Mathematics Notes/Ebook: A Comprehensive Guide for Secondary Schools

Introduction

This ebook is designed to provide a clear and concise understanding of key mathematical concepts typically covered in secondary schools. It is structured into five main sections: Number and Numeration, Algebra, Geometry and Trigonometry, Statistics, and Revision. Each section builds upon foundational knowledge, progressively introducing more complex topics.

SECTION D: STATISTICS

This section introduces the fundamental concepts of statistics, focusing on the collection, organization, analysis, interpretation, and presentation of data.

1. Representation of Data

- **Definition of Data:** Raw facts and figures collected for a specific purpose.
- **Types of Data:**
    - **Qualitative Data:** Describes characteristics or categories (e.g., colors, genders, types of cars).
    - **Quantitative Data:** Deals with numbers and can be measured or counted.
        - **Discrete Data:** Can only take specific, distinct values (e.g., number of students, shoe sizes).
        - **Continuous Data:** Can take any value within a given range (e.g., height, weight, temperature).
- **Methods of Data Collection:**
    - Questionnaires
    - Interviews
    - Observations
    - Experiments
    - Existing records
- **Organization of Data:**
    - **Raw Data:** Data as it is collected, before any organization.

    - **Frequency Distribution Table:** A table that lists categories or classes of scores and the number of times each occurs (frequency).
        - **Ungrouped Frequency Distribution:** For discrete data with a small range.
        - **Grouped Frequency Distribution:** For continuous data or discrete data with a large range. Involves creating class intervals.
            - **Class Interval:** A range of values within which data points fall.
            - **Class Limit:** The upper and lower bounds of a class interval.
            - **Class Boundary:** The true limits of a class interval, used to ensure continuity (e.g., for 10-19, class boundaries are 9.5-19.5).
            - **Class Mark (Midpoint):** The midpoint of a class interval (Lower limit + Upper limit) / 2.
            - **Cumulative Frequency:** The running total of frequencies.
            - **Relative Frequency:** The proportion of observations falling into a given category or interval (frequency / total frequency).
- **Graphical Representation of Data:** Visual methods to display data, making it easier to interpret trends and patterns.
    - **Bar Charts:** Used for discrete or categorical data. Bars are separated.
        - Simple Bar Charts
        - Compound Bar Charts (for comparing multiple categories)
    - **Histograms:** Used for continuous data. Bars are contiguous (touching). The area of each bar is proportional to the frequency.
        - Frequency density = Frequency / Class width (for unequal class widths).
    - **Pie Charts:** Used to show parts of a whole. The area of each sector is proportional to the frequency.
        - Angle of sector = (Frequency / Total frequency) × 360^{0}.
    - **Frequency Polygon:** Formed by connecting the midpoints of the tops of the bars in a histogram. Can also be drawn from a frequency table using class marks.
    - **Ogive (Cumulative Frequency Curve):** A graph of cumulative frequency against the upper class boundaries. Used to find medians, quartiles, percentiles, etc.
    - **Pictograms:** Uses symbols or pictures to represent data. Each symbol represents a certain quantity.

2. Measure of Location (Measures of Central Tendency)

These statistics describe the "center" or "typical" value of a dataset.

- **Mean ($\bar{x}$):** The sum of all values divided by the number of values.
  - **For Ungrouped Data:** $\bar{x} = \sum x/n$
  - **For Grouped Data:** $\bar{x} = \sum fx/\sum f$ (where 'x' is the class mark/midpoint for grouped data).
  - **Assumed Mean Method:** A quicker way to calculate the mean for large datasets, especially grouped data.
- **Median:** The middle value of a dataset when it is ordered from least to greatest.
  - **For Ungrouped Data:**
    - If 'n' is odd, median is the $((n+1)/2)^{th}$ term.
    - If 'n' is even, median is the average of the $(n/2)^{th}$ and $((n/2)+1)^{th}$ terms.
  - **For Grouped Data:** Found using the Ogive or the formula:
    Median $= L + [(N/2 - CF_b)/f_m] \times c$
    where:
    - L = Lower class boundary of the median class.
    - N = Total frequency ($\sum f$).
    - CF_b = Cumulative frequency of the class before the median class.
    - f_m = Frequency of the median class.
    - c = Class width of the median class.

- **Mode:** The value that appears most frequently in a dataset.
    - **For Ungrouped Data:** The value with the highest frequency. A dataset can have no mode, one mode (unimodal), or more than one mode (multimodal).
    - **For Grouped Data:** The class interval with the highest frequency (modal class). An approximate mode can be calculated using a formula or found from a histogram.
- **Comparison of Mean, Median, and Mode:**
    - **Mean:** Most common, sensitive to outliers.
    - **Median:** Less affected by outliers, good for skewed distributions.
    - **Mode:** Useful for categorical data, indicates the most popular item.

3. Measure of Dispersion (Measures of Spread)

These statistics describe how spread out or varied the data points are.

- **Range:** The difference between the highest and lowest values in a dataset.
    - Range = Maximum value - Minimum value.
    - Simple but highly affected by outliers.
- **Interquartile Range (IQR):** The difference between the upper quartile (Q3) and the lower quartile (Q1).
    - IQR = Q3 - Q1.
    - **Quartiles:** Divide the data into four equal parts.
        - Q1 (Lower Quartile): The median of the lower half of the data (25th percentile).
        - Q2 (Median): The median of the entire data (50th percentile).
        - Q3 (Upper Quartile): The median of the upper half of the data (75th percentile).
    - **Finding Quartiles from Ogive or Formulas (similar to median formula).**
    - Less affected by outliers than the range.

- **Mean Deviation:** The average of the absolute differences between each data point and the mean (or median/mode).
  - **For Ungrouped Data:** Mean Deviation = $\sum |x - \bar{x}|/n$
  - **For Grouped Data:** Mean Deviation = $\sum f|x - \bar{x}|/\sum f$
- **Variance (σ^2 or s^2):** The average of the squared differences from the mean.
  - **For Ungrouped Data:** $\sigma^2 = \sum (x - \bar{x})^2/n$ or $\sum x^2/n - (\bar{x})^2$
  - **For Grouped Data:** $\sigma^2 = \sum f(x - \bar{x})^2/\sum f$ or $\sum fx^2/\sum f - (\bar{x})^2$
- **Standard Deviation (σ or s):** The square root of the variance. It's the most common measure of spread and has the same units as the data.
  - $$\sigma = \sqrt{\text{Variance}}$$
  - A smaller standard deviation indicates data points are closer to the mean, while a larger one indicates they are more spread out.

1. Permutation and Combination

- **Factorial (!):** The product of all positive integers less than or equal to a given positive integer.
  - $n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$.
  - $0! = 1$.
- **Permutation:** The number of ways to arrange a set of items where the order of arrangement matters.
  - **Permutation of n distinct items taken r at a time:** $P(n, r) = n!/(n-r)!$
  - **Permutation of n items with repetitions:** $n!/(n_1!n_2!\ldots n_k!)$ (where $n_1, n_2, \ldots, n_k$ are the counts of repeated items).
  - **Circular Permutations:** $(n-1)!$ for distinct items.
- **Combination:** The number of ways to select a set of items from a larger set where the order of selection does not matter.
  - **Combination of n distinct items taken r at a time:** $C(n, r) = n!/(r!(n-r)!)$
  - Also written as $\binom{n}{r}$.
- **Relationship between Permutation and Combination:** $P(n, r) = r! \times C(n, r)$.
- **Problem Solving involving Permutations and Combinations:** Differentiating when order matters (permutation) and when it does not (combination).

2. Probability

- **Definition:** The likelihood or chance of an event occurring.
- **Basic Terms:**
    - **Experiment:** A process with a well-defined set of possible outcomes.
    - **Outcome:** A single possible result of an experiment.
    - **Sample Space (S):** The set of all possible outcomes of an experiment.
    - **Event (E):** A subset of the sample space (a specific outcome or set of outcomes).
    - **Favorable Outcome:** An outcome that satisfies the condition of a particular event.
- **Classical (Theoretical) Probability:**

- Classical (Theoretical) Probability:
  - $P(E) = \text{Number of favorable outcomes/Tota}$
  - $0 \leq P(E) \leq 1$.
  - $P(\text{certain event}) = 1$.
  - $P(\text{impossible event}) = 0$.
- **Empirical (Experimental) Probability:** Based on observed frequencies from experiments.
  - $P(E) = \text{Number of times event E occurs/To}$
- Types of Events:
  - **Mutually Exclusive Events:** Events that cannot occur at the same time ($P(A \cap B) = 0$).
    - Addition Rule for Mutually Exclusive Events: $P(A \cup B) = P(A) + P(B)$.
  - **Non-Mutually Exclusive Events:** Events that can occur at the same time.
    - General Addition Rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

    - **Independent Events:** The occurrence of one event does not affect the probability of the other event.
        - Multiplication Rule for Independent Events: $P(A \cap B) = P(A) \times P(B)$.
    - **Dependent Events:** The occurrence of one event affects the probability of the other event.
        - Conditional Probability: $P(A|B) = P(A \cap B)/P(B)$ (Probability of A given B has occurred).
        - Multiplication Rule for Dependent Events: $P(A \cap B) = P(A) \times P(B|A)$.
- **Complement of an Event:** The event that A does not occur, denoted as A'.
    - $P(A') = 1 - P(A)$.
- **Tree Diagrams:** Used to visualize and calculate probabilities for sequential events.