

End of Topic Test – Chapter 1 (Test 1: Topics 1–3) — Mark Scheme

Total: 47 marks

1. Types of data

(a) B1 Membership type (Basic/Plus/Premium) identified as categorical. B1 for a valid second categorical variable with justification – e.g. Age, *if* recorded in age bands/groups (e.g. 18–25, 26–35), can be treated as an ordinal categorical variable. (Note: strictly only one of the four listed variables – Membership type – is inherently categorical; award the second mark for any answer that correctly explains why no other listed variable is naturally categorical, or for a well-justified grouped-age argument.) [2]

(b) B1 Discrete: weekly number of gym visits (counted in whole numbers). B1 Continuous: resting heart rate *or* age (measured on a continuous scale). [2]

(c) B1 Age is discrete if recorded in completed whole years (e.g. “23 years old”). B1 Age is continuous if measured precisely (e.g. exact age including months/days, or as time elapsed since birth), since it can take any value on a scale. [2]

2. Averages and spread

Data: 12, 18, 15, 21, 14, 19, 16 ($n = 7$)

(a) M1 $\sum x = 115$. A1 mean = $115 \div 7 = 16.4$ (3 s.f.). [2]

(b) B1 data ordered: 12, 14, 15, **16**, 18, 19, 21. B1 median = 16. [2]

(c) M1 range = $21 - 12$. A1 range = 9. [2]

(d) B1 the median is not affected by outliers/extreme values, unlike the mean. B1 for skewed data the mean is pulled towards the tail, so the median better represents a “typical” value. [2]

3. Interpreting a bar chart

Frequencies: Car 70, Bus 35, Train 25, Cycle 10, Walk 10 (total 150).

(a) B1 Car. B1 correctly supported by reading off the tallest bar (70 people). [2]

(b) M1 $25/150$. A1 = 16.7% (3 s.f.), accept $16\frac{2}{3}\%$. [2]

(c) B2 (or B1+B1) any valid limitation, e.g. a bar chart does not show proportions of the whole as clearly as a pie chart; it can be visually misleading if the vertical axis does not start at 0 or has a broken scale; exact frequencies are not always easy to read; it gives no indication of change over time. [2]

4. Stem-and-leaf diagram

Data (as read from the diagram): 12, 14, 15, 18, 20, 21, 23, 27, 31, 34, 35, 40, 42 ($n = 13$; note the question states 15 students but only 13 leaves are shown – mark using the 13 values given, allowing equivalent method if a candidate’s own consistent count is used).

(a) M1 data already ordered by stem-and-leaf; median is the middle (7th of 13) value. A1 median = 23 hours. [2]

(b) B1 stem ‘1’ (10–19 hrs) and stem ‘2’ (20–29 hrs) both occur 4 times each – data is bimodal. B1 modal classes = 10–19 and 20–29 hours. [2]

(c) B2 (or B1+B1) valid advantage, e.g. a stem-and-leaf diagram retains every original data value (unlike a bar chart, which groups data into categories), while still showing the shape of the distribution; exact values can be read off directly. [2]

5. Pie chart interpretation

Meetings 120° (33.3%), Emails 90° (25.0%), Breaks 60° (16.7%), Projects 90° (25.0%).

(a) M1 120/360. A1 = $\frac{1}{3}$. [2]

(b) M1 0.25×8 . A1 = 2 hours. [2]

(c) B2 (or B1+B1) valid reason, e.g. a pie chart clearly shows each category as a proportion of the whole day, which is appropriate since the categories are parts of one fixed total (a full working day); easy to compare relative sizes of categories at a glance. [2]

6. Stratified sampling

Sampling fraction = $60/1200 = \frac{1}{20} = 0.05$.

(a) M1 sampling fraction = 0.05 identified/used. M1 each cell value calculated as (cell frequency) $\times 0.05$ (unrounded), e.g. North: Detached 10, Semi 9, Flats 5; East: Detached 4.5, Semi 6, Flats 4.5; South: Detached 3, Semi 5, Flats 4; West: Detached 1.5, Semi 3.5, Flats 4. A1 whole-number samples given for all 12 cells (allow any consistent, sensible rounding). B1 rounding explained clearly, e.g. round to the nearest whole household, and where a subtotal would otherwise be wrong (East, West) round one value up and one down (or similar) so that each neighbourhood subtotal is preserved (East = 15: e.g. 5 + 6 + 4; West = 9: e.g. 2 + 3 + 4 or 1 + 4 + 4). A1 all sample sizes correct/consistent and sum to 60 overall. [5]

(b) B1 obtain/compile a list (sampling frame) of all East households of that type. B1 number each household on the list and use a random number generator/table (or calculator) to select the required quantity. B1 select without replacement, repeating for each household type separately within East. [3]

(c) B1 ensures proportional representation of every neighbourhood and household type (avoids a simple random sample missing a whole group by chance). B1 reduces sampling variability/increases precision of estimates compared with an SRS of the same size. B1 (second distinct advantage) allows valid comparisons to be made between neighbourhoods/household types. [3]

(d) B1 drawback, e.g. requires an accurate, up-to-date list of households by neighbourhood and type, which may be costly or difficult to obtain/classify. B1 development, e.g. some cells are very small (e.g. West Detached $\approx 1-2$ households), giving unreliable estimates for that group. B1 suggested improvement, e.g. set a minimum sample size per stratum, or combine very small strata. B1 development of the improvement, e.g. this still allows meaningful analysis of small groups without needing an excessively large overall sample. [4]