

**PART 1 DUE BY JULY 31<sup>ST</sup>!!**

**PART 2 DUE FIRST DAY OF CLASS!!**

Dear IB Math HL Student,

Welcome back to Year 2 IB Math HL! The focus this year will be mostly on Calculus. Others topics include Vectors, Probability, Statistics, Exponential and Logarithmic Functions. With the main branch of our content in Calculus we will need to take all the mathematical skills we already have to a deeper level.

To help prepare us for Calculus, there are certain algebraic skills that we will have to have under our belts to ensure understanding of the extensions within the Calculus content. Part 1 of your summer packet will review some of those foundational skills as well as review some content we learned in Year 1. Please resource all of your Year 1 guided notes to help you through this portion of the summer packet. Part 2 focuses on new content presented in Chapter 5 of your textbook. It will be presented with guided notes along with videoed lessons. Each section will include a practice problem set to be completed after the guided notes and videos are done.

**Please print both parts of the summer packet and complete all work and notes on the printed copies. Problems set in Chapter 5 should be done on a separate piece of paper.**

**SUBMIT PART 1 BY JULY 31<sup>ST</sup> ON GOOGLE CLASSROOM AND BRING PART 2 TO THE FIRST DAY OF CLASS!!!**

**BOTH PARTS OF THE SUMMER PACKET WILL BE GRADED!**

**AN ASSESSMENT ON CHAPTER 5 WILL TAKE PLACE WITHIN THE FIRST WEEK OF SCHOOL!!**

**HAVE A GREAT SUMMER!!!!**

**Mrs Steere**

Part 1: Due July 31st

Name: \_\_\_\_\_

1. Consider the sequence 6, 17, 28, 39, 50,....

a. Find the formula for the nth term:

b. Find the 50th term:

c. Is 325 a term in the sequence?

d. Is 761 a term?

e. Find the sum of the first 20 terms.

2. Consider the sequence 12, -6, 3, -1.5, ....

a. Find the formula for the nth term:

b. Hence, find the 13th term:

c. Find the sum of the first 10 terms.

d. Find the infinite sum of the sequence.

3. Find the sum

$$\sum_{r=1}^4 (3r - 5)$$

4. Find the sum (hint: use a summation formula)

$$\sum_{i=1}^{15} 50(8)^{i-1}$$

5. Simplify without using a calculator. Leave exponents positive.

a.  $(-3a^7)^2$

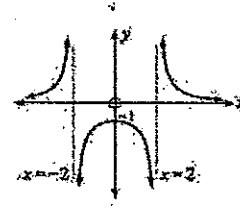
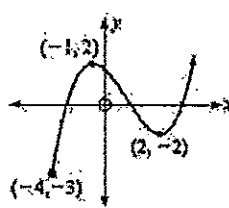
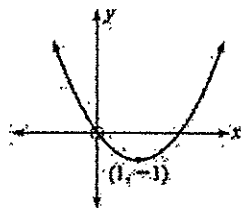
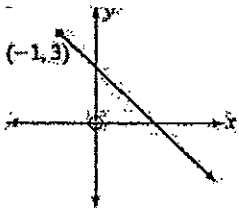
b.  $\frac{8ab^5}{2a^2b^4}$

c.  $\frac{2^{x+1}}{2^{1-x}}$

d.  $27^{\frac{4}{3}}$

e.  $(\frac{2a^{-1}}{b^2})^3$

6. For each of the following graphs find the domain and range:



7. For each find the domain without a calculator.

a.  $f(x) = \sqrt{4-x}$

b.  $f(x) = 5x - 3x^{\frac{1}{2}}$

c.  $f(x) = \frac{x^2-4}{x-2}$

8. Given  $f(x) = x^2 + 1$  and  $g(x) = 3 - x$  find in the simplest form

a.  $(f \circ g)(x) =$

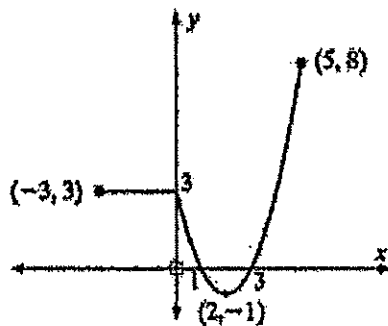
b.  $(g \circ f)(x) =$

c.  $(f \circ g)(4) =$

9. For the graph of  $y = f(x)$ , sketch graphs of:

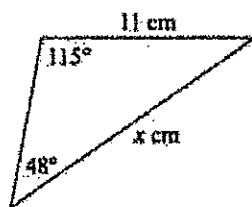
a.  $y = f(-x)$       b.  $y = -f(x)$

c.  $y = f(x+2)$     d.  $y = f(x) + 2$

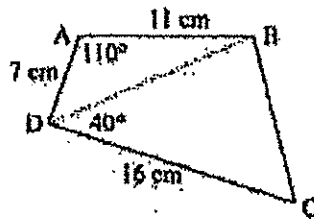


10. Given the point (3,6) on a function  $f(x)$ . If  $g(x) = 2f(x-3) - 5$ , find the point (3,6) in  $g(x)$

11. Find  $x$



12. Find the area of quadrilateral ABCD:



13. A sector has an angle of  $68.2^\circ$  and an area of  $20.8 \text{ cm}^2$ . Find: (perimeter = arc length + radius + radius)  
 a its radius                      b its perimeter.



14. Evaluate *without* a calculator (extremely important concept)

- a)  $\sin 60^\circ$     b)  $\cos 45^\circ$     c)  $\tan 150^\circ$     d)  $\sin 300^\circ$     e)  $\cos 210^\circ$     f)  $\tan 315^\circ$   
 g)  $\sin \pi/6$     h)  $\cos \pi$     i)  $\tan 5\pi/6$     j)  $\sin 3\pi/2$     k)  $\cos 5\pi/4$     l)  $\tan 11\pi/6$

15. Without using a calculator, evaluate:

- a)  $\sin 30^\circ \cos 60^\circ$     b)  $\cos^2(\frac{\pi}{4}) - \sin(\frac{7\pi}{6})$     c)  $1 - \cos^2(\frac{\pi}{6})$

16. Without using a calculator, find:

- a)  $\cos \theta$  if  $\sin \theta = -\frac{1}{\sqrt{2}}$ ,  $\pi < \theta < \frac{3\pi}{2}$ .                      b)  $\sin \theta$  if  $\cos \theta = \frac{2}{3}$ ,  $0 < \theta < \frac{\pi}{2}$ .

c) Find  $\sin x$  and  $\cos x$  given that:  $\tan x = -\frac{4}{3}$  and  $\frac{\pi}{2} < x < \pi$

17. Solve *without* a calculator: Find  $\theta$  in radians if  $0 \leq \theta \leq 2\pi$  and:

a)  $\cos \theta = \frac{1}{2}$       b)  $\sin \theta = 1$       c)  $\cos \theta = -1$       d)  $\sin 2\theta = \frac{1}{2}$

18. Use technology to solve for  $0 < x < 8$ ; give answers to 3 significant figures

a)  $\sin 2x = 0.162$       b)  $\cos(x - 1.3) = -0.609$

19. Solve by factoring. Find  $x$  in radians if  $0 \leq x \leq 2\pi$

a)  $\sin^2 x + \sin x - 2 = 0$       b)  $2\sin^2 x + 1 = 3\sin x$

20. If  $\sin A = \frac{4}{5}$  and  $\cos A = \frac{3}{5}$  find the values of:

a)  $\sin 2A$       b)  $\cos 2A$

21. If  $\cos \beta = \frac{2}{5}$  where  $\frac{3\pi}{2} < \beta < 2\pi$ , find the value of  $\sin \beta$  and hence the value of  $\sin 2\beta$ .

22. Without using technology draw the graphs of the following for  $0 \leq x \leq 2\pi$ :

i)  $y = 3\sin x$        $y = \cos 2x$        $y = -\frac{3}{2}\sin x$

$y = \sin x + 2$

$y = \sin(x + \frac{\pi}{4})$



Please watch the following video lessons on youtube as you complete the chapter 5 guided notes. Also use your textbook to help complete the guided notes on Chapter 5. All PROBLEM SETS need to be completed on a SEPARATE PIECE OF PAPER THAT IS CLEARLY MARKED FOR EACH SECTION.

<https://youtube.com/playlist?list=PLmrVBibuhI6m1I1sM4anomSMXC2ooCJb5>

## Chapter 5 (Statistics and probability)

Each section has two parts: guided notes to help you through the section, and a short problem set.

### 5.1 Sampling

#### Notes

#### Data

- Data can be either *qualitative* or *quantitative*.
  - Qualitative data: \_\_\_\_\_
  - Ex: \_\_\_\_\_
  - Quantitative data: \_\_\_\_\_
  - Ex: \_\_\_\_\_
- Quantitative data can be *discrete* or *continuous*.
  - Discrete: \_\_\_\_\_
  - Ex: \_\_\_\_\_
  - Continuous: \_\_\_\_\_
  - Ex: \_\_\_\_\_
- The trick:
  - Discrete → \_\_\_\_\_ → \_\_\_\_\_
  - Continuous → \_\_\_\_\_ → \_\_\_\_\_

#### Sampling

- Definitions: The *population* consists of \_\_\_\_\_. A \_\_\_\_\_ is a subset of the population used to draw conclusions about the population.
- Five considerations when developing a sample:

	population from which the sample is taken
Sampling frame	
	a single member of the sampling frame chosen to be sampled
Sampling variable	
Sampling values	

## Sampling techniques

Simple random sample (SRS)	
	Members of the population are listed. Participants are chosen based on a random starting point and fixed interval.
Stratified sampling	
Quota sampling	
	You select members of the population that are most readily available or easily accessible.

## Bias and reliability

- **Definition:** *Bias* refers to \_\_\_\_\_
- The purpose of sampling is to create a faithful and accurate representation of the population. So, when and how is this distorted?



– 2.

– 3.

– 4.

- **Definition:** Data is \_\_\_\_\_ if you can repeat the collection process and produce similar results.
- **Definition:** Data is \_\_\_\_\_ if there is enough data available to support your conclusions.
- Two factors can lead to unreliable data:
  - Missing data: \_\_\_\_\_ and \_\_\_\_\_
  - \_\_\_\_\_: Typos and outside influences on the participant

### Problem Set

1. For each scenario, define **i.** the target population, **ii.** the sampling frame, **iii.** the sampling unit, **iv.** the sampling variable, and **v.** the sampling value.
  - (a) The weight of ball bearings manufactured by a company
  - (b) The volume, to the nearest milliliter, that a soft drink factory fills its 1 liter bottles of soda
2. Describe how you could choose a systematic sample of 40 books from a library of 2,000 books. Identify any bias that may be present.
3. A non-governmental organization (NGO) would like to take a sample from its various world-wide bases to give some information about its results globally. Below is a list of the number of bases the NGO has in each continent.

Europe	57
Africa	35
Antarctica	2
Asia	35
Oceania	57
North America	85
South America	35

- (a) Describe how you could use this information to conduct a stratified sample.
  - (b) Identify any bias that may be present.
4. A company produced 1,000 batteries. The manufacturer claims they have a life of 4,000 years. Explain why testing the population would not be possible. Suggest a sampling technique which may be beneficial to help test these batteries.

## 5.2 Descriptive statistics

### Notes

#### Measures of central tendency

- This should hopefully be prior knowledge for you.
- Measures of central tendency describe what is going on \_\_\_\_\_.
- Mean ( $\bar{x}$  if \_\_\_\_\_,  $\mu$  if \_\_\_\_\_)
  - Define mean as you remember it:

– Let's update our knowledge of mean:  $\bar{x}$  or  $\mu = \frac{\sum_{i=1}^k f_i x_i}{n}$ , where  $\sum_{i=1}^k f_i x_i$  is \_\_\_\_\_, and  $n$  is \_\_\_\_\_. More specifically,  $f_i$  is the \_\_\_\_\_,  $x_i$ .

– Practically speaking, you DO NOT need to know the scary-looking formula! But if you do see it on an IB exam, you know what it means.

– For data grouped by intervals, take the \_\_\_\_\_ as each  $x_i$ .

- Median

– Define median as you remember it:

– If  $n$  is odd, the median occurs at the \_\_\_\_\_th term.

– If  $n$  is even, the median is the \_\_\_\_\_.

- Mode

– Define mode as you remember it:

## Measures of dispersion

- Some of this will be prior knowledge, some will not.
- Range = \_\_\_\_\_
- *Quartiles* separate the data set into \_\_\_\_\_ sections.
  - $Q_1$  has \_\_\_\_\_% of the data below it. It's the \_\_\_\_\_ of the lower half of the data set.
  - $Q_2$  is the \_\_\_\_\_ of the data set.
  - $Q_3$  has \_\_\_\_\_% of the data below it. It's the \_\_\_\_\_ of the upper half of the data set.
  - $Q_4$  is the \_\_\_\_\_ of the data set.
  - *Interquartile range* (\_\_\_\_\_) = \_\_\_\_\_
  - A major advantage of the IQR is \_\_\_\_\_  
\_\_\_\_\_. We call this a \_\_\_\_\_ measure.

## Histograms

- Histograms are best for \_\_\_\_\_ data. For \_\_\_\_\_ data, a simple bar graph is enough.
- Differences between a histogram and a bar graph:
  - Our textbook suggests having \_\_\_\_\_ interval classes for a histogram.
  - For discrete data, use a \_\_\_\_\_. If that discrete data is grouped into interval classes (i.e., the range 1, 2, 3 is represented by the interval 1–3), bars begin \_\_\_\_\_ above and below the interval class. For example, the bar for 1–3 would be drawn from \_\_\_\_\_.
  - When asked to comment on the "distribution of the data", consider four things:
    - 1.

– 2.

– 3.

– 4.

- You can compare data sets of different sizes by using a \_\_\_\_\_.
  - The *relative frequency* of a given class interval is given as a \_\_\_\_\_.
  - The relative frequency is given by \_\_\_\_\_, where  $f$  is \_\_\_\_\_ and  $n$  is \_\_\_\_\_.
- Unequal class widths
  - Using \_\_\_\_\_ where the data is more spread out and \_\_\_\_\_ where the data is more concentrated can help more accurately pinpoint central tendencies but not \_\_\_\_\_.

### Problem Set

1. The number of people traveling in each of 33 cars was counted, and the results are shown in the table below.

Number of people	1	2	3	4	5	6
Frequency	8	11	6	4	2	2

- (a) Find the limits of each category (meaning the bounds of the interval that will go onto the histogram).
- (b) Complete the following frequency table.

Number of people	Frequency	Interval on histogram
1	8	$0.5 < x \leq 1.5$
2		
3		
4		
5		
6		

- (c) Draw a frequency histogram to represent the data.  
Note: You may find the hint on page 323 of our textbook helpful for this problem.

2. The data given has been produced from the masses of 50 koalas, in kilograms, in an Australian nature reserve.

33 19 24 35 36 24 29 29 29 34  
 38 35 35 35 36 60 35 50 34 48  
 41 41 51 42 35 36 32 61 30 40  
 41 19 33 34 17 35 35 38 35 42  
 20 29 50 33 37 28 49 58 45 40

Construct a frequency histogram of this distribution. Describe the distribution, commenting on the shape, center, and spread.

3. You are given the following frequency tables on the length of time males and females spent on their mobile phones during a period of one day.

Time in minutes, male	Frequency	Time in minutes, female	Frequency
$0 \leq x < 15$	5	$0 \leq x < 15$	4
$15 \leq x < 30$	8	$15 \leq x < 30$	5
$30 \leq x < 45$	10	$30 \leq x < 45$	4
$45 \leq x < 60$	5	$45 \leq x < 60$	14
$60 \leq x < 75$	2	$60 \leq x < 75$	2

- (a) Explain why it is necessary to use a relative frequency histogram to compare the data in this context.
- (b) By adding an additional column to the table, write down the relative frequencies.
- (c) Draw relative frequency histograms for the two sets of data.
- (d) Analyze each distribution.
- (e) Compare the length of time per day spend on the phone by the male and female subjects.
4. A survey of 90 mothers was taken in New Zealand to inquire about their age when giving birth to their first child.

Age $A$ in years	Frequency
$15 < A \leq 20$	5
$20 < A \leq 23$	15
$23 < A \leq 25$	20
$25 < A \leq 30$	20
$30 < A \leq 40$	30

- (a) Determine which type of histogram would give the best representation.
- (b) Construct the histogram.
- (c) Calculate the mean and determine the modal class.

### 5.3 The justification of statistical techniques

#### Notes

#### Box-and-whisker diagrams (boxplots)

- A box-and-whisker plot is used to visualize a summary of the data set using five numbers. The “five-number summary” is made up of \_\_\_\_\_.
- General diagram of a boxplot:

- Recall that  $x$  is an outlier if \_\_\_\_\_ or \_\_\_\_\_.

#### Variance and standard deviation

- Variance and standard deviation both use all data values in a set to \_\_\_\_\_. Standard deviation is the more useful of the two.
- **Definition:** The \_\_\_\_\_, denoted by \_\_\_\_\_ for a population and \_\_\_\_\_ for a sample, gives a mean average of the distance between each data point and the mean.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x - \mu)^2}{n}}$$

- **Definition:** \_\_\_\_\_ is equal to  $\sigma^2$ .
- The \_\_\_\_\_ the data are to each other, the \_\_\_\_\_ the value of  $\sigma$ .
- Fortunately, IB expects you to use a GDC for this, so we will not cover it by hand!

#### Moving from sample to population

- We assume that our sample is an accurate representation of the population, and therefore is an \_\_\_\_\_. So, our sample mean  $\bar{x}$  is an \_\_\_\_\_ for the population mean  $\mu$ .

- It's a little trickier for standard deviation.
  - We know that the “true” population standard deviation is divided by  $n$  under the radical.
  - However, if you want to *estimate* the population standard deviation, the denominator changes to \_\_\_\_\_.
  - Why do we do this?

– On the TI-83/84 models,  $Sx$  represents the \_\_\_\_\_, and  $\sigma x$  represents the \_\_\_\_\_.

### Cumulative frequency

- With raw data, it's easy to find the median: just \_\_\_\_\_.  
With data grouped into intervals, however, it's harder to determine where the median or a quartile lies exactly.
- The cumulative frequency curve is called an \_\_\_\_\_.
- How to draw a cumulative frequency curve:
  - The \_\_\_\_\_ should be placed along the  $y$ -axis, and the \_\_\_\_\_ along the  $x$ -axis.
  - The initial point of the curve is  $(x, 0)$ , where  $x$  is the \_\_\_\_\_.
  - Each cumulative frequency value is plotted using the \_\_\_\_\_ of each interval.
  - Connect all points using a \_\_\_\_\_.

### Problem Set

1. Consider the data from our discussion of the estimated population standard deviation. A sample of the weights of 25 apples, in grams, is shown below.

132	122	132	125	134
129	130	131	133	129
126	132	133	133	131
133	138	135	135	134
142	140	136	132	135

- (a) Write down the minimum and four quartiles of the data set.
- (b) Hence, construct a boxplot of the data set.

(c) Calculate the IQR of the data set. Are there any outliers?

2. In a biscuit factory a sample of 10 packets of biscuits were weighed. The data is given below in grams.

196, 197, 199, 200, 200, 200, 202, 203, 203, 205

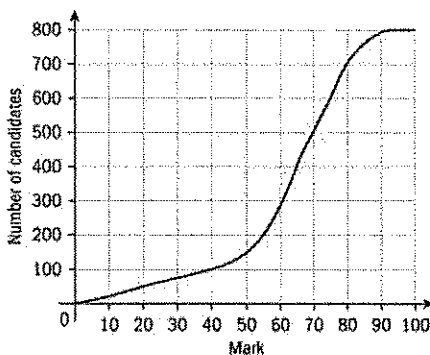
Calculate the mean and standard deviation of this data.

3. 25 rabbits were born in one week on a farm. Their weights, in grams, are recorded below.

450	453	452	480	501
462	475	460	470	430
485	435	425	465	456
475	435	466	482	455
462	435	462	478	455

From this information, make a prediction about the mean and standard deviations of the weights for the whole year.

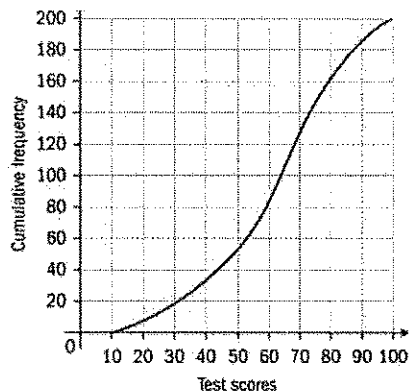
4. A quiz marked out of 100 is written by 800 students. The cumulative frequency graph for the marks is given below.



(a) Estimate the number of students who scored 25 marks or less on the quiz.

(b) The middle 50% of the quiz results lie between the marks  $a$  and  $b$ , where  $a < b$ . Find the values of  $a$  and  $b$ .

5. 200 vehicles are tested for their air pollution efficiency. The results are given in the cumulative frequency graph below.





- (a) Estimate the median test score.
- (b) The top 10% of vehicles receive a lower insurance premium price A and the next best 20% of the vehicles receive price B. Estimate the minimum scores required to obtain both prices.

## 5.4 Correlation, causation, and linear regression

### Notes

- **Definition:** \_\_\_\_\_ shows a relationship between two variables. they can be expressed as an ordered pair \_\_\_\_\_.
- **Definition:** A \_\_\_\_\_ (our book calls it a \_\_\_\_\_) is a graph composed of \_\_\_\_\_ to show the relationship between two quantitative variables from an individual in the data set.
- We should consider three things when looking at a scatter plot:
  - 1.
  - 2.
  - 3.
- We can use \_\_\_\_\_ to model the trend shown by a scatter plot. However, this should only be done when there is clearly a \_\_\_\_\_ relationship present, meaning that one variable clearly impacts the value of the other.
  - If this type of relationship is present, then we call  $x$  the \_\_\_\_\_ variable and  $y$  the \_\_\_\_\_ variable.
- The \_\_\_\_\_ gives us a quantitative measure of the strength of the trend present in the scatter plot. You may see it called the \_\_\_\_\_ (PMCC) or simply \_\_\_\_\_.
- Use this space to draw the chart showing the different strengths and directions of  $r$ .

### Making a scatter plot on your GDC

- Before we begin, make sure the GDC's capability to find the PMCC ( $r$ ) is turned on. On the TI-83/84 models, press: (write in the steps)
  
- Enter your data: (write in the steps)
  
- Calculate the equation of the linear regression line and the correlation coefficient: (write in the steps)
  
- OPTIONAL: to view your scatterplot... (write in the steps)

### Making predictions using a scatter plot

- **Definitions:** Suppose the  $x$ -values in a set of bivariate data range from  $a$  to  $b$ . Let  $p$  be the  $x$ -value for which you are trying to predict a  $y$ -value.
  - If  $p \in [a, b]$ , then this is called \_\_\_\_\_.
  - Otherwise ( $p \notin [a, b]$ ) this is called \_\_\_\_\_.
- The accuracy of interpolation depends on...

- The accuracy of extrapolation depends on this too, but also...

### Correlation fallacies

Correlation vs. causation	
Correlation is only linear	
Lurking variables	
Artificial mathematical relationships	
Use of separate populations	
Poor sampling	

### Problem Set

1. Complete Exercise 5I #1abd in your textbook. (I didn't want to reproduce the graphs.)
2. The following table gives the heights and weights of 14 race horses.

Height (m), $X$	1.48	1.51	1.23	1.57	1.29	1.30	1.37	1.17	1.20	1.34	1.42	1.42	1.37	1.44
Weight (kg), $Y$	329	314	185	356	228	230	257	171	185	214	315	271	242	285

- (a) Use technology to draw a scatter diagram.
  - (b) Calculate the line of best fit.
  - (c) Comment on the correlation coefficient in the context of the question.
  - (d) Use the equation generated by your calculator to predict the weight of a race horse with height of 1.38 m.
3. Ten pairs of twins take an intelligence test. For each pair of twins, one is female and the other is male. The bivariate data obtained is given in the table below.

Female	100	110	95	90	103	120	37	105	89	111
Male	98	107	95	89	100	112	99	101	89	109

- (a) Find the Pearson product moment correlation coefficient,  $r$ .
- (b) State, in two words, a description for this linear correlation.
- (c) Letting the male score be represented by  $x$  and the female score by  $y$ , find the equation of the
  - i.  $y$  on  $x$  line of best fit
  - ii.  $x$  on  $y$  line of best fit

Hint: " $y$  on  $x$ " means that the linear regression equation is in the form  $y = ax + b$ . The same logic is true for " $x$  on  $y$ ". On your GDC, simply pay attention to whether  $L_1$  or  $L_2$  is entered for "XList" and "YList"; this way, you avoid having to re-enter everything.

- (d) Another pair of female/male twins are discovered ("discovered"...as if this is a place where female/male twins don't exist?) and the male twin scored a 105 on the test. The female twin was too sick to take it. Estimate the score that she would have obtained, giving your answer to the nearest integer.
- (e) Yet another pair of female/male twins are "discovered". The female twin scored a 95 on the test but the male twin refused to take it. Estimate the score that he would have obtained, giving your answer to the nearest integer.
- (f) If, for a further pair of male/female twins, the male scored a 140 on test, explain why it would be unreliable to use a line of best fit to estimate the female's score.