# Answers - Sampling and scatter graphs

Year: Year 11

## Answer to Question 1

a. The first 80 pupils to arrive may all be from the same year or the same bus, so the sample would be biased. Stratified sampling makes sure that every year group is fairly represented.

b. Fraction sampled $= \frac{80}{800} = \frac{1}{10}$. Year 7: **18**, Year 8: **17**, Year 9: **16**, Year 10: **15**, Year 11: **14**.

c. $18 + 17 + 16 + 15 + 14 = 80$. ✓

## Answer to Question 2

a. $\frac{18}{50} \times 1200 =$ **432 pupils**. The fraction of walkers in the sample, $\frac{18}{50} = 0.36$, is used as an estimate for the whole school.

b. It is based on a sample, not on all pupils, so it could be different from the true value.

c. Combined: $18 + 41 = 59$ walkers out of $50 + 100 = 150$. $\frac{59}{150} \times 1200 =$ **472 pupils** (a larger sample gives a better estimate).

## Answer to Question 3

a. People at a gym exercise more than average, so the sample is not typical. Sample people at random from the electoral roll or a mixture of places.

b. The choir members like singing and are not representative of all pupils. Choose pupils at random from the school list.

c. Many people are at work and cannot answer, so the sample excludes workers. Call at different times, including evenings and weekends.

## Answer to Question 4

a. $\frac{8}{40} = \frac{1}{5}$ of the second sample is marked.

b. If $N$ is the population, $\frac{60}{N} = \frac{8}{40}$, so $N = \frac{60 \times 40}{8} =$ **300 frogs**.

c. The marked frogs mix evenly with the others, and no frogs are born, die or move in or out between the two catches (marks do not come off).

## Answer to Question 5

a. **Strong positive correlation**: more revision, higher marks. As $x$ increases, $y$ increases and the points lie close to a straight line.
b. Gradient $= \dfrac{41-15}{8-2} = \dfrac{26}{6} = 4.3$. Using $y - 15 = 4.33(x-2)$: $y = 4.33x + 6.33$. Use two points that lie close to the trend and are far apart. The gradient is the change in $y$ divided by the change in $x$.
c. $4.33 \times 5.5 + 6.33 =$ **30.2** (about 30). Substitute $x = 5.5$ into the equation. The estimate is reliable because 5.5 hours is inside the range of the data.

## Answer to Question 6

a. $\dfrac{80-40}{10-2} =$ **5**. The gradient is the change in $y$ divided by the change in $x$: $\frac{40}{8}$.
b. $y = 5x + c$. At $(2, 40)$: $40 = 10 + c$, so $c = 30$. $y = 5x + 30$. Substitute one point to find $c$. Check with $(10, 80)$: $50 + 30 = 80$.
c. $5 \times 7 + 30 =$ **65**. Substitute $x = 7$ into $y = 5x + 30$.
d. Each extra hour of revision adds about **5 marks**. The gradient is the increase in $y$ for each increase of 1 in $x$.

## Answer to Question 7

a. $\dfrac{80}{N} = \dfrac{12}{60}$, so $N = \dfrac{80 \times 60}{12} =$ **400 fish**.
b. The sample contains only large fish, so the tagged proportion might be wrong, and the estimate is **unreliable** (it may give too few if small fish are never counted).
c. The tagged fish need time to mix with the rest of the population.

## Answer to Question 8

a. Fraction $= \dfrac{40}{240} = \dfrac{1}{6}$. Office: $150 \div 6 =$ **25**. Factory: $60 \div 6 =$ **10**. Drivers: $30 \div 6 =$ **5**.
b. Drivers have different travel needs, so missing them would give biased conclusions.
c. Number the 60 factory workers and use random numbers (or draw names from a hat) to select 10.

## Answer to Question 9

a. $N = \dfrac{45 \times 30}{6} =$ **225 deer**. The fraction marked in the second sample estimates the fraction marked in the whole forest: $\frac{45}{N} = \frac{6}{30}$.
b. Deer may move in or out of the forest, or marked deer may be easier or harder to see.
c. If 11 of 30 were marked, $N = \dfrac{45 \times 30}{11} = 123$, so the estimate would be **much lower**.

## Answer to Question 10

a. A **strong positive correlation**; arm span is almost equal to height.
b. Between 160 (158) and 170 (169): about **164 cm** (arm span is close to the height).
c. 120 cm is outside the data range, so this is extrapolation and may not follow the same pattern.
d. No. Both are related to body size; correlation does not prove one causes the other.

## Answer to Question 11

a. $\dfrac{260 - 80}{30 - 15} =$ **12**. Each extra degree adds about 12 sales.
b. $y = 12x + c$; $80 = 180 + c$, so $c = -100$: $y = 12x - 100$. Substitute the point $(15, 80)$ to find $c$. Check with $(30, 260)$: $360 - 100 = 260$.
c. $12 \times 27 - 100 =$ **224 sales**. 27°C is inside the range of the data, so this is interpolation.
d. 40°C is outside the range of the data: extrapolation. The trend may not continue outside the data.

## Answer to Question 12

a. $600 \div 15 =$ **40 pupils**. Divide the size of the list by the gap between chosen names.
b. **7th, 22nd, 37th, 52nd**. Start at 7 and add 15 each time.
c. Advantage: quick and fair to carry out. Disadvantage: if the list has a pattern (for example, every 15th is a monitor) the sample may be biased.

## Answer to Question 13

a. **Negative correlation**.
b. Gradient $-1.2$: the car loses £1 200 a year. Intercept 15: the value when new is £15 000.
c. $15 - 1.2 \times 5 =$ **£9 000**.
d. $15 - 1.2x = 0$ gives $x = 12.5$ years. In practice a car is worth something even when old, so the line is not suitable for such a long time.

## Answer to Question 14

a. **505** (more than 480) and **0** (there is no pupil 0) cannot be used.
b. **47** appears twice; ignore the repeat because a pupil cannot be chosen twice.
c. **47, 312, 479, 188, 260**.
d. Every pupil has the same chance of being chosen, because the numbers are generated by chance.

## Answer to Question 15

a. **Sample B**: it is larger, so it varies less from the true population value.
b. Total hours $= 20 \times 3.5 + 100 \times 4.1 = 70 + 410 = 480$. Mean $= \dfrac{480}{120} =$ **4.0 hours**.
Find the total hours for each sample by multiplying the mean by the sample size, then divide by the combined size.
c. $900 \times 4.0 =$ **3 600 hours**. Multiply the number of pupils by the mean time.

## Answer to Question 16

a. $N = \dfrac{50 \times 80}{10} =$ **400**.
b. If tags fall off, fewer tagged fish are found in the second sample, so the fraction tagged looks smaller and $N = \dfrac{50 \times 80}{\text{fewer}}$ is **too high**.
c. The population has changed, so the proportion in the first sample no longer represents the whole population.

## Answer to Question 17

a. $(6, 40)$ is far from the pattern of the others.
b. Gradient $= \dfrac{28 - 12}{7 - 1} = \dfrac{16}{6} = 2.67$. $y - 12 = 2.67(x - 1)$, so $y = 2.67x + 9.33$.
c. $2.67 \times 4.5 + 9.33 =$ **21.3**. Substitute $x = 4.5$ into the equation.

d. A mistake in measuring or recording, or an unusual case that really does not follow the trend.

## Answer to Question 18

a. Fraction $= \frac{60}{1200} = \frac{1}{20}$. Boys: $720 \div 20 = \mathbf{36}$. Girls: $480 \div 20 = \mathbf{24}$. The sample is $\frac{60}{1200} = \frac{1}{20}$ of the school, so each group is divided by 20. Check: $36 + 24 = 60$.

b. Boys: $\frac{21}{36} \times 720 = 420$. Girls: $\frac{9}{24} \times 480 = 180$. Total **600 pupils**. Use the fraction of boys and the fraction of girls in the sample who prefer football, then scale up each group separately and add.

c. $\frac{9}{24} = \mathbf{37.5\%}$. 9 of the 24 girls in the sample prefer football, so this fraction estimates the proportion of all girls.

## Answer to Question 19

a. Only readers took part, and only those with strong opinions replied (self-selection), so the sample is biased.

b. **No**: the sample is not random and not representative of all adults.

c. Take a random sample of adults, for example from the electoral roll, and ask the same question of each person.

d. The **random sample of 1 000**: it is larger and unbiased.

## Answer to Question 20

a. $N = \frac{50 \times 40}{5} = \mathbf{400}$.

b. $400 + 100 = \mathbf{500}$.

c. The proportion ringed is $\frac{50}{500} = 0.1$. Expected ringed $= 0.1 \times 60 = \mathbf{6}$.

d. The formula would give $N = \frac{50 \times 60}{6} = 500$, which is the new total; it would be wrong to say that it is still the original population of 400.