Evaluating a statistical investigation

Grade 9 | Worked solutions

GCSE Statistics topic practice

1 Dev wants to test the hypothesis "Students who walk to school are late less often than students who travel by bus." He plans to count the number of times each student was late last year using the school's registers, and to use a questionnaire to find out how each student travels to school.

a. Give one advantage of using the school's registers instead of collecting the data on lateness himself. (1)

b. Give one reason why the registers may not be reliable for Dev's investigation. (1)

c. Dev tries out his questionnaire on 10 students before he uses it. Explain why this is a good idea. (1)

(3 marks)

a) One of these

The data have already been collected, so it is quick

It covers every student for the whole year

b) One of these

He does not know how accurately it was recorded

Teachers may record lateness in different ways

c) This is a pilot study

It shows if the questions are clear and easy to answer

He can fix any problems before the full survey

a) quick, lots of data b) accuracy unknown c) finds problems

2 Priya goes to a school with 1200 students. Her hypothesis is "Year 11 students get less sleep than Year 7 students." One Monday morning she asks 10 of her friends in Year 11 and the 10 students in her brother's Year 7 tutor group who arrive first, "How many hours did you sleep last night?" Her results are summarised in the table.

	Mean (hours)	Standard deviation (hours)
Year 7	8.6	0.7
Year 11	6.9	1.8

a. Give two criticisms of the way Priya collected her data. (2)

b. Priya says, "Every Year 11 student sleeps less than every Year 7 student." Use the table to explain why this may not be true. (1)

c. Priya concludes that her hypothesis is true for all secondary school students in the UK. Give one reason why this conclusion is not valid. (1)

(4 marks)

a) Any two of these

Her friends and the first 10 to arrive are not random samples,

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

so the samples may be biased

Each sample is small (only 10 students)

She asked about one night only,

and a Sunday night may not be typical

b) Look at the standard deviations

The Year 11 standard deviation is large (1.8 hours)

so the Year 11 sleep times are spread out

Some Year 11 students may sleep more

than some Year 7 students

c)

All of her data come from one school

It does not represent every school in the UK

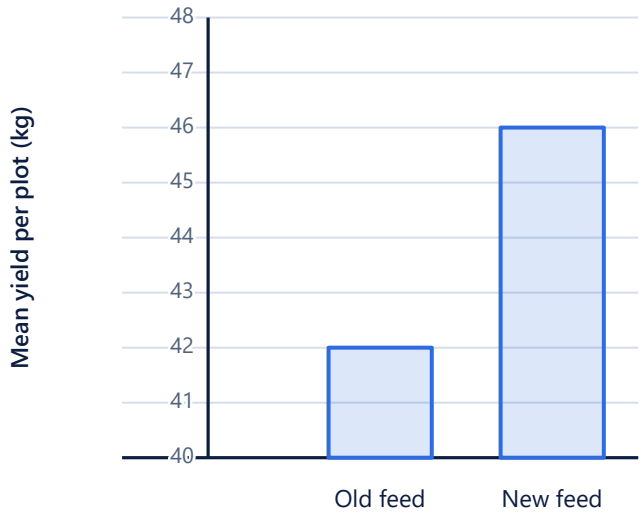
a) not random, small sample b) large sd c) one school

3 A company sells plant feed. Its advert says, "Our new feed gives a far bigger crop of tomatoes." The advert shows this bar chart of the mean yield per plot for some plots given the old feed and some plots given the new feed.

a. Explain why the bar chart is misleading. (1)

b. Work out the actual percentage increase in the mean yield. Give your answer correct to 1 decimal place. (2)

c. Write down one other piece of information you would need before deciding whether the new feed is better than the old feed. (1)



(4 marks)

a)

The yield axis starts at 40, not at 0

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

The new feed bar looks 3 times as tall,

but the yield is only a little higher

b) Read the yields from the chart

Old feed 42 kg, new feed 46 kg

$$46 - 42 = 4$$

$$\frac{4}{42} \times 100 = 9.523...$$

percentage increase = 9.5%

c) One of these

How many plots were used for each feed

Whether the plots had the same soil, water and light

How spread out the yields were

a) axis starts at 40 b) 9.5% c) number of plots

4 Nadia investigates whether a revision app improves test marks. The students in her year sat two tests. Between the tests, each student chose whether or not to use the app. The table summarises the improvement in marks from the first test to the second test.

	Number of students	Mean improvement	Median improvement
Used the app	15	6.2	3
Did not use the app	60	4.1	4

Nadia concludes, "The mean is higher for the app group, so the app improves marks."

a. For the students who used the app, the mean is much higher than the median. What does this suggest about their improvements? (1)

b. State which average is better for comparing the two groups, and what it suggests about Nadia's conclusion. (2)

c. Explain how letting students choose whether to use the app could bias the results. (1)

d. Suggest one way Nadia could improve her investigation. (1)

(5 marks)

a) The mean is pulled up by extreme values

A few students had very large improvements

(the data have a positive skew)

b) The median is better

It is not affected by the extreme values

The median for the app group is lower (3 against 4)

so the data do not support her conclusion

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.

c)

Students who chose the app may be more motivated

They might have improved more without the app

d) One of these

Put the students into the two groups at random

Use two groups of the same size

a) a few large values b) median, not supported

5 A journalist collects the number of libraries and the number of crimes recorded last year in each of 12 towns. Spearman's rank correlation coefficient for her data is 0.82. She writes the headline "Libraries cause crime". The table shows her data for two of the towns.

Town	Population	Number of libraries	Number of crimes
Ashby	250 000	12	20 000
Morton	40 000	2	3600

a. Interpret the value 0.82 in context. (1)

b. Explain why the journalist's headline is not justified. Suggest what is likely to be the real reason for her result. (2)

c. By working out the number of crimes per 1000 people, show that the town with more crimes does not have the higher crime rate. (2)

(5 marks)

a) 0.82 is close to 1

There is strong positive correlation

Towns with more libraries tend to have more crimes

b) Correlation does not mean one thing causes the other

Bigger towns have more people,

so they have more libraries and also more crimes

c) crimes per 1000 people = $\frac{\text{crimes}}{\text{population}} \times 1000$

Ashby

$$\frac{20000}{250000} \times 1000 = 80$$

Morton

$$\frac{3600}{40000} \times 1000 = 90$$

Ashby has more crimes (20000 against 3600)

but Morton has the higher crime rate (90 against 80)

a) strong positive correlation b) population c) 80 and 90

Free for classroom and home use: print, photocopy and share. Not for resale or re-hosting.