

Statistics and probability

IB Maths AA HL · Knowledge organiser

KEY DEFINITIONS

Bayes' theorem

Reverses a conditional probability: $P(B | A) = \frac{P(B)P(A | B)}{P(A)}$.

Probability density function

For a continuous variable, $f(x) \geq 0$ and the total area under f is 1.

Expected value

$E(X)$ is the long-run mean of X : $\sum x P(X = x)$, or $\int x f(x) dx$ for a continuous variable.

Variance

$\text{Var}(X) = E(X^2) - [E(X)]^2$ measures spread around the mean.

YOU SHOULD BE ABLE TO...

- Distinguish populations from samples and discrete from continuous data, compare simple random, systematic, stratified, quota and convenience sampling, and identify sources of bias.
- Find least-squares regression lines of y on x and x on y with technology, choose the right one for a prediction and explain why correlation is not causation.
- Find the variance of a discrete random variable, model situations with $B(n, p)$ and find binomial probabilities, mean and variance with technology.
- Standardise normal values with $z = (x - \mu) / \sigma$ and use the inverse normal to find an unknown mean or standard deviation from given probabilities.
- Work with piecewise probability density functions and use $E(aX + b)$ and $\text{Var}(aX + b)$ to find the mean and variance of transformed variables.
- Use Bayes' theorem with tree diagrams in contexts such as medical testing and interpret the result.

COMMON MISTAKES

- Continuous random variables: mode, median and expectation from a pdf
- Discrete probability: independence assumed, $E(X^2)$ and Var misused
- Using the class boundary instead of the midpoint when estimating the mean from grouped data
- Plotting cumulative frequency at class midpoints instead of upper class boundaries

KEY FORMULAS

B = given in the exam (formula booklet or list); learn the rest.

IQR; mean B	$IQR = Q_3 - Q_1, \quad \bar{x} = \frac{\sum f_i x_i}{n}$
Outliers	$\backslash \text{ or } > Q_3 + 1.5 IQR$
Variance	$\sigma^2 = \frac{\sum f(x - \mu)^2}{n} = \frac{\sum f x^2}{n} - \mu^2$
Probability B	$P(A) = \frac{n(A)}{n(U)}, \quad P(A) + P(A') = 1$
Combined events B	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Independent B	$P(A \cap B) = P(A)P(B)$
Conditional B	$P(A B) = \frac{P(A \cap B)}{P(B)}$
Bayes B	$P(B A) = \frac{P(B)P(A B)}{P(B)P(A B) + P(B')P(A B')}$
Bayes, partition B_1, B_2, B_3 B	$P(B_i A) = \frac{P(B_i)P(A B_i)}{\sum_j P(B_j)P(A B_j)}$
Discrete X B	$E(X) = \sum x P(X = x),$ $\text{Var}(X) = E(X^2) - [E(X)]^2$
Binomial pmf	$P(X = r) = \binom{n}{r} p^r (1 - p)^{n-r}$
Binomial mean, variance B	$np, \quad np(1 - p)$
Standardised normal B	$z = \frac{x - \mu}{\sigma}$
Continuous X B	$E(X) = \int x f(x) dx, \quad \text{Var}(X) = \int x^2 f(x) dx - \mu^2$
pdf; median m	$\int f(x) dx = 1; \quad \int_{-\infty}^m f(x) dx = \frac{1}{2}; \quad \text{mode: max of } f$
Linear transformation B	$E(aX + b) = aE(X) + b,$ $\text{Var}(aX + b) = a^2 \text{Var}(X)$

More on the full formula sheet.

WORKED EXAMPLE

X has probability density function $f(x) = kx$ for $0 \leq x \leq 2$. Find k and $E(X)$.

- $\int_0^2 kx dx = 2k = 1$, so $k = \frac{1}{2}$
- $E(X) = \int_0^2 x \times \frac{1}{2}x dx = [x^3/6]_0^2 = 8/6$

Answer: $k = \frac{1}{2}$ and $E(X) = 4/3$