

Mean, Mode & Median

Mean

How do you find it?

- you need to find the sum of the numbers...all of them added together, you then need to divide this answer by the number of numbers in the set.

The mean should be used...

- if the data is numerical...just numbers,
- if there are no extreme values...unusually large or small numbers.

Mode

How do you find it?

- You need to find the number that appears most often in the set of data.
- There may be more than one mode!

The mode should be used...

- if the data is categorical...not numbers.

Median

How do you find it?

- you need to put all of the numbers in order starting with the smallest, you then need to pick out the middle one or average of the middle two.

The mode should be used...

- if the data is numerical...just numbers,
- if there are extreme values.

Question 12

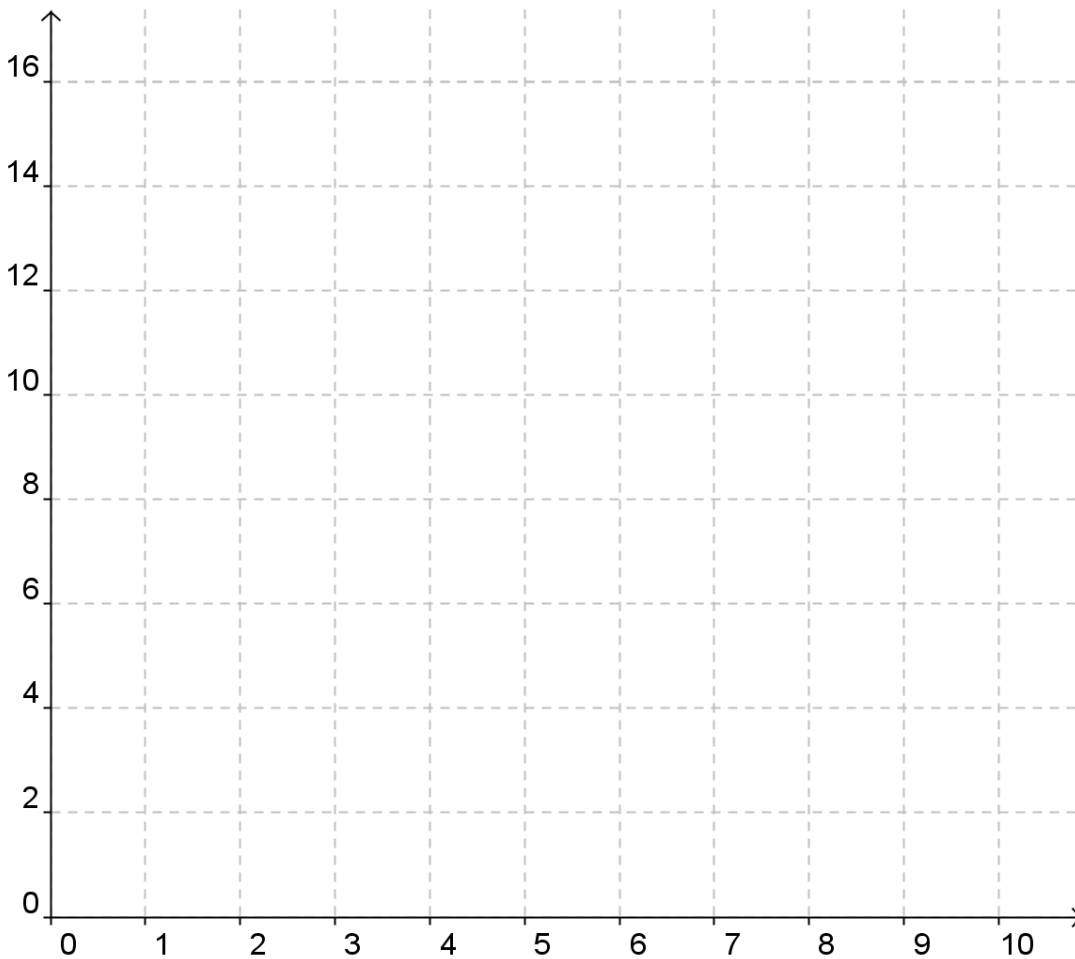
The number of raffle tickets bought by a group of 40 visitors to a school fair is shown in the table below.

0	3	6	1	2	8	2	7	6	2
1	2	3	8	3	0	6	1	2	0
3	7	3	0	1	2	6	2	3	1
6	2	6	1	9	8	3	0	0	2

(i) Complete the following table. *Note: 0 – 2 includes 0 but does not include 2, and so on.*

Number of tickets bought	0 – 2	2 – 6	6 – 8	8 – 10
Number of visitors				

2. Represent this data on a histogram.



Range, and Interquartile range

Range

How do you find it?

you need to subtract the smallest number on the list from the largest one.

The range should be used...

if the data is numerical...just numbers,

if there are no extreme values...unusually large or small numbers.

Interquartile range

How do you find it?

you need to put all of the numbers in order starting with the smallest,

you then need to find a quarter of the total number of numbers in the set,

you remove this many numbers from the bottom of the list,

you then remove the same number of numbers from the top of the list,

finally you subtract the biggest number you are left with from the smallest number you are left with.

Note: it is just finding the range of the middle half of the numbers!

The interquartile range should be used...

if the data is numerical...just numbers,

if there are extreme values...unusually large or small numbers.

You can be given...

- List of numbers...an array,
- Frequency table,
- Grouped frequency table.

Range

To find the range, just subtract the smallest number from the biggest one.

Interquartile Range

1. Put the numbers in ascending order
2. Use the formula $\frac{1}{4}(n + 1)$ to tell you which number on the list is the **lower quartile**.
3. Then use the formula $\frac{3}{4}(n + 1)$ to tell you which number is the **upper quartile**.
4. The difference between these two numbers is the interquartile range.

Example 1

If the list of numbers was 2, 3, 4, 8, 9, 10 12... there are 7 numbers on the list... $n = 7$.

$$\text{So } \frac{1}{4}(n + 1) = \frac{1}{4}(7 + 1) = \frac{1}{4}(8) = 2 \dots \text{so the 2}^{\text{nd}} \text{ number is the lower quartile!}$$

2, 3, 4, 8, 9, 10 12...so 3 is the lower quartile.

$$\text{And } \frac{3}{4}(n + 1) = \frac{3}{4}(7 + 1) = \frac{3}{4}(8) = 6 \dots \text{so the 6}^{\text{th}} \text{ number is the upper quartile!}$$

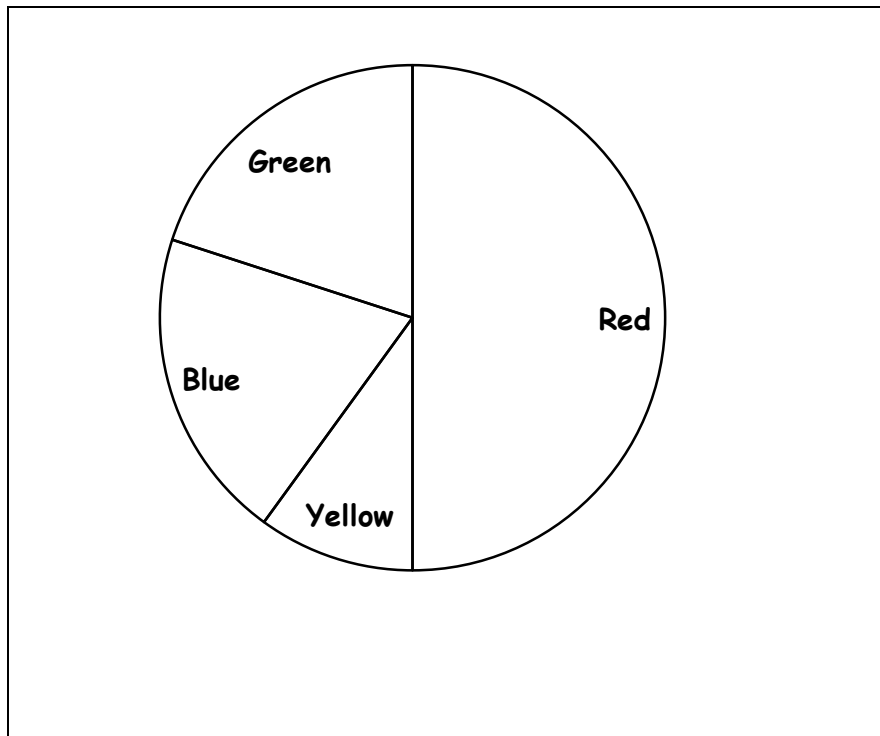
2, 3, 4, 8, 9, 10 12...so 10 is the upper quartile. The interquartile range is upper quartile – lower quartile = $10 - 3 = 7$

Pie Charts

INTERPRETING PIE CHARTS

Question 16

90 people's favourite colour

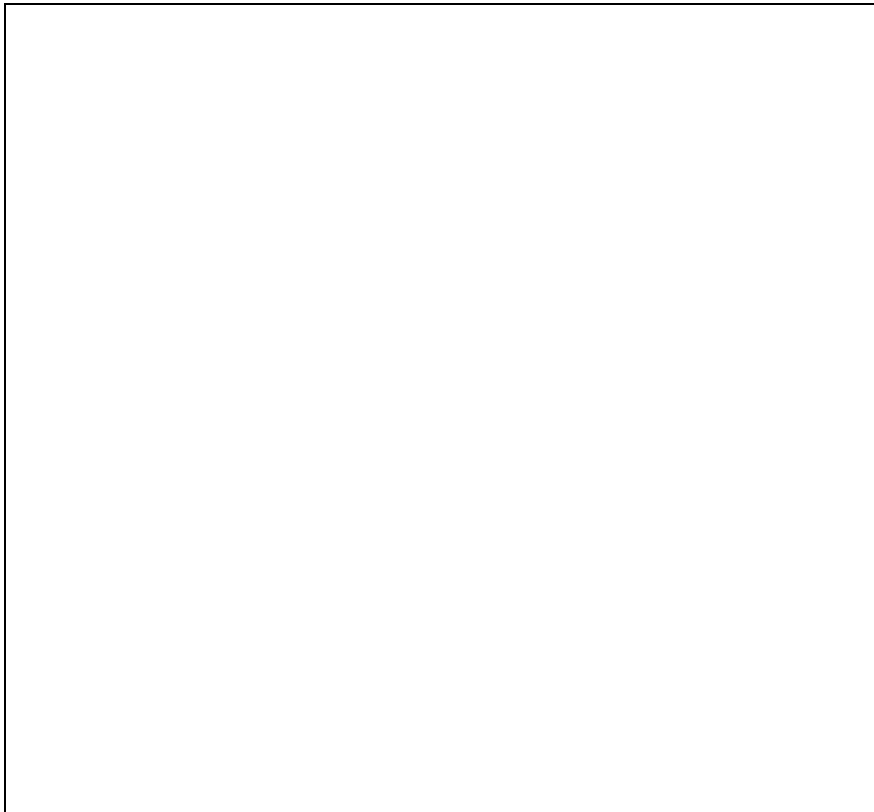


Favourite colour	Degrees	Value
Red	180°	
Yellow	36°	9
Blue	72°	
Green	72°	
1 person = ___ degrees		

Question 17

Type of pet that 60 people had

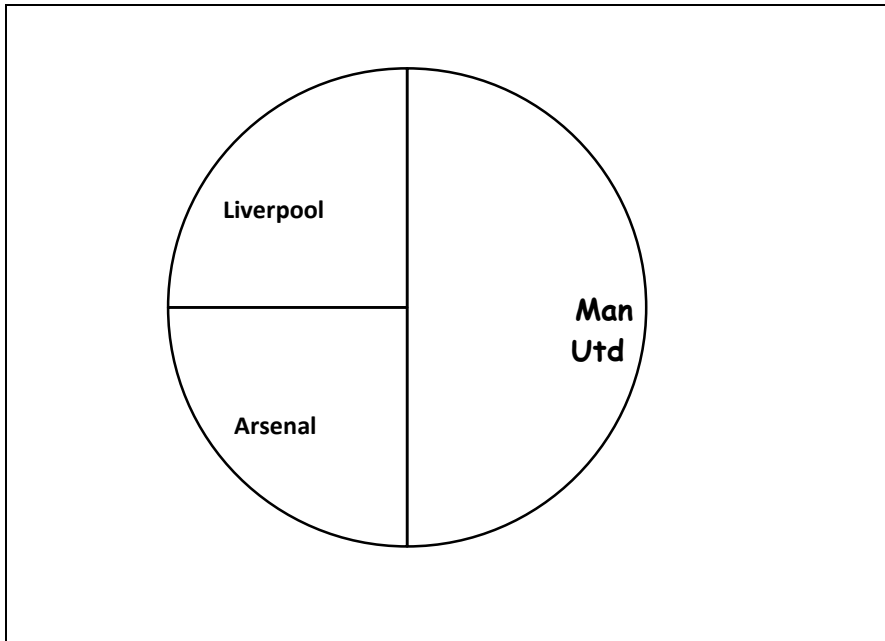
Based on the data provided construct a pie chart in the circle given



Pet	Degrees	Value
Dog	180 ⁰	
Cat	36 ⁰	
Budgie	36 ⁰	
Goldfish	36 ⁰	
Other		
1 person = ____ degrees		

Question 18

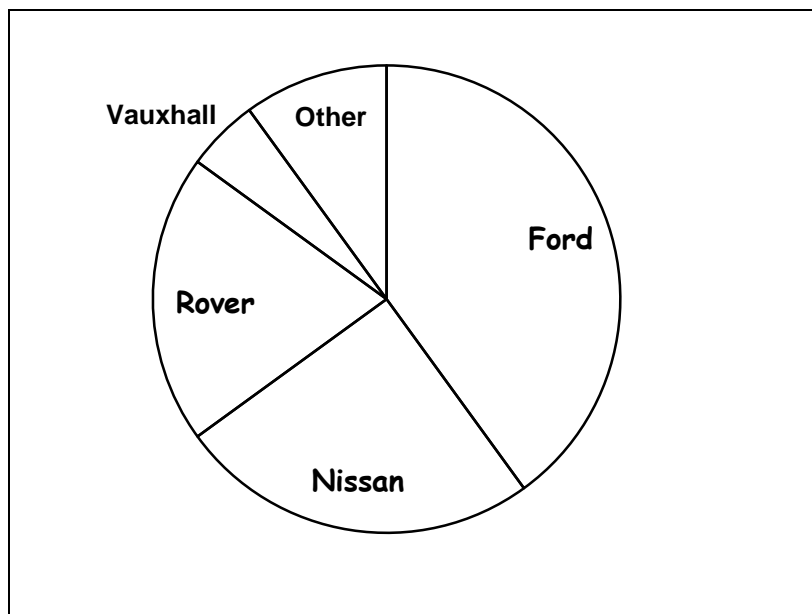
Favourite football team of 120 people



Favourite colour	Degrees	Value
Liverpool		
Man Utd		
Arsenal		
1 person = ___ degrees		

Question 19

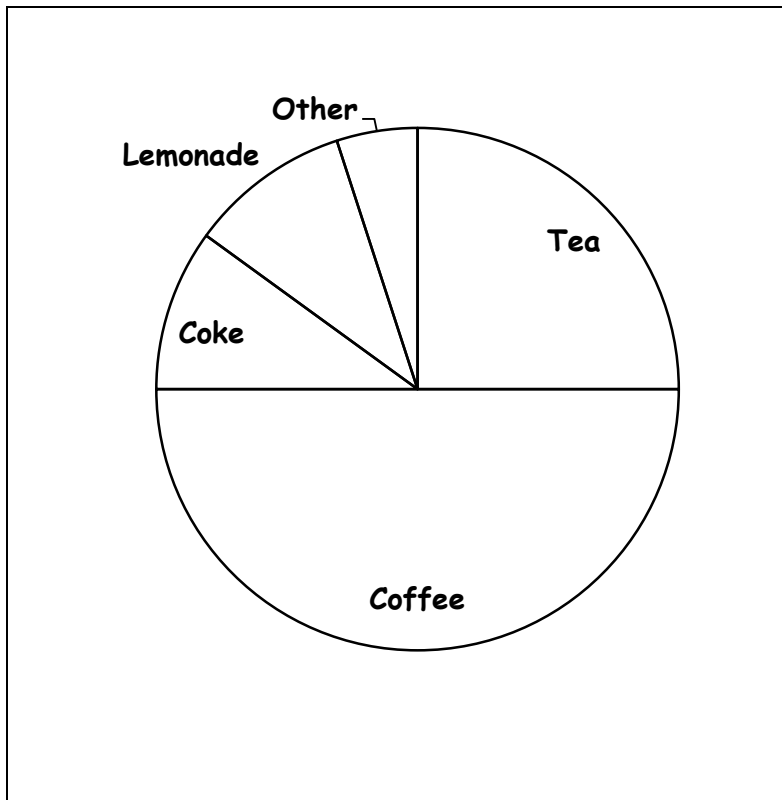
Make of ___ cars



Make of car	Degrees	Value
Ford	144 ⁰	
Nissan	90 ⁰	
Rover	72 ⁰	8
Vauxhall		
Other	36 ⁰	
1 person = ____ degrees		

Question 20

Favourite drink of 80 people



Drink	Degrees	Value
Tea	90 ⁰	
Coffee	180 ⁰	
Coke	36 ⁰	
Lemonade		
Other	18 ⁰	
1 person = ____ degrees		

- (i) What fraction of people said Tea was their favourite drink?

- (ii) What fraction of people said Lemonade was their favourite drink?

- (iii) What percentage of people said Coffee was their favourite drink?

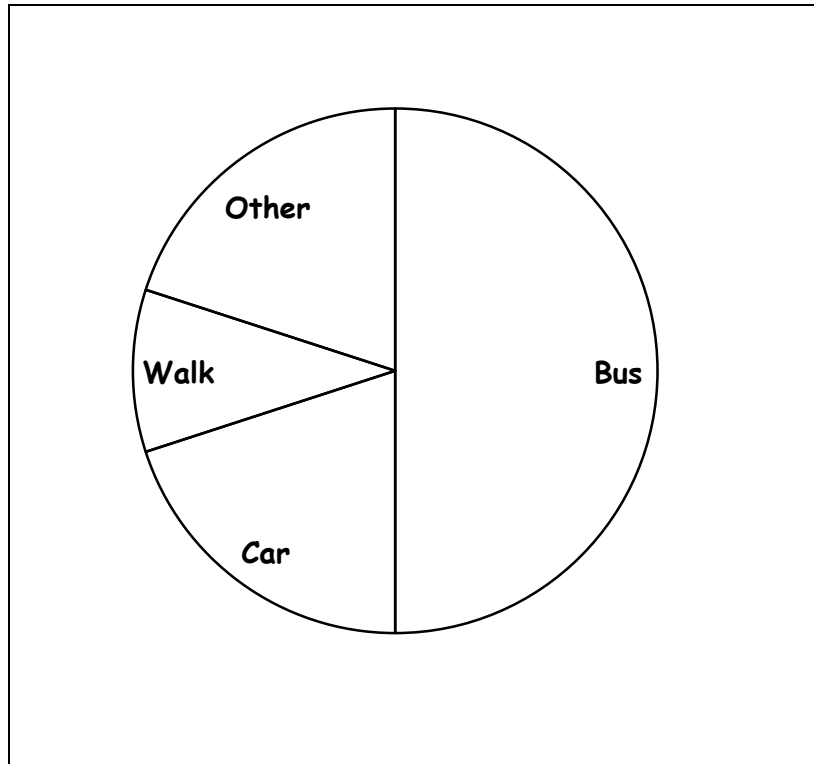
- (iv) What percentage of people said Other was their favourite drink?

- (v) What is the probability that a person, chosen at random, said their favourite drink is Coke?

- (vi) How many of the people asked said that Coke or Lemonade was their favourite drink?

Question 21

How 30 people travel to school



Transport	Degrees	Value
Bus		15
Walk		3
Car		6
Other		
1 person = ____ degrees		

- (i) Explain how you know how many people travel to school in 'Other' ways
- (ii) What is the probability that a person, chosen at random, travels to school by Car? Simplify your answer if possible.
- (iii) What percentage of people travel to work by walking?
- (iv) If 300 people had completed the survey, how many people would you expect to travel to work by Bus? Show your working.
- (v) If 150 people had completed the survey, how many people would you expect to travel to work by Car?

STANDARD DEVIATION

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

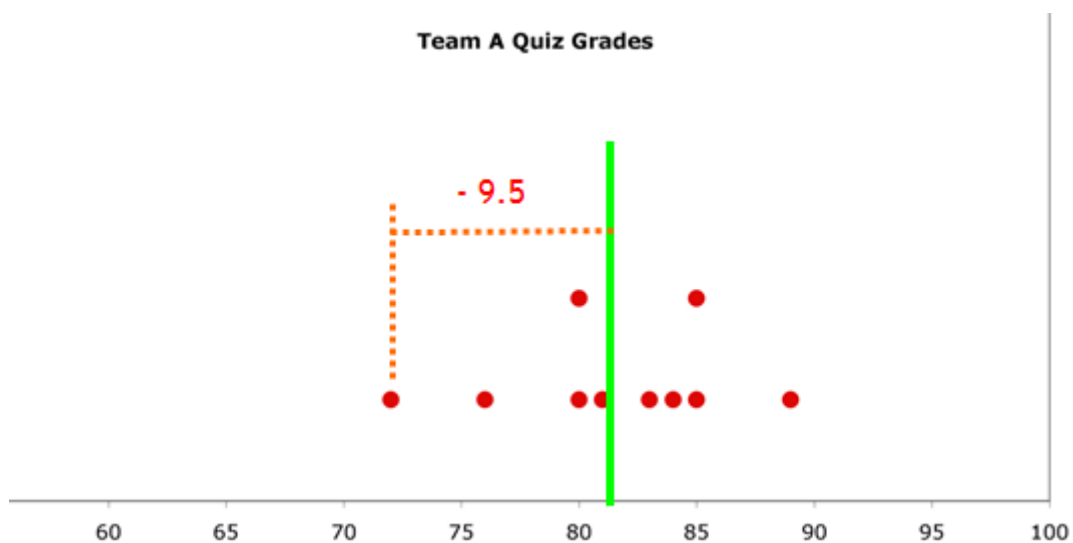
σ = lower case sigma
 \sum = capital sigma
 \bar{x} = x bar

- It tells us what is happening between the minimum and maximum scores
- It tells us how much the scores in the data set vary around the mean
- It is useful when we need to compare groups using the same scale

The Standard Deviation measures how far away each number in a set of data is from their mean.

For example, the lowest score, 72. How far away is 72 from the mean of 81.5?

$$72 - 81.5 = -9.5$$




How to Find the Standard Deviation

1. Find the mean of the data.
2. Subtract the mean from each value – called the deviation from the mean.
3. Square each deviation of the mean.
4. Find the sum of the squares.
5. Divide the total by the number of items – result is the variance.
6. Take the square root of the variance – result is the standard deviation.

Question 22

Three siblings are setting off for the One Direction concert tonight on in the O2 arena, they are aged 12, 17 and 31.

- i. Find their mean age.
- ii. Find the standard deviation from the mean correct to the nearest degree.
- iii. Write down the mean age and the standard deviation from the mean of these three people, five years later.



Question 23

The set S is the number of goals scored by both teams over five games = $\{2, 4, 5, 7, 7\}$

Find:

- i. the mode of S
- ii. the median of S
- iii. the mean of S
- iv. the standard deviation from the mean.

Question 24

The number of Easter eggs twenty students received on Easter Sunday last year.

No. of eggs	0	1	2	3	4
$f(x)$	2	4	8	4	2

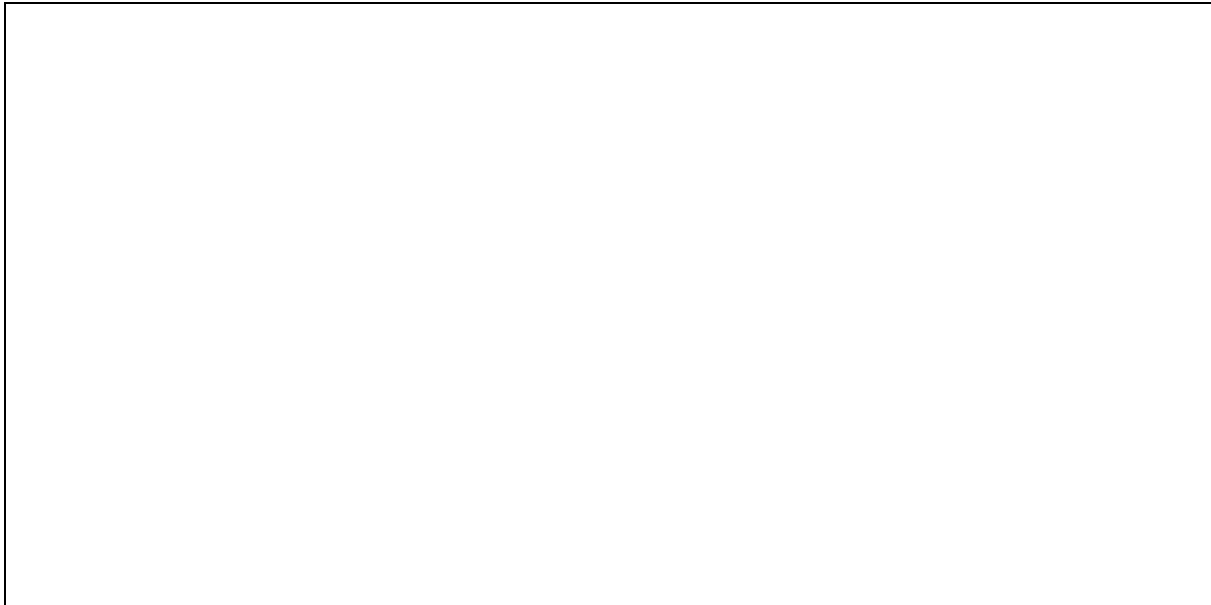
Find the mean and the standard deviation from the mean, correct to one decimal place.

Question 25

Twenty students are asked for how many minutes they spend on facebook or twitter each day. This table shows their replies:

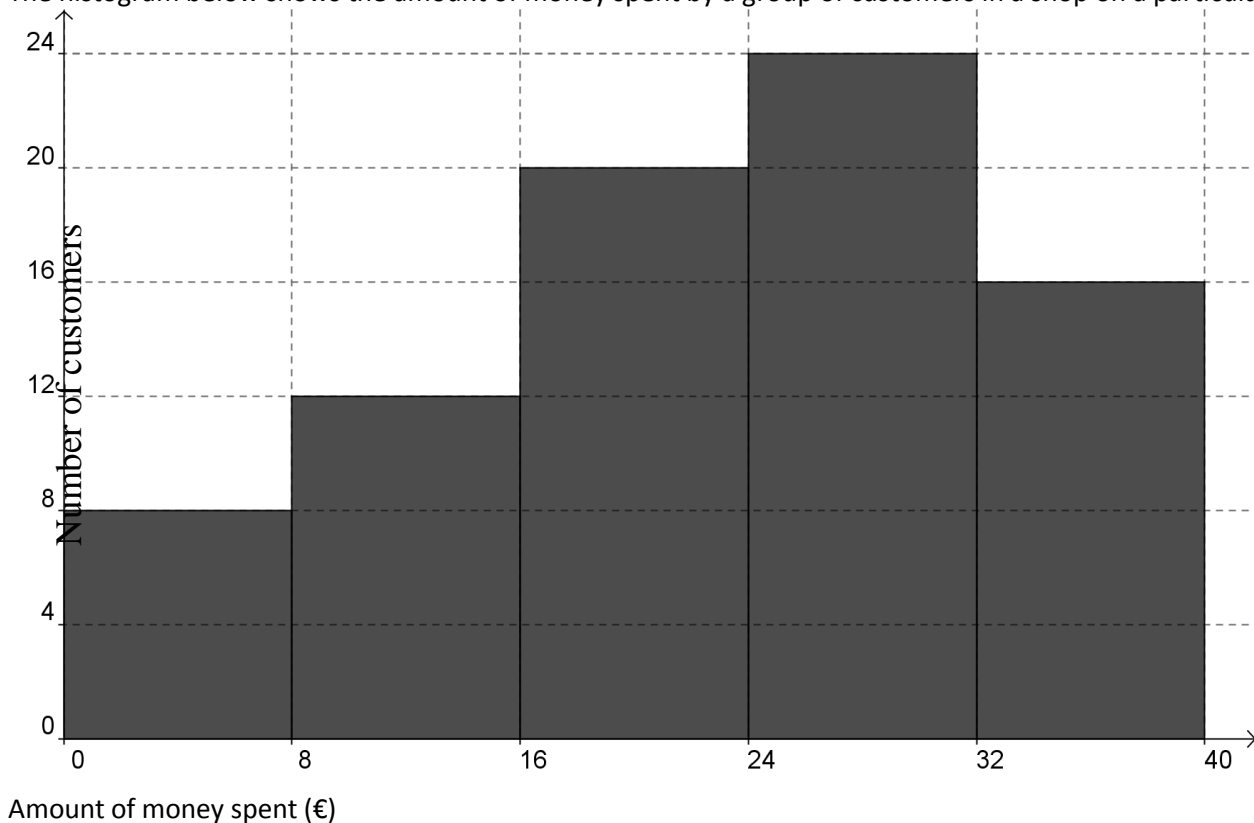
Time (minutes)	0 - 40	40 - 60	60 - 80	80 - 100	100 - 120
Frequency	2	6	5	3	4

- (i) Using the mid – intervals values, estimate the mean viewing time.
- (ii) Find the standard deviation to the nearest minute.



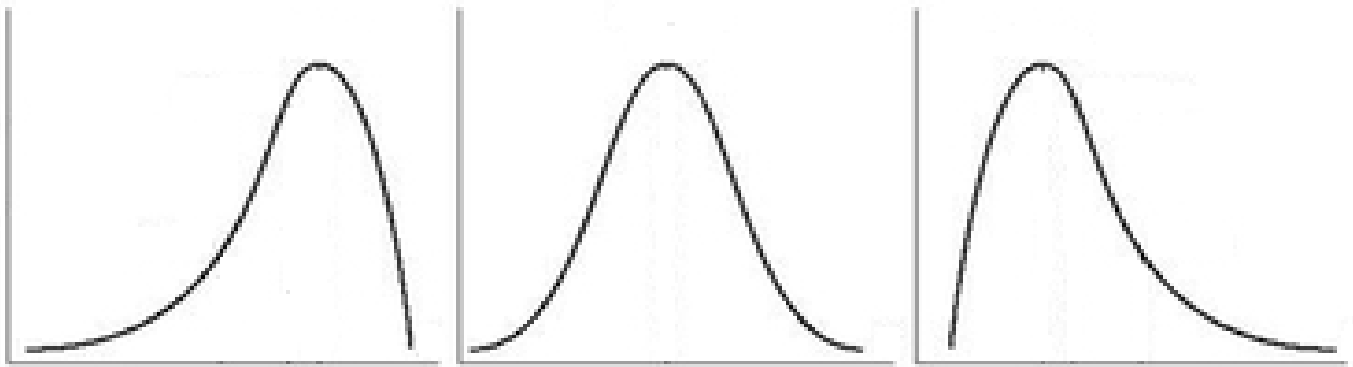
Question 26

The histogram below shows the amount of money spent by a group of customers in a shop on a particular day.



Normal Distribution & Empirical Rule

When describing the **shape** of a graph, say whether it is **symmetrical**...not leaning to one side, or **skewed**...leaning to one side.

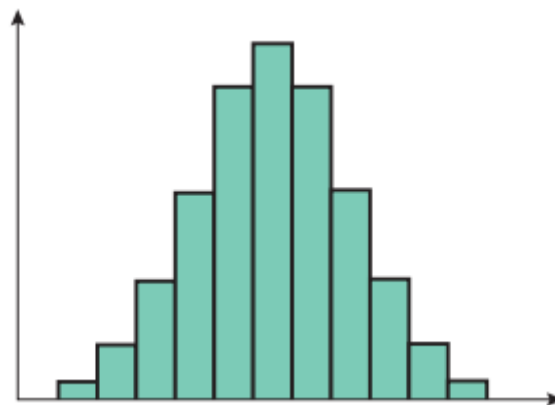


Skewed to the **left**
(look at your left foot)

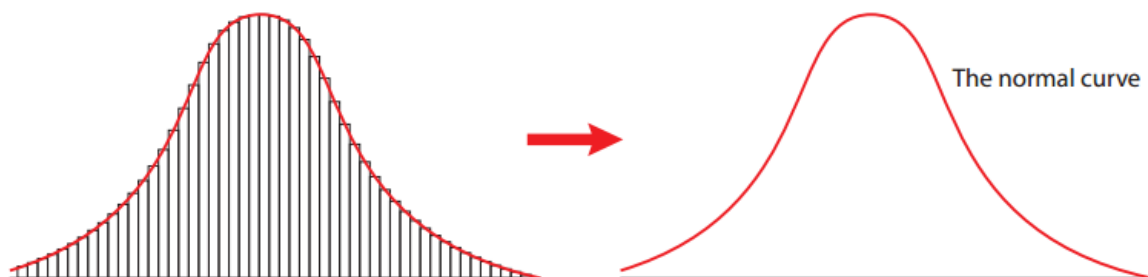
Symmetrical
aka Normal

Skewed to the **right**
(look at your right foot)

When the physical characteristics, such as height or weight, of a large number of individuals are arranged in order, from lowest to highest, in a frequency distribution, the same pattern shows up repeatedly. This pattern shows that a large number of values cluster near the middle of the distribution, as illustrated by the symmetrical histogram shown below:

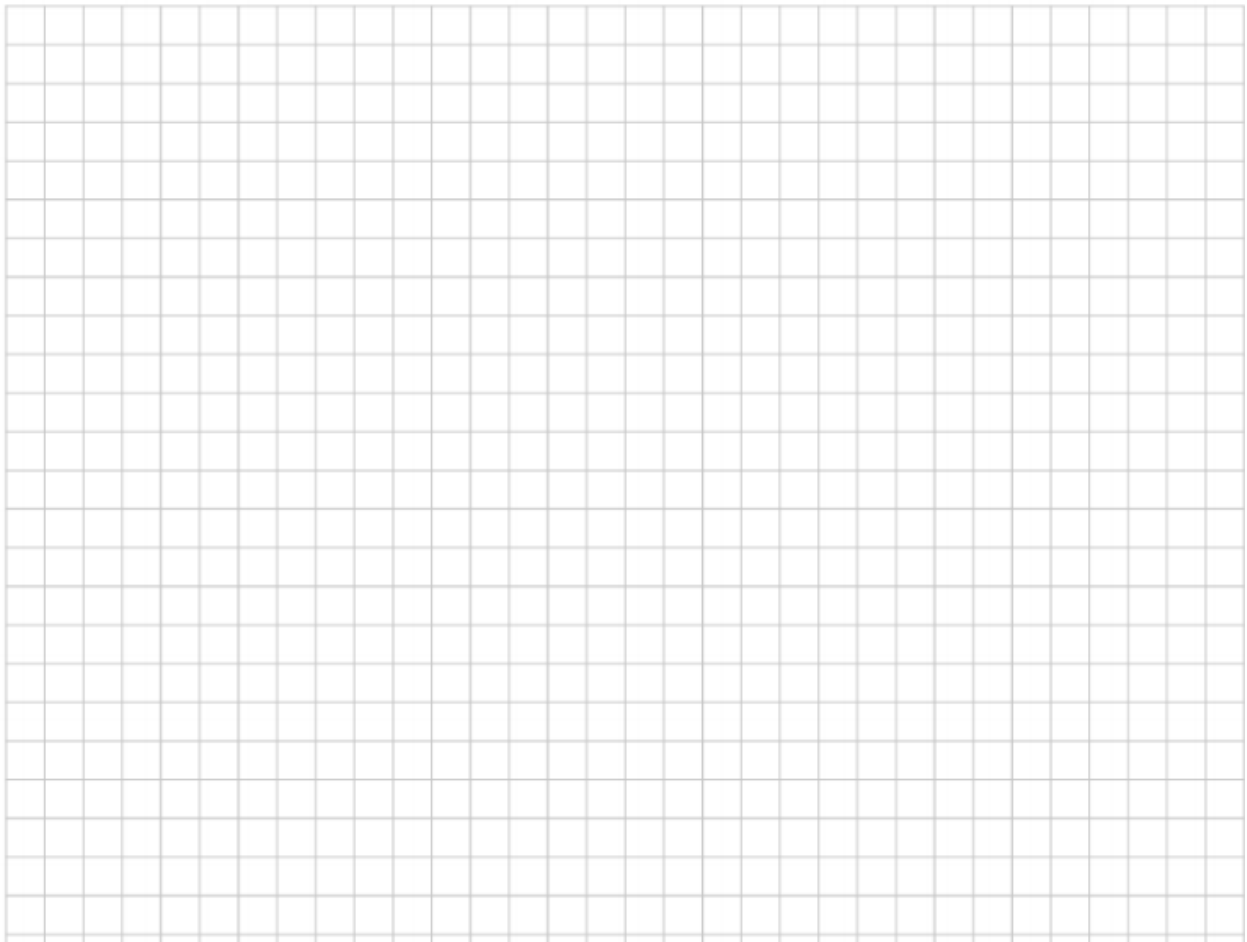


If the distribution is very large and continuous, and the class intervals become sufficiently small, the distribution forms a symmetrical bell-shaped smooth curve called the curve of normal distribution or simply the normal curve, as shown.



Previously, we dealt with a measure of spread called standard deviation, which gives an indication of the distance the data is from the mean.

(v) Display John's data in a pie chart. Show all of your calculations clearly.

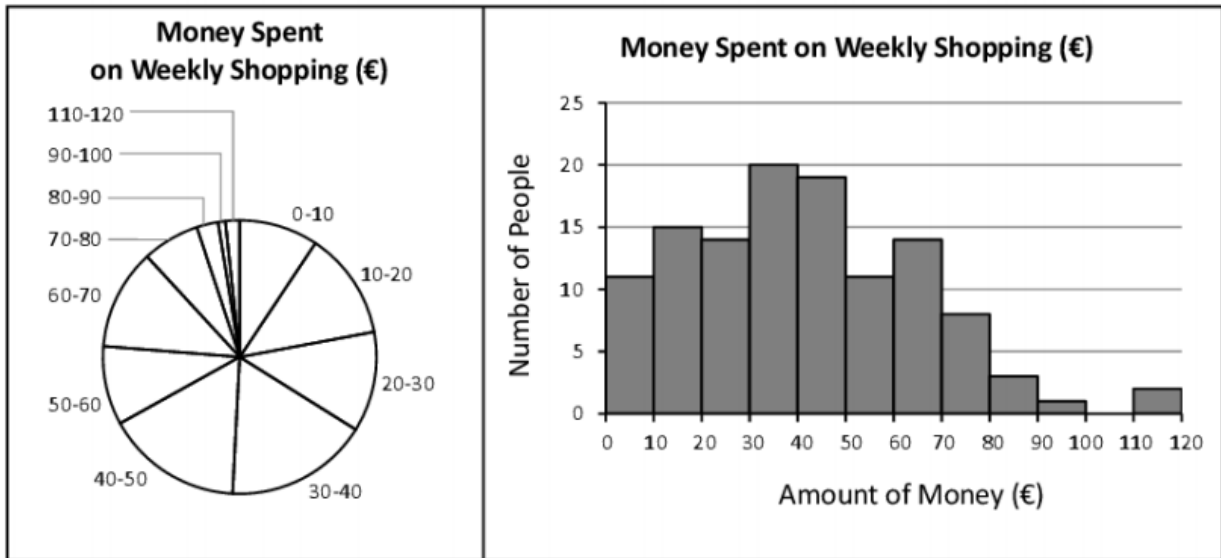
A large grid consisting of 20 columns and 30 rows, intended for students to perform calculations and draw a pie chart.

(b) Margaret wants to examine if people prefer to do their weekly shopping in *Tesco*, *Dunnes Stores*, *SuperValu*, or *Lidl*. She stands outside her local *Lidl* shop for one day, and asks everyone as they leave the shop where they prefer to do their weekly shopping.

Give one reason why Margaret's data may be biased.

A grid consisting of 20 columns and 10 rows, intended for students to write a reason why Margaret's data may be biased.

- (c) Mary is interested in the amount of money people spend on their weekly shopping. She surveys people as they leave the local supermarket on a Saturday morning, and displays her results in the two graphs below.



- (i) Mary wants to show that about half of her sample spent less than €40 on their weekly shopping. Which graph do you think she should use? Give a reason for your answer.

Answer: _____

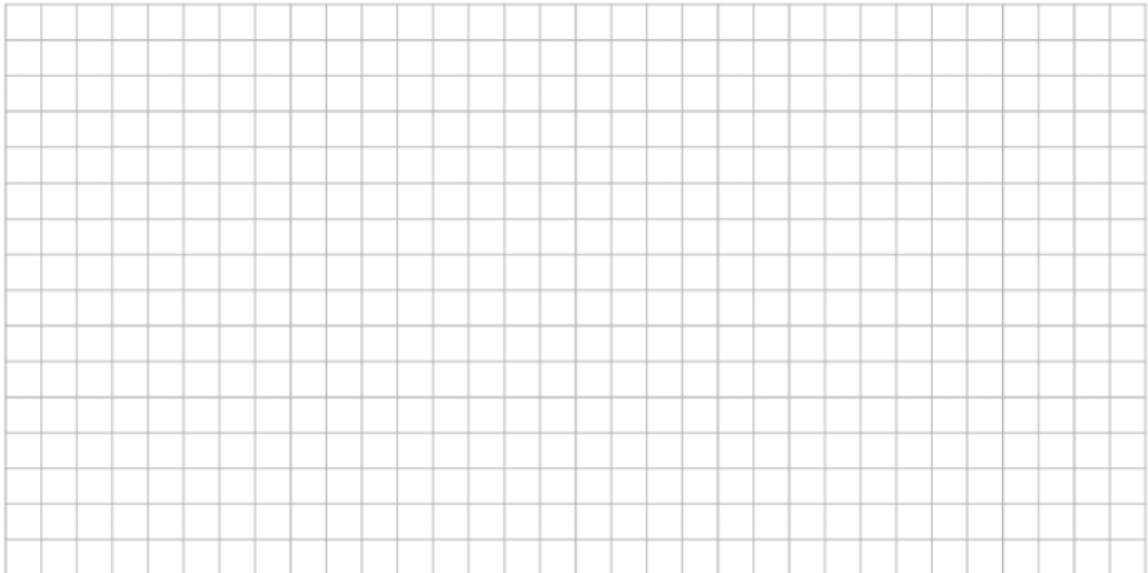
Reason: _____

- (ii) Mary wants to show that there were more people in the 30–40 group than in any other. Which graph do you think she should use? Give a reason for your answer.

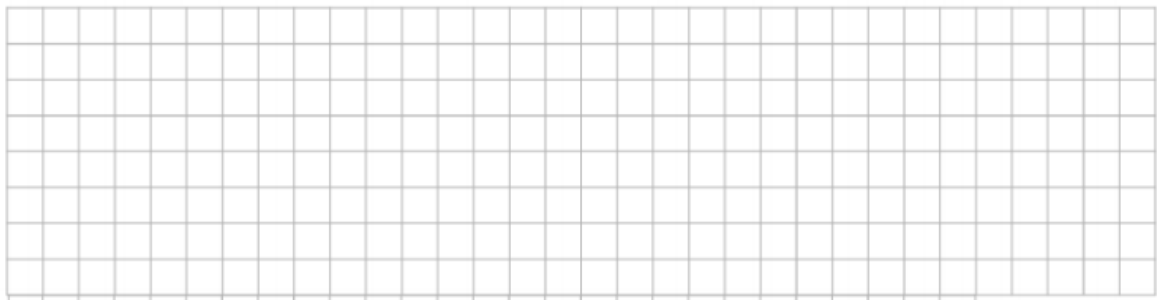
Answer: _____

Reason: _____

(b) Calculate the overall percentage of students who kept their mobile phones under their pillows.



(c) A new pie chart is to be drawn showing the mobile phone location for all students. Calculate the measure of the angle that would represent the students who kept their mobile phones under their pillows.



Question 38

The ages of the Academy Award winners for best male actor and best female actor (at the time they won the award) from 1992 to 2011 are as follows:

Male actor 54 52 37 38 32 45 60 46 40 36 47 29 43 37 38 45 50 48 60 50

Female actor 42 29 33 36 45 49 39 26 25 33 35 35 28 30 29 61 32 33 45 29

(a) Represent the data on a back-to-back stem-and-leaf diagram.

Male actors		Female actors
	2	
	3	
	4	
	5	
	6	
		Key:

(c) (i) Outline another method which could have been used to calculate the mean salary.

(ii) Which method is more accurate? Explain your answer.

Answer																					
Reason																					

Question 40

The ages of the 30 people who took part in an aerobics class are as follows:

18 24 32 37 9 13 22 41 51 49
15 42 37 58 48 53 27 54 42 24
33 48 56 17 61 37 63 45 20 39

The ages of the 30 people who took part in a swimming class are as follows:

16 22 29 7 36 45 12 38 52 13
33 41 24 35 51 8 47 22 14 24
42 62 15 24 23 31 53 36 48 18

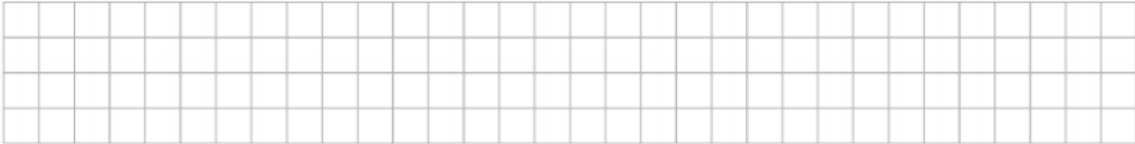
(a) Represent this data on a back-to-back stem-and-leaf diagram.

Aerobics class			Swimming class	
		0		
		1		
		2		
		3		
		4		
		5		
		6		
			Key:	

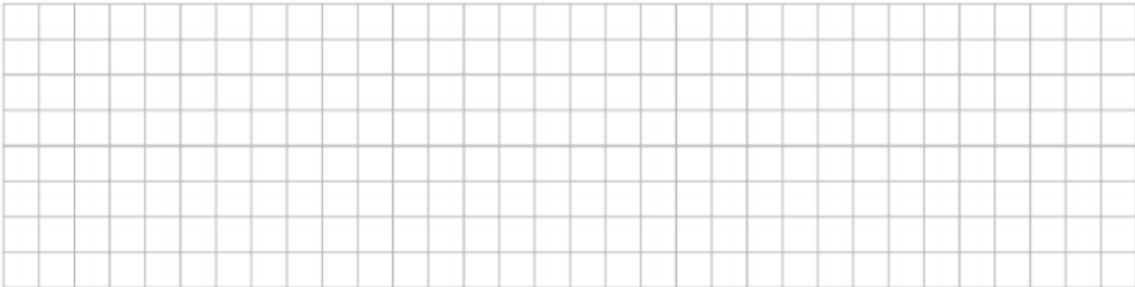
(b) Use your diagram to identify the median in each case.



(c) What other measure of central tendency could have been used when examining this data?



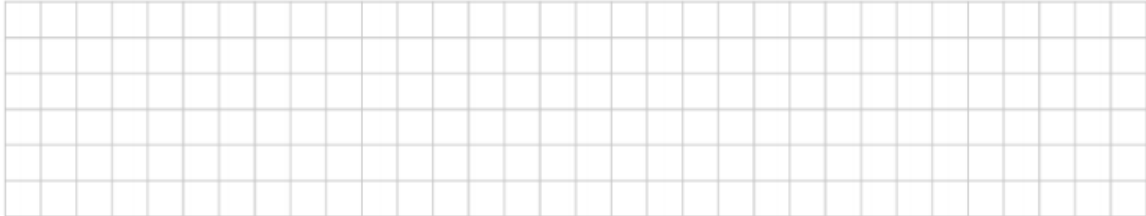
(d) Based on the data make one observation about the ages of the two groups.



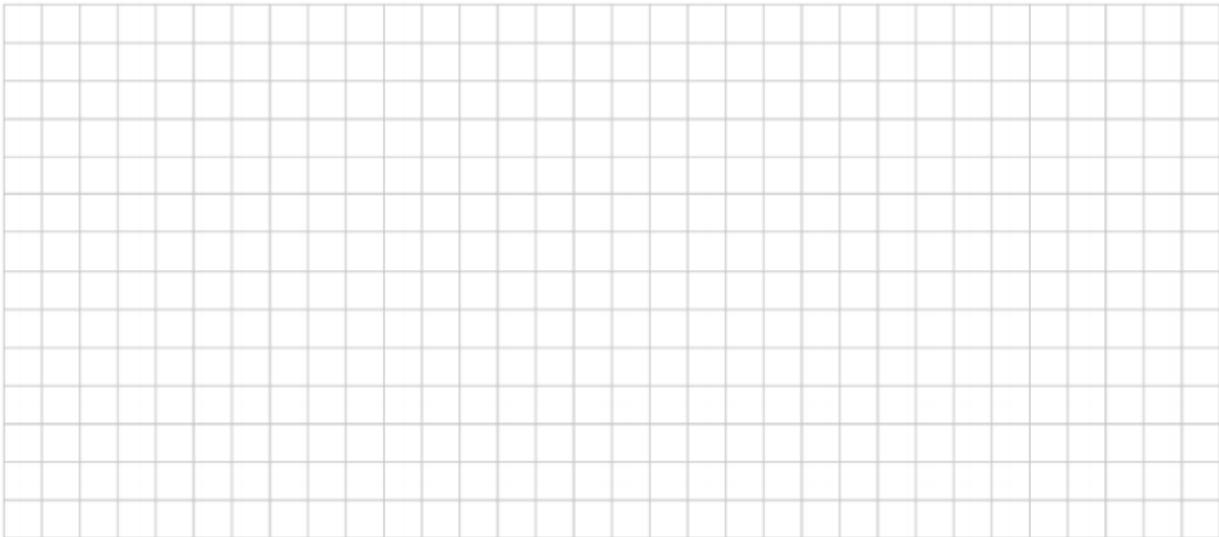
Question 41

A survey is being conducted of voters’ opinions on several different issues.

(a) 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?



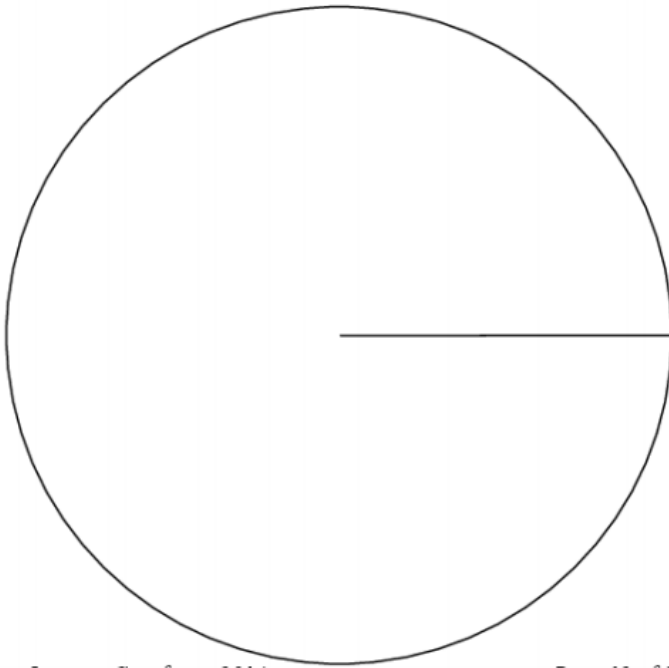
(b) A political party had claimed that it has the support of 24% of the electorate. Of the voters in the sample above, 243 stated that they support the party. Is this sufficient evidence to reject the party’s claim, at the 5% level of significance?



- (b) Homeowners, who registered, were required to value their property in one of a number of given Valuation Bands. The percentage who had valued their properties in each Valuation Band is given in the table below.

Valuation Band	€0 - €100 000	€100 001 - €150 000	€150 001 - €200 000	€200 001 - €250 000	€250 001 - €300 000	Over €300 000
Percentage of registered homeowners	24.9	28.6	21.9	10.4	4.9	9.3

Represent the data in the table using the pie chart below. Label each sector you create and show the angle in each sector clearly.



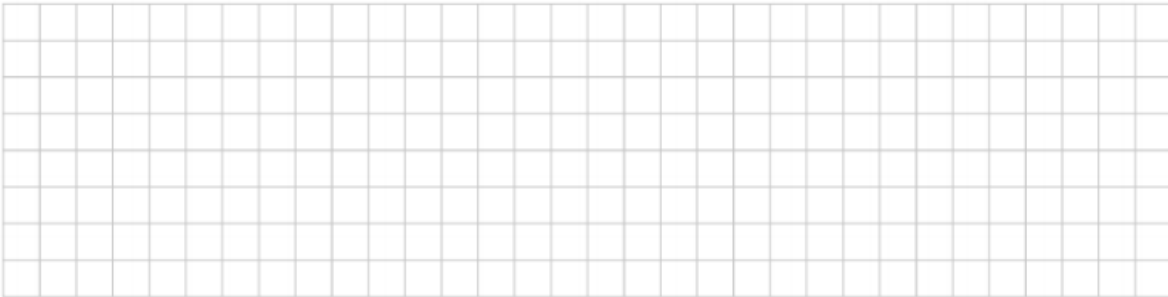
- (c) (i) Use the data in the table above and your answer to part (a) (i) above to complete the following table.

Valuation Band	Tax per property	Number of properties	Total tax due (€)
€0 - €100 000	€45	425 790	19 160 550
€100 001 - €150 000	€112	489 060	
€150 001 - €200 000	€157		
€200 001 - €250 000	€202		
€250 001 - €300 000	€247		
Over €300 000	NA		NA

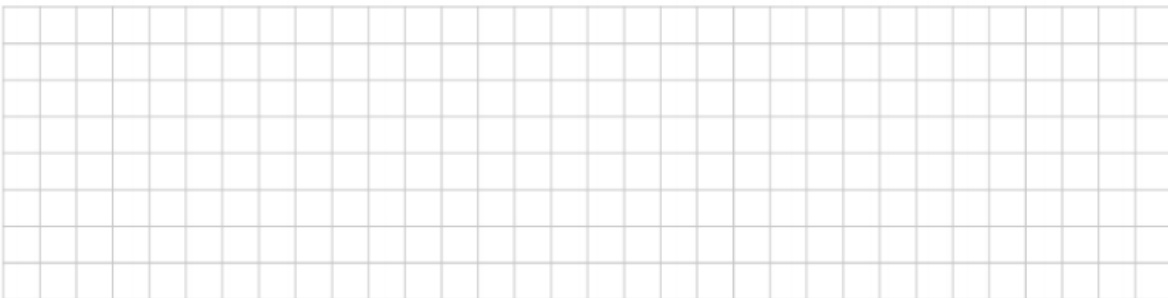
NA = Not Available



(ii) Find the total tax due on those properties, registered by October 2013, with a valuation up to €300 000.



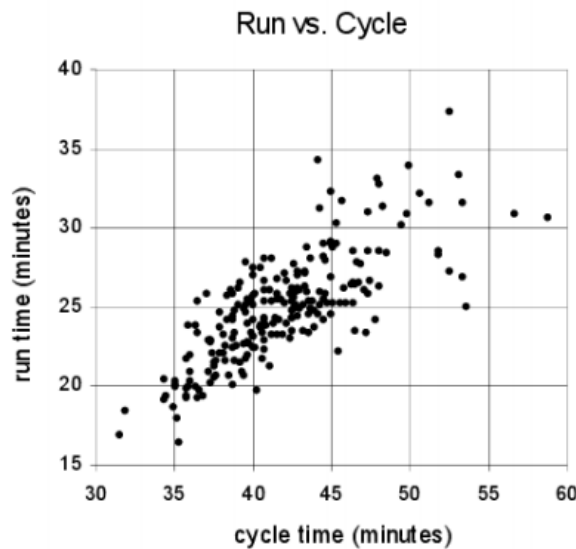
(iii) The total tax due on all the properties that were registered was estimated at €241 million. Find the total tax due on those properties with a valuation over €300 000.



(a) Use the histograms to complete the following sentences:

- (i) The event that, on average, takes longest to complete is the _____.
- (ii) In all three histograms, the times are grouped into intervals of _____ minutes.
- (iii) The time of the fastest person in the swim was between _____ and _____ minutes.
- (iv) The median time for the run is approximately _____ minutes.
- (v) The event in which the times are most spread out is the _____.

(b) Máire is interested in the relationship between the athletes' performance in the run and in the cycle. She produces the following scatter diagram.



(i) The correlation coefficient between the times for these two events is one of the numbers below. Write the letter corresponding to the correct answer in the box.

- A. 0.95
- B. 0.77
- C. 0.13
- D. -0.13
- E. -0.77
- F. -0.95

(ii) Frank was the slowest person in the run. How many people took longer to complete the cycle than Frank did?

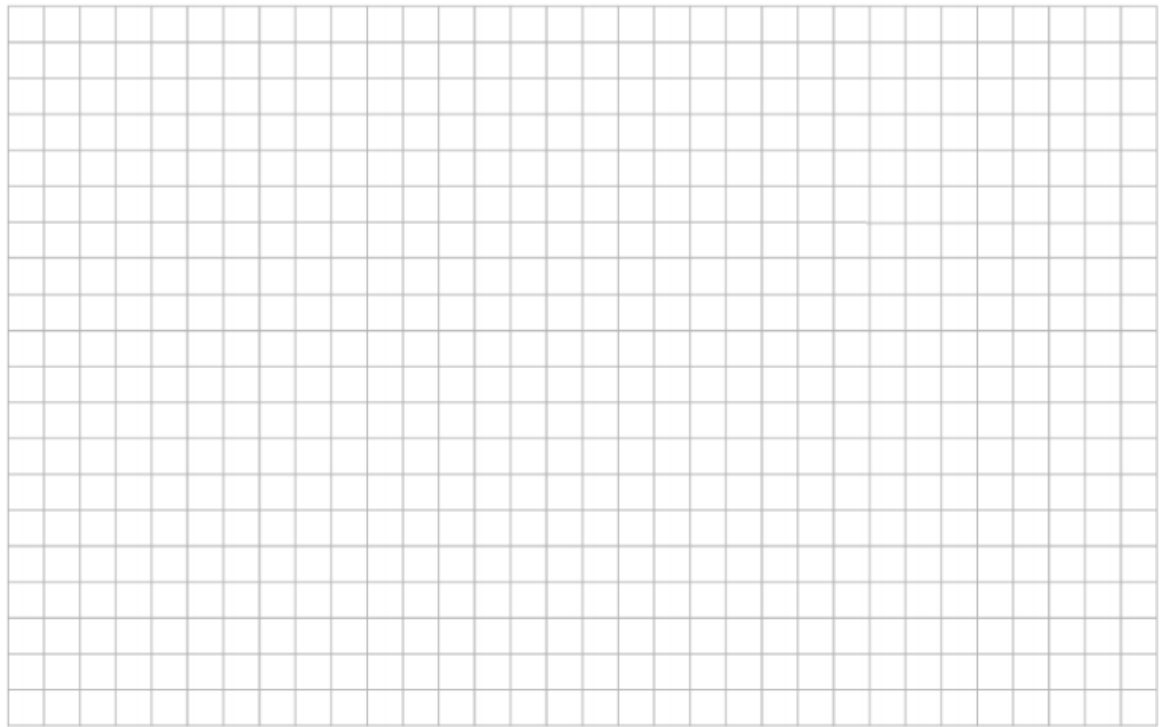
Answer: _____

Question 48

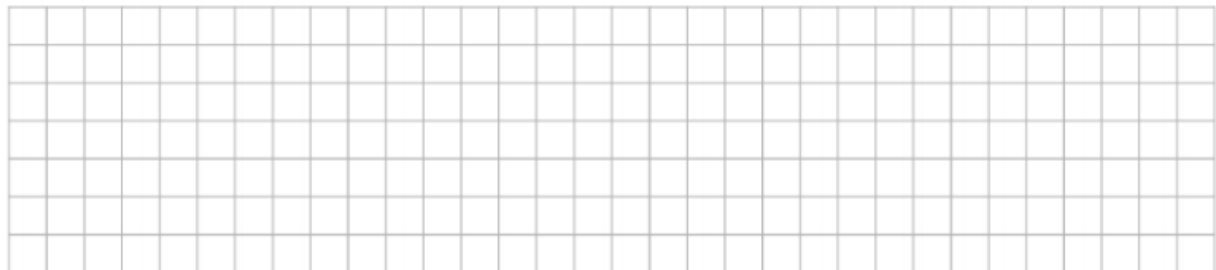
The following table gives data on new private cars sold in Ireland in each quarter of each year from 2006 to 2011.

New private cars sales								
Year	Number of cars sold					Engine type of cars sold		
	January to March	April to June	July to Sept.	October to Dec.	Annual Total	Petrol	Diesel	Other
2006	75 769	54 572	32 873	10 059	173 273	128 634	44 010	629
2007	81 750	57 124	32 418	9 462	180 754	128 346	50 560	1 848
2008	77 441	37 128	27 361	4 540	146 470	92 298	50 283	3 889
2009	27 140	15 225	9 049	3 018	54 432	22 802	30 645	985
2010	34 555	26 806	17 011	6 535	84 907	27 124	53 998	3 785
2011	39 484	29 770	13 467	4 211	86 932	23 246	61 730	1 956

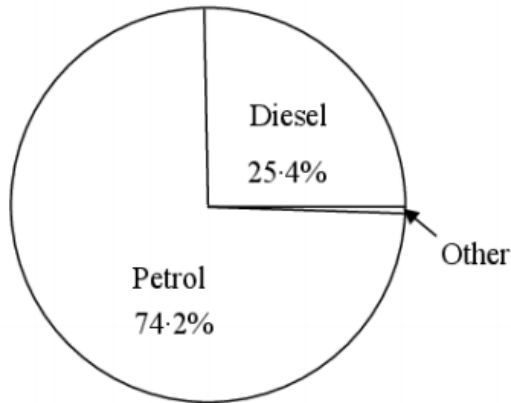
- (a) (i) Show the *annual total* sales of cars over the six years, using a suitable chart.



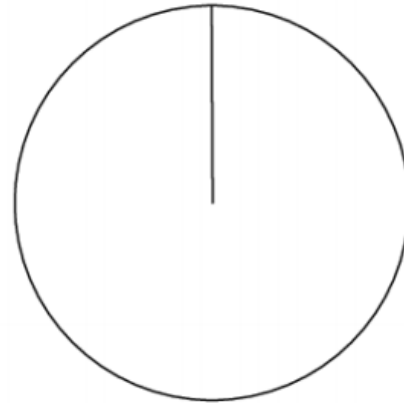
- (ii) Find the mean number of cars sold per year over the six years.



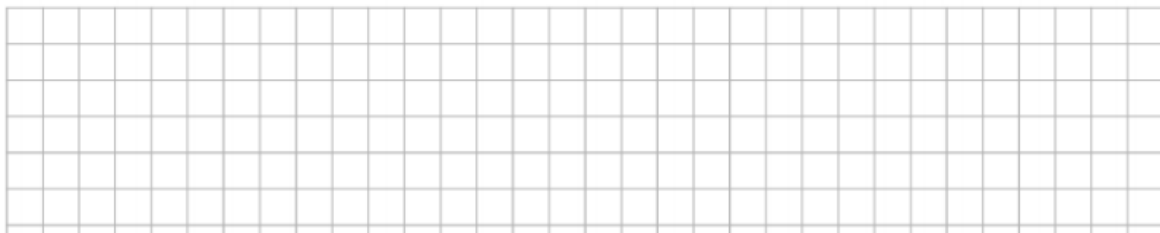
- (c) (i) Two pie charts are being used to show the change from 2006 to 2011 in the popularity of petrol and diesel cars. Complete the second pie chart.



2006



2011



- (ii) Which of the following statements best describes the change over time in the popularity of diesel cars as a percentage of the total?

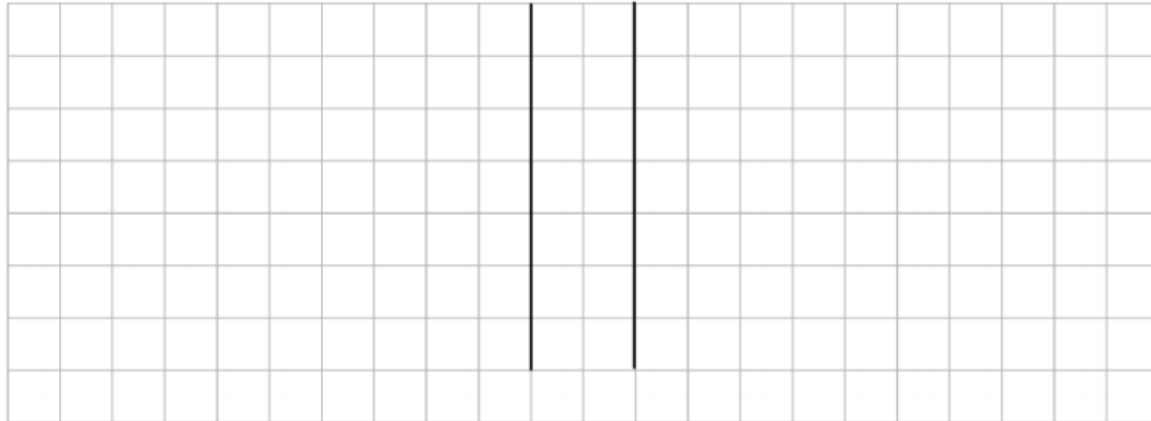
- A. Diesel cars have suddenly become very popular in the last year or two.
- B. Diesel cars have increased very steadily in popularity over the last six years.
- C. Diesel cars have become very popular since car sales started to improve.
- D. Diesel cars got more popular each year, with an especially big increase in 2009.
- E. Diesel cars became popular as car sales fell but have been getting less popular as they rise again.

Write the letter corresponding to the correct answer in the box.

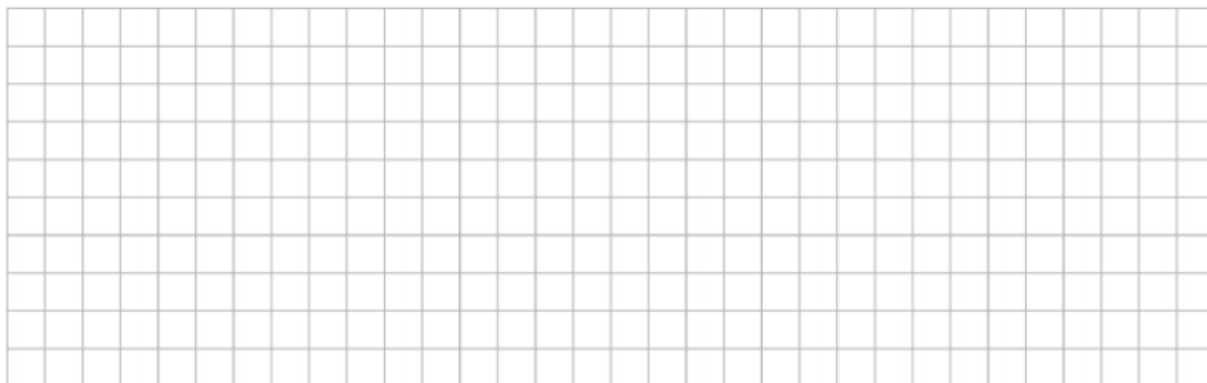
- (d) A survey of some of the most popular models of private cars sold in 2011 examined the CO₂ emissions in g/km from diesel engines and petrol engines. The data are as follows:

Diesel engines	Petrol engines
117, 125, 120, 125, 134, 110,	139, 133, 150, 157, 138, 159,
118, 114, 119, 119, 116, 107.	129, 138, 134, 129, 129, 136.

- (i) Construct a back-to-back stem-and-leaf plot of the above data.



- (ii) Does the information suggest that diesel engines produce lower CO₂ emissions than petrol engines? In your answer you should refer to the stem-and-leaf plot and to an appropriate measure of central tendency.



- (iii) Does the information suggest that there is a greater variation in the CO₂ emissions of diesel engines than petrol engines? In your answer you should refer to the stem-and-leaf plot and an appropriate measure of variability.

