Mathematics of Data Management, Grade 12

University Preparation MDM4U

This course broadens students' understanding of mathematics as it relates to managing data. Students will apply methods for organizing and analysing large amounts of information; solve problems involving probability and statistics; and carry out a culminating investigation that integrates statistical concepts and skills. Students will also refine their use of the mathematical processes necessary for success in senior mathematics. Students planning to enter university programs in business, the social sciences, and the humanities will find this course of particular interest.

Prerequisite: Functions, Grade 11, University Preparation, or Functions and Applications, Grade 11, University/College Preparation

MATHEMATICAL PROCESS EXPECTATIONS

The mathematical processes are to be integrated into student learning in all areas of this course.

Throughout this course, students will:

Problem Solving

- develop, select, apply, compare, and adapt a variety of problem-solving strategies as they pose and solve problems and conduct investigations, to help deepen their mathematical understanding;

Reasoning and Proving

- develop and apply reasoning skills (e.g., use of inductive reasoning, deductive reasoning, and counter-examples; construction of proofs) to make mathematical conjectures, assess conjectures, and justify conclusions, and plan and construct organized mathematical arguments;

Reflecting

- demonstrate that they are reflecting on and monitoring their thinking to help clarify their understanding as they complete an investigation or solve a problem (e.g., by assessing the effectiveness of strategies and processes used, by proposing alternative approaches, by judging the reasonableness of results, by verifying solutions);

Selecting Tools and Computational Strategies

- select and use a variety of concrete, visual, and electronic learning tools and appropriate computational strategies to investigate mathematical ideas and to solve problems;

Connecting

- make connections among mathematical concepts and procedures, and relate mathematical ideas to situations or phenomena drawn from other contexts (e.g., other curriculum areas, daily life, current events, art and culture, sports);

Representing

- create a variety of representations of mathematical ideas (e.g., numeric, geometric, algebraic, graphical, pictorial representations; onscreen dynamic representations), connect and compare them, and select and apply the appropriate representations to solve problems;

Communicating

- communicate mathematical thinking orally, visually, and in writing, using precise mathematical vocabulary and a variety of appropriate representations, and observing mathematical conventions.

A. COUNTING AND PROBABILITY

OVERALL EXPECTATIONS

By the end of this course, students will:

1. solve problems involving the probability of an event or a combination of events for discrete sample spaces;
2. solve problems involving the application of permutations and combinations to determine the probability of an event.

SPECIFIC EXPECTATIONS

1. Solving Probability Problems Involving Discrete Sample Spaces

By the end of this course, students will:

1.1 recognize and describe how probabilities are used to represent the likelihood of a result of an experiment (e.g., spinning spinners; drawing blocks from a bag that contains different-coloured blocks; playing a game with number cubes; playing Aboriginal stick-and-stone games) and the likelihood of a real-world event (e.g., that it will rain tomorrow, that an accident will occur, that a product will be defective)

1.2 describe a sample space as a set that contains all possible outcomes of an experiment, and distinguish between a discrete sample space as one whose outcomes can be counted (e.g., all possible outcomes of drawing a card or tossing a coin) and a continuous sample space as one whose outcomes can be measured (e.g., all possible outcomes of the time it takes to complete a task or the maximum distance a ball can be thrown)

1.3 determine the theoretical probability, P_i (i.e., a value from 0 to 1), of each outcome of a discrete sample space (e.g., in situations in which all outcomes are equally likely), recognize that the sum of the probabilities of the outcomes is 1 (i.e., for n outcomes, $P_1 + P_2 + P_3 + \ldots + P_n = 1$), recognize that the probabilities P_i form the probability distribution associated with the sample space, and solve related problems

Sample problem: An experiment involves rolling two number cubes and determining the sum. Calculate the theoretical probability of each outcome, and verify that the sum of the probabilities is 1.

1.4 determine, through investigation using class-generated data and technology-based simulation models (e.g., using a random-number generator on a spreadsheet or on a graphing calculator; using dynamic statistical software to simulate repeated trials in an experiment), the tendency of experimental probability to approach theoretical probability as the number of trials in an experiment increases (e.g., "If I simulate tossing two coins 1000 times using technology, the experimental probability that I calculate for getting two tails on the two tosses is likely to be closer to the theoretical probability of $\frac{1}{4}$ than if I simulate tossing the coins only 10 times")

Sample problem: Calculate the theoretical probability of rolling a 2 on a single roll of a number cube. Simulate rolling a number cube, and use the simulation results to calculate the experimental probabilities of rolling a 2 over 10, 20, 30, …, 200 trials. Graph the experimental probabilities versus the number of trials, and describe any trend.

1.5 recognize and describe an event as a set of outcomes and as a subset of a sample space, determine the complement of an event, determine whether two or more events are mutually exclusive or non-mutually exclusive (e.g., the events of getting an even number or getting an odd number of heads from tossing a coin 5 times are mutually exclusive), and solve related probability problems [e.g., calculate $P(\sim A)$, $P(A \text{ and } B)$, $P(A \text{ or } B)$] using a variety of strategies (e.g., Venn diagrams, lists, formulas)

1.6 determine whether two events are independent or dependent and whether one event is conditional on another event, and solve related probability problems [e.g., calculate $P(A \text{ and } B)$, $P(A \text{ or } B)$, $P(A \text{ given } B)$] using a variety of strategies (e.g., tree diagrams, lists, formulas)

2. Solving Problems Using Counting Principles

By the end of this course, students will:

2.1 recognize the use of permutations and combinations as counting techniques with advantages over other counting techniques (e.g., making a list; using a tree diagram; making a chart; drawing a Venn diagram), distinguish between situations that involve the use of permutations and those that involve the use of combinations (e.g., by considering whether or not order matters), and make connections between, and calculate, permutations and combinations

Sample problem: An organization with 10 members is considering two leadership models. One involves a steering committee with 4 members of equal standing. The other is an executive committee consisting of a president, vice-president, secretary, and treasurer. Determine the number of ways of selecting the executive committee from the 10 members and, using this number, the number of ways of selecting the steering committee from the 10 members. How are the calculations related? Use the calculations to explain the relationship between permutations and combinations.

2.2 solve simple problems using techniques for counting permutations and combinations, where all objects are distinct, and express the solutions using standard combinatorial notation [e.g., $n!$, $P(n, r)$, $\binom{n}{r}$]

Sample problem: In many Aboriginal communities, it is common practice for people to shake hands when they gather. Use combinations to determine the total number of handshakes when 7 people gather, and verify using a different strategy.

2.3 solve introductory counting problems involving the additive counting principle (e.g., determining the number of ways of selecting 2 boys *or* 2 girls from a group of 4 boys and 5 girls) and the multiplicative counting principle (e.g., determining the number of ways of selecting 2 boys *and* 2 girls from a group of 4 boys and 5 girls)

2.4 make connections, through investigation, between combinations (i.e., n choose r) and Pascal's triangle [e.g., between $\binom{2}{r}$ and row 3 of Pascal's triangle, between $\binom{n}{2}$ and diagonal 3 of Pascal's triangle]

Sample problem: A school is 5 blocks west and 3 blocks south of a student's home. Determine, in a variety of ways (e.g., by drawing the routes, by using Pascal's triangle, by using combinations), how many different routes the student can take from home to the school by going west or south at each corner.

2.5 solve probability problems using counting principles for situations involving equally likely outcomes

Sample problem: Two marbles are drawn randomly from a bag containing 12 green marbles and 16 red marbles. What is the probability that the two marbles are both green if the first marble is replaced? If the first marble is not replaced?

B. PROBABILITY DISTRIBUTIONS

OVERALL EXPECTATIONS

By the end of this course, students will:

1. demonstrate an understanding of discrete probability distributions, represent them numerically, graphically, and algebraically, determine expected values, and solve related problems from a variety of applications;
2. demonstrate an understanding of continuous probability distributions, make connections to discrete probability distributions, determine standard deviations, describe key features of the normal distribution, and solve related problems from a variety of applications.

SPECIFIC EXPECTATIONS

1. Understanding Probability Distributions for Discrete Random Variables

By the end of this course, students will:

1.1 recognize and identify a discrete random variable X (i.e., a variable that assumes a unique value for each outcome of a discrete sample space, such as the value x for the outcome of getting x heads in 10 tosses of a coin), generate a probability distribution [i.e., a function that maps each value x of a random variable X to a corresponding probability, $P(X = x)$] by calculating the probabilities associated with all values of a random variable, with and without technology, and represent a probability distribution numerically using a table

1.2 calculate the expected value for a given probability distribution [i.e., using $E(X) = \sum xP(X = x)$], interpret the expected value in applications, and make connections between the expected value and the weighted mean of the values of the discrete random variable

Sample problem: Of six cases, three each hold \$1, two each hold \$1000, and one holds \$100 000. Calculate the expected value and interpret its meaning. Make a conjecture about what happens to the expected value if you add \$10 000 to each case or if you multiply the amount in each case by 10. Verify your conjectures.

1.3 represent a probability distribution graphically using a probability histogram (i.e., a histogram on which each rectangle has a base of width 1, centred on the value of the discrete random variable, and a height equal to the probability associated with the value of the random variable), and make connections between the frequency histogram and the probability histogram (e.g., by comparing their shapes)

Sample problem: For the situation involving the rolling of two number cubes and determining the sum, identify the discrete random variable and generate the related probability histogram. Determine the total area of the bars in the histogram and explain your result.

1.4 recognize conditions (e.g., independent trials) that give rise to a random variable that follows a binomial probability distribution, calculate the probability associated with each value of the random variable, represent the distribution numerically using a table and graphically using a probability histogram, and make connections to the algebraic representation

$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$

Sample problem: A light-bulb manufacturer estimates that 0.5% of the bulbs manufactured are defective. Generate and graph the probability distribution for the random variable that represents the number of defective bulbs in a set of 4 bulbs.

1.5 recognize conditions (e.g., dependent trials) that give rise to a random variable that follows a hypergeometric probability distribution, calculate the probability associated with each value of the random variable (e.g., by using a tree diagram; by using combinations), and represent the distribution numerically using a table and graphically using a probability histogram

1.6 compare, with technology and using numeric and graphical representations, the probability distributions of discrete random variables (e.g., compare binomial distributions with the same probability of success for increasing numbers of trials; compare the shapes of a hypergeometric distribution and a binomial distribution)

Sample problem: Compare the probability distributions associated with drawing 0, 1, 2, or 3 face cards when a card is drawn 3 times from a standard deck with replacement (i.e., the card is replaced after each draw) and without replacement (i.e., the card is not replaced after each draw).

1.7 solve problems involving probability distributions (e.g., uniform, binomial, hypergeometric), including problems arising from real-world applications

Sample problem: The probability of a business person cancelling a reservation at La Place Pascal hotel is estimated to be 8%. Generate and graph the probability distribution for the discrete random variable that represents the number of business people cancelling when there are 10 reservations. Use the probability distribution to determine the probability of at least 4 of the 10 reservations being cancelled.

2. Understanding Probability Distributions for Continuous Random Variables

By the end of this course, students will:

2.1 recognize and identify a continuous random variable (i.e., a variable that assumes values from the infinite number of possible outcomes in a continuous sample space), and distinguish between situations that give rise to discrete frequency distributions (e.g., counting the number of outcomes for drawing a card or tossing three coins) and situations that give rise to continuous frequency distributions (e.g., measuring the time taken to complete a task or the maximum distance a ball can be thrown)

2.2 recognize standard deviation as a measure of the spread of a distribution, and determine, with and without technology, the mean and standard deviation of a sample of values of a continuous random variable

2.3 describe challenges associated with determining a continuous frequency distribution (e.g., the inability to capture all values of the variable, resulting in a need to sample; uncertainties in measured values of the variable), and recognize the need for mathematical models to represent continuous frequency distributions

2.4 represent, using intervals, a sample of values of a continuous random variable numerically using a frequency table and graphically using a frequency histogram and a frequency polygon, recognize that the frequency polygon approximates the frequency distribution, and determine, through investigation using technology (e.g., dynamic statistical software, graphing calculator), and compare the effectiveness of the frequency polygon as an approximation of the frequency distribution for different sizes of the intervals

2.5 recognize that theoretical probability for a continuous random variable is determined over a range of values (e.g., the probability that the life of a lightbulb is between 90 hours and 115 hours), that the probability that a continuous random variable takes any single value is zero, and that the probabilities of ranges of values form the probability distribution associated with the random variable

2.6 recognize that the normal distribution is commonly used to model the frequency and probability distributions of continuous random variables, describe some properties of the normal distribution (e.g., the curve has a central peak; the curve is symmetric about the mean; the mean and median are equal; approximately 68% of the data values are within one standard deviation of the mean and approximately 95% of the data values are within two standard deviations of the mean), and recognize and describe situations that can be modelled using the normal distribution (e.g., birth weights of males or of females, household incomes in a neighbourhood, baseball batting averages)

2.7 make connections, through investigation using dynamic statistical software, between the normal distribution and the binomial and hypergeometric distributions for increasing numbers of trials of the discrete distributions (e.g., recognizing that the shape of the hypergeometric distribution of the number of males on a 4-person committee selected from a group of people more closely resembles the shape of a normal distribution as the size of the group from which the committee was drawn increases)

Sample problem: Explain how the total area of a probability histogram for a binomial distribution allows you to predict the area under a normal probability distribution curve.

2.8 recognize a *z*-score as the positive or negative number of standard deviations from the mean to a value of the continuous random variable, and solve probability problems involving normal distributions using a variety of tools and strategies (e.g., calculating a *z*-score and reading a probability from a table; using technology to determine a probability), including problems arising from real-world applications

Sample problem: The heights of 16-month-old maple seedlings are normally distributed with a mean of 32 cm and a standard deviation of 10.2 cm. What is the probability that the height of a randomly selected seedling will be between 24.0 cm and 38.0 cm?

C. ORGANIZATION OF DATA FOR ANALYSIS

OVERALL EXPECTATIONS

By the end of this course, students will:

1. demonstrate an understanding of the role of data in statistical studies and the variability inherent in data, and distinguish different types of data;
2. describe the characteristics of a good sample, some sampling techniques, and principles of primary data collection, and collect and organize data to solve a problem.

SPECIFIC EXPECTATIONS

1. Understanding Data Concepts

By the end of this course, students will:

1.1 recognize and describe the role of data in statistical studies (e.g., the use of statistical techniques to extract or mine knowledge of relationships from data), describe examples of applications of statistical studies (e.g., in medical research, political decision making, market research), and recognize that conclusions drawn from statistical studies of the same relationship may differ (e.g., conclusions about the effect of increasing jail sentences on crime rates)

1.2 recognize and explain reasons why variability is inherent in data (e.g., arising from limited accuracy in measurement or from variations in the conditions of an experiment; arising from differences in samples in a survey), and distinguish between situations that involve one variable and situations that involve more than one variable

Sample problem: Use the Census at School database to investigate variability in the median and mean of, or a proportion estimated from, equal-sized random samples of data on a topic such as the percentage of students who do not smoke or who walk to school, or the average height of people of a particular age. Compare the median and mean of, or a proportion estimated from, samples of increasing size with the median and mean of the population or the population proportion.

1.3 distinguish different types of statistical data (i.e., discrete from continuous, qualitative from quantitative, categorical from numerical, nominal from ordinal, primary from secondary, experimental from observational, microdata from aggregate data) and give examples (e.g., distinguish experimental data used to compare the effectiveness of medical treatments from observational data used to examine the relationship between obesity and type 2 diabetes or between ethnicity and type 2 diabetes)

2. Collecting and Organizing Data

By the end of this course, students will:

2.1 determine and describe principles of primary data collection (e.g., the need for randomization, replication, and control in experimental studies; the need for randomization in sample surveys) and criteria that should be considered in order to collect reliable primary data (e.g., the appropriateness of survey questions; potential sources of bias; sample size)

2.2 explain the distinction between the terms *population* and *sample*, describe the characteristics of a good sample, explain why sampling is necessary (e.g., time, cost, or physical constraints), and describe and compare some sampling techniques (e.g., simple random, systematic, stratified, convenience, voluntary)

Sample problem: What are some factors that a manufacturer should consider when determining whether to test a sample or the entire population to ensure the quality of a product?

2.3 describe how the use of random samples with a bias (e.g., response bias, measurement bias, non-response bias, sampling bias) or the use of non-random samples can affect the results of a study

2.4 describe characteristics of an effective survey (e.g., by giving consideration to ethics, privacy, the need for honest responses, and possible sources of bias, including cultural bias), and design questionnaires (e.g., for determining if there is a relationship between a person's age and their hours per week of Internet use, between marks and hours of study, or between income and years of education) or experiments (e.g., growth of plants under different conditions) for gathering data

Sample problem: Give examples of concerns that could arise from an ethical review of surveys generated by students in your school.

2.5 collect data from primary sources, through experimentation, or from secondary sources (e.g., by using the Internet to access reliable data from a well-organized database such as E-STAT; by using print sources such as newspapers and magazines), and organize data with one or more attributes (e.g., organize data about a music collection classified by artist, date of recording, and type of music using dynamic statistical software or a spreadsheet) to answer a question or solve a problem

D. STATISTICAL ANALYSIS

OVERALL EXPECTATIONS

By the end of this course, students will:

1. analyse, interpret, and draw conclusions from one-variable data using numerical and graphical summaries;
2. analyse, interpret, and draw conclusions from two-variable data using numerical, graphical, and algebraic summaries;
3. demonstrate an understanding of the applications of data management used by the media and the advertising industry and in various occupations.

SPECIFIC EXPECTATIONS

1. Analysing One-Variable Data

By the end of this course, students will:

1.1 recognize that the analysis of one-variable data involves the frequencies associated with one attribute, and determine, using technology, the relevant numerical summaries (i.e., mean, median, mode, range, interquartile range, variance, and standard deviation)

1.2 determine the positions of individual data points within a one-variable data set using quartiles, percentiles, and *z*-scores, use the normal distribution to model suitable one-variable data sets, and recognize these processes as strategies for one-variable data analysis

1.3 generate, using technology, the relevant graphical summaries of one-variable data (e.g., circle graphs, bar graphs, histograms, stem-and-leaf plots, boxplots) based on the type of data provided (e.g., categorical, ordinal, quantitative)

1.4 interpret, for a normally distributed population, the meaning of a statistic qualified by a statement describing the margin of error and the confidence level (e.g., the meaning of a statistic that is accurate to within 3 percentage points, 19 times out of 20), and make connections, through investigation using technology (e.g., dynamic statistical software), between the sample size, the margin of error, and the confidence level (e.g., larger sample sizes create higher confidence levels for a given margin of error)

Sample problem: Use census data from Statistics Canada to investigate, using dynamic statistical software, the minimum sample size such that the proportion of the sample opting for a particular consumer or voting choice is within 3 percentage points of the proportion of the population, 95% of the time (i.e., 19 times out of 20).

1.5 interpret statistical summaries (e.g., graphical, numerical) to describe the characteristics of a one-variable data set and to compare two related one-variable data sets (e.g., compare the lengths of different species of trout; compare annual incomes in Canada and in a third-world country; compare Aboriginal and non-Aboriginal incomes); describe how statistical summaries (e.g., graphs, measures of central tendency) can be used to misrepresent one-variable data; and make inferences, and make and justify conclusions, from statistical summaries of one-variable data orally and in writing, using convincing arguments

2. Analysing Two-Variable Data

By the end of this course, students will:

2.1 recognize that the analysis of two-variable data involves the relationship between two attributes, recognize the correlation coefficient as a measure of the fit of the data to a linear model, and determine, using technology, the relevant numerical summaries (e.g., summary tables such as contingency tables; correlation coefficients)

Sample problem: Organize data from Statistics Canada to analyse gender differences (e.g., using contingency tables; using correlation coefficients) related to a specific set of characteristics (e.g., average income, hours of unpaid housework).

2.2 recognize and distinguish different types of relationships between two variables that have a mathematical correlation (e.g., the cause-and-effect relationship between the age of a tree and its diameter; the common-cause relationship between ice cream sales and forest fires over the course of a year; the accidental relationship between the consumer price index and the number of known planets in the universe)

2.3 generate, using technology, the relevant graphical summaries of two-variable data (e.g., scatter plots, side-by-side boxplots) based on the type of data provided (e.g., categorical, ordinal, quantitative)

2.4 determine, by performing a linear regression using technology, the equation of a line that models a suitable two-variable data set, determine the fit of an individual data point to the linear model (e.g., by using residuals to identify outliers), and recognize these processes as strategies for two-variable data analysis

2.5 interpret statistical summaries (e.g., scatter plot, equation representing a relationship) to describe the characteristics of a two-variable data set and to compare two related two-variable data sets (e.g., compare the relationship between Grade 12 English and mathematics marks with the relationship between Grade 12 science and mathematics marks); describe how statistical summaries (e.g., graphs, linear models) can be used to misrepresent two-variable data; and make inferences, and make and justify conclusions, from statistical summaries of two-variable data orally and in writing, using convincing arguments

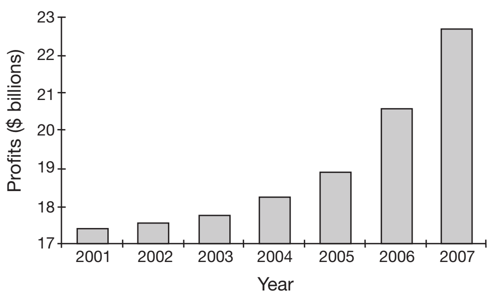
3. Evaluating Validity

By the end of this course, students will:

3.1 interpret statistics presented in the media (e.g., the UN's finding that 2% of the world's population has more than half the world's wealth, whereas half the world's population has only 1% of the world's wealth), and explain how the media, the advertising industry, and others (e.g., marketers, pollsters) use and misuse statistics (e.g., as represented in graphs) to promote a certain point of view (e.g., by making a general statement based on a weak correlation or an assumed cause-and-effect relationship; by starting the vertical scale at a value other than zero; by making statements using general population statistics without reference to data specific to minority groups)

3.2 assess the validity of conclusions presented in the media by examining sources of data, including Internet sources (i.e., to determine whether they are authoritative, reliable, unbiased, and current), methods of data collection, and possible sources of bias (e.g., sampling bias, non-response bias, cultural bias in a survey question), and by questioning the analysis of the data (e.g., whether there is any indication of the sample size in the analysis) and conclusions drawn from the data (e.g., whether any assumptions are made about cause and effect)

Sample problem: The headline that accompanies the following graph says "Big Increase in Profits". Suggest reasons why this headline may or may not be true.



3.3 gather, interpret, and describe information about applications of data management in occupations (e.g., actuary, statistician, business analyst, sociologist, medical doctor, psychologist, teacher, community planner), and about university programs that explore these applications

E. CULMINATING DATA MANAGEMENT INVESTIGATION

OVERALL EXPECTATIONS

By the end of this course, students will:

1. design and carry out a culminating investigation* that requires the integration and application of the knowledge and skills related to the expectations of this course;

2. communicate the findings of a culminating investigation and provide constructive critiques of the investigations of others.

SPECIFIC EXPECTATIONS

1. Designing and Carrying Out a Culminating Investigation

By the end of this course, students will:

1.1 pose a significant problem of interest that requires the organization and analysis of a suitable set of primary or secondary quantitative data (e.g., primary data collected from a student-designed game of chance, secondary data from a reliable source such as E-STAT), and conduct appropriate background research related to the topic being studied

1.2 design a plan to study the problem (e.g., identify the variables and the population; develop an ethical survey; establish the procedures for gathering, summarizing, and analysing the primary or secondary data; consider the sample size and possible sources of bias)

1.3 gather data related to the study of the problem (e.g., by using a survey; by using the Internet; by using a simulation) and organize the data (e.g., by setting up a database; by establishing intervals), with or without technology

1.4 interpret, analyse, and summarize data related to the study of the problem (e.g., generate and interpret numerical and graphical statistical summaries; recognize and apply a probability distribution model; calculate the expected value of a probability distribution), with or without technology

1.5 draw conclusions from the analysis of the data (e.g., determine whether the analysis solves the problem), evaluate the strength of the evidence (e.g., by considering factors such as sample size or bias, or the number of times a game is played), specify any limitations of the conclusions, and suggest follow-up problems or investigations

2. Presenting and Critiquing the Culminating Investigation

By the end of this course, students will:

2.1 compile a clear, well-organized, and detailed report of the investigation

2.2 present a summary of the culminating investigation to an audience of their peers within a specified length of time, with technology (e.g. presentation software) or without technology

2.3 answer questions about the culminating investigation and respond to critiques (e.g., by elaborating on the procedures; by justifying mathematical reasoning)

2.4 critique the mathematical work of others in a constructive manner

*This culminating investigation allows students to demonstrate their knowledge and skills from this course by addressing a single problem on probability and statistics or by addressing two smaller problems, one on probability and the other on statistics.