6.3.1 Standard deviation, comparisons and graphs

Name: _____ Class: _____ Date: _____ **Total: 21 marks**

KEY WORDS error bars • null hypothesis

Objective

Process IA2 data: complete a standard-deviation table, compare groups with means and SDs, state a null hypothesis, and draw the line graph with a best-fit line that survives an anomaly. Original practice for the Unit 6 analysis criteria.

1 Key knowledge and worked examples

■ Completing the SD table

The formula is given; the marks are in the table. For masses 18.1, 19.7, 16.2, 13.8, 17.1 g with mean 17.0: each deviation squared gives 1.21, 7.29, 0.64, **10.24**, 0.01 – the blank cell is $(13.8 - 17.0)^2 = 10.24$ – summing to 19.39:

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{19.39}{4}} = 2.2 \text{ g}$$

Round with the data: one decimal place matches the measurements.

■ Comparing two groups

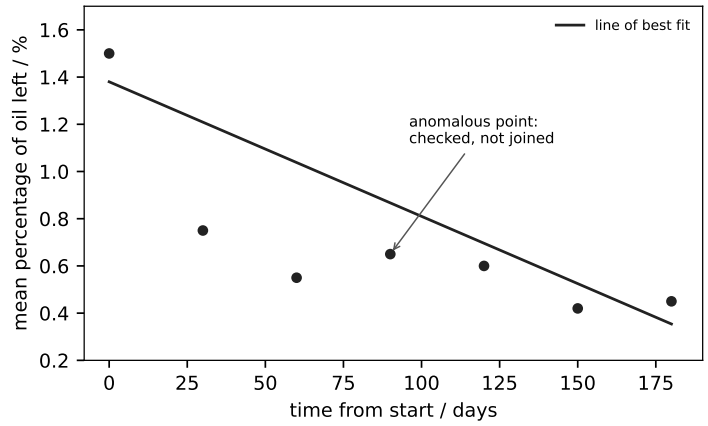
Plot each mean with its SD as an **error bar** . A 5-year plantation at 12.8 ± 2.0 g and a 15-year at 17.0 ± 2.2 g: the bars do not overlap, so the difference is likely significant; the t-test decides formally. Overlapping bars – no secure difference.

■ Null hypothesis before the test

"There is **no correlation** between the percentage of oil remaining and the time of storage" – stated before calculating; the test either rejects it or fails to reject it. The conclusion sentence names both variables.

■ The graph with an anomaly

Time (days) is continuous → line graph, axes labelled with units, linear scales, all points plotted, one best-fit line. The 90-day point (0.65 %) breaks the trend: anomalous. The best fit ignores it; check by repeating that measurement – never delete silently and never bend the line to visit it.



2 Practice

2.1 Five sites give masses 16.0, 18.5, 17.2, 15.9, 19.4 g. Calculate the mean and $\sum(X - \bar{X})^2$. [3]

2.2 With $\sum(X - \bar{X})^2 = 12.70$ and n = 5, calculate the standard deviation. [2]

2.3 Group A: 12.8 ± 2.0 g; Group B: 14.5 ± 2.6 g (mean ± SD). What can and cannot be concluded from these figures alone? [2]

2.4 Write a null hypothesis for an investigation of the effect of soil moisture on the number of earthworms per quadrat. [1]

3 Exam-style questions

3.1 The table shows mean percentage of crude oil left in soil trays: day 0 → 1.50, 30 → 0.75, 60 → 0.55, 90 → 0.65, 120 → 0.60, 150 → 0.42, 180 → 0.45. Describe the graph you would draw, including the axis decision, the treatment of each point, and what the best-fit line shows; identify and deal with any anomaly. [5]

3.2 Calculate the standard deviation of the five earthworm masses 18.1, 19.7, 16.2, 13.8, 17.1 g (mean 17.0 g), showing the missing squared deviation, and describe how the means and standard deviations are used to compare two plantations. [5]

3.3 Explain why a raw-data table must record times (not rates), and how means arise from the repeats, in a recovery-time investigation. [3]

Solutions

Original-task marking: M1 method, A1 accuracy, B1 reasoning.

2.1 Mean $= (16.0 + 18.5 + 17.2 + 15.9 + 19.4)/5 = 87.0/5 = 17.4$ g. Squared deviations: $1.96, 1.21, 0.04, 2.25, 4.00$; $\sum(X - \bar{X})^2 = 9.46$.

2.2 $s = \sqrt{12.70/4} = \sqrt{3.175} = 1.78 \approx 1.8$ g.

2.3 B's mean is higher, and the SD ranges (10.8–14.8 and 11.9–17.1) overlap, so the difference between the means cannot be judged significant from these figures alone; a t-test (or SD inspection without overlap) is needed for a conclusion.

2.4 There is no correlation between soil moisture and the number of earthworms per quadrat.

3.1 Line graph – time is continuous – with days on x and mean % oil left on y, axes labelled with units and linear scales; all seven points plotted correctly; a single best-fit line falling from about 1.4 to about 0.45, not bent through every point; the 90-day value 0.65 sits above the trend – anomalous: the line ignores it and the measurement is checked by repeating that sampling.

3.2 The missing squared deviation is $(13.8 - 17.0)^2 = 10.24$; $\sum(X - \bar{X})^2 = 1.21 + 7.29 + 0.64 + 10.24 + 0.01 = 19.39$; $s = \sqrt{19.39/4} = 2.2$ g. Compare: the means show which plantation holds more earthworm mass on average; SDs show the spread – plotted as error bars or used in a t-test, non-overlap (or a significant t) means the difference between plantations is significant.

3.3 Rates are derived after measurement; the raw table must hold what was actually read – each student's recovery time in seconds with units; means are calculated from repeats within each group after measurement; any rate or summary belongs in the analysis step, not the raw-data record.