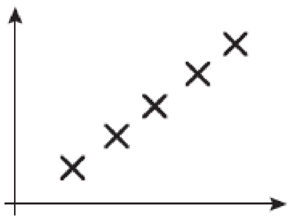
Age group	20–29	30–39	40–49	50–59	60–69	70+
x	12.5	5.9	18.5	19.4	31.2	31
y	3.7	9	10.2	19	13	18
Rank, x	5	6	4	3	1	2
Rank, y	6	5	4	1	3	2
d	−1	1	0	2	−2	0
d^2	1	1	0	4	4	0

$$\sum d^2 = 10$$

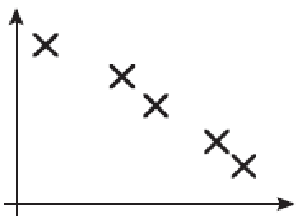
$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 10}{6(6^2-1)} = 1 - \frac{10}{35} = \frac{5}{7} = 0.714 \text{ (3 s.f.)}$$

c $H_0: \rho = 0,\ H_1: \rho > 0$
Sample size = 6
Significance level = 0.05
The critical value for r_s for a 0.05 significance level with a sample size of 6 is $r_s = 0.8286$.
As $0.714 < 0.8286$, accept H_0. There is insufficient evidence at the 5% significance level to suggest a positive association between the rates of deaths from pneumoconiosis and lung cancer.

15 a i As $r = 1$, there is a perfect positive correlation. The points form a straight line, with a positive gradient.



ii As $r_s = -1$ but $r > -1$ there is an imperfect negative correlation. The points approximate a straight line, with a negative gradient.



15 b i Rearranging the data first, the table d and d^2 for each pair of ranks.

Dog	***A***	***B***	***C***	***D***	***E***	***F***	***G***
Judge 1	1	4	2	3	5	6	7
Judge 2	1	2	4	3	5	7	6
d	0	2	−2	0	0	−1	1
***d*²**	0	4	4	0	0	1	1

$\sum d^2 = 10$

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)} = 1 - \frac{6\times 10}{7(7^2-1)} = 1 - \frac{10}{35} = \frac{5}{7} = 0.821 \text{ (3 s.f.)}$$

ii $H_0: \rho = 0$, $H_1: \rho > 0$

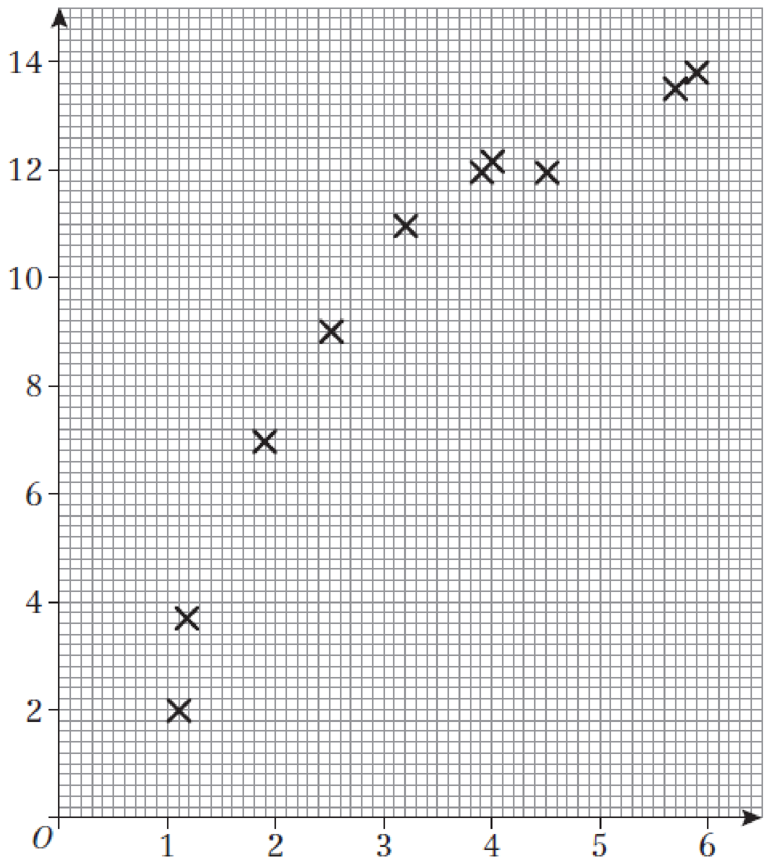
Sample size = 7

Significance level = 0.05

The critical value for r_s for a 0.05 significance level with a sample size of 7 is $r_s = 0.7143$.

As $0.821 > 0.7143$, r_s lies within the critical region, so reject H_0. There is sufficient evidence at the 5% significance level that the judges are in agreement, i.e. there is evidence of a (positive) correlation between the ranks awarded by the judges.

16 a



b The product moment correlation coefficient measures the linear correlation between two variables, i.e. it is a measure of the strength of the linear link between the variables.

16 c The summary statistics for t and p are:

$\sum t = 33.9 \qquad \sum p = 96.4$

$$S_{tt} = \sum t^2 - \frac{\left(\sum t\right)^2}{n} = 141.51 - \frac{33.9^2}{10} = 26.589$$

$$S_{pp} = \sum p^2 - \frac{\left(\sum p\right)^2}{n} = 1081.74 - \frac{96.4^2}{10} = 152.444$$

$$S_{tp} = \sum tp - \frac{\sum t \sum p}{n} = 386.32 - \frac{33.9 \times 96.4}{10} = 59.524$$

d $r = \dfrac{59.524}{\sqrt{152.444 \times 26.589}} = 0.93494\ldots = 0.935$ (3 s.f.)

e $H_0: \rho = 0,\ H_1: \rho > 0$

Sample size = 10

Significance level = 0.01

The critical value for r for a 0.01 significance level with a sample size of 10 is $r = 0.7155$.

As $0.934 > 0.7155$, r lies within the critical region, so reject H_0. There is evidence at the 1% significance level of a positive correlation between the reactant and the product in the chemistry experiment.

f The test for linear correlation is significant but the scatter diagram suggest that there is a non-linear relationship between the variables. The product–moment correlation coefficient should not be used here since the association/relationship is not linear.

17 a $H_0: \rho = 0,\ H_1: \rho < 0$

Sample size = 7

Significance level = 0.01

The critical value for r_s for a 0.01 significance level with a sample size of 7 is $r_s = -0.8929$.

As $-0.93 < -0.8929$, r_s lies within the critical region, so reject H_0. There is sufficient evidence at the 1% significance level that the speed of flow gets slower the wider the river is.

b i This would have no effect on the coefficient since the rank of the flow at G stays the same.

ii Spearman's rank correlation will decrease (i.e. get closer to –1) since the new observation is further supporting the hypothesis.

c Where two or more data values are equal (so there is a tied rank), these observations should be assigned a rank equal to the mean of the tied ranks. Then the product moment correlation coefficient formula should be used to find the Spearman's rank correlation coefficient.

18 a The area under the probability distribution function curve must equal 1, so:

$$\int_0^2 k(4x - x^3)\mathrm{d}x = 1$$

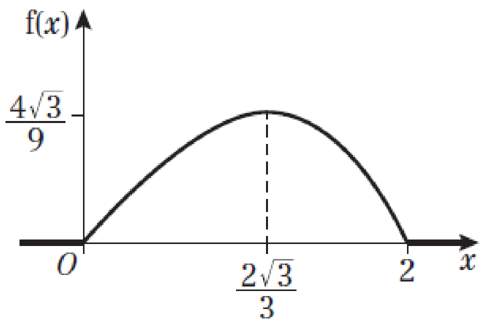
$$\Rightarrow k\left[2x^2 - \frac{1}{4}x^4\right]_0^2 = 1$$

$$\Rightarrow k(8-4) = 1$$

$$\Rightarrow 4k = 1$$

$$\Rightarrow k = \frac{1}{4}$$

b Between (0, 0) and (0, 2) the function is a cubic equation with negative x^3 coefficient. In this region the function is positive, with a local maximum. (The value of x where this local maximum occurs is found in part **d**.)



c $\mathrm{E}(X) = \int_{-\infty}^{\infty} x\,\mathrm{f}(x)\mathrm{d}x = \int_0^2 x \times \frac{1}{4}(4x - x^3)\mathrm{d}x$

$$= \left[\frac{1}{3}x^3 - \frac{1}{20}x^5\right]_0^2 = \frac{8}{3} - \frac{32}{20} = \frac{160-96}{60} = \frac{64}{60} = \frac{16}{15} = 1.07 \text{ (3 s.f.)}$$

d The mode is the value of x at the maximum of f(x), i.e. the highest point of the graph.

At the mode $\mathrm{f}'(x) = 0$, so $1 - \frac{3}{4}x = 0$

$$\Rightarrow x = \sqrt{\frac{4}{3}} = \frac{2}{\sqrt{3}} = \frac{2\sqrt{3}}{3} = 1.15 \text{ (3 s.f.)}$$

e Find the cumulative distribution function, F(x):

$\mathrm{F}(x) = \int_0^x t - \frac{1}{4}t^3\mathrm{d}t = \left[\frac{1}{2}t^2 - \frac{1}{16}t^4\right]_0^x = \frac{1}{2}x^2 - \frac{1}{16}x^4$ for $0 \leqslant x \leqslant 2$, and $\mathrm{F}(x) = 0$ otherwise

Let m be the median, then F(m) = 0.5. This gives:

$$\frac{1}{2}m^2 - \frac{1}{16}m^4 = \frac{1}{2} \Rightarrow m^4 - 8m^2 + 8 = 0$$

$$\Rightarrow m^2 = \frac{8 \pm \sqrt{64-32}}{2} = 4 \pm \sqrt{8}, \text{ so } m^2 = 4 - \sqrt{8} \text{ as } 0 \leqslant m^2 \leqslant 4$$

$$\Rightarrow m = \sqrt{1.1715\ldots} = 1.08 \text{ (3 s.f.)}$$

f Mean (1.07) < median (1.08) < mode (1.15) ⇒ negative skew

19 a The area under the probability distribution function curve must equal 1, so:

$$\int_2^3 kx(x-2)\mathrm{d}x = 1$$

$$\Rightarrow k\left[\frac{1}{3}x^3 - x^2\right]_2^3 = 1$$

$$\Rightarrow k\left(9-9-\frac{8}{3}+4\right) = 1$$

$$\Rightarrow \frac{4}{3}k = 1$$

$$\Rightarrow k = \frac{3}{4}$$

b Using $\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 \mathrm{f}(x)\mathrm{d}x$ gives:

$$\mathrm{E}(X^2) = \int_2^3 \frac{3}{4}x^3(x-2)\mathrm{d}x = \frac{3}{4}\int_2^3 (x^4 - 2x^3)\mathrm{d}x = \frac{3}{4}\left[\frac{1}{5}x^5 - \frac{1}{2}x^4\right]_2^3$$

$$= \frac{3}{4}\left(\frac{243}{5} - \frac{81}{2} - \frac{32}{5} + \frac{16}{2}\right) = \frac{3}{4}\left(\frac{211}{5} - \frac{65}{2}\right) = \frac{3}{4} \times \frac{(422-325)}{10} = \frac{291}{40}$$

So $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{291}{40} - \left(\frac{43}{16}\right)^2 = \frac{291}{40} - \frac{1849}{256}$

$$= \frac{1}{8}\left(\frac{291}{5} - \frac{1849}{32}\right) = \frac{1}{8}\left(\frac{9312-9245}{160}\right) = \frac{67}{8\times 160} = \frac{67}{1280} = 0.0523 \text{ (3 s.f.)}$$

c $\mathrm{F}(x) = \int_2^x \frac{3}{4}(t^2 - 2t)\mathrm{d}t = \left[\frac{3}{4}\left(\frac{1}{3}t^3 - t^2\right)\right]_2^x$

$$= \left(\frac{3}{4}\left(\frac{1}{3}x^3 - x^2\right) - \frac{3}{4}\left(\frac{1}{3}\times 2^3 - 2^2\right)\right) = \frac{1}{4}(x^3 - 3x^2 + 4)$$

So $\mathrm{F}(x) = \begin{cases} 0 & x < 2 \\ \frac{1}{4}(x^3 - 3x^2 + 4) & 2 \leqslant x \leqslant 3 \\ 1 & x > 3 \end{cases}$

d $\mathrm{F}(2.70) = \frac{1}{4}(2.7^3 - 3\times 2.7^2 + 4) = 0.453 \text{ (3 s.f.)}$

$\mathrm{F}(2.75) = \frac{1}{4}(2.75^3 - 3\times 2.75^2 + 4) = 0.527 \text{ (3 s.f.)}$

So $\mathrm{F}(2.70) < 0.5 < \mathrm{F}(2.75)$

As $\mathrm{F}(m) = 0.5$, therefore the median lies between 2.70 and 2.75.

20 a Find the probability density function by differentiating: $F_1'(y) = f_1(y) = 13 - 8y$

If $f_1(y)$ is a probability density function, $f_1(y) \geqslant 0$ on the interval $1 \leqslant y \leqslant 2$

But $f_1(y) < 0$ when $y > \frac{13}{8}$, so $f_1(y) < 0$ when $1.625 < y \leqslant 2$

So f_1 is not a probability density function and therefore F_1 cannot be a cumulative distribution function.

b $F_2(2) = 1$, so $k(2^4 + 2^2 - 2) = 1$

$\Rightarrow 18k = 1 \Rightarrow k = \frac{1}{18}$

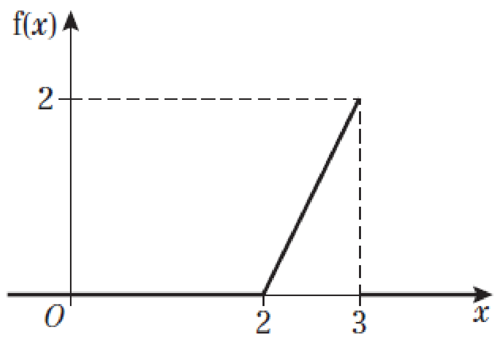
c $P(Y > 1.5) = 1 - P(Y \leqslant 1.5) = 1 - F_2(1.5) = 1 - \frac{1}{18}(1.5^4 + 1.5^2 - 2)$

$$= 1 - \frac{1}{18}\left(\frac{81}{16} + \frac{9}{4} - 2\right) = 1 - \frac{1}{18}\left(\frac{81 + 36 - 32}{16}\right) = 1 - \frac{85}{288} = \frac{203}{288} = 0.705 \text{ (3 s.f.)}$$

d $f_2(y) = \frac{dF_2(y)}{dy} = \frac{1}{18}\frac{d}{dy}(y^4 + y^2 - 2) = \frac{1}{18}(4y^3 + 2y) = \frac{1}{9}(2y^3 + y)$

Hence $f_2(y) = \begin{cases} \frac{1}{9}(2y^3 + y) & 1 \leqslant y \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$

21 a Between (2, 0) and (3, 2), f(x) is a straight line. For $x < 2$ and $x > 3$, f(x) = 0. The graph is:



b The mode is the value of x at the maximum of f(x), i.e. the highest point of the graph.
So, in this case, the mode is 3.

c Using $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$ gives:

$$E(X^2) = \int_2^3 2x^2(x-2)dx = \int_2^3 (2x^3 - 4x^2)dx = \left[\frac{1}{2}x^4 - \frac{4}{3}x^3\right]_2^3$$

$$= \frac{81}{2} - 36 - 8 + \frac{32}{3} = \frac{32}{3} - \frac{7}{2} = \frac{64}{6} - \frac{21}{6} = \frac{43}{6}$$

So $\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{43}{6} - \left(\frac{8}{3}\right)^2 = \frac{43}{6} - \frac{64}{9} = \frac{129}{18} - \frac{128}{18} = \frac{1}{18} = 0.0556$ (3 s.f.)

21 d $\text{F}(m) = \int_2^m 2(x-2)\,\text{d}x = \left[x^2 - 4x\right]_2^m = m^2 - 4m - (4-8) = m^2 - 4m + 4$

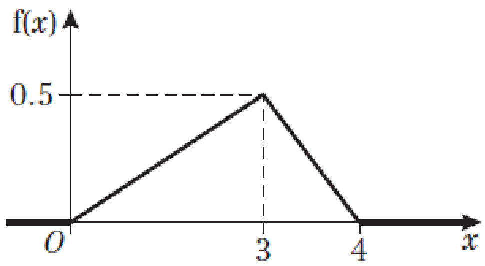
As $\text{F}(m) = 0.5$, this gives $m^2 - 4m + 4 = 0.5 \Rightarrow 2m^2 - 8m + 7 = 0$

So $m = \dfrac{8 \pm \sqrt{64-56}}{4} = \dfrac{4 \pm \sqrt{2}}{4}$

As $\dfrac{4-\sqrt{2}}{4} < 2$, it is outside the range, so $m = \dfrac{4+\sqrt{2}}{4} = 2.71$ (3 s.f.)

e Mean (2.66) < median (2.71) < mode (3) $\Rightarrow$ negative skew
It can also be inferred that the distribution is negatively skewed from the sketch (part **a**).

22 a Between (0, 0) and (3, 0.5), f(x) is a straight line with a positive gradient; between (3, 0.5) and (4, 0), f(x) is a straight line with a negative gradient; and for $x < 0$ and $x > 4$, f(x) = 0.
The graph is:



b The mode is the value of x at the maximum of f(x), i.e. the highest point of the graph.
So, in this case, the mode is 3.

c For $0 \leqslant x < 3$

$\text{F}(x) = \int_0^x \frac{1}{6}t\,\text{d}t = \frac{1}{12}x^2$

For $3 \leqslant x \leqslant 4$

$\text{F}(x) = \int_3^x (2 - \frac{1}{2}t)\,\text{d}t + \int_0^3 \frac{1}{6}t\,\text{d}t = \left(2x - \frac{1}{4}x^2\right) - \left(2 \times 3 - \frac{1}{4} \times 3^2\right) + \frac{1}{12} \times 3^2 = 2x - \frac{1}{4}x^2 - 3$

$$\text{F}(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{12}x^2 & 0 \leqslant x < 3 \\ 2x - \frac{1}{4}x^2 - 3 & 3 \leqslant x \leqslant 4 \\ 1 & x > 4 \end{cases}$$

d $\text{F}(3) = \dfrac{9}{12} = 0.75$. As $\text{F}(3) > 0.5$, the median must be between 0 and 3.

So $\text{F}(m) = \dfrac{1}{12}m^2 = 0.5$

$\Rightarrow m^2 = 6 \Rightarrow m = \sqrt{6} = 2.45$ (3 s.f.)

22 e From part **d**, P_1 lies in $0 \leqslant x < 3$ and P_9 lies in $3 \leqslant x \leqslant 4$

$$\mathrm{F}(P_9) = 0.9, \text{ so } 2P_9 - \frac{1}{4}P_9^2 - 3 = \frac{9}{10}$$

$$\Rightarrow 5P_9^2 - 40P_9 + 78 = 0$$

$$\Rightarrow P_9 = \frac{40 \pm \sqrt{1600 - 1560}}{10} = \frac{40 \pm \sqrt{40}}{10} = 4 \pm \sqrt{0.4}$$

As $4 + \sqrt{0.4} > 4$ and outside the range, $P_9 = 4 - \sqrt{0.4} = 3.36754\ldots$

$$\mathrm{F}(P_1) = 0.1, \text{ so } \frac{1}{12}P_1^2 = 0.1$$

$$\Rightarrow P_1^2 = 1.2 \Rightarrow P_1 = 1.09544\ldots$$

So $P_9 - P_1 = 3.36754 - 1.09544 = 2.27$ (3 s.f.)

23 a $\mathrm{P}(X > 0.3) = 1 - \mathrm{P}(X \leqslant 0.3) = 1 - \mathrm{F}(0.3) = 1 - (2 \times 0.3^2 - 0.3^3) = 0.847$

b $\mathrm{F}(0.59) = 2 \times (0.59)^2 - (0.59)^3 = 0.491$ (3 s.f.)

$\mathrm{F}(0.60) = 2 \times (0.6)^2 - (0.6)^3 = 0.504$

So $\mathrm{F}(0.59) < 0.5 < \mathrm{F}(0.60)$

As F(m) = 0.5, therefore the median lies between 2.70 and 2.75.

As F(m) = 0.5, therefore the median lies between 0.59 and 0.60.

c $\mathrm{f}(x) = \frac{\mathrm{dF}(x)}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}(2x^2 - x^3) = 4x - 3x^2 \qquad$ for $0 \leqslant x \leqslant 1$

$$\mathrm{f}(x) = \begin{cases} 4x - 3x^2 & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

d $$\mathrm{E}(X) = \int_0^1 x\,\mathrm{f}(x)\,\mathrm{d}x = \int_0^1 (4x^2 - 3x^3)\,\mathrm{d}x = \left[4\frac{x^3}{3} - 3\frac{x^4}{4}\right]_0^1 = \frac{7}{12} = 0.583 \text{ (3 s.f.)}$$

e To find the mode, solve $\mathrm{f}'(x) = 0$

$$\frac{\mathrm{df}(x)}{\mathrm{d}x} = 4 - 6x, \text{ so } \frac{\mathrm{df}(x)}{\mathrm{d}x} = 0 \Rightarrow x = \frac{2}{3} = 0.667 \text{ (3 s.f.)}$$

f Mean (0.583) < median ($0.59 < m < 0.6$) < mode (0.667) ⇒ negative skew

24 a The area under the probability distribution function curve must equal 1, so:

$$\int_0^2 k\,\mathrm{d}x + \int_2^4 \frac{k}{x}\,\mathrm{d}x = 1$$

$$\Rightarrow \left[kx\right]_0^2 + \left[k\ln x\right]_2^4 = 1$$

$$\Rightarrow 2k + k(\ln 4 - \ln 2) = 1$$

$$\Rightarrow 2k + k\ln 2 = 1$$

$$\Rightarrow k = \frac{1}{2 + \ln 2}$$

24 b $E(X)=\int_{-\infty}^{\infty} x\,f(x)\,dx=\frac{1}{2+\ln 2}\int_0^2 x\,dx+\frac{1}{2+\ln 2}\int_2^4 1\,dx$

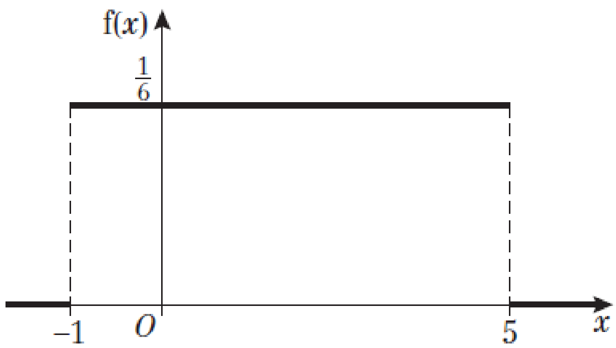
$$=\frac{1}{2+\ln 2}\left[\frac{1}{2}x^2\right]_0^2+\frac{1}{2+\ln 2}\left[x\right]_2^4=\frac{1}{2+\ln 2}(2+4-2)$$

$$=\frac{4}{2+\ln 2}=1.49\text{ (3 s.f.)}$$

25 a $\frac{1}{b-a}=\frac{1}{5-(-1)}=\frac{1}{6}$

So between $\left(-1,\frac{1}{6}\right)$ and $\left(5,\frac{1}{6}\right)$, f($x$) is a horizontal straight line, for $x<-1$ and $x>5$, f(x) = 0.

The graph is:



b $E(X)=\frac{b+a}{2}=\frac{5+(-1)}{2}=2$

c $\text{Var}(X)=\frac{(b-a)^2}{12}=\frac{(5+1)^2}{12}=3$

d $P(-0.3<X<3.3)=(3.3-(-0.3))\times\frac{1}{6}=\frac{3.6}{6}=0.6$

26 a $\frac{1}{b-a}=\frac{1}{6-2}=\frac{1}{4}$

So:

$$f(x)=\begin{cases}\frac{1}{4} & 2\leqslant x\leqslant 6\\ 0 & \text{otherwise}\end{cases}$$

b $E(X)=\frac{b+a}{2}=\frac{6+2}{2}=4$

c $\text{Var}(X)=\frac{(b-a)^2}{12}=\frac{(6-2)^2}{12}=\frac{16}{12}=\frac{4}{3}$

26 d $\text{F}(x)=\int_2^x \frac{1}{4}\text{d}t=\left[\frac{1}{4}t\right]_2^x=\frac{1}{4}(x-2)$

So:

$$\text{F}(x)=\begin{cases}0 & x<2\\ \frac{1}{4}(x-2) & 2\leqslant x\leqslant 6\\ 1 & x>6\end{cases}$$

e $\text{P}(2.3<X<3.4)=\frac{1}{4}(3.4-2.3)=\frac{11}{40}=0.275$

Alternative method:

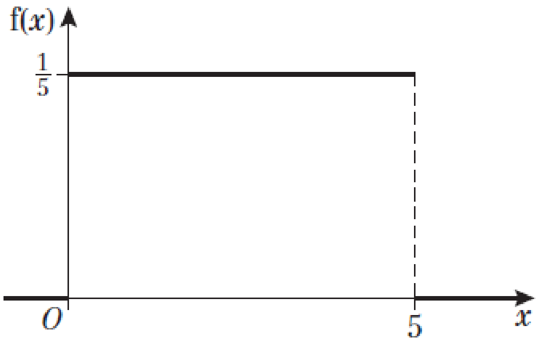
$\text{P}(2.3<X<3.4)=\text{F}(3.4)-\text{F}(2.3)=\frac{1}{4}(3.4-2)-\frac{1}{4}(2.3-2)=0.275$

27 a The probability distribution of X is a continuous uniform distribution

$\frac{1}{b-a}=\frac{1}{5-0}=\frac{1}{5}$

So between $\left(0,\frac{1}{5}\right)$ and $\left(5,\frac{1}{5}\right)$, f($x$) is a horizontal straight line, for $x<0$ and $x>5$, f(x) = 0.

The graph is:



b $\text{E}(X)=\frac{b+a}{2}=\frac{5+0}{2}=2.5\,\text{cm}$

$\text{Var}(X)=\frac{(b-a)^2}{12}=\frac{5^2}{12}=\frac{25}{12}=2.083$ (3 d.p.)

c $\text{P}(X>3)=(5-3)\times\frac{1}{5}=\frac{2}{5}$

d $\text{P}(X=3)=0$

28 a $$\text{f}(x)=\begin{cases}\frac{1}{\beta-\alpha} & \alpha\leqslant x\leqslant\beta\\ 0 & \text{otherwise}\end{cases}$$

28 b $\text{E}(X) = 2 \Rightarrow \dfrac{\beta + \alpha}{2} = 2 \Rightarrow \beta + \alpha = 4 \Rightarrow \beta = 4 - \alpha$

$\text{P}(X < 3) = \dfrac{5}{8} \Rightarrow \dfrac{3-\alpha}{\beta-\alpha} = \dfrac{5}{8} \Rightarrow 24 - 8\alpha = 5\beta - 5\alpha \Rightarrow 5\beta = 24 - 3\alpha$

Substituting for from β the first equation gives:

$5(4-\alpha) = 24 - 3\alpha \Rightarrow 20 - 24 = 2\alpha \Rightarrow \alpha = -2$

So $\beta = 4 - \alpha = 4 - (-2) = 6$

Solution is $\alpha = -2,\ \beta = 6$

29 a $\text{E}(X) = \dfrac{150 + 0}{2} = 75\,\text{cm}$

b $\text{Var}(X) = \dfrac{150^2}{12} \Rightarrow \text{Standard deviation} = \dfrac{150}{\sqrt{12}} = 43.3 \text{ (3 s.f.)}$

c If the piece of string with the ring on it is longer than 150 cm, then the shorter piece of wire is at most 30 cm, and if the piece of string with the ring on it is shorter than or equal to 30 cm, then by definition the shorter length of wire is at most 30 cm long. So the required probability is:

$\text{P}(X \leqslant 30) + \text{P}(X \geqslant 120) = \dfrac{30}{150} + \dfrac{30}{150} = \dfrac{60}{150} = \dfrac{2}{5}$

30 a $\text{P}(F < 3B) = \text{P}(F - 3B < 0)$

Let $X = F - 3B$

$\text{E}(X) = \text{E}(F) - 3\text{E}(B) = 238 - 3 \times 82 = -8$

$\text{Var}(X) = \text{Var}(F) + 3^2\text{Var}(B) = 7^2 + 3^2 \times 3^2 = 130$

Hence $X \sim \text{N}(-8, 130)$

$\text{P}(X < 0) = 0.7586$ (4 d.p.) (from calculator or tables)

b The assumption made is that the duration of the two rides are independent.
Validity: this is likely to be the case – two separate control panels operate each ride.

c $D = B_1 + B_2 + B_3$

$\text{E}(D) = \text{E}(B) + \text{E}(B) + \text{E}(B) = 3 \times 82 = 246$

$\text{Var}(D) = \text{Var}(B) + \text{Var}(B) + \text{Var}(B) = 3 \times 3^2 = 27$

Hence $D \sim \text{N}(246, 27)$

d $\text{P}(|F - D| < 10) = \text{P}(-10 < F - D < 10)$

Let $Y = F - D$

$\text{E}(Y) = 238 - 246 = -8$

$\text{Var}(Y) = 49 + 27 = 76$

Hence $Y \sim \text{N}(-8, 76)$

$\text{P}(-10 < Y < 10) = \text{P}(Y < 10) - \text{P}(Y < -10) = 0.5713$ (4 d.p.)

31 a $\text{E}(R) = \text{E}(X) + \text{E}(Y) = 20 + 10 = 30$

b $\text{Var}(R) = \text{Var}(X) + \text{Var}(Y) = 4 + 0.84 = 4.84$

31 c $R \sim N(30, 4.84)$ from parts **a** and **b**
$P(28.9 < R < 32.64) = P(R < 32.64) - P(R < 28.9) = 0.8849 - 0.3085 = 0.5764$ (4 d.p.)

32 a Let X be the total weight of 4 randomly chosen adult men.
$X = M_1 + M_2 + M_3 + M_4$
$E(X) = 4 \times 84 = 336$
$\text{Var}(X) = 4 \times 11^2 = 484$
$X \quad N(336, 484)$
$P(X < 350) = 0.7377$ (4 d.p.)

b Let $M \sim N(84, 121)$ and $W \sim N(62, 100)$ and $Y = M - 1.5W$
$E(Y) = 84 - 1.5 \times 62 = -9$
$\text{Var}(Y) = \text{Var}(M) + 1.5^2\text{Var}(W) = 11^2 + 1.5^2 \times 10^2 = 346$
So $Y \sim N(-9, 346)$, and $P(Y < 0) = 0.6858$ (4 d.p.)

33 a $E(D) = E(A) - 3E(B) + 4E(C) = 5 - 3 \times 7 + 4 \times 9 = 20$
$\text{Var}(D) = \text{Var}(A) + 3^2\ \text{Var}(B) + 4^2\ \text{Var}(C) = 2^2 + 9 \times 3^2 + 16 \times 4^2 = 341$
So $D \sim N(20, 341)$, and $P(D < 44) = 0.9031$ (4 d.p.)

b $E(X) = E(A) - 3E(B) + 4E(C) = 5 - 3 \times 7 + 4 \times 9 = 20$
$\text{Var}(D) = \text{Var}(A) + \text{Var}(B) + \text{Var}(B) + \text{Var}(B) + 4^2\ \text{Var}(C) = 2^2 + 3 \times 3^2 + 16 \times 4^2 = 287$
So $X \sim N(20, 287)$, and $P(X > 0) = 1 - P(X \leqslant 0) = 1 - 0.1189 = 0.8811$ (4 d.p.)

34 a Let $W = C_1 - C_2$
$E(W) = 350 - 350 = 0$
$\text{Var}(W) = 8 + 8 = 16$
So $W \sim N(0, 16)$
$P(|W| > 6) = 1 - P(W < 6) + P(W < -6) = 0.0668 + 0.0668 = 0.1336$ (4 d.p.)

b Let $X = C - L$
$E(X) = 350 - 345 = 5$
$\text{Var}(X) = 8 + 17 = 25$
So $X \sim N(5, 25)$
$P(X > 0) = 1 - P(X < 0) = 1 - 0.1587 = 0.8413$ (4 d.p.)

c Let $Y = \sum_{i=1}^{24} C_i + B$
$E(Y) = 24 \times 350 + 100 = 8500$
$\text{Var}(Y) = 24 \times 8 + 2^2 = 196$
So $Y \quad N(8500, 196)$
$P(8510 < Y < 8520) = P(Y < 8520) - P(Y < 8510) = 0.92343 - 0.76247 = 0.1610$ (4 d.p.)

d All random variables (each can of cola and the box) are independent and normally distributed.

Challenge

1 a i From calculator using suitable degrees of accuracy: $y = -2.63 + 2.285x$

ii From calculator using suitable degrees of accuracy: $y = 1.04 + 0.1206x + 0.2353x^2$

iii From calculator using suitable degrees of accuracy: $y = 1.1762e^{0.3484x}$

b This table sets out the residuals for each data point (x, y) for each of the three models:

x	y	**Linear**	ε	**Quadratic**	ε	**Exponential**	ε
1	1.5	–0.345	1.845	1.3959	0.1041	1.66644	–0.16644
3	3.3	4.225	–0.925	3.5195	–0.2195	3.34507	–0.04507
4	5.3	6.51	–1.21	5.2872	0.0128	4.739297	0.560703
5	7.5	8.795	–1.295	7.5255	–0.0255	6.714631	0.785369
7	13.8	13.365	0.435	13.4139	0.3861	13.478407	0.321593
8	16.8	15.65	1.15	17.064	–0.264	19.09619	–2.29619

The quadratic model is the most suitable: the residuals are smaller and randomly scattered about zero.

2 a The area under the probability distribution function curve must equal 1, so:

$$\int_0^\infty ke^{-x}\,dx = k\left[-e^{-x}\right]_0^\infty = k(0-(-1)) = k \Rightarrow k = 1$$

b $\int_0^x e^{-t}\,dt = \left[-e^{-t}\right]_0^x = -e^{-x} - (-1) = 1 - e^{-x}$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geqslant 0 \end{cases}$$

c $P(1 < X < 4) = P(X < 4) - P(X < 1) = F(4) - F(1)$

$$= (1 - e^{-4}) - (1 - e^{-1}) = e^{-1} - e^{-4} = \frac{e^3 - 1}{e^4}$$

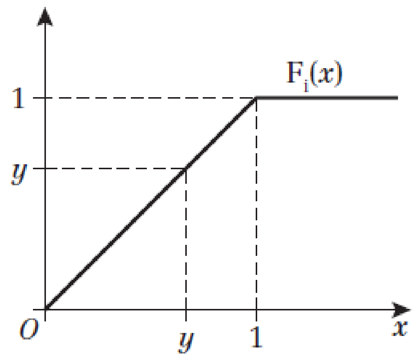
Challenge

3 $X_i \sim \text{U}[0,1]$ $\quad Y = \max(X_i)$

a

$$\text{f}_i(x) = \begin{cases} 0 & x < 0 \\ 1 & 0 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$$

$$\text{F}_i(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leqslant x \leqslant 1 \\ 1 & x > 1 \end{cases}$$



$$\text{F}_Y(y) = \text{P}\big((X_1 \leqslant y) \cap (X_2 \leqslant y) \cap \ldots \cap (X_n \leqslant y)\big)$$
$$= y^n$$

So $\text{f}_Y(y) = \dfrac{\text{d}}{\text{d}y}(y^n) = ny^{n-1}$

Hence, $\text{E}(Y) = \int_{-\infty}^{\infty} y\text{f}_Y(y)\text{d}y = \int_0^1 ny^n \text{d}y = \dfrac{n}{n+1}\left[y^{n+1}\right]_0^1 = \dfrac{n}{n+1}$

b Median of Y is the value m such that $\text{F}_Y(m) = 0.5$

$\Rightarrow \text{F}_Y(m) = m^n = 0.5 \Rightarrow m = \sqrt[n]{0.5}$

c Let $Y = X_1 + X_2$

$\text{f}_Y(z) = \int_{-\infty}^{\infty} \text{f}_1(z-t)\text{f}_2(t)\text{d}t = \int_0^1 \text{f}_1(z-t)\text{d}t$ since $\text{f}_2(t) = 0$ for values of t outside the interval $[0,1]$

$\text{f}_1(z-t)$ is only non-zero for $0 \leqslant z - t \leqslant 1 \Rightarrow z - 1 \leqslant t \leqslant z$

Since X_i takes values in [0, 1], Y takes values in [0, 2].

Consider $0 \leqslant z \leqslant 1$ so $z - 1 \leqslant 0 \leqslant t \leqslant z \leqslant 1$:

$$\text{f}_Y(z) = \int_0^z \text{f}_1(z-t)\text{d}t = \int_0^z \text{d}t = [t]_0^z = z$$

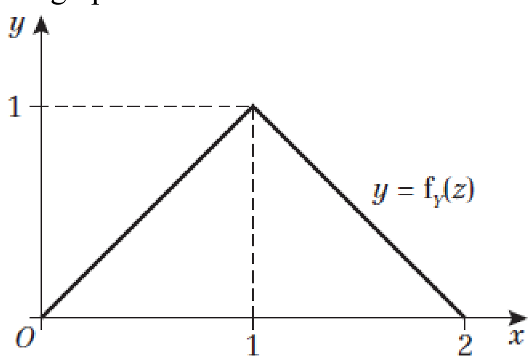
Consider $1 \leqslant z \leqslant 2$ so $0 < z - 1 \leqslant t \leqslant 1 \leqslant z$:

$$\text{f}_Y(z) = \int_{z-1}^1 \text{f}_1(z-t)\text{d}t = \int_{z-1}^1 \text{d}t = [t]_{z-1}^1 = 1 - (z-1) = 2 - z$$

So:

$$\text{f}_Y(z) = \begin{cases} z & 0 \leqslant z \leqslant 1 \\ 2 - z & 1 < z \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

The graph is:



Challenge

3 d Let $Z = Y_1 + X_3$ where Y is defined as in part **c**. So:

$$f_Y(z) = \begin{cases} z & 0 \leqslant z \leqslant 1 \\ 2 - z & 1 < z \leqslant 2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathrm{f}_3(x) = \begin{cases} 1 & 0 \leqslant x \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

$\mathrm{f}_Z(z) = \int_{-\infty}^{\infty} \mathrm{f}_Y(z-t)\mathrm{f}_3(t)\mathrm{d}t = \int_0^1 \mathrm{f}_Y(z-t)\mathrm{d}t$ since $\mathrm{f}_3(t) = 0$ for values of t outside $[0,1]$

$\mathrm{f}_Y(z-t)$ is only non-zero for $0 \leqslant z - t \leqslant 2 \Rightarrow z - 2 \leqslant t \leqslant z$

Since Y takes values in [0, 2] and X_3 takes values in [0, 1], Z takes values in [0, 3].

Consider $0 \leqslant z \leqslant 1$ so $z - 2 \leqslant -1 < 0 \leqslant t \leqslant z \leqslant 1$:

$$\mathrm{f}_Z(z) = \int_0^z (z-t)\mathrm{d}t = \left[zt - \tfrac{1}{2}t^2\right]_0^z = \frac{1}{2}z^2$$

Consider $1 \leqslant z \leqslant 2$ so $z - 2 \leqslant 0 \leqslant t \leqslant 1 < z$:

$$\mathrm{f}_Z(z) = \int_0^{z-1} (2 - z + t)\mathrm{d}t + \int_{z-1}^1 (z-t)\mathrm{d}t = \left[2t - zt + \frac{1}{2}t^2\right]_0^{z-1} + \left[zt - \frac{1}{2}t^2\right]_{z-1}^1$$

$$= 2(z-1) - z(z-1) + \frac{1}{2}(z^2 - 2z + 1) + z - \frac{1}{2} - \left(z(z-1) - \frac{1}{2}(z^2 - 2z + 1)\right)$$

$$= 2z - 2 - z^2 + z + \frac{1}{2}(z^2 - 2z + 1) + z - \frac{1}{2} - z^2 + z + \frac{1}{2}(z^2 - 2z + 1)$$

$$= 5z - \frac{5}{2} - 2z^2 + z^2 - 2z + 1 = -z^2 + 3z - \frac{3}{2}$$

$$= -\left(z - \frac{3}{2}\right)^2 + \frac{9}{4} - \frac{3}{2} = \frac{3}{4} - \left(z - \frac{3}{2}\right)^2$$

Consider $2 \leqslant z \leqslant 3$ so $0 < z - 2 \leqslant t \leqslant 1 < 2 < z$:

$$\mathrm{f}_Z(z) = \int_{z-2}^1 (2 - z + t)\mathrm{d}t = \left[2t - zt + \frac{1}{2}t^2\right]_{z-2}^1$$

$$= 2 - z + \frac{1}{2} - \left(2(z-2) - z(z-2) + \frac{1}{2}(z^2 - 4z + 4)\right)$$

$$= 2 - z + \frac{1}{2} - 2z + 4 + z^2 - 2z - \frac{1}{2}z^2 + 2z - 2$$

$$= \frac{1}{2}z^2 - 3z + \frac{9}{2} = \frac{1}{2}(z^2 - 6z + 9) = \frac{1}{2}(z-3)^2$$

So:

$$\mathrm{f}_Z(z) = \begin{cases} \frac{1}{2}z^2 & 0 \leqslant z \leqslant 1 \\ \frac{3}{4} - (z - \frac{3}{2})^2 & 1 < z \leqslant 2 \\ \frac{1}{2}(z-3)^2 & 2 < z \leqslant 3 \\ 0 & \text{otherwise} \end{cases}$$

Challenge

3 d (continued)

Between (0, 0) and (1, 0.5), $f_Z(z)$ is a positive quadratic; between (1, 0.5) and (2, 0.5), $f_Z(z)$ is a negative quadratic with a maximum at (1.5, 0.75); between (2, 0.5) and (3, 0), $f_Z(z)$ is a positive quadratic; otherwise $f_Z(z)$ is zero.
The graph is: