Year 9 Statistics – Revision Notes

HappyKids.fit

Year 9 · Maths · Statistics · Key ideas and worked examples

For grown-ups and learners: read each part and try the worked example. Then work through the sheets: Foundation, Developing, Challenge, then Mixed Practice to revise everything. Show your working: it earns marks in tests and helps you spot mistakes.

1. Planning: types of data, sampling and bias

Start with a question and a prediction, then decide **what data** answers it. **Qualitative** data is described in words. **Quantitative** data is numerical: **discrete** data is counted (number of pets, shoe size); **continuous** data is measured (time, mass, height). Trial your method on a few people first. In a **random** sample everyone has the same chance of being chosen. A sample is **biased** if some groups are missed or over-represented: only gym members, one time or place, a **leading question**, a **self-selected** online vote. Then ask **further questions**.

Example: Data is collected on the number of people on each bus that passes. Is the data qualitative or quantitative? Is it categorical, discrete or continuous?
→ It is counted, so it can only take separate values
Answer: quantitative, discrete

2. Recording and representing data

Tally charts and **frequency tables** record data as you collect it. A **two-way table** counts two categories together; a **Venn** or **Carroll diagram** shows overlap or two yes/no properties. **Dual bar charts** put two groups side by side and **compound bar charts** stack them. **Pie charts** show parts of a whole and **time series graphs** show change over time. A **frequency polygon** plots each frequency at its class **midpoint**. **Scatter graphs** show relationships, **back-to-back stem-and-leaf diagrams** compare two small sets, and **infographics** present key facts.

Example: The two-way table shows what 89 students chose for lunch. Some values are missing. How many girls chose salad?
→ Boys, Salad: 44 − 10 − 14 = 20
→ Girls, Salad: 41 − 20 = 21
Answer: 21 girls

	Pasta	Salad	Soup	Total
Boys	10	?	14	44
Girls	5	?	?	?
Total	15	41	33	89

A **back-to-back stem-and-leaf diagram** puts two sets on one stem. The leaves on the left are read outwards from the stem. In the diagram below, 7 | 4 | 9 means 47 for Team X and 49 for Team Y.

Team X	Stem	Team Y
7 4	**4**	9
8 5 2	**5**	6 8
6 3	**6**	1 2 3 5

3. Grouped data: estimated mean, modal class and median class

In a table with classes such as $10 < t \leq 20$, the exact values are lost, so the mean is an **estimate**. Use each class **midpoint** (15 here): estimated mean = sum of (midpoint × frequency) ÷ total frequency. The **modal class** has the highest frequency. To find the class containing the **median**, work out its position (the 20th and 21st of 40 values) and add up the frequencies until you reach it.

Example: 25 students recorded their homework time. The classes were $0 < t \leq 10$, $10 < t \leq 20$ and so on up to $40 < t \leq 50$ minutes, with frequencies 3, 4, 10, 7 and 1. The midpoints are 5, 15, 25, 35 and 45. Sum of midpoint × frequency = 15 + 60 + 250 + 245 + 45 = 615. Estimated mean = 615 ÷ 25 = **24.6 minutes**. The modal class is **$20 < t \leq 30$**. The median is the 13th time. 3 + 4 = 7 and 7 + 10 = 17, so it is also in the class **$20 < t \leq 30$**.

4. Comparing distributions and interpreting data

Compare two groups with an **average** and the **range**, in context. For example, B has the higher median, so it did better on average. A has the smaller range, so its marks were more consistent. Look for **trends** and **relationships**, and watch for **misleading** information: axes not starting at 0, uneven scales, tiny samples, and claims that correlation proves cause.

Example: In the back-to-back diagram in section 2, Team X has 7 scores with median 55 and range 66 – 44 = 22. Team Y has median 61 and range 65 – 49 = 16. So Team Y scored more on average, and its scores were more consistent.

Think like a mathematician: test a few cases (**specialising**), look for the rule that always works (**generalising**), and explain why you are sure (**convincing**). When you check someone's working, say what is wrong and how to **improve** it.

Common mistakes: using class upper bounds instead of midpoints. Dividing by the number of classes instead of the total frequency. Choosing the middle class as the median class. Counting the overlap of a Venn diagram twice. Comparing averages without the spread, or without saying what it means.