IB Math HL, Year 2: Summer Packet

Due on the first day of class

Welcome back to IB Math HL!

As you can imagine, this class is a continuation of the content you learned in Year 1, with a greater emphasis placed on preparing you for the IB exam this coming May. You have already begun to develop the deep quantitative reasoning and problem solving skills necessary to excel in this course–take a step back for a minute and think about just how far you have come as a math student in just one year! I hope this perspective helps build your confidence when tackling Year 2.

The assignment has two parts. The first part is designed to be a review of content you learned in Year 1. Some of these problems, where possible, are designed to reflect the IB Exam. **Do not throw away your notes from Year 1! Use them to help you.** The other part of this assignment focuses on new content; specifically the content in Chapter 5 of your textbook, which covers the basics of statistics. It's relatively straightforward. This section is a combination of guided notes and practice problems. Videos have been created and posted to my YouTube channel covering the content from Chapter 5, just as we've done all year. You are encouraged to use them to help you through the content. As we know, the book isn't the most helpful, and this chapter is no different.

Before the assignment begins here, you'll find an extensive list of IB notations and commands, and formulas that we learned in Year 1. These formulas come from your formula sheet you received at the beginning of Year 1 but I reprinted them for your convenience. If you need resources beyond that, you are welcome to use any online resource at your disposal. You must show all work to receive full credit, and work should be completed neatly and thoroughly. In the interest of saving some paper, I didn't provide a lot of room to complete these problems, so **please work on separate paper.** I indicated in each problem whether or not you should use your GDC to complete it.

As of right now, I'm supposed to teach this class in the fall, and **I plan to count this assignment as a series of homework grades. There will also be a test on Chapter 5 within the first few days of class.** However, things often change over the summer; no matter who your teacher is in the fall, assume that this assignment will count as a grade and that the content will be assessed in some way.

If you have any questions at all over the summer, please reach out to me! My email is **rcox@theproutschool.org**, and I will leave our Remind from Year 1 open. I would be happy to meet with you on Zoom for help on an as-needed basis. Have a great summer!

- Mr. Cox

Note: Problems start on page 6, Chapter 5 starts on page 10. This looks daunting, but the packet is mostly formula sheets and notes to help you.

Some helpful resources: (These are all hyperlinks- click to follow the link)

- Khan Academy for online tutorials
- Printable Paper for free graph paper printouts
- My YouTube channel with the content videos. A playlist of the Chapter 5 content will be created soon to help you find the right videos.

A review of IB notations and commands

Notation

IB states that they will accept most conventional forms of notation. However, you may see these used on official IB papers and prompts.

Absolute value	$\|x\|$	IB will refer to this as *modulus*
Angles		We typically write angle A as $\angle A$. IB will use the notation $\hat{A}$, or $B\hat{A}C$ if the angle can be defined by three vertices.
Graphing calculator		IB will refer to this as a *GDC* (graphic display calculator). The TI-83 Plus/TI-84 Plus, as well as similar Casio models, are recommended. The TI-Nspire is allowed but must be switched to test mode for IB exams because of the computer algebra system (CAS) installed.
Interval notation	$]a,b[$ $[a,b]$	Refers to the *open* interval $a < x < b$, also denoted (a,b). Refers to the *closed* interval $a \leq x \leq b$.
Line segments		Line segments $\overline{AB}$ will be written as $[AB]$
Number sets	$\mathbb{N}$	The set of positive integers and zero (natural numbers), $\{0,1,2,3,\dots\}$
	$\mathbb{Z}$	The set of integers, $\{0,\pm1,\pm2,\pm3,\dots\}$
	$\mathbb{Z}^+$	The set of positive integers, $\{1,2,3,\dots\}$
	$\mathbb{Q}$	The set of rational numbers, any number that can be written as a fraction in simplest form
	$\mathbb{Q}^+$	The set of positive rational numbers, $\{x\|x \in \mathbb{Q}, x > 0\}$
	$\mathbb{R}$	The set of real numbers
	$\mathbb{R}^+$	The set of positive real numbers, $\{x\|x \in \mathbb{R}, x > 0\}$
	$\mathbb{C}$	The set of complex numbers of the form $a+bi$, where $a,b \in \mathbb{R}$.
Repeating decimals		Standard notations: $0.\overline{3} = 0.3333\dots$, $0.\overline{123} = 0.123123\dots$ IB notation: $0.\dot{3}$, $0.\dot{1}2\dot{3}$
Set notation	$\in$	An element is a member of a given set. Ex: $x \in \mathbb{R}$
	$\subseteq$	A set is contained within, or equal to, another set. Ex: $\mathbb{Z} \subseteq \mathbb{R}$
	$\subset$	A set is contained within, but not equal to, another set. This notation is less common. Ex: $\mathbb{Z} \subset \mathbb{R}$
	$\cup$	The union of two sets. Ex: $A \cup B$
	$\cap$	The intersection of two sets. Ex: $A \cap B$
Slope		IB will refer to this as the *gradient*

Command Terms

Calculate	Obtain a numerical answer showing the relevant stages in that working.
Determine	Obtain the only possible answer.
Draw	Represent by means of a labeled, accurate diagram or graph, using a pencil. A ruler should be used for straight lines. Diagrams should be draw to scale. Graphs should have points correctly plotted and joined in a straight line or curve.
Find	Obtain an answer, showing relevant stages in that working.
Hence	Use the preceding work to obtain the required result.
Hence or otherwise	It is suggested that the preceding work is used, but other methods could also receive credit.
Show that	Obtain the required result (possible using the information given) without the formality of proof. These questions do not generally require the use of a calculator.
Sketch	Represent by means of a diagram or graph (labelled as appropriate). The sketch should give a general idea of the required shape or relationship, and should include relevant features.
Solve	Obtain the answer(s) using algebraic and/or numerical and/or graphical methods.
Write down	Obtain the answer(s), usually by extracting information. Little to no calculation is required. Working does not need to be shown.

Formulas and equations from Year 1

Topic 1: Number and algebra

nth term, arithmetic seq.	$u_n = u_1 + (n-1)d$
Sum, arithmetic seq.	$S_n = \frac{n}{2}(u_1 + u_n)$
nth term, geometric seq.	$u_n = u_1 r^{n-1}$
Sum, finite geometric seq.	$S_n = \frac{u_1(1-r^n)}{1-r}, r \neq 1$
Sum, infinite geometric seq.	$S_n = \frac{u_1}{1-r}, \lvert r \rvert < 1$
Combinations	$\binom{n}{r} = {}^nC_r = \frac{n!}{r!(n-r)!}$
Permutations	${}^nP_r = \frac{n!}{(n-r)!}$
Binomial Theorem (BT)	$(a+b)^n = \sum_{r=0}^{n} {}^nC_r \cdot a^{n-r}b^r$
BT for negative exponents	$(1-x)^{-n} = 1 + nx + \frac{n(n+1)}{2!}x^2 + \cdots + \frac{n(n+1)(n+2)\cdots(n+r-1)}{r!}x^r \ldots,$ $n \in \mathbb{Z}^+,\ -1 < x < 1$
BT for rational exponents	$(1+x)^{\alpha} = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots + \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-r+1)}{r!}x^r \ldots,$ $\alpha \in \mathbb{Q},\ -1 < x < 1$

Complex numbers	$z = a + bi$
Modulus of a complex number	$\lvert z \rvert = \sqrt{a^2 + b^2}$
Conjugate of a complex number	$z^* = a - bi$

Topic 2: Functions

Quadratic formula	$x = \dfrac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
Discriminant	$\Delta = b^2 - 4ac$
Sum of roots x of a polynomial equation	$\sum_{i=0}^{n} x_i = \dfrac{-a_{n-1}}{a_n}$
Product of roots x of a polynomial equation	$\prod_{i=0}^{n} x_i = \dfrac{(-1)^n a_0}{a_n}$

Topic 3: Geometry and trigonometry

Distance between 2 points $(x_1, y_1, z_1), (x_2, y_2, z_2)$	$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$
Midpoint between 2 points $(x_1, y_1, z_1), (x_2, y_2, z_2)$	$\left(\dfrac{x_1 + x_2}{2}, \dfrac{y_1 + y_2}{2}, \dfrac{z_1 + z_2}{2}\right)$
Volume, right pyramid	$V = \dfrac{1}{3}Ah$, where A is base area, h is height
Volume, right cone	$V = \dfrac{1}{3}\pi r^2 h$
Volume, sphere	$V = \dfrac{4}{3}\pi r^3$
Surface area, cone	$S = \pi r^2 + \pi r l$, where l is the slant height
Surface area, sphere	$S = 4\pi r^2$
Law of sines	$\dfrac{\sin A}{a} = \dfrac{\sin B}{b} = \dfrac{\sin C}{c}$ (reciprocals also valid)
Law of cosines	$c^2 = a^2 + b^2 - 2ab\cos C$ (switching variables also valid)
Area, triangle	$\text{Area} = \dfrac{1}{2}ab\sin C$ (switching variables also valid)
Arc length	$l = r\theta$, where θ is in radians
Sector area	$A = \dfrac{1}{2}r^2\theta$, where θ is in radians
Quotient identities	$\tan\theta = \dfrac{\sin\theta}{\cos\theta}$, $\cot\theta = \dfrac{\cos\theta}{\sin\theta}$
Reciprocal identites	$\csc\theta = \dfrac{1}{\sin\theta}$, $\sec\theta = \dfrac{1}{\cos\theta}$, $\cot\theta = \dfrac{1}{\tan\theta}$

Pythagorean identities	$\sin^2\theta + \cos^2\theta = 1$
	$\tan^2\theta + 1 = \sec^2\theta$
	$1 + \cot^2\theta = \csc^2\theta$
Sum and difference formulas	$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$
	$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$
	$\tan(\alpha \pm \beta) = \dfrac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$
Double-angle formulas	$\sin(2\theta) = 2\sin\theta\cos\theta$
	$\cos(2\theta) = \cos^2\theta - \sin^2\theta \quad = 1 - 2\sin^2\theta \quad = 2\cos^2\theta - 1$
	$\tan(2\theta) = \dfrac{2\tan\theta}{1 - \tan^2\theta}$

Topic 4: Statistics and probability

You'll cover this topic in this packet! So, no formulas here.

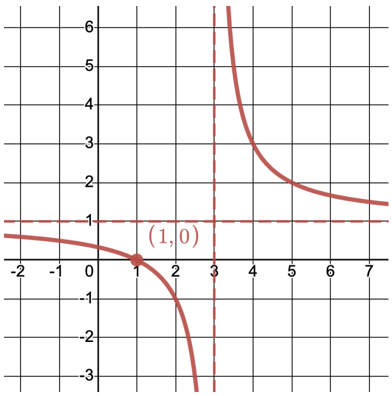
Topic 5: Calculus

Derivative of $f(x)$ from first principles	$f'(x) = \lim_{h\to 0} \dfrac{f(x+h) - f(x)}{h}$	
Power rule	$\dfrac{d}{dx}(x^n) = nx^{n-1}$	
Sum and difference rule	$\dfrac{d}{dx}[f(x) \pm g(x)] = f'(x) \pm g'(x)$	
Product rule	$\dfrac{d}{dx}(uv) = u'v + v'u$	
Quotient rule	$\dfrac{d}{dx}\left(\dfrac{u}{v}\right) = \dfrac{vu' - uv'}{v^2}$	
Kinematics	$v(t) = s'(t),\ a(t) = v'(t) = s''(t)$	
Derivatives of trig functions	$\dfrac{d}{dx}\sin x = \cos x$	$\dfrac{d}{dx}\csc x = -\csc x\cot x$
	$\dfrac{d}{dx}\cos x = -\sin x$	$\dfrac{d}{dx}\sec x = \sec x\tan x$
	$\dfrac{d}{dx}\tan x = \sec^2 x$	$\dfrac{d}{dx}\cot x = -\csc^2 x$
Derivatives of inverse trig functions	$\dfrac{d}{dx}\arcsin x = \dfrac{1}{\sqrt{1-x^2}}$	
	$\dfrac{d}{dx}\arccos x = -\dfrac{1}{\sqrt{1-x^2}}$	
	$\dfrac{d}{dx}\arctan x = \dfrac{1}{1+x^2}$	

Section 1: Year 1 Review

1. (No GDC) Prove by contradiction that if $n^3 + 3$ is odd then n is even, $\forall n \in \mathbb{Z}^+$.
2. (No GDC) Use induction to prove that $5^{2n-1} + 1$ is divisible by 6, $\forall n \in \mathbb{N}$.
3. (GDC) For each sequence, find the general term and sum.
   (a) $4, -1, -6, \ldots, -66$
   (b) $5, -\frac{5}{2}, \frac{5}{4}, -\frac{5}{8}, \ldots$
4. (GDC) A delegation of five students is to be selected for a Model United Nations conference. There are 10 boys and 13 girls to choose from.
   (a) In how many different ways can a delegation be chosen if there are no restrictions?
   (b) If they team is to include at least one boy and one girl, in how many ways can the delegation be selected?
5. (GDC?) Find the term in x^6 in the expansion of $\left(x^2 - \frac{1}{x}\right)^6$. (Hint: write $\frac{1}{x}$ using an exponent.)
6. (No GDC) Consider the functions f and g where $f(x) = 3x - 5$ and $g(x) = x - 2$.
   (a) Find the inverse function f^{-1}.
   (b) Given that $g^{-1}(x) = x + 2$, find $(g^{-1} \circ f)(x)$.
   (c) Given also that $(f^{-1} \circ g)(x) = \frac{x+3}{x}$, solve $(f^{-1} \circ g)(x) = (g^{-1} \circ f)(x)$.
7. (No GDC) The function f is defined by $f(x) = \frac{ax+b}{cx+d}, x \neq -\frac{d}{c}$. The function g is defined by $g(x) = \frac{2x-3}{x-2}, x \neq -2$.
   (a) Find the inverse function f^{-1} and the domain of f^{-1}.
   (b) Express $g(x)$ in the form $A + \frac{B}{x-2}$ where A, B are constants.
   (c) Sketch the graph of $y = g(x)$. State the equations of any asymptotes and the coordinates of any intercepts with the axes.
   (d) The function h is defined by $h(x) = \sqrt{x}, x \geq 0$. State the domain and range of $h \circ g$.

8. (No GDC) A rational function is defined by $f(x) = a + \dfrac{b}{x-c}$ where the parameters $a, b, c \in \mathbb{Z}$ and $x \in \mathbb{R}, x \neq c$. The graph below is the graph of $y = f(x)$.



   (a) Using the information on the graph, state the values of a and c.

   (b) Express the translation of the parent function $y = \dfrac{1}{x}$ using the column vector $\begin{pmatrix} h \\ k \end{pmatrix}$.

   (c) Find the value of b.

9. (No GDC) Given $f(x) = x^2 + x(2-k) + k^2$, find the range of values of k for which $f(x) > 0$ for all real numbers x. (Hint: If $f(x) > 0\ \forall x \in \mathbb{R}$, how many real zeroes are there?)

10. (No GDC) The cubic polynomial $3x^3 + px^2 + qx - 2$ has a factor $(x+2)$ and leaves a remainder 4 when divided by $(x+1)$. Find the values of p and q.

11. (No GDC)

   (a) Show that the complex number i is a root of the equation $x^4 - 5x^3 + 7x^2 - 5x + 6 = 0$.

   (b) Hence or otherwise, find the other roots of this equation.

12. (GDC to check only) Let $z_1 = 2 + 3i$, $z_2 = 1 - 4i$, and $z_3 = 2 + i$. Find $\dfrac{z_1 + z_2}{z_3^*}$.

13. (No GDC) Consider the function defined by $f(x) = 2x^2 + 3x - 4$.

   (a) By differentiating from first principles, show that $f'(x) = 4x + 3$.

   (b) Hence, find an equation for the tangent to the graph of f at $x = 1$.

   (c) Find an equation for the normal line to the graph of f at $x = 1$.

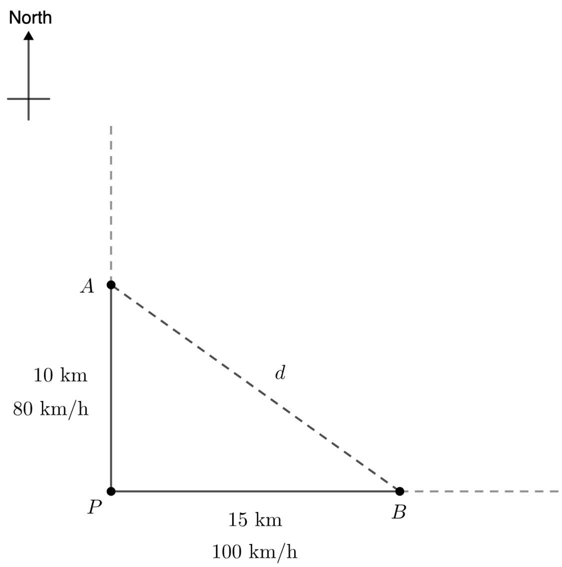
14. For each relation, find $\dfrac{dy}{dx}$.

   (a) $y \cos x = x^2 + y^2$

   (b) $\cos(xy) = 1 + \sin y$

15. (GDC) You are inflating a spherical balloon at a rate of 7 $\text{cm}^3\ \text{s}^{-1}$. How fast is its radius increasing when the radius is 4 cm?

16. (GDC) A road running north to south crosses a road going east to west at point P. Car A is driving north along the first road, and car B is driving east along the second road. At a particular time car A is 10 km to the north of P and traveling at 80 km h^{-1}, while car B is 15 km to the east of P and traveling at 100 km h^{-1}. How fast is the distance between the two cars changing? A diagram has been provided.



17. (GDC) Find the dimensions of the rectangle of largest area with perimeter 100 ft.

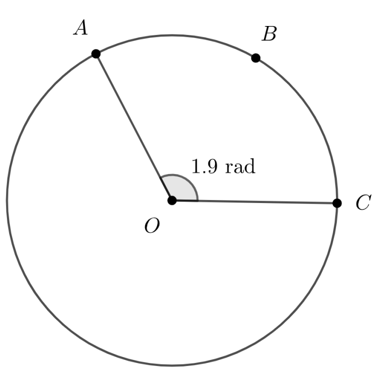
18. (GDC) A piece of cardboard is 1 m by 0.5 m. A square is cut from each corner and the remaining sides are folded up to make an open-top box. What are the dimensions of the box with maximum volume?

19. (GDC) A metal sphere has a radius 12.7 cm.

    (a) Find the volume of the sphere, expressing your answer in the form $a \times 10^k$, $1 \leq a < 10, k \in \mathbb{Z}$.

    (b) The sphere is to be melted down and remolded into the shape of a cone with a height of 14.8 cm. Find the radius of the base of the cone, correct to 3 significant figures.

20. (GDC) The following diagram shows a circle with center O and radius 40 cm. The points A, B, C are on the circumference of the circle and $A\hat{O}C = 1.9$ rad.

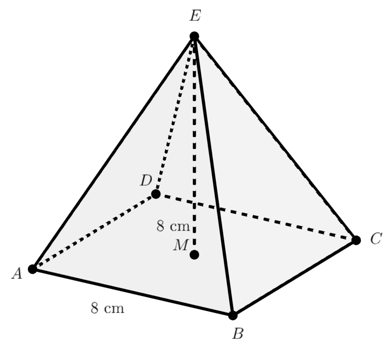


(a) Find the length of arc ABC.

(b) Find the perimeter of sector $OABC$.

(c) Find the area of sector $OABC$.

21. (GDC) $ABCDE$ is a square-based right pyramid with $AB = 8$ cm. The point E is 8 cm directly above the point M at the center of the base, as shown below.



(a) Find CA.

(b) Find AM.

(c) Find EA.

(d) Find the angle between EA and the base.

22. (No GDC) Verify the following identities.

(a) $\sin 4x = 8\cos^3 x \sin x - 4\cos x \sin x$

(b) $\dfrac{1}{\tan\theta + \cot\theta} = \sin\theta\cos\theta$

23. Find the derivative of each of the following functions.

(a) $f(x) = \sin(\cos x)$

(b) $g(x) = \csc^3 x$

(c) $h(x) = \arctan(2x + 1)$

24. An observer watched a rocket launch from a distance of 2 km. The angle of elevation θ is increasing at 3° per second at the instant when $\theta = 45°$. How fast is the rocket climbing at that instant?

25. Two sides of a triangle have lengths 12 m and 15 m. The angle between them is increasing at a rate of 2°/min. How fast is the length of the third side increasing when the angle between the sides of fixed length is 60°?

Section 2: 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?
  - 1.

- 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 smple 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 ______________________, μ if ______________________)
  - Define mean as you remember it:

  - Let's update our knowledge of mean: $\bar{x}$ or $\mu = \dfrac{\sum_{i=1}^{k} f_i x_i}{n}$, 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 show 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 σ^2.
- The ____________________ the data are to each other, the ____________________ the value of σ.
- 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 μ.

- 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 σx 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.

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

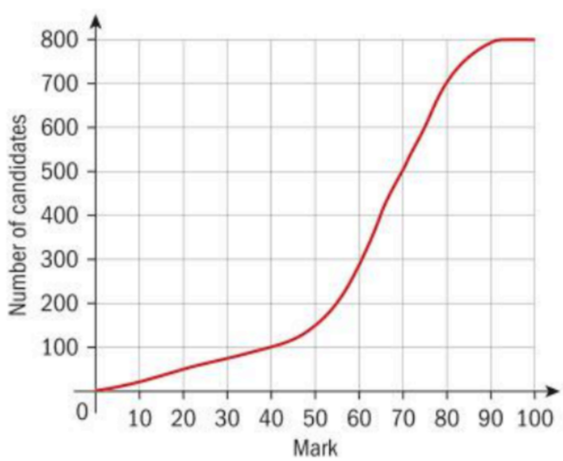
Calculate the mean and standard deviation of this data.

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.

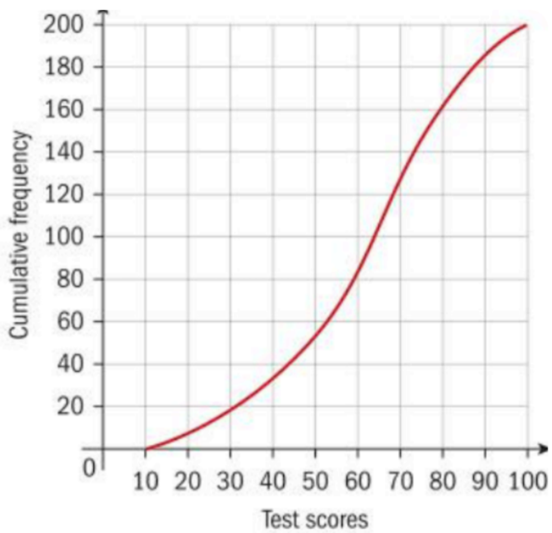
4. A quiz marked out of 100 is written by 800 students. The cumulative frequency graph for the marks is given below.



(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.