

NCEA Level 2 · AS 91264: Making an Inference from Samples

Name: _____

Date: _____

This standard is all about taking samples from populations and using them to make a fair claim (an inference) about the wider group. Work through the sheet building box plots, calculating informal confidence intervals, and writing clear justified conclusions, that's what earns the marks.

Revision summary

- The statistical enquiry cycle (PPDAC): Problem, Plan, Data, Analysis, Conclusion. Your investigative question should compare a variable between two populations, e.g. 'I wonder if Year 12 boys at our school are taller than Year 12 girls, on average?'
- We take a **SAMPLE** to make an inference about the whole **POPULATION** (we don't have all the data). Use the median as the measure of centre for comparisons.
- Five-number summary: minimum, lower quartile (Q1), median (Q2), upper quartile (Q3), maximum. $IQR = Q3 - Q1$. Box plot: box from Q1 to Q3, line at median, whiskers to min and max.
- Informal confidence interval for a median: $median \pm 1.5 \times (IQR \div \sqrt{n})$. This estimates where the true population median likely sits.
- To make a call: if the two informal confidence intervals do NOT overlap, you can make the call back to the population that one median is likely bigger. If they overlap, you cannot make a call.
- Always describe features in context: centre (median), spread (IQR/range), shape (symmetry, skew), and any unusual/outlying values.
- A good conclusion answers the question, refers to the sample evidence, makes a call back to the population (or says you can't), and mentions sampling variability.
- Bigger sample sizes give narrower confidence intervals, so a more reliable inference.

Key skills

1. The reaction times (in seconds) of 9 students on a computer test are: 0.28, 0.31, 0.33, 0.35, 0.36, 0.39, 0.41, 0.44, 0.52. Find the five-number summary.
2. For a sample the lower quartile is 152 cm and the upper quartile is 168 cm. Calculate the interquartile range (IQR).
3. A sample of 25 kiwifruit has a median weight of 92 g and an IQR of 20 g. Calculate the informal confidence interval for the population median.
4. Explain the difference between a sample and a population, using the example of investigating the weekly screen time of Year 11 students at your school.

Achieved

5. A student investigates whether Year 13 students at Avondale College sleep more on weekends than on weeknights. Write a suitable investigative comparison question and identify the variable and the two groups being compared.

6. A sample of 16 avocados from a Kaitaia orchard has weights (g): 180, 190, 195, 200, 205, 210, 210, 215, 220, 225, 230, 235, 240, 250, 260, 280. Find the median, Q1 and Q3, then describe how you would draw the box plot.
7. A survey of 36 café flat whites finds a median price of \$5.20 and an IQR of \$0.90. Calculate the informal confidence interval for the median price and interpret it in context.
8. Two box plots show the heights of sampled netball players (median 172 cm) and basketball players (median 179 cm). The netball confidence interval is 168–176 cm and the basketball interval is 175–183 cm. State whether you can make a call about the populations and why.

Merit

9. A student compares reaction times (seconds) of a sample of 30 Year 12 students who play video games regularly (median 0.34, IQR 0.09) with 30 who do not (median 0.41, IQR 0.12). Calculate both informal confidence intervals, compare them, and make a call.
10. A sample of 50 commute times (minutes) to an Auckland CBD workplace has median 34, Q1 = 26, Q3 = 46, minimum 15, maximum 78. Describe the distribution referring to centre, spread and shape, and explain what the shape suggests about commute times.
11. Two students take separate samples of the same population of Year 9 running times. Sample A has $n = 20$ (median 62 s, IQR 10 s); Sample B has $n = 80$ (median 62 s, IQR 10 s). Calculate both confidence intervals and explain how sample size affects the reliability of the inference.

Excellence

12. A student concludes: 'Because the median height of my sample of boys (170 cm) is bigger than my sample of girls (163 cm), all boys at the school are taller than all girls.' Critique this conclusion and explain what a statistically valid conclusion should include.
13. Explain why the informal confidence interval uses $\text{median} \pm 1.5 \times (\text{IQR} \div \sqrt{n})$, and why we compare intervals rather than simply comparing the two sample medians directly. Reason in terms of what would happen if the samples were taken again.
14. Two samples of tomato weights give overlapping confidence intervals, so a call cannot be made. A student says 'This proves the two populations have equal median weights.' Explain why this is incorrect, and describe two changes to the investigation that could allow a valid call to be made if a real difference exists.

Answer key (for parents and tutors)

1. The reaction times (in seconds) of 9 students on a computer test are: 0.28, 0.31, 0.33, 0.35, 0.36, 0.39, 0.41, 0.44, 0.52. Find the five-number summary.
Ordered already. $n = 9$. Minimum = 0.28, Maximum = 0.52. Median is the 5th value = 0.36. Lower half (below median): 0.28, 0.31, 0.33, 0.35 $\rightarrow Q1 = (0.31 + 0.33) \div 2 = 0.32$. Upper half: 0.39, 0.41, 0.44, 0.52 $\rightarrow Q3 = (0.41 + 0.44) \div 2 = 0.425$. Five-number summary: 0.28, 0.32, 0.36, 0.425, 0.52.
Achieved: applies the method for finding a five-number summary.
2. For a sample the lower quartile is 152 cm and the upper quartile is 168 cm. Calculate the interquartile range (IQR).
 $IQR = Q3 - Q1 = 168 - 152 = 16$ cm.
Achieved: correctly computes IQR.
3. A sample of 25 kiwifruit has a median weight of 92 g and an IQR of 20 g. Calculate the informal confidence interval for the population median.
Interval = median $\pm 1.5 \times (IQR \div \sqrt{n}) = 92 \pm 1.5 \times (20 \div \sqrt{25}) = 92 \pm 1.5 \times (20 \div 5) = 92 \pm 1.5 \times 4 = 92 \pm 6$. So the interval is from 86 g to 98 g.
Achieved: applies the informal confidence interval formula.
4. Explain the difference between a sample and a population, using the example of investigating the weekly screen time of Year 11 students at your school.
The population is ALL Year 11 students at the school, every student we are interested in making a claim about. The sample is the smaller group of Year 11 students we actually collect screen-time data from (for example, 40 randomly chosen students). We use the sample because measuring everyone is impractical, and we use it to make an inference about the whole Year 11 population.
Achieved: distinguishes sample and population in context.
5. A student investigates whether Year 13 students at Avondale College sleep more on weekends than on weeknights. Write a suitable investigative comparison question and identify the variable and the two groups being compared.
Investigative question: 'I wonder if the median weekend sleep time of Year 13 students at Avondale College is longer than their median weeknight sleep time?' The variable is sleep time (hours per night, a numerical variable). The two groups compared are 'weekend nights' and 'weeknights' for Year 13 students. It is a comparison question because it compares a numerical variable across two groups and refers to the population.
Achieved: poses a valid comparison investigative question.
6. A sample of 16 avocados from a Kaitia orchard has weights (g): 180, 190, 195, 200, 205, 210, 210, 215, 220, 225, 230, 235, 240, 250, 260, 280. Find the median, Q1 and Q3, then describe how you would draw the box plot.
 $n = 16$. Median = average of 8th and 9th values = $(215 + 220) \div 2 = 217.5$ g. Lower 8 values: 180, 190, 195, 200, 205, 210, 210, 215 $\rightarrow Q1 = \text{average of 4th and 5th} = (200 + 205) \div 2 = 202.5$ g. Upper 8 values: 220, 225, 230, 235, 240, 250, 260, 280 $\rightarrow Q3 = \text{average of 4th and 5th} = (235 + 240) \div 2 = 237.5$ g. Box plot: draw a box from 202.5 to 237.5 with a line at 217.5, whisker left to minimum 180, whisker right to maximum 280.
Achieved: calculates quartiles and constructs a box plot.
7. A survey of 36 café flat whites finds a median price of \$5.20 and an IQR of \$0.90. Calculate the informal confidence interval for the median price and interpret it in context.
Interval = $5.20 \pm 1.5 \times (0.90 \div \sqrt{36}) = 5.20 \pm 1.5 \times (0.90 \div 6) = 5.20 \pm 1.5 \times 0.15 = 5.20 \pm 0.225$. So the interval is from about \$4.98 to \$5.43. This means we are fairly confident the true median price of all flat whites in this population lies between \$4.98 and \$5.43.
Achieved: computes and interprets an informal confidence interval.
8. Two box plots show the heights of sampled netball players (median 172 cm) and basketball players (median 179 cm). The netball confidence interval is 168–176 cm and the basketball interval is 175–183 cm. State whether you can make a call about the populations and why.
The two confidence intervals overlap (netball goes up to 176 cm, basketball starts at 175 cm, the range 175–176 cm is shared). Because the intervals overlap, we CANNOT make a call that one population median is greater than the other. Based on this sample we cannot claim basketball players are taller than netball players in the population.
Achieved: applies the overlap rule to decide if a call can be made.

9. A student compares reaction times (seconds) of a sample of 30 Year 12 students who play video games regularly (median 0.34, IQR 0.09) with 30 who do not (median 0.41, IQR 0.12). Calculate both informal confidence intervals, compare them, and make a call.

Gamers: $0.34 \pm 1.5 \times (0.09 \div \sqrt{30})$. $\sqrt{30} \approx 5.477$, so $0.09 \div 5.477 \approx 0.01643$; $1.5 \times 0.01643 \approx 0.0246$. Interval ≈ 0.315 to 0.365 s. **Non-gamers:** $0.41 \pm 1.5 \times (0.12 \div 5.477)$. $0.12 \div 5.477 \approx 0.0219$; $1.5 \times 0.0219 \approx 0.0329$. Interval ≈ 0.377 to 0.443 s. The intervals (0.315–0.365 and 0.377–0.443) do NOT overlap. Since the gamers' interval sits entirely below the non-gamers' interval, I can make the call that back in the population, Year 12 students who play video games regularly tend to have faster (lower) reaction times than those who do not.

Merit: links the two confidence intervals to a justified population call.

10. A sample of 50 commute times (minutes) to an Auckland CBD workplace has median 34, Q1 = 26, Q3 = 46, minimum 15, maximum 78. Describe the distribution referring to centre, spread and shape, and explain what the shape suggests about commute times.

Centre: the median commute is 34 minutes. **Spread:** IQR = $46 - 26 = 20$ minutes, and the total range is $78 - 15 = 63$ minutes. **Shape:** the distance from the median to Q3 ($46 - 34 = 12$) is larger than from Q1 to the median ($34 - 26 = 8$), and the upper whisker (46 to $78 = 32$) is much longer than the lower whisker (26 to $15 = 11$). This shows the distribution is skewed to the right (positively skewed). This suggests most commuters get to work in a moderate time, but a smaller group has much longer commutes (perhaps from outer suburbs or heavy traffic), pulling the upper tail out.

Merit: connects boxplot features (skew) to a contextual interpretation.

11. Two students take separate samples of the same population of Year 9 running times. Sample A has $n = 20$ (median 62 s, IQR 10 s); Sample B has $n = 80$ (median 62 s, IQR 10 s). Calculate both confidence intervals and explain how sample size affects the reliability of the inference.

Sample A: $62 \pm 1.5 \times (10 \div \sqrt{20})$. $\sqrt{20} \approx 4.472$, $10 \div 4.472 \approx 2.236$, $1.5 \times 2.236 \approx 3.35$. Interval ≈ 58.6 to 65.4 s (width ≈ 6.7 s). **Sample B:** $62 \pm 1.5 \times (10 \div \sqrt{80})$. $\sqrt{80} \approx 8.944$, $10 \div 8.944 \approx 1.118$, $1.5 \times 1.118 \approx 1.68$. Interval ≈ 60.3 to 63.7 s (width ≈ 3.4 s). Both have the same centre but Sample B's interval is about half as wide. A larger sample gives a narrower confidence interval, so we can pin down the population median more precisely and make a more reliable inference. This is because $\sqrt{n}$ is in the denominator, as n increases, the margin decreases.

Merit: links sample size to interval width and reliability using the formula structure.

12. A student concludes: 'Because the median height of my sample of boys (170 cm) is bigger than my sample of girls (163 cm), all boys at the school are taller than all girls.' Critique this conclusion and explain what a statistically valid conclusion should include.

The conclusion has several flaws. Firstly, it compares only the sample medians without considering sampling variability, a different sample could give a different result, so a raw difference in medians is not enough. It should be supported by informal confidence intervals: only if the intervals do not overlap can a call be made back to the population. Secondly, the claim 'all boys are taller than all girls' is an overgeneralisation, an inference is about the population MEDIAN (or 'boys tend to be taller on average'), never about every individual, since the distributions clearly overlap (some girls are taller than some boys). A valid conclusion would: answer the investigative question, state the sample evidence, use the confidence intervals to justify making (or not making) a call, phrase the claim as being about the population tendency/median, and acknowledge that another sample could vary. For example: 'Because the confidence intervals do not overlap, I am confident that back in the population, boys at the school tend to be taller than girls, on average.'

Excellence: critiques flawed reasoning and justifies the features of a valid inference.

13. Explain why the informal confidence interval uses $\text{median} \pm 1.5 \times (\text{IQR} \div \sqrt{n})$, and why we compare intervals rather than simply comparing the two sample medians directly. Reason in terms of what would happen if the samples were taken again.

Any single sample is just one of many possible samples from the population, so its median is an estimate that would vary if we resampled, this is sampling variability. The term $\text{IQR} \div \sqrt{n}$ captures how much the sample median is likely to bounce around: it is larger when the data is more spread out (bigger IQR) and smaller when the sample is bigger (dividing by $\sqrt{n}$), because larger samples give more stable estimates. The factor 1.5 widens this into an interval that plausibly contains the true population median. If we simply compared the two sample medians, a small difference could easily be caused by chance sampling variation and disappear in a new sample. By comparing intervals, we only make a call when the difference is larger than the natural variation we'd expect from

resampling (non-overlap), meaning the difference is likely real in the population rather than a fluke of which people happened to be sampled.

Excellence: generalises the reasoning behind the interval formula and the non-overlap decision rule.

14. Two samples of tomato weights give overlapping confidence intervals, so a call cannot be made. A student says 'This proves the two populations have equal median weights.' Explain why this is incorrect, and describe two changes to the investigation that could allow a valid call to be made if a real difference exists.

Overlapping intervals do NOT prove the medians are equal. They only mean that, with this data, the observed difference could plausibly be due to sampling variability, so we don't have enough evidence to claim a difference, 'no call' is not the same as 'no difference'. A real difference may exist but be hidden by wide intervals. Two changes: (1) Increase the sample size, because the margin is $1.5 \times (IQR \div \sqrt{n})$, a larger n makes both intervals narrower, so if a real difference exists the intervals are more likely to separate and allow a call. (2) Improve sampling quality/reduce variability, use a well-designed random sampling method and consistent, accurate measurement so the IQR reflects true variation rather than measurement error or bias; a cleaner, less spread sample also narrows the intervals. Either change increases the chance of detecting a genuine population difference.

Excellence: distinguishes 'no call' from 'no difference' and generalises how design affects detecting a real effect.