NCEA Level 1 · AS 91944: Exploring Data with the Statistical Enquiry Cycle

Name: Date:

Hey! This standard is all about running a proper statistical investigation from start to finish, asking a good question, sorting your data, making sensible graphs, and writing conclusions that the numbers actually support. Work through each section in order, and always back up what you say with evidence from your data.

Revision summary

- The statistical enquiry cycle (PPDAC): Problem (pose a comparison/summary question), Plan (how to get data), Data (collect/clean/source), Analysis (graphs + summary stats), Conclusion (answer the question with evidence).
- A good comparison question names the two groups, the numeric variable, and asks which tends to be bigger, e.g. 'Do Year 11 boys tend to be taller than Year 11 girls at our school?'
- Measures of centre: mean = sum of values ÷ number of values; median = middle value when ordered; mode = most common value.
- Measures of spread: range = max − min; interquartile range IQR = upper quartile (Q3) − lower quartile (Q1). IQR describes the middle 50% and resists outliers.
- Shape: describe as symmetric, right-skewed (tail to high values) or left-skewed (tail to low values); note any clusters, gaps or outliers.
- Best displays: dot plot or box plot for one numeric variable (great for comparing two groups), bar chart for categories, back-to-back or paired displays for comparisons.
- Comparing groups: compare medians (centre), overlap of boxes/IQRs (spread), and shape. If the difference in medians is large relative to the spread, you can make a stronger 'back to the population' claim.
- Conclusions must be evidence-based: quote actual values from your graphs/stats, and only generalise beyond the sample cautiously if the data was collected sensibly.

Key skills

1. The number of texts sent by 9 students at Avondale College in one hour: 3, 5, 5, 8, 10, 12, 14, 18, 22. Find the median and the range.

2. Calculate the mean of these weekend earnings (in $): 40, 55, 60, 60, 85.

3. For the data 12, 15, 15, 18, 20, 21, 25, 30, find the lower quartile (Q1), upper quartile (Q3) and the interquartile range.

4. A dot plot of daily bus wait times (minutes) has most dots between 2 and 6, with one dot far out at 20. Describe the shape and identify any outlier.

Achieved

5. A student wants to investigate reaction times at Vertex Academy. Write a suitable investigative comparison question, and state what data would need to be collected.

6. Given the shoe sizes of 10 netball players: 6, 7, 7, 8, 8, 8, 9, 9, 10, 11, choose a suitable display and describe the shape and centre.

7. The times (minutes) for 12 people to complete a 3 km parkrun in New Lynn: 16, 18, 18, 19, 20, 21, 22, 23, 24, 25, 27, 30. Find the median and the IQR.

8. A café records the number of coffees sold each hour: 14, 22, 18, 30, 25, 19. Calculate the mean and state whether the mean or median better represents a 'typical' hour if one hour had a huge one-off order of 90 coffees added.

Merit

9. Two Year 11 classes did the same 40-mark test. Class A: median 26, Q1 20, Q3 31. Class B: median 24, Q1 22, Q3 27. Compare the two distributions, linking centre and spread, and comment on which class did better.

10. A student collected commuting distances (km) from a survey posted only in the school's cycling club group chat, then concluded 'most students at our school live within 5 km of school'. Explain two problems with this conclusion and how the plan could be improved.

11. Weekly pocket money ($) for a sample: 5, 10, 10, 12, 15, 15, 15, 20, 25, 60. A student says 'the average is $18.70 so a typical student gets about $19'. Using the shape of the data, evaluate this claim and suggest a better summary.

Excellence

12. Two box plots comparing sleep hours of Year 9 and Year 13 students overlap: Year 9 median 8.5 h (IQR 7.5–9.5), Year 13 median 6.5 h (IQR 5.5–7.5). Explain, with justification, whether it is reasonable to claim that in the wider student population Year 13 students tend to sleep less than Year 9 students.

13. Prove that adding a single new value equal to the current mean of a data set leaves the mean unchanged, and explain what this shows about the mean's sensitivity.

14. A student surveyed 200 people at Avondale train station one Monday morning and found the median daily screen time was 1.5 hours, concluding 'New Zealanders spend about 1.5 hours per day on screens'. Critically evaluate every stage of this enquiry (problem, plan, data, analysis, conclusion) and describe a redesign that would support a valid national claim.