## **Types of Data**

#### **Question 1**

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

(i)		Eye co	lours i	n a gro	oup of	childre	en.										
(ii	)	Length	ns of tii	ne tak	en by	compe	etitors	to fini	sh a m	aratho	on.						
(iii	i)	Numb	ers of s	studer	its atte	ending	a part	icular	school								
(iv	()	Counti	ies wh	ere a s	ample	of 100	) babie	es wer	e born								
(						0.100											
(v)	)	Severi	ty of p	ain exp	berien	ced by	patier	nts afte	er surg	ery.		1		1	1	1	

#### **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.

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

Height	s of th	e trees	s in a p	articu	ar wo	odland	۱.									
onthly re	ents (ir	n €) for	brobe	erties i	n Dubl	in.										
	Height	Heights of th	Heights of the trees	Heights of the trees in a p	Heights of the trees in a particul Heights of the trees in a particul ponthly rents (in €) for properties in Heights of the trees in a particul	Heights of the trees in a particular wo	Heights of the trees in a particular woodland Description of the trees in a particular woodland Heights of the trees in a particular woodland Description of the trees in a particular woodland Description of the trees in a particular woodland Heights of t	Heights of the trees in a particular woodland.         Heights of the trees in a particular woodland.         Donthly rents (in €) for properties in Dublin.	Heights of the trees in a particular woodland.         Heights of the trees in a particular woodland.         Donthly rents (in €) for properties in Dublin.	Heights of the trees in a particular woodland.  Heights of the trees in a particular woodland.  Donthly rents (in €) for properties in Dublin.	Heights of the trees in a particular woodland.         Heights of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in a particular woodland.         Image: state of the trees in trees in a particular woodland.         Image: state of the trees in t	Heights of the trees in a particular woodland.         Heights of the trees in a particular woodland.         Image: Stress of the tree	Heights of the trees in a particular woodland.         Heights of the trees in a particular woodland.         Image: Second seco	Heights of the trees in a particular woodland.         Heights of the trees in a particular woodland.         Image: Stress of the tree	Heights of the trees in a particular woodland.         Heights of the trees in a particular woodland.         Image: Second Seco	Heights of the trees in a particular woodland.         Heights of the trees in a particular woodland.         Image: Stress of the tree

(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	od groi	ups of	a num	ber of	patien	its.											
(ii)	Qua	lity of	service	in a r	estau	ant.												
(iii)	Cou	ntries	where	the to	p 50 c	ar moo	dels ar	e mad	e.	1	1	1	1	1	1	1	1	1
(iv)	Gra	des ob	tained	by a g	roup o	of stud	ents ir	i a ma	ths tes	t.								
(v)	Hair	colou	r of a g	group (	of chilo	dren.						,						

#### **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.

Listai	iy unc	c p033	nucs n	51 this	variab	ic.						

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


(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?

	0														
(iii)	Cate or ca	gorise ategori	each i ical (no	metho ominal	d of so or orc	oring a linal).	as num	nerical	(discr	ete or	contin	uous)			

## 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 59	62 65	57 64	60 28 73	3 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.

## **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°C	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?

## **STANDARD DEVIATION**



- 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.

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.

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.

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

Time	0 - 40	40 - 60	60 - 80	80 - 100	100 -120
(minutes)					
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.



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



Amount of money spent (€)

(i)		Com	plete	e the	follo	owin	g tab	ole.	-															
	Am	noun	t of r	none	ey sp	ent (	€)		C	)-8			8-16			16-24			24-32			32-40		
	Nu	Number of customers								12														
(ii)		Usin	g mi	d-int	erva	l valu	ues, e	estima	ate	the r	near	n am	ount	of n	none	ey spo	ent.				I		1	 
(iii) 																								
																								<u> </u>
(111)		Wha	t wa	s the	star	ndaro	d dev	/iatior	1 of	the	data	?												
									_															
(iv)		 Wha	t wa	s the	e mo	dal ir	l nterv	al?																
• •																								
	L	I	1	1	1	1	1			<u> </u>	1	1	1	1	I	1	1	1	1		1			

#### What was the maximum number of customers that could have spent more than the mean amount of (v) money in the shop?


## **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.

There is a very important relationship between the normal curve and standard deviations. It is called the Empirical Rule and it is given below:

For any large population with mean  $\overline{x}$  and standard deviation  $\sigma$ 

- (i) about 68% of the values will lie within one standard deviation of the mean, that is, between  $\bar{x} \sigma$  and  $\bar{x} + \sigma$
- (ii) about 95% of the values will lie within two standard deviations of the mean, that is, between  $\bar{x} 2\sigma$  and  $\bar{x} + 2\sigma$
- (iii) almost all (99.7%) of the values will lie within three standard deviations of the mean, that is,  $\bar{x} 3\sigma$  and  $\bar{x} + 3\sigma$

Based on the Empirical Rule, the probability that a score, selected at random, will be within one standard deviation of the mean is 68% or 0.68. As illustrated below:



#### **Question 27**

Copy and complete these sentences:

- (i) Approximately \_\_\_% of the data lies within one standard \_\_\_\_\_\_ of the mean.
- (ii) Approximately 95% of the data lies within \_\_\_\_\_\_ standard deviations of the \_\_\_\_\_\_.

#### **Question 28**

For each of the following normal curves, find the percentage of all the values that are in the shaded area:



The mean speed of vehicles on a given road can be modelled by a normal distribution with mean 55 km/h and standard deviation 9 km/h. What would be the speed of a vehicle that was travelling at:

- (i) One standard deviation below the mean
- (ii) two standard deviations above the mean
- (iii) three standard deviation above the mean?

## **Margin of Error & Confidence Intervals**

The purpose of sampling is to gain information about the whole population by surveying a small part of the population. This small part is called a sample. If data from a sample is collected in a proper way, then the sample survey can give a fairly accurate indication of the population characteristic that is being studied.

- The notation  $\widehat{p}$  is used to denote sample proportion.
- The notation **p** is used to represent population proportion.

Since p is generally not known,  $\hat{p}$  is used as an estimator for the true population proportion, p.

Sample Surveys are rarely accurate so account for this; we generate the *margin of error*. This number is based on the sample size, the larger the sample size the more accurate the sample is to the population.

$$E = \frac{1}{\sqrt{n}}$$

- E= Margin of Error for the Sample
- *n* = The number of samples

If the sample size is 400, then

$$E = \frac{1}{\sqrt{n}} = E = \frac{1}{\sqrt{400}} = 0.05\%$$

#### **Question 30**

Work out the margin of error for each of the following random samples:

(i) 900 (ii) 1200 (iii) 2025 (iv) 800

Give your answer correct to 2 decimal places where necessary.

#### **Question 31**

In a random sample of 500 households, 80 said that they had at least one pet.

- (i) What is the sample size?
- (ii) What is the margin of error?
- (iii) What is the sample proportion?



In a random sample of 200 students, 48 said that they spend at least one hour each day watching television.

- (i) Write down the sample size.
- (ii) What is the margin of error?
- (iii) What is the sample proportion, p<sup>?</sup>?
- (iv) If you increase the sample size to 400, what effect would this have on the margin of error?

## **Confidence Interval**

The confidence interval of a sample is a combination of the sample survey, the margin of error and how this reflects the population.

The confidence interval gives a range of values for which we are 95% confident that the population will lie within this range.

$$\hat{\mathbf{p}} - \frac{1}{\sqrt{n}}$$

#### **Question 33**

A manufacturer tests a random sample of 300 items and finds that 45 are defective.

- (i) Write down the sample size, n.
- (ii) Calculate the margin of error.
- (iii) Work out the sample proportion, p<sup>^</sup>.
- (iv) Work out a confidence interval for the proportion of defective items produced.



## **Hypothesis Testing**

When making statements mathematician will 'check' these statements against the statistics. This statement is known as the hypothesis, we then test this/these hypothesises to be true or false.

#### How to Carry out a Hypothesis Test

1. Write down H0, the null hypothesis, and H1, the alternative hypothesis

For example, to test whether a coin is biased if we get 7 heads in 10 tosses, we could formulate the following hypotheses: H0: The coin is not biased. H1: The coin is biased.

- 2. Write down or calculate the sample proportion, p<sup>^</sup>.
- 3. Find the margin of error

$$E = \frac{1}{\sqrt{n}}$$

4. Write down the confidence interval for the population proportion

$$\hat{\mathbf{p}} - \frac{1}{\sqrt{n}}$$

- 5. Examine the Value
  - (i) If the value of the population proportion stated is within the confidence interval, accept the null hypothesis H0 and reject H1.
  - (ii) If the value of the population proportion is outside the confidence interval, reject the null hypothesis H0 and accept H1.

#### **Question 34**

The manufacturer of Chummy Bits claims that 80% of dog owners choose this product for their dogs. In a random sample of 200 dog owners, 155 chose Chummy Bits.

- (i) Write down the null hypothesis, H0.
- (ii) Write down the alternative hypothesis, H1.
- (iii) What is the sample size, n?
- (iv) Calculate the margin of error.
- (v) Calculate the sample proportion.
- (vi) Work out the confidence interval.
- (vii) At the 95% confidence level, is the manufacturer's claim correct?

## **Past Exam Questions**

#### **Question 35**

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

	j	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				

														 	_	_

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

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

(.)	Disping	bound b	catter III	a pro ema	. onon	un or you	eareanations	creary.	

(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.

- 7	-															

#### **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:

# (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.

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

Male	Female

#### **Question 39**

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	24500	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)

(€1000)	) - 10	10 - 20	20 - 30	30-40	40 – 50	50 - 60	60 – 70
No. of Employees							

[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.



) (i)	Outline another method which could have been used to calculate the mean salary.
_	
(11)	Which method is more accurate? Eurolain your answer
(11)	which method is more accurate? Explain your answer.
A	iswer
F	ason

#### 1.11 . . . . . . . . . .

## Question 40

The ages of	of the 3	30 peop	le who t	ook part	in an ac	robics c	lass are	as follov	VS:	
1	8	24	32	37	9	13	22	41	51	49
1	5	42	37	58	48	53	27	54	42	24
3	33	48	56	17	61	37	63	45	20	39
The ages o	of the 3	30 peop	le who t	ook part	in a swi	mming	class are	as follo	ws:	
1	6	22	29	7	36	45	12	38	52	13
3	3	41	24	35	51	8	47	22	14	24
4	2	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.

															_	
	_	_											_			

#### **Question 41**

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

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

A political party had claimed that it has the support of 24% of the electorate. Of the voters in (b) 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?



(a) A widget-manufacturing company repeatedly asserts that 80% of traders recommend their brand of widget. In a survey of 40 traders, 24 said that they would recommend the company's widget. Use a hypothesis test at the 5% level of significance to decide whether there is sufficient evidence to reject the company's claim. State clearly the null hypothesis and your conclusion.

		_					_						_	_	_		_		_	_					_		
	 _	_					_							_			_	_	_	_	_		_	_	_		
	_		_											_				_	_	_	_				_	_	
	 _	-	-											_				_	-	_	_			_	_	_	
	-	_	-		_	_		_	_	_	_	_		_	_	_	_	_	-	_	-	_	_	_	_		
	 -	-	-		_			_	 _		_	_		_	_	_	_	_	-	-	+	_		_	-		
	 -	-	+			_								-	_		_	-	+	+	+		_	-	-		
_	+	+	+		_	_			_	_	_	-		-		_	_		+	+	+	-	 _		-	+	
	-	-	+																	-					-		
	_											_		_					_	_					_		
	_		_											_				_	_	_	_	_		_	_		
	 _	-	-					_	 					_				_	-	_	_	_	 	_		_	
	-	-	-	-										_				_	-	_	-			_			
	_		_															_	_	_	_			_	_		

(b) A large group of students has a mean height of 170 cm with a standard deviation of 14 cm. The heights of these students are normally distributed. Use the empirical rule to find a height interval that will contain the heights of approximately 95% of the students.

		_							_		_		_			_	_	
											_						_	

#### **Question 43**

A newspaper report in October 2013 stated that 90% of homeowners who were liable for property tax had registered for it. The total number of properties liable for the tax was estimated at 1.9 million.

(a) (i) Estimate the number of properties that were registered.

(ii) Suggest one reason why some properties were not registered.

(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	€0 -	€100001 -	€150 001 -	€200 001 -	€250 001 -	Over
Band	€100 000	€150000	€200 000	€250 000	€300 000	€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 - €100000	€45	425 790	19160550
€100001 - €150000	€112	489 060	
€150001 - €200000	€157		
€200 001 - €250 000	€202		
€250001 - €300000	€247		
Over €300000	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.



#### (iv) Find the mean estimated tax per property on those properties with a valuation over €300 000.

						-		-						_
				-		-		-	-	-				

(v) Some homeowners may under-value their property in order to pay less tax. For example, one estimate stated that 20% of properties in the €100 001 - €150 000 band should have been valued in the €150 001 - €200 000 band. Based on this estimate, find the amount of extra tax that would be raised if these properties were registered in the correct Valuation Band.



#### **Question 44**

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 produces histograms of the times for the three events. Here are the three histograms.



- (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:

(iii) Brian did not enter this race. Suppose that he had, and suppose that he completed the cycle in 52 minutes and the run in 18 minutes. Explain why this performance would have been very unusual.

		-	-	-				_													
	-	-	-	+	-	_	 _	 _							-		_	 			_
	_	-	-	+	-	_	 	 		-	_						_	 			
		-	-	+	-	_	 	 	 -		_	-	-	-	-	-	_	 		-	_
				_			 		 							 				 	

(c) Máire knows already that the male athletes tend to be slightly faster than the female athletes. She also knows that athletes can get slower as they get older. She thinks that male athletes in their forties might be about the same as female athletes in their thirties. She decides to draw a back-to-back stem-and-leaf diagram of the times of these two groups for the swim. There were 28 females in their thirties, and 32 males in their forties. Here is the diagram:

Female	e, 3	30	-:	39	ye	ars		Ma	ale	, 4	0 -	- 4	9 years
						4	13						•
						0020	14	9					
					1	0	15	1	3	4	5	6	
9	8	8	7	3	2	2	16	3	4	6	7	7	8
			6	4	3	2	17	6	7	7			
						1	18	0	1	3	8	9	
	9	6	3	1	0	0	19	0	0	1	2	3	4
						ACC-16	20	3	9	9			
				3	3	2	21	2	2				
						4	22	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1					
						192.23	23	0					
						8	24	80					
							25						
						5	26						
							27						
							28						
						7	29						
				K	ev		14	9	,	me	an	s 1	4.9 minutes

(i) Describe what differences, if any, there are between the two distributions above.

(ii) Máire drew the diagram because she thought that these two groups would be about the same. Do you think that the diagram would cause Máire to confirm her belief or change it? Give reasons for your answer.

#### **Question 45**

The table below shows the rates of births, marriages and deaths in Ireland from 1990 to 2010. The rates are per 10 000 of the estimated population.

Number of Births, Marriages and Deaths in Ireland (per 10 000 of the estimated population)												
Year	Births	Marriages	Deaths									
1990	151	51	90									
1991	150	49	89									
1992	144	47	87									
1993	138	47	90									
1994	135	46	86									
1995	135	43	90									
1996	140	45	87									
1997	144	43	86									
1998	146	45	85									
1999	144	50	87									
2000	145	51	83									
2001	150	50	79									
2002	155	52	76									
2003	155	51	73									
2004	153	52	71									
2005	148	52	68									
2006	154	52	67									
2007	163	52	64									
2008	168	50	63									
2009	167	48	63									
2010	165	46	61									
		(Source: Central S	tatistics Office, http://www.cso.ie)									

(a) Complete the back to back stem and leaf plot below to show the marriage rate and death rate in Ireland during the period covered in the table above.

Marriage rate	Death rate
	4
	5
	6
	7
	8
	9
	Key:

(b) State one difference that can be observed between the distributions of the marriage rate and the death rate in your plot.



(c) Find the median and interquartile range of the yearly marriage rates in Ireland from 1990 to 2010.

Median	Interquartile range

# (d) (i) Find the mean of the death rate in Ireland from 1990 to 2010. Give your answer correct to one decimal place.



(ii) The standard deviation of the death rates in the table over is 10.3. List all of the death rates that are within 1 standard deviation of the mean.

(e) In 2010, the number of children born in Ireland was 75174. Use this number to estimate the total population of Ireland in 2010.



(f) Use your answer to (e) to estimate the number of people who died in Ireland in 2010.



(g) "More children were born in Ireland in 1990 than in 2000." Give a reason, based on the data, why this statement is not necessarily true.



(h) Find the ratio, Birth rate : Death rate, for the two years 1990 and 2010. Based on your answers for the two years, what would you predict about the population of Ireland in future years. Give a reason for your answer.

1990 Ratio	2010 Ratio
Prediction	
<b>n</b>	

(i) The birth rate and death rate over the 21 years are plotted against each other in the scatter plot below. The correlation coefficient between the two sets of data is -0.85. Describe the relationship between the two sets of data and suggest a reason why this might be the case.





The heights in 2011 of Irish males born in 1992 are normally distributed with mean 178.8 cm and standard deviation 7.9 cm.

(i) Use the *empirical rule* to complete the following sentence:

"95% of ninetee	n-year-old Irish men are between	and	in height."

(ii) Use the empirical rule to make one other statement about the heights of nineteen-yearold Irish men.

The male students in part (b) are a *sample*. The males in part (c) are a *population*. Is this is a suitable population to compare the sample to? Give a reason for your answer.

Answer:	_																				
Reason:	eason:																				
			-		_																

Would you say that the males in the class are taller, smaller, or about the same as the population? Use the data to justify your answer.

Answer:

#### Justification:



An economics student wants to find out whether the length of time people spend in education affects how much they earn. The student carries out a small study. She asks twelve adults to state their annual income and the number of years they spent in full-time education. The data are given in the table below, and a partially completed scatter plot is given.

	Years of education	Income /€1,000	70						
Γ	11	28		-					
	12	30	0 60						
	13	35	3100	-		-			
	13	43	ອ50						
	14	55	COL						
	15	38	. <u>ii</u> 40			·			
	16	45	nual			'	••		
	16	38	UV 30						
	17	55							
	17	60							
	17	30	20	10	12	14	16	18	20
	19	58			Ve	ars of e	ducatio	'n	

(i) The last three rows of data have not been included on the scatter plot. Insert them now.

(ii) What can you conclude from the scatter plot?

(iii) The student collected the data using a telephone survey. Numbers were randomly chosen from the Dublin area telephone directory. The calls were made in the evenings, between 7 and 9 pm. If there was no answer, or if the person who answered did not agree to participate, then another number was chosen at random.

Give one possible problem that might make the results of the investigation unreliable. State clearly why the issue you mention could cause a problem.

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

			New	v private ca	rs sales			
		Number	of cars sold	ł		Engine	e type of ca	rs sold
Vear	January to	April to	July to	October	Annual	Detrol	Discol	Othor
rear	March	June	Sept.	to Dec.	Total	Petroi	Diesei	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	9049	3 018	54 432	22 802	30 645	985
2010	34 555	26 806	17011	6 535	84 907	27 124	53 998	3 785
2011	39 484	29 770	13 467	4 2 1 1	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.

#### (iii) Calculate the percentage increase in annual car sales between 2009 and 2011.

(iv) Aoife says that this increase shows car sales are currently going well. Paul says that car sales are currently going badly. He says that sales have fallen by 52% since 2007 and that they are well below average.

Complete the sentences below to give a criticism of each argument.

Aoife's argumen	nt does not re	cognise that		
Paul's argument	t does not rec	ognise that		

(v) Give a more balanced description of the pattern of car sales over the six years.

_		_			_	_	_	_	_	_	_	_	 		_

(ii) Suggest a reason for this pattern of sales.

(iii) The sales for the first quarter of 2012 are 36 081.Find, with justification, an estimate for the total annual sales for 2012.

ſ																

(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.



- (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 CO<sub>2</sub> 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.

			 			- 1			 		

(ii) Does the information suggest that diesel engines produce lower CO<sub>2</sub> 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 CO<sub>2</sub> emissions of diesel engines than petrol engines? In your answer you should refer to the stem-andleaf plot and an appropriate measure of variability.

