# **Types of Data**

#### **Question 1**

Characterise each of the following variables as <u>numerical</u> or <u>categorical</u>. In each case, list any three possible values for the variable.

(i) Eve colours in a group of children

(i)	ii) Lengths of time taken by competitors to finish a marathon.																
(ii)	Length	ns of tir	ne tak	en by	compe	etitors	to fini	sh a m	aratho	on.							
(iii)	Numb	ers of s	studen	ts atte	ending	a part	icular	school									
(iv)	Counti	es whe	ere a s	ample	of 100	) babie	es wer	<u>e born</u>	. <u> </u>								
(v)	Severi	ty of pa	ain exp	perient	ced by	patier	nts afte	er surg	ery.								

#### **Question 2**

Characterise each of the following numerical variables as <u>discrete</u> or <u>continuous</u>. In each case, give a reason for your choice.

(i) Numbers of texts sent by a boy in a particular week.

1 1	
1	

(v)	Atte	ndanc	e at or	ne tear	n's ma	tches	during	one s	eason.					

Characterise each of the following categorical variables as ordinal or nominal.

In each case, give a reason for your choice.

Also list three possible values for the variable in each case.

(i)	Bloc		ups of a													
(ii)	Oua	lity of	service	in a r	estaur	ant										
(,																
(iii)	Cou	ntries	where	the to	p 50 c	ar moo	dels ar	e mad	e.		 		 			
(iv)	Grad	des ob	tained	by a g	roup c	of stud	ents ir	a mat	ths tes	t.						
(v)	 Hair	colou	r of a g	roup	) of chile	lren.										
(*)																

#### **Question 4**

(i) Give one example of an <u>ordinal</u> categorical variable. List any three possible values for this variable.

#### (ii) Give one example of a <u>nominal</u> categorical variable. List any three possible values for this variable.

LISUAL	yune	e poss	indes it	JI UIIS	variau	ie.						

#### (iii) Give one example of a <u>discrete</u> numerical variable. List the range of possible values for this variable.

	2221010	vulue		ubic.						

(iv) Give one example of a <u>continuous</u> numerical variable. List the range of possible values for this variable.

### **Question 5**

A class test in English poetry consists of 20 questions.

The resulting score from the test reflects work rate and aptitude.

#### (i) List any three ways a student's score could be reported.

• •	,	,							

# (ii) What are the possible values (or range of values) for each of these scoring methods?

300111	gineu	1003:													 
(iii)	Cate	gorise	each i	metho	d of sc	oring	as num	nerical	(discr	ete or	contin	uous)			
	or ca	ategor	ical (no	ominal	or or	linal).							 	 	 


## 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.

## **Finding Averages from Lists of Numbers**

#### **Question 6**

Look at the following list of numbers.

10, 4, 5, 4, 12, 2, 8, 5, 4, 7, 4

- (i) What is the modal number (another term for mode)?
- (ii) What is the median?
- (iii) Calculate the mean of the list.


#### **Question 7**

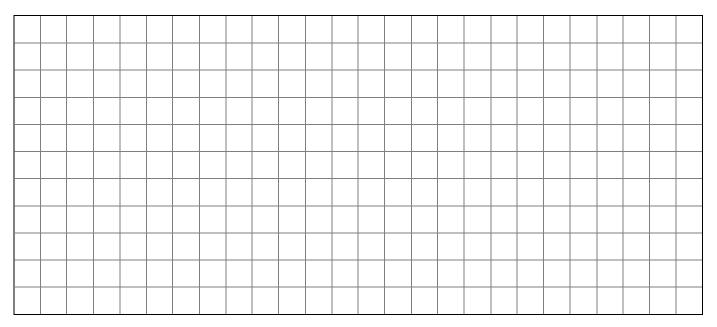
The following list shows the results obtained by a student in 12 science tests over the course of a particular school year.

65 9	E0	62	65	57	64	60	28	73	70	68	61
------	----	----	----	----	----	----	----	----	----	----	----

(i) Find the student's mean result.

- (ii) Find the median result.
- (iii) Write down the student's modal result.
- (iv) Which of the three averages best represents the set of results?

Explain your reasoning.



Look at the frequency table shown below.

Number	1	2	3	4
Frequency	4	3	2	1

(i) Calculate the mean of this set of data.

- (ii) Find the median of this set of data.
- (iii) Write down the modal number from this set of data.

#### **Question 9**

An employer counted the number of days missed by his 21 employees during a particular week. The results are shown below.

1	4	2	3	2	2	5
4	2	4	5	3	4	2
3	5	3	2	1	4	2

#### (i) Complete the following table.

Number of days missed	1	2	3	4	5
Number of workers					

(ii) Calculate the mean number of days missed per employee.

(iii) What is the modal number of days missed?

(iv) What is the median number of days missed?

A test consisting of five questions was given to a class of 25 students. The number of correct answers given by the students is given in the table below.

Number of questions answered correctly	1	2	3	4	5
Number of students	5	2	9	6	3

1. Find the mean number of correct answers given.

## Finding averages from a grouped frequency table

When calculating the mean from a grouped frequency table, you have to use mid-interval values. These values are the mean of the two end-point numbers.

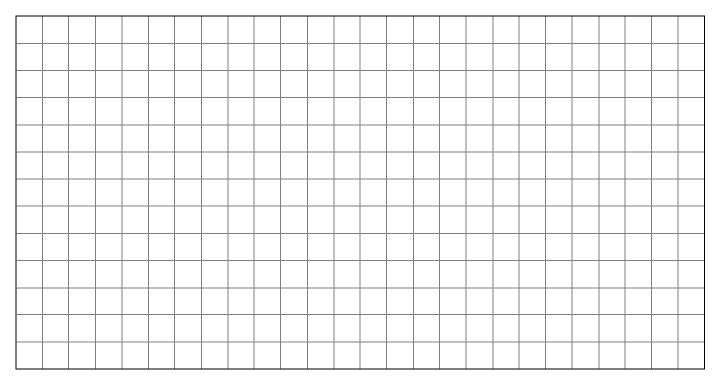
Mid-interval value 
$$=\frac{0+8}{2}=\frac{8}{2}=4$$

#### **Question 11**

The length of time (in hours) taken by a group of workers to complete a given task was recorded. The results are summarised in the table below.

Time taken (hours)	0 – 2	2-4	4 – 6	6 – 8
Number of workers	12	9	6	3

- (i) How many workers were observed altogether?
- (ii) Use mid-interval values to calculate the mean amount of time taken.
- (iii) In which interval does the median lie?
- (iv) What is the modal interval?
- (v) "The length of time taken is a continuous numerical data". Do you agree with this statement?



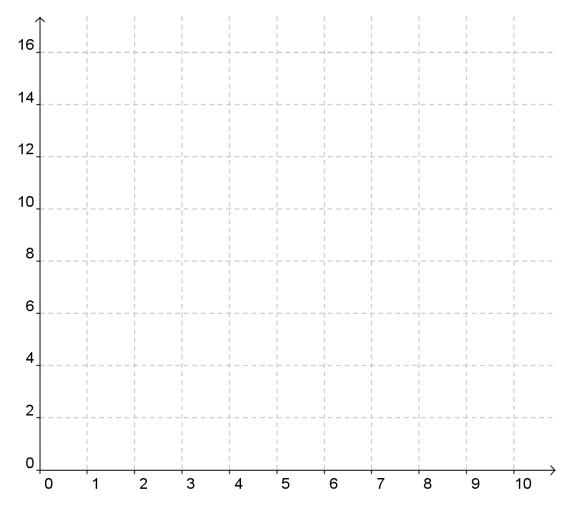
The number of raffle tickets bought by a group of 40 visitors to a school fair is shown in the table below.

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.

So 
$$\frac{1}{4}(n+1) = \frac{1}{4}(7+1) = \frac{1}{4}(8) = 2$$
...so the 2<sup>nd</sup> number is the lower quartile!

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

And 
$$\frac{3}{4}(n+1) = \frac{3}{4}(7+1) = \frac{3}{4}(8) = 6$$
...so the 6<sup>th</sup> number is the upper quartile!

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

#### **Example 2**

- 1. If the list of numbers was 1, 1, 2, 2, 3, 4, 5, 7... there are 8 numbers on the list... n = 8.
- 2. So  $\frac{1}{4}(n+1) = \frac{1}{4}(8+1) = \frac{1}{4}(9) = 2.25$ ...so the average of the 2<sup>nd</sup> and 3<sup>rd</sup> numbers (the numbers on either side of 2.25) on the list is the law are questiled.

side of 2.25) on the list is the lower quartile!

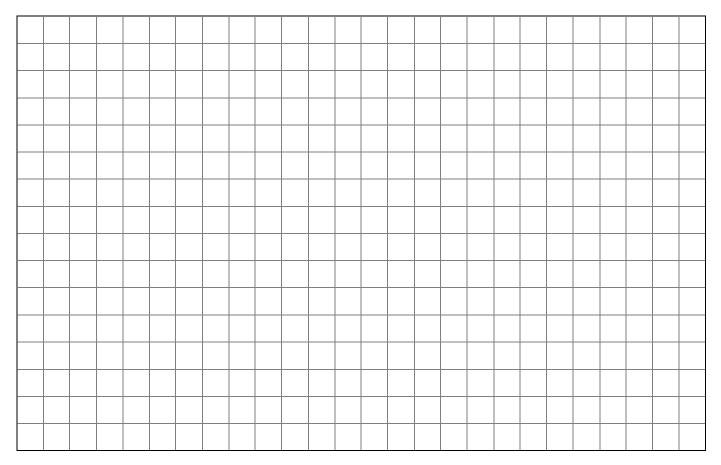
- 1, 1, 2, 2, 3, 4, 5, 7...so  $\frac{1+2}{2} = \frac{3}{2} = 1.5$  is the lower quartile. 3.  $\frac{3}{4}(n+1) = \frac{3}{4}(8+1) = \frac{3}{4}(9) = 6.75$ ...so the average of the 6<sup>th</sup> and 7<sup>th</sup> numbers (the numbers on either side of 6.75) on the list is the upper quartile!
  - 1, 1, 2, 2, 3, 4, 5, 7...so  $\frac{4+5}{2} = \frac{9}{2} = 4.5$  is the upper quartile.
- 4. The interquartile range is upper quartile lower quartile = 4.5 1.5 = 3

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

#### **Question 13**

Look at the following list of numbers. 10, 4, 5, 4, 12, 2, 8, 5, 4, 7, 4

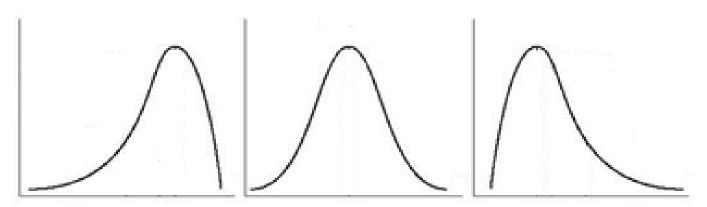
- (i) Find the range of the list.
- (ii) Find the interquartile range of the list.



Look at the following list of numbers. 3, 9, 6, 4, 8

- (i) Find the range of the list.
- (ii) Find the interquartile range of the list.


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)

## **Stem & Leaf Diagrams**

When drawing a stem and leaf diagram...

- (i) everything except the last digit goes down the centre...in the "stem",
- (ii) the last digit only goes out to the side...in the "leaf",
- (iii) digits in the leaves should be **ordered**...arranged in order of size with the smallest closest to the stem,
- (iv) if there are two sets of data (which you are comparing), one set of data forms a leaf on one side of the stem and the other forms a leaf on the other side,
- (v) Note: this is called a back-to-back stem and leaf diagram
- (vi) a key must be included...on **both** sides if there are two sets of data.

#### Example 1

If the set of data was 108, 105, 121, 123, 128, 125, 134 and 132...

10	5	8			
11					
12	1	3	5	8	
13	2	4			Key:  10 5 = 105
"stem"		"leaf"			

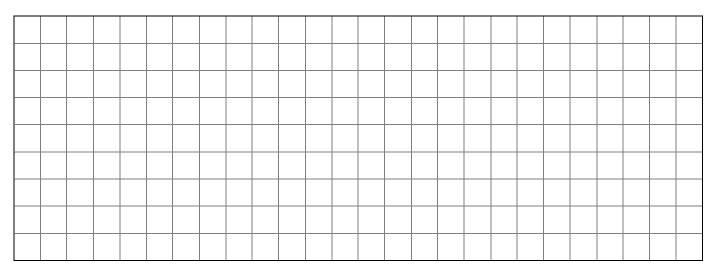
#### **Question 15**

The temperature was recorded every day for two weeks in a particular town.

The results are shown below.

14∙3℃	11.5℃ 11.2℃		10∙8℃	11∙7℃	13∙6℃	13∙9℃		
14·0℃	13·6℃	11·4℃	11·4℃	10·2℃	13·5℃	13∙1℃		

Represent this data on a stem and leaf diagram.

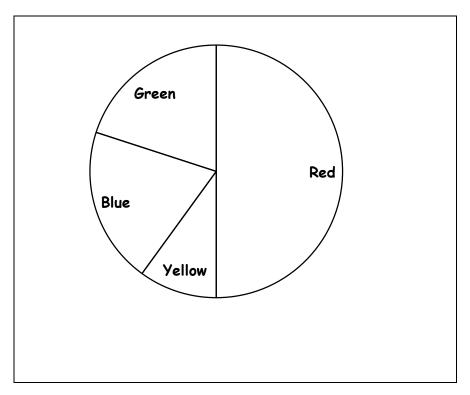


# **Pie Charts**

INTERPRETING PIE CHARTS

## **Question 16**

90 people's favourite colour

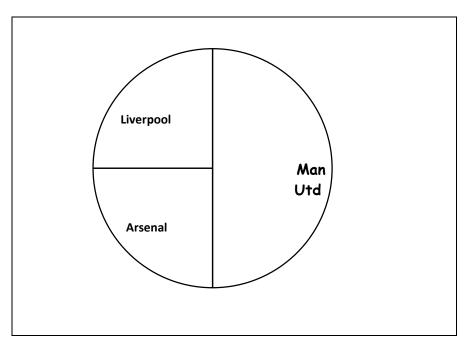


Favourite colour	Degrees	Value
Red	180 <sup>0</sup>	
Yellow	36 <sup>0</sup>	9
Blue	72 <sup>0</sup>	
Green	72 <sup>0</sup>	
1 person = degrees		

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


Pet	Degrees	Value
Dog	180 <sup>0</sup>	
Cat	36 <sup>0</sup>	
Budgie	36 <sup>0</sup>	
Goldfish	36 <sup>0</sup>	
Other		
1 person = degrees		

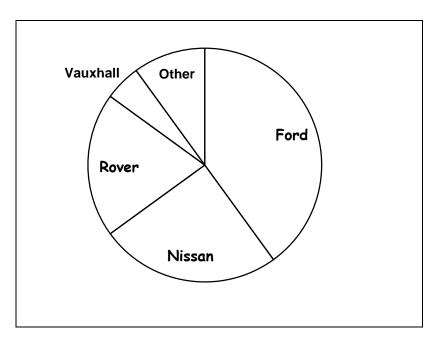
## **Question 18** Favourite football team of 120 people



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

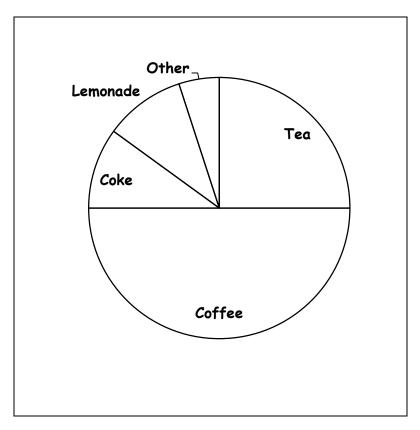


Make of \_\_\_\_\_ cars



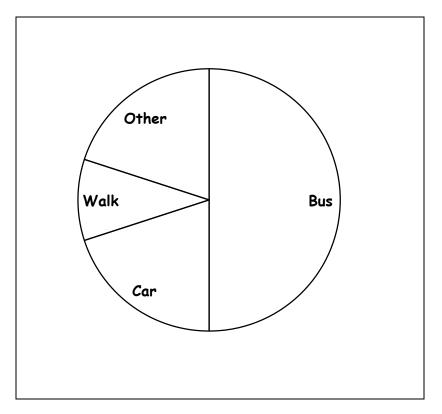
Make of car	Degrees	Value
Ford	144 <sup>0</sup>	
Nissan	90 <sup>0</sup>	
Rover	72 <sup>0</sup>	8
Vauxhall		
Other	36 <sup>0</sup>	
1 person = degrees		

Favourite drink of 80 people



Drink	Degrees	Value
Теа	90 <sup>0</sup>	
Coffee	180 <sup>0</sup>	
Coke	36 <sup>0</sup>	
Lemonade		
Other	18 <sup>0</sup>	
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?



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?

## **Past Exam Questions**

#### **Question 22**

All of the students in a class took *IQ Test 1* on the same day. A week later they all took *IQ Test 2*. Their scores on the two IQ tests are shown in the tables below.

	IQ Test 1								
86	104	89	105	96					
96	103	94	104	119					
115	79	97	111	108					

IQ Test 2												
83	120	105	111	114								
99	111	108	106	97								
97	102	94	108	117								

(i) Draw a back-to-back stem-and-leaf plot below to display the students' scores.

IQ Test 1		IQ Test 2
	7	
	8	
	9	
	10	
	11	
	12	

(ii) Find the range of scores for each IQ test.

IQ Test 1:	IQ Test 2:

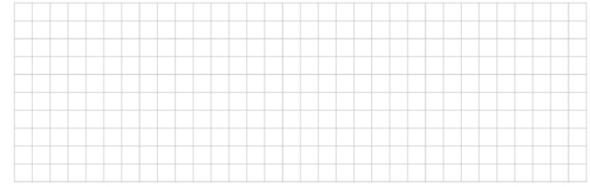
#### (iii) Find the median score for each IQ test.

IQ Test 1:	IQ Test 2:

#### (iv) Find the mean score for each IQ test.

IQ Test 1:	IQ Test 2:

(v) Compare the scores on the two IQ tests. Refer to at least one measure of central tendency and at least one measure of variability (spread) in your answer.



(vi) Marshall says that every student in the class must have done better on *IQ Test 2* than on *IQ Test 1*. Is Marshall correct? Explain your answer.

Students in a class are investigating spending in their local area. They each carry out a different survey, and display the results.

(a) John is investigating whether people pay for their weekly shopping with Credit Card, Debit Card, Cash, or Cheque. When people tell him which one of these they usually use, he writes it in a table. His results are shown below.



Credit Card	Debit Card	Debit Card	Cash	Debit Card
Credit Card	Cash	Cash	Credit Card	Debit Card
Debit Card	Debit Card	Cheque	Cash	Cash
Cash	Cash	Debit Card	Cash	Credit Card

(i) What type of data has John collected? Put a tick ( $\checkmark$ ) in the correct box below.

Numerical	Numerical	Categorical	Categorical
Continuous	Discrete	Nominal	Ordinal

(ii) Fill in the frequency table below.

Method of Payment	Credit Card	Debit Card	Cash	Cheque
Frequency				

		_														

e mode of John's data? Mode =	What is the mode of John's dat
-------------------------------	--------------------------------

#### (iv) John says that he cannot find the mean of his data. Explain why this is the case.

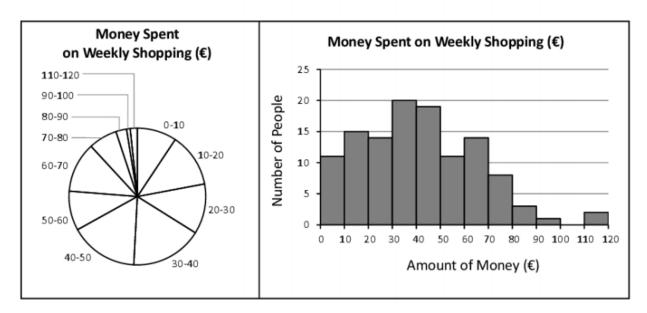
(v)	Display	John S	uata in a	pie chart.	SHOW	your	carcun	ations	ciear	iy.		
												_
												_
_												
_						 						
_												_
_												_
_											_	_
											_	_
_											_	_
_												_
						 						_
_			_			 						_
_												_
_												_
_												
_												
_						 						
_					++-					+		
-										+++		
-												

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

(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.

(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.



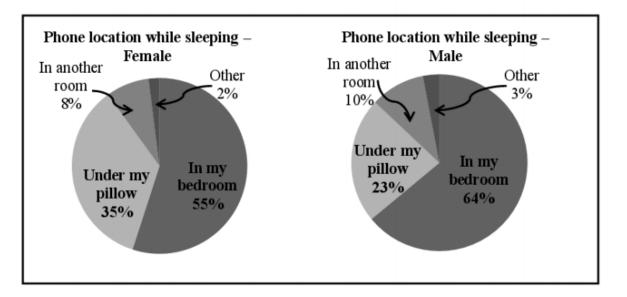
(i) Mary wants to show that about half of her sample spent less than €40 on their weekly shopping. Which graph do you think she should use? Give a reason for your answer.

Answer:									
Reason:								-	

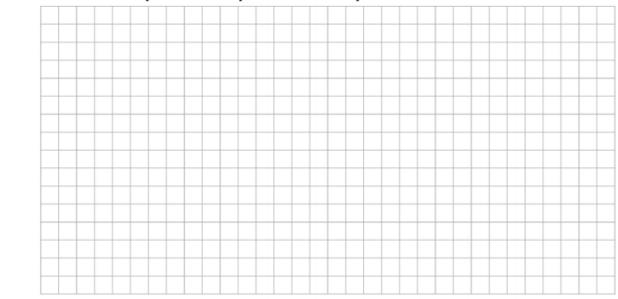
(ii) Mary wants to show that there were more people in the 30–40 group than in any other. Which graph do you think she should use? Give a reason for your answer.

Answer:								
Reason:								
				_		_		
				-				

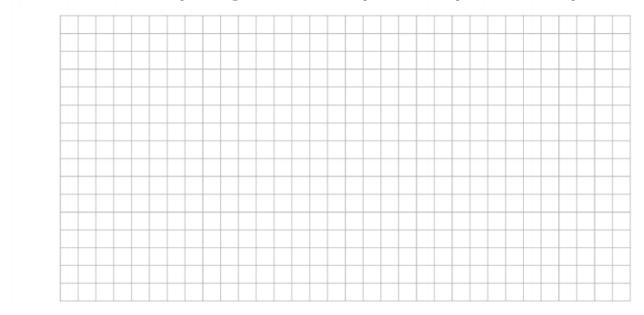
In total 7150 second level school students from 216 schools completed the 2011/2012 phase 11 *CensusAtSchool* questionnaire. The questionnaire contained a question relating to where students keep their mobile phones while sleeping.



(a) Given that this question was answered by 4171 girls and 2979 boys, calculate how many female students kept their mobile phones under their pillows.



#### (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 25**

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:

# (b) State one similarity and one difference that can be observed between the ages of the male and female winners.

			_	-																			
	 _	_	-	_	_			 	-	 	-	-	 _	 		_	_	_	-	-		_	
	 	_	-			_	-	 		 	-	-	 	 	 	_	_				 	-	-
	 _	_	-			_	_			 _	-		 	 		_	_				 		-
	 		-					 	_	 	_	_	 	 	 	_					 	_	_

#### (c) Mary says "The female winners were younger than the male winners." Investigate this statement in relation to:

#### (i) The mean age of the male winners and mean age of the female winners.

-	

## (ii) The median age of the male winners and the median age of the female winners.

				 	- 1												
					- 1												
					-	 _											_
				 	- 1												
		_	-		-	 -	 -	_		-	-	-		-		-	
					- 1												
	$\vdash$	_				 -	 _	_			 _	 	 		_		 
					- 1												
	$\vdash$	_		 		 -	 	_	_		 _	 	 		_		 
				 	- 1												
 		_				 	 				 	 	 		 		 _
					- 1												
		_				 	 				 	 	 		 		 
					- 1												
				_	_	 						 					
					- 1												
					_												

#### (d) Find the interquartile ranges of the ages of the male winners and of the female winners.

Male	Female	

#### **Question 26**

The salaries, in €, of the different employees working in a call centre are listed below.

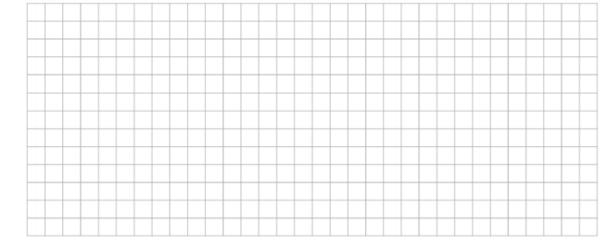
22 000	16500	38000	26 500	15 000	21 000	15500	46000	
42 000	9500	32000	27000	33 000	36 000	24000	37000	
65 000	37000	24 500	23 500	28 000	52 000	33000	25000	
23 000	16500	35000	25 000	33 000	20 000	19500	16000	

Use this data to complete the grouped frequency table below. (a)

Salary (€1000)	0-10	10 - 20	20 - 30	30-40	40 - 50	50 - 60	60 – 70			
No. of Employees										
Note: $10 - 20$ means $\notin 10000$ or more but less than $\notin 20000$ etc.]										

[Note: 10 - 20 means  $\in 10000$  or more but less than  $\in 20000$ , etc.]

## (b) Using mid-interval values find the mean salary of the employees.



(c)	(i)	Outline another method which could have been used to calculate the mean salary.
	(ii)	Which method is more accurate? Explain your answer.
	A	nswer
	R	leason
	-	

The age	s of th	e 30 peo	ple who	took pai	rt in an a	erobics	class are	e as follo	WS:		_
	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 age	s of th	e 30 peop	ple who	took pa	rt in a sv	vimming	g class ai	e as foll	ows:		
	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 tendancy could have been used when examining this data?

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

	 _		_	_	_		_		_		_	_	_	 _			_	_	_	_	_			_			-
																											Ī
																											Î
																										_	İ
	-					-	-	 _	_	-		-		 -	_		-			_	-	_	_	-	-	_	t
-				_	-	-	-	 _		-		_	_	 -		_	_	_	_		-			_	_		ł
-				_	-			 						 _					_								ł
_	_	_			_		_	 											_								1