Your notes

GCSE Statistics: revision notes

Skewness and outliers

Contents

1. Shapes of distributions
2. Measuring skewness
3. Outliers
4. Key terms
5. Quick questions
6. What to do next

Shapes of distributions

- A **symmetrical distribution** has mean = median = mode, and looks the same on both sides.
- **Positive skew**: the tail is on the right (high values). Mean > median > mode. On a box plot, UQ – median is greater than median – LQ.
- **Negative skew**: the tail is on the left (low values). Mean < median < mode. On a box plot, median – LQ is greater than UQ – median.
- Incomes are usually positively skewed: most earn moderate amounts, and a few earn a lot.

Measuring skewness

- **Pearson's coefficient of skewness** = 3 × (mean – median) ÷ standard deviation.
- A positive answer means **positive skew**; a negative answer means **negative skew**; close to 0 means roughly symmetrical.
- Example: mean 30, median 27 and standard deviation 6 give 3 × 3 ÷ 6 = 1.5, so the data is positively skewed.
- The quartile measure uses (UQ – 2 × median + LQ) ÷ (UQ – LQ).

Outliers

- An **outlier** is a value that is very different from the rest.
- A value is an **outlier** if it is below LQ – 1.5 × IQR or above UQ + 1.5 × IQR. With LQ 20 and UQ 30, the IQR is 10, so **outliers** are below 5 or above 45.
- Another rule treats values more than 3 standard deviations from the mean as **outliers**.
- If an **outlier** is a mistake, remove it when cleaning the data. If it is genuine, keep it and explain it.

Key terms

Symmetrical distribution	A distribution where mean, median and mode are equal.
Positive skew	A distribution with a tail of high values.
Negative skew	A distribution with a tail of low values.
Pearson's coefficient of skewness	3 × (mean − median) ÷ standard deviation.
Outlier	A value very different from the rest of the data.

Quick questions

Cover the answers and test yourself: remembering is what makes it stick.

1. In a positively skewed distribution, how do the mean and median compare?
2. Where is the tail in negatively skewed data?
3. In a symmetrical distribution, which is true?
4. Mean 30, median 27 and standard deviation 6. What is Pearson's coefficient of skewness?
5. A coefficient of skewness is −0.8. What does this show?
6. LQ is 20 and UQ is 30. Above what value is a data point an outlier?
7. LQ is 20 and UQ is 30. Below what value is a data point an outlier?
8. What should you do with an outlier that is a recording mistake?

Answers

1. The mean is greater than the median
2. On the left, at low values
3. Mean = median = mode
4. 1.5
5. Negative skew
6. 45
7. 5
8. Remove it when cleaning the data

What to do next

Practise 10 questions on this topic

Try 3 exam questions, with mark schemes

Read the lesson, one step at a time

Free revision for GCSE and KS3 at lycolearn.co.uk. Some notes are for one exam board only, and say so in brackets.

Your notes