Statistics and probability

International Baccalaureate · IB Diploma · Mathematics: Applications and Interpretation · SL

Scope and prerequisites

This handout covers selected SL topics for the course first assessed in 2021; it is not the whole syllabus.

Prerequisites: read the stated quantities and units, and use arithmetic and the model conditions below.

The examples are original and fictional: they are not real measurements or real student investigations.

Data summaries, histograms and interpretation

Can one average tell the whole story?

- Two groups have the same median but different spread. One summary cannot describe both location and consistency.
- This lesson studies **frequency density** : Frequency divided by class width, used as histogram height.

The method

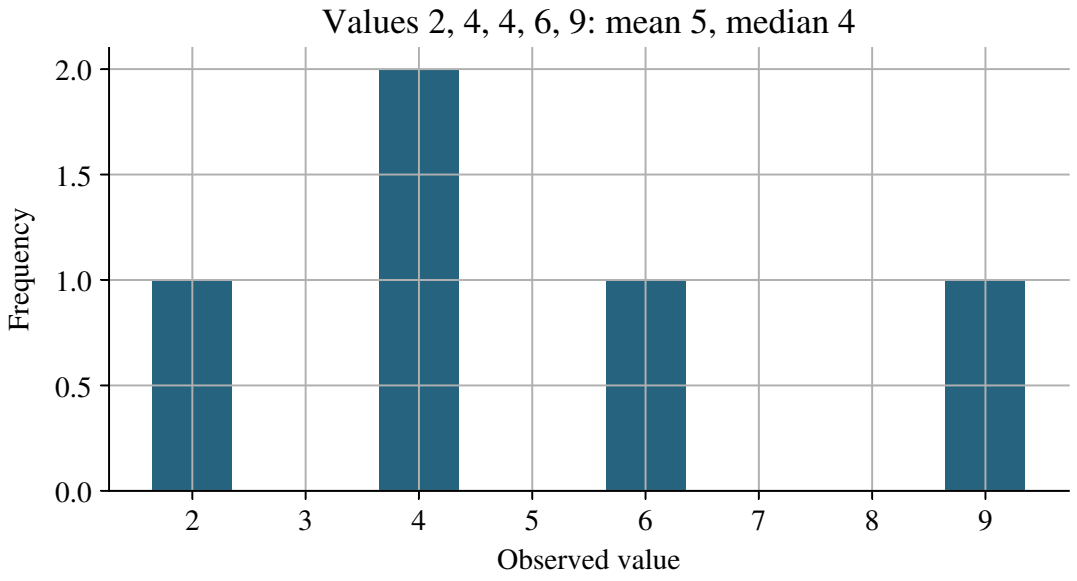
- Compare an appropriate average and spread in context.
- A histogram uses area for frequency, so height=frequency/class width.
- Grouped estimates assume representative values within intervals.

A worked case

$$\overline{x} = \frac{\sum x_i}{n}, \qquad \text{density} = \frac{\text{frequency}}{\text{class width}}$$

Example:

A class from 10 to 20 with frequency 30 has density 30/10=3. A class from 20 to 40 with frequency 20 has density 20/20=1. Its wider bar must not be mistaken for a larger density. For values 2,4,4,6,9, the median is 4 and mean is 5.



For 2, 4, 4, 6, 9: mean = 25 ÷ 5 = 5, median = 4

■ Where it goes wrong

- The tallest histogram bar need not contain the most observations.
- A grouped mean is an estimate.
- Correlation does not prove causation, and extrapolation extends beyond the observed range.

Not true: A histogram bar's height always equals its frequency, even with unequal class widths. With unequal class widths the height is frequency density; frequency is the area of the bar.

■ Applying it

- Choose a display that fits the data type.
- Give both a numerical comparison and what it means for the population; do not infer more precision than the sample supports.

■ In your course

- Applications and Interpretation SL, the course first assessed in 2021.
- Every paper allows a calculator, but method marks still need the expression or equation written before its value.

Key:

- **Frequency density**: Frequency divided by class width, used as histogram height.
- Compare an appropriate average and spread in context.
- Watch for: the tallest histogram bar need not contain the most observations.

Write from a mean of 8

- A class of width 3 has frequency 18.
- The values 4, 4, 7, 9 and 16 have a mean and a median. Bar height is not always the frequency.

Probability, trees and conditional reasoning

What changes after the first draw?

- A bag has 3 red and 2 blue counters. Taking two without replacement changes the chance of the second colour.
- This lesson studies **conditional probability** : The probability of an event after restricting the sample space to a stated condition.

The method

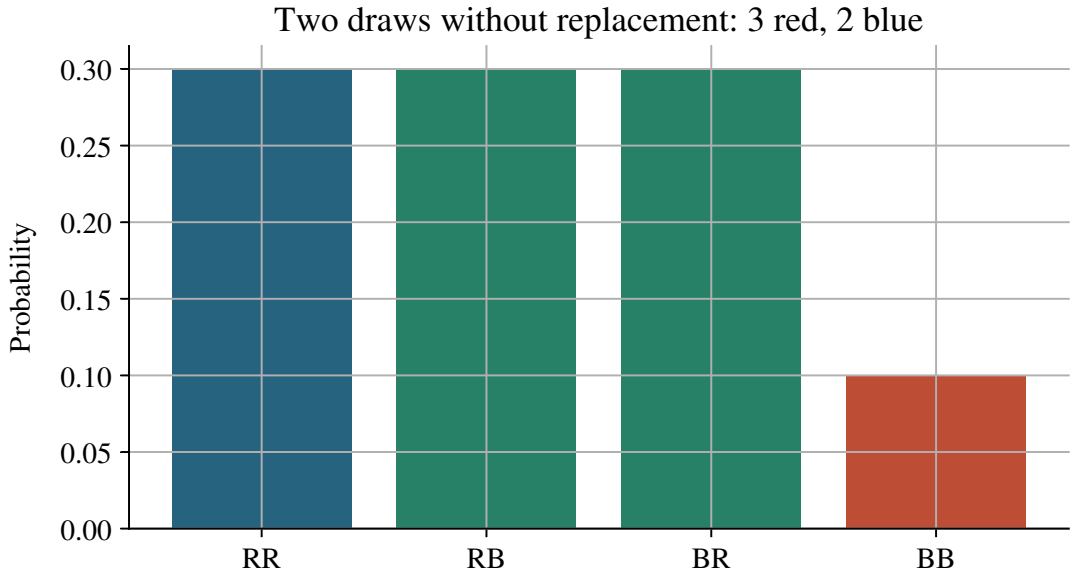
- Multiply along a tree branch and add disjoint branches.
- With replacement, the composition stays fixed.
- Conditional probability is P(A given B)=P(A∩B)/P(B), for P(B)>0.

A worked case

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0$$

Example:

Without replacement, P(two red)=3/5×2/4=3/10. P(one of each)=3/5×2/4+2/5×3/4=3/5. If P(A∩B)=0.12 and P(B)=0.3, P(A given B)=0.4.



Without replacement: P(RR) = 3/5 × 2/4 = 0.3 and P(BB) = 2/5 × 1/4 = 0.1

Where it goes wrong

- Mutually exclusive means no overlap; independent means that knowing one event does not change the other's probability.
- Two disjoint events with positive probability are not independent.

Not true: Mutually exclusive events with positive probabilities must be independent. Mutually exclusive events with positive probabilities are dependent: if one happens, the other cannot, so P(A and B)=0≠P(A)P(B).

Applying it

- A two-way table makes the restricted denominator visible.
- Before using a product P(A)P(B), justify independence from the context or the supplied information.

In your course

- Applications and Interpretation SL, the course first assessed in 2021.
- Every paper allows a calculator, but method marks still need the expression or equation written before its value.

Key:

- **Conditional probability**: The probability of an event after restricting the sample space to a stated condition.
- Multiply along a tree branch and add disjoint branches.
- Watch for: mutually exclusive means no overlap; independent means that knowing one event does not change the other's probability.

■ Write from a conditional probability of 0.70

- P(A and B) is 0.28 and P(B) is 0.40.
- Find P(A given B). Mutually exclusive events that can happen are not independent.

Binomial, normal and Poisson models

■ What makes a count predictable?

- A quality inspector counts defective items. The number is random, but a model can describe its likely range.
- This lesson studies **expected value** : The probability-weighted mean of a random variable.

■ The method

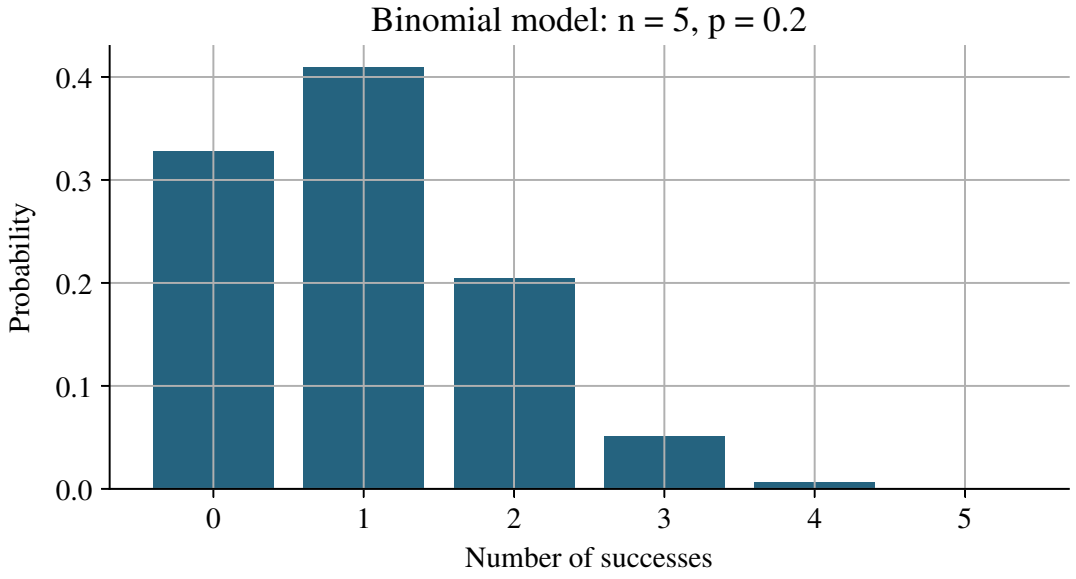
- A binomial model needs fixed n, independent trials, two outcomes and constant p.
- E(X)=np and Var(X)=np(1-p).
- For a normal model use z=(x-μ)/σ.
- A Poisson model describes counts with a constant rate and appropriate independence assumptions.

■ A worked case

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

Example:

For X binomial(5,0.2), P(X=0)=0.8^5=0.32768, E(X)=1 and Var(X)=0.8. For a normal quantity with μ=100,σ=15, the value 130 has z=2. A Poisson mean of 3 per hour gives mean 6 over two hours.



$X \sim B(5,\ 0.2)$: $P(X = 1) \approx 0.41$ is the largest bar; $P(X = 5)$ is tiny

---

Where it goes wrong

- Not every count is binomial: changing p or dependence can invalidate it.
- For a continuous variable, the probability of one exact value is zero.
- Continuity correction matters when approximating a discrete distribution by a normal one.

Not true: Every count of successes has a binomial distribution regardless of dependence. The binomial model needs independent trials with a fixed probability; dependent trials break it.

Applying it

- Write the event as an inequality before using calculator distribution functions.
- Distinguish $P(X<k)$, $P(X\leq k)$ and a tail complement.
- State assumptions in context.

In your course

- Applications and Interpretation SL, the course first assessed in 2021.

Key:

- **Expected value**: The probability-weighted mean of a random variable.
- A binomial model needs fixed n, independent trials, two outcomes and constant p.

- Watch for: not every count is binomial: changing p or dependence can invalidate it.

■ Write from a z-score of 4

- A value 90 comes from a distribution with mean 70 and standard deviation 5.
- Find the z-score. A binomial count needs independent trials.

Hypothesis testing and contextual conclusions

■ Could chance explain the result?

- A factory claims that only 10% of items are defective. A sample contains more defects, but chance alone may explain some difference.
- This lesson studies **significance level** : The chosen probability threshold for rejecting a null hypothesis.

■ The method

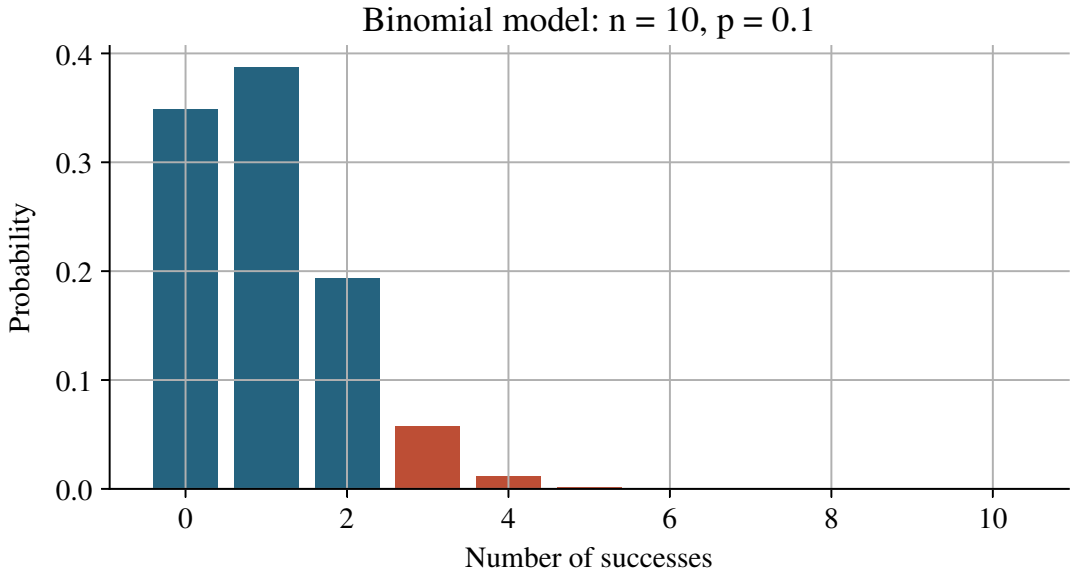
- State H_0 and H_1 in population parameters before inspecting the outcome.
- Calculate the appropriate tail probability under H_0.
- Reject H_0 when the evidence meets the specified significance rule; otherwise say there is insufficient evidence.

■ A worked case

$$H_0 : p = p_0, \qquad H_1 : p > p_0$$

Example:

For H_0:p=0.1 and H_1:p>0.1 with n=10, observing at least 3 defects has probability 1-(0.9^10+10×0.1×0.9^9+45×0.1^2×0.9^8)≈0.070191. At 5%, this is insufficient evidence that the defect rate exceeds 10%.



Under B(10, 0.1), $P(X \geq 3) \approx 0.07$: the shaded upper tail is the evidence region for the test

---

■ Where it goes wrong

- Failing to reject H_0 is not proof that H_0 is true.
- Choose the tail from H_1, not from whichever tail gives a small result.
- Statistical significance does not measure the practical size of an effect.

Not true: Failing to reject a null hypothesis proves it is true. Not rejecting the null hypothesis only means there is not enough evidence against it.

■ Applying it

- Finish with a sentence about the population and the original claim.
- State the model's assumptions and consider whether the sampling procedure supports them.

■ In your course

- Applications and Interpretation SL, the course first assessed in 2021.
- Every paper allows a calculator, but method marks still need the expression or equation written before its value.

Key:

- **Significance level**: The chosen probability threshold for rejecting a null hypothesis.
- State H_0 and H_1 in population parameters before inspecting the outcome.
- Watch for: failing to reject H_0 is not proof that H_0 is true.

Write from a p-value of 0.08

- A one-tailed p-value is 0.08 and the significance level is 5 percent.
- Decide whether to reject the null hypothesis. Failing to reject does not prove it true.

Write from a p-value of 0.03

- A one-tailed p-value is 0.03 against a 5 percent significance level.
- 0.03 is below 0.05, so reject H_0 – the result is significant at the chosen level.

Write from a two-tailed p-value of 0.08

- A two-tailed p-value of 0.08 is tested at the 10 percent level.
- Reject at 10 percent; the identical result at 5 percent would not – the threshold is chosen BEFORE the data are seen.