Summer 2025

Dear Incoming Student,

Congratulations on accepting the challenge of taking International Baccalaureate Mathematics Higher Level (IB Math HL). I have prepared this two section packet to give you additional information about the course and to help you prepare to be successful in IB Math HL.

IB Math HL is a rigorous course that covers topics in Algebra, Functions, Trigonometry, Vectors, Statistics and Probability, and Calculus. In your second year as a HL student, you will take an IB Math HL exam certifying that you have mastered the content of a college-level curriculum. Depending on where you go to school, you may be granted college credit based on the score you receive on the exam. You will also be writing an Internal Assessment. The Internal Assessment is worth 20% of your IB Exam grade.

IB Math HL is a demanding course. You should be proud of the work you have done to prepare yourself thus far. My goal is to help you grow as a mathematician as much as possible over the next year, but I can only be a guide. It is up to you to put forth the consistent effort necessary to succeed in this course. Note: If you struggled in Algebra 2 Honors, this course will be extremely difficult for you. If you have trouble with this packet you might want to see your counselor about a schedule change.

I have developed a **summer assignment** that includes **two sections**. The **Section 1** is to refresh some skills necessary for success in this course. Please complete the entire packet, showing all of your work carefully, **without use of a calculator wherever possible**. Many of our exams this year will have a non-calculator component in order to prepare for that aspect of the IB exam. Everything in this packet should be review for you. If you need help with any of the topics, check out online resources like Khan Academy or You Tube. Be sure to use the attached formula sheet as well. The full IB Formula Packet is posted on Google Classroom. Whenever giving approximate answers, use 3 significant figures. IB uses significant figures not a specified number of decimal places. I also included a resource sheet with some notation and command terms that you should start becoming familiar with.

Section 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. **Problems set in Chapter 5 should be done on a separate piece of paper.**

Please print both sections of the summer packet and complete all work and notes on the printed copies.

SUBMIT Section 1 BY JULY 31ST ON GOOGLE CLASSROOM (Code: rs4cfmcv) AND BRING Section 2 TO THE FIRST DAY OF CLASS!!! You must show all work to receive credit. Please make sure that you have mastered the material in this packet to ensure your success. All the material in this packet is content that will be on the IB Exam as well as on exams periodically throughout the year

BOTH PARTS OF THE SUMMER PACKET WILL BE GRADED!

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

I look forward to getting to know you and working with you next year. Please feel free to email me at any time with questions. Have a relaxing, enjoyable summer. I look forward to studying mathematics with you next year!

Sincerely,

Mrs. Kimberly Steere

ksteere@theproutschool.org

Equations given in Class and on the IB Exam

Topic 1—Algebra

1.1	The n^{th} term of an arithmetic sequence	$u_n = u_1 + (n-1)d$
	The sum of n terms of an arithmetic sequence	$S_n = \frac{n}{2}(2u_1 + (n-1)d) = \frac{n}{2}(u_1 + u_n)$
	The n^{th} term of a geometric sequence	$u_n = u_1 r^{n-1}$
	The sum of n terms of a finite geometric sequence	$S_n = \frac{u_1(r^n - 1)}{r-1} = \frac{u_1(1-r^n)}{1-r},\ r \neq 1$
	The sum of an infinite geometric sequence	$S_\infty = \frac{u_1}{1-r},\ \lvert r \rvert < 1$
1.2	Exponents and logarithms	$a^x = b \Leftrightarrow x = \log_a b$
	Laws of logarithms	$\log_c a + \log_c b = \log_c ab$ $\log_c a - \log_c b = \log_c \frac{a}{b}$ $\log_c a^r = r \log_c a$
	Change of base	$\log_b a = \frac{\log_c a}{\log_c b}$

Topic 2—Functions and equations

2.4	Axis of symmetry of graph of a quadratic function	$f(x) = ax^2 + bx + c \Rightarrow$ axis of symmetry $x = -\frac{b}{2a}$
2.6	Relationships between logarithmic and exponential functions	$a^x = e^{x \ln a}$ $\log_a a^x = x = a^{\log_a x}$
2.7	Solutions of a quadratic equation	$ax^2 + bx + c = 0 \Rightarrow x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},\ a \neq 0$
	Discriminant	$\Delta = b^2 - 4ac$

Interval Notation (Not given in class or on the exam... just a refresher)

Description	Interval notation	Description	Interval notation	Description	Interval notation
$x > a$	(a, ∞)	$x \leq a$	$(-\infty, a]$	$a \leq x < b$	$[a, b)$
$x \geq a$	$[a, \infty)$	$a < x < b$	(a, b) - open interval	$a < x \leq b$	$(a, b]$
$x < a$	$(-\infty, a)$	$a \leq x \leq b$	$[a, b]$ - closed interval	All real numbers	$(-\infty, \infty)$

MEMORIZE NOTATION/COMMAND TERMS LIST

NOTATION

<table>
<tr><td>Number Sets</td><td>

$\mathbb{N}$	the set of positive integers and zero, $\{0,1,2,3,\ldots\}$
$\mathbb{Z}$	the set of integers, $\{0,\pm1,\pm2,\pm3,\ldots\}$
$\mathbb{Z}^+$	the set of positive integers, $\{1,2,3,\ldots\}$
$\mathbb{Q}$	the set of rational numbers (any # that can be written as a fraction)
$\mathbb{Q}^+$	the set of positive rational numbers, $\{x \mid x \in \mathbb{Q}, x > 0\}$
$\mathbb{R}$	the set of real numbers
$\mathbb{R}^+$	the set of positive real numbers, $\{x \mid x \in \mathbb{R}, x > 0\}$

</td></tr>
<tr><td>Absolute Value</td><td>IB will refer to this as modulus</td></tr>
<tr><td>Line Segments</td><td>Line segment, $\overline{AB}$, will be written as $[AB]$</td></tr>
<tr><td>Angles</td><td>We write Angle A as $\angle A$.
IB will use the following notation: $\hat{A}$</td></tr>
<tr><td>Repeating Decimals</td><td>

		IB Notation
$\frac{1}{3}$	$0.\overline{3}$	$0.\dot{3}$
.123123123123	$0.\overline{123}$	$0.\dot{1}2\dot{3}$

</td></tr>
<tr><td>Slope</td><td>IB will refer to this as the gradient</td></tr>
</table>

COMMAND TERMS

Calculate	Obtain a numerical answer showing the relevant stages in the 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 drawn 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 proceeding 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 information given) without the formality of proof. "Show that" 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 or no calculation is required. Working does not need to be shown.

QUADRATICS

Section A: Factor each quadratic. If the quadratic cannot be factored, write "prime."

1. $x^2 - x - 2$
2. $x^2 + 3x - 4$
3. $8x^2 - 50y^2$
4. $3x^2 - 5x + 2$
5. $2x^2 - x - 6$
6. $x^3 - 3x^2 - 18x$

Section B: Solve each equation using any method except graphing or guess and check.

1. $x^2 + 25 = 10x$
2. $x^2 + 3x - 1 = 0$
3. $x + \frac{12}{x} = 7$
4. $x^2 + 2 = 9$
5. $x^2 - 5x = 0$
6. $36x^2 - 25 = 0$

Section C: State the following for each of the given equations: axis of symmetry, vertex, direction of opening, x-intercepts, and y-intercepts. Then sketch the graph using that information.

1. $y = -2(x+2)(x-1)$
2. $y = 0{,}5(x-2)^2 - 4$
3. $y = 2x^2 + 6x - 3$

Section D: Find the values of p such that the equations below have the given characteristics. Hint; Use the discriminant.

1. Two different real roots
   $px^2 + 5x + 2 = 0$
2. Two equal real roots
   $2x^2 - 3x + p = 0$
3. No real roots
   $px^2 - 4px + 5 - p = 0$

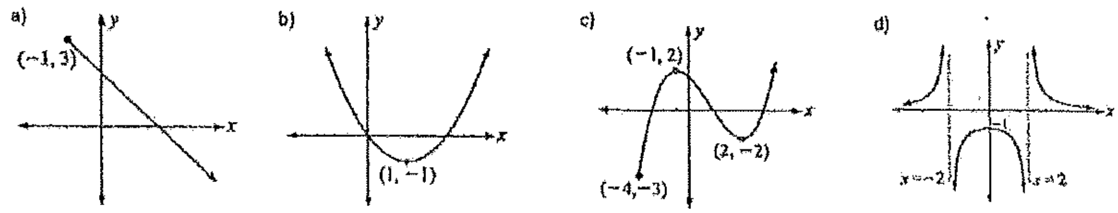
Section E: Use your graphing calculator to find the following

1. Solve $3x^2 - x - 5 = 0$
2. Intersection points of $y = -x^2 - 5x + 3$ and $y = x^2 + 3x + 11$

FUNCTIONS

Section A: For each of the following find the domain and range without using a calculator.

1.



2.

a) $f(x) = \sqrt{x}$ b) $f(x) = \sqrt{4 - x}$ c) $y = 5x - 3x^2$ d) $y = \frac{x+4}{x-2}$

Section B: Find the inverse of each function.

1. $f(x) = 2x + 1$

2. $f(x) = \frac{x^2}{3}$

3. $g(x) = \frac{5}{x-2}$

4. $g(x) = \sqrt{4 - x} + 1$

5. If the point (2, 7) is on the graph of $f(x)$, what point must be on the graph of $f^{-1}(x)$?

6. Explain, in complete sentences, the relationship between a function and its inverse.

Section C: Let $f(x) = 2x^2 - 1$; $g(x) = 3x$ and $h(x) = 5 - x$. Find the following.

1. $f(-3)$

2. $(f \circ g)(x)$

3. $(h \circ f)(x)$

4. $(f \circ h)(x + 1)$

5. $(g \circ h)(4)$

6. $(f \circ f)(-1)$

Section D: Answer the following questions concerning equations of lines

1. What is the slope, x-intercept, and y-intercept of the equation $5x - 4y = 8$?

2. What is the slope-intercept form of the equation of the line between the points (4, 3) and (7, –2) ?

3. What is the slope-intercept form of a line perpendicular to $y = -2x + 9$ passing through the (4, 7) ?

Section E: Find the horizontal & vertical asymptotes and holes (if applicable) of the following.

1. $y = \frac{1}{2x-5}$

2. $y = \frac{x^2-5}{2x^2-12}$

3. $y = \frac{x^2+2x-3}{x^3+6x^2-7x}$

Section F: For each pair of functions $f(x)$ and $g(x)$, describe the transformations that would transform $f(x)$ into $g(x)$.

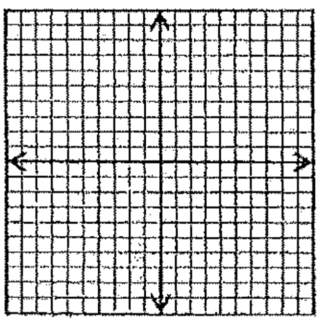
1. $f(x) = x^2$;
   $g(x) = (x-5)^2 + 2$

2. $f(x) = \sqrt{x}$;
   $g(x) = \sqrt{3x} - 10$

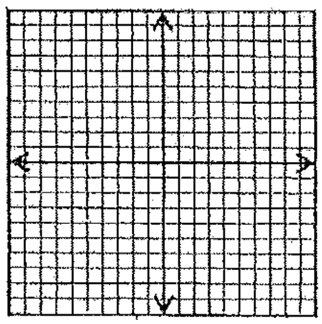
3. $f(x) = e^x$;
   $g(x) = -5(e)^{x-1}$

Section G: Graph each function, clearly showing its key features (maxima, minima, and intercepts). Identify its domain and range. (Remember: No calculator!)

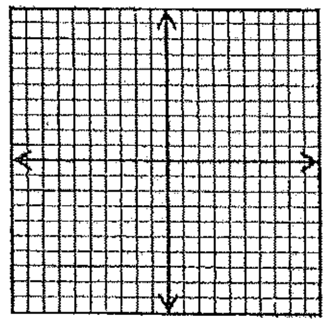
1. $f(x) = x^2 - 5$



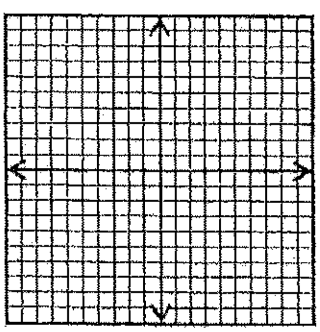
2. $f(x) = 3x - 4$



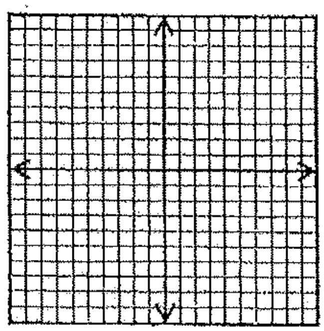
3. $f(x) = x^3 + 1$



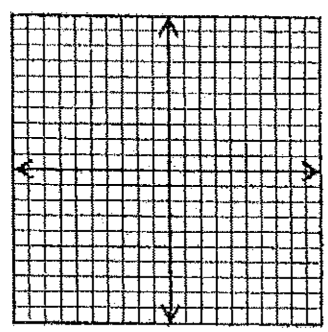
4. $f(x) = \sqrt{x+6}$



5. $f(x) = |x-1| + 3$



6. $f(x) = 2^x - 4$



ALGEBRA

Section A: Simplify the following without a calculator.

1. $(2x^5)^{-3}$ 2. $8^{2/3}$ 3. $81^{-3/4}$ 4. $\sqrt[3]{16x^3}$

5. $\sqrt{10x^2} \cdot \sqrt{70x^6}$ 6. $\frac{\sqrt{72x^4}}{\sqrt{3x}}$ 7. $\frac{5}{7-\sqrt{5}}$ 8. $\sqrt{5} - 5\sqrt{125} - 7\sqrt{180}$

Section B: Solve using algebra.

1. $\begin{aligned} 3x + 7y &= 36 \\ x &= 5y - 10 \end{aligned}$

2. $\begin{aligned} 6x + 10y &= 32 \\ 4x - 2y &= 4 \end{aligned}$

3. $\begin{aligned} x &= y^2 \\ x - y &= 6 \end{aligned}$

4. $\begin{aligned} x^2 + y^2 &= 25 \\ y &= x^2 - 13 \end{aligned}$

Section C: Solve for x. Eliminate any extraneous solutions, if any.

1. $\sqrt{37 - 3x} = x - 3$ 2. $-3(2x + 1)^3 = -192$ 3. $\frac{x}{3} - \frac{5}{2} = \frac{-3}{x}$

4. $\frac{4x-1}{x+1} = x - 1$ 5. $2|3x - 1| + 5 = -2x + 8$ 6. $5(x - 3) \leq 8(x + 5)$

EXPONENTIAL AND LOGARITHMIC EQUATIONS

Section A: Find the following without using a calculator.

1.

a) $\log_4 64$ b) $\log_2 1/4$ c) $\log_8 1$ d) $\log_9 3$ e) $\log_m m^6$ f) $\ln(e^{2x})$

Section B: Solve each equation for x or y.

1. $7 = 5^x$
2. $25e^{x/2} = 750$
3. $\log_2 y = 3$
4. $3\ln x + 2 = 0$
5. $\log_2 y + \log_2(y+1) = 1$
6. $4^y = 32$ (Solve without a calculator)

Section C: Answer the following questions about the equation $W = 2500(3^{-t/3000})$ where W is the weight in gram of a radioactive substance after t years.

1. a) Find the initial weight
   b) Find the weight after 1500 years
2. Find how many years it takes to reduce its value 30%

Rational Operations

Section A: Simplify each expression.

(Remember to factor where necessary and state restricted domain values)

1. $\frac{9x+18}{x^2-2x-8} \cdot \frac{3x-12}{6x}$

2. $\frac{m^2-2m-8}{8m+24} \div \frac{2m-8}{m^2+7m+12}$

3. $\frac{6x-30}{x^2-7x+10} \cdot \frac{7x-14}{6x}$

4. $\frac{x^2-16}{9-x} \cdot \frac{x^2+x-90}{x^2+14x+40}$

5. $\frac{1}{5p^2} \div \frac{9p-36}{5p^3-35p^2}$

6. $\frac{x-1}{3x+4} + \frac{2x+9}{3x+4}$

7. $\frac{3}{y+5} + \frac{y}{y^2+7y+10}$

8. $\frac{6x-7}{x^2+6x+5} + \frac{4}{x+5}$

Section 2: Due the First Day of School **Name:_______________________**

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 completed on a SEPARATE PIECE OF PAPER THAT IS CLEARLY MARKED FOR EACH SECTION.

https://youtube.com/playlist?list=PLmrVBIbuhI6m1I1sM4anomSMXC2ooCJb5

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 $\mu = \frac{\sum_{i=1}^{k} f_i x_i}{n}$, 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 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
$0 \le x < 15$	5	$0 \le x < 15$	4
$15 \le x < 30$	8	$15 \le x < 30$	5
$30 \le x < 45$	10	$30 \le x < 45$	4
$45 \le x < 60$	5	$45 \le x < 60$	14
$60 \le x < 75$	2	$60 \le x < 75$	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
$15 < A \le 20$	5
$20 < A \le 23$	15
$23 < A \le 25$	20
$25 < A \le 30$	20
$30 < A \le 40$	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.

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

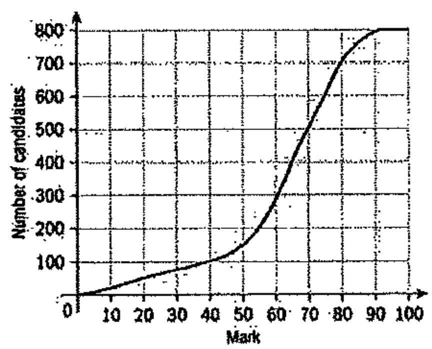
Calculate the mean and standard deviation of this data.

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.

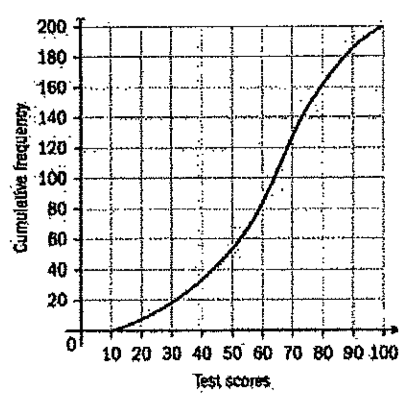
4. A quiz marked out of 100 is written by 800 students. The cumulative frequency graph for the marks is given below.



(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.