CIE Biology International A-level

Analysis and Interpretation of Data

Practical Notes

Analyse and interpret data to reach conclusions

Processing results

Data can be processed in different ways depending on the aim of the practical. Simple processing methods include calculating **means**, **changes**, and **rates**.

Means are calculated when the practical is repeated or multiple samples are taken. Taking a mean value of many data points **allows the variability** of the results to be assessed. Using means, **standard deviation** can also be calculated, which is a measure of the **spread of data**. **Standard error**, which is a measure of the **reliability** of the **mean** i.e. how close the sample mean is to the real mean, can be calculated from standard deviation. The formula for standard error is:

$$\text{Standard deviation} / \sqrt{n}$$

Standard error can be presented on **graphs** in **error bars**, where the range is + and - 2x standard error. Generally, if the error bars on two values do not overlap, the values are significantly different.

Change may be calculated for change of mass, length or temperature eg. in osmosis-related practicals. **Percentage change** may sometimes be required, as this allows comparison when the starting point is different. The formula of percentage change is:

$$(\text{final value} - \text{initial value}) / \text{initial value} \times 100\%$$

Rates can be **directly** or **inversely proportional** to the results collected. If time to complete a certain reaction is measured, often in the case of enzyme reactions, rate is calculated by **1/time**. However, rate can also be indicated by the release of a product, eg. volume of gas produced in a fixed time, then rate is directly proportional to the volume and may not require further processing.

For certain practicals (often involving field investigations and genetic crosses), statistical tests will have to be used to analyse the significance of the results. There are four statistical tests that students are expected to use:

- Spearman's rank coefficient
- Pearson's linear correlation coefficient
- T-test
- Chi-squared test

The first step to performing a statistical test is to write a **null hypothesis**. A null hypothesis states that there is **no significant correlation** or **difference** between the two data sets students have collected. Next, whether the data sets are being **compared** or **correlated**, and the nature of the data - whether it is **continuous** or **discontinuous**. Select the appropriate statistical test using the table below.

Statistical test	Purpose	Type of data	Degrees of Freedom
T-test	Compare, find if difference is significant	Continuous, normally distributed eg. means	$(n_1 - 1) + (n_2 - 1)$
Chi-squared	Compare observable and expected values, find if difference is significant	Frequency, discrete and categorical	No. of categories - 1
Spearman's rank	Correlate	Ordinal data	No. of pairs of data - 2
Pearson's linear correlation	Correlate	Continuous, linear, normally distributed	No. of pairs of data - 2

Use the formula provided to calculate a value. Find the critical value in a table provided at **p=0.05** and the calculated **degrees of freedom**, where p=0.05 indicates a **5% probability** that the difference or relationship in the data is **due to chance alone**. If the **calculated value** is **larger** than the **critical value**, the **null hypothesis is rejected**, and there is a smaller than 5% probability that the difference or relationship is due to chance alone.

Another statistical test called **Simpson's Index of Biodiversity** is specific to field investigations on **biodiversity**. It gives a value between **0 and 1**, where values closer to 1 have higher biodiversity. The formula may or may not be provided in the exam, hence it should be memorised.

$$D = 1-(\Sigma\ (n/N)^2)$$

Patterns and trends

To describe patterns and trends, first describe the **overall trend**, **quoting data** to support your observations. Afterwards, move onto describing any particular **features of the graph**, eg. peaks and troughs, or data points that do not fit the overall trend.

Students may be asked to **identify anomalies**. Anomalies are data that **do not fit the trend**. When calculating a mean, the anomalies in the data set should be excluded.

Finding unknowns

Certain practicals require students to **estimate an unknown value**, eg. concentration of an unknown solution. This often requires students to produce **a set of standards** to compare against, which may be colour standards, time taken for the reaction of a known solution, or a graph plotted using standard solutions.

To increase the **accuracy** of the estimate, the number of different standards used should be increased at **smaller intervals** around the estimate.