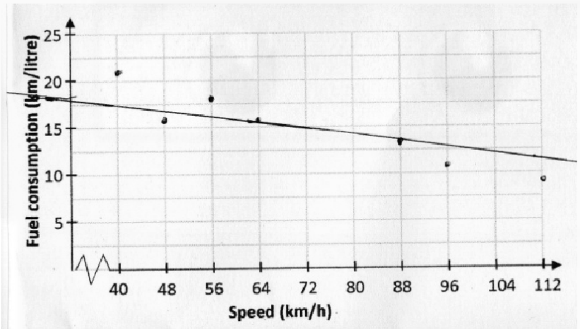
MarkingScheme

StatisticsPart1H

Question 1 (2017)

(a)	Correlation coefficient $= -0{\cdot}957$	**Scale 5A (0, 5)**
(b)	Plots Line of Best Fit on Graph 	**Scale 5C (0, 2, 4, 5)** *Low Partial Credit:* • 3 correct plots *High Partial Credit:* • All plots correct with an incorrect line of best fit • All plots correct and no line of best fit
(c)	As speed increases by 1 km/h the average distance travelled on 1 litre of fuel decreases by $0{\cdot}15$km **or** The rate at which fuel consumption in km/l is decreasing as the speed in km/h increases	**Scale 5B (0, 3, 5)** *Partial Credit:* • Some reference to speed and fuel consumption • Reference to rate of change
(d) **(i)**	$\frac{260}{96} - \frac{260}{112} = \frac{65}{168} = 0{\cdot}3869 \text{ hrs}$ $= 23{\cdot}21$ $= 23 \text{ to nearest minute}$	**Scale 5C (0, 2, 4, 5)** *Low Partial Credit:* • $\frac{260}{96}$ or $\frac{260}{112}$ *High Partial Credit:* • $\frac{260}{96} - \frac{260}{112}$ or equivalent • Answer in hours
(d) **(ii)**	$\left(\frac{260}{9} \times 1{\cdot}329\right) - \left(\frac{260}{11} \times 1{\cdot}329\right)$ $= €6{\cdot}98$	**Scale 5C (0, 2, 4, 5)** *Low Partial Credit:* • Amount of Mary's fuel or amount of Jane's fuel *High Partial Credit:* $\left(\frac{260}{9} - \frac{260}{11}\right) \times 1{\cdot}329$ or equivalent

(a) (i)	$\mu = 63{\cdot}5 \qquad \sigma = 10$ $z = \frac{50 - 63{\cdot}5}{10} = -1{\cdot}35$ $P(z > -1{\cdot}35) = P(z < 1{\cdot}35)$ $= 0{\cdot}9115$ $91{\cdot}15\%$	**Scale 10D (0, 3, 5, 8, 10)** *Low Partial Credit:* • μ or σ identified *Mid Partial Credit:* • z found *High Partial Credit:* • $P(z < 1{\cdot}35)$and stops
(a) (ii)	$P(x > Z) = 0{\cdot}015$ $P(x < Z) = 0{\cdot}985$ $Z = 2{\cdot}17$ $\frac{x - 63{\cdot}5}{10} = 2{\cdot}17$ $x = 85{\cdot}2 \text{ kg}$	**Scale 5D(0, 2, 3, 4, 5)** *Low Partial Credit:* • identifies 0·985 *Mid Partial Credit:* • identifies 2·17 *High Partial Credit:* • formula for x fully substituted
(a) (iii)	$n = 150, \qquad \bar{x} = 62, \qquad s = 10 \text{ kg}$ $H_o \rightarrow$ mean weight has not changed $H_1 \rightarrow$ mean weight has changed $z = \frac{62 - 63{\cdot}5}{\frac{10}{\sqrt{150}}}$ $= -1{\cdot}8371 > -1{\cdot}96$ Mean weight has not changed **or** Confidence interval: $\bar{x} \pm 1{\cdot}96 \frac{\sigma}{\sqrt{n}}$ $62 \pm 1{\cdot}96 \frac{10}{\sqrt{150}}$ $62 \pm 1{\cdot}96(0{\cdot}8165)$ $62 \pm 1{\cdot}6003$ $[60{\cdot}3997, 63{\cdot}6003]$ 63·5 falls within this interval $\therefore$ insufficient evidence to reject the null hypothesis The mean weight has not changed	**Scale 15D (0, 5, 7, 9, 15)** *Low Partial Credit:* • z formulated with some substitution • states null/alternative hypothesis only • reference to $\pm 1{\cdot}96$ *Mid Partial Credit:* • z fully substituted *High Partial Credit:* • $z = -1{\cdot}8371 > -1{\cdot}96$ • fails to contextualise the answer

<table>
<tr><th>Q9</th><th>Model Solution – 50 Marks</th><th>Marking Notes</th></tr>
<tr><td>(a)
(i)</td><td>$\mu = 39400, \quad \sigma = 12920$
$z = \frac{x-\mu}{\sigma} = \frac{60000-39400}{12920}$
$z = 1{\cdot}59$
$P(z > 1{\cdot}59) = 1 - P(z < 1{\cdot}59)$
$= 1 - 0{\cdot}9441 = 0{\cdot}0559$
$= 5{\cdot}59\%$
= 5·6%</td><td>Scale 10D (0, 3, 5, 8, 10)
Low Partial Credit
• μ and σ identified

Mid Partial Credit
• $z = 1{\cdot}59$

High Partial Credit
• identifies 0·9441</td></tr>
<tr><td>(a)
(ii)</td><td>$P(z \le z_1) = 0{\cdot}9$
$z_1 = 1{\cdot}28$
$\Longrightarrow z_2 = -1{\cdot}28$
$\Longrightarrow \frac{x-39400}{12920} = -1{\cdot}28$
$x = 22862{\cdot}40$
= €22 862</td><td>Scale 5C (0, 2, 4, 5)
Low Partial Credit
• identifies 1·28 but fails to progress

High Partial Credit
• formula for x fully substituted</td></tr>
<tr><td>(a)
(iii)</td><td>$\mu = 39400, \quad \sigma = 12920,$
$\bar{x} = 38280, \quad n = 1000$

$H_0 \Rightarrow \mu = 39400$
$H_1 \Rightarrow \mu \ne 39400$

$z = \frac{38280-39400}{\frac{12920}{\sqrt{1000}}} = -2{\cdot}74$

$-2{\cdot}74 < -1{\cdot}96$

Result is significant. There is evidence to reject the null hypothesis
The mean income has changed.</td><td>Scale 15D (0, 4, 7, 11,15)
Low Partial Credit
• z formulated with some substitution
• states null and/or alternative hypothesis only
• reference to 1·96

Mid Partial Credit
• z fully substituted

High Partial Credit
• $z = -2{\cdot}74$ and stops
• fails to state the null and alternative hypothesis correctly
• fails to contextualise the answer</td></tr>
</table>

or

Confidence Interval:

$$\bar{x} \pm 1{\cdot}96\frac{\sigma}{\sqrt{n}}$$

$$39400 \pm 1{\cdot}96\frac{12920}{\sqrt{1000}}$$

$$[38599{\cdot}2,\ 40200{\cdot}8]$$

38280 outside range

Result is significant. There is evidence to reject the null hypothesis

The mean income has changed.

or

Confidence Interval:

$$\bar{x} \pm 1{\cdot}96\frac{\sigma}{\sqrt{n}}$$

$$38280 \pm 1{\cdot}96\frac{12920}{\sqrt{1000}}$$

$$38280 \pm 1{\cdot}96(408{\cdot}57)$$

$$[37479{\cdot}2,\ 39080{\cdot}8]$$

39400 outside range

Result is significant. There is evidence to reject the null hypothesis

The mean income has changed.

(b)	$26974 - 1{\cdot}96\left(\frac{5120}{\sqrt{400}}\right) \le \mu$ $\le 26974 + 1{\cdot}96\left(\frac{5120}{\sqrt{400}}\right)$ $26472{\cdot}24 \le \mu \le 27475{\cdot}76$	Scale 10C (0, 3, 7, 10) *Low Partial Credit* • interval formulated with some correct substitution *High Partial Credit* • interval formulated with fully correct substitution
(c)	The distribution of sample means will be normally distributed	Scale 5B (0, 2, 5) *Partial Credit* • mentions 30 (or more) but not contextualised
(d)	$\frac{1}{\sqrt{n}} = 0{\cdot}045$ $\frac{1}{0{\cdot}045} = \sqrt{n}$ $n = \left(\frac{1}{0{\cdot}045}\right)^2 = 493{\cdot}827$	Scale 5C (0, 2, 4, 5) *Low Partial Credit* • $\frac{1}{\sqrt{n}}$ *High Partial Credit* • n formulated with fully correct substitution **Note:** Accept 493 farmers or 494 farmers

(a) Find a 95% confidence interval for the mean amount spent in a supermarket on that Saturday.

$$\frac{\sigma}{\sqrt{n}} = \frac{20\cdot73}{\sqrt{100}} = 2\cdot073$$

C. I.= $\bar{x} \pm 1\cdot96\frac{\sigma}{\sqrt{n}} = 90\cdot45 \pm 4\cdot06$

We can be 95% confident that the mean amount spent was in the range
€86·39 < μ < €94·51

(b) A supermarket has claimed that the mean amount spent by shoppers on a Saturday is €94. Based on the survey, test the supermarket's claim using a 5% level of significance. Clearly state your null hypothesis, your alternative hypothesis, and your conclusion.

H_0: Mean spend is €94
H_1: Mean spend is not €94

METHOD 1:
$\bar{x} = 90\cdot45, \quad \sigma = 20\cdot73, \quad \mu = 94, \quad n = 100$

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{90\cdot45 - 94}{2\cdot073} = -1\cdot71$$

$-1\cdot71 > -1\cdot96$

Fail to reject null hypothesis (Not enough evidence to reject the null hypothesis)

or

METHOD 2:
M€94 is inside the confidence interval for the mean spend in the population
€86·39 < μ < €94·51 worked out in part **(i)** etc.
Fail to reject null hypothesis (Not enough evidence to reject the null hypothesis)

Or

METHOD 3:
C.I. based on a sample of 100 based on the claim is:

$$89\cdot94 < \bar{x} < 98\cdot06$$

€90·45 is inside this interval.
Fail to reject null hypothesis (Not enough evidence to reject the null hypothesis)

(c) Find the p-value of the test you performed in part **(b)** above and explain what this value represents in the context of the question.

$$P(z < -1{\cdot}71) = 1 - P(z < 1{\cdot}71)$$
$$= 1 - 0{\cdot}9564$$
$$= 0{\cdot}0436$$

p-value: $= 0{\cdot}0436 \times 2 = 0{\cdot}0872$

Meaning: If the mean amount spent really was €94, then the probability that the sample mean would be €90·45 by chance is 8·72%. It is because this is more than a 5% chance that we do not reject the null hypothesis.

Question 5 (2015)

(a) Find the 95% confidence interval for the mean mark in the subject, in the Dublin region. Interpret this interval.

The 95% confidence interval is

$$\left(\bar{x} - \frac{1.96s}{\sqrt{n}}, \bar{x} + \frac{1.96s}{\sqrt{n}}\right) = \left(374 - \frac{1.96(45)}{\sqrt{50}}, 374 + \frac{1.96(45)}{\sqrt{50}}\right)$$
$$= (374 - 12.473, 374 + 12.473)$$
$$= (361.527, 386.473)$$

We can say with 95% confidence that the mean mark in the subject, in the Dublin area lies in the interval $(361.527, 386.473)$. Having '95% confidence' means that were we to repeat this procedure (i.e sampling and calculating a confidence interval) many times, the true mean would lie inside the calculated interval 95% of the time.

MODEL ANSWER BY thematstutor.ie ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(b) The mean mark in the subject for all Leaving Certificate candidates, in 2014, was 385 and the standard deviation was 45. John suggests that the mean mark in the Dublin region is not the same as in the whole country. Test this hypothesis using a 5% level of significance. Clearly state your null hypothesis, your alternative hypothesis and your conclusion.

H_0: The mean mark in Dublin is the same as the mean mark in the whole country.
H_1: The mean mark in Dublin is different from the mean mark in the whole country.
The mean mark in the whole country is 385 which is inside the confidence interval from the previous part.
That means that we do <u>not</u> reject the null hypothesis at the 5% level of significance.
In other words we have not found significant evidence to suggest that the mean mark in Dublin is different from the mean mark in the whole country.

MODEL ANSWER BY thematstutor.ie ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(a) Test the principal's claim using a 5% level of significance. Clearly state your null hypothesis, your alternative hypothesis and your conclusion.

H_0: The mean of the distances from the students' homes to the school is 3.5km.
H_1: The mean of the distances from the students' homes to the school is differnt from 3.5km.
We calculate the z-score corresponding to 3.7 given a population mean of 3.5 and a standard error of $\frac{\sigma}{\sqrt{n}} = \frac{0.5}{\sqrt{60}} = 0.0645$ (correct to 3 significant places). So

$$z = \frac{3.7 - 3.5}{0.0645} = 3.1$$

At the 5% level of significance, the rejection region is $z \leq -1.96, z \geq 1.96$ and 3.1 lies in this rejection region (since $3.1 > 1.96$).
Therefore we reject the null hypothesis at the 5% level of significance.

MODEL ANSWER BY thematshstutor.ie ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(b) In the above sample of 60 students, 20% of them lived within 2.5 km of the school. Find the 95% confidence interval for the proportion of students from that school who live within 2.5 km of the school.

$\hat{p} = 20\% = 0.2$ and $n = 60$. So the confidence interval is

$$\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

which is

$$\left(0.2 - 1.96\sqrt{\frac{0.2(0.8)}{60}}, 0.2 + 1.96\sqrt{\frac{0.2(0.8)}{60}}\right)$$

That is $(0.2 - 0.1012, 0.2 + 0.1012)$ or

$$(0.0988, 0.3012).$$

MODEL ANSWER BY thematshstutor.ie ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(c) Data from 10 years ago shows that, at that time, 26% of the student population lived within 2.5 km of the school. Based on your answer to part (b) is it possible to conclude, at the 5% level of significance, that the proportion of students living within 2.5 km of the school has changed since that time? Explain your answer.

H_0: The propotion of students living within 2.5km of the school is 26%.
H_1: The proportion of students living within 2.5km of the school is different from 26%.
In this case the z-score corresponding to 0.2 is

$$\frac{0.2-0.26}{\sqrt{\frac{0.26(1-0.26)}{60}}} = -1.06$$

Now

$$-1.96 \leq -1.06 \leq 1.96$$

so we cannot reject the null hypothesis at the 5% level of significance.
Therefore it is not possible to conclude, at the 5% level of significance that the proportion of students living within 2.5km of the school has changed.

MODEL ANSWER BY
themathstutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(d) A statistician wishes to estimate, with 95% confidence, the proportion of students who live within a certain distance of the school. She wishes to be accurate to within 10 percentage points of the true proportion. What is the minimum sample size necessary for the statistician to carry out this analysis?

We need Margin of Error $\leq 10\% = 0.1$ (since the Margin of Error is the maximum radius of a 95% confidence interval). Now

$$\text{Margin of Error} = \frac{1}{\sqrt{n}}$$

So we have

$$\frac{1}{\sqrt{n}} \leq 0.1$$

This is equivalent to $\sqrt{n} \geq \frac{1}{0.1}$ or $\sqrt{n} \geq 10$, or $n \geq 100$. So the minimum sample size is 100.

MODEL ANSWER BY
themathstutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(a) The mean lifetime of light bulbs produced by a company has, in the past, been 1500 hours. A sample of 100 bulbs, recently produced by the company, had a mean lifetime of 1475 hours with a standard deviation of 110 hours. Test the hypothesis that the mean lifetime of the bulbs has not changed, using a 0.05 level of significance.

H_0: The mean lifetime of the lightbulbs is 1500 hoursor $\mu = 1500$.
H_1: The mean lifetime of the lightbulbs is different from 1500 hours or $\mu \neq 1500$.

Under the null hypothesis $\mu = 1500$ and the standard error is approximately (for a large sample)

$$\frac{s}{\sqrt{n}} = \frac{110}{\sqrt{100}} = 11.$$

So the z-score corresponding to the observed value 1475 is

$$\frac{1475 - 1500}{11} = -2.27$$

correct to 2 decimal places.
Now

$$-2.27 < -1.96$$

so we can reject the null hypothesis at the 5% level of significance.
In other words, there is significant evidence that the mean lifetime of the bulbs has changed.

(b) Find the p-value of the test you performed in part **(a)** above and explain what this value represents in the context of the question.

As in the solution for part **(a)** above, the z-score corresponding to 1475 is -2.27 correct to 2 decimal places.
Now $P(z \leq -2.27) = P(z \geq 2.27)) = 1 - P(z \leq 2.27) = 1 - 0.9884 = 0.0116$ (using the tables p36).
So the required p-value is $2(0.0116) = 0.0232$ or 2.32%.

This p-value represents the probability of randomly selecting a sample of 100 bulbs with mean lifetime that is less than or equal to 1475hrs or greater than or equal to 1525hrs, given that the mean lifetime of the bulbs is 1500hrs.

(a) A car from the economy fleet is chosen at random. Find the probability that the tyres on this car will last for at least 40 000 km.

The z-score for 40 000 is

$$\frac{40000-45000}{8000} = -0.625$$

We want the probability that a randomly selected tyre will last for at least 40 000km, so we need to calculate $P(z \geq -0.625)$.
Now $P(z \geq -0.625) = P(z \leq 0.625) = 0.734$ from Formula and Tables p.36. So the probability that the tyres will last at least 40 000km is 0.734.

(b) Twenty cars from the economy fleet are chosen at random. Find the probability that the tyres on at least eighteen of these cars will last for more than 40 000 km.

We have 20 Bernoulli trials with $p = 0.734$ (from part **(a)**). So

$$\begin{aligned} P(20 \text{ successes}) &= (0.734)^{20} = 0.0021 \\ P(19 \text{ successes}) &= \binom{20}{1}(0.734)^{19}(1-0.734) = 0.0149 \\ P(18 \text{ successes}) &= \binom{20}{2}(0.734)^{18}(1-0.734)^2 = 0.0514 \end{aligned}$$

correct to 4 decimal places. So

$$P(\text{at least 18 successes}) = 0.0021 + 0.0149 + 0.0514 = 0.0684$$

Therefore the answer is $0.0684 = 6.84\%$.

(c) The company is considering switching brands from *Evertread* tyres to *SafeRun* tyres, because they are cheaper. The distributors of *SafeRun* tyres claim that these tyres have the same mean lifespan as *Evertread* tyres. The car rental company wants to check this claim before they switch brands. They have enough data on *Evertread* tyres to regard these as a known population. They want to test a sample of *SafeRun* tyres against it.

The company selects 25 cars at random from the economy fleet and fits them with the new tyres. For these cars, it is found that the mean life span of the tyres is 43 850 km.

Test, at the 5% level of significance, the hypothesis that the mean lifespan of *SafeRun* tyres is the same as the mean of *Evertread* tyres. State clearly what the company can conclude about the tyres.

The null hypothesis is H_0: The mean lifespan of *SafeRun* is 45 000.
We calculate the z-score corresponding to 43850 given a population mean of 45 000 and a standard error of $\frac{\sigma}{\sqrt{25}} = \frac{8000}{\sqrt{25}} = 1600$ (*). Thus

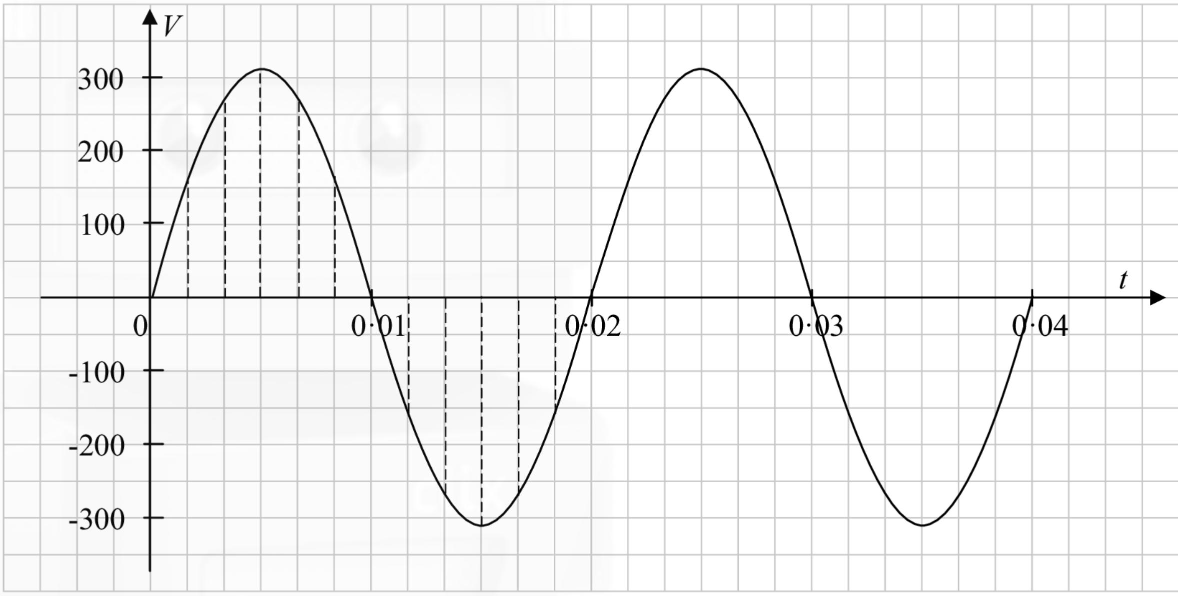
$$z_1 = \frac{43850 - 45000}{1600} = -0.71875$$

Now $-1.96 < -0.71875 < 1.96$ so we cannot reject the null hypothesis at the 5% level of significance.

The company can conclude that they do not have significant evidence to suggest that *SafeRun*'s claim is false.

(*): We assume that the standard deviation of the lifespan of the *SafeRun* tyres is the same as that of the *Evertread*. We must make some assumption about the standard deviation in order to solve the problem.

The graph below shows the voltage, V, in an electric circuit as a function of time, t. The voltage is given by the formula $V = 311\sin(100\pi t)$, where V is in volts and t is in seconds.



(a) **(i)** Write down the range of the function.

Range: $[-311, 311]$

(ii) How many complete periods are there in one second?

$$\frac{100\pi}{2\pi} = 50 \text{ periods per second}$$

Or

Time for 1 period $= 0{\cdot}02$ seconds

Number of periods in 1 second $= \dfrac{1}{0\cdot 02} = 50$

(b) **(i)** The table below gives the voltage, correct to the nearest whole number, at equally spaced intervals from t_0 to t_{12} over one complete period (as shown by the dashed lines on the diagram). Use the entries given in the table and the properties of the function to complete the table.

T	t_1	t_2	t_3	t_4	t_5	$t_6 = 0{\cdot}01$	t_7	t_8	t_9	t_{10}	t_{11}	$t_{12} = 0{\cdot}02$
V	156	269	311	269	156	0	−156	−269	−311	−269	−156	0

(ii) Using a calculator, or otherwise, calculate the standard deviation, σ, of the twelve values of V in the table, correct to the nearest whole number.

$$\sigma = 219 \cdot 89 = 220$$

(c) **(i)** The standard deviation, σ, of closely spaced values of any function of the form $V = a\sin(bt)$, over 1 full period, is given by $k\sigma = V_{\max}$, where k is a constant that does not depend on a or b, and $V_{\max}$ is the maximum value of the function. Use the function $V = 311\sin(100\pi t)$ to find an approximate value for k correct to three decimal places.

$$k = \frac{V_{\max}}{\sigma} = \frac{311}{220} \approx 1 \cdot 414$$

(ii) Using your answer in part **(c) (i)**, or otherwise, find the value of b required so that the function $V = a\sin(bt)$ has 60 complete periods in one second and the approximate value of a so that it has a standard deviation of 110 volts.

$$\frac{b}{2\pi} = 60 \Rightarrow b = 120\pi = 377$$

$$k\sigma = V_{\max} \quad \Rightarrow \quad V_{\max} = 1{\cdot}414 \times 110 = 155{\cdot}54 \Rightarrow a = 156$$

(a) Suggest two categories of people, aged 15 years and over, who might not be in the labour force.

Students, Retired, Stay at home persons, Disabled

(b) Find the median and the interquartile range of the total persons at work over the period.

Median: $\frac{1}{2}(1828\cdot 6+1867\cdot 0)=1847\cdot 8$

IQR: $1954\cdot 9-1803\cdot 4=151\cdot 5$

(c) The following data was obtained from Table 1. The percentages of persons aged 15 years and over at work, unemployed, or not in the labour force for the year 2006 are given below.

		At work	Unemployed	Not in the labour force
Persons aged 15 years and over	2006	57·9%	3·5%	38·6%
	2011			

(i) Complete the table for the year 2011. Give your answers correct to one decimal place.

		At work	Unemployed	Not in the labour force
Persons aged 15 years and over	2006	57·9%	3·5%	38·6%
	2011	50·4%	10·1%	39·5%

(ii) A census in 2006 showed that there were 864 449 persons in the population aged under 15 years of age. The corresponding number in the 2011 census was 979 590. Assuming that none of these persons are in the labour force, complete the table below to give the percentages of the *total population* at work, unemployed, or not in the labour force for the year 2011.

		At work	Unemployed	Not in the labour force
Total population	2006	46·1%	2·8%	51·1%
	2011	39·6%	8·0%	52·4%

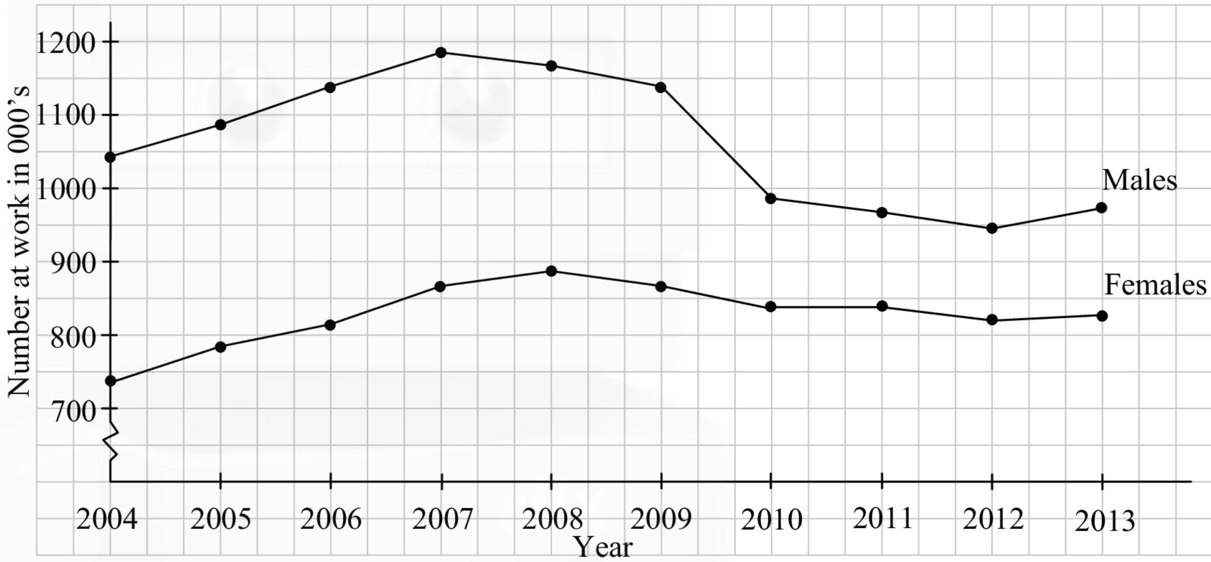
(iii) A commentator states that "The changes reflected in the data from 2006 to 2011 make it more difficult to balance the Government's income and expenditure."
Do you agree with this statement? Give two reasons for your answer based on your calculations above.

Yes
Percentage at work down, so reduced taxes collected, so income reduced.
Percentage not in workforce up, so increased expenditure on supports, pensions etc.

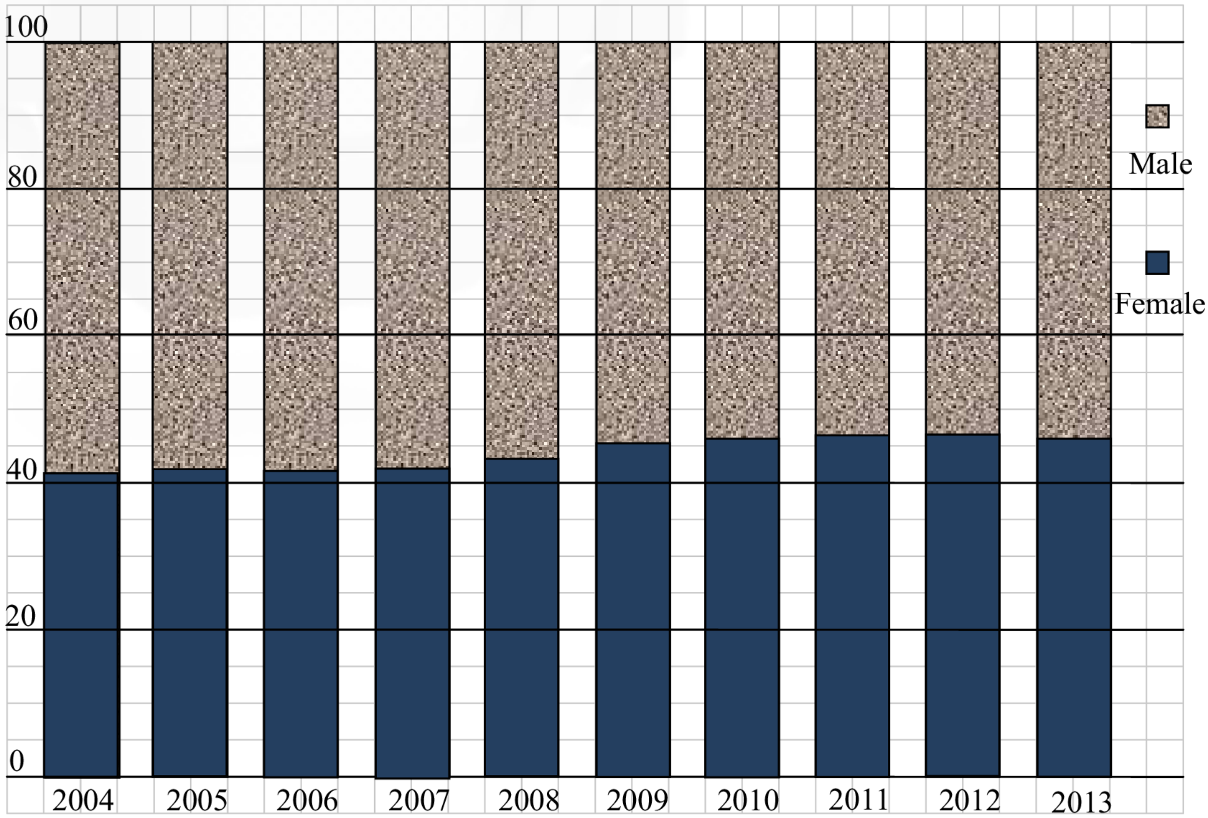
Note: Answer here depends on candidate's answers in previous sections

(d) Liam and Niamh are analysing the number of males and the number of females at work over the period 2004 to 2013.

Liam draws the following chart using data obtained from Table 1.



Niamh also uses data from Table 1 and calculates the number of females at work as a percentage of the total number of persons at work and then draws the following chart.



(i) Having examined both charts, a commentator states "females were affected just as much as males by the downturn in employment." Do you agree or disagree with this statement? Give a reason for your conclusion.

> Disagree.
> Male unemployment declined from 2007, female from 2008.
> Greater decline in number of males employed.

(ii) Which, if any, of the two charts did you find most useful in reaching your conclusion above? Give a reason for your answer.

> Liam's graph shows trend over time as well as the numbers.
> Niamh's graph only shows percentage in workforce and gives no information about actual numbers.

(iii) Use the data in Table 1, for the years 2012 and 2013 only, to predict the percentage of persons, aged 15 years and over, who will be at work in 2014.

2012	49·4% at work	
2013	50·3% at work	(+0·9%)
2014	51·2% at work	

Note: Candidates not required to round to any particular number of places of decimals

Question 11 (2014)

(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.
>
>

(ii) The mean and the median time for the run are approximately equal. Estimate this value from the corresponding histogram.

mean $\approx$ median $\approx$ 25 minutes.

From the histogram it seems that about half of the data lies to the left of 25, so that is our estimate for the median.

(iii) Estimate from the relevant histogram the standard deviation of the times for the swim.

standard deviation $\approx$ 3 minutes

The empirical rule says that 95% percent of the data lies within 2 standard deviations of the mean.
A reasonable guess from the histogram is that 95% of the data lies in between 12 minutes and 24minutes, so we might guess that the standard deviation is about $\frac{24-12}{4} = 3$.

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

The data is continuous numerical data, so the frequency of any possible value is either 0 or 1. In this situation the mode would be meaningless.

(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

$$\frac{24.528 + 21.416}{2} = 22.972 \text{ minutes.}$$

MODEL ANSWER BY
themathstutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

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 | 67.5 | and | 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 $88.1 + 2(10.3)$.

MODEL ANSWER BY thethemathstutor.ie ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(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.

MODEL ANSWER BY thethemathstutor.ie ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

(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.

MODEL ANSWER BY
themathstutor.ie
ONLINE SUPPORT SYSTEM FOR PROJECT MATHS

Question 12 (2013)

(i) Sample space

The set of all possible outcomes of an experiment.

(ii) Mutually exclusive events

Events E and F are mutually exclusive if they have no outcomes in common.
i.e $P(E \cup F) = P(E) + P(F)$

(iii) Independent events

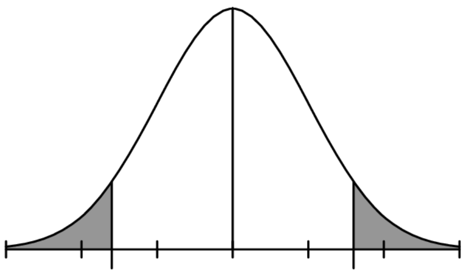
Two events are independent if the outcome of one does not depend on the outcome of the other.
i.e $P(E \cap F) = P(E).P(F)$ or $P(E|F) = P(E)$ or $P(F|E) = P(F)$

$$P(X \le 68) = P\left(Z \le \frac{68-60}{5}\right) = P(Z \le 1 \cdot 6) = 0 \cdot 9452$$

(ii) Find $P(52 \le X \le 68)$.

$$P(52 \le X \le 68) = P\left(\frac{52-60}{5} \le Z \le \frac{68-60}{5}\right)$$

$= P(-1 \cdot 6 \le Z \le 1 \cdot 6)$



$P(Z \le -1 \cdot 6) = P(Z \ge 1 \cdot 6)$

$= 1 - P(Z \le 1 \cdot 6)$

$= 1 - 0 \cdot 9452 = 0 \cdot 0548$

$P(-1 \cdot 6 \le Z \le 1 \cdot 6) = P(Z \le 1 \cdot 6) - P(Z \le -1 \cdot 6)$

$= 0 \cdot 9452 - 0 \cdot 0548 = 0 \cdot 8904$

OR

$$P(52 \le X \le 68) = P\left(\frac{52-60}{5} \le Z \le \frac{68-60}{5}\right)$$

$= P(-1 \cdot 6 \le Z \le 1 \cdot 6)$

$= 1 - 2P(Z \ge 1 \cdot 6)$

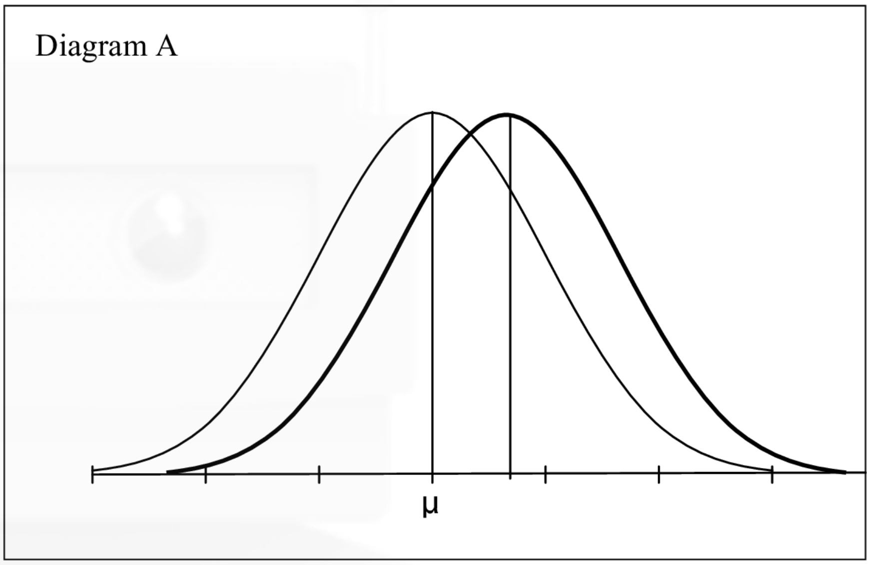
$= 1 - 2(1 - P(Z \le 1 \cdot 6))$

$= 1 - 2(1 - 0 \cdot 9452) = 1 - 2(0 \cdot 0548) = 1 - 0 \cdot 1096 = 0 \cdot 8904$

The effect, on plant growth, of each of the hormones is described. Sketch, on each diagram, a new distribution to show the effect of the hormone.

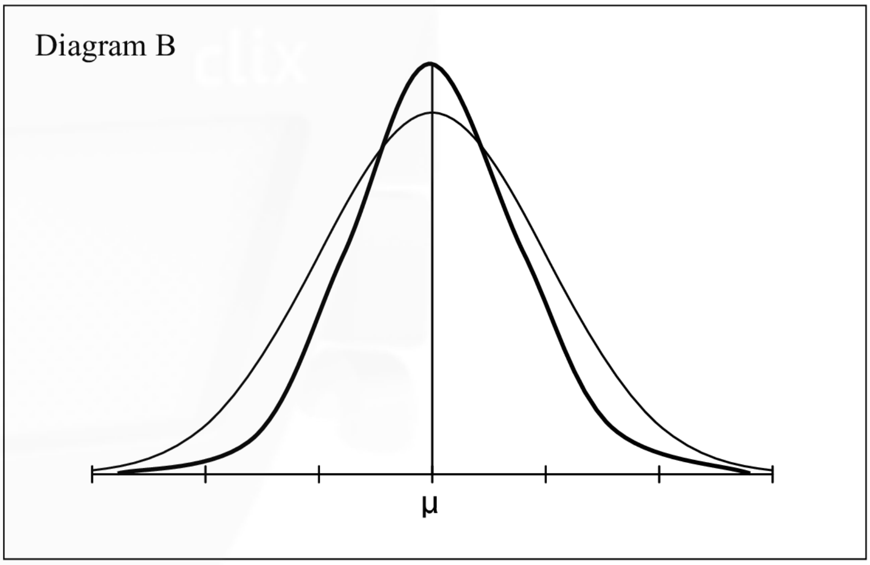
Hormone A

The effect of hormone A was to increase the height of all of the plants.



Hormone B

The effect of hormone B was to reduce the number of really small plants and the number of really tall plants. The mean was unchanged.



Hormone C

The effect of hormone C was to increase the number of small plants and the number of tall plants. The mean was unchanged.