

Statistics and probability

IB Maths AA SL · Knowledge organiser

KEY DEFINITIONS

Independent events

A and B are independent when $P(A \cap B) = P(A) \times P(B)$: one happening does not change the chance of the other.

Conditional probability

$P(A | B)$ is the probability of A given that B has happened: $P(A \cap B) / P(B)$.

Discrete random variable

A variable that takes separate values, each with a probability; the probabilities add to 1.

Normal distribution

A symmetric, bell-shaped distribution set by its mean μ and standard deviation σ .

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.
- Draw a line of best fit through the mean point, find the regression line of y on x with technology and use it to predict within the data range.
- Use tree diagrams and sample space diagrams to find probabilities of combined events, with and without replacement, using $P(A \cup B)$ for overlapping events.
- Find normal probabilities with technology and use the inverse normal to find the value with a given probability below it, such as the top 5% boundary.
- Use $P(A | B) = P(A \cap B)/P(B)$ and test whether events are independent by checking $P(A \cap B) = P(A)P(B)$.
- Combine standardisation and the binomial distribution in multi-part questions, such as the number of items outside a tolerance.

COMMON MISTAKES

- ✗ Probability: assuming independence, and conditional probability with complements
- ✗ Binomial and normal: "at least", inequalities, rounding bounds and which distribution
- ✗ 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.

Interquartile range **B** $IQR = Q_3 - Q_1$

Mean ($n = \sum f_i$) **B** $\bar{x} = \frac{\sum f_i x_i}{n}$

Outliers $\backslash (x \text{ or } x > Q_3 + 1.5 IQR$

Variance (s.d. σ) $\sigma^2 = \frac{\sum f(x-\mu)^2}{n} = \frac{\sum f x^2}{n} - \mu^2$

Data $\times a$ then $+b$ mean $\rightarrow a\bar{x} + b$, s.d. $\rightarrow |a|\sigma$

Regression: use y on x to predict y, x on y to predict x; interpolate, don't extrapolate

Probability **B** $P(A) = \frac{n(A)}{n(U)}$, $P(A) + P(A') = 1$

Combined events **B** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Mutually exclusive **B** $P(A \cup B) = P(A) + P(B)$

Independent **B** $P(A \cap B) = P(A)P(B)$

Conditional **B** $P(A | B) = \frac{P(A \cap B)}{P(B)}$

Expected value **B** $E(X) = \sum x P(X = x)$

Valid distribution; fair game $\sum P(X = x) = 1$; $E(\text{gain}) = 0$

Binomial $X \sim B(n, p)$ $P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$

Binomial mean, variance **B** $E(X) = np$, $\text{Var}(X) = np(1-p)$

Standardised normal **B** $z = \frac{x-\mu}{\sigma}$

More on the full formula sheet.

WORKED EXAMPLE

$X \sim B(10, 0.3)$. Find $P(X = 2)$.

1. $P(X = 2) = {}^{10}C_2 \times 0.3^2 \times 0.7^8$

2. $= 45 \times 0.09 \times 0.0576...$

Answer: $P(X = 2) = 0.233$ (3 s.f.)