Test answers - Statistical sampling and data

Year: Year 12 · Duration: 60 minutes · /20

Answer to Question 1 (4 marks)

a. In a stratified sample each stratum is sampled in proportion to its size; the sampling fraction is $\frac{80}{800} = \frac{1}{10}$. Year 12: $\frac{300}{800} \times 80 = 30$; Year 13: 50 (2) Check: $30 + 50 = 80$ ✓.

b. An opportunity sample uses whoever is available, which may not reflect the population. It is representative of the population and removes the selection bias (2) Each year group is represented in the correct proportion, so the results are more reliable.

Answer to Question 2 (4 marks)

a. Multiply each number of goals by its frequency, add, and divide by 20. $\sum fx = 0 + 6 + 14 + 9 + 4 = 33$, mean $= 1.65$ (2) The mean is 1.65 goals per match.

b. Use $\sigma^2 = \dfrac{\sum fx^2}{\sum f} - \bar{x}^2$. $\sum fx^2 = 0 + 6 + 28 + 27 + 16 = 77$, variance $= 3.85 - 2.7225 = 1.1275$, sd $= 1.06$ (2) The standard deviation is about 1.06 goals.

Check: the frequencies add up to $3 + 6 + 7 + 3 + 1 = 20$ ✓. Rounding the mean before squaring changes the variance, so keep 1.65 exactly. With the $n - 1$ divisor, $s \approx 1.09$, which is also accepted.

Answer to Question 3 (4 marks)

An outlier lies more than 1.5 IQR above Q_3 or below Q_1. Find the IQR, then both limits, and compare the maximum and minimum.

IQR $= 8$ (1). Upper limit $= 22 + 12 = 34$ (1). Lower limit $= 14 - 12 = 2$ (1). 41 is an outlier; 3 is not (1).

So the maximum 41 is an outlier because it is above 34, but the minimum 3 is not, because it is above the lower limit 2.

Answer to Question 4 (4 marks)

In a histogram the area of each bar represents frequency: frequency = frequency density × class width.

$3.2 \times 5 = 16$ (1). $1.5 \times 10 = 15$ (1). Total $= 31$ (2).

So the class 10-15 has 16 values, the class 15-25 has 15 values, and together they contain 31 values.

The most common error is to use the heights 3.2 and 1.5 as frequencies. Check: the frequency densities are $16 \div 5 = 3.2$ and $15 \div 10 = 1.5$ ✓.

Answer to Question 5 (4 marks)

a. The gradient -85 is the change in y for each increase of 1 in x. The price falls by about £85 for each extra year (1) So the laptops lose about £85 of value each year.
b. Substitute $x = 3$ into the line. $620 - 255 = £365$ (1) A 3-year-old laptop is worth about £365.
c. Substitute $x = 9$ and compare with the data range. $620 - 765 < 0$ (negative price) and 9 years is outside the data range (2) Using the line outside the range of the data (extrapolation) is unreliable.

In (c), one mark is given for each reason.