

Comparing Data

Notes

Statistics tries to describe data by looking at the central value and at how the data is spread out

Central value (*measure of central tendency*) — median (or mean)

How spread out (*measure of dispersion*) — interquartile range (or range)

When comparing data we need to make two comments

1. State which group has the highest *central tendency* — i.e. the highest *median* (or mean)
2. State which group has the least *spread* — so use the *interquartile range* (or range if the interquartile ranges are equal)

Higher median ⇒ *higher average* — this is usually a good thing (e.g. test scores) but could also be bad (e.g. golf scores — the least number number of shots is best)

Lower IQR (or range) ⇒ *lower spread* — this is usually a good thing because it means greater consistency

N.B. You must use numbers to support your statements

E.g. 1 Kate and Meghan went for 10 bike rides for 40 km and recorded their time taken, in minutes.

Here is a summary of the data. Compare their times.

Class	Lowest time	Lower quartile	Median	Upper quartile	Highest time
Kate	70	73	80	82	87
Meghan	38	75	77	81	87

Working:

1st comment = central tendency

Meghan is fastest on average because her median time is lower than Kate's median (77 vs 80)

2nd comment = spread

However, Kate is a more consistent cyclist results because her interquartile range is lower than Meghan's (6 vs. 9)

Video: [Comparing data](#)

[Solutions to Starter and E.g.s](#)

Exercise

9-1 class textbook: p492 E14.6 Qu 1-5
A*-G class textbook: p448 E14.3 Qu 2-5
9-1 homework book: p170 E14.6 Qu 1-4
A*-G homework book: p127 E14.3 Qu 1-3

Summary

Central value (*measure of central tendency*) — median (or mean)

How spread out (*measure of dispersion*) — interquartile range (or range)

When comparing data we need to make two comments, with numbers to support the statements:

1. State which group has the highest *central tendency* — i.e. the highest *median* (or mean)

2. State which group has the least **spread** — so use the **interquartile range** (or range if the interquartile ranges are equal)

Higher median \Rightarrow **higher average** — this is usually a good thing (e.g. test scores) but could also be bad (e.g. golf scores — the least number number of shots is best)

Lower IQR (or range) \Rightarrow **lower spread** — this is usually a good thing because it means greater consistency

Homework book answers (only available during a lockdown)

