Question 1 (25 marks)

The random variable X has a discrete distribution. The probability that is takes a value other than 13, 14, 15 or 16 is negligible.

(a) Complete the probability table below and hence calculate $E(X)$, the expected value of X.

x	13	14	15	16
$P(X = x)$	0.383	0.575	**0.038**	0.004

Note that the sum of the probabilities must be 1. Therefore $P(15) = 1 - 0.383 - 0.575 - 0.004 = 0.038$.

$E(X) = 13(0.383) + 14(0.575) + 15(0.038) + 16(0.004) = 13.663.$

(b) If X is the age, in complete years, on 1 January 2013 of a student selected at random from among all second-year students in Irish schools, explain what $E(X)$ represents.

$E(X)$ represents the mean of the ages of all second-year students in Irish schools on 1 January 2013.

(c) If ten students are selected at random from this population, find the probability that exactly six of them were 14 years old on 1 January 2013. Give your answer correct to three significant figures.

We have 10 Bernoulli trials with $p = P(X = 14) = 0.575$. So the probability of exactly six successes is given by

$$
\begin{aligned}
\binom{10}{6} p^6 (1-p)^4 &= \frac{10!}{6!4!}(0.575)^6(1-0.575)^4 \\
&= 210(0.575)^6(0.425)^4 \\
&= 0.248
\end{aligned}
$$

correct to three significant places.
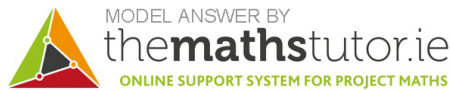
Question 2 (25 marks)

(a) Explain what is meant by *stratified sampling* and *cluster sampling*. Your explanation should include:

- a clear indication of the difference between the two methods
- one reason why each method might be chosen instead of simple random sampling.

In *stratified sampling* the population is divided into strata and a sample is taken by applying random sampling within each of the strata. We usually ensure that the size of the sample from each strata is proportional to the size of the strata relative to the overall population. One reason why this method might be chosen is to ensure that each strata of the population is represented proportionately in the overall sample.

In *cluster sampling* the population is divided in groups and then we choose a ia simple random sample of the groups. Then we create our sample by choosing simple random samples only from the selected groups (sometimes we might create the sample by choosing all of the selected groups). One reason why this method might be chosen instead of a simple random sample is that it can reduce the cost of sampling, since we restrict our sample to only the selected groups rather than the entire population.
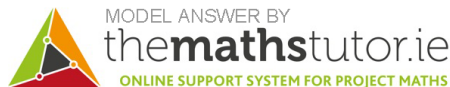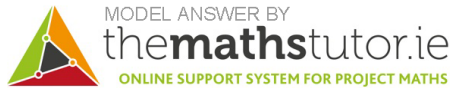
(b) A survey is being conducted of voters' opinions on several different issues.

(i) What is the overall margin of error of the survey, at 95% confidence, if it is based on a simple random sample of 1111 voters.

The margin of error is

$$\frac{1}{\sqrt{1111}} = 0.03000$$

correct to 4 significant places, or 3.000%.

(ii) A political party had claimed that it has the support of 23% of the electorate. Of the voters in the sample above, 234 stated that they support the party. Is this sufficient evidence to reject the party's claim, at the 5% level of significance?

H_0: The party has the support of 23% of the electorate, or $p = 0.23$.
The standard error of the sampling distribution, assuming that H_0 is true, is given by

$$s = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.23(0.77)}{1111}} = 0.01263$$

correct to four significant places. Now the observed sample proportion is

$$\hat{p} = \frac{234}{1111} = 0.2106.$$

The z-score corresponding to 0.2106 is

$$\frac{\hat{p} - p}{s} = \frac{0.2106 - 0.23}{0.012623} = -1.5360.$$

At the 5% level of significance, we reject the null hypothesis if this z-score lies outside the interval $[-1.96, 1.96]$. However, the calculated z-score lies inside this interval. This means that we do not have enough evidence at the 5% level of significance to reject the party's claim.
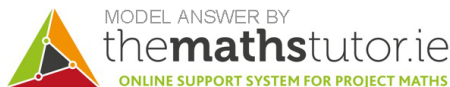
Question 3 (25 marks)

(a) Show that, for all $k \in \mathbb{R}$, the point $P(4k-2, 3k+1)$ lies on the line $l_1 : 3x - 4y + 10 = 0$.
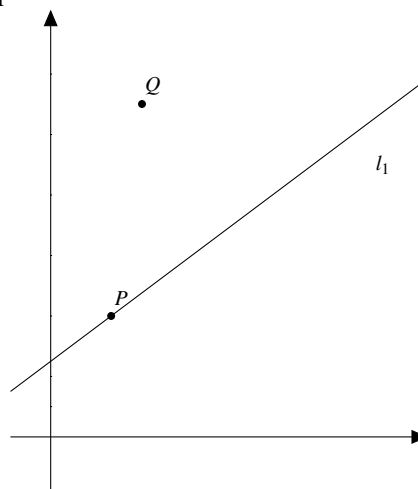
If $(x, y) = (4k-2, 3k+1)$ then

$$
\begin{aligned}
3x - 4y + 10 &= 3(4k-2) - 4(3k+1) + 10 \\
&= 12k - 6 - 12k - 4 + 10 \\
&= 0
\end{aligned}
$$

So the equation of l_1 is satisfied. Therefore $(4k-2, 3k+1)$ lies on l_1.

(b) The line l_2 passes through P and is perpendicular to l_1. Find the equation of l_2 in terms of k.

We have

$$
\begin{aligned}
3x - 4y + 10 &= 0 \\
\Leftrightarrow \\
-4y &= -3x - 10 \\
\Leftrightarrow \\
y &= \frac{3}{4}x + \frac{5}{2}
\end{aligned}
$$

Therefore the slope of l_1 is $\frac{3}{4}$. Therefore the slope of l_2 is $\frac{1}{\frac{3}{4}} = -\frac{4}{3}$. So l_2 has slope $-\frac{4}{3}$ and passes through $(4k-2, 3k+1)$. So it has equation

$$
y - (3k+1) = -\frac{4}{3}(x - (4k-2))
$$

or

$$
3y - 3(3k+1) = -4(x - (4k-2)).
$$

Rearranging this gives

$$
4x + 3y - 25k + 5 = 0.
$$

(c) Find the value of k for which l_2 passes through the point $Q(3, 11)$.

The equation of l_2 is
$$4x + 3y - 25k + 5 = 0.$$

Now $(3, 11)$ lies on l_2 if and only if $4(3) + 3(11) - 25k + 5 = 0 \Leftrightarrow 25k = 50 \Leftrightarrow k = 2$. So the $k = 2$ is the required value.

(d) Hence, or otherwise, find the co-ordinates of the foot of the perpendicular from Q to l_1.

When $k = 2$ the equation of l_2 is

$$4x + 3y - 45 = 0.$$

So to find the required point, we solve

$$3x - 4y + 10 = 0$$
$$4x + 3y - 45 = 0$$

simultaneously.
This is equivalent to

$$12x - 16y + 40 = 0$$
$$12x + 9y - 135 = 0$$

Subtracting yields
$$-25y + 175 = 0.$$

Therefore $25y = 175$ and $y = \frac{175}{25} = 7$.
Now $3x - 4(7) + 10 = 0 \Leftrightarrow 3x = 4(7) - 10 = 18 \Leftrightarrow x = 6$. So the foot of the perpendicular from Q to l_1 has co-ordinates $(6, 7)$.
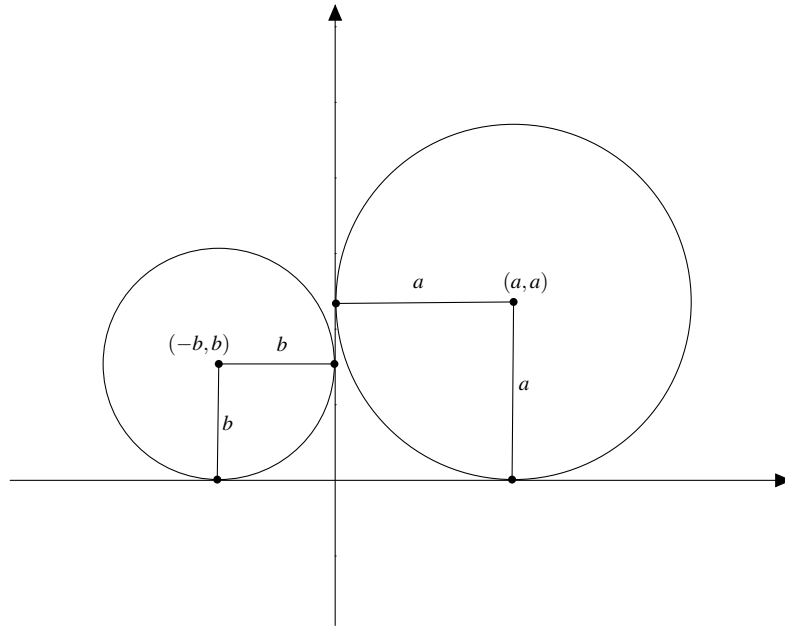
Question 4 (25 marks)

The centre of a circle lies on the line $x + 2y - 6 = 0$. The x-axis and the y-axis are tangents to the circle. There are two circles that satisfy these conditions. Find their equations.

Consider the diagram below:

We can see from this diagram that if (x, y) is the centre of a circle that has both the x-axis and the y-axis as tangents, then either

- Case 1: $y = x$

- Case 2: $y = -x$

In either case the radius r is $\pm x$. Since the radius is positive, we must have $r = |x|$.
Case 1: $y = x$. We are also told that $x + 2y - 6 = 0$. Subsituting x for y in the latter equation gives

$$x + 2x - 6 = 0 \Leftrightarrow 3x - 6 = 0 \Leftrightarrow x = 2.$$

Now $y = x$ so $y = 2$. Therefore the centre of the circle has co-ordinates $(2, 2)$ and the radius is 2. Therefore in this case the circle has equation

$$(x - 2)^2 + (y - 2)^2 = 4.$$

Case 2: $y = -x$. As before we use this to substitute $-x$ for y in the equation $x + 2y - 6 = 0$. This gives

$$x + 2(-x) - 6 = 0 \Leftrightarrow -x - 6 = 0 \Leftrightarrow x = -6.$$

It follows that $y = -(-6) = 6$. So in this case the centre has co-ordinates $(-6, 6)$ and the radius is 6. So this circle has equation

$$(x + 6)^2 + (y - 6)^2 = 36.$$

Question 5 (25 marks)

The diagram below shows the graph of the function $f : x \mapsto \sin 2x$. The line $2y = 1$ is also shown.

The graph of f is in blue. The graph of g is in green and the graph of h is in red.
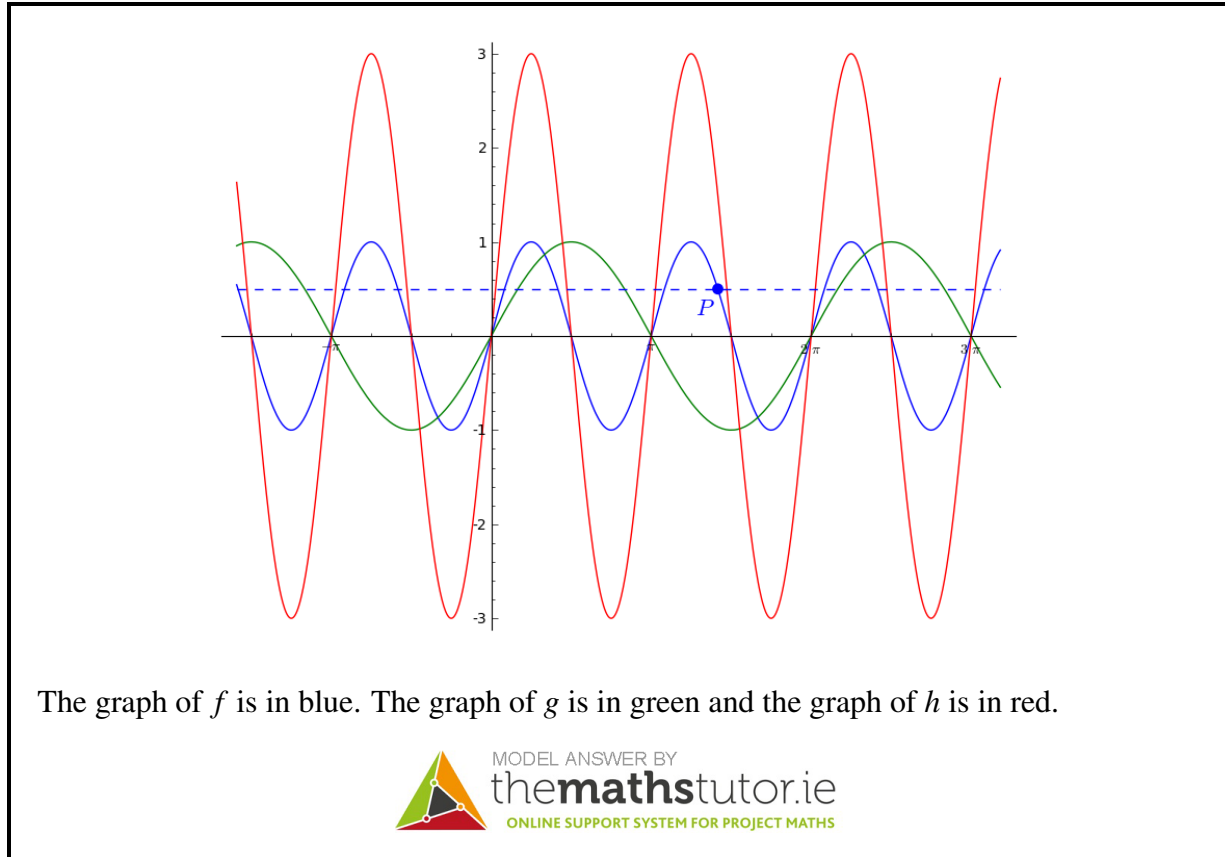
(a) On the same diagram above sketch the graphs of $g : x \mapsto \sin x$ and $h : x \mapsto 3\sin 2x$. Indicate clearly which is g and which is h.

(b) Find the co-ordinates of the point P in the diagram.

P is a point of intersection of the line $y = \frac{1}{2}$ and the curve $y = \sin 2x$. So the x-co-ordinate of P is a solution of the equation $\sin 2x = \frac{1}{2}$.

Therefore either

$$2x = \frac{\pi}{6} + 2n\pi, n \in \mathbb{Z}$$

or

$$2x = \frac{5\pi}{6} + 2n\pi, n \in \mathbb{Z}.$$

In the first case, we get

$$x = \frac{\pi}{12} + n\pi, n \in \mathbb{Z}$$

and in the second case we get

$$x = \frac{5\pi}{12} + n\pi, n \in \mathbb{Z}.$$

So x is one of the following numbers

$$\ldots, \frac{\pi}{12}, \frac{5\pi}{12}, \frac{13\pi}{12}, \frac{17\pi}{12}, \frac{25\pi}{12}, \frac{29\pi}{12}, \ldots$$
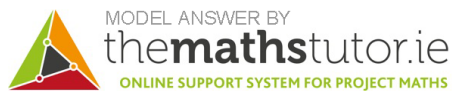
From the diagram we see that

$$\frac{5\pi}{4} \le x \le \frac{3\pi}{2}$$

so the only possibility is that

$$x = \frac{17\pi}{12}.$$

Clearly the y-co-ordinate of P is $\frac{1}{2}$ since it lies on the line $2y = 1$. So the co-ordinates of P are $\left(\frac{17\pi}{12}, \frac{1}{2}\right)$.

Question 6A (25 marks)

Explain, with the aid of an example, what is meant by *proof by contradiction*.

Note: you do not need to provide the full proof involved in your example. Give sufficient outline to illustrate how contradiction is used.

Explanation:

To prove a statement by contradiction, we assume that the statement is false and then prove that this assumption contradicts another statement that is known to be true.
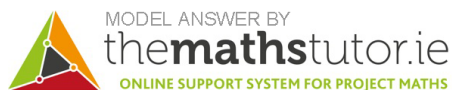
Example:

Consider the statement

> The lines $y = x$ and $y = x + 1$ do not intersect.

We can prove this by contradiction as follows.
Assume that that the statement is false. So there is some point (a, b) that lies on both lines.
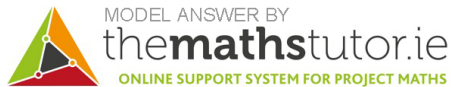Therefore $b = a$ since the point is on the line $y = x$ and $b = a + 1$ since the point is on the line $y = x + 1$.
Combining these equation, we get

$$a = a + 1$$

which implies that

$$0 = 1.$$

However this <u>contradicts</u> that fact that $0 \neq 1$. Therefore our original assumption is false.
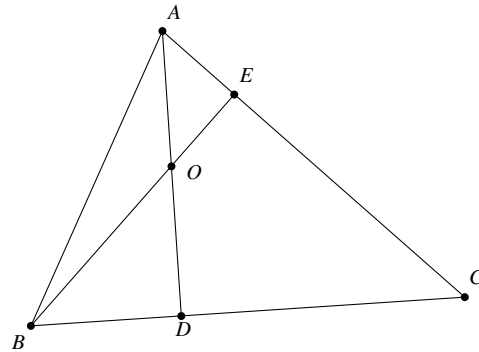So we have proved that the lines $y = x$ and $y = x + 1$ do not intersect.

Question 6B (25 marks)

ABC is a triangle.

D is the point on *BC* such that $AD \perp BC$. *E* is the point on *BC* such that $BE \perp AC$.

AD and *BE* intersect at *O*.

Prove that $|\angle DOC| = |\angle DEC|$.

Consider the quadrilateral *DOEC*. We have

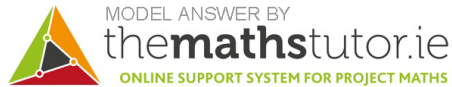$$|\angle CDO| + |\angle OEC| = 90° + 90° = 180°.$$

Therefore *DOEC* is a cyclic quadrilateral (by the converse of Corollary 5).
Therefore $\angle DOC$ and $\angle DEC$ are angles standing on the same arc of a circle (the circumcircle of *DOEC*).
Therefore, by Theorem 19

$$|\angle DOC| = |\angle DEC|$$

as required.

Section B	**Contexts and Applications**	**150 marks**

Question 7   **(75 marks)**

The *King of the Hill* triathlon race in Kinsale consists of a 750 metre swim, followed by a 20 kilometre cycle, followed by a 5 kilometre run.

The questions below are based on the data from the 224 athletes who completed this triathlon in 2010.

Máire is analysing data from the race, using statistical software. She has a data file with each competitor's time for each part of the race, along with various other details of the competitors.


Lizzie Lee, winner of the women's event

Máire gets the software to produce some summary statistics and it produces the following table. Three of the entries in the table have been removed and replaced with question marks (?).

|	Swim	Cycle	Run
Mean	18.329	41.927	?
Median	17.900	41.306	?
Mode	#N/A	#N/A	#N/A
Standard Deviation	?	4553	3409
Sample Variance	10.017	20.729	11.622
Skewness	1.094	0.717	0.463
Range	19.226	27.282	20.870
Minimum	11.350	31.566	16.466
Maximum	30.576	58.847	37.336
Count	224	224	224

Máire produces histograms of the times for the three events. Here are the histograms, without their titles.
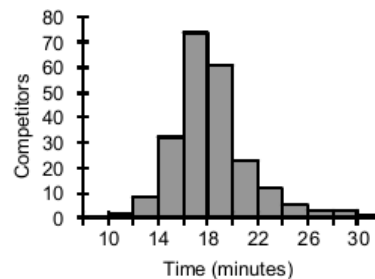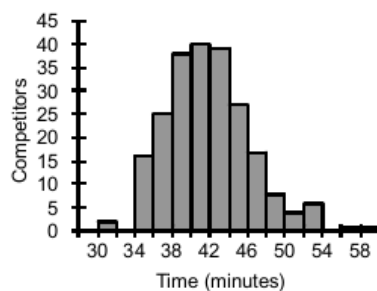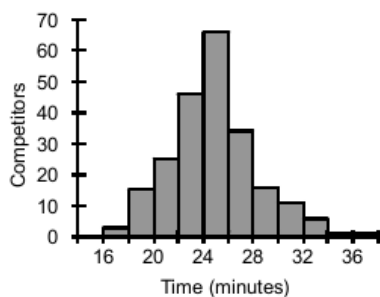
(a) (i) Use the summary statistics in the table to decide which histogram corresponds to each event. Write the answers above the histograms.

We can use the minimum and maximum values to match the histograms to the events.
The leftmost histogram has a minimum between 16 and 18. From the table we see that the only event that matches this is the run.
The rightmost histogram has a minimum value somewhere between 10 and 12, so that must be the swim
Therefore the middle histogram must be the cycle.

MODEL ANSWER BY
the**maths**tutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS
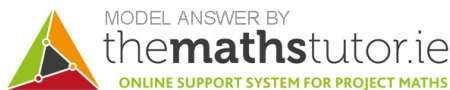
(ii) The mean and the median time for the run are approximately equal. Estimate this value from the corresponding histogram.
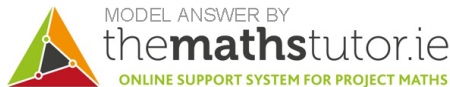
mean ≈ median ≈ [25 minutes.]
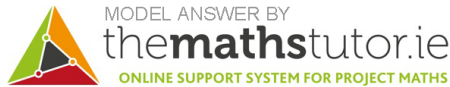
(iii) Estimate from the relevant histogram the standard deviation of the times for the swim.
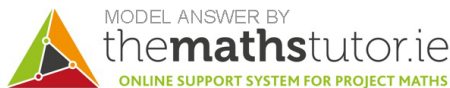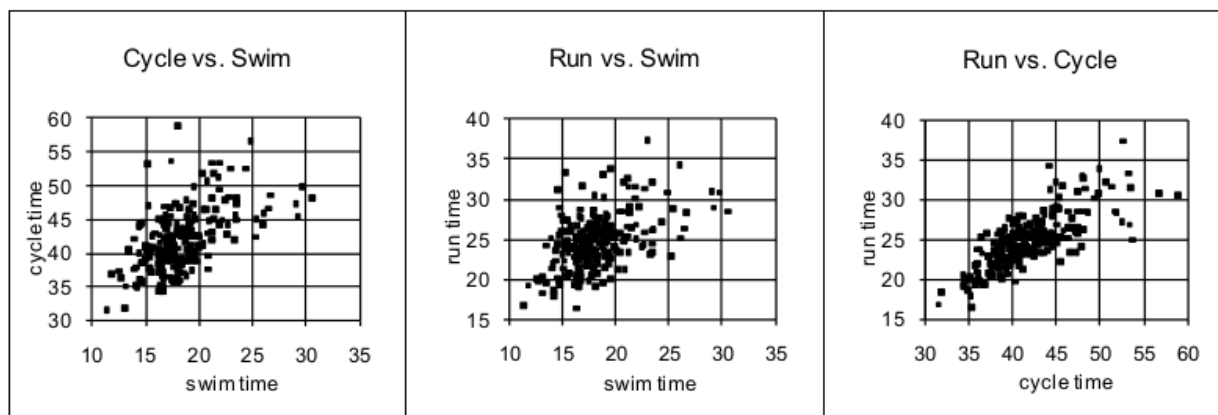
standard deviation ≈ | 3 minutes |

(iv) When calculating the summary statistics, the software failed to find a *mode* for the data sets. Why do you think this is?

Máire is interested in the relationships between the athletes' performance in the three different events. She produces the following three scatter diagrams.

(b) Give a brief summary of the relationship between performance in the different events, based on the scatter diagrams.

> There is a strong positive correlation between the run times and the cycle times.
> There is a positive correlation between the run times and the swim times.
> There is a positive correlation between the cycle times and the swim times.
>
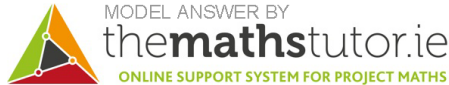
(c) The best-fit line for the run-time based on swim-time is $y = 0.53x + 15.2$. The best-fit line for run-time based on cycle-time is $y = 0.58x + 0.71$. Brian did the swim in 17.6 minutes and the cycle in 35.7 minutes. Give your best estimate of Brian's time for the run, and justify your answer.

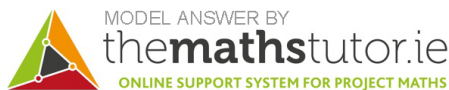> Based on his swim-time we estimate his run-time to be
>
> $$0.53(17.6) + 15.2 = 24.528.$$
>
> Based on his cycle-time we estimate his run-time to be
>
> $$0.58(35.7) + 0.71 = 21.416.$$
>
> It seems reasonable to take the mean of these two estimates as our best guess. Therefore we estimate his run-time to be
>
> $$\frac{24.528 + 21.416}{2} = 22.972 \text{ minutes.}$$
>

The mean finishing time for the overall event was 88.1 minutes and the standard deviation was 10.3 minutes.

(d) Based on an assumption that the distribution of overall finishing times is approximately normal, use the *empirical rule* to complete the following sentence:

"95% of the athletes took between $\boxed{67.5}$ and $\boxed{108.7}$ minutes to complete the race."

The empirical rule says that 95% of the data lies within two standard deviations of the mean. In this case that means that 95% of the data lies between $88.1 - 2(10.3)$ and $88.1 + 2(10.3)$.
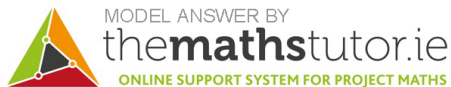
(e) Using the normal distribution tables, estimate the number of athletes who completed the race in less than 100 minutes.

Given a mean of 88.1 minutes and a standard deviation of 10.3 minutes, the standardised z-score corresponding to 100 minutes is

$$\frac{100 - 88.1}{10.3} = 1.155$$

correct to three decimal places. Now using the tables we see that $P(z \leq 1.155) = 0.8760$ correct to four decimal places. So we estimate that the number of athletes who completed the race in less than 100 minutes is $0.8760(224) \approx 196$. So about 196 athletes.
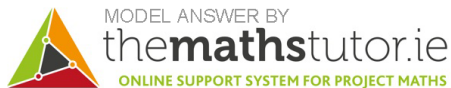
(f) After the event, a reporter wants to interview two people who took more than 100 minutes to complete the race. She approaches athletes at random and asks them their finishing time. She keeps asking until she finds someone who took more than 100 minutes, interviews that person, and continues until she finds a second such person. Assuming the athletes are cooperative and truthful, what is the probability that the second person that she interviews will be the sixth person she approaches?

From part (e), we know that the probability of a randomly selected athlete having a finish time of more than 100 minutes is $1 - P(\text{less than 100 minutes}) = 1 - 0.876 = 0.124$.
So we model this problem as a sequence of Bernoulli trials where the probability of success in each trial is

$$p = 0.124.$$

The probability that the second success occurs on the sixth trial is

$$P(\text{exactly one success in the first 5}) \times P(\text{success in trial 6})$$

Now

$$
\begin{aligned}
P(\text{exactly one in first 5}) &= 5 \times p(1-p)^4 \\
&= 5(0.124)(0.876)^4 \\
&= 0.365
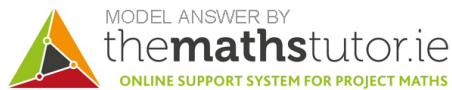\end{aligned}
$$

correct to three decimal places.
Also,

$$P(\text{success in trial 6}) = p = 0.124.$$

So the answer is

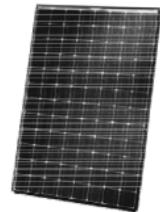$$0.365 \times 0.124 = 0.0453$$

correct to three decimal places.

Question 8 (50 marks)

A stand is being used to prop up a portable solar panel. It consists of a support that is hinged to the panel near the top, and an adjustable strap joining the panel to the support near the bottom.

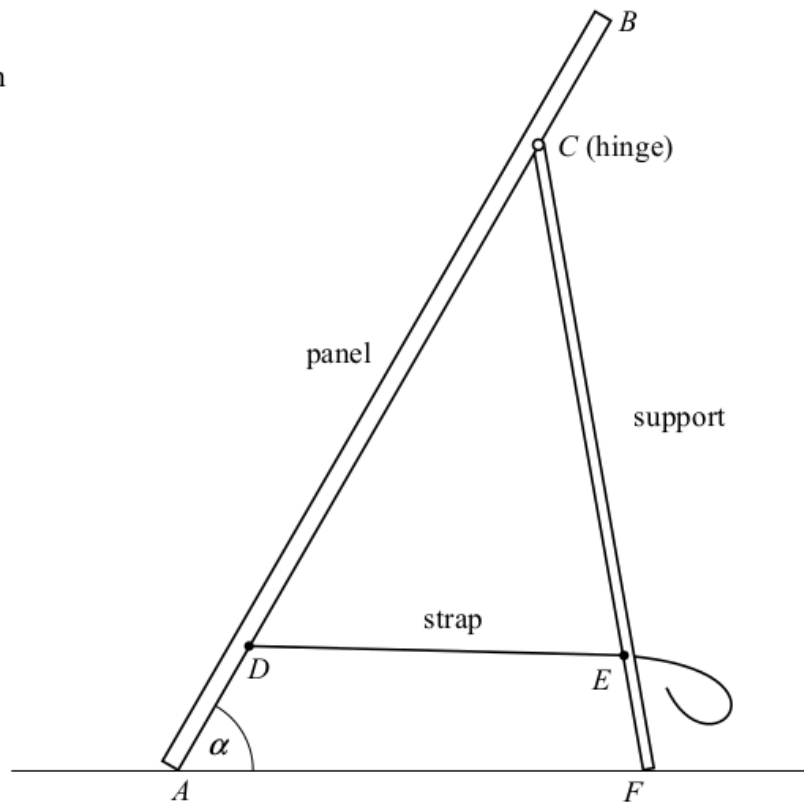By adjusting the length of the strap, the angle between the panel and the ground can be changed.

The dimensions are as follows:

$|AB| = 30$ cm

$|AD| = |CB| = 5$ cm

$|CF| = 22$ cm

$|EF| = 4$ cm.

B

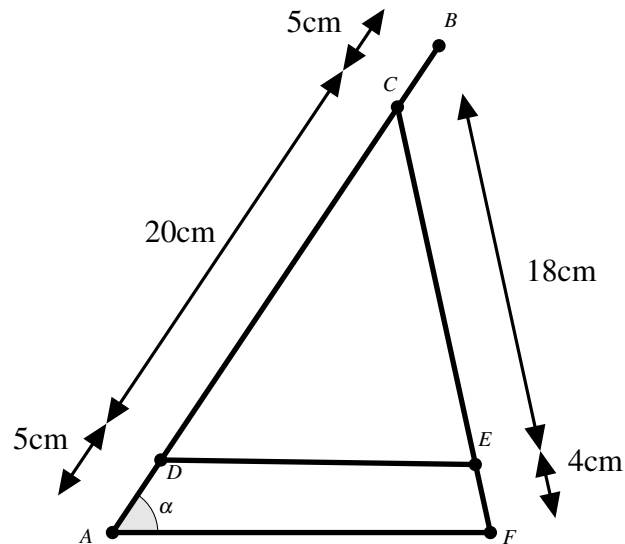C (hinge)

panel

support

strap

D

E

α

A

F

(a) Find the length of the strap $|DE|$ such that the angle α between the panel and the ground is $60°$.

Consider the diagram below in which we have marked some of the lengths. Note that $|DC|$ must be 20cm since we are told that $|AB| = 30$cm. Similarly $|CE| = 18$cm since we are told that $|CF| = 22$cm.

Now suppose that $\alpha = 60°$. Then we apply the Sine Rule to the triangle $\triangle ACF$. So

$$\frac{\sin 60°}{22} = \frac{\sin(\angle F)}{25}$$

Therefore

$$\sin(\angle F) = \frac{25\sin(60)}{22} = 0.9841$$

correct to four decimal places. Therefore $|\angle F| = \sin^{-1}(0.9841) = 79.77°$. Now we use this to calculate $\angle C$. So

$$|\angle C| = 180 - 60 - 79.77 = 40.23°.$$

Now we apply the Cosine Rule to the triangle $\triangle CDE$. Thus,

$$\begin{aligned} |DE|^2 &= 20^2 + 18^2 - 2(20)(18)\cos(40.23°) \\ &= 400 + 324 - 549.7 \\ &= 174.3 \end{aligned}$$

Therefore

$$|DE| = \sqrt{174.3} = 13.20\text{cm}.$$

So the length of the strap when $\alpha = 60°$ is 13.20cm.

(b) Find the maximum possible value of α, correct to the nearest degree.
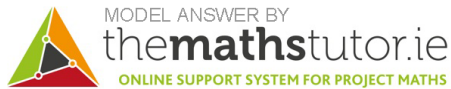
The maximum possible value of α will occur when the stand is set so that CF is vertical. In that case $\triangle ACF$ is a right angled triangle with the hypotenuse $|AC| = 25$cm. The side opposite the angle α is $|CF| = 22$cm. Therefore in this case,

$$\sin\alpha = \frac{22}{25} = 0.88.$$

So

$$\alpha = \sin^{-1}(0.88) = 61.64 = 62°$$

correct to the nearest degree.

Question 9 (25 marks)

A regular tetrahedron has four faces, each of which is an equilateral triangle.

A wooden puzzle consists of several pieces that can be assembled to make a regular tetrahedron. The manufacturer wants to package the assembled tetrahedron in a clear cylindrical container, with one face flat against the bottom.

If the length of one edge of the tetrahedron is $2a$, show that the volume of the smallest possible cylindrical container is $\left(\dfrac{8\sqrt{6}}{9}\right)\pi a^3$.
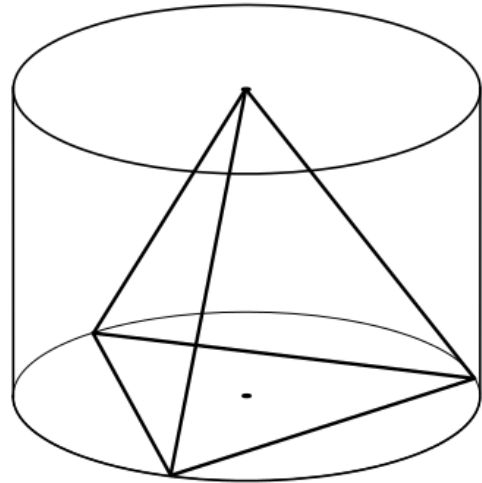
Consider the base of the cylindrical container together with the base of the tetrahedron drawn in the diagram below:
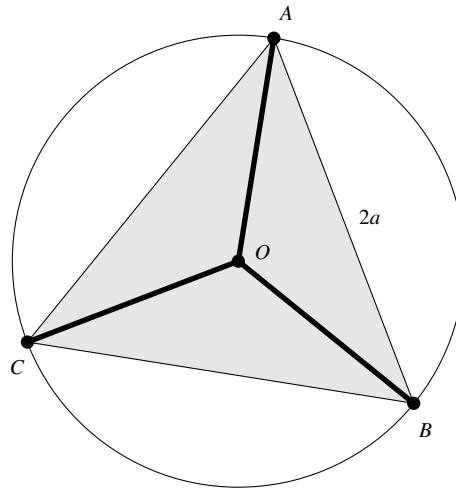
Now O is the circumcentre of an equilateral triangle. So

$$|\angle BOA| = |\angle AOC| = |\angle COB|$$

and since these angles sum to $360°$, we must have $|\angle BOA| = 120°$.
Now consider the triangle $\triangle AOB$. It is an isosceles triangle, so $|\angle ABO| = |\angle BAO|$. Since
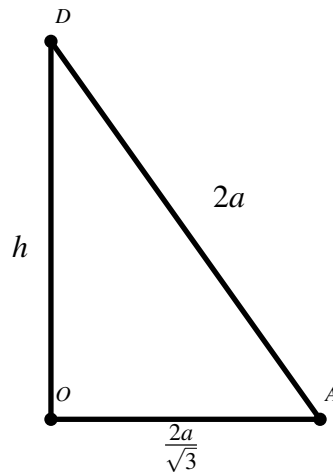
$$|\angle ABO| + |\angle BAO| + |\angle BOA| = 180°$$

we get $|\angle ABO| = 30°$. Now we apply the Sine Rule to the triangle $\triangle AOB$ to get $\frac{|OA|}{\sin 30°} = \frac{2a}{\sin 120°}$ or

$$|OA| = 2a\frac{\sin 30°}{\sin 120°} = 2a\frac{\frac{1}{2}}{\frac{\sqrt{3}}{2}} = \frac{2a}{\sqrt{3}}.$$

So the radius of the cylinder is $\frac{2a}{\sqrt{3}}$.

Solution continued on next page ...

Now let D be the top point of the tetrahedron and drop a vertical line from D to O to create a right angled triangle as shown below. Let h be the height of the cylinder.

Therefore, by Pythagoras' Theorem,

$$h^2 + \left(\frac{2a}{\sqrt{3}}\right)^2 = (2a)^2$$

which implies that $h^2 = 4a^2 - \frac{4a^2}{3} = \frac{8a^2}{3}$. Therefore

$$h = a\sqrt{\frac{8}{3}} = 2a\sqrt{\frac{2}{3}}.$$

Now the volume of a cylinder with height $2a\sqrt{\frac{2}{3}}$ and radius $\frac{2a}{\sqrt{3}}$ is

$$\pi\left(\frac{2a}{\sqrt{3}}\right)^2 2a\sqrt{\frac{2}{3}} = \pi\frac{8a^3\sqrt{2}}{3\sqrt{3}}.$$

Now multiply this last expression above and below by $\sqrt{3}$ to obtain a volume of

$$\pi\frac{8a^3\sqrt{6}}{9}$$

as required.