Year 10 · Statistics: Sampling Smarter, Inferring Better

Name: _____ Date: _____

Hi! This sheet helps you think about how we collect good data through sampling, why random samples matter, and when to use the mean or median. Take your time and show your reasoning, statisticians explain their thinking!

Warm-up

1. What is the difference between a census and a sample?

2. In simple random sampling, what is true about each item's chance of being chosen?

3. A survey measures the height (in cm) of students, a numerical variable. What is the smallest sample size recommended for a reliable inference?

4. Name one problem that biased sampling can cause.

Practice

5. Mere wants to know the favourite lunch food of students at Avondale College (about 1,900 students). Should she use a census or a sample? Explain.

6. A student surveys only her netball team about how many hours of sport students play each week, then claims her result represents the whole school. Why is this biased, and how could she fix it?

7. Explain how you could use technology (such as a random number generator) to choose a simple random sample of 30 students from a roll of 450 students.

8. For measuring students' reaction times (a numerical variable), why is a sample of 30 considered enough, but for finding what percentage of students prefer a certain phone brand (a categorical variable) about 1,000 is recommended?

9. The weekly pocket money of 9 students (in $) is: 5, 8, 10, 10, 12, 15, 20, 25, 90. Find the mean and the median. Which is the better measure of the typical amount, and why?

10. A statistical question asks: 'What is the typical time Year 10 students take to travel to Vertex Academy?' Most students take 5–20 minutes, but a few travel over an hour. Would the mean or median better answer this question? Explain.

Challenge

11. A local café owner wants to know what proportion of Avondale residents would buy a new plant-based menu. She surveys 40 customers who are already in her café on a Sunday morning. Identify TWO sources of bias, and describe a better sampling plan including a suitable sample size.

12. Two random samples of daily rainfall (in mm) are taken in Auckland. Sample A: 0, 0, 2, 3, 5, 5, 8, 12, 15 (9 days). Sample B is the same but adds one very wet day of 60 mm (10 days). Calculate the mean and median for each sample. Explain what happens to each measure and what this shows about choosing a measure of central tendency.

Answer key (for parents and tutors)

1. What is the difference between a census and a sample?
   A census collects data from every member of the population, while a sample collects data from only a selected part of the population. A sample is used to make inferences about the whole population when a census is impractical.
2. In simple random sampling, what is true about each item's chance of being chosen?
   Every item in the population has an equal probability (an equal chance) of being selected. This is what makes the sample fair and free from bias.
3. A survey measures the height (in cm) of students, a numerical variable. What is the smallest sample size recommended for a reliable inference?
   At least 30. For numerical variables, a sample of at least 30 is recommended to make a reliable inference about the population.
4. Name one problem that biased sampling can cause.
   Biased sampling leads to data that does not fairly represent the population, producing inaccurate and misleading statistics (for example, over-representing one group and drawing a wrong conclusion).
5. Mere wants to know the favourite lunch food of students at Avondale College (about 1,900 students). Should she use a census or a sample? Explain.
   A sample. Surveying all 1,900 students would take too much time and effort, so a well-chosen sample can give a reliable estimate. Favourite lunch food is a categorical variable, so she should aim for a large sample (about 1,000) collected using technology, or at least the largest practical random sample.
6. A student surveys only her netball team about how many hours of sport students play each week, then claims her result represents the whole school. Why is this biased, and how could she fix it?
   It is biased because netball players are more likely to play a lot of sport than the average student, so the team is not representative of all students. To fix it, she should take a simple random sample from the entire school roll, giving every student an equal chance of selection.
7. Explain how you could use technology (such as a random number generator) to choose a simple random sample of 30 students from a roll of 450 students.
   Number every student on the roll from 1 to 450. Use a random number generator to produce 30 different whole numbers between 1 and 450 (ignoring any repeats). The students matching those numbers form the simple random sample. Because each number is equally likely, each student has an equal chance of being chosen.
8. For measuring students' reaction times (a numerical variable), why is a sample of 30 considered enough, but for finding what percentage of students prefer a certain phone brand (a categorical variable) about 1,000 is recommended?
   Numerical variables give a summary such as a mean or median, and a sample of at least 30 is usually enough to estimate these reliably. Categorical variables report proportions or percentages, which vary more from sample to sample, so a much larger sample (about 1,000, obtained using technology) is needed to estimate a percentage reliably.
9. The weekly pocket money of 9 students (in $) is: 5, 8, 10, 10, 12, 15, 20, 25, 90. Find the mean and the median. Which is the better measure of the typical amount, and why?
   Mean = (5+8+10+10+12+15+20+25+90) ÷ 9 = 195 ÷ 9 ≈ $21.67. Median = the 5th value (middle of 9) = $12. The median ($12) is the better measure because the $90 value is an outlier that pulls the mean upward, making $21.67 higher than what most students actually receive. The median is not affected by the outlier, so it better represents the typical amount.
10. A statistical question asks: 'What is the typical time Year 10 students take to travel to Vertex Academy?' Most students take 5–20 minutes, but a few travel over an hour. Would the mean or median better answer this question? Explain.
    The median would be better. The few very long travel times are outliers that would inflate the mean, making the 'typical' time seem longer than it is for most students. The median is resistant to these extreme values and gives a fairer picture of the typical travel time.

11. A local café owner wants to know what proportion of Avondale residents would buy a new plant-based menu. She surveys 40 customers who are already in her café on a Sunday morning. Identify TWO sources of bias, and describe a better sampling plan including a suitable sample size.
Source 1: The sample only includes people already in her café, who are more likely to enjoy her food and may not represent all Avondale residents. Source 2: Sunday-morning customers may differ from the wider population (for example, in age, diet, or income), so the timing adds further bias. Better plan: The variable (would buy / would not buy) is categorical, so she should aim for about 1,000 responses obtained using technology. She could obtain a list or numbering of Avondale residents and use a random number generator to select participants (simple random sampling), so every resident has an equal chance of being chosen. This removes the café-only and timing bias and gives a large enough sample for a reliable percentage estimate.

12. Two random samples of daily rainfall (in mm) are taken in Auckland. Sample A: 0, 0, 2, 3, 5, 5, 8, 12, 15 (9 days). Sample B is the same but adds one very wet day of 60 mm (10 days). Calculate the mean and median for each sample. Explain what happens to each measure and what this shows about choosing a measure of central tendency.
Sample A: Mean = (0+0+2+3+5+5+8+12+15) ÷ 9 = 50 ÷ 9 ≈ 5.6 mm. Median = 5th value = 5 mm. Sample B: Sum = 50 + 60 = 110, Mean = 110 ÷ 10 = 11 mm. Ordered data: 0,0,2,3,5,5,8,12,15,60; median = average of 5th and 6th values = (5+5) ÷ 2 = 5 mm. Result: Adding the 60 mm day nearly doubled the mean (from 5.6 to 11 mm) but left the median unchanged at 5 mm. This shows the mean is strongly affected by outliers while the median is resistant. For rainfall data with occasional extreme wet days, the median often better represents the typical daily rainfall, though the mean is useful if you need a total (for example, total rainfall = mean × number of days).