Year 10 · Statistics: Reading Between the Data

Name: _____ Date: _____

Hi! This sheet helps you interpret data like a real statistician, comparing groups fairly, predicting from scatter plots, and spotting graphs that try to fool you. Take your time and always explain your thinking in context.

Warm-up

1. A scatter plot at a Mission Bay dairy shows daily temperature (from 15°C to 28°C) against ice-cream sales. If you use the graph to predict sales at 22°C, is that interpolation or extrapolation?

2. Using the same graph, you predict sales at 34°C. Is that interpolation or extrapolation, and why should you be cautious?

3. A line of best fit for that dairy is: sales = 12 × temperature + 30. Predict the number of ice creams sold at 20°C.

4. In the 75%-to-50% comparison rule, which value of Group A and which value of Group B do we compare?

Practice

5. Box plots compare Year 10 student heights at School A and School B. School A: min 150, Q1 158, median 163, Q3 168, max 178 (cm). School B: min 160, Q1 167, median 172, Q3 178, max 188 (cm). Using the 75%-to-50% rule, can you claim that School B students tend to be taller back in the population?

6. A news website shows a bar chart of two companies' yearly profit: Kea Ltd $42 million and Tui Ltd $45 million. The bar for Tui looks about twice as tall as Kea's. Explain how the graph misrepresents the data and how it should be fixed.

7. A study of kauri trees relates trunk circumference to age. The line of best fit is: age = 0.8 × circumference + 5, where circumference ranges from 50 cm to 200 cm. (a) Predict the age of a tree with circumference 120 cm. (b) Would you trust a prediction for a 400 cm tree?

8. A survey of 30 students at one Avondale school found 60% preferred rugby to football. A newspaper headline says: 'New Zealand teenagers prefer rugby.' Comment on this claim, taking uncertainty into account.

9. Two Year 10 classes sit the same maths test (out of 100). Class A: Q1 55, median 62, Q3 70. Class B: Q1 50, median 65, Q3 72. Using the 75%-to-50% rule, can you claim Class B tends to score higher in the population?

10. A pictogram about recycling shows one rubbish-bin symbol = 100 kg, but the bin drawn for 2023 is made twice as wide AND twice as tall as the 2022 bin to show 200 kg. Why does this mislead readers?

Challenge

11. A scatter plot relates weekly study hours (0 to 15 hours) to test score for Year 10 students. It shows a moderate positive relationship with line of best fit: score = 4 × hours + 40. (a) Predict the score for a student who studies 8 hours. (b) A student claims: 'If I study 30 hours, this predicts a score of 160, so I'm guaranteed top marks.' Give two reasons this reasoning is flawed.

12. A researcher wants to investigate: 'Do Year 10 students at Avondale schools who play a sport tend to have faster reaction times than those who don't?' She measures reaction times (seconds) for a sample of each group. Sport group: Q1 0.24, median 0.27, Q3 0.30. Non-sport group: Q1 0.29, median 0.33, Q3 0.37. (a) Apply the 75%-to-50% rule to compare the groups (remember: smaller time = faster). (b) Explain what elements of chance mean for how confident she can be.

Answer key (for parents and tutors)

1. A scatter plot at a Mission Bay dairy shows daily temperature (from 15°C to 28°C) against ice-cream sales. If you use the graph to predict sales at 22°C, is that interpolation or extrapolation?
   Interpolation. 22°C lies inside the range of the temperature data (15°C to 28°C), so the prediction is within the data and is called an interpolation.

2. Using the same graph, you predict sales at 34°C. Is that interpolation or extrapolation, and why should you be cautious?
   Extrapolation, because 34°C is outside the data range (15°C–28°C). We have no evidence the same pattern continues past 28°C, so the prediction is less reliable.

3. A line of best fit for that dairy is: sales = 12 × temperature + 30. Predict the number of ice creams sold at 20°C.
   sales = 12 × 20 + 30 = 240 + 30 = 270 ice creams.

4. In the 75%-to-50% comparison rule, which value of Group A and which value of Group B do we compare?
   We compare the upper quartile (Q3, the 75% mark) of Group A with the median (the 50% mark) of Group B. If Group B's median is larger than Group A's upper quartile, then more than 50% of B is larger than 75% of A.

5. Box plots compare Year 10 student heights at School A and School B. School A: min 150, Q1 158, median 163, Q3 168, max 178 (cm). School B: min 160, Q1 167, median 172, Q3 178, max 188 (cm). Using the 75%-to-50% rule, can you claim that School B students tend to be taller back in the population?
   Compare School B's median (172 cm) with School A's upper quartile Q3 (168 cm). Since 172 > 168, more than 50% of B's data is larger than 75% of A's data. The rule is satisfied, so we CAN make the claim that School B students tend to be taller than School A students back in the population. (We still note there is overlap and variability, so it is a tendency, not a certainty.)

6. A news website shows a bar chart of two companies' yearly profit: Kea Ltd $42 million and Tui Ltd $45 million. The bar for Tui looks about twice as tall as Kea's. Explain how the graph misrepresents the data and how it should be fixed.
   The vertical axis does not start at 0 (it likely starts around $40m). This 'truncated axis' exaggerates the small difference: $45m is only about 7% more than $42m, not double. To represent the data fairly, the y-axis should start at $0, which would show two bars of nearly the same height.

7. A study of kauri trees relates trunk circumference to age. The line of best fit is: age = 0.8 × circumference + 5, where circumference ranges from 50 cm to 200 cm. (a) Predict the age of a tree with circumference 120 cm. (b) Would you trust a prediction for a 400 cm tree?
   (a) age = 0.8 × 120 + 5 = 96 + 5 = 101 years. This is an interpolation (120 cm is inside 50–200 cm), so it is reasonably reliable. (b) A 400 cm circumference is far outside the data range (extrapolation), so the prediction would be unreliable, the growth pattern may change for very old trees, and we have no data there.

8. A survey of 30 students at one Avondale school found 60% preferred rugby to football. A newspaper headline says: 'New Zealand teenagers prefer rugby.' Comment on this claim, taking uncertainty into account.
   The claim over-reaches. The data comes from only 30 students at ONE school, which is a small, non-random sample and may not represent all NZ teenagers. Because of sampling variability (a different sample could give a different result), we cannot confidently generalise to the whole population. A fairer statement is that in this sample, more students preferred rugby, but a larger, more representative survey would be needed to make a claim about all NZ teenagers.

9. Two Year 10 classes sit the same maths test (out of 100). Class A: Q1 55, median 62, Q3 70. Class B: Q1 50, median 65, Q3 72. Using the 75%-to-50% rule, can you claim Class B tends to score higher in the population?
   Compare Class B's median (65) with Class A's upper quartile Q3 (70). Since 65 is NOT greater than 70, less than 50% of B's data lies above 75% of A's data. The rule is not satisfied, so we cannot claim that Class B tends to score higher back in the population. The two groups' scores are too similar to call.

10. A pictogram about recycling shows one rubbish-bin symbol = 100 kg, but the bin drawn for 2023 is made twice as wide AND twice as tall as the 2022 bin to show 200 kg. Why does this mislead readers?
    Doubling the amount should double the area, but making the picture twice as wide and twice as tall makes the 2023 bin 2 × 2 = 4 times the area of the 2022 bin. The eye compares areas, so 200 kg looks like 4 times 100 kg

instead of 2 times. The symbols should only double in one direction (or use two identical bin symbols) so the visual matches the actual doubling.

11. A scatter plot relates weekly study hours (0 to 15 hours) to test score for Year 10 students. It shows a moderate positive relationship with line of best fit: score = 4 × hours + 40. (a) Predict the score for a student who studies 8 hours. (b) A student claims: 'If I study 30 hours, this predicts a score of 160, so I'm guaranteed top marks.' Give two reasons this reasoning is flawed.
(a) score = 4 × 8 + 40 = 32 + 40 = 72. This is an interpolation (8 hours is within 0–15), so it is reasonably reliable. (b) Reason 1: 30 hours is far outside the data range (0–15), so it is an extrapolation and the linear pattern is unlikely to continue. Reason 2: the model gives 160, but scores are capped at 100, which is impossible, showing the line cannot be extended that far. Also, the relationship is only 'moderate', so there is scatter/variability and study hours do not perfectly determine the score. The prediction is not a guarantee.

12. A researcher wants to investigate: 'Do Year 10 students at Avondale schools who play a sport tend to have faster reaction times than those who don't?' She measures reaction times (seconds) for a sample of each group. Sport group: Q1 0.24, median 0.27, Q3 0.30. Non-sport group: Q1 0.29, median 0.33, Q3 0.37. (a) Apply the 75%-to-50% rule to compare the groups (remember: smaller time = faster). (b) Explain what elements of chance mean for how confident she can be.
(a) Faster means a SMALLER time, so we test whether the sport group tends to be faster (smaller). Compare the sport group's median (0.27) with the non-sport group's LOWER quartile Q1 (0.29). Since 0.27 < 0.29, more than 50% of the sport group's times are smaller (faster) than 75% of the non-sport group's times. So we CAN claim that sporty students tend to have faster reaction times back in the population. (b) Because the data comes from a sample, chance affects which students were chosen and how they performed on the day (tiredness, guessing, distraction). A different sample could give slightly different quartiles. So she should state her finding as a tendency 'back in the population', acknowledging variability, rather than as a certain fact, and note that a larger, well-chosen sample would strengthen the conclusion.