Question 1

Question 2 **(25 marks)**

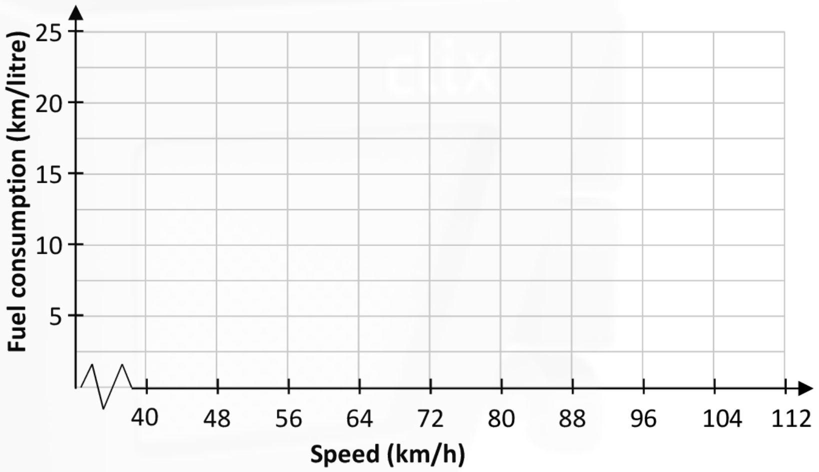
An experiment measures the fuel consumption at various speeds for a particular model of car.
The data collected are shown in Table 1 below.

Table 1							
Speed (km/hour)	40	48	56	64	88	96	112
Fuel consumption (km/litre)	21	16	18	16	13	11	9

(a) Find the correlation coefficient of the data in Table 1, correct to 3 decimal places.

Correlation Coefficient =	

(b) Plot the points from the table on the grid below **and** draw the line of best fit (by eye).



(c) The slope of the line of best fit is found to be $-0{\cdot}15$.
What does this value represent in the context of the data?

(d) Mary drove from Cork to Dublin at an average speed of 96 km/h.
Jane drove the same journey at an average speed of 112 km/h.
Each travelled 260 km and paid 132·9 cents per litre for the fuel.
Both used the model of car used to generate the data in Table 1.

(i) Find how much longer it took Mary to complete the journey.
Give your answer correct to the nearest minute.

(ii) Based on the data in Table 1 and their average speeds, find how much more Jane spent on fuel during the course of this journey.

Question 8 **(60 marks)**

(a) In 2015, in a particular country, the weights of 15 year olds were normally distributed with a mean of 63·5 kg and a standard deviation of 10 kg.

(i) In 2015, Mariska was a 15 year old in that country. Her weight was 50 kg.
Find the percentage of 15 year olds in that country who weighed more than Mariska.

(ii) In 2015, Kamal was a 15 year old in that country.
1·5% of 15 year olds in that country were heavier than Kamal.
Find Kamal's weight.

(iii) In 2016, 150 of the 15 year olds in that country were randomly selected and their weights recorded. It was found that their weights were normally distributed with a mean weight of 62 kg and a standard deviation of 10 kg. Test the hypothesis, at the 5% level of significance, that the mean weight of 15 year olds, in that country, had not changed from 2015 to 2016. State the null hypothesis and your alternative hypothesis. Give your conclusion in the context of the question.

Question 3

Data on earnings were published for a particular country. The data showed that the annual income of people in full-time employment was normally distributed with a mean of €39 400 and a standard deviation of €12 920.

(a) **(i)** The government intends to impose a new tax on incomes over €60 000.
Find the percentage of full-time workers who will be liable for this tax, correct to one decimal place.

(ii) The government will also provide a subsidy to the lowest 10 % of income earners.
Find the level of income at which the government will stop paying the subsidy, correct to the nearest euro.

(iii) Some time later a research institute surveyed a sample of 1000 full-time workers, randomly selected, and found that the mean annual income of the sample was €38 280.
Test the hypothesis, at the 5 % level of significance, that the mean annual income of full-time workers has changed since the national data were published.
State the null hypothesis and the alternative hypothesis.
Give your conclusion in the context of the question.

(b) The research institute surveyed 400 full-time farmers, randomly selected from all the full-time farmers in the country, and found that the mean income for the sample was €26 974 and the standard deviation was €5120.
Assuming that annual farm income is normally distributed in this country, create a 95 % confidence interval for the mean income of full-time farmers.

(c) It is known that data on farm size are not normally distributed.
The research institute could take many large random samples of farm size and create a sampling distribution of the means of all these samples.
Give one reason why they might do this.

(d) The research institute also carried out a survey into the use of agricultural land.
n farmers were surveyed.
If the margin of error of the survey was 4·5 %, find the value of n.

Question 2 **(25 marks)**

A survey of 100 shoppers, randomly selected from a large number of Saturday supermarket shoppers, showed that the mean shopping spend was €90·45. The standard deviation of this sample was €20·73.

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

(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.

(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-value:

Explanation:

Question 1 **(25 marks)**

A survey of 50 Leaving Certificate candidates in 2014, randomly selected in the Dublin region, found that they had a mean mark of 374 in a certain subject. The standard deviation of this sample was 45.

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

(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.

Question 2 **(50 marks)**

The principal of a large school claims that the average distance from a student's home to the school is 3·5 km. In order to test this claim, a sample of 60 students from the school was randomly selected. The students were asked how far from the school they lived. The mean distance from these students' homes to the school is 3·7 km with a standard deviation of 0·5 km.

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

(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.

(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.

Answer:

Reason:

(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?

Question 3 **(25 marks)**

(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.

(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.

p-value:

Meaning:

Question 4 **(50 marks)**

A car rental company has been using *Evertread* tyres on their fleet of economy cars. All cars in this fleet are identical. The company manages the tyres on each car in such a way that the four tyres all wear out at the same time. The company keeps a record of the lifespan of each set of tyres. The records show that the lifespan of these sets of tyres is normally distributed with mean 45 000 km and standard deviation 8000 km.

(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.

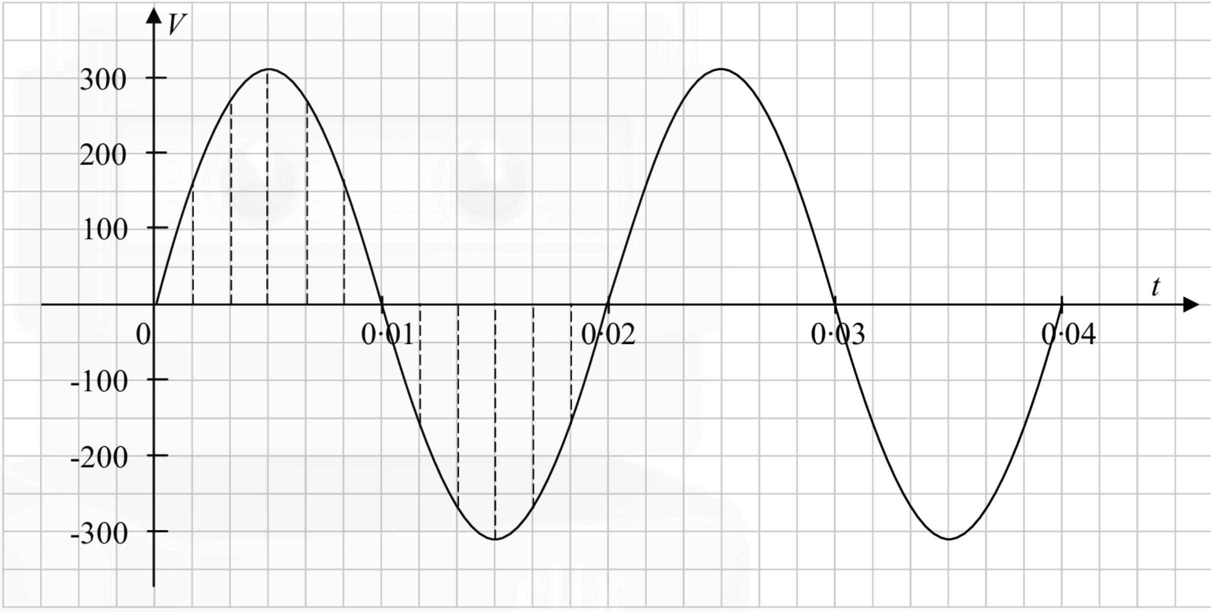
(b) Twenty cars from the economy fleet are chosen at random. Find the probability that the tyres on at least eighteen of these cars will last for more than 40 000 km.

(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 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.

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

(b) **(i)** The table below gives the voltage, correct to the nearest whole number, at equally spaced intervals from t_1 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									

(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.

(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.

(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.

Table 1 below gives details of the number of males (M) and females (F) aged 15 years and over at work, unemployed, or not in the labour force for each year in the period 2004 to 2013.

Table 1

Labour Force Statistics 2004 to 2013 - Persons aged 15 years and over (000's)

Year	At work			Unemployed			Not in labour force			Total
	M	F	Total	M	F	Total	M	F	Total	
2004	1045.9	738.9	1784.8	79.6	31.6	111.2	457.1	854.2	1311.3	3207.3
2005	1087.3	779.7	1867.0	81.3	33.5	114.8	459.5	846.6	1306.1	3287.9
2006	1139.8	815.1	1954.9	80.6	38.1	118.7	457.6	844.9	1302.5	3376.1
2007	1184.0	865.6	2049.6	84.3	39.2	123.5	472.4	852.7	1325.1	3498.2
2008	1170.9	889.5	2060.4	106.3	41.0	147.3	494.8	872.5	1367.3	3575.0
2009	1039.8	863.5	1903.3	234.0	82.4	316.4	505.6	874.9	1380.5	3600.2
2010	985.1	843.5	1828.6	257.6	98.2	355.8	529.2	884.6	1413.8	3598.2
2011	970.2	843.2	1813.4	260.7	103.4	364.1	540.1	881.5	1421.6	3599.1
2012	949.6	823.8	1773.4	265.2	108.0	373.2	546.5	896.9	1443.4	3590.0
2013	974.4	829.0	1803.4	227.7	102.3	330.0	557.8	895.0	1452.8	3586.2

(*Source: Central Statistics Office http://www.cso.ie*)

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

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

(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.

(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.

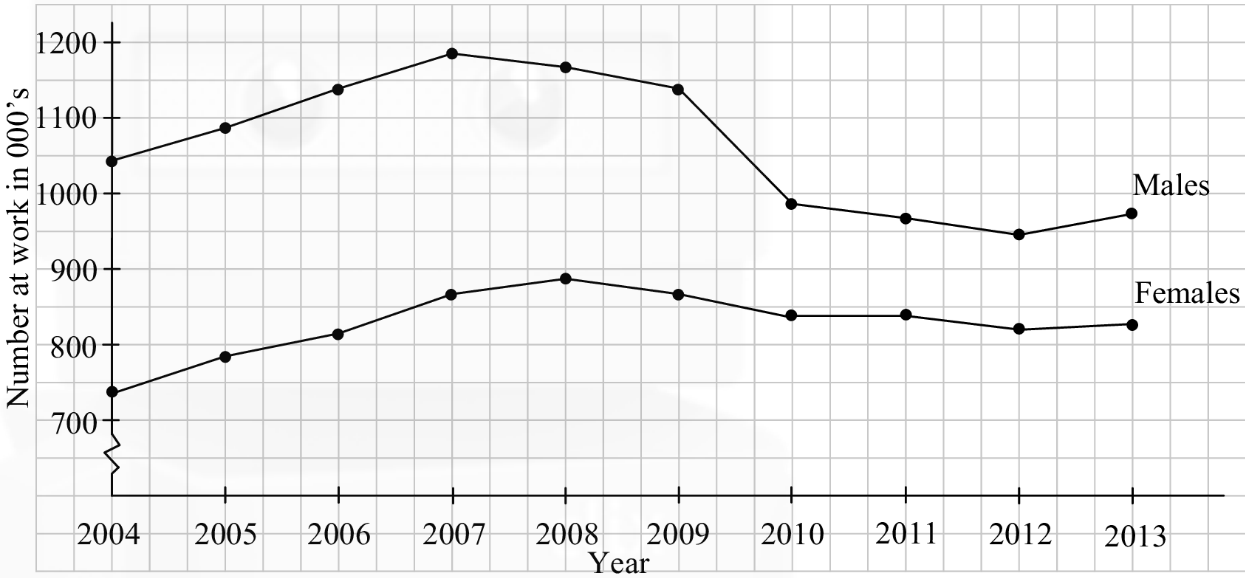
		At work	Unemployed	Not in the labour force
Total population	2006	46·1%	2·8%	51·1%
	2011			

(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.

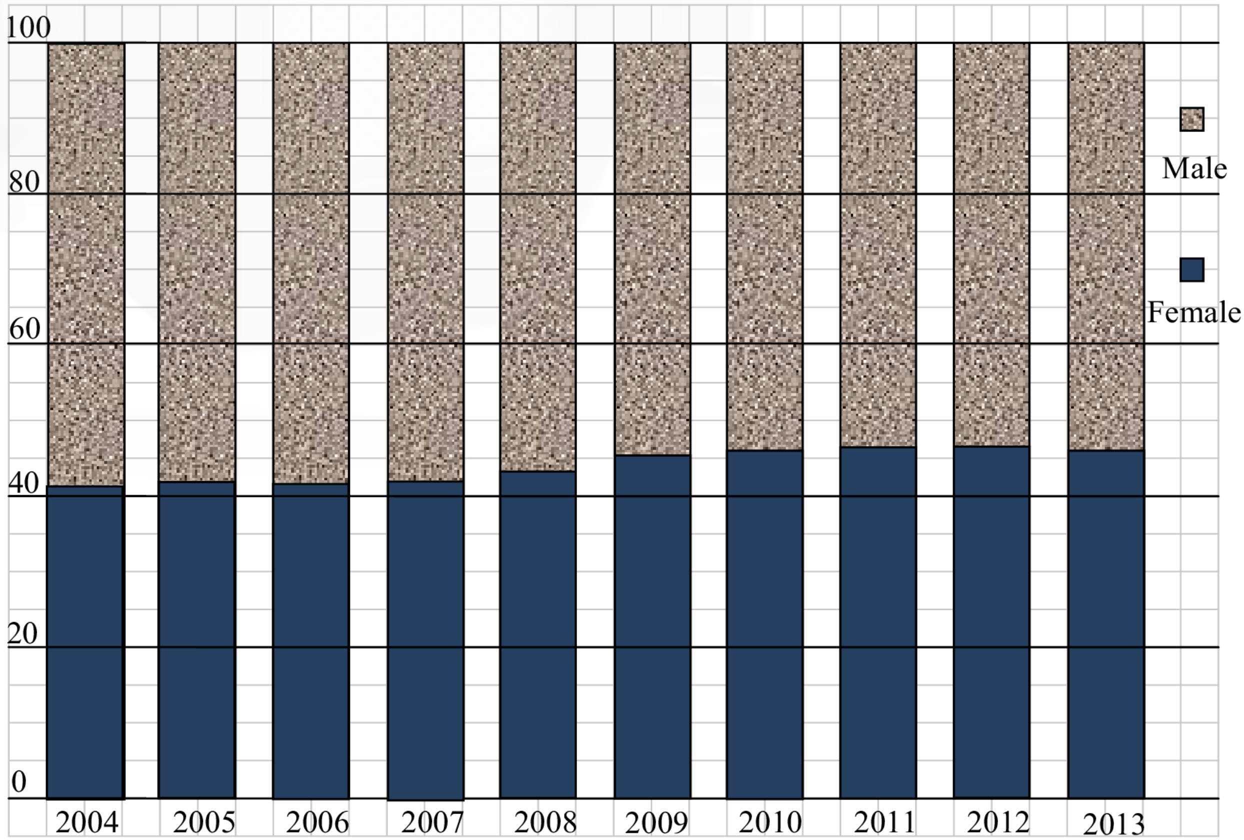
page	running

(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 from Table 1.



Niamh uses the same data 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.

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

(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.

page	running

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 data from 224 athletes who completed this triathlon in 2010.

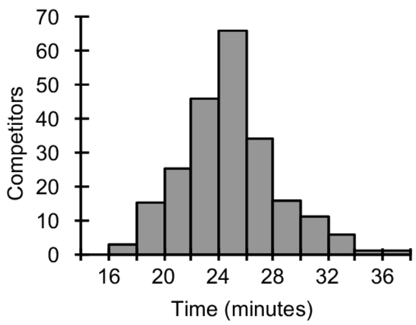
Máire is analysing data from the race, using statistical software. She has a data file with each competitor's time for each part of the race, along with various other details of the competitors.

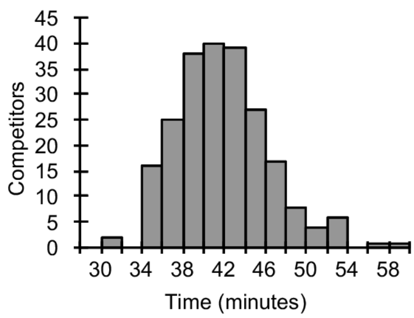
Lizzie Lee, winner of the women's event

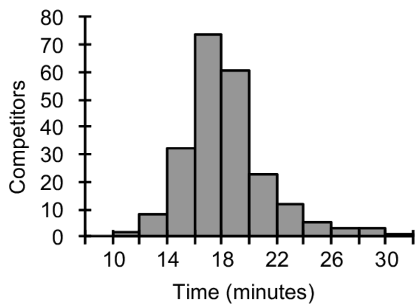
Máire gets the software to produce some *summary statistics* and it produces the following table. Three of the entries in the table have been removed and replaced with question marks (?).

	Swim	*Cycle*	*Run*
Mean	18.329	41.927	**?**
Median	17.900	41.306	**?**
Mode	#N/A	#N/A	#N/A
Standard Deviation	**?**	4.553	3.409
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 three 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.

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

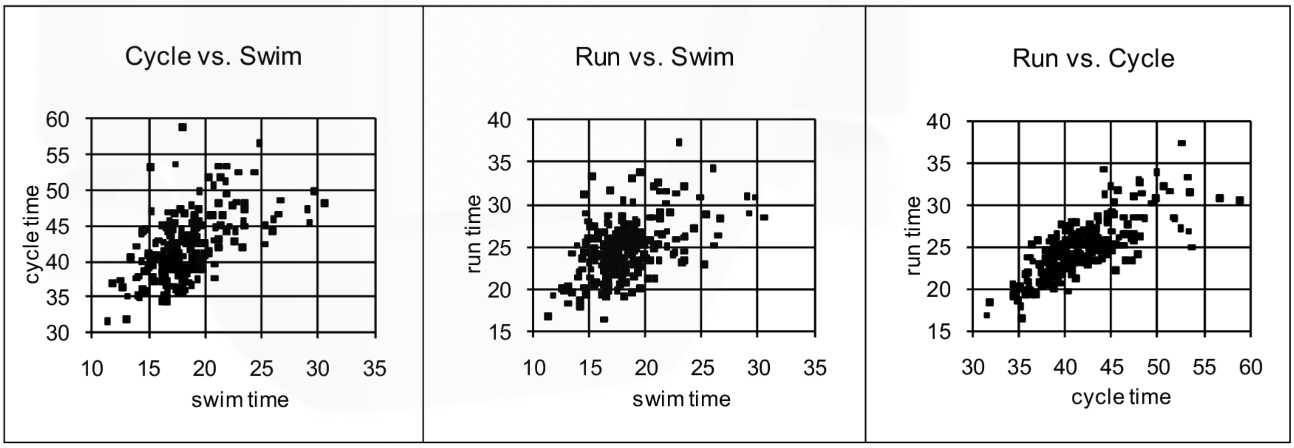
mean ≈ median ≈ ______________

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

standard deviation ≈ ___________

(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.

page running

(c) The best-fit line for run-time based on swim-time is $y = 0 \cdot 53x + 15 \cdot 2$. The best-fit line for run-time based on cycle-time is $y = 0 \cdot 58x + 0 \cdot 71$. 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.

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 ________ and ________ minutes to complete the race."

(e) Using normal distribution tables, estimate the number of athletes who completed the race in less than 100 minutes.

(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 she interviews will be the sixth person she approaches?

(a) Explain each of the following terms:

(i) Sample space

(ii) Mutually exclusive events

(iii) Independent events.

Question 2 **(25 marks)**

(a) A random variable X follows a normal distribution with mean 60 and standard deviation 5.

(i) Find $P(X \leq 68)$.

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

(b) The heights of a certain type of plant, when ready to harvest, are known to be normally distributed, with a mean of μ. A company tests the effects of three different growth hormones on this type of plant. The three hormones were used on a different large sample of the crop. After applying each hormone, it was found that the heights of the plants in the samples were still normally distributed at harvest time.

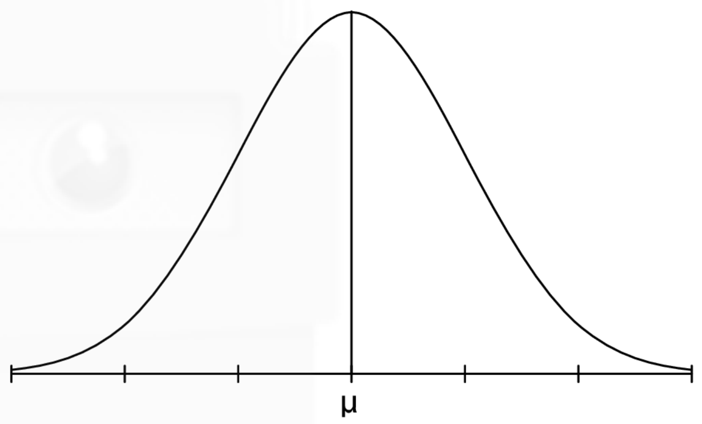
The diagrams A, B and C, on the next page, show the expected distribution of the heights of the plants, at harvest time, without the use of the hormones.

The effect, on plant growth, of each of the hormones is described on the next page. 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.

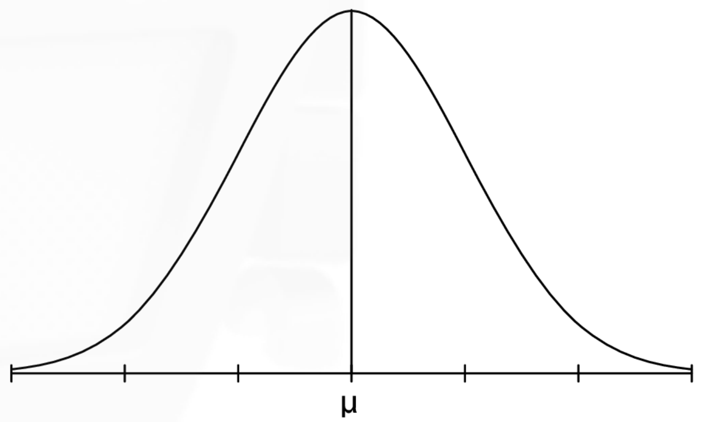
Diagram A



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.

Diagram B



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.

Diagram C