GCSE Statistics: Grade 9

Worked solutions

GCSE Statistics practice

1 **CHANGING DATA: EFFECT ON THE MEAN AND STANDARD DEVIATION**

The 20 workers at a factory have a mean weekly wage of £480 with a standard deviation of £36.

a. Every worker is given a 5% pay rise and then a further £10 a week. Work out the new mean and the new standard deviation of the weekly wages.

b. Instead of the pay rise, a new worker joins the original 20 workers on a weekly wage of exactly £480. Calculate the standard deviation of the weekly wages of the 21 workers. Give your answer correct to the nearest penny.

(4 marks)

a) A 5% rise multiplies every wage by 1.05

$480 \times 1.05 = 504$

$36 \times 1.05 = 37.8$

Adding £10 moves every wage up by the same amount

The mean goes up by 10, the spread does not change

$\text{new mean} = 504 + 10$

$\text{new mean} = £514$

$\text{new sd} = £37.80$

b) The new wage is equal to the mean

so the mean stays at £480

$$sd^2 = \frac{\text{sum of } (x - \text{mean})^2}{n}$$

$\text{sum of } (x - \text{mean})^2 = 20 \times 36^2$

$\text{sum of } (x - \text{mean})^2 = 25920$

The new worker adds $(480 - 480)^2 = 0$ to this sum

$$\text{new sd} = \sqrt{\frac{25920}{21}}$$

$\text{new sd} = \sqrt{1234.28...}$

$\text{new sd} = £35.13$

a) mean £514, sd £37.80 b) £35.13

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

2 BINOMIAL PROBABILITY PROBLEMS

A factory makes light bulbs and sells them in packs of 4. Each bulb works with probability p, independently of the other bulbs. The probability that all 4 bulbs in a pack work is 0.6561

You may use $(p + q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$

a. Show that $p = 0.9$

b. Work out the probability that a pack has at most one faulty bulb.

c. A shop buys 500 packs. Work out an estimate for the number of these packs with two or more faulty bulbs.

(4 marks)

a) All 4 work is the term p^4

$p^4 = 0.6561$

$p = \sqrt{(\sqrt{0.6561})}$

$p = \sqrt{0.81}$

$p = 0.9$

b) $q = 1 - 0.9 = 0.1$ (faulty)

At most one faulty means none faulty or exactly one faulty

$p^4 = 0.6561$

$4p^3q = 4 \times 0.9^3 \times 0.1$

$4p^3q = 4 \times 0.729 \times 0.1$

$4p^3q = 0.2916$

$0.6561 + 0.2916 = 0.9477$

c) Two or more faulty is everything else

$1 - 0.9477 = 0.0523$

$0.0523 \times 500 = 26.15$

About 26 packs

a) $0.9^4 = 0.6561$ b) 0.9477 c) 26 packs

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

3 NORMAL DISTRIBUTION AND CONTROL LIMITS

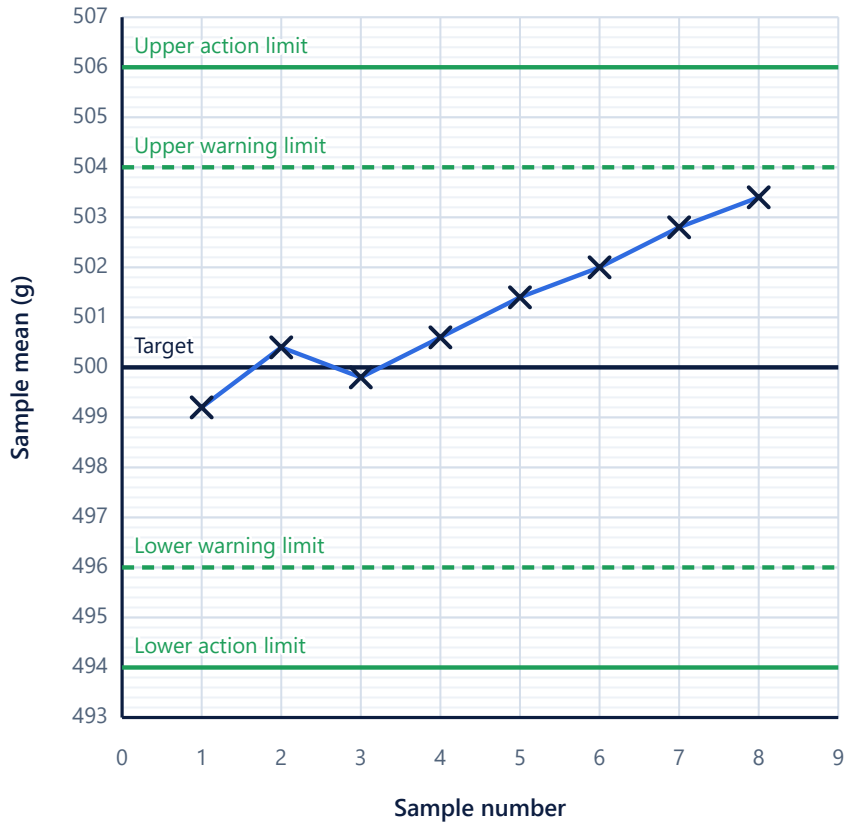
A machine fills bags with sugar. When the machine is working correctly, the mean masses of samples of bags are normally distributed with mean 500 g and standard deviation 2 g.

The warning limits on a control chart are set 2 standard deviations from the target, and the action limits are set 3 standard deviations from the target. The control chart shows the target line and the means of 8 samples.

a. Work out the warning limits and the action limits, and draw them on the control chart.

b. When the machine is working correctly, what percentage of sample means would you expect to be outside the warning limits?

c. Chidi says, "Every sample mean is inside the warning limits, so the machine does not need checking." Comment on Chidi's statement, using the control chart.



(5 marks)

a) Warning limits are 2 sd from the target

$500 + 2 \times 2 = 504$

$500 - 2 \times 2 = 496$

Action limits are 3 sd from the target

$500 + 3 \times 2 = 506$

$500 - 3 \times 2 = 494$

Draw horizontal lines at 494, 496, 504 and 506

b) 95% of values are within 2 sd of the mean

$100\% - 95\% = 5\%$

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

c) Chidi is not right

The sample means have gone up five times in a row

and the last five are all above the target

This trend shows the process is drifting

The machine should be checked now,

before a mean goes outside the limits

a) 496, 504 and 494, 506 b) 5% c) rising trend, check

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

4 **HISTOGRAM PROBLEMS WITH MISSING VALUES**

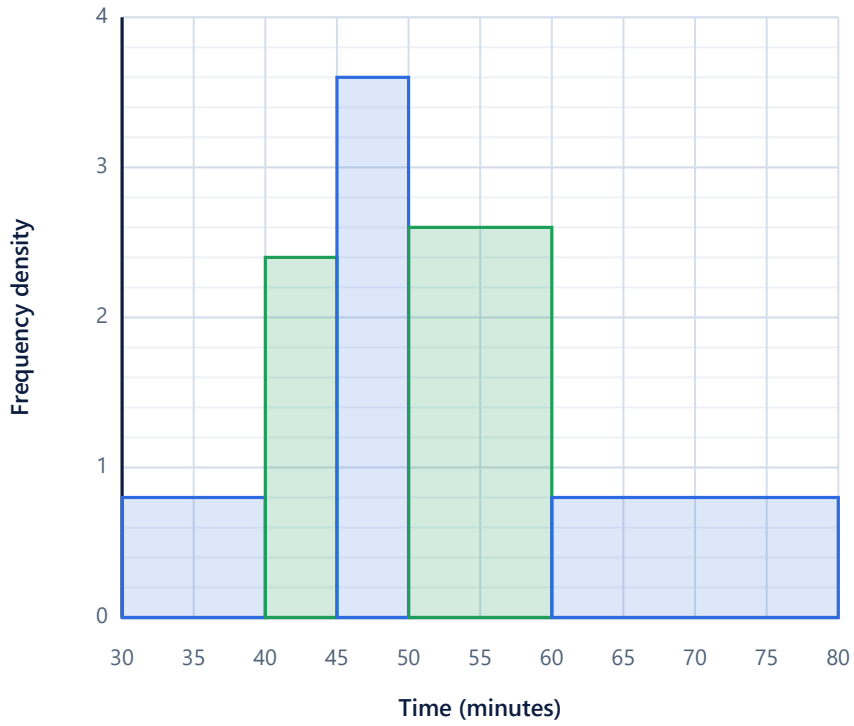
The incomplete table and the incomplete histogram show the times, t minutes, that some runners took to finish a 10 km race.

Time (t minutes)	Frequency
$30 < t \leq 40$	8
$40 < t \leq 45$	12
$45 < t \leq 50$	
$50 < t \leq 60$	26
$60 < t \leq 80$	

a. Use the histogram to complete the table.

b. Use the table to complete the histogram.

c. Work out an estimate for the median time. Give your answer correct to 1 decimal place.



(5 marks)

a) frequency = frequency density x class width

Check with the first bar

$0.8 \times 10 = 8$

45 to 50

$3.6 \times 5 = 18$

60 to 80

$0.8 \times 20 = 16$

b) frequency density = frequency ÷ class width

40 to 45

$12 \div 5 = 2.4$

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

50 to 60

$26 \div 10 = 2.6$

Draw bars of height 2.4 and 2.6

c) Find the total frequency

$8 + 12 + 18 + 26 + 16 = 80$

$80 \div 2 = 40$

The median is the 40th value

Cumulative frequencies

8, 20, 38, 64, 80

The 40th value is in the class $50 < t \leq 60$

It is 40 - 38 = 2 values into a class of 26

$$\text{median} = 50 + \frac{2}{26} \times 10$$

$$\text{median} = 50 + 0.769...$$

$$\text{median} = 50.8 \text{ minutes}$$

a) 18 and 16 b) heights 2.4 and 2.6 c) 50.8 minutes

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

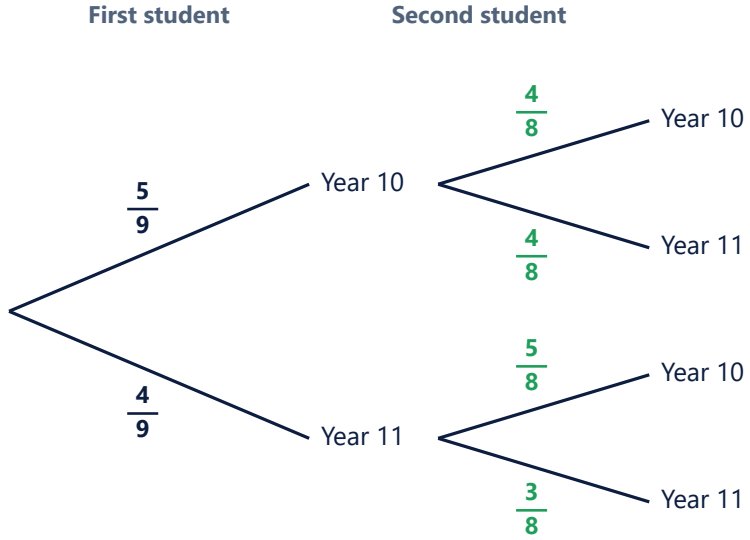
5 CONDITIONAL PROBABILITY WITHOUT REPLACEMENT

A focus group is made up of 5 students from Year 10 and 4 students from Year 11. A researcher picks two of these 9 students at random, one after the other, to interview. No student can be picked twice.

a. Complete the probability tree diagram.

b. Work out the probability that the two students picked are from the same year group.

c. Given that the two students picked are from the same year group, work out the probability that they are both from Year 11.



(5 marks)

a) After the first pick there are 8 students left

After Year 10 there are 4 Year 10 and 4 Year 11 Left

$\frac{4}{8}$ and $\frac{4}{8}$

After Year 11 there are 5 Year 10 and 3 Year 11 Left

$\frac{5}{8}$ and $\frac{3}{8}$

b) Multiply along the branches

$$P(\text{both Year 10}) = \frac{5}{9} \times \frac{4}{8}$$

$$P(\text{both Year 10}) = \frac{20}{72}$$

$$P(\text{both Year 11}) = \frac{4}{9} \times \frac{3}{8}$$

$$P(\text{both Year 11}) = \frac{12}{72}$$

Add the two ways

$$\frac{20}{72} + \frac{12}{72} = \frac{32}{72}$$

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

$$\frac{32}{72} = \frac{4}{9}$$

c) Only look at the outcomes with the same year group

$$P(\text{both Year 11 given same}) = \frac{12}{72} \div \frac{32}{72}$$

$$P(\text{both Year 11 given same}) = \frac{12}{32}$$

$$P(\text{both Year 11 given same}) = \frac{3}{8}$$

a) $\frac{4}{8}, \frac{4}{8}, \frac{5}{8}, \frac{3}{8}$ b) $\frac{4}{9}$ c) $\frac{3}{8}$

6 **SAMPLING AND CAPTURE-RECAPTURE ASSUMPTIONS**

A ranger wants to estimate the number of fish in a lake. She catches 60 fish, tags them and puts them back. One week later she catches 80 fish and finds that 12 of them are tagged.

a. Work out an estimate for the number of fish in the lake.

b. Write down one assumption the ranger has made.

c. The ranger later finds out that 15 of the tagged fish died before the second catch, and that no other fish died, were born, or entered or left the lake. Work out a new estimate for the number of fish in the lake at the time of the second catch, and state whether the estimate in part (a) was an overestimate or an underestimate of this number.

(5 marks)

a) M = 60 tagged, n = 80 caught later, m = 12 of these tagged

$$N = \frac{M \times n}{m}$$

$$N = \frac{60 \times 80}{12}$$

$$N = \frac{4800}{12}$$

$$N = 400$$

b) One of these

The tagged fish mixed fully with the other fish

No tags fell off

The population did not change between the catches

c) Tagged fish alive at the second catch

$$60 - 15 = 45$$

$$N = \frac{45 \times 80}{12}$$

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

$$N = \frac{3600}{12}$$

$$N = 300$$

400 is greater than 300

so the estimate in part a) was an overestimate

a) 400 b) tagged fish mixed fully c) 300, overestimate

7 **RELIABILITY OF REGRESSION PREDICTIONS**

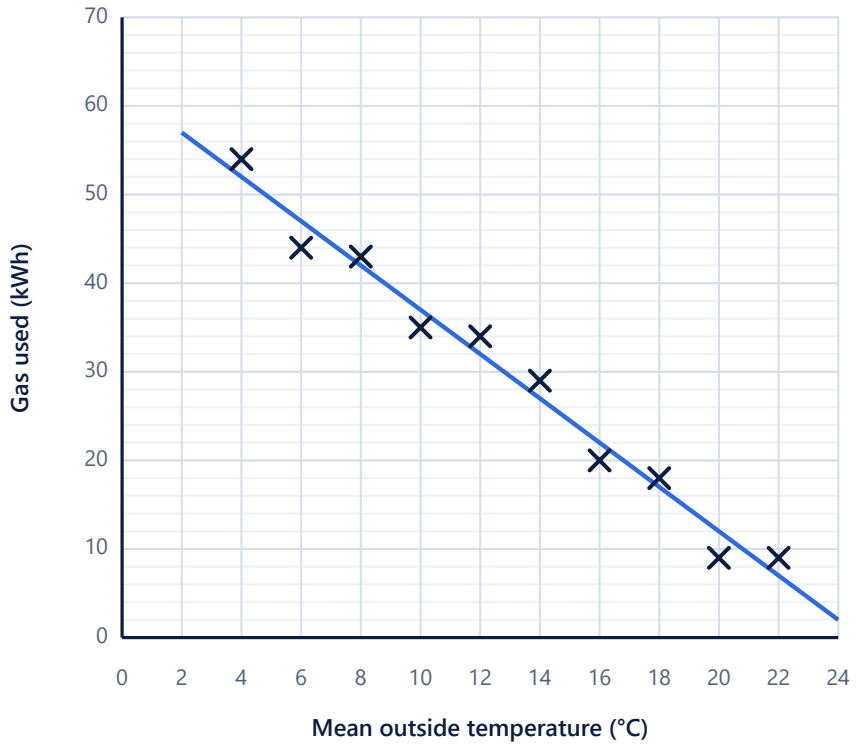
Imran records the mean outside temperature, x °C, and the amount of gas used to heat his house, y kWh, on each of 10 days. The scatter diagram shows his data and the regression line.

The equation of the regression line is $y = 62 - 2.5x$

a. Use the equation to estimate the gas used on a day when the mean outside temperature is 15 °C.

b. Interpret the value −2.5 in the equation of the regression line.

c. Imran uses the equation to predict the gas used on a day when the mean outside temperature is 30 °C. Work out his prediction and give two reasons why it is not reliable.



(4 marks)

a) Put $x = 15$ into the equation

$y = 62 - 2.5 \times 15$

$y = 62 - 37.5$

$y = 24.5$ kWh

b) −2.5 is the gradient

For each 1°C rise in outside temperature

the gas used goes down by 2.5 kWh, on average

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

c) Put $x = 30$ into the equation

$y = 62 - 2.5 \times 30$

$y = 62 - 75$

$y = -13$ kWh

Reason 1

He cannot use a negative amount of gas

Reason 2

30°C is outside the range of his data (4°C to 22°C)

This is extrapolation, the pattern may not continue

a) 24.5 kWh b) 2.5 kWh less per 1°C c) −13 kWh

8 **EVALUATING A STATISTICAL INVESTIGATION**

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.

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.

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.

(4 marks)

a) Any two of these

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

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

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

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