

IB DP MATHEMATICS

AA Topic 4 · Statistics and Probability

TOPICAL WORKSHEET

4.1 Descriptive Statistics · 4.2 Correlation & Regression
4.3 Probability · 4.4 Probability Distributions

Standard and Higher Level · Syllabus 2025

Fahad H. Ahmad

+92 323 509 4443 | Megalecture.com

Section A — Multiple Choice (15 marks)

Circle the letter of the best answer to each question. Each question is worth 1 mark.

1. For a dataset, the interquartile range (IQR) is calculated as [1]
 - A $Q3 + Q1$
 - B $Q3 - Q1$
 - C maximum – minimum
 - D $(Q1+Q3) \div 2$

2. In a box-and-whisker plot, the line inside the box represents the [1]
 - A mean
 - B median
 - C mode
 - D range

3. A dataset has mean 20 and standard deviation 4. A value of 28 is how many standard deviations above the mean? [1]
 - A 1
 - B 2
 - C 4
 - D 8

4. An outlier in a dataset is most appropriately identified using [1]
 - A the mean and standard deviation only
 - B a value more than $1.5 \times \text{IQR}$ beyond $Q1$ or $Q3$
 - C any value greater than the median
 - D the mode of the dataset

5. Pearson's product-moment correlation coefficient, r , always lies in the range [1]
 - A $0 \leq r \leq 1$
 - B $-1 \leq r \leq 1$
 - C $-\infty \leq r \leq \infty$
 - D $0 \leq r \leq 100$

6. A correlation coefficient of $r = -0.95$ indicates [1]
 - A a strong positive linear correlation
 - B a strong negative linear correlation
 - C no linear correlation
 - D a weak positive linear correlation

- 7.** The line of best fit (linear regression line) for a set of bivariate data is generally used to [1]
 A show correlation is the same as causation
 B make predictions for values within or near the range of the data
 C calculate the mode of the dataset
 D guarantee an exact value for any prediction
- 8.** If two variables have a correlation coefficient of $r = 0.05$, this suggests [1]
 A a strong positive correlation
 B a strong negative correlation
 C essentially no linear correlation
 D the variables are causally related
- 9.** For any two events A and B, the addition rule of probability states [1]
 A $P(A \cup B) = P(A) + P(B)$
 B $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 C $P(A \cup B) = P(A) \times P(B)$
 D $P(A \cup B) = P(A \cap B)$
- 10.** Two events A and B are independent if and only if [1]
 A $P(A \cap B) = 0$
 B $P(A \cap B) = P(A) + P(B)$
 C $P(A \cap B) = P(A) \times P(B)$
 D $P(A) = P(B)$
- 11.** The conditional probability of A given B, $P(A|B)$, is defined as [1]
 A $P(A \cap B) \div P(B)$
 B $P(A) \div P(B)$
 C $P(A \cap B) \times P(B)$
 D $P(B) \div P(A \cap B)$
- 12.** In a Venn diagram showing two overlapping events A and B, the region where the circles overlap represents [1]
 A $A \cup B$
 B $A \cap B$
 C neither A nor B
 D A only

- 13.** A random variable X follows a binomial distribution $X \sim B(n, p)$ when **[1]**
- A there are a fixed number of independent trials, each with two outcomes and a constant probability of success
 - B the data is continuous and symmetric about the mean
 - C there are an infinite number of possible outcomes
 - D the probability of success changes with each trial
- 14.** For $X \sim B(n, p)$, the mean (expected value) of X is given by **[1]**
- A $E(X) = n + p$
 - B $E(X) = np$
 - C $E(X) = np(1-p)$
 - D $E(X) = p \div n$
- 15.** For a normally distributed random variable, approximately what percentage of values lie within ± 1 standard deviation of the mean? **[1]**
- A 50%
 - B 68%
 - C 95%
 - D 99.7%

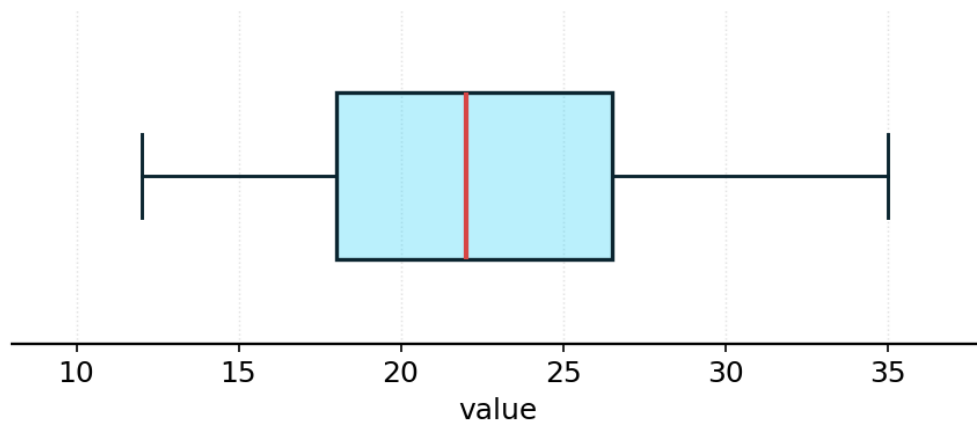
Section B — Structured Questions (65 marks)

Answer all questions in the space provided. Working must be shown for full marks.

4.1 Descriptive Statistics

Question 1 [7 marks]

The box-and-whisker plot shows a dataset of 11 values (computed).



Box-and-whisker plot of an 11-value dataset (computed).

(a) Using the box plot, write down the median, the lower quartile (Q1), and the upper quartile (Q3) of the dataset. [2]

(b) Calculate the interquartile range (IQR) of the dataset. [2]

(c) Using the box plot, write down the minimum and maximum values, and hence calculate the range of the dataset. [3]

Question 2 [6 marks]

Measures of central tendency and spread summarise the key features of a dataset.

(a) State one advantage of using the median (rather than the mean) to describe the centre of a dataset that contains an outlier. [2]

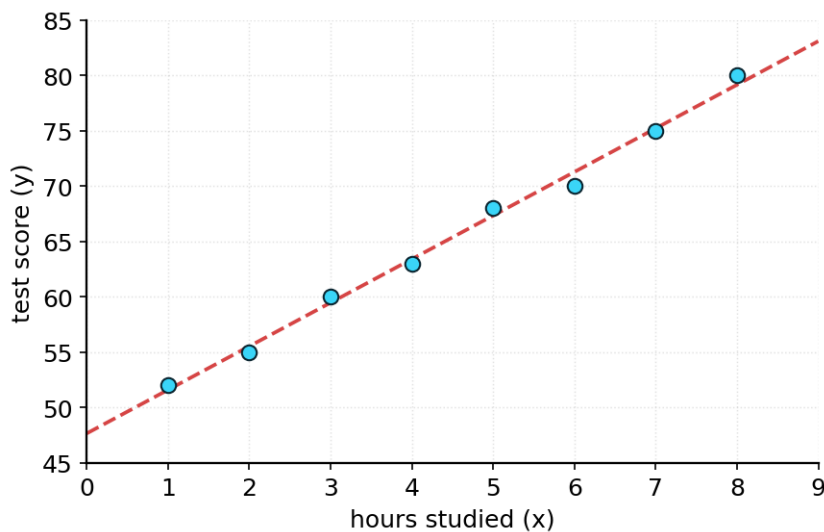
(b) A dataset has mean 45 and standard deviation 6. Calculate how many standard deviations the value 57 is above the mean. [2]

(c) State the rule commonly used to identify an outlier, in terms of $Q1$, $Q3$, and the IQR. [2]

4.2 Correlation & Regression

Question 3 [7 marks]

The scatter diagram shows the number of hours studied (x) and test score (y) for 8 students, with a line of best fit (computed).



Hours studied vs test score, with line of best fit (computed).

(a) Using the scatter diagram, describe the correlation between hours studied and test score. [2]

(b) The value of Pearson's correlation coefficient for this data is $r = 0.997$. Interpret this value. [2]

- (c) The equation of the line of best fit is $y = 3.94x + 47.64$. Use this equation to predict the test score for a student who studies for 9 hours. [3]

Question 4 [6 marks]

Pearson's correlation coefficient, r , quantifies the strength and direction of a linear relationship between two variables.

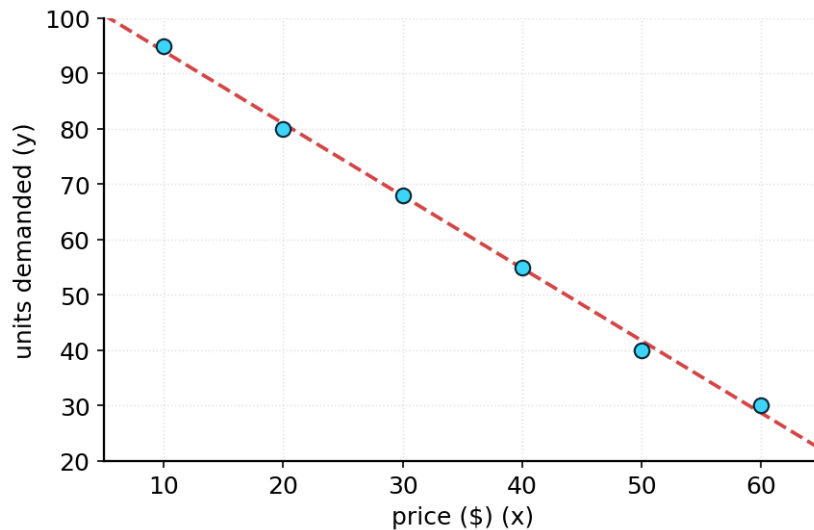
- (a) State the range of possible values of Pearson's correlation coefficient, r , and state what the values $r = -1$, $r = 0$, and $r = 1$ each represent. [2]

- (b) A dataset has a correlation coefficient of $r = 0.85$. Describe the strength and direction of this correlation. [2]

- (c) Explain why a strong correlation between two variables does not necessarily imply that one causes the other. [2]

Question 5 [7 marks]

The scatter diagram shows the price (\$) of a product (x) and the number of units demanded (y), with a line of best fit (computed).



Price vs units demanded, with line of best fit (computed).

- (a) Using the scatter diagram, describe the correlation between price and units demanded. [2]

- (b) The value of Pearson's correlation coefficient for this data is $r = -0.999$. Interpret this value. [2]

- (c) State the sign of the gradient of the line of best fit shown, and explain how this relates to the value of r found in part (b). [3]

4.3 Probability

Question 6 [6 marks]

Two events, A and B, have $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.2$.

- (a) State the addition rule of probability for $P(A \cup B)$. [2]

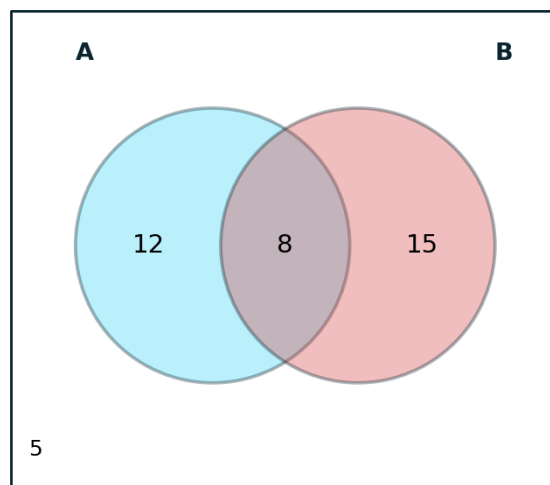
- (b) Calculate $P(A \cup B)$. [2]

(c) Determine, showing your working, whether events A and B are independent.

[2]

Question 7 [7 marks]

The Venn diagram shows the number of students (out of 40) who study French (A) and/or Spanish (B) (given data).



Venn diagram showing the number of students studying French (A) and/or Spanish (B) (given data).

(a) Using the Venn diagram, calculate $P(A)$, the probability that a randomly selected student studies French.

[2]

(b) Calculate $P(A \cap B)$, the probability that a randomly selected student studies both French and Spanish.

[2]

(c) Calculate $P(A \cup B)$, the probability that a randomly selected student studies French or Spanish (or both).

[3]

Question 8 [6 marks]

Using the same events A and B as in Question 6, where $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.2$.

(a) State the formula for the conditional probability $P(A|B)$.

[2]

(b) Calculate $P(A|B)$.

[2]

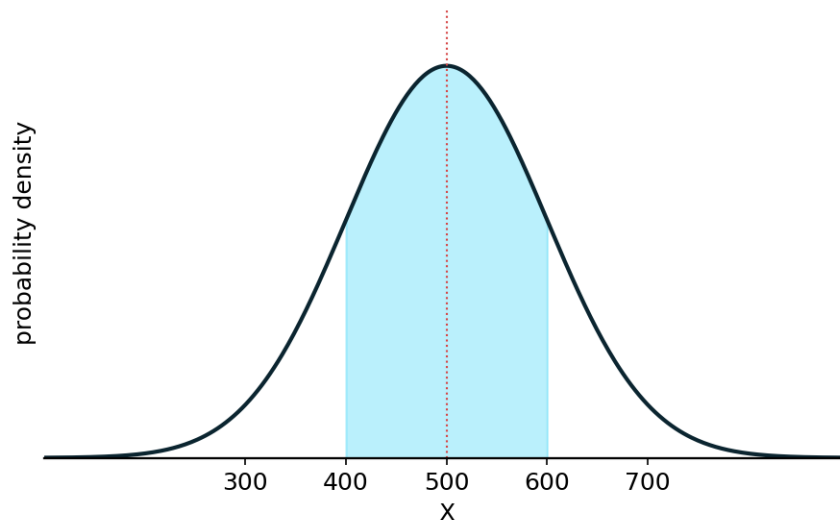
(c) Compare your answer to part (b) with $P(A)$, and explain what this confirms about events A and B.

[2]

4.4 Probability Distributions

Question 9 [7 marks]

The graph shows a normal distribution with mean $\mu = 500$ and standard deviation $\sigma = 100$, with the region between 400 and 600 shaded (computed).



Normal distribution, $\mu=500$, $\sigma=100$, with $P(400 < X < 600)$ shaded (computed).

(a) Using the graph, state the mean and standard deviation of this distribution, and state how the shaded region relates to μ and σ .

[2]

(b) Using the empirical rule (or z-scores), estimate $P(400 < X < 600)$, giving your answer to 3 significant figures. [2]

(c) Calculate $P(X > 600)$, giving your answer to 3 significant figures. [3]

Question 10 [6 marks]

A random variable X follows a binomial distribution, $X \sim B(10, 0.3)$.

(a) State three conditions required for a random variable to follow a binomial distribution. [2]

(b) Calculate the mean and the variance of X . [2]

(c) Calculate $P(X = 3)$, giving your answer to 3 significant figures. [2]