NCEA Level 3 · AS 91582: Making the Call with Confidence Intervals

Name: _____ Date: _____

Hi! This standard is about using a bootstrap confidence interval to make a formal inference comparing two groups back in their populations. Work through each section in order, and always finish by writing your 'call' in context.

Revision summary

- Formal inference = making a claim about a POPULATION difference using a SAMPLE.
- Bootstrap: resample each sample WITH replacement (same size), record the difference in statistics (e.g. median A − median B), repeat ~1000 times to build a bootstrap distribution.
- Confidence interval (CI): the middle 90% of the bootstrap distribution, from the 5th percentile to the 95th percentile of the resampled differences.
- Making the call: if the whole CI is above 0, infer group A's population parameter is larger; if entirely below 0, it is smaller; if the CI contains 0, you cannot make a call (the difference could plausibly be zero).
- Interpretation: 'It is a fairly safe bet that the median [variable] for [population A] is between X and Y greater/less than for [population B].'
- Sampling variability: different samples give different statistics; larger samples give narrower CIs and more reliable calls.
- Always compare like with like, use the sample size, spread, shape and any outliers to support or temper your inference.
- Good justification links the CI position relative to 0, the sample sizes, and the practical context.

Key skills

1. A bootstrap confidence interval for the difference in median weekly grocery spend (Avondale households − Blockhouse Bay households) is (−$5, $42). Does this interval let you make a call about which population has the higher median spend? Briefly say why.

2. Explain what 'resampling with replacement' means when bootstrapping from a sample of 30 commute times.

3. A bootstrap distribution of 1000 differences in mean rainfall (mm) is sorted. The 50th value is 2.1 and the 950th value is 18.4. Write down the 90% bootstrap confidence interval.

4. A CI for the difference in median screen time (Year 13 students − Year 9 students) is (0.4 h, 2.7 h). Write the formal call in context.

Achieved

5. A researcher samples 40 kūmara from a Dargaville farm and 40 from a Kaipara farm, recording weight (g). A bootstrap gives a 90% confidence interval for the difference in median weight (Dargaville − Kaipara) of (12 g, 58 g). Make a formal inference in context.

6. From samples of daily bus wait times (minutes) in Onehunga and Ōtāhuhu, a bootstrap CI for the difference in medians (Onehunga − Ōtāhuhu) is (−9.2, −1.5). What can you conclude?

7. Explain why we make an inference about the POPULATION and not just report which SAMPLE median is higher.

8. A CI for the difference in mean weekly earnings (part-time workers aged 18–20 − aged 21–24) in a survey is (−$85, −$10). Give the call and state one reason a wider sample might change your confidence.

Merit

9. Two samples of marathon finish times (minutes) are taken: Auckland runners (n = 60, sample median 248, IQR 42) and Wellington runners (n = 55, sample median 261, IQR 39). A bootstrap 90% CI for the difference in medians (Auckland − Wellington) is (−24, −2). Discuss what this tells you, linking the sample summaries to the inference.

10. A student takes two samples from the same population by mistake and bootstraps a CI for the difference in medians, getting (−3.1, 3.4). They conclude 'there is no difference in the populations.' Explain why this wording is not quite right and give a better statement.

11. Explain, with reference to how the bootstrap distribution is built, why increasing each sample size from 20 to 80 tends to give a narrower confidence interval.

Excellence

12. A council claims 'water use per household is the same in Titirangi and Glen Eden.' Samples give a 90% bootstrap CI for the difference in median daily water use (Titirangi − Glen Eden) of (18 L, 61 L). A councillor argues this 'proves' the council is wrong. Critique both the council's claim and the councillor's argument, and state what can properly be concluded.

13. Two analysts study the same two populations. Analyst A uses samples of size 25 and gets CI (−2, 30) for the difference in medians; Analyst B uses samples of size 200 and gets CI (6, 22). Explain how both results can be consistent, and discuss which supports a stronger inference and why.

14. Explain why a bootstrap confidence interval built by resampling can be trusted to reflect population uncertainty, even though we only ever resample from a single original sample and never from the population itself.

Answer key (for parents and tutors)

1. A bootstrap confidence interval for the difference in median weekly grocery spend (Avondale households − Blockhouse Bay households) is (−$5, $42). Does this interval let you make a call about which population has the higher median spend? Briefly say why.
   No. The interval (−$5, $42) contains 0, so it is plausible that there is no difference in the population medians. Because zero lies inside the interval, you cannot confidently say which population has the higher median spend.
   Achieved: applies the 'contains 0' rule to decide whether a call can be made.

2. Explain what 'resampling with replacement' means when bootstrapping from a sample of 30 commute times.
   You draw 30 values one at a time from the original sample of 30, but after each draw the value is put back so it can be chosen again. This produces a new sample of size 30 in which some original values may appear more than once and others not at all. This models the variability we would see if we took fresh samples from the population.
   Achieved: states the mechanics of bootstrap resampling.

3. A bootstrap distribution of 1000 differences in mean rainfall (mm) is sorted. The 50th value is 2.1 and the 950th value is 18.4. Write down the 90% bootstrap confidence interval.
   The middle 90% runs from the 5th percentile (the 50th of 1000 values) to the 95th percentile (the 950th value). So the 90% CI is (2.1 mm, 18.4 mm).
   Achieved: reads a confidence interval from percentiles of a bootstrap distribution.

4. A CI for the difference in median screen time (Year 13 students − Year 9 students) is (0.4 h, 2.7 h). Write the formal call in context.
   Because the whole interval is above 0, it is a fairly safe bet that the median daily screen time for the population of Year 13 students is between 0.4 and 2.7 hours greater than for the population of Year 9 students.
   Achieved: writes a contextual call from a CI lying entirely above zero.

5. A researcher samples 40 kūmara from a Dargaville farm and 40 from a Kaipara farm, recording weight (g). A bootstrap gives a 90% confidence interval for the difference in median weight (Dargaville − Kaipara) of (12 g, 58 g). Make a formal inference in context.
   The entire interval (12 g, 58 g) is positive, so 0 is not a plausible value for the difference. Therefore it is a fairly safe bet that the median weight of kūmara from the Dargaville farm population is between 12 g and 58 g greater than the median weight of kūmara from the Kaipara farm population.
   Achieved: applies the CI method to make a straightforward population call.

6. From samples of daily bus wait times (minutes) in Onehunga and Ōtāhuhu, a bootstrap CI for the difference in medians (Onehunga − Ōtāhuhu) is (−9.2, −1.5). What can you conclude?
   The whole interval is negative, so it is a fairly safe bet that the median bus wait time for the Onehunga population is between 1.5 and 9.2 minutes LESS than for the Ōtāhuhu population. In other words, Onehunga tends to have shorter waits.
   Achieved: interprets an entirely negative CI as a directional call.

7. Explain why we make an inference about the POPULATION and not just report which SAMPLE median is higher.
   Sample medians vary from sample to sample (sampling variability), so a difference between two sample medians might just be due to chance. A confidence interval accounts for this variability and lets us say what the difference is likely to be back in the populations, which is what we actually care about, not the specific samples we happened to collect.
   Achieved: distinguishes sample statistic from population parameter and references sampling variability.

8. A CI for the difference in mean weekly earnings (part-time workers aged 18–20 − aged 21–24) in a survey is (−$85, −$10). Give the call and state one reason a wider sample might change your confidence.
   The interval is entirely negative, so it is a fairly safe bet that the mean weekly earnings for the population of 18–20 year-old part-time workers is between $10 and $85 less than for the 21–24 group. A larger sample would generally produce a narrower interval, giving a more precise (and possibly more confident) estimate of the difference.
   Achieved: makes a call and links sample size to interval width.

9. Two samples of marathon finish times (minutes) are taken: Auckland runners (n = 60, sample median 248, IQR 42) and Wellington runners (n = 55, sample median 261, IQR 39). A bootstrap 90% CI for the difference in medians (Auckland – Wellington) is (−24, −2). Discuss what this tells you, linking the sample summaries to the inference.
The sample median for Auckland (248) is 13 minutes lower than Wellington (261), suggesting Auckland runners finish faster. The bootstrap CI (−24, −2) is entirely below 0, so it is a fairly safe bet that the median finish time for the Auckland population is between 2 and 24 minutes less than for the Wellington population. The similar IQRs (42 vs 39) show comparable spread, so the median difference is a fair comparison. The fairly large samples (60, 55) mean the estimate is reasonably reliable, though the width of the interval (22 minutes) shows there is still meaningful uncertainty about exactly how big the difference is.
Merit: links sample statistics, spread and sample size to interpret the CI.

10. A student takes two samples from the same population by mistake and bootstraps a CI for the difference in medians, getting (−3.1, 3.4). They conclude 'there is no difference in the populations.' Explain why this wording is not quite right and give a better statement.
The CI contains 0, so we cannot make a call about a difference, but 'no difference' is too strong. We have not proved the population medians are equal; we have only found that 0 is a plausible value for the difference, along with a whole range of small positive and negative values. A better statement: 'Because the confidence interval contains 0, it is plausible there is no difference, so we cannot make a call about which population median is larger.'
Merit: distinguishes 'cannot make a call' from 'proved no difference' and reasons about plausibility.

11. Explain, with reference to how the bootstrap distribution is built, why increasing each sample size from 20 to 80 tends to give a narrower confidence interval.
Each bootstrap resample is drawn from the original sample, so its variability reflects the sample size. With larger samples, each resampled median is based on more data and is closer to the true population median, so the resampled differences cluster more tightly. The bootstrap distribution of differences becomes narrower, and since the CI runs from its 5th to 95th percentile, that interval also narrows. This means a larger sample gives a more precise estimate of the population difference and often lets us make a call when a small sample could not.
Merit: connects sample size to the spread of the bootstrap distribution and hence CI width.

12. A council claims 'water use per household is the same in Titirangi and Glen Eden.' Samples give a 90% bootstrap CI for the difference in median daily water use (Titirangi – Glen Eden) of (18 L, 61 L). A councillor argues this 'proves' the council is wrong. Critique both the council's claim and the councillor's argument, and state what can properly be concluded.
The councillor is closer to correct but overstates it. The CI (18, 61) lies entirely above 0, so 0 is not a plausible value for the difference; this gives strong evidence AGAINST the council's claim of equal median use, so the claim is not supported. However, a confidence interval never 'proves' anything with certainty, it is a 90% procedure, meaning about 10% of such intervals would fail to capture the true difference, and different samples would give slightly different intervals. What can properly be concluded: it is a fairly safe bet that the median daily water use for the Titirangi population is between 18 L and 61 L higher than for Glen Eden, so the evidence contradicts the council's equal-use claim, but this is a probabilistic inference, not proof.
Excellence: critically evaluates competing claims and articulates the probabilistic nature of inference.

13. Two analysts study the same two populations. Analyst A uses samples of size 25 and gets CI (−2, 30) for the difference in medians; Analyst B uses samples of size 200 and gets CI (6, 22). Explain how both results can be consistent, and discuss which supports a stronger inference and why.
Both intervals are consistent because they overlap and both are compatible with a true difference somewhere around the low-to-mid teens. Analyst A's small samples give a wide interval (width 32) that includes 0, so A cannot make a call, the sampling variability is too large to rule out a zero difference. Analyst B's much larger samples give a narrower interval (width 16) sitting entirely above 0, so B CAN confidently infer the first population's median is 6 to 22 higher. B's inference is stronger: the larger sample reduces sampling variability, tightens the bootstrap distribution, and produces a more precise estimate that excludes 0. The two are not contradictory, A simply lacked the precision to detect a difference that B's data reveals.
Excellence: generalises the effect of sample size on precision and reconciles apparently different results.

14. Explain why a bootstrap confidence interval built by resampling can be trusted to reflect population uncertainty, even though we only ever resample from a single original sample and never from the population itself.
    The key idea is that a randomly collected sample is our best available picture of the population, its shape, spread and centre approximate the population's. Resampling with replacement from this sample mimics the process of drawing many fresh samples from a population that looks like our sample. Each resample varies because of which values are (re)selected, and that variation stands in for true sampling variability. By repeating this many times, the spread of resampled differences estimates how much the difference-in-statistics would vary across genuine repeated sampling, which is exactly the uncertainty a confidence interval must capture. The method works well when the sample is reasonably large and representative; if the sample is small or biased, the resamples inherit those flaws, so the interval can be misleading. Thus the bootstrap is trustworthy to the extent the original sample faithfully represents the population.
    Excellence: justifies the logical foundation of bootstrapping and states its assumptions and limitations.